HAL
archives-ouvertes.fr

La classification à facettes pour la gestion des connaissances métier : méthodologie d'élaboration de FolkClassifications à facettes.

Orélie Desfriches Doria

To cite this version:

Orélie Desfriches Doria. La classification à facettes pour la gestion des connaissances métier : méthodologie d'élaboration de FolkClassifications à facettes.. Sciences de l'information et de la communication. Conservatoire national des arts et metiers - CNAM, 2013. Français. <NNT : 2013CNAM0903>. <tel-01015606>

HAL Id: tel-01015606

https://tel.archives-ouvertes.fr/tel-01015606

Submitted on 26 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

le cnam

ÉCOLE DOCTORALE Abbé Grégoire

Laboratoire DICEN

THÈSE présentée par :

Orélie DESFRICHES DORIA

soutenue le : **26 Novembres 2013**

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Discipline/ Spécialité : Sciences de l'Information et de la Communication

La Classification à facettes pour la Gestion des Connaissances Métier

Méthodologie d'élaboration de FolkClassifications à facettes

THÈSE dirigée par :

Monsieur ZACKLAD Manuel	Professeur, CNAM

RAPPORTEURS :

Monsieur CHAUDIRON Stéphane	Professeur, Université Lille III
Monsieur COTTE Dominique	Professeur, Université Lille III

JURY :

Madame SIMONNOT Brigitte	Professeur, Université de Lorraine
Madame BROUDOUX Evelyne	Maître de Conférences, CNAM

A Gaby, à ma mère,

Remerciements

Je tiens en tout premier lieu à remercier Manuel Zacklad, mon directeur de recherche, qui m'a permis de me lancer dans l'aventure de la thèse, pour sa confiance, la qualité de ses conseils et les échanges scientifiques fructueux dont il a m'a fait bénéficier.

Je tiens à remercier les membres du Jury, Brigitte Simonnot, Stéphane Chaudiron, Dominique Cotte, Evelyne Broudoux, pour leur intérêt et le temps passé à lire mon travail, la richesse des échanges lors de la soutenance, ainsi que leurs corrections attentives pour la version finale de ce document.

Je remercie également les participants au projet Miipa-Doc, qui sont intervenus dans différents moments du déroulement de ce projet : Gilles Bertin du CNAM, Camille Vaucelle, L'Hédi Zaher, Jean-Pierre Cahier et Aurélien Bénel de l'Université de Technologie de Troyes, cela a été un réel plaisir de travailler avec eux. Je les remercie pour la qualité des échanges scientifiques que nous avons pu avoir et pour leur bienveillance. Je tiens à remercier également les partenaires du projet à EDF : Philippe Haïk, et les deux participants à mon expérimentation, Benoît Ricard et Sylvain Mahé.

Je veux aussi remercier ici Yves Marleau, Inge Alberts, Jennifer Schellinck de la société Cogniva, qui m'ont accueillie dans leur locaux à Gatineau, au Québec, et qui m'ont formée à leur méthodologie.

Mes vifs remerciements sont aussi adressés aux membres du Laboratoire Dicen, notamment Brigitte Guyot, pour ses conseils méthodologiques, Haud Gueguen, Françoise Rouard, Cécile Payeur, Claire Scopsi, Emma Bester, et Noémie Musnik, pour leur soutien, et leurs encouragements si précieux.

Je remercie aussi Michèle Hudon, pour sa gentillesse et son professionnalisme, ainsi que Sylvie Rouxel, Thomas Bottini et Benjamin Diemert, qui m'ont permis d'aller à la soutenance en étant plus sereine.

Un grand merci tout particulièrement à mes relecteurs et amis, Kevin Perros, Anne-Solènne Marroulle, Gérald Kembellec et Audrey Saint-Georges, sans vos critiques et votre travail de correction, ce travail aurait été de moins bonne qualité.

Je veux aussi remercier chaleureusement ma famille et notamment ma mère, Anne Doria, et mon frère, Eric Desfriches Doria, leur soutien indéfectible représente une des choses les plus rares, et des plus précieuses.

Enfin, je remercie aussi mes amis rennais, Angélique Tartarin, Romain Thomas, Mathilde Lucas, Lucille Sandré pour les bouffées d'air frais qu'ils m'ont procurées, pendant ce travail de quatre années, pour leur soutien, et leur accueil. Merci aussi aux amis de Paris, notamment Claire Vauléon, Julia Lamassourre, Valérie Dhiver et Karim Leleu pour leurs encouragements et leur présence.

Résumé

Dans le cadre du projet ANR Miipa-Doc (Méthode et services Intégrés Institutionnels et Participatifs pour la Classification à Facettes des Contenus Documentaires Complexes) en partenariat avec EDF R&D, l'UTT (Université de Technologie de Troyes), l'entreprise Cogniva, et la JEU (Jeune Entreprise Universitaire) Sémiotag, nous avons développé une méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances métier dans les organisations.
Mon rôle dans ce projeta consisté, en dialogue avec les développeurs, à définir les concepts qui structurent les principes des FolkClassifications à Facettes implémentés dans le prototype, et à articuler l'analyse des scénarios d'usage avec leur mise en œuvre dans l'Interface Homme-Machine.

Au commencement de cette recherche, nous nous sommes interrogés sur la pertinence de l'utilisation des classifications à l'heure où les moteurs de recherche sont largement déployés. Le savoir étant contenu en partie dans les documents, les questions d'organisation du savoir sont des questions d'accès au contenu du savoir, des questions de représentation et de description du contenu de ces documents. Ce sont ces fonctions que remplissent les Systèmes d'Organisation des Connaissances (SOC).

Traditionnellement, les classifications ont été des outils d'organisation du savoir. Dans le numérique, les moteurs de recherche sont largement déployés, mais les classifications transférées sous forme numérique restent utiles pour situer un sujet dans une discipline, et pour la découverte d'autres sujets. Elles ont une organisation systématique et prévisible que les moteurs ne peuvent pas proposer. Enfin, elles sont généralement construites par des experts du monde des bibliothèques ou de la documentation qui ont une bonne connaissance de leur fond documentaire.

Une organisation dispose en général d'un Système d'Information (SI) qui peut contenir un ou plusieurs SOC pour des visées différentes qui peuvent être la GED, l'archivage, la gestion personnelle de l'information, le travail collaboratif…Les SOC de nature institués, personnel et collaboratifs contribuent à l'activité de gestion de l'information, qui est un moyen pour la réalisation des missions des métiers de l'entreprise, et la pérennité de l'entreprise dépend de la bonne conduite de ces activités métier. Les approches traditionnelles des SOC sont dans une forme d'implicite de la portée des activités informationnelles sur les activités métier.

D’autre part, l’appropriation des SOC par les utilisateurs est difficile, elle nécessite des efforts de la part des utilisateurs, en termes d’apprentissage, d’acquisition d’une vision globale du SOC et de connaissance de la couverture du champ. Elle nécessite aussi des compromis sur le sens et l’usage des termes du SOC, en particulier l’apprentissage des conventions d’usage du SOC.

Des études démontrent que la cohérence entre plusieurs utilisateurs dans l’usage des SOC est peu satisfaisante. Ainsi, le classement ou l’ajout de métadonnées sont des activités qui impliquent des aspects interprétatifs, qu’il faut instrumenter, et l’activité d’utilisation d’un SOC est constituée d’une myriade de micro-décisions, sur lesquelles l’utilisateur doit arbitrer.

Les questions de recherche que nous avons traitées consistent donc à évaluer en quoi des caractéristiques spécifiques de dispositifs contributifs permettent de lever ces limites des SOC traditionnels, et sous quelles conditions un système basé sur la contribution et l’implication des utilisateurs dans sa conception peut contribuer à un cercle vertueux de la gestion de l’information.

La croissance de la quantité d’information numérique nécessite des outils de gestion qui se limitent souvent au système de gestion de fichiers, concernant les documents produits et utilisés dans le cadre de l’activité de travail (proto-documents). Les problématiques abordées dans ce travail, consistent, concernant l’indexation, à étudier les formes de conciliation de conceptions de Systèmes d’Organisation des Connaissances formels, institués (thésaurus, ontologies formelles), opposées en général aux conceptions folksonomiques et à celles issues du Web socio-sémantique. Il s’agissait aussi d’étudier quelles formes d’émergence de consensus terminologique sont possibles. Concernant la distinction entre travail réel et travail prescrit, nous avons travaillé sur les aspects imposés, négociés, individuels de la charge de travail liée à la gestion de l’information. Enfin, sur la tension entre travail individuel ou collectif, nous avons cherché quelles seraient les définitions des collectifs considérés, et quelles articulations des dimensions institutionnelles, organisationnelles, individuelles des activités de gestion de l’information sont possibles.

Selon nous, les SOC multidimensionnels permettent de représenter des éléments contextuels afin de guider l’interprétation du sens des termes insérés dans le SOC. Mais aucun objet numérique ne peut porter avec lui tous les éléments de son contexte. Il y a donc des choix à effectuer sur la pertinence des informations contextuelles qu’il est nécessaire de conserver, pour faciliter la recherche d’information et le travail d’appropriation ou de réappropriation d’un document.

Les cadres théoriques que nous avons mobilisés concernant l’organisation des connaissances sont composés de trois champs : nous avons exploré l’approche du *Domain Analysis* qui implique que la

description du sujet d'un document devrait prendre en compte l'usage potentiel des documents. Selon cette approche, la pertinence de la description par sujet devrait être dirigée par la connaissance approfondie du domaine, des méthodes de recherche, ainsi que des besoins des utilisateurs cibles. Nous avons rapproché cette approche de Hjørland du projet du Web Socio-Sémantique qui vise à créer des ontologies basées sur les contributions d'utilisateurs distribués qui portent des points de vue. Enfin, nous nous sommes intéressés aux approches de Gestion des connaissances, notamment celles orientées-document.
Concernant l'analyse des activités métier et informationnelles, les théories que nous avons convoquées sont la théorie de l'Activité, qui nous a permis d'obtenir une grille de lecture de l'activité au sens large. Ensuite nous nous sommes intéressés à la Sémiotique des Transactions Coopératives (Zacklad), qui révèle la nature dynamique et transactionnelle des activités liées à l'information. Enfin, les approches de la psychologie du travail et de l'ergonomie françaises nous ont permis d'étayer notre approche orientée-utilisateur.

De ces approches issues de la psychologie du travail, et de l'ergonomie, nous avons pu mobiliser différents concepts, dont les suivants qui nous ont permis de développer notre point de vue : le sens du travail, les référentiels communs, et les activités métafonctionnelles, qui nous ont permis d'affirmer que l'activité de maintenance d'un système documentaire personnel, comme la réorganisation des dossiers d'une arborescence de fichiers, ou la construction, ou la maintenance d'un SOC, constituent une forme d'activité méta-fonctionnelle.

Dans le premier chapitre, nous présentons les problématiques liées à l'organisation des connaissances, à l'indexation par sujet, à la documentarisation, ainsi que les principes qui guident l'élaboration des classifications, des classifications classiques jusqu'aux classifications à facettes. Nous proposons ensuite une revue de la littérature sur les systèmes de tagging collaboratifs existants, notamment ceux qui offrent une structuration du système de tags à l'aide de facettes. Nous présentons enfin le projet ANR Miipa-Doc, dans le cadre duquel cette recherche a été menée.

Le deuxième chapitre est consacré à une synthèse des méthodes existantes d'élaboration de classifications à facettes. Ainsi, nous commençons par présenter celle de Ranganathan, qui a pour objectif le classement des livres dans une bibliothèque. Puis nous exposons la méthode du CRG (Classification Research Group) porté par Brian C. Vickery, pour des fonds bibliographiques spécifiques. Ensuite, nous poursuivons en présentant celle de l'entreprise Cogniva qui s'applique au Record Management dans les organisations. Enfin, nous décrivons la méthode de Denton (2003) basée sur les apports de Spiteri (1998) pour l'organisation des informations dans des sites Web.

Dans notre troisième chapitre, nous commençons par élaborer une typologie des Systèmes d'Information (SI), et nous poursuivons en réalisant une analyse de l'impact de ces types de SI sur l'activité de gestion de l'information. Les notions de contexte, de métier, d'activité, et de sens de l'activité, ainsi que le champ du PIM (Personal Information Management) sont explorées pour proposer une conception contextuelle de la structuration et de l'organisation des connaissances. Nous présentons de manière détaillée le prototype Hypertagging, développé dans le cadre du projet Miipa-Doc, en fin de ce chapitre.

Le quatrième chapitre est consacré à la présentation de notre expérimentation, réalisée avec deux Ingénieurs-Chercheurs de la R&D d'EDF, dont nous avons étudié les postes de travail individuels afin de proposer une classification à facettes pour la gestion de leurs documents de travail. Nous avons procédé à des entretiens individuels et en groupe pour tenter de comprendre leur activité afin d'élaborer une structuration de la classification à facettes basée sur les activités. Les hypothèses de travail sont présentées, et validées ou invalidées.

Dans notre cinquième et dernier chapitre, nous présentons notre méthodologie d'élaboration de la FolkClassifications à facettes pour la gestion des connaissances métier. Elle s'inspire des résultats de notre expérimentation et vise à appuyer une démarche de gestion des connaissances (orientée-document), sur la gestion personnelle de l'information dans son flux quotidien. Nous cherchons donc à proposer un modèle d'Organisation des Connaissances selon une approche contextuelle de la gestion de l'information (SOC multidimensionnels), intégrant une implémentation des principes de la classification à facettes, structuré par l'analyse des activités métier, pour des collectifs restreints (orienté-utilisateur), dans un IHM ergonomique qui vise à équiper les activités de gestion de l'information.

De futures expérimentations restent à mener pour la généralisation des résultats de recherche, tandis que la détermination des dimensions collaboratives des FolkClassifications à facettes, pourrait être outillée. Il apparaît nécessaire d'élaborer des critères d'évaluation de ce type de Système d'Organisation des Connaissances, tout en continuant à faire évoluer l'IHM du prototype.

Mots-clés : Organisation des connaissances, Système d'Organisation des Connaissances, Gestion des connaissances, Classification, Classification à facettes, Tagging, Indexation, Classement, Méthodologie, Système d'Information, Activité, Métier, Contexte, Gestion de l'information, Système Orienté-utilisateur, Collaboration, Contribution.

Abstract

We address Knowledge Organization and Knowldege Management general issues and we present the basic principles for the development of faceted classifications. Then we introduce a synthesis of existing methods to develop this kind of classification. Next we propose an analysis of effects of Information Systems on information management activities. By doing this, we produce an analysis of the concept of activity inspired by Activity Theory, occupational psychology and french ergonomics. A presentation of Hypertagging prototype developed in the frame of Miipa-Doc research project, based on users tagging and structured by faceted classification principles, is provided. Our experiment about generating faceted classification for document management activities is detailed. Finally, we expose our method for the development of faceted classification for Knowledge Management purposes in the context of specific trades. This Knowledge Management approach is incorporated in Knoweldge Organization activities.

Keywords : Knowledge Organization, Knowledge Organization System, Knowldege Management, Classification, Faceted Classification, Tagging, Indexing, Ranking, Methodology, Information System, Activity, Trade, Context, Information Management, User-oriented System, Collaboration, Participation.

Table des matières

Remerciements 4
Résumé 6
Abstract 10
1 Organisation des connaissances et classifications à facettes 30
1.1 Introduction 30
1.2 Organisation des connaissances et Gestion des connaissances 30
1.2.1 Gestion des connaissances 2.0 35
1.2.2 Systèmes d'Organisation des Connaissances (SOC) 37
1.2.3 L'indexation 40
1.2.3.1 Evaluation de l'indexation 43
1.2.4 La notion de sujet 44
1.2.5 La Documentarisation 48
1.3 Définition des classifications au sens général 50
1.3.1 Classifications génériques et spécifiques 51
1.3.2 Les classifications bibliographiques 52
1.3.2.1 Caution bibliographique 55
1.3.3 Présentation générale de la classification à facettes 56
1.3.3.1 Principes originels de la classification à facettes selon Ranganathan 58
1.3.3.2 Présentation historique du « Prolegomena to Library Classification » 58
1.3.3.3 Les classifications selon Ranganathan 60
1.3.3.3.1 Les classifications énumératives 60
1.3.3.3.2 Les classifications quasi-énumératives 60
1.3.3.3.3 Les classifications quasi-à-facettes 60
1.3.3.3.4 Les classifications à facettes rigides 61
1.3.3.3.5 Les classifications à facettes libres 61
1.3.3.4 L'analyse par facettes 62
1.3.3.5 Qu'est-ce qu'une facette ? 64
1.3.3.6 Les avantages et les limites des classifications à facettes 66
1.3.4 Impact du numérique sur les principes des classifications 69
1.4 Le tagging collaboratif 72
1.4.1 Présentation d'un outil de *bookmarking* social 73
1.4.2 Motivations, rôles des utilisateurs et indexation à l'aide de systèmes de tagging 77

1.4.3 Le tagging, quel processus de construction de sens partagé ? 79
1.4.4 Nature et fonctions des tags 84
1.4.5 Les avantages et les limites des systèmes de tagging..... 87
1.4.6 La structuration dans les systèmes de tagging 90
1.4.7 La structuration des systèmes de tagging par facettes..... 92
1.4.7.1 TagMe ! 92
1.4.7.2 Classification collaborative de collection évolutive..... 93
1.4.7.3 FacetMap 93
1.4.7.4 Facetag 95
1.4.7.5 Facette 96
1.4.8 Méthodes de classification à facettes pour les outils de tagging sur le Web 98
1.4.9 Les autres projets proches de Miipa-Doc 100
1.4.9.1 Flamenco 100
1.4.9.2. Projets visant le classement de l'expertise 103
1.5 Le projet Miipa-Doc 104
1.5.1 Rôles des partenaires 105
1.5.2 Les questions de recherche..... 106
1.5.3 Evolution du projet..... 106
1.6 Conclusion du Chapitre 1 108
2 Chapitre 2 : Méthodologies d'élaboration de classifications à a facettes 110
2.1 Introduction 110
2.2 La méthodologie de Ranganathan 110
2.2.1 Préalable sur les différents types de sujets et leur formation 111
2.2.1.1 Les types de sujets..... 111
2.2.1.2 Les méthodes de formation des sujets..... 112
2.2.2 Les cinq lois des sciences des bibliothèques..... 113
2.2.3 Lois sur la structure syntaxique d'une classification 114
2.2.4 Les Canons pour le plan des idées 115
2.2.4.1 Les canons pour le choix des caractéristiques..... 116
2.2.4.2 Les canons pour la succession des caractéristiques 116
2.2.4.3 Les canons pour les ensembles de classes..... 117
2.2.4.4 Les canons pour les chaînes de classes 118
2.2.4.5 Les canons pour les séquences de classes 118

2.2.4.6 Les principes pour réaliser des séquences efficaces 118
2.2.5 Les canons pour le travail sur le plan terminologique 119
2.2.6 Sur le plan de la notation 120
2.2.6.1 Canons de notation pour des ensembles d'entités en expansion 122
2.2.7 Différences entre les classifications bibliographiques hiérarchiques et les classifications à facettes libres 123
2.2.8 Le modèle à cinq facettes de Ranganathan 125
2.2.9 Principes d'indexation des sujets à l'aide du modèle à cinq facettes 127
2.2.9.1 Identification de la composition des sujets en facettes 127
2.2.9.2 Constitution syntaxique des séquences de facettes : principes théoriques . 129
2.2.9.3 Principes supplémentaires pour les séquences de facettes 130
2.2.10 Méthode de transformation d'un sujet en indice de classification à facettes . 131
2.2.11 Critiques de la méthodologie de Ranganathan 132
2.3 La méthodologie du CRG 135
2.3.1 Historique du CRG 135
2.3.1.1 Des classifications spécifiques plutôt que génériques 136
2.3.2 La conception de la classification à facettes selon le CRG 137
2.3.2.1 Le point de vue du CRG sur la Colon Classification 138
2.3.3 Méthodologie d'élaboration de classification à facettes du CRG 138
2.3.3.1 La construction de la liste de la classification 139
2.3.3.1.1 Définir les entités d'intérêt pour les utilisateurs 139
2.3.3.1.2 Collecte dans les sources 140
2.3.3.1.3 Elaboration des facettes 141
2.3.3.1.4 Structuration de chaque facette 142
2.3.3.1.5 Le modèle à treize facettes du CRG 142
2.3.3.1.6 Catégories fondamentales et facettes du modèle 143
2.3.3.1.7 Structuration des facettes entre elles 145
2.3.3.1.8 Ajout d'une notation appropriée 147
2.3.3.1.9 Processus d'indexation à l'aide du modèle à facettes du CRG 149
2.3.3.1.10 Recherche à l'aide de la classification dans un catalogue sur fiches 149
2.3.4 Différences entre classifications à facettes et classifications traditionnelles 151
2.3.5 Comparaison des méthodologies de Ranganathan et du CRG 153
2.4 La méthodologie GenISIS (Cogniva) 155

2.4.1 Une méthodologie au service d'une application 156
2.4.1.1 Les fonctionnalités d'ISIS 157
2.4.1.2 Aux origines de l'application ISIS 158
2.4.1.3 Classification fonctionnelle du Ministère des Finances 159
2.4.1.4 Objectifs d'ISIS 160
2.4.1.5 Objectifs théoriques de la méthodologie GenISIS 160
2.4.1.6 Objets de la méthodologie 161
2.4.2 Notions de base de la méthodologie GenISIS 161
2.4.2.1 La notion d'information 161
2.4.2.2 Le cycle de vie de l'information 162
2.4.2.3 Les métadonnées 162
2.4.2.4 La classification fonctionnelle 163
2.4.2.5 Les Records 164
2.4.2.6 La conservation et l'élimination 164
2.4.2.7 La valeur opérationnelle des documents 165
2.4.2.8 Les modèles 165
2.4.2.9 La décomposition fonctionnelle 165
2.4.2.10 Modèle de Processus d'affaires 167
2.4.2.11 Décomposition fonctionnelle et modèle de Processus d'Affaires 170
2.4.3 Les concepts génériques dans la structure ISIS 171
2.4.3.1 Les Facettes 172
2.4.3.2 Les types de facettes 172
2.4.3.2.1 Distinction entre Facettes de Contenu et Facettes de Contexte 172
2.4.3.2.2 Facettes Entreprise et facettes spécifiques 173
2.4.3.2.3 Les relations 174
2.4.3.2.4 Thésaurus 174
2.4.3.2.5 Conservation, élimination et valeur opérationnelle des documents 174
2.4.4 Règles pour guider la création des facettes 175
2.4.5 Le modèle de facettes GenISIS 176
2.4.5.1 Les facettes de contenu du modèle 176
2.4.5.2 Les facettes de contexte du modèle 177
2.4.5.3 Autres facettes de contexte 178
2.4.5.4 Précisions sur les facettes du modèle ISIS 179

2.4.6 La méthodologie GenISIS de construction du schéma de facettes 180
2.4.6.1 Les Ressources 180
2.4.6.2 Les étapes de la première phase : l'analyse initiale 181
2.4.6.3 Les étapes détaillées de la deuxième phase : l'analyse détaillée..... 182
2.4.6.3.1 Les Interviews de haut niveau 182
2.4.6.3.2 Les interviews de niveau intermédiaire 183
2.4.6.3.3 Les *Workshops*..... 185
2.4.6.3.4 Les questionnaires 186
2.4.6.3.5 Méthode H 186
2.4.7 Observations sur la méthodologie GenISIS 187
2.5 Classifications à facettes pour les sites Web 190
2.5.1 Le modèle simplifié de Spiteri (1998)..... 193
2.5.1.1 Principes pour le choix des facettes 193
2.5.1.2 La succession des facettes et des valeurs de facettes 194
2.5.1.3 Canons pour le plan terminologique 195
2.5.1.4 Canons pour le plan de notation 195
2.5.1.5 Les principes éliminés des deux méthodologies précédentes 196
2.5.1.5.1 Les principes éliminés des canons pour les ensembles de classes 196
2.5.1.5.2 Les principes éliminés des canons pour les chaines de classes 196
2.5.1.5.3 Les principes éliminés des canons pour les séquences de classes..... 197
2.5.1.5.4 Les Canons de notation..... 197
2.5.1.5.5 Les principes éliminés des séquences de facettes..... 197
2.5.1.6 Observations sur le modèle de Spiteri (1998) 198
2.5.2 Méthodologie de classifications à facettes de Denton pour les sites Web 198
2.5.3 Observations sur la méthode de Denton (2003) 201
2.6 Synthèse des caractéristiques des méthodologies présentées..... 202
3 Chapitre 3 : Systèmes d'Information et Activités 205
3.1 Introduction 205
3.2 Les systèmes d'informations 205
3.2.1 Notion de document 208
3.2.2 Quels critères de différenciation des systèmes d'information ?..... 212
3.2.2.1 Le degré de contrôle exercé sur les utilisateurs..... 212
3.2.2.2 Les fonctions du système 213

3.2.2.3 Les bénéfices du système selon les utilisateurs 213
3.2.2.4 Les types de documents que le système héberge 214
3.2.2.5 Les modalités de participation des utilisateurs finaux 214
3.2.2.6 Le type de sémantique induit par le système 215
3.2.2.7 Les SOC disponibles dans ces SI. 216
3.3 Qu'est-ce qu'une activité ? 218
3.3.1 Les composants de l'activité 218
3.3.2 Les artefacts et le contexte 219
3.3.3 Le développement de l'activité 220
3.3.4 La théorie de la cognition distribuée 221
3.3.5 La théorie de l'action située 222
3.3.6 Points de différentiation pour le choix d'une théorie de l'activité 223
3.3.6.1 La structure de l'activité 223
3.3.6.2 Les invariants de l'activité 224
3.3.6.3 Conception symétrique ou asymétrique des artefacts et des humains 225
3.3.6.4 Interaction et communication 226
3.3.6.5 Rôle du contexte 227
3.3.6.6 Sur le plan méthodologique 228
3.4 Théorie de l'Activité : une conception dynamique de l'activité 228
3.4.1 Travail prescrit et travail réel 229
3.4.2 L'activité et le sujet : des éléments hétérogènes 230
3.4.2.1 Les sources du développement conjoint de l'activité et du sujet 231
3.4.3 Sémiotique des transactions coopératives 233
3.5 L'activité de gestion de l'information dans les organisations 237
3.5.1 Pourquoi gérer l'information ? 238
3.5.2 Les pratiques informationnelles 240
3.5.3 Distinction entre les usages et les pratiques 242
3.5.4 Que sont les activités de gestion de l'information ? 243
3.5.5 Systèmes d'Information Personnels et *Personal Information Management* 247
3.5.5.1 Comment l'information est-elle personnelle ? 249
3.5.5.2 PIM et contexte 249
3.5.5.3 Catégorisation et PIM 251
3.5.5.4 Le problème du « basic level » : une question de catégorisation 253

3.5.6 Le contexte : une notion complexe 255
3.5.6.1 Contexte et catégorisation 258
3.5.6.2 Contexte et collectif 260
3.5.7 L'évolution de l'organisation du travail 263
3.5.8 Le concept de métier réhabilité 265
3.5.8.1 Définition du concept de métier 266
3.5.8.2 Le passage de l'individuel au collectif à travers le prisme du métier 268
3.5.8.3 Référentiel commun : une représentation collective du contexte 269
3.5.9 Présentation du prototype Hypertagging 271
3.5.9.1 Présentation des deux modules d'Hypertagging 272
3.5.9.2 Les options de tagging 274
3.5.9.3 Les options de classement 275
3.5.9.4 Les options de paramétrage des facettes 276
3.5.10 Systèmes d'information structurés et Systèmes d'information 2.0 279
3.5.10.1 Coexistence des SIP et des SI structurés 279
3.5.10.2 Les PGI et les Workflows 282
3.5.10.3 Les systèmes d'information semi-structurés et 2.0 285
3.6 Conclusion du chapitre 3 : 286
4 Chapitre 4 : Expérimentation 289
4.1 Introduction 289
4.2 Présentation des participants à l'expérimentation et de leurs activités 289
4.3 Présentation de l'expérimentation 290
4.3.1 Objectifs principaux de la première phase en deux volets 291
4.3.2 Positionnement de l'enquêteur 293
4.3.3 Contraintes 293
4.3.4 Les matériaux des observations et objets des entretiens 294
4.3.5 Qualification théorique des sources 295
4.4 Déroulement du premier volet de la phase I de l'expérimentation 295
4.4.1 Les entretiens et observations 295
4.4.2 Objectifs des observations des libellés des dossiers 297
4.4.2.1 Données collectées durant le 1er volet de la phase I 297
4.4.2.1.1 Observation des documents 297
4.4.2.1.2 Observations des arborescences 298

4.4.2.2 Observations et analyse du contenu des libellés des dossiers 299
4.4.2.3 Observations et analyse de la forme des libellés des dossiers..... 301
4.4.2.4 Elaboration itérative de la classification à facettes 304
4.4.2.4.1 Construction des facettes et tags..... 305
4.4.2.4.2 La structuration en onglets de la classification à facettes..... 306
4.4.3 Résultats 308
4.4.3.1 Evaluation par les participants de la classification à facettes proposée 312
4.4.4 Déroulement du deuxième volet de la phase I de l'expérimentation 315
4.4.5 Evolution des objectifs du deuxième volet de la phase I de l'expérimentation . 316
4.4.5.1 Présentation du mode opératoire des participants 318
4.4.5.2 Le positionnement des participants 318
4.4.5.3 Les sources 318
4.4.5.4 Critères d'évaluation des classifications produites 319
4.4.6 Présentation des données..... 322
4.4.6.1 Observations et description et analyse des classifications à facettes 322
4.4.6.1.1 La classification produite par l'agent A 322
4.4.6.1.2 Observations générales de la classification de l'agent A 328
4.4.6.2 Classification produite par l'agent B..... 329
4.4.6.2.1 Observations générales de la classification de l'agent B..... 337
4.5 Phase II de l'expérimentation..... 339
4.5.1 Objectifs de la phase II..... 340
4.5.2 Observations des éléments communs entre les deux classifications..... 340
4.5.3 Résultats 341
4.6 Conclusion..... 342
5 Chapitre 5 : Méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances..... 345
5.1 Introduction 345
5.2 Objectifs méthodologiques..... 347
5.3 Contraintes 348
5.4 Définitions..... 349
5.4.1 Facette 349
5.4.2 Sémiotag..... 349
5.4.3 Onglet 352

5.5 Procédures de collecte de la méthodologie 353
5.6 Déroulement de la méthodologie 354
5.6.1 Première phase : entretiens – collecte – élaboration du support 354
5.6.2 Deuxième phase : élaboration individuelle de FolkClassification avec Hypertagging 358
5.6.3 Troisième phase : détermination des dimensions partagées 359
5.7 Conclusion 361
6 Bibliographie 369
7 Annexes 395

Liste des tableaux

Tableau 1. Synthèse des différences entre les classifications à facettes et les classifications énumératives..... 69
Tableau 2. Comparaison des modèles de facettes du CRG et de Ranganathan 155
Tableau 3. Synthèse des caractéristiques des méthodologies présentées..... 203
Tableau 4. Typologie des problématiques traitées par les Systèmes d'Information..... 207
Tableau 5. Types d'information collectés..... 300
Tableau 5. Caractéristiques et fonctions des onglets..... 312

Liste des figures

Figure 1. Classe Histoire de la Bibliothèque du Congrès, (extrait de Shirky, 2005) 71
Figure 2. Schéma des divisions dans une classification traditionnelles 152
Figure 3. Schéma des divisions dans une classification à facettes 153
Figure 4. Schéma théorique de modélisation d'un processus 169
Figure 5. Représentation d'un modèle de processus d'une activité de développement de logiciel 170
Figure 6. Illustration du croisement des approches Top-Down et Bottom-Up 171
Figure 7. Organisation des facettes dans le gabarit du modèle ISIS 180
Figure 8. Schéma des composants de la méthode H 187
Figure 9. Facettes disponibles pour le filtrage des vidéos sur Youtube 191
Figure 10. Facettes rigides pour filtrer les résultats de recherches sur Google 191
Figure 11. Représentation de l'activité d'Engeström (1999) 221
Figure 12. Schématisation d'une transaction coopérative de (Zackald, 2006) 234
Figure 13. Schématisation d'un flux transactionnel de (Zacklad, 2006) 235
Figure 14. Représentation du cercle vertueux des métiers d'Yves Clot (2008) 269
Figure 15. Module Saver d'Hypertagging 273
Figure 16. Module Finder d'Hypertagging 274
Figure 17. Panneau de droite du module Saver 274
Figure 18. Paramétrage d'une facette dans Hypertagging 277
Figure 19. Exemple de Facette 278
Figure 20. Extrait des activités listées 302
Figure 21. Extrait de la collecte des types de documents 302
Figure 22. Extrait des thèmes relevés 303
Figure 23. Extrait des autres termes collectés 304
Figure 24. Synthèse des activités des participants 305
Figure 25. Premier onglet de la proposition de classification à facettes 308
Figure 26. Deuxième onglet de la classification à facettes 309
Figure 27. Quatrième onglet de la classification à facettes 309
Figure 28. Fonctionnement des contributions des agents dans les projets 313
Figure 29. Contre-exemple de classification à facettes 319
Figure 30. Exemple correct de classification à facettes 320
Figure 31. Facette Lieu 1 320

Figure 32. Facette Lieu 2 321
Figure 33. Facette Lieu 3 321
Figure 34. Facette Lieu 4 321
Figure35. Onglet « Caractérisation du document » de la classification de l'agent A 323
Figure 36. Onglet « Contenu du document » de la classification de l'agent A. 324
Figure37. Onglet « Réalisation de l'activité » de la classification de l'agent A 325
Figure 38. Onglet « Préparation activité future » de la classification de l'agent A. 326
Figure 39. Onglet « Gestion de la compétence et Veille » de la classification de l'agent A . 327
Figure 40. Partie haute de l'onglet « Rôle (Co) Auteur Destinataire » de la classification de l'agent B 330
Figure 41. Partie basse de l'onglet « Rôle (Co) Auteur Destinataire » de la classification de l'agent B 331
Figure 42. Onglet « Activité et Usage » de la classification de l'agent B 332
Figure 43. Onglet « Etat contenu et forme » de la classification de l'agent B..... 333
Figure 44. Onglet « Thème technique » de la classification de l'agent B..... 334
Figure 45. Onglet « IK Recueil Méthodo outils » de la classification de l'agent B 335
Figure 46. Partie haute de l'onglet « SI Dév Pilotage Appli » de la classification de l'agent B 336
Figure 47. Partie basse de l'onglet « SI Dév Pilotage Appli » de la classification de l'agent B 336
Figure 48. Facette Lieu 1 350
Figure 49. Facette Lieu 2 351
Figure 50. Facette Lieu 3 351
Figure 51. Facette Lieu 4 351
Figure 52. Exemple de modèle pré-paramétré à des fins pédagogiques 358

Liste des annexes

Annexe 1 : Exemples de types d'information mentionnés dans les intitulés des dossiers :... 396
Annexe 2 : Liste des facettes issues de la collecte 397
Annexe 3 : Classification à facettes proposée dans l'expérimentation 402

Introduction

Dans le contexte du numérique, les outils de gestion de l'information se sont multipliés, complexifiés, jusqu'à devenir des outils d'encadrement des processus de travail.

Il existe une profusion d'outil de gestion de l'information qui cohabitent dans nos espaces numériques, (bookmarks, mails, messageries instantanées, systèmes de gestion de fichiers, base de données, Web, bases de GED), basés sur une diversité de Systèmes d'Organisation des Connaissances comme les moteurs de recherche, les classifications, les annuaires, les outils de filtrages, de cartographies, les folksonomies, les thésaurus ou les champs de bases de données. D'après (Zacklad *et al*., 2011), la diversité des écosystèmes documentaires est un frein à la documentarisation cohérente des ressources documentaires.

Cette variété de SOC a un impact également sur la mise en œuvre de ces activités de gestion de l'information, tandis que de nombreuses recherches (Guyot, 2002, 2004, 2006, 2009), (Kolmayer et Peyrelong, 1999), (Peyrelong, 2002), (Autissier et Lahlou, 1999) viennent confirmer le constat que les activités de gestion de l'information numérique sont devenues indispensables à la bonne exécution des missions des acteurs dans une organisation, tout en restant largement implicites.

Dans le cadre du projet Miipa-Doc (Méthodes et Services intégrés Institutionnels et Participatifs pour la Classification à Facettes des Contenus Documentaires Complexes), nous avons étudié plus en profondeur certains des outils de gestion de l'information numérique, comme les systèmes de gestion de fichiers par arborescences, les classifications, les espaces de bookmarks en ligne et les folksonomies, dans le cadre des usages individuels. De plus, ces recherches nous ont menée à nous intéresser à une forme de classification inventée par le bibliothécaire et mathématicien Ranganathan en 1924, en Inde, la classification à facettes. Elle est souvent considérée comme l'apport le plus important des 50 dernières années en sciences de la documentation, bien que les auteurs aient des difficultés à s'accorder sur une définition de la notion de facette. De la bibliothèque, espace auquel elle se destinait en premier lieu, nous l'avons transposée aux espaces de gestion personnelle de l'information numérique. Ainsi, nous avons voulu étudier la manière dont elle s'adapte aux postes de travail numériques et quel impact le numérique produit sur ce type de classification.

Dans ce cadre, nous avons étudié ces activités de gestion de l'information numérique dans une perspective inspirée par les théories de l'activité et les apports scientifiques de champs comme l'ergonomie ou la psychologie du travail. Ces disciplines nous ont fourni des cadres d'analyse de l'activité que nous avons appliqués aux activités de gestion de l'information, pour les caractériser.

Ainsi, l'approche que nous proposons de ces activités les considère comme des activités secondaires, coûteuses cognitivement, qui nécessitent des microdécisions de manière récurrente, et qui sont au service d'autres activités principales, dans la plupart des cas. Nous avançons qu'il est possible de valoriser ces activités de gestion de l'information, de les exploiter dans une vision prospective de l'activité, et qu'il serait nécessaire de leur (re)donner du sens, de la valeur, afin d'ouvrir leurs usages à des formes d'intelligences collectives constructives. Ainsi, il serait possible de leur reconnaître un impact potentiellement positif sur les pratiques métiers, en leur redonnant des perspectives à moyen ou long terme.

A partir de l'étude d'un outil de gestion de l'information, la classification à facettes, nous proposons une méthode d'élaboration de classification à facette, qui intègre dans son processus itératif, une vision contextuelle des métiers et de l'activité. Elle vise à replacer l'activité de gestion de l'information au centre des activités métier, dans le cadre du développement du numérique, qui est incontestablement un outil de communication, et un outil qui affecte les pratiques métier notamment, celles de partage et de transmission d'information, de connaissances, et plus génériquement celles liées au savoir.

Ces activités de gestion de l'information sont largement langagières, or, les communautés métier développent des formes de discours endogènes qui participent à l'acquisition d'une expertise. Ainsi la méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances métier, que nous proposons dans le dernier chapitre de ce document, vise à l'élaboration de classification à facettes émergentes pour des collectifs métier, et conçues afin de soutenir une démarche de gestion des connaissances métier contextuelle, appuyée sur les activités de gestion de l'information courantes.

Le premier chapitre présente donc les différents SOC que nous avons étudiés. Les principes de l'indexation, de la documentarisation, des classifications bibliographiques et à facettes y

sont abordés. Nous proposons également une étude des systèmes de tagging et de leur mode de structuration, notamment ceux qui exploitent les principes de la classification à facettes. A la fin de ce chapitre nous présentons le projet Miipa-Doc.

Dans le deuxième chapitre, nous détaillons les différentes méthodologies qui existent en matière d'élaboration de classifications à facettes. En premier lieu, nous abordons celle de Ranganathan, pour les bibliothèques, et celle du Classification Research Group (CRG) porté par B.C. Vickery, qui vise à la construction de classification pour des domaines spécifiques. Ensuite, nous présentons celle proposée par l'entreprise Cogniva qui a pour objectif le Record Management dans les organisations. Enfin, nous exposons celle élaborée par (Denton, 2003), inspirée des études et propositions de (Spiteri, 1998) et qui s'applique à la structuration de sites Web. Une synthèse des caractéristiques de ces méthodologies est présentée à la fin de ce chapitre.

Le troisième chapitre est consacré à l'analyse des activités de gestion de l'information d'après les approches de la Théorie de l'activité, et de celles issues de la psychologie du travail et de l'ergonomie. Nous avons choisi ce cadre d'analyse en référence aux travaux qui révèlent le manque de visibilité et de reconnaissance de ces activités de gestion de l'information. Nous nous sommes particulièrement intéressés aux aspects personnels de celles-ci, à travers la littérature sur le Personal Information Management (PIM) et les notions de contexte et de métier que nous développons. Le prototype Hypertagging développé dans le cadre du projet Miipa-Doc y est également présenté. Enfin, nous caractérisons l'impact des choix de types de Systèmes d'Information sur les activités de gestion de l'information.

Le quatrième chapitre présente notre expérimentation, qui avait comme objectifs, d'une part, la construction d'une classification à facettes pour les participants à partir de l'étude de leurs activités métier, et d'autre part, la production de classifications à facettes, pour leurs activités de gestion de l'information, de manière émergente, par ces acteurs. Enfin, nous dégageons les aspects communs de ces classifications, qui nous semblent liés au cœur de métier des acteurs.

Le cinquième chapitre propose une formalisation d'une méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances métier, qui s'appuie sur les

activités de gestion de l'information. Cette méthodologie s'inspire des apports de notre expérimentation.

Enfin, dans notre conclusion nous synthétisons nos apports et les prolongements souhaitables de cette recherche.

Chapitre 1 : Organisation des connaissances et Classifications à facettes

1 Organisation des connaissances et classifications à facettes

1.1 Introduction

L'organisation des connaissances est un champ large qui concerne de nombreuses communautés : Sciences de l'information et de la documentation, informatique, Knowledge Management, sociologie des sciences et au-delà. Si l'on considère les espaces où se concentrent les problématiques liées à l'organisation des connaissances, on en distingue trois principaux, le Web, les bibliothèques et les organisations de type entreprises. Les objectifs, les processus et les principes d'organisation des connaissances dans ces trois types d'espaces diffèrent à plusieurs niveaux : quantité d'information, dispositifs de classement et recherche d'information disponibles, types d'utilisateurs, activités mises en œuvre, finalité des espaces évoqués. Dans ce chapitre nous abordons les notions liées à l'organisation des connaissances dans les bibliothèques et sur le Web principalement. Ainsi nous détaillons les principes de l'indexation, des classifications énumératives et des classifications à facettes, ainsi que ceux des outils de tagging collaboratif sur le Web. A la fin de ce chapitre, nous présentons le projet Miipa-Doc (Méthodes et Services intégrés Institutionnels et Participatifs pour la Classification à Facettes des Contenus Documentaires Complexes) dans le cadre duquel cette recherche a été menée.

1.2 Organisation des connaissances et Gestion des connaissances

L'organisation des connaissances fait écho à des expressions variées dans la littérature, comme la gestion des connaissances, l'ingénierie des connaissances, l'organisation du savoir, la capitalisation des connaissances, les systèmes d'organisation des connaissances, le partage des savoirs, le management des connaissances, la mémoire organisationnelle. Hjørland distingue deux sens principaux pour cette expression. En effet, dans son sens restreint, l'organisation des connaissances évoque des activités comme la description de documents,

l'indexation et la classification, exercées dans les bibliothèques, les bases de données bibliographiques, les archives […] par les bibliothécaires, les archivistes, les spécialistes de l'information, les spécialistes de certains sujets, ainsi que par les algorithmes informatiques et les profanes. Dans son sens plus large, Hjørland considère que l'organisation des connaissances a à voir avec l'organisation sociale du travail intellectuel comme l'organisation des universités et autres institutions de recherche et d'éducation, la structure des disciplines et des métiers, l'organisation sociale des médias, la production et la dissémination du savoir. (Hjørland, 2008, p.2-3).

Pour Cotte (2007a, p.69), la définition « qui recouvrirait le plus la notion d'organisation des connaissances au sens strict est que les organisations se préoccupent de recenser, qualifier, classifier, organiser, partager le volume global de connaissances que les salariés produisent, brassent, manipulent. »

Il est communément admis que les connaissances représentent un capital pour les organisations. Ainsi, on trouve, dans la littérature, des expressions qui indiquent que nous sommes entrés dans une « société de la connaissance », où la source économique de base n'est plus le capital, ni les ressources naturelles, mais la connaissance, (De Medeiros Franco *et al.*, 2000). Par conséquent, la valorisation des connaissances peut générer des profits par l'amélioration de la rentabilité du travail (Cotte, 2007a). Pour Teulier et Girard (2005, p.4), qui voient l'organisation comme « le produit de l'activité des acteurs, le savoir-être ensemble pour produire quelque chose », les connaissances cruciales pour les entreprises sont les connaissances pour l'action. Ces dernières sont considérées comme telles lorsque l'on peut rendre compte de « leur mise en œuvre, leur concrétisation dans une action pour atteindre un but » (Teulier et Girard, 2005, p.3).

Une grande entreprise dispose généralement d'un organigramme, voire d'un réseau social qui permet aux acteurs de situer les services et les personnes dans la hiérarchie. Dans le cas où elle possède un réseau social d'entreprise, celui-ci permet de situer et d'identifier les compétences de personnes ressources. Pour autant, cela ne permet pas au centre de décision de l'entreprise de disposer de toutes les connaissances disponibles dans l'entreprise. Plus précisément, celles-ci reposent dans les pratiques de tous les acteurs de l'entreprise, dans les documents stockés sur divers supports, dont leurs postes informatiques individuels, ainsi que dans leur créativité et leur expertise, soutenues par ces documents, pour inventer des stratégies afin de répondre aux objectifs fixés par la hiérarchie. Dans cette perspective, les documents

fournissent donc un support à l'expertise des acteurs et nous pensons que leur contextualisation améliore les capacités d'adapter le savoir qu'ils possèdent déjà.

Les connaissances formalisées, celles qui sont inscrites dans des documents, ne constituent, malgré leur volume, qu'une partie seulement de l'ensemble des connaissances produites dans l'exercice d'un métier. Des méthodologies existent pour faire émerger les connaissances tacites comme celle proposée par Nonaka et Takeushi (1997). Pour ces auteurs, l'innovation est conditionnée par l'émergence de nouvelles connaissances organisationnelles. Cette création de nouvelles connaissances se caractérise par la conversion de connaissances. Ils distinguent les connaissances tacites qui sont personnelles, spécifiques au contexte, difficiles à formaliser et à communiquer, des connaissances explicites qui sont transmissibles dans un langage formel voire systématique. Il existe plusieurs modes de conversion des connaissances (socialisation, extériorisation, combinaison, intériorisation) mais celui le plus efficace est l'extériorisation c'est-à-dire la conversion de connaissances tacites en connaissances explicites, à l'aide de métaphores, analogies, hypothèses, concepts, modèles. Ce dernier processus prend effet dans un dialogue et/ou une réflexion collective.

Cependant, même la partie formalisée des connaissances d'une organisation et consignée dans des documents reste fragile. En effet, la mondialisation des marchés, la libéralisation de l'économie et l'impact des Technologies de l'Information et de la Communication (TIC) accélèrent les processus de décisions, en même temps que l'entreprise évolue dans un espace à trois dimensions, d'après Grunstein (2002). Une dimension globale, une dimension locale et enfin une dimension d'influence caractérisée par des interactions avec d'autres organisations. Cela rend nécessaire les supports informatisés car ils sont dé-corrélés de la présence physique, les connaissances sont donc disséminées sur de multiples postes et autres supports physiques dont la pérennité est toujours à surveiller. D'autre part, ces mêmes facteurs engendrent une instabilité dans l'environnement des entreprises, qui aboutit à des besoins de contextualisation des informations. Ainsi la forme d'incertitude que génère cette situation se répercute sur les Systèmes d'Organisation des Connaissances qui décrivent une réalité évolutive et sujette à interprétation. (Boydens, 2010). Ainsi, d'après (Dudezert, 2012, p.27) :

> *« Les connaissances de l'entreprise dans tous ses domaines de compétences (recherche et développement, production, achats, marketing, comptabilité, finances, etc.) deviennent une ressource à gérer au même titre qu'un parc de machines ou que les compétences des salariés. Pour faire face à cet enjeu, les entreprises sont alors amenées à mettre en place des démarches spécifiques de*

gestion appelées démarches de gestion des connaissances (GC) ou knowledge management (KM). »

Dans les années 2000, la littérature issue de l'ingénierie des connaissances se concentre sur des approches formelles, qui tendent par exemple à construire une mémoire organisationnelle définie comme « une représentation persistante, explicite, désincarnée des connaissances et des informations dans une organisation ». (Gandon et Dieng-Kuntz, 2005), ou encore à mettre en place des approches basées sur des modélisations successives de l'organisation, des tâches et des agents (Tort *et al.*, 2005), ou des méthodes comme MASK, basée elle aussi sur des modélisations (Ermine, 2000).

Dans le cadre de la capitalisation de la mémoire projet, caractérisée par des champs d'expertise multiples, une variété de méthodes ont été produites, comme la méthode IBIS qui cherche à représenter la logique de conception à l'œuvre dans les projets, ou l'approche QOC, le système DRCS et le logiciel DRAMA, qui ont les mêmes objectifs. Le méta-modèle EMMA vise à représenter le processus de conception de logiciel sous forme de buts à atteindre, tandis que la méthode SAGACE modélise les connaissances qui décrivent le système de production. D'autres méthodes de capitalisation existent comme la méthode REX, MEREX, SYGMA, l'approche « *componential framework* », ou la méthodologie CommonKADS. (Dieng *et al.*, 2000). Les méthodes comme MASK, REX, ou CommonKADS, visent à extraire les connaissances et à en faire des modèles conceptuels, orienté vers la résolution de problèmes. Ainsi, la structure de ce type de mémoire décrit généralement (Oswaldo *et al.*, 2004) :

- la définition du problème, ou le processus prescrit ;
- la méthode et la stratégie de résolution de problèmes suivie par l'expert ;
- une description des concepts manipulés dans cette résolution de problèmes.

Mais de nombreux auteurs se sont aussi intéressés à une approche contextualisée des connaissances (Mahé, 2005 ; Teulier et Girard, 2005 ; Sauvagnac et Falzon, 2000 ; Argyris, 1995, Grundstein, 2003) qui souligne l'importance de connaître le contexte de création des connaissances lors de leur réutilisation dans un nouveau contexte. Ainsi, pour (Teulier et Girard, 2005, p.12), « C'est l'activité elle-même qui est observée, en temps réel, ou des traces de l'activité (compte-rendu, plannings (…), qui révèlent des comportements, des ensembles

de choix ou d'actes inscrits dans un tissu social, commercial, technique ou organisationnel ». Grundstein (2003), propose une vision du Knowledge Management (KM) avec deux visées : une visée patrimoniale, qui a pour but la préservation, la réutilisation et l'actualisation ; et une visée d'innovation (intensive) qui s'applique à rendre possible l'apprentissage organisationnel et la création active de connaissances individuelles et collectives. Dans cette démarche, Grundstein (2002) souligne que « le poste de travail informatisé devient l'instrument incontournable des membres de l'entreprise étendue ». Ainsi l'enjeu se concentre sur le système d'information numérique nécessaire à l'acteur pour réaliser ses activités métier.
(Oswaldo *et al.*, 2004), adoptent une vision plus large de la gestion des connaissances, qui intègre la capitalisation des connaissances, mais aussi le partage et l'apprentissage. Pour (Poitou, 2007), gérer des connaissances correspond à mettre en œuvre des pratiques sur l'utilisation des connaissances, connaissances qui peuvent connaître trois états :

- un état inerte, où elles se trouvent simplement disponibles dans les appareils, qui constituent leurs dépôts (outillage, documents, site de travail) ;
- un état de transition, où elles sont sélectionnées et extraites par l'agent de leurs dépôts, pour être converties de connaissances inertes en savoirs productifs ;
- un état actif, où les connaissances sont engagées dans des actes productifs.

Ainsi, une distinction importante soulignée par (Wilson, 2002), qui adopte une vision critique de ce qui est entendu à travers l'expression « Knowledge Management », consiste à distinguer le savoir de son support. Ainsi, pour lui, le savoir se définit comme suit : « *knowledge involves the mental processes of comprehension, understanding and learning that go on in the mind and only in the mind, however much they involve interaction with the world outside the mind, and interaction with others.* » (Wilson, 2002). Si le savoir est vu comme les processus immatériels et conceptuels de compréhension, d'incorporation dans les structures de savoir existantes et d'apprentissage, alors, *« data and information may be managed, and information resources may be managed, but knowledge (…) can never be managed* » (Wilson, 2002). Ainsi, il est possible de transférer des données, supports, ressources, documents à propos de ce que l'on sait dans une base de données, mais pas le savoir, selon cette définition. De plus, d'après (Teulier et Girard, 2005), dans le cadre d'une démarche de gestion des connaissances, « il ne faut pas oublier que ce qui est réellement formalisé, quelle qu'en soit la forme, ne peut être qu'une information qui ne prend sens que pour des personnes ayant des schémas

d'interprétation partagés». Par ailleurs, le savoir professionnel combine des connaissances liées à un domaine spécialisé et un savoir pratique acquis par l'expérience (Osty, 2006). Une partie du savoir pratique est implicite entre les membres d'un même métier. Nous employons le terme « implicite » en référence à (Wilson, 2002), qui signale une confusion entre savoir tacite et implicite. Pour ce dernier, le savoir tacite est caché, inaccessible à la conscience de celui qui sait, tandis que le savoir implicite est exprimable.

1.2.1 Gestion des connaissances 2.0

Ainsi une démarche de gestion des connaissances peut viser à expliciter le savoir implicite à travers des documents, des bases de connaissances, des systèmes experts, des « groupwares » comme cela s'est mis en place dans la première décennie de l'existence de ces démarches, entre 1990 et 2000. D'après (Dudezert, 2012), d'autres modes de capture et de capitalisation des connaissances plus liés à l'individu et à sa pratique de travail ont vu le jour entre 2000 et 2007, comme les communautés de pratiques, les annuaires et les plateformes collaboratives. Depuis ces dernières années, on observe l'émergence d'un « nouvel idéal type » de l'organisation centrée « connaissance » ou organisation 2.0. Ainsi, d'après (Dudezert, 2012, p.28) :

> *« Pour les entreprises qui s'inspirent de ce modèle, la gestion des connaissances ne doit pas chercher à contrôler cette ressource, mais plutôt à agir sur ses voies de circulation pour faciliter son développement. Il s'agit de favoriser la création de relations entre les individus et entre les individus et l'organisation. Dans ce contexte, la structuration des flux d'information est déterminante. »*

Ce dernier auteur définit l'organisation 2.0 comme suit : « modèle idéal d'organisation d'entreprise dans le contexte de l'économie de la connaissance qui s'appuie sur cinq caractéristiques :

- un salarié devenu knowledge entrepreneur, convaincu que sa connaissance est son capital de travail et dont l'objectif est de valoriser ce capital en s'appuyant sur son réseau social ;
- une structure organisationnelle fondée sur les réseaux sociaux de l'organisation qui supposent une égalité des acteurs vis-à-vis de la connaissance et l'absence de hiérarchie ;

- une animation de l'action collective qui repose sur une auto-organisation entre acteurs, fondée sur la collaboration ;
- un manager qui n'est plus en charge du contrôle de l'exécution des tâches mais qui devient l'animateur de la collaboration ;
- une régulation de l'action collective qui ne repose plus sur l'autorité mais sur la persuasion (négociation et/ou endoctrinement) et le contrôle social » (Dudezert, 2012).

Cette orientation récente du KM profite de plusieurs outils orientés ou soutenus par les principes du Web 2.0, sur lesquels s'appuient les démarches de KM, comme les réseaux sociaux (Prevel, 2012), les cartographies des connaissances (Sellin, 2012), les dispositifs organisationnels comme des communautés de pratiques (Guillaume, 2012) et les Réseaux d'Echanges Réciproques du Savoir (RERS), développés notamment à La Poste (Van Den Abeele, 2012).

L'ergonomie aussi s'est intéressée aux pratiques de capitalisation des connaissances, notamment (Falzon et Sauvagnac, 2001) qui proposent une classification des théories de la construction des connaissances. A travers une caractérisation de la démarche adoptée, en fonction des connaissances visées par la démarche, du traitement des connaissances effectué et de la place des connaissances et activités dans le processus de capitalisation, ils proposent quatre profils de démarches de KM. Ainsi, ils présentent les quatre profils suivants de démarches de gestion des connaissances :

- un profil orienté technologie : il est basé sur le système d'information qui est nourri par le recueil de connaissances
- un profil dit « novateur » : il est basé sur le changement organisationnel et technologique comme moteur de la gestion des connaissances
- un profil « ergonomie » : il met l'accent sur le progrès individuel des opérateurs
- un profil « méta-fonctionnel » : il est fondé sur la reconnaissance du processus de capitalisation comme une activité à part entière, demandant à être outillée.

A propos de ce dernier profil, (Falzon et Sauvagnac, 2001) précisent qu'il est « souhaitable d'intégrer (…) l'activité méta-fonctionnelle à l'activité fonctionnelle. D'une part, il est plus

efficace d'accomplir ces deux activités simultanément et plus profitable à chacun de mener une activité réflexive aussi souvent que possible ». Les activités fonctionnelles sont « directement orientées vers la production immédiate », tandis que les activités méta-fonctionnelles, peuvent correspondre à « des activités de construction d'outils (outils matériels ou cognitifs), destinés à une utilisation ultérieure *éventuelle et* visant à faciliter l'exécution de la tâche ou à améliorer la performance » (Falzon, 1994). Elles sont souvent effectuées en marge du travail (en marge des activités fonctionnelles) et trouvent leur source dans le travail. Elles sont rarement formalisées et reconnues et plutôt spontanées et ignorées. Elles sont orientées vers le futur (Falzon, 1994). Ainsi, nous pensons que l'activité de maintenance d'un système documentaire personnel comme la réorganisation des dossiers d'une arborescence de fichiers sur un poste informatique peut constituer une forme d'activité méta-fonctionnelle.

Si le champ de la gestion des connaissances se focalise actuellement sur des démarches axées sur les outils issus du Web 2.0, comme les réseaux sociaux d'entreprise et les plateformes collaboratives, cela ne l'affranchit pas des problématiques liées à la structuration de ces espaces de gestion des connaissances. D'après (Thévenot, 1998), cette structuration est indispensable pour le partage d'information, car ne peut être transmis, que ce qui est structuré. Dans les démarches de gestion des connaissances axées sur les documents, cette structuration repose largement, sur les Systèmes d'Organisation des Connaissances (SOC) qui permettent la description des ressources contenant le savoir à capitaliser. Ainsi, d'après (Zacklad, 2012), « l'organisation des connaissances correspond aux procédés permettant la structuration de ces connaissances de manière à faciliter leur mémorisation, leur évocation, leur exploitation ». Les différentes méthodes de gestion des connaissances évoquées plus haut mettent en œuvre des SOC variés. Dans la partie suivante, nous présentons différentes approches existantes des SOC.

1.2.2 Systèmes d'Organisation des Connaissances (SOC)

Pour (Hodge, 2000), les SOC, dont le but est d'organiser les documents, pour faciliter la recherche et la gestion d'une collection, tandis qu'ils servent de pont entre les besoins d'information des utilisateurs et le contenu de la collection, englobent :

> *« all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification and*

> *categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontologies. »* (Hodge, 2000, p.3)

Pour (Hodge 2000), qui propose une revue de tous les SOC existants pour les bibliothèques numériques, ils présentent les caractéristiques communes suivantes :

- ils imposent une vision particulière du monde sur la collection et les items contenus en son sein ;
- une même entité peut être caractérisée différemment selon le SOC utilisé ;
- il doit y avoir une communauté suffisante entre les concepts exprimés dans le SOC et les objets du monde réel auxquels les concepts font référence pour rendre possible leur utilisation.

Il distingue trois catégories générales de SOC : les listes de termes (qui contiennent par exemple les glossaires ou les dictionnaires), les classifications et catégories et les listes de relations qui correspondent par exemple aux thésaurus ou aux ontologies.

(Zacklad, 2010c, p.2) souligne la difficulté de les comparer entre eux et propose cinq critères d'évaluation des SOC. Il en donne la définition suivante :

> *« Les SOC sont des ensembles de termes codifiés constituant un lexique disposant ou non de règles d'association explicites et permettant de faciliter les opérations de gestion portant sur des contenus documentaires ou sur les caractéristiques de situations non préalablement documentées. »*

Enfin, (Souza *et al.* 2010), quant à eux, en donnent une définition encore plus générique : « *knowledge representations based on concepts and with different degrees of relationships among them* ». Ces derniers auteurs font une revue de la littérature sur les SOC qui les conduit à affirmer que les comparaisons entre les SOC sont souvent basées sur un seul critère. Dans ces comparaisons, les ontologies sont souvent présentées comme la forme ultime, ou la plus élaborée des SOC. Il y a d'après eux, une confusion sur la terminologie précise des différents types de SOC (par exemple sur le terme « Taxonomie » ou « Ontologie »). Ces confusions tendent à obscurcir les détails qui différencient les SOC et les différents buts pour lesquels ils ont été conçus. Ainsi, pour (Souza *et al.*, 2010), on peut affirmer que tous les SOC sont le

produit d'une sorte de modélisation ontologique, mais l'usage du terme « ontologie » peut causer des confusions. En effet, les ontologies sont théoriquement capables d'accueillir n'importe quel type de modèle de relation et n'importe quel type de concept, ainsi, un thésaurus ou une taxonomie sont aisément représentable dans une structure ontologique. Sur ce point, (Hjørland, 2012, p.15) formule la question comme suit : les autres types de SOC sont-ils nécessaires pour la théorie de l'organisation des connaissances, ou bien les autres SOC doivent-ils être considérés comme des cas particuliers d'ontologies avec des jeux de relations sémantiques plus restreints ?

Sur un autre plan, (Souza *et al.*, 2010) mentionnent qu'il peut exister des variations non négligeables dans des instances particulières de ces types idéaux de SOC, qui peuvent faire varier certaines de leurs caractéristiques de manière importante. Ainsi la comparaison des SOC n'est pas chose facile, puisque certains SOC comparés sont construits pour des buts différents, avec des portées différentes et des applications variables. (Souza *et al.*, 2010), proposent une cartographie des SOC qu'ils classent selon la typologie suivante :

- les textes non structurés : cette catégorie inclut par exemple les résumés ;
- les listes de termes ou de concepts : elles incluent, par exemple, les folksonomies, les glossaires, les dictionnaires ;
- les structures de concepts et de relations : elles incluent, par exemple, les taxonomies, les classifications, les thésaurus et les ontologies ;
- les structures de concepts, de relations et les agencements ou présentations de ces structures : ils incluent, par exemple, les cartographies et les modèles.

On constate qu'il n'y a donc pas de consensus sur une typologie des SOC dans la communauté des chercheurs en SIC, francophones, ou anglophones. Par ailleurs, (Souza *et al.*, 2010) proposent une taxonomie des dimensions des SOC qu'il est possible de prendre en compte dans une comparaison entre ceux-ci, et qui permettrait de comparer des formes hybrides de SOC.

Pour nous, les SOC comprennent donc tous les dispositifs qui servent à renseigner des métadonnées pour décrire des aspects formels (attributs standards des documents comme la date, l'auteur, le titre, le format ou la taille), ou des aspects plus liés au contenu des documents (comme les métadonnées concernant le sujet du document). Ces informations de contenu sont traduites, dans le cadre de l'utilisation d'un SOC, dans le langage représenté dans le SOC, et ce, pour pouvoir retrouver les documents ultérieurement. Ainsi pour les

thésaurus, on décrit le contenu des documents à l'aide de descripteurs. Dans un SOC, il y a plusieurs niveaux d'analyse, d'une part, le niveau syntaxique qui concerne les types de relations entre les termes du langage et qui constitue la différence la plus importante entre les différents types de SOC, selon (Hjørland, 2012), mais aussi, l'organisation des termes du langage (hiérarchique, multidimensionnelle, réseau sémantique, ou présentation à plat comme les nuages de tags), les principes de structuration et de construction du langage documentaire. Ces deux derniers niveaux d'analyse concernent aussi le niveau sémantique (qui crée les termes ? Est-ce que ce sont des concepts ou des termes ? Sont-ils extraits des textes ou construits par une analyse humaine ? Comment réfèrent-ils au texte décrit ?). D'autre part, concernant les modalités d'utilisation (pré-coordonné, ou post-coordonné), d'alimentation, ou de maintenance, on distingue les langages documentaires contrôlés, des langages dits libres.

Ainsi, la structure d'un SOC peut être formelle (un modèle) mais les termes peuvent être alimentés par l'utilisateur librement. On reconnait généralement que les thésaurus, les ontologies, les classifications sont des langages contrôlés, et les folksonomies relèvent de langages documentaires libres. Enfin, il faut aussi prendre compte le dispositif dans lequel est inséré le SOC, c'est-à-dire, la base de données, l'interface Web, ou l'IHM d'un logiciel fonctionnant en local ou en réseau, les aspects liés au travail collaboratif (système distribué) et les buts pour lequel le SOC est conçu.

Sans passer en revue les définitions de tous les SOC, nous nous intéresserons, dans la suite de cette partie à certains d'entre eux, comme les classifications, et en particulier aux classifications à facettes, et aux folksonomies, mais il nous faut d'abord préciser l'approche que nous envisageons d'une des fonctions des SOC : l'indexation.

1.2.3 L'indexation

L'opération d'utilisation d'un SOC pour décrire un document est communément appelée indexation et celle qui concerne la recherche de document, est qualifiée selon les caractéristiques du SOC utilisé et du type de recherche effectué, par des termes qui diffèrent notamment dans le monde anglo-saxon. Ainsi en français, on parle de recherche d'information de manière générique, et depuis peu de recherche exploratoire, tandis que dans le monde anglo-saxon, on distingue l'*information retrieval*, de *l'information seeking*, de *l'information searching*, et du *browsing*.

Selon Chaumier (2002, p13), l'indexation est la méthode d'analyse documentaire la plus utilisée. L'analyse documentaire, selon cet auteur, vise « à représenter le contenu d'un

document sous une forme différente de sa forme originelle afin d'en faciliter la consultation ou le repérage dans un stade ultérieur. » (Chaumier, 2002, p.11). D'après l'ADBS, l'indexation peut se définir comme un « processus destiné à représenter, au moyen des termes ou indices d'un langage documentaire ou au moyen des éléments d'un langage libre, les notions caractéristiques du contenu d'un document (ressource, collection) ou d'une question, en vue d'en faciliter la recherche, après les avoir identifiées par l'analyse ». Ainsi, comme le remarque Chaumier (2002, p.13), l'indexation correspond à la phase où l'on détermine « les données ou critères à mémoriser et sur lesquels les opérations de recherche s'effectueront ». Cette indexation est réalisée à l'aide d'un langage documentaire. Dans le cas des langages libres, le choix des termes est laissé à l'indexeur ou à l'utilisateur du système.

Traditionnellement l'indexation qui concerne prioritairement le sujet d'un document, est vue comme un processus en deux étapes (Mai, 2005 ; Albrechtsen, 1993) : la première est l'analyse du sujet du document et la seconde consiste à le traduire dans les termes du langage documentaire utilisé. Mais il existe plusieurs type d'approches de ce processus, tandis qu'il en existe plusieurs modalités, comme l'indexation automatique ou humaine, qui constituent pour (Fidel, 1994) les approches de l'indexation orientées sur le document (« *document-oriented indexing* »), par opposition à l'approche orientée requêtes (« *request-oriented indexing* »), qui elle-même dérive d'une approche centrée sur les utilisateurs (« *user-centered indexing* »).
(Mai, 2005) aborde aussi ces différentes approches, mais un peu différemment. Il en existe deux principales, selon lui :

- l'approche « orientée-document » : elle se focalise sur l'entité et sa description fidèle. L'indexeur doit s'en tenir au texte et aux affirmations de l'auteur, il doit baser le sujet uniquement sur l'analyse des docs dans une vision à long terme ;
- l'approche « orientée-utilisateur » : elle implique que l'indexeur doit se demander « Comment rendre ce doc visible à des utilisateurs potentiels ? », ou bien,
- « Quels termes utiliser pour transmettre ce savoir à ceux intéressés ? ».
- L'indexeur doit avoir en tête les besoins des utilisateurs et leur terminologie quand il détermine le sujet et choisit les termes.

L'approche « orientée document » connaît une variante, l'approche « centrée document » (*document-centered*), dans laquelle l'indexeur prend le document comme point de focalisation pour l'analyse mais dans le choix des termes, il doit garder à l'esprit les questions qui peuvent

être posées dans le système. Elle inclut donc la prise en compte des besoins et questions des utilisateurs dans le choix des termes. Ces deux approches (« orientée document » et « centrée-document ») assument que le sujet d'un document peut être déterminé en dehors de tout contexte ou usage particulier.

Enfin, une dernière approche étend l'approche « orientée-utilisateur », c'est l'approche « centrée sur le domaine » (*domain-centered*). Dans celle-ci, le domaine est le point focal pour l'analyse. Cette analyse du document est basée sur l'analyse et la compréhension du domaine et des utilisateurs. L'analyse du sujet doit être placée dans le discours du domaine et des utilisateurs. Cette approche est issue d'un courant des Sciences de l'Information et de la Communication intitulé le « domain analysis » développé, notamment, par Hjørland. D'après (Mai, 2005), ce courant met en avant l'étude des activités et des produits d'un domaine pour comprendre les structures et significations sous-jacentes. Ainsi, dans ce courant de recherches, l'organisation et la représentation d'information devraient commencer par une analyse du contexte (vu comme le domaine), des discours et des activités. Ainsi, dans l'approche « domain-centered », le sujet d'un document est dépendant du domaine dans lequel le document est utilisé. L'indexeur doit, pour être capable de comprendre le rôle d'un document dans un domaine, connaître la nature du domaine, son histoire, les écoles de pensée, les paradigmes, les objets de recherche, les activités et les objectifs en cours dans ce domaine. Selon cette approche, la première phase de l'indexation, l'analyse, est une indétermination dépendante du contexte et est aussi un processus de décision. Il y a donc plus que deux étapes en indexation d'après ce courant, ce sont les suivantes, d'après (Mai, 2005) :

1. analyse du domaine ;
2. analyse des utilisateurs ;
3. analyse du rôle de l'indexeur ;
4. à quel discours le document participe ;
5. comment va être utilisé le document ;
6. quel est le sujet ;
7. traduction dans les termes de l'index.

Ces différentes conceptions du processus d'indexation impactent la conception adoptée de la notion de sujet, ainsi que l'évaluation qui peut être faite du processus.

1.2.3.1 Evaluation de l'indexation

Concernant l'évaluation, d'après (Bakewell, 1983), la plus connue des tentatives pour évaluer les performances des index est constituée des tests de Cranfield qui produisent quatre repères :

- la mesure de l'exhaustivité : il s'agit de la mesure selon laquelle le contenu d'un document est analysé par un indexeur
- la précision : il s'agit de la mesure de l'efficacité relative du système à travers la comparaison entre le nombre de documents pertinents retrouvés par rapport au nombre total de documents pertinents ou non, retrouvés par la recherche. C'est l'indicateur de la mesure du bruit.
- le rappel : il s'agit de la mesure de l'efficacité du système à partir du ratio entre le nombre de documents pertinents trouvés lors d'une recherche documentaire et le nombre total de documents pertinents existant dans le système. C'est l'indicateur de la mesure du silence.
- pertinence : la mesure selon laquelle le système permet de retrouver des documents pertinents

Les études de Cranfield sont importantes car elles ont ouvert des champs de recherche sur l'évaluation des systèmes d'indexation & de recherche d'information, parce qu'elles ont apporté de l'attention aux problématiques impliquées dans l'évaluation de la qualité des résultats de ces systèmes.

Une autre forme d'évaluation du processus d'indexation est réalisée par la mesure la cohérence inter-indexeurs, qui est bien souvent peu élevée. D'après (Giroux *et al.*, 1995), les études sur la cohérence inter-indexeurs se focalisent sur les variations des termes mais pas sur la nature des termes ou sur les processus qui les produisent. Pour (Fidel, 1994), les évaluations de la cohérence inter-indexeurs montrent la part éminemment subjective de cette opération. Elles concluent à un faible agrément sur les termes, entre indexeurs, et pour un seul indexeur dans des temps différents. Ces incohérences ont donc conduit certains auteurs (Gastaldy *et al.* 1994) à approcher l'indexation sous l'angle des processus cognitifs qu'elle met en cause, ou à approcher ce processus comme une activité de résolution de problèmes. (Giroux, *et al.*, 1995). Pour (Milstead, 1994), le haut niveau de décalage dans l'indexation provient en partie des différences de perception au sujet des ressources et de la valeur d'une ressource. A ce sujet, une étude comparative menée par (Gastaldy *et al.* 1994) sur la cohérence entre indexeurs à l'aide d'un langage contrôlé et d'un langage libre, conclut que les

recouvrements de termes choisis pour décrire le sujet des documents, sont plus élevés dans l'utilisation d'un langage libre que dans celle d'un langage contrôlé. Ceci est dû à la part d'interprétation nécessaire dans l'utilisation des termes des langages contrôlés.
Ainsi, dans le processus d'indexation à l'aide d'un langage contrôlé, il y a deux phases d'interprétation qui peuvent conduire à des incohérences entre indexeurs, celle qui s'opère lors de la définition du sujet du document, et celle qui est à l'œuvre lors du choix des termes utilisés pour décrire ce sujet dans le langage documentaire utilisé. De plus, cette variation des termes d'indexation peut être compliquée en fonction des objectifs de l'indexation. Ainsi, selon que l'indexation vise à déterminer le sujet principal, ou à distinguer le sujet principal du sujet secondaire, ou encore qu'elle vise à décrire exhaustivement tous les sujets abordés dans un document, ou enfin qu'elle vise à faire sens pour les utilisateurs, selon les approches « orientées-utilisateur », elle se trouve contrainte par des facteurs qui peuvent augmenter les occasions d'interprétation, ou d'incertitude quant à la cohérence entre indexeurs et aussi pour un même indexeur.

1.2.4 La notion de sujet

La problématique de la définition du sujet d'un document est complexe et ne peut pas être traitée sur le même plan que les autres propriétés du document, comme la date, ou le format, ou encore la langue. (Albrechtsen, 1993) et (Hjørland, 1992), parmi d'autres chercheurs, réintroduisent le concept de « sujet » comme un élément déterminant dans l'indexation, tout en recommandant aux indexeurs de ne pas se focaliser uniquement sur le contenu des documents, mais d'anticiper l'impact et la valeur d'un document pour un usage potentiel. (Albrechtsen 1993) identifie trois approches du concept de sujet qui sont les suivantes :

Une première conception simpliste basée sur l'extraction automatique dans laquelle le sujet est considéré comme une entité absolument objective qui peut être extraite des documents comme une abstraction linguistique directe en utilisant des méthodes d'indexation statistiques. D'après cette approche, l'indexation peut être complètement automatisée. Cette approche consiste à extraire des termes du texte intégral. Elle présente l'avantage d'apporter des gains économiques mais ne facilite pas le transfert du savoir.

Une deuxième conception « orientée contenu » qui implique une interprétation subjective. L'analyse du sujet, dans cette approche implique une identification de sujets qui ne sont pas

explicitement contenus dans la surface textuelle du document mais qui peuvent être perçus par un indexeur humain. Elle est la plus commune des approches de l'indexation par sujet. Une critique qui peut être apportée, est qu'elle considère le document comme une unité isolée et sans considérer les usages possibles du document. Elle a le mérite de faire émerger du document les aspects explicites et implicites des documents.

Une troisième conception « orientée besoin » implique une conceptualisation et une anticipation basées sur les utilisateurs, les usages, les besoins. Elle considère les données de sujet comme des instruments pour le transfert du savoir et vise à faciliter la recherche d'informations pragmatiques et de savoir. D'après cette approche, les documents sont créés pour communiquer du savoir, ainsi les données relatives au sujet devraient être adaptées pour fonctionner comme des instruments de médiation étant capables de rendre visible ce savoir pour toute personne intéressée. Elle est liée à l'approche « *domain analysis* ». Lors de l'indexation, l'indexeur, au lieu de dégager le sens implicite et explicite du document, doit se poser les questions : « Comment puis-je rendre ce document visible à des utilisateurs potentiels ? Quel vocabulaire utiliser pour transmettre le savoir contenu à ceux intéressés ? ». Ce sont les recherches de savoir de l'utilisateur qui déterminent la méthode d'indexation. Un document est analysé dans le but de prédire son utilité potentielle pour des groupes particuliers d'utilisateurs.

Pour (Buckland, 2012), l'indexation, vue comme la description du sujet d'un document, est une activité langagière et a donc à voir avec la conception adoptée du langage. Pour lui, l'indexation, réalisée à l'aide d'un langage libre ou contrôlé, est imbriquée dans une culture et cette activité partage les problématiques liées au langage naturel. Les métadonnées prennent effet et reflètent un contexte culturel. L'indexation est réalisée en vue d'un usage futur, tandis que la détermination du sujet du document est basée sur le savoir déjà construit, et donc sur le passé. Le sens des expressions est issu des usages du langage, culturellement fondés et situés temporellement dans le passé, ce qui fait dire à (Buckland, 2012), que l'indexation est dans une tension entre les discours passés et les utilisations futures des documents. Ainsi, les termes assignés aux documents pour les décrire seraient obsolètes de manière inhérente, ceci étant lié à la nature du langage. D'autre part, la détermination du sujet d'un document ne se fait pas uniquement en rapport avec un seul discours, mais avec une myriade de discours hétérogènes et simultanés. Ainsi le langage évolue au sein de communautés de discours, il est

le produit et le reflet de ces communautés. (Buckland, 2012) souligne ici l'instabilité et la multiplicité du sens des mots et donc l'ambiguïté du langage, qui est contradictoire avec la recherche d'une description stabilisée des documents en indexation.

On constate ainsi que l'indexation, vue comme une activité langagière, nous amène à nous interroger aussi sur les aspects sémantiques du langage. (Hjørland, 1992) aborde cette problématique sous l'angle de la question de l'existence de l'objectivité. D'une part, on peut dire que si une personne X et une personne Y exercent correctement l'analyse, elles arriveront au même sujet pour un document. Il s'agit d'une conception objective du sujet. L'indexation pourrait donc être réalisée indépendamment du contexte. Or, nous avons constaté précédemment que la cohérence entre indexeurs est souvent faible dans la description des sujets d'un document. Ainsi, si aucune instance subjective, dans son rôle relatif au document ne peut garantir une analyse correcte du sujet du document, cette analyse est toujours subjective et cela peut mener à une conception agnostique du sujet : il est impossible de dire ce qu'est un sujet et comment faire pour le déterminer. (Hjørland, 1992) propose une conception pragmatique du sujet pour trouver une issue à ces problématiques, en réfutant le fait que le point de vue des auteurs, des lecteurs, des spécialistes de l'information ou de toute autre personne (comme l'éditeur), ou leur compréhension subjective peuvent produire une vision certaine ou objective du sujet d'un document, ni définir le concept de sujet. Mais chacun de ces points de vue contribue à la détermination de ce sujet. Sa position s'oppose à la conception idéaliste platonicienne qui voit le sujet comme une idée et conçoit le concept comme une entité abstraite psychique ou mentale, qui existe par elle-même et pour elle-même. Les concepts généraux sont considérés comme représentant quelque chose d'universel, qui existe en dehors et indépendamment de toute conscience humaine, et qui existe en même temps prioritairement ou primairement (dans le monde des idées, séparé, comme Dieu dans la cognition à priori de Kant). Ainsi les idées auraient des propriétés universelles ou fixes. Le sujet pourrait donc être une idée qui pourrait être analysée en parties individuelles.
Pour (Hjørland, 1992), aucune procédure fixée ne peut garantir une analyse correcte du sujet et il propose de revenir à des aspects pragmatiques de l'indexation, qui a une fonction instrumentale ou pragmatique dans les bibliothèques ou les systèmes d'information. Il préconise ainsi de prendre en compte l'usage potentiel des documents comme un facteur déterminant dans la description du sujet des documents. Ainsi, pour lui, un sujet est toujours un sujet pour quelqu'un ou pour quelque chose, c'est un moyen orienté vers un but.

L'approche de Hjørland est basée sur une conception pragmatique du sujet et vise à en faire une théorie matérialiste. Le sujet se construit dans une évaluation des propriétés des documents en prenant en compte l'objectif d'optimiser la perception potentielle du document, qui dépend du contexte. Les sujets en eux mêmes doivent donc être définis comme « le potentiel épistémologique des documents ». Un potentiel est une possibilité objective (il ne s'agit pas d'idée subjective ou objective). Ainsi pour (Hjørland, 1992, p.185):

> *« A subject description is thus a prognosis of future potentials. This prognosis can be based on positive as well as negative judgements. The subject description can be seen as both a kind of vision and as an evaluation in relation to current research. The most important prerequisite in subject description is not a special kind of method, but is maturity in judgement. The use of subject systems thus also assumes interpretation. The user must enter into the universe of the system and its devising. »*

D'autre part, pour (Hjørland, 2012), le fait d'indexer un document, de lui assigner des termes issus d'un langage contrôlé, est en même temps une assignation de ce document à la classe des documents indexés par ce terme. Ainsi, il écrit :

> *« the act of labeling a document (say by assigning a term from a controlled vocabulary to a document) is as the same time to assign that document to the class of documents indexed by that term (all documents indexed or classified as X belong to the same class of documents). » (Hjørland, 2012, p.13)*

Cette conception de l'indexation du document, qui considère que les valeurs entrées dans un champ de base de données, créent des ensembles de document basés sur ces valeurs, amène Hjørland à affirmer que tous les SOC sont des systèmes de classification. Cela signifie que des documents indexés avec les mêmes termes forment un ensemble, et les termes sont vus comme des classes. Ceci nous conduit (Hjørland, 2012) à s'interroger, sur les critères de classification pour assigner un document A à une classe X. Dans ce cadre il promeut une vision interprétative des classifications, issue d'un mouvement interdisciplinaire intitulé le « *Evidence Based Practice* » (EBP) qui part de la médecine. Ce mouvement soutient que les classifications doivent améliorer et accorder la pertinence de la recherche d'information avec les critères de pertinence. En effet, on pense souvent que les outils sont neutres et objectifs mais dans tous les systèmes il y a des biais, comme par exemple dans les moteurs de recherche, qui peuvent être paramétrés pour fournir différents résultats et différents ordres dans les résultats. Pour faire ce paramétrage et/ou pour évaluer le système on a besoin, d'après

Hjørland, d'avoir une sorte de classification de ce qui devrait être trouvé. La pertinence utilisateur dans les LIS (Library and Information Sciences) est la principale évaluée mais, d'après ce courant de recherche, il vaudrait mieux, en médecine, baser l'évaluation de la pertinence sur des critères scientifiques plutôt que sur le jugement ou l'opinion des utilisateurs. Ainsi, classer devrait consister à faire des distinctions pertinentes en relation avec les buts du système, ce qui implique la considération et la négociation entre différentes vues et intérêts. D'après ce mouvement, les documents doivent être classés en fonction de critères explicites, basés sur une hiérarchie de méthodes de recherches. Ainsi pour (Hjørland, 2012), « it is important not just to consider classification to be about logical decisions, but to be based on interpretation of discourses and on the negociation of different interests ».

1.2.5 La Documentarisation

La définition d'une autre notion proche de l'indexation, mais plus adaptée selon nous à l'information sous forme numérique, est celle proposée par exemple dans (Zacklad, 2005), qui concerne la notion de « documentarisation ». La documentarisation qui consiste « à transcrire ou à enregistrer une production sémiotique sur un support pérenne puis à l'équiper d'attributs spécifiques visant à faciliter les pratiques liées à son exploitation ultérieure dans le cadre de la préservation de transactions communicationnelles distribuées » (Zacklad, 2005) constitue un enjeu majeur dans la conservation, la circulation et la transmission du savoir, en permettant l'accès aux documents et aux contenus numériques. En effet, dans cette définition, l'auteur précise que le support doit être pérenne, ce qui implique qu'à l'heure où l'information circule prioritairement sous forme numérique, cette condition n'est pas nécessairement remplie. La notion d'indexation ne questionne pas la forme ni le statut du document qui a pourtant évolué avec la numérisation des supports et la généralisation d'Internet. A l'époque où les bibliothèques et centre de ressources étaient les endroits privilégiés de conservation du savoir, le support le plus répandu, pour ce qui concerne les documents textuels, était l'imprimé, donc d'une certaine manière, le contenu était fixé une fois pour toutes, la forme était figée et la notion de document pouvait s'apparenter à la notion d'œuvre. Les collections étaient relativement homogènes. La forme numérique des documents permet, entre autres, de les rendre accessibles dans des versions successives, de les rédiger de manière collaborative, de les fragmenter pour en réinvestir des parties dans d'autres documents, de conserver une version d'un document à plusieurs endroits du réseau… Ainsi la pérennité de ces formes numériques des documents dépend donc aussi de leur documentarisation. La mémorisation

des attributs spécifiques liés au processus de documentarisation qui peuvent être, entre autres, d'ordre technique (contenu), organisationnel (coordination), de contexte (activité) ou de localisation (accès à un document), représente un effort conséquent, que les SOC viennent soutenir (Pikas, 2007). Les SOC jouent donc ici un rôle fondamental d'articulation entre les utilisateurs, les dispositifs de stockage et les connaissances contenues dans les documents numériques caractérisés, notamment, par leur fragmentation. Ainsi d'après (Cotte, 2005) :

> *« (…) ce qui se joue aujourd'hui, c'est la circulation des formes relevant de registres différents, et le recyclage de bribes, d'éléments d'information d'un système à l'autre. La matière première de l'information est destinée à être transmise, recyclée, rhabillée dans des modèles différents. » (Cotte, 2005, p.111)*

Dans les environnements 2.0, où l'utilisateur est devenu l'indexeur-concepteur-utilisateur des SOC, les métadonnées de contenu et les autres attributs standards (par exemple, la date, l'auteur, le format de fichier) et spécifiques (relatifs à l'usage en cours ou prévu du document, comme par exemple, l'intérêt, le type de contenu, l'activité) des documents sont renseignés par les utilisateurs, selon l'usage ou l'intérêt qu'ils ont pour la ressource. Les SOC ne sont pas contrôlés dans ces environnements, ainsi les problématiques relatives au sujet en indexation, peuvent être transposées aux autres attributs (standards ou spécifiques) descriptibles par les dispositifs 2.0.

Il n'y pas de cohérence intrinsèque aux attributs standards dans les dispositifs 2.0, ce qui signifie que la date apposée sur un documents numérique dans un dispositif 2.0, par des utilisateurs variés, peut relever de plusieurs fonctions : un tag relatif à une date peut évoquer la date de remise d'un document officiel, pour lequel la ressource décrite constitue une source d'information, ou il peut s'agir de la date à laquelle le document a été utilisé comme ressource, ou encore de la date d'édition du document (attribut standard). Ainsi l'indexation ou la documentarisation à l'aide des dispositifs 2.0, relève plutôt de l'annotation au sens que lui confère (Prié, 2000). Ainsi, pour cet auteur,

> *« Dans le cas le plus général, annoter un document, c'est attacher à l'une de ses parties une description qui correspond à un usage que l'on – ou que toute autre personne – souhaitera en faire plus tard.*
>
> *(…)*
>
> *Indexer, c'est avant tout décrire un document pour le retrouver. Annoter, c'est décrire son interprétation du document, en vue de n'importe quelle tâche d'exploitation de ce document. On indexe pour rechercher plus tard, on annote pour donner des traces de son interprétation, pour documenter la tâche que*

l'on est en train d'accomplir. Ces traces pourront alors être destinées à soi-même, ou bien partagées, et leur mise en place peut se faire de façon plus ou moins automatique, et plus ou moins collaborative. » (Prié, 2000, p.15)

Dans la partie suivante nous allons nous intéresser à deux types de SOC spécifiquement, qui se situent dans des conceptions plutôt opposées du travail de documentarisation, ou d'indexation. Ainsi nous abordons les classifications, et plus précisément les classifications à facettes, et les systèmes de tagging de ressources.

1.3 Définition des classifications au sens général

D'après (Hudon et Mustapha El Hadi, 2010, p.11),

« La classification est l'opération qui consiste à organiser des entités en classes, de sorte que les entités semblables ou parentes soient regroupées et séparées des entités non semblables ou étrangères. La classe est donc un ensemble d'entités (objets physiques ou virtuels, individus, attributs, etc.) présentant une ou plusieurs caractéristiques communes. »

Dans le même sens, selon Mills (2004), la classification est l'opération la plus fondamentale en indexation et correspond au fait de reconnaitre et d'établir des classes d'objets, et des sous classes, dont les membres manifestent une caractéristique particulière ou un jeu de caractères particuliers.

Selon Vickery (2008), c'est une liste de termes présentés dans une forme structurée initialement basée sur la relation générique et d'inclusion de classes. Les classifications générales sont basées prioritairement sur des disciplines enseignées, qui sont des méthodologies et des points de vues spécifiques qui correspondent à des activités et se concentrant sur des entités et phénomènes qui peuvent être observés du point de vue de plusieurs disciplines.

Pour Hjørland (2007), une classification est un ordonnancement d'objets (ou de processus ou d'idées) en classes sur la base de propriétés. Les propriétés des objets ne sont pas juste « données » mais sont disponibles sur la base de descriptions et d'une compréhension en amont de ces objets.

Pour Kwasnik (1999), la classification est un des processus de construction du savoir et des théories. Elle souligne l'importance des décisions sur les règles de distinction, dans les classifications hiérarchiques, dont la plus importante concerne le premier critère de division, car il détermine la forme et la portée représentationnelle de la classification.

1.3.1 Classifications génériques et spécifiques

Les objectifs d'un système de classification sont des facteurs déterminants pour la construction du système. En effet, pour Beghtol (2001), toute classification est créée pour remplir un objectif, il n'y a pas de classification neutre vis-à-vis d'un but. Les systèmes de classification reflètent un point de vue, les assomptions qui gouvernent chaque système donnent le contexte général du système et jouent un rôle pour sa fondation sémantique. Pour Hjørland (2007), les critères pour la classification devraient être basés sur une compréhension des buts spécifiques, valeurs et intérêts en jeu. La construction d'une classification doit suivre ses buts, et différents objets peuvent être classés de différentes manières pour des buts différents, requérant des structures différentes (Denton, 2003). Beghtol (2001) affirme que la complexité du monde du savoir ne permet plus de considérer un domaine de savoir comme stable ou unifié, les structures du savoir ont changé, avec la globalisation grandissante des échanges d'information. Il est donc nécessaire de repenser tous les aspects des systèmes de classification. Nous nous intéressons particulièrement à l'aspect contextuel ou spécifique des classifications. En effet, le *Classification Research Group* (CRG), créé en 1952, a produit une littérature fournie sur les classifications à facettes et notamment sur des classifications spécifiques. Nous entendons par ce dernier terme, des classifications dont l'objectif n'est pas d'englober tout l'univers du savoir, comme les classifications des bibliothèques tentent généralement de le réaliser.

Ainsi, le CRG préfère observer une collection de documents et construire une classification appropriée à cette collection. (Vickery, 1966). Ce dernier auteur, s'oppose à l'application mécanique et contrainte à un sujet, de catégories fondamentales (Vickery, 1960), qui doivent plutôt servir de guide pour suggérer des caractéristiques possibles qui ne devraient pas être ignorées. (La Barre, 2010). L'interdisciplinarité remet également en cause la viabilité de tout schéma de savoir universel implémenté dans une classification hiérarchique (Mills, 2004). En effet, dans les classifications bibliographiques, les premières divisions correspondent souvent à des disciplines.

Ainsi, le CRG propose de commencer par organiser les concepts eux-mêmes plutôt que de diviser l'univers du savoir en classes, puis de les diviser pour en dégager d'autres concepts selon une approche descendante. En effet, il objecte aux approches basées sur les disciplines qu'elles présentent les écueils suivants (Spiteri, 1995) : les mises à jour sont difficiles, notamment l'insertion d'un nouveau sujet dans le système de notation établi, certaines idées qui pourraient entrer dans plusieurs disciplines (catégories) ne sont pas représentables dans

ces classifications, enfin, elles présentent des collections d'items groupés rigidement. Cet exemple de l'interdisciplinarité est emblématique des besoins de pouvoir construire des classifications multi-points de vue.

Cela va dans le sens de la théorie des niveaux intégrés qui considère que le monde se développe du simple au complexe par accumulation de nouvelles propriétés divergentes et à certains points, des changements apparaissent qui transforment l'entité membre d'un groupe en membre d'un nouveau groupe. Cette approche propose à la place de l'approche descendante souvent utilisée pour construire les classifications des bibliothèques à partir de domaines de savoirs prédéterminés, une approche ascendante qui se propose de former les domaines de savoirs après avoir reconstitué ensemble les concepts et déterminé les domaines qu'ils constituent (Vickery, 2008). Cette théorie dans laquelle chaque niveau émerge du précédent mais possède ses propres propriétés, sans relation d'héritage, a été présentée dans le LeonManifesto (http://www.iskoi.org/ilc/leon.php) qui s'intéresse particulièrement aux problèmes d'interdisciplinarité. Il y est mentionné que les classifications devraient encourager l'interdisciplinarité en se basant sur les phénomènes étudiés plutôt que sur les disciplines et en combinant les phénomènes avec les théories et méthodes adaptées pour les étudier à l'aide de techniques analytico-synthétiques comme la classification à facettes. Cette théorie renforce l'idée que les domaines de savoirs ne peuvent être déterminés qu'après une analyse de leur composition plutôt que par des prédéfinitions des domaines de savoirs pour décider ensuite comment les décomposer. (Spiteri, 1995).

1.3.2 Les classifications bibliographiques

Les classifications bibliographiques sont généralement énumératives, organisées sous forme hiérarchique et basées uniquement sur la relation générique/spécifique, qui peut se traduire par « partie de » ou « sorte de » (Beghtol, 2001). Un schéma énumératif est caractérisé par le fait qu'il construit une vue d'ensemble de l'univers des pensées et des choses, dans une séquence continue avec une notation développée pour chaque terme, du plus générique au plus spécifique (Sayers, 1955 cité par Begthol, 2008). Les hiérarchies divisent et subdivisent les choses en groupes dans lesquels chaque nouveau groupe est une sous-espèce de son groupe parent. Ainsi, le principe d'héritage des propriétés se traduit par le fait que tout ce qui est vrai d'un groupe est aussi vrai de ses sous-groupes. Elles sont généralement dotées d'un système de notation. Le problème principal de ces classifications est l'indexation par sujets (Mills, 2004) et la difficulté à représenter des sujets composés ou pouvant appartenir à

plusieurs branches d'une arborescence, ou dépendants de plusieurs disciplines. Les principes de la classification à facettes sont élaborés spécifiquement pour pouvoir décrire des sujets complexes ou composés.

Dans le cadre du classement de documents, les systèmes de classifications visent trois fonctions d'après (Hudon et Mustapha El Hadi, 2010) : une fonction cognitive d'organisation des concepts contenus dans les documents, une fonction bibliographique qui permet la systématisation des représentations documentaires, enfin une fonction bibliothéconomique qui vise l'organisation physique des objets. Mais un schéma de classification ne peut remplir ces objectifs que s'il est complet, flexible et hospitalier (Hudon et Mustapha El Hadi, 2010). L'hospitalité d'un schéma de classification signifie qu'il doit pouvoir accueillir de nouvelles classes sans devoir reconcevoir toute la classification. Plus précisément, d'après (Kwasnik, 1999) les caractéristiques des classifications hiérarchiques sont les suivantes :

- l'intégralité : la première classe est la plus générique et décrit le domaine de la classification, elle inclut toutes les subdivisons placées sous celle-ci ;
- la relation genre/espèce : il n'y a qu'un seul type de relation entre les classes et sous classes qui est la relation générique/spécifique, ou genre/espèce ou « sorte de » ;
- le principe d'héritage : tout ce qui est vrai d'une classe est vrai aussi de ses sous classes ;
- la transitivité : comme les attributs sont hérités, les sous classes sont des membres de la classe supérieure mais aussi de toutes les classes supérieures ;
- des règles d'association systématiques et prédictibles : les règles de division sont déterminées à l'avance, ainsi toutes les entités d'une classe donnée se ressemblent de manière prédictible, et sont différentes des autres classes du même niveau de hiérarchie de manière prédictible ;
- Exclusivité mutuelle : une entité ne peut appartenir qu'à une seule classe ;
- La règle des critères nécessaires et suffisants : l'appartenance à une classe est déterminée par des règles d'inclusion qui dictent les attributs à posséder pour appartenir à la classe, ainsi si une entité possède ces attributs, c'est une caution

suffisante et elle appartient à la classe, même si elle peut posséder d'autres attributs.

(Kwasnik, 1999) évoque les avantages suivants des classifications hiérarchiques : elles tendent à l'exhaustivité, tandis que le formalisme de la hiérarchie, qui implique le principe d'héritage, permet de traiter des attributs complexes, de manière économe. Ainsi, chaque attribut n'a pas besoin d'être répété à chaque niveau, ils sont portés par la structure hiérarchique. De plus, cette structure et le principe d'héritage permettent d'effectuer des inférences entre les classes, pour découvrir des caractéristiques des entités lors d'une navigation dans la classification. Ce type de classification parvient aussi à donner des définitions des entités qu'elle contient, basées sur des attributs qui pointent les similarités et les différences entre les membres des classes. Enfin, elles permettent d'avoir une vue d'ensemble du domaine classé, si la représentation visuelle de la classification est adaptée.

Enfin, (Kwasnik, 1999) évoque les limites des classifications hiérarchiques, bien qu'elles constituent, pour elle, d'excellentes représentations du savoir pour des domaines matures dans lesquels la nature des entités et la nature des relations importantes sont connues. Ces limites sont les suivantes : Elles adoptent un point de vue unique sur les entités classées, ainsi, par exemple, un amphithéâtre grec pourra être considéré comme un lieu, mais aussi comme un site archéologique, ou un lieu touristique, ou encore comme un lieu de spectacle, ou enfin du point du point de vue de son architecture, de sa forme ou de ses matériaux. Il y a un point de vue largement prédominant dans ce type de classification. Ainsi, dans le cas où une entité décrit ou porte plusieurs concepts, ce type de structure est restrictif, puisque l'entité ne peut être positionné qu'en relation à un seul concept spécifique ou catégorie. (Wild *et al.*, 2009). Par ailleurs, elles nécessitent une bonne connaissance du domaine à classer pour déterminer les principes de division adoptés, ce qui peut apparaître préjudiciable dans des domaines où les connaissances émergent de manière continue. Enfin, (Kwasnik 1999) juge que les règles d'inclusion (critères nécessaires et suffisants) sont trop strictes, ainsi les entités d'une classe peuvent avoir des attributs communs entre eux, mais tous ne partagent pas nécessairement les mêmes attributs.

Si pour (Broughton, 2006), il n'est pas possible de créer une classification qui soit vraiment objective ou neutre, ou absolument correcte, pour (Wild *et al*., 2009), les classifications énumératives sont plus sujettes à la subjectivité que les schémas analytico-synthétiques puisqu'elles sont liées à l'identification de critères de division qui peuvent être dépendants d'un point de vue.

La DDC (Dewey Decimal Classification) et l'UDC (Universal Decimal Classification) sont des exemples de classification énumératives. La DDC comprend 10 classes principales, avec des subdivisions via 10 sous-classes. Les classes vont de la philosophie, des religions à la science, aux mathématiques, à la technologie, à la géographie et à l'histoire. Elle n'est pas sans défauts, et ceci provient du fait qu'elle manque d'objectivité, étant biaisée par les conceptions américaines du début du 20ème siècle (Wild *et al.*, 2009). La Classification Décimale Universelle (UDC) est un développement du système de Dewey, adaptant la structure des classes et adoptant la notation décimale. Mais elle diffère de la DDC en vertu de sa nature synthétique. La DDC autorise seulement des indices singuliers, quand l'UDC autorise les indices à être combinés. De plus, des indices prédéterminés peuvent y être annexés. La combinaison est permise par la mise à disposition d'une gamme d'opérateurs, permettant une composition de descriptions de classes. Ces opérateurs incluent des relations (par exemple statistiques de production automatisée), des extensions (par exemple l'ingénierie de la manufacture), et des additions (par ex. production et marketing). Les indices supplémentaires sont proposés sous la forme de tables auxiliaires qui proposent des indices prédéfinis pour des concepts qui réapparaissent souvent, par exemple, pour indiquer qu'un document est écrit dans une langue différente de celle du corpus principal. La description de types communs de concepts (par l'usage des tables auxiliaires) et la combinaison de classes suggèrent qu'un document pourrait être décrit par une combinaison de types de classes communs. Cette approche est exploitée dans la classification à facettes qui décrit cette notion par le fait qu'une entité donnée (un sujet par exemple) est composée d'un certain nombre de concepts différents. En isolant des concepts dans des facettes cohérentes, il est possible de reconstruire toute entité à partir d'une combinaison appropriée de classes à l'intérieur de chaque facette. (Wild *et al.*, 2009).

1.3.2.1 Caution bibliographique

Ces classifications bibliographiques sont la plupart du temps basées sur une notion inventée par Hulme en 1911, la caution bibliographique, qui consiste à fonder le système d'accès sur le volume de documents sur des thèmes contenus dans la littérature. Une classification, selon ce principe, doit refléter les sujets des livres qui existent déjà et ne devrait pas chercher à énumérer tous les sujets potentiels. (Beghtol, 2001). Hjørland (2007) précise au sujet de cette notion qu'elle ne consiste pas à identifier un texte pour en tirer une relation sémantique, mais à négocier entre les différentes affirmations présentes dans les textes et à en sélectionner une qui a le plus haut degré d'autorité ou qui est considérée comme la meilleure en fonction des

buts de la classification. D'autres types de caution sont mentionnés, comme la caution culturelle, qui correspond au domaine socioculturel dans laquelle la classification est développée et pour lequel elle est opératoire et qui influence le développement du système. Dans cette perspective, une classification peut être vue comme un document qui révèle les préoccupations, croyances, opinions de la culture qui l'a créée. (Beghtol, 2001). Hjørland (2007) affirme aussi que toute classification sera toujours un reflet d'un certain point de vue ou d'une approche des objets à classer.

1.3.3 Présentation générale de la classification à facettes

Créée en 1924, par S.R. Ranganathan, un bibliothécaire et mathématicien indien, la classification à facettes facilite l'expression des sujets composés, ce que permettent difficilement les classifications traditionnelles énumératives, qui n'autorisent pas facilement la combinaison de termes provenant de différentes parties d'un schéma de classification. (Spiteri, 1998). Dans ce nouveau modèle de classification, « les sujets des documents ne sont plus exprimés par un indice global tiré d'une liste existant à priori mais par une combinaison personnalisée d'indices-concepts. » (Maniez, 1999). Dans sa théorie de l'analyse à facettes, publiée en 1933 et actualisée en 1967, Ranganathan a proposé des jeux de critères, de postulats et de principes pour la construction de schémas de classification à facettes, qu'il a appliqués dans la Colon Classification. Elle est composée d'un modèle à cinq facettes qui constituent le modèle PMEST (Personnalité, Matière, Energie, Espace, Temps) qui est censé pouvoir décrire n'importe quel sujet, bien que Ranganathan reconnaissait qu'il n'y a pas de classification idéale ou universelle. (Hudon & Mustapha El Hadi, 2010).

La notion de facettes apparait souvent comme « l'apport théorique le plus important de ce siècle en science de l'information » (Maniez, 1999). La classification à facettes est créditée de nombreux avantages dans la littérature. Ceux qui reviennent le plus souvent sont l'expressivité, la souplesse, la simplicité, la cohérence, l'adaptabilité. (Maniez, 1999), (Ali Du, 2004), (Marleau *et al.*, 2009). Elle est aussi reconnue comme favorisant le browsing, la navigation et la recherche d'information (Brougthon, 2006). Ce dernier auteur explique notamment que la classification à facettes permet le *browsing* (qui consiste à parcourir rapidement un corpus à l'aide, éventuellement d'un SOC, pour savoir ce qu'il contient) grâce à sa structure logique et à sa capacité d'exprimer synthétiquement des sujets complexes ou composés. Sa structure compatible avec une interface utilisateur et le fait qu'elle offre des

points d'accès multiples vers les documents, favorisent la navigation à l'intérieur d'un corpus. Enfin la recherche est facilitée par le filtrage progressif basé sur des critères multiples (les facettes) (Broughton, 2006).

Pour Denton, (2003) la classification à facettes peut être définie ainsi : « un jeu de catégories mutuellement exclusives et conjointement exhaustives, chacune construite en isolant une perspective sur les items (une facette) qui se combinent pour décrire entièrement tous les objets en question, et que l'utilisateur peut utiliser en cherchant ou en parcourant pour trouver ce dont il a besoin ». Pour Beghtol (2008), elle correspond à une approche analytico-synthétique qui s'apparente à un dispositif syntaxique et non sémantique. Ainsi pour Beghtol (2001), qui précise que l'on peut trouver des hiérarchies dans les classifications à facettes, au même titre que les classifications énumératives sont construites sur ce principe, les critères qui différencient les classifications énumératives des classifications à facettes, résident dans la présentation des hiérarchies, et également, dans la manière dont la notation synthétique est prise en charge. Ainsi, la fonction de la notation synthétique est de pouvoir exprimer la combinaison de sujets qui ne sont pas listés dans la classification.

Pour Kwasnik (1999), la classification à facettes repose sur la croyance qu'il y a plus qu'une seule manière de voir le monde, et que même les classifications qui sont vues comme stables se révèlent provisoires et dynamiques. La classification à facettes est obtenue par le biais de l'analyse par facettes, qui est une technique détaillée, pour la littérature la plus récente, dans (Hjørland, 2013 ; La Barre, 2010 ; Spiteri, 1998). Globalement, il s'agit d'une technique en deux étapes, la première correspond à l'analyse, qui consiste à déconstruire un sujet en ses diverses composantes, et la seconde, la synthèse, qui concerne la génération d'un indice significatif intégrant les différents composants du sujet en fonction de règles syntaxiques. (Hudon & Mustapha El Hadi, 2010).

Une facette peut être définie comme l'une des dimensions d'analyse d'un sujet dans une classification multi-dimensionnelle, mais il est difficile pour les auteurs de s'accorder sur une définition, comme le montre Maniez, (1999). Nous y revenons en partie 1.3.3.5.

La classification à facettes fait usage du principe de coordination des concepts, qui permet de présenter le sujet d'un texte par la combinaison ou la synthèse de deux ou plus de symboles, chacun représentant un terme conceptuel indépendant. L'apport de la classification à facettes se situe au niveau du tri sur les regroupements de termes. (Vickery, 1966).

Selon les environnements où elle est déployée, (bibliothèque, Web, services de *bookmarks*, ou entreprise) ses caractéristiques évoluent, ainsi que les méthodes d'élaboration de ces

classifications à facettes. Nous revenons plus en détail sur ces méthodes d'élaboration de classifications à facettes en chapitre 2.

1.3.3.1 Principes originels de la classification à facettes selon Ranganathan

Dans cette partie, nous présentons les définitions données par Ranganathan (1892 – 1972) des différents types de classifications qui existent au moment de l'édition du « Prolegomena to Library Classification » (1967). Il mentionne, en effet, que les facettes peuvent être intégrées de manière plus ou moins fondamentale dans ces types de classifications. Le degré d'intégration des facettes impacte la construction des indices et l'évolutivité de la classification. Nous abordons également l'analyse par facettes et la définition du concept de facette. Enfin, nous présentons les avantages et les limites des classifications à facettes ; une synthèse des différences entre les classifications bibliographiques hiérarchiques et les classifications à facettes est proposée.

1.3.3.2 Présentation historique du « Prolegomena to Library Classification »

Dans le « Prolegomena to Library Classification » de Ranganathan (1967), on trouve les principes d'élaboration de classifications à facettes, conçus par Ranganathan lui-même. Ce livre édité pour la première fois en 1937 et réédité trois fois, (en 1957 et 1967) présente les différentes versions de la classification à facettes selon Ranganathan. Celle-ci a connu de nombreuses éditions, et dans cette édition la plupart des exemples sont donnés en référence à la 6ème édition de la classification à facettes, la 7ème édition étant en préparation lors de la publication du livre (même si l'on y trouve de nombreux exemples déjà aboutis faisant référence à la 7ème édition). On constate donc que l'élaboration d'une méthode de classification originale comme celle de la classification à facettes a nécessité un travail continu et itératif. Cet ouvrage, auquel nous ferons référence de nombreuses fois dans ce travail, est conçu comme un manuel de formation et fait usage d'une structure extrêmement rigoureuse à la manière des codes de lois. Dans ce volume, les différentes lois, principes, canons, postulats présentés ont une importance et une influence décroissante. Dans la partie consacrée aux principes normatifs, les 5 Lois Fondamentales de la science des bibliothèques, formulées pour la première fois en 1928 et publiées en 1931, sont abordées en premier et présentées comme une référence de portée supérieure en cas de conflit entre les autres règles annoncées ensuite. Pour Ranganathan le cœur de la classification en bibliothèques est la classification par sujets. Il distingue plusieurs types de sujet (nous reviendrons sur ce point

dans le chapitre 2) et plusieurs types de classification en œuvre dans les bibliothèques (que nous détaillons dans la suite de cette partie) :

- les classifications énumératives
- les classifications quasi-énumératives
- les classifications quasi-a-facettes
- les classifications à facettes rigides
- les classifications à facettes libres

Il distingue plusieurs plans dans l'élaboration des classifications : le plan des idées, le plan terminologique et le plan de la notation avec une nette dominante, dans l'ouvrage, des indications étant en lien avec le plan notationnel. Ceci s'explique par le fait qu'au moment de l'invention de la classification à facettes, la micro-informatique n'était pas développée ni diffusée, les problématiques des systèmes de notation étaient donc bien plus prégnantes qu'aujourd'hui. Elles sont maintenant plus liées à des aspects informatiques, la gestion et la génération des indices étant largement assistées voire complètement réalisées par les systèmes informatiques installés dans les bibliothèques.

Dans son livre de 1951, « Classification and Communication », Ranganathan révèle des préoccupations plus générales, qu'il relie à celles des classifications. Ainsi, s'intéressant aux aspects cognitifs il voit la classification comme une nécessité neurologique. (p.25). Sur le plan de la communication, il considère que celle-ci est meilleure si l'interlocuteur est doué en classification. La classification a à voir avec l'efficacité. Il explique que l'efficacité de la communication dépend de la clarté de l'expression qui dépend des standards de classement, mesurés par le degré d'utilité *(helpfulness)* et la rapidité de la performance. Il affirme aussi l'importance de l'étude de la classification en abstraction pour soutenir la construction d'une discipline de la Classification Pure qui faciliterait la pratique des classifications appliquées aux situations particulières et améliorerait le développement de l'humanité. (p.26). Pour lui, cette discipline de la Classification Pure devrait donc être enseignée dès l'enfance. Ainsi, il écrit :

> *« In every sphere of life, advancement of humanity depends on the extent to which such pure disciplines are established, cultivated, and transmitted to the growing children of successive generations. If human progress depends upon communication, as communication depends on classification for its efficiency, cultivation of Pure Classification and the practice of Applied Classification becomes truly necessary for the advance of man. » (Ranganthan, 1951, p.26)*

1.3.3.3 Les classifications selon Ranganathan

Ranganathan distingue les différents types de classification énoncés dans la partie précédente capables de traiter des types de sujets variables.

1.3.3.3.1 Les classifications énumératives

Les classifications énumératives, quasi-énumératives et quasi-à-facettes partagent la caractéristique de présenter essentiellement une liste *(schedule)* qui énumère tous les sujets décrivant les documents à classer (du passé, du présent et du futur anticipé). La classification de la Library of Congress en est un exemple. Les limites de ce type de classification sont les suivantes :

- elles sont difficiles à alimenter,
- la liste des sujets est longue à parcourir,
- les indices des classes sont rigides (monolithiques),
- elles sont dépassées rapidement, leur réactivité ou résilience est limitée.

1.3.3.3.2 Les classifications quasi-énumératives

La classification quasi-énumérative inclut une distinction par rapport à la précédente qui consiste à ajouter à la liste principale de sujets, des catégories communes qui énumèrent des composés de sujet communs. Dans ce type de classification, le potentiel de combinaison pour former des classes composées est accru. Elle présente les mêmes limites que les classifications énumératives à la différence que les combinaisons possibles pour les indices des classes les rendent non monolithiques. La Classification Décimale de Dewey (CDD) en est un exemple.

1.3.3.3.3 Les classifications quasi-à-facettes

La principale différence entre cette forme de classification et les précédentes réside dans le fait que la liste de sujets énumère des sujets basiques et composés (pour les définitions de ces termes, voir Chapitre 2, Partie 2.2.1.1) et qu'elle est complétée de catégories composés de sujets communs et spécifiques. Dans ce type de classification l'indice des classes est moins rigide, les indices ne sont pas monolithiques. La liste des sujets reste longue. La résilience est meilleure grâce à la combinaison de classes possible. La Classification Décimale Universelle (CDU) en est un exemple.

1.3.3.3.4 Les classifications à facettes rigides

La différence dans cette classification vis-à-vis de la précédente est que la liste de sujets ne contient pas de sujets composés. Ce type de classification est composé d'une liste de sujets basiques, de catégories de composés de sujets communs et spécifiques. L'indice des classes est une combinaison des trois types de listes. Dans cette forme de classification, les listes sont courtes, l'ajout de nouveaux sujets peut se trouver limité, leur résilience est améliorée, les indices des classes sont non monolithiques. La Colon Classification (CC) inspirée par les principes du mécano et conçue par Ranganathan en est un exemple. Les indices sont construits par assemblage et un symbole, le « : » (colon en anglais), connecte les différentes partie de l'indice. Dans ce type de classification à facettes rigide, les facettes et leur séquence (l'ordre dans lequel elles sont enchainées) sont prédéterminées pour tous les sujets allant avec une classe basique. Les trois premières éditions de la CC prévoyaient une formule de facettes pour chaque classe basique. Plus précisément, la rigidité provient du fait que les classes de composés spécifiques éligibles à l'utilisation avec une classe basique sont prédéterminées.

1.3.3.3.5 Les classifications à facettes libres

C'est la 4ème édition de la CC en 1952 qui aboutit à l'élaboration de ce type de classification, aussi appelée classification analytico-synthétique. Cette forme de classification reprend les principes énoncés pour les classifications à facettes rigides mais en omettant les formules prédéterminées de facettes et en proposant des jeux de postulats et de principes pour guider l'élaboration de la séquence de facettes appropriée. D'après l'auteur, quelles que soient les facettes apparaissant dans un sujet composé, elles sont retrouvées grâce à l'analyse par facettes. D'après nous, dans ce type de classification, la difficulté peut provenir de la formation nécessaire non seulement aux bibliothécaires mais aussi aux utilisateurs de la bibliothèque puisque la recherche d'un sujet dans une classification à facettes va nécessiter l'analyse par facettes du sujet de la recherche. Dans cette forme de classification un sujet n'a donc pas d'indice préétabli dans le schéma de classification, il est construit par les étapes suivantes :

- décomposition du sujet en ses concepts constituants fondamentaux,
- prélèvement dans les listes appropriées du schéma de classification des indices appropriés,

- synthèse des indices prélevés selon une méthode précise.

Une facette est donc un terme générique utilisé pour décrire tout composant, que ce soit un sujet basique, ou un composé de sujet.

1.3.3.4 L'analyse par facettes

L'analyse par facettes est une technique, ou méthode, par laquelle il est possible de construire une classification à facettes, cette dernière étant définie comme une structure. (La Barre, 2010) insiste sur cette distinction entre structure et technique. Cette méthode itérative est décrite comme le processus mental par lequel un jeu de caractéristiques, qui forme la base d'une classification, peut être énuméré. (Ranganathan, 1951). (Vickery, 1963), précise que l'analyse par facettes, en partie analogue aux règles traditionnelles de division logique diffère des classifications énumératives en trois points :

1. des règles de division strictes : « toute catégorie logique distincte doit être isolée, toute nouvelle caractéristique de division doit être clairement formulée, toute relation nouvelle doit être définie » ;
2. « les facettes ne sont pas intégrées dans des tables rigides et énumératives, mais ont la possibilité de se combiner, de telle sorte que chaque type de relation entre les termes et entre les sujets puisse s'exprimer » ;
3. « la classification à facettes rompt avec la restriction qui imposait à la classification traditionnelle la relation hiérarchique du type : genre/ espèce » (Vickery, 1963)

D'après (La Barre, 2010), il n'est pas possible de créer une classification à facettes sans conduire une analyse par facettes. Les principales étapes de création d'une classification à facettes sont présentées dans (La Barre, 2010) comme suit :

1. Définir le champ de la classification en posant la question : quelles sont les entités d'intérêt pour les utilisateurs ?

2. Formuler des facettes : examiner une gamme de matériau qui exprime les intérêts des utilisateurs, généralement des documents, pour en extraire des termes candidats, puis ranger ces termes candidats dans des groupes (des facettes) en définissant chaque terme en référence au champ de la classification

3. Construire un ordre hiérarchique des termes collectés dans chaque facette, et éliminer pendant cette étape, les doublons, réaliser les fusions nécessaires,

4. Créer des notes d'application pour les termes qui le nécessitent,

5. Ordonner les facettes,
6. Créer le système de notation,
7. Appliquer le système de notation aux facettes.

L'analyse par facettes ne concerne que la première et la deuxième étape. Cette dernière doit être réalisée selon un modèle de facettes préexistant, parmi ceux qui ont été élaborés, nous nous intéressons plus précisément à celui proposé par le CRG, à 13 facettes, dans le chapitre 2. Cette liste de facettes préexistante, sert de point de départ, de manière non prescriptive, pour rendre l'analyse à facettes complète et prévenir les omissions (La Barre, 2010). On peut s'interroger sur la validité d'un tel procédé, puisque selon les domaines, les facettes utiles aux utilisateurs peuvent varier, en nature et en degré de granularité. Cette procédure suppose que ce modèle est valable pour tous les champs, ce qui va à l'inverse de la volonté du CRG de produire des classifications spécifiques. Nous citons ici (Hjørland, 2013) qui a une vision critique de l'analyse par facettes du point de vue sémantique :

> *« It is problematic if FA is based on the view that meanings are given in a language, and that subjects can be constructed by a universal set of semantic primitives given by natural languages. (…) The decomposing of complex concepts into primitive concepts cannot simply be done by a priori, conceptual analysis. (…) The elements in classification are not given by the logical analysis of the philosopher or the information specialists, but are given by scientific discoveries and theories. (…) In faceted classification a given document is classified by taking one or more symbols from the appropriate facets and combining them according to certain rules. (…) The idea is that the same building blocks can be used for all purposes (cf. Meccano). The underlying philosophical assumption is that elements do not change their meaning in different contexts. According to modern theories of meaning, this is a rather problematic assumption » (Hjørland, 2013, p.29).*

Nous discutons les différentes méthodes d'élaboration de classifications à facettes plus longuement dans le chapitre 2.

Enfin, pour (Hjørland, 2013), il semble que certaines questions ne trouvent pas encore de réponse comme : les facettes sont-elles découvertes ou construites ? Comment sont-elles précisément identifiées ? Comment peuvent-elles être évaluées ? Ces incertitudes ou approximations au sujet de la classification à facettes peuvent provenir du fait qu'il n'y a pas de consensus sur la définition de ce qu'est une facette.

1.3.3.5 Qu'est-ce qu'une facette ?

Pour (Vickery, 1963), la définition d'une facette correspond à l'ensemble des divisions (valeurs de facettes), produit par la division d'un sujet à partir d'une seule caractéristique. Dans un article de 1999, Maniez tente d'en donner une définition. Il distingue des types de facettes catégorielles et structurelles en fonction du type d'entité auquel on applique cette notion (concept, ou sujet et objet). Ainsi, il commence par mentionner que le terme de facette a été emprunté au vocabulaire courant, ce qui ne facilite pas la recherche d'une définition. Puis il recourt à une analyse linguistique, selon les deux axes paradigmatiques et syntagmatiques du discours. Selon cette approche, l'axe syntagmatique correspond à la syntaxe ou à l'enchaînement admis ou correct des termes dans une phrase. L'axe paradigmatique correspond quant à lui à « la série des substitutions possibles de termes de la phrase » (Maniez, 1999). Par exemple, une phrase comme : « le chien mange un jouet frénétiquement » a une construction Sujet + Verbe + complément d'objet + complément de manière. Cette structure de la phrase correspond à l'axe syntagmatique en linguistique. Les termes de la phrase pourraient être remplacés par « l'enfant déguste une glace en plissant les yeux ». La structure de cette dernière phrase est la même que celle de la première phrase, mais les termes qui la composent ont changé, il s'agit de l'axe paradigmatique. Certaines catégories conceptuelles comme les fonctions grammaticales, (sujet, complément) peuvent être compatibles avec des catégories grammaticales variées, (pronom, substantif, nom propre). Ainsi il n'y a pas de véritable correspondance dans le langage naturel entre la fonction grammaticale et la catégorie grammaticale.

Dans la classification à facettes, notamment pour Ranganathan, les valeurs de facettes correspondent à l'axe paradigmatique, on choisit des valeurs dans des facettes, qui peuvent être interchangeables, selon les sujets à décrire. Et l'enchainement des facettes pour former un indice de bibliothèque, correspond à l'axe syntagmatique car les facettes sont ordonnées selon une syntaxe très précise. (Nous revenons sur la méthodologie de Ranganathan plus longuement en chapitre 2.) Mais, d'une part, le modèle PMEST (Personnalité, Matière, Energie, Espace, Temps) à cinq facettes de Ranganathan qui est un modèle très abstrait, entretient des ambiguïtés, sur le sens des catégories qui le composent, et d'autre part, l'unité sémantique contenue dans les facettes sont des concepts, d'après (Maniez, 1999). Ainsi, pour cet auteur, Ranganathan a conçu « un modèle d'expressions des sujets dans lequel les

catégories syntaxiques et paradigmatiques sont confondues sous le terme de facette ». Dans le même sens, (Beghtol, 2001), considère que, dans la classification à facettes, la structure est mise au service du sens. Ainsi la distinction entre la syntaxe et le sens est faible.

Si le CRG a proposé un modèle plus détaillé de facettes, il reste une abstraction qui contient des catégories fondamentales, comme Entité, Agent, Opération. Elles sont construites en remontant de valeurs spécifiques (comme, par exemple, la fragmentation) vers un niveau plus générique (une opération). Ainsi d'après (Maniez, 1999), les catégories fondamentales « sont les niveaux supérieurs d'une hiérarchisation des concepts obtenue par réduction progressive du spécifique au générique ». Il s'agit de l'opération inverse de celle effectuée dans une classification énumérative. Ces catégories fondamentales correspondent à un type de facette, qu'il appelle facettes fondamentales, ou catégorielles. Elles servent à classer des concepts et sont analytiques et applicables à toute classification indépendamment des domaines (Maniez, 1999). Pour le classement des sujets ou des objets, il préfère utiliser un type distinct de facettes, moins abstraites, plus spécialisées, qu'il qualifie de structurelles, et qu'il définit comme suit : « composantes essentielles de la structure d'une entité ou d'un sujet composite, dont la séquence ordonnées permet de classer par affinités ces entités ou ce sujets. » (Maniez, 1999). Enfin il propose une définition applicable à ces deux types de facettes, qui appelle les « facettes classificatoires », puisqu'elles servent au classement de ressources documentaires : « gamme d'attributs communs fondamentaux en nombre limité utilisé comme technique d'analyse et de classification des concepts et sujets » (Maniez, 1999).

Vickery, (1966) propose une liste de cinq questions pour déterminer des facettes :

- Quelles sont les entités d'intérêt pour les utilisateurs ?
- Quels aspects de ces entités les intéressent ?
- Quel concept un terme représente-t-il ?
- Dans quelle catégorie conceptuelle ce concept devrait-il être inclus ?
- Quelles sont les relations de classe entre ce concept et les autres concepts dans la même catégorie ?

(Mills, 2004) distingue des types de facettes distribuées selon les deux modèles de facettes de Ranganathan et du CRG : « *facet analysis might be said to be the assignment of terms to true categories (Time, Space, Matter, etc.) and to relationnal categories (Kind, Part, Agent, etc.)* ». Ainsi pour Mills, un document peut faire l'objet d'une infinité de descriptions de

sujets, dont la plupart seront des sujets composés. Ils requerront plus d'un terme pour les décrire. Les catégories issues du modèle PMEST de Ranganathan sont vues comme vraies ou fausses, et celles du modèle du CRG sont vues comme relationnelles, parce qu'une partie, une sorte ou un agent est relatif à quelque chose ou quelqu'un. Nous allons maintenant tenter de recenser les avantages et les limites de la classification à facettes dans la partie suivante.

1.3.3.6 Les avantages et les limites des classifications à facettes

Pour (Kwasnik, 1999), comme nous l'avons déjà mentionné, « la notion de facette repose sur la croyance qu'il y a plus qu'une seule manière de voir le monde et que même les classifications qui sont vues comme stables sont en fait provisoires et dynamiques». Les avantages des classifications à facettes selon (Kwasnik, 1999) et (Hudon et Mustapha El Hadi, 2010) sont les suivants :

- Leur conception ne nécessite pas un savoir complet sur le domaine, ce qui est utile pour des champs émergents ou qui évoluent ;
- Elles peuvent s'accommoder d'une variété de structures théoriques et de modèles, ce qui autorise une variété de perspectives. Ainsi, elles établissent des liens ponctuels, fonctionnels et pragmatiques entre concepts et/ou sujets, plutôt que des relations strictement taxinomiques permanentes, et l'ordre de présentation des facettes d'un sujet est ajustable au contexte et aux besoins ;
- L'expressivité : une approche par facettes peut être plus expressive car chaque facette est libre d'incorporer le vocabulaire et la structure qui convient le mieux au savoir représenté par chaque facette. De plus, il n'existe théoriquement aucune limite quant au nombre et à la spécificité des sujets qui peuvent être représentés ;
- La flexibilité : les composantes de la structure restent indépendantes les unes des autres jusqu'au moment de la synthèse d'un sujet et de l'indice qui le représente, et à la recherche également ;
- L'hospitalité : l'ajout de nouvelles entités est facile, la structure classificatoire est lâche et expansible à l'infini.

Les limites des classifications à facettes selon (Kwasnik, 1999) et (Hudon et Mustapha El Hadi, 2010) sont les suivantes :

- La difficulté à établir les facettes appropriées ;
- Le manque de relation entre facettes ;
- Les difficultés éventuelles de visualisation de la classification ;
- Le risque de vision trop locale, d'un degré de précision trop élevé dans la classification, qui la rendrait non interopérable ;

Vickery (1966) considère que les avantages de la classification à facettes pour des champs spécifiques du savoir produit les avantages suivants :

1. Par rapport à un système comme la classification UDC, elle rapproche plus facilement tous les aspects d'un champ de connaissances, qui dans UDC seraient des éléments éparpillés. Elle donne la possibilité de détailler plus avant les concepts d'un domaine de connaissances avec une notation plus courte. Les combinaisons de termes sont plus flexibles. Si elles sont produites pour un champ spécifique, elles peuvent accueillir des modifications liées à des changements de granularité spécifique.
2. Par rapport à des systèmes moins structurés, l'analyse par facettes produit un vocabulaire qui est plus cohérent pour l'indexation et la formulation de requêtes, tandis que les hiérarchies dans les facettes rendent plus faciles les recherches génériques.
3. Par rapport à des systèmes plus structurés, elles sont moins coûteuses et plus faciles à construire, et plus faciles à utiliser pour l'indexation et la recherche.

En résumé, le degré de structuration de la classification à facettes est vu comme optimal par rapport à d'autres systèmes par Vickery (1966). Les limites de la classification à facettes d'après (Vickery, 1966) sont les suivantes :

1. Par rapport à d'autres types de classification comme l'UDC, elle demande davantage d'efforts pour sa conception et doit être testée et améliorée, itérativement. Si elle est construite pour un champ spécifique, il est plus difficile d'intégrer des sujets marginaux dont l'intérêt émerge.
2. Par rapport à des systèmes moins structurés, elles sont plus lourdes à concevoir. L'ajout d'un système de notation alourdit aussi les processus d'indexation et de recherche.
3. Par rapport à des systèmes plus structurés, elle propose moins de facilités pour la recherche générique.

On constate, que ces deux dernières remarques sont à replacer dans le contexte de la publication de ce travail.

La Barre (2006) relève les questions théoriques suivantes qui restent à explorer et qui nous paraissent pertinentes :

- Quelle est la nature basique d'une facette ?
- Quelle est la valeur de l'exclusivité mutuelle dans une facette ?
- Quelle est la valeur de l'homogénéité dans les facettes ?
- Est-ce la notation apporte actuellement encore une valeur ?
- Les facettes peuvent-elles être universelles ?

Par ailleurs, Wild *et al.* (2009) soulèvent les difficultés suivantes :

- les interprétations différentes de la notion de facettes ;
- la confusion entre l'analyse à facettes et la classification à facettes ;
- le manque de guide méthodologique ;
- les exemples basés sur des domaines simplifiés dans la littérature ;
- l'impression que l'identification des facettes n'est pas problématique ;
- la question méthodologique sur la conception descendante (top down) ou ascendante (bottom up).

Nous présentons ci-après un tableau qui récapitule les différences entre les classifications à facettes et les classifications énumératives.

Limites des classifications énumératives	Solutions potentielles proposées dans les principes des classifications à facettes
Elles proposent un point de vue unique sur les objets à classer	Les liens sont ponctuels, fonctionnels et pragmatiques entre concepts ou sujets (possibilité de varier les perspectives ou points de vue)
Un seul critère de division y est toléré	Le principe de la classification à facettes est d'appliquer plusieurs critères de division simultanément
Pour les élaborer, le créateur de la classification a besoin de disposer d'un savoir global et complet sur le domaine	Elle ne nécessite pas un savoir complet sur le domaine, ce qui est utile pour des champs émergents ou qui changent
Toutes les entités d'une classe doivent présenter le même niveau conceptuel de granularité	En principe la classification à facettes applique aussi cette règle
La transitivité peut devenir une contrainte (comme les attributs d'une classe sont hérités, les sous classes sont des membres de la classe supérieure mais aussi de toutes les classes supérieures)	Les composantes restent indépendantes dans une classification à facettes
L'exclusivité mutuelle : une entité ne peut appartenir qu'à une seule classe	En principe la classification à facettes applique aussi cette règle
Les règles d'inclusion dans les classes peuvent s'avérer trop rigides	La structure classificatoire est lâche et extensible

Tableau 1. Synthèse des différences entre les classifications à facettes et les classifications énumératives.

1.3.4 Impact du numérique sur les principes des classifications

Nous allons ici aborder les évolutions que le numérique produit sur les systèmes de classification.

En préalable, il nous faut mentionner, dans le sens de (Beghtol, 2001) qu'une classification n'est jamais neutre, elle est culturellement imprégnée, tandis qu'il n'y a pas de système de classification parfait, ni complètement cohérent (Shirky, 2005). Dans un article de 2012, intitulé « *Is classification necessary after Google ?* », Hjørland, dans le cadre de son approche basée sur l'analyse par domaines, affirme que nous n'avons, à l'heure du numérique, besoin

que des meilleures classifications. Les moteurs de recherche sont devenus des outils très performants, ainsi, la plus-value d'une classification dans le cadre du Web, proviendrait du fait qu'elle pourrait donner à voir les points de vue à l'œuvre dans un domaine. Pour cela, l'indexation devrait être réalisée par des spécialistes dans une perspective historique. Il faudrait donc une indexation plus fine, experte d'un domaine, qui permettrait une indexation plus analytique ou critique plutôt que celle qui s'en tient à une vue objective du savoir, factuelle.

Traditionnellement, l'indexation des livres dans les bibliothèques est réalisée à l'aide des classifications énumératives et d'après une règle selon laquelle, si 20% du livre est à propos d'un sujet, alors il est assigné à une classe correspondant à ce sujet (Hjørland, 2012). Or, un livre peut comporter plusieurs sujets. Shirky (2005) à ce propos, explique que c'est l'objet « Livre » qui doit se trouver à un seul endroit dans la bibliothèque, mais pas les idées contenues dans ce livre, ainsi les principes des classifications énumératives dans les bibliothèques, sont construits d'après la contrainte physique du classement des objets. Dans le cadre du numérique ce type de contrainte ne devrait pas être à l'œuvre, puisque dans le numérique, «*there is no shelf* » (Shirky, 2005). Il y a ainsi, moins de restrictions sur les lieux de stockage dans le numérique (Wild *et al.*, 2009).

Ensuite, une classification bibliographique assure que pour chaque livre, sa place logique existe déjà dans le système, même avant que le livre soit publié. Cette stratégie consiste à définir à l'avance les catégories pour couvrir les cas possibles. (Shirky, 2005) estime que cette stratégie est dépassée à l'heure du numérique. Ainsi des classifications construites itérativement, voire collaborativement, grâce au numérique permettraient d'assurer le classement des sujets émergents de manière pertinente. Ceci rappelle le projet du Web socio-sémantique qui pourrait être rapproché des approches de Hjørland évoquées ci-dessus. Le Web Socio-Sémantique combine les approches du Web Sémantique et du Web Social. Il vise à documentariser par les utilisateurs eux-mêmes, des ressources auxquelles ces membres de communautés accèdent de manière distribuée, spatialement, temporellement, voire socialement sur le Web. (Zacklad, 2005). Plus précisément, « le web socio-sémantique s'oppose à la vision logiciste du web sémantique initialement proposés par T. Berners-Lee tout en s'inscrivant dans le projet d'une structuration des ressources documentaires partagées sur le web facilitant la réalisation d'activités coopératives distribuées » (Zacklad 2007).

Le processus de documentarisation des ressources, dans le cadre du Web Socio-Sémantique repose sur « un format de représentation de l'information, la métasémiotique HyperTopic,

permettant de construire et de partager aisément des ontologies sémiotiques de type cartes de thèmes (Cahier 2005) ou réseaux de description » (Zacklad, 2007a). Cette approche va dans le sens de l'évolution du modèle d'une bibliothèque reposant sur l'idée de stabilité du savoir (Hjørland, 2012) et de maturité du contenu (Wild, *et al.*, 2009), vers l'informatisation des contenus et la mise à disposition d'un savoir dont la pérennité est plus mouvante sur le Web.

De plus, les classifications énumératives ont été conçues d'après le principe de la caution bibliographique, qui implique que si peu de documents existent sur un thème dans le fond de la bibliothèque, alors il sera représenté minoritairement dans la classification. Ainsi dans la classification de la Bibliothèque du Congrès, la rubrique sur l'histoire, par exemple, comporte des classes pour la plupart des pays Européens, mais ne contient qu'une classe générique pour des continents comme l'Afrique, l'Asie, ou pour des zones géographiques comme la péninsule des Balkans (voir Figure 1). D'après (Shirky, 2005) mettre sur le même plan les zones géographiques en gras dans la figure 1 et les autres, est le signe que cette classification de l'histoire ne reflète pas la géographie, ni la population, mais qu'elle est construite pour optimiser le nombre de livres dans les étagères. Avec le numérique, ce type de contraintes n'a pas lieu de diriger la conception d'une classification et c'est la raison pour laquelle les problèmes d'inégalité de la représentation de certains sujets dans des classifications comme celui que l'on observer ici, ne devraient plus se produire. Dans ce cas précis, la conception du schéma de classification, en ignorant une partie des zones géographiques, dans la rubrique pour l'histoire, devient un parti pris politique. Mettre aujourd'hui, sur le même plan des pays comme la France et une zone géographique comme l'Afrique serait aujourd'hui perçu comme une conception indélicate politiquement.

D: History (general)

DA: Great Britain	DK: Former Soviet Union
DB: Austria	DL: Scandinavia
DC: France	DP: Iberian Peninsula
DD: Germany	DQ: Switzerland
DE: Mediterranean	**DR: Balkan Peninsula**
DF: Greece	**DS: Asia**
DG: Italy	**DT: Africa**
DH: Low Countries	DU: Oceania
DJ: Netherlands	DX: Gypsies

Figure 1. Classe Histoire de la Bibliothèque du Congrès, (extrait de Shirky, 2005)

Enfin, les systèmes de notation à l'œuvre dans les classifications énumératives pour les bibliothèques peuvent être simplifiés voir supprimés, dans les bibliothèques numériques, puisque comme l'explique (Wild *et al.*, 2009), un document numérique n'a pas besoin de telles exigences grâce aux possibilités de relier un document à une classe spécifique par le calcul avec des méthodes comme les URI (Uniform Resource Identifier).

Concernant la classification à facettes, elle commence à être déployée, de manière expérimentale, sur le Web, notamment comme un moyen de structuration des espaces de tagging, (ce que nous abordons dans la partie suivante) et dans les entreprises. Dans ce dernier cadre, (Wild *et al.*, 2009) note que l'on passe de classifications à facettes analytiques, au sens de la résolution d'éléments complexes en éléments simples comme pour des sujets ; à des classifications à facettes descriptives, qui relèvent de la description de caractéristiques des documents.

Dans la partie suivante, nous présentons les systèmes de tagging, les moyens de structuration des espaces de tagging et enfin le projet de recherche dans le cadre duquel cette recherche a été réalisée.

1.4 Le tagging collaboratif

Les discussions sur les outils de « tagging collaboratif » ont émergé sur le Web à la fin de l'année 2004 et en 2005, utilisant différents libellés pour les désigner, dans les articles et les post de blogs. Dans la littérature, concernant les définitions, il y a une tendance à combiner les concepts et à mettre sur le même plan le *bookmarking* social, le *tagging* et les folksonomies. (Trant, 2008). Ainsi, dans la littérature on trouve différentes expressions, telles que « folksonomies », « social tagging », « tagging collaboratif » ou « classification sociale », ou encore « Web 2.0 », pour désigner cette sorte d'outils que nous allons définir en détails ici. Certains parlent de « social tagging », expression qui se réfère à la pratique de nommage publique ou de catégorisation publique de ressources dans un environnement en ligne ou partagé. L'assemblage de tags qui en résulte forme une folksonomie, une combinaison du terme « folk » (peuple) et du terme « taxonomie » pour désigner un assemblage organique et informel de terminologie de la même famille. (Vander Wal, 2005).

> *« Social Tagging" refers to the practice of publicly labeling or categorizing resources in a shared, on-line environment. The resulting assemblage of tags*

form a "folksonomy": a conflation of the worlds 'folk' and 'taxonomy' used to refer to an informal, organic assemblage of related terminology » (Vander Wal, 2005).

Le Web 2.0 est orienté vers la conception par les utilisateurs et vers la participation sociale, le tagging par les utilisateurs – permettant la contribution des utilisateurs sous forme de mots-clés, pour l'organisation et la recherche d'information personnelle ou partagée. (O'Reilly, 2005) :

D'autres auteurs (Lin *et al.* 2006), décrivent le même type d'outil en le qualifiant de classification sociale. Elles dépendent des *inputs* des utilisateurs et de leur consensus, (…) leur construction fondamentale dépend de tags choisis librement. Il n'y a pas de hiérarchie, pas de relations spécifiées directement entre les termes, pas de groupes de tags prédéterminés ou de noms de catégorie pour organiser les tags. Golder et Huberman (2006) définissent le tagging collaboratif comme : « *the practice of allowing anyone – especially consumers – to freely attach keywords or tags to content* ». Ainsi le *bookmarking* social correspond à la pratique dans laquelle les utilisateurs décrivent des ressources qu'ils ont sélectionnées sur le Web, sous forme d'URL, les *bookmarks* (ou marques-pages ou encore favoris selon les terminologies des navigateurs, respectivement, Firefox et Internet Explorer) par des mots clés, librement choisis, des tags, pour pouvoir les retrouver ensuite et aussi les partager.

1.4.1 Présentation d'un outil de *bookmarking* social

Les outils de *bookmarking* social se substituent donc aux traditionnels marques-pages ou favoris, stockés dans le navigateur des utilisateurs individuels, en permettant un hébergement en ligne des *bookmarks*, dans un espace privé, protégé par une identification. Ces espaces de sauvegarde des liens d'intérêt pour les utilisateurs, existent sous différentes formes autorisant des variantes dans les autorisations liées à l'utilisation du service, notamment en termes de partage des liens tagués. Certains outils (comme www.Diigo.com) offrent une interface très riche en liens améliorant la navigation. Diigo propose une gestion fine des groupes d'utilisateurs permettant, d'une part, de partager des liens avec toute la communauté Diigo, d'autre part, d'avoir un espace strictement privé de sauvegarde des liens, et enfin, de partager des liens avec un groupe. L'adhésion au groupe peut être complètement paramétrée (par envoi d'une demande, ou groupe ouvert à tous, administration du groupe, et possibilité d'exclure des membres) et par conséquent maitrisée. Il existe aussi une fonctionnalité permettant de

suivre l'activité des membres qui, par exemple, partagent nos centres d'intérêts, ce qui tend à rapprocher ce type d'outil des réseaux sociaux. On pourrait ainsi parler de ces outils en termes de « réseau social fondé sur le partage de *bookmarks* ». Il est également possible d'effectuer des « recherches sociales », en composant une requête à base de tags, dans la base de bookmarks des utilisateurs, pour trouver des liens déjà repérés par d'autres membres de la communauté Diigo. On peut aussi chercher des liens dans la communauté à l'aide d'un dispositif de navigation dans les tags communément appelé « nuage de tags » qui présente les tags, distribués visuellement (souvent par ordre alphabétique), et dont la taille de police augmente en fonction de la fréquence d'utilisation du tag (sa popularité). Ce nuage de tags est aussi proposé pour chaque utilisateur individuel dans son espace personnel. Il agit comme un filtre pour la recherche, ne présentant que les liens tagués avec le (ou les) tags sélectionnés dans le nuage. Enfin, cet outil donne accès aux *bookmarks* de la communauté, en mettant en avant ceux qui sont les plus populaires, autorisant ainsi la détection de tendances dans les intérêts des utilisateurs.

D'autres outils du même type existent pour des types de ressources variés comme par exemple : Del.icio.us (www.delicious.com) pour des *bookmarks*. CiteULike (http://www.citeulike.org/) et Bibsonomy (http://www.bibsonomy.org), ces deux derniers sites servant plutôt à taguer des articles académiques, tandis que Flickr (http://www.flickr.com/) sert à taguer des photos et Technorati (http://technorati.com/) qui permet le tagging de post et de blogs.

Ces outils possèdent généralement les caractéristiques fonctionnelles suivantes :

- Ils indexent un type de ressource particulier (photos, URL, morceaux de musiques, articles de recherche académique...) ;
- Ils proposent de taguer les ressources à l'aide d'un dispositif de saisie des tags ;
- Généralement, il existe un système de recommandation de tags à la saisie ;
- Ils proposent un système de visualisation des tags sous forme de nuage de tags ;
- Ils proposent un système de recherche de ressource à l'aide d'un formulaire de recherche (qui peut être une barre de recherche simple - soit un moteur de recherche) ;
- Ils donnent accès aux ressources des autres utilisateurs ;

- Ils intègrent une fonction d'espace privatif ;
- Certains ont des fonctionnalités de gestion des groupes ;
- Certains présentent des moyens de catégorisation des tags (sous forme de « bundles » dans par exemple Delicious) ;
- Certains présentent des formes de catégorisation des liens (sous forme de listes de liens spécifiques dans Diigo par exemple) ;

La catégorisation des tags a donné lieu à de nombreuses publications dans le champ de la recherche (prolifique) sur ce type d'outil. Nous y revenons dans la suite de cette partie. (voir partie 1.4.6). D'après Marlow *et al.*, (2006a, 2006b), la conception des systèmes de tagging a un effet considérable sur le contenu et l'utilité des tags, et impacte fortement le comportement des utilisateurs. Ainsi ces systèmes de tagging doivent-ils répondre à différentes questions de conception. La liste suivante est élaborée d'après (Marlow *et al.* 2006) et (Voss, 2007) :

Droits de tagging :

- Qui peut tagguer ?
- Est-ce-que tout le monde peut tout tagguer ?
- Y-a-t-il des restrictions basées sur les ressources, les tags, ou les utilisateurs ?
- Qui décide des restrictions ?
- Y-a –t-il des distinctions entre différents tags en fonctions de différents types d'utilisateurs ? Ou de ressources ?

Source des ressources :

- Est-ce que les utilisateurs contribuent aux ressources ou collectent des ressources déjà créées, ou juste fournies par eux ?
- Est-ce que les utilisateurs taguent des ressources déjà présentes dans le système ?
- Qui décide de quelles ressources taguer ?

Représentation des ressources :

- Quel type de ressources est tagué ?
- Comment sont présentées les ressources pendant le tagging ? Et une fois taguées ?

Feedback du tagging :

- Comment l'interface supporte l'entrée d'un tag ?
- Est-ce que les utilisateurs voient les tags assignés aux autres ressources par d'autres utilisateurs ?
- Voient-ils les autres ressources taguées avec le même tag ?
- Est-ce que le système recommande des tags et sur quel algorithme sont basées ces recommandations ?
- Est-ce que le système refuse des tags inappropriés ?

Agrégation des tags :

- Est-ce qu'un tag ne peut être assigné qu'une seule fois à une ressource ? (le « set model », élimine les doublons)
- Ou un tag peut-il être assigné plusieurs fois à une ressource (« bag model » autorise cette action et permet d'effectuer des statistiques sur l'opinion des utilisateurs sur une ressource)

Contrôle du vocabulaire :

- Y a-t-il des restrictions sur quel tag utiliser et ne pas utiliser ?
- Les tags sont-ils créés lors de l'opération de tagging ou la gestion du vocabulaire est-elle une tâche séparée ?
- Qui gère le vocabulaire ?
- A quelle fréquence est-il mis à jour ?
- Comment les changements sont-ils enregistrés ?

Connectivité du vocabulaire :

- Les tags sont-ils connectés par des relations ?
- Ces relations sont-elles associatives ou monohiérarchiques, ou multi-hiérarchiques, ou typées (ontologies) ?
- D'où viennent les relations ?
- Sont-elles limitées au vocabulaire commun (pré-coordination) ou peuvent-elles être utilisées dynamiquement dans le tagging (post coordination et indexation syntaxique) ?

Connectivité des ressources :

- Comment les ressources sont-elles connectées ensemble : avec des liens ou par groupes hiérarchiques ?
- Les ressources peuvent-elles être taguées sur plusieurs niveaux de hiérarchie ?
- Comment sont faites les connexions ?

Tagging automatique :

- Le tagging est-il enrichi de tags, ou de relations, créés automatiquement ?

Connectivité sociale :

- Y-a-t-il des liens entre les utilisateurs ?
- Comment se font ces liens (groupe, suivi de l'activité des membres etc.) ?

Dans la littérature, les principaux axes d'approche de ces outils de tagging se distribuent entre les études qui s'intéressent aux applications variées du tagging (Web, images, *boorkmarks*, bibliothèques, entreprise, identification de l'expertise), et des études qui se focalisent plus sur des analyses statistiques le datamining sur des données issues du tagging. On trouve aussi des études qui ont pour centre d'intérêt le comportement des utilisateurs, (Kipp, Campbell, 2006 ; Feinberg, 2006), ainsi que des études sur les communautés créées par ces outils. Ensuite, on trouve des études qui portent sur la nature des systèmes de tagging vus comme des Systèmes d'Organisation des Connaissances (SOC) pour l'indexation et la recherche d'information. Enfin, on trouve des études qui portent sur la structuration automatique des espaces de tags, basée sur des algorithmes (Heymann, Garcia Molina, 2006).

Nous avons exploré les questions de conception liées à ces outils de tagging collaboratif. Nous allons ici nous intéresser à d'autres dimensions d'analyse de ces outils selon cinq axes complémentaires : 1) les outils de tagging en tant que SOC dans le processus d'indexation et leur impact sur ce processus, 2) le tagging et les processus de construction du sens partagé, puis 3) la nature et la fonction des tags, ensuite 4) les avantages et les limites que présentent les systèmes de tagging, et enfin 5) les projets existants concernant la structuration des systèmes de tagging.

1.4.2 Motivations, rôles des utilisateurs et indexation à l'aide de systèmes de tagging

Les systèmes de tagging étant largement basés sur la participation des utilisateurs, il faut ici mentionner quelles sont les motivations évoquées dans la littérature et quels rôles sont assignés aux utilisateurs. D'après Marlow *et al.* (2006), une grande part de la motivation des utilisateurs est déterminée par la conception du système et par la méthode avec laquelle ils sont exposés aux pratiques inhérentes aux tagging. Les motivations des utilisateurs sont divisées en motivations organisationnelles et sociales, elles comprennent :

- la possibilité future de retrouver des ressources,
- la contribution et le partage,
- le fait de pouvoir attirer l'attention sur une ressource,
- l'aspect ludique et compétition présents dans certains outils,
- la représentation de soi,
- la possibilité d'exprimer une opinion.

D'autres auteurs, comme Weick *et al.* (2005) et Golder et Huberman (2005) considèrent que le tagging est principalement lié au « *sensemaking* ». Zollers (2007) ajoute à cette liste la possibilité de faire de l'activisme, tandis que Kipp (2006), évoque le fait de gagner du temps et de pouvoir organiser ses propres ressources. Tennis (2006) considère que le tagging social est principalement bénéfique pour l'organisation personnelle de ressources, de la même manière que Vander Wall (2005) explique que le bénéfice immédiat pour le *Personal Information Management* (PIM) est une des principales motivations pour l'investissement dans ces systèmes. Cependant, il est important de noter, comme Tennis (2006) l'indique, que

le tagging constitue une remise en cause des systèmes d'indexation classiques basés sur des langages contrôlés et effectués par des catalogueurs experts. Par ailleurs, il écrit : « Social tagging highlights the interstices of authorship, intertextuality, and context in indexing, and asks us to fill in the gaps. It is a catalyst for improvement and innovation in indexing» (Tennis, 2006).

Considérant l'indexation dans son acception traditionnelle, vue comme une délégation par une institution à des catalogueurs, du processus d'analyse du contenu documentaire d'un fond et de la représentation de cette analyse à l'aide d'un outil (langage contrôlé), dans un système d'information, pour des utilisateurs, il pose ainsi la question suivante : « What's the point in dismissing the totalizing discourse in indexing? » (Tennis, 2006). Dans cet article, Tennis (2006) conçoit un système d'indexation classique avec un langage contrôlé comme un discours dominant ou totalisant, qui empêcherait des formes alternatives d'indexation et s'interroge sur les conséquences du rejet de cette pratique traditionnelle de l'indexation. Un des effets produit est que les utilisateurs des systèmes de tagging endossent plusieurs rôles, traditionnellement bien distincts. D'après Voss (2007), qui considère le tagging comme une indexation manuelle, les utilisateurs peuvent occuper quatre rôles :

- Auteur de la ressource : il crée ou modifie la ressource.
- Collecteur de ressource : il ajoute une ressource au système de tagging.
- Indexeur ou taggueur : il taggue une ressource.
- Rechercheur : il utilise des tags pour chercher des ressources.

Nous notons que cet auteur ne mentionne pas le (ou les) rôles liés au partage de ressources ou à l'administration de groupes sur ces plateformes de tagging social. Ainsi, les fonctions traditionnelles des catalogueurs et des langages contrôlés se voient remises en cause par l'usage des systèmes de tagging social. Pour autant, dans l'évocation des avantages et des inconvénients des deux systèmes d'indexation (l'indexation classique et le tagging social), on trouve des conceptions mesurées, comme celle de Tennis (2006) qui à travers une comparaison fouillée des deux systèmes, conclut que les similarités entre ces systèmes sont plutôt superficielles. Voss (2007), quant à lui, pense que la frontière entre les deux systèmes est moins stricte qu'il n'y paraît, puisque les systèmes de tagging intègrent des systèmes de recommandation qui agissent comme les langages contrôlés, et qui permettent de réguler, par les suggestions de tags, la variété des tags apposés sur un lien, même s'il n'y a pas de contrainte dans les systèmes basés sur le tagging. Dans le même sens, Golder et Huberman

(2005) démontrent, à travers une expérimentation, une convergence des tags et une stabilisation des tags liées à la pratique d'imitation et au partage de savoir, matérialisés dans les systèmes de recommandation de tags, selon le principe de preuve sociale qui induit qu'une action est vue comme correcte si l'on voit les autres la faire. Ainsi, choisir un tag utilisé par les autres est un choix sûr et ne requiert pas d'effort. Enfin Voss (2007) affirme que l'organisation du savoir nécessitera toujours des manipulations manuelles couteuses. Il pense qu'une interface de tagging bien élaborée produit de meilleurs résultats que n'importe quel algorithme élaboré. Il en tient pour preuve le phénomène Wikipédia, qui démontre, selon lui, qu'un groupe de volontaires peut créer de larges ressources de savoir, si un but commun et un système de droits adaptés existent.

> *« a clever interface to support tagging can be worth much more than any elaborated algorithm. Anyway the art of creating interfaces for developed tagging systems is still in its infancy. Knowledge organization will always need manual input so it is costly to manage – but Wikipedia showed that groups of volunteers can create large knowledge resources if a common goal and the right toolkit exist! » (Voss, 2007, p.7)*

Les outils de bookmarking social peuvent donc apparaître comme des méthodes alternatives d'indexation, de classification et de recherche de ressources issues du Web (Kipp, 2011). Il semble que l'on tente d'assigner aux folksonomies le même rôle que celui que jouent les thésaurus dans les bibliothèques et les bases de données, en termes d'indexation et de recherche d'information, mais appliqué à des ressources issues du web.

1.4.3 Le tagging, quel processus de construction de sens partagé ?

Certains auteurs s'intéressent aux folksonomies et aux tags dans une perspective de complémentarité ou de substitution avec des langages contrôlés existants, de type thésaurus, ou dans une perspective de construction d'ontologies à partir de la matière potentiellement très riche que fournissent les folksonomies.

En amont de ces réflexions sur la complémentarité de ces outils et sur leur éventuelle compatibilité, il nous semble important de nous intéresser à ces outils et à leurs unités constituantes (les tags) du point de vue des relations sémantiques entre les tags et les ressources décrites. En effet, chacun des langages documentaires que nous connaissons (thésaurus, taxonomies, ontologies) possède des règles de fonctionnement, de combinaison de ses éléments et de relation à ce qui est décrit, qui font l'objet d'un consensus relatif au sein

des communautés qui les utilisent. Il nous apparait nécessaire dans le cadre de cette recherche de tenter de clarifier les relations au réel (et aux ressources) qu'entretiennent les tags avec ce qu'ils désignent en fonction de leur différents usages. A travers ces réflexions, la dimension des usages très présente dans la littérature du domaine, nous renseigne sur les variations des utilisations et des formes de tags, mais ces outils, présentant la plupart du temps la possibilité de discriminer deux espaces (l'un privé et l'autre public), nous emmènent sur un terrain peu exploré, à savoir l'usage individuel et/ou partagé d'un système de description voire de catégorisation et d'accès aux ressources par le biais d'un langage personnel que nous qualifions de « laconique » ou « lacunaire » et éventuellement partagé. Les questions auxquelles nous tentons de trouver ici des réponses sont les suivantes:

- Les systèmes de tagging social sont-ils vraiment utilisés pour des usages partagés ?
- Comment un consensus sur les termes d'indexation, conditionnant le partage, peut-il advenir ?

Les études publiées dans la littérature sur les usages, la nature et la fonction des tags ainsi que la littérature sur les usages, les champs d'application des outils de tagging nous aideront à répondre à ces questions. Quelques lectures issues du champ de l'étude de la sémantique ont construit notre réflexion sur la relation référentielle des mots dans le langage en général, afin de comprendre celle qui est à l'œuvre dans les systèmes de tagging. Dans chacun des langages documentaires traditionnellement utilisés dans l'indexation, les relations sémantiques sont définies. Par exemple dans les thésaurus, les relations à l'intérieur du langage lui-même sont de type hyperonymie, hyponymie, relations d'association et de synonymie (Employé Pour). On n'observe rien de tel dans une folksonomie puisque la seule relation entre les termes est celle qui reflète la complémentarité par la combinaison de plusieurs tags pour décrire une ressource et leur éventuelle fréquence de cooccurrence.

En ce qui concerne la relation d'un descripteur (issu d'un thésaurus) à la ressource à décrire, M. Amar (2000) propose une caractérisation de ce type de termes liée aux deux fonctions de l'indexation (d'une part la caractérisation et la catégorisation du contenu des documents et d'autre part la recherche d'information). Ainsi un descripteur est l'expression linguistique d'un concept préexistant et stable (les concepts sont reconnus et extraits du document). Il est constitué d'une expression linguistique stable (sa formulation est stable). Il est aussi vu sous la forme d'un condensateur textuel (il réduit le texte à l'essentiel). Il est doté d'autonomie (il

peut être compris et utilisé en dehors du langage documentaire), et enfin, il sert de relai textuel (permet de faire le lien avec d'autres documents), (Amar, 2000).

Dans les systèmes de tagging de ressources, le producteur et le récepteur sont en premier lieu la même personne, ce qui n'est pas le cas dans la pratique des langages documentaires. Une autre divergence entre ces deux types d'outils consiste à observer que la nature du corpus est variable dans les systèmes de tagging de ressources issues du Web, (on peut taguer des vidéos, des liens, des photos) ; par opposition à l'usage des langages documentaires en bibliothèques et bases de données qui gèrent en général des types de ressources restreints et identifiés. La fonction de relai textuel est un principe de fonctionnement avéré dans les thésaurus tandis qu'on constate qu'il n'est pas toujours applicable dans les folksonomies à cause des variations orthographiques et des autres problèmes de synonymies et homonymies des tags (cf Infra).

Pour les mêmes raisons, auxquelles s'ajoute le nombre élevé d'utilisateurs dans ces plateformes de tagging, la formulation du tag n'est pas stable et la fonction de condensateur du texte n'est pas nécessairement le but recherché. D'après Smith (2007), les tags sont meilleurs que les mots clés contrôlés pour l'identification de sujets latents (ceux qui ne sont pas couverts par les règles d'indexation et d'analyse classiques du sujet dans les bibliothèques), comme les sujets secondaires.

L'autonomie d'un tag est aussi une caractéristique des mots du langage mais il faut parfois des éléments contextuels proprement personnels pour comprendre le sens d'un tag si sa formulation n'est pas correcte ou conforme à l'usage courant du langage. Ainsi un tag comme « ouv2.8Foc22obt60 » sera difficilement déchiffrable en dehors du contexte d'un site d'amateurs de pratique photographique. En l'occurrence cet exemple de tag servirait de mémo pour se souvenir des réglages liés à une photographie (ouverture du diaphragme 2.8, focale 22 mm, vitesse d'obturation 1/60ème de seconde) sur un site de tagging de photographies.

Concernant la fonction de représentation des concepts, une étude de Kipp (2011) révèle des catégories de tags qui décrivent, outre le sujet du document, des éléments liés au temps et aux tâches en lien avec la ressource pour l'utilisateur, des informations géographiques, des contenus spécifiques d'une ressource, des éléments affectifs, la priorité ou l'importance du document, le régime de droit associé au document, et enfin des éléments non compréhensibles pour un autre utilisateur que le créateur des tags qui se traduisent par l'utilisation de codes. Heckner (2007) propose un modèle de catégories linguistiques des tags et analyse d'où proviennent les tags, en termes de similarité avec le texte des ressources et leur position dans

les ressources (texte, titre ou résumé). Pour lui le tag typique provient du titre de l'article et est directement lié à son sujet, sous la forme d'un uniterme. Monnin (2009) insiste sur la variabilité des formes prises par les tags, décrivant le libellé des tags comme « un espace vide inscriptible susceptible d'accueillir les entités symboliques voire informatiques » laissées au choix de l'utilisateur. Ainsi pour cet auteur, les formes variées de tags comprises comme des chaines de caractères interdisent de déterminer une approche sémantique unifiée des tags.

Ainsi les tags servent à décrire des éléments plus variés que les langages documentaires, comme les thésaurus et par conséquent ne représentent pas nécessairement des concepts.

Par définition, l'usage individuel d'un tag issu du langage naturel, apparait comme pouvant potentiellement recouvrir plusieurs sens. Un terme polysémique dans le langage naturel restera polysémique dans une folksonomie, la désambiguïsation relative provenant de la combinaison avec d'autres tags, par une relation d'inférence entre les termes. Ceci est observable aussi dans le langage naturel et souvent, la clarification du sens d'un terme polysémique se fait grâce à notre connaissance de la langue (signification). Elle s'effectue également grâce au contexte, entendu comme les autres mots de la phrase et la reconnaissance des conditions de l'acte de communication, la situation, dont on déduit le sens (Nyckees, 1998). A ce propos, Hayes (2008) insiste sur le caractère fondamentalement ambigü de la relation de référence passant par la description. Citant Russel (1911), il rappelle que la relation référentielle peut se produire selon deux modalités, la connaissance et la description. La première concerne les objets désignés avec lesquels nous pouvons nous trouver en relation directe dans le monde. La deuxième a lieu lors de la désignation d'entités non accessibles, ce qui est le cas pour les ressources issues du Web, puisqu'il s'agit avant tout d'URLs et de textes (appréhendables par le langage), ou d'autres ressources non directement accessibles dans l'expérience sensible. En effet, nous considérons que les tags entrent dans le cadre d'une relation à la référence sous forme de description et nous nous bornerons ici à étudier les tags issus du langage naturel, prenant donc la forme de mots identiques à ceux que l'on trouve dans la langue, pouvant toutefois présenter des variabilités d'orthographe et de syntaxe pour les mots composés, et décrivant des ressources textuelles.

La relation référentielle dans le langage en général, se caractérise différemment selon les courants de pensée dans le champ de la sémantique. La notion de référence désigne une relation générale entre une expression linguistique et l'aspect de la réalité auquel elle renvoie. (Recanati, 2008). Le courant de la sémantique référentielle associe aux mots des choses, la sémantique cognitive associe aux mots des représentations mentales, enfin, la sémantique

pragmatique associe aux mots l'usage, l'emploi qui en est fait (Recanati, 2008). Plus précisément, la relation référentielle peut être envisagée sous l'angle suivant : si la référence d'une expression peut varier d'une occurrence à l'autre et être fixée par les relations que le référent doit entretenir avec l'occurrence, ce qui réfère n'est pas l'expression type mais l'occurrence de l'expression en contexte par le biais d'une relation contextuelle entre l'expression et le référent. La « signification linguistique de l'expression reflète la relation qui doit s'établir en contexte entre l'occurrence (token) de l'expression et son référent (Recanati, 2008).

Par exemple, un tag générique comme « évaluation » peut renvoyer à l'évaluation des performances d'un système technique, à l'évaluation au travail, à l'évaluation budgétaire d'un projet, à l'évaluation scientifique d'un article par la communauté scientifique, à l'évaluation quantitative d'une méthodologie par des experts, à une estimation, à la mesure de la température d'un liquide dans un filtre. A travers ces exemples, aux niveaux de précision variables, on voit le caractère fondamentalement relationnel du langage, lié aux différents usages possibles des expressions qui le composent. Si nous considérons l'aspect social du contexte, les acteurs qui emploient le terme générique « évaluation » vont l'utiliser dans le contexte de l'exercice d'une activité dans laquelle l'agrément sur le sens du terme, par les intervenants de la communauté liée au champ d'activité, inclut une convention de nommage et une acceptation de la norme concernant le sens de ce terme. Ceci décrit l'aspect fondamentalement social du langage qui consiste d'après Hayes (2008) à accepter l'agrément historique conséquent concernant la compréhension de noms et de phrases, et à accepter les normes de la communauté dans l'usage de ces noms et de ces phrases.

On remarque dans la typologie des tags dégagée par Kipp (2011) que les tags prennent des fonctions variées et pas seulement celle de description du contenu d'une ressource. Nous admettons que le fait qu'assigner un tag à une ressource est un acte de communication, mais celui-ci est d'une part et prioritairement orienté vers le taggueur lui-même, et d'autre part, éventuellement, vers la communauté qui utilise le service de bookmarking social. Cet acte de communication consiste à décrire le contenu d'une ressource, mais aussi à « baliser le chemin », afin d'en retrouver l'accès, afin de naviguer à travers le corpus. Il consiste aussi à organiser les contenus et à associer aux ressources des éléments affectifs, de priorité ou d'autres éléments organisationnels.

La pratique du tagging social se trouve donc pour une grande partie orientée vers l'utilisateur – tagueur lui-même et l'usage du langage n'apparait pas dans ces conditions

fondamentalement social mais tourné vers un usage prioritairement individuel et personnel. On constate donc l'aspect paradoxal des outils de bookmarking *social* dans lesquels *le langage perd sa caractéristique sociale* par le biais de l'appropriation individuelle des tags et du dispositif de tagging. En effet, dans cette optique le langage ne fait pas l'objet de conventions de nommage ni de consensus sur le sens des termes. Il s'agit d'un usage du langage naturel qui se révèle fonctionnel et qui pourrait se rapprocher d'un jargon. Dans ces conditions, le tagging fonctionne comme un langage, mais pas comme la langue naturelle. Il semble que l'on puisse en dégager des caractéristiques spécifiques suivantes :

- un tag ne suffit généralement pas à décrire une ressource et les cas d'usage d'un tag unique s'apparentent plutôt à une pratique de catégorisation qu'à une description de la ressource
- les tags révèlent leur sens par leur combinaison et leur relation à la ressource qu'ils décrivent
- il existe des types de tags liés à des types d'usage du tagging : par exemple description du contenu d'une ressource (tags décrivant le contenu) ou de l'usage que l'on veut en faire (tag organisationnels)
- ce langage est synthétique

Cependant, si au début, l'activité de tagging d'un utilisateur est majoritairement orientée vers lui-même, pour son usage personnel du système de gestion de ressources, les possibilités de partage de ressources produisent des effets de glissement comme l'adoption d'un tag d'un utilisateur à l'autre, ou au sein d'une communauté. Les recommandations proposées par ces systèmes de tagging social peuvent favoriser le consensus sur le sens d'un tag et le partage de ce tag. Enfin, comme l'explique Shirky (2005), les systèmes de recommandations de tags étant basés sur les cooccurrences de tags, et pas sur leur signification, la sémantique des tags se trouve chez les utilisateurs de ces systèmes et dans les communautés qui s'en servent.

1.4.4 Nature et fonctions des tags

Il est utile de préciser ici la nature et la fonction des tags. Ainsi, un tag est un mot-clé librement choisi par l'utilisateur, en langage naturel, pour décrire une ressource en termes de contenu, de forme, de nature, de tâche à effectuer en relation avec la ressource, ou encore, en termes d'appréciation de la ressource. L'agrégation des tags en un ensemble de tags d'un utilisateur ou d'une communauté constitue une folksonomie, présentée sous forme de nuage

de tags. Dans ces folksonomies, ces différents types de tags correspondant à des usages divers du dispositif, ne sont généralement pas discriminables.

Les tags servent à inscrire des métadonnées associées aux ressources décrites. Les métadonnées sont souvent définies comme des données à propos des données. D'après Mathes (2004), les métadonnées sont des informations souvent très structurées, à propos de livres, d'articles, de photographies, ou d'autres supports conçus pour supporter des fonctions spécifiques. Ces fonctions facilitent généralement l'organisation et l'accès aux informations. Traditionnellement, les métadonnées sont créées par des professionnels dédiés à cette tâche. Une alternative consiste à ce qu'elles soient créées par les auteurs des ressources ou encore par les utilisateurs des ressources. D'après Golder et Huberman (2005)**,** tandis que les tags sont descriptifs, ils réalisent les fonctions suivantes :

- identifier à propos de quoi ou de qui est la ressource,
- identifier ce qu'est la ressource,
- identifier à qui elle appartient,
- raffiner des catégories,
- identifier des qualités ou caractéristiques,
- ils servent de référence pour soi,
- ils servent à l'organisation des tâches.

On pourrait rapprocher les tags de la notion d'annotation que Van Setten *et al.* (2006) définissent comme toute assertion ou métadonnée à propos d'une ressource électronique. Ces auteurs ajoutent que « *knowing who (or what) provided the annotations can help people in determining the relevance of the annotations for their own goals* » (Van Setten *et al.* 2006). Sen *et al.* (2006) définissent les tags ainsi : « *short free form labels used to describe items in a domain* ». De plus ils conçoivent le fait d'appliquer des tags comme un investissement dans une ontologie personnelle pour organiser les items. Cette notion d'ontologie personnelle nous apparaît un peu excessive étant donné la nature en général formelle des ontologies, qui induisent des relations formalisées entre les concepts d'un domaine. Cependant la mention d'une ontologie personnelle nous paraît intéressante pour refléter le rapport que nous entretenons avec les artefacts personnels de classement et d'organisation des connaissances. Ainsi nous pensons qu'une partie des informations relationnelles entre les données, les métadonnées, les ressources et les tâches que nous réalisons, sont mémorisées dans un réseau

de liens que nous maintenons en permanence, et dont les SOC sont une partie visible. La partie invisible étant liée à toute sorte d'indices et de traces que nous laissons sur les documents, dans les métadonnées, et également dans le sens que nous associons à un tag et qui peut être hermétique à la compréhension d'un autre utilisateur.

Toujours d'après Sen *et al.* (2006), il existe des sortes de tags. Dans une expérimentation sur le tagging de films, ils tentent de lier des classes de tags aux tâches réalisées par les utilisateurs. Ainsi, ils considèrent que les tags « factuels » sont utiles pour apprendre et trouver des films, les tags « subjectifs » sont utiles pour l'expression individuelle d'une opinion, et les tags « personnels » sont utiles pour l'organisation. Leur expérimentation révèle aussi que les utilisateurs n'apprécient pas le fait de voir les tags personnels des autres utilisateurs. Aussi les auteurs recommandent de créer des fonctionnalités d'espaces publiques et privés dans les applications de tagging, pour cacher les tags personnels qui servent à l'organisation personnelle des ressources. Enfin, ils déduisent de leur expérimentation un autre résultat intéressant qui démontre que l'attitude des utilisateurs vis-à-vis de classes de tags suggère qu'il peut être valable de classer les tags dans les outils de tagging.

Dans une autre étude qui vise à comparer les termes d'indexation issus d'un langage documentaire, du tagging social, et d'une indexation automatique à partir des titres, Lin *et al.* (2006) arrivent à la conclusion qu'il y a peu de recouvrement entre les termes issus du tagging et ceux issus du thésaurus. Ainsi, les tags sont aussi utilisés pour d'autres usages que pour décrire uniquement le contenu des ressources tandis que ceux issus du thésaurus décrivent exclusivement le contenu. D'après ces auteurs, les tags ne décrivent généralement pas le contenu global mais plutôt des contenus spécifiques. Ils sont plus proches du langage naturel que les vocabulaires contrôlés. Le recouvrement entre tagging et indexation automatique sur les titres donne plus de similarités entre les termes, les tags étant souvent inspirés du titre de la ressource (selon le principe du moindre effort). Dans ce même article, les auteurs tentent aussi de trouver comment catégoriser les tags dans des groupes prédictibles. Dans leur expérimentation sur ce sujet, ils utilisent des photographies comme ressources pour le tagging, et ils concluent que certaines catégories sont presque universellement utilisées pour ces ressources, lesquelles catégories consistent en : Lieu, Personne, Sujet et Evénement. Une catégorie reste à part, mais hautement utilisée aussi, celle des tags composés, comme par exemple le tag « rockstarglasses ». Il s'agit d'une catégorie construite d'après la forme syntaxique du tag. Les utilisateurs produisent très fréquemment des tags composés plutôt que des unitermes. Nous pensons que cette pratique permet d'ajouter du sens contextuel au tag, ou

de donner plus d'information qu'à l'aide d'un uniterme, et éventuellement de catégoriser le tag à l'aide de sa formulation elle-même. Des recherches supplémentaires nous apparaissent nécessaires pour étudier la sémantique de ces tags composés pour en comprendre davantage sur leur construction. Enfin, ils trouvent un grand nombre de tags difficilement catégorisables, à cause de l'usage personnel du sens qui en est fait, cela démontre, selon nous, l'importance de la connaissance du contexte dans la catégorisation des tags.

1.4.5 Les avantages et les limites des systèmes de tagging

Comme le fait justement remarquer Mathes (2004), la plupart des problèmes auxquels pallient les vocabulaires contrôlés sont présents dans les folksonomies. Ainsi on trouvera chez de nombreux auteurs les critiques classiques, inhérentes aux folksonomies, correspondant aux problèmes d'homonymie (un mot peut désigner plusieurs choses différentes), aux problèmes d'ambiguïté des tags (un terme peut être utilisé par plusieurs utilisateurs avec des sens variables, au manque de gestion de synonymes (plusieurs mots différents peuvent désigner la même chose), et aux problèmes d'orthographes variables, ainsi qu'au manque d'harmonisation des formes au singulier ou au pluriel des termes.

Les forces des systèmes folksonomiques identifiées par Mathes (2004) sont les suivantes : ils renforcent la sérendipité par le « *browsing* », l'accès est simple, le coût cognitif est faible pour les utilisateurs, le feedback est présent et resserré (grâce aux suggestions de tags). Il y a la possibilité de partager des tags, et enfin, ils suivent les « *desirelines* » des utilisateurs. Concernant ce dernier avantage évoqué, il s'agit d'une notion mise au jour par Merholz (2004) qui considère que le fait de laisser les utilisateurs initier le travail de construction d'un langage d'indexation, à travers un système folksonomique, pour le rendre ensuite contrôlé, peut rendre celui-ci plus pertinent. Mathes (2004) considère que les folksonomies sont plus faciles à utiliser en regard de la nécessité de prendre des décisions sur le degré selon lequel une catégorie pré-définie convient. Il indique que les utilisateurs ne sont pas assez formés pour utiliser des langages contrôlés.

Golder et Huberman (2005) notent que le tagging est le plus utile quand il n'y a personne dans le rôle du documentaliste et qu'il y a trop de contenus à classer pour une seule autorité. Les désavantages des folksonomies évoqués par Quintarelli (2005) sont le manque de précision, le manque de hiérarchie, un quotient faible de retrouvabilité et des problèmes de montée en charge. Les bénéfices évoqués par cet auteur, consistent en le fait qu'elles reflètent les modèles conceptuels des populations et sont donc capables de convenir aux besoins réels

des utilisateurs en partie parce qu'elles sont ouvertes à tous. Un des bénéfices dérivés des folksonomies est de donner la possibilité d'agréger des données pour l'analyse et la création de communautés.

Kroski (2005) décrit des avantages, pour partie, liés à la « sagesse des foules » :

- Elles sont ouvertes, tout le monde peut participer,
- Elles sont actuelles (elles sont réactives et en constante évolution),
- Elles offrent la possibilité de découvrir des contenus,
- Elles ne sont pas binaires,
- Elles sont démocratiques et autorégulées,
- Elles suivent les « desirelines » (Merholz, 2004),
- Elles offrent un aperçu du comportement des utilisateurs,
- Elles permettent de créer des communautés,
- Elles offrent une alternative aux langages contrôlés à moindre coût,
- Elles présentent des qualités d'utilisabilité.

Du coté des désavantages, Kroski (2005) note, outre les problèmes classiques des folksonomies de synonymie et d'homonymie, le manque de hiérarchie, le faible taux de rappel lors de la recherche (elles ne garantissent pas l'exhaustivité des résultats), les possibilités de contournements (spam) dont elles peuvent faire l'objet. Enfin il évoque le problème du « basic level » qui se manifeste dans les différents degrés de précision des tags utilisés par les utilisateurs pour une même ressource. (Nous revenons à ce problème dans le chapitre 3 en partie 3.5.6.1).

Pour Peterson (2006), cité par Trant (2008), le relativisme des folksonomies présente des défauts inhérents, et ne pourront jamais produire l'exactitude, la précision, la justesse, des classifications formelles, au détriment des utilisateurs tandis que Weinberger (2006), affirme que le relativisme est une bonne chose, et qu'il faut s'accommoder des défauts précités, (le manque de cohérence et de précision). Il met en valeur la possibilité d'intégrer de multiples perspectives, de refléter des concepts émergents, de faciliter l'émergence de groupes sociaux via une sémantique partagée, et préconise ainsi de s'accommoder de la diversité. (Weinberger, 2005).

Dans Shirky (2004) on trouve, des arguments contre les langages contrôlés. Ainsi, le contrôle des synonymes réduirait les nuances, et sacrifierait le sens, la hiérarchie étant souvent forcée et fausse, (l'utilisateur doit s'adapter au système). La poly-hiérarchie serait essentielle pour la nature multidimensionnelle du sens. Dans Shirky (2005), l'auteur écrit que les classifications formelles ne conviennent pas aux ressources électroniques. Les classifications conviennent quand le domaine à organiser possède un petit corpus, des catégories formelles, des entités stables, des entités restreintes, des liens clairs, avec des participants qui sont des catalogueurs experts et qui sont aussi des sources de jugement d'autorité fiables. D'après cet auteur, sur le Web, le corpus est large et indéfini, aussi, "the only group that can categorize everything is everybody".

D'après Smith (2007) les tags sont meilleurs que les mots clés contrôlés pour l'identification de sujets latents (ceux qui ne sont pas couverts par les règles d'indexation et d'analyse classiques du sujet dans les bibliothèques), comme les sujets secondaires. Enfin, les avantages des folksonomies avancés par Hayman (2007), sont les suivants :

- elles sont multidimensionnelles : plusieurs tags peuvent être assignés pour exprimer un concept et il est possible de les combiner,
- elles reflètent le langage des utilisateurs,
- les utilisateurs peuvent souligner ce qui est important pour eux dans le document grâce aux tags,
- les possibilités de partage des tags et de création de savoir par agrégation,
- il n'y a pas besoin de stocker une ressource dans un seul dossier, chaque tag constitue une collection instantanément,
- elles permettent le développement de communautés,
- elles autorisent la détection de tendances sur le web,
- le tagging est une méthode d'indexation rapide, simple, directe et il est possible de l'utiliser sans formation à l'indexation.

Les limites énoncées par Hayman (2007) sont, outre les critiques classiques, les suivantes :

- les tags peuvent être pauvres,
- les différents niveaux de spécificité des tags (qui correspond au problème du « basic level »),

- les significations personnelles sont fréquentes dans les tags,
- il peut y avoir des ambiguïtés, par exemple, dans un cas où un lien est tagué avec le tag « Video », on ne sait pas s'il s'agit d'une vidéo ou bien si cela est à propos de vidéo,
- le choix des tags peut changer en fonction des tendances,
- la forme des tags autorisée dans les systèmes de tagging est souvent des unitermes : il en résulte des formes de tags idiosyncratiques, composées de plusieurs tags attachés (par exemple, « rockstarglasses »),
- les utilisateurs ont tendance à favoriser les tags les plus populaires,
- Le manque de relations entre les tags à la manière des thésaurus (Terme Générique ou Terme Spécifique) que l'on pourrait rapprocher de la critique qui pointe le manque de catégorisation des tags.

1.4.6 La structuration dans les systèmes de tagging

Les principaux champs qui font usage, ou projettent de faire usage, des systèmes de tagging sont le Web et les entreprises. Diverses thématiques traversent ces deux champs. Parmi celles que nous avons relevées dans la littérature, certains projets tentent d'utiliser les systèmes de tagging pour élaborer des langages contrôlés, d'autres s'orientent vers le Knowledge Management et la gestion de l'expertise, enfin beaucoup de projets tendent à trouver des moyens de structurer les folksonomies existent.

Une partie de la littérature sur les folksonomies est en lien avec celle sur la construction des ontologies pour le Web Sémantique. En effet, certains auteurs (Marchetti *et al.* 2007, Dix *et al.* (2006), (Monnin, 2009), considèrent que les folksonomies constituent une matière riche et adaptée pour servir de support à l'élaboration d'ontologies. D'autres auteurs, comme Hayman (2007) dans le champ du e-learning projettent de construire des « taxonomies – directed folksonomies », dans lesquelles les utilisateurs se voient suggérer des termes issus de thésaurus à la saisie des tags, sous forme de recommandations. Cet auteur évoque aussi l'approche qui consiste à construire un langage contrôlé à partir des tags de la folksonomie, notamment dans le contexte de l'entreprise. Elle écrit d'ailleurs:

> *« Tags might reflect emerging concepts that are of interest in the enterprise, and serve to capture employees' knowledge. Tagging could enable the development of a corporate folksonomy that could be integrated into an enterprise-wide vocabulary. » (Hayman, 2007)*

Les avantages liés au tagging en entreprise, qu'elle relève sont reportés ici :

- Il évite la frustration des utilisateurs liées aux langages contrôlés et rigides, et à des systèmes inadéquats de RI ;
- Ils sont peu chers, et simples à implémenter ;
- Ils facilitent la démocratie dans les lieux de travail ;
- Le tagging répartit le processus chronophage de classement du contenu des ressources ;
- Il favorise la création de réseaux sociaux autour de sujets spécifiques ;
- Ces réseaux peuvent refléter les intérêts et l'expertise des utilisateurs ;
- Ils peuvent servir à capturer le savoir des employés.

Egalement dans le domaine des entreprises, John et Seligman (2006) présentent un scénario pour l'identification d'experts dans l'entreprise à travers un système de tagging de personnes basé sur l'expertise et des clusters de tags reliés. Dans ce projet, on ne taggue pas des bookmarks ou des documents mais des collègues. Parmi la littérature qui vise à trouver des moyens de structuration des folksonomies, une majorité des projets est focalisée sur des dispositifs à facettes. A ce propos, Schraefel (2009) met en avant les avantages de la recherche à facettes. Selon lui, la recherche à facettes améliore les possibilités d'explorer un domaine via ses attributs, puisqu'il affirme : « the metadata is the message ». Il s'intéresse à la recherche exploratoire qu'il définit comme la démarche qui vise à harmoniser la conception des systèmes d'interaction Homme-Machine (IHM), avec des modèles issus des recherches sur la RI, pour supporter des types alternatifs de recherches et de construction du savoir. Selon lui, l'avantage des catégories est de donner un sens de la portée de l'espace d'information. Une opportunité pour la construction du savoir réside dans la mise en avant des attributs de ces espaces et dans la possibilité de mettre au premier plan les associations entre ces attributs. Ces cinq dernières années, les paradigmes de la recherche exploratoire se sont largement focalisés sur comment faciliter ces types vues associatives. Ces modèles sont devenus connus sous le nom de recherche à facettes.

> *« While requiring human intervention to construct domain categories, their advantage is—like looking at library shelves—to give a person a sense of the scope of an information space. An opportunity for knowledge building in foregrounding such attributes of a space is the ability to foreground associations among these attributes. Over the past five years, exploratory search paradigms have largely focused on how to facilitate this kind of associative view. These models have become known as facetted search. » (Schraefel, 2009, p.54)*

Les avantages des recherches à facettes exposés par Schraefel (2009) sont reportés ici :

- les possibilités de découvertes : l'exploration par facettes permet aux utilisateurs de faire de nouvelles connexions à propos d'un domaine, ou à propos des attributs d'un domaine ;
- elle permet de reconsidérer un domaine ;
- elle permet de construire du savoir en présentant les facettes, les attributs du domaine et donne une manière d'interpréter le domaine, à travers ses attributs ;
- elle représente aussi souvent la quantité d'items de chaque genre (par facette) ;
- elle est comparable au fait de passer en revue les livres sur une étagère : comme dans une stratégie de butinage qui fonctionne en partie par sérendipité.

1.4.7 La structuration des systèmes de tagging par facettes

Nous présentons ici, les différents espaces ou dispositifs de tagging qui proposent aux utilisateurs de structurer l'espace de tagging à l'aide de facettes.

1.4.7.1 TagMe !

Il s'agit d'un système de tagging pour les images couplé au site de tagging de photographies Flickr. La désambiguïsation des tags se fait par le moyen de liens créés automatiquement et manuellement vers des URI issues de DBPédia. Un système de zonage permet de tagguer non seulement le contenu principal mais aussi d'autres zones de l'image. Ce dispositif de tagging de zones identifiées de l'image est utilisé pour corréler les tags. De plus, il existe dans cette application, une fonctionnalité de catégorisation des tags par facettes qui ajoutent un contexte additionnel. Dans (Abel, Kawase, Krause, 2010), les créateurs cherchent à évaluer si l'exploitation du contexte additionnel (les facettes) améliore la recherche et les performances

du classement. Ils affirment que l'assignation de tags dans des catégories peut relier des ressources. En élargissant la recherche aux autres tags de la catégorie du tag sélectionné, ils obtiennent des résultats pertinents qui n'auraient pas été trouvés autrement. Cette méthode améliore le rappel et les informations contextuelles améliorent les résultats de recherches d'images.

1.4.7.2 Classification collaborative de collection évolutive

Wu, Zubair, Kurt (2007) présentent une expérimentation de construction de classification à facettes de manière itérative, construite d'abord par des experts, pour créer un brouillon de classification à partir d'une collecte de métadonnées dans la collection d'items et qui sera validé progressivement par les utilisateurs à qui l'on demande de classer un corpus restreint avec ce brouillon de classification. Elle est ensuite complémentée par un système d'apprentissage basé sur le panel d'items classés et le brouillon de classification, pour arriver à un classement automatique des items et un enrichissement automatique des catégories. Les utilisateurs font la maintenance de la classification. L'évaluation porte sur un corpus d'environ 500 items et peu de détails sont donnés sur la méthode d'élaboration de la classification à facettes par les experts. Leur approche est basée sur l'hypothèse qu'un groupe de personnes peut faire mieux qu'une petite équipe d'experts. D'après ces auteurs, la classification à facettes permet l'assignation de classifications multiples à un objet, supportant des perspectives multiples d'utilisateurs dans l'exploration, ce qui rend cette forme de classification appropriée aux approches collaboratives. La classification à facettes produit une recherche d'information basée sur une reconnaissance des catégories plutôt que sur l'élaboration de requêtes avec des mots-clés. Elle favorise, selon eux, l'exploration et la découverte.

1.4.7.3 FacetMap

Pour Wilson (2006), la classification à facettes implique l'orthogonalité, ce qui signifie que chaque facette est indépendante des autres. Il distingue le modèle strict de classification à facettes du modèle proposé par Ranganathan. Aussi, le modèle strict de classification à facettes interdit l'assignation de plusieurs valeurs d'une même facette à un item (ce que nous nommons sous le terme de « multivalué »). Le tagging permet l'assignation de tags sans préoccupation des autres tags assignés. Ainsi les tags portent l'orthogonalité à l'extrême,

chaque tag étant orthogonal aux autres tags. La contribution de Ranganathan, selon Wilson (2006) consiste à affirmer que l'on peut définir plusieurs classifications pour une ressource. La Colon Classification sépare en unités lexicales le sujet en exprimant chaque composant par un code. Les composants du code sont séparés par des virgules et les codes sont regroupés de manière, selon lui, intuitive. La classification à facettes stricte, au lieu de grouper les tags en groupes intuitifs, range les attributs orthogonaux des ressources elles-mêmes. Le mérite de Wilson (2006) est de tenter de formaliser les règles de divisions logiques à l'œuvre dans les principes de la classification à facettes. Ainsi, il postule que :

- Quand une caractéristique de division est exclusive d'une autre, celles-ci appartiennent à la même facette : par exemple, dans une facette type de gâteau, un cake ne peut pas être à la fois un cake et un cookie.
- Quand deux caractéristiques sont indépendantes elles appartiennent à des facettes différentes : par exemple, un cake peut être au chocolat ou à la cerise, la saveur est indépendante du type de gâteau.

Des valeurs sont orthogonales quand ces deux caractéristiques sont indépendantes, et alors elles appartiennent à des facettes différentes. Si entre deux facettes, les valeurs sont orthogonales, alors à l'intérieur des facettes, les valeurs ne sont pas orthogonales entre elles. Si ces deux propositions ne sont pas vérifiées, alors chaque tag est orthogonal aux autres et il s'agit d'un simple système de tagging et non d'une classification à facettes (stricte). Il propose donc la définition suivante de la notion de facette : un jeu de valeurs dans lequel l'assignation d'une valeur à une ressource limite l'assignation à cette ressource d'autres valeurs de ce jeu. Il s'agit du principe d'exclusivité mutuelle des valeurs issues d'une même facette, ou pour le dire autrement, de l'interdiction d'assigner deux valeurs d'une même facette à une ressource qui est un principe définitoire de la notion de facette. En termes méthodologiques, il préconise de penser la facette comme une question et les valeurs de facettes sont les réponses possibles à cette question.

FacetMap, groupe les tags et les organise en facettes intuitives grâce à un algorithme. Cet algorithme fonctionne en passant chaque ressource tagguée en revue et il range les tags en fonction du principe selon lequel chaque ressource ne peut pas avoir plus d'un tag provenant de la même facette. Un des défauts de ce dispositif réside dans le fait que l'algorithme ne fait

pas de traitement sémantique et n'est donc pas en mesure de donner un titre aux facettes. Il ne génère que des facettes numérotées.

1.4.7.4 Facetag

D'après Qintarelli, Rosati et Resmini (2007), Facetag est conçu pour améliorer l'architecture de l'information d'un système de tagging social de bookmarks. Il vise notamment pallier aux problèmes aux problèmes des nuages de tags, lesquels ne sont pas suffisants pour donner une information riche sémantiquement ni pour apporter une navigation multidimensionnelle.
Les problèmes des nuages de tags sont les suivants d'après ces auteurs :

- ils ne traitent pas les problèmes de variabilité du langage, ce qui réduit le quotient de recherchabilité et rend l'extensibilité du système faible ;
- les tags les plus fréquents étant les plus visibles, les tags dominants sont favorisés ;
- l'ordre généralement alphabétique de présentation des tags réduit les possibilités de navigation rapide, et rend difficile la construction d'une représentation mentale globale et cohérente des tags présents, la portée globale de la folksonomie est difficile à saisir ;
- un nuage sans hiérarchisation ne peut pas supporter visuellement des relations hiérarchiques.

Facetag propose des hiérarchies optionnelles de tags et donne la possibilité d'organiser les tags avec des relations de type Père/Fils. Les hiérarchies de tags peuvent être assignées dans des facettes prédéfinies qui servent ensuite à la recherche. Le système mixe les modes de recherche pour maximiser la pertinence des résultats de recherche, faciliter la navigation et la découverte.
Les facettes proposées dans Facetag sont inspirées du modèle à 13 facettes du CRG (évoqué en partie 2.3 du chapitre 2), mais le système n'en propose que 4 : Type de ressource, Thème, Personne, But/Objectif.
Le nombre de *bookmarks* indexés avec un tag apparaît à coté du tag et ce de manière dynamique pendant les recherches. Le filtrage avec les tags classés dans des facettes est complémenté d'une barre de recherche de tags avec une fonction d'auto-complétion. Après

une première sélection de termes, ce qui est présenté à l'utilisateur est un sous-ensemble des ressources disponibles dans la base, fondé sur les critères de la recherche en cours.

La définition de la notion de facette donnée par ces auteurs est la suivante : c'est une catégorie orthogonale de termes (tags) dans un système de métadonnées. Chaque facette a un nom et désigne des dimensions conceptuelles différentes ou des types de caractéristiques pertinentes pour la collection. Les facettes peuvent être hiérarchiques ou plates. Le multivalué est possible dans ce système.

1.4.7.5 Facette

Facette est un logiciel développé par Peter Lai en 2009 au MIT (Massachussetts Institute of Technology), qui vise améliorer le *bookmarking* social effectué sur Delicious à l'aide de facettes. Il combine le tagging à facettes et la création collaborative de facettes, pour classer les tags. Les tags sont donc organisés dans des facettes et l'Interface Homme-Machine (IHM) facilite la navigation dans les bookmarks en utilisant les tags. Elle assiste aussi la création de tags insérés dans des facettes. La création collaborative de facettes repose sur des processus de recommandation ou de suggestion de facettes aux utilisateurs. Le tagging collaboratif est assisté par des recommandations de tags qui sont suggérés sur la base des tags déjà créés. L'assistance à la création de facettes est basée sur les tags déjà insérés dans des facettes, il suggère des couples de forme « Facette + tag ».

Le système de classement à facettes des tags est vu comme un guide pour la navigation à travers l'addition d'un niveau supplémentaire d'organisation des tags. D'après Lai (2009), la fonction des tags est d'ajouter des métadonnées de contenu mais leur ambiguïté a pour conséquence que : « *without the presence of additional context the implied meaning of a tag (…) may be difficult to determine* » (Lai, 2009, p.15). Ainsi les facettes apportent un contexte additionnel pour l'interprétation du sens des tags tout en apportant un niveau de structuration supplémentaire. Elles contiennent des mots-clés qui décrivent des propriétés des objets.

Les avantages énoncés par Lai (2009) de la classification à facettes sont les suivants :

- elle excelle à organiser des informations qui peuvent être naturellement décrites par un jeu de catégories orthogonales ;
- la navigation à travers les facettes améliore les capacités des utilisateurs à explorer des collections d'items non familières ;

- ainsi quand un utilisateur parcoure un jeu d'items non familiers, il effectue des « *patterns* de *browsing* » aussi appelés « *orienteering* ». Ce terme est défini par Lai (2009) comme le comportement de l'utilisateur qui utilise des indices de navigation (cherche des repères) pour explorer itérativement la structure des informations. Un utilisateur qui explore de nouvelles informations manque souvent d'un but et de mots-clés pour la recherche, la qualité de l'exploration dépend des indices de navigation.
- elle permet des filtres sur des dimensions multiples pour la recherche d'information, ce qui accroit le sens du contrôle des utilisateurs et fait diminuer la désorientation, elle réduit donc les difficultés d'exploration puisque chaque valeur de facette offre une nouvelle option de navigation

Nous ajoutons aux avantages cités ici, auxquels nous souscrivons, la possibilité qu'offre la classification à facettes d'obtenir une vision globale de la structuration du champ couvert par la classification. Par le scan visuel rapide des catégories et valeurs présentes dans la classification, l'utilisateur peut situer la portée conceptuelle de la classification.

Concernant la méthode d'élaboration de classification à facettes, l'auteur mentionne que cette opération est difficile et qu'il existe deux méthodes, l'une automatique et l'autre manuelle. Les méthodes automatiques sont fondées sur des statistiques pour l'assignation de tags dans des facettes comme dans le projet FacetMap cité précédemment qui fonctionne avec un algorithme basé sur l'exclusivité des valeurs à l'intérieur des facettes. Ces méthodes automatiques ne traitent pas les aspects sémantiques ni les problèmes d'ambiguïté des tags. Ils produisent des classifications perfectibles par des interventions humaines.

Les méthodes manuelles, comme celle utilisée dans le projet Facetag cité ci-dessus, qui est basé sur le modèle à 13 facettes du CRG, produisent une flexibilité que les systèmes automatisés ou semi automatisés ne peuvent pas fournir. Les systèmes semi-automatisés autorisent une saisie manuelle restreinte qui réduit la flexibilité du système et rend le processus d'assignation des tags plus difficile. Ainsi certains tags peuvent entrer dans des catégories de manière forcée ou artificielle. Concernant les méthodes manuelles, Lai (2009) qualifie les classifications ainsi produites de « *free faceted classification* » qui laissent à l'utilisateur le choix de créer des tags et des facettes au fur et à mesure du processus de tagging des ressources et aussi de laisser des tags non classés dans des facettes.

1.4.8 Méthodes de classification à facettes pour les outils de tagging sur le Web

Dans un article daté de 2010, Louise Spiteri effectue une revue de la littérature sur les méthodes existantes pour l'usage des facettes pour le tagging social. Elle analyse ces méthodes en cherchant les réponses apportées par les auteurs des 17 articles traités aux questions suivantes, qui selon elle, doivent être éclaircies dans le cadre d'une méthode d'élaboration de classification à facettes pour le tagging social :

1. Etant donné le large champ de sujets couvert par les applications de tagging, comment choisir les facettes qui s'appliquent à tous les sujets ou domaines ?
2. Quel est le nombre optimal de facettes ?
   Nous pensons que cette question est liée au degré de précision des termes qui servent de valeurs de facettes et ainsi elle est liée au problème du « *basic level* » (voir partie 3.5.6.1 chapitre 3) ;
3. Qui doit choisir les facettes ? les utilisateurs du site, l'éditeur ou encore des chercheurs ?
4. Comment s'assurer que les facettes choisies reflètent les besoins des utilisateurs ?
5. Enfin, comment maintenir et imposer un contrôle de la qualité des facettes ?

Globalement, Spiteri (2010) mentionne que la plupart des auteurs des articles traités dans cette revue de littérature omettent d'expliquer clairement comment constituer des ensembles de tags en facettes. Elle note aussi le manque de définition claire de la notion de facette. Elle insiste sur le manque d'attention porté à l'exclusivité mutuelle, excepté chez Wilson (2006). Aussi, elle affirme :

> *« If facets are to serve as efficient tools by which to organize tags and provide them with a meaningful context, it is important that each facet be defined clearly, represent only one attribute, or characteristic of division, and not overlap with any other facet. » (Spiteri, 2010, p.103)*

Plus précisément, commentant les travaux de Hearst (2006) sur un navigateur à facettes, « Flamenco » et auxquels nous reviendrons dans la suite de cette partie, elle note que l'auteur ne donne pas d'informations sur les personnes habilitées à construire la classification à facettes, ni sur comment elle est élaborée, ni sur sa maintenance. Par la suite, citant les travaux de Lichtblau, Trice et Wartik (2006), qui soutiennent que la structuration des tags en

facettes est plus efficace pour des domaines spécifiques plutôt que pour le Web en entier, elle souligne l'intérêt de leur approche de la construction de classification à facettes, que nous pourrions rapprocher des travaux de Wilson (2006) cités précédemment. En effet, ils proposent de dériver les facettes d'un jeu de 7 questions, ce qui rappelle la recommandation de Wilson (2006) de considérer les facettes comme des questions et les valeurs de facettes comme des réponses à ces questions. Ces 7 « W » questions sont reportées ici, mais il est à noter qu'elles sont conçues pour classer des services :

1. Qui utilise le service ?
2. Que fait le service ?
3. Sur quoi agit le service ?
4. Pour qui est-il conçu ?
5. Où est-il utilisé ?
6. Quand est-il utilisé ?
7. Pourquoi est-il utilisé ?

Citant ensuite les travaux de Peter & Stock (2007), elle souligne qu'ils affirment que les utilisateurs d'un service de tagging sur le Web agissent dans des contextes différents, pour différentes tâches et selon des motivations variables. En effet, selon ces auteurs, les tags peuvent décrire le contenu d'une ressource, mais également des caractéristiques plus formelles (Auteur, Origine, Format), ils peuvent aussi être utilisés à des fins de gestion personnelle de l'information (PIM) ou pour exprimer un jugement. Ils préconisent de créer des catégories selon les fonctions des tags, à savoir, la description du contenu, de caractéristiques formelles, d'aspects organisationnels, ou affectifs. Nous pensons que ces préconisations sont justes car elles sont conformes aux observations de Kipp (2011). Spiteri (2010) note cependant que Peter & Stock (2007) n'expliquent pas comment créer les facettes ni qui doit les créer.

A propos des travaux de Quintarelli, Rosati et Resmini (2007), cités précédemment, sur Facetag, Spiteri (2010) note que ces auteurs qui basent leur classification à facettes sur le modèle à 13 facettes du CRG, n'expliquent pas comment ils la dérivent de ce modèle.

Dans la plupart des articles mentionnés, il n'y a pas d'indication sur un nombre de facettes optimal, ni sur la méthode d'évaluation de la pertinence et de la validité des facettes. Elle propose un cadre pour l'évaluation des classifications à facettes pour le tagging social qui est composé des éléments suivants à évaluer :

- le choix des facettes ;
- dans quelle mesure elles reflètent les besoins des utilisateurs ;

- l'aisance d'utilisation ;
- le nombre de facettes ;
- dans quelle mesure elles sont utilisées ;
- si elles sont utilisées correctement : par exemple l'assignation correcte correspond au classement de noms de lieux dans une facette « Lieu » ;
- comment les facettes sont gérées, maintenues.

1.4.9 Les autres projets proches de Miipa-Doc

Nous présentons ici les autres projets que nous avons pu relever dans la littérature, qui peuvent être rapprochés du projet dans le cadre duquel la recherche que nous présentons a pu être menée et qui sera présenté en dernière partie de ce chapitre.

1.4.9.1 Flamenco

Dans la même direction que les démarches de structuration par facettes d'outils de gestion de l'information (comme les systèmes de tagging), Hearst (2006) présente un navigateur à facettes : « Flamenco ». Il est dédié au classement des résultats de recherche. Il se distingue des méthodes de *clustering* qui d'après Hearst (2006) consistent à grouper des items en fonction d'une mesure de similarité entre ces items. Pour les documents, les techniques de clustering fonctionnent d'après des associations et points communs entre des caractéristiques, qui sont généralement des mots et des phrases. Ce processus de clustering est automatisable, mais les algorithmes sont imparfaits, ils font des erreurs de classement, leur prédictibilité est limitée et ils ne peuvent pas nommer les clusters.

Flamenco exploite un système de catégories à facettes hiérarchiques. D'après Hearst (2006), tandis qu'un système de catégories est un jeu de labels significatifs organisés de manière à refléter les concepts pertinents du domaine et est généralement créé manuellement (même si l'assignation des items aux catégories peut être automatisée), un système à facette, au lieu de créer une large hiérarchie de catégories, est construit comme un jeu de hiérarchies de catégories chacune correspondant à une facette pertinente pour la collection.

Flamenco est donc annoncé comme un navigateur qui propose une interface de navigation intuitive pour l'exploration et la découverte d'informations. Ses fonctionnalités principales

sont de guider la navigation (il montre simultanément le stade précédent de la navigation et les stades suivants possibles) et de mixer les modes de recherches (il intègre le recherche en texte libre dans la structure des catégories). Cette approche réduit le travail mental en favorisant la reconnaissance des catégories plutôt que l'appel à la mémoire, en suggérant des alternatives logiques (et non nécessairement prévues) à chaque stade de la recherche. Il évite aussi les résultats vides.

Les concepteurs de Flamenco ont également mené plusieurs études d'utilisabilité des IHM pour les métadonnées à facettes. Ainsi, English, Hearst, Sinha, Swearingen, et Yee (2002) tentent d'évaluer la démarche qui consiste à utiliser des métadonnées à facettes pour classer les résultats de recherches conduites dans des sites Web. Les métadonnées à facettes sont des descripteurs orthogonaux à l'intérieur d'un système de métadonnées. Ils mesurent la propension à utiliser le système, l'aisance d'usage, l'utilité pour le travail de conception, d'exploration d'une collection, pour trouver un item particulier, pour voir les relations entre les items. Ils concluent leur étude réalisée sur des recherches d'images uniquement, que des métadonnées générées dynamiquement, à facettes, avec des hyperliens, constituent une une alternative à la recherche dans des sites web basée sur des requêtes de mots clés.

Dans une autre étude de 2002, les mêmes auteurs (English, Hearst, Sinha, Swearingen et Yee, 2002) comparent une interface de recherche simple par mots clés sans classement des résultats, avec une interface de recherche avancée qui présente les métadonnées avec des cases à cocher et des menus déroulants sans classement des résultats, et aussi avec une interface de browsing qui permet à l'utilisateur de naviguer à travers la collection, et de construire implicitement des requêtes en « ET » par la sélection de valeurs dans les facettes.

L'approche utilisée dans la dernière interface citée ci-dessus, est construite sur des métadonnées à facettes hiérarchiques comme base pour la structuration de la navigation, indiquant les choix suivants (à chaque stade), proposant des vues alternatives, tout en permettant le raffinement et l'expansion de la requête et en préservant un représentation cohérente de la structure de la collection.

Ils cherchent à évaluer si les utilisateurs apprécient et comprennent une organisation flexible des métadonnées à partir de différentes hiérarchies et si les hiérarchies à facettes sont préférables à des hiérarchies uniques. L'étude révèle que les utilisateurs préfèrent l'interface à facettes. Elle leur donne un sens accru du contrôle. Le design itératif de l'interface à facettes, réalisé d'après les retours apportés par les utilisateurs, produit une interface finale évaluée par les utilisateurs comme étant utile et compréhensible. Les hiérarchies à facettes permettent aux

utilisateurs de modifier de manière flexible leurs requêtes plutôt qu'ils ne cherchent à trouver les mots clés appropriés. Le système est reconnu comme présentant les avantages suivants :

- il cherche la cohérence ;
- offre des feedbacks informatifs ;
- offre un traitement simple des erreurs ;
- permet d'inverser des actions facilement ;
- supporte le contrôle de l'utilisateur ;
- réduit la charge de mémorisation à court terme.

Enfin, une troisième étude du même type, menée en 2003, par Yee, Swearingen, Li et Hearst, a pour objet d'évaluer le design et l'utilisabilité de l'interface de métadonnées à facettes à partir de test basés sur des collections d'images. Les facettes de métadonnées, présentent la structure de la collection. Les résultats qu'ils annoncent sont plutôt positifs et sont énoncés ci-après :

- Meilleur satisfaction de la recherche et meilleur succès de la recherche ;
- Meilleure confiance dans les résultats ;
- Haut niveau de rappel ;
- Peu de requêtes vides (sans résultats) ;
- Plus utile et flexible que les bases d'images traditionnelles ;
- Les utilisateurs se sentent plus familiers du contenu de la collection ;
- Possibilités de requêtes multi-facettes ;
- Enclenchement de la recherche plus facile (donne une idée de quoi chercher) ;
- Donne le chemin à suivre (quel est le stade suivant) ;
- Sur le niveau de détails des facettes: le leur est jugé utile, clair, guidant ;

L'interface de métadonnées à facettes, par rapport aux bases de données traditionnelles d'images où l'on cherche par requêtes de mots clés, est préférée fortement par la plupart des participants à l'étude. Les limites de cette étude restent le fait qu'elle est effectuée sur un seul type de ressources (les images), les types de documents, selon nous, varient plus que les

images en termes de description formelle, et aussi, le fait qu'aucune information n'est présentée sur la détermination du niveau de détail des facettes.

1.4.9.2. Projets visant le classement de l'expertise

D'autres projets, un peu similaires au projet Miipa-Doc, que nous allons présenter dans la suite de cette partie, sont évoqués ici. Ils se distribuent autour des thématiques du repérage de l'expertise dans le champ de l'entreprise et de la construction de communautés à l'aide d'outils de tagging. Ainsi on relève les travaux de Bogers *et al.* (2006) qui étudient le classement automatique *Versus* collaboratif de l'expertise. En l'occurrence, ils classent automatiquement l'expertise des personnes travaillant dans l'entreprise, par extraction de termes issus des documents et des courriers électroniques des utilisateurs. Ils définissent l'expertise comme la possession de talents créatifs et d'analyse supérieurs et la capacité à traiter et intégrer de nouvelles informations plus vite que les non-experts. Souvent les experts sont aussi reconnus comme tels dans une communauté parce qu'ils remplissent certains critères comme la possession de diplômes spécifiques. John et Seligman (2006), proposent un système de reconnaissance automatique d'experts, en temps réel, basé sur les tags associés à des bookmarks dans un outil de tagging et ils élaborent un scénario qui intègre ces données, ainsi que d'autres éléments plus contextuels comme la disponibilité des experts au moment de la demande ou leur localisation. Il s'agit d'une démarche qui intègre toutes les données disponibles sur les individus dans le Système d'Information de l'entreprise.

Dans le projet « Fringe Contacts » de Farrell et Lau (2006) d'IBM, les individus travaillant dans une entreprise se tagguent entre eux, et eux-mêmes, pour décrire leurs compétences, rôles, pour la localisation de l'expertise dans l'entreprise et également pour la construction de communautés (Farrell, Lau, Nusser, 2007). Serdyukov *et al.* (2011) travaillent sur le tagging automatique de l'expertise des individus d'une entreprise, à des fins d' «expertise profiling ».

Quant à Wu *et al.* (2006) ils tentent de récolter le savoir socialement construit à partir de folksonomies.

D'autres auteurs, dans des démarches plus classiques, conçoivent un système de bookmarking social pour les entreprises (Millen *et al.*, 2006), tandis que Heyman et Garcia-Molina (2006) élaborent un algorithme pour construire automatiquement des hiérarchies de tags à partir des données des systèmes de tagging. Ils précisent les caractéristiques saillantes des vocabulaires de tagging qu'ils prennent en considération: l'agrément entre les utilisateurs sur les tags appropriés, la densité qui correspond à la fréquence à laquelle les utilisateurs taguent des

ressources et le recouvrement, c'est- à-dire la fréquence à laquelle les utilisateurs taguent les mêmes ressources. Enfin, nous avons relevé le projet de Begelman *et al.* (2006) qui cherche à appliquer des techniques de clustering aux espaces de tags pour une meilleure exploration et utilisation de ces derniers.

1.5 Le projet Miipa-Doc

Elaboré en partenariat avec Cogniva (l'entreprise qui a proposé la méthodologie GenISIS de création de classification à facettes), l'Université de Technologie de Troyes (UTT), EDF R&D, et la Jeune Entreprise Universitaire (JEU) Cogniva Europe, en réponse à un appel à projet ANR Contint, en 2008, le Projet ANR Miipa-Doc (Méthodes et Services intégrés Institutionnels et Participatifs pour la Classification à Facettes des Contenus Documentaires Complexes) visait à concevoir de nouvelles méthodes d'ingénierie socio-cognitives ascendantes pour l'organisation des contenus documentaires complexes et à développer l'architecture logicielle correspondante.

Notre travail prend effet dans le cadre de ce projet de recherche, notre rôle a été focalisé sur la tentative présentée ici d'étudier les méthodologies existantes de création de classification à facettes, d'en dégager les éléments congruents et ceux moins adaptés aux objectifs méthodologiques du projet, afin de proposer des spécifications, recommandations méthodologiques et d'usage pour le projet d'architecture logicielle.

L'association cohérente des acteurs impliqués dans ce projet de recherche s'est heurtée à quelques difficultés d'ordre pragmatique, qui ont quelque peu fait évoluer les objectifs du projet, son déroulement et ses résultats. Ainsi, initialement, les acteurs impliqués dans le projet avaient pour objectif de complémenter l'approche institutionnelle de la gestion documentaire proposée par l'entreprise Cogniva (présentée dans le chapitre 2), qui est elle-même centrée sur le champ du record management pour les documents des organisations.

Le projet Miipa-Doc était donc conçu en complémentarité de l'infrastructure sémantique et du logiciel ISIS, et enfin de la méthodologie GenISIS, avec pour objectif de développer les modules et extensions méthodologiques de services participatifs, coopératifs et folksonomiques, inspirés des recherches menées à l'UTT par le Laboratoire de recherche Tech-CICO, sur le Web Socio-sémantique et les approches par points de vue pour la documentarisation des ressources documentaires. Le projet visait aussi à développer des interfaces pour la recherche et la consultation des documents afin de faciliter l'offre d'une

conception plus subjective des documents et favorisant éventuellement l'émergence d'une vision orientée métier des ressources documentaires.

Alors que le projet se centrait sur une forme de complémentarité effective entre les approches institutionnelles et les approches ascendantes, issues des recherche citées précédemment, et itératives impliquant les acteurs dans la conception et la maintenance des Systèmes d'Organisation des Connaissances, (SOC), qui sont essentiels au déploiement des pratiques du record management, divers éléments ont orienté le projet de manière un peu différente.

1.5.1 Rôles des partenaires

Le partenaire Cogniva apportait l'infrastructure logicielle formant la base du projet, tandis que Cogniva Europe devait développer les modules participatifs (folksonomiques) et coopératifs (multi-experts), ainsi que la méthode d'alimentation terminologique des classifications à facettes et la méthode de consultation. Cette Jeune Entreprise Universitaire avait aussi pour objectif de développer la méthode d'articulation entre les parties institutionnelles et expertes des classifications, et d'apporter des spécifications concernant la gestion des rôles et des compétences. Les objectifs scientifiques poursuivis par les partenaires EDF R&D du projet étaient les suivants dans l'agencement initial :

- faciliter l'accès à des contenus de natures différentes (textes, images, plans etc.)
- combiner différentes techniques d'indexation et permettre un enrichissement progressif des index (afin d'améliorer en continu les performances du système de recherche)
- améliorer la pertinence des réponses faites par le système de recherche d'information par une prise en compte du profil de l'utilisateur (notion de point de vue)
- permettre à l'utilisateur de comprendre et de s'approprier la connaissance contenue dans les documents.

En effet, EDF étant l'énergéticien (fournisseur d'énergies sous différentes formes) majeur en France et l'exploitant des centrales de production nucléaire, le partenaire EDF R&D (EDF Recherche et Développement), apporte un appui technique et fonctionnel à la réflexion globale menée pour la rénovation du Système d'Information concernant la production

nucléaire, et en particulier sur les systèmes d'accès aux documents pertinents en contexte opérationnel, notamment pendant les opérations de maintenance des centrales, sessions pendant lesquelles de nombreuses personnes de métiers différents sont mobilisées, et pendant lesquelles il peut être produit jusqu'à plus de trois mille documents, que l'opérateur énergétique cherche à capitaliser. Différents terrains d'expérimentations étaient prévus, notamment des terrains opérationnels sur site (dans les centrales nucléaires).
Enfin, l'UTT apportait un soutien scientifique et technique au développement et à l'intégration technique des modules folksonomiques basés sur des méthodes d'ingénierie cognitives ascendantes ainsi que pour les expérimentations à mener sur les différents terrains envisagés chez EDF.

1.5.2 Les questions de recherche

Les questions de recherche pour ce qui a concerné notre contribution dans le projet se concentraient autour des problématiques suivantes :

- Aspects normatif Versus spontané de l'indexation : dimensions et dispositifs de conciliation des conceptions liées aux problématiques des SOC formels institués (thésaurus, ontologies formelles) opposées en général aux conceptions folksonomiques et à celles issues du Web socio-sémantique concernant les vocabulaires d'indexation ; et quelles formes d'émergence de consensus terminologique ;
- Différences entre le travail prescrit et le travail réel : Aspects imposés, négociés, individuels de la charge de travail liée à la gestion de l'information ;
- Travail Individuel ou collectif, avec quelles définitions des collectifs considérés, et dimensions institutionnelles, organisationnelles, individuelles de l'indexation: quelle articulation possible ?

1.5.3 Evolution du projet

Suite à une conjonction de facteurs, distribués dans le temps, que nous évoquons rapidement ici, le projet s'est finalement orienté sur le développement d'un outil autonome de gestion de

documents basé sur la classification à facettes et le tagging de documents. Le prototype de cet outil est présenté en chapitre 3 (en partie 3.5.9).

Tout d'abord, la distance géographique importante entre les deux partenaires Cogniva et Cogniva Europe, a rendu les échanges plus difficiles, sur le plan technique, notamment, pour les développeurs et les personnes impliquées de l'UTT pour la coopération scientifique, mais aussi sur le plan de la coopération concernant la méthodologie d'articulation des dimensions institutionnelles et individuelle ou collectives (au sens restreint) de l'indexation. De plus, les retards dans le développement des différents modules, qui sont des événements courants dans le développement logiciel, ont rendu difficiles à tenir les échéances liées au projet. Enfin, l'accès aux terrains pour les expérimentations a été lui aussi difficile pour l'autre doctorante, Noémie Musnik, en charge de cette partie du projet. Elle écrit d'ailleurs : « En raison des difficultés d'accès aux CNPE, inhérentes à ce type de terrain sensible et hautement sécurisé, de la disponibilité réduite des acteurs que nous envisagions de rencontrer ainsi que du calendrier des projets portés par la R&D auxquels nous étions amenés à contribuer, nous nous sommes orientés vers un terrain plus proche géographiquement parlant : une équipe d'ingénieurs-chercheurs d'EDF-R&D, le groupe P1B du Département STEP (Simulation et Traitement de l'information pour l'Exploitation des systèmes de Production). » (Musnik, 2012)

Ainsi, constatant que plusieurs éléments venaient à l'encontre du déroulement et du cadre initiaux du projet, ainsi que certaines divergences scientifiques notamment sur la conception de la notion de métier, et sur l'approche des systèmes d'information, le projet a connu une réorientation féconde et non conflictuelle pour les partenaires. Le projet a donc connu plusieurs stades d'évolution, avant de se concentrer sur le développement et la recherche d'un nouveau modèle de classification à facettes pour la gestion quotidienne des documents, basée sur la tagging des documents par leurs utilisateurs. Suite à ces réorientations, les problématiques se sont concentrées sur des questions autour de la méthode de conciliation entre des approches descendantes (Top-Down) et des approches ascendantes (Bottom-Up) basées sur les contributions des utilisateurs et sur la nature des collectifs qui ont des intérêts à partager des ressources. Nous avons orienté les recherches sur ce mode collectif de gestion informationnelle vers la notion de métier.

1.6 Conclusion du Chapitre 1

Dans le cadre du projet Miipa-Doc, un prototype Hypertagging de logiciel de classifications à facettes pour l'organisation et la gestion des connaissances basé sur le tagging des utilisateurs a été développé, qui est présenté en Chapitre 3. Ce projet nous a ainsi conduits à étudier en détails la littérature sur les classifications à facettes et les systèmes basés sur le tagging par les utilisateurs. Par ailleurs, il a nécessité une étude approfondie des problématiques liées à l'indexation, à la documentarisation, aux classifications en général, et au concept de sujet, présentée dans ce chapitre 1.

Dans le chapitre suivant, nous proposons une présentation et une analyse des méthodologies existantes de classifications à facettes. Ces méthodologies diffèrent selon les environnements dans lesquels les classifications visent à être utilisées : bibliothèques, fonds documentaires spécifiques, organisations de type administratif ou sites Web.

Chapitre 2 : Méthodologies d’élaboration de classifications à facettes

2 Chapitre 2 : Méthodologies d'élaboration de classifications à a facettes

2.1 Introduction

Plusieurs approches de l'élaboration de classifications à facettes existent et ont été publiées. Elles diffèrent en fonction des environnements dans lesquelles elles sont déployées. Dans ce chapitre, nous présentons la première méthodologie, celle de Ranganathan, associée au monde des bibliothèques, puis celle du CRG (Classification Research Group), associée à des domaines spécifiques, qui aboutit à des classifications spécifiques, (pour la distinction entre les classification génériques et spécifiques, voir le chapitre 1, (partie 1.3.1). Ensuite nous nous intéressons au modèle ISIS proposé pour Record Management dans les organisations. Enfin nous décrivons la méthodologie proposée par Denton (2003), inspirée des approches de Spiteri (1998) et qui est dédiée à l'espace des sites Web.

2.2 La méthodologie de Ranganathan

Tout d'abord, il faut mentionner que cette méthodologie est largement influencée par les problématiques liées au système de notation. Elle a été développée pour une bibliothèque, et est donc déterminée par les contraintes de la culture papier. Nous considérons que celle-ci constitue une des seules tentatives, avec la classification de Bliss (2ème édition), (Mills, 2004), de réaliser un modèle qui cherche à classer l'ensemble des champs du savoir contenu dans une bibliothèque. Elle se confronte aux problèmes de l'universalité des classifications puisqu'elle doit couvrir tous les champs possibles d'une bibliothèque. Ces problématiques pourraient, selon nous, être rapprochées de celles afférentes à l'espace du Web, qui couvre une variété non mesurable de domaines de savoirs et, de plus, une variété de formats et de types d'informations dont la fiabilité et la pérennité n'est pas assurée. Mais contrairement aux contraintes induites par l'espace du Web, cette classification ne classe que des sujets. Le corpus classé est relativement homogène. Cette méthodologie correspond à celle qui a présidé à l'élaboration de la Classification de Colon (CC).

2.2.1 Préalable sur les différents types de sujets et leur formation

Ranganathan (1967) distingue des types de sujets en fonction de leur construction.

2.2.1.1 Les types de sujets

Les différents types de sujets qu'il évoque sont les suivants :

Les isolats : Un isolat est une idée ou un complexe d'idées convenant à former un composant d'un sujet, mais ne pouvant pas lui-même former un sujet. Nous pouvons illustrer cette notion par quelques exemples comme enfant, l'or, la notion de structure ou l'Inde. Dans ces exemples, on constate que ces catégories peuvent concerner un état du développement humain, une matière ou un minerai, une notion générique, ou encore une zone géographique. Cette catégorie de composés de sujets peut inclure des isolats communs ou spécifiques. (Maniez, 1999) note qu'ils correspondent à des concepts.

Les sujets basiques : ce sont des sujets non composés, ne contenant pas d'isolat. Quelques exemples fournis par Ranganathan (1967) nous ferons constater que, bien souvent, ils correspondent aux classes génériques de premier niveau des classifications bibliographiques, qui reflètent les disciplines : mathématiques, géométrie, chimie, ingénierie, biochimie, littérature, linguistique, religion, psychologie, éducation, sciences politiques…

Les sujets composés : il s'agit d'un sujet composé de la combinaison d'un sujet basique avec un ou plusieurs isolats. Par exemple, la chimie des halogènes, la littérature dramatique, l'enseignement aux adultes… que l'on peut décomposer comme suit : « chimie » est le sujet basique et « halogène » est un composé chimique - l'isolat -, la « littérature » est le sujet et le terme « dramatique » est une caractéristique générique des œuvres artistiques - l'isolat -, « l'enseignement » est le sujet basique et « adultes » est une catégorie générique pour classer le développement humain.

Les sujets complexes : ce sont des sujets basés sur une relation entre deux sujets, par exemple : les mathématiques pour la physique, les différences entre la physique et la chimie… Ranganathan fait aussi une distinction entre la portée générale ou spécifique des sujets et la longueur de leur développement dans les documents. Il introduit ainsi les notions d'extension et d'intention des sujets, l'extension correspond au degré de développement (la grandeur du

champ couvert par le sujet) et l'intention au degré de généralité du sujet (plus l'intention est grande plus le sujet est spécifique).

Les macro-sujets : sont des sujets de grande extension et de petite intension, c'est-à-dire, que le développement est long mais le sujet très spécifique. Il s'agit soit d'un sujet basique, soit d'un sujet composé, souvent son extension prend la forme d'un livre.

Les micro-sujets : sont des sujets de petite extension (la forme est courte) mais de grande intentio. Le sujet est générique ; il peut s'agir de pamphlet ou d'un chapitre de livre par exemple.

2.2.1.2 Les méthodes de formation des sujets

Les sujets peuvent être construits, structurés ou développés selon cinq processus.
La dissection : on découpe les entités en parties coordonnées (de même niveau). Par exemple pour les continents, cela donne l'énumération suivante : Asie, Afrique, Europe, etc.

La stratification : on construit le sujet par superposition de facettes, (comme un sandwich, d'après Ranganathan), par exemple : la culture du riz, à Madras, dans les années 1960.

La dénudation : on fait décroître l'extension du sujet (le champ couvert) et s'accroître l'intention du sujet (la précision). Cela revient à aller du plus générique au plus spécifique, comme dans l'exemple suivant : Philosophie, logique, logique déductive

L'assemblage lâche : cela consiste à assembler plusieurs sujets basiques, ou isolats, et sert à exprimer une ou plusieurs relations entre les composants de l'ensemble. Chaque composant correspond à une « phase ». Les relations entre les phases sont de cinq types : relation générale, relation de biais, relation de comparaison, relation de différence, relation d'influence. Les exemples suivants illustrent notre propos : influence de la géographie sur l'histoire, différence entre cellule animale et cellule végétale.

La super-imposition : on utilise deux ou plusieurs isolats provenant de la même classe. Par exemple : agriculture à Hong Kong et Macao (qui sont deux régions administratives spéciales de la République Populaire de Chine).

2.2.2 Les cinq lois des sciences des bibliothèques

Les cinq lois des sciences de la bibliothèque qui régissent les interactions entre les autres principes et postulats présentés, en cas de conflit, sont les suivantes :

Loi n°1 : *Books are for use*

Elle implique que la recherche par auteur est utile, mais qu'il faut aussi pouvoir rechercher par sujet. « Books are for use as embodied idea, not as physical commodities, not even as production of particular individuals […] » (Ranganathan, 1967, p. 119)

Dans cette partie, Ranganathan présente les fonctions des bibliothèques, qui consistent à

- ranger les livres dans un ordre aidant l'utilisateur à les trouver (*helpful sequence*) ou à mécaniser un tel rangement ;
- aider à trouver la place la plus utile pour un nouveau livre parmi ceux qui sont déjà présents ;
- aider à trouver la place la plus appropriée aux nouveaux sujets qui émergent parmi ceux déjà reconnus.

Loi n°2 : Every reader his book

Cette loi indique que le rangement des livres est utile au lecteur qui recherche de l'information sur un sujet si :

- les livres sur un même sujet sont rangés ensemble ;
- s'ils sont classés par langues ;
- s'ils sont classés chronologiquement à l'intérieur du classement par langues.

Ranganathan mentionne aussi que les utilisateurs ayant, en général, une idée vague de ce qu'ils cherchent, les livres doivent donc être rangés par sujets, selon la précision du sujet et le degré de filiation par rapport à un sujet très général. Le rangement dans les étagères doit, non seulement rendre lisible, mais aussi présenter le champ entier d'intérêt d'un utilisateur, ce qui permet qu'il découvre, rende conscient et fasse émerger le vrai sujet de sa recherche.

Loi n°3 : *Every book his reader*

Cette loi suppose qu'il est possible de donner un lecteur approprié à chaque livre, si les sujets sont rangés entre eux selon le degré de leur relation mutuelle. Elle indique aussi que pour les

livres ayant plusieurs sujets, chaque sujet doit être classé et catalogué. Une notice qui fournit à l'utilisateur l'exposition d'un sujet subsidiaire d'un livre est appelée « Entrée-Sujet analytique ».

Loi n°4 : *Save the time of the reader*
Cette loi traite du besoin d'un ordre utile pour les sujets spécifiques des livres. Ranganathan considère que le rangement alphabétique des sujets conduit à leur éparpillement. Il faut donc, selon lui, les classer selon une logique d'enchainement des rubriques qu'il décrit plus loin dans l'ouvrage.

Loi n°5 : *Library is a growing organism*
Le rangement doit faciliter le service aux utilisateurs, à un moment précis, de manière exhaustive et rapide, peu importe la taille de la bibliothèque, son taux d'expansion, ou d'émergence de nouveaux sujets.

2.2.3 Lois sur la structure syntaxique d'une classification

Les lois qui régissent la structure syntaxique de la classification à facette libre dans la méthode de Ranganathan, et qui nous paraissent déterminantes d'un point de vue global, sont les suivantes :

Loi d'Interprétation : chaque élément d'une séquence de facettes a une place déterminée : on ne peut pas la remplacer impunément.

Loi d'Impartialité : l'ordre dans une séquence de facettes ne doit pas être déterminé aléatoirement, ni arbitrairement, mais d'après des principes applicables à tous les sujets. Cela correspond à des préoccupations notationnelles.

Loi de Parcimonie : il faut penser la classification à facettes en termes d'économie de place, de force, d'énergie, de matériel, d'argent, et de temps.

Loi de Variation locale : Ranganathan fait une différence entre les usages locaux et généraux des bibliothèques. Il considère que si une bibliothèque parmi celles accessibles dans son voisinage choisit un schéma de classification alternatif, cela peut devenir un problème pour

les usagers (usage local). L'usage général correspond pour lui à l'aspect international des classifications et aux problématiques aujourd'hui bien reconnues d'interopérabilité.

Il précise que, pour des intérêts spéciaux (vus comme étant de nature nationale), ou pour des collections spécifiques, ou des usages spécialisés et répétés de certaines parties du fond d'une bibliothèque, on peut adapter, par exemple, la facette « Lieu » à un collection locale. On peut créer des catégories favorites comme pour la langue majoritairement utilisée dans un fond, ou des classes favorites de sujets, en utilisant des dispositifs de notation artificiels comme une classe ø afin de les positionner en premier dans une liste, et ainsi de raccourcir l'indice.

Ranganathan distingue plusieurs plans dans l'élaboration des classifications : le plan des idées, le plan terminologique, et le plan de la notation. Nous les présentons ci-après dans cet ordre de succession.

2.2.4 Les Canons pour le plan des idées

L'élaboration d'un schéma de classification, sur le plan des idées, implique un schéma de classes qui correspond à cinq concepts inhérents. A chacun de ces concepts sur le plan des idées, correspondent des canons. Ces cinq concept qui sont présentés dans cette partie, sont les suivants :

1. Les caractéristiques
2. La succession des caractéristiques
3. Des ensembles de classes
4. Des chaines de classes
5. Des séquences de classes (filiatory sequence)

Point terminologique

Une division est le processus de tri des entités d'un ensemble en sous catégories sur la base de caractéristiques, ainsi que le résultat de cette division des entités.

Une caractéristique de division est une caractéristique utilisée comme base pour la division des entités d'un ensemble.

Une caractéristique est un attribut ou un complexe d'attributs, en référence auquel la similarité ou la différence des entités peut être déterminée et en référence auquel au moins deux des entités sont différentes.

Un attribut est une propriété ou une mesure qualitative ou quantitative d'une entité.

Une entité est un « existant » concret ou conceptuel, par exemple, une chose ou une idée.

2.2.4.1 Les canons pour le choix des caractéristiques

Différenciation : Les caractéristiques utilisées comme base de la classification doivent différencier certaines des entités de l'univers et doivent aboutir à la formation d'au moins deux classes ou isolats distincts.

Pertinence : une caractéristique utilisée comme base pour une classification doit être pertinente pour l'objectif de la classification.

L'auteur remarque que, pour des considérations pratiques, il faut restreindre le nombre de caractéristiques. Si on n'en choisit que certaines, cela signifie qu'il y a plusieurs schémas possibles pour un même ensemble d'entités, en fonction des objectifs de la classification.
Comment choisir les caractéristiques les plus pertinentes ? Ranganathan répond à cette question : « There is yet no definite answer to this question. No *à priori* rules for hitting upon the most helpful set of characteristics have been found as yet. Generally it depends on genius ; but other things being equal, persons with knowledge and experience are likely to develop the flair to reject the less helpful characteristics » (Ranganathan, 1967, p.147).
Il peut paraître surprenant de s'en remettre à la perspicacité des concepteurs, qui sont souvent aussi les utilisateurs dans les outils actuels. Il s'agit d'une question ouverte et à laquelle nous tentons de répondre dans la suite de ce travail, dans le cadre des expérimentations menées, et dans un contexte particulier. Il nous apparaît que cette question se pose à chaque nouveau contexte dans lequel il faut créer une classification et qu'elle est fortement corrélée aux objectifs du dispositif.

Vérifiabilité : une caractéristique doit être vérifiable. Par exemple, la date de naissance d'un auteur est vérifiable.

Persistance : une caractéristique doit rester inchangée tant que le but de la classification ne change pas. Ainsi, les facettes doivent refléter des propriétés permanentes des entités.

2.2.4.2 Les canons pour la succession des caractéristiques

La concomitance : deux caractéristiques ne doivent pas être concomitantes, par exemple, l'âge et la date de naissance. En d'autres termes, deux caractéristiques ne doivent pas aboutir à deux classes identiques, ou qui se recouvrent. Il s'agit ici du principe d'exclusivité mutuelle des facettes.

La succession pertinente : la succession des caractéristiques doit être pertinente en fonction des objectifs de la classification. La mise en œuvre de ce canon est abordée plus en détails dans la partie 2.2.9.3).

La cohérence : la succession des caractéristiques doit susciter l'approbation aussi longtemps qu'il n'y a pas de changement dans les objectifs de la classification. Elle doit être durable pour la cohérence de la classification dans la durée.

2.2.4.3 Les canons pour les ensembles de classes

L'exhaustivité : les classes dans un ensemble de classes doivent être exhaustives vis-à-vis de la classe immédiatement supérieure. L'auteur mentionne la possibilité d'ajouter une classe « Autre » pour satisfaire ce canon.

L'exclusivité : il s'agit ici de l'exclusivité mutuelle des classes. Cela signifie qu'une entité ne peut appartenir qu'à une seule classe d'un ensemble de classes de même niveau hiérarchique.

Des séquences efficaces : la séquence de classes d'un ensemble de classes doit être efficace pour les buts de ceux à qui elle est destinée. Cela peut donc varier en fonction du public. (voir partie 2.2.4.6)

Des séquences cohérentes : quand des classes similaires apparaissent dans des ensembles de classes différents, elles doivent être parallèles dans tous ces ensembles.
Exemple de jeu parallèle :

Médecine	Psychologie
Yeux	Vision
Oreilles	Audition
Nez	Odorat

Main	Toucher
Bouche	Goût

2.2.4.4 Les canons pour les chaînes de classes

L'extension décroissante : quand on passe du premier niveau de la classification vers le dernier, l'extension de la chaîne doit aller en décroissant, tandis que l'intention doit s'accroître. Il s'agit ici d'aller du plus générique au plus spécifique dans l'application des divisions successives.
L'extension a pour mesure le nombre d'entités comprises dans la classe. Elle correspond au champ couvert par la classe. L'intention a pour mesure le nombre de caractéristiques de division nécessaires pour dériver une classe depuis l'ensemble original d'entités.

La modulation : une chaîne de classes doit inclure une classe de chaque niveau compris entre le premier et le dernier niveau. Ranganathan qualifie aussi ce canon par l'expression « Resolving power » qui correspond à la capacité à reconnaître la classe inférieure appropriée à celle immédiatement supérieure. Par conséquent, une chaîne de classe, ou hiérarchie de classes, doit être dérivée depuis la classe immédiatement supérieure en utilisant le « resolving power » le moins fort possible à chaque stade de la division.

2.2.4.5 Les canons pour les séquences de classes

Les classes subordonnées : quand A1, A2, A3 sont des sous-classes d'une classe A, alors elles doivent suivre immédiatement la classe A dans la succession, sans être séparées de A par une autre classe, ni entre elles.

Les classes coordonnées : si les classes A et B sont originaires du même ensemble et étaient consécutives à l'intérieur de celui-ci, alors, elles ne doivent pas êtres séparées par une autre classe, excepté A1, A2, A3, qui ont pour classe immédiatement supérieure, A.

2.2.4.6 Les principes pour réaliser des séquences efficaces

Dans cette partie, il s'agit de l'énumération d'une liste de principes énoncés pour mettre en œuvre le canon énoncé en partie 2.2.4.3, qui concerne les séquences efficaces. Nous énonçons ici seulement ceux qui nous paraissent les plus importants. Ils servent à organiser les éléments d'un ensemble ou d'une classe dans un ordre utile, que ceux-ci soient des listes de sujets basiques ou d'isolats.

Principe chronologique : si les sujets d'un ensemble ont leur origine dans des temps différents, ils doivent être rangés chronologiquement.

Principe du « dernier dans l'évolution » : si les sujets d'un ensemble appartiennent à des stades différents de l'évolution, ils doivent être rangés en référence à cette évolution. Par exemple, pour des grandes périodes historiques, on rangerait ces périodes dans l'ordre suivant : préhistoire, antiquité, moyen-âge, époque moderne.

Principe de mesure quantitative : si les sujets d'un ensemble admettent une distinction quantitative, on peut les ranger par ordre croissant ou décroissant.

Principe de complexité croissante : si les sujets d'un ensemble présentent des degrés différents de complexité, il faut les ranger dans une séquence de complexité croissante.

Principe de séquence canonique : quand des sujets sont présentés traditionnellement dans une séquence spécifique, on doit s'y conformer.

Principe de caution bibliographique : les sujets d'un ensemble peuvent être organisés en fonction de la quantité décroissante de documents publiés ou anticipés sur les sujets.

Principe de séquence alphabétique : si aucune autre séquence n'est plus utile, on peut ranger les sujets d'un ensemble alphabétiquement par leur nom d'usage actualisé internationalement.

2.2.5 Les canons pour le travail sur le plan terminologique

Dans ces canons, on trouve des indications pour le choix des termes de la classification.

Canon de contexte : ce qu'un terme de la classification désigne doit être déterminé en fonction des classes supérieures appartenant à la même chaîne. Un terme a un contexte basé sur sa position dans le système de classification.

Canon d'énumération : c'est le même principe que le précédent mais appliqué aux sous-classes, le sens d'une classe est rendu clair par ses sous-divisions.

Canon d'actualité : le terme utilisé pour désigner une classe doit être celui utilisé couramment, actuellement, par les spécialistes du champ du sujet. Ceci implique que certains termes peuvent devenir obsolètes et doivent être remplacés.

Canon de retenue : un terme utilisé dans une classification doit être neutre, il ne doit pas exprimer d'opinion.

2.2.6 Sur le plan de la notation

Ce que nous retenons des indications données par Ranganathan au sujet de la notation, et qui sont très détaillées dans (Ranganthan, 1967, p. 218 – 290 et p.308- 323) est présenté ici.

Il souligne la nécessité d'utiliser des nombres ordinaux qui permettent de définir la position d'un élément dans une série, pour la notation. Selon lui, les qualités d'un système de notation sont les suivantes :

- Les indices de classes doivent être courts ;
- Les indices et leur mode de composition doivent être simples : ceci dépend du type de symboles utilisés et de la concision des indices ;
- Pour les micro-sujets, qui apparaissent notamment dans l'indexation des périodiques, l'indice ne peut pas être court.

Dans la Colon Classification, l'indice est découpé, chaque bloc correspondant à une facette, et les connecteurs indiquant le type d'idée représenté par la facette. Nous reviendrons plus loin en détails sur le modèle de classification à facettes, PMEST à cinq facettes (Personality,

Matter, Energy, Space et Time) proposé par Ranganathan. Ce que nous pouvons en dire au sujet de la notation, sur son principe, c'est que chaque type de facette du modèle PMEST est traduit par un nombre ordinal et celui-ci est précédé d'un signe de ponctuation dans l'indice.

- Ainsi, le nombre ordinal de la facette Temps est la virgule inversée : « ˌ »
- Le symbole qui précède la facette Espace est le point : « . »
- Celui qui précède la facette Energie est les deux points : « : »
- Celui qui précède la facette Matière est le point virgule : « ; »
- Celui qui précède la facette Personnalité est la virgule : « , »

Il s'agit d'une notation multipartite avec des blocs de nombres ordinaux connectés par des symboles ayant un sens, chaque symbole connecteur indiquant le caractère distinctif de l'idée représentée dans la succession des blocs. Dans la Colon Classification, les indices sont construits avec plusieurs types de symboles, notamment, par exemple, les lettres capitales, les lettres minuscules, les nombres ordinaux, la ponctuation, des parenthèses, les fractions décimales. La combinaison d'un nombre de symboles par bloc (par facette) conduit à pouvoir calculer le nombre potentiel de classes d'un système de notation. Un système de notation par groupes, comme celui proposé par Ranganathan, peut être conçu sur une base qui énoncerait que chaque bloc contient deux symboles, qui sont des lettres capitales (sauf I, O, et Z qui sont des classes vides), les classes successives seraient alors de type :
AA, AB…AY, BB, BA, BB… BY… YA, YB… YY, ce qui permet de former 529 classes.

Dans un système qui énonce que chaque bloc contient trois symboles, qui sont des nombres ordinaux, les classes successives présenteraient la forme suivante, qui permet d'augmenter la capacité de la classification :
111, 112…119, 121, 122… 129, 191, 192…199,
211, 212… 219, 221, 222… 229… 291, 292…299…

Enfin, un indice peut être de nature mixte, c'est-à-dire utiliser des types de symboles différents, par exemple, des lettres capitales, et des nombres ordinaux, ou bien, n'être constitué que d'un type de symbole comme dans les exemples donnés ci-dessus.

Ranganathan précise qu'un indice est unique, que chaque sujet doit être représenté par un seul indice, et que deux sujets ne peuvent pas avoir le même indice. Cela correspond au canon concernant la synonymie des indices. La longueur limitée des indices peut conduire à des problèmes d'homonymie des indices. Concernant leur uniformité, deux solutions sont possibles : soit l'indice est uniforme quel que soit la précision du sujet, soit il est relatif, le

nombre de symboles d'un indice correspond alors au nombre de division successives à effectuer pour l'obtenir. C'est cette dernière solution qui est choisie par Ranganathan dans la Colon Classification.
Le principe d'associer des signes de ponctuation aux symboles de l'indice pour leur donner du sens, ici, de signifier de quelle facette il s'agit, correspond au canon de la notation à facettes. La succession des blocs de symboles dans l'indice représente le degré d'importance des caractéristiques, dans la succession des caractéristiques pour représenter une classe.

2.2.6.1 Canons de notation pour des ensembles d'entités en expansion

Le canon énoncé ici cherche à résoudre les problèmes au niveau de la notation et ceux relatifs à l'ajout d'une classe dans un schéma existant, en prenant en compte l'imprévisibilité des nouvelles entités.

Canon d'Hospitalité : il doit être possible de faire des ajouts à la fin d'un ensemble ou entre deux indices consécutifs.
L'extrapolation correspond à la possibilité d'ajouter un nouvel indice au début ou à la fin d'un ensemble. L'interpolation correspond à la possibilité d'ajouter un nouvel indice entre deux indices consécutifs.

Pour l'extrapolation, et l'interpolation, plusieurs solutions non exclusives les unes des autres sont possibles :

- prévoir des symboles vides : laisser des symboles vides au début et à la fin. Mais cette solution est limitée, quand tous les symboles vides sont utilisés, on ne peut plus extrapoler ou interpoler ;
- prendre un autre type de symbole : on peut ajouter une sorte de symbole non utilisé pour les ajouts, mais quand tous les symboles disponibles sont utilisés, cette méthode aussi est limitée ;
- sectoriser les symboles : ce principe consiste à prévoir une classe finale comme une classe numéro 9, dans un système d'indice à un seul symbole numérique ordinal, pour en faire une classe « Autres ».

Dans la CC, chaque espèce de symbole forme une zone dans un ensemble, cela revient à dire que chaque zone est sectorisée. Pour les symboles des types suivants, voici quelques exemples de zones sectorisées :

- pour les lettres capitales : la lettre « Z » matérialise une zone sectorisée ;
- pour les lettres minuscules : la lettre « z » matérialise une zone sectorisée ;
- pour les symboles numérique ordinaux à un seul symbole : le chiffre « 9 » fait de même.
- Dans la CC, on utilise aussi la notation par paquet : on peut ainsi extrapoler une zone entière à la fin d'un indice en ajoutant un bloc de l'indice entre parenthèses.
- On peut aussi ajouter une classe « Ø » (zéro) au début d'un ensemble, comme un symbole vide.
- Enfin, on peut utiliser le principe des fractions décimales, c'est-à-dire considérer chaque indice comme une fraction décimale. Une nouvelle classe est donc créée en subdivisant la classe précédant celle dont on veut faire l'ajout, sur la base de l'ajout d'une nouvelle caractéristique, matérialisée par un symbole supplémentaire dans l'indice.
- Exemple : K97 = Mammifères

  K972 = Marsupiaux

  K9721 : Diprotodontes

2.2.7 Différences entre les classifications bibliographiques hiérarchiques et les classifications à facettes libres

Nous relevons trois différences fondamentales entre ces deux types de classification. La première concerne la prédétermination des indices dans les classifications bibliographiques hiérarchiques, tandis que les indices désignent des sujets uniques dans la classification à facettes libres. La deuxième différence concerne le processus d'indexation. Enfin la troisième différence vise les processus de constitution des hiérarchies.

La première différence fondamentale entre la classification à facettes libres de Ranganathan et les classifications bibliographiques, se manifeste dans le fait que les indices des ressources que l'on classe avec la classification à facettes sont uniques tandis que ceux des classifications bibliographiques hiérarchiques, sont des indices prédéterminés, et ils désignent des classes de sujets.

Ainsi, dans la Classification de Colon, les indices désignent des sujets uniques, les classes possèdent des symboles, et sont combinables à l'infini pour construire ces indices. Pour le dire autrement, d'après Spiteri (1998), chaque sujet est représenté par un seul indice, et chaque indice représente un seul et unique sujet.

La CC possède toutefois une liste des indices contenus dans la classification présentant la terminologie correspondant à chaque indice comme un dictionnaire. Sur le plan structurel, elle est également composée d'un index de classification qui présente chaque élément terminologique et l'indice correspondant. Ces éléments terminologiques sont rangés alphabétiquement.

D'après (Maniez, 1999, p.253) :

> *« L'objectif premier de la CC est de parvenir à un nouveau système de classification des sujets. Les termes des listes décrites ci-dessus sont hiérarchisés, et chaque concept est exprimé à la fois par un terme linguistique et par un indice alphanumérique. Ceci constitue une première classification, celle des termes-concepts, qui est de type traditionnel et intégrée au système (les tables des facettes par domaines). D'autre part, l'ordre rigoureux des facettes (citation order) dans les formules d'indexation et l'expressivité de la notation génèrent des indices globaux cohérents que l'on peut classer, et ce classement des sujets constitue la véritable classification à facettes de la CC, de type analytico-synthétique. »*

Cette forme de classification à facettes fonctionne sur la base d'une liste de sujets basiques (qui correspondent aux disciplines, donc à la division de premier niveau dans les classifications bibliographiques classiques), complétée de nombreuses listes qui contiennent des isolats (des composés de sujets). La présentation des listes (de sujets basiques, et d'isolats) de la classification à facettes est bien hiérarchique dans les listes elles-mêmes, mais toutes les listes sont placées sur le même plan. Lorsque l'on veut classer un livre par son sujet, on va prélever dans les listes de termes, ceux qui correspondent à la composition du sujet, les

ordonner par priorité dans la constitution du sujet, puis effectuer la transformation de cette composition de termes en indice.

Schématiquement, le processus d'indexation à l'aide d'une classification bibliographique classique, après avoir analysé le sujet pour en déterminer la classe principale (à quelle division de premier niveau faut-il s'intéresser), nécessite de parcourir la classification et d'associer un indice existant au sujet du document, le plus proche possible, en fonction de la précision de la classification. Ceci implique que, par défaut, il peut se produire qu'une classe adaptée n'existe pas, et par conséquent le classement est approximatif dans ces cas. Ce genre de problème se produit plus rarement dans une classification à facettes libres. En effet, celle-ci est plus souple pour ajouter de nouveaux sujets éventuellement manquants, grâce aux combinaisons inédites permises par la structure de la CC. Ceci nous conduit à considérer les problématiques d'alimentation de la classification à facettes libres, pour un univers d'entités en expansion, comme celui des sujets.

D'autre part, dans la partie 2.2.4.6, qui porte sur la constitution de séquences efficaces, il s'agit de conseils pour construire des listes hiérarchiques, selon des critères variables, par exemple chronologiquement, par quantité croissante ou décroissante ou encore par complexité croissante. Dans les classifications bibliographiques classiques, les hiérarchies ne sont construites que d'après le degré de généralité du sujet qui doit aller en décroissant. Les hiérarchies dans ces derniers systèmes doivent aller du plus générique au plus spécifique, et selon les règles de division logiques (évoquées au chapitre 1). Cela démontre une forme d'ouverture des principes de construction des hiérarchies qui peuvent être construites toujours de manière systématique, mais selon des critères propres et les plus adaptés possibles à chaque liste de sujets basiques et d'isolats.

2.2.8 Le modèle à cinq facettes de Ranganathan

Cinq catégories fondamentales constituent le modèle de facettes PMEST évoqué plus haut. Elles servent à décrire tous les éléments composant un sujet. Nous reviendrons plus loin sur leur syntaxe ou l'art de les combiner. Nous allons ici nous intéresser à leur contenu. Ce modèle à cinq facettes, d'après Ranganathan, permet d'aborder l'indexation des sujets sans aucun *à priori* :

> *« The postulate of the five Fundamental Ideas and the other postulates and principles going with it have been tried out from 1955 outward. While basing Practical Classification on them, we do not begin with any pre-determined*

ideas about the sequence of the coumpound or complex subjects going with a basic subject, except that it should be helpful to the majority of readers. And the subjects do fall in a helpful sequence if classified by the postulational method. Moreover, a subtle consistency –a consistency at a deeper hidden level – pervades not only the sequence of the subjects but also the sequence of the facets in all subjects and even the sequence of the successive characteristics used in going down the chain of isolates ideas within any facet. Above all, classification of a subject on the basis of the Postulates and Principles is done without any pre-determined ideas about the facets it should or should not have, or about their number, or about the sequence. The resulting scheme for classification is a Freely Faceted Classification in every sens of the term. » (Ranganathan, 1967, p. 397).

Nous énumérons ci-après ce que décrit chaque facette du modèle.

Facette Temps (Time)

Elle désigne un ensemble d'isolat décliné comme suit : millénaire, siècle, décennie, année, etc., et ces valeurs constituent« une manifestation » de cette facette. Il s'agit de donner les indications relatives à des éléments temporels d'un sujet, dans cette facette.

Facette Espace (Space)

Elle sert à décrire des éléments relatifs à la géographie, à l'espace, aux lieux. Par exemple, un ensemble d'isolats géographiques concernant la planète Terre pourrait être décliné comme suit : continents, océans, mers, etc. Un ensemble d'isolats concernant la géographie des milieux serait décliné comme suit : désert, prairie, montagne, plaines, rivières, lacs, etc.

Concernant les regroupements de population, un ensemble d'isolat pourrait être décliné comme suit : mégalopoles, villes, villages, hameaux, lieux-dits, etc.

Facette Energie (Energy)

Elle sert à décrire une action effectuée par tout type d'entité (inanimée, animée, conceptuelle, intellectuelle, intuitive).Par exemple : prévention, soin, distribution, etc.

Facette Matière (Matter)

Cette facette sert à décrire des matériaux ou des propriétés, comme les protéines, le bois, l'acier, la hauteur, la douceur ou la fermeté.

Facette Personnalité

Elle est obtenue par le biais de la « méthode du résidu ». Cela signifie que tout élément qui ne peut pas entrer dans les autres catégories (facettes) doit être placé dans cette facette.

On remarque ici l'ambiguïté entre les trois facettes suivantes : Energie, Matière et Personnalité. Au sujet du postulat de départ des cinq facettes fondamentales, Ranganathan reconnaît qu'il peut y avoir des doutes sur la distinction entre ces trois facettes dans certains cas. Mais selon lui, cela se produit rarement (Ranganathan, 1967, p.401).

2.2.9 Principes d'indexation des sujets à l'aide du modèle à cinq facettes

Nous présentons ci-après le processus de décomposition d'un sujet en facettes, ainsi que les principes, énoncés par Ranganathan, pour constituer des séquences de facettes. Enfin nous précisons la méthode de transformation d'un sujet en indice de classification à facettes.

2.2.9.1 Identification de la composition des sujets en facettes

Pour identifier la facette basique, il faut connaître la liste des sujets basiques. Celle-ci, d'après Ranganthan, est souvent perceptible dans le titre du document. L'indication de la facette basique dans le titre d'un document peut être explicite, implicite ou absente.

Les facettes basiques pour les sujets composés

Une facette basique contient des sujets basiques. Tout sujet composé possède un sujet basique. Un sujet peut contenir deux ou plusieurs facettes basiques. Il s'agit alors d'un cas d'une « relation de phase » entre :

- des facettes basiques ;
- des sujets composés, dont les facettes basiques sont les facettes basiques de chaque sujet composé ;
- entre un sujet basique et un sujet composé : un des sujets est contenu dans une facette basique et un autre sujet composé possède aussi une ou plusieurs facettes basiques.

Par exemple :

- « Traité sur l'exploitation minière du charbon » : la facette basique est ici l'exploitation minière ;
- « Manuel d'histoire de la France » : la facette basique est ici l'histoire ;
- « Structure des protéines » : la facette basique est ici implicite et se traduit par « chimie » ;
- « Impôts sur le revenu » : la facette basique est ici implicite et se traduit par « Economie ».

Les facettes d'isolats pour les sujets composés

Les facettes d'isolats contiennent des isolats (c'est-à-dire des composants de sujets). Chaque facette d'isolat d'un sujet composé ne peut être considérée comme une manifestation que d'une et seulement une des cinq catégories fondamentales. Pour l'identification des isolats dans le titre du document, il faut prêter attention au fait que les composants du sujet peuvent être implicites, explicites, cachés, ou absents.

Considérons à titre d'exemple : « structure des protéines et microscope électronique ». Ici, le sujet basique est la « chimie », la « structure » est une facette isolat du type Matière, « protéine » est une facette isolat de type Personnalité, « Détermination » est une facette isolat implicite de type Energie, et le « microscope électronique » est une facette isolat de type Matière. Cela donne une notation temporaire, car non ordonnée, qui est la suivante, avec les sigles suivants :

BF = Basique Facette

M = Facette Matière

P= Facette Personnalité

E= Facette Energie

Chimie (BF). Structure (M). Protéine (P). Détermination (E). Microscope électronique (M).

Dans certains cas, le contexte du sujet détermine dans quelle catégorie fondamentale placer des sujets qui peuvent entrer dans plusieurs facettes. Dans un sujet sur la « Fibre de coton », si le sujet basique est « Agriculture », le composant du sujet « Fibre de coton » sera du type de facette Personnalité, mais si le sujet basique est la « Technologie du textile » alors, le composant du sujet « Fibre de coton » sera du type de facette Matière.

2.2.9.2 Constitution syntaxique des séquences de facettes : principes théoriques

Dans une séquence de facette décrivant un sujet, une même facette peut apparaître plusieurs fois pour la description du sujet. Un sujet peut donc être composé de plusieurs séries de facettes ordonnées pour le décrire de manière exhaustive.

Les séries

Une série est toujours composée de la même manière selon les règles suivantes :

- Les facettes Espace (Space) et Temps (Time) sont toujours placées dans la dernière série ;
- La facette Energie (Energy) ne doit apparaître qu'une seule fois dans une série et à la fin de cette série ; elle joue donc le rôle de séparateur de série ;
- Les facettes Personnalité et Matière sont placées dans l'ordre suivant : d'abord la facette Personnalité, puis la facette Matière ;
- La première facette est la Facette Basique ;

On matérialise la série en ajoutant un chiffre ordinal devant le code de la facette. Par exemple, (1P) pour une facette Personnalité apparaissant dans la première série, et (2M) pour une facette Matière apparaissant dans la deuxième série dans un même sujet.

Les niveaux

Dans une même série, les facettes Personnalité et Matière peuvent aussi apparaître plusieurs fois. Elles doivent, dans ce cas être regroupées entre elles, placées les unes à la suite des autres dans une même série. On matérialise le niveau en apposant un chiffre ordinal après le code de la facette. Cela signifie que si deux facettes « Personnalité » et deux facettes « Matière » se trouvent dans la première série, alors la notation est de ce type : (1P1). (1P2). (1M1). (1M2)

Les facettes Espace (Space) et Temps (Time) apparaissent dans la dernière série seulement. Il n'y a donc pas besoin d'indiquer leur numéro de série, mais si elles ont plusieurs occurrences dans le sujet, alors il faut indiquer leur niveau. Par exemple, S1 est une facette Espace de niveau 1, et S2 est une facette Espace de niveau 2. La facette Energie n'est pas concernée par les niveaux puisqu'elle ne peut apparaître qu'une seule fois dans une série.

Exemple

Pour le sujet : « Prévention de la virulence des maladies de la tige du riz avec des produits chimiques pendant la période de sécheresse de 1967 à Madras » la décomposition en facettes ordonnées est la suivante :

Agriculture (BF). Riz (1P1). Tige (1P2). Maladie (1M1). Virulence (1M2). Prévention (E1). Produit chimique (2M1). Madras (S1). 1967 (T1). Période de sécheresse (T2).

2.2.9.3 Principes supplémentaires pour les séquences de facettes

Ces principes visent à guider l'ordonnancement des éléments dans une séquence de facettes.

Principe de la Vision d'Ensemble

Soient deux facettes A et B, si l'existence de B est conditionnée par l'existence de A, alors A doit précéder B dans la série. (Ranganthan, 1967, p.425). Par exemple, dans les sujets suivants :

« Prévention des maladies », l'ordre sera le suivant : Maladie. Prévention.

« Président de l'Inde », l'ordre sera la suivant : Inde. Président.

Principe du Tout-Organe

Un organe dans ce sens est une partie fonctionnelle d'une entité mais, si elle est séparée du tout, sa fonction cesse. Si B est un organe de A, A doit précéder B dans la série.

Principe « Vache-Veau » (Ranganthan, 1967, p.427)

Quand deux facettes peuvent être séparables mais ne devraient pas l'être. Cela ne donne pas l'ordre mais permet de déterminer les limites des séries. Pour déterminer l'ordre des éléments il faut en référer au principe de la Vision d'Ensemble. Par exemple, considérons le sujet suivant, « renforcement des pouvoirs du président en Inde ». « Renforcement » correspond à la première série, et « pouvoirs du président en Inde » correspond à la deuxième série.

Principe de l'Actant - Action – Acteur - Outil

Si une Facette B décrit une Action sur un Objet décrit par une Facette A, par un Acteur décrit par une Facette C, avec un Outil décrit par une Facette D, alors l'ordre adapté est le suivant :

ABCD. Par exemple pour « le filage du coton avec un fuseau par les femmes », l'ordre prescrit par ce principe est le suivant : Coton – Filage – Femmes – Fuseau

2.2.10 Méthode de transformation d'un sujet en indice de classification à facettes

Ce processus contient huit étapes successives que nous décrivons ici.

1. **Préparation du titre expressif**

A cette étape, il faut observer le titre du document, il peut manquer le sujet basique, ou des termes isolats. Dans ce cas, il faut inférer les termes absents et les insérer dans le titre expressif du document. Si le titre contient des parties composites, il faut les découper en constituants fondamentaux. Cette étape correspond à l'identification puis à la décomposition du sujet d'un document en éléments fondamentaux en langage naturel.

2. **Préparation du titre en « Termes Kernel »**

Un « Terme Kernel » est un substantif singulier qui va constituer un terme contenu dans une facette. A cette étape, il faut effectuer une lemmatisation, c'est à dire, réduire chaque terme du langage naturel de l'étape précédente en terme substantif singulier, et en supprimant les mots vides (les conjonctions de coordination, prépositions, articles).

3. **Préparation pour l'analyse du titre**

A cette étape, il s'agit de trouver les catégories fondamentales (les facettes) pour chaque Terme Kernel :

- Déterminer les séries, et les niveaux de ces termes de facettes.
- Ensuite, il faut inscrire les symboles des facettes à coté de chaque Terme Kernel (BF pour Facette Basique, T pour la facette Temps, S pour la facette Espace, P pour la facette Personnalité, M pour la facette Matière, et E pour la facette Energie).
- Enfin, il faut ajouter le numéro de série et de niveau autour des symboles des facettes. Par exemple, pour une facette Personnalité de série 1, puis une facette

Matière de série et de niveau 1 et une Facette Matière de série 1 et niveau 2, la notation sera la suivante : Terme Kernel (1P), Terme Kernel (1M1), Terme Kernel (1M2).

4. **Préparation de la transformation du titre**

Cette étape consiste à remettre les Termes Kernel dans l'ordre en fonction des priorités dans le sujet et d'après les canons énoncés en partie 2.2.9.3.

5. **Préparation du titre dans des termes standard**

Il s'agit de remplacer les Termes Kernel par des termes du langage classificatoire.

6. **Préparation du titre avec les numéros des classes**

Remplacer chaque terme standard par son numéro de classe avec l'index de la classification.

7. **Préparation de l'indice**

Remplacer les symboles des facettes (BF, P, M, E, S, T) par les signes de ponctuation appropriés.

8. **Vérification**

Il faut retransformer l'indice dans le langage naturel pour vérifier la pertinence de l'indice.

2.2.11 Critiques de la méthodologie de Ranganathan

Pour Mills (2004), l'importance du travail de Ranganathan réside dans le fait d'avoir été un pionnier dans le domaine de la classification à facettes, et d'avoir réalisé des expérimentations pour son développement. Mills, affirme que la classification à facettes est l'apport le plus important pour la science des bibliothèques depuis cinquante ans. Le caractère central de la classification à facettes se manifeste dans l'observation rigoureuse des règles de division logiques, à la différence des classifications générales, selon (Mills, 2004), tandis que Ranganathan lui-même n'a jamais mentionné cette caractéristique fondamentale.

Nous rappelons ici ces caractéristiques de division logiques, qui se résument en trois règles :

1. une seule caractéristique de division peut être appliquée à la fois : cela rappelle le canon d'exclusivité énoncé dans la partie qui porte sur les Canons pour les ensembles de classes (partie 2.2.4.3) ;
2. les divisions ne doivent pas avoir de manque dans une succession de classes et doivent se suivre de manière proche : cela rappelle le canon de Modulation (Résolving power) énoncé dans la partie qui porte sur les Canons pour les chaînes de classes (partie 2.2.4.4) ;
3. les divisions doivent être exhaustives : cela correspond au canon d'exhaustivité énoncé dans la partie qui porte sur les Canons pour les ensembles de classes (partie 2.2.4.3).

Cependant, il semble que cette méthodologie ne réponde pas à un certain nombre de questions :

- Comment choisir les caractéristiques de division ?

  A cette question, Ranganathan s'en remet à la perspicacité des concepteurs de la classification.

- Comment concevoir la classification elle-même en termes de méthodologie ?

  En effet, cette méthodologie orientée par des règles apparaît difficilement en mesure de guider la mise en place d'une classification entière avec les différentes étapes à suivre : elle est fondée sur des principes théoriques et peu opérationnelle. Par ailleurs, elle n'évoque pas quelles sont les ressources utilisées pour la création d'une classification : est-ce le principe de la caution bibliographique, ou une connaissance générique de tous les champs couverts par la classification, ou encore un recueil de textes reconnus ?

Ainsi, cette méthodologie est très précise sur le plan de la notation mais pas sur le plan de l'élaboration complète de la classification. Plus précisément, cette méthodologie n'explique pas comment réaliser les listes de sujets basiques et d'isolats. Elle commence à l'étape suivant l'élaboration des listes, comme si elles étaient déjà données. Ses préoccupations sont organisationnelles (au sens de l'organisation des connaissances) et sont plutôt focalisées sur la notation, mais également sur l'usage du système classificatoire par les indexeurs. Cependant elle ne constitue pas un guide complet pour l'élaboration du schéma. Sur ce point, d'après

Mills (2004), la Classification de Bliss (2ème ed.) donne des conseils pour l'élaboration d'une classification spécifique à facettes, en six étapes, qui sont les suivantes :

1. division des sujets (ou champ du savoir traité par la classification) en grandes catégories (Facettes) ;
2. division de chaque facette en liste (de valeurs de facettes) ;
3. décision concernant l'ordre de citation des facettes et des valeurs de facettes dans la classification (constitution du corps organisé de la classification) ;
4. décision concernant l'ordre de citation des facettes pour la description des sujets ;
5. ajout d'un système de notation ;
6. ajout d'un index de sujets alphabétique.

Elle est orientée vers l'utilisation et la compréhension des principes fondamentaux de la classification à facettes, qui consistent à décomposer les sujets en éléments fondamentaux, puis à les traduire dans le langage de classification, c'est à dire sur l'analyse par facettes.

Elle ne précise pas certains aspects du fonctionnement des facettes. Ainsi, outre le fait que le sens de certaines facettes comme la facette Personnalité ou la facette Matière est difficile à déterminer clairement, le sens des facettes subordonnées à la Facette Basique est inféré du fait que cette facette basique encadre la compréhension du contexte. Cette facette est une forme d'ordre hiérarchique, puisqu'elle vient en premier. Elle est du niveau de généricité le plus élevé et détermine le sens des autres facettes. Ces dernières reflètent les différentes relations que les concepts peuvent avoir avec la facette basique. Cette facette basique spécifie ainsi l'usage des autres facettes, tandis que la syntaxe de la classification décrit le type de facette (Mills, 2004). Ces relations sémantiques ne sont pas explicitement décrites par Ranganathan.

Louise Spiteri (1998) présente elle aussi les principes de la classification à facettes, en comparant les approches de Ranganathan et du CRG pour proposer un modèle simplifié d'élaboration de cette sorte de classification. Nous revenons en partie 2.5.1 sur ses travaux, car Denton (2003), les reprend pour élaborer des classifications pour les sites Web. Ce qu'il est intéressant de noter de la lecture critique de Spiteri (1998) est qu'elle relève les incohérences, les affirmations conflictuelles ou redondantes, dans les propositions de Ranganthan.

Tout d'abord, la complexité syntaxique et sémantique du langage de Ranganathan rend ses travaux difficiles d'accès et, à certains égards, obscurs. Elle souligne aussi que la structure du « Prolegomena to Library Classification » (Ranganthan 1967) que nous rapprochons de celle

d'un code de loi, avec ses 46 Canons, 13 Postulats et 22 Principes, peut le rendre intimidant, notamment pour des étudiants.

Elle relève plusieurs incohérences, que nous listons ici. En partie 2.2.4.2, nous avons présenté les canons pour la succession des caractéristiques, Spiteri (1998) note que celui qui concerne « La succession pertinente » peut se trouver en conflit avec le fait que Ranganathan prescrit un ordre précis pour l'utilisation des facettes. Dans cette même partie, le canon de « Cohérence » peut aussi entrer en conflit avec la Loi d'impartialité (Partie 2.2.3), puisque cette dernière énonce qu'il y a un ordre pour toutes les facettes qui doit être le même pour tous les sujets. Dans la partie 2.2.4.3, nous avons présenté les canons pour les ensembles de classes, et sur celui qui concerne « L'exhaustivité », Spiteri (1998) émet un doute sur le fait qu'on peut être certain d'avoir représenté tous les aspects d'une entité. Dans cette même partie, le canon qui concerne « Les séquence cohérentes » et qui explique les jeux de classes parallèles, Spiteri (1998) affirme que la similarité des classes concernées n'est pas expliquée clairement, et également que ce canon peut entrer en conflit avec le fait, que selon Ranganathan, il y a un ordre de citation des facettes qui doit être utilisé pour tous les sujets.

Dans la partie 2.2.4.4, nous avons présenté les canons pour les chaînes de classes, et sur celui qui concerne « l'extension décroissante », Spiteri (1998) note qu'il implique une division hiérarchique d'entités, qui ne correspond pas aux principes de l'analyse par facettes.

2.3 La méthodologie du CRG

Dans cette partie nous présentons la conception proposée par le Classification Research Group (CRG), quelque peu différente de la classification à facettes de Ranganathan, et qui vise principalement l'élaboration de classifications pour des champs spécifiques. A la différence de Ranganathan dont le projet était de proposer un modèle à facettes pour décrire tous les champs du savoir contenu dans une bibliothèque classique, les membres du CRG ont un projet plus limité en termes d'étendue du domaine couvert par une classification.

2.3.1 Historique du CRG

D’après (Vickery, 1966), la composition de ce groupe d’étude, créé en 1952, a fluctué mais a toujours gardé un noyau d’une douzaine de personnes régulières comme, par exemple, parmi les plus renommés, les membres suivants : Jean Aitchison, Eric Coates, Jason Farradane, Douglas Foskett, Barbara Kyle, Jack Mills, Bernard Palmer et Brian C. Vickery.

Leurs thèmes de recherche, entre 1952 et 1960, se sont concentrés sur la construction et l’usage de schémas spécifiques de classification. Dans les années qui ont suivi, ils se sont plutôt intéressés aux relations entre les classifications spécifiques et génériques, et aux problèmes liés à la construction d’une nouvelle classification générique.

A cette époque, l’influence de groupe est notoire, puisque d’après (Vickery 1966), au Royaume-Uni et dans l’Europe tout entière des membres du CRG sont sollicités par des institutions ayant des besoins de classification. Lors de leurs interventions, ils recommandent des schémas à facettes, ce qui contribue à la notoriété de la classification à facettes. Ce groupe de chercheurs a aussi alerté le fond de recherche en Sciences de l’OTAN sur les problématiques des classifications génériques et spécifiques, et cette institution a octroyé une bourse à la Library Association de Londres pour les étudier.

2.3.1.1 Des classifications spécifiques plutôt que génériques

Dans les classifications traditionnelles, le cadre général est la liste des disciplines traditionnelles, qui constituent les classes principales. Chaque discipline forme les frontières d’une classe principale, et ce sont ces classes principales que l’on soumet à l’analyse par facettes. Pour un schéma spécifique la frontière est le champ d’intérêt des utilisateurs. Cela en fait une classification orientée-utilisateurs, dont le champ peut être pluridisciplinaire. Pour le CRG, un domaine peut être un champ plus ou moins détaillé. Ainsi ses membres ont travaillé sur des schémas dont le titre révèle en partie le contenu comme : technologie du diamant, sciences des bibliothèques, technologie agro-alimentaire, ou science du sol.

Le modèle utilisé dans ces classifications spécifiques est du type Objet-Processus, ou Entité-Activité. Dans chaque cas, les intérêts des utilisateurs guident un choix d’entité (diamants, bibliothèques, alimentation, sols), et parmi les relations possibles (propriétés, processus, opérations, comportement, activités, usages) certaines sont sélectionnées comme pertinentes. Le degré de détail du champ couvert par la classification spécifique varie en fonction des besoins des utilisateurs.

D'après Alan Rees, qui commente les travaux du CRG, et que Vickery cite :

> *« faceted schemes are analytico-synthetic rather than enumerative. They pré-coordinate rather than post-coordinate. They operate at the present time with manual rather than mechanical searches and are, (…), mission-oriented rather than discipline-oriented. What we mean by that is that they are designed for user groups whose interests cut across the traditional fields.» (Vickery, 1966, p.14-15).*

2.3.2 La conception de la classification à facettes selon le CRG

Le CRG considère la classification à facettes comme importante pour trois raisons :

1. pour son impact sur les schémas de classification génériques existants : la grande classification bien reconnue, l'UDC (Universal Decimal Classification), a intégré une structure à facettes. D'autre part, le Bristish Standard Institution a choisi un membre du CRG, Jack Mills, pour préparer le nouveau guide de l'UDC.

2. Pour la notoriété des membres du CRG qui sont sollicités dans les institutions européennes, et qui préconisent ce type de classification, avec un bon écho.

3. Car des besoins de vocabulaires contrôlés apparaissent dans les années 50, besoins auxquels la classification à facettes répond en fournissant un jeu de principes et de techniques qui ont démontré leur efficacité.

Ces arguments avancés dans les années 60 sont à replacer dans le contexte de l'émergence de nouveaux systèmes documentaires comme l'invention des thésaurus qui date des années 50.

La classification à facettes, en divisant les termes en catégories ou facettes avec une notation pour lier les facettes entre elles, a donné la possibilité de rassembler des termes dans une relation coordonnée pour désigner une idée complexe.

Du point de vue des concepteurs de classifications, beaucoup de termes qui paraissent simples de prime abord (voir partie 2.2.3 Méthodologie du CRG, page suivante), sont en réalité des concepts complexes. C'est dans le traitement de ces termes que les catégories conceptuelles (les facettes) peuvent être utilement appliquées. Elles aident à révéler les contradictions, les omissions, l'implicite ou les inférences nécessaires.

Du point de vue des indexeurs, la structure de chaque champ de sujet est présentée. Chaque facette étant relativement courte, elle peut être scannée visuellement en entier pour choisir le bon terme.

Du point de vue des utilisateurs, d'après (Vickery, 1966), ceux qui cherchent à connaître le langage d'indexation sont rares, la plupart commencent leurs recherches avec une requête simple, correspondant rarement au sujet composé complet. Il considère que l'utilisateur tente de déverrouiller le système à l'aide de cette requête simple (à un mot), qui peut être un mot d'un sujet composé, ou un terme lié à tel mot d'un sujet composé. Cette requête va déverrouiller le système seulement si elle dirige vers le sujet composé recherché ou si elle oriente l'utilisateur vers un terme plus spécifique qui donnera accès au sujet recherché. Ces sujets recherchés sont atteints plus facilement si la structure de l'index est basée sur des relations génériques qui sont clairement présentées dans les listes de facettes.

Une des différences fondamentales, entre les classifications traditionnelles et la Colon Classification (CC) d'une part, et les classification proposées par le CRG d'autre part, réside dans le fait que celles du CRG sont « orientées utilisateur ». Elles sont spécifiques à un domaine, pour servir les intérêts d'utilisateurs particuliers, et non ceux d'un public large comme celui d'une bibliothèque municipale ou universitaire. Cette caractéristique influence leur usage et leur structure. (Vickery, 1966, p.40).

2.3.2.1 Le point de vue du CRG sur la Colon Classification

De manière générale, la classification à facettes est vue par le CRG comme un outil pour l'usage par des indexeurs humains, et son application n'est pas limitée à une forme de matériel enregistrée, tandis qu'il est précisé qu'avec un code adapté, la recherche pourrait être assistée par des machines. Si l'on convient qu'elle est développée pour des champs spécifiques, son application est réservée aux intérêts de groupes limités et homogènes d'utilisateurs. A ce titre, la CC est regardée comme un jeu de schémas spécifiques dans une variété de champs, dont certains ont été développés en détail. Ainsi, chaque discipline qui est détaillée en facettes dans la CC est vue comme une classification indépendante des autres, et constitue, pour le CRG, une classification à part entière dont certaines parties sont communes aux autres parties de la classification (les listes d'isolats communs).

2.3.3 Méthodologie d'élaboration de classification à facettes du CRG

La classification à facettes est vue comme un schéma de classification, obtenu par l'analyse à facettes. La classification se compose des éléments suivants :

- Une liste des termes d'indexation inclus dans la classification, classés, avec les symboles de notation pour chaque terme. Le but de cette liste est de donner une carte visuelle des concepts inclus dans le schéma et leurs interrelations pour guider le choix des termes pour les indexeurs et utilisateurs.
- Un index alphabétique de ces termes avec leur notation. Le but de cet index est de guider les utilisateurs et indexeurs depuis le premier mot choisi comme entrée du catalogue possible vers la zone de la classification la plus à même de proposer un choix de terme associé au premier mot sélectionné, pour faciliter le choix final du mot le plus approprié.
- Un jeu de règles qui :
    - décrit la structure de la liste des termes d'indexation et l'index alphabétique ;
    - indique comment les utiliser pour la sélection d'un terme ;
    - indique comment combiner les symboles de notation pour créer des entrées dans le catalogue ;
    - décrit si une entrée et sa notation doivent être ajoutées pour un sujet particulier ;
    - indique la séquence dans laquelle les entrées avec notation doivent être rangées ;
    - indique comment les entrées de l'index sont dérivées ;
    - explique comment les utiliser le plus judicieusement possible.

2.3.3.1 La construction de la liste de la classification

Nous résumons ici le processus d'élaboration de classification à facettes proposé par le CRG, qui se déroule en six étapes : la définition des entités d'intérêt pour les utilisateurs, la collecte de termes, l'élaboration des facettes, leur structuration individuelle, la structuration des facettes entre elles et enfin, l'ajout d'une notation appropriée.

2.3.3.1.1 Définir les entités d'intérêt pour les utilisateurs

La première étape consiste à déterminer les entités d'intérêt pour les utilisateurs et quels aspects de ces entités représentent un intérêt. Peu de détails sont fournis dans la littérature sur

cette étape, notamment sur les modes de recueil des besoins des utilisateurs. Sont-ils découverts par le moyen d'entretiens, ou par d'autres procédés ?

2.3.3.1.2 Collecte dans les sources

Les classifications à facettes spécifiques proposées par le CRG sont basées sur la connaissance détaillée des besoins des utilisateurs, des glossaires existants, ainsi que d'autres langages d'indexation existants, de manuels sur le champ traité par la classification et de la littérature sur le champ. La deuxième étape consiste donc en une collecte terminologique. A ce sujet, Vickery (1966, p. 45) écrit : « The most effective way of doing this is to examine a representative range of material that directly expresses the interests of the user group ; for example, their own reports and papers on topics of interest. »

Pour Vickery (1966), la classification à facettes est un langage contrôlé et structuré de descripteurs. Il distingue plusieurs niveaux de termes concernant les langages contrôlés traditionnels et les langages de requêtes des utilisateurs :

- Le niveau du besoin d'information qui est du langage naturel ;
- Le niveau des requêtes utilisateurs qui est un langage d'interrogation ;
- Le niveau des termes extraits des textes est terminologique ou verbal ;
- Le niveau des descripteurs qui est conceptuel ou symbolique. Les descripteurs sont des symboles verbaux contrôlés pour des concepts (d'après Mooers, cité par Vickery, 1966).

L'approche classique en classification consiste à s'intéresser à une idée, et à chercher ensuite le symbole qui peut la représenter. Dans la classification à facettes, selon le CRG, il est plus pertinent de collecter la terminologie depuis les dictionnaires, les thésaurus, et la littérature, pour ensuite l'organiser. Ils collectent alors des mots, ils utilisent des termes. (Vickery, 1966, p. 37). Le fait que ces mots incarnent des entités conceptuelles n'est qu'une conséquence de ce processus. D'après Vickery, la différence entre un mot en général (en langage naturel) et un descripteur est une différence de degré, non une différence de type, et cette différence de degré concerne le degré de contrôle et de structure. Pour lui, les deux sortes de mots représentent des concepts.

Le niveau terminologique correspond donc à la collecte des termes et à l'expression des relations présentes entre les termes dans les textes. La structure introduite dans les

vocabulaires contrôlés est révélée uniquement par l'analyse conceptuelle. L'analyse par facettes est une sorte d'analyse conceptuelle.

2.3.3.1.3 Elaboration des facettes

La troisième étape est l'inspection des termes à inclure dans la classification. Elle consiste, après la collecte des termes, qui forment un matériau brut, à examiner chaque terme, en l'interrogeant avec les questions suivantes :

- quel concept représente-t-il ?
- dans quelle catégorie conceptuelle ce concept devrait-il être placé ?
- quelles sont les relations de classes entre ce concept et les autres concepts inclus dans la même catégorie ?

Cette étape consiste à ranger les termes collectés dans des groupes correspondant à des facettes. Ce sont des groupes homogènes de termes, dérivés en définissant chaque terme en fonction des centres d'intérêt de la classification. Par exemple en Chimie, « l'alcool » est une « substance chimique », « liquide » est un « état » de cette substance, la « volatilité » en est une « propriété », « combustion » est une « réaction », et « l'analyse » est une « opération » réalisée dessus par un humain. Dans cet exemple, les mots « substance », « état », « propriété », « réaction », « opération », sont, d'une part, des caractéristiques de division par lesquelles la classe Chimie est dérivée, ce sont, d'autre part, des catégories logiques dans lesquelles on peut assembler des termes. Enfin, elles expriment certaines relations générales entre les termes. Par exemple, la relation entre « alcool » et « combustion » est une relation générale de type Substance/ Réaction. Ces relations sont indiquées par un opérateur relationnel dans les classifications du CRG.

Concernant les divisions logiques, chaque division distinctive doit être isolée, chaque nouvelle caractéristique de division doit être clairement formulée, chaque nouvelle relation doit être reconnue. L'application stricte de ces règles fait dire à Vickery (1966), qu'il s'agit là d'une différence entre les classifications traditionnelles et les classifications à facettes spécifiques.

D'autres différences mentionnées par Vickery (1966) résident dans la souplesse du système à facettes, qui permet aux catégories de ne pas être figées dans des listes énumératives, car toutes les combinaisons de termes de catégories différentes sont autorisées, et toute relation

entre termes ou entre sujets peut être exprimée. La classification à facettes se dégage de la restriction des classifications traditionnelles qui s'en tiennent à des relations hiérarchique de type Genre/ Espèce, de nouvelles relations logiques pouvant être introduites, permettant de refléter plus justement la complexité du savoir.
Le résultat final de ce processus est une liste de facettes, qui est un schéma conceptuel, une structure de catégories conceptuelles, dans laquelle les concepts exprimés terminologiquement ont été organisés. La méthodologie utilise aussi les techniques suivantes : élimination de synonymes, descripteurs génériques, lemmatisation, chaînes hiérarchiques, coordination de descripteurs, formes de liens et contrôle des rôles.

2.3.3.1.4 Structuration de chaque facette

Cette quatrième étape de structuration consiste à trouver un ordre hiérarchique pour les termes collectés et insérés dans chaque facette. Le résultat de cette opération est un jeu de facettes structurées. Le degré de détail de chaque facette dépend des usages des utilisateurs. Vickery (1966, p.52) précise que si l'on prévoit qu'une facette sera peu utilisée, alors elle peut être peuplée de termes plutôt génériques. Cette étape permet de faire les éventuelles fusions de termes nécessaires, d'éliminer les synonymes, de replacer des termes éventuellement insérés dans une facette inappropriée, et d'ajouter des termes manquants. Des notes d'application peuvent aussi être ajoutées à certaines facettes, si nécessaire.

2.3.3.1.5 Le modèle à treize facettes du CRG

La présentation du modèle à treize facettes du CRG est ici nécessaire afin de pouvoir poursuivre la présentation des dernières étapes de la méthodologie.

Ce modèle a été conçu d'après l'expérience des membres du CRG, qui ont constaté qu'une liste de catégories récurrentes se retrouvait souvent dans les différents schémas à facettes qu'ils ont pu élaborer. Ce modèle rassemble des jeux de facettes en catégories fondamentales de facettes. Ces catégories fondamentales de facettes permettent d'identifier la relation au sujet principal énoncé dans la première facette, comme chez Ranganathan, avec la facette basique. Les sept catégories fondamentales de facettes sont énoncées ci-après :

- Entité / Objet

- Partie/ Composant/ Structure
- Matériau
- Attributs
- Opérations
- Agent
- Lieu/ Condition

Chaque catégorie fondamentale de facettes contient des facettes qui composent le modèle du CRG. Dans la liste qui les présente ci-dessous, les termes du premier niveau peuvent décrire une catégorie fondamentale de facette ou constituer directement des facettes. Ces cas sont illustrés dans les exemples qui suivent la présentation de ce modèle de facettes. Ainsi d'après cette présentation, une facette pourrait être intitulée « Composant » ou, plus spécifiquement si cela est pertinent, « Organe ». Nous donnons des exemples de valeurs de facettes après chaque titre de facette.

2.3.3.1.6 Catégories fondamentales et facettes du modèle

- Entité / Objet :
    - Naturel : minéraux, animaux, plantes, sols
    - Produit : ponts, moteur, fibre
    - Construction mentale : équations, rectangles, formules
- Partie/ Composant/ Structure : rayon, roue, aile
    - Organe : cœur, graine
- Matériau : métal, verre, azote
- Attributs
    - Qualités/ Propriétés : cohésion, couleur, solubilité
    - Processus / Comportement : vibration, inflammation
- Opérations
    - Expérimentales : taille, élevage, éducation
    - Mentales : calcul, raisonnement
- Agent : tout objet ou entité pouvant agir comme un agent
- Lieu/ Etat : Bombay, achevé

L'exemple suivant nous aidera à comprendre que ce modèle est adaptable en fonction du champ traité par la classification, il sert de guide. En effet, dans (Vickery, 1960, p.20), on trouve cette mise en garde contre une application systématique du modèle :

> *« Toute liste de catégories fondamentales, ne devrait pas être utilisée mécaniquement et appliquée arbitrairement au sujet, mais il peut être utile de l'employer à titre de guide provisoire en abordant un nouveau domaine. Elle fournit un cadre déterminé qui peut convenir au sujet et servir de guide en suggérant des caractéristiques possibles qui ne devraient pas être négligées. Cette liste, d'autre part, ne doit jamais être considérée comme épuisant le sujet et comme nécessairement applicable à tout sujet. »*

Il apparaît logique, par ailleurs, que ce modèle soit contextuellement adapté puisqu'il est dédié à être appliqué à des domaines différents, qui peuvent ne pas posséder toutes les caractéristiques énoncées dans le modèle de facettes, ou en posséder certaines supplémentaires. Il ne vise pas à constituer une syntaxe systématique pour la description du réel, mais forme un guide pour appuyer la construction de schémas de classifications spécifiques. Les catégories fondamentales de facettes peuvent ainsi contenir des facettes spécifiques au domaine traité, mais le principe consiste à tenter de les faire entrer dans ces catégories fondamentales pour garder une cohérence vis-à-vis des autres classifications à facettes créées d'après ce modèle. Nous proposons ici un exemple pour illustrer le principe des catégories fondamentales.

Exemple 1 : Catégories fondamentales pour la science de sols

N° de Facette	Nom de la facette	Exemple de valeur de facette
1	Entité centrale de la classification	Sols
2	Partie du sol	Gravier
3	Composant	Azote
4	Structure du sol	Coupe
5	Couche du sol	Sédimentaire
6	Organismes dans le sol	Bactérie
7	Matériau Parent du sol	Boue
8	Processus dans le sol	Minéralisation
9	Propriétés du sol	Cohésion
10	Mesure de propriété	Point d'adhésion
11	Operations sur le sol	Amendement
12	Equipement pour les opérations	Charrue

Dans cet exemple, tiré de (Vickery, 1967, p.47), la facette 1 correspond au sujet de la classification et est une Entité. Les facettes 2 à 5 sont des Parties, Composants, Structure. Les

facettes 6 et 7 appartiennent à la catégorie fondamentale qui concerne les Matériaux. Les facettes 8, 9 et 10 sont des facettes issues de la catégorie fondamentale Attributs. La facette 11 est une Opération, et la facette 12 est un Agent.

2.3.3.1.7 Structuration des facettes entre elles

Cette cinquième étape s'intéresse à l'ordre des facettes entre elles. Il dépend de l'usage prévu de la classification, en termes de post-coordination, ou de pré-coordination. La post-coordination implique que chaque terme peut être assigné indépendamment des autres, avec un ordre des facettes qui présente d'abord, selon les préconisation du CRG, les facettes Entités, puis les facettes Parties et Composants, puis les facettes Matériau, ensuite les facettes Attributs, Opérations, Agents, Lieu, tel que présenté dans le modèle précédemment.

La pré-coordination, qui est la plus utilisée au moment de l'édition du manuel de Vickery (1966), implique des séquences de facettes précises, car l'ordre des facettes dans la classification est usuellement le même que l'ordre des facettes dans les indices. L'ordre des facettes dans les indices est appelé « citation order » par les concepteurs de classifications à facettes, et décrit la séquence dans laquelle les facettes sont combinées pour former des indices pré-coordonnés.

Quand le modèle de Ranganathan proposait un « citation order » basé sur le modèle des cinq facettes, Personnalité, Matière, Energie, Espace et Temps, en le complétant des dispositifs de séries et de niveaux (voir partie 2.2.9.2.), le CRG évite ces dispositifs en détaillant davantage les catégories.
L'ordre de citation des facettes, dans le modèle du CRG, est donc le suivant:

- Entité / Objet
- Partie/ Composant/ Structure
- Matériau
- Attributs
- Opérations
- Agent

- Lieu/ Condition

Si bien que, dans la liste qui présente la classification, ces éléments sont disposés ainsi :

Entité – Partie – Matériau
Entité – Attributs
Entité – Opérations
Entité – Agent

Dans l'exemple suivant, qui est un extrait d'une classification sur les avions, la facette 1 est l'entité principale. Les facettes 2, 3 et 4 sont des facettes de la catégorie fondamentale des Partie/ Composant/ Structure. Les facettes 5, 6 et 7 sont des facettes de la famille Matériau. Les facettes 8, 9 et 10 sont des facettes de la famille Attributs. Les facettes 11, 13, 14 sont des facettes Opérations, tandis que les facettes 12 et 15 sont des facettes Agent.

N° de Facette	Nom de la facette	Exemple de valeur de facette
1	Entité centrale de la classification	Avion
2	Partie	Aile, Hélice
3	Composant	Moteur
4	Composant de Machine	Piston
5	Matériau spécifié par usage	Fuel
6	Matériau spécifié par constitution	Métal
7	Matériau spécifié par forme	Liquide
8	Propriété des matériaux	Elasticité
9	Comportement de l'atmosphère	Humidité
10	Comportement de l'avion	Maniabilité
11	Opération de l'avion	Décollage
12	Instrument opératoire	Altimètre
13	Opérations d'ingénierie	Test
14	Opérations mathématiques	Analyse
15	Equipement pour Operations	Balance

Nous pouvons faire deux remarques sur cette classification concernant l'ordre de citation. La première découle de l'observation des facettes 5, 6 et 7 dans le tableau, qui montrent une possibilité de préciser la spécification d'une facette. Ainsi, les matériaux qui composent un avion ou qui servent à son fonctionnement, peuvent être considérés selon plusieurs aspects. La possibilité de préciser la spécification d'une facette constitue donc selon nous un assouplissement du modèle, que nous intitulons « principe de spécification de facettes ». De la

même manière, un avion peut être considéré sous divers aspects, dans son ensemble.Par exemple du point de vue :

- de son usage : fret ou transport de personnes ;
- de son propriétaire : compagnie aérienne, état, propriétaire privé, militaire ;
- de son type de propulsion :propulsion nucléaire, turboréacteur, combustion/explosion, hélice, solaire ;
- de son nombre de moteurs : moteur unique, bimoteur ;
- de son type de train d'atterrissage : en nacelles, fixe, rétractable.

La deuxième observation consiste à remarquer que la facette 12 qui appartient à la catégorie fondamentale Agent, est placée entre deux facettes Opérations, alors qu'elle devrait se situer plutôt dans les dernières facettes de la liste, si l'on suit l'ordre de citation évoqué ci-dessus. Vickery (1967, p.55) explique qu'il est possible d'insérer des facettes « dépendantes » qui s'appliquent, qui sont en relation, uniquement avec certaines facettes, celles qui les précèdent, dans l'ordre de citation logique. Ainsi, la facette 12 est en relation uniquement avec les deux facettes précédentes 10 et 11 dans le tableau. Nous appellerons ce dispositif le principe d'intercalation des facettes.

2.3.3.1.8 Ajout d'une notation appropriée

La sixième étape de la méthodologie consiste à ajouter un système de notation approprié. La notation consiste en un symbole ou un jeu de symboles, utilisés pour représenter les sujets des documents, dans l'organisation matérielle des ouvrages. Elle permet de créer les indices qui sont les références codées des ouvrages. Elle permet de mécaniser l'organisation matérielle des ouvrages.

L'argument principal pour justifier l'ajout de la notation est l'organisation matérielle des ouvrages par sujets, du plus générique au plus spécifique, ce qui permet à l'utilisateur qui vient chercher dans la collection avec un sujet peu précis, de parcourir la gamme des ouvrages dans le catalogue et de les retrouver ordonnés de façon correspondante, dans les étagères. L'organisation des sujets n'est pas de l'ordre du langage naturel mais plutôt de l'ordre conceptuel. La notation, comme chez Ranganathan, doit pouvoir accueillir de nouveaux sujets. Ainsi, n'importe quelle partie de l'indice doit pouvoir être étendue.

Il est souvent mentionné que l'indice doit remplir ces fonctions élégamment ; il doit être immédiatement compréhensible par l'utilisateur, en termes de relations ordinales dans la structure de l'indice. Enfin, il doit être facile à mémoriser, et doit par conséquent être le plus économe possible en longueur. Ses caractéristiques doivent être définies en fonction des caractéristiques des utilisateurs les plus intensifs de ce dispositif, donc soit les lecteurs, soit les indexeurs. Il peut être souhaitable que la notation reflète les caractéristiques structurelles des sujets classés. La structure renvoie à deux aspects :

- la structure d'un sujet composé individuel : l'indice doit être conçu pour éclairer les utilisateurs sur les combinaisons de termes réalisées ;
- la structure hiérarchique d'une facette : l'indice peut être conçu pour indiquer que deux termes proviennent du même ensemble, ou de la même facette. S'il est conçu ainsi, l'indice est dit « hiérarchique », sinon il est purement ordinal.

Un indice pour une classification spécifique peut contenir :

- un symbole pour chaque terme issu d'une facette ;
- un symbole pour décrire de quelle facette il est issu ;
- un dispositif syntaxique pour lier deux termes issus de la même facette ;
- un dispositif pour les appels à des termes issus d'éventuels autres systèmes de classification complémentaires (interopérabilité) ;
- un symbole pour représenter des relations comme « comparaison » ou « effets ».

Exemple d'utilisation d'indices

Considérons :

- une facette Entité notée A qui décrit le type de bibliothèque ;
- une facette Matériau notée B qui décrit le type de document ;
- une facette Processus notée C qui décrit les opérations ou activités ;
- une facette Agent notée D qui décrit les agents des opérations.

Alors, dans une classification sur les bibliothèques, par exemple, les indices pour des sujets intitulés comme suit, seraient :

- « Stockage des pamphlets dans une bibliothèque commerciale » : A9B3C5
- « Stockage des pamphlets dans une bibliothèque universitaire » : A65B3C5

- « Effets du manque de personnel sur le catalogage des coupures de presse dans une bibliothèque commerciale » : A9B4C6/D4

Dans ces exemples, A9 correspond au type de Bibliothèque commerciale, tandis que A65 décrit une bibliothèque universitaire. B3 symbolise les Pamphlets, tandis que B4 est utilisé pour les coupures de presse. C5 décrit l'opération de stockage, tandis que C6 indique l'opération de catalogage. Dans le dernier exemple, la notation inclue un dispositif incarné dans le « slash » pour décrire la relation « Effets », tandis que D4 est utilisé pour indiquer l'agent « manque de personnel ».

2.3.3.1.9 Processus d'indexation à l'aide du modèle à facettes du CRG

L'indexation, pour le CRG consiste à observer un document, puis à déterminer de quel(s) sujet(s) il traite et lesquels, parmi ceux repérés, sont pertinents pour les utilisateurs. Dans cette décision, il peut être choisi d'indexer seulement le sujet majeur, mineur, ou les deux. Chaque sujet sélectionné comme pertinent doit être formulé en une phrase, en langage naturel, qui pourrait servir de titre au document. Il faut ensuite chercher chaque mot dans l'index alphabétique des termes de la classification. Dans le cas où les mots ne se trouvent pas dans l'index, il faut alors procéder à l'analyse par facettes de ces mots, c'est-à-dire les assigner à des facettes présentes dans la classification, et chercher le terme correspondant, qui sera soit un synonyme, soit une subdivision d'un terme existant, soit absent de la classification. Dans ce dernier cas, une décision quant à l'insertion d'un nouveau terme doit être prise. Cette étape requiert une connaissance du sens des termes et de la structure de la classification. Pour chaque sujet, il faut ensuite associer les symboles de notation, et les combiner selon les règles de la classification. D'après (Vickery, 1967, p.69), une règle devrait être invariable : combiner les éléments dans l'ordre dans lequel ils apparaissent dans la classification.

2.3.3.1.10 Recherche à l'aide de la classification dans un catalogue sur fiches

Dans un processus de recherche dans un catalogue papier sur fiches, dans un usage pré-coordonné, l'usager de la classification peut avoir à manipuler trois sortes de documents :

- La classification des ouvrages par sujet : dans ce fichier, sont contenues toutes les acquisitions du fond documentaire. Chaque entrée ne réfère qu'à un document, il présente d'abord l'indice, déterminant son classement par sujet, puis des données bibliographiques classique, et enfin la localisation du document. Il peut y avoir plusieurs entrées par document pour donner accès à des sujets différents traités par le même document. Ce fichier permet de présenter les références dans un ordre qui va placer les documents ayant des sujets proches, les uns à coté des autres.

- L'index alphabétique des ouvrages par sujets : dans ce fichier, les sujets sont classés alphabétiquement, chaque entrée réfère à une seule entrée du fichier de classification des ouvrages. Il présente les sujets exprimés en langage naturel et le symbole de notation correspondant. Il peut y avoir plusieurs entrées pour le même sujet pour représenter un sujet avec des termes qui varient. Ce fichier sert à retrouver un sujet dans un langage proche de celui des utilisateurs et à guider, à l'aide du fichier de classification des ouvrages, vers ceux qui sont susceptibles d'être pertinents.

- Un index alphabétique des auteurs des documents : ce fichier présente une liste des auteurs et les documents correspondants, chaque entrée renvoie vers un seul document. Il reprend les informations du fichier de classification des ouvrages.

Plusieurs cas peuvent se présenter :

Cas 1 : le lecteur cherche avec un terme qu'il trouve dans l'index alphabétique. Il peut alors se référer au fichier de la classification des ouvrages par sujets par le biais du symbole notationnel qu'il a pu prélever dans l'index alphabétique.

Cas 2 : le lecteur cherche un terme qui est à la fois une entité (donc le début d'un indice), et qui est aussi un composé de sujet. Alors, il trouvera les documents dont le terme est le sujet principal par la méthode précédente, mais pas les sujets dont il est un composé, excepté s'il existe un fichier d'entrées multiples.

Cas 3 : le lecteur cherche avec deux termes qui sont des composés de sujets, (qui n'arrivent pas en premier dans un indice), il peut alors trouver des documents sur chacun des deux termes, et parcourir la classification depuis un terme plus générique, mais il risque d'oublier des documents, excepté s'il utilise un fichier d'entrées multiples. Il se peut aussi que la composition des deux termes ait une signification précise et qu'elle ne soit pas contenue dans les références indexées au niveau de chaque terme. Là encore, un fichier d'entrées multiples se révèle très utile. Un fichier d'index alphabétique permuté est aussi reconnu comme étant d'un usage utile mais difficile à utiliser.

Dans ce type de dispositif, le catalogue n'est pas informatisé, à l'inverse de la plupart des catalogues de bibliothèques, aujourd'hui. Cela complique la recherche. En effet, au lieu de

parcourir la classification et de sélectionner des facettes et des termes dans ces facettes dans un dispositif doté de liens hypertextes, qui pourrait donner lieu à des modifications multiples des sélections pour faire le tour d'un sujet, et ceci très rapidement grâce à la puissance des systèmes informatiques actuels, l'utilisateur doit manipuler des fichiers papier, comprendre le système de notation, le mémoriser et ensuite aller chercher les documents sur les étagères. Même si le fond documentaire n'est pas informatisé, nous pensons que le processus de fouille à l'aide de la classification peut être grandement simplifié et soutenu par un système à facettes informatisé.

2.3.4 Différences entre classifications à facettes et classifications traditionnelles

D'après Vickery (1966), dans les classifications traditionnelles, les divisions effectuées pour arriver à un arbre du savoir commencent depuis un univers original de connaissances, qui est le sommet de l'arbre. Les divisions immédiatement inférieures sont les classes principales (ou disciplines traditionnelles). Chaque document doit se trouver en un seul point de l'arbre. Théoriquement, le processus de division est logique. Chaque classe est subdivisée en classes spécifiques au moyen d'une caractéristique de division particulière. Chaque terme d'une classe est divisé d'une seule manière, et toutes les subdivisions d'une classe sont des vraies espèces du genre parent. Il s'agit ici d'appliquer la même caractéristique de division à un même ensemble à subdiviser, à chacun des stades de division, pour obtenir des classes homogènes, au niveau des classes coordonnées. Cette pratique implique la rigidité du système des classifications traditionnelles, car il n'est pas possible, selon cette règle, de diviser un même ensemble selon différentes caractéristiques qu'il peut éventuellement posséder. Le fait qu'un document ne se trouve qu'à un seul endroit de l'arbre participe également à cette rigidité.

Au lieu de commencer par le sommet de l'arbre, comme une classification à priori (les classifications traditionnelles), l'analyse par facettes rassemble les termes d'un champ, (elle commence à un niveau hiérarchiquement inférieur), pour les grouper dans des catégories : substance, état, propriété, réaction, opération, système, etc., puis range les termes de chaque catégorie selon l'ordre le plus approprié, pour construire une carte classificatoire.

Les divisions ne sont pas toujours logiques (au sens selon des classifications classiques) dans la CC. Par exemple, les classes de premier niveau d'une classe Chimie Physique sont ainsi subdivisées dans la CC :

1. Action et combinaison chimiques
2. Solution
3. Mélange
4. Thermochimie
5. Photochimie
6. Electrochimie
7. Magnétochimie

Les classes 2 et 3 sont des types de substrats, les classes 4 à 7 sont des disciplines de la chimie classées en fonction de différentes formes d'énergie, la classe 1 est discipline transversale.

On voit ainsi que dans l'analyse par facettes les divisions d'un univers original peuvent se faire sur la base de caractéristiques hétérogènes, mais toujours uniques pour chacune des subdivisions du même niveau. Il est donc possible pour les classes d'un même niveau de subdivision d'être basées sur des caractéristiques différentes, mais on ne peut en appliquer qu'une pour chaque classe de ces classes subdivisées. Les différences de conception des divisions logiques sont illustrées dans les schémas suivants.

La différence avec les classifications traditionnelles réside aussi dans le fait que le choix de la première caractéristique de divisions est l'élément le plus déterminant puisqu'il décide de la portée conceptuelle de la classification. On constate donc que les auteurs ne s'accordent pas sur l'application stricte des règles de divisions logiques dans les classifications à facettes. (cf Mills 2004, partie 2.2.11)

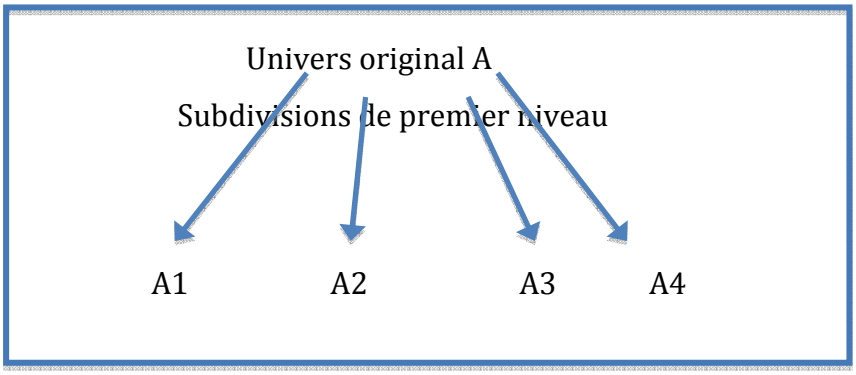


Figure 2. Schéma des divisions dans une classification traditionnelles

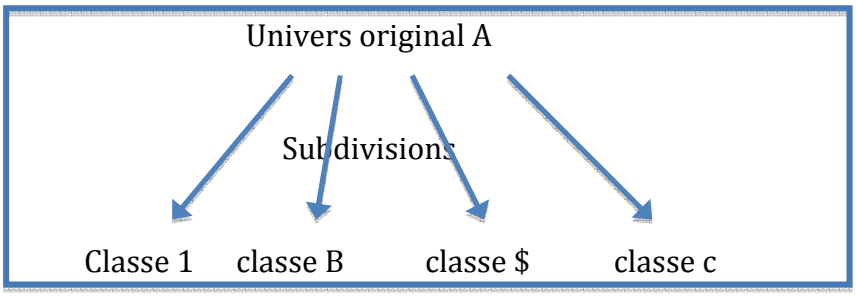


Figure 3. Schéma des divisions dans une classification à facettes

Une autre différence entre les classifications spécifiques à facettes et une classification traditionnelle comme la UDC, mentionnée par (Vickery, 1967, p.87) réside dans le fait que les premières donnent plus de détails que ne le fait la UDC, et c'est une des raisons pour lesquelles, les schémas de classification spécifiques sont préconisés pour des domaines dans lesquels il y a des besoins de classifications détaillées.

Les classifications à facettes spécifiques permettent, par ailleurs, plus de souplesse dans la combinaison des termes pour décrire des sujets composés, et elles sont plus synthétiques car elles intègrent dans les indices les indications précises des facettes spécifiques dont sont tirés les termes. De plus, les termes de certaines facettes peuvent être plus éparpillés dans la UDC que dans les classifications construites sur le modèle du CRG.

Enfin, la notation est plus courte dans les classifications à facettes proposées par le CRG que dans une classification traditionnelle. Elle est donc plus lisible, et plus facilement mémorisable.

2.3.5 Comparaison des méthodologies de Ranganathan et du CRG

Tout d'abord, chez Ranganathan, la classification est conçue à priori, comme la majorité des classifications bibliographiques, de manière descendante, par la division de l'univers du savoir en grandes disciplines, et de manière théorique. Les sources d'inspiration et de connaissances pour la construction de la CC ne sont pas mentionnées dans le manuel de Ranganathan. La méthodologie du CRG, à l'inverse, propose de partir de sources précises, comme la littérature d'un domaine et les langages d'indexation existants pour effectuer une collecte de termes, qui seront ensuite à organiser en facettes.

L'application des principes de l'analyse par facettes paraît plus souple et plus détaillée dans la méthodologie du CRG, que dans celle de Ranganathan. En effet, des indications sont données sur le processus d'analyse de chaque terme collecté dans celle du CRG, que nous rappelons ici :

- quel concept représente-t-il ?
- dans quelle catégorie conceptuelle ce concept devrait-il être placé ?

- quelles sont les relations de classes entre ce concept et les autres concepts inclus dans la même catégorie ?

C'est la raison pour laquelle nous affirmons que la méthodologie du CRG se rapproche nettement plus d'une méthodologie générative (ou opérationnelle) que celle de Ranganathan, que nous considérons comme basée sur des principes théoriques. Ainsi, la méthodologie proposée par le CRG est plus précise et concrète pour guider l'élaboration d'une classification à facettes.

Il faut préciser que l'objectif des deux méthodes n'est pas le même, d'un côté Ranganathan cherche à élaborer une classification pour l'ensemble des sujets d'une bibliothèque, et s'adresse potentiellement dans ce cadre à des interlocuteurs qui pourraient avoir le même objectif ou d'autres objectifs liés, comme l'élaboration d'un standard ou d'une norme. D'un autre côté, le CRG vise à construire ces schémas à facettes pour des organisations spécifiques sur des domaines spécifiques.

Ensuite, l'ordre de citation des facettes du modèle est plus souple dans l'approche du CRG que dans celle de Ranganathan. Chez ce dernier, l'ordre de citation des facettes dans les indices est rigide, et doit être le même pour tous les sujets, tandis que l'on peut intercaler des facettes dans l'ordre préconisé des catégories fondamentales de facettes adaptées dans l'approche du CRG (cf partie 2.3.3.1.7) en fonction des besoins contextuels du domaine à organiser. Il s'agit du principe d'intercalation.

Enfin, si l'on cherche à comparer les modèles des deux méthodologies, on constate que le CRG utilise un modèle conceptuellement plus détaillé, moins générique, et moins abstrait que Ranganathan, ce qui le rend plus compréhensible, accessible, et utilisable. Ainsi, la facette basique « Entité » dans le modèle du CRG nous paraît plus claire que celle intitulée « Personnalité » du modèle de Ranganathan qui joue approximativement le même rôle.

D'après nous, si l'on cherche à faire entrer le modèle du CRG dans celui de Ranganathan, on constate bien une filiation, et l'on pourrait ainsi considérer que ce dernier a créé un méta-modèle de facettes pour expliquer les principes de la classification à facettes, que le CRG a opérationnalisés.

Familles de facette du Modèle du CRG	Facettes du Modèle de Ranganathan
Entité /Objet	Personnalité
Partie/ Composant/ Structure	
Matériau	Matière
Attributs	
Opérations	Energie
Agent	Matière
Lieu/ Etat	Lieu & Temps

Tableau 2. Comparaison des modèles de facettes du CRG et de Ranganathan

Dans le tableau ci-dessus, l'agent d'une opération est considéré comme une facette Matière dans le modèle de Ranganathan, car la facette Energie est réservée aux actions. Les catégories fondamentales de facettes dans le modèle du CRG correspondent donc à un méta-modèle plus détaillé. L'intercalation possible dans le modèle du CRG, correspond aux dispositifs des Séries et Niveaux pour les indices chez Ranganathan. (voir partie 2.2.9.2).

Enfin, à propos des modèles présentés par le CRG et Ranganathan, Spiteri (1998) souligne une différence fondamentale entre les deux approches. En effet, pour les membres du CRG, il n'existe pas de catégories fondamentales valables pour tous les sujets, les catégories doivent être déduites de la nature des sujets à classer. Le CRG préfère identifier les catégories en référence au contexte du sujet lui-même. Ainsi, les membres du CRG pensent qu'aucune liste ne peut être exhaustive ou applicable à tous les sujets.

2.4 La méthodologie GenISIS (Cogniva)

La méthodologie GenISIS est une composante d'une structure plus large composée d'un modèle conceptuel, de l'application logicielle ISIS (Information Semantic Infrastructure Services), et de cette méthodologie. Cette méthodologie permet de concevoir le modèle conceptuel relatif à une organisation dans laquelle le déploiement du logiciel est prévu. Elle a pour fonction principale de soutenir le processus d'adaptation du modèle et d'élaboration de celui-ci pour une organisation particulière afin de l'injecter ensuite dans l'application. Ainsi, elle encadre le travail conséquent de collecte d'information et d'élaboration théorique, qui est nécessaire pour la création du modèle de métadonnées adapté à l'organisation, avant le

paramétrage de l'application. Nous avons bénéficié de plusieurs sessions de formation à GenISIS organisées dans le cadre du projet Miipa-Doc, en France. La méthodologie GenISIS est dérivée des règles d'archivage. Ses principes reposent sur des notions qui sont définies dans la méthodologie, et qui sont les suivantes :

- L'information
- Le cycle de vie de l'information
- Les métadonnées
- La classification fonctionnelle
- Les records
- La conservation et l'élimination
- La valeur opérationnelle des documents
- Les modèles
- Décomposition fonctionnelle
- Modèle de processus
- La classification à facettes

Nous revenons sur les définitions de ces notions dans la suite de cette partie.

2.4.1 Une méthodologie au service d'une application

D'après les concepteurs de GenISIS, la variété des standards et des applications de gestion de l'information cause des problèmes de redondances entre les applications, et des problèmes pour la recherche et le partage d'informations. ISIS se positionne comme une solution unique et intégrée de gestion de l'information concernant l'architecture de l'information.
L'architecture de l'information est la structure utilisée pour organiser, gérer une collection d'informations dans un contexte particulier. Il s'agit d'une description abstraite de la façon dont l'information d'une organisation est gérée. Elle décrit comment les ressources informationnelles peuvent être groupées, comment elles sont reliées entre elles, comment aller d'une ressource à la suivante. Un plan de fichiers associé au cycle de vie des documents ou au processus de recherche est un des composants de l'architecture de l'information, tandis qu'un

site web ou le système de fichier en font aussi partie. Le schéma de métadonnées utilisé en est également un composant. Dans une organisation, une architecture unifiée et exhaustive qui inclue tous ces composants et assure la cohérence entre ces composants, est appelée « Enterprise Information Architecture » (EIA). Dans le cadre de l'utilisation de l'application ISIS, le modèle ISIS est l'unique source pour les composants de l'architecture de l'information de l'organisation qui choisit de l'utiliser.

D'autre part, les descriptions insuffisamment détaillées des sujets ou contenus des documents, conduisent à des problèmes de localisation de documents, de pertes d'informations et à des difficultés lors des recherches documentaires. Pour traiter ces problèmes, ISIS constitue un moyen d'associer des métadonnées aux documents traités par l'organisation. Ces métadonnées sont issues du modèle ISIS adapté à l'organisation par le biais de la méthodologie GenISIS. Le modèle de métadonnées est élaboré d'après des modèles de la décomposition fonctionnelle et des processus d'affaire de l'organisation, qui permettent de définir la valeur des documents et ainsi de leur appliquer les règles d'archivage appropriées. Nous revenons plus en détails sur ces notions dans la suite de cette partie.

2.4.1.1 Les fonctionnalités d'ISIS

Nous décrivons ici succinctement les fonctionnalités de l'application ISIS.

Le service de métadonnées est le service principal évoqué ci-dessus. Il consiste à associer à chaque facette un jeu de valeurs prédéfinies à l'aide de la méthodologie et à contraindre leur utilisation par des jeux de relations entre les valeurs de facettes. Une facette dans ce cadre est une taxonomie conceptuelle hiérarchique. Le système consiste en un jeu de valeurs suffisantes pour décrire tous les objets informationnels selon leur contenu et le contexte dans lequel ils sont créés ou utilisés. Les relations entre valeurs de facettes sont préétablies, et servent à filtrer les combinaisons autorisées de valeurs. Elles standardisent l'usage des métadonnées qui peuvent être renseignées automatiquement sur la base de ces relations définies.

Le service de thésaurus traite l'aspect linguistique des valeurs de facettes, en permettant d'associer une valeur à plusieurs dénominations en fonction des préférences des utilisateurs ou du vocabulaire qui leur est le plus familier. Il intègre aussi les relations classiquement présentes dans un thésaurus en présentant des spécifications des valeurs de facettes en

fonction de ces relations de Termes Générique, Terme Spécifique, Terme Associé, Employé Pour.

Le service d'archivage lie les métadonnées au plan de fichier de l'entreprise, qui est basé sur les règles d'archivage et qui permet de placer les documents dans les dossiers appropriés en fonction de leur durée de conservation.

Le service de recherche propose aux utilisateurs de parcourir les ressources selon plusieurs perspectives, en fonction des facettes auxquelles ils ont accès. Il permet aussi lors d'une recherche par sujet de découvrir le contexte lié au sujet.

Le service de carte du savoir est une fonctionnalité qui lie les personnes travaillant dans l'organisation avec une description de leur domaine de compétences en langage naturel. Cela étend les informations liées aux métadonnées Auteur.

Le service de gouvernance distribue les responsabilités liées à la maintenance d'une facette à des personnes appropriées dans l'organisation, qui peuvent ainsi suivre l'activité d'une facette, approuver ou rejeter une modification. Ce service sert aussi à produire des diagnostics et des rapports de performance.

2.4.1.2 Aux origines de l'application ISIS

Il faut ici mentionner que l'application ISIS a été déployée dans certains services du gouvernement Canadien. Ainsi, en 2006, l'institution Bibliothèque Archives du Canada publiait à propos des systèmes d'archivage : « *Fundamentally, the Regime is based on the development of documentation standards for recordkeeping within institutions linked to program activities, business lines, and assigned roles and responsibilities* »

En 2009, la directive de Bibliothèque Archives du Canada concernant l'archivage était approuvée. Les objectifs de cette directive étaient les suivants : « *Ensure effective recordkeeping practices that enable departments to create, acquire, capture, manage and protect the integrity of information resources of business value in the delivery of Government of Canada programs and services* ». Les responsabilités définies par cette directive consistent en particulier à :

- identifier, établir, mettre en œuvre et maintenir des entrepôts dans lesquels les ressources d'information à valeur opérationnelle sont stockées et préservées dans des mémoires physiques ou électroniques ;
- établir, utiliser et maintenir des structures de taxonomies ou classifications pour faciliter le stockage et la recherche d'information, (*search & retrieval*), ayant une valeur opérationnelle dans tous les formats ;
- établir, mettre en œuvre et maintenir les périodes de conservation pour les informations ayant une valeur opérationnelle, de manière appropriée au format ;
- développer et mettre en œuvre un processus documenté de suppression pour toutes les ressources d'information ;
- réaliser des activités régulières de suppression pour toutes les ressources d'information.

2.4.1.3 Classification fonctionnelle du Ministère des Finances

Au sein du gouvernement du Canada, le Ministère des Finances a besoin de savoir quelles sont les fonctions des départements du gouvernement et comment ils mettent en œuvre ces fonctions. Parce que ce Ministère veut faire le lien entre le financement, et le rapport départemental des dépenses, directement avec les fonctions que les départements doivent effectuer, des descriptions fonctionnelles ont été créées qui sont des « décompositions fonctionnelles » des départements du gouvernement. Elles sont appelées « Program Activity Architecture » (PAA)

Selon les concepteurs de GenISIS, alors que les PAA sont créées afin de mettre en œuvre le suivi financier, elles sont aussi pertinentes pour le management de l'information, car elles apportent un bon point de départ pour la décomposition fonctionnelle d'une organisation. En effet, une PAA peut être ensuite utilisée pour assister l'établissement d'une classification fonctionnelle des documents pour une organisation gouvernementale. Nous revenons dans la suite de cette partie sur la notion de classification fonctionnelle.

Isis a ainsi été construit spécifiquement dans le but de permettre à des organisations de déterminer quels documents doivent être traités comme des « records », et pour apporter des recommandations sur la conservation et les mesures d'élimination. En particulier, elle est conçue pour permettre à l'organisation de déterminer la valeur opérationnelle des documents,

basée sur les métadonnées associées à ceux-ci. Ceci permet en retour de déterminer si ce sont des records ou non, et quelles règles de conservation doivent leur être appliquées.

« *Élimination* est le terme en usage chez les archivistes pour désigner l'acte de retirer aux documents leur statut d'archives et donc les charges qui y sont attachées, à savoir l'obligation de conservation et le droit d'accès. La *destruction* est l'acte physique d'anéantissement de l'objet matériel, par déchiquetage, broyage, recyclage ou tout autre moyen approprié. Élimination et destruction visent respectivement la décision et sa mise en œuvre. » (Marie-Anne Chabin, 2007)[1].

2.4.1.4 Objectifs d'ISIS

D'après les concepteurs de GenISIS, les objectifs d'ISIS sont les suivants du point de vue organisationnel :

- déterminer les ressources d'information qui ont une valeur opérationnelle et déterminer combien de temps conserver ces ressources ;
- assurer l'association de métadonnées de qualité aux ressources ;
- documenter sous forme de preuves les activités liées aux transactions de l'organisation ;
- améliorer la capacité à réutiliser l'information ;
- accroître l'efficacité et améliorer les prises de décision.

Les objectifs du point de vue individuel consistent à :

- réduire le poids des tâches d'association de métadonnées ;
- permettre aux individus de trouver, suivre, garder, et découvrir des ressources d'information pertinentes ;
- permettre aux utilisateurs d'organiser l'information selon leur propre perspective.

2.4.1.5 Objectifs théoriques de la méthodologie GenISIS

La cohérence : elle consiste en une procédure formelle qui produit des résultats cohérents et simplifie les processus d'émergence : plusieurs personnes qui feraient l'analyse indépendamment les unes des autres dans une organisation devraient arriver aux mêmes résultats dans le modèle.

[1] http://www.marieannechabin.fr/archiver-et-apres/6-desarchiver/

La prévisibilité : comme la méthodologie est précise et exhaustive, il est possible de prévoir les résultats, le temps et les efforts à investir, et donc les coûts à engager.

2.4.1.6 Objets de la méthodologie

A la fin du processus d'analyse, doivent être produits :

- Les facettes de métadonnées
- Les valeurs de facettes
- Les hiérarchies de valeurs pour chaque facette
- Les relations entre les valeurs des différentes facettes

2.4.2 Notions de base de la méthodologie GenISIS

La méthodologie GenISIS tente de répondre à des questions centrales dans les problématiques du domaine du Record Management :

- combien de temps faut-il conserver une ressource d'information ?
- comment conserver des ressources d'information pour pouvoir les retrouver ?
- quelle valeur attribuer à une ressource d'information ? comment décider de sa valeur ?

La première question concerne les règles d'archivage, tandis que la deuxième est orientée sur les problématiques d'organisation des connaissances. Enfin, la troisième question porte sur la détermination de l'importance d'une ressource, ce qui détermine comment et combien de temps la conserver. Nous présentons ci-après les définitions proposées dans la méthodologie GenISIS des notions de base sur lesquelles elle est fondée.

2.4.2.1 La notion d'information

Cette méthodologie repose sur une définition de l'information qui est formulée comme suit : « *Any entity, electronic or otherwise, capable of conveying or supporting intelligence or knowledge; e.g. a book, a letter, a picture, a sculpture, a database, a person.* » Il est aussi précisé qu'une ressource d'information peut être un agrégat d'items, et qu'elle est le résultat d'un ensemble composé d'un contenu, d'un contexte et d'une structure. Ce dernier élément est important dans le cadre de l'étude de cette méthodologie, car nous verrons plus loin qu'elle impacte directement le modèle à facettes proposé, puisque celui-ci est aussi composé de trois types de facettes : des facettes de contenu, des facettes de contexte, et d'une facette

qui articule les deux autres types de facettes et que nous pouvons considérer comme une facette de structure (voir partie 2.4.5).

2.4.2.2 Le cycle de vie de l'information

L'application ISIS est en mesure de gérer le cycle de vie de l'information. Le cycle de vie d'un document correspond à l'ensemble des phases par lesquelles il passe durant sa durée de validité, en partant de sa création jusqu'à son élimination. Le concept de cycle de vie de l'information ne décrit pas comment stocker ou organiser les informations. D'après cette méthodologie, « The way information can be organized is dependent on the semantic nature of the information being organized. ». Cela signifie que l'organisation des informations dépend du contenu du document, de son sujet ou du sens porté par le document. Il faut donc en tenir compte dans la conception d'un système pour une organisation particulière, ainsi que des propriétés particulières des informations qui sont gérées. Ces propriétés particulières des informations peuvent relever des aspects suivants :

- Matériel privé ou confidentiel
- Matériel de valeur particulière, unique ou de grande utilité
- Propriétés matérielles
- Matériel qui procure des preuves, des faits ou des preuves d'événements

Ces ressources d'information peuvent nécessiter une organisation ou un traitement particuliers, en référence à des besoins spécifiques, des propriétés ou fonctions spécifiques.

2.4.2.3 Les métadonnées

Elles sont définies, dans cette méthode, comme des données ou informations sur l'information primaire. Il y est ajouté qu'elles peuvent décrire le contexte, le contenu ou la structure des documents à archiver, ainsi que leurs règles de gestion dans le temps. Cette définition des métadonnées est directement en lien avec la définition de l'information en termes de contenu, contexte et structure, donnée en partie 2.4.2.1). Cette définition permet de décrire des types de métadonnées. Ainsi dans la méthodologie GenISIS, on trouvera :

- Des métadonnées de contenu qui décrivent le sujet des informations contenues dans les documents ;

- Des métadonnées qui décrivent le contexte dans lequel la ressource d'information existe ;
- Des métadonnées de structure qui décrivent la structure de la ressource d'information, vue comme un contenant d'information.

Le choix des métadonnées est important dans le sens où celles que l'on choisit de représenter dans un système d'information orientent leur usage. Ainsi, pour ISIS, elles conditionnent les durées de conservation, les règles d'archivage associées, et les espaces de stockage des ressources. Le modèle ISIS que nous présentons dans la suite de cette partie, propose un modèle de métadonnées adaptable au contexte de l'organisation.

2.4.2.4 La classification fonctionnelle

Ce concept a été créé pour faire se rejoindre deux buts :

1. Classer les documents, sur la base de cette classification, intégrer des règles de conservation et d'élimination
2. Créer une classification de documents stable et pertinente sur période de temps étendue

Dans une classification fonctionnelle, les documents sont classés selon les fonctions endossées par les personnes qui créent les documents, et qui motivent la création des documents. Les avantages de ce type de classification sont de deux ordres, d'après les concepteurs de GenISIS : d'une part, la distribution des fonctions dans une organisation est vue comme étant relativement stable, ce qui permet de garder une cohérence dans la durée. Ce point nous semble discutable. En effet, dans de nombreuses organisations de grande taille, la structure des fonctions est remaniée dans une logique de rentabilisation des locaux ou du personnel et l'on assiste bien souvent à des fusions des services supports, à des redécoupages de la structure organisationnelle globale, à tel point que les acteurs peuvent en être désorientés. Cela peut aussi impacter la recherche d'information. Ainsi, d'après nos expérimentations, une recherche d'information peut être conduite d'après la structure organisationnelle en place au moment de l'élaboration du document recherché. Elle constitue plutôt, selon nous, un repère, ou un critère de recherche, même si elle est remaniée. Une application qui serait capable de proposer un historique de la structure organisationnelle comme critère de recherche, éventuellement sous forme de cartographie, gagnerait à être exploitée dans ces organisations dont la structure des services est régulièrement modifiée.

D’autre part, un argument en faveur des classifications fonctionnelles, d’après les concepteurs de GenISIS, réside dans le fait que la fonction relatée dans le document constitue un bon indicateur de sa valeur comme Record, et peut être utilisée pour déterminer les règles de conservation et d’élimination.

Les désavantages sont de deux ordres : ils concernent le fait que ce type de classification est généralement construit sur une vision de haut niveau de l’organisation, qui peut être difficile à synchroniser avec le point de vue des individus travaillant dans l’organisation. Ainsi, ces individus peuvent rencontrer des difficultés à se situer et faire correspondre leurs activités quotidiennes avec celles décrites dans la classification fonctionnelle.

2.4.2.5 Les Records

Dans un environnement d’archivage, un record est une preuve documentaire, sans considération de forme ou de médium, créée, reçue, maintenue, ou utilisée dans une organisation (privée ou publique) ou par un individu dans la poursuite de ses obligations légales ou des transactions d’affaires. Les ressources d’information sont désignées ou reconnues comme des records selon certaines conditions. Les propriétés que doivent satisfaire ces ressources d’information pour être reconnues comme Records sont les suivantes :

- l’exactitude
- l’exhaustivité
- l’inaltérabilité
- elles sont recherchables

Les différences entre les Records et des documents au sens commun du terme sont, d’une part, que les records, à la différence des documents plus classiques, ne sont pas accessibles au public, et d’autre part, que les records n’existent qu’en un seul exemplaire, et donc dans un seul lieu de stockage.

2.4.2.6 La conservation et l’élimination

La conservation est la durée pendant laquelle un record est gardé par son créateur (personne ou organisation). Ils sont conservés sur la base de leur usage (s’ils sont en usage actif ou non), et si non, sur la base de leur valeur.

L’élimination se rapporte à ce qui advient des records à la fin de leur période de conservation. Même si ce terme implique souvent la destruction du record, ce n’est pas toujours le cas. S’il

a toujours de la valeur après sa période de conservation hors du contexte de l'organisation, il peut être archivé.

2.4.2.7 La valeur opérationnelle des documents

Les documents ayant une valeur opérationnelle sont des documents publiés ou non, sans considération de forme ou de médium, qui sont créés ou acquis parce qu'ils facilitent ou documentent les prises de décisions en support aux programmes, services et opérations en cours, et qu'ils supportent l'évaluation des départements et les besoins en termes de performance et d'exactitude.

2.4.2.8 Les modèles

Il existe plusieurs sortes de modèles : les modèles statiques comme les cartes, les modèles dynamiques comme les simulations, et les modèle d'organisation qui sont de deux sortes, d'après la méthodologie GenISIS, d'une part les modèles des Processus d'affaires, et d'autre part, les décompositions fonctionnelles. Ces derniers peuvent être complémentaires, et c'est l'approche adoptée dans GenISIS. Un modèle est une description abstraite, ou une représentation, de phénomènes particuliers. Un modèle inclut :

- les objets impliqués dans le phénomène
- les relations qui existent entre ces objets
- la nature de ces relations

Un modèle permet d'avoir une vue d'ensemble, en une seule fois, des phénomènes représentés. Il peut contenir plusieurs parties et relations entre les parties. Seuls les détails pertinents et utiles pour les objectifs du modèle sont représentés ; il n'a pas besoin d'être exhaustif.

2.4.2.9 La décomposition fonctionnelle

Une décomposition fonctionnelle est une description fonctionnelle d'une organisation. A l'origine, ce concept provient de l'informatique et du monde de l'ingénierie. Les organisations sont vues de plus en plus comme des systèmes, et empruntent donc certaines notions à ces disciplines pour l'analyse des affaires. Elle se base sur les objectifs de l'organisation, soit l'activité principale de l'organisation, comme par exemple, produire des pesticides, fabriquer des passeports, gérer une clientèle pour un commanditaire ou produire un logiciel de logistique. Le but principal ou activité principale, peut être décomposé en sous-objectifs ou en

fonctions qui sont au service du but principal ou des buts intermédiaires. La décomposition fonctionnelle est hiérarchique par nature, les fonctions sont décomposées en sous fonctions entre 2 et 4 niveaux de hiérarchie. Dans la hiérarchie, les fonctions sont indépendantes et sont isolées, plutôt que rangées dans une séquence particulière.

Dans le cas de la production d'un logiciel, il faut connaître les fonctionnalités que le logiciel doit fournir. La modularité et la décomposition de la conception peuvent être soutenues par la compréhension fonctionnelle des exigences qui doivent être remplies par le logiciel. En ingénierie, le but peut être de fournir des systèmes pour prendre en charge des fonctions.

Pour définir les fonctions, il est conseillé d'utiliser des verbes d'action et de se détacher de la structure organisationnelle. Pour la construction de la décomposition fonctionnelle, il faut commencer par définir le but principal de l'organisation, c'est-à-dire sa mission principale. Ensuite il faut considérer les différentes façons selon lesquelles ce but peut être réalisé. Quand on approche des fonctions, on dit : « je fais 'fonction' en faisant 'sous-fonction' et 'sous fonction'». Ces sous-fonctions doivent être autonomes. Les fonctions émergent lors du processus de construction du modèle. La méthodologie détaillée de création de la décomposition fonctionnelle est abordée dans la suite de cette partie mais nous en donnons un aperçu ici. Elle se déroule en deux phases.

1ère phase d'analyse : obtenir une première vision globale de l'organisation

La décomposition fonctionnelle permet de regarder l'organisation d'un point de vue stratégique et se traduit par des entrevues de haut niveau avec plusieurs personnes, en général des personnes assez haut placées dans la hiérarchie (dirigeants) afin d'avoir un niveau d'abstraction suffisant, dans un premier temps. Ceci est motivé par le fait qu'en général les personnes qui travaillent à un niveau moins élevé dans la hiérarchie n'ont pas nécessairement une vue globale de l'organisation et de leur place dans celle-ci. Elle est complétée par l'analyse de documents :

- Organigramme ;
- Listes de tâches ;
- Rapports stratégiques ;
- Décomposition fonctionnelle existantes ;
- Eventuels processus de qualité et de normalisation des procédures de travail, qui facilite alors le travail de l'analyste ISIS, comme le processus ISO 9001.

Cette première analyse donne lieu à la réalisation d'un premier modèle, pour pouvoir le présenter au client et le retravailler avec lui. Ce modèle peut se présenter sous forme de « cartes cognitives » ou cartes heuristiques. Avec ces cartes, on peut ensuite conduire des entretiens en groupe de travail, ou en face à face, pour la correction, l'ajout de définitions, la complétion de la carte ou l'ajout de notes sur la carte.

2ème phase d'analyse : développement des sous fonctions

Elle se compose d'entretiens avec des cadres intermédiaires qui seront désignés par les dirigeants interrogés dans la première phase. Ceci permet d'intégrer des acteurs qui, désignés, peuvent être coopératifs, mais aussi probablement motivés pour participer. Dans l'ordre, on tente de dégager : Fonctions/ Sous-fonctions/ Activités/ Tâches

Principalement, on distingue en général trois types de fonctions :

1. Actions métier/ Fonction opérationnelle : production de biens ou services

2. Fonction décisionnelle : pilotage/ contrôle de l'action métier, évaluation de la performance, par exemple : contrôle, audit, gestion des risques, planification

3. Support administratif : par exemple Ressources Humaines, Finances

Une aide pour cette étape est l'utilisation de la « H-Méthod » (détaillée dans la partie 2.4.6.3.5). Elle permet d'analyser les documents en termes d'input (intrants) et d'output (extrants) dans des flux documentaires horizontaux ponctués d'actions. On peut également analyser les dossiers, comment ils sont organisés, selon quels critères d'évaluation, afin d'identifier différentes perspectives de l'action, et commencer à identifier les processus.

Le deuxième niveau d'analyse sert donc à raffiner les sous-fonctions et activités et à identifier les processus transverses multifonctions afin de briser les silos informationnels. C'est l'idée de « Fonction Versus Processus ». Sans l'étude des processus, d'après les concepteurs de GenISIS, certaines informations restent confinées dans le modèle par fonctions.

2.4.2.10 Modèle de Processus d'affaires

Connecté avec la décomposition fonctionnelle, il détermine comment une organisation mène à bien ses fonctions, et les détails spécifiques à une organisation dans son fonctionnement. Cela se fait en identifiant les entités (acteurs, personnes, ou groupes) impliqués dans une organisation, en déterminant comment ils sont en relation et comment ils interagissent les uns avec les autres. Il faut aussi identifier les inputs (intrants) et les outputs (extrants) et comment

les acteurs (entités) utilisent les inputs, produisent les outputs, et comment ils traitent les inputs.

Une entité est une personne, un lieu, ou objet d'intérêt pour une organisation, capable d'être décrit avec le langage naturel, et pertinent pour le contexte de l'environnement spécifique de l'organisation.

Un attribut est un aspect, une qualité, une caractéristique, ou un descripteur d'une entité ou d'une relation. Il doit présenter un intérêt pour l'organisation, être descriptible en langage naturel, et être pertinent pour le contexte de l'environnement spécifique de l'organisation.

Une Relation représente toute association, lien, connexion, entre les entités d'intérêt de l'organisation, et possède les mêmes caractéristiques linguistiques et de pertinence que les entités et attributs, définis précédemment.

Un Processus est une séquence d'activités en relation**,** ou séquence de tâches reliées qui construit et compose une activité. Ce sont généralement des tâches interdépendantes, et il y a un enchainement bien défini pour une activité ou d'une tâche à l'autre. Nous réfèrerons à ce concept dans la suite de ce mémoire à l'aide du terme « Processus ».

Cogniva développe une banque de processus standards (génériques) qui aide les acteurs a modéliser leur propre activité. Nous donnons quelques exemples types de processus standards :

- Transaction : Question / Réponse
- Conception de produit
- Recrutement
- Organisation de réunion

De même, nous donnons un exemple de décomposition du processus de réalisation d'une tâche : Analyse => Planification => Exécution => Evaluation => Complétion

La modélisation des processus se fait grâce à un logiciel de Gestion par Processus Métier, comme le logiciel QPR, qui permet d'injecter des informations dans ISIS directement.

Le processus méthodologique d'élaboration du modèle de processus est détaillé, aussi, nous y revenons dans la suite de cette partie, mais nous allons commencer à le présenter ici.

Pour créer un modèle de processus, il faut suivre les étapes suivantes, d'après GenISIS :

- Identifier les fonctions et sous fonctions de l'organisation, ce qui correspond à la décomposition fonctionnelle

- Identifier les objets dans l'organisation qui soutiennent la réalisation des fonctions principales
- Pour chaque fonction, il faut déterminer :
- Un point de départ et un point d'arrivée
    - Les Inputs et Outputs (intrants et extrants)
    - Les acteurs et objets impliqués dans cette fonction
- Enfin, il faut décrire la façon dont les objets et les acteurs vont du départ à l'arrivée, en utilisant les inputs et en créant des outputs.

Le schéma suivant présente un exemple théorique de modélisation de processus.

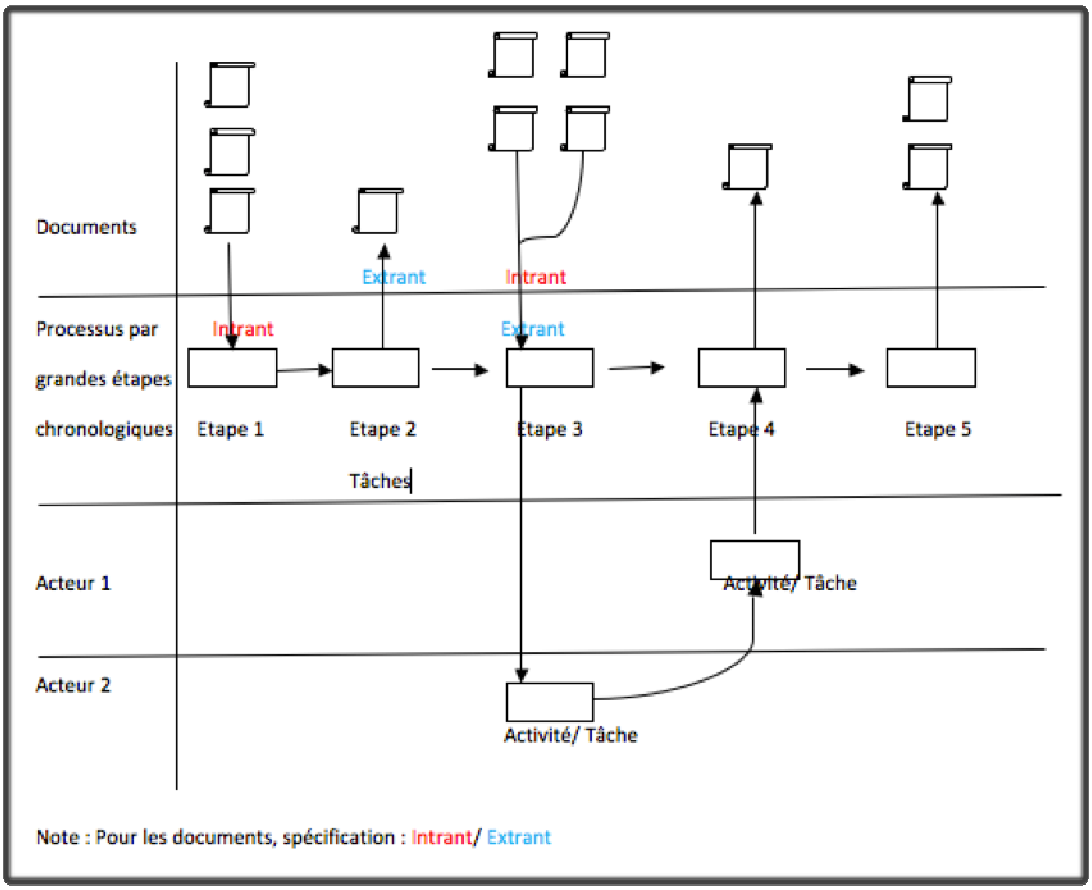


Figure 4. Schéma théorique de modélisation d'un processus

Le schéma suivant présente un exemple de modèle de processus pour une activité de développement de logiciel.

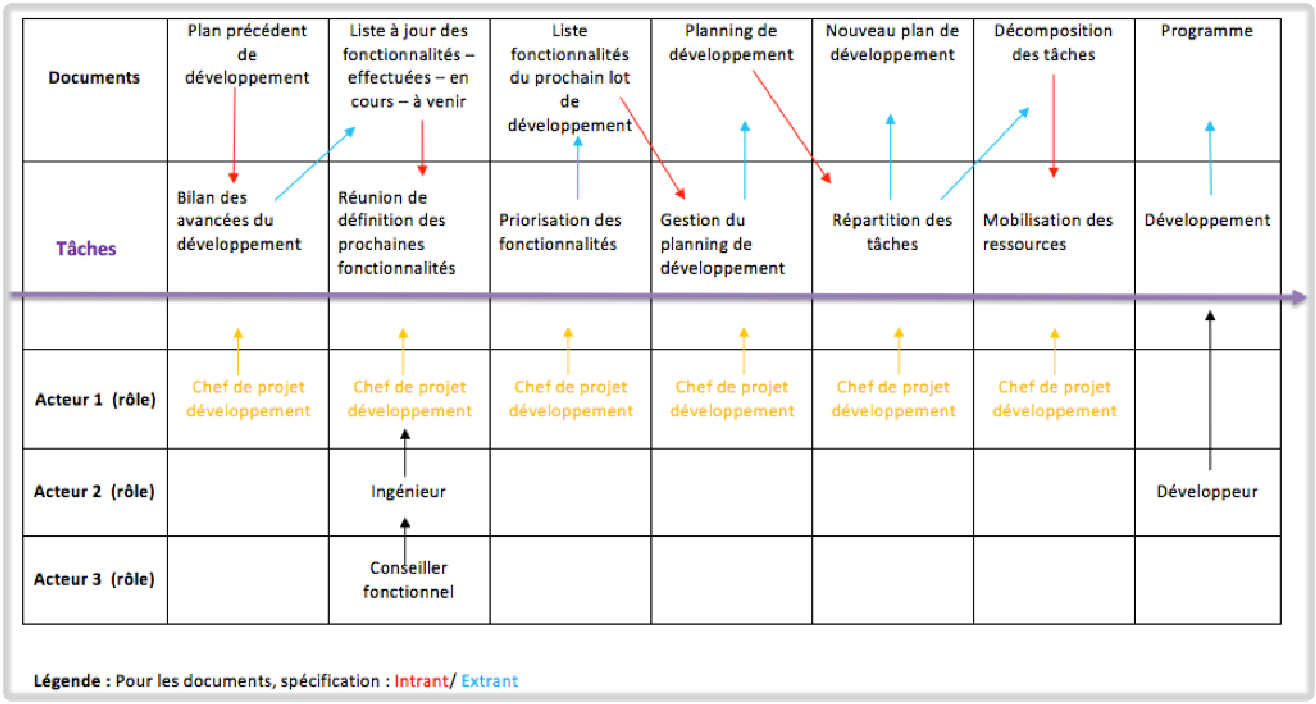


Figure 5. Représentation d'un modèle de processus d'une activité de développement de logiciel

2.4.2.11 Décomposition fonctionnelle et modèle de Processus d'Affaires

Globalement, la décomposition fonctionnelle et l'analyse des tâches déterminent les rôles. Les tâches sont dégagées dans les processus que l'on commence à identifier par l'analyse des référentiels disponibles (organigrammes, listes de tâches, fiches de poste…). L'articulation entre ces deux techniques d'analyse se construit en commençant par la décomposition fonctionnelle qui permet d'identifier les grandes activités, les identifiants et les noms des personnes intervenant dans le périmètre concerné par l'étude. De ces éléments, on déduit les sous-activités, dont on déduit ensuite les processus à étudier dans l'analyse par processus d'affaires.

D'après les concepteurs de GenISIS, l'approche top down (descendante) correspond à l'analyse des fonctions d'affaires, et l'approche bottom-up (ascendante) à l'analyse des processus. Les fonctions, qui décrivent les missions d'une organisation, peuvent être décomposées en processus qui décrivent comment elles accomplissent ces missions. Les fonctions peuvent ainsi être divisées en sous-fonctions ou en processus. Un processus peut être divisé seulement en sous-processus. Si une activité peut être réalisée, il s'agit d'un processus, sinon c'est une fonction. Par exemple, on ne peut pas réaliser une fonction de Marketing, mais on peut accomplir le processus de contacter un client. Ainsi, deux

perspectives se croisent pour dégager les documents à valeur opérationnelle dans le contexte d'ISIS, comme illustré dans la figure suivante :

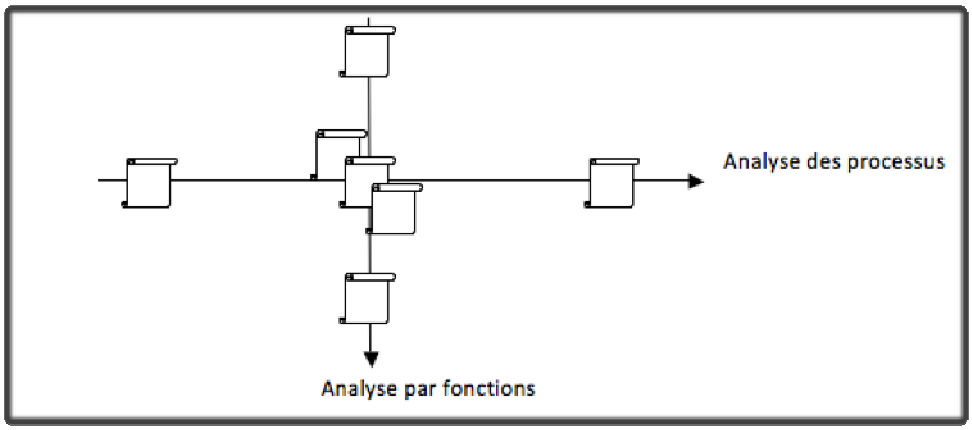


Figure 6. Illustration du croisement des approches Top-Down et Bottom-Up

2.4.3 Les concepts génériques dans la structure ISIS

L'objectif d'ISIS n'est pas l'analyse de processus, mais bien de dégager les documents à valeur opérationnelle (les records) qui doivent être archivés, conservés et accessibles. Dans la perspective de l'archivage, l'approche par le contexte de création du document est privilégiée par ISIS. Ainsi, c'est à la création ou à la réception du document que sa nature de Record ou non sera déterminée.

Le mode de collecte des informations est déterminant dans le processus de création du modèle. La méthodologie GenISIS est donc très détaillée sur cet aspect. Avant d'aborder cette partie de la méthodologie, nous allons nous intéresser aux différents concepts portés par la méthodologie GenISIS et l'application ISIS, puis nous présenterons le gabarit du modèle de facettes à partir duquel on construit un modèle de facettes adapté à chaque organisation. Enfin, nous présenterons le déroulé de la collecte d'information et de la construction du modèle adapté, et nous ferons quelques observations sur cette méthodologie.

Une hypothèse générale dans ISIS est que les ressources d'information ou les types de ressources d'information identifiés comme étant les principaux inputs (intrants) et outputs (extrants) dans les processus auront une valeur opérationnelle. Ces types de ressources d'information sont identifiés durant la création du modèle ISIS pour l'organisation.

2.4.3.1 Les Facettes

Ce sont les éléments sémantiques dans ce modèle GenISIS. Elles sont organisées hiérarchiquement, comme chez Ranganathan, et une partie de l'ensemble des facettes du modèle se réfère au contenu du document. Cette partie du modèle est donc comparable aux classifications à facettes de Ranganathan et du CRG, précédemment évoquées dans cette partie, qui classent les documents par leur sujet. Les caractéristiques des facettes sont les suivantes :

- chaque facette a un nom associé et un jeu de valeurs définies ;
- les jeux de valeurs sont organisés en hiérarchie dans chaque facette ;
- dans chaque facette particulière, certaines valeurs sont sélectionnables, et d'autres non : les « valeurs-rallonges », placées à la fin de la liste de valeurs seront toujours disponibles à la sélection, et celle en haut de la liste font l'objet de contraintes par le jeu de relations établies entre les valeurs de facettes pour encadrer leur usage (leur disponibilité à la sélection).

2.4.3.2 Les types de facettes

Il existe, dans le modèle ISIS, plusieurs types de facettes : les facettes de contenu, les facettes de contexte, les facettes Entreprise, et les facettes spécifiques. Nous présentons ici chacun de ces types de facettes.

2.4.3.2.1 Distinction entre Facettes de Contenu et Facettes de Contexte

Les Facettes de contenu constituent un sous-ensemble des facettes du modèle ISIS ; elles représentent le sujet du document, et elles présentent les métadonnées de contenu.

Les Facettes de contexte présentent des métadonnées de contexte, elles servent à capturer des informations sur le contexte dans lequel le document est créé. Une majorité de ces facettes est liée aux activités qui sont menées pendant la création du document, ou bien motivant la création du document. Il s'agit donc ici d'une conception du contexte essentiellement basée sur les activités distinguées pendant l'analyse par décomposition fonctionnelle et par processus d'affaires. Cette analyse est considérée comme produisant un cadre stable, puisque d'après les concepteurs de GenISIS, les fonctions de l'organisation changent moins que d'autres aspects de l'organisation. Nous revenons sur ce point dans notre partie sur les observations à propos de la méthodologie ISIS (partie 2.4.7).

Les facettes de contexte ne relèvent pas du contenu du document. La plupart des concepts de contexte pourraient être réinterprétés comme des concepts liés au contenu. Ainsi, pour une facette Temps, les valeurs de cette facette pourraient concerner la date de création ou de modification du document, tandis qu'elles peuvent aussi se rapporter à la date ou la période évoquée dans le document. A cause de cette ambiguïté entre contenu et contexte pour certains aspects de la description par facettes, il est important de savoir quelles facettes traitent le contenu et le contexte.

Le troisième type de métadonnées, évoqué en partie 2.4.2.3, qui concerne la structure du document, vu comme un contenant d'information, et qui peut concerner le format de fichier, ou la taille du fichier, par exemple, n'est pas traité par ISIS. Il est mentionné que ces métadonnées structurelles sont attachées aux documents par d'autres systèmes comme un système de gestion de fichiers, ou l'application de création de la ressource.

2.4.3.2.2 Facettes Entreprise et facettes spécifiques

Dans un modèle ISIS spécifique à une organisation, certaines facettes du gabarit de modèle sont plus ou moins pertinentes pour les différents composants de l'organisation.

Les facettes Entreprise

Ce sont des facettes proposées pour l'ensemble des personnes travaillant dans l'organisation. Si elles sont conservées durant la transformation du gabarit en modèle spécifique, cela implique qu'elles sont pertinentes pour chaque personne de l'organisation. Elles sont aussi appelées Facette- Noyau (*Core Facet*).

Les facettes Spécifiques

Elles ne sont pertinentes que pour un groupe de l'organisation ou pour certains groupes. Ces facettes sont rendues disponibles aux utilisateurs pertinents en fonction de la sélection de valeur effectuée dans les facettes Entreprise. Par exemple, un numéro d'employé est utile au service des Ressources Humaines et au service Finances, mais pas au département Qualité. Si un employé sélectionne une valeur qui indique qu'il appartient au service des Ressources Humaines il aura alors accès à la facette « Numéro d'employé ».

Les facettes de contenu sont généralement des facettes spécifiques et elles n'apparaissent donc que pour certains groupes d'utilisateurs. Elles peuvent être conçues sous forme de facettes Entreprise si elles sont pertinentes pour un large pourcentage d'utilisateurs dans l'organisation. Les facettes de contexte sont toujours des Facettes Entreprise.

2.4.3.2.3 Les relations

Une instanciation du modèle ISIS adapté à une organisation particulière, représente le contexte et le contenu des ressources d'information de l'organisation. Le modèle est une « coquille vide » à remplir avec les valeurs adaptées au contexte de l'organisation. Le modèle fait des références directes ou indirectes à beaucoup d'entités (clients, employés, projets) et aux attributs de ces entités (nom, lieu, sujet) qui sont pertinents pour l'organisation.

Le modèle ISIS est lié de manière très étroite au modèle de processus, et une majorité des valeurs de facettes peut être dérivée de ce modèle de processus. Comme le modèle représente beaucoup d'entités de l'organisation, au niveau global, il contient beaucoup de facettes et de valeurs. Les relations sont donc très utiles pour simplifier l'utilisation du logiciel, mais elles compliquent la création du modèle.

Les relations sont spécifiées entre les valeurs de facettes, comme on spécifie des relations dans un modèle de processus. Les combinaisons de plusieurs valeurs de facettes sont considérées comme valides lorsqu'elles représentent un fait à propos de l'organisation. Il y a aussi des combinaisons invalides, lorsqu'une valeur est associée à une autre, et que la combinaison est possible mais fausse, au sens logique. On ne peut pas choisir plusieurs valeurs d'une même facette pour décrire un document.

2.4.3.2.4 Thésaurus

Ce service donne la possibilité de proposer des valeurs de facettes additionnelles, équivalentes ou associées à une valeur de facette particulière dans le modèle. On peut donc intégrer deux mots synonymes dans le modèle. Un des termes est choisi comme le terme principal et les valeurs additionnelles sont associées avec cette valeur. On peut aussi faire des jeux de valeurs selon les langues : « valeurs en français » ou « valeurs en anglais » et il est ensuite possible de choisir d'afficher les valeurs, par défaut, dans un jeu particulier correspondant à une langue.

2.4.3.2.5 Conservation, élimination et valeur opérationnelle des documents

Le programme de conservation et d'élimination sert à déterminer combien de temps conserver une ressource d'information et à savoir comment la traiter, une fois que la durée de conservation est terminée. Dans ISIS, la période de conservation et les spécifications d'élimination sont déterminées par les métadonnées associées aux ressources d'information. Le schéma de métadonnées est dérivé du modèle ISIS. Ainsi, les valeurs de facettes constituent les métadonnées. Les métadonnées déterminent la valeur opérationnelle des documents, qui à son tour détermine quels documents et quelles catégories de documents sont

concernées par les problématiques de conservation et d'élimination, et quelles règles de conservation et d'élimination doivent leur être appliquées.

2.4.4 Règles pour guider la création des facettes

Le nom de la facette : il doit être unique, cela signifie que deux facettes ne peuvent pas avoir le même intitulé.

Les valeurs de facettes : à l'intérieur d'une facette, le nom d'une valeur doit être unique et significatif, cela signifie qu'il ne peut pas y avoir deux mêmes valeurs dans une facette.

L'organisation des facettes : une facette ne doit pas avoir plus de trois niveaux de hiérarchie.

Les valeurs sélectionnables : les valeurs en haut de la hiérarchie sont généralement celles qui font l'objet du jeu de contraintes et de relations entre valeurs de facettes, les « valeurs-rallonges » sont placées à la fin de la liste de valeurs d'une facette et sont toujours sélectionnables.

Auto-complétion des facettes : la règle générale édictée est que dans 75 % du temps, les utilisateurs ne doivent pas avoir à renseigner plus de cinq facettes manuellement pour un document.

Nombre de facettes : la règle édictée dans GenISIS, limite à quinze le nombre de facettes en tout, dix facettes Entreprise, et cinq facettes spécifiques.

Taille des facettes : si elles contiennent trop de valeurs, il faut découper les facettes de contenu en multiples facettes de contenu plus économes.

Efficacité des valeurs de facettes : les facettes doivent inclure seulement les valeurs utiles aux personnes pour décrire et retrouver leurs documents.

Séparation des valeurs de facette : si deux valeurs peuvent être appliquées à un même document, alors elles doivent provenir de facettes distinctes, cela signifie qu'on ne peut appliquer deux valeurs provenant d'une même facette à un même document.

Pertinence sémantique : les valeurs de facettes rassemblées dans une facette doivent être en lien au niveau sémantique avec le nom de la facette

Facettes de contexte non Entreprise : les facettes de contexte qui ne seraient pas des facettes-Noyau doivent être cachées à l'utilisateur et leurs valeurs doivent être dérivées automatiquement grâce aux relations entre facettes. Cela signifie que l'on peut ajouter des valeurs automatiquement, de manière transparente pour l'utilisateur, systématiquement en fonction d'autres valeurs sélectionnées dans d'autres facettes.

2.4.5 Le modèle de facettes GenISIS

La méthodologie GenISIS propose un jeu de facettes-Noyau qui constitue le modèle de facettes proposé, le gabarit du modèle, qui est considéré comme pertinent pour toutes les organisations, avec des adaptations selon les besoins des organisations. En effet, des facettes additionnelles peuvent être ajoutées à ce schéma de base pour refléter les besoins particuliers d'une organisation. Le modèle est composé de facette de contenu et de facettes de contexte.

2.4.5.1 Les facettes de contenu du modèle

Elles contiennent des données à propos des informations contenues dans les documents. Le processus de recherche associé s'apparente alors à chercher des documents par leur sujet, comme dans une bibliothèque. Rappelons que d'après Ranganathan, un sujet peut être décrit selon plusieurs points de vue. Le modèle contient trois facettes-Noyau de contenu pour décrire le sujet d'un document, même si elles ne sont pas toutes pertinentes pour chaque ressource. Ce sont les suivantes :

La Facette Lieu = elle décrit le lieu évoqué dans le document ;

La Facette Temps = elle décrit l'aspect temporel évoqué dans le document ;

La Facette Thème = elle décrit le sujet du document.

Les deux premières facettes de contenu sont identiques aux facettes de Temps et de Lieu du Modèle PMEST, de Ranganathan. Deux autres facettes de contenu qui ne sont pas techniquement des Facettes-Noyau mais qui peuvent être pertinentes pour une organisation ou une partie de cette organisation sont proposées :

La Facette Personne: elle est intégrée si le sujet contenu dans le document est à propos d'une personne ;

La Facette Evénement : elle est utilisée si le sujet du document est à propos d'un événement, par exemple : les Jeux Olympiques de Pékin.

2.4.5.2 Les facettes de contexte du modèle

Elles contiennent des données à propos du contexte de création ou d'utilisation des documents. Dans ce type de classification, les objectifs de haut niveau et les fonctions sont souvent décomposés selon un grain de plus en plus fin, jusqu'à ce qu'ils englobent toutes les activités ou processus, réalisés dans leur ensemble.

La Facette Fonction : elle correspond à une hiérarchie de fonctions et sous fonctions de l'organisation ;

La Facette Activité : elle propose une description générique des activités qui représente les tâches ou activités impliquées par une fonction ou une sous-fonction.

Ces facettes Fonction et Activité sont déduites de la décomposition fonctionnelle. Ce sont des valeurs clés, car elles déterminent les règles de conservation et de suppression des records associés.

La Facette Mes tâches : Elle présente la liste des processus décomposés avec un degré de finesse élevé, décrivant les processus dans le langage des personnes qui en sont en charge. Pour cette facette, s'il n'y a pas de documents produits dans une tâche, elle n'est alors pas mentionnée. La problématique du niveau de granularité des processus est cruciale, et déterminée par la valeur opérationnelle ou non des documents produits dans le cadre d'une tâche. Cette facette « Mes tâches » a été créée pour compenser la difficulté des personnes qui travaillent dans l'organisation à se situer dans la perspective de l'organisation à haut niveau comme dans la classification fonctionnelle. Elle représente les Processus d'affaires de l'organisation, du point de vue des acteurs qui les réalisent.

Facette Type de contenu : elle présente le type de la ressource d'information utilisée, c'est-à-dire le type de document ; il peut y avoir des types génériques et des types spécifiques, identifiés et nommés conventionnellement ;

Facette Rôle : elle représente le rôle ou le poste associé avec la personne qui crée la ressource. Dans certaines organisations, une personne peut avoir plusieurs rôles, par exemple : chef de projet sur un projet, et contributeur ou consultant sur d'autres projets ;

Facette Objectif du document : la définition des valeurs de cette facette est en lien avec la théorie linguistique des actes de langage ; elle décrit l'acte de communication majeur pour lequel un document est créé, comme par exemple : poser une question, faire une requête, ou demande, fournir une information, faire une recommandation, obtenir un agrément, appliquer une directive ;

Facette Organisation du travail : cette facette est considérée comme une facette-relais. Elle réfère à la manière particulière dont une organisation organise son travail et ses employés. L'exemple principal est l'organisation par projet mais il peut se produire que les organisations soient agencées selon les cas qu'elles traitent, ou sous forme de comités. Il s'agit ici d'un type spécial de facette de contexte. C'est une facette-relais car elle connecte les facettes de contenu et de contexte du point de vue des relations entre facettes.

2.4.5.3 Autres facettes de contexte

Facette Nom : elle présente le nom de la personne qui crée la ressource ;
Facette Position : elle mentionne le titre du poste de la personne désignée par la Facette Nom ;
Facette Structure organisationnelle : elle reflète la position occupée dans la structure organisationnelle, de la valeur sélectionnée dans la Facette Position.

Les Relations-Noyau dans le modèle

- La facette Rôle détermine la facette Position ;
- La facette Positon détermine la facette Structure organisationnelle ;
- Il y a un lien unitaire entre les Types de contenu spécifiques et la facette Mes tâches ;
- Il y a une connexion entre la facette Type de contenu et la facette Objectif du document ;

- Chaque valeur de la facette Mes tâches est reliée à une valeur de la facette Activité et de la facette Fonction ;
- La facette Nom est connectée à la facette Position ;
- La facette Position est connecté la Facette-Relai (Organisation du travail) ;
- La Facette-Relai est connectée aux Facettes de Contenu.

2.4.5.4 Précisions sur les facettes du modèle ISIS

Il existe des ambiguïtés entre les facettes Rôle et Position. En effet, les différences entre celles-ci sont difficiles à distinguer. Le Rôle est défini par les tâches, qui peuvent être équivalentes pour un rôle donné, et par les documents à valeur opérationnelle produits de façon équivalente dans chaque rôle. C'est par l'analyse décrite dans la suite de cette partie que l'on détermine les rôles. La position, quant à elle, correspond, la plupart du temps à un numéro d'identification, elle peut être complémentée du titre du poste occupé qui est considéré comme pérenne, il reste le même si une personne évolue et change de poste dans l'organisation. Ainsi, si Grégory et Annette occupent tous deux le rôle d'adjoint administratif, ils auront une Position qui sera respectivement, « adm 1550 », et « adm 2140 ».

Concernant les facettes « Objectif du document » et « Type de contenu », pour chaque processus, il faut essayer de trouver des valeurs pour la facette « Objectif du document », pour éviter de créer trop de valeur pour la facette « Type de contenu », qui sont considérées comme peu représentatives de la valeur opérationnelle.

Pour la facette de contenu « Sujet », une liste de sujets est demandée aux acteurs sur le terrain, et l'on peut aussi s'aider des éventuels thésaurus existants, ou de programmes de Traitement Automatique du Langage (TAL) pour les faire émerger.

Le schéma suivant présente l'organisation des facettes dans le modèle ISIS.

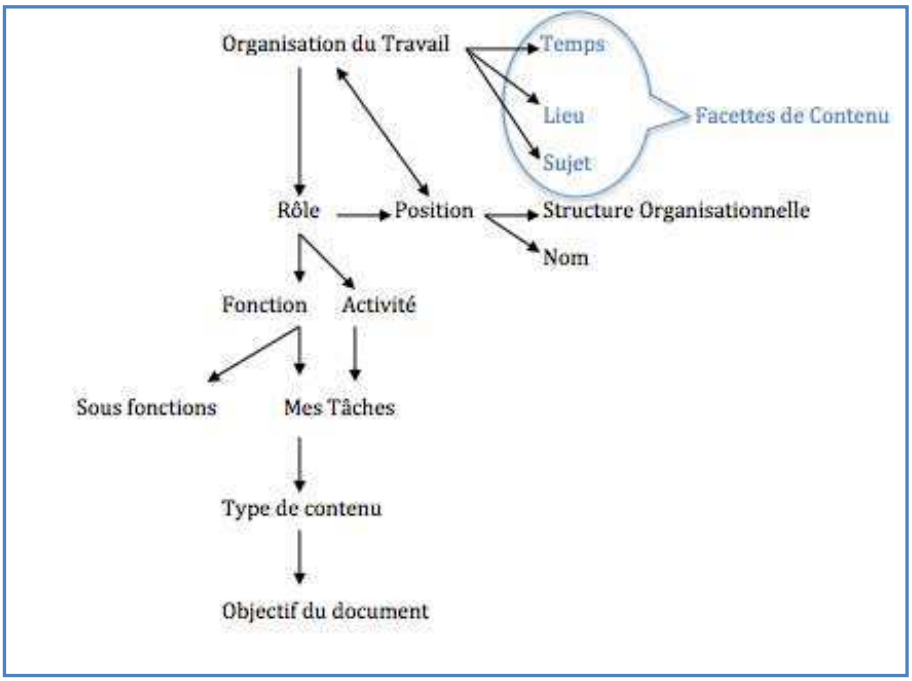


Figure 7. Organisation des facettes dans le gabarit du modèle ISIS

2.4.6 La méthodologie GenISIS de construction du schéma de facettes

Le processus général de création du modèle s'effectue en deux phases, une phase d'analyse initiale, et une phase d'analyse détaillée. Chacune de ces phases produit des supports d'information intermédiaires qui ne sont pas des parties du modèle final mais qui contiennent les informations nécessaires pour la création du modèle final.

Le résultat principal de la première phase est la décomposition fonctionnelle, une représentation fonctionnelle de l'organisation. Le produit principal de la deuxième phase est le modèle de processus de l'organisation. Le modèle de processus est une description active du fonctionnement de l'organisation. Il connecte les objets clé et les rôles avec les actions et fonctions et considère les flux d'information à travers l'organisation. Le degré de détail procuré dépend de l'organisation.

2.4.6.1 Les Ressources

Des ressources en termes de temps et de personnes sont nécessaires, pour le succès du projet, de la part de l'organisation, et de la part des intervenants qui créent le modèle ISIS.

Du coté de l'organisation, il est nécessaire de fournir les ressources suivantes :

- des employés clé qui assistent aux réunions et aux interviews

- des collecteurs d'information qui vont récolter des informations dans l'organisation ou qui vont demander à d'autres personnes de les récolter pour eux
- des employés qui vont remplir des questionnaires en ligne et qui vont assister à des workshops

Les engagements du coté des concepteurs sont les suivants :

- conduire des Interviews
- conduire des réunions et des sessions d'information
- planifier et conduire les workshops
- fournir des questionnaires
- traiter et analyser l'information
- produire des résultats intermédiaires

2.4.6.2 Les étapes de la première phase : l'analyse initiale

En général l'analyste commence par rassembler les données, en faisant des Interviews, en prenant des notes, et en enregistrant les interactions. L'étape suivante est la collecte et l'analyse de ces informations, qui transforment ces données brutes en données intermédiaires. Ces structures de données sont ensuite transmises aux concepteurs qui créent le modèle pour encoder les facettes, les valeurs, l'organisation des facettes et les relations entre les valeurs. L'analyse initiale est décrite ci-dessous succinctement :

- Session d'information initiale et Interviews de prise de contact
- Création du comité de direction du projet (soutien structurel pour la gestion du projet)
- Définition des termes de référence pour le comité de direction
- Collecte préliminaire d'information
- Collecte d'information initiales (contenu de l'intranet, définition des missions, mandats, organigramme)
- Analyse de l'information initiale
- Interviews de Haut niveau
- Trouver des personnes pour ces Interviews
- Conduite des Interviews
- Traitement et analyse de ces Interviews
- Création de la décomposition fonctionnelle
- Plan du projet
- Développement du Plan du projet
- Validation du plan du projet par le comité
- Sélection des groupes pour la deuxième phase

2.4.6.3 Les étapes détaillées de la deuxième phase : l'analyse détaillée

Pour chaque groupe sélectionné dans la fin de la première phase, le cycle de collecte d'information est le suivant

- Interviews de niveau intermédiaire
- Workshops
- Questionnaires en ligne
- Création du modèle de processus
- Créer un modèle de processus pour chaque groupe
    - Identifier la valeur opérationnelle
    - Identifier les périodes de conservation
- Modèle ISIS sur papier

2.4.6.3.1 Les Interviews de haut niveau

Dans ce type d'interview les personnes interrogées sont censés avoir une vision globale, une vision d'ensemble, elles peuvent ne pas connaître tous les détails mais doivent avoir un sens des différentes parties et de la manière dont elles sont connectées. Il s'agit de poser des questions particulières dont on va pouvoir extraire les informations requises pour la création de la décomposition fonctionnelle, et pour commencer à remplir le modèle de processus. Comme il y a plusieurs Interviews le processus est cumulatif. Le but global est d'obtenir les informations requises pour la fin du processus.

Traitement et analyse des données des interviews de haut niveau

A l'issue de cette série d'interviews on possède des enregistrements et des notes. L'étape suivante consiste à produire des données intermédiaires, à partir de ces données brutes qui seront complétées après, grâce aux interviews de niveau intermédiaire, et les workshops.

Les produits de données intermédiaires de la phase de l'analyse initiale

Suite au traitement des données récoltées pendant la phase de l'analyse initiale, certains documents de données intermédiaires doivent être produits, qui sont les suivants :

- La Table des facettes **:** ce tableau contient une liste des facettes, incluant :

- les Facettes-Noyau
- les Facettes Entreprise non-noyau
- les Facettes spécifiques possibles
- avec leur intitulé préféré
- des définitions
- des exemples
- pour chaque facette, il peut y avoir des notes additionnelles
- des informations pour le module de thésaurus, avec des valeurs possibles

– la décomposition fonctionnelle : ce schéma de décomposition fonctionnelle avec des notes sur les contacts potentiellement intéressants pour la suite de l'étude doit être réalisé

– le modèle de la Table des processus : elle va être utilisée pour capturer les informations des futures interviews de niveau intermédiaire et des workshops. A cette étape, il reste des zones d'incertitudes, avec des éléments à vérifier comme les noms des facettes.

– La Liste de documents : une liste des documents contenant des informations pertinentes sur l'organisation, et qu'il faut obtenir, doit être dressée.

– Le rapport : il s'agit d'un rapport résumant les résultats, et contenant toute information additionnelle jugée pertinente.

– Les fichiers audio et notes d'interviews : elles doivent être fournies aux autres personnes qui travaillent sur le projet.

– Transfert d'information entre les interviews de haut niveau et celles de niveau intermédiaire : dans certains cas, si ce n'est pas la même personne qui fait les deux séries d'interviews, il faut veiller à transmettre les produits intermédiaires de l'étude.

2.4.6.3.2 Les interviews de niveau intermédiaire

A cette étape, l'organisation est divisée en groupes plus petits, et l'analyste doit obtenir une image plus précise des processus d'affaires pour chaque groupe. L'objectif est de continuer à raffiner et remplir les vides dans les informations récoltées pour chaque groupe. A la fin de

cette étape, les interviews de niveau intermédiaire, les workshops et les questionnaires aboutissent à l'élaboration du modèle de processus d'affaires, à un tableau des processus et aux autres produits intermédiaires complets pour chaque groupe. L'analyste va ensuite combiner les éléments spécifiques de chaque groupe, et construire le modèle ISIS final. Elles doivent être menées avec des personnes qui ont une vision d'ensemble du groupe étudié. L'objectif est de raffiner et compléter les informations déjà récoltées, en posant les mêmes questions que durant les autres interviews mais à un niveau différent, pour obtenir une vision plus détaillée des processus et des flux d'informations de ce groupe, ainsi que des informations supplémentaires correspondantes sur les valeurs de facettes et les relations. On développe à ce stade les modèles des processus d'affaires pour chaque groupe étudié, et pour ce faire, il est possible d'utiliser la Méthode H, présentée en partie 2.4.6.3.5.

Traitement et analyse des données des interviews de niveau intermédiaire

Les informations qui doivent être obtenues pendant cette étape :

- Une décomposition des facettes complète et un modèle de processus pour le groupe contenant les inputs, les outputs et les types de contenus.
- Une liste de Facettes spécifiques pour le groupe (avec celle du modèle ISIS, et celles spécifiques à l'organisation).
- Les valeurs pour les Facettes Entreprise et spécifiques qui sont pertinentes pour le groupe ainsi que les valeurs éventuelles pour le module de thésaurus.
- L'organisation spécifique des facettes pour le groupe.
- La logique spécifique des facettes, avec les indices qui doivent être étudiés pour révéler les facettes spécifiques
- Les relations entre valeurs de facettes
- Les valeurs opérationnelles qui constituent les informations concernant l'archivage
- Un rapport qui résume les résultats

Modèle de processus et table des processus

Ce sont les principaux résultats en termes de données intermédiaires. Le schéma du modèle de processus décrit en une forme visuelle les processus impliqués dans le groupe de travail. A ce stade il faut croiser les résultats de cette analyse avec la décomposition fonctionnelle issue des interviews de haut niveau.

La table de processus contient la majorité des informations nécessaires aux concepteurs du modèle. C'est un tableau, la première colonne contient la liste des processus du groupe, tandis que les autres colonnes donnent des détails additionnels pour chaque processus mentionné. Les colonnes correspondent aux facettes du modèle ISIS. Pour chaque processus identifié, il faut entrer une liste des inputs et outputs et signaler par un symbole ceux qui auront une valeur opérationnelle.

Les autres produits d'information intermédiaires

La Table des nœuds de facettes : c'est une table listant chaque nœud de facettes, une description du nœud, les nœuds père et fils potentiels de chaque nœud, les termes liés sémantiquement et les termes préférés. Ce sont des tableaux supplémentaires, qui organisent et définissent les relations entre les valeurs. Pour obtenir les nœuds des valeurs de facettes (leur hiérarchisation), on utilise soit la table des processus et le modèle de processus, soit on utilise les résultats des workshops et des questionnaires. Avant ces workshops et questionnaires, il faut renseigner le plus possible ces tableaux en se basant sur les interviews pour pouvoir repérer les vides et les combler à l'aide des deux dernières étapes.

Table des Positions/ Noms/ Rôles : c'est une table qui liste pour chaque position, le nom de la personne qui l'occupe, et le rôle de cette personne.

Table de l'organisation du travail : table qui liste les projets en cours dans le groupe et les personnes qui travaillent pour ces projets.

Table du contenu de l'organisation du travail : une table qui liste tous les thèmes pertinents pour chaque valeur de la table précédente.

2.4.6.3.3 Les *Workshops*

Comme il n'est pas possible d'interviewer toutes les personnes de l'organisation et de leur présenter tous les résultats et analyses, le workshop est utile pour vérifier qu'aucun aspect de l'organisation n'a été ignoré et que les personnes dans l'organisation sont bien représentées dans les informations collectées pendant les interviews. Les personnes présentes sont celles recommandées pendant les interviews de niveau intermédiaire.

Le but des *workshops* est de s'assurer que les activités de travail sont bien décrites par les informations collectées. Le *workshop* est orienté, à travers la présentation de la décomposition fonctionnelle et des modèles de processus, sur la vérification des informations récoltées, et sur la complétion des informations manquantes ou erronées. Diverses stratégies pour mener les workshops sont abordées dans la méthodologie ISIS.

2.4.6.3.4 Les questionnaires

Après les *workshops*, les personnes présentes à ces sessions sont sollicitées pour remplir un questionnaire. Le but est d'obtenir des valeurs détaillées pour chaque facette et des relations entre ces valeurs ainsi que de finir de récolter les détails qui auraient pu être manqués. Les valeurs spécifiques de facettes qui sont difficiles à obtenir par la technique des *workshops* doivent être collectées grâce à cet outil, et les noms exacts des documents à consulter, éventuellement manquant, aussi. Un modèle de questionnaire existe et doit être adapté.

2.4.6.3.5 Méthode H

Développée par Project Perfect, c'est une stratégie simple pour obtenir des informations de la part des personnes pour faire une analyse des processus d'affaires. Elle passe en revue les *inputs, outputs*, fonctionnalités, règles des affaires et les données impliquées dans le travail d'une personne ou d'un groupe, permettant à l'analyste de développer une compréhension des processus et fonctions impliqués dans le travail. Cette représentation peut ensuite être transformée en modèle de processus d'affaires.

Les composants de la Méthode H sont les suivants :

- Inputs
- Fonctionnalités
- Règles des affaires
- Données
- Outputs

D'après cette méthode, on peut modéliser un processus de la façon suivante :

- Un individu reçoit des *inputs*,
- Il les utilise pour exercer ses fonctions dans le travail
- En suivant les règles
- En utilisant les données existantes
- Pendant ce processus, il créé des *outputs*
- Qui à leur tour deviennent des inputs pour les autres individus

Le schéma suivant illustre cette modélisation et ses composants.

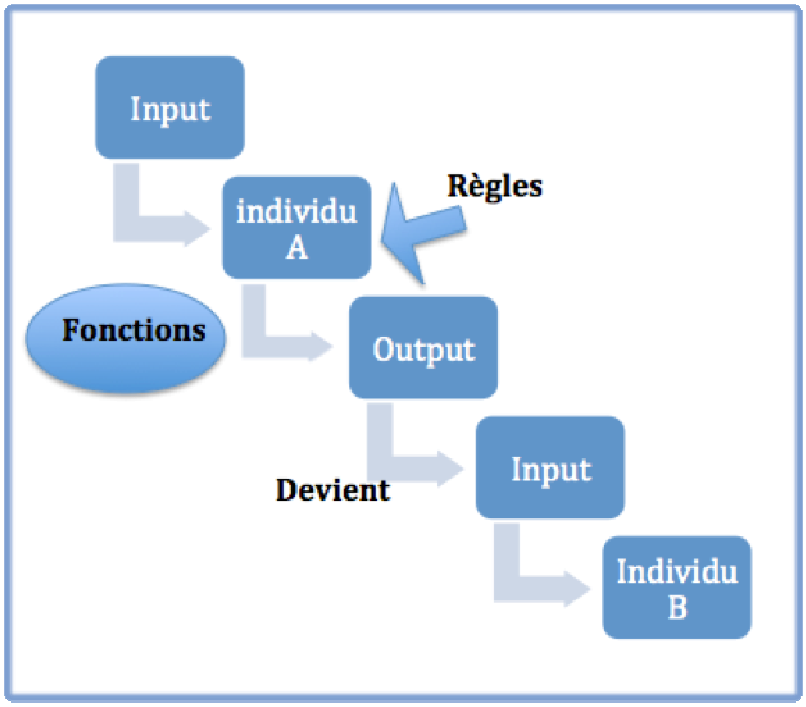


Figure 8. Schéma des composants de la méthode H

2.4.7 Observations sur la méthodologie GenISIS

D'un point de vue global la méthodologie GenISIS est plus orientée conception que les autres méthodologies évoquées. Elle nous paraît plus opérationnelle. Elle est très structurée et basée sur l'explicitation de ses notions clés. Elle vise un objectif de certification ISIS pour les consultants.

Les objectifs de GenISIS sont clairement orientés vers le Record Management. Elle vise à déterminer la valeur opérationnelle des documents qui à son tour détermine quels documents et quelles catégories de documents sont concernées par les problématiques de conservation, et quelles règles de conservation et d'élimination doivent leur être appliquées. Ceci implique qu'elle cherche un modèle économe, et qu'elle fixe un nombre de facettes limité à 15 facettes en tout.

Isis est marquée, du point de vue de sa conception, par le fait qu'elle a été conçue pour des organisations gouvernementales, cadrées, où les fonctions et les rôles sont bien définis, et où les procédures sont rigoureuses. Il se peut que dans certaines entreprises, les décompositions fonctionnelles soient plus difficiles à réaliser, plus délicates en termes d'enjeux liés aux jeux d'acteurs. Ainsi nous pouvons poser la question suivante : le coté très rigoureux de la méthode est-il assoupli dans son application à des organisations moins formelles que celles relevant d'un gouvernement, comme par exemple, pour des centres de recherche, des laboratoires, des groupes de travail, des groupes métier ou des communautés de pratiques ?

Ce modèle est globalement adapté pour des organisations structurées sur ce modèle constitué de services supports et de services opérationnels pour les missions principales de l'organisation. Ainsi, dans une organisation de taille plus petite, où plusieurs rôles ou missions peuvent être accomplies par le même service, comment le modèle peut-il s'adapter ?

D'autre part, comme le modèle doit être structurellement le même pour tous les services de l'entreprise, et présenter de manière exhaustive les listes de fonctions, il est conçu globalement, de manière top- down. La structure de l'application est la même pour tous les acteurs de l'organisation. Ainsi, le service rendu par ISIS n'est peu ou pas adaptable par les individus à leurs propres pratiques.

A propos de la décomposition fonctionnelle qui est fondée sur la stabilité des fonctions dans une organisation. Ce point nous semble discutable. En effet, dans de nombreuses organisations de grande taille, la structure des fonctions est remaniée dans une logique de rentabilisation des locaux ou du personnel et l'on assiste bien souvent à des fusions des services supports, à des redécoupages de la structure organisationnelle globale, à tel point que les acteurs peuvent en être désorientés. Cela peut aussi impacter la recherche d'information. Ainsi, d'après nos expérimentations, une recherche d'information peut être conduite d'après la structure organisationnelle en place au moment de l'élaboration du document recherché. Elle constitue plutôt, selon nous, un repère, ou un critère de recherche, même si elle est remaniée. Une application qui serait capable de proposer un historique de la structure

organisationnelle comme critère de recherche, éventuellement sous forme de cartographie, gagnerait à être exploitée dans ces organisations dont la structure des services est régulièrement modifiée

Globalement, l'approche de GenISIS est de concevoir le modèle de facettes d'une organisation depuis celui existant dans cette méthode. Nous sommes donc de nouveau dans un cas où l'on cherche à faire entrer des dimensions contextuelles de description des documents, propres à une organisation, dans un modèle préexistant.

De plus, l'adaptation relative du schéma de facettes à l'organisation est basée sur les modélisations issues de la décomposition fonctionnelle et des processus d'affaires. Il s'agit, selon nous, et comme nous le démontrons plus en détails dans le chapitre 3, d'une conception figée de l'activité, qui n'intègre pas le caractère dynamique et évolutif des activités réalisées dans une organisation. Le modèle des processus d'affaires sert de guide procédural et détaille chaque action à réaliser pour l'accomplissement de ce qui est appelé une activité. Mais, ce sont les aspects formels des activités qui sont représentés dans ce modèle, qui aboutissent donc à une standardisation de l'activité dans et par l'application, standardisation qui nous paraît pouvoir être rapprochée du fonctionnement induit par les PGI (Progiciel de Gestion Intégrés), que nous abordons dans le chapitre 3 également. Enfin, cette conception standardisée de l'activité nous semble révéler un décalage entre le travail réel et la vision de l'activité fournie par les concepteurs de ces activités, qui sont les premières sources interviewées chronologiquement dans GenISIS dans les interviews de haut niveau (voir partie 2.4.6.3.1).

Par ailleurs, la rigidité du système induit que certaines facettes ne peuvent suivre les évolutions des situations de travail. Ainsi en est-il de la facette « Position », qui mentionne le titre des postes occupés par les acteurs, qui peuvent être amenés à changer en fonction des évolutions des profils professionnels des acteurs. Les valeurs de la facette « Position » ne peuvent pas être modifiées. Ceci peut nous conduire à réfléchir à l'évolutivité limitée du schéma de classification en lui-même qui apparaît rigidifié par les relations entre les valeurs de facettes. Ainsi la maintenance du schéma de classification semble délicate, tandis que les relations entre valeurs de facettes limitent les possibilités de nouvelles combinaisons de valeurs et brident la flexibilité de système, flexibilité considérée comme un des avantages liés à l'usage de la classification à facettes. La présence des jeux de relations entre valeurs de facettes tend à faire ressembler le système ISIS plus à une ontologie qu'à une classification à facettes.

Ainsi, les valeurs portées par ISIS sont la simplicité d'utilisation, l'efficacité, et le contrôle, reposant sur une semi-automatisation, qui est vue comme un moyen d'allègement de la charge de travail. Cette approche pourrait être rapprochée des conceptions du courant de la cognition distribuée. (voir chapitre 3, partie 3.3.4). La semi-automatisation de la saisie des métadonnées, proposée dans ISIS, peut être vue d'une manière positive quand il s'agit d'appliquer des procédures formelles. Mais dans des situations moins formelles, ou moins routinières comme dans les processus d'innovation par exemple, ce procédé aurait, selon nous, tendance à brider ou à freiner les processus créatifs.

De plus, cette simplicité d'utilisation, basée sur semi-automatisation est en désaccord avec les approches de l'ergonomie qui fondent les progrès de l'efficacité au travail, en partie sur la responsabilisation des acteurs et sur les marges de manœuvres nécessaires et concédées pour intégrer l'imprévu, qui constitue une part non négligeable du travail. En effet, souvent, dans une organisation qui n'est pas gouvernementale, mais plutôt de type entreprise, il s'agit d'adapter les procédures aux cas particuliers rencontrés. La semi-automatisation prive les acteurs de ces marges de manœuvre et de leur autonomie, qui est elle-même perçue, dans le cadre du déploiement d'ISIS, comme génératrice d'erreurs, de la même manière que dans les approches proposées par les PGI, comme nous le mentionnons dans le chapitre 3.

2.5 Classifications à facettes pour les sites Web

En dehors des bibliothèques, des collections spécifiques, et de l'entreprise, un autre champ de déploiement de la classification à facette se manifeste sur le Web, pour l'organisation des informations contenues dans des sites web, à dominante commerciale. Ainsi on peut visiter les sites suivants pour se faire une idée de l'application de la classification à facettes au Web.

Un site qui propose des facettes pour caractériser les vins à des fins commerciales :
http://www.wine.com/v6/search/advancedsearch.aspx
Un site de vente en ligne, consulter par exemple la rubrique des chaussures pour femmes, qui propose des critères de filtrage que nous tenons pour des facettes : http://www.ebay.fr/
On remarquera que www.youtube.com, propose depuis peu, un filtrage des vidéos disponibles par facettes rigides dont une copie d'écran est reportée ci-dessous :

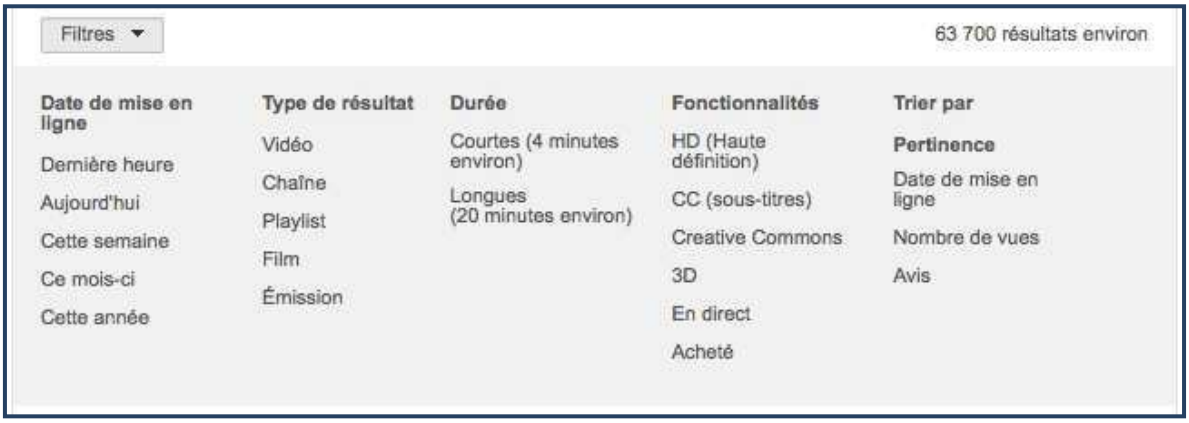


Figure 9. Facettes disponibles pour le filtrage des vidéos sur Youtube

Par ailleurs, www.google.com propose aussi des outils de filtrage, qui rappellent les facettes et dont on pourra avoir un aperçu dans la copie d'écran suivante, et qui sont accessible par le bouton « Outils de recherche » :

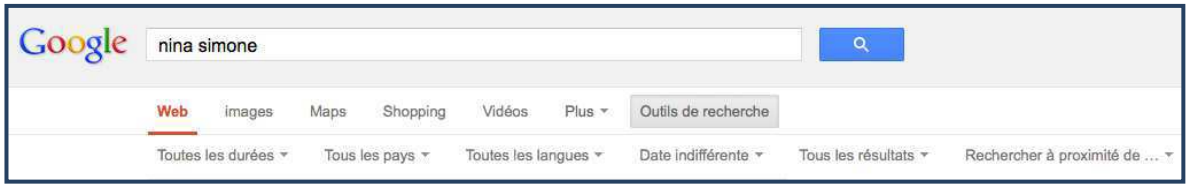


Figure 10. Facettes rigides pour filtrer les résultats de recherches sur Google

Il s'agit ici de facettes rigides qui portent sur des éléments formels des ressources. Ce que nous désignons comme des facettes rigides sont des facettes qui ne sont pas contextuelles, ce sont toujours les mêmes qui sont présentées, pour n'importe quelle requête dans ces deux outils de recherche généralistes.

Denton (2003) est un des seuls auteurs à notre connaissance, avec La Barre (2004, 2006) et Vickery (2008) à avoir approché la problématique de la classification à facettes pour les sites Web. La Barre (2006) a étudié en détails la construction des sites web, leur conception, les options de recherches et de navigation proposées, en faisant un travail conséquent de repérage de sites Web qui utilisent des formes plus ou moins recherchées de classifications à facettes ou admises comme telles. En effet, pour cette recherche, elle part du postulat que la tendance des concepteurs de sites Web est de construire des classifications à facettes qui ne sont pas complètes, qui n'ont pas de notation ni d'ordre de citation pour les facettes (La Barre, 2006,

p.51). Elle souligne que la classification à facettes et l'analyse par facettes relèvent pour l'une de ce que nous appelons ici un SOC et pour l'autre d'une technique d'analyse. Sur l'analyse par facettes, elle propose de se fier à ce que Vickery (1966) a défini et qu'elle adapte pour l'étude et la conception de sites Web. Nous reportons ici les propositions de Vickery (1966), tandis que l'étude de la conception de classifications pour les sites Web dépasse le cadre de ce travail.

Vickery, (1966) liste cinq questions pour déterminer des facettes :

- quelles sont les entités d'intérêt pour les utilisateurs ?
- quels aspects de ces entités les intéressent ?
- quel concept un terme représente-t-il ?
- dans quelle catégorie conceptuelle ce concept devrait-il être inclus ?
- quelles sont les relations de classe entre ce concept et les autres concepts dans la même catégorie ?

Vickery (2008), revient sur la première de ces questions, en affirmant que toute entité a un nombre indéterminé de caractéristiques, il repose la question de savoir lesquelles sont significatives et selon quel point de vue. Anciennement, la réponse était fournie par la caution bibliographique mais Vickery propose d'utiliser plutôt une approche « orientée-utilisateur ». Si une classification est conçue pour un groupe d'utilisateurs spécifiques, alors une étude de leurs centres d'intérêts, (comme par exemple à travers l'étude de leurs documents, ou de leurs requêtes dans un système de RI), devrait indiquer quels points de vue incorporer dans la classification.

Denton (2003) pose les principes d'élaboration de classification à facettes pour les contenus de sites Web, en se basant sur la méthodologie et le modèle simplifié proposé par Spiteri (1998), qui sont eux-mêmes fondés sur les méthodologies de Ranganathan et du CRG.

Nous présentons ici le modèle simplifié de Spiteri (1998), ensuite nous nous intéressons à celui de Denton (2003) en sept étapes. Ce dernier précise que les principes qu'il propose sont conçus pour des corpus de taille moyenne, ou petite, et que son modèle peut être utilisé pour les sites Web privés ou publiques.

2.5.1 Le modèle simplifié de Spiteri (1998)

Il est précisé par Spiteri (1998) que cette simplification du modèle initial est élaborée pour le rendre accessible aux étudiants.

2.5.1.1 Principes pour le choix des facettes

Elle conserve les premiers principes énoncés par Ranganathan pour le choix des facettes (voir partie 2.2.4.1.) :

Différenciation : les caractéristiques utilisées comme base de la classification doivent différencier certaines des entités de l'univers, et doivent aboutir à la formation d'au moins deux classes ou isolats distincts.

Pertinence : une caractéristique utilisée comme base pour une classification doit être pertinente pour l'objectif de la classification.

Vérifiabilité : une caractéristique doit être vérifiable. Par exemple, la date de naissance d'un auteur est vérifiable.

Persistance : une caractéristique doit rester inchangée tant que le but de la classification ne change pas. Ainsi, les facettes doivent refléter des propriétés permanentes des entités.

Dans cette catégorie qui concerne le choix des facettes, elle inclut aussi certains canons énoncés pour la succession des caractéristiques dans la méthodologie de Ranganathan, en les reformulant.

Au canon de *concomitance* qui énonce que deux caractéristiques ne doivent pas être concomitantes, elle préfère le *canon d'homogénéité* des facettes et celui *d'exclusivité mutuelle*. Elle scinde donc en deux principes le canon de concomitance énoncé par Ranganathan, et elle considère que ces deux principes procèdent au choix des facettes, plutôt qu'à l'élaboration de leur succession.

Principe de catégories fondamentales : Elle ajoute dans cette catégorie le principe du CRG qui est le principe des *catégories fondamentales* qui énonce qu'il n'existe pas de catégories fondamentales valables pour tous les sujets et que la nature des facettes doit être déterminée en fonction du domaine à classer.

2.5.1.2 La succession des facettes et des valeurs de facettes

Elle conserve les deux canons énoncés par Ranganathan pour la succession des caractéristiques, et elle préconise d'étendre leur application à la succession des valeurs de facettes à l'intérieur des facettes.

La succession pertinente : la succession des caractéristiques doit être pertinente en fonction des objectifs de la classification.

La cohérence : la succession des caractéristiques doit susciter l'approbation aussi longtemps qu'il n'y a pas de changement dans les objectifs de la classification, elle doit être durable pour la cohérence dans la durée de la classification.

De la même manière, les canons énoncés pour la succession des valeurs de facettes qui correspondent au canon pour des séquences efficaces (détaillé dans la partie 2.2.4.6) sont repris et affectés aussi aux séquences de facettes elles-mêmes. Il s'agit des principes suivants :

Principe chronologique : si les sujets d'un ensemble prennent leur origine dans des temps différents, ils doivent être rangés chronologiquement.

Principe du « dernier dans l'évolution » : si les sujets d'un ensemble appartiennent à des stades différents de l'évolution, ils doivent être rangés en fonction de cette évolution, par exemple, pour des grandes périodes historiques, on rangerait ces périodes dans l'ordre suivant : préhistoire, antiquité, moyen-âge, époque moderne.

Principe de mesure quantitative : si les sujets d'un ensemble admettent une distinction quantitative, on peut les ranger par ordre croissant ou décroissant.

Principe de complexité croissante : si les sujets d'un ensemble présentent des degrés différents de complexité, il faut les ranger dans une séquence de complexité croissante.

Principe de séquence canonique : quand des sujets sont présentés traditionnellement dans une séquence spécifique, on doit s'y conformer.

Principe de caution bibliographique : les sujets d'un ensemble peuvent être organisés en fonction de la quantité décroissante de documents publiés ou anticipés sur les sujets.

Principe de séquence alphabétique : si aucune autre séquence n'est plus utile, on peut ranger les sujets d'un ensemble alphabétiquement par leur nom d'usage actualisé internationalement.

2.5.1.3 Canons pour le plan terminologique

Sur le plan terminologique, Spiteri (1998) reprend les deux premiers principes énoncés par Ranganathan et élimine le dernier à savoir le canon de retenue qui interdit l'expression d'une opinion dans les termes d'une classification. Les deux principes qu'elle reprend sont les suivants :

Canon de contexte : ce qu'un terme de la classification désigne doit être déterminé en lien avec les classes supérieures appartenant à la même chaine. Un terme a un contexte basé sur sa position dans le système de classification.

Canon d'actualité : le terme utilisé pour désigner une classe doit être celui utilisé couramment, actuellement, par les spécialistes du champ du sujet. Ceci implique que certains termes peuvent devenir obsolètes et doivent être remplacés.

2.5.1.4 Canons pour le plan de notation

Spiteri (1998), reprend les principes suivants qui assurent l'exclusivité mutuelle dans la notation :

Canon de synonymie : Ranganathan précise qu'un indice est unique, que chaque sujet doit être représenté par un seul indice, et que deux sujets ne peuvent pas avoir le même indice.

Elle retient aussi le *canon d'hospitalité :* il doit être possible de faire des ajouts à la fin d'un ensemble ou entre deux indices consécutifs, par interpolation ou extrapolation.

Enfin, elle reprend le principe du *classement ordonné* du CRG, et qui n'est pas présent dans la méthodologie de Ranganathan : le système de notation doit refléter le classement ordonné des sujets. Ainsi, lors du classement des livres sur une étagère ou dans un catalogue, la notation doit permettre de suivre l'ordre des sujets du système de classification (par exemple du plus général au plus complexe).

2.5.1.5 Les principes éliminés des deux méthodologies précédentes

Les principes, issus des méthodologies de Ranganathan et du CRG, éliminés dans ce modèle simplifié sont présentés ici.

2.5.1.5.1 Les principes éliminés des canons pour les ensembles de classes

L'exhaustivité : les classes dans un ensemble de classes doivent être exhaustives vis-à-vis de la classe immédiatement supérieure. L'auteur mentionne la possibilité d'ajouter une classe « Autre » pour satisfaire ce canon. Ce principe est éliminé parce qu'il est difficile de savoir si toutes les caractéristiques des entités ont été prises en compte.

Des séquences efficaces : la séquence de classes d'un ensemble de classes doit être efficace pour les buts de ceux à qui elle est destinée. Cela peut varier en fonction du public. Ce principe est éliminé car il est présent dans le canon de succession pertinente pour les facettes et valeurs de facettes du présent modèle.

Des séquences cohérentes : quand des classes similaires apparaissent dans des ensembles de classes différents, elles doivent être parallèles dans tous ces ensembles. Ce canon est éliminé car il peut être difficile de déterminer selon quel degré certaines classes ont des structures parallèles.

2.5.1.5.2 Les principes éliminés des canons pour les chaines de classes

L'extension décroissante : quand on passe du premier niveau de la classification vers le dernier, l'extension de la chaine doit aller en décroissant, tandis que l'intention doit s'accroitre. Il s'agit ici d'aller du plus générique au plus spécifique dans l'application des divisions successives. Ce canon est supprimé car il implique une division du savoir Genre-Espèce qui n'est pas intrinsèquement pertinent pour les classifications à facettes.

La modulation : une chaine de classes doit inclure une classe de chaque niveau compris entre le premier et le dernier niveau. Ranganathan qualifie aussi ce canon par l'expression « Resolving power », qui correspond à la capacité à reconnaître la classe inférieure appropriée à celle immédiatement supérieure. Par conséquent, une chaine de classes, ou hiérarchie de classes, doit être dérivée depuis la classe immédiatement supérieure en utilisant le « resolving power » le moins fort possible à chaque stade de la division. Ce canon est supprimé car il

implique un ordre prescrit des facettes et qu'il est donc en conflit avec le principe de Succession pertinente du modèle simplifié.

2.5.1.5.3 Les principes éliminés des canons pour les séquences de classes

Les classes subordonnées : quand A1, A2, A3 sont des sous –classes d'une classe A, alors elles doivent suivre immédiatement la classe A dans la succession, sans être séparées de A par une autre classe, ni entre elles.

Les classes coordonnées : si les classes A et B sont originaires du même ensemble et étaient consécutives à l'intérieur de celui-ci, alors, elles ne doivent pas êtres séparées par une autre classe, excepté A1, A2, A3, qui ont pour classe immédiatement supérieure, A.

Ces canons sont supprimés parce qu'ils sont contenus dans le principe d'exclusivité mutuelle de ce modèle simplifié.

Canon d'énumération : le sens d'une classe est rendu clair par ses sous-divisions. Ce principe est éliminé parce qu'il est contenu dans le principe de contexte du présent modèle.

Canon de retenue : un terme utilisé dans une classification doit être neutre, ne doit pas exprimer d'opinion. Ce principe est supprimé car il peut être en conflit avec le principe de d'actualité qui est conservé.

2.5.1.5.4 Les Canons de notation

Nous n'avons pas détaillé les canons qui traitent de la notation dans la partie qui présente la méthodologie de Ranganathan parce qu'ils traitent des types de notations possibles plutôt que des qualités fondamentales des systèmes de notation. Spiteri (1998) les a supprimés pour ces mêmes raisons.

2.5.1.5.5 Les principes éliminés des séquences de facettes

Les principes des séries et des niveaux ont été supprimés parce qu'ils sont en conflit avec le principe des Catégories fondamentales du présent modèle, ainsi qu'avec le principe de l'ordre

de citation du CRG qui ne se conforme pas aux prescriptions de Ranganathan, en affirmant qu'il n'y a pas d'ordres universellement adaptés aux facettes.

Ainsi Spiteri (1998) a aussi supprimé les principes supplémentaires pour les séquences de facettes qui pour rappel, visent à guider l'ordonnancement des éléments dans une séquence de facettes, et sont les suivants : le Principe de la Vision d'Ensemble, le Principe du Tout-Organe, Principe « Vache-Veau », Principe de l'Actant - Action – Acteur – Outil (voir partie 2.2.9.3).

2.5.1.6 Observations sur le modèle de Spiteri (1998)

Comme on peut le constater, ce modèle est effectivement simplifié par rapport à celui de Ranganathan. Il énonce des règles fondamentales pour la mise en place de classifications à facettes mais, selon nous, il constitue un modèle prescriptif, et non un guide opérationnel pour l'analyse par facettes et la construction d'une classification à facettes. Il s'agit d'une concaténation des deux modèles de Ranganathan et du CRG. Ainsi, nous considérons que la méthode proposée par Denton est complémentaire aux propositions de Spiteri (1998) parce qu'elle opérationnalise leurs principes.

2.5.2 Méthodologie de classifications à facettes de Denton pour les sites Web

Tout d'abord, la définition de la notion de facette qu'il adopte correspond à la suivante :
« a set of mutually exclusive and jointly exhaustive categories, each made by isolating one perspective on the items (the facets), that combine to completely describe all objects in question, and which user can use, by searching and browsing, to find what they need. » (Denton, 2003).

Il ajoute que tous les bénéfices des classifications à facettes peuvent être mis à profit sur le Web et que la conception d'une classification doit se faire d'après ses buts. Ainsi différents objets peuvent être classés de différentes manières pour des objectif différents, requérant des structures diverses. (Denton, 2003).

Il préconise d'utiliser la classification à facettes quand le but recherché est de décrire des items selon plus de deux dimensions. Il propose une procédure à sept étapes pour la construction de classifications à facettes, qu'il base sur les apports de Vickery (1960), et Spiteri (1998).

Les sept étapes sont présentées ci-dessous :

1. **Collecte dans le domaine :** collecter un jeu représentatif des entités à classer ;
2. **Listage des entités :** lister les entités et décomposer les descriptions en parties, réarranger les mots dans des listes, en décomposant les phrases en concepts basiques qu'il faut isoler ;
3. **Création des facettes :** examiner les listes qui sont le résultat du travail effectué à l'étape 2 et extraire les catégories générales, de haut niveau qui apparaissent dans cette matière. Etudier ces catégories relevées et les restreindre en un jeu de facettes mutuellement exclusives et conjointement exhaustives, dans lesquelles tous les termes de l'étape précédente peuvent entrer. Pour cette étape Denton préconise d'utiliser les principes présentés par Spiteri (1998) présidant au choix des facettes, ainsi que de s'inspirer de classifications à facettes existantes comme le modèle de Ranganathan ou celui de Vickery. Il propose aussi de se fier à un jeu de questions proposé par (Broughton, 2001) pour l'analyse d'un champ dans le domaine des technologies:
    - Qu'est ce qui est produit ?
    - Quelles sont ses parties, ses composants ?
    - Comment est-ce produit ?
    - Par quels moyens est-ce produit ?
    - Par qui ?
    - Quand ?
    - Et Où ?

Il rappelle qu'il faut garder à l'esprit les besoins des utilisateurs et le but de la classification, à l'aide des questions suivantes :

- qui va l'utiliser ?
- pourquoi ?

- est ce que les utilisateurs vont faire des requêtes ou du browsing à l'aide de la classification, ou les deux ?
- dans quelle mesure connaissent-ils le domaine ?

4. **Arrangement des facettes :**

Il commence par faire un test pour ordonner tous les termes dans les facettes. Denton préconise d'utiliser les principes pour la succession des facettes et valeurs de facettes proposés par Spiteri (1998). Il peut rester des termes difficilement ordonnables dans des facettes. A ce stade il faut tester si toutes les entités originales peuvent être décrites par les valeurs de facettes ordonnées issues de cette opération d'ordonnancement des valeurs dans les facettes. Si certaines valeurs manquent, il faut revenir en arrière et procéder à une nouvelle analyse et réordonner les termes. Quand ce brouillon préliminaire de classification démontre son efficacité pour le classement des entités originales, alors il faut procéder à l'ordonnancement final des facettes en utilisant les principes pour la succession des facettes et valeurs de facettes proposés par Spiteri (1998). Dans différentes facettes, différentes sortes d'ordre peuvent être utilisés pour ranger les valeurs de facettes. Enfin, il faut harmoniser le vocabulaire pour en faire un langage contrôlé. Denton ne donne pas de détails sur cette opération.

5. **Ordre de citation :** d'après Denton, sur le Web, il est possible de changer autant que nécessaire l'ordre de citation mais il faut quand même décider d'un ordre de présentation par défaut pour le site. Cet ordre par défaut doit être compréhensible et utile à tous les utilisateurs, tandis que les plus expérimentés doivent pouvoir le réarranger.
6. **La classification :** la conception à ce stade est terminée, il faut ensuite utiliser la classification pour classer tous les items du domaine.
7. **Révision, test et maintenance :** si des problèmes sont rencontrés à l'étape 6, il faut alors revenir en arrière autant que nécessaire, pour les corriger. Les problèmes suivants peuvent se produire : le classement des valeurs doit être modifié, des valeurs manquent, le choix des facettes s'avère inapproprié. Il faut ensuite tester avec des utilisateurs, et faire les modifications nécessaires en fonction de leurs retours d'expérience.

   La maintenance consiste à mettre à jour la terminologie, à vérifier que les nouvelles entités peuvent être classées correctement, et à s'assurer que les facettes sont suffisantes

pour les évolutions du domaine. Il peut falloir ajouter ou réarranger des valeurs de facettes, ou des facettes et éventuellement concevoir la classification en entier à nouveau.

2.5.3 Observations sur la méthode de Denton (2003)

Un de ses mérites est d'être applicable, opérationnelle. On peut observer qu'elle est conçue pour être itérative, la construction de la classification dans cette méthode n'est pas linéaire, elle peut nécessiter des retours en arrière, et des révisions.

L'étape 2 revient à identifier les composants d'un sujet en concept basique et ses composants.

Sur l'étape 3, Denton ne précise pas ce que signifie une catégorie, ou une catégorie générale. Dans cette même étape, on constate que les questions que Denton recommande de garder à l'esprit sur les besoins des utilisateurs et les buts de la classification concernent des aspects différents : les utilisateurs, leurs connaissances, le but de la classification, et les problématiques d'utilisabilité et de design de l'IHM.

Sur le troisième point de l'étape 4, qui concerne l'harmonisation du vocabulaire pour en faire un langage contrôlé, peu de détails sont apportés par Denton, nous supposons qu'il inclut dans cette partie, l'élimination des doublons, le traitement de la synonymie, etc. Sur ce même point, nous pensons que même si ces processus de traitement de la terminologie sont effectués, cela ne suffit pas nécessairement à produire le consensus nécessaire dans un langage contrôlé.

Sur l'ordre de citation par défaut dans l'étape 5, Denton ne donne pas d'indication sur la façon de l'élaborer. L'étape 6 consiste à classer les items avec la classification élaborée, il s'agit donc d'une classification élaborée à priori, même si sa conception est itérative, on peut se demander si dans de très larges collections, une conception au fil de l'eau ne pourrait pas être plus adaptée. Nous rappelons ici que cette méthode convient pour des corpus de taille moyenne ou petite.

Enfin, sur les tests (étape 7), Denton ne formule pas précisément ce qu'il faut évaluer avec les utilisateurs. Pour conclure cette partie nous pouvons aussi observer que Denton ne donne pas d'indication sur le choix du degré de précision des valeurs de facettes et des facettes, ce qui nous apparaît comme une des questions à résoudre dans le cadre de ce travail.

2.6 Synthèse des caractéristiques des méthodologies présentées

Nous présentons une synthèse des caractéristiques des méthodologies présentées dans ce chapitre dans le tableau suivant. La dernière colonne concerne la méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances métier, que nous présentons en chapitre 5.

Tableau 3. Synthèse des caractéristiques des méthodologies présentées

Critères d'analyse	Méthodologie de Ranganathan	Méthodologie du CRG	Méthodologie GenISIS	Méthodologie de Denton inspiré de Spiteri	Méthodologie pour des "FolkClassifications"
Objectif généraux de la classification	Bibliothèques	Classification spécifiques	Record Management dans les organisations	Sites Web	Gestion personnelle de l'information et Gestion collaborative des connaissances dans les entreprises
Visée Universelle ou Contextuelle	Universelle	Contextuelle	Universelle	Contextuelle	Contextuelle
Type de méthodologie : Basée sur des principes théorique ou Opérationnelle	Basée sur des principes théoriques	Hybride Basée sur des principes théorique et Opérationnelle	Opérationnelle	Opérationnelle	Opérationnelle
Type de méthodologie : top-Down ou Bottom up	Top- Down	Hybride Top-Down / Bottom-Up	Top-Down	Bottom-Up	Bottom-Up
Modèle de facettes prédéfini ou émergent	Modèle de facettes prédéfini	Modèle de facettes prédéfini	Modèle de facettes prédéfini	Modèle de facettes émergent	Modèle de facettes émergent
Sources	Non précisé	Connaissance des besoins des utilisateurs, des glossaires et autres langages d'indexation existants, et de la littérature sur le champ	Entretiens de haut niveau/ de niveau intermédiaires/ documents référentiels/ workshops/ questionnaires	Collecte d'un jeu représensentatif des d'entités à classer	Collecte de documents/connaissance du métier/ modèle émergent du consultant/ entretiens/ confrontations
Présence ou Absence d'un ordre de citation des facettes	Présence d'ordre de citation des facettes	Présence de principes pour l'ordre de citation des facettes	Présence d'ordre de citation des facettes sous forme de relations entre les valeurs de facettes	Absence d'ordre de citation des facettes	Absence d'ordre de citation des facettes
Présence ou Absence d'un système de notation formel	Présence d'un système de notation formel	Présence d'un système de notation formel	Absence d'un système de notation formel	Absence d'un système de notation formel	Absence d'un système de notation formel
Respect de l'exclusivité mutuelle des facettes entre elles	Oui	Oui	Oui	Oui	Oui
Possibilité de multivalué	Oui, mais avec un dispositif de notation ordonné pour distinguer leurs niveaux et séries	Oui, avec un dispositif de notation	Non	Oui	Oui
Présence de Hiérarchies dans les valeurs de facettes	Oui	Oui	Oui	Oui	Non
Présence de principes pour l'ordre dans les valeurs de chaque facette	Oui	Oui	Oui	Oui	Non
Présence d'une limite normée dans le nombre de facettes	Oui	Non	Oui	Non	Non
Homogénéité des types de propriétés traitées dans les valeurs dans chaque facette	Oui	Oui	Oui	Oui	Oui
Homogénéité des types de valeurs de facettes dans chaque facette	Oui en principe, mais en réalité Non vu le degré de généralité du modèle	Oui	Non, il y a des types de valeurs appliquées variablement (valeurs-rallonges)	Oui en principe	Non

Chapitre 3 : Systèmes d'Information et Activités

3 Chapitre 3 : Systèmes d'Information et Activités

3.1 Introduction

Ce troisième chapitre est consacré à l'analyse des activités de gestion de l'information d'après les approches de la Théorie de l'activité, et de celles issues de la psychologie du travail et de l'ergonomie. Ainsi, ces cadres d'analyse nous permettent de caractériser les activités de gestion de l'information. Nous les abordons plus particulièrement sous leurs aspects personnels à travers l'étude de la littérature sur le Personal Information Management (PIM) et les notions de contexte et de métier que nous développons pour ses aspects collectifs. Le prototype Hypertagging développé dans le cadre du projet Miipa-Doc est également présenté. Enfin, nous tentons de caractériser l'impact sur les activités de gestion de l'information et sur les activités métier, des choix de Systèmes d'Information (SI) en fonction d'une typologie de ces systèmes que nous proposons.

3.2 Les systèmes d'informations

Selon les définitions adoptées du Système d'Information (SI), celles-ci recouvrent des réalités variables. On trouve dans la littérature différentes acceptions de cette expression allant dans une progression du macro au micro, d'après (Mitev, 2008), qui propose une revue de la littérature anglophone sur les SI dans les organisations de la conception d'une société de l'information ou société en réseau, à une conception centrée sur les domaine d'application comme la médecine ou la banque, avec une qualification des SI en fonction de ces domaines, comme les « Health Information Systems ». Enfin, au niveau meso, on trouve des approches du SI dans les organisations qui s'intéressent aux relations entre les TIC (Technologies de l'Information et de la Communication) et les organisations, avec des thématiques qui sont orientées sur le développement des SI pour la gestion des projets, des données, du savoir, ou des approches plus socio-techniques centrées sur la participation des utilisateurs, ou encore des approches gestionnaires (aspects stratégiques des SI) ou managériales avec des

applications dédiées à la gestion de la relation client (CRM), ou des systèmes de gestion intégrée des ressources de l'entreprise (PGI). (Mitev, 2008)

Dans la littérature française dont nous avons connaissance, ces derniers sont aussi mentionnés par l'expression « Systèmes d'Information Organisationnels » qui a été utilisée comme titre d'un numéro de la revue Sciences de la Société, intitulé « Système d'Information organisationnels ? » (N°63, 2004).

On trouve aussi, dans la littérature sur le SI, des notions comme celle de dispositif informationnel. (Roux, 2004). D'après cet auteur, qui cite (Reix, 2002), un SI peut se définir comme « un ensemble organisé de ressources: matériel, logiciel, personnes, données, procédures, (…) permettant d'acquérir, de traiter, stocker, communiquer des informations (sous forme de données, textes, images, sons) dans les organisations». Cependant, elle précise qu'en sciences de l'information, le SI n'est pas un système informatique bien qu'il y soit souvent assimilé. Elle préfère réserver l'emploi de l'expression « Système d'Information » pour la description de la combinaison des outils techniques, tandis que celle de « dispositif informationnel » inclut l'articulation avec les acteurs et leurs interactions avec les outils. (Roux, 2004).

Dans le même sens, (Zacklad, 2006) établit une distinction entre deux composantes du SI, « les réseaux informatiques, de télécommunication et stockage (…), qui véhiculent les signaux le plus souvent numérisés », et « le système documentaire qui à partir de certains flux de données issus du précédent, est le résultat d'une documentarisation des fichiers pour permettre leur ré-exploitation ultérieure. Le premier [composant] garantit l'écoulement d'un flux continu d'informations d'un point à un autre, relié par le réseau, depuis l'acquisition et le traitement du signal jusqu'à sa restitution sur des équipements adaptés. Le second [composant] organise le contenu de l'information, son stockage et son accès. » (Zacklad, 2006). Nous étudierons ici plutôt le deuxième composant évoqué.

D'après (Guyot, 2006, p.53), « un système d'information est un ensemble d'informations/ représentations organisé selon certaines fins, à l'aide d'un ou plusieurs dispositifs avec, éventuellement, des outils. Il représente quelque chose et peut être fermé ou ouvert (porosité des frontières). C'est un système d'acteurs produit et tenu par des pratiques, des stratégies, un système de règles, une organisation et des outils, « disposés » selon une finalité et renvoyant à un univers symbolique. »

Elle distingue deux types de dispositifs, ceux qui gèrent l'activité et ceux qui gèrent l'information. Ceux qui gèrent l'activité incluent les outils de workflow et les PGI (Progiciel de Gestion Intégrée) ; parmi ceux qui gèrent l'information, on trouve les dispositifs de veille, de Record Management, de gestion des connaissances, les groupwares, les intranets, les bases de données, les forums. Il nous semble cependant que cette distinction entre gestion de l'activité et gestion de l'information est trop généraliste, notamment pour les PGI. En effet, ceux-ci nous semblent gérer les deux. Plus précisément, et d'un point de vue global, comme le précise (Guyot, 2006), différentes problématiques traversent le champ des outils actuellement disponibles. Elle distingue six grandes problématiques que nous présentons de manière schématique dans le tableau suivant avec les outils qui leur correspondent.

Problématique majoritairement traitée	Outils correspondants
Mémoire : accumuler, archiver	GED, outils de Record Management
Gestion des connaissances : capitaliser, expliciter, construire du sens, contextualiser	Bases de connaissances, intranets, forums
Exploitation : rechercher, rendre accessible	Systèmes documentaires, Systèmes de recherche d'information (SRI), dispositifs de veille, Systèmes Interactifs d'Aide à la Décision (SIAD), Bases de données (BDD), Systèmes d'Information Personnels (SIP)
Organisation du travail et des processus : cadrer, coordonner, rationnaliser l'organisation des processus de travail	ERP, Workflows, Systèmes d'Information Personnels (SIP)
Communication des flux : faire circuler l'information courante	Messageries électroniques, listes de diffusion, dispositifs de veille, Groupware, SIAD, Forums
Mettre en réseau : mettre en relation, rassembler, unifier	Réseaux sociaux, Groupware, annuaires d'entreprise

Tableau 4. Typologie des problématiques traitées par les Systèmes d'Information

Si cette catégorisation des outils n'est pas optimale car beaucoup des types d'outils pourraient entrer dans plusieurs catégories; celle-ci reste cependant utile pour situer les outils dans leurs usages attendus. Cette présentation des outils pourrait, en effet, s'articuler selon d'autres

points de vue, par exemple selon le degré de structuration des systèmes évoqués, selon le type de document qu'ils gèrent spécifiquement, leur impact sur l'activité de gestion de l'information, et l'activité des opérateurs ou selon la conception des activités métier et des métiers. C'est ainsi que nous allons explorer ce que nous appelons communément le système d'information, qui regroupe et rassemble dans cette catégorie vague tous les systèmes présents dans le tableau qui précède, et que nous décrirons sous différentes modalités, en catégorisant le SI en fonction de son degré de structuration, qui sera décliné sous trois formes : le système d'information structuré, semi-structuré, et le système d'information 2.0.
Il nous semble en effet que ce degré de structuration du SI produit des implications sur les modalités d'usages, en termes d'activité de gestion documentaire, mais aussi sur la conception de l'activité des opérateurs, et, par conséquent, sur la conception des métiers afférente. Les différents degrés de structuration précédemment cités du SI sont généralement déterminés en référence à la structuration des données, qu'ils servent à gérer ou à mettre à disposition. Ainsi, on parle de données structurées, semi-structurées ou non structurées.

Nous allons tenter ici de donner des éléments d'analyse des systèmes d'information précédemment présentés afin d'être en mesure de caractériser ces outils en fonction de leur degré de structuration. Nous notons que dans la littérature en SIC sur les systèmes d'information, on trouve indifféremment les termes de TIC (Technologies d'Information – Communication), NTIC (Nouvelles Technologies d'Information et de Communication) et dispositifs informationnels pour les désigner.

3.2.1 Notion de document

Aussi, pour commencer nous allons nous intéresser aux documents qu'ils permettent de supporter. A cet effet, nous ne donnerons pas de définition généraliste de la notion de document, la littérature étant prolifique sur ce sujet mais les auteurs ayant des difficultés à s'accorder sur une définition univoque. A ce propos, il est pertinent de mentionner les travaux de Lund (2009) qui soulignent la complexité des implications d'une définition de cette notion, ou encore ceux de Buckland (1998) qui recueillent un certain nombre de définitions du document électronique, par opposition au document papier. On constate un large spectre de propositions, qui aux débuts de la science de la documentation cherchaient à définir le document en référence à la culture de l'imprimé, avec, notamment, une recherche pour définir les limites de ce qui peut être considéré comme un document. Dans cette démarche on trouve

les travaux bien connus d'Otlet (1934), qui indiquent que le document peut être un objet tel qu'une pièce de musée, des traces archéologiques, ou une représentation graphique.

On trouve également les travaux de Briet (1951) qui mentionnent qu'un « document est une preuve à l'appui d'un fait ». Elle a aussi proposé une autre définition : « Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène physique ou intellectuel ». D'après Buckland (1998), les propositions de Suzanne Briet supposent qu'un document a une matérialité, et son statut de preuve (et de document) doit être construit, reconnu. A l'autre extrême de ces définitions en référence à la culture de l'imprimé, on trouve des propositions très générales et pragmatiques de définition du document électronique, comme « *anything that can be given a file name and stored on eletronic media* » (Buckland, 1998). Selon nous, une des difficultés de ce travail de définition réside dans la coexistence de modalités de formes anciennes (l'imprimé) et modernes (électronique) ainsi que dans la pluralité des rôles et fonctions associée à cette notion. Dans ce sens, (Pédauque, 2006) propose de regrouper les fonctionnalités d'un document en quatre propriétés : mémorisation, organisation, création et transmission.

Dans une précédente publication, Pédauque (2003) tente de caractériser le document numérique en fonction de trois grandes dimensions issues des sciences du langage : le document comme forme, comme signe et comme médium. Le document est considéré dans sa forme éditoriale initiale, le livre ou la revue, sous la dimension de la forme, comme composé d'un support manipulable sur lequel est fixée une trace interprétable, l'inscription. Le passage au document numérique induit un entrelacement des supports et des signaux, la notion de support devenant ambiguë, l'inscription relève alors du codage et le document devient composé d'une structure et de données, à la manière d'un programme informatique. Le contrat de lecture classique entre l'auteur et le récepteur devient un contrat de lisibilité, lié à la structure inhérente au document, comme dans les fichiers XML.

Lorsque l'on observe le document comme un signe, il s'agit de l'organiser parmi d'autres signes, et de comprendre son contenu. Le document peut donc, selon cette dimension, être défini comme composé d'un « texte informé » (augmenté de marqueurs comme des métadonnées), et de connaissances. Lorsque le document est considéré comme un médium, le contrat de lecture relève de la sociabilité, et le document peut être vu comme composé soit d'une inscription et d'une légitimité, qui lui confère le statut de document du point de vue d'une dynamique qui serait plutôt concernée par l'aspect de preuve d'un document ; soit il

peut être considéré comme composé d'un texte et d'une procédure du point de vue de la communication organisationnelle.

Ainsi, (Pédauque, 2003) écrit-il :

> *« La communication organisationnelle étudie le document immergé dans des pratiques et des situations professionnelles caractérisées, et donc contraintes, par des systèmes de règles. Elle le regarde alors à plusieurs niveaux : tout d'abord comme un écrit identifié dans un contexte, formalisé par des règles d'écriture, de circulation, d'usage, inscrivant une intention liée à l'action, et gardant trace des négociations socio-techniques menées autour de lui. Cela amène, notamment, à étudier les processus de fabrication et de gestion des documents, activités qui ne sont plus seulement le fait d'acteurs spécialistes, mais qui se distribuent dans l'ensemble de l'organisation. Elle le regarde ensuite comme un élément structurant l'organisation, en tant que support de coordination. Enfin, elle le voit comme un moyen utilisé par les acteurs individuels ou collectifs dans leurs différentes stratégies. Sur le plan méthodologique, le document est pris comme un « observable » permettant d'étudier à la fois les relations entre les acteurs, (il est un médiateur), les modes de régulation (il est un outil de management) et les recompositions organisationnelles (il en est l'un des révélateurs). » (Pédauque, 2003, p.20-21).*

Suite à cette revue rapide des définitions possibles du document, on constate qu'il est loin d'être un objet simple et qu'il porte différentes fonctions. Au niveau organisationnel, il s'agit d'un support de communication, qui garde des traces des échanges, d'un support de coordination, d'un moyen individuel et collectif pour les opérateurs de remplir leurs missions, et il est aussi une sorte de témoin de l'activité.

A propos de l'activité de gestion de l'information, nous avons mentionné les travaux de (Pédauque, 2006) qui décrivent quatre fonctionnalités des documents, nous paraissant pouvoir être rapprochées des problématiques des Systèmes d'Information présentées dans le tableau 4. Ainsi la fonction de mémorisation des documents correspond-elle aux problématiques de Mémoire et de Gestion des connaissances, la fonction de création des documents est assurée par fonction d'exploitation des SI, la fonction organisatrice des documents correspond à celle d'organisation du travail et des processus des SI, enfin, la fonction de transmission des documents peut être rapprochée de la fonction de communication des flux et de mise en réseau des SI. On voit ainsi que l'on peut dresser un parallèle entre des fonctions portées par les documents et des fonctions actuellement opérationnalisées et formalisées dans les SI (qui gèrent des documents et des activités).

(Pédauque, 2006) mentionne les activités de gestion de l'information comme des activités autrefois confiées à des médiateurs professionnels, mais qui sont « désormais à la charge des utilisateurs ».

D'autre part, toujours d'après (Pédauque, 2006), « Tout document est désormais, à un stade de son existence, un document numérique et toute réalisation matérielle du document est marquée par la forme qu'imprime aux logiques documentaires le programme informatique ». Le vaste déploiement de l'informatique dans les organisations, qu'on appelle aussi le passage au numérique, suppose de nouveaux savoir-faire, impacte les activités métier, le volume documentaire, et la manière de rendre compte de son travail. Concernant le volume croissant de documents produits et à gérer, « On peut se demander si une des conséquences de l'informatisation n'est pas de produire des unités d'informations autrefois proto-documentaires directement comme des documents. » (Pédauque, 2006, p.20). On voit ainsi que le numérique impacte aussi la nature des documents eux-mêmes et qu'il crée des besoins d'outils pour gérer (classer, nommer, décrire et retrouver) le volume de ces proto-documents, définis comme une « collection de matériels documentaires, plus ou moins cohérents et organisés » (Pédauque, 2006) et qui constituent des moyens pour la production de documents. Nous considérons que la croissance de cette masse de proto-documents constitue des réservoirs de connaissances et d'informations utiles à l'organisation, ou à ses acteurs individuels et collectifs, et qu'il est donc nécessaire de les organiser pour pouvoir les exploiter, tandis qu'ils deviennent le corollaire de l'activité. Pour le dire autrement, toute activité est productrice d'informations, sinon de documents, et nécessite des prises d'informations qui peuvent être contenues dans des documents, ou disséminées dans le contexte de la situation de travail. (Guyot, 2006, p.19)

Nous reviendrons dans la suite de cette partie sur la notion d'activité que nous décrirons sous plusieurs angles, notamment du point de vue de la psychologie du travail et de l'ergonomie. Mais ce qui nous préoccupe d'abord ici est focalisé sur la manière de définir les critères de description des Systèmes d'Information, qui nous permettrons d'explorer les impacts des différents types de SI sur l'activité de gestion d'information, qui sera analysée par la suite selon une grille issue de la psychologie du travail et la théorie de l'activité. Ce faisant, une description plus générale de l'activité sera proposée d'après cette même théorie, en incluant des éléments d'analyse issus de la psychologie du travail.

3.2.2 Quels critères de différenciation des systèmes d'information ?

Nous avons défini un jeu de sept critères de différenciation des systèmes d'information, qui vont être présentés en détails et qui nous permettent de proposer trois grandes catégories de SI : les systèmes d'information structurés, semi-structurés, et les SI « d'inspiration 2.0 ». Ils seront ensuite synthétisés dans un tableau. Ils comprennent : 1) le degré de contrôle exercé sur les utilisateurs, 2) les fonctions du système, 3) les bénéfices du système selon les utilisateurs, 4) les types de documents que le système héberge, 5) les modalités de participation des utilisateurs finaux dans la conception du système, 6) le type de sémantique induit par le système, 7) les SOC disponibles dans ces SI.

3.2.2.1 Le degré de contrôle exercé sur les utilisateurs

Il s'agit, pour ce premier critère, d'examiner les modalités d'utilisation du point de vue du contrôle exercé sur les données, qui correspondent au degré de contrainte exercé sur les utilisateurs, ou autrement dit, des modalités de régulations associées.

Dans un SI structuré, la régulation repose sur la standardisation des données, des processus d'alimentation du système, et sur une gestion des droits rigoureuse et standardisée, en fonction des statuts des utilisateurs, pour ce qui concerne l'accès.

Zacklad (2010a, p.4), dans une approche transactionnelle et documentaire des SI, considère que la régulation relève des processus de coordination, et que dans ce type de régulation par la standardisation, « on tente de procéder à un traitement anticipé des interdépendances qui a pour conséquence une standardisation du déroulement de la transaction primaire ».

D'après (Zacklad, 2006, p.9), la coordination correspond à « l'ensemble des artefacts et des méthodes permettant de résoudre les problèmes liés à l'articulation d'activités interdépendantes complexes, (…), et d'assurer une continuité entre les situations d'activités appartenant à même flux transactionnel ». Nous reviendrons à l'Approche Transactionnelle de l'Action dans la suite de ce chapitre, lorsque nous aborderons l'analyse de l'activité.

Dans un SI semi-structuré, d'après (Zacklad, 2010a, p.4), les transactions avec le système relèvent d'une régulation en situation. Ce type de SI correspond à des dispositifs où un des objectifs est de « faciliter le travail de régulation par les personnes engagées dans la transaction primaire durant le cours de l'action ». Il s'agit donc aussi d'une régulation inter-

personnelle. Les SI « d'inspiration 2.0 » sont plutôt caractérisés sur ce point par l'autorégulation, par les acteurs eux-mêmes, individuellement.

3.2.2.2 Les fonctions du système

Nous distinguons trois grands types de fonctions. Dans les SI structurés, les fonctions sont liées à des besoins de formaliser les processus, les connaissances, les procédures, ainsi qu'à des besoins d'archivage. Les SI semi-structurés d'après (Zacklad, 2010a, p.3), permettent d'assurer « la transition entre des situations d'activités distribuées qui permettra aux personnes de prolonger ou de relancer de la manière la plus cohérente possible les transactions en cours ». Ce type de système a donc principalement pour fonction de rendre possible la coopération et la collaboration entre des acteurs distribués.
Les SI 2.0 sont orientés sur le partage, la diffusion, la promotion de contenus et d'éléments de profils personnels qui deviennent en partie collectifs au cours de l'utilisation du système. C'est pour cette raison que nous considérons que ce système fait coexister deux espaces, l'espace personnel et l'espace collectif (au sens restreint), et qu'il est donc orienté sur la gestion personnelle de contenus.

3.2.2.3 Les bénéfices du système selon les utilisateurs

Les systèmes d'information structurés permettent d'obtenir rapidement des données de reporting fiables, par la rationalisation, l'uniformisation, la normalisation et le contrôle des données, procédures, et activités de travail. Ils offrent une vision globale de l'activité de l'entreprise, en favorisant la transparence des données (la saisie des données est contrainte, ce qui les rend immédiatement disponibles pour un autre service, en évitant également les saisies en doublons). Théoriquement, ces pratiques sont censées être un facteur d'économie et de rentabilité pour l'organisation, de fluidité des processus, et de rationalisation des moyens.
Un SI semi-structuré fournit un cadre plus souple que les SI structurés, orienté vers l'échange, la construction et l'enrichissement des connaissances d'acteurs distribués à travers l'organisation en mode synchronisé ou non. Il peut favoriser le travail en mode projet ou les échanges entre praticiens d'un même métier, il est au service des utilisateurs, et des collectifs de travail.

Les bénéfices d'un SI 2.0, du point de vue des utilisateurs finaux, favorise leur autonomie et la valorisation de leurs connaissances ou expertises, il est moins formel que les deux systèmes précédents mais peut contribuer à une organisation par projets. Du point de vue des gestionnaires, le contrôle de ces systèmes est plus ardu, ces systèmes pouvant être perçus comme opaques, et donc difficiles à exploiter par des gestionnaires.

3.2.2.4 Les types de documents que le système héberge

Dans (Zacklad, 2010a, p.4), on trouve une caractérisation des documents en fonction de leur degré de codification. C'est cette perspective que nous adoptons ici pour désigner les types de documents. Ainsi, dans cette typologie, les documents « fortement codifiés » correspondent aux « documents de codage » (liste de nomenclature, fichiers de programmes) et aux « documents articulatoires » (bases de données informatiques codifiées, et mécanismes de coordination). Ce sont ces documents qui sont majoritairement traités par des Si structurés.
Les documents « faiblement ou non codifiés » sont des « documents ressources » (œuvres fixées sur un support : livre, rapport, transcription) et les « Document pour l'Action » sont définis comme tels : ils peuvent être potentiellement inachevés de manière prolongée (disponibles sous plusieurs versions), et posséder une certaine pérennité liée à l'engagement des différents rédacteurs et au caractère distribué (réparti dans le temps et l'espace) des différentes contributions grâce au réseau, ils peuvent être fragmentés, du fait de cette rédaction potentiellement à plusieurs mains, ils mettent à contribution des réalisateurs aux statuts variables dans l'activité, et enfin leurs parties s'articulent de manière non triviale. (Zacklad, 2005). Ils incluent les fichiers texte, image, sons, annotés collectivement, les messages électroniques, les systèmes de forum, blogs et wikis (Zacklad, 2010a, p.4), et sont majoritairement gérés par des systèmes d'information semi-structurés.
Enfin, les documents principalement gérés par des systèmes d'information d'inspiration 2.0 consistent généralement en des liens hypertextes vers des ressources issues du Web, des informations de profils, et des fils RSS.

3.2.2.5 Les modalités de participation des utilisateurs finaux

Les SI structurés sont généralement installés dans une approche descendante (Top-down), ce sont des experts de ces systèmes, souvent prestataires externes à l'organisation, qui réalisent

le déploiement du SI. Cependant, de petits groupes de travail, pour la spécialisation des applications métier, sont organisés. Ils intègrent quelques utilisateurs finaux, mais souvent, cette phase du déploiement du système, qui garantit pourtant l'atteinte des fonctions énoncées plus haut, n'est pas réalisée, à cause des coûts prohibitifs de ces systèmes.

Le fonctionnement des Si semi-structurés est largement basé sur la participation et l'adhésion aux outils par les utilisateurs. Cette participation se fait selon une modalité de modération, et la richesse du système dépend aussi de l'animation qui en est faite par les responsables de certaines parties du système.

Dans les SI 2.0, l'ensemble du fonctionnement du dispositif est basé lui aussi sur la participation des utilisateurs, qui alimentent le système, renseignent des informations de profil, suivent des flux RSS. La matière de l'information contenue dans ces systèmes est ainsi dépendante du bon vouloir des utilisateurs. Le système et son succès repose sur ces derniers, c'est ce que l'on appelle une démarche orientée selon une méthode ascendante (ou Bottom-up).

3.2.2.6 Le type de sémantique induit par le système

D'après la distinction entre différents types de sémantique proposée dans (Zacklad, 2010a), appliquée à la définition des types de langages de capture sémantique, on distinguera la sémantique référentielle à l'œuvre dans les systèmes structurés, de la sémantique rhétorico-herméneutique qui est effective dans des systèmes semi-structurés, et éventuellement dans les systèmes d'information 2.0. Ainsi, d'après (Zacklad, 2010a) :

> *« La signification référentielle est basée 1) sur la dénotation du syntagme, un objet tangible auquel il se réfère « dans le monde », et 2) sur la dénotation d'expressions associant plusieurs syntagmes organisés en propositions, une valeur de vérité exprimée de manière binaire (vrai ou faux).*
>
> *(…)*
>
> *Le présupposé de la signification référentielle est essentiel dans les langages de programmation et pour interpréter le contenu des bases de données à vocation technique (…), les bases de données logistiques, financières, administratives (…). Ainsi dans un entrepôt industriel, la vocation principale de la base de données est de permettre d'attester de la présence d'un article dans un rayon donné. » (Zacklad, 2010a, p.9-10).*

Dans l'approche référentielle de la signification, les informations contenues dans le système sont fortement codifiées, et elles tendent à « épuiser » la signification du document, tandis que dans les documents « faiblement ou non codifiés », cette codification « ne concerne que des zones restreintes dédiées au référencement et aux liaisons explicites entre fragments » (Zacklad, 2006, p.9). Ainsi, dans une approche rhétorico-herméneutique de la signification, celle-ci est dépendante de ses aspects logiques, contextuels, et situationnels. Le contexte étant compris comme l'environnement textuel des termes à expliciter, et la situation correspondant à la situation d'énonciation (les personnes en présence, leurs relations, les objectifs poursuivis) (Zacklad, 2010a, p.10).

L'interprétation des documents contenus dans les SI semi- structurés ou 2.0 « ne peut pas se faire sans considérer les caractéristiques toujours pour partie variables de la situation transactionnelle » (Zacklad, 2006, p.9). Dans les systèmes semi-structurés, les éléments contextuels et situationnels sont portés par le système, qui permet des dispositifs de discussion instantanée, d'échanges désynchronisés comme dans les forums, et de classifications construites collaborativement, qui permettent de désambiguïser les aspects situationnels et contextuels des termes employés dans les échanges. Dans les SI 2.0, le même type de relation à la signification est à l'œuvre, alors que l'usage strictement personnel permis autorise une sémantique proprement personnelle, intégrant des relations référentielles aux objets décris quasi-confidentielles.

3.2.2.7 Les SOC disponibles dans ces SI.

Les SOC d'après (Zacklad, 2010c, p.2), « sont des ensembles de termes codifiés constituant un lexique disposant ou non de règles d'association explicites et permettant de faciliter les opérations de gestion portant sur des contenus documentaires ou sur les caractéristiques de situations non préalablement documentées (modèle documentaire ou modèle de situation). »

On peut étudier ces différents SOC selon divers critères (Zacklad, 2007a, 2011) mais cela dépasse le cadre de cette partie. Ainsi, en accord avec les différentes caractéristiques présentées précédemment, dans les SI structurés, on trouvera plus fréquemment des ontologies, des thésaurus, des index automatiques au service de moteurs de recherche, des classifications hiérarchiques classiques, ou à facettes rigides et des systèmes de modélisation généralistes et spécialisés (bien que ces derniers fassent, d'après (Zacklad, 2010a), plutôt

partie d'un ensemble plus large dénommé « appareils de capture sémantique », car ils présentent généralement un degré de structuration plus élevé par rapport aux autres SOC que nous allons évoquer. Dans les SI semi-structurés, on trouvera des SOC comme les classifications à facettes souples, les ontologies sémiotiques, des cartes de thèmes multi-points de vue. Enfin, dans les SI 2.0, on trouvera également des classifications à facettes souples, et des folksonomies.

Cette catégorisation des SI et des dispositifs informationnels correspondant, reste une schématisation. Il faut garder à l'esprit qu'un système de GED peut, par exemple, intégrer un système de *groupware*. Ainsi, la tendance la plus répandue est la coexistence de plusieurs dispositifs informationnels. Cette coexistence peut produire des difficultés pour les utilisateurs, en termes d'accessibilité technique (par exemple l'utilisation d'un VPN), de manipulation, ou de démarche cognitive, chaque outil façonnant « une certaine attente de comportement qui exige de la personne une connaissance de ses codes techniques, sémiotiques, sémantiques et organisationnels ». (Guyot, 2009)

Cependant, d'après la dernière caractéristique, on constate que la majorité des SOC est destinée à entrer dans des SI structurés, et ceci logiquement, puisqu'une de leurs missions est bien de normaliser les termes d'indexation et de recherche à l'aide d'un langage contrôlé, afin d'assurer un service de recherche d'information le plus précis et exhaustif possible.

On voit apparaître la place prépondérante des systèmes structurés qui remplissent des fonctions spécifiques de rationalisation des moyens par la normalisation et le contrôle. Cependant, ce formalisme des SI structurés ne présente pas uniquement des avantages, et nous allons nous intéresser dans la partie 3.5.10.2 à leur impact sur l'activité de gestion de l'information, comprise comme une activité de soutien aux activités principales correspondant au métier (qui ne concerne pas directement la gestion de l'information, et qui constitue un moyen de l'exercer).

Ainsi, si l'activité d'information peut être considérée comme secondaire (Guyot, 2002, 2004, 2006), elle conserve « toutes les caractéristiques d'une activité à ceci près qu'elle fait partie des activités secondaires de soutien au travail, même si dans certains métiers elle en constitue le cœur ». (Bailly *et al.*, 2002, cité par Guyot, 2004).

3.3 Qu'est-ce qu'une activité ?

Nous tentons ici de donner des pistes d'analyse de l'activité à partir de la Théorie de l'Activité et des apports de la psychologie du travail et de l'ergonomie. La Théorie de l'Activité trouve ses origines dans la philosophie allemande classique, les écrits de Marx et Engels, et dans les travaux soviétiques russes de la psychologie historico-culturelle de Vygostki, Leontiev et Luria qui date des années 1920 (Engeström, 1999).

Il nous faut préciser ici que l'accès aux textes du chef de file de cette théorie, Vygotski, étant difficile (sa thèse sur la Psychologie de l'Art n'a été traduite en français qu'en 2005, par exemple), nous nous sommes basés sur la littérature d'auteurs comme (Engeström, 1999), ou (Nardi 1996) qui tente d'utiliser cette théorie pour la conception de systèmes d'interaction homme-machine ; (Liccope, 2008) qui propose un travail positionnant les autres théorie qui traitent de l'activité. Nous avons aussi bénéficié des recherches publiées dans un numéro spécial de la revue « Activités » intitulé « Activité et action/ Cognition située », de 2004, qui rassemble des auteurs comme Theureau, Béguin et Clot, Salembier et Zouinar, et enfin des écrits de (Clot, 2008, 1995) qui s'inspirent des travaux de Vygotski.

Dans cette théorie, l'objectif est de comprendre les interrelations entre individus, et avec les artefacts dans les activités quotidiennes. Elle n'offre ni technique ni procédure finalisées pour la recherche, mais plutôt des outils conceptuels qui doivent être concrétisés en accord avec la nature spécifique des objets que l'on observe (Engeström, 1993). La notion de conscience y est centrale et est définie comme le phénomène qui unifie l'attention, l'intention, la mémoire, le raisonnement et le langage.

Dans la partie suivante, nous présentons succinctement la Théorie de l'activité, qui sera détaillée davantage en partie 3.4, ainsi que deux autres théories de l'activité : les courants de la Cognition Distribuée et de l'Action Située.

3.3.1 Les composants de l'activité

Dans cette Théorie de l'activité, l'unité d'analyse est l'activité. Elle est composée d'un sujet (individu ou groupe engagé dans une activité), d'un objet (qui est détenu par le sujet et qui motive l'activité en lui donnant une direction spécifique), d'actions et d'opérations. Les actions sont des processus orientés par des buts qui doivent être entrepris pour remplir l'objet

(la mission). Ces buts sont conscients, et différentes actions peuvent être entreprises pour réaliser un même but. Les actions ont un aspect opérationnel, elles peuvent devenir routinières et inconscientes au fur et à mesure de la pratique. Les opérations dépendent des conditions dans lesquelles les activités sont réalisées.

Les objets (motivations) peuvent se transformer au cours de l'activité. Ce ne sont pas des structures immuables. Ils ne changent pas d'un instant à l'autre: il y a une forme de stabilité dans les objets, et leurs changements ne sont pas triviaux.

Cependant, les composants de l'activité peuvent changer dynamiquement quand les conditions de l'activité changent.

D'après (Licoppe, 2008, p.291) :

> *« À travers son objet, l'activité fait système. L'objet de l'activité est en effet organisé hiérarchiquement selon trois niveaux, celui de l'activité individuelle ou collective dans son ensemble orientée vers un besoin ou un mobile, celui des actions intermédiaires qui participent de son accomplissement et sont orientées par des buts partiels et conscients, et enfin, celui des opérations, c'est-à-dire des accomplissements non réflexifs gouvernés par les contingences de la situation. (…) Dans cette perspective, les artefacts ne sont pas traités sur le même plan que les humains (au contraire de ce qui se passe avec la cognition distribuée), mais ils jouent le rôle essentiel de médiation entre le sujet de l'activité et l'objet visé par celle-ci. »*

3.3.2 Les artefacts et le contexte

Une idée centrale de cette théorie s'exprime dans les médiations assurées par les artefacts.

Ces artefacts sont des instruments, signes, langages, machines, créés par les individus pour contrôler leur propre comportement. Ils portent avec eux une culture et une histoire et possèdent des structures persistantes à travers les activités qui se déroulent dans le temps et l'espace.

La notion de contexte, est elle aussi importante dans la Théorie de l'Activité. Ainsi l'activité elle-même est le contexte. Le contenu d'un système d'activité est constitué de l'objet, d'actions, d'opérations qui composent le contexte.

Le contexte se constitue par la mise en œuvre de l'activité qui implique des personnes et des artefacts. Le contexte n'est pas vu comme un conteneur extérieur ou comme une coquille à

l'intérieur de laquelle les personnes se comportent d'une certaine manière. Les personnes le génèrent de manière consciente et délibérée, en partie à travers l'objet. Il est à la fois intérieur (objets) et extérieur aux personnes, en impliquant des artefacts, les autres personnes, et des cadres spécifiques. Il ne peut pas être réduit à une énumération de personnes et d'artefacts.

> *« Il y a une relation transformative entre les personnes et les artefacts. Le sujet n'est pas donné à l'avance, ainsi ses fonctions psychologiques et le plan de sa conscience se construisent dans le cadre de l'activité et du type de médiation à travers lequel s'accomplit celle-ci. La transformation ou le déplacement des médiations symboliques ou artefactuelles employées par le sujet de l'activité est donc inséparable d'une reconfiguration de celui-ci et de ses fonctions psychologiques. » (Licoppe, 2008, p.291).*

3.3.3 Le développement de l'activité

Ce développement est double: il concerne non seulement l'activité mais aussi le sujet qui est affecté par l'activité.

> *« C'est ce qui se passe lorsqu'une activité est tout d'abord accomplie aux moyens d'instruments externes (qui étendent les capacités du sujet) et dans un contexte de socialisation, avant que les compétences correspondantes ne soient internalisées. Cette internalisation signale la capacité du sujet à accomplir l'activité par lui-même et sans recourir à des moyens externes, c'est-à-dire son développement. » (Licoppe, 2008, p.291)*

Dans les années 90, Engeström étend le formalisme de cette théorie au fonctionnement de systèmes d'activité collectifs. Ainsi, il propose le modèle suivant de ce système d'activités collectif (Engeström, 1999).

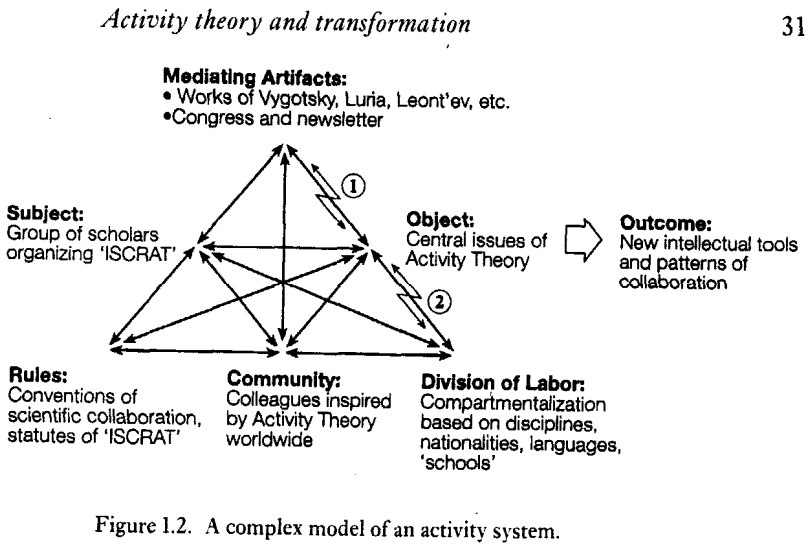


Figure 11. Représentation de l'activité d'Engeström (1999)

Nous avons présenté ici les fondamentaux de la Théorie de l'Activité. Nous allons poursuivre en présentant brièvement deux autres théories existantes de l'activité pour expliquer notre choix, puis nous prolongerons cette partie avec les développements apportés par (Clot, 2008, 1995) à cette démarche d'analyse de l'activité.

3.3.4 La théorie de la cognition distribuée

D'après Licoppe (2008), le modèle de la cognition distribuée est issu de la rencontre, en Californie du Sud, entre anthropologie et sciences cognitives, respectivement incarnées par E. Hutchins et D. Norman.

L'unité d'analyse dans cette théorie, d'après (Nardi, 1996), est le système cognitif composé de personnes et d'artefacts. Elle vise à étudier le fonctionnement de ce système cognitif. Il s'agit d'une vision fonctionnelle de ce système. Elle cherche à comprendre comment se fait la coordination entre les individus et les artefacts.

Ainsi, la cognition distribuée considère que les fonctions et les capacités cognitives sont des propriétés caractéristiques de collectifs hétérogènes composés d'humains et d'artefacts. Les humains et les artefacts sont des agents dans un système et sont mis sur le même plan (Licoppe, 2008). Dans ce cadre, le modèle qui est l'œuvre est celui du stimuli-réponse, ou réflexe, qui peut être conditionné pour une recherche d'économies cognitives. La recherche de

ces dernières implique un lien direct et non déterministe entre action et perception d'une part et, que, d'autre part, entre action et cognition il n'y aurait pas nécessairement de processus délibératif ou évaluatif complexe. Il s'agit d'accélérer l'effectuation de la « réaction » en exploitant des jugements de bas niveau permettant des économies cognitives (par exemple des aiguilles qui sont des repères pour l'ouverture du train d'atterrissage d'un avion). « La décision d'engager une action est rapportée à un jugement de perception spatiale qui allège la charge cognitive des pilotes et, en particulier, leur mémoire ». (Licoppe, 2008, p.290). Dans cette théorie, c'est « la recherche de ces économies cognitives [qui] guide la coévolution des structures sociales et procédurales qui caractérisent l'activité et de l'environnement dans lequel celle-ci se déroule. » (Licoppe, 2008, p.290) et c'est la disponibilité des informations qui déclenche l'action.

Cette conception de l'activité a donc un impact sur la conception et l'évolution des artefacts, vus comme des composants d'un ensemble fonctionnel couplé à un opérateur humain, dans une recherche d'économie des ressources de l'opérateur humain, éventuellement par le biais de l'usage d'automatismes machiniques ou de réflexes humains conditionnés.

3.3.5 La théorie de l'action située

« L'ethnométhodologie développée par H. Garfinkel en Californie dans les années 1960 (…) s'est tournée assez rapidement vers les activités de travail » (Licoppe, 2008, p.292).

Dans cette théorie, l'activité humaine est considérée comme ayant une nature émergente, contingente. La structure de l'activité ne la précède pas, mais apparaît au fil de l'activité dans l'immédiateté de l'activité. L'unité d'analyse ne correspond ni à l'individu, ni à l'environnement, mais plutôt à la relation entre les deux. La théorie de l'action située se focalise sur les activités quotidiennes de personnes agissant dans un cadre, ce qui oblige l'analyste à observer les flux de l'activité en cours, le déroulement le l'activité réelle, dans ses cadres réels (Nardi, 1996). Elle met donc l'emphase sur l'étude de l'improvisation et des contingences, ce qui ne permet pas d'étudier en même temps les phénomènes stables, durables qui persistent à travers les situations. L'accent est mis sur les réponses individuelles créatives des individus dans des situations considérées comme uniques.

Au niveau méthodologique,

elle « prône « l'indifférence » quant à la validité des descriptions qui sont produites par les participants (…). Les descriptions (comme d'ailleurs les règles et les plans ou encore les artefacts) sont traitées sous l'angle de leur rapport « indexical » à l'action : leur élaboration et leurs usages sont sujets aux mêmes contingences circonstancielles et interprétatives que les actions et les circonstances qu'elles décrivent. (…). Pour l'ethnométhodologie, il est essentiel que l'analyste devienne lui-même un membre compétent dans l'accomplissement de l'activité qu'il étudie. » (Licoppe, 2008, p.292).

3.3.6 Points de différentiation pour le choix d'une théorie de l'activité

Dans ce qui suit, nous évoquons les perspectives de chacune des trois théories présentées, sur les thématiques en lien avec l'analyse de l'activité, et nous présentons les arguments qui nous ont conduits à choisir la Théorie de l'Activité comme cadre de notre travail.

3.3.6.1 La structure de l'activité

La Théorie de l'Activité et la Cognition Distribuée offrent une prise en compte et un traitement des buts et motivations, tandis que dans l'Action Située, l'activité est contingente, improvisée, les buts n'y sont pas des conditions de l'action. Selon la théorie de l'Action située, une analyse de l'expérience directe dans le monde réel mène à considérer que les buts sont construits, souvent par une interprétation verbale, après la réalisation de l'activité, rétrospectivement et réflexivement, comme une justification de cette activité.

Dans la Théorie de l'activité, l'activité est formée premièrement et principalement par l'objet, détenu par le sujet(le but), qui implique des dimensions psychologiques et une intentionnalité. Le but précède l'action et la motive. Différents objets mènent à des activités différentes. Intérêts, intentions et savoir antérieur participent de l'objet et du développement de l'activité.

Dans la Cognition Distribuée, le but du système n'implique pas de conscience, il s'agit du concept abstrait systémique de recherche d'économie.

Dans la théorie de l'Action située, concernant la planification, les plans sont vus comme des artefacts de notre raisonnement à propos de l'action, et non comme des mécanismes génératifs de l'action. Ce sont des rationalisations post hoc de l'action dont le sens ne peut émerger que dans l'immédiateté d'une situation donnée. (Nardi, 1996).

Cela revient à poser le problème de l'intentionnalité, l'activité et ses valeurs étant générées simultanément. C'est la situation en cours dans laquelle on est littéralement immergé qui

donne une direction à nos actions, et non une réflexion, ou l'usage de représentations mentales durables. Chaque activité est par définition unique, constituée par la confluence de facteurs particuliers, qui ensemble forment une situation. Le modèle de l'Action Située se confine aux niveaux des actions et opérations par rapport au modèle de la Théorie de l'Activité. Le modèle de la relation entre acteur et environnement est de type réaction ou encore de type stimulus (situation) - réponse (actions et opérations). Il n'y a pas de tentative de prédire ou de cataloguer des réactions invariantes dans la théorie de l'Action Située, les situations sont vues comme évoluant de manière imprédictible.

On constate l'absence de contrôle de l'activité par les opérateurs dans les situations de travail étudiées selon cette théorie, ainsi que leur absence d'autonomie dans le contrôle, que l'on distingue de l'externalisation des structures de contrôle dans des supports matériels, ou organiques (contrôle par les objets, par les artefacts organisationnels normatifs, ou par les autres agents). (Clot & Béguin, 2004).

On constate aussi une forme de passivité des acteurs qui sont privés de tout crédit à l'égard des plans ou des formes d'anticipations qu'ils peuvent élaborer avant d'agir. L'acteur est privé d'intention, et ne fait que réagir à des contingences dans le cadre de cette approche de l'activité. Il ne peut donc construire aucun repère.

3.3.6.2 Les invariants de l'activité

Ces invariants de l'activité peuvent se traduire dans les artefacts, les institutions, les valeurs culturelles, les manières d'effectuer une opération (les routines). Ce sont des motifs dans l'activité qui font l'objet d'une forme de répétition, à travers plusieurs situations étudiées.

Dans la Théorie de l'Activité et la Cognition Distribuée, les structures persistantes sont une valeur centrale. La Théorie de l'Activité est orientée vers le développement historique et culturel de l'activité et du rôle de médiateur des artefacts, ce qui implique une recherche des formes répétitives et aussi de ce qui varie dans des situations variables.

D'après (Clot & Béguin, 2004), dans le cadre de la Théorie de l'Activité, au sujet des invariants, on retiendra ce qui suit :

> *« (…) pour le dire à la manière de Leontiev, la formation d'une action ne se ramène nullement à la formation d'invariants, et à plus forte raison n'en est pas le résultat. C'est plutôt le contraire : ils sont le produit d'actions répétées au cours d'activités non réitérables (…) Du coup, les invariants n'engendrent pas l'activité, ils ne l'expliquent pas. C'est elle qui s'explique avec eux, dans tous les sens du terme. En ce sens, le processus de répétition est une répétition sans répétition (…). D'une part parce qu'il n'y a pas d'invariant mis en œuvre*

sans une histoire chaque fois singulière de l'activité. D'autre part parce que chaque activité (singulière) affecte, que ce soit sous la forme d'une consolidation ou d'un développement, le répertoire des fonctionnalités de l'invariant (…).

Regardée ainsi, l'action est moins située dans la situation que dans l'activité. Ce sont les obstacles, les discordances, les conflits objectifs, subjectifs ou intersubjectifs rencontrés dans l'activité, qui génèrent une intensité plus ou moins forte de tension, et qui invitent le sujet à mobiliser et développer les invariants. » (Clot & Béguin, 2004, p.45)

Dans la Cognition Distribuée, la recherche des invariants est une question centrale, notamment celles qui concerne les artefacts. Les propriétés des artefacts sont vues comme persistantes à travers les situations étudiées. Ils peuvent être conçus et reconçus, en respect de leur propre structure, mais aussi en respect de situations spécifiques d'usage. A ce titre, les études issues des travaux de la Cognition Distribuée s'intéressent aux pratiques de travail qui couvrent plusieurs situations.

On a pu remarquer que la recherche d'invariants dans la théorie de l'Action Située n'est d'une part, pas une priorité étant donnée la conception contingente de l'activité qui est sa spécificité. D'autre part, l'introduction de la notion de routine dans cette théorie sème le trouble. (Nardi, 1996).

3.3.6.3 Conception symétrique ou asymétrique des artefacts et des humains

Dans la Théorie de l'Activité, l'accent est mis sur l'importance de la conscience et de la motivation, qui ne s'observent que chez l'humain. Les artefacts sont donc différents des humains. Les artefacts sont des médiateurs des pensées humaines et de leurs comportements. Humains et artefacts ne sont donc pas équivalents. Les artefacts dans ce cadre sont définis comme des instruments au service des activités, ou encore comme médiateurs de la cognition.

Au lieu d'envisager les gens comme des nœuds dans un système, à la manière de la Cognition Distribuée, les humains sont actifs dans la Théorie de l'Activité : ils contrôlent leurs outils pour des buts créatifs, plutôt qu'ils ne sont considérés comme des automates dont les opérations doivent être automatisées, ou comme des nœuds pour lesquels le droit à la dignité et au privé ne sont pas garantis. (Nardi, 1996). Dans le courant de la Cognition Distribuée, les humains et les artefacts sont mis sur le même plan, ce sont des agents dans un système. L'agent humain est vu comme une chaîne mécanique prédictible.

Dans la théorie de l'Action Située, les humains sont vus dans le cadre des études menées comme des « numéros réactifs », (Nardi, 1996), et passifs.

3.3.6.4 Interaction et communication

Dans le courant de l'Action Située, certains présupposés sont à l'œuvre concernant les interactions, d'après (Salembier et Zouinar, 2004, p.66) :

> *1. « la réussite de l'interaction repose sur une production de l'intelligibilité mutuelle, production favorisée par l'accès mutuel aux ressources disponibles dans un environnement matériel, social et culturel particulier et plus ou moins stabilisé ;*
>
> *2. les significations portées par le langage mis en œuvre au cours de cette production de l'intelligibilité mutuelle ne reposent pas seulement sur une interprétation des unités linguistiques utilisées, mais aussi sur une interprétation d'autres éléments de la situation. »*

Dans l'Action Située, « une condition fondamentale de la production d'intelligibilité mutuelle tient à l'externalisation des représentations, des actions, et de leurs supports, qui va permettre un accès conjoint à ces ressources. De ce point de vue, la situation idéale est la co-présence physique d'agents coopérants. L'accès conjoint aux ressources, malgré son importance, n'est néanmoins pas suffisant pour garantir les conditions d'une coordination des activités individuelles et d'une coopération entre agents efficaces. » (Salembier et Zouinar, 2004, p.70).

Les autres facteurs qui vont influer sur l'intelligibilité mutuelle sont les suivants :

- la sensibilité à l'accomplissement en temps réel de l'activité en train de se faire des autres acteurs
- la capacité des agents à interpréter correctement les événements de la situation
- le niveau de disponibilité des agents
- la capacité à déterminer quelles sont les informations utiles à transmettre

Dans l'école historico – culturelle (la Théorie de l'Activité), on insiste sur la communication comme un aspect inhérent à toutes les activités orientées-objet (Engeström, 1999), tandis que dans la Cognition Distribuée, la disponibilité des informations, qui est elle-même dépendante de la structuration sociale des activités, est le déclencheur de l'action. (Licoppe, 2008).

3.3.6.5 Rôle du contexte

De manière générale, la psychologie ergonomique subit une double influence, celle du paradigme cognitiviste d'une part, et celle de la Théorie de l'Activité d'autre part. Dans le paradigme cognitiviste, il est postulé que « l'environnement social, culturel, ainsi que la nature physique du cerveau ne posent aucune contrainte sur les structures de la pensée et [il] en [est] déduit que l'objet d'étude de la psychologie cognitive est le *traitement de l'information*. » (Italiques des auteurs), (Darses, Hoc, Chauvin, 2004, p.223).

Concernant la Théorie de l'Activité :

> *Elle « introduit de facto les caractéristiques situées de la signification – finalisée par des objectifs d'action, matériellement inscrite dans des artefacts et culturellement empreinte d'une histoire collective – comme des composantes centrales pour comprendre l'activité psychologique humaine et pour modéliser les processus de construction de signification. » (…) Aujourd'hui, on s'accorde sur le fait que les structures de la pensée sont biologiquement, culturellement et socialement contraintes. (…) Le rôle dimensionnant du contexte et de l'environnement social et culturel dans la détermination de l'action fait donc partie intégrante des cadres d'interprétation de la psychologie ergonomique. » (Darses, Hoc, Chauvin, 2004, p.223, 224)*

L'objet de la psychologie ergonomique est donc « l'interaction entre ces éléments de l'environnement (le « contexte ») et la dynamique des activités mentales », d'après (Darses, Hoc, Chauvin, 2004, p.224).

Dans l'Action Située, « l'activité se construit donc par un processus interactionnel et circonstanciel, sur la base d'interactions locales situées dans un contexte donné. » (Darses, Hoc, Chauvin, 2004, p.225).

La problématique de la « validité écologique » des analyses fondées sur le courant de l'Action Située prête à quelques réflexions. Ainsi la validité écologique vise à vérifier, du point de vue méthodologique, que l'objet étudié est bien inclus dans la situation expérimentale. Etant donné que les situations observées par cette méthode sont des construits du chercheur qui définit le cadre de ce qui est observé et ce qu'il observe, il est possible que la validité écologique ne soit pas optimale. Ce sont des situations construites, cadrées par le chercheur, sans inclusion des discours des observés dans les expérimentations. Ce qui nous conduit à réfléchir sur la reproductibilité des études qui permet de valider des apports scientifiques et de faire des généralisations sur cette base, processus qui nous paraît difficilement compatible avec l'Action Située, la méthodologie des études étant propre à chaque chercheur.

Les études menées dans le cadre de la cognition distribuée « portent sur des situations professionnelles complexes où la dimension technique du système est prépondérante » (Darses, Hoc, Chauvin, 2004, p.226) comme dans le cas du pilotage d'avion ou de navire. Ainsi, dans cette théorie, le contexte est constitué du système technique que l'on cherche à optimiser.
Nous reviendrons plus en détail dans la suite de cette partie sur une définition plus complète du contexte.

3.3.6.6 Sur le plan méthodologique

Au niveau méthodologique, le courant de l'Action Située autorise pour ses études les matériaux suivants : vidéo (pour enregistrer les comportement et conversations), la surveillance en présence des personnes étudiées pour observer leurs mouvements, le traçage des artefacts et instruments (pour auditer l'usage), l'analyse basée sur les événements (pour traquer les taches individuelles), mais elle considère les Interviews avec prudence, le discours étant vu comme idéalisé, rationnalisé, tandis que le chercheur lui-même construit en partie le phénomène étudié en posant les limites de l'observation et en projetant des cadres d'analyse sur les situations étudiées. Les buts sont quant à eux perçus comme subjectivement rapportés, comme une interprétation verbale. Comme évoqué plus haut, les plans sont des constructions rétrospectives, et ce qui constitue une situation à étudier est un construit jugé pertinent par le chercheur. Ni la généralisation, ni la comparaison des observations tirées de l'étude des situations n'est possible.
La Théorie de l'Activité, considère les interviews comme propres à refléter l'activité. La capacité à rendre conscient, même partiellement, le niveau des opérations est un aspect du dynamisme des niveaux de l'activité proposé par cette théorie.

Ces différents éléments nous ont amenés à choisir la Théorie de l'activité comme cadre théorique pour l'analyse de l'activité de gestion de l'information, et plus globalement pour tenter de caractériser les activités métier.

3.4 Théorie de l'Activité : une conception dynamique de l'activité

Dans cette partie nous présentons les développements récents des approches de la psychologie du travail d'Yves Clot, et de l'ergonomie, basées sur les fondamentaux de la Théorie de l'Activité telle que caractérisée précédemment.

3.4.1 Travail prescrit et travail réel

Une distinction bien admise aujourd'hui, dans la recherche en psychologie du travail et en ergonomie cognitive, consiste à décrire les prescriptions concernant les activités de travail comme faisant l'objet d'une interprétation de la part des opérateurs.

Ainsi, une activité fait souvent l'objet d'une description formelle, notamment de son objet (objectif), qui a été élaborée par les concepteurs de l'activité. Ces derniers peuvent être vus comme les managers, tandis que celui qui réalise l'activité doit s'approprier cette description pour l'opérationnaliser, c'est à dire identifier les actions à mener et les opérations à conduire pour le succès de l'activité, selon la décomposition de l'activité en actions et opérations proposée précédemment.

Ainsi, d'après (Leplat, 1997), le travail prescrit correspond à l'ensemble des activités et tâches demandées à un opérateur sous forme de consignes ou d'objectifs à atteindre dans des conditions données en échange de sa rétribution.

Comme nous avons pu le remarquer de nombreuses théories tentent de définir les concepts d'activité et de tâche sans les épuiser. Dans la conception de l'activité et de la tâche proposée par Leplat (1997), l'activité se décompose en étapes, en tâches et en sous tâches, ou tâches intermédiaires, selon des degrés de raffinement variables. « La tâche est communément définie comme un but à atteindre dans des conditions déterminées. » Leplat (1997). Nous évoquerons donc dans la suite de cette partie, la tâche comme étant équivalente aux actions définies dans la Théorie de l'Activité.

La tâche prescrite est une « formulation », le fruit du travail, en amont, des concepteurs de la tâche. Elle peut « être conçue comme le modèle de la tâche à réaliser » destiné à un opérateur. Elle comporte donc une part d'implicite et nécessite une interprétation de la part de celui qui va la réaliser, une redéfinition qui va consister « à opérationnaliser celle-ci en fonction des conditions présentes » pour aboutir, selon Leplat (1997), à une tâche redéfinie. La tâche redéfinie prend en compte les objectifs de la tâche prescrite mais aussi d'autres buts, notamment des buts personnels (promotion, intégration sociale, santé...). Elle constitue un

compromis et va guider l'action qui caractérisera la tâche effective correspondant au travail réel. Le travail réel inclut aussi les imprévus et les contingences.

3.4.2 L'activité et le sujet : des éléments hétérogènes

Selon (Clot, 2008), « le sujet n'est pas tout entier dans son activité précisément parce qu'il est hétérogène, sujet de plusieurs activités psychiques ou pratiques qui se recoupent en lui, plus ou moins liées ou déliées, plus ou moins disponibles aux réorganisations fonctionnelles que supposent son développement. » (Clot, 2008, p.27)

Le sujet s'inscrit dans une perspective historique: dans l'activité, il « dénoue et renoue de façon unique les liens tissés entre toutes les activités qu'il a dû tenter de rendre compatibles en dehors de lui et de son histoire. » (Clot, 2008, p.25)

Il propose aussi une conception dialogique de l'activité et du sujet, en intégrant les autres sujets que chacun porte en soi comme objet et moyen d'une activité propre. Ainsi, « il est préoccupé par l'activité d'autrui (…) Autrui est dans l'objet. Dans la pratique l'objet de l'activité du ou des sujets est le lieu d'une collision entre activités ou, au moins, d'un échange. » (Clot, 2008, p.19). L'auteur affirme également que l'altérité est constitutive de la subjectivité.

On remarquera que l'activité est aussi un élément hétérogène : elle a des fonctions psychologiques et sociales, elle est vue comme médiatisée et médiatisante, elle a du sens. Enfin, elle fait l'objet d'un développement.

Concernant ses fonctions psychologiques et sociales, (Clot, 2008, p10) précise que « c'est l'activité qui lie ou délie l'individuel et le social, le sujet et l'organisation du travail, les sujets entre eux et ces sujets avec les objets qui les mobilisent. Elle est l'arène et le siège où ils passent l'un dans l'autre, la plus petite unité de l'échange social. »

L'activité est médiatisée par des artefacts, et est médiatisante par son pouvoir de relancer d'autres liaisons. C'est ici l'apport de (Clot, 2008), qui consiste à étudier comment l'activité peut se développer dans une dynamique permanente de transformation des situations sur des plans qui peuvent être divers, sur le plan psychologique personnel, sur le plan de l'apprentissage ou sur la réalisation de l'objet de l'activité. Cette dynamique peut être orientée négativement, comme nous le verrons ensuite, ou positivement pour le sujet. C'est ce que l'auteur appelle le développement du « pouvoir d'agir ». Ce dernier dépend, en partie, de la

capacité de l'activité à en relancer d'autres, même si ce n'est pas sur le même objet ou sur le même plan, et de la plasticité subjective et psychologique du sujet, de sa « mobilité inter-fonctionnelle ».

Ce développement du pouvoir d'agir prend sa source dans le sens de l'activité, qui est entendu comme « le rapport de valeur que le sujet instaure entre cette action et les autres actions possibles » (Clot, 2008, p.9 ; 1995, p.140). Il s'agit d'un coût subjectivement estimé de la tâche. Il indique que :

> *« le sens de l'activité réalisée, de l'action en cours, se perd trop souvent quand disparaît dans le travail du sujet ou des sujets le rapport entre les buts auxquels il faut se plier, les résultats auxquels il faut s'astreindre et ce qui compte vraiment pour eux. » (Clot, 2008).*

3.4.2.1 Les sources du développement conjoint de l'activité et du sujet

Pour Yves Clot, l'activité n'est jamais une répétition qui relève du pur automatisme : elle est une « répétition sans répétition ». Il questionne l'opposition classique entre un fonctionnement ordinaire, comme un invariant ou une routine, et son opposé, l'activité créative. Il écrit ainsi :

> *« Le développement d'une activité ne se ramène nullement à la construction d'invariants par le sujet, à plus forte raison n'en est pas le résultat. (…) les invariants sont construits par l'expérience d'actions répétées dans des situations (…) jamais identiques. (…)*
>
> *Dans l'exercice professionnel un fonctionnement n'est jamais identique, même en situation stabilisée. Une répétition est toujours unique et le même fonctionnement accueille dans l'activité réelle des fonctionnalités différentes. (…) C'est dans le jeu entre fonctionnalités différentes du même fonctionnement dans des activités variées et répétées que le développement prend sa source. » (Clot, 2008, p.25)*

La construction ou la préservation du sens du travail passe par la possibilité de contribuer pour les opérateurs, par leur initiative, à la réalisation du travail, à l'organisation du travail, et éventuellement à la définition des tâches et des référentiels communs.

> *« vivre au travail, c'est donc pouvoir développer son activité, ses objets, ses instruments, ses destinataires, en affectant l'organisation du travail par son initiative. L'activité des sujets se trouve au contraire* ***désaffectée*** *lorsque les*

choses, en milieu professionnel se mettent à avoir entre elles des rapports qui leur viennent indépendamment de cette initiative possible. (…) cette désaffection diminue le sujet, le déréalise, non sans effets quant à l'efficacité même de son action, au-delà des effets sur sa santé » (Clot, 2008, p.7)

Clot insiste sur la préservation des marges de manœuvre des sujets qui se situent au niveau des actions dans la Théorie de l'Activité, pour la conservation d'un réservoir de possibilités de prises d'initiatives qui caractérisent les situations où le « pouvoir d'agir » est effectif et préserve le sujet d'une forme d'aliénation. Ces marges de manœuvre se situent au niveau des actions, puisque l'objet de l'activité provient des prescriptions qui définissent les objectifs du travail, tandis que les actions correspondent aux différentes étapes de réalisation de l'activité.
Du coté du sujet, les conditions de son développement à travers l'activité de travail, qui est vu comme une « métamorphose des fonctions psychologiques », sont liées à sa capacité à « être affecté ». Ainsi, à propos de la subjectivité qui est le lieu de ce pouvoir d'être affecté, Yves Clot écrit :

*« La subjectivité est sans doute moins une disposition constitutive du sujet que le pouvoir d'être affecté dont chacun dispose plus ou moins en fonction de son histoire singulière. C'est son incomplétude qui rend le sujet disponible au développement de l'activité et non une puissance d'agir autonome. C'est sa vitalité dialogique interne qui le prépare à supporter ou à saisir les inattendus du réel devant lesquels il doit **se** déterminer. » (Clot, 2008, p.27)*

Un sujet est donc capable d'être affecté, « grâce à sa mobilité inter-fonctionnelle ». Il s'agit de :

« sa capacité à mettre une fonction au service d'une autre, ses émotions, par exemple au service de sa pensée et inversement. Ces migrations inter-fonctionnelles sont, dans le psychisme, un instrument décisif pour donner un destin dynamique à l'affect. Celui –ci « défait » le sujet, le déconstruit s'il ne se métamorphose pas. Et il ne peut le faire qu'en se réalisant dans un jeu ouvert entre expression des émotions et représentations mentales, entre intellect et sentiment. Ce mouvement mesure la plasticité fonctionnelle du sujet. » (Clot, 2008, p.27)

Le pouvoir d'agir concerne l'activité.

« Il mesure le rayon d'action effectif du sujet ou des sujets dans leur milieu professionnel habituel, ce qu'on peut aussi appeler le rayonnement de l'activité, son pouvoir de recréation. (…)

Il augmente ou diminue en fonction de l'alternance fonctionnelle entre le sens et l'efficience de l'action où se joue le dynamisme de l'activité, son efficacité.

Cette efficacité n'est pas seulement l'atteinte des buts poursuivis. C'est tout autant la découverte de buts nouveaux. » (Clot, 2008, p.13)

Ainsi, du point de vue méthodologique, ce que tente de réaliser la psychologie du travail, consiste à trouver des moyens de développer le pouvoir d'agir des opérateurs. Ainsi, ce courant de recherche vise à :

> *« inventer des dispositifs techniques permettant aux sujets de transformer leur expérience vécue d'un objet en objet d'une nouvelle expérience vécue. (…) C'est le fondement de théorique de la méthodologie historico-développementale : permettre au sujet de transformer des fonctionnements réalisés en objet d'un nouveau fonctionnement afin d'étudier le développement réel – possible et impossible – et ses principes. » (Clot, 2008, p.171)*

Nous avons présenté les fondements et les développements de la Théorie de l'activité afin de pouvoir clarifier un des objectifs scientifiques du projet de recherche évoqué dans la suite de cette partie, et qui consiste à déterminer les aspects du travail prescrit et du travail réel dans les activités de gestion de l'information. Nous poursuivons ici en mettant en relation la Théorie de l'Activité avec la théorie de la Sémiotique des transactions coopératives proposée par (Zacklad, 2005, 2006, 2008, 2009) dans le champ de la discipline des Sciences de l'Information et de la Communication. Elle propose, en effet, une approche pouvant être rapprochée, sur plusieurs points, de la Théorie de l'Activité.

3.4.3 Sémiotique des transactions coopératives

Cette théorie pose le concept de « transaction coopérative » comme central pour l'analyse des activités individuelles et collectives. Les transactions sont définies comme « des interactions productives, le plus souvent associées à des rencontres (mais pouvant aussi être largement asynchrones), permettant la transformation d'un artefact médiateur et des personnes partie prenantes afin de satisfaire une visée transactionnelle », (Zacklad, 2008, p.2).

On constate, d'après cette définition des transactions, la dynamique inhérente aux activités, qui, au cours de leur réalisation, transforment non seulement l'artefact, (qui joue aussi un rôle central dans cette théorie), mais également les personnes. Plus précisément, cette

transformation opère simultanément sur deux plans : les caractéristiques des artefacts médiateurs et le contenu expérientiel des personnes engagées dans la transaction. (Zacklad, 2008).

Cette théorie envisage les personnes engagées dans la réalisation d'une activité comme des bénéficiaires ou réalisateurs (ou co-bénéficiaire et co-réalisateur), « toutes les configurations de symétrie et d'asymétrie entre ces positions étant possibles. Une personne individuelle peut être engagée dans une transaction avec elle-même. » (Zacklad, 2006, 2008). On retrouve la réflexivité du sujet affirmée dans la Théorie de l'Activité et dans les approches d'Y. Clot - réflexivité admise dans cette théorie grâce à cette configuration individuelle de la transaction possible.

D'autre part, le rôle des artefacts, dans la Sémiotique des transactions coopérative est celui de médiateur. Ce rôle qui peut être rapproché de la médiation assurée par les artefacts dans la Théorie de l'Activité, médiation qui est définie comme telle : « toute transaction doit être médiatisée, soit par des gestes et les paroles des personnes en présence, soit « à distance », par l'intermédiaire d'artefacts pérennes circulant d'un corps à l'autre. » (Zacklad, 2006, 2008). Le schéma suivant rappelle ces différentes caractéristiques de la transaction coopérative.

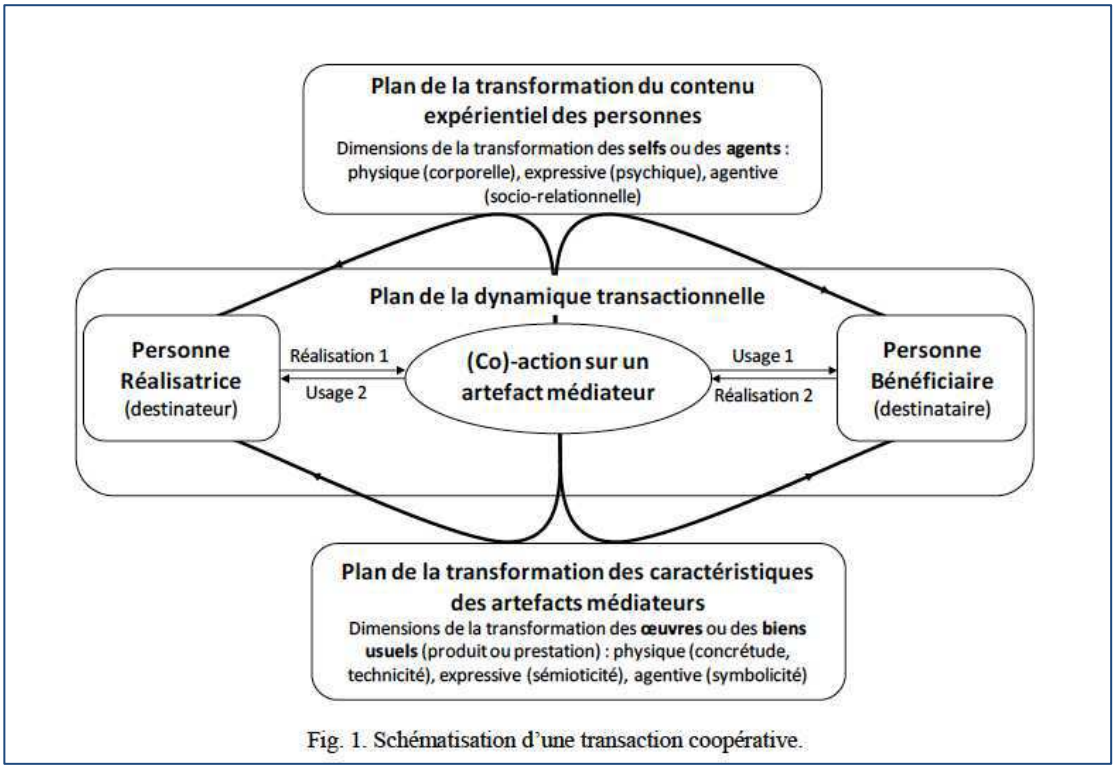


Figure 12. Schématisation d'une transaction coopérative de (Zackald, 2006)

Par ailleurs, « les transactions coopératives sont conduites à l'intérieur de situations transactionnelles elles-mêmes reliées par des flux transacstionnels. » (Zacklad, 2008, p.3).

Ces flux ont une « configuration rhizomatique » (Zacklad, 2006) qui les rend interdépendants, une transaction pouvant en relancer une autre à un point de sa réalisation. Cet historique de l'activité, ou « généalogie » signale les éléments de la transaction qui produisent un autre flux dans une sorte de « réaction en chaîne ». Ainsi :

> *« l'analyse de l'action dans une perspective transactionnelle est rendue particulièrement difficile par le caractère fragmenté et réparti des procès transactionnels. En effet, non seulement la réalisation d'une transaction est-elle distribuée selon différents paramètres spatio-socio-temporels, mais une transaction donnée est susceptible d'en relancer d'autres d'une manière plus ou moins systématiquement dérivée de la première. Pour reprendre les métaphores géologiques de Deleuze et Guattari (1980), nous considérerons que les transactions constituent des flux qui s'écoulent à des vitesses variables en s'entrecroisant à divers endroits. Toute situation d'activité effective peut ainsi être analysée en fonction de sa « généalogie », qui l'inscrit à la fois dans un procès transactionnel direct et manifeste, mais également en fonction de procès transactionnels antérieurs auxquels les participants à la situation ne font pas nécessairement référence de manière explicite » (Zacklad, 2007b, p.3).*

Le schéma suivant illustre cette configuration rhizomatique des flux transactionnels dans lesquels les situations transactionnelles se relancent entre elles, et en relancent de nouvelles.

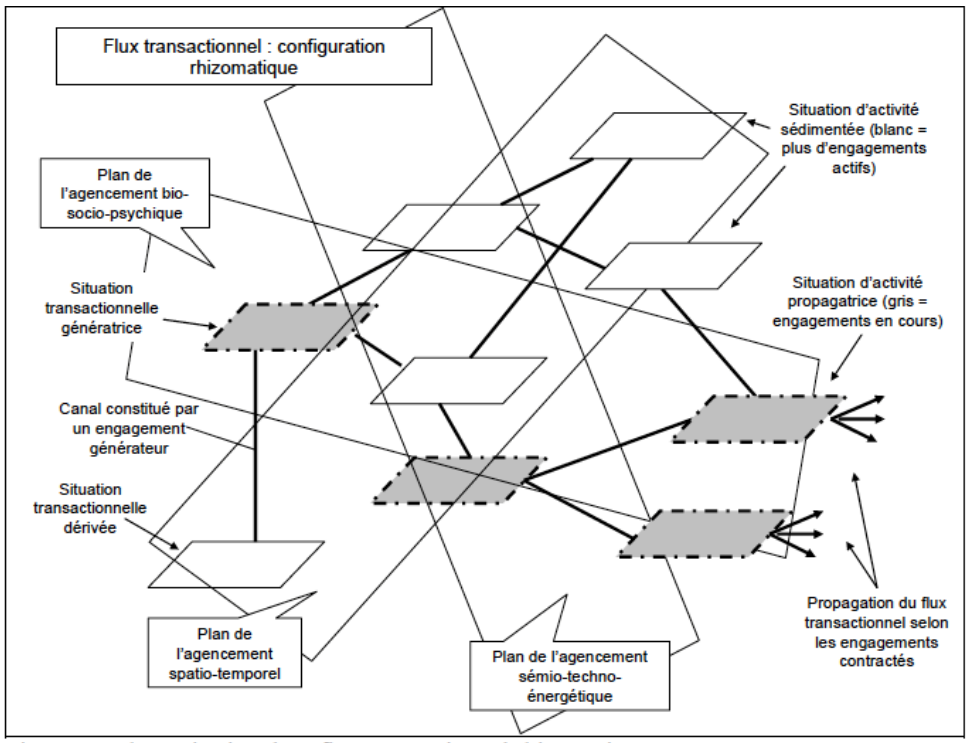


Figure 2. Schématisation d'un flux transactionnel rhizomatique.

Figure 13. Schématisation d'un flux transactionnel de (Zacklad, 2006)

Dans ce schéma, les nœuds sont des situations d'activité effectives, c'est à dire les différentes phases d'une ou plusieurs transactions. Les liens sont les relations « généalogiques » entre ces activités. On peut rapprocher cette vision de la genèse d'une activité de celle de la Théorie de l'Activité, qui replace la compréhension d'une activité dans l'histoire du sujet et de celle de ses autres activités.

(Zacklad, 2006) distingue deux types de situations d'activités : celles qui sont propagatrices et qui relancent de nouvelles activités, et celles qui sont sédimentées, qui n'en relancent pas. Concernant les situations d'activité propagatrices, nous les rapprochons de ce que (Clot, 2008) appelle le « pouvoir d'agir » bien que celui-ci ne concerne pas nécessairement une activité à proprement parler, mais aussi des aspects de prises de conscience psychologiques, ou des développements sur d'autres plans que celui de la réalisation d'une activité.

Les composants d'une situation transactionnelle sont donc les suivants, d'après (Zacklad, 2005) :

- un ou des réalisateurs
- un ou des bénéficiaires
- des paramètres de la situation (que nous détaillons ci-dessous)
- une (co)-production sémiotique

Remarquons, tout d'abord, que la présence d'un (ou des bénéficiaires) comme élément définitoire d'une situation transactionnelle rappelle la vision de (Clot, 2008, p.201), qui affirme qu'une activité est toujours dirigée, adressée à un destinataire, « elle fait d'autrui, non un contexte mais un subtexte de l'action ». La (co)-production sémiotique véhicule « un *contenu sémiotique* via un *médium* élaboré par les réalisateurs à l'intention des bénéficiaires » (Zacklad, 2005). Les paramètres de la situation transactionnelle, qui peut s'apparenter à ce que nous désignerons sous le terme de « contexte », sont composés des éléments énoncés ci-dessous, d'après (Zacklad, 2005). Ils nous permettront de commencer à définir cette notion particulièrement importante dans notre travail :

- un projet commun
- des relations sociales entre le bénéficiaire et le réalisateur
- un cadre spatio-temporel et des conditions contingentes

- le choix d'un médium
- des équipements techno-informationnels (les artefacts)
- les terrains représentationnels commun et individuel
- les compétences des réalisateurs et des bénéficiaires

Nous allons ici définir plus précisément certains des éléments de la liste présentée ci-dessus.

Les artefacts

Ils doivent offrir une « (1) représentation externe des buts, (2) des ressources documentaires contribuant à alimenter la production sémiotique et (3) des supports pérennes associés au médium et facilitant la circulation et le partage de celui-ci. » (Zacklad, 2005)

Les terrains représentationnels

Ils correspondent à « l'ensemble des représentations évoquées durant la transaction issues (1) des compétences et (2) des autres paramètres de la situation transactionnelle » (Zacklad, 2005). Ces terrains représentationnels qui portent les représentations des agents de la transaction, sont étudiés par l'ergonomie sous les termes de « référentiels communs ». Ils nous intéressent tout particulièrement, étant donné qu'ils conditionnent l'intercompréhension des buts, des propositions, des missions, des compétences, de l'histoire, et des représentations de ces derniers, pour les agents engagés dans une transaction, et pour le succès de la transaction.

Nous reviendrons sur ces « référentiels communs en partie 3.5.8.3. Nous allons, dans la suite de cette partie, tenter d'appliquer certains des éléments présentés de la Théorie de l'Activité, des travaux d'Y. Clot, et de la théorie de la sémiotique des transactions coopératives, à l'analyse des activités de gestion de l'information.

3.5 L'activité de gestion de l'information dans les organisations

D'après (Zacklad, 2008), dans le cadre des transactions coopératives, une organisation « peut être définie comme un ensemble de flux transactionnels récurrents obéissant à un ensemble de règles explicites et de conventions ». Tandis que pour (Guyot, 2006, p.17) une organisation correspond à la définition suivante :

Elle « - se structure autour de buts qu'elle se donne : système contraint par des objectifs internes de rentabilité, de compétitivité, et par des normes internes, inséré dans un environnement qui lui fixe une partie de ses contraintes (macro-économiques) ; - c'est aussi un système social organisé autour de processus, correspondant à une activité métier (de vente, de production, etc.) »

D'après (Bernoux, 2009, p.132), il s'agit d'un « construit par des acteurs intégrant des contraintes comme éléments des stratégies. »Tandis que Crozier et Friedberg (1977) « ont été parmi les premiers à proposer aussi clairement d'étudier l'organisation comme un phénomène autonome, c'est-à-dire obéissant à ses propres règles de fonctionnement et non déterminé par des contraintes extérieures. » (Bernoux, 2009).

En effet, ils voient le « phénomène organisationnel comme un phénomène totalement autonome et artificiel, dont il faut expliquer l'existence comme celle d'un construit contingent et non pas comme celle d'une conséquence de lois plus générales » (Crozier et Friedberg, 1977). De manière plus classique, une organisation est caractérisée par les traits suivants (Bernoux, 2009, p.135) :

- une division des tâches,
- une distribution des rôles,
- un système d'autorité,
- un système de communications,
- un système de contribution- rétribution.

3.5.1 Pourquoi gérer l'information ?

Traditionnellement, les bibliothèques étaient le lieu dont la mission était de conserver, de répertorier le savoir et d'y donner accès, bien que, d'après Chaumier (2002), la prise de conscience de l'intérêt d'exploiter les ressources des bibliothèques, plutôt que de les cantonner à la conservation ne date que de la fin de la première guerre mondiale. Dans les années 1960, on parle « d'explosion documentaire » car « le document ne comprend plus seulement le livre, mais les articles de périodiques, [lesquels se multiplient], les comptes rendus de congrès, les rapports de recherche. » (Chaumier, 2002, p. 8).

En parallèle de ce phénomène de croissance du volume de documents, le développement rapide de la micro-informatique, qui se concrétise aujourd'hui avec le succès d'Internet, qui

dépend lui-même du développement des infrastructures de réseaux, et qui est corrélé au mouvement de mondialisation des marchés et de financiarisation des économies, aboutit à des stratégies de dématérialisation des documents, certains auteurs préférant le terme de déterritorialisation, à des formes numériques des Systèmes d'Organisation des Connaissances. Par ailleurs, la judiciarisation de la société qui peut se définir comme le processus au cours duquel « un traitement juridique ou judiciaire se substitue à un autre mode de régulation sociale»[2], participe à ce besoin devenu capital de gérer les documents dans les organisations de tous les types (TGE, administrations, états).

Enfin, la configuration générationnelle de notre population active en lien avec la démographie aboutit à de nombreux départs en retraite de séniors, qui, s'ils ne sont pas anticipés par les organisations en termes de gestion des connaissances, produisent des pertes de connaissances qui peuvent s'avérer préjudiciables, pour les entreprises tout particulièrement. En effet, si la transmission du savoir acquis par ces séniors tout au long de leurs carrières n'est pas assurée, c'est tout un ensemble de savoirs, savoirs-faires, de connaissances liées à l'expérience qui se trouvent en dehors de l'entreprise lorsque les acteurs l'ont quittée. Une enquête, en partie qualitative, sur l'emploi des salariés selon l'âge (enquête ESSA, 2001), réalisée par la DARES, révèle que les services de ressources humaines se sentent peu concernés par le vieillissement de leurs effectifs, vieillissement qui concerne plus les grands établissements (plus de 200 salariés). De plus, d'après (Guyot, 2006) :

> *« une partie de l'avantage concurrentiel d'une entreprise est liée aux façons dont elle surveille son marché, et surtout à la façon dont elle gère ses propres informations, celles qui sont liées à son métier. C'est ainsi qu'on parle de management stratégique de l'information et d'information stratégique pour signaler qu'une information acquise au bon moment fera la différence dans son positionnement, ou de capital de connaissances, pour souligner l'importance d'une exploitation des savoir-faire acquis par la pratique. (…) l'information est désormais considérée comme un potentiel possédant une valeur stratégique. » (Guyot, 2006, p.17)*

Ainsi, l'introduction des TIC et la croissance du volume des documents électroniques à gérer par chaque individu à son poste de travail, et en particulier dans le travail tertiaire, conduit à ce que certains auteurs qualifient de « surcharge informationnelle » (Isaac, Campoy, Kalika, 2007). En effet, une étude réalisée par Autissier et Lalhou (1999) démontre que dans

[2]Jean, Jean-Paul, « La judiciarisation des questions de société », *Après-demain*, 398, oct-nov 1997, p.

l'entreprise étudiée, « les managers consacrent un tiers de leur temps en moyenne à ces tâches de manutention de l'information ».

La manutention de l'information est définie par Vacher (1998) comme « l'ensemble des tâches de manipulation, manuelles ou mécaniques, des supports d'information : tri de documents, classement dans des dossiers, indexation de bases de données, mise en forme de comptes-rendus, recherche de papiers, disquettes, adresses électroniques, etc. ». La surcharge évoquée plus haut « est directement liée à la notion de maîtrise du temps et au fait que les TIC contribuent à augmenter le temps de traitement de l'information au détriment des activités liées à l'exercice du métier. » (Isaac, Campoy, Kalika, 2007)

D'après l'étude menée par Isaac, Campoy, et Kalika entre 2001 et 2005, sur les salariés d'une grosse entreprise, il semble que la perception par les opérateurs eux-mêmes de ces tâches de gestion de l'information soit celle d'une tâche envahissante au regard de leurs activités métier. En effet, dans leur enquête « la surcharge cognitive, estimée au travers du temps requis pour classer l'information augmente et passe de 43 % en 2001 à 57% en 2005 ». La perception des salariés enquêtés, concernant le temps estimé à effectuer des tâches de gestion de l'information, confirme l'impact prévisible de l'expansion du volume des documents à traiter. Dans le même temps, les systèmes d'information n'apportent pas de solution simple et efficace pour faciliter cet aspect du travail, et le temps alloué pour ces tâches diminue. D'après (Isaac, Campoy, Kalika, 2007), Meier, en 1963 identifie déjà « la surcharge d'information comme source de stress chez les employés, productrice de dysfonctionnements opérationnels et de pertes d'efficacité ».

3.5.2 Les pratiques informationnelles

Dans le champ des Sciences de l'Information et de la Communication, on trouve une littérature qui s'accroît sur la thématique des pratiques informationnelles. Mais de quoi s'agit-il ? Pouvons-nous tenir pour équivalents cette expression et celle « d'activité de gestion de l'information » ?

Quelle différence entre une pratique, un usage, et une activité en matière d'information ?

La définition des pratiques informationnelles proposée par (Paul, Perriault, 2004), cités par (Fabre, *et al.*, 2011) nous semble pouvoir décrire des pratiques qui concernent tout autant l'information, qu'une activité :

> *« conduites finalisées, individuelles ou collectives, figées ou adaptatives, socialement situées, inscrites dans une temporalité, sous tendues par des représentations, des savoirs, une logique et un raisonnement, marquées par une appréciation de soi et des autres, et révélatrices d'une culture qu'elles enrichissent éventuellement en retour » (Paul, Perriault, 2004)*

D'après (Fabre, *et al.*, 2011) :

> *« Les pratiques informationnelles peuvent se définir comme une manière concrète d'exercer une activité sur l'information (Rey, 1995) visant des résultats concrets sans intention d'expliquer comment le résultat a été atteint. Les pratiques informationnelles sont des procédures, des manières de faire sur et avec l'information (recherche, évaluation, exploitation, traitement, stockage, diffusion) permettant à une activité, par exemple professionnelle, de se dérouler. »*

D'autre part, d'après (Ihadjadene et Chaudiron, 2009) les pratiques informationnelles correspondent à :

> *« la manière dont un ensemble de dispositifs, de sources formelles ou non, d'outils, de compétences cognitives sont effectivement mobilisés, par un individu ou un groupe d'individus, dans les différentes situations de production, de recherche, d'organisation, de traitement, d'usage, de partage et de communication de l'information. Nous englobons dans ce terme de « pratique » aussi bien les comportements, les représentations que les attitudes informationnelles de l'humain (individuel ou collectif) associés à ces situations » (Ihadjadene et Chaudiron, 2009).*

Cette dernière définition nous semble correspondre à ce que nous souhaitons étudier ici : les activités de gestion de l'information en tant que processus individuel et collectif, qui relève de la communication, qui met en jeu des représentations, des compétences, des comportements, des attitudes sociales, des compétences techniques, des opérations cognitives, et qui implique et transforme des états psychologiques, ainsi que l'état des connaissances des individus ou des collectifs.

Elles concernent, selon nous, dans le champ des activités professionnelles, les opérations de manipulation sur les documents (création, nommage, modification, enregistrement, classement, recherche), mais aussi les opérations d'organisation des documents, (gestion du système de classement, création, nommage et modification des dossiers contenant les documents), enfin, les opérations de recherche d'information (qui incluent la formulation d'un besoin d'information, le choix d'une stratégie de recherche, le choix des outils et des sources,

la mise en œuvre effective de la recherche par requête ou browsing, la lecture, le tri, la sélection des résultats, l'organisation, l'enregistrement et la gestion des informations jugées utiles, voire réutilisables, et ce de manière itérative).

3.5.3 Distinction entre les usages et les pratiques

Concernant cette distinction, nous adoptons le point de vue proposé par (Chaudiron et Ihadjadene, 2010, p.4), qui est résumé par ce qui suit :

> *(…) les usages sont envisagés comme l'expression d'un processus constitué d'interactions complexes mettant en relation un individu et un dispositif qui peut être, ou non, un artefact technique. Une caractéristique de la dimension d'usage est donc sa (relative) stabilité, à la fois diachronique et synchronique. (…)*
>
> *Il s'agit de deux notions qui permettent d'appréhender la réalité selon deux angles différents et complémentaires. La distinction que nous proposons suggère de réserver le terme d'usage pour désigner les travaux portant sur les dispositifs, techniques ou non, et leurs interactions avec les usagers ; le terme de pratique sera réservé pour caractériser les approches centrées sur le « comportement composite » à l'œuvre dans les différentes sphères, informationnelles, culturelles, journalistiques, etc. L'observation des pratiques, qu'elles soient individuelles ou collectives, nécessite alors d'adopter une approche de l'action envisagée comme un processus en tension entre les savoirs mobilisables, les compétences immédiates, les habitus, les arts de faire, les désirs d'agir, etc.*
>
> *On parlera donc des usages des moteurs de recherche ou des téléphones portables, des usages de l'internet ou des usages des bibliothèques numériques pour désigner la façon dont on utilise le dispositif en tenant compte du cadre socio-professionnel, culturel, politique, économique... Les usages des dispositifs techniques s'inscrivent en ce sens dans différents contextes, professionnel, privé, familial... À l'inverse, nous parlerons de pratiques pour désigner les études qui sont centrées sur l'humain (individuel ou collectif) et qui analyse ses modalités d'action (ou d'inaction), ses représentations, ses attitudes. (…)*
>
> *On parlera alors de pratiques informationnelles pour désigner la manière dont l'ensemble de dispositifs, des sources, des outils, des compétences cognitives sont effectivement mobilisés dans les différentes situations de production, de recherche, traitement de l'information.*

D’après cette citation, on constate que le contexte, notion qui sera plus précisément définie dans la suite de cette partie, est plus large dans l’approche qui consiste à étudier les pratiques plutôt que les usages, étant donné qu’elle inclut les aspects humains, sociaux, psychologiques, cognitifs et relationnels.

Dans le cadre de ce travail, nous utiliserons donc indifféremment les termes de « pratiques informationnelles », et « d’activités de gestion de l’information ».

3.5.4 Que sont les activités de gestion de l’information ?

Les activités de gestion de l’information sont générées par l’exécution d’une tâche, activité ou mission. Ainsi on suppose à la manière de (Guyot, 2006) que toute activité génère et nécessite de l’information. Ceci nous conduit à proposer, qu’en observant l’organisation personnelle et collective des informations et documents, nous pouvons obtenir des traces de ces activités qui nous renseignent sur la façon dont les acteurs réalisent leurs activités.

Le plus souvent, dans une organisation, on observe une coexistence de dispositifs informationnels, plus ou moins formels. Ainsi, si l’on considère que gérer l’information ne constitue pas l’activité principale des acteurs dont on parle, cette activité de gestion de l’information s’exerce dans une discontinuité et dans des formes multiples. D’après (Guyot, 2002), il peut s’agir de :

- chercher de l’information,
- fournir ou demander une information à un collègue,
- se procurer un document,
- produire un rapport, une note,
- consulter un dossier,
- classer des documents de travail

Mais il peut également s’agir de retrouver un document précédemment créé, enregistrer un document, nommer ou renommer un document, lui ajouter des métadonnées, des annotations.

Toutes ces actions peuvent être réalisées avec divers outils. Ainsi, on peut rechercher de l’information dans un intranet, ou sur Internet, ou encore dans une base de données interne.

On peut fournir une information à un collègue par courrier électronique, se procurer un document dans un dossier partagé, classer des dossiers de travail dans son poste personnel.

En plus des problématiques spécifiques de forme de l'information (données, statistiques, documents textuels), (Guyot, 2002), on se trouvera fréquemment confronté à des problématiques de format de fichier et de choix du bon outil de lecture d'un document, ainsi qu'à des problématiques d'usages des outils (moteurs de recherche, accès à des bases de données, VPN) et à des questions liées au lieu de classement des documents, et à leur nommage ou à leur indexation.

Dans le cas de la recherche d'information, spécifiquement, certaines tâches seront nécessaires, par exemple, dans le but de rassembler de l'information pour la rédaction d'une synthèse.

Il faudra donc procéder par étapes :

- faire l'inventaire des documents déjà en notre possession
- lire ou relire ces documents déjà disponibles et retrouvés
- il faudra ensuite formuler un besoin informationnel
- élaborer une stratégie de recherche
- inventorier les sources possibles et disponibles
- choisir un outil de recherche
- mettre en œuvre la recherche effective par requête ou browsing, ou les deux
- trier, sélectionner, lire des résultats, (téléchargement de documents, enregistrement, consultation de pages, insertion dans un système de bookmarks de pages sélectionnées, impression)
- raisonner: inventorier les sources d'information, les informations, et identifier de nouveaux besoins d'information
- redéfinir une stratégie de recherche
- inventorier les sources possibles et disponibles
- choisir un outil de recherche
- mettre en œuvre la recherche effective, et ce, de manière itérative…
- exploiter, insérer les informations dans la tâche qui a déclenché la recherche d'information

D'après la Théorie de l'Activité présentée précédemment, dans le cadre de l'exemple cité ci-dessus, l'activité correspond aux buts que l'on se donne, ici, la rédaction d'une synthèse, les actions qui sont réalisées pour remplir le but, et qui sont, d'après cette théorie, des processus conscients correspondent à la plupart des étapes mentionnées pour la recherche d'information : formulation d'un besoin d'information, choix des sources, des outils, élaboration de la stratégie, etc.

Les opérations qui sont des processus non conscients, dans cet exemple, correspondent aux activités de lecture, de saisie des mots sur un clavier pour une requête dans un moteur, de scan des résultats, les différents clics requis pour la navigation. Pendant ces processus, l'acteur ne réalise pas de microdécisions ou de choix, sinon concernant le contenu, (ce qui ne correspond pas à ce que nous évoquons précisément ici), et uniquement si la situation le nécessite (par exemple en cas de besoin d'une requête avancée). Les choix concernant le contenu des pages

visitées, par exemple, relèvent d'opérations d'analyse du contenu, donc d'actions dans le langage de la Théorie de l'Activité, mais ce n'est pas le cas du clic sur un lien en lui-même.

On constate bien qu'il s'agit, en matière d'activité de gestion d'information concernant les actions, d'une « série de microdécisions qui s'enchainent avec des actions dont certaines sont longues et coûteuses. » (Guyot, 2002). Cette mise en œuvre des activités de gestion de l'information suppose de la part de l'acteur d'être en mesure de juger de la valeur d'une information, pour la tâche qui la génère, mais éventuellement aussi dans une perspective à plus long terme de réutilisation, notamment dans les démarches qui incluent des objectifs de gestion des connaissances. Il peut donc s'agir de séries de micro-jugements, orientés différemment selon les tâches présentes dans l'esprit de l'acteur et les objectifs qu'il prend en compte dans son activité de gestion de l'information.

L'acteur pourra respecter des objectifs directs, comme la réalisation d'une note ou d'un rapport, correspondant au travail prescrit, la série d'activités de recherche d'information que ce travail nécessite étant comprise implicitement comme nécessaire. Mais bien souvent, elle ne fait l'objet d'aucune formalisation, ni prescription. Il s'agit donc d'une tâche redéfinie. Elle appartient au travail réel.

L'acteur peut aussi, au cours de ce travail, souhaiter capitaliser pour lui-même des informations qu'il juge susceptibles d'être réutilisables pour une autre mission faisant partie de son activité au sens large (son activité métier). Il exerce alors son pouvoir d'autonomie, ou ses marges de manœuvres qu'il négocie lui-même, et éventuellement avec lui-même, afin de s'autoriser à conserver du temps pour capitaliser des documents ou des informations.

Dans ces exemples, l'activité d'information est une activité vue comme un moyen. Elle entre dans une catégorie dite « secondaire », voire « banale » ou « souterraine » (Guyot, 2002). On voit l'opacité de cette relation peu contractualisée, peu reconnue, peu explicite, entre l'activité principale et les activités de gestion de l'information.

En effet, si en matière d'archivage et de gestion des documents à valeur probante, « engageante », (Chabin, 2009) ou patrimoniale, bien souvent les organisations élaborent des prescriptions pour la conservation et l'indexation de ces documents, le champ des documents intermédiaires (mails, rapports en cours de rédaction, document ressources), des proto-documents, et des DopA (Documents pour l'Action) (Zacklad, 2006) n'est pas traité par ces prescriptions. Leur gestion n'est pas prescrite, et peu outillée, laissant les utilisateurs livrés à eux-mêmes quant à l'organisation, au classement et à l'indexation de leurs documents de travail.

Cette gestion des documents de travail n'apparait donc pas nécessairement dans les prescriptions, tout en restant une tâche incontournable pour la réalisation des missions principales des agents. Cette tâche de gestion de l'information, comme nous l'avons vu occupe de plus en plus les agents. Mais n'est pas formulée dans les prescriptions du travail, elle n'est donc pas non plus reconnue dans l'activité qui détermine les rétributions. Elle souffre par conséquent d'un manque de visibilité et le travail réel effectué dans ce sens souffre d'un manque de reconnaissance par les prescripteurs. Ainsi, la gestion des informations, considérée comme une tâche secondaire, ni prescrite, ni formulée, en regard de l'activité principale (ou activité métier) des agents, contribue à leur surcharge cognitive et engendre une mauvaise considération de cette tâche par ceux-ci. De plus, la non-reconnaissance de l'empiètement de ces activités de gestion de l'information sur les activités principales des agents par les prescripteurs risque bien souvent d'engendrer un équipement mal adapté à cet aspect de leur travail. (Desfriches Doria et Zacklad, 2010).

Si l'on tente de synthétiser les caractéristiques des activités de gestion de l'information, on constate qu'elles touchent à plusieurs aspects du travail et qu'elles revêtent les aspects suivants, d'après (Guyot, 2002) :

- c'est une activité de soutien ou vue comme secondaire,
- elle est discontinue,
- elle transversale,
- elle est invisible, implicite,
- elle nécessite un apprentissage des outils,
- elle se situe entre prescription et autonomie,
- elle implique une (ou plusieurs) interprétations,
- elle fait l'objet d'une tension entre individuel et collectif,
- elle est difficile à mesurer, à observer, à évaluer,
- elle nécessite des indicateurs qualitatifs

Cette activité d'information a aussi des liens avec :

- la production de règles: procédures de travail, vocabulaires, pratiques individuelles, développement de méthodes, d'indicateurs d'usage et du volume de l'activité d'information,
- les activités de communication,
- sur le plan cognitif : elle implique des savoir-faire, connaissances, représentations collectives pour le partage,
- elle touche à des questions d'identités, de motivation, d'univers psychologique.

Enfin, du point de vue de l'organisation :

- Elle concourt à la coordination : c'est un repère par rapport à l'activité des autres acteurs
- Elle participe à une vision synthétique de l'activité
- Sa traçabilité contribue à la pérennité de l'entreprise

3.5.5 Systèmes d'Information Personnels et *Personal Information Management*

Du point de vue des acteurs, une difficulté supplémentaire est qu'ils endossent plusieurs rôles dans cette activité de gestion de l'information, d'après (Guyot, 2002) :

- décideur : choix des sources les plus adaptées
- assembleur : mobilise des moyens, outils, personnes
- gestionnaire : met en œuvre une démarche selon la situation et la finalité, évalue la performance du dispositif ou qualité de l'information et sa pertinence

Du point de vue des usages individuels, les acteurs fonctionnent selon une logique d'économie de moyens, ils ont tendance à reproduire des opérations déjà expérimentées, les outils d'information nécessitant un investissement cognitif et un apprentissage, le développement de compétences documentaires que l'acteur ne peut pas réaliser, par manque de temps, et de disponibilité intellectuelle ou cognitive. Pour résoudre ses problèmes d'information, l'acteur met donc en place des routines, mobilise des savoir-faire immédiatement disponibles (acquis et n'a pas le temps d'en acquérir d'autres), met au point

des stratégies très personnelles, qui restent largement implicites, de gestion de ses dossiers numériques. (Guyot, 2002). Ceci produit des Systèmes d'Information Personnels (SIP) (Guyot, 2002) qui sont au service des besoins de leur concepteur-gestionnaire-utilisateur, et qui en rendent donc l'accès ou la compréhension pour un autre acteur très difficile.

Dans la littérature anglophone, ce thème est abordé par les auteurs qui travaillent sur le Personal Information Management (PIM). Le PIM est le champ qui s'intéresse au système d'information développé ou créé par les personnes, individuellement, pour un usage personnel dans un environnement de travail (Barreau, 1995). Il s'agit pour nous d'une sorte spécifique d'activité de gestion de l'information.

Dans les années 80, (Malone 1983) s'est intéressé à l'organisation physique des dossiers dans les bureaux. Il ressort de son étude que les dossiers sont organisés soit en piles, soit en dossiers thématiques. Les dossiers thématiques proviennent de la catégorisation de l'information, et les piles sont des assemblages de documents hétérogènes, qui n'ont en commun que la caractéristique d'être difficilement classables. De plus, (Malone, 1983) met en évidence que l'organisation des dossiers et document a aussi une fonction de mémorisation et organisationnelle. Il note:

> *« The mechanical difficulty of finding and labeling anew file folder is often non trivial (…) When a filing system has several hierarchical levels (…), the mechanical difficulty of creating new categories can be a major barrier to keeping the file system current. » (Malone, 1983).*

Il ajoute que dans les classifications multidimensionnelles, la possibilité de pouvoir simplement assigner un document à plusieurs catégories pourrait être un apport de l'informatique aux problèmes de classification, ainsi que le fait de différer le classement, en permettant aux utilisateurs de laisser des documents non-classés, en attendant de leur trouver un nom et une place appropriée. Il faut noter qu'à l'époque, les possibilités offertes par l'informatique étaient loin d'être aussi largement développées qu'aujourd'hui. Cette dernière possibilité, correspondrait à la classe « Divers » trouvée dans de nombreuses classifications et dans de nombreuses études sur les systèmes d'information personnels (Khoo *et al.*, 2006).

3.5.5.1 Comment l'information est-elle personnelle ?

William Jones (2010) aborde cette question de manière détaillée. Ainsi, il distingue plusieurs dimensions complémentaires à prendre en compte dans la description de l'information vue comme personnelle dans le champ du PIM.

L'information personnelle peut correspondre à :

- de l'information contrôlée par moi : celle que je garde, comme des emails, des documents de travail, etc. ;
- de l'information à propos de moi, mais sous le contrôle d'une autre entité : cela concerne les dossiers médicaux, les profils en ligne et les réseaux sociaux par exemple ;
- de l'information dirigée vers moi : qui peut correspondre aux emails, aux publicités, et autres informations de recommandations basée sur une récolte de données ;
- de l'information envoyée par moi : celle que j'envoie par mail, ou que je publie sur un blog par exemple ;
- de l'information « expérimentée » par moi : comme des pages Web que je visite, ou mon historique de navigation ;
- enfin la dernière dimension évoquée est celle de la pertinence, une information personnelle dans ce cadre est une information pertinente pour moi.

3.5.5.2 PIM et contexte

Pour donner une définition plus précise de ce qui est appelé le « Personal Information Management » nous nous référons à (Barreau, 1995) :

> *« It includes a person's methods and rules for acquiring the information which becomes part of the system, the mechanisms for organizing and storing the information, the rules and procedures for maintaining the system, the mechanisms for retrieval, and the procedures for producing the various outputs required. » (Barreau, 1995).*

Il s'agit donc de la gestion de l'information quotidiennement générée et reçue par les utilisateurs, liée au fil des activités poursuivies. L'organisation de ces documents est

maintenue par l'utilisateur lui-même avec des outils, souvent rudimentaires, mais très largement répandus, comme les systèmes de gestion de fichiers arborescents.

Un système de gestion de fichier arborescent propose la même structure des informations qu'une classification classique, c'est-à-dire, un arbre de dossiers contenant des dossiers, et des fichiers, organisés hiérarchiquement. Dans le cadre d'une classification classique, les classes sont organisées du plus générique au plus spécifique. Dans le cadre du PIM, cela est rarement le cas. En effet, dans le PIM, les arborescences de dossiers et de fichiers, la plupart du temps, ne sont pas construites du plus générique au spécifique, car celles-ci sont mises à jour beaucoup plus souvent que les classifications bibliographiques. Elles sont construites en suivant le fil de l'activité métier, qui est constamment en mouvement : les priorités sont réorganisées, les missions changent ou évoluent.

L'acteur découvre en la créant, au fur et à mesure, son arborescence de fichier, au gré des évolutions de son ou de ses activités métier. Il ne peut pas la créer à priori. Il dispose de plusieurs solutions théoriquement pour la maintenance de l'organisation de son système de fichiers, il peut soit la réorganiser à chaque ajout, soit périodiquement, ce qui paraît peu réaliste puisque cette activité est réalisée dans les interstices et considérée comme secondaire, soit à l'occasion d'un changement de poste, ce qui paraît plus pertinent, soit encore raisonner dans une logique d'accumulation. Dans ce dernier cas, qui nous semble être le plus répandu, le système de fichier croît avec le temps passé dans l'organisation, sans que les acteurs n'aient l'occasion d'en faire l'objet d'une rationalisation de l'organisation ou d'une maintenance en raison du coût cognitif de cette opération et du peu de temps alloué à ces activités de gestion de l'information.

De nombreuses études ont été menées afin d'observer la structure des arborescences de fichiers d'utilisateurs, et de comprendre les usages et pratiques de ces systèmes de gestion de fichiers, ainsi que les logiques à l'œuvre dans ces activités de catégorisation (Lee, 2011 ; Barreau, 1995, Bergam *et al.*, 2003 ; Kwasnik, 1989 ; Khoo *et al.*, 2006). Ainsi, les problématiques liées aux classifications classiques et à la catégorisation, bien étudiées dans la littérature, peuvent être rapprochées de celles à l'œuvre dans le champ du PIM, mais avec des aspects plus cognitifs et psychologiques, puisqu'il s'agit de classifications personnelles.

Nous y revenons en partie 3.5.5.3.

Une observation récurrente dans le champ du PIM, est que ces systèmes de classement personnels sont fortement corrélés aux activités (au sens métier du terme) (Guyot, 2002, 2006). Ils sont donc fortement personnalisés et contextuels, puisqu'ils s'adaptent à la nature

changeante et constamment renégociée des tâches à effectuer dans un métier. Ainsi, « PIM systems are more intimate and more in tune with the way people think and work than information systems designed for general use. » (Barreau, 1995). Dans le même sens, Pour Kwasnik (1989), les classifications sont plutôt centrées sur la personne et la situation que sur les objets.

La plupart de l'information acquise dans un Système d'Information Personnel (SIP) est générée par l'utilisateur et a de la valeur pour celui-ci. A l'intérieur d'un SIP, « each person's vocabulary is itself "controlled" since it is unique for that person » (Barreau, 1995). Dans son étude sur l'influence du contexte sur les processus employés par les individus pour classer et organiser leurs propres documents dans leur propre espace, (Barreau, 1995) démontre que les conditions dans lesquelles les documents sont créés et utilisés changent, et que ces conditions impactent ces pratiques informationnelles. Ces conditions comprennent pour (Barreau, 1995) : le contexte de travail, qui inclut la nature de la tâche, le sujet traité, les outils disponibles, le public concerné par le travail, et l'usage attendu de l'information.
Certains des problèmes liés à ces SIP sont caractérisés comme suit : ils sont difficiles d'accès pour une autre personne que leur concepteur, il est difficile d'en avoir une vue d'ensemble, et il est difficile d'en partager des éléments.

3.5.5.3 Catégorisation et PIM

Dans un article sur la psychologie du PIM, (Lansdale, 1988) s'intéresse au processus de catégorisation et au fonctionnement de la mémoire.
Concernant la catégorisation, il affirme que les informations n'entrent pas facilement dans des structures de catégorisation bien ordonnées et il avance deux raisons pour ce fait. D'une part, l'ambiguïté inhérente au langage qui implique que l'utilisateur, même s'il est le créateur du nom d'une catégorie, pourra lors de la recherche ne pas mettre le même sens sur le nom donné à la catégorie. D'autre part, l'information dans le monde réel peut entrer dans des catégories floues ou qui se chevauchent, ce qui implique que toute catégorisation d'un item est pertinente seulement pour certains aspects de celui-ci. (Lansdale, 1988).
Concernant la mémoire, il affirme qu'il est difficile de se souvenir du nom d'un document ou d'une catégorie, mais que notre mémoire retient beaucoup plus d'éléments à propos des documents, relativement aux critères de recherche généralement proposés.

Nous pouvons ainsi intégrer à notre mémoire d'autres éléments comme le moment où nous avons reçu le document, son visuel, le sens de son contenu, ou ce que nous étions en train de faire à ce moment donné. Ces éléments sont des informations contextuelles. Lansdale (1988) nous semble énoncer une hypothèse trop peu mentionnée dans la littérature, en proposant de s'intéresser à ce dont se souviennent les personnes, afin de proposer des systèmes qui exploitent ces caractéristiques de la mémoire, caractéristiques qui peuvent être proprement personnelles ou partagées par un groupe de personnes.

> *« It seems reasonable to conclude that if we could determine what people could remember and what they could not, then we are in a position to ask (a) how existing methods exploit or fail to exploit memory; and (b) what we can do within the constraints of new technology which would provide the most usable system. » (Lansdale, 1988).*

Il ajoute que les fonctionnements de la mémoire sont tels que nous retenons plus difficilement les détails que le sens général d'un événement. Il affirme qu'il y a une relation forte entre la façon dont une information est gardée en mémoire et ce à quoi nous pensons quand nous essayons de la retrouver. Le processus de mémorisation semble être lui aussi un phénomène contextuel. Ainsi pour Lansdale, l'informatique et les critères de recherche proposés dans les Systèmes de Recherche d'Information (SRI) ne sont pas adaptés de manière optimale pour exploiter les caractéristiques de la mémoire humaine.

> *« As we shall see, computer systems do not use this information but rely upon the user remembering filenames and/or the categorization of required information. What we are good at is being ignored, and what we are required to do is a difficult and flawed psychological process. » (Lansdale, 1988).*

Dans le même sens, une étude sur les pratiques personnelles de gestion de l'information d'ingénieurs menée par (Pikas, 2007), démontre que ces derniers n'utilisent pas tous les mêmes stratégies pour retrouver leurs documents, et notamment en termes de type d'information. Cette étude affirme que l'élément le plus important lors de la recherche d'un document est la dimension temporelle, qui peut elle-même se traduire par des occurrences variables (pour certains cas il peut s'agir de la saison, ou de la date précise, ou d'une période, ou encore d'une phase d'un projet) (Rapport de Recherche Décembre, 2011).

3.5.5.4 Le problème du « basic level » : une question de catégorisation

Du point de vue de la théorie de la catégorisation, les problématiques évoquées précédemment reviennent à tenter de définir d'après quels critères pertinents faut-il construire une catégorisation pour rendre la recherche plus efficace et à comprendre comment ce processus se construit.

Comme l'explique Elin K. Jacob (2004), dans la théorie classique de la catégorisation, la formation des concepts n'est pas vue comme un processus de création, mais comme un processus de reconnaissance. Ainsi, dans la philosophie empiriste anglaise, les personnes distinguent les objets en fonction de leurs attributs, propriétés, ou qualités (couleur, forme, fonctions ou parties). Il y a eu beaucoup d'hypothèses sur la nature des objets, leurs qualités, sur leurs relations, et sur la relation avec l'idée que les gens s'en font. Pour Locke, (1632 - 1704), un objet est constitué d'une intension (signification) et d'une extension (l'objet dans la classe). L'intention est une spécification des qualités que doit posséder un objet pour appartenir à une classe. L'extension consiste en objets qui possèdent ces qualités (Mervis et Rosch, 1981).

Dans cette théorie classique, les catégories sont déterminées par des jeux de critères définissant l'appartenance à une catégorie. Les attributs (qualités) connectent les objets au monde. Cette théorie repose sur trois propositions basiques :

1. L'intension d'une catégorie est une représentation résumée d'une catégorie entière d'entités :

   Cette proposition postule que la définition d'une catégorie est l'union des caractéristiques essentielles qui permettent l'identification de l'appartenance à cette catégorie (les extensions). Si tous les membres d'une catégorie partagent ce jeu de caractéristiques, alors chaque membre de la catégorie est également représentatif de la catégorie vue comme un tout. Pour cette raison, la structure interne d'une catégorie est dite sans classement, car aucun de ses membres ne peut être plus typique qu'un autre, ou plus représentatif de la catégorie qu'un autre membre.

2. Les caractéristiques essentielles qui composent l'intension d'une catégorie sont individuellement nécessaires et conjointement suffisantes pour déterminer l'appartenance à une catégorie :

   Cette proposition signifie que comme chaque membre de la catégorie doit posséder toutes les caractéristiques essentielles qui composent l'intension de la catégorie, la possession de ces caractéristiques est suffisante pour déterminer l'appartenance à cette catégorie. Il existe donc une relation binaire entre une entité et une catégorie, telle

qu'une entité est ou n'est pas un membre de la catégorie. Pour cette raison, les limites des catégories sont vues comme fixées et rigides.

3. Si une catégorie A est emboîtée dans une catégorie super-ordonnée B, les caractéristiques qui définissent B sont contenues dans le jeu de caractéristiques qui définit A.

   Cette proposition définit le principe d'héritage des propriétés qui existe entre les catégories dans une structure hiérarchique. Ainsi, chaque membre d'une catégorie qui est un sous-ensemble d'une catégorie super-ordonnée doit posséder non seulement les caractéristiques qui déterminent l'appartenance au sous-ensemble mais aussi le jeu de caractéristiques essentielles qui détermine l'appartenance à toute catégorie super-ordonnée, dans laquelle le sous-ensemble est emboîté.

Les critiques de la théorie classique de la catégorisation ont opposé à ces conceptions les difficultés des individus à identifier les caractéristiques des entités. Ceci ébranle l'affirmation selon laquelle le jeu des caractéristiques essentielles déterminant une catégorie est un absolu et pose aussi la question de la disponibilité de ces caractéristiques essentielles pour les membres d'une même communauté linguistique. Les auteurs de ces critiques, qui ont émergé dans les années 70, dont Mervis et Rosch (1981) tentent de démontrer et de bâtir une théorie de la classification prototypique dont une des affirmations est que certains membres d'une catégorie seraient plus représentatifs de la catégorie que d'autres membres de cette même catégorie.

Ainsi, des expérimentations menées démontrent qu'un moineau est plus représentatif de la classe « oiseau » qu'une autruche ou un pingouin. Selon ces mêmes auteurs, une catégorie qui correspond au « basic level » est celle qui est la plus efficace cognitivement. Comme elle correspond au niveau le plus basique de catégorisation, la valeur informationnelle des groupes d'attributs est maximisée. C'est le niveau auquel les catégories maximisent les similarités intra-catégorielles, relativement aux similarités inter-catégorielles. Il se définit aussi comme le niveau le plus général pour lequel les personnes utilisent les mêmes actions-moteur pour interagir avec les membres de la catégorie. Les membres de cette catégorie de niveau basique ont des formes générales similaires. Une image mentale peut refléter cette catégorie en entier, par opposition à la conception basée sur la possession de certaines caractéristiques essentielles. Enfin, les objets sont reconnus comme appartenant à cette catégorie de niveau basique, plus rapidement, que les membres des catégories des autres niveaux. Le niveau basique est celui pour lequel les adultes et les enfants peuvent spontanément nommer des objets appartenant à cette catégorie. Les principes qui sous tendent la détermination du niveau

basique sont supposés universels, mais des facteurs culturels et le niveau d'expertise individuels peuvent jouer un rôle dans cette tâche.

Ainsi on peut constater que même si l'on admet l'existence d'un niveau basique (basic level) dans le processus de catégorisation, certains facteurs contextuels reviennent dans le jeu des critères qui le déterminent.

3.5.6 Le contexte : une notion complexe

Ainsi, le contexte, dans l'analyse de l'activité du point de vue de la Théorie de l'Activité, est admis de manière théorique comme jouant un rôle dimensionnant dans la détermination de l'action, (Darses, Hoc, Chauvin, 2004). Le contexte y est vu comme généré par la mise en œuvre de l'activité. Les personnes génèrent le contexte de manière délibéré et consciente, en partie à travers l'objet (le but de l'activité). Il est à la fois intérieur (objectifs de l'activité) et extérieur aux personnes, impliquant les artefacts, les autres personnes, et des cadres spécifiques.
Dans la sémiotique des transactions communicationnelles, les paramètres de la situation transactionnelle (le contexte) sont les suivants :

- un projet commun
- des relations sociales entre le bénéficiaire et le réalisateur
- un cadre spatio-temporel et des conditions contingentes
- le choix d'un médium
- des équipements techno-informationnels (les artefacts)
- les terrains représentationnels commun et individuel
- les compétences des réalisateurs et des bénéficiaires

Nous pensons, comme Dourish (2004), que le contexte est une notion très importante pour comprendre, contextualiser, désambigüiser, des formes d'activités et d'informations, tandis que sa nature est d'être continuellement négociée et redéfinie. Pour Wild *et al.* (2009), les

documents sont embarqués dans un contexte et contiennent des représentations des aspects tangibles et intangibles du monde.

Dans une approche sociologique de cette notion, pour Dourish (2004), le monde tel que nous le percevons est essentiellement un consensus quant à son interprétation. Les catégories abstraites sont des éléments qui s'imposent au monde par le biais de nos interactions avec celui-ci et avec chacun, plutôt que des éléments qui existent à l'intérieur du monde. Il voit donc la notion de contexte comme un problème d'interaction. Ainsi, pour lui, la contextualité est une propriété relationnelle qui se tient entre les objets et les activités, et elle est définie dynamiquement. Pour Dourish (2004), le contexte est particulier à chaque occasion d'activité et d'action. C'est une propriété occasionnelle, pertinente en fonction de paramètres particuliers, de cas particuliers d'actions, et de parties particulières de cette action. Enfin, le contexte résulte de l'activité. Il est activement produit, maintenu, et ordonné dans le cours de l'activité. Cette conception se rapproche de celle de la Théorie de l'Activité concernant le contexte.

C'est une caractéristique émergente de l'interaction déterminée dans le moment et dans l'action (en la réalisant). Cela mène (Dourish, 2004) à affirmer que le contenu et le contexte ne devraient pas être séparés. Le contexte ne peut être une description extérieure stable de caractéristiques dans lesquelles l'activité survient. A l'inverse, le contexte découle de et est soutenu par l'activité elle-même. Ainsi, la distinction effectuée entre contexte et contenu dans la méthodologie GenISIS peut-elle être vue comme artificielle. Le contexte y est envisagé sous l'angle du contexte dans lequel une ressource informationnelle est créée, incluant les activités durant laquelle la ressource est créée, ou les motivations pour sa création. Mais dans cette méthodologie, le contexte et le contenu sont considérés comme indépendants.

(Lee, 2011), propose une vision plus structurée de la notion de contexte au sens large. Ainsi, il distingue trois modalités complémentaires selon lesquelles le contexte peut être compris, ou encore pris en compte, tandis que « A great deal of human communication takes place at the intersection between the three types of context. » (Lee, 2011). Il définit le contexte en référence à une Entité Cible (Target Entity), comme étant relationnel de manière inhérente. Il affirme que c'est toujours « un contexte de, à propos de, entourant quelque chose » qu'il décrit comme l'Entité Cible. Ainsi, dans sa formulation la plus générique, en relation à une Entité Cible, le contexte serait « tout le reste » ou tout (états, objets, faits, relations) ce qui dans un univers n'est pas l'Entité Cible. Il écrit d'ailleurs :

> *« In short, context is a set of things, factors or attributes that are related to a TE [Target Entity] in important ways (e.g. operationally, semantically, conceptually, pragmatically) but are not so closely related to the TE that they are considered to be exclusively part of the TE itself. » (Lee, 2011)*

Les trois types de contexte qu'il définit sont les suivants :

1. le contexte linguistique ou sémantique: le jeu d'expressions symboliques ou de représentations, entourant une Entité Cible (Target Entity), qui aide quelqu'un à exprimer, construire du sens, traduire ou à agir sur ou avec l'Entité Cible.

   Ce premier sens du contexte traite la place de l'Entité Cible dans un discours plus large ou un système d'information. Cela peut inclure, par exemple, le discours dans lequel une affirmation est imbriquée, les autres documents classés dans la même catégorie, ou encore la théorie formelle dans laquelle un concept doit être compris.

2. les caractéristiques objectives ou socialement construites et les conditions de la situation dans laquelle apparaît ou se produit l'Entité Cible.

   Ce deuxième sens du contexte traite le jeu de facteurs objectifs ou inter-subjectivement reconnus qui entourent une Entité Cible.

3. les aspects de l'état mental ou physique, la tendance, l'intention, l'identité ou les expériences récentes d'un acteur qui pèsent sur la manière dont il interprète, comprend, agit à l'intérieur de la situation ou sur ce qu'il en retient.

   Ce troisième sens traite le statut subjectif d'un agent particulier.

D'après cette conception du contexte, les SI structurés agissent à l'intersection de ces trois sortes de contextes. La plupart des conventions pour l'activité qui construisent ou maintiennent des systèmes relatifs à la connaissance sont des tentatives d'englober l'expression d'aspects du monde (le contexte défini selon la deuxième proposition comme caractéristiques objectives ou socialement construites et comme les conditions de la situation), avec des informations suffisantes et appropriées, (le contexte linguistique et sémantique)afin qu'elles puissent être interprétées par des lecteurs ou récepteurs futurs (contexte subjectif d'un agent particulier), dans des conditions relativement prédictibles. (Lee, 2011).

Dans le champ de l'ergonomie, d'après (Leplat, 2008), le contexte correspond à « ce qui joue un rôle dans l'activité sans avoir été explicitement prévu à cet effet ». La distinction entre travail prescrit et travail réel (voir partie 3.4.1), suggère que le contexte est constitué par les

conditions prises en compte dans l'activité qui ne figurent pas explicitement dans la tâche prescrite.

Le contexte peut être défini comme interne ou externe. Le contexte externe correspond aux conditions externes de l'activité comme les traits physiques (lumière), les objets matériels, les marqueurs temporels, le milieu technique ou organisationnel, les traits des interlocuteurs, tandis que le contexte interne comprend les représentations par le sujet du contexte externe (l'univers mental dans lequel se développe l'activité, les événements mémorisés, les connaissances, l'expérience, la valeur subjective de la tâche) (Leplat, 2008). Ainsi, le contexte ne peut pas être défini indépendamment de l'activité, mais il se développe parallèlement à l'activité. Il varie au cours de l'action et a besoin d'être redéfini selon le moment considéré.

Nous avons tenté de définir le contexte en référence à la Théorie de l'Activité pour le concept d'activité et également dans le cadre des activités de gestion de l'information, vues comme secondaires, comme un moyen d'exercer une activité principale. Puis nous nous sommes intéressés au contexte dans le cadre des approches du PIM, la gestion personnelle de l'information étant fortement contextuelle.

3.5.6.1 Contexte et catégorisation

Le contexte, dans le cadre du PIM, joue un rôle lié à la mémorisation des informations liées à l'usage des documents, comprenant l'activité principale, les règles de l'organisation (les prescriptions en matière de gestion personnelle de l'information, si elles existent),des aspects organisationnels personnels liés à l'activité, des aspects psychologiques, cognitifs, temporels, et liés aux artefacts, vus comme le système d'information, les outils disponibles au sein de celui-ci, ainsi que leurs affordances.

Tandis que le contexte est considéré comme un élément dynamique qui influe sur la gestion personnelle de l'information, « Disposer du contexte d'un document facilite le travail d'appropriation de celui-ci ». (Kolmayer et Peyrelong, 1999). Il peut s'agir, pour partie, du contexte d'élaboration du document, et pour partie de celui que crée l'utilisateur.

Pour (Lee, 2011), il est important de distinguer la notion large de contexte, constituée selon lui par les interactions et les relations entre une Entité Cible et son environnement, du jeu plus spécifique d'informations contextuelles qui est reflété dans un système d'information. Aucun objet numérique ne peut porter avec lui tous les éléments de son contexte. Ainsi, « *No set of*

information objects, no matter how detailed, can fully capture and reflect everything about a given situation. » (Lee, 2011).

Il y a donc des choix à effectuer sur la pertinence des informations contextuelles qu'il est nécessaire de conserver, pour faciliter la recherche d'information et le travail d'appropriation ou de réappropriation d'un document. Cette dernière problématique peut être rapprochée du problème du « basic level » (Mervis et Rosch, 1981). Puisqu'un système d'information ne peut pas garder toutes les informations contextuelles, lesquelles faut-il choisir ? Comment les individus réalisent-ils ces choix ? Y a-t-il des moyens de soutenir ces opérations de catégorisation en fonction de la pertinence des informations contextuelles ?

Dans ce sens, une étude de (Kwasnik, 1989) sur les pratiques classificatoires quotidiennes des individus au travail nous éclaire. Il en ressort que l'usage prévu ou les buts pour lesquels un document est conçu est le premier critère de catégorisation invoqué. Ainsi, « *the use to which something is put is often the basic level at which it will be classified* », (Kwasnik, 1989).Pour illustrer le sens de ce critère, elle donne l'exemple d'une personne qui pourrait utiliser des critères de distinction, parmi des livres sur des sujets variés, les livres acquis à une certaine période, le format et la taille de certains livres, tandis que tous ces livres pourraient être physiquement autant que cognitivement groupés comme des « livres utilisés pour l'enseignement ». Ainsi Kwasnik utilise le mot « use » pour désigner ce que nous appellerions une « activité métier ». On voit ici les limites de la traduction, car si nous sommes capables tel que nous l'avons fait précédemment, de distinguer en langue française, les nuances entre des notions comme « l'usage » et « les pratiques », il serait hasardeux de se lancer dans ce genre de considérations appliquées à la littérature anglophone. Aussi, nous nous fierons plutôt à l'exemple donné par (Kwasnik, 1989) pour considérer que l'activité est un des critères prégnants de catégorisation dans les choix classificatoires des individus au travail.

Concernant la pertinence des informations contextuelles à choisir, (Lee, 2011) pense que l'arbitrage final est toujours effectué par des agents humains. Ainsi, pour un contexte de recherche scientifique, il écrit : « *The majority of rich contextual information about data takes the form of tacit knowledge of researchers (…) and is often conveyed through individuals and communities of practice (…) rather than formal information systems.* »

(Bergman *et al.*, 2003), affirment quant à eux, de manière plus radicale, que les concepteurs de systèmes de PIM échouent à reconnaître les besoins spécifiques des systèmes de PIM. Ils affirment que ces systèmes de PIM devraient prendre en compte la valeur ajoutée subjective

des attributs donnés aux informations par les utilisateurs, pour réduire la charge cognitive et faciliter la recherche d'information. La même information peut être en relation avec différents sujets pour différents utilisateurs :

> *« PIM system serves only one user, it should make use of subjective attributes; those that are given to the information by the user while interacting with it. These attributes are often episodic and idiosyncratic in nature, meaningful to the person himself but possible meaningless to others (…). » (Bergman et al., 2003)*

Pour (Barreau, 1995), dans le champ du PIM, le contexte correspond au contexte de travail, qui inclut la nature de la tâche, le sujet traité, les outils disponibles, le public concerné par le travail, et l'usage attendu de l'information. Elle souligne le bénéfice qu'apporteraient des systèmes de points d'accès multiples, tout en soutenant que les systèmes de PIM doivent être flexibles et adaptables aux besoins des tâches et des idiosyncrasies des individus qui les utilisent.

> *« For PIM systems, output must be criterion-based, subject to the needs and objectives of the individual rather than normalized to generic tasks for multiple users as in general ISAR [Information Storage And Retrieval] Systems.*
>
> *(…)*
>
> *Flexibility and variability of output become important criteria in evaluating PIM systems.*
>
> *(…)*
>
> *Successful PIM systems, and software used to implement these systems, should accommodate situational factors to facilitate system use » (Barreau, 1995).*

3.5.6.2 Contexte et collectif

Le contexte jusqu'ici a été considéré dans le champ de la gestion personnelle de l'information, donc dans une acception axée plutôt sur les pratiques individuelles, mais dans une organisation, les individus travaillent en interaction. Aussi, afin de tenter de saisir toute la complexité de cette notion, nous proposons ici de rassembler les éléments qui nous paraissent

pertinents, contribuant à l'émergence d'un contexte en situation de travail dans une organisation. Il s'agit ici d'une liste d'éléments pouvant influer sur la définition du contexte dans une situation donnée. Ainsi, tous ces éléments n'entrent pas nécessairement en jeu lors de la caractérisation d'une activité ou d'un événement, mais ce sont les différentes dimensions que nous admettons comme pouvant y contribuer.

Du point de vue de l'organisation, qui est souvent perçue par les individus comme apportant un certain nombre de contraintes, la définition du fonctionnement de l'organisation n'incluant pas nécessairement une prise en compte leur point de vue, nous retiendrons comme critères pouvant éventuellement contribuer à définir le contexte sur un plan conceptuel :

- le type d'organisation : entreprise, administration, institution, association
- le mode d'organisation : hiérarchique, horizontale
- le type de division du travail : tayloriste, en réseau, par projet
- les règles de l'organisation et le (ou les) système(s) de régulation à l'œuvre en son sein

(Ces derniers éléments sont énoncés ici, dans un souci d'exhaustivité mais ils dépassent le cadre de notre propos, et ne seront donc globalement pas traités dans la suite de ce travail.)

Du point de vue des collectifs en jeu dans une organisation, et dont le fonctionnement dépend des éléments énoncés ci-dessus, nous considérons qu'il existe trois sortes de collectifs plus ou moins étendus :

- le métier
- le collectif de travail : projet, équipe, service, département
- le réseau de professionnels étendu au delà de l'organisation et entretenant des liens plus ou moins distendus

Au sein de ces collectifs, des systèmes de représentations plus ou moins partagées sont à l'œuvre dans les processus d'échanges, de coopération, collaboration ou coordination. Nous revenons dans la partie suivante sur la définition du concept de métier et sur les systèmes de représentation qui peuvent éventuellement être médiatisés par des artefacts. Il faudrait aussi pouvoir établir une typologie des liens sociaux ou aspects relationnels, en fonction de la fréquence des échanges, de leur teneur, des habitudes, des normes, et des propriétés qualitatives de ces relations (relations professionnelles, amicales…).

Au niveau individuel, les éléments qui jouent un rôle dans une définition d'un contexte à un moment donné sont les suivants, (qui ont été, pour la plupart, détaillés précédemment) :

- L'activité
- Le métier
- Les artefacts (SI, application présentes dans le SI et leurs affordances)
- Le sujet individuel (facteurs psychologiques, cognitifs, sociaux, relationnels, affectifs, conatifs et les représentations des individus)
- Les compétences
- Le niveau d'expertise
- La maitrise des artefacts

Enfin, nous pouvons identifier des facteurs extérieurs de contrainte, plus ou moins prononcés, et qui varient selon le type d'organisation, le mode d'organisation, le type d'organisation du travail et les règles de l'organisation, qui sont les suivants :

- L'aspect temporel : degré d'urgence, temps disponible, temps alloué, durée du travail, activités collectives synchrones ou asynchrones
- L'aspect géographique : distribué, dans un même lieu, ou les deux alternativement (télétravail).

Nous allons maintenant nous intéresser au concept de métier, qui constitue, selon nous, non seulement un point d'articulation entre l'exercice individuel et collectif des activités dans les organisations, mais incarne aussi le lieu de convergence d'une tension qui a été mise en évidence entre l'activité métier et les activités secondaires, vues comme des moyens pour réaliser les premières, comme l'activité de gestion de l'information. Ainsi, d'après (Osty, 2002, p.19) :

> *« Pour les individus, le métier désigne l'horizon de projection donnant sens à l'expérience de travail. L'accomplissement de soi est rendu possible par une confrontation problématique de l'homme avec une matière résistante à transformer. Le savoir de métier relève d'un art non réductible à l'application des règles formelles de travail. La référence au métier est bien souvent teintée d'amertume, d'une quête de reconnaissance non assouvie.*

Pour le management de l'entreprise, la référence au métier constitue une manière de revalorisation symbolique de certaines fonctions et d'organisation des filières d'emplois. Mais elle ne présage pas d'une certaine forme d'émancipation sociale, vis-à-vis de la dépendance contractuelle de l'entreprise. La vigueur des modes de gestion des hommes montre à l'inverse une reprise en main de la gestion des savoirs de métiers, là où les anciennes communautés professionnelles en faisaient leur domaine réservé et pesaient dans les relations sociales. »

3.5.7 L'évolution de l'organisation du travail

Les modèles contemporains d'organisation du travail, caractérisés par le développement de la flexibilité pour s'adapter à la fluctuation de la demande du marché, qui s'accompagne de l'externalisation de certains services de l'entreprise, de l'organisation par projet au sein des organisations, de la décentralisation des pôles de décision, et du développement de l'organisation en réseau, aboutissent à des mutations des formes de travail qui nécessitent des adaptations notoires de la part des acteurs des organisations adhérant à ces phénomènes. Ainsi, une plus grande polyvalence est nécessaire, pour le développement de leur employabilité, de même qu'une autonomie accrue, qui se traduit par le développement de l'auto-contrôle, et qu'une disponibilité croissante, due à l'exigence d'être joignable à tout moment et en n'importe quel lieu, impliquant donc d'être connecté en permanence (organisation en réseau). Ces évolutions des modèles d'organisation du travail impactent aussi la gestion des ressources humaines, qui s'individualisent et deviennent de nouveaux lieux de négociations (salaires, fiches de postes, définition des compétences mobilisées) (Boltanski et Chiapello, 1999).

La recherche de la flexibilité, qui trouve son origine dans des motifs économiques s'est traduite par « des techniques d'organisation et de gestion efficientes de type zéro stock, juste-à-temps, gestion en flux tendus » (Pesqueux, 2005). Il s'agit d'une flexibilité appliquée à la production des biens, principalement. La flexibilité se caractérise aussi par la capacité d'apprentissage de l'organisation nécessitant de la mémoire, si bien qu'elle est liée aux performances des hommes et des systèmes d'information. La flexibilité se matérialise en prenant pour variable d'ajustement les effectifs, en recourant à l'externalisation pour répondre

aux fluctuations de l'activité, phénomène soutenu par les réseaux de télécommunication. (Pesqueux, 2005).

D'après (Osty, 2002, p.36) concernant la flexibilité :

> *« L'introduction des nouvelles technologies absorbant la contrainte de coût rend obsolète une conception de l'efficacité reposant sur la productivité directe du travail humain. Les nouvelles modalités de la concurrence supposent davantage de flexibilité de l'appareil de production et s'accomodent mal de la rigidité d'une production de masse, peu apte au changement permanent de modèles de produits, de procédés de production ou de délais de production raccourcis. (…)*
>
> *La mondialisation des échanges, l'accentuation de la concurrence, la demande de produits plus variés et de qualité, l'émergence de nouveaux services, le renchérissement du loyer de l'argent, l'accélération du progrès technique, tels sont désormais les éléments de la nouvelle donne de l'environnement, qui introduisent de nouveaux modes de compétition, fondés sur la différentiation. Le facteur coût dans la compétitivité des entreprises s'efface au profit d'équations plus complexes combinant innovation, qualité, flexibilité, réactivité et rentabilité. » (Osty,2002, p.36)*

L'organisation en réseau repose elle aussi sur les infrastructures de communication qui impactent l'économie, la société, et l'organisation du travail. L'organisation en réseau se singularise par les idées suivantes selon (Pesqueux, 2005) : le changement permanent (qui engendre la réactivité face à la concurrence), un espace-temps en constant remodelage (par l'immédiateté des communications et la connexion permanente de ses acteurs), et l'innovation organisationnelle permanente (par le foisonnement des outils de coordination, de transmission, partage, communication instantanée, et de réseaux sociaux qui évoluent continuellement). L'organisation par projet se caractérise comme suit :

> *Elle «évoque une entreprise dont la structure est faite d'une multitude de projets associant des personnes variées dont certaines participent à plusieurs projets. La nature même de ce type de projets étant d'avoir un début et une fin, les projets se succèdent et se remplacent, recomposant au gré des priorités et des besoins, les groupes ou équipes de travail. » (Boltanski et Chiapello, 1999).*

Cette organisation par projets modifie la nature des rapports de pouvoir et de hiérarchie à l'intérieur des organisations, l'exercice d'un métier par un acteur ne prenant plus effet

seulement dans la sphère d'un seul service d'une organisation et les acteurs n'étant plus subordonnés uniquement à l'autorité d'un responsable de ce service.

Mais, d'après (Osty, 2002, p.37), ces modifications ne touchent pas que les acteurs au niveau individuel, elles impactent aussi la nature et le fonctionnement des collectifs de travail :

> *« Les caractéristiques des situations de travail contemporaines engagent tout à la fois l'individu, sollicité dans sa capacité d'initiative, d'intelligence de la situation, d'une vision globale et systémique du procès de production, mais aussi le collectif de travail dans ses capacités d'ajustement relationnel, d'intercompréhension et de mutualisation des savoirs théoriques et expérientiels. »*

Dans le cadre de ces évolutions de l'organisation du travail, le concept de métier est-il encore pertinent pour l'analyse de l'activité des personnes dans les organisations ?

3.5.8 Le concept de métier réhabilité

La notion de métier est ancienne et renvoie notamment aux corporations de métiers, au compagnonnage, au syndicalisme. On pourrait penser que l'éclatement des structures du travail évoquées plus haut est propice à la disparition de cette conception de l'activité rémunératrice. Pourtant, on observe, d'après (Osty, 2002), qu'un « phénomène social d'affirmation des métiers pourrait bien s'amplifier, sous le coup des nouveaux enjeux de production, réhabilitant le métier comme une configuration sociale et organisationnelle de la modernité ». Elle fonde cette affirmation, qu'elle résume dans l'expression de l'émergence d'un « désir de métier », sur l'observation de trois processus dans des milieux de travail variés (Osty, 2002, p.50) :

- l'élaboration d'une compétence spécifique à travers la production, l'actualisation et la transmission de savoirs au sein d'un collectif
- la construction d'une identité collective, le collectif assurant la socialisation et l'identification à une communauté de métier
- l'édification de règles sociales, à travers la régulation : le groupe professionnel, stabilisant les règles, cherche à obtenir une reconnaissance légitime

On constate donc que le métier, d'après (Osty, 2002), ne se conçoit que dans une dimension collective, d'une part à l'intérieur du collectif métier et d'autre part au sein des organisations.

3.5.8.1 Définition du concept de métier

Commençons par souligner qu'il est possible de retrouver des éléments communs aux métiers, que leur domaine d'exercice soit l'industrie ou le tertiaire : « la même revendication d'un groupe à faire reconnaître sa différence et sa compétence sur un périmètre de techniques » et la « quête d'une stabilité mêlée au souci de statuts garantis » (Clot, 2008). Clot souligne aussi que « autrui » joue un rôle essentiel dans le sentiment d'appartenance à un métier par le biais de la reconnaissance par la hiérarchie et par les pairs. Le métier, vu comme un ensemble de personnes partageant des pratiques, des savoirs, et savoir-faire, « devient synonyme de communauté ou de collectif d'appartenance », et, par extension implique que « le jugement des pairs octroie l'appartenance au métier ».

De la même manière, d'après les trois processus observés par F. Osty, avoir un métier c'est tout d'abord « s'inscrire dans une dynamique de production d'un savoir opératoire érigeant l'expérience en mode d'apprentissage privilégié ». En effet, on assiste à un glissement de la conception des savoirs correspondant initialement aux connaissances explicites, formelles, acquises et sanctionnées par un diplôme attestant des qualifications, vers un modèle plus axé sur les compétences, lié à la reconnaissance de l'activité réelle des individus dans le travail, individus dont les connaissances implicites et les savoirs pratiques sont fondés sur l'expérience. Ces compétences reconnues au sein des organisations ont vocation à être partagées et évaluées par les pairs. Ainsi, « La stabilisation des modes opératoires repose sur la capacité d'un groupe social à organiser et leur circulation et leur transmission. L'explicitation d'un référentiel commun est la condition nécessaire de la mise en visibilité des savoirs engagés dans l'action » (Osty, 2002, p205).
Le deuxième processus qui permet de définir la notion de métier, le sentiment d'une identité collective, se caractérise par le partage de valeurs, normes, système de représentations, et par une culture professionnelle (règles techniques, comportements, langage spécifique) qui marque l'appartenance à une communauté métier. La quête de reconnaissance sociale accompagnant cette identification à l'activité de travail se double d'une identification plus ou

moins prononcée à l'organisation. Cette identité collective se construit dans les interactions quotidiennes de travail.

Enfin, à travers l'édification de règles et de compétences pour la régulation, à travers sa capacité à les transmettre et les diffuser assurant ainsi leur pérennité, une communauté métier joue sa reconnaissance au sein de l'entreprise ainsi que la légitimité de sa professionnalisation. D'après (Osty, 2002, p205), « La construction de règles locales légitimes émerge comme forme de régulation conjointe à part entière, même si la question de son articulation avec le niveau de l'entreprise reste posée. »

La notion de référentiel commun, qui porte des représentations partagées par les acteurs qui interagissent et le construisent, est très importante dans la constitution d'une pratique en métier. « Le métier s'échafaude sous sa forme collective par stabilisation d'un référentiel commun de pratiques, de valeurs partagées et d'un accord sur le sens de la mission. » (Osty, 2002, p. 212). Ce référentiel se construit à travers les interactions dans les situations de travail et sur la mise en relation des savoirs individuels faisant du savoir « un bien commun ».

> *« la circulation verticale et horizontale des savoirs d'action forme l'épicentre de l'identité collective de métier. En ce sens l'appartenance collective est ancrée sur cette capacité communicationnelle d'échanges d'expériences et de transformation de l'expérience en savoirs pratiques. La mutualisation des savoirs est la condition de construction d'un collectif de métier car elle conditionne la fiabilité des modes d'ajustement. » (Osty, 2002, p.215)*

Chez (Osty, 2002), la constitution d'un référentiel commun, ou référentiel métier, est vue comme la construction d'une compétence d'ajustement, collective. Il permet la coordination, l'intercompréhension, la circulation et la rencontre des savoirs.
Etant donné que ces outils sont constitutifs de l'existence d'un métier, dans une organisation, voire au delà, et qu'ils relèvent d'une dynamique sociale de métier, ils doivent conduire à reconsidérer la place des métiers, pour offrir des outils de gestion différenciés, sans mettre en péril la cohésion de l'entreprise.

Ainsi, le soutien des collectifs de travail à l'aide d'outils adaptés à leurs configurations organisationnelles, pour la transmission des compétences et du savoir apparaît être un enjeu

fort de la transformation actuelle des métiers. De ce fait, la gestion des documents, utile à la transmission des connaissances et compétences d'un métier se révèle cruciale.

3.5.8.2 Le passage de l'individuel au collectif à travers le prisme du métier

(Clot, 2008) propose une conception dynamique de l'expérience du métier, c'est-à-dire, le processus d'appropriation du métier et de sa maîtrise. D'après lui, lorsqu'un opérateur commence dans son métier, il respecte au maximum la dimension impersonnelle (prescrite) du travail. Petit à petit, il s'approprie cette dimension, notamment à travers les échanges avec ses pairs, puis devient un expert. La dimension personnelle dans le métier atteint alors ses limites, le genre professionnel dans lequel il est exercé étant trop restreint. Le sujet va donc devoir inventer, c'est -à- dire faire l'inventaire de ce qui existe déjà, identifier les limites et trouver avec le collectif de nouveaux moyens pour exercer son métier. Cette stylisation du genre professionnel rend l'expert auteur de son métier. La dimension personnelle du métier devient plus présente, et ce processus peut aboutir à un enrichissement des tâches prescrites initiales à partir des tâches réalisées au niveau de maîtrise de l'expert, et à une dimension impersonnelle (prescrite) plus forte.

« C'est que le travail collectif soutenu par le collectif de travail a rendu possible ces développements mutuellement indépendants » (Clot, 2008). C'est aussi par ce biais que la mémoire transpersonnelle du genre professionnel s'affranchit de chacun pour être mieux disponible pour tous.

Ce processus d'évolution de l'exercice et de la maîtrise d'un métier, qualifié de « Cercle vertueux » des métiers par Y. Clot, est représenté dans la figure suivante.

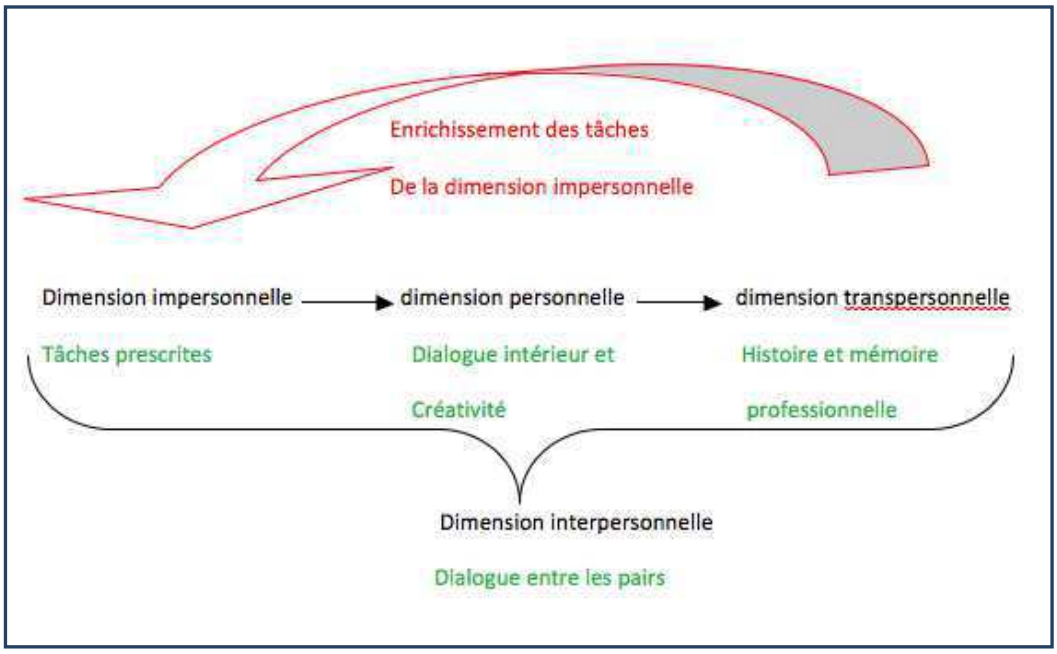


Figure 14. Représentation du cercle vertueux des métiers d'Yves Clot (2008)

3.5.8.3 Référentiel commun : une représentation collective du contexte

Ainsi, le contexte peut aussi être considéré comme un élément de spécification du sens des communications, notamment dans les processus de partage de représentations individuelles ou de connaissances. Un des outils qui peut rendre ce service de mise à disposition du contexte des communications consiste en la mise en place de référentiels communs, pour les activités collectives.

En ergonomie, les référentiels communs sont vus comme construisant une représentation commune du contexte pour plusieurs opérateurs, comme des modèles de systèmes techniques, des règles d'utilisation, ce sont des moyens de communication et de mise en œuvre pour rendre les activités compatibles. Grâce à un référentiel commun, « les opérateurs perçoivent le contexte de leur activité de la même manière et peuvent ainsi interpréter plus facilement et plus correctement les actions et communications de l'autre. » (Leplat, 2008).

Un référentiel commun est « une représentation fonctionnelle commune aux opérateurs, qui oriente et contrôle l'activité que ceux-ci exécutent collectivement. » (Leplat, 1991).

Il est commun à groupe d'acteurs et dépend de l'activité, des opérateurs et de la situation :

Comme il sert à préparer, réaliser, orienter, et à contrôler l'activité, il est dit *opératif* ou *fonctionnel*.

Cette expression de « référentiel commun » désigne aussi les représentations des acteurs qui travaillent ensemble, dans les mêmes lieux ou des lieux différents, de manière synchrone ou asynchrone. Ainsi, on parle aussi dans la littérature sur ce sujet, de *représentations distribuées.* Enfin, comme « il dépend de la situation dans laquelle évoluent les opérateurs (…) le référentiel est *circonstanciel* ou *situé*. (Giboin, 2004).

Dans un référentiel commun, il est important de noter que ce qui est considéré comme commun correspond « à des propriétés communes des représentations construites et utilisées par les opérateurs » (Leplat, 2001 cité par Giboin, 2004), et repose sur les référentiels individuels. Ici encore, nous notons des rapprochements possibles avec la problématique du « basic level », puisque les référentiels individuels doivent être en partie partagés pour la construction d'un référentiel commun. Ainsi, (Giboin, 2004) souligne l'importance de la question du « degré de communauté entre les représentations des acteurs » ou de « l'intersection des référentiels individuels ». La question des catégories d'information organisées selon le critère du « basic level », c'est –à-dire le plus efficace cognitivement, et celui qui est partagé le plus aisément entre les personnes, nous apparaît donc de premier ordre dans la construction d'un référentiel commun.

En termes de contenu, cet outil peut contenir différentes formes d'informations, concernant soit la tâche à accomplir (buts, étapes, contraintes), le problème à résoudre, (les solutions ou hypothèses), l'équipe de travail (statut, compétences, préférences des interlocuteurs),ou les échanges passés. Il peut aussi contenir des éléments conatifs (en lien avec la motivation), affectifs, sensori-moteurs, ou sociaux. (Giboin, 2004). Un référentiel commun, selon (Giboin, 2004), sert plusieurs objectifs :

- la reconstruction de l'implicite dans la définition d'une tâche ;
- la reconnaissance des intentions ;
- l'adaptation aux autres ;
- la communication ;
- la coordination ;
- la répartition de la charge de travail.

Concernant la maintenance d'un dispositif comme un référentiel commun, ses mises à jour doivent être fréquentes : étant donné que l'activité est un élément dynamique, cet outil de soutien aux activités doit lui aussi être dynamique. Sa construction doit avoir pour objectif la

construction d'une compétence collective. Celle-ci s'avère d'autant plus difficile que les membres du groupe participant ont des spécialités ou formations différentes.
Le prototype élaboré dans le cadre du projet Miipa-doc, qui sera présenté dans ce qui suit peut être considéré comme un outil informatique de construction et d'alimentation d'un référentiel commun, même s'il remplit également d'autres missions. Du point de vue de la typologie des systèmes d'information, il s'agit d'un SI « d'inspiration 2.0 ».

3.5.9 Présentation du prototype Hypertagging

Hypertagging correspond à la combinaison de deux types d'outils qui existent indépendamment et notamment sur le Web, d'une part les outils de tagging, et d'autre part, la classification à facettes. Comme nous l'avons vu au chapitre 1 (partie 1.4.6), divers projets tentent de structurer les espaces de tagging à l'aide de facettes, selon des méthodes plus ou moins automatisées et pour des ressources issues du Web, la plupart du temps. L'originalité d'Hypertagging est de proposer de taguer manuellement des documents de travail au sein de communautés métier pour une démarche de gestion de connaissances appuyée sur la gestion personnelle de l'information.
La méthodologie d'élaboration de « FolkClassifications à facettes pour la gestion des connaissances métier » est présentée en chapitre 5. Nous présentons ici le prototype Hypertagging, développé dans le cadre du projet Miipa-Doc, et ses fonctionnalités. Les objectifs premiers de ce prototype consistent à faciliter et équiper les opérations de nommage, indexation (ou documentarisation), et de classement des documents de travail, gérés au jour le jour dans l'exercice d'une activité de travail. Ces opérations de nommage, indexation et classement sont qualifiées de coûteuse cognitivement, puisqu'elles nécessitent des micro-prises de décisions très rapprochées, et récurrentes.
Lorsqu'on tague un document à l'aide d'Hypertagging, ces trois opérations peuvent être fusionnées. Ainsi les tags associés à un document, constituent à la fois son titre et des métadonnées. Le prototype dispose d'un système d'apprentissage automatique des règles qui président au classement des documents dans les dossiers en fonction des associations de tags. Ce dispositif propose également le partage de certains espaces de tagging pour permettre le partage de documents et le travail collaboratif. Enfin, le prototype possède des fonctionnalités

de recherches de documents basées sur le filtrage à l'aide des tags sélectionnés par l'utilisateur dans les facettes. Les tags sont classés dans des facettes manuellement.

3.5.9.1 Présentation des deux modules d'Hypertagging

Hypertagging est structuré en deux modules, le module du Saver (pour le tagging et le classement) et le module du Finder (pour la recherche). Nous présentons ces deux modules dans les copies d'écran suivantes (Figures 15 et 16).

Dans chacun de ces deux modules on peut observer deux panneaux. Le panneau de gauche présente les facettes, et est identique dans les deux modules, puisque les facettes et leurs mises à jour sont synchronisées entre les deux modules.

Dans ce volet de gauche, qui présente les facettes, on peut remarquer trois onglets : un onglet intitulé « Général », un deuxième onglet intitulé « Document », et un troisième onglet intitulé « Loisirs ». Chacun de ces onglets rassemble des facettes en fonction d'éléments de la méthodologie présentés en chapitre 5. Les onglets sont la représentation d'un groupement de facettes en fonction d'éléments contextuels, qui fournit la structuration globale de la classification.

Une facette peut être affichée dans plusieurs onglets par la duplication. Sous l'onglet « Général », un bouton « créer Facette vide » est présent, il sert à créer de nouvelles facettes au fur et à mesure des besoins.

Le partage d'espaces de tagging se matérialise par le partage d'onglets. Un bouton est prévu à cet effet, dans la partie haute, à droite du volet des facettes dans chaque module.

Le volet de droite du module Saver est détaillé dans la Figure 17.

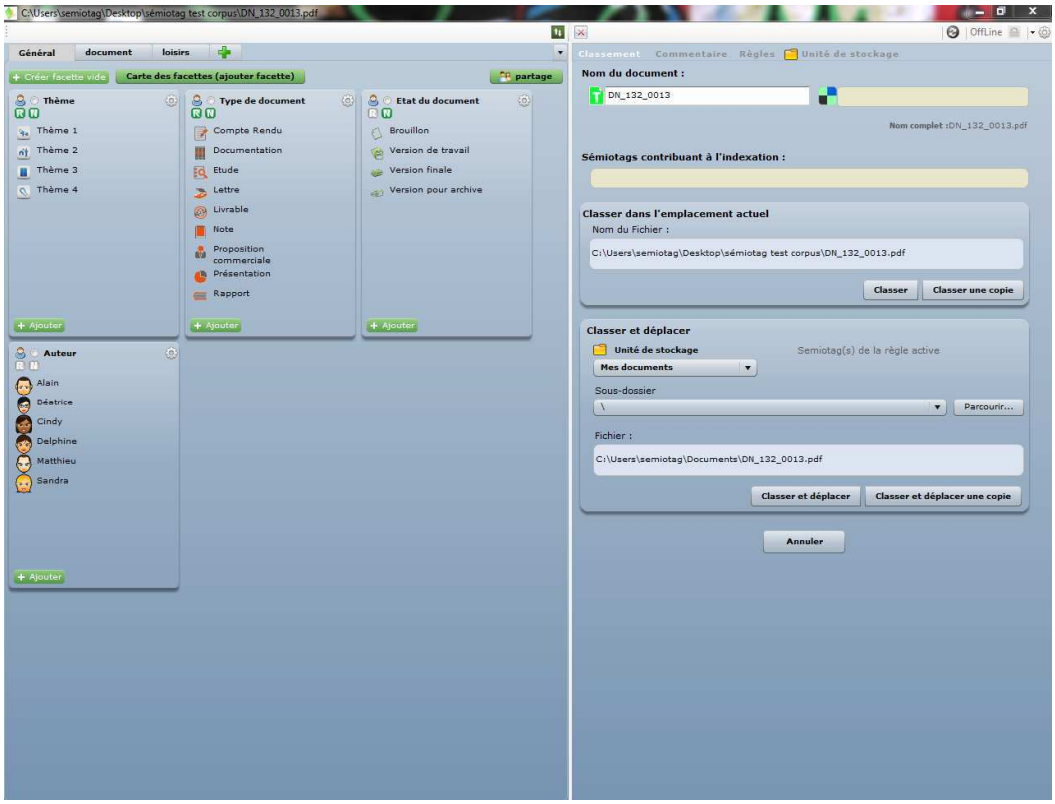


Figure 15. Module Saver d'Hypertagging

Dans le module du Finder (Figure 15), un tag est en surbrillance, cela signifie qu'il a été sélectionné comme filtre de recherche. Ainsi dans le volet de droite du Finder, la liste des résultats de cette recherche de documents affiche les titres (composés entre autres des tags) des documents retrouvés. On peut ainsi remarquer que les documents présentés dans la liste des résultats du module Finder, contiennent tous dans leur titre le tag « Compte-Rendu » qui a été sélectionné comme filtre de recherche (en surbrillance) dans le volet de gauche. Ces documents ont été préalablement tagués avec ce tag « Compte-Rendu ».

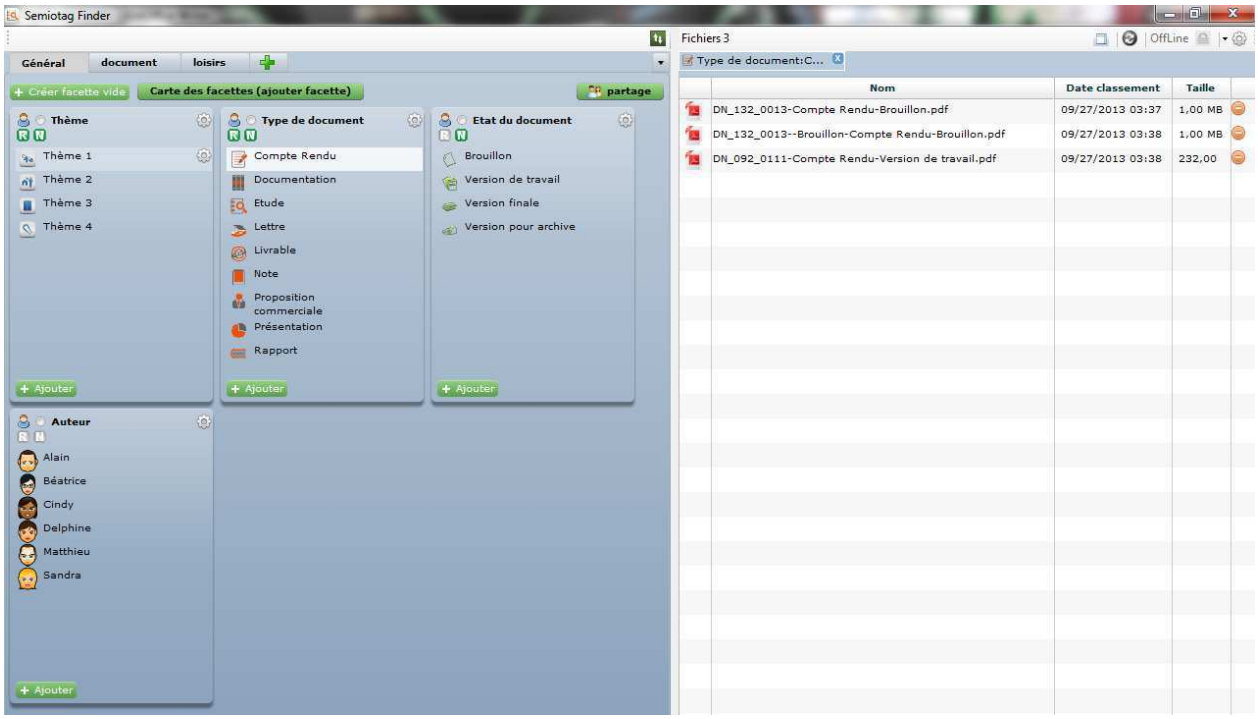


Figure 16. Module Finder d'Hypertagging

3.5.9.2 Les options de tagging

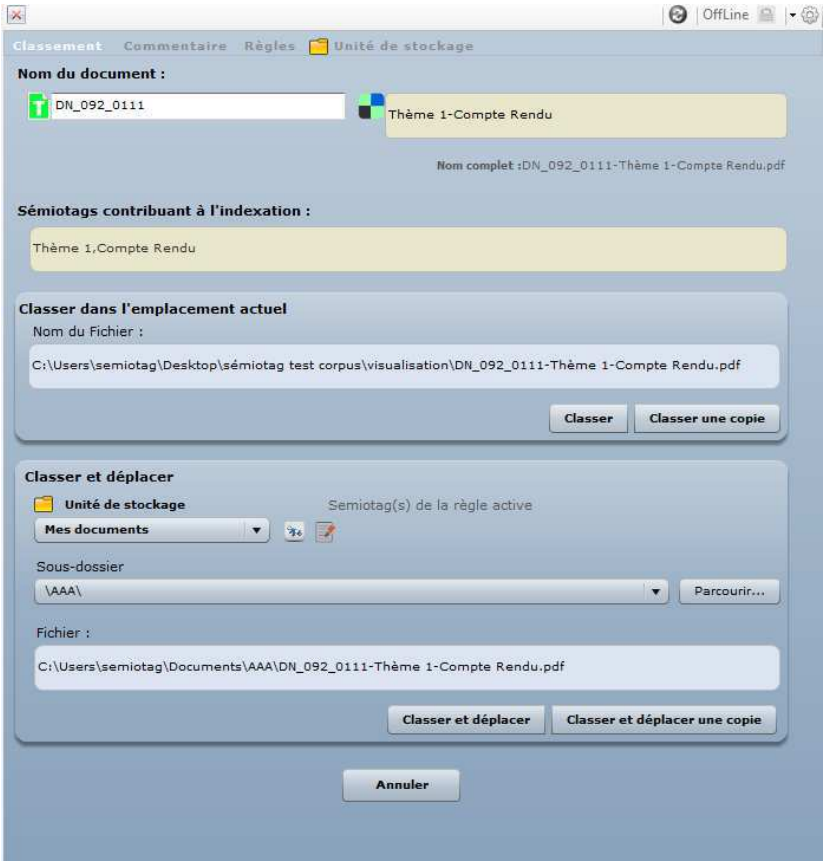


Figure 17. Panneau de droite du module Saver

Dans cette Figure 17, le bouton T vert en haut à gauche sert à déterminer si le nom original du document est conservé durant le tagging du document. S'il est en vert le nom original du document est conservé comme c'est le cas ici. Il est possible d'inverser ce paramètre en cliquant une fois sur ce bouton, alors il devient rouge et le nom original n'est pas conservé.

Le deuxième rectangle en haut à droite (en grisé) rappelle les tags qui sont sélectionnés pour un document. Sous ce rectangle le nom complet du document est indiqué en fonction de ces deux paramètres. En effet, les tags peuvent être insérés comme métadonnées mais ne pas contribuer au nommage du document, et ainsi ne pas apparaître dans le nom complet du document. Ceci est paramétrable dans les propriétés de chaque facette, et peut être modifié en cours de tagging. C'est la raison pour laquelle un champ grisé est prévu, qui rappelle les Sémiotags contribuant à l'indexation.

Il est aussi possible de paramétrer globalement l'usage du dispositif pour que les tags ne contribuent pas au nommage du document, ceci est utile dans l'archivage de documents, par exemple. En effet, cette option permet de ne pas modifier le nom du document. Pour ce faire, il faut effectuer un clic sur le bouton composé de quatre petits carrés, qui précède le premier champ grisé (depuis le haut de l'écran).

3.5.9.3 Les options de classement

Une fois les tags associés au document, il existe plusieurs options de classement. Ainsi, l'on peut choisir de classer le document en cours d'indexation dans le même emplacement que celui où il est stocké au moment du classement. C'est le sens du bouton simple « Classer ». En effet, pour le moment l'appel de l'application se fait depuis un explorateur de fichier à l'aide d'un menu contextuel. Les fichiers sont donc préexistants.

Il est également possible de classer une copie de ce document en cours d'indexation, au même emplacement. C'est le sens du bouton simple « classer une copie ». Cette action duplique le document, et seul le document dupliqué possède des tags.

On peut aussi classer le document dans un autre dossier. C'est à cette fonction qu'est dédiée la partie basse « Classer et Déplacer » de la Figure 17. Cela fonctionne selon les mêmes principes que ceux évoqués ci-dessus mais avec la possibilité de parcourir un système de gestion de fichier, grâce au bouton « Parcourir ».

Dans le cas où l'on souhaite classer le document dans un autre dossier, trois cas peuvent se produire qui font intervenir le système d'apprentissage des règles de classement en fonction des associations de tags. C'est dans le paramétrage des facettes que l'on décide si les tags

d’une facette vont servir ou non à déterminer des règles de classement, en fonction de l’usage que l’on souhaite faire du dispositif. Nous y revenons dans la suite de cette partie.

Cas N°1 : il n’existe pas de règles : on parcoure le système de fichier pour trouver l’emplacement où classer ce document. Une fois le bon dossier identifié, on classe soit le document, soit une copie du document.

Cas N°2 : il existe une règle déjà créée pour l’association des tags sélectionnés pour le document : la règle s’applique, et le document (ou la copie du document) est classé dans le dossier indiqué dans le champ « Sous- Dossiers ». C’est le cas illustré dans la figure 17, le sous-dossier est intitulé « /AAA/ ». Les icones présentes au dessus du champ « Sous-Dossiers » rappellent les Sémiotags qui fondent la règle en cours d’application.

Cas N°3 : il n’existe pas encore de règle de classement mais certaines facettes, dont nous avons sélectionné des tags, sont paramétrées pour créer des règles de classement : une nouvelle règle de classement va être créée, et elle est paramétrable à cette étape. Ainsi, il est possible de classer le document dans un dossier spécifique en parcourant le système de fichier, et c’est ce sous-dossier qui sera intégré dans la nouvelle règle en association avec la combinaison des tags issus des facettes paramétrées pour créer des règles. Il est aussi possible de classer le document dans l’espace de stockage défini par défaut, qui est « Mes Documents », alors la règle associera cet espace de stockage « Mes Documents » à la combinaison de tags sélectionnés.
Pour visualiser les règles de stockage créées il faut sélectionner l’onglet « Règles » tout en haut de ce panneau droit du module Saver. Les autres onglets de cette zone, servent à visualiser les éventuels commentaires ajoutés sur un document, et les différents espaces de stockage disponibles selon la configuration du programme. Ces éléments ne sont pas détaillés ici, et sont sujets à des évolutions en cours. L’interface de tagging – classement, présentée en Figure 17 pourrait éventuellement subir des modifications pour être simplifiée.

3.5.9.4 Les options de paramétrage des facettes

Les options de paramétrage des facettes sont présentées en Figure 18.

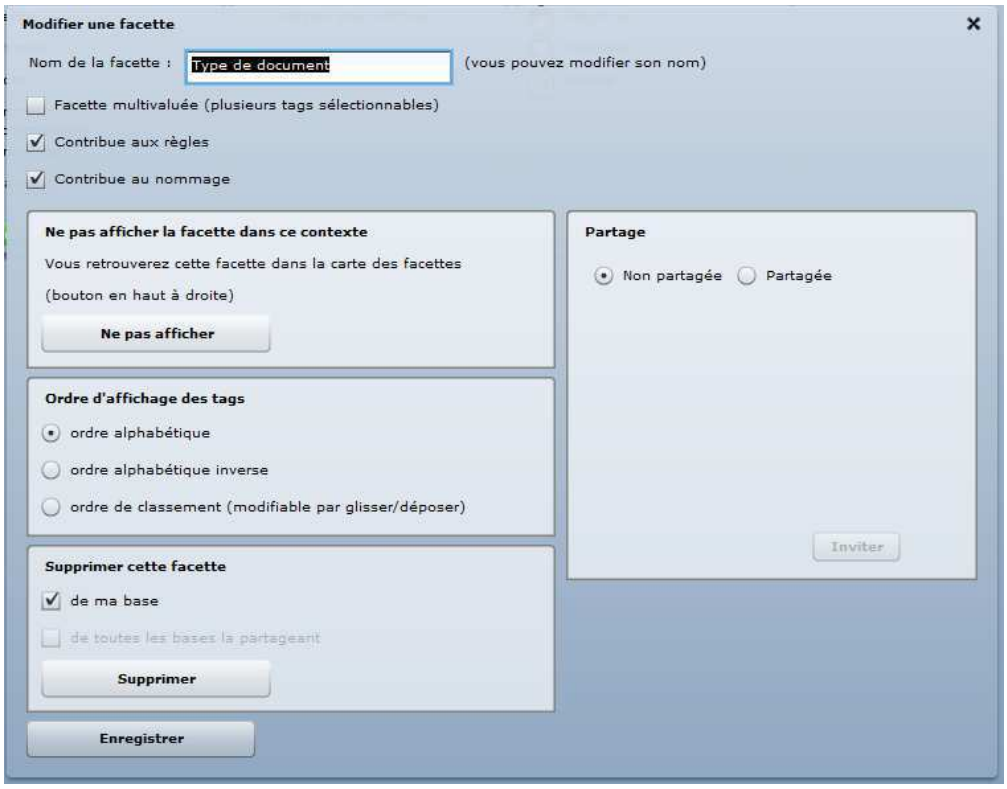


Figure 18. Paramétrage d'une facette dans Hypertagging

Dans cette figure 18, on remarque que l'on peut paramétrer le nom de la facette, ainsi que quelques autres options (de haut en bas) :

- si l'on coche la case « Facette multivaluée », alors il est possible de choisir plusieurs valeurs de cette facette pour taguer les documents, dans le cas contraire ce n'est pas possible.
- si l'on coche la case « Contribue aux règles », alors les tags de cette facette vont créer des règles de classement, dans le cas contraire, les tags ne créent pas de règle.
- si l'on coche la case « Contribue au Nommage », alors les tags de cette facette seront visibles dans le nom du document, dans le cas contraire les tags ne contribuent pas au nommage.

Il est également possible de supprimer l'affichage de cette facette de l'onglet (appelé dans cette version « contexte ») où elle est insérée, elle ne sera pas pour autant supprimée de la base des facettes. On peut également paramétrer l'ordre d'affichage des tags et supprimer

totalement la facette. Ces options sont accessibles via la roue située tout en haut à droite de la facette présentée en Figure 19.

Figure 19. Exemple de Facette

Dans cette Figure 19, qui présente un exemple de facette, on peut remarquer les lettres « R » et « N » sous le nom de la facette. Ces lettres indiquent visuellement le paramétrage des facettes concernant les règles de classement et le nommage des documents. Ainsi, quand une de ces lettres est verte, la fonctionnalité est activée, quand elle est grisée, elle est désactivée. Ici, la facette « Etat du document » contribue au nommage mais ne sert à pas à créer des règles de classement. Le bouton situé en bas de la facette « + Ajouter », sert à ajouter des tags.

Figure 20. Ajout d’un tag

Lorsque l'on souhaite ajouter un tag ou le modifier, une fenêtre telle que présentée en Figure 20, s'affiche. Il est nécessaire de choisir un libellé pour le tag. On peut choisir d'ajouter une icône en sélectionnant le bouton « Choisir dans la bibliothèque » qui donne accès à des jeux d'icônes proposés dans le prototype, ou en sélectionnant « Parcourir votre ordinateur » pour choisir parmi ses propres images. Pour terminer il faut enregistrer. Il est aussi possible dans cette fenêtre de supprimer le tag.

3.5.10 Systèmes d'information structurés et Systèmes d'information 2.0

Après la présentation du prototype Hypertagging, qui constitue un exemple de Système d'Information « d'inspiration 2.0 » , nous nous intéressons à leur positionnement vis-à-vis des systèmes d'information structurés, ainsi qu'aux limites de chacun de ces systèmes, dans cette partie.

Dans les entreprises, ce qui est observé est plutôt une juxtaposition des systèmes d'informations formels et personnels, (Cotte, 2007b ; Guyot, 2009). Cette juxtaposition incite à repenser les rapports de complémentarité ou de concurrence entre ceux-ci (Guyot, 2009), et la possibilité de l'existence d'une logique cohérente dans les usages complémentaires de ces systèmes d'information contrastés.

3.5.10.1 Coexistence des SIP et des SI structurés

Ainsi, d'après (Cotte, 2007b), « Il se pratique ainsi une forme d'organisation spontanée de l'espace documentaire qui ne ressemble certes pas à celle qui est recommandée par les spécialistes de la discipline, les documentalistes, les archivistes, mais qui n'en est pas pour autant inefficace ». Poursuivant au sujet des systèmes d'information, l'auteur soutient que ces organisations personnelles de l'information, pas toujours cohérentes ni en règle avec les recommandations des professionnels de l'information, sont à intégrer dans les stratégies de gestion globale de l'information des organisations. Ainsi, il écrit, « (…) on peut aussi penser que l'accomplissement des promesses initiales des systèmes ne peut réellement advenir que par le truchement de ces ajustements qui du même coup, ne sont pas à côté des systèmes, mais en font partie de manière intégrale. » (Cotte, 2007b). Cette perspective de coexistence des SIP

et des SI structurés se double de d'une tendance à ce que (Guyot, 2009) appelle « un déversement des systèmes d'information personnels (SIP) dans un espace collectif ».

Si l'existence des SIP juxtaposés peut produire des redondances, des coûts importants en termes de temps passé à maintenir le SIP ou à y rechercher de l'information, la mutualisation de ces « micro- systèmes hétérogènes non interopérables » (Guyot, 2009), paraît poser plusieurs types de problématiques.

Ainsi, nous relevons la tension entre l'individuel et le collectif, déjà évoquée, cette question concernant aussi le type de collectif pour lequel une mutualisation des SIP peut fonctionner. Des problèmes d'interopérabilité des SI personnels (par exemple les systèmes de gestion de fichiers par dossiers arborescents personnels), comme les variations de vocabulaires, des classements, des critères de mémorisation ou de recherche variables selon les individus, ainsi que la maîtrise variable des outils. Enfin, s'agit-il d'une intégration totale, partielle, ou d'une cohabitation, et dans quelle mesure cette « intégration » demande-t-elle une adaptation de la part de leur concepteur- gestionnaire-utilisateur ?

Dans cette coexistence, d'un point de vue schématique, on trouve donc d'une part des systèmes formels collectifs, comme par exemple, les intranets, dans lesquels les individus vont chercher l'information institutionnelle, administrative, technique, ou relative aux normes ; et les espaces personnels dans lesquels le classement est plus aléatoire, puisque lié au tissage progressif de l'activité, et qui contiennent éventuellement des copies des documents issus de l'intranet, des proto-documents, des DOcuments Pour l'Action (DOPA) (Zacklad, 2006), et des documents ressources issus de sources variées, telle que le Web, les bases de données accessibles, les bases de connaissances, les application dédiées etc., ainsi que des documents finalisés et leurs versions intermédiaires.

D'un point de vue expérimental, ce que l'on remarque d'après une étude menée dans le cadre des expérimentations qui seront présentées dans le chapitre suivant, dans une démarche de construction d'une typologie théorique des types de documents manipulés par les acteurs observés, - typologie issue d'observations empiriques - (Desfriches Doria, 2012a, 2012b), les types de documents les plus manipulés sont les documents de travail collaboratif, définis comme les « documents rédigés collaborativement, au sein d'un groupe où le travail des individus est fortement dépendant de celui des autres, comme c'est le cas dans l'organisation par projets » (Desfriches Doria, 2012b). Ces types de documents sont typiquement ceux dont on trouve des copies dans les SIP et sur les serveurs partagés.

Dans les problématiques associées à la gestion de ces documents de travail collaboratif, deux approches existent avec des niveaux de contrainte variables. Ainsi, on trouve des approches descendantes (top-down), prescriptives, pour le nommage et le classement des documents, ce qui permet de normaliser les métadonnées associées aux ressources et présente les avantages classiques des langages documentaires.

Les approches ascendantes (bottom-up) sont basées sur la participation et l'enrichissement des métadonnées par les utilisateurs, de manière émergente. Cette dernière approche est beaucoup moins prescriptive que la précédente, mais ne permet pas l'application de conventions de nommage, ni la normalisation des métadonnées par un Système d'Organisation des Connaissances structuré et formel comme, par exemple, un thésaurus. Ainsi, dans une approche ascendante, on peut trouver le même document à plusieurs endroits de stockage, avec des noms différents et des métadonnées elles aussi variées, ce qui peut poser des problèmes de redondances, d'expansion inutile du volume des données, des problèmes de coordination, de gestion des versions, et des problèmes de doublons lors de la recherche d'information. Lors d'un choix de ce type de solution ascendante, il faut donc chercher à réduire ces obstacles au bon fonctionnement de la coordination et du travail collaboratif en général.

D'après (Guyot, 2002), dans cette tension entre l'individuel et la collectif, on trouve aussi des problématiques qui tournent autour des intérêts divergents entre le local et le collectif. Ainsi, « Rechercher un dénominateur commun au nom d'un intérêt général et supérieur qui unirait les membres d'une même entité ne peut simplement s'appuyer sur le fait que chacun partage les mêmes buts ou partage le même contexte. » (Guyot, 2002)

Cette affirmation pose la question de l'intérêt général et de l'adhésion au dispositif de mutualisation.

D'autre part, cela pose la question du statut des documents et des individus, de leur légitimité, et de la propriété des documents. Enfin, sortir de l'implicite d'un système individuel suppose un travail de négociation sur les formats, la portée des données, leur conditions d'usages. (Guyot, 2002).

Concernant l'adaptation nécessaire des acteurs aux systèmes formels, (Guyot, 2006, p.146) note :

> *« Interconnecter ou centraliser ces univers privatifs touche à la personne et à l'idée que celle-ci se fait de son propre travail, de sa place, de son statut face aux autres. (…) il s'agit (…) de sortir ces systèmes personnels de leur statut privé en développant des espaces communs dans le but d'optimiser une partie des opérations de stockage et de recherche d'informations détenues par les*

individus mais potentiellement utiles à d'autres. Cela suppose de définir ce qu'est une information utile à plusieurs personnes, et engager chacun à penser aux autres en même temps qu'à lui. »

Ainsi, le passage d'un SIP à un système collectif ou semi-formel, implique des modifications du milieu de travail, des positionnements des acteurs, de la perception de leur place, rôle et statut. Ces modifications peuvent être assumées et voulues, ou au contraire, constituer des effets collatéraux qu'il faut anticiper, et sur lesquels des actions de pédagogie et de formation peuvent tenter de jouer pour amplifier l'adhésion au système, en exposant aux acteurs les bénéfices personnels qu'ils peuvent trouver dans l'usage du système et les objectifs poursuivis dans le déploiement des systèmes en question.

En effet, dans le déploiement d'un Progiciel de Gestion Intégré (PGI), soit d'un système d'information structuré, une partie du succès du système dépend du changement organisationnel qui l'accompagne. D'ailleurs, « la thématique du changement organisationnel est fortement liée à celle des technologies devenues vecteur des réorganisations des façons de gérer l'information. » (Guyot, 2006, p.37).

3.5.10.2 Les PGI et les Workflows

Nous donnons ici une brève description du fonctionnement de quelques SI structurés d'après (Guyot, 2006), puis nous aborderons les impacts du déploiement de ces systèmes en termes organisationnels et en termes de modifications des activités, des métiers, et des représentations des acteurs.

« Une plate-forme intègre ou interconnecte, désormais tous les sous-systèmes régulant un processus particulier. Il [le PGI] centralise à l'échelle de l'entreprise ce qui était auparavant localisé, cette centralisation étant censée offrir une transversalité d'application et d'informations et les rendre disponibles à tous les acteurs. Chacun travaille avec cet outil commun comme si c'était le sien. Utilisateur des informations et des applications qui le concernent, il en est également contributeur, car toutes les données qu'il donne seront prises en compte par différents modules du système général. Cela lui évite de saisir plusieurs fois les mêmes informations. Chaque utilisation renseigne le système.

Ces progiciels gèrent à la fois des données techniques, celles qui relèvent du fonctionnement, des données d'approvisionnement, de commande, puisqu'il y a interfaçage d'applications développées par chaque entité.

(…)

> *Sur le plan pragmatique, ils réduisent les marges de manœuvre des acteurs, les bricolages possibles, par leur rigidité, contraignant parfois ceux-ci à contourner le système par la saisie de données inexactes, pour pouvoir avancer dans une tâche. » (Guyot, 2006, p. 68)*

Dans un article sur l'évolution des SI et les reconfigurations des situations de travail Nicotri, (2004), qui intègre dans cette catégorie que nous appelons les « systèmes d'information structurés », les PGI, mais aussi les outils de workflow, de datamining, les CRM (Customer Relationship Management)et les intranets, note que « les modalités de publicisation des données sont intégrées dans la structuration même du système d'information ».

Les systèmes de workflows sont définis comme suit par (Guyot, 2006, p.67) :

> *« La démarche d'un workflow est de découper les tâches, de les modéliser et d'indiquer les liens qu'elles ont entre elles, selon des « routes » séquentielles, parallèles ou en boucle. Des diagrammes représentent le processus en nommant les différents rôles (de manageur, de rédacteur), en leur associant des acteurs, des règles de coordination (par exemple nature de l'information à transmettre ; personne à contacter) ; des ressources et les documents à mobiliser ou à produire. »*

Au sujet des workflows, on peut ajouter, d'après (Nicotri, 2004), qu'il n'y a pas de négociation possible avec des procédures informatisées, par opposition aux situations interpersonnelles de recherche de compromis ou aux procédures papier, que l'on peut contourner, adapter, ou complémenter. Ainsi, « les acteurs sont contraints au respect des procédures, sous peine de ne pas pouvoir travailler » (Nicotri, 2004).

Les SI structurés sont donc basés sur la normalisation des processus, facteur d'économies d'échelle, qui produit une vision globale cohérente de l'organisation (Guyot, 2006, p.69). La coordination et l'uniformité des situations de travail recherchée est alignée sur le paramétrage technique du SI, lui même basé sur des modélisations. Les ajustements possibles pour les acteurs y sont donc très limités, l'interprétation des situations de travail étant de ce fait cadrée par le système. Ainsi,

> *« Les contraintes et orientations intégrées au SIO [Système d'Information Organisationnel] modifient la signification de certaines situations de travail dans lesquelles, pour agir en conformité avec ce que l'organisation attend d'eux, les utilisateurs doivent relever de nouvelles pertinences, modifier les modes de compréhension et orienter différemment leurs actions. » (Nicotri, 2004).*

Enfin, les SI structurés se caractérisent aussi par une forme évaluative unique et intégrée au SI.

Les limites évoquées dans la littérature sont constituées de critiques quant aux coûts de mise en œuvre de ces systèmes, aux problèmes de maintenance difficile (Guyot, 2006), et aux contraintes qu'ils font peser sur les acteurs en termes organisationnels et en termes d'adaptation au système. En effet, la logique descendante du déploiement de ces systèmes les rend coercitifs et les logiques de généralisation des situations d'activités pour leur uniformité a tendance à effacer les spécificités propres aux métiers.

Enfin, ils véhiculent les représentations imposées de l'organisation, des activités, des métiers, qui sont recomposées, en fonction du système technique. Ce qui va jusqu'à impacter le jugement que les acteurs portent sur leur travail, (Nicotri, 2004).

Plus globalement, nous pouvons dire que ces systèmes fonctionnent d'après une logique qui passe par l'augmentation de la contrainte pesant sur les utilisateurs, pour une réduction de l'incertitude quant à la gestion de l'organisation. Nous pensons, malgré les avantages qu'ils présentent pour des objectifs de visibilité gestionnaire, que ces systèmes basés sur des situations standardisées, convoquent une conception figée de l'activité, qui tend à réduire les capacités d'adaptation de l'organisation à son environnement concurrentiel.

Du point de vue de l'activité, il s'agit de considérer ces systèmes comme des équipements cognitifs qui se rapprochent de la conception de la cognition distribuée, où l'humain et l'artefact forment un système et où ces derniers sont mis sur le même plan. Nous affirmons donc que ces systèmes sont peu adaptés à une conception de l'activité en développement, comme présentée au début de ce chapitre, et qu'ils peuvent générer des situations de travail pathologiques, au sens de la psychologie du travail. Les marges de manœuvres des opérateurs sont quasiment absentes de ces dispositifs, le système d'évaluation restant quantitatif, alors que l'activité de gestion de l'information devrait faire l'objet de développements de nouveaux indicateurs qualitatifs. (Guyot, 2002).

Enfin, d'un point de vue plus politique, (Durampart, 2004) pose la question de savoir si ces Systèmes d'Information Organisationnels (SIO) « ne sont pas un moyen de transformer le statut et la définition du rôle de l'acteur (mais aussi sa propre représentation du rôle qui lui est attribué) dans l'entreprise sans passer par les usages habituels de la négociation ».

3.5.10.3 Les systèmes d'information semi-structurés et 2.0

Nous présentons ici les points forts des Si semi-structurés et 2.0, qui peuvent se rapprocher de ce qui est appelé « Groupware » dans la littérature.

D'après (Guyot, 2002), dans ces types de SI « l'acteur est fortement impliqué : il classe et organise sa propre information, tout en considérant ce qui peut profiter à ses collègues ; il est tour à tour, intervenant, producteur, et parfois même prestataire, s'il administre le système ».

Nous avons évoqué précédemment les limites des approches ascendantes (*bottom-up*) utilisées dans ces types de SI, qui sont basées sur la participation et l'enrichissement des métadonnées par les utilisateurs, de manière émergente. Nous rappelons ici certaines des caractéristiques de ces SI, énoncées en début de ce chapitre.

A l'inverse des SI structurés, les Si semi-structurés et 2.0 ne reposent pas sur une formalisation des processus et activités par la standardisation, leur socle commun étant d'être appuyés sur la participation des utilisateurs, de favoriser leur autonomie, d'être orientés vers le partage ou la construction collective des connaissances par les acteurs qui sont généralement distribués dans des lieux et temps distants. Ils peuvent contribuer à la coopération, la collaboration, la diffusion et la promotion de contenus et d'éléments de profils personnels. Les SI 2.0 intègrent un espace personnel et un espace partagé, entre lesquels les transferts sont facilités.

Dans ces types de SI, comme pour les groupwares, l'individu organise son usage en l'insérant dans d'autres activités. Le raisonnement est basé sur l'appropriation de l'outil par l'utilisateur. Ainsi, le propre du groupware est de caler ses modes de fonctionnement sur un système d'information personnel dont il démultiplie et amplifie les caractéristiques: il est un support et une aide au travail, c'est aussi un outil de suivi de l'activité. Enfin, il procure un regard sur l'activité des autres. (Guyot, 2002). Ces systèmes sont donc plus propices à s'adapter à l'usage de l'utilisateur, et ils n'entraînent pas de modification de la structure organisationnelle.

Deux logiques différentes peuvent être à l'œuvre dans ces types de SI, d'après (Guyot, 2002):

D'une part, on peut procéder selon une démarche qui vise à appuyer le système collectif sur la spécificité des systèmes personnels dans un mécanisme de montée en généralité à partir des comportements réels de chacun, et de sa capacité à partager son propre système. C'est dans cette démarche que se situe le prototype présenté précédemment. Ce processus est basé sur la

recherche d'un consensus entre des intérêts personnels qui peuvent devenir un bien commun. Il s'agit d'un processus d'explicitation progressif, long et itératif.

A l'inverse, l'autre démarche procède d'une vision politique et globalisante du bien commun comme la somme des intérêts particuliers qui convergeraient naturellement vers l'intérêt de l'organisation, dans une logique d'accumulation. Il n'y a pas de reconnaissance de l'existence de la valeur des systèmes personnels hétérogènes dans cette optique.

Nous pensons que la première démarche est plus favorable à l'adhésion au système proposé et à la recherche d'un consensus riche, c'est-à-dire porteur d'échanges qui peuvent mener les participants produire du savoir à travers la recherche de ce consensus. Cette négociation qui peut avoir lieu sur les termes utilisés dans le Systèmes d'Information 2.0, constituant des métadonnées, fait aussi vivre et exister les communautés qui les utilisent, leur donne une visibilité, un lieu et un support d'échanges, tout en en gardant une trace pérenne.

Ces communautés doivent selon nous être comprises comme des communautés métiers, qui à travers la participation au système, peuvent contribuer au « cercle vertueux » de l'exercice d'un métier (Clot, 208), c'est-à-dire à l'enrichissement d'une part des prescriptions, mais aussi aux interactions entre les membres du collectif, et à l'apprentissage du métier pour les novices et pour eux-mêmes.

Les Si semi-structurés et 2.0, qui permettent un forme de marge de manœuvre, d'appropriation et d'usage personnel du système pour les utilisateurs, nous paraissent mieux correspondre à l'approche théorique que nous avons choisie pour l'analyse de l'activité de gestion de l'information et aux approches qui permettent aux points de vue différents existants au sein d'une organisation, de s'exprimer et de produire des échanges riches.

3.6 Conclusion du chapitre 3 :

A travers l'intérêt que nous avons porté à l'activité de gestion de l'information, nous avons pu relever des tensions qui fondent l'intérêt de cet objet d'étude.

D'une part, nous avons mentionné la tension qui se noue au niveau des acteurs individuels entre leurs activités métier et leurs activités de gestion de l'information. Ainsi, ce qui semble se jouer dans cette analyse de l'activité de gestion de l'information la dépasse largement. Il s'agit, dans le choix des systèmes d'information, et d'après la théorie de l'activité et des approches issues de la psychologie du travail, d'être en mesure d'éclairer des choix plus

fondamentaux sur la manière dont les acteurs vont exercer leur métier, l'activité de gestion de l'information étant alors vue comme un moyen pour la réalisation des activités métier.

Ainsi, dans une perspective d'utilisation d'un SI structuré, il apparaît qu'il faut rester vigilant sur la standardisation des activités, qui réduit d'autant les marges de manœuvre des opérateurs, marges de manœuvres qui constituent pourtant à la fois un levier d'action pour la réussite des activités, puisqu'elles sont un moyen de remplir ses missions en fonction des contextes toujours renouvelés, ainsi qu'un moyen de développement de l'activité et du sujet. Ces derniers sont vus comme des conditions de la santé au travail, et comme une garantie de l'intérêt, pour celui qui réalise ces tâches, à continuer de les effectuer le mieux possible.

Dans ce cadre, cette tension entre activité métier et activité de gestion de l'information se déplace et peut devenir une tension entre l'organisation et le métier. En effet, si le système d'information ne donne pas aux acteurs les moyens de réaliser leur activité métier, il s'en suit que le système d'information devient un obstacle au bon fonctionnement de l'organisation.

Dans les dimensions partagées du SI, nous avons pu relever cette tension survenant entre les espaces individuels et collectifs, qui se matérialise dans des conflits entre systèmes d'information personnels et systèmes d'information collectifs et/ou structurés. Ici, nous avons pu nous intéresser à des approches participatives des processus de documentarisation, qui peuvent constituer des supports et des occasions d'échanges de pratiques métier et de débats, par exemple, sur le sens d'un terme ou le sujet d'un document, pratiques qui rendent l'exercice d'un métier plus vivant et dynamique et permettent le développement du métier.

Enfin, à travers la Théorie de l'Activité et les approches issues de la psychologie du travail, nous avons tenté de proposer un cadre théorique d'analyse de l'activité de gestion de l'information.

Chapitre 4 : Expérimentation

4 Chapitre 4 : Expérimentation

4.1 Introduction

Dans ce chapitre, nous présentons l'expérimentation que nous avons menée afin d'évaluer plusieurs objectifs relatifs au projet Miipa-Doc. Son déroulement s'est distribué en deux phases, dont la première contient deux volets. Les objectifs premiers ont quelque peu évolué au cours de cette expérimentation, comme nous le précisons dans cette présentation. Nous formulons des hypothèses dont la validité est évaluée. L'expérimentation s'est déroulée sur une durée totale de 26 mois, de manière discontinue.

4.2 Présentation des participants à l'expérimentation et de leurs activités

Les deux participants aux expérimentations sont des agents du même service, du même métier, dans la même organisation, il s'agit donc de l'unité minimale pour l'observation d'un groupe. Plus précisément, il s'agit de deux ingénieurs chercheurs du Groupe EDF, de la division R&D, du département STEP (Simulation et Traitement de l'information pour l'Exploitation des systèmes de Production) dans le groupe P1B et de la compétence IK (Ingénierie des Connaissances). Le groupe P1B est une équipe d'ingénieurs-chercheurs appartenant à l'un des quinze départements d'R&D-EDF. Ce groupe est composé de vingt-cinq ingénieurs-chercheurs, répartis sur trois compétences : traitement du signal et des images (TSI), systèmes dynamiques (SD) et ingénierie des connaissances (IK), (Musnik, 2012). Par ailleurs, ces participants sont des membres du projet Miipa-Doc (présenté en chapitre 1), EDF étant un partenaire industriel dans ce projet. La compétence centrale chez ces deux participants est l'Ingénierie des Connaissances.

Ces deux derniers éléments peuvent constituer des biais méthodologiques, mais un des avantages de cette configuration de l'expérimentation réside dans le fait qu'ils sont experts

sur les problématiques traitées dans le projet. On peut également noter d'autres avantages: ils sont membres du projet, ils sont donc intéressés pour participer de manière active aux expérimentations, l'accès pour les observations devrait donc s'en trouver facilité. Ils ont des niveaux d'expertise différents, l'un possède une expérience de 15 années dans son métier, tandis que l'autre bénéficie du titre d'Expert.

Globalement, leurs activités sont structurées en deux parties du point de vue organisationnel : il s'agit des deux grands types d'activités structurellement différentes.

- Un type d'activités qui s'intègre dans des projets, dont le cœur est la gestion des connaissances : Travail sur des outils, méthodes, modèles pour capitalisation, partage et gestion des connaissances. Ces activités donnent lieu à l'élaboration d'études, de recueils de connaissances, d'analyses de sujet, et à des propositions d'outils de gestion des connaissances.
- Un autre type d'activité qui est plutôt de l'ordre de la contribution dans des projets qui ne sont pas centrés sur la Gestion des connaissances (appui expert), dans une logique de prestation : Appui technique pour la préservation, le transfert, l'exploitation, la collaboration, l'émergence de connaissances. Ces activités se caractérisent par un accent mis sur la structuration des connaissances, et des contributions à l'analyse des besoins.

4.3 Présentation de l'expérimentation

L'expérimentation a connu deux phases successives durant lesquelles les objectifs, les hypothèses, et le positionnement des participants ainsi que de l'enquêteur ont évolué.
Dans la description de cette expérimentation, il faut commencer par mentionner qu'elle se fonde sur un présupposé qui est énoncé ci-après. Certaines des hypothèses présentées ne seront pas validées au cours de l'expérimentation.

Un présupposé : peu d'activité ne sont pas du tout documentées sur un poste de travail individuel, même les activités temporaires, transitoires, de courte durée, ou d'occurrence

unique laissent des traces sur les postes informatisés, individuels. Ainsi, on considèrera que toute activité produit des documents numériques et donne lieu à un classement.
Une première hypothèse apparaît suite à l'énoncé de ce présupposé.

Hypothèse 1 :
Si toute activité génère des documents qui sont classés, plus ou moins rigoureusement, alors, l'observation des documents et de leur classement peut constituer un moyen d'observer l'activité et de la comprendre. Ainsi, les arborescences classificatoires (qui servent aux acteurs à classer leurs documents) sont représentatives des activités.

4.3.1 Objectifs principaux de la première phase en deux volets

L'objectif du premier volet de la phase 1 consistait pour nous à construire une classification à facettes basée sur les activités métier pour les participants, en se basant sur la compréhension de leur activité, provenant des observations et des entretiens menés.

Pour un deuxième volet de cette première phase, nous avions ensuite comme objectif d'obtenir que les participants produisent eux-mêmes des classifications à des fins de comparaison entre notre proposition et les leurs pour évaluer la pertinence et les éléments communs ou divergents en fonction de la connaissance de l'activité. En d'autres termes, nous voulions évaluer la convergence des classifications proposées par l'intervenant, et par les participants, pour la validation, la critique et l'amélioration d'une méthode d'observation de l'activité et d'intervention méthodologique sous forme de prestation. Les participants étaient alors vus comme étant en capacité de juger de la pertinence de la classification à facettes.
L'intervenant était, lui, considéré comme légitime pour produire des classifications à facettes avec le prototype, pour d'autres acteurs, tout en étant extérieur au contexte d'activité. Ainsi cela constituait une deuxième hypothèse.
Il nous fallait aussi dans le cadre de ce deuxième volet de la phase 1, repérer dans quelle mesure les classifications à facettes se situaient dans une transposition des arborescences en classification à facettes.

Hypothèse 2 :

Un intervenant consultant est en mesure de produire des classifications pour des acteurs, qui travaillent dans des contextes d'activité qui sont extérieurs à l'activité de l'intervenant.

Dans le cadre de l'évaluation de la similarité entre les arborescences de départ et les classifications produites, il nous fallait tenter de répondre aux questions suivantes :

a) Les classifications à facettes reprennent-elles des éléments des arborescences de manière systématique ?

b) S'agit-il d'une re-conception totale de l'organisation de l'information concernant le classement du poste individuel ?

c) Comment la compréhension des principes de la classification à facettes impacte-t-elle la mise en œuvre de l'élaboration de classifications personnelles à facettes ?

Ceci conduit à formuler deux autres hypothèses qui sont les suivantes :

Hypothèse 3 :

Les agents peuvent reconsidérer en globalité leur système de classement.

Hypothèse 4 :

Les agents sont en mesure de s'emparer des principes de la classification à facettes implémentés dans le prototype.

Pour juger de cette 4ème hypothèse formulée, il nous faudra préciser des critères d'appréciation de la bonne compréhension des principes de la classification à facettes.

Résumé des hypothèses formulées

Hypothèse 1 :

Si toute activité génère des documents qui sont classés, plus ou moins rigoureusement, alors, l'observation des documents et de leur classement peut constituer un moyen d'observer l'activité et de la comprendre.

Hypothèse 2 :

Un intervenant consultant est en mesure de produire des classifications pour des acteurs, dont les contextes d'activité lui sont étrangers.

Hypothèse 3 :

Les agents peuvent reconsidérer en globalité leur système de classement.

Hypothèse 4 :

Les agents sont en mesure de s'emparer des principes de la classification à facettes implémentés dans le prototype.

4.3.2 Positionnement de l'enquêteur

Nous nous sommes positionnés, dans cette première phase, dans un rôle de consultant, dans une logique de prestation, parce que nous pensions pouvoir produire une classification à facettes pertinente, en nous basant sur la compréhension des activités des participants (Hypothèse 2).

Nous ne proposons pas de modèle théorique de facettes préexistant. La classification à facettes, dans ce cadre est élaborée en fonction des documents, des activités, des collectifs impliqués, des fonctions pour lesquelles on utilise l'outil contextuellement à chaque organisation.

4.3.3 Contraintes

La classification à facettes produite par l'intervenant, avec le prototype, ne doit pas contenir trop d'onglets, ni trop de facettes dans les onglets. De plus, le nombre de tags est limité lui

aussi. Si une facette contient trop de tags, alors elle ne constitue pas un critère suffisamment discriminant et son intégration dans la classification à facettes doit être remise en cause.
D'autre part, les problématiques liées, en général, aux IHM (Interface-Homme Machine) nous contraignent en termes d'espace dans l'écran. Ainsi, si l'on veut garder les avantages liés à la visualisation globale de la classification afin de favoriser la découverte et l'exploration des informations, ainsi que l'augmentation du sens du contrôle pour l'utilisateur, et enfin faire diminuer la désorientation (Lai, 2009), il faut limiter à un nombre raisonnable les onglets.
Ainsi, si l'utilisateur crée plus d'une dizaine d'onglets, l'usage de la classification à facettes implémentée dans l'outil devient inconfortable et perd ses avantages ergonomiques. En effet, l'espace occupé par l'IHM dans l'écran rend l'utilisation de la classification à facettes plus difficile au-delà de 10 onglets, en termes de visualisation et de navigation dans l'interface.

4.3.4 Les matériaux des observations et objets des entretiens

Dans cette optique, nous avons observé les arborescences de classement des postes de travail individuels des ingénieur-chercheurs, ainsi que les titres des dossiers et des documents qu'ils contiennent. La plupart du temps nous n'avons pas eu accès aux contenus des documents, sauf quand les participants les ouvraient, rapidement pour se les remémorer, mais sans avoir la possibilité de les consulter, pour des raisons de confidentialité.
De manière complémentaire, nous avons mené des entretiens pour obtenir des descriptions des activités réalisées par les participants, et nous avons conduit des observations sur les postes de travail pendant ces entretiens, pour nous faire expliquer les liens entre les activités décrites et les arborescences de dossiers et de documents.
Les arborescences de dossiers ont été observées seulement partiellement (nous revenons sur ce point ultérieurement) et nous n'avons pas obtenu de copies d'écran systématiques de ces matériaux, pour les mêmes raisons de confidentialité.
L'élaboration de la classification à facettes et de l'analyse de l'activité ont été réalisées sur la base des comptes-rendus des entretiens, et sur les commentaires sur ces données, par les participants à l'expérimentation.

4.3.5 Qualification théorique des sources

D'après la distinction entre travail réel et travail prescrit que nous avons présentée dans le chapitre 3 (partie 3.4.1), nous considérons que les descriptions des activités données par les participants pendant les entretiens correspondent, en termes théoriques, à des activités qui se rapprochent de l'aspect prescrit du travail. En effet, pour les participants, ces descriptions ont consisté à rendre accessible à un interlocuteur extérieur au milieu de travail, l'activité telle qu'elle peut être décrite comme dans des fiches métier normées ou des fiches de postes, sans détailler la mise en œuvre effective. Les observations ont permis de découvrir les aspects réels du travail, ignorés par les entretiens sur l'activité, tout en fournissant un support de discussion sur les activités effectives.

4.4 Déroulement du premier volet de la phase I de l'expérimentation

4.4.1 Les entretiens et observations

Nous avons réalisé deux entretiens collectifs, un au début du processus et un à la fin, ainsi que deux séries d'entretiens individuels. Les quatre entretiens individuels ont pris effet sur une durée d'un mois, leur durée a été limitée à 3h maximum par entretien. Il nous est apparu important de réaliser ces entretiens avec des intervalles de temps limités entre eux, pour accélérer le processus de récolte des données, et pour favoriser une meilleure mémorisation des détails.

Le premier entretien collectif a été une prise de contact. Il a servi à observer la structure globale de l'environnement informationnel, à commencer à déterminer le périmètre des observations dans les arborescences, et à récolter des informations génériques sur les activités des agents.

Le dernier entretien collectif a servi à évaluer la structure de la base de facettes proposée, ainsi qu'à affiner les valeurs proposées. Lors de ce dernier entretien, nous avons utilisé une application de cartographie afin d'obtenir une vue globale de la classification sous forme de carte des facettes.

Le premier entretien individuel était dédié à la description de son activité par l'agent et aussi à l'observation générale de l'arborescence (système de classement, intitulé des dossiers), tandis que le deuxième entretien a servi à une exploration systématique des documents présents dans

les arborescences selon le périmètre déterminé d'observation du classement. Ici, ce périmètre s'est borné aux projets en cours, nous n'avons pas observé les archives, ni les projets passés.
Entre chaque entretien, un compte-rendu a été réalisé à partir des notes prises pendant les entretiens pour servir de base à la suite du travail. En effet, une relecture était effectuée collectivement (l'agent concerné et le responsable de l'élaboration de la classification), pour apporter des précisions voire des corrections, au début de l'entretien suivant. Ces notes ont servi au travail d'analyse pour la création du schéma de classification et l'élaboration des valeurs de facettes. Lors du deuxième entretien, la description des activités en lien avec l'observation de la structuration et du libellé des dossiers et des titres des documents, a nécessité de la précision dans les activités décrites. Ceci étant dû à l'extériorité de l'intervenant au milieu de travail observé.

Quatre étapes dans le déroulement des entretiens :

1. Description de l'activité
2. Observation de la structure globale de l'arborescence
3. Identification du périmètre de l'observation
4. Observation et description systématique des dossiers et types de documents en lien avec l'activité à laquelle ils sont rattachés

Synthèse des objectifs des entretiens :

- Faire décrire son activité par l'agent
- Reformuler
- Lister les activités telles qu'elles sont évoquées
- Lister les types de documents évoqués au fil de l'eau en fonction des activités
- Étude visuelle et structurelle de l'arborescence

En effet, c'est l'activité qui constitue le premier point d'entrée dans les observations. Ceci est lié à plusieurs facteurs :

- les agents ont parlé de leurs activités spontanément

- les spécificités de leurs activités varient en fonction des projets
- les activités génèrent des documents spécifiques à une activité, mais aussi des types de documents redondants
- il n'y a pas d'activité qui ne soit pas documentée (présupposé évoqué précédemment).

4.4.2 Objectifs des observations des libellés des dossiers

Ce sont les libellés des dossiers qui nous renseignent sur le contenu des dossiers, le mode d'organisation de l'arborescence adopté, en même temps qu'ils nous servent de support pour récolter des informations sur l'activité.
Les dossiers sont observés dans l'ordre hiérarchique dans lequel ils sont conçus, on observe d'abord le dossier à la racine, les dossiers contenus dans ce dossier à la racine, et les sous-dossiers contenus dans ces dossiers, successivement. En effet, un dossier à la racine est plus susceptible de contenir dans son libellé des informations plus génériques et éventuellement sur les activités, que les sous-dossiers contenus dans ce dossier placé à la racine.

4.4.2.1 Données collectées durant le 1er volet de la phase I

Nous présentons ici les données que nous avons collectées pendant le premier volet de la première phase, à l'aide des observations et entretiens.

4.4.2.1.1 Observation des documents

- Observer et lister les différents types de documents, et éventuellement leur fréquence
- Faire décrire les types de documents par l'agent :
- Poser la question : « Dans ce dossier que trouve-t-on comme type de document ? »

Souvent l'agent se souvient des documents présents dans le dossier étudié en observant les titres des documents, mais parfois il a besoin de les ouvrir pour s'en souvenir.

4.4.2.1.2 Observations des arborescences

Les résultats concernant l'identification des activités des agents participants seront abordés plus loin dans ce chapitre. Concernant les observations sur les arborescences, on observe des divergences dans les modes de classement des documents et en termes de régularité du classement. En effet, chez l'agent A, les critères de classement sont plus récurrents et rigoureux, tandis que pour l'agent B, le classement est moins systématique. On remarque que le classement de l'agent A produit un type d'arborescence moins étendu que l'agent B, c'est-à-dire qu'il y a moins de dossiers à la racine du disque où sont stockés les documents chez l'agent A que chez l'agent B. L'arborescence de l'agent A sera dite « synthétique », celle de l'agent B sera dite « étendue ». Les définitions suivantes sont donc produites :

- Arborescence Synthétique : elle contient peu de dossiers à la racine, contient beaucoup de niveaux de profondeur hiérarchique (classement généralement plus systématique dans ce type d'arborescences)
- Arborescence Etendue : elle contient beaucoup de dossiers à la racine, mais moins de niveaux de profondeur hiérarchique (classement généralement plus évolutif en fonction d'éléments conjoncturels, des facteurs extérieurs viennent justifier la création des dossiers comme la commodité de l'accès, l'urgence d'un dossier, la priorité d'un travail, des contraintes techniques, ou des stratégies d'usages). On observe aussi une multiplication des types de libellés, dans ce type d'arborescences.

Un troisième type que nous avons repéré dans d'autres observations pour d'autres objectifs vient compléter cette typologie des arborescences proposée :

- Arborescence Sélective : la plupart des dossiers de premier niveau sont peu profonds et il y a quelques dossiers de premier niveau qui renferment une arborescence à part entière qui est de type étendue. Il s'agit d'un type d'arborescence dans lequel certaines zones seulement sont développées en termes de niveaux de profondeur.

Ceci nous conduit à une première reformulation méthodologique pour traiter le problème des observations dans les arborescences étendues, il est nécessaire de :

- Identifier le type d'arborescence que l'on étudie ;

- Déterminer le périmètre des observations. Cela revient à dire qu'il est nécessaire, puisque la durée des entretiens/ observations est limitée, de délimiter une zone à étudier représentative de l'arborescence.

Pour l'agent A, nous avons observé toute l'arborescence des projets en cours, et les archives partiellement. Pour l'agent B, nous n'avons exploré que quelques zones comme un dossier qui renferme des dossiers et des documents sur un gros projet, un dossier qui traite une affaire, et un dossier qui traite une action. Les projets, affaires, défis et actions sont, dans le cadre des activités des agents EDF, des types de projets avec des priorités et durées variables.

En fonction du type d'arborescence identifié, le périmètre à observer est donc le suivant :

- Si l'arborescence est synthétique :
    - déterminer les zones/ régions de l'arborescence qui servent à classer les documents de travail sur les projets en cours
    - il faut noter le type d'organisation : par date/ projet/ type de contribution
- Si l'arborescence est étendue :
    - observer globalement le nombre de niveaux et le type d'organisation : par priorité/ commodité/ urgence/ projet/en fonction de contraintes techniques
    - Déterminer la zone la plus actuellement utilisée dans l'arborescence en fonction des travaux en cours et se focaliser sur cette partie (il peut s'agir de quelques dossiers disséminés)
- Si l'arborescence est sélective :
    - trouver les parties de l'arborescence qui contiennent les sous-arborescences étendues imbriquées
    - Explorer ces parties

4.4.2.2 Observations et analyse du contenu des libellés des dossiers

Nous avons listé les activités et les types de documents, et collecté d'autres sortes de termes à cette étape. On remarquera que cette méthode peut être rapprochée de la méthode de

construction d'un thésaurus, qui commence par une collecte des termes issus de la littérature sélectionnée comme représentative du domaine du thésaurus. On a aussi pu remarquer que l'on retrouve toujours les mêmes types d'information dans le libellé des dossiers. On en propose ici un tableau récapitulatif :

Type d'information auquel ils correspondent	Exemple d'intitulés issus des observations
Projet (et date)	Actions 2007, TIREX,
Activité	Transfert + nom d'entité, formation, retour d'expérience, réalisation outil, recueil de données,
Thématique	Moteurs de recherche, multimédia
Entité	Comp, SEPTEN[3], …
Personne	Confidentiel
Type d'activité	Accompagnement, administratif, Appui expert, suivi, modélisation, quantification, publication, Principal
Outil	Nom d'outil, maquette carto, nom de BDD
Type de document	état de l'art, nom de norme, présentation, thèse, ITV, Livrables, contrat, fichiers VME,
Partenaire	Confidentiel
Evénement	Confidentiel
Lieu	Confidentiel
Etat du document/ cycle de vie du document	Archive, ancien
Type de contenu	Visuel
Phase du travail / lot / échéance, étape du projet	Travail, en cours, test, synthèse, analyse résultats, évolution, nom de lot,
Mode de réalisation/ modalité	Fil de l'eau
Destinataire	Nom de personnes

Tableau 5. Types d'information collectés

[3] division de l'organigramme

4.4.2.3 Observations et analyse de la forme des libellés des dossiers

Les libellés des dossiers peuvent être composés de combinaisons des éléments suivants :

- Abréviation / Code
- Noms de projets
- Échéance/ étape du projet
- Nom d'un lot
- Contenu d'un lot
- Composante de l'organisation
- Division organisationnelle
- Type de documents
- Activité
- Date (année)
- Thème de recherche
- Objet de l'étude
- Cycle de vie du document
- Application spécifique
- Tâche spécifique
- Types de fichiers
- Type de code (en programmation)
- Mode de recueil de données
- Nom de partenaire
- Nom de personne
- Pense-bête/ mémo
- Titre explicite
- Nom de Base de Données

Des exemples de libellés de ces types d'informations relevés sont donnés en annexe 1.

La pratique du code (abréviations, acronymes) pourrait se rapprocher de la pratique du tag :

- Le titre d'un dossier est significatif pour l'agent et quasiment pour lui seul
- Ils donnent peu d'information sur le contenu des dossiers

Ceci confirme l'intérêt des entretiens pour expliciter le sens des intitulés et connaitre le lien entre les libellés, le contenu des dossiers et les activités.

Les types les plus courants de combinaisons de ces éléments sont les suivants :

- Contenu + acteurs : exemple : REX sites
- Mode de réalisation + activité : exemple : recueil de données_ fil de l'eau
- Nom d'outil + fonctionnalité
- Type de doc+ activité : exemple : pré-étude_architecture

Liste des activités

Nous présentons ci-après un extrait de la liste des activités que nous avons relevées durant les entretiens et observations.

- Étude de contexte
- recueil de besoins
- Observations
- Interview
- structuration des connaissances et de données
- spécifications fonctionnelles pour des outils de gestion des connaissances
- Analyse
- modélisation
- recueil de connaissances
- études bibliographiques / états de l'art
- benchmarks sur les solutions
- synthèses
- création d'outils de fonctionnalités avancées
- Développement

- Suivi des développements
- méthodologie
- livraison aux clients
- accompagnement
- Suivi
- suivi de prestations
- recettes
- Restitution
- Elaboration de livrables
- Séminaires
- Publication
- Partenariats
- Préparation affaire

Figure 20. Extrait des activités listées

Liste des types de documents

Nous avons relevé plus de 110 types de documents différents qui constituent une typologie empirique. Nous en présentons un extrait dans la figure suivante.

- Documentation
- Cours
- Article
- Charte graphique
- Méthodologie
- Bibliographie thématique
- Bibliographie transverse
- Benchmark
- État de l'art
- Document administratif
- Entretien
- Étude
- Copie d'écran
- Rapport
- Présentation
- Bibliographie générale
- Contribution
- Entretien
- Formation
- Budget

- Bilan annuel
- Exports
- CDC
- Proposition commerciale
- Rapports de test
- Annexes
- Test algorithme
- Recette
- Extraction
- Notes
- Thèse
- Lettre
- Tutoriels
- Etude d'opportunité
- Note d'opportunité
- Etude de besoins
- Formulaire
- Fiches de visites
- Données d'entrée
- Norme
- Tableau de bord
- Contrat

- Scénario d'usage
- Analyse du besoin
- Spécification
- Suivi qualité
- Fiche
- CR
- Description de livrable
- Fichiers paramétrage
- Modèles de données
- Description de processus
- Note de synthèse
- document de travail
- Document intermédiaire
- Benchmark
- Pré-étude
- Document transverse
- Note de cadrage
- Maquettes
- Accord de consortium
- Tableau des objectifs
- Documents de pilotage

Figure 21. Extrait de la collecte des types de documents

Cette typologie constitue déjà une réduction de la complexité réelle des observations (Coenen-Huther, 2007), étant donné qu'il peut y avoir plusieurs occurrences d'un même type de document dans les arborescences, cette situation étant due au fait que les agents travaillent sur plusieurs projets en même temps avec parfois des rôles qui varient selon les projets.
Il n'est pas envisageable d'aboutir à une taxonomie de 110 valeurs pour une facette, avec nos objectifs de réduction des coûts cognitifs des opérations de gestion des documents, aussi nous avons tenté d'en faire une typologie théorique dans (Desfriches Doria, 2012a, 2012b).

Autres données collectées

Nous avons, au cours des entretiens et observations, collecté d'autres listes de données comme les thèmes abordés, dont nous présentons un extrait dans la figure suivante.

- Partenariat/ partenaires/ Contact
- Query suggestion
- Anglais
- Langages de programmation (Python, Java, Plone, Zop)
- Travail coopératif assisté par ordinateur (TCAO)
- Outil / Logiciel/ Application
- Thèse
- Stage
- Département STEP
- Bonnes pratiques
- Pérennisation supports
- Lien consommation/météo
- Offre de solutions internes
- IK
- Système qualité EDF

- Organisation client (Données d'entrée/ reçues)
- Remontée des informations
- Diffusion sélective
- Mécanismes
- Mode débug
- Pérennité entreprise
- Vieillissement
- Matériaux
- Cryptage
- Déplacement
- Disponibilité
- Personnel
- Concertation inter-domaine
- Contexte
- Compétence SI
- Sujet (de stage/ de thèse)
- Rédaction/diffusion des FAV

- Rex
- organisation des AT
- Activité humaines
- Pilotage AT
- Reconnaissance vocale
- BDK
- Architecture informatique industrielle
- Fonctionnalités
- Prototypes
- Algorithmes
- Editeur
- Déploiement
- Industrialisation
- Composants Progiciel
- Guide de diagnostic

Figure 22. Extrait des thèmes relevés

Nous avons relevé d'autres termes au cours des entretiens, observations et dans nos comptes-rendus, qui se sont révélés comme des facettes potentielles. Nous en donnons ici un aperçu, dans la figure suivante.

- Données d'entrées/ sortie
- Date (année)
- Versions : Final/ doc de travail/ V. Intermédiaire/ Validation/ Archive/ Ancien
- Source (pour les Documents reçus) : R&D, IRSN, Client interne, LDAP, CEA, UNIE, SEPTEN, entité de production, entité d'ingénierie, INPO, Pilotage des AT, ergonomes, SDIN, PSRP, DSI
- Stade / étape du projet : avancement/ ...
- État du document (en travail, intermédiaire, pré-diffusion/ diffusion/ Officiel)
- Doc partagés
- Cycle de l'activité (analyse de besoins/ développement de modèles/ spécification/ exploitation)
- En fonction des réunions (réunion de suivi/ d'avancement/ revue de conception des livrables/ de synthèse/ de bilan...)
- En fonction du cycle du projet : conception/ implémentation/ maquettage/ évaluation / raffinage
- Nom du projet

Figure 23. Extrait des autres termes collectés

A partir de ces données, nous sommes entrés dans une démarche itérative d'élaboration de classification à facettes, contrainte par les caractéristiques de l'outil. Ainsi, il nous a fallu déterminer les tags, les facettes et la structuration des regroupements de facettes.

4.4.2.4 Elaboration itérative de la classification à facettes

Nous avons procédé successivement à l'élaboration des facettes et des valeurs de facettes, puis à la structuration de la classification en onglets. Nous ne présentons ici que la dernière version de notre proposition de classification, qui est la plus aboutie, et que nous avons soumise à l'appréciation des participants.

Description synthétique de l'activité des participants

L'encadré suivant présente une synthèse des différentes activités des participants.

L'activité des agents se découpe comme suit :

1. **Préparation du futur**
2. **Veille**
3. **Animation de la compétence**
   - Savoirs faires
   - Partenaires
   - Formations
   - Conservation de la compétence pour permettre de monter en compétence
4. **Réalisation des activités :**
   - Les aspects au cœur de cette activité :
     - Production de livrables (+ ou – détaillés)
     - Conseils / avis
     - Expertise (on les consulte)

L'expertise implique :

- Collecte d'information
- Données détaillées pour :
  - Réutilisation
  - Analyse spécifique
  - Appui Ingénierie des connaissances

Plus précisément cette activité d'expertise peut se décliner comme suit :

- Elaboration de scénarios/ spécifications
- Etude du rôle des acteurs (avec l'équipe Facteur Humain)
- Contribution à l'analyse des besoins (L'analyse des besoins justifie les dispositifs mis en place)
- Ingénierie des besoins
- Modélisation de structure
- Elaboration de Cartographie pour représenter des éléments
- Elaboration de modèles de processus

Figure 24. Synthèse des activités des participants

4.4.2.4.1 Construction des facettes et tags

Nous avons déterminé les facettes et les tags, en fonction des éléments listés ci-dessous :

- les besoins exprimés dans les entretiens
- la relecture des comptes-rendus
- le relevé systématique des types de documents mentionnés
- le relevé des types d'activités

- le relevé des autres types d'information mentionnés, et le regroupement en informations de même nature
- les recoupements entre les entretiens
- le relevé des redondances et caractéristiques du travail communes

A partir d'une synthèse des informations évoquées pendant les entretiens et récoltées pendant les observations, nous avons élaboré un tableau avec les listes de valeurs potentielles que l'on catégorise au fur et à mesure dans des facettes potentielles. Pour ce faire, il faut identifier quelle est la propriété décrite par chaque valeur pour la placer dans une catégorie dont le titre décrit le type de propriété. Ainsi dans une classification, et selon différents points de vue, un tag qui serait intitulé « Couteau » pourrait être classé dans plusieurs catégories, comme les suivantes : outil, instrument, arme, objet de valeur, objet artisanal. Un tag peut donc appartenir à plusieurs catégories de propriétés, comme un mot est polysémique. Son sens peut être appréhendé selon différents points de vue. La détermination des facettes est réalisée en fonction des objectifs de la classification et en fonction du domaine traité dans lequel elle va être utilisée. Il est nécessaire d'expliciter ces éléments contextuels pour réaliser une classification dans laquelle les facettes et les tags sont pertinents.
Les objectifs sont donc constitués de l'activité de gestion documentaire quotidienne, individuelle, dans le cadre du métier d'Ingénieur-chercheur, expert en Gestion des connaissances, dans le domaine de l'énergie, chez EDF. Dans cette démarche, nous avons visé l'exhaustivité dans le relevé des termes, pour en faire des valeurs de facettes potentielles, qu'il est possible de réduire lorsqu'on a obtenu suffisamment de matière. La liste des facettes et leurs valeurs dégagées à ce stade sont disponibles en Annexe 2.

4.4.2.4.2 La structuration en onglets de la classification à facettes

A ce stade, nous avons obtenu un jeu de facettes potentielles, et leurs valeurs sont listées, sans être nécessairement exhaustives. Elles peuvent être complétées par la suite, selon un processus itératif. D'après la présentation du prototype en chapitre 3 (partie 3.5.9), les facettes doivent être rassemblées en groupes cohérents en fonction d'une logique, qui peut dépendre des activités du contexte ou d'autres éléments structurants, qui peuvent aussi relever de scénarii d'utilisation. Ainsi pendant cette phase d'expérimentation, nous avons déterminé des onglets en fonction :

- des besoins exprimés lors des entretiens
- des activités des participants
- des propositions de facettes trouvées et de leur quantité

Il s'agit de regroupements de facettes pouvant avoir des caractéristiques communes. A cette étape, les éléments communs que nous avons identifiés peuvent être :

- Les activités auxquelles les documents se rapportent
- Le type de collectif investi dans la création ou l'utilisation des documents : niveau de l'organisation/ niveau métier/ groupe projet/ individuel
- Les usages
- La structuration des activités
- La structure organisationnelle de l'entreprise
- Les phases d'un projet
- Les scénarii d'utilisation

Cette conception de la structuration en onglets nous mène à formuler une 5ème hypothèse qui est la suivante :

Hypothèse 5 :

La structuration de la classification à facettes en onglets (ensembles de facettes), implémentée dans le prototype, peut se faire à l'aide d'éléments hétérogènes. Les onglets ne reflètent pas nécessairement seulement les activités, mais peuvent être construits sur d'autres éléments du contexte. Ainsi on pourra trouver par exemple, un onglet « Projet », un onglet « Document », et un onglet « Métier ». (voir Annexe 2)

Dans cette optique, les projets relèvent de la structuration des activités, un onglet « Document » concernerait des aspects purement liés au document, et contiendrait donc des facettes comme « Type de document », « Opération sur les documents », « Version du document » et « Accessibilité ». On trouvera en Annexe 2 la liste des valeurs de chacune des facettes de cet onglet, et des autres onglets évoqués ici.

Enfin, un onglet « Métier » pourrait contenir des facettes qui concerneraient l'activité, les objectifs, les outils, ou les thèmes, d'après les relevés de termes effectués.

Cette hypothèse ne sera évaluée qu'à l'issue de l'expérimentation, tandis que nous avons mis en œuvre dans les résultats, une structuration hétérogène des onglets dans la classification que nous avons proposée.

4.4.3 Résultats

Nous avons donc listé les activités (27 valeurs au total), les types de documents (plus de 110 types au total), les thématiques (entre 30 et 50 valeurs par agent), et les autres types d'informations qui pouvaient potentiellement devenir des facettes que nous avons présentés précédemment.

La facette Type de document qui potentiellement pouvait contenir plus de 110 valeurs nous a contraints à dégager des types de documents génériques et d'autres plus spécifiques à certaines activités. Cette abondance de types de documents a été une contrainte forte dans l'élaboration de cette maquette.

Nous présentons, dans les figures qui suivent, une maquette de la version finale de la classification à facettes que nous avons proposée aux agents, et ses principes de structuration.

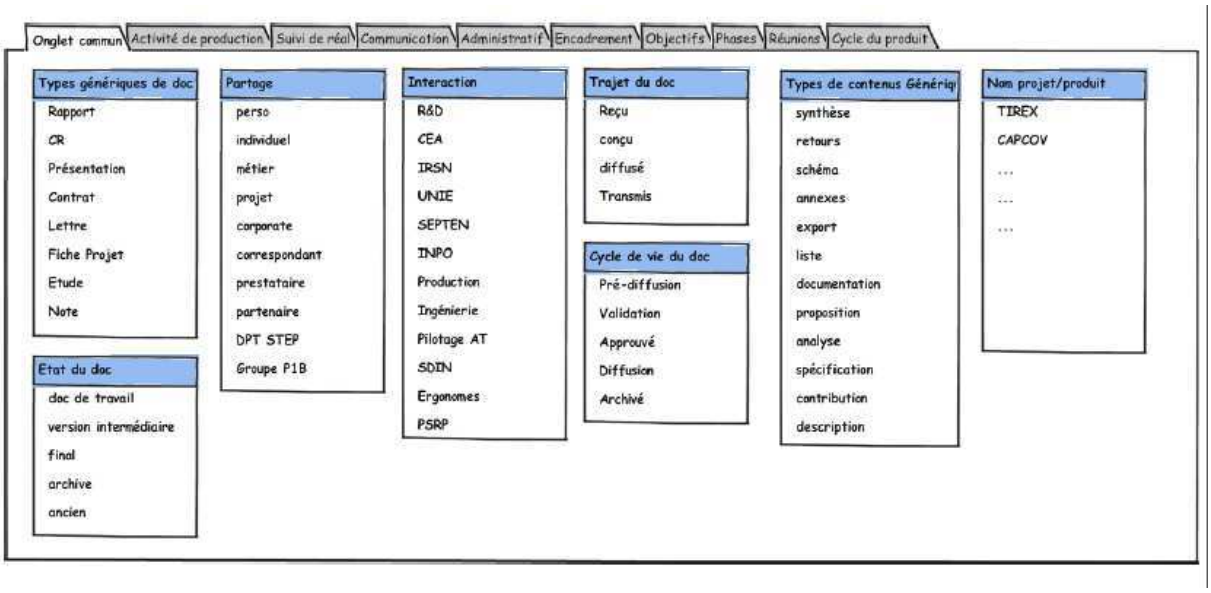


Figure 25. Premier onglet de la proposition de classification à facettes

Le premier onglet est un onglet commun où l'on trouve des facettes avec des valeurs génériques, valables pour de nombreux cas d'indexation de documents.

- Facette Type de document générique : permet d'indiquer un type courant de document
- Facette Partage : indique le degré de partage du document
- Facette Interaction : indique à qui ou quelle entité est lié le document
- Facette Trajet du document : indique son statut en termes de réalisation

- Facette Cycle de vie du document : indique son degré de validation
- Facette Type de contenus générique : qualifie le type de contenu
- Facette Nom du projet : indique le projet dans lequel le document est actif
- Facette Etat du document : indique le degré d'achèvement du document

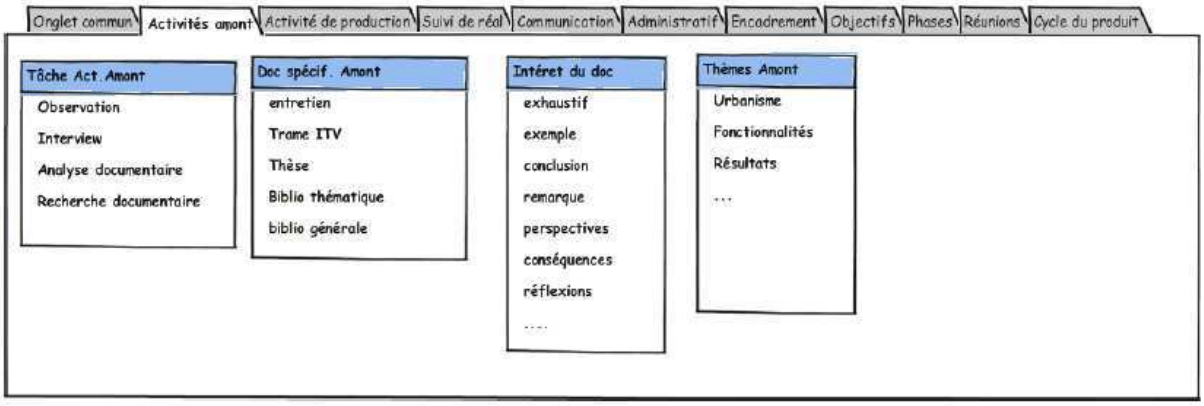


Figure 26. Deuxième onglet de la classification à facettes

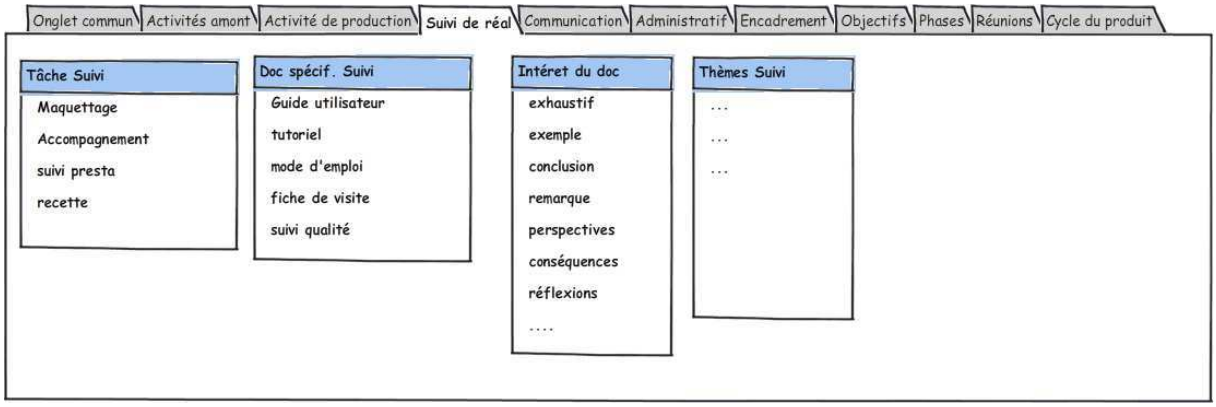


Figure 27. Quatrième onglet de la classification à facettes

Dans les trois figures présentées ci-dessus, on peut compter 11 onglets (contextes) qui sont issus soit de l'analyse des activités, en termes de décomposition des activités, soit d'autres éléments d'organisation de l'activité comme les 4 derniers onglets. La structuration de la classification à facettes est donc réalisée avec des onglets qui sont hétérogènes.

Cette structuration de la classification à facettes détermine aussi un scénario d'utilisation. Ainsi lors de l'indexation d'un document avec Hypertagging, l'utilisateur est contraint de renseigner un minimum d'information à propos du document en utilisant le premier onglet intitulé « Onglet Commun ». Si l'utilisateur le juge opportun, il peut utiliser les autres onglets pour renseigner d'autres métadonnées à propos du document. Dans ce cas, il doit passer sur un autre onglet pour compléter le renseignement des métadonnées. L'ensemble des onglets qui composent cette proposition de classification à facettes est présenté en annexe 3.

Les autres onglets sont plus spécifiques et présentent la possibilité d'indexer un document en fonction des spécificités des activités. Leur intitulé correspond, pour les 5 onglets qui sont positionnés après l'onglet commun, à l'activité en question, et sont listés ci-après :

- Activité de production : cet onglet correspond à des tâches de production de documents comme des états de l'art, des benchmarks, ou des recueils de connaissances ;

- Suivi de réalisation : cet onglet correspond à des tâches de suivi de l'élaboration d'un outil informatique, plus précisément, il concerne les tâches de maquettage, de suivi de la prestation, d'accompagnement, de recette ;

- Communication : cet onglet correspond à des activités liées à des présentations, des séminaires ou publications ;

- Administratif : cet onglet correspond à des tâches de gestion administrative, liées aux différentes entités dans lequel les agents sont intégrés, par exemple, la Compétence, le Groupe, ou le Département ;

- Encadrement : cet onglet correspond à des tâches d'encadrement de stagiaires ou doctorants

Les 4 derniers onglets correspondent à des découpages de l'activité selon d'autres critères, et en fonction d'autres contraintes. Ils sont intitulés :

- Objectifs : cet onglet correspond à une caractérisation des activités en fonction des objectifs de celle-ci ; ce peut être, par exemple, la préparation d'une affaire, la modélisation ou la restitution ;

- Phases : cet onglet correspond à un découpage en fonction des phases standard des projets, par exemple : étude de besoin, conception, expérimentation, validation ;

- Réunions : il existe des types de réunions, qui produisent des types de compte-rendu précis et sont organisées systématiquement dans la mise en œuvre des projets comme les réunions de suivi, les réunions techniques, les réunions de conception de livrables, de synthèse, ou de bilan.

- Cycle du produit : un dernier repère que nous avons relevé dans l'activité des Ingénieurs-chercheurs est le cycle du produit, quand l'activité consiste à élaborer et livrer un produit, nous avons donc créé un onglet, pour ne pas changer les repères réguliers des utilisateurs, il s'agit donc d'un onglet créé en fonction de l'usage. Il concerne les différentes étapes de l'élaboration du produit, à savoir, la préparation, la spécification, la conception, le développement, le maquettage, l'évaluation et le raffinage.

Dans ces 10 onglets spécifiques, on trouve toujours les mêmes facettes proposée avec des valeurs contextuelles pour chaque facette, déterminées en fonction du découpage de l'activité réalisé dans l'onglet:

- Une facette Tâche : permet de raffiner les tâches qui apparaissent dans la réalisation d'une activité. Cette facette est contextualisée.

- Une Facette Type de document spécifique à cette activité : permet de préciser le type spécifique du document lorsque les types génériques sont trop larges, et que le document est en lien avec une activité spécifique. Cette facette est contextualisée.

- Une facette Intérêt du document : permet à l'utilisateur de préciser ce qui l'intéresse dans un document. Cette facette est une facette dupliquée.

- Une facette Thème spécifique à l'activité : permet d'associer un thème au document. Cette facette est une facette dupliquée.

Cette organisation et structuration de la classification à facette permettait de prendre en compte l'aspect multiforme de l'activité, tout en conservant une certaine simplicité puisque dans chaque onglet autre que l'onglet commun, on retrouve les mêmes types de facettes contenant pour certaines des valeurs contextualisées. Les facettes contextualisées reprennent ici le nom de l'onglet dans leur titre pour permettre d'en faire des facettes différentes. En effet, deux facettes qui possèdent le même titre sont identiques et doivent posséder également les mêmes valeurs.

En termes de modèle, ce ne sont pas les mêmes facettes, mais des facettes traitant le même type de propriété. Quand deux facettes sont identiques, elles possèdent le même titre, et les mêmes valeurs, elles sont donc dans ce cas, dupliquées. Pour obtenir deux facettes différentes, il est nécessaire de produire un titre de facette différent, et des valeurs non identiques. Pour autant, nous considérons que deux tags identiques dans deux facettes différentes ne sont pas des tags identiques, le sens du tag étant déterminé par la spécification apportée par le titre de la facette qui précise le type de propriété contenu dans ses valeurs (dans les valeurs de cette facette).

Ainsi, dans ce scénario, le contexte commun est valable pour tous les membres de l'organisation, il correspond à l'effort minimal d'indexation demandé aux agents. Les contextes spécifiques aux activités présentent plutôt des facettes métiers et proposent à une communauté de pratiques, ou au sein d'un service de l'organisation, des dimensions d'indexation facultatives, relevant d'un effort plus conséquent d'indexation, mais permettant de la rendre plus précise.

On rassemble dans le tableau suivant les caractéristiques de cette fonction des onglets.

Onglet	Commun	Spécifique
Type de facettes	Facettes communes	Facettes facultatives
Type d'indexation	Standard	+ précise + spécifique + personnelle
Modalités organisationnelles	Transverses	Autres modalités organisationnelles : activité/ temporelle…
Degré d'investissement demandé aux utilisateurs	Effort minimum	Effort + conséquent
Degré de contrainte	Formel, obligatoire pour tous les documents	Facultatif en fonction des besoins
Exemples de facettes	Type de document générique, Lot, Année, Entité concernée, Etat du document, Nom du projet, Cycle de vie du document	Type de document spécifique, Type de contenu spécifique, Valeur de l'info, Intérêt du document, Période concernée, Thématique

Tableau 5. Caractéristiques et fonctions des onglets

4.4.3.1 Evaluation par les participants de la classification à facettes proposée

La structuration basée en partie sur l'activité avait été retenue étant donné le présupposé énoncé au début de ce chapitre. Ainsi, si toute activité produit des ressources documentaires, les documents et leur classement se situent dans un lien intime avec l'activité, au sens où ils reflètent de manière rapprochée l'activité réelle. Les autres éléments qui structurent l'activité ne sont pas endogènes, comme les arborescences et les documents, ils sont produits en dehors de l'activité. Ils sont des moyens de structurer globalement l'organisation, comme l'organisation par services, ou en phases de projets, mais ils ne correspondent pas à l'organisation individuelle des documents dans les arborescences. Ainsi, la proposition de classification à facettes semble se rapprocher du modèle prescrit de l'activité et des éléments exogènes à l'individu dans l'organisation du travail. Ceci est lié en partie, selon nous, à une forte tendance chez les participants à parler de leurs activités, en référence au travail prescrit, mais également au temps limité des observations des arborescences et des documents, tandis

que très peu de copies d'écran ont été fournies. Nous avons donc intégré dans la classification proposée, une forte dimension liée au cadrage organisationnel, correspondant à l'organisation des activités telle qu'elle est réalisée par les concepteurs de leurs activités.

Par ailleurs, il a été nécessaire, au cours des entretiens, de revenir sur la définition des activités des acteurs, à chaque entrevue. Il est apparu que nous avons finalement rencontré des difficultés à cerner l'activité et le mode de fonctionnement de l'organisation dans laquelle ils l'exercent, ceci étant lié à la complexité de la structuration de l'organisation et des activités ; ainsi qu'au fait nous avons eu un accès très réduit au terrain pour observer. De plus, nous avons obtenu des explications contradictoires sur leurs activités pendant les entretiens collectifs avec les participants. Enfin, entre les activités programmées et non programmées (en % de temps de travail), les projets, sous-projets, les programmes, les actions, les défis, les contributions dans les projets, et la structuration en poupées russes des projets, il est apparu assez ardu d'avoir une vision claire des activités exercées par les participants à l'expérimentation.

Dans d'autres réunions, il nous a été précisé la nature du fonctionnement des contributions des agents de la compétence IK dans les projets. Nous la schématisons dans le graphique suivant à l'aide d'un exemple :

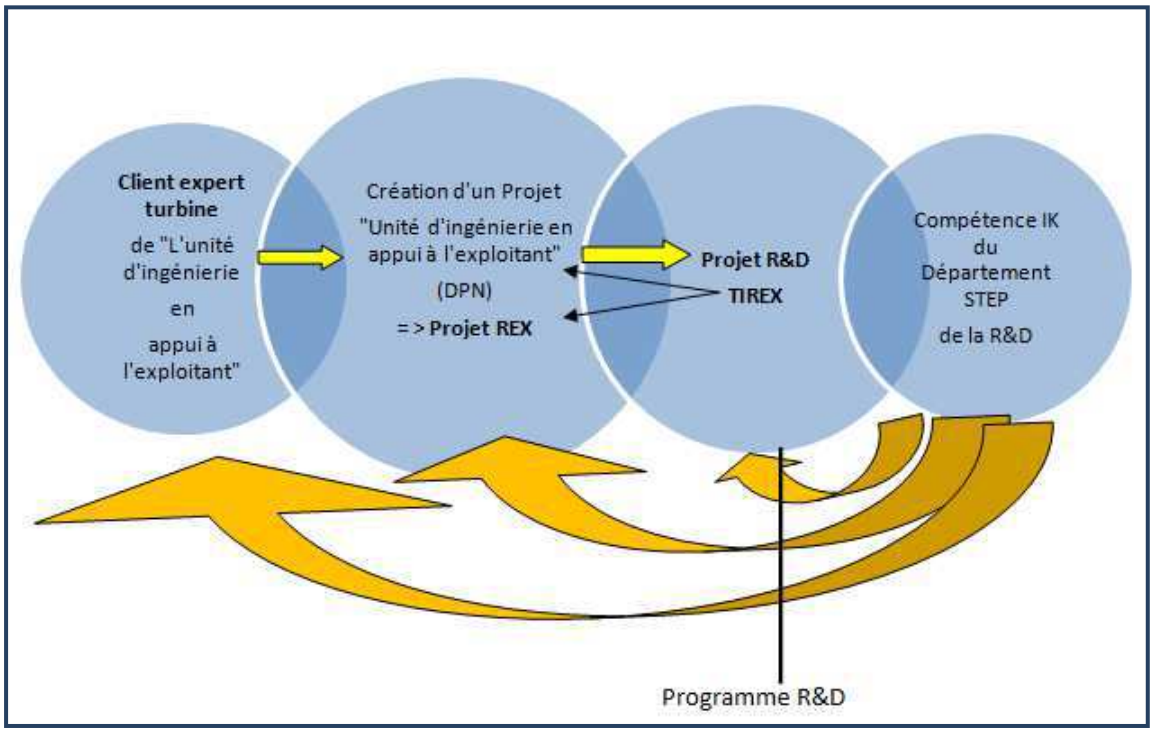


Figure 28. Fonctionnement des contributions des agents de la R&D dans les projets

Ce schéma reflète un exemple d'initiation d'une contribution des agents la compétence IK de la R&D d'EDF, à une autre entité d'EDF. Chronologiquement, il est à lire de la gauche vers la droite : le Client expert Turbine de « L'Unité d'Ingénierie en Appui à l'exploitant » est à l'origine d'un projet parce qu'il a un besoin, ce qui engendre la création d'un projet au sein de « L'Unité d'Ingénierie en Appui à l'exploitant » qui fait partie de la DPN (Division de la Production Nucléaire), un projet qui s'intitule « REX ». La R&D créée un projet TIREX pour répondre à son niveau au besoin du projet REX de « L'Unité d'Ingénierie en Appui à l'exploitant »; et la Compétence IK du département STEP de la R&D fait des contributions principalement au niveau du Projet TIREX mais aussi aux autres niveaux éventuellement. Les flèches orange représentent les contributions des agents de la compétence IK aux différents niveaux d'un tel projet. Ces contributions se font sous forme d'interaction. Les flèches jaunes reflètent l'ordre chronologique dans lequel le projet est initié depuis la demande vers les différents niveaux organisationnels qui interviennent. Les flèches noires indiquent les contributions au niveau des départements de la R&D vers les acteurs métiers. A travers l'observation de ce schéma on peut constater la complexité des niveaux de structuration des projets et des interactions entre les différents acteurs intervenant dans les projets.

Les participants ont jugé que la structuration de la classification, telle que présentée précédemment, pourrait être améliorée, en termes d'utilisabilité. Le scénario d'utilisation n'a pas été retenu comme pertinent à ce stade. Ainsi les hypothèses 1 et 2, qui concernent le premier volet de la phase 1 de l'expérimentation, n'ont pas été validées de façon complète.

Nous rappelons ici ces hypothèses :

Hypothèse 1 :

Si toute activité génère des documents qui sont classés, plus ou moins rigoureusement, alors, l'observation des documents et de leur classement peut constituer un moyen d'observer l'activité et de la comprendre.

Ainsi, nous avons obtenu une vision peu claire de l'organisation des activités et des activités elles-mêmes à travers les observations. De nombreuses explications sont nécessaires pour les complémenter, ce qui justifie les entretiens, qui nous paraissent indispensables pour éclairer la compréhension partielle des activités fournie par les observations. Ce matériau est jugé trop brut pour être traité de manière autonome par l'intervenant. Cependant, il est quand même possible à partir de ces matériaux (observations et entretiens) de produire une proposition de

classification à facettes en partie basée sur les activités, en guise d'amorçage du processus d'élaboration de la classification à facettes. Cette hypothèse est donc validée seulement partiellement.

Hypothèse 2 :

Un intervenant consultant est en mesure de produire des classifications pertinentes pour des acteurs, dont les contextes d'activité lui sont étrangers.

Cette hypothèse n'est pas validée, puisque la classification n'a pas été adoptée en l'état. En effet, il apparaît extrêmement difficile d'entrer dans l'activité réelle d'agents dont le contexte d'activité est extérieur à l'intervenant. Les participants ont une vision plus fine de leur activité comparée à celle qu'il est possible d'obtenir à travers ce dispositif. Il faut donc prolonger l'expérimentation.

A ce stade, et puisque notre proposition n'a pas été adoptée, nous avons remis en cause notre méthode d'observation, en nous interrogeant sur les approches du courant de l'Action Située, (évoqué dans le chapitre 3). Il est apparu que les méthodes d'observation de ce courant auraient pu être pertinentes si l'accès au terrain n'avait pas été aussi restreint. De plus, dans les activités que nous cherchions à observer et pour lesquelles les activités de gestion de l'information sont intensives et omniprésentes, ces méthodes d'observation directe nous auraient conduits à gêner les agents dans la réalisation de leurs missions. D'autres facteurs comme la confidentialité des données ont aussi joué un rôle en faveur d'un maintien de notre position méthodologique.

4.4.4 Déroulement du deuxième volet de la phase I de l'expérimentation

Ce deuxième volet de la phase I, durant lequel les participants devaient produire une proposition individuelle de classification à facettes, devait prendre effet dans la suite de la proposition réalisée par l'intervenant. Finalement, les deux participants ont produit chacun une classification à facettes au 23ème mois de l'expérimentation. Ceci étant en partie dû à la difficulté à disposer d'un prototype suffisamment stable et souple d'utilisation, en lien avec les réorientations du projet Miipa-Doc. Nous avions prévu initialement de comparer notre proposition avec celles effectuées par les participants pour évaluer la pertinence de cette dernière et la convergence entre celles des participants et celle de l'intervenant. Dans cette

comparaison, une étude détaillée des tags et facettes utilisés par les participants devait être conduite. Nous voulions également tenter de repérer dans quelle mesure la classification à facettes se constituait selon une forme de transposition des arborescences observées.

4.4.5 Evolution des objectifs du deuxième volet de la phase I de l'expérimentation

Etant donné que notre proposition n'a pas été jugée suffisamment pertinente, et que nous avons conclu de cette appréciation que l'intervenant n'était pas en mesure, ni en position, de comprendre l'activité réelle des agents observés, et donc de produire une classification pertinente, comme énoncé ci-dessus, l'objectif de fonder une méthode d'intervention pour l'élaboration de la classification à facettes en contexte métier, a subi un glissement vers un objectif de prise en main autonome de la classification à facettes à l'aide du prototype.

En effet, les conclusions du premier volet de l'expérimentation nous ont conduits à privilégier plutôt une tentative d'évaluation de la prise d'autonomie des participants quant à l'élaboration de leur classification à facettes. Un biais méthodologique est lié au fait que nous ne pourrons pas déterminer l'influence de notre proposition élaborée durant le premier volet sur les classifications produites par les participants.

Cependant, cette évolution des objectifs de certaines parties de l'expérimentation n'évacue pas les problématiques de transposition des arborescences, et de comparaison entre les principes de structuration de la proposition de classification réalisée par l'intervenant, et ceux des participants.

D'autre part, une comparaison entre les deux classifications produites par les participants paraît, dans ce cadre, pertinente, en termes de structuration des groupes de facettes (des onglets) ainsi qu'en termes d'étude des facettes, pour évaluer dans quelle mesure des individus, qui exercent le même métier, dans le même service et dans la même organisation, bien qu'ayant des activités qui varient, peuvent partager des dimensions de description de leurs documents.

En effet, le prototype intègre une fonctionnalité de partage d'onglets pour l'aspect collaboratif de la gestion documentaire au sein d'un métier. Cette évaluation des potentialités de partage des classifications fait l'objet de la deuxième phase de l'expérimentation.

Nous rappelons ici les questions concernant le deuxième volet de la phase I évoquées au début de la présentation de l'expérimentation :

Dans le cadre de l'évaluation de la similarité entre les arborescences de départ et les classifications produites, il nous faut tenter de répondre aux questions suivantes :

a) Les classifications à facettes reprennent-elles des éléments des arborescences de manière systématique ?

b) S'agit-il d'une re-conception totale de l'organisation de l'information concernant le classement du poste individuel ?

c) Comment la compréhension des principes de la classification à facettes impacte-t-elle la mise en œuvre de l'élaboration de classifications personnelles à facettes ?

Ceci nous conduisait à formuler deux autres hypothèses suivantes que nous notifions ici:

Hypothèse 3 :

Les agents peuvent reconsidérer en globalité leur système de classement.

Hypothèse 4 :

Les agents sont en mesure de s'emparer des principes de la classification à facettes implémentés dans le prototype.

Les autres objectifs nouveaux de ce deuxième volet de l'expérimentation sont les suivants :

Sur l'appropriation des principes de la classification à facettes nous avons cherché à mesurer comment sont compris les principes de la classification à facettes, c'est à dire, comment ils sont mis en œuvre par des experts de la gestion des connaissances ? Ce métier est proche des sciences de l'information et de la communication. En effet, les participants connaissent bien les problématiques liées à la gestion des documents puisqu'ils sont aussi partie prenante du projet Miipa-Doc.

Sur l'appropriation du prototype, nous ne développons pas cette partie de l'expérimentation qui n'a pas pu avoir lieu, en raison de l'avancement retardé du développement informatique. Ce dernier facteur explique aussi la fourniture tardive, dans le calendrier de l'expérimentation, des classifications individuelles par les participants. Cependant, les participants étant membres du projet, ils ont été informés précisément et régulièrement du fonctionnement et des évolutions du prototype.

4.4.5.1 Présentation du mode opératoire des participants

Cette présentation du mode opératoire choisi par les participants nous permet de répondre aux deux questions a) et b), énoncées ci-dessus. En effet, leur mode opératoire a consisté à partir d'un jeu de documents qu'ils ont déterminé individuellement, et à classer et renseigner les métadonnées successivement, en créant les facettes et les tags au fil de l'eau. La classification était vide au début de ce processus.

4.4.5.2 Le positionnement des participants

Leur discours vis-à-vis de cette méthode est le suivant : ils se positionnent comme des « cogniticiens » et selon ce positionnement, ils affirment qu'il est préférable de partir du terrain plutôt que de construire des modèles à priori. Ceci est en accord avec les objectifs du projet de créer un système qui permette l'émergence des classifications à facettes selon une approche ascendante (bottom-up).

4.4.5.3 Les sources

Les jeux de documents utilisés sont différents et propres à chacun des participants, ils comptent pour l'agent A, 12 documents pour l'élaboration de la classification à facettes, et environ 30 documents pour la validation. L'agent B a utilisé entre 30 et 50 documents pour créer la classification à facettes progressivement.

Ainsi à la question a) nous pouvons apporter une réponse claire : les participants ne se sont pas servis de leurs arborescences pour concevoir leur classification à facettes.

Et à la question b), nous pouvons répondre que les participants ont donc envisagé la conception d'une classification à facettes comme une re-conception de l'organisation du classement de leurs documents.

A ce stade, nous pouvons dire que l'hypothèse 3 est validée, c'est à dire qu'ils sont en mesure de reconsidérer leur système de classement en globalité puisqu'ils ne se sont pas inspirés de leurs arborescences.

Pour répondre à la question c), il faut élaborer un jeu de critères de validation de la classification à facettes. Ainsi ce que nous observerons dans cette partie de l'expérimentation dans les classifications fournies par les participants, sera constitué de la structure des classifications en onglets, et les facettes. Nous n'avons pas fait d'analyse sémantique systématique des tags. Nous cherchons à déterminer comment le contexte et les activités sont intégrés dans les classifications en termes structurels.

4.4.5.4 Critères d'évaluation des classifications produites

Pour commencer cette analyse, nous allons donner un contre-exemple de classification à facettes.

Figure 29. Contre-exemple de classification à facettes

Ainsi dans la figure ci-dessus, on peut observer 3 facettes, qui ont des titres différents, mais qui contiennent les mêmes valeurs.

Or, un des principes de la classification à facettes est d'optimiser le classement, aussi, si deux facettes possèdent le même jeu de valeurs, alors il faut les fusionner, et créer une autre facette qui rassemblera les titres des deux facettes qui possédaient les mêmes valeurs. Dans la création de cette nouvelle facette, les titres des deux facettes qui possèdent des valeurs identiques deviennent les valeurs de cette nouvelle facette. Ainsi pour les informations contenues dans cet exemple on préfèrera, pour respecter les principes de la classification à facettes, la structuration suivante, en termes de facettes. Ce procédé permet de réduire le nombre de facettes, et des économies d'espace.

Figure 30. Exemple correct de classification à facettes

Pour poursuivre, nous énonçons un deuxième principe qui est le suivant : dans une classification considérée, si deux facettes possèdent le même titre, alors elles doivent contenir les mêmes valeurs. Deux facettes identiques ne peuvent provenir que d'une duplication, elles ne peuvent pas contenir des valeurs qui diffèrent de l'une à l'autre.

Une facette doit être unique, en ce sens, le titre d'une facette l'identifie comme étant unique, même si dans le prototype elle peut être dupliquée dans plusieurs onglets.

Ainsi une propriété contenue dans les valeurs d'une facette, et dont le titre de la facette décrit le type, est homogène, et toutes les occurrences de ce type de propriété doivent être rassemblées dans la même facette. Ainsi on pourra obtenir, pour une facette qui décrit des lieux, les exemples suivants qui varient selon le degré de granularité de la propriété décrite par les valeurs :

Cet exemple de Facette Lieu, compte parmi ses valeurs, des noms de villes, de pays, de capitales, et de hameaux et de lieudits. Le titre de la facette est assez générique pour englober dans sa formulation toutes les valeurs qu'elle contient, pour former le type de toutes ses valeurs.

Figure 31. Facette Lieu 1

Ce deuxième exemple de Facette traite également une propriété des items qui est un lieu, mais la propriété exacte est le « pays ». Le titre de la facette décrit donc le type de ses valeurs, qui consistent donc uniquement en des noms de pays.

Figure 32. Facette Lieu 2

Ce troisième exemple qui traite d'une propriété qui est un lieu mais dont le type exact est la ville, rassemble des noms de villes, qui pour certaines, sont des capitales, et d'autres non, la sorte de ville est variable dans cet exemple.

Figure 33. Facette Lieu 3

Dans ce quatrième exemple, on traite toujours une propriété qui est un lieu, mais on a décrit le type de ses valeurs comme étant la capitale dans le titre, par conséquent on ne trouve que des noms de capitales dans cette facette.

Figure 34. Facette Lieu 4

Un même type d'information (le lieu) peut donc être transformé en facettes variées selon différents critères qui dépendent de l'objectif de la classification et du contexte. Ainsi le titre de la facette doit décrire le type de propriété contenu dans ses valeurs.

Nous veillerons donc dans cette partie de l'expérimentation, aux deux critères suivants :

- deux facettes ne peuvent pas posséder le même titre si elles ne renferment pas les mêmes valeurs
- un jeu de valeurs identiques dans deux facettes différentes (dont les titres sont différents) est une « erreur » du point de vue des principes de la classification à facettes
- un tag particulier peut se trouver présent dans deux facettes différentes et dans ce cas, il ne s'agit pas du même tag puisque le sens du tag provient du couplage entre le titre de la facette qui énonce à quel type de propriété le tag se rapporte, et le tag désigne une occurrence du type de propriété décrit dans le titre de la facette.

4.4.6 Présentation des données

4.4.6.1 Observations et description et analyse des classifications à facettes

Nous présentons ici des copies d'écran des classifications à facettes produites par les participants, que nous commentons. Nous procédons à une analyse de la structuration des classifications produites et à une évaluation du respect des deux principes énoncés ci-dessus.
Il est à noter que celles fournies par l'agent A ne sont pas produites à l'aide du prototype Hypertagging, en raison des délais de développement informatique. L'agent a donc codé lui-même une interface dans un navigateur pour pouvoir réaliser l'objectif de l'expérimentation.

4.4.6.1.1 La classification produite par l'agent A

La classification produite par l'agent A contient 5 onglets, que nous allons présenter, commenter et analyser.

Onglet 1 : « Caractérisation du document »

Le premier onglet contient des éléments génériques de description du document, il est intitulé « Caractérisation du document » :

Liste des facettes :

- Mon rôle
- Fonction des auteurs
- Auteur EDF R&D
- Auteurs hors EDF R&D
- Auteurs Hors EDF
- Type de document
- Epoque
- Etat du document
- Public visé par le document

L'intérêt de ce contexte est qu'il permet d'avoir une vision synthétique des éléments principaux relatifs au document en une seule vue (dans un seul contexte), il s'agit d'éléments de description externes (par opposition aux éléments relatifs au contenu, qui sont des éléments de description internes). Il est présenté dans la figure suivante.

Du point de vue du scénario d'utilisation, lors de la recherche on peut se limiter à ce contexte pour retrouver un document, ou faire un premier filtrage sur les éléments qui sont présents dans cet onglet, l'utilisateur n'aura pas nécessairement besoin d'aller dans plus de deux onglets pour retrouver le document qu'il cherche dans le cas où il sait assez précisément ce qu'il cherche. Du point de vue de l'ergonomie cette structuration est intéressante.

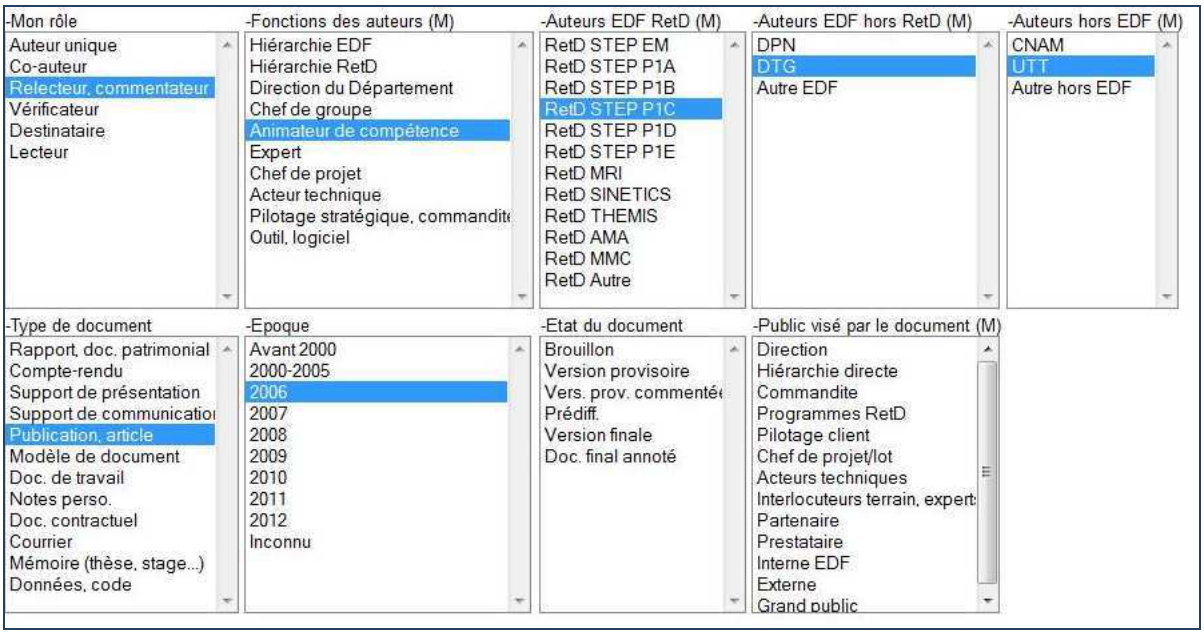


Figure35. Onglet « Caractérisation du document » de la classification de l'agent A

Onglet 2 : « Contenu du document »

Le deuxième onglet est intitulé « Contenu du document », et il traite les aspects liés au contenu d'un document.

Liste des facettes :

- Nature du contenu
- Forme du contenu
- Partie d'un autre document
- Domaine d'application
- Thématique

Ce contexte traite uniquement les informations liées au contenu des documents. Il vient en deuxième après le 1er onglet générique, ce qui permet en termes de recherche de compléter le filtrage réalisé à l'aide du 1er onglet, en passant directement dans le suivant pour ajouter des filtres sur des éléments de contenu. Ceci nous fait dire que dans la hiérarchisation des informations retenues en général par cet agent, les éléments génériques sont les plus aisés à revenir dans la mémoire, et ceux de contenu viennent juste après. L'agent a confirmé cette affirmation, pour environ 90 % des cas. Ce deuxième onglet est présenté dans la figure suivante.

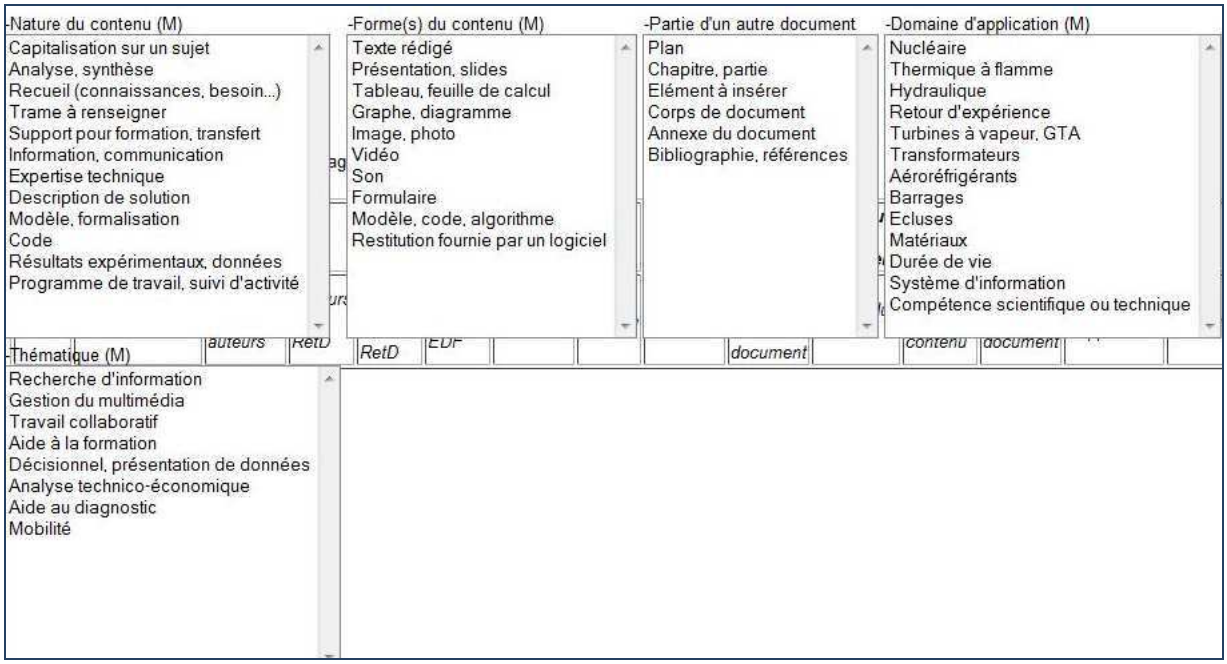


Figure 36. Onglet « Contenu du document » de la classification de l'agent A.

Sur le choix des facettes, les deux facettes suivantes peuvent prêter à confusion : « Nature du contenu » et « Forme du contenu ». Mais lorsqu'on observe les valeurs de facettes, on identifie bien les différentes informations traitées. Ainsi la désambiguïsation du sens du titre de la facette, et par conséquent de la détermination du type d'entité contenu dans la facette, peut provenir de l'observation des valeurs des facettes, qui doivent être homogènes, en référence au titre de la facette.

Onglet 3 : « Réalisation de l'activité »

Le troisième onglet de l'agent A, intitulé « Réalisation de l'activité », sert à renseigner les dimensions relatives aux projets dans le cadre desquels les documents sont produits.

Liste des facettes

- Projet
- Produit ou sous projet
- Phase de l'étude
- Fonction du doc dans le projet

Il vient en complément des autres onglets, notamment du 1er onglet « caractérisation du document ». Cet onglet est présenté dans la figure suivante.

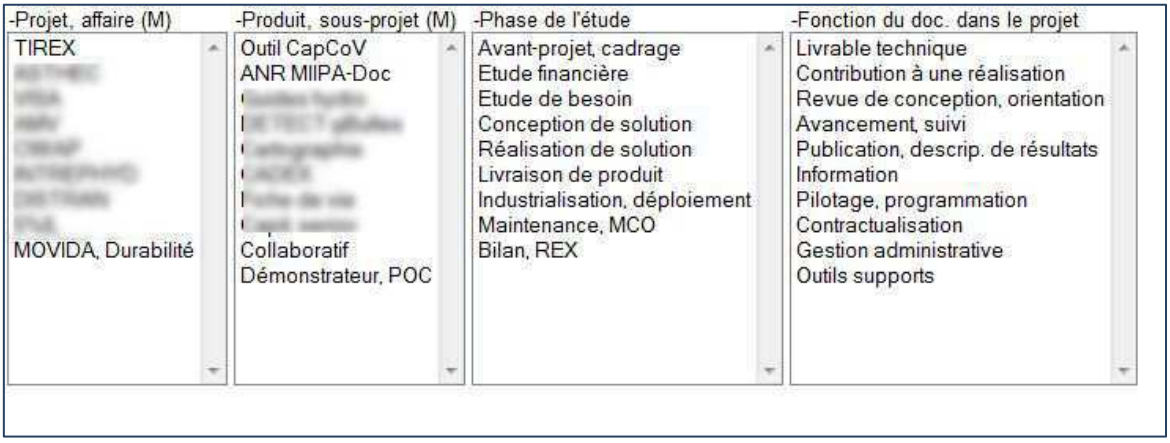


Figure37. Onglet « Réalisation de l'activité » de la classification de l'agent A

En termes de scénario, l'organisation des contextes est ici optimale du point de vue de la structuration globale de la classification. Chaque contexte est bien identifié et a une fonction claire, l'utilisateur devra, pour renseigner les aspects pragmatiquement nécessaires à la bonne indexation et à l'injection des métadonnées, passer par le premier contexte, puis s'il est nécessaire pourra renseigner le Nom du projet, la phase du projet, ainsi que la fonction du document dans ce projet grâce à ce troisième onglet.

A ce stade on peut remarquer que le principe du scénario proposé dans le premier volet de la phase I de l'expérimentation a été ici repris, c'est à dire que toutes les informations génériques relatives au document sont présentées dans le premier onglet, les autres onglets servent à raffiner l'indexation ou les filtres lors de la recherche.

Onglet 4 : « Préparation activité future »

Le quatrième onglet est intitulé « Préparation activité future » et il correspond à une des activités des agents, mentionnées dans l'encadré (partie 4.4.2.4).

Ce contexte trouve donc sa légitimité en référence à cette description de leur activité. Il est présenté dans la figure suivante

Liste des facettes

- Nature du document
- Client possible

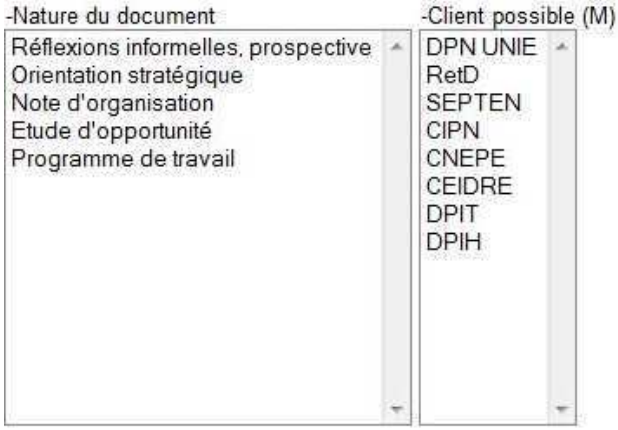


Figure 38. Onglet « Préparation activité future » de la classification de l'agent A.

Une distinction est ici opérée entre des types de document génériques dans l'onglet 1 et des types de document relatifs à cette activité spécifiquement (Etude d'opportunité). Mais on remarque des valeurs hétérogènes dans la facette Nature du document qui comprend des valeurs telles que : Réflexions informelles/ prospective ou Programme de travail ou encore Etude d'opportunité. On constate ici que le principe qui consiste créer une facette pour des types de documents génériques, dans le premier onglet, pour des raisons d'accessibilité et de facilité d'utilisation, en lien avec les autres facettes présentes dans le premier contexte, qui traite les éléments génériques de description des documents, et de créer une autre facette qui sert à décrire des types de documents spécifiques à des activités précises ;ce principe qui était présenté dans la proposition faite par l'intervenant durant le premier volet de cette phase de l'expérimentation a été reproduit ici par l'agent A. Ce contexte vient donc lui aussi en

complément du premier onglet, qui sert par exemple à décrire les éléments relatifs aux auteurs des documents indexés à l'aide de l'onglet 4, produits dans le cadre de l'activité décrite dans le titre de cet onglet.

Onglet 5 : « Gestion de la compétence et Veille »

En référence à l'encadré (partie 4.4.2.4) qui présente les activités des agents, une partie des activités des agents est constituée par la veille et éventuellement l'animation de la compétence. Ces deux aspects de leur activité semblent liés, puisqu'ils sont traités dans le même contexte. Ainsi d'après eux « ils sont dépositaires d'une compétence qui implique une animation. Dans ce cadre, ils ont aussi besoin de savoir où ils en sont, et aussi de savoir quels sont les nouveaux thèmes. La « veille au service de la compétence » permet d'instruire des sujets émergents, de positionner et faire fructifier la veille, le positionnement est un des objectifs de l'animation de la compétence. L'animation concerne la dimension relationnelle de cette activité, tandis que la veille concerne un contenu technique. » (extrait d'un compte-rendu d'une réunion du 18 décembre 2012). Cet onglet est présenté dans la figure suivante.

Liste des facettes

- Compétence concernée
- Aspect traité
- Type de partenaire
- Thème

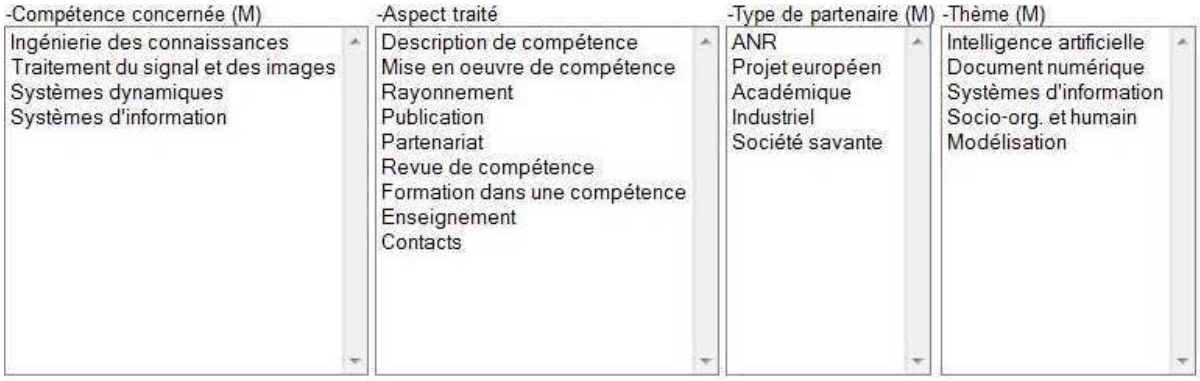


Figure 39. Onglet « Gestion de la compétence et Veille » de la classification de l'agent A

Il apparait que cet onglet vient lui aussi en complément de l'onglet 1, notamment pour le traitement des aspects relatifs au type de document puisqu'avec cette organisation des contextes, si l'utilisateur veut pouvoir renseigner le type de document dans le cadre des activités traitées par l'onglet 5, alors il doit passer par l'onglet 1, l'onglet 5 ne présentant pas

de facette « Type de document » (ni de facette « Type de document spécifique » à cette activité).

Une suggestion sur ce point, pourrait être de dupliquer la facette « Type de document », ainsi que la facette « Epoque » pour faciliter la tâche d'indexation pour l'utilisateur, en lui évitant de devoir passer par l'onglet 1, puis l'onglet 5, pour renseigner des métadonnées sur un document en lien avec l'activité traitée par l'onglet 5.

4.4.6.1.2 Observations générales de la classification de l'agent A

On constate qu'il est possible de mieux comprendre l'activité telle qu'elle est structurée, grâce à la classification à facettes produite puisque la gestion de la compétence et la veille qui apparaissaient dans un premier temps comme deux activités distinctes, se révèlent fortement liées.

Globalement, on observe, deux modes d'entrées dans l'utilisation de cette classification, l'un focalisé sur les documents, l'autre en lien avec les activités des agents observés. Cela signifie que la structuration de la classification est basée en partie sur les documents et sur l'activité de gestion documentaire qui est intensive et omniprésente dans le métier des participants, et sur les activités opérationnelles qu'ils réalisent, activités opérationnelles qui se reflètent dans les trois derniers onglets.

Cette structuration de la classification nous apparait très économe en termes de charge cognitive. Si l'on fait une synthèse des informations que l'on peut traiter à l'aide de chaque contexte, et de manière complémentaire, cela est distribué comme suit :

- Onglet 1 : description générale du document
- Onglet 2 : description du contenu du document
- Onglet 3 : informations relatives à l'activité de gestion de projet
- Onglet 4 : informations relatives à l'activité de préparation de l'activité future
- Onglet 5: informations relatives aux activités de veille et d'animation de la compétence

Enfin, le scénario d'utilisation, qui consiste à passer d'un onglet à l'autre au cours de l'indexation ou de la recherche pour choisir les métadonnées, en fonction de la structuration de la classification, qui avait été jugé peu pertinent dans un premier temps, a été ici reproduit.
Un autre scénario consisterait en l'insertion des facettes du premier onglet dans tous les autres onglets, mais en termes d'ergonomie, et d'IHM, ce scénario semble se révéler peu efficace, ceci étant dû à la surcharge d'informations présentes dans chaque onglet, et à la multiplication des facettes au sein d'un onglet, qui ne garantit pas que l'IHM resterait utilisable facilement.
De plus, la structuration de la classification par le biais des activités nous semble plus apte à se rapprocher des pratiques réelles du travail, et des besoins réels des utilisateurs dans la gestion quotidienne des documents.

Du point de vue du respect des critères de validité de cette classification, ils sont bien respectés, il n'y a pas de facettes qui aient deux titres identiques, ni de jeux de valeurs identiques dans des facettes différentes. Les facettes qui traitent des types d'entités proches comme la facette « Thématique » dans l'onglet 2 et la facette « Thème » dans l'onglet 5, sont différenciées, syntaxiquement leur titre n'est pas identique, et au niveau de la sémantique de la facette, le sens de ces titres est rapporté au contexte reflété dans le titre de l'onglet. Ainsi la facette « Thématique » dans l'onglet « Contenu du document » concerne le thème décrit dans les documents, tandis que la facette « Thème » dans l'onglet « Gestion de la compétence et Veille » est un thème de veille, de niveau plus générique.

4.4.6.2 Classification produite par l'agent B

La classification de l'agent B contient 6 onglets, que nous allons présenter, commenter et analyser. Elle a été réalisée à l'aide du prototype Hypertagging, les icônes qui accompagnent les tags pour former des Sémiotags n'ont pas été renseignées, c'est pourquoi on trouve un carré orangé à côté de chaque tag. Dans les onglets 1 et 6, les listes de valeurs dans les facettes étant trop longues, nous avons séparé en deux copies d'écran distinctes, la présentation de ces onglets.
L'onglet « Général » est un onglet d'amorçage créé par défaut à l'installation du prototype, et ne sera donc pas pris en compte.

Onglet 1 : « Rôle (Co) Auteur Destinataire »

Cet onglet intitulé « Rôle (Co) Auteur Destinataire » est dédié à la description des informations relatives aux rôles de l'utilisateur, de l'auteur et du destinataire du document, ainsi qu'à la description de leur entité d'appartenance. On peut remarquer que ces informations relatives à l'auteur et au destinataire sont plus détaillées que dans la classification de l'agent A, puisque l'agent B a consacré un onglet entier à l'identification des auteurs, des destinataires, et de leurs rôles. Cette pratique révèle la complexité des interactions entre les acteurs, et de leur positionnement vis-à-vis des documents et dans l'activité.

Cet onglet est présenté dans les deux copies d'écran suivantes.

Liste des facettes

- Mon rôle
- Auteur équipe
- Auteur R&D
- Auteur EDF
- Auteur externe
- Rôle auteur
- Destinataire R&D
- Destinataire EDF
- Destinataire externe
- Rôle Destinataire

Figure 40. Partie haute de l'onglet « Rôle (Co) Auteur Destinataire » de la classification de l'agent B

Figure 41. Partie basse de l’onglet « Rôle (Co) Auteur Destinataire » de la classification de l’agent B

Onglet 2 : « Activité et Usage »

Cet onglet intitulé « Activité et Usage » est dédié à traiter les éléments de description des aspects organisationnels et du rôle des documents en lien avec l’activité et les projets. On observe deux facettes nommées de manière identique mais avec des tags différents, ce qui est problématique vis-à-vis de la méthodologie ; c’est un contre exemple de ce à quoi l’utilisateur est censé arriver. Il est présenté dans la copie d’écran suivante.

Liste des facettes

- Activité
- Domaine Progr. Activité
- Projet – Affaire
- Activité
- Année
- Statut du doc pour réaliser
- Brique de travail
- Statut données d’entrées

Nous apportons ici quelques précisions sur les facettes :

La facette « Domaines Prog-activité » correspond aux domaines d’activité.

La facette « Projet Affaire » correspond à un regroupement d’activités.

La deuxième facette « Activité » (la plus à droite de l’écran) détaille les contributions précises de l’agent B.

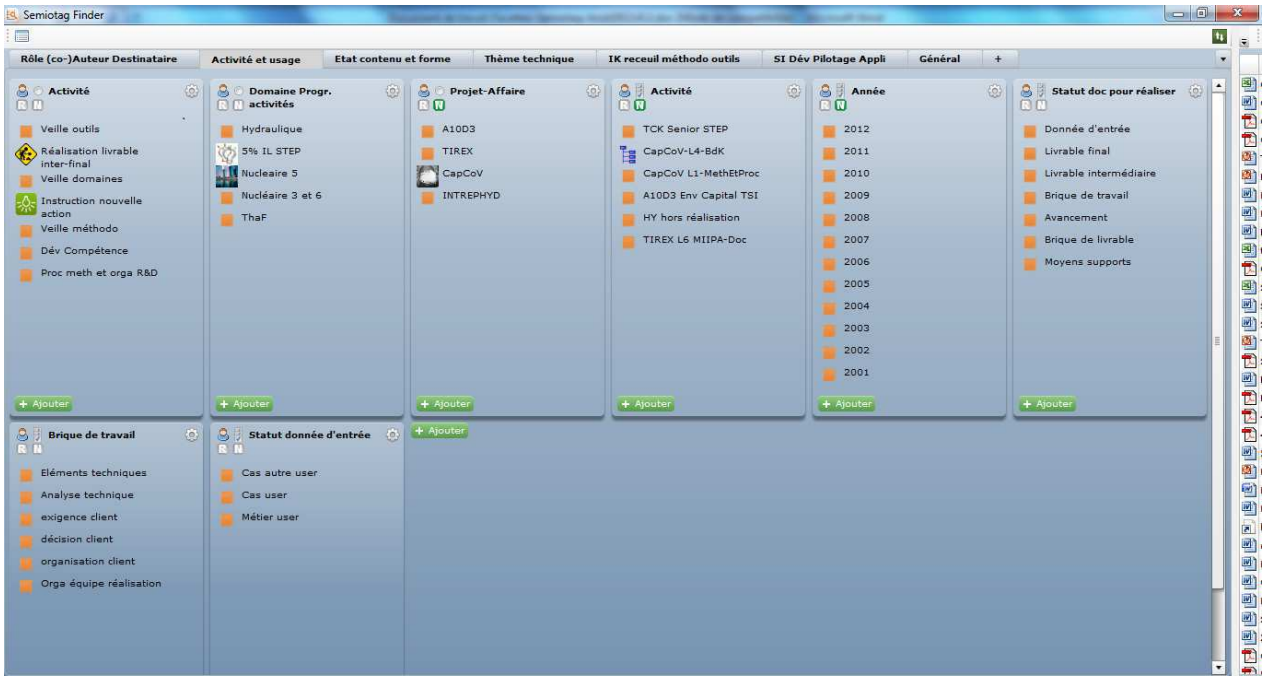


Figure 42. Onglet « Activité et Usage » de la classification de l'agent B

Onglet 3: « Etat contenu et forme »

Cet onglet intitulé « Etat contenu et forme » décrit les éléments relatifs à l'état du document, aux objectifs, au thème, au type de contenu, et à la forme du document. Il est présenté dans la copie d'écran suivante.

Le fait de préciser « Objectif *principal* » dans le titre d'une des facettes présentes dans cet onglet, implique ou sous entend qu'il pourrait y avoir une facette « Objectif secondaire », qui n'est pas présente.

Même remarque pour la facette « Thème *Principal* », mais d'après l'agent B elle correspond à un méta-thème qui sert à définir comment le document va lui servir, cela lui permet de donner une catégorie « dimensionnante » pour son activité.

Une confusion peut apparaître entre les facettes suivantes : « Type d'éléments contenu » et « Objet contenus ».

On observe aussi la présence d'une facette Type de document spécifique, mais on ne trouve à aucun endroit (ou dans un autre onglet) une facette « Type de document générique » qui aurait pourtant été pressentie puisque l'utilisateur a choisi d'intituler la facette « Type de document *spécifique* ».

D'après l'agent B, une facette « Type de document générique » n'apparait pas assez discriminante pour lui, il préfère gérer des états des documents.

Cette dernière pratique évacue le problème des types de documents trop nombreux dans nos études préalables (Desfriches Doria, 2012a, 2012b), tandis que dans la classification de l'agent A, celui-ci a créé des « méta-types » de documents.

Liste des facettes

- Etat du document
- Objectif principal du doc
- Thème principal
- Type d'éléments contenus
- Objets contenus
- Modalités
- Nature
- Types de doc spécifiques

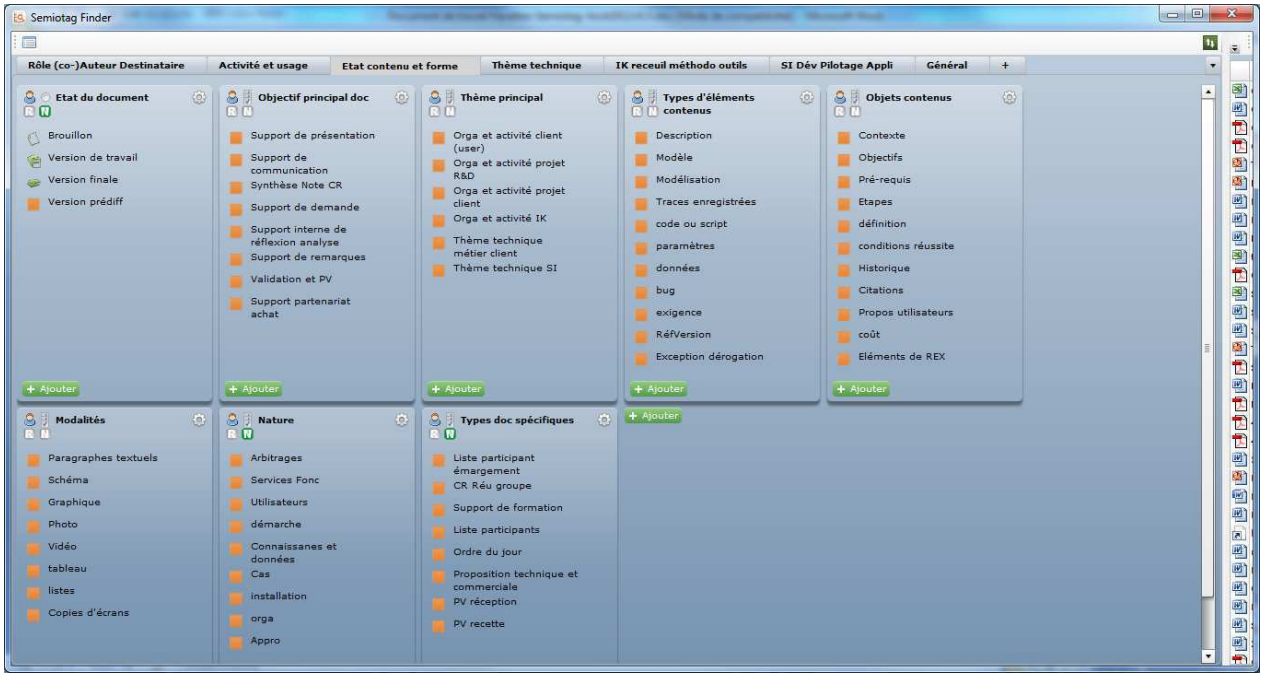


Figure 43. Onglet « Etat contenu et forme » de la classification de l'agent B

Onglet 4: « Thème technique »

Cet onglet intitulé « Thème technique » est consacré la gestion des documents relatifs à des aspects techniques, liés à l'expertise de l'utilisateur et à des aspects individuels et spécifiques de son métier. On constate que la facette « Thème principal » a été dupliquée de l'onglet 3, cette fonctionnalité de duplication de facette est donc apparue utile. Cet onglet est présenté dans la figure suivante.

Le fonctionnement de cet onglet est particulier, il y a une relation entre la première facette, et les autres. Ainsi, le renseignement d'une valeur dans la facette « Thème principal » dans cet onglet, situe le contexte des autres facettes techniques.

Liste des facettes

- Thème principal
- Matériaux
- Composant
- Domaine métier
- Paramètres influents
- Sujet

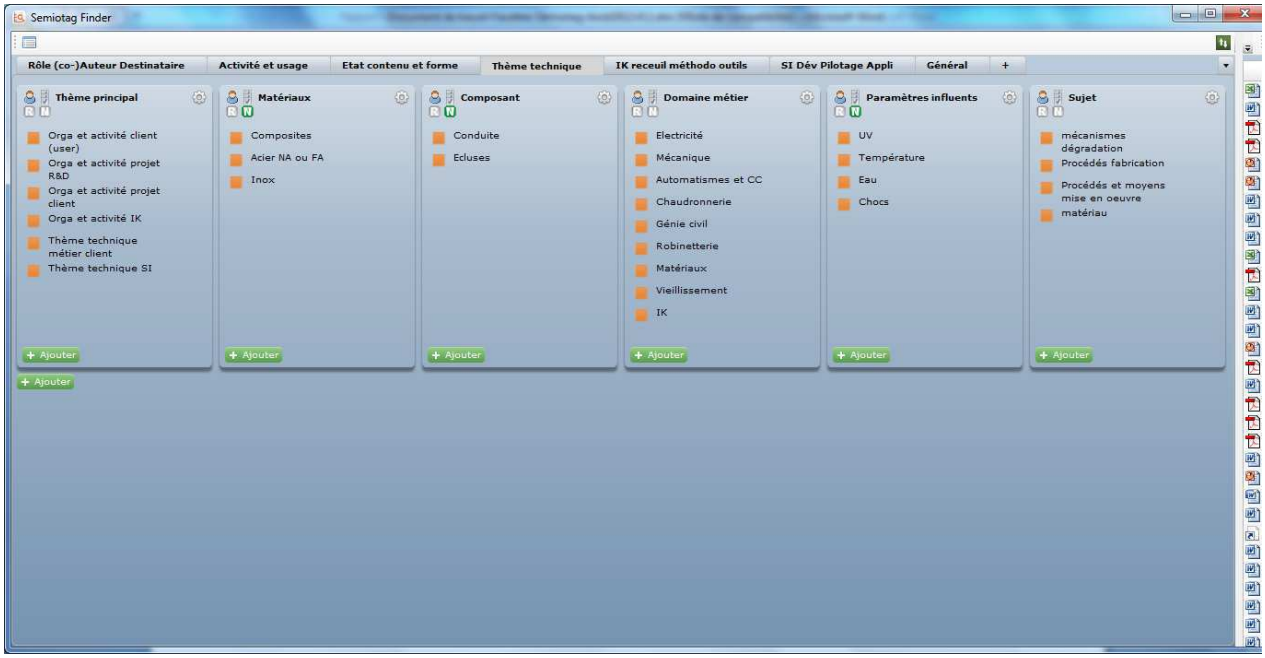


Figure 44. Onglet « Thème technique » de la classification de l'agent B

Onglet 5 : « IK Recueil Méthodo outils »

Cet onglet est intitulé « IK Recueil Méthodo outils ». Il est présenté dans la figure suivante.

Il s'agit de l'onglet qui sert à traiter les documents relatifs à l'activité principale de l'agent B, il concerne le cœur de métier de l'agent.

Liste des facettes

- Type réalisation
- Phase
- Contenu
- Axe IK
- Focalisation K

- Référence
- Usages
- Traitements connaissances
- Thèmes spécifiques

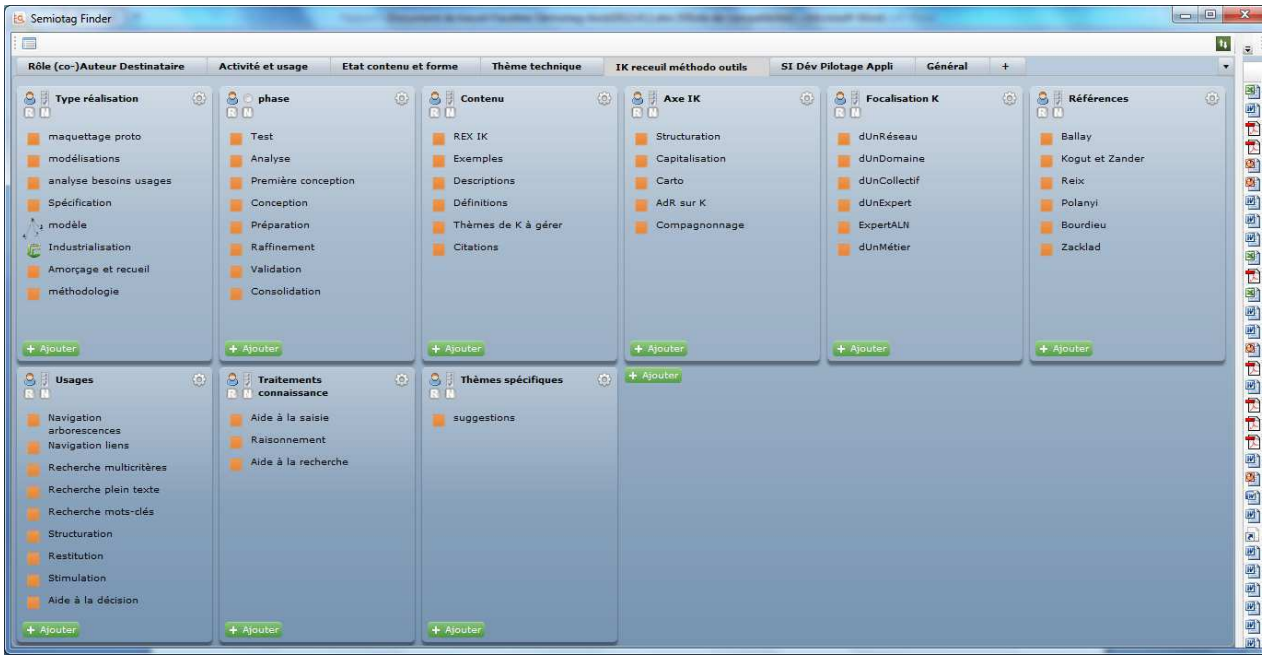


Figure 45. Onglet « IK Recueil Méthodo outils » de la classification de l'agent B

Onglet 6 : « SI Dév Pilotage Appli »

Cet onglet intitulé « SI Dév Pilotage Appli » est clairement orienté pour la gestion des documents relatifs au pilotage des activités de développement d'applications d'après les valeurs de facettes observées.

L'ensemble des facettes de cet onglet est utile particulièrement pour un rôle dans le métier de l'agent B, il traite tous les aspects liés à l'intégration, la définition, le développement et l'intégration dans le SI, il gère aussi les aspects urbanisation (compatibilité avec une autre appli au sein du SI global, architecture).

Cet onglet est présenté dans les deux copies d'écran suivantes.

Liste des facettes :

- Thématique
- Jalon
- Industrialisation
- Thème techno
- Objets techniques
- CAW

- Environnement
- Version recette
- Doc technique SI
- Contrainte et exigence

Figure 46. Partie haute de l'onglet « SI Dév Pilotage Appli » de la classification de l'agent B

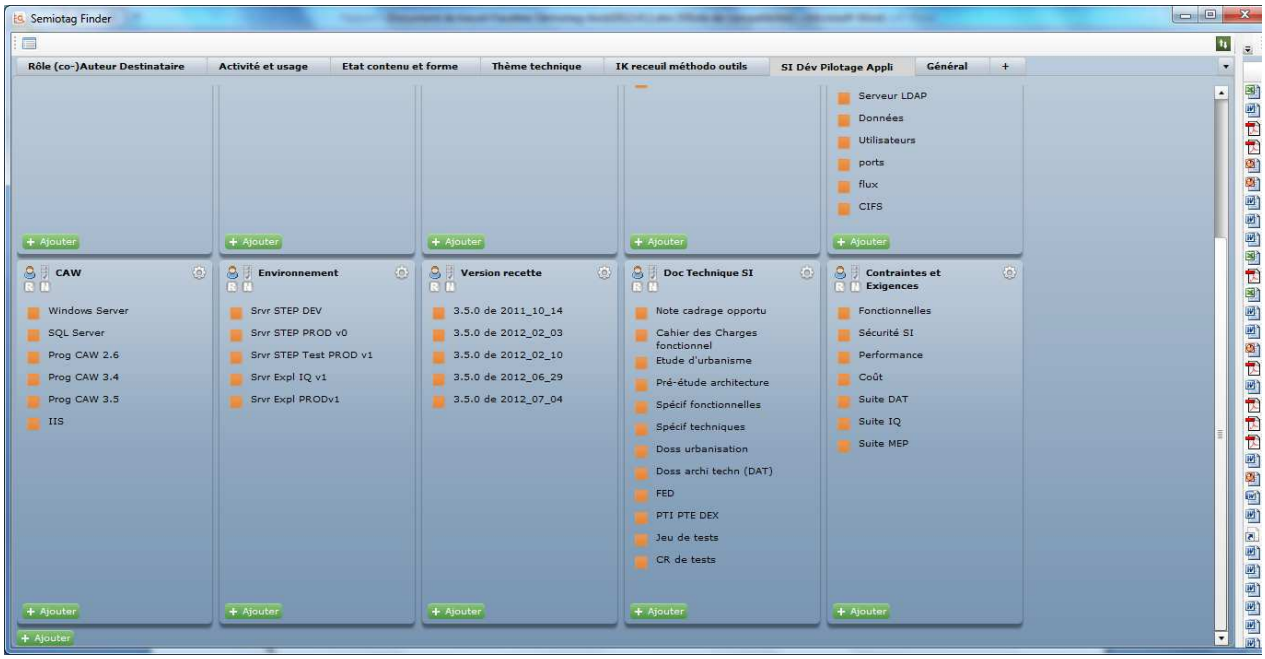


Figure 47. Partie basse de l'onglet « SI Dév Pilotage Appli » de la classification de l'agent B

4.4.6.2.1 Observations générales de la classification de l'agent B

On constate de manière générale une granularité fine dans cette classification, certaines facettes contenant de nombreuses valeurs, ce qui peut éventuellement la rendre difficile à utiliser.

Concernant le scénario d'utilisation son fonctionnement est basé sur le même principe que pour l'agent A et que celui qui a été proposé par l'intervenant dans le volet 1 de cette première phase de l'expérimentation. Il consiste à passer d'un onglet à l'autre pour l'indexation d'un seul document mais, dans cette classification, ce scénario peut potentiellement être plus difficile à utiliser puisqu'il y a un onglet entier consacré aux auteurs/destinataires, un onglet projet, un onglet Etat- Forme -Contenu du document, puis des onglets plus orientés métier. Si l'utilisateur veut renseigner les métadonnées complètes à associer à un document, il doit passer par les trois premiers onglets pour renseigner les informations associées à l'auteur et au destinataire, puis les informations relatives aux aspects organisationnels, et ensuite les informations relatives à l'état, la forme, ou le contenu du document. Enfin, il choisira un onglet métier parmi ceux qui traitent le cœur de métier, ou les aspects techniques, ou encore le pilotage du développement d'une application.

Dans cette classification, on constate qu'il y a des modes d'entrées divers dans la gestion et la recherche des documents. Ainsi, si l'on fait une synthèse des éléments d'information que l'on peut traiter avec ces onglets, il s'agit de :

- Onglet 1 : description très précise des éléments liés aux auteurs et aux destinataires
- Onglet 2 : description des aspects organisationnels du document (projet)
- Onglet 3 : informations relatives à l'état, la forme ou au contenu du document
- Onglet 4 : informations relatives à des aspects techniques d'une partie de l'activité de l'agent B
- Onglet 5 : informations relatives à l'activité qui constitue le cœur de métier de l'agent B
- Onglet 6 : informations relatives à l'activité de pilotage du développement d'applications, qui constitue une partie de l'activité métier de l'agent B.

Concernant la validité de l'application des principes de la classification à facettes, plusieurs observations peuvent être produites :

Sur le premier principe, on a relevé deux facettes qui possèdent le même titre (« Activité ») dans l'onglet 2. Ces deux facettes « Activité » désignent deux composants différents de l'activité de l'agent. La première facette « Activité » décrit des activités-types, que l'on retrouve dans la structuration des onglets de la classification de l'agent A. Dans la première facette « Activité » de l'agent B, on trouve des valeurs comme « Veille outils », « Veille méthodo », « Veille domaine » et une valeur « dév Compétence » qui signifie « Développement de la Compétence ».

Ainsi, ces activités de Veille (qui servent aussi à la préparation de l'activité future) et de développement de la compétence, ne sont pas absentes de la classification de l'agent B, elles sont moins mises en avant, et moins pratiquées, que les activités techniques, qui font l'objet de deux onglets chez l'agent B. Chez l'agent A, on n'observe pas d'onglet purement technique. La deuxième facette « Activité » de l'agent B détaille des contributions précises de l'agent B.

Ces deux facettes ne désignent donc pas le même type d'entité, il s'agit d'une erreur du point de vue syntaxique, mais les entités décrites ont des degrés de granularité différents, la deuxième étant plus précise, et centrée sur l'utilisateur, cette « erreur » peut donc être corrigée en changeant le titre d'une de ces deux facettes.

Sur le deuxième principe, qui énonce qu'un jeu de valeurs identiques dans deux facettes différentes (dont les titres sont différents) est une « erreur », nous avons relevé quelques observations :

Dans le premier onglet de l'agent B, qui traite de manière détaillée les aspects liés aux auteurs et destinataires, certaines facettes ont des valeurs qui pourraient potentiellement être identiques. Ainsi en est-il des facettes « AuteurEDF » et « Destinataire EDF » d'une part, et d'autre part des facettes « Rôle auteur » et « Rôle Destinataire ».

Nous considérons qu'il ne s'agit, ici, pas d'une erreur, même si avec l'évolution dans le temps de cette classification, les valeurs pourront être identiques entre ces deux couples de facettes étudiés, car les facettes dans cet onglet tentent de refléter les interactions autour du document et son trajet, les facettes ici ont un rôle particulier, une fonction de traçage des entités d'origine des auteurs ou des destinataires. Cela révèle que les agents étudiés sont en interaction avec beaucoup d'interlocuteurs différents dont les rôles varient au fil des activités.

Globalement on remarque que le scénario choisi est le même entre les deux agents, il consiste à passer d'un onglet à un autre pour indexer de manière de plus en plus précise le document.
La forme globale de la structuration est la même chez les deux agents, avec des variantes. Ainsi, on a pu observer que les agents dédient un ou plusieurs onglets aux aspects formels et au contenu du document, les autres onglets sont consacrés à renseigner des informations relatives à des activités, conçues selon des degrés de granularité variables. L'agent B a ajouté un onglet pour les informations relatives aux auteurs et destinataires des documents (Onglet 1). A la vue des observations qui concernent la validité des classifications produites, nous pouvons dire que nous avons validé l'hypothèse 4. Nous la reportons ici :

Hypothèse 4 :
Les agents sont en mesure de s'emparer des principes de la classification à facettes implémentés dans le prototype.

Enfin, à travers les observations de la structuration des classifications produites par les participants, nous pouvons dire que nous avons validé l'hypothèse 5. Nous la reportons ici :

Hypothèse 5 :
La structuration de la classification à facettes en onglets, implémentée dans le prototype, peut se faire à l'aide d'éléments hétérogènes. Les onglets ne reflètent pas nécessairement seulement les activités, mais peuvent être construits sur d'autres éléments du contexte. Ainsi on pourra par exemple, trouver un onglet « Projet », un onglet « Document », et un onglet « Métier »

4.5 Phase II de l'expérimentation

Cette deuxième phase vise à évaluer les éléments qui sont potentiellement partageables au sein de ces classifications des deux agents, qui exercent le même métier, dans le même service et la même organisation. Si nous avons vu précédemment que les modes de structuration peuvent être rapprochés, qu'en est-il des facettes ?

4.5.1 Objectifs de la phase II

Cette phase de l'expérimentation vise à tenter de déterminer comment partager, entre des acteurs du même métier, dans le même service et la même organisation (ce qui constitue l'unité minimale d'observation d'un groupe métier), des facettes et des modes de structuration cohérents de la classification, pour permettre le partage de documents.

Ainsi, le système d'injection de métadonnées contextuelles proposé par Hypertagging ne peut pas prendre en compte tous les aspects potentiellement pertinents pour une personne, parce que l'inflation du nombre de facettes rendrait le système inutilisable, il faut donc effectuer des choix avertis, pour sélectionner les facettes les plus pertinentes. Cette problématique se pose de manière encore plus prégnante pour le partage de facettes et d'onglets au sein d'un groupe.

Nous tentons d'apporter des réponses, à la question suivante : comment faire ces choix de facettes en fonction des éléments suivants :

- l'outil s'adapte aux utilisateurs, les facettes ne sont pas imposées, mais émergentes
- l'outil vise à équiper l'activité réelle des utilisateurs, et notamment l'activité réelle de gestion des documents

4.5.2 Observations des éléments communs entre les deux classifications

Ainsi nous repartirons de la présentation des onglets de chaque classification qui a précédé pour la comparaison du contenu des classifications. Nous allons énumérer dans cette partie les éléments communs en termes de types d'entités traités qui sont communs. Ainsi nous ne produirons pas ici d'analyse détaillée des tags utilisés par les agents observés.

Dans les deux classifications, les aspects qui concernent les auteurs et destinataires sont bien traités, mais selon des modalités qui diffèrent. Nous considérerons donc que ces aspects sont communs aux deux agents. Seules leurs modalités nécessitent des ajustements.

L'aspect temporel est lui aussi traité par les deux classifications, mais les modalités varient là aussi. Ainsi l'un utilise des périodes de temps tandis que l'autre utilise exclusivement des années comme valeurs de facettes pour ce type d'entité.

Le domaine d'application (thermique, nucléaire, hydraulique) est présent dans les deux classifications. Pour l'un il sert à décrire le contenu d'un document, tandis que pour l'autre il est inséré dans des aspects organisationnels.

Le nom des projets est lui aussi un élément commun aux deux classifications, pour l'un des agents il est indiqué dans l'onglet qui traite le cœur de métier, tandis que pour l'autre il est intégré dans un onglet qui traite des aspects organisationnels.

Concernant le document en lui-même, l'état du document est traité par les deux classifications dans un onglet qui gère le contenu ou la forme du document chez les deux agents.
La thématique des documents est intégrée dans les deux classifications dans des onglets qui gèrent le contenu ou la forme du document chez les deux agents. Leurs valeurs varient.
La « nature du contenu » ou le « type d'éléments contenus » (description, modélisation, analyse, synthèse, données, code) sont des aspects communément intégrés dans des onglets qui gèrent le contenu ou la forme du document. Les valeurs sont variables et l'intitulé de la facette aussi.
La « forme du contenu » ou « modalités » du contenu (texte, schéma, graphique, tableau, liste, vidéo, son) sont des facettes toutes les deux relevées dans les classifications, dans des onglets qui servent à traiter le contenu ou la forme du document chez les deux agents. Leurs valeurs varient et leur intitulé également.
La « fonction du document dans le projet » présente dans l'onglet 3 chez l'agent A (onglet qui traite le cœur de métier) peut se rapprocher des deux facettes suivantes chez l'agent B : « statut du doc pour réaliser » et « brique de travail » dans l'onglet 2 qui traite des aspects organisationnels. Les valeurs entre ces facettes sont, en effet, assez proches et contiennent des tags comme par exemple : livrable, contribution, avancement, support.
La facette « phase de l'étude » qui est présente dans l'onglet 3 chez l'agent A (onglet qui traite le cœur de métier) pourrait être rapprochée de deux facettes présentes dans l'onglet 5 de l'agent B, qui traite lui aussi le cœur de métier. Ces dernières sont intitulées « Type de réalisation » et « Phase ». Leurs valeurs peuvent être rapprochées même si elles varient.

4.5.3 Résultats

Ainsi, nous pouvons affirmer que les aspects qui sont orientés autour du document en lui-même et de son contenu, sont globalement partageables puisqu'ils sont relevés en commun dans les deux classifications. Nous les listons ci-après :

- ceux qui concernent l'auteur et les destinataires
- l'aspect temporel ;
- le domaine d'application ;
- l'état du document ;
- la thématique du document ;
- la nature du contenu ;

- la forme du contenu.

Concernant les activités, seulement certaines parties de l'activité peuvent être partagées, et notamment ce qui constitue le cœur du métier. Ainsi, nous avons relevé les facettes suivantes qui pourraient être partagées avec des ajustements :

- le nom des projets ;
- la fonction du document dans le projet ;
- la phase de l'étude.

On a pu constater ainsi, que l'agent B a une orientation plus technique, que l'agent A qui travaille sur des tâches moins techniques, comme la préparation de l'activité future, et la gestion de la compétence (du service). Ainsi, les classifications produites reflètent les missions particulières dans l'activité des agents, qui conduisent à des besoins différents. En effet, l'agent A, en plus des missions classiques de R&D en ingénierie des connaissances a été chargé de la mission d'animer la compétence, ce qui entraine des besoins de classification de l'ensemble des travaux réalisés dans le service, et par conséquent, des besoins de classification plus universelle, tandis que l'agent B compte parmi ses missions celle qui correspond au pilotage opérationnel du développement d'une application industrialisée dans le système d'information de l'entreprise, ce qui implique de gérer de nombreuses versions de l'application, et des aspects techniques et organisationnels associés au développement, à ses évolutions, à son maintien en conditions opérationnelles, ou encore à son intégration dans le SI.

Les acteurs d'un même métier, dans le même service, et la même organisation, ont donc des activités qui sont différentes, ce qui avait été difficile à percevoir lors des entretiens collectifs. Ce qui peut être partagé correspond à ce qui est le plus proche du cœur de l'activité métier.

4.6 Conclusion

Au cours de cette expérimentation, nous avons tenté de produire une classification à facettes pour deux acteurs du même métier, en nous basant sur des entretiens et des observations de leur classement personnel de l'information. Il ressort de la première phase de l'expérimentation, que les méthodes utilisées n'ont pas permis d'observer suffisamment l'activité réelle de gestion de l'information, tandis qu'elle a permis d'obtenir une vision globale de l'activité métier des participants. Ainsi, globalement, les classifications proposées

par les acteurs sont plus pertinentes, plus complètes, plus détaillées, et contiennent plus de jargon et d'implicite que celles que nous avons proposées.

Le deuxième volet de la première phase de l'expérimentation nous a permis de constater que la structuration individuelle des classifications produites par les participants présente des éléments communs. De même que dans les arborescences observées, le système à facettes est utilisé pour un usage personnel et individuel mais on y retrouve des éléments qui pourraient être partagés, même si certains ajustements sont nécessaires.

D'autre part, l'expérimentation révèle que deux acteurs du même métier, dans le même service, au sein de la même organisation peuvent avoir des activités en partie différentes, ce qui se retrouve dans les classifications produites individuellement par chaque participant.

Cette deuxième phase nous a permis de mieux comprendre la distribution des activités, et ce qu'il y a de commun dans l'activité de ces deux participants, et qui constitue le cœur de métier, c'est à dire l'expertise en gestion des connaissances. Le « basic level » peut donc se situer autour des aspects qui traitent le cœur de métier.

Finalement, nous pensons que les utilisateurs futurs du prototype Hypertagging doivent être les auteurs principaux de leurs classifications à facettes. L'appropriation correcte des principes de la classification à facettes est le signe du succès de la démarche. L'intervention méthodologique reste nécessaire pour la pédagogie qui a pour objectif de permettre et de garantir l'appropriation de l'outil, et pour l'élaboration d'un socle commun de facettes et de tags partagés au niveau de collectifs restreints, pour le travail collaboratif, et la mutualisation des ressources.

Le chapitre suivant sera donc consacré à la présentation d'une méthodologie générative de classification à facettes implémentée dans le prototype, basée sur l'apprentissage des principes de la classification facettes et la prise en main autonome de l'outil. La question méthodologique principale à laquelle nous tenterons d'apporter des réponses sera la suivante : Comment guider des utilisateurs dans la conception de leur propre système d'organisation des connaissances à facettes ?

Chapitre 5 : Méthodologie d’élaboration de FolkClassifications à facettes pour la gestion des connaissances

5 Chapitre 5 : Méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances

5.1 Introduction

D'une manière générale, nous proposons ici une méthodologie qui vise à faciliter les activités récurrentes de gestion de l'information, en exploitant leur potentiel informatif afin d'appuyer sur celles-ci une démarche de gestion des connaissances métier dans les organisations de type entreprise. Il s'agit ici de faire reposer la démarche de gestion des connaissances sur l'activité de gestion de l'information, qui est nécessaire à l'activité métier, quotidiennement.

Elle s'appuie sur les processus de documentarisation des documents, les pratiques personnelles de gestion de l'information (PIM) et le prototype de SOC que nous avons présenté en chapitre 3. Le processus est ainsi moins lourd et moins couteux en termes d'investissement personnel, en termes financiers et de temps que dans les démarches de gestion des connaissances basées sur des modélisations successives évoquées dans le chapitre 1. Il s'agit d'une démarche de gestion des connaissances inspirée des méthodes du Web 2.0, puisque le prototype présenté est un outil de documentarisation basée sur le tagging par les utilisateurs, structuré à l'aide de facettes, qui sont elles-mêmes élaborées, ainsi que la structure globale du dispositif, d'après la méthodologie que nous présentons.

Dans cette méthodologie, nous sommes focalisés sur les connaissances métiers, car il nous apparaît que dans l'organisation par projets, qui est très répandue dans les organisations actuellement, et qui rassemble des acteurs de métiers différents, les projets fonctionnent d'après des processus et des déroulements très encadrés voire systématiques, qui sont largement équipés par des outils informatiques, gérant les plannings, le phasage des projets, ainsi que des aspects comme le chemin critique ou le partage des documents relatifs aux projets.

Il nous semble que le métier et l'activité, tels que définis dans le chapitre 3, et comme nous avons pu le constater d'après nos expérimentations, possèdent des dimensions personnelles plus ou moins semblables, et des dimensions collectives, qui gagnent à être explorées à partir

des pratiques de gestion de l'information, souvent réalisées dans un relative solitude (Guyot, 2009), qui constituent pourtant un des moyens de la réalisation des objectifs de l'activité métier, et forment une sorte de dénominateur commun.

Ainsi, tous les acteurs d'une organisation ont parmi leurs activités, des activités de gestion de l'information. Tous les acteurs d'un projet également, mais dans le cadre des métiers, nous avons pu découvrir que certaines dimensions de l'activité et certaines représentations sont communes. Ces dernières constituent le socle d'un partage pérenne, qui permet de partager et diffuser des informations à des acteurs de l'organisation exerçant le même métier (parfois en des lieux distants), d'améliorer les pratiques métier en faisant profiter les autres acteurs du métier d'une expertise acquise par l'expérience, de favoriser l'émulation, et de permettre le développement du métier (Clot, 2008) pour des acteurs dont l'expertise peut être hétérogène, tout en s'appuyant sur les pratiques indispensables de gestion de l'information.

Ainsi, si deux acteurs du même métier dans la même organisation, n'exercent pas toujours leur métier de manière identique, il subsiste des éléments communs, comme la culture de métier, les représentations, et des significations contextuelles associées à des termes. Nous nous sommes donc orientés vers une méthodologie axée sur ce concept de métier et fondée sur un SOC.

Dans ce cadre, nous proposons d'exploiter le temps passé à la gestion de l'information pour mettre en place une pratique qui intègre une dimension supplémentaire, explicitée, non contrainte, de gestion des connaissances, et qui permet de redonner un sens et une reconnaissance à ces activités de gestion de l'information (le sens du travail) (Clot, 2008).

Dans ce sens, c'est ponctuellement que l'utilisateur, qui est intégré au processus de création du dispositif pour une meilleure adhésion et compréhension de la démarche, partagera des ressources, que lui-même jugera utile aux autres membres de son métier. En effet, il ne s'agit pas de tout partager, mais seulement ce qui peut être jugé potentiellement utile par les utilisateurs. Ainsi, les espaces personnels de tagging des documents correspondent à la gestion de l'information dans son flux quotidien, tandis que l'espace de tagging, partagé au sein des acteurs d'un métier, sert à la capitalisation des connaissances métier.

A terme, ce SOC individuel et collectif, construit itérativement, constitue un référentiel commun, au sens où nous l'avons défini dans le chapitre 3 (partie 3.5.8.3). Du point de vue de l'ergonomie, les activités de construction, d'alimentation et de maintenance de ce SOC sont vues comme des activités méta-fonctionnelles, puisqu'elles visent la construction et la maintenance d'un référentiel commun.

Par ailleurs, cette démarche de gestion des connaissances suit les recommandations formulées par (Falzon et Sauvagnac, 2001), qui précisent qu'il est « souhaitable d'intégrer (…) l'activité méta-fonctionnelle à l'activité fonctionnelle (…) [et qu'il] est plus efficace d'accomplir ces deux activités simultanément, et plus profitable à chacun de mener une activité réflexive aussi souvent que possible. » Dans ce cadre, l'activité fonctionnelle correspond à la gestion du flux permanent de l'information, tandis que l'activité méta-fonctionnelle se rapporte à la gestion des connaissances, et correspond concrètement à l'anticipation des usages possibles des documents gérés ainsi qu'à une vision prospective de l'activité fondée sur l'expérience.

La méthodologie que nous présentons est donc une méthodologie d'élaboration de FolkClassifications à facettes comme soutien d'une démarche de gestion des connaissances métier. Il s'agit d'une première version de cette méthodologie, issue de nos expérimentations. Aussi, des expériences supplémentaires restent à mener pour sa validation et son perfectionnement. Par ailleurs, le travail présenté dans ce chapitre constitue la matière d'un livrable du projet Miipa-Doc.

Elle se démarque des méthodologies abordées en chapitre 2 par le fait qu'elle ne vise pas à faire entrer le modèle de facettes dans un modèle préexistant, mais à élaborer un modèle de facettes contextuel adapté aux dimensions fluctuantes des activités métier, dans des collectifs restreints, au sein d'organisation considérées comme uniques. En effet, les modèles de facettes élaborés ne visent pas à l'universalisation des schémas de facettes. Ils sont propres à chaque groupe métier, dans chaque organisation, même si certains éléments peuvent être transférables d'une organisation à l'autre.

5.2 Objectifs méthodologiques

D'une manière générale, la méthodologie doit permettre la gestion personnelle de l'information, la gestion collaborative des connaissances métier de manière émergente, ainsi que l'élaboration itérative d'une classification à facettes.

Concernant la gestion personnelle de l'information, le succès de la démarche dépend de la prise d'autonomie des acteurs vis-à-vis de l'utilisation de l'outil et de sa maintenance. Cette autonomie recherchée des utilisateurs est construite en partie sur le fait que la démarche méthodologique les intègre au processus d'élaboration de la classification. L'autonomie des

acteurs est un enjeu fort puisque le dispositif est basé sur les principes du Web 2.0, et, par conséquent, sur la participation des utilisateurs et sur leur appropriation de l'outil.

En termes de gestion des connaissances, le succès de la démarche relève de la pédagogie qui accompagne le processus méthodologique. Celle-ci s'appuie sur les conceptions de l'activité issues de l'ergonomie et de la psychologie du travail, qui fournissent un cadre théorique et pragmatique d'analyse de l'activité, vue comme en développement. L'activité, dans ce cadre, n'est pas figée, mais bien dans une dynamique permanente d'évolution et d'ajustement. Par ailleurs, à travers cette démarche, il s'agit d'équiper une activité de gestion de l'information, secondaire vis-à-vis des activités métiers, bien qu'indispensable, et de lui redonner une place explicitement reconnue au service de l'activité métier ainsi qu'une fonction supplémentaire d'enrichissement des pratiques métiers.
Les études révèlent une activité de gestion de l'information largement à la charge de chaque utilisateur de l'information. Cette activité est mise au service de la réalisation du travail au sens du travail réel. Elle relève en partie du « bricolage », ceci étant lié au manque d'outils adaptés et de temps à y consacrer. Concernant la gestion des connaissances, cette méthode doit permettre de découvrir les aspects communs de l'activité au sein d'un métier, exercé dans une organisation identifiée.
L'élaboration de la FolkClassification est réalisée dans une compromis entre les approches ascendantes et descendantes (top down et bottom up), selon un processus itératif. Elle se déroule en trois phases qui seront détaillées dans la suite de cette partie.

5.3 Contraintes

Les contraintes auxquelles il faut se plier dans l'élaboration de la classification sont liées principalement aux problématiques génériques des IHM (Interface-Homme Machine). Elles nous contraignent en termes d'espace dans l'écran. Ainsi, si l'on veut garder les avantages de la classification à facette dans le prototype, il faut limiter à un nombre raisonnable les onglets, les facettes et les tags pour garantir la pérennité et la pertinence de la classification. D'autres contraintes peuvent émerger comme celles liées à des utilisateurs trop peu disponibles, ou peu coopératifs, pendant les entretiens et la suite de la méthodologie.

5.4 Définitions

Nous proposons ici des définitions des principaux concepts de cette méthodologie liée au prototype Hypertagging.

5.4.1 Facette

Les facettes découpent l'univers original des items à classer dans son ensemble, selon plusieurs critères de caractérisation de cet ensemble simultanément, chaque item de l'univers n'étant pas tenu de posséder toutes les propriétés décrites par l'ensemble des facettes disponibles dans le schéma de classification. Ainsi, une valeur de facettes correspond à une occurrence d'une propriété des items classés, propriété qui est mentionnée dans le titre de la facette. Le titre de la facette décrit donc le type des valeurs contenues dans la facette. Ainsi, deux règles sont à respecter :

- deux facettes ne peuvent pas posséder le même titre si elles ne renferment pas les mêmes valeurs ;
- un jeu de valeurs identiques dans deux facettes différentes (dont les titres sont différents) est une « erreur » du point de vue des principes de la classification à facettes.

Les facettes sont donc mutuellement exclusives mais il est possible d'attribuer à un document plusieurs valeurs provenant de la même facette (multi-valué).

5.4.2 Sémiotag

Association d'un tag, qui est un mot clé décrivant les items à classer en langage naturel, qui représente une occurrence d'une propriété des items à classer, et d'une icône représentant de manière visuelle le sens du tag, de manière contextuellement significative pour l'utilisateur ou les utilisateurs. Ce sont les valeurs inscrites dans une facette.

Un Sémiotag particulier peut se trouver présent dans deux facettes différentes (facettes qui ont deux titres différents). Dans ce cas, il ne s'agit pas du même tag puisque le sens du tag provient du couplage entre le titre de la facette, qui énonce à quel type de propriété des items le tag se rapporte, et le tag désigne une occurrence du type de propriété décrit dans le titre de la facette.

Les valeurs de facettes sont donc homogènes au sein d'une facette, en termes de type de propriété des items qu'elles décrivent, mais leurs valeurs peuvent connaître des variations aux niveaux suivants :

- variation grammaticale : noms, verbes, adjectifs ;
- variation sémantique : une facette « Colloques » pourrait avoir pour valeurs des noms de lieux, des noms de colloques, des noms de personnes organisatrices de colloques, tant que cela reste cohérent pour l'utilisateur ;
- variation utilitaire : les mêmes noms de villes pourraient être utilisés dans deux facettes différentes pour désigner, pour une facette, des informations relatives à des colloques, et pour une autre facette, des informations de type purement géographique, cela dépend du titre attribué aux deux facettes, qui doit être différent pour chacune d'elles ;
- variation du niveau de granularité : une facette géographique, par exemple, pourrait avoir pour valeurs des noms de pays, de villes, de rues, tant que le titre de la facette est cohérent et englobe toutes ces valeurs dans son niveau de généralité.

Nous reprenons ici une explication exposée au chapitre 4 pour plus de clarté de notre propos. Ainsi, on pourra obtenir, pour une facette qui décrit des lieux, les exemples suivants, qui varient selon le degré de granularité de la propriété décrite par les valeurs :

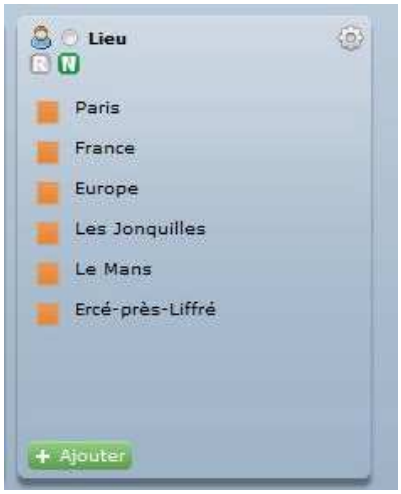


Figure 48. Facette Lieu 1

Cet exemple de Facette Lieu compte parmi ses valeurs des noms de villes, de pays, de capitales, de hameaux et de lieux-dits. Le titre de la facette est assez générique pour englober dans sa formulation toutes les valeurs qu'elle contient, pour former le type de toutes ses valeurs.

Figure 49. Facette Lieu 2

Ce deuxième exemple de Facette traite également une propriété des items qui est un lieu, mais la propriété exacte est le « pays ». Le titre de la facette décrit le type de ses valeurs qui consistent donc uniquement en des noms de pays.

Figure 50. Facette Lieu 3

Ce troisième exemple qui traite d'une propriété qui est un lieu mais dont le type exact est la ville. Elle rassemble donc des noms de villes, qui pour certaines sont des capitales, et d'autres non. La sorte de ville est variable dans cet exemple.

Figure 51. Facette Lieu 4

Dans ce quatrième exemple, on traite toujours une propriété qui est un lieu, mais on décrit le type de ses valeurs comme étant la capitale dans le titre. Par conséquent on ne trouve que des noms de capitales dans cette facette.

Un même type d'information (le lieu) peut donc être transformé en facettes variées selon différents critères qui dépendent de l'objectif de la classification et du contexte. Ainsi, le titre de la facette doit décrire le type de propriété contenu dans ses valeurs.

5.4.3 Onglet

Représentation d'un groupement de facettes, en fonction d'éléments contextuels, qui fournit la structuration globale de la classification. Ces onglets peuvent être déterminés en fonction des éléments contextuels suivants :

- les besoins : variables à ajuster en fonction d'aspects contingents, comme la commodité, l'urgence, et en fonction des propriétés choisies des items que l'on cherche à décrire
- les usages : en fonction des choix et habitudes des utilisateurs à étudier pendant les phases d'entretiens
- le corpus : en fonction du type d'items à indexer : une homogénéité relative est nécessaire dans le corpus à classer à l'aide de l'outil. En effet, si l'on classe principalement des documents textuels, il est quand même possible d'avoir des documents vidéo à classer ; mais dans ce cas, la vidéo sera considérée comme un document et les propriétés spécifiques de ce type d'items ne seront pas décrites.

  Ainsi, si l'on classe principalement des films, ce type d'item possède des attributs spécifiques qui seront reflétés dans le schéma. Par exemple, un film a un réalisateur, un producteur, un monteur, un compositeur de musique, des acteurs, un genre (fantastiques, épouvante, comédie, etc.), il appartient éventuellement à un courant cinématographique, il relève d'un type (court, moyen, long métrage). Ces propriétés constitueront des facettes potentielles.

  Si l'on cherche à décrire des films avec des attributs propres à ces ressources, tandis que d'autres types d'items doivent aussi être classés (comme des documents textuels par exemple), il est possible de structurer, en partie, la classification en fonction des types d'items, ce qui aboutit à l'élaboration d'un onglet dédié à la documentarisation de ce type d'items particulier (les films).
- la structuration des activités : un métier suppose en général plusieurs grands types d'activité distincts. Par exemple, le métier d'enseignant-chercheur comporte des

activités administratives, d'enseignement, de recherche, des activités liées à des projets de recherches, et des activités liées à des événements scientifiques. Ainsi, il est possible de structurer la classification par onglets qui se rapportent à des grands types d'activités.

- la structure organisationnelle globale : en fonction des grands services dans l'organisation

5.5 Procédures de collecte de la méthodologie

Les types de procédures de collecte d'information utilisées lors de la mise en place de cette démarche sont les suivants :

- **Réunion en collectif global (tous les acteurs concernés par le déploiement)**

  Objectifs : les actions de pédagogie et d'explication du programme d'élaboration du dispositif et des principes de la classification à facettes

- **Réunion en collectif métier**

  Objectifs **:**
  - comprendre l'activité des acteurs, les pratiques de gestion et de circulation des documents
  - étudier ensemble la proposition de modèle de la classification proposée par le consultant, qui devra être adaptée par chaque acteur à ses besoins
  - recueillir des retours d'expérience sur les affordances de l'outil et les difficultés éventuellement rencontrées
  - obtenir des données, sur la base desquelles on détermine quelles dimensions seront partageables
  - finaliser les classifications pour chaque collectif considéré et préparer l'utilisation future, notamment le processus d'alimentation
  - optimiser les classifications

- **Entretiens individuels**
  Objectifs :
  - comprendre les différences dans les activités des différents acteurs d'un même métier,
  - comprendre les fonctionnements et habitudes de classement personnels adoptés

- **Elaboration individuelle de la classification à facettes avec l'outil**
  Objectifs :
  - évaluer la prise en main de l'outil
  - évaluer ce qui est retenu des efforts de pédagogie sur les principes de la classification à facettes

Chaque type de procédé donne lieu à la rédaction de compte-rendu et à l'élaboration de modèles itérativement construits, jusqu'à ce que les acteurs soient convaincus de la pertinence du schéma, de l'exhaustivité relative, et de la capacité de la classification à accueillir leurs nouveaux documents potentiels.

5.6 Déroulement de la méthodologie

Cette démarche méthodologique comprend trois phases successives, qui visent, d'une part à proposer un modèle pour les objectifs pédagogiques, d'autre part, à faire élaborer une classification à facettes individuellement par les acteurs impliqués dans la démarche, et enfin, à déterminer les dimensions structurelles et partagées des classifications pour la gestion des connaissances.

5.6.1 Première phase : entretiens – collecte – élaboration du support

Cette première phase est constituée de réunions, d'entretiens ou interviews semi-dirigés, afin d'identifier le domaine d'activité (secteur : énergie, écologie, etc.) et les types d'activités effectuées (conseil, veille, gestion de projet, etc.). Cette description du domaine global et des métiers de l'organisation se fait lors du premier entretien collectif, (réunion en collectif global) et continue durant les suivants (réunions en collectifs métiers et entretiens individuels), la description de l'activité devenant au fil des séances de plus en plus précise.
Ce que l'on cherche à comprendre correspond aux éléments suivants pour le type d'études B en collectif métier, et le type d'étude C en entretiens individuels :

- le métier ;
- les missions ;
- les objectifs ;
- les habitudes de travail.

Cette première phase peut se dérouler comme suit :

- Une première réunion introductive en collectif global

Cette réunion vise à commencer les actions de pédagogie sur Hypertagging et les actions d'explication sur le programme d'élaboration du système et sur les principes de la classification à facettes.

- Série de réunion en collectif métier : une réunion par métier

Pour comprendre le métier, les activités des acteurs et les pratiques de gestion et de circulation des documents, on tente de répondre aux questions suivantes et d'observer :

- en quoi consiste le métier :
  - que faites vous de manière récurrente dans votre journée de travail ?
  - et plus occasionnellement ?
  - quels sont vos interlocuteurs les plus réguliers ?
  - qu'échangez-vous avec eux ?
  - quels types de documents, d'informations ?
  - dans le cadre de quelle activité ?
  - quels problèmes rencontrez-vous de manière régulière et plus occasionnelle ?
- que partagent les acteurs et comment ?
  - y a-t-il des types de documents habituellement partagés ?
  - par quels moyens : mails, listes de diffusion, dossiers partagés ?
- observation des espaces éventuels de partage de documents :
  - leur organisation ;
  - comment est gérée la participation ;
  - les libellés de dossiers ;
  - les thèmes concernés ;
  - les types de documents ;
  - les titres des documents.

Ce type de réunion au début de la démarche vise aussi le recrutement des personnes motivées par la responsabilité d'animation de l'espace métier partagé.

- Série d'entretiens individuels avec des personnes-pivots, qui pourront être les responsables locaux de facettes, ou les correspondants pour la maintenance, des personnes-relais, (par ex, documentaliste, personnel responsable des archives, des personnels de la DSI). On mènera un entretien par personne. Cette série vise à :
  - continuer la pédagogie à l'aide de démonstrations avec un modèle générique à facettes, spécifique à un métier, déjà paramétré ;
  - préparer l'autonomie des acteurs, et préparer les fonctions d'animateur d'une partie de la classification (fonctionnalités avancées) ;
  - comprendre les fonctionnements et habitudes personnelles de classement et de gestion de l'information adoptés par les personnes interviewées.

- observation des postes informatiques individuels de travail :
    - leur organisation ;
    - les libellés de dossiers ;
    - les thèmes concernés ;
    - les types de documents ;
    - les titres des documents
- pendant l'observation, il faut relever des termes sur les activités évoquées, les types de documents, et d'autres sortes de termes (thèmes, cycles de l'activité, état des documents, sources etc.) à partir de ce qu'on observe dans le classement des postes individuels.

- Série d'entretiens individuels avec des personnes de l'entreprise à des postes gérant de l'information, et idéalement, avec au moins deux personnes de chaque métier.
  Ces entretiens visent à :
    - continuer la pédagogie à l'aide de démonstrations avec un modèle générique à facettes, spécifique à un métier, déjà paramétré ;
    - pour comprendre les différences dans les activités des différents acteurs d'un même métier ;
    - comprendre les fonctionnements et habitudes de classement personnels adoptés.
    - observation des postes informatiques individuels de travail :
        - leur organisation ;
        - les libellés de dossiers ;
        - les thèmes concernés ;
        - les types de documents ;
        - les titres des documents
    - pendant l'observation, il faut relever des termes sur les activités évoquées, les types de documents et d'autres sortes de termes (thèmes, cycles de l'activité, état des documents, sources etc.) à partir de ce qu'on observe dans le classement des postes individuels.

Chaque réunion ou entretien donne lieu à des prises de notes ou enregistrements, et à la rédaction d'un compte-rendu.

Cette première phase vise à récolter des informations sur l'activité et le langage utilisé par les acteurs des différents métiers observés, pour expliquer leurs activités. Il s'agit de comprendre aussi la logique de l'articulation des activités entre elles.

La prise de note doit permettre de réaliser des listes des activités évoquées par métier, ainsi que des listes des types de documents et d'autres sortes de termes (thèmes, cycles de l'activité, état des documents, sources, etc.) qu'il faudra faire entrer dans des catégories sur la base de similarités entre les éléments de ces listes.

Suite à cette première phase de la méthodologie, le consultant doit faire un travail de synthèse, et procéder à l'élaboration d'une proposition de schéma de classification pour chaque groupe métier.

En définitive, l'objectif de cette phase est la collecte d'informations sur l'activité et la collecte de termes pour produire une matière brute à analyse, rationnaliser, et organiser sous forme de facettes et de tags pour l'élaboration du modèle contextuel à chaque métier, qui sert de support à la suite du déroulement de la démarche méthodologique.

Cette phase d'entretiens peut se dérouler en omettant les observations des postes informatiques s'il existe, par exemple, une folksonomie concernant leurs documents de travail alimentée par les acteurs participants.

Cette phase pourrait aussi être réalisée à partir de l'extraction de termes issus de documents métier, bien que les entretiens restent nécessaires pour comprendre l'activité. Dans ce dernier cas, elle pourrait se rapprocher des méthodes de construction de thésaurus, qui commencent par une collecte de termes à partir de documents représentatifs du domaine. Cette modalité peut se rapprocher d'une expérimentation d'élaboration de thésaurus ad hoc que nous avons menée, qui est décrite dans (Desfriches Doria et Zacklad, 2010b) et qui est construite d'après une conception de l'indexation – extraction inspirée de (Amar, 2000).

Le modèle pré-paramétré peut se présenter comme suit, ici, pour une activité de veille. (Figure.52).

Figure 52. Exemple de modèle pré-paramétré à des fins pédagogiques

5.6.2 Deuxième phase : élaboration individuelle de FolkClassification avec Hypertagging

La deuxième phase consiste en l'élaboration individuelle d'une classification à facettes émergente à partir d'un jeu de documents avec Hypertagging par les futurs utilisateurs

Cette deuxième phase est élaborée d'après une conception ascendante reposant sur les principes du Web 2.0. Durant celle-ci, les participants vont construire individuellement au fil de l'eau, en même temps qu'ils réalisent leur activité de travail habituelle, une classification à facettes pour le classement de leurs documents. Ils doivent classer entre trente et cinquante documents (pour que les classifications soient significatives), avec l'outil et élaborer des facettes, des tags, ainsi que des onglets assurant la structuration globale de la classification au cours du processus de classement. Au départ, la classification est vide pour chaque utilisateur. Le modèle est fourni comme exemple aux participants de chaque métier, ainsi qu'un mode d'emploi et un guide méthodologique qui rappelle les principes de la classification à facettes

et les règles d'élaboration de facettes correctes. Ceci permet d'évaluer ce qui est compris ou non de l'usage de l'outil, en même temps que cela fournit des classifications individuelles au consultant lui permettant d'effectuer un travail de recoupement et de rapprochement des facettes qui peuvent traiter le même type de propriétés, et qui seront celles qui permettront un partage.

Cette phase est réalisée en autonomie par les acteurs, et sans intervention du consultant. Elle peut prendre entre deux et trois semaines. A l'issue de cette phase d'élaboration de classifications à facettes individuelles, le consultant récupère les bases des classifications, pour les étudier.

5.6.3 Troisième phase : détermination des dimensions partagées

Dans cette phase, le consultant recherche les dimensions communes des classifications individuelles produites par les acteurs.

Avant cette analyse, il faut vérifier que les facettes sont correctement construites, c'est à dire qu'elles respectent les deux règles suivantes :

- deux facettes ne peuvent pas posséder le même titre si elles ne renferment pas les mêmes valeurs ;
- un jeu de valeurs identiques dans deux facettes différentes (dont les titres sont différents) est une « erreur » du point de vue des principes de la classification à facettes.

Dans le cas où il y a deux facettes possédant le même titre mais avec des valeurs différentes, il faut modifier le titre d'une de ces facettes en cohérence avec les tags qu'elles contiennent et le type de propriété décrit par les valeurs de ces facettes.

Ensuite, le consultant étudie la structuration globale des classifications en onglets, une par une et en rapprochant les différents onglets pour les typer, selon qu'ils désignent, soit des activités, soit des dimensions de description des documents (par exemple des dimensions de description externe des documents, ou du contenu des documents), soit en fonction de types d'items spécifiques (films, documents textuels, photographies), ou encore des services de l'entreprise. D'autres éléments de structuration des onglets peuvent émerger également.

Le consultant observe ensuite les facettes et tente de déterminer le type de propriété décrite dans chaque facette de chaque onglet, de chaque classification, et il élabore des tableaux contenant des listes des propriétés décrites dans chaque classification et par onglet. Il faut prévoir des colonnes pour les titres des la facettes et pour mentionner les types d'onglets.
Si la propriété décrite par une facette est difficilement discernable, il faut alors observer les tags dans une démarche déductive, afin de trouver quelle est la propriété relative aux items décrite dans la facette.
Ce processus se poursuit par un travail de recoupement dans les tableaux, pour distinguer les types d'onglets communs et différents. En fonction des informations récoltées sur l'activité pendant les entretiens, qui ont donné une connaissance assez précise au consultant des activités métier par acteur, il faut tenter de donner un sens aux divergences concernant les activités, et conserver celles qui sont communes et qui constituent les dimensions réellement partagées du cœur de métier. Ce travail de recoupement est effectué sur les facettes également, pour proposer des facettes partageables par le collectif métier.
A ce stade, il n'est pas possible d'harmoniser les tags manuellement. Cette opération reste à effectuer pendant les réunions de finalisation, en dialogue avec les acteurs du métier. Un document de synthèse des aspects partageables est rédigé pour guider les prochaines réunions.
Après ce travail d'analyse et de synthèse, des réunions en collectif métier sont organisées, pendant lesquelles le consultant présente les classifications à facettes métier aux collectifs métiers concernés. Ces classifications sont des propositions ; il peut y en avoir plusieurs versions. Elles contiennent des onglets, des facettes, et quelques valeurs de facettes (celles jugées les plus pertinentes par le consultant) pour chaque facette, afin d'amorcer le processus d'alimentation.
Ces réunions en collectif métier ont donc pour objectif :

- de finaliser les classifications pour chaque collectif considéré, et préparer l'utilisation future dans le cadre de la gestion quotidienne individuelle ;
- de mettre en place le processus d'alimentation des dimensions partagées de la classification pour la gestion des connaissances ;
- d'optimiser les classifications.

Le déploiement final peut ensuite avoir lieu.

5.7 Conclusion

La définition de la facette que nous proposons implique une approche de la classification à facettes vue comme ascendante, puisque le processus d'élaboration des facettes commence par une collecte de termes candidats qui sont des valeurs de facettes potentielles, que l'on regroupe en fonction, d'une part, de ce qu'ils décrivent en commun parmi les propriétés des items à classer et d'autre part, en fonction de leurs caractéristiques communes, qu'on choisit pragmatiquement, en référence aux variations utilitaires des valeurs de facettes. Les facettes sont donc élaborées, dans un deuxième temps, après la collecte des termes candidats, et elles décrivent une propriété commune des items du corpus, bien que chaque item ne soit pas tenu de posséder toutes les propriétés décrites dans les facettes de la classification.

Ainsi, plutôt que d'appliquer un modèle de facettes préexistant à un corpus considéré comme unique, nous pensons que l'élaboration des classifications à facettes selon notre méthode permet de découvrir des dimensions non standardisées des items, qui sont pertinentes dans le contexte d'activités métier, exercées de manière particulière dans une organisation spécifique.

Il nous semble qu'à travers l'élaboration de cette méthodologie hybride entre une démarche descendante et ascendante (top-down et bottom-up), nous avons pu répondre à certaines questions théoriques relatives à la classification à facettes, soulevées dans la littérature. Ainsi, La barre (2006), mentionnait les questions théoriques suivantes à explorer :

1. Quelle est la nature basique d'une facette ?
2. Quelle est la valeur de l'exclusivité mutuelle dans une facette ?
3. Quelle est la valeur de l'homogénéité dans les facettes ?
4. Est-ce la notation apporte actuellement encore une valeur ?
5. Les facettes peuvent-elles être universelles ?

Ces questions nous ont guidés dans la conduite de ce travail. Nous avons pu proposer une définition de la facette, ainsi que nous positionner concernant l'universalité des facettes, que nous concevons comme essentiellement contextuelles, même si certaines structures de la classification, ou certaines facettes peuvent émerger de façon récurrente. De plus amples expérimentations sont à mener pour évaluer cette dernière affirmation. Concernant l'exclusivité mutuelle des facettes, elle est un des principes fondateurs de la classification à facettes et doit être conservée comme guide méthodologique. L'homogénéité des facettes est relative, elle doit se rapporter au type de propriété décrit dans la facette, mais doit tolérer des

variations sur les aspects sémantiques, pragmatiques, et grammaticaux, et enfin sur le niveau de granularité, tant que le titre de la facette est cohérent et englobe toutes les valeurs de la facette dans son niveau de généralité. Sur la notation, nous considérons que la numérisation des SOC ne la rend pas indispensable.

Nous discutons, finalement, les limites des classifications à facettes évoquées par (Kwasnik, 1999) et (Hudon et Mustapha El Hadi, 2010) en chapitre 1. La difficulté à établir les facettes appropriées nécessite un guidage méthodologique précis, auquel nous espérons avoir contribué. Concernant le manque de relations entre facette, nous considérons que cet aspect n'est pas nécessaire et fait perdre les qualités d'autonomie et d'indépendance des facettes entre elles, qualités qui permettent des croisements inédits de dimensions d'indexation et de critères de recherche. Des systèmes de classification à facettes intégrant des relations entre facettes existent (Marleau *et al.*, 2009), comme le modèle ISIS, mais nous considérons qu'il bride les possibilités de combinaison de valeurs de facettes. De plus, ce modèle formel ne laisse pas l'utilisateur libre de créer son schéma de classification, puisque celui-ci prend effet dans une démarche top-down, qui consiste en une analyse précise de la structure organisationnelle de l'entreprise ainsi que des business process qui sont modélisés.

Une autre difficulté avancée par les mêmes auteurs correspond à des problèmes de visualisation du schéma global. Grâce aux onglets, qui rassemblent les facettes en fonction d'un principe d'organisation et de structuration logique pour l'utilisateur et son activité telle qu'il l'exerce, l'utilisateur est en mesure de visualiser toutes les facettes dont il a besoin au moment de la documentarisation ou de la recherche de documents, si le schéma est bien élaboré.

Enfin, une autre problématique relevée par les mêmes auteurs est le risque de vision trop locale. Il s'agit d'un des problèmes auxquels la méthodologie que nous proposons tente d'apporter des réponses. Pour choisir un degré de précision ou de profondeur de l'indexation approprié, la recommandation principale consiste à tenter de garder la cohérence de la classification comme objectif lors de sa maintenance. Pour ce faire, l'utilisateur ne devrait pas se focaliser sur un document, mais l'insérer dans la dimension du corpus. Idéalement, lors de la création d'un tag, l'utilisateur doit d'abord vérifier qu'il n'y a pas un autre terme équivalent ou qui se rapproche de celui qu'il veut créer. La question type qu'il devrait se poser lors de la création d'un tag est la suivante : si je crée ce tag, est ce qu'il va me resservir pour d'autres documents ? Si la réponse à cette question est négative, il faut alors tenter de trouver une

formulation plus générique, car nous considérons qu'un cas d'utilisation unique d'un tag est d'un degré de précision trop élevé dans une classification à facettes. Le risque, si l'utilisateur ne respecte pas ce principe, est d'aboutir à une classification trop détaillée, trop précise, et inutilisable sur un moyen terme par sa densité et son ampleur.

La difficulté dans l'utilisation d'une classification à facettes adaptée au milieu de l'entreprise, telle qu'Hypertagging l'implémente, réside dans le fait de passer d'un niveau global d'organisation des connaissances (le niveau du schéma de la classification), à un niveau plus détaillé pour le renseignement des métadonnées propres à un document ou à une situation d'activité de travail, tout en veillant à ne pas générer trop de facettes, ou de valeurs de facettes, afin de garder la cohérence globale du schéma. En effet, deux opérations différentes sont demandées à l'utilisateur. Elles consistent à utiliser le système et à le maintenir en même temps, pour la pertinence de la classification. Cette dernière activité correspond à une activité méta-fonctionnelle.

Conclusion

Dans ce mémoire, nous avons cherché, d'une part, à caractériser les activités de gestion de l'information à partir des cadres théoriques portés par la Théorie de l'activité et les approches de la psychologie du travail et de l'ergonomie. Nous considérons ces activités, entre autres, comme des activités en souffrance ; elles manquent d'outils adaptables au contexte de travail évolutif, afin des les équiper. Du point de vue de l'acteur, elles constituent pourtant un moyen de remplir ses missions, tandis qu'elles composent le socle sur lequel repose l'efficacité au travail, la performance de l'entreprise, son développement, ainsi que la pérennisation de ses connaissances et sa durabilité juridique.

Ainsi se nouent différentes problématiques autour de ces activités de gestion de l'information, qui ont des objectifs différents. Globalement, du point de vue de l'acteur, il s'agit d'organiser son corpus de documents le mieux possible pour remplir ses missions et éventuellement de développer ses connaissances ; tandis que du point de vue de l'entreprise, les objectifs de contrôle, de traçabilité et de performance déterminent le plus souvent le choix des outils pour équiper ces activités.

Nous avons choisi d'étudier ces problématiques du point de vue de l'acteur, sans ignorer les problématiques gestionnaires, mais en cherchant plutôt à caractériser l'impact des Systèmes d'Information formels, de plus en plus répandus dans les organisations, sur les situations de travail des individus y travaillant. A travers cette caractérisation, nous avons pu déterminer les conceptions de l'activité impliquées par les choix de Systèmes d'Information.

Ainsi, nous pensons que les pratiques dans ces activités de gestion de l'information, qui sont vues comme des moyens au service des métiers d'une organisation, sont influencées par leurs équipements et l'environnement de travail. Par exemple, comme nous l'avons constaté dans le chapitre 2, le type d'organisation dans lequel est déployé un SOC influe sur les principes qui président à son élaboration. Dans ce chapitre, nous avons détaillé les méthodologies existantes pour les classifications à facettes, qui visent des objectifs différents selon les environnements dans lesquels elles sont déployées.

A cet égard, la méthodologie de Ranganathan apparaît systématique ou formelle, en partie parce que le corpus concerné est homogène. L'objectif unique est de permettre des recherches dans le fonds d'une bibliothèque. Celle de Vickery est plus adaptable puisqu'elle vise à

l'élaboration de classifications pour des fonds spécifiques. Les principes de celle-ci tendent donc vers une forme d'universalité de la méthode, tout en ménageant des possibilités de spécialisation. Les objectifs de la méthodologie GenISIS, s'orientent vers le Record Management, que nous rapprochons des impératifs et des problématiques de contrôle et de traçabilité. Elle se situe dans l'application d'un modèle générique d'organisation à des organisations spécifiques. Ainsi, le modèle de facettes est institué, il est adapté par le biais des modélisations qui sont des représentations schématiques des situations de travail. Par ailleurs, ces modélisations sont élaborées, en premier lieu, d'après des entretiens de haut niveau, avec les personnes qui conçoivent le travail prescrit. Les entretiens individuels servent à raffiner le modèle. La méthodologie de Denton qui vise la structuration de sites Web, semble adaptée à des corpus variés et de taille moyenne ou petite.

Les objectifs de notre méthodologie pour des FolkClassifications à facettes sont différents puisqu'ils visent à soutenir une démarche de gestion collective des connaissances métier, appuyée sur les activités de gestion de l'information personnelles, à l'aide de classifications à facettes émergentes, dans les entreprises.

De ce point de vue, la méthodologie la plus proche de celle que nous avons proposée est GenISIS, puisque l'environnement de déploiement visé est celui des organisations pour ces deux approches.

Mais la méthodologie GenISIS est conçue comme l'application d'un modèle de facettes génériques à n'importe quelle organisation, tandis que les FolkClassifications ambitionnent de refléter les besoins contextuels des acteurs d'un métier, exercé dans une organisation donnée. Les dimensions partagées dans Hypertagging, qui soutiennent la démarche de gestion des connaissances, proviennent des classifications individuelles et émergentes, mises en commun. La méthodologie que nous avons proposée se situe dans une logique d'apprentissage et de prise d'autonomie vis-à-vis de l'outil, quand GenISIS est dans une élaboration complexe, longue et coûteuse, basée sur des modélisations multiples dans une logique de prestation. Par ailleurs, GenISIS est fondée sur la stabilité des fonctions dans l'entreprise, tandis que les FolkClassifications prennent leurs racines dans les métiers.

Enfin, nous pensons que la nature émergente et progressive de la construction des FolkClassifications les rend plus aptes à intégrer les évolutions rapides de l'environnement dans lequel les connaissances sont produites dans les entreprises. Ainsi, dans GenISIS la maintenance et l'évolution de la classification nous apparaît constituer un problème complexe et coûteux.

Concernant la validité de notre méthodologie, d'autres expérimentations restent à conduire afin d'apporter des améliorations aux processus de collecte d'information, et de rapprochement des classifications individuelles.

Spiteri (2010) propose des critères d'évaluation pour des classifications à facettes dans les outils de tagging social. De la même manière, nous pensons que de futures recherches pourraient être menées sur les critères d'évaluation des FolkClassifications à facettes.

6 Bibliographie

Abel, F., Kawase, R. et Krause, D. (2010). Leveraging multi-faceted tagging to improve search in folksonomy systems (p. 299-300). Dans *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*. Récupéré le 16 septembre 2013 du site : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.4817&rep=rep1&type=pdf

Albrechtsen, H. (1993). Subject analysis and indexing: from automated indexing to domain analysis. *Indexer*, *18*. Récupéré le 1er octobre 2013 du site : http://www.theindexer.org/files/18-4/18-4_219.pdf

Ali, F. M., et Du, W. (2004). Toward reuse of object-oriented software design models. *Information and Software Technology*, *46*(8), 499-517.

Amar, M. (2000). *Les fondements théoriques de l'indexation: une approche linguistique*. Paris, France : ADBS.

Argyris C. (1995*). Savoir pour agir. Surmonter les obstacles à l'apprentissage organisationnel*, Paris, France : InterÉditions.

Autissier, D. et Lahlou, S. (1999). Les limites organisationnelles des TIC : Emergence d'un phénomène de saturation cognitive (p. 121 – 130). Dans *Actes de communication à la 4ème Conférence de l'Association Information et Management (AIM).*

Bakewell, K. G. B. (1993). Research in indexing: more needed?. *Indexer*, *18*, 147-151. Récupéré le 1er octobre 2013 du site : http://www.theindexer.org/files/18-3/18-3_147.pdf

Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y. et Shachak, A. (2008). Structured versus unstructured tagging : a case study. *Online Information Review*, *32*(5), 635-647.

Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science*, *46*(5), 327-339.

Begelman, G., Keller, P. et Smadja, F. (2006). Automated tag clustering : Improving search and exploration in the tag space. Dans *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Royaume-Uni,* (p. 15-33). Récupéré le 15 septembre 2013 du site : http://www.ra.ethz.ch/cdstore/www2006/www.rawsugar.com/www2006/20.pdf

Beghtol, C. (2008). From the universe of knowledge to the universe of concepts: the structural revolution in classification for information retrieval. *Axiomathes*, *18*(2), 131-144.

Beghtol, C. (2001). Relationships in classificatory structure and meaning, Dans C.A. Bean et R. Green (Dir.), *Relationships in the organization of knowledge.* Norwell, *MA:* Ed. Kluwer Academic Publishers.

Béguin, P. et Clot, Y. (2004). L'action située dans le développement de l'activité, *Activités, 1(2),* 27-49. Récupéré le 15 septembre 2013 du site de la revue : http://www.activites.org/v1n2/beguin.fr.pdf

Bergman, O., Beyth-Marom, R. et Nachmias, R. (2003).The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology*, *54*(9), 872-878.

Bernoux, P. (2009). *La sociologie des organisations* (p. 466). Paris, France : Ed. Seuil.

Bertrand-Gastaldy, S., Giroux, L., Lanteigne, D. et David, C. (1994). Les produits et processus cognitifs de l'indexation humaine (p. 29-40). *ICO Québec Printemps 1994, La gestion de l'information textuelle*, *6*(1). Récupéré le 1er octobre 2013 du site : http://www.ling.uqam.ca/sato/publications/bibliographie/Ico-94.htm

Bogers, T., Thoonen, W. et Van Den Bosch, A. (2006). Expertise classification: Collaborative classification vs. automatic extraction. *Advances in Classification Research Online*,

17(1), 1-20. Récupéré le 15 septembre 2013 du site de la revue : https://digital.lib.washington.edu/ojs/index.php/acro/article/viewFile/12488/10988

Boltanski, L. et Chiapello, E. (1999). *Le nouvel esprit du capitalisme.* Paris, France : Gallimard.

Boydens, I. (2011). Hiérarchie et anarchie. Dépasser l'opposition entre organisation centralisée et distribuée ? *Les Cahiers du numérique*, *6*(3), 77-101.

Briet, S. (1951). *Qu'est-ce que la documentation ?* (Vol. 1). Paris, France : Editions documentaires, industrielles et techniques.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. Dans *Aslib Proceedings*, *New Information Perspectives*, *(58)*1-2, 49-72

Broughton, V. (2001). Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multidimensional knowledge structures. *New review of hypermedia and multimedia*, *7*(1), 67-102.

Buckland, M. K. (2012). Obsolescence in subject description. *Journal of documentation*, *68*(2), 154-161. Récupéré le 1er octobre 2013 du site : http://people.ischool.berkeley.edu/~buckland/obsolsubject.pdf

Buckland, M. (1998). What is a digital document ?. *Document numérique*, *2*(2), 221-230.

Cahier, J. P. (2005). *Ontologies sémiotique pour le Web socio sémantique* (Thèse de Doctorat, Université de Technologie de Troyes, France). Récupéré le 1er octobre 2013 du site : http://tel.archives-ouvertes.fr/docs/00/14/91/93/PDF/cahier-phd-1.pdf

Castillo, O., Matta, N. et Ermine, J. L. (2004). De l'appropriation des connaissances vers l'acquisition des compétences (p. 1-8). Dans *Actes du 2ème Colloque du groupe de travail C2EI du GDR-MACS*.

Chaudiron, S. et Ihadjadene, M. (2010). De la recherche de l'information aux pratiques informationnelles. *Études de communication. Langages, information, médiations (35),* 13-30. Récupéré le 17 septembre 2013 du site : http://edc.revues.org/2257

Chaumier, J. (2002). Les systèmes documentaires. *Que sais-je ? (1419).* Paris, France : Presses Universitaires de France.

Christopher, A. (2011). A framework for contextual information in digital collections. *Journal of Documentation*, *67*(1), 95-143.

Clot, Y. (2008). *Travail et pouvoir d'agir*. Paris, France : Presses Universitaires de France.

Clot, Y. (1995). *Le travail sans l'homme. Pour une psychologie des milieux de travail et de vie, 2ème ed.* Paris, France : La Découverte.

Cotte D. (2007a). L'organisation des connaissances entre le formalisme des outils et la complexité des représentations : une illustration par le cas des portails d'entreprise (p.63-78). Dans *Actes du Sixième Congrès du chapitre français de l'ISKO, « Organisation des connaissances et société des savoirs : concepts, usages, acteurs »*. Récupéré le 30 septembre 2013 du site : http://www.isko-france.asso.fr/pdf/isko2007/Actes%20ISKO%20FR%202007%20p%2063-78.pdf

Cotte, D. (2007b). Espace de travail et logique documentaire. *Études de communication. Langages, information, médiations*, (30), 25-38. Récupéré le 17 septembre 2013 du site : http://edc.revues.org/436

Cotte, D. (2005). La structure comme index : le rôle des interfaces documentaires dans la construction du sens (p.105-114). Dans I. Timimi et S. Kovacs (Coord.), *Indice, index, indexation.* Lille, France : ADBS.

Crozier, M. et Friedberg, E. (1977). *L'acteur et le système.* Paris, France : Ed. Seuil.

Dalbin, S. et Guyot, B. (2007). Documents en action dans une organisation : des négociations à plusieurs niveaux. *Études de communication. Langages, information, médiations*, (30), 55-70. Récupéré le 17 septembre 2013 du site : http://edc.revues.org/467

Darses, F., Hoc, J. M. et Chauvin, C. (2004). Cadres théoriques et méthodes de production de connaissances en psychologie ergonomique (p.221-251). Dans J.M. Hoc et F. Darses (Dir.), *Psychologie ergonomique: tendances actuelles*. Paris, France : Presses Universitaires de France

David, C., Giroux, L., Bertrand-Gastaldy, S., Lanteigne, D. et Bertrand, A. (1995). Indexing as problem solving: a cognitive approach to consistency (p. 49-55). Dans *Proceedings of the annual meeting-American Society for Information Science (32).* Récupéré le 1er octobre 2013 du site : http://www.ualberta.ca/dept/slis/cais/david.htm

De Medeiros Franco, E., Aguiar Dutra A, et Dos Santos N. (2000). La gestion de la connaissance dans la construction civile : cartes cognitives comme stratégies pour la création de la connaissance (p.250 – 253). Dans B. Mélier et Y. Quéinnec (Coord.), *Communication et Travail, XXXVe congress de la SELF*. Toulouse, France : Ed. Octares.

Denton, W. (2003). *How to make a faceted classification and put it on the web.* Récupéré le 17 septembre 2013 du site : http://www.miskatonic.org/library/facet-web-howto.html

Desfriches Doria, O. et Zacklad M. (2010). Apports de psychologie du travail pour caractériser l'activité de gestion de l'information (p.53-69). Dans M. Ihadjadene, M. Zacklad et K. Zreik (Dir.), *Actes du 13ème Colloque International sur le Document Electronique (Cide 13), « Document numérique, entre permanence et mutations »*. Paris, France : Europia.

Desfriches Doria, O. et Zacklad M. (2010b). Améliorer la recherche d'information à l'aide de thésaurus « ad hoc », Expérimentations et réflexions méthodologiques. *Document Numérique (13)* 2, 13-40.

Desfriches Doria, O. (2012a). The role of activities awareness in faceted classification development, *Knowledge Organization (39)4*, 283-291.

Desfriches Doria, O. (2012b). Systems d'Organisation des Connaissances hétérogènes et pratiques de gestion de l'information. (p. 235-245). Dans Mustapha El Hadi W. (Dir.), *L'organisation des connaissances, dynamisme et stabilité.* Cachan, France : Lavoisier

Dieng, R., Corby, O., Golebiowska, J., Matta, N. et Ribiere, M. (2000). *Méthodes et outils pour la gestion des connaissances* (Vol. 2). Paris, France : Dunod.

Dix, A., Levialdi, S. et Malizia, A. (2006). Semantic halo for collaboration tagging systems. Dans *Proceedings of the Workshop on Social Navigation and Community-based Adaptation Technologies, at the Int. Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 06).* Récupéré le 16 septembre 2013 du site: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.2306&rep=rep1&type=pdf

Dourish, P. (2004). What we talk about when we talk about context. Dans *Personal and ubiquitous computing*, *8*(1), 19-30. Récupéré le 17 septembre 2013 du site : http://www.csc.kth.se/utbildning/kth/kurser/DH2622/mdifk11/inlamningsuppgifter/Dourish.pdf

Dudezert, A. (2012). Le KM au cœur de la stratégie d'entreprise. *Documentaliste-Sciences de l'Information*, *49*(2), 26-43.

Durampart, M. (2004). Les services informationnels – SIO et organisation du travail. *Sciences de la société (63).*151-164

Engeström, Y. (1999). Activity theory and individual and social transformation (p. 19-38). Dans Y. Engeström, R. Miettinen et R.L. Punamaki (Dir.), *Perspectives on activity theory*. New York, NY : Cambridge University Press.

English, J., Hearst, M., Sinha, R., Swearingen, K. et Yee, K. P. (2002a). Hierarchical faceted metadata in site search interfaces. Dans *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (p. 628-639). ACM. Récupéré le 15 septembre 2013 du site : http://flamenco.sims.berkeley.edu/papers/flamenco-shortpaper02.pdf

English, J., Hearst, M., Sinha, R., Swearingen, K. et Lee, K. P. (2002b). *Flexible search and navigation using faceted metadata*. Technical report, University of Berkeley, School of Information Management and Systems, 2003. Submitted for publication. Récupéré le 15 septembre 2013 du site : http://oc.ac.ge/pluginfile.php/451/mod_resource/content/0/_2_English_et_al_2003_Flexible_Search_and_Navigation_using_Faceted_Metadata.pdf

Ermine J.L. (2000), *Les systèmes de connaissances*, Paris, France : Hermes Science publications.

Fabre, I., Liquete, V. et Gardiès, C. (2011). Pratiques informationnelles et construction des savoirs dans une communauté professionnelle. *Les Enjeux de l'Information et de la Communication, (10),* 1-14. Récupéré 17 septembre 2013 du site : http://hal.archives-ouvertes.fr/docs/00/67/05/28/PDF/LesenjeuxGresec.pdf

Falzon, P. (1994). Les activités méta-fonctionnelles et leur assistance. *Le Travail Humain, 57 (1),* 1-23. Récupéré le 15 septembre 2013 du site : http://ergonomie.cnam.fr/equipe/falzon/articles_pf/activ_meta_fonct_assist_94.pdf

Falzon, P. et Sauvagnac, C. (2001). Mémoire organisationnelle : du recueil des savoirs à leur construction coopérative (p. 27-48). Dans M. Zacklad et M. Grundstein (Dir.), *Management des connaissances – Modèles d'entreprise et applications.* Paris, France : Hermès Science.

Farrell, S., Lau, T. et Nusser, S. (2007). Building communities with people-tags. Dans *Human-Computer Interaction–INTERACT 2007* (p. 357-360). Springer Berlin Heidelberg. Récupéré le 17 septembre 2013 du site : http://www.claymountain.com/svn/gradu/L%C3%A4hdeartikkelit/Johdanto/FLN07.pdf

Farrell, S. et Lau, T. (2006). Fringe contacts : People-tagging for the enterprise. Dans *Proceedings of the WWW'06 Collaborative Web Tagging Workshop*. Récupéré le 17 septembre 2013 du site : http://domino.watson.ibm.com/library/CyberDig.nsf/papers/53299B30AD986C78852571B0004F46A9/$File/rj10384.pdf

Feinberg, M. (2006). An examination of authority in social classification systems. *Advances in Classification Research Online*, *17*(1), 1-11.

Fidel, R. (1994). User-centered indexing. *Journal of the American Society for Information Science*, *45*(8), 572-576. Récupéré le 1er octobre 2013 du site : http://lis571-help.wdfiles.com/local--files/websites/user%20centered%20indexing.pdf

Gandon F. et Dieng-Kuntz R. (2005), Ontologie pour un système multi-agents dédié à une mémoire d'entreprise (Chapitre 7, p.141-163). Dans R. Teulier, J. Charlet et P. Tchounikine (Dir.), *Ingénierie des connaissances*. Paris, France : L'Harmattan.

Giboin, A. (2004). La construction de référentiel communs dans le travail coopératif (p. 119-139). Dans J.M. Hoc et F. Darses F. (Dir.), *Psychologie ergonomique: tendances actuelles*. Paris, France : Presses Universitaires de France.

Golder, S. A. et Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32(2)*, 198-208, Récupéré le 16 septembre 2013 du site : http://jis.sagepub.com/content/32/2/198.full.pdf

Grundstein, M. (2003). *De la capitalisation des connaissances au management des connaissances dans l'entreprise, Les fondamentaux du knowledge management*, Récupéré le 1er octobre 2013 du site : http://basepub.dauphine.fr/bitstream/handle/123456789/2620/INTKM0304_finalversion_modifiee.pdf?sequence=2

Grundstein, M. (2002). De la capitalisation des connaissances au renforcement des compétences dans l'entreprise étendue. Communication présentée au *1er Colloque du groupe de travail Gestion des Compétences et des Connaissances en Génie Industriel « Vers l'articulation entre Compétences et Connaissances »*. Récupéré le 1er octobre 2013 du site : http://basepub.dauphine.fr/bitstream/handle/123456789/4143/GCCGINantes0212v2.pdf?sequence=1

Guillaume, L.P. (2012). Les communautés de pratiques (CoP), Exploiter l'intelligence collective, *Documentaliste-Sciences de l'Information, (49)*2, 51-52

Guyot, B. (2009). Se mouvoir au sein du monde de l'information: comment les personnes parlent de leur activité d'information. *Études de communication. Langages, information, médiations.* (33), 101-118.

Guyot, B. (2006). *Dynamiques informationnelles dans les organisations*. Paris, France : Hermès Science publications.

Guyot, B. (2004). Eléments pour une approche informationnelle dans les organisations. *Sciences de la société*, (63), 11-26. Récupéré le 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/24/30/PDF/sic_00001094.pdf

Guyot, B. (2002). Une activité de travail méconnue: l'activité d'information. Dans *International Conference on Uses of Services and Telecommunications (ICUST)*. Récupéré 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/20/77/PDF/sic_00000095.pdf

Hayes, P. J. et Halpin, H. (2008). In defense of ambiguity. *International Journal on Semantic Web and Information Systems (IJSWIS)*, *4*(2), 1-18. Récupéré le 16 septembre 2013 du site : http://www.ltg.ed.ac.uk/~ht/HayesSlides.pdf

Hayman, S. (2007). Folksonomies and Tagging: New developments in social bookmarking. Communication présentée à la *Ark Group Conference: Developing and Improving*

Classification Schemes, Sydney, Australie. Récupéré du site le 15 septembre 2013 : http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=3A69AB07435F6E851A3331905535817E?doi=10.1.1.138.8884&rep=rep1&type=pdf

Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, *49*(4), 59-61. Récupéré le 15 septembre 2013 du site de la revue : http://dl.acm.org/citation.cfm?id=1121983

Heckner, M., Mühlbacher, S. et Wolff, C. (2008). Tagging tagging: analysing user keywords in scientific bibliography management systems. *Journal of digital information (JODI)*, *9*(2). Récupéré le 16 septembre 2013 du site : http://epub.uni-regensburg.de/6839/1/HeckneretalTaggingTaggingJoDI2008.pdf

Heymann, P. et Garcia-Molina, H. (2006). Collaborative creation of communal hierarchical taxonomies in social tagging systems. *InfoLab Technical Report*. Récupéré le 15 septembre 2013 du site : http://ilpubs.stanford.edu:8090/775/1/2006-10.pdf

Hjørland, B. (2013). Facet analysis: The logical approach to knowledge organization. *Information Processing & Management*. *49*(2), 545-557. Récupéré le 1er octobre 2013 du site : http://pure.iva.dk/ws/files/34910652/The_facet_analytic_approach_final.doc

Hjørland, B. (2012). Is classification necessary after Google ?. *Journal of Documentation*, *68*(3), 299-317. Récupéré le 1er octobre 2013 du site : http://pure.iva.dk/ws/files/33076611/Is_classification_necessary.doc

Hjørland, B. (2008). What is knowledge organization (KO) ?. *Knowledge Organization*, *35*(2-3), 86-101. Récupéré le 30 septembre 2013 du site: http://arizona.openrepository.com/arizona/bitstream/10150/106183/1/What_is_Knowledge_Organization_Finalrev_correctedC.doc

Hjørland, B. (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, *41*(1), 367-405. Récupéré le 1er octobre 2013 du site :

http://arizona.openrepository.com/arizona/bitstream/10150/105225/1/Semantics_and_KO_K.doc

Hjørland, B. (1992). The concept of 'subject'in information science. *Journal of documentation*, *48*(2), 172-200. Récupéré le 1er octobre 2013 du site : http://arizona.openrepository.com/arizona/bitstream/10150/105962/1/the%2520concept%2520of%2520subject.pdf

Hodge, G. (2000). Systems of knowledge organization for digital libraries. *Digital library federation, council on library and information resources.* Récupéré le 1er octobre 2013 du site : http://www.kevenlw.name/downloads/working/%E5%85%83%E6%95%B0%E6%8D%AE%E4%B8%8E%E4%BF%A1%E6%81%AF%E8%B5%84%E6%BA%90%E7%BB%84%E7%BB%87/reference/pub91.pdf

Hudon, M. et Mustapha El Hadi, W. (2010). Organisation des connaissances et des ressources documentaires. *Les Cahiers du numérique*, *6*(3), 9-38. Récupéré le 1er octobre 2013 du site : http://www.grico.fr/wp-content/uploads/2011/09/LCN6_3_04_Intro2011.pdf

Ihadjadene, M. et Chaudiron, S. (2009). Des processus aux pratiques: quels modèles informationnels pour analyser l'accès à l'information en contexte professionnel ?. Dans *Actes du Colloque Evolutions technologiques et information professionnelle: pratiques, acteurs et documents*. Récupéré le 17 septembre 2013 du site : http://halshs.archives-ouvertes.fr/docs/00/46/87/28/PDF/Ihadjadene-Chaudiron_GRESEC.pdf

Isaac, H., Campoy, E. et Kalika, M. (2007). Surcharge informationnelle, urgence et TIC. L'effet temporel des technologies de l'information, *Revue Management & Avenir (12)*, 153 – 172.

Jacob, E. K. (2004). Classification and categorization : a difference that makes a difference. *Library trends*, *52*(3), 515-540. Récupéré le 15 septembre 2013 du site : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.255&rep=rep1&type=pdf

John, A. et Seligmann, D. (2006). Collaborative tagging and expertise in the enterprise. Dans *Collab. Web Tagging Workshop in conj. with WWW2006*. Récupéré le 15 septembre 2013 du site : http://wwwiti.cs.uni-magdeburg.de/iti_db/lehre/Seminar%20Virtual%20Engineering/Paper/folksonomy2.pdf

Jones, W. (2010). *Keeping Found Things Found: The Study and Practice of Personal Information Management: The Study and Practice of Personal Information Management*. Burlington, MA: Ed. Morgan Kaufmann.

Khoo, C., Luyt, B., Ee, C., Osman, J., Lim, H.H. et Yong, S. (2007). How users organize electronic files on their workstations in the office environment : a preliminary study of personal information organization behaviour. *Information Research*, *12*(2), Récupéré le 17 septembre 2013 du site : http://dialnet.unirioja.es/servlet/articulo?codigo=2209542&orden=100441&info=link

Kipp, M. E. (2011). User, Author and Professional Indexing in Context : An Exploration of Tagging Practices on CiteULike. *Canadian Journal of Information and Library Science*, *35*(1), 17-48. Récupéré le 16 septembre 2013 du site : http://eprints.rclis.org/14867/1/kipp-2011-userauthorindexerlis-cjils351.pdf

Kipp, M. E. et Campbell, D. G. (2006). Patterns and inconsistencies in collaborative tagging systems : An examination of tagging practices. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1-18.

Kolmayer, E. et Peyrelong, M. F. (1999). L'émergence du document dans un processus de capitalisation des connaissances. Communication présentée aux *Journées d'Etudes Org&Co de la SFSIC, Les recherches en communication organisationnelle. Concepts et théorisations*. Récupéré le 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/20/81/PDF/sic_00000108.pdf

Kroski, E. (2005). *The Hive Mind: Folksonomies and User-Based Tagging*. Récupéré le 16 septembre 2013 du site : http://www.educause.edu/library/resources/hive-mind-folksonomies-and-user-based-tagging

Kwasnik, H.B**.** (1989). How a personal document's intended use or purpose affects its classification in an office (p. 207-210). Dans N.J. Belkin et C.J. van Rijsbergen (Eds.), *Proceedings of the 12th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '89).* New York, NY: ACM.

La Barre, K. (2011). A Semantic (Faceted) Web ?. *Les Cahiers du numérique*, *6*(3), 103-131.

La Barre, K. (2010). Facet analysis. *Annual Review of Information Science and Technology*, *44*(1), 243-284.

La Barre, K. (2006). *The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design* (Doctoral dissertation, faculty of the University Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in the School of Library and Information Science, Indiana University).

La Barre, K. (2004). *Faceted Maps of Knowledge and Domains :"Peeling the Onion of an Idea,"An Examination of the Use of Facet Analysis in Website Design* (Doctoral dissertation, Indiana University, Bloomington).

Lai, P. (2009). *Facette: using facets to improve tag-based bookmarking.* (Doctoral dissertation, Massachusetts Institute of Technology). Récupéré 17 septembre 2013 du site: http://dspace.mit.edu/bitstream/handle/1721.1/61303/702658480.pdf?sequence=1

Lansdale, M. W. (1988). The psychology of personal information management. *Applied ergonomics*, *19*(1), 55-66.

Leplat, J. (2008). *Repères pour l'analyse de l'activité en ergonomie*. Paris, France : Presses Universitaires de France.

Leplat, J. (1997). *Regards sur l'activité en situation de travail : contribution à la psychologie ergonomique*. Paris, France : Presses Universitaires de France.

Licoppe, C. (2008). Dans le «carré de l'activité»: perspectives internationales sur le travail et l'activité. *Sociologie du travail*, *50*(3), 287-302.

Lin, X., Beaudoin, J. E., Bui, Y. et Desai, K. (2006). Exploring characteristics of social classification. *Advances in Classification Research Online*, *17*(1), 1-19. Récupéré le 15 septembre 2013 du site : https://digital.lib.washington.edu/ojs/index.php/acro/article/viewFile/12491/10990

Lund, N. W. (2009). Document theory. *Annual Review of Information Science and Technology*, *43*(1), 1-55.

Mai, J. E. (2005). Analysis in indexing : document and domain centered approaches. *Information processing & management*, *41*(3), 599-611. Récupéré le 1er octobre 2013 du site : https://ualearn.blackboard.com/bbcswebdav/pid-197685-dt-content-rid-501286_1/courses/49608.201240/readings/Mai.pdf

Marchetti, A., Tesconi, M., Ronzano, F., Rosella, M. et Minutoli, S. (2007). Semkey: A semantic collaborative tagging system. Dans *Workshop on Tagging and Metadata for Social Information Organization at WWW* (Vol. 7, p. 8-12). Récupéré 16 septembre 2013 du site: http://wwwconference.org/www2007/workshops/paper_45.pdf

Marlow, C., Naaman, M., Boyd D. et Davis, M. (2006). HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, To Read. Dans *Proceedings of the seventeenth conference on Hypertext and hypermedia* (p.31-40). ACM. Récupéré le 15 septembre 2013 du site : http://pagesperso.lina.univ-nantes.fr/~prie-y/ens/06-07/M2R/articles/p31-marlow.pdf

Maniez, J. (1999). Des classifications aux thésaurus : du bon usage des facettes. *Documentaliste-Sciences de l'Information*, *36*(4-5), 249-260.

Marleau Y., Mas S. et Zacklad M. (2009). Exploitation des facettes et des ontologies sémiotiques pour la gestion documentaire. Dans E. Broudoux et G. Chartron (Dir.), *Actes de la deuxième conférence Document numérique et Société, Traitements et pratiques documentaires : vers un changement de paradigme ?*. Paris, France : ADBS.

Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication*, *47*(10). Récupéré le 15 septembre 2013 du site de l'auteur : http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf

Merholz, P. (2004). Metadata for the masses. *Adaptive path*, *19*. Récupéré le 16 septembre 2013 du site de l'auteur : http://www.adaptivepath.com/ideas/e000361

Mervis, C. B. et Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, *32*(1), 89-115. Récupéré le 15 septembre 2013 du site de la revue : http://www.annualreviews.org/doi/abs/10.1146/annurev.ps.32.020181.000513?journalCode=psych

Millen, D. R., Feinberg, J. et Kerr, B. (2006). Dogear : Social bookmarking in the enterprise. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (p. 111-120). ACM. Récupéré le 15 septembre 2013 du site de la revue : http://dl.acm.org/citation.cfm?id=1124792

Mills, J. (2004). Faceted classification and logical division in information retrieval. *Library trends*, *52*(3), 541-570. Récupéré le 17 septembre 2013 du site : https://www.ideals.illinois.edu/bitstream/handle/2142/1035/L?sequence=2#page=173

Milstead, J. L. (1994). Needs for research in indexing. *Journal of the American Society for Information Science*, *45*(8), 577-582.

Mitev, N. (2008). Un tour d'horizon de la recherche anglophone sur les systèmes d'information dans les organisations (p. 15-29). Dans *L'information dans les organisations : dynamique et complexité.* Tours, France : Presse Universitaires François-Rabelais.

Monnin, A. (2009). Qu'est-ce qu'un tag ? Entre accès et libellés, l'esquisse d'une caractérisation. Dans *Actes de la conférence Ingénierie des Connaissances (IC) 2009 :*

20es Journées Francophones d'Ingénierie des Connaissances, "Connaissance et communautés en ligne". Récupéré le 16 septembre 2013 du site : http://halshs.archives-ouvertes.fr/docs/00/37/74/10/PDF/Monnin_IC2009_41.pdf

Musnik, N. (2012). *La gestion de l'information en contexte : Enquête sur les pratiques informationnelles des ingénieurs-chercheurs d'EDF-R&D* (Thèse de Doctorat, Conservatoire National des Arts et Métiers - CNAM, Paris, France.).

Nardi, B. A. (1996). Activity theory and human-computer interaction. Dans *Context and consciousness: Activity theory and human-computer interaction*, 7-16. Cambridge : Mass. : MIT Press. Récupéré 17 septembre 2013 du site : http://www.ics.uci.edu/corps/phaseii/nardi-ch1.pdf

Nicotri, V.B. (2004). Systèmes d'information organisationnels et jugement situé. *Sciences de la société (63)*. 57-74.

Nonaka, I. et Takeuchi, H. (1997). *La connaissance créatrice : la dynamique de l'entreprise apprenante*. Bruxelles, Belgique : De Boeck.

Nyckees, V. (1998). *La sémantique.* Paris France : Belin.

O'Reilly, T. (2007).What is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 65(1), 17-38, Récupéré 16 septembre 2013 du site: http://ssrn.com/abstract=1008839

Otlet, P. (1934). *Traité de documentation : le livre sur le livre, théorie et pratique*. Bruxelles, Belgique : Ed. Mundaneum.

Osty, F. (2002). *Le désir de métier : engagement, identité et reconnaissance au travail*. Rennes, France : Presses Universitaires de Rennes.

Osty, F. 2006. Sociologie des professions et des métiers. Dans (Dir.) Alter N., *Sociologie du monde du travail.* (p. 103 – 119). Paris, France : Presses Universitaires de France.

Pédauque, R. T. (2006). *Document et modernités*. Récupéré le 15 septembre 2013 du site: http://archivesic.ccsd.cnrs.fr/docs/00/06/28/26/PDF/Pedauque3-V4.pdf

Pédauque, R. T. (2003). *Document : forme, signe et médium, les re-formulations du numérique*. Récupéré le 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/21/99/PDF/sic_00000511.pdf

Pesqueux, Y. (2002) *Organisations : modèles et représentations*. Paris, France : Presses Universitaires de France.

Peterson, E. (2006). Beneath the metadata: Some philosophical problems with folksonomy. *D-Lib Magazine*, *12*(11). Récupéré le 16 septembre 2013 du site : http://dare.uva.nl/document/448330

Peyrelong, M. F. (2002). Apports et implications de l'approche par le document (p.195-202). Dans *Actes du 13ème congrès de la SFSIC (Société Française des Sciences de l'Information et de la Communication)*. Récupéré le 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/26/42/PDF/sic_00001516.pdf

Pikas, C. K. (2007). Personal information management strategies and tactics used by senior engineers. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1-21. Récupéré le 1er octobre 2013 du site : http://arizona.openrepository.com/arizona/bitstream/10150/106236/1/Final_Pikas_PIM_Senior_Engineers_ASIST_2007.pdf

Poitou, J. P. (2007). Des techniques de gestion des connaissances à l'anthropologie des connaissances. *Revue d'anthropologie des connaissances*, *1*(1), 11-34.

Prével, P. (2012). Pratiques et méthodes du KM. *Documentaliste-Sciences de l'Information*, *49*(2), 44-61.

Prié, Y. (2000). Sur la piste de l'indexation conceptuelle de documents. *Document numérique*, *4*(1-2), 11-35. Récupéré le 1er octobre 2013 du site : http://pagesperso.lina.univ-nantes.fr/~prie-y/download/prie-dni.pdf

Quintarelli, E. (2005). Folksonomies: power to the people. *International Society for Knowledge Organization Italy - UniMIB Meeting*. Récupéré le 16 septembre 2013 du site : http://www.iskoi.org/doc/folksonomies.htm

Quintarelli, E., Resmini, A. et Rosati, L. (2007). Information architecture : Facetag : Integrating bottom-up and top-down classification in a social tagging system. *Bulletin of the American Society for Information Science and Technology*, *33*(5), 10-15. Récupéré le 15 septembre 2013 du site de l'auteur : http://facetag.org/download/facetag.pdf

Ranganathan, S. R. (1967). *Prolegomena to library classification. Vol. 1, (3ème ed.).* Dehli, Inde : Ess Ess Publication pour Sarada Ranganathan Endowment for Library Science.

Ranganathan, S. R. (1951). *Classification and communication.* Dehli, Inde : Ess Ess Publication pour Sarada Ranganathan Endowment for Library Science.

Recanati, F. (2008). *Philosophie du langage (et de l'esprit).* Paris, France : Gallimard.

Reix, R. (2002). Systèmes d'information et performance de l'entreprise étendue. Dans F. Rowe (éd.), *Faire de la recherche en systèmes d'information.* Paris, France : Vuibert.

Rosch, E. (1999). Principles of categorization. Dans E. Margolis et S. Laurence (dir.) *Concepts : core readings*, 189-206. Cambridge, Mass : MIT Press

Roux, A. (2004). Système d'information et dispositif informationnel: quelle articulation ?. *Sciences de la société*, (63), 27-39.

Salembier, P. et Zouinar, M. (2004). Intelligibilité mutuelle et contexte partagé, Inspirations conceptuelles et réductions technologiques, *Activités*, *1(2)*, 64-85. Récupéré le 15 septembre 2013 du site de la revue : http://www.activites.org/v1n2/salembier.pdf

Sauvagnac, C., Falzon, P. et Leblond, R. (2000). La mémoire organisationnelle : reconstruction du passé, construction du futur (p. 442-450). Dans M. Zacklad, M., G. Kassel G., *et al.* (Dir.), *Ingénierie des connaissances: évolutions récentes et nouveaux défis*. Paris, France : Eyrolles.

Schraefel, M.C. (2009). Building Knowledge: what's Beyond Keyword search ?. 52-59. IEEE. Récupéré le 2 octobre 2013 du site : http://stono.cs.cofc.edu/~bowring/classes/csis%20633/readings/BuildingKnowledgeBeyondSearch.pdf

Sellin K. (2012). Les cartographies de connaissances pour une organisation 2.0 : la démarche de « transfert de savoir-faire » de Total, *Documentaliste-Sciences de l'information, (49)*2, 48-50.

Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J. et Riedl, J. (2006). Tagging, communities, vocabulary, evolution (p.181-190). Dans *Proceedings of the 2006 20th anniversary conference on Computer Supported Cooperative Work* ACM. Récupéré le 15 septembre 2013 du site : http://cleo.ics.uci.edu/teaching/Winter10/231/readings/5b-Sen-TaggingCommunities-CSCW.pdf

Serdyukov, P., Taylor, M., Vinay, V., Richardson, M. et White, R. W. (2011). Automatic people tagging for expertise profiling in the enterprise (p. 399-410). Dans *Advances in Information Retrieval*. Springer Berlin Heidelberg. Récupéré le 15 septembre 2013 du site : http://research.microsoft.com/en-us/um/people/ryenw/papers/SerdyukovECIR2011.pdf

Shirky, C. (2005). *Ontology is Overrated : Categories, Links, and Tags.* Récupéré du site de l'auteur le 15 septembre 2013 : http://www.shirky.com/writings/ontology_overrated.html

Shirky, C. (2004). *Folksonomy*. Récupéré du site de l'auteur le 15 septembre 2013 : http://www.corante.com/many/archives/2004/08/25/folksonomy.php

Smith, T. (2007). *Cataloging and you : Measuring the efficacy of a folksonomy for subject analysis.* Récupéré le 16 septembre 2013 du site : http://arizona.openrepository.com/arizona/handle/10150/106434

Souza, R. R., Tudhope, D. et Almeida, M. B. (2012). Towards a taxonomy of KOS : Dimensions for classifying Knowledge Organization Systems. *Knowledge Organization, 39*(3), 179-192. Récupéré le 1er octobre 2013 du site : http://mba.eci.ufmg.br/downloads/Souza_Tudhope_Almeida_-_KOS_Taxonomy.Submitted.pdf

Spiteri, L. F. (2010). Incorporating facets into social tagging applications : An analysis of current trends. *Cataloging & Classification Quarterly, 48*(1), 94-109.

Spiteri, L. F. (1998). A simplified model for facet analysis. *Canadian journal of information and library science, 23* (1-30). Récupéré 17 septembre 2013 du site: https://courses.washington.edu/is530/win09b/IMT%20530b_09/spiteri.pdf

Spiteri, L. F. (1995). The Classification Research Group and the theory of integrative levels. *The Katharine Sharp Review, 1*.

Tennis, J. T. (2006). Social tagging and the next steps for indexing. *Advances in Classification Research Online, 17*(1), 1-15. Récupéré le 15 septembre 2013 du site de la revue : http://journals.lib.washington.edu/index.php/acro/article/viewFile/12493/10992

Teulier, R. et Girard, N. (2005). Modéliser les connaissances pour l'action dans les organisations (Chapitre 18, 389-412). Dans R. Teulier, J. Charlet et P. Tchounikine (Dir.), *Ingénierie des connaissances.* Paris, France : *L'Harmattan*. Récupéré le 1er octobre 2013 du site : http://hal.archives-ouvertes.fr/docs/00/26/30/19/PDF/2006-01-23-979.pdf

Theureau, J. (2004). L'hypothèse de la cognition (ou action) située et la tradition d'analyse du travail de l'ergonomie de langue française, *Activités, 1(2),* 11-25. Récupéré le 15 septembre 2013 du site de la revue : http://www.activites.org/v1n2/vol1num2.book.pdf

Thevenot, D. (1998). *Le partage des connaissances: une mémoire interactive pour la compétitivité de l'entreprise*. Paris, France : SEP.

Tort F., Teulier R., Grozs G. et Charlet J. (2005). Ingénierie des besoins, ingénierie des connaissances : complémentarité des approches de modélisation (p. 207-228). Dans R. Teulier, J. Charlet et P. Tchounikine (Dir.), *Ingénierie des connaissances.* Paris, France : *L'Harmattan.* Paris, France: Eds, L'Harmattan.

Trant, J. (2009). Studying Social Tagging and Folksonomy : A Review and Framework. *Journal of Digital Information, 10(1),* récupéré du site de la revue : http://journals.tdl.org/jodi/index.php/jodi/article/view/269

Vacher, B. (1998). Les enjeux de la manutention de l'information. *Système d'information et Management, 3*(2), 65-83.

Van Den Abeele, M. (2012). Partage réciproque de savoirs en entreprise, *Documentaliste-Sciences de l'Information, (49)*2, p. 40

Vander Wal, T. (2005). *Folksonomy Definition and Wikipedia*. Récupéré du site de l'auteur le 15 septembre 2013 : http://www.vanderwal.net/random/entrysel.php?blog=1750

Van Setten, M., Brussee, R., van Vliet, H., Gazendam, L., van Houten, Y. et Veenstra, M. (2006). On the importance of "who tagged what". Dans *Proceedings of the Workshop on the Social Navigation and Community based Adaptation Technologies at AH* (Vol. 6). Récupéré le 16 septembre 2013 du site : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.3609&rep=rep1&type=pdf

Vickery, B. C. (2008). Faceted classification for the web. *Axiomathes, 18*(2), 145-160.

Vickery, B. C. (2008). *The structure of subject classification for document retrieval*, Récupéré le 1er octobre 2013 du site : http://www.iskoi.org/ilc/vickery.php

Vickery, B. C. (1966). *Faceted Classification schemes.* New Brunswick, New Jersey : The Rutgers University Press.

Vickery, B. C. (1963). *La classification à facettes, Guide pour la construction de schémas spéciaux* (Salvan, P., trad.). Paris, France : Ed. Gauthier-Villars (Ouvrage original publié en 1960 sous le titre *Faceted classification, a guide to construction and use of special schemas).*

Voss, J. (2007). Tagging, folksonomy & co-renaissance of manual indexing ?. Récupéré le 15 septembre 2013 du site : http://arxiv.org/abs/cs/0701072

Weick, K. E., Sutcliffe, K. M. et Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, 16(4), 409–421.

Weinberger, D. (2006). What is a folksonomy anyway ?. *Joho the Blog.* Récupéré le 16 septembre 2013 du site: http://www.hyperorg.com/blogger/2006/11/19/beneath-the-metadata-a-reply/

Wilson, T. (2006). The strict faceted classification model. *Information architecture summit, Vancouver*, 23-27. Récupéré le 15 septembre 2013 du site de l'auteur : http://facetmap.com/pub/strict_faceted_classification.pdf

Wilson, T. D. (2002). The nonsense of knowledge management. *Information research*, *8*(1). Récupéré le 1er octobre 2013 du site : http://informationr.net/ir/8-1/paper144.html?referer=www.clickfind.com.au

Wild, P. J., Giess, M. D. et Mc Mahon, C. A. (2009). Describing engineering documents with faceted approaches: observations and reflections. *Journal of Documentation*, *65*(3), 420-445.

Wu, H., Zubair, M. et Maly, K. (2006). Harvesting social knowledge from folksonomies (p. 111-114). Dans *Proceedings of the seventeenth conference on Hypertext and Hypermedia* ACM. Récupéré le 15 septembre 2013 du site de la revue : http://dl.acm.org/citation.cfm?id=1149962

Yee, K. P., Swearingen, K., Li, K. et Hearst, M. (2003). Faceted metadata for image search and browsing (p.401-408). Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. Récupéré le 15 septembre 2013 du site :http://kevinli.net/flamenco-chi03.pdf

Zacklad, M. (2012). Organisation et architecture des connaissances dans un contexte de transmédia documentaire : les enjeux de la pervasivité. *Études de communication. Langages, information, médiations*, (39), 41-65.

Zacklad, M. (2011). Évaluation des systèmes d'organisation des connaissances. *Les Cahiers du numérique*, *6*(3), 133-166. Récupéré le 17 septembre 2013 du site : http://www.zacklad.org/articles_soc_web2/Evaluation%20des%20SOC.pdf

Zacklad, M. (2010a). Introduction aux appareils de capture sémantique dans les TIC et les SI. Dans B. Juanals et J.M. Noyer (Dir.), *Technologies de l'information et intelligences collectives*, Paris, France : Hermès Science Publications. Récupéré le 17 septembre 2013 du site : http://www.zacklad.org/pages/appareil%20capture%20V7.doc

Zacklad, M. (2010b). Les écritures de l'information dans les TIC et le SI. Communication présentée au *Colloque « Information Science: Synergy through Diversity » de la Canadian Association for Information Science CAIS/ACSI.* Récupéré le 17 septembre 2013 du site de l'auteur: http://www.cais-acsi.ca/proceedings/2010/CAIS104_Zacklad_Final.pdf

Zacklad, M. (2008). Ordre du discours : cadrage et recadrage des transactions communicationnelles dans les processus de changement et d'innovation au sein des organisations. *Matérialité, action et discours, (21).* Récupéré le 17 septembre 2013 du site :

http://www.zacklad.org/articles_transactions_cooperatives/Taylor%20Court%208%20REPRISE-12.pdf

Zacklad, M. (2007a). Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI). Arsenaul et K. Dalkir (Dir), Actes du *35e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté : Franchir les frontières.* Récupéré le 1er octobre 2013 du site : http://hal.archives-ouvertes.fr/docs/00/20/24/40/PDF/cais-acsi_zacklad_-_avec_ref.pdf

Manuel Zacklad (2007b), Une théorisation communicationnelle et documentaire des TIC. Dans Reber, B., Brossaud, C., (Dir), *Humanités numériques 2 : socio-informatique et démocratie cognitive (Traité IC2, série cognition et traitement de l'information)*, Hermes Science Publications, Londres-Paris.

Zacklad, M. (2006). *Une approche communicationnelle et documentaire des TIC dans la coordination et la régulation des flux transactionnels*. Récupéré le 17 septembre 2013 du site : http://hal.archives-ouvertes.fr/docs/00/11/32/72/PDF/working_paper_flux_transactionnels.pdf

Zacklad, M. (2005). Processus de documentarisation dans les Documents pour l'Action (DopA): statut des annotations et technologies de la coopération associées (nouvelle version corrigée). Dans *Actes du colloque « Le numérique: Impact sur le cycle de vie du document pour une analyse interdisciplinaire », 13-15 Octobre 2004, Montréal*, Québec : Ed. ENSSIB. Récupéré le 17 septembre 2013 du site : http://archivesic.ccsd.cnrs.fr/docs/00/06/26/25/PDF/sic_00001459.pdf

Zacklad, M., Desfriches-Doria, O., Bertin, G., Mahé, S., Ricard, B., Musnik, N.(...) Lewkowicz, E. (2011). Miipa-Doc: Gestion de l'hétérogénéité des classifications documentaires en entreprise. Communication présentée à *Conférence Hypertextes et hypermédias. Produits, Outils et Méthodes (H²PTM) 2011*. Récupéré le 1er octobre 2013 du site : http://mipadoc.org/wp-content/uploads/2012/01/H2PTM-Miipa-Doc-ANR-Project.pdf

Zacklad, M., Bénel A., Zaher L.H., Lejeune C., Cahier, J.P. et Zhou C. (2007). Hypertopic: une métasémiotique et un protocole pour le Web socio-sémantique (p. 217-228). Dans F. Trichet (Coord.), *Actes des 18èmes journées francophones d'ingénierie des connaissances (IC2007)*, Paris, France : Cépaduès.

Zollers, A. (2007). Emerging motivations for tagging: Expression, performance, and activism. Dans *Workshop on Tagging and Metadata for Social Information Organization, présenté à the 16th International World Wide Web Conference*. Récupéré le 16 septembre 2013 du site : http://wwwconference.org/www2007/workshops/paper_55.pdf

7 Annexes

Annexe 1 : Exemples de types d'information mentionnés dans les intitulés des dossiers :

- Thème de recherche : moteur de recherche/ multimédia
- Type de document : présentation, thèse, ITV, livrables, contrat
- Objet de l'étude : outils
- Cycle de vie du document : archives, ancien
- Application spécifique : « maquette Carto »
- Tâche spécifique : « réalisation outil »
- Types de fichiers : « fichiers VME »
- Type de code : « affichage »
- Nom d'un lot
- Contenu d'un lot
- Mode de recueil de données : « Fil de l'eau »
- Nom partenaire
- Division organisationnelle
- Abréviation / Code
- Échéance/ étape du projet
- Pense-bête/ mémo: « autre biblio_a_voir »
- Nom de personne
- Titre explicite : « charte_graphique_edf_2008 »
- Type de document/ activité: « pré_etude_architecture »
- Nom de BDD

Annexe 2 : Liste des facettes issues de la collecte

FACETTES pour l'onglet métier

Mots outils	Activités	objectifs	Outils	catégories de thèmes	Thèmes
méthode	capitalisation de connaissances	collaboration	ontologies	aspects temporels	génération
outils	partage de connaissances	soutien	PAGODDE		vieillissement
solution	optimisation système	performance	CAPCOV		pérennité
indicateurs	appui à exploitation	gestion	visio-conférence		durée de vie
usages	animation	inspection	supports interactifs		cycle
enjeu	expérimentation	conduite	wikis		perspectives
objectif	industrialisation	génération	RS	apsects contextuels	aspects socio-organisationnels
gain	planification	Expertise	SI décisionnels		aspects humains
proces	automatisation	prévision	CASSIOPEE		pratiques métier
procédure	simulation	économie			usages
pratques	diagnostic	compétitivité			contexte
cible	formation	flexibilité		information	recherche d'information
besoin	modelisation	efficacité			indexation
savoir faire	organisation	programmation			connaissances
expérience	anticipation	réduction			datamining
	spécification	renouvellement			mutlimédia
	essai	stockage			SI
	maintenance	mise à jour			NTIC
	production	surveillance			informatique
	communication	standardisation			évolution terminologie
	décision	restitution			information
	prospection	rédaction		technologie	nouvelles technologies
	évaluation	élaboration			neutronique
	veille	qualification			ultrason
	estimation	approbation			terminal
	faisabilité	emission		enjeux	mobilité
	régulation	diffusion			actifs
	contrôle	vérification			accessibilité
	amélioration	précision			marché
	préparation	mesure			règlementation
	réalisation	commande			patrimoine
	analyse	évolution			disponibilité
	configuration	échange			opportunité
	test	prévention			commerce
	ajustement	réfection			consommation
	innovation	reconstruction		dysfonctionnements	risques
	intégration	direction			surcharge
	traitement	déconstruction			criticité
	fonctionnement	ingénierie			confort
	suivi				coupure/ panne
	sécurisation			matériel	parc nucléaire
	pérennisation				ouvrage
	rénovation				installation
	pilotage				tranche
	accompagnement				réseau électrique
	développement				centrales
					rareté
					équipement
					combustible
					transformateur
					cuves REP
					fours
					piles
					liaison souterraines

				environnement	renouvelable
					environnement
					sûreté
					développement durable
					sédiments
					pollution
					risques climatiques
					risques fuite
					contraintes climatiques
					eau
					impact environnemental
					local
					batiment
					impact sanitaire
					pays en développement
				entreprise EDF	marketing
					groupe
				qualité	expérience de terrain
					pertinence
					critère
					qualité
					effet
				concret	mécanique
					chaleur
					procédés industriels
					manœuvre
				physique	systèmes dynamiques
					radio-protection
					signal
					images
					pays en développement
					rareté
					renouvelable

s

onglet Projet

FACETTES POUR L'ONGLET PROJET

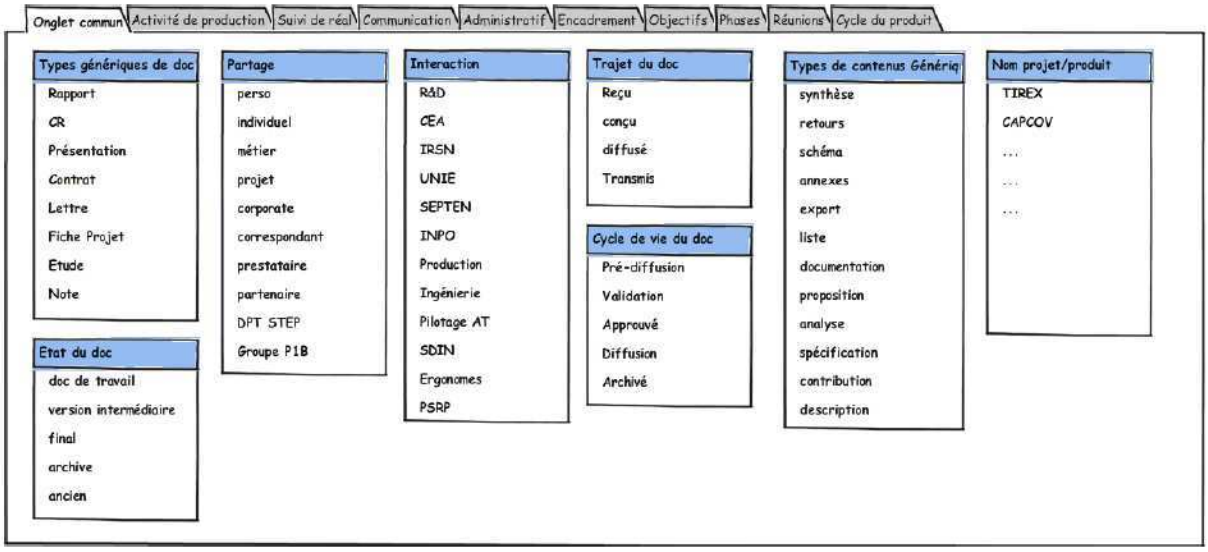
Defi	métier (à détailler)	projet	phases du projet	Nom et numéro de lot	compétences	statut (de l'indexeur)	type d'installation	sites	terrains	partenariat académiques	partenariat industriel	partenariat interne	clients
D01	production électricité	[illegible]	enclenchement		TSI	ingénieur	hydraulique	[illegible]	arret de tranche	ANR	AIRBUS	LNHE	RTE
D02	management énergie	[illegible]	conception		IK	ingénieur chercheur	nucléaire	[illegible]	REX	SUPELEC	ANDRA	ICAME	DPI
D03	réseaux électriques/transport	images aéroportées	clôture		SD	ingénieur expert	photovoltaïque	[illegible]	usine	Ircyn	CEA	EIFER	DIN
D04	développement commercial/distribution	CapCoV	installation			stagiaire	thermique	[illegible]		EPM		EPI	CNEN
D05	énergies renouvellables/environnement	[illegible]	intervention			prestataire		[illegible]		GIPSA-LAB		EFESE	CNPE
D06	technologies de l'information	[illegible]				management		[illegible]		INRIA		LME -TSI	CIPN
D07		[illegible]						[illegible]		ENST		MFEE	SEPTEN
D08		[illegible]						[illegible]		CNAM		MIRE	CIDEN
D09		[illegible]						[illegible]		UTT		MMC	CEIDRE
D10		[illegible]						[illegible]		OSAKA		MRI	DPN
D11		[illegible]						[illegible]		KOBE		OSIRIS	UTO
D12		[illegible]						[illegible]		KYOTO		SINETICS	UNI
D13		[illegible]								HALDEN		STEP	DCN
D14		[illegible]								EPRI		compétence SI	DPIH
D15		[illegible]								CNRS			DPIT
D16		[illegible]											DTG
D17		[illegible]											CIH
D18		[illegible]											COIT
D19		[illegible]											DAIP
D20		[illegible]											DCRG
D21		[illegible]											DOAAT
D22		[illegible]											ERDF
D23		[illegible]											DC
D24 INTEP		prévision production photovoltaïque											

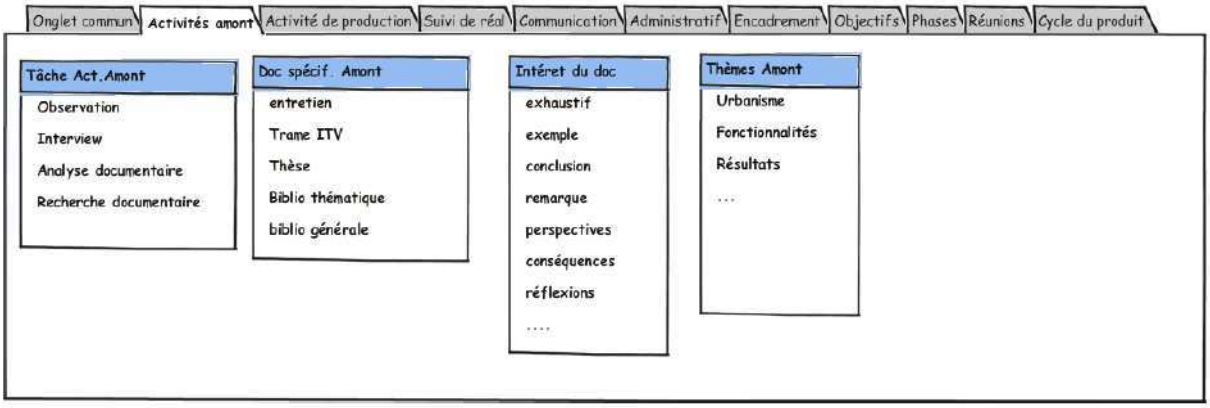
FACETTES pour l'onglet document

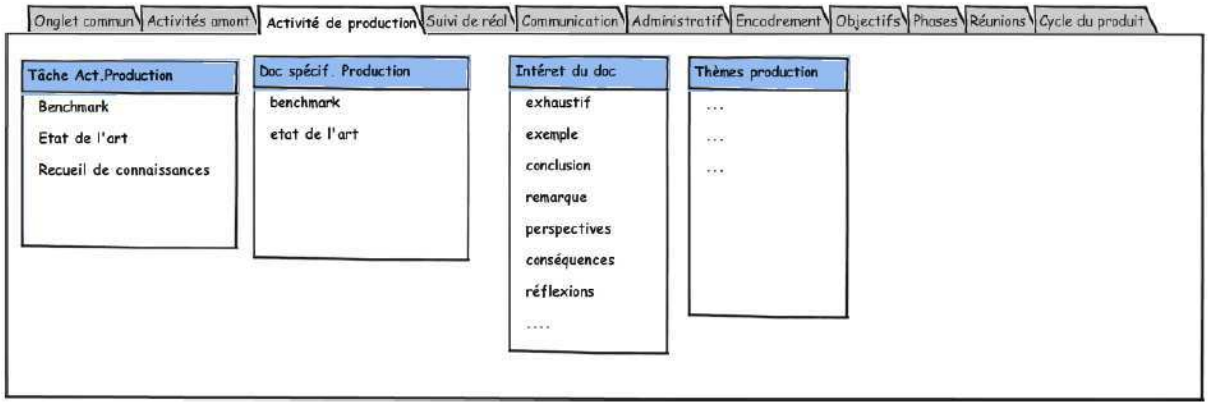
types de docs	opérations sur les docs	version	accessibilité
schémas d'exploitation	validation	ancienne	EDF
guide opératoire	modification	??	Libre
programme de travail	extension		Restreint
tableau de bord	confirmation		Confidentiel
revue	révision		
analyse qualitative			
contrat			
programme			
planning			
livrable			
chiffres			
étude			
étude de besoin (EB)			
programme informatique			
benschmark			
appel à projet			
proposition			
trame			
photo			
relevé de valeur			
enregistrement vocal			
note de synthèse			
guide			
CR			
courrier			
relevé de décision			
cahier d'essai			
fiche d'essai			
graphe			
tableau			
liste de taches			
observation			
Retour d'expérience			
CR d'essai			
phasage			
bilan synthétique			
bilan complet			
bilan expertise			
diagnostic partiel			
statistiques			
calcul			
algorithme			
etude d'opportunité			

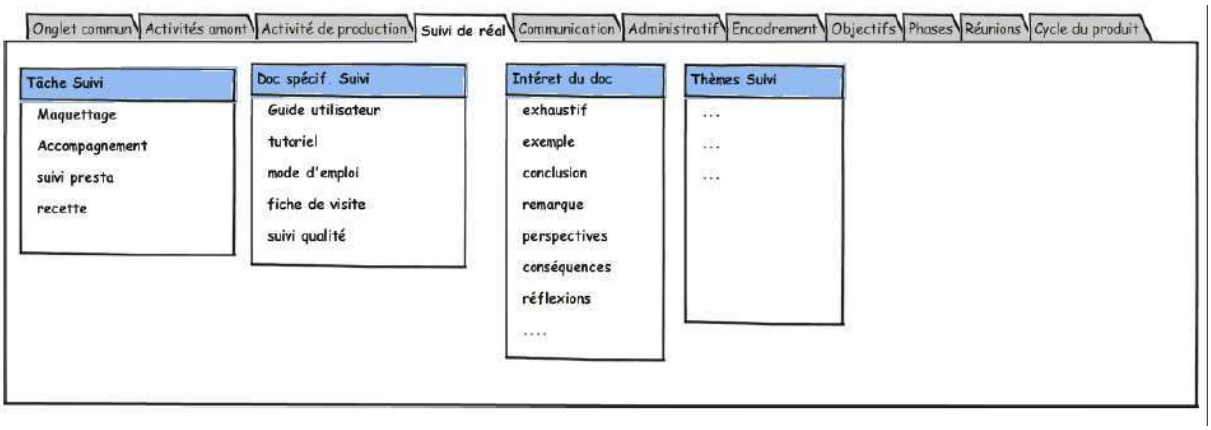
etude de faisabilité			
Note H			
Publication			
Document de travail			
Biblio			
Note Eureka (S)			
Présentation			
document de suivi			
OTM			

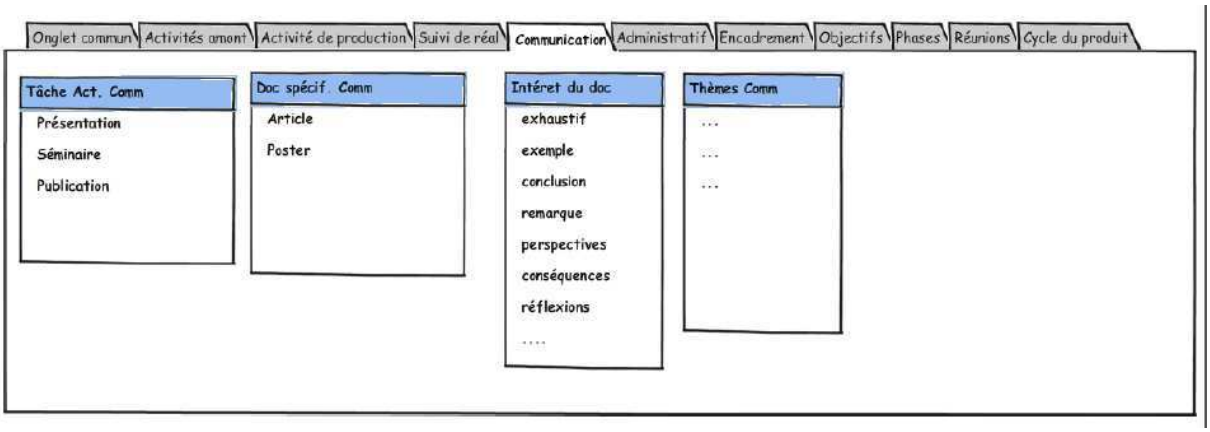
Annexe 3 : Classification à facettes proposée dans l'expérimentation

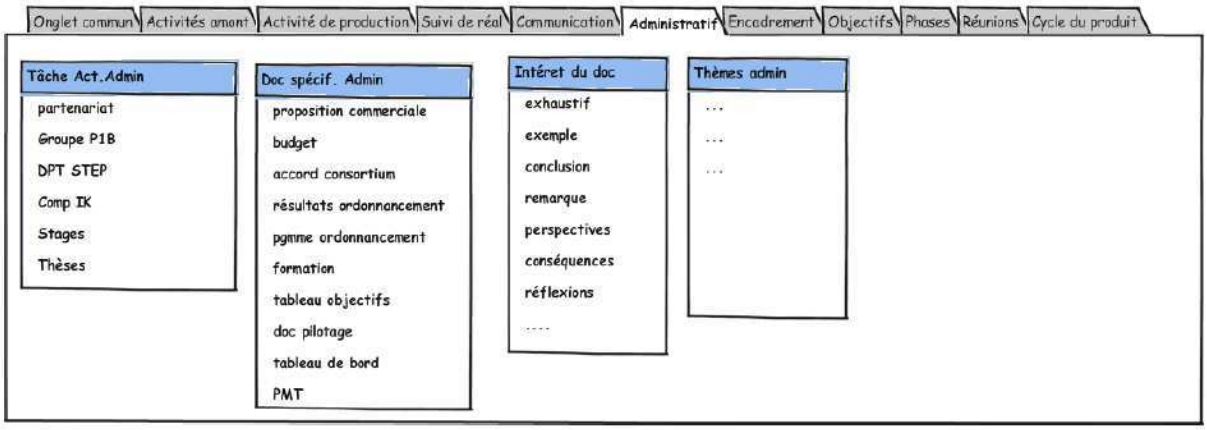


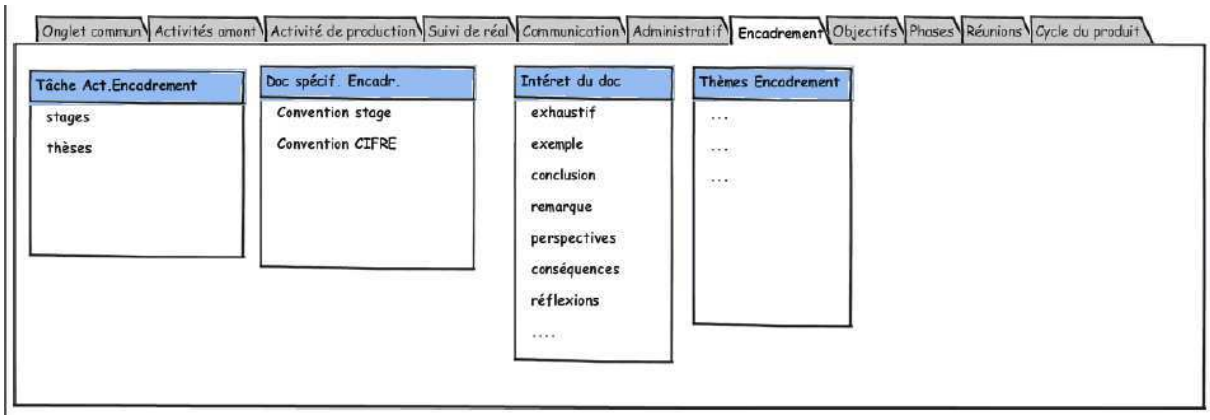


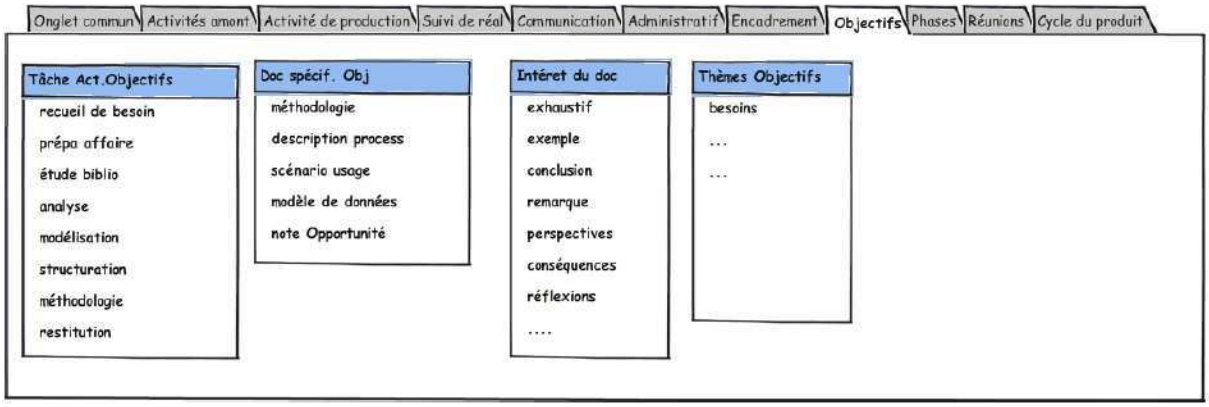


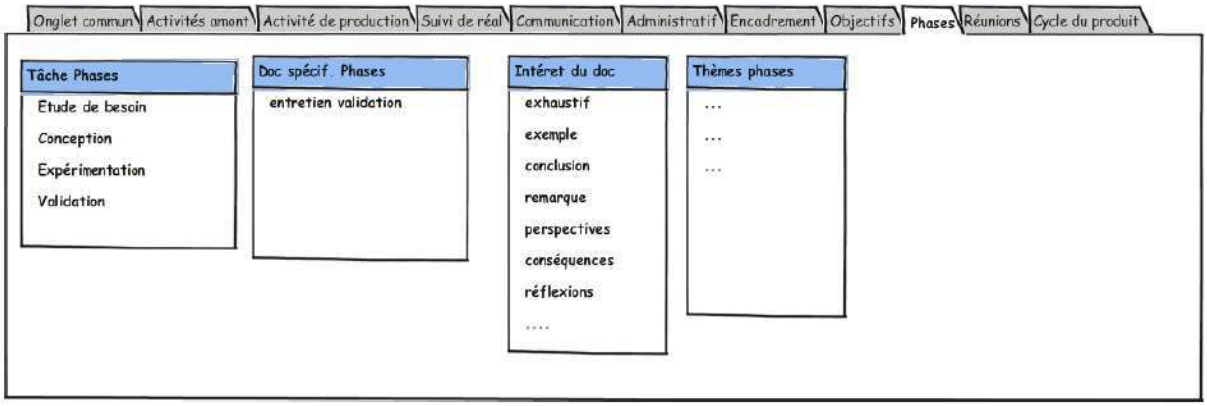

Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Suivi
Maquettage
Accompagnement
suivi presta
recette
Doc spécif. Suivi
Guide utilisateur
tutoriel
mode d'emploi
fiche de visite
suivi qualité
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Suivi
...
...
...

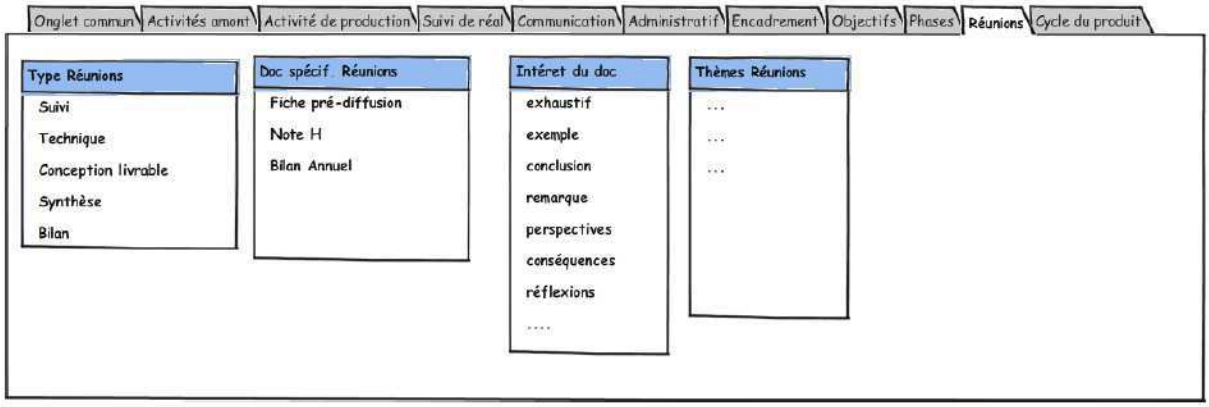

Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Act. Comm
Présentation
Séminaire
Publication
Doc spécif. Comm
Article
Poster
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Comm
...
...
...

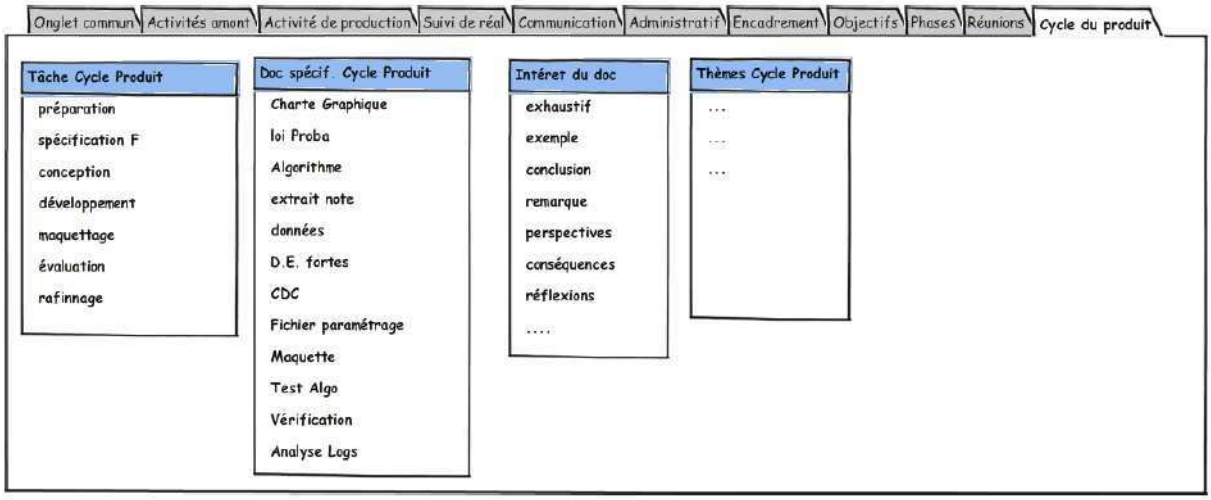

Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Act.Admin
partenariat
Groupe P1B
DPT STEP
Comp IK
Stages
Thèses
Doc spécif. Admin
proposition commerciale
budget
accord consortium
résultats ordonnancement
pgmme ordonnancement
formation
tableau objectifs
doc pilotage
tableau de bord
PMT
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes admin
...
...
...


Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Act.Encadrement
stages
thèses
Doc spécif. Encadr.
Convention stage
Convention CIFRE
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Encadrement
...
...
...


Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Act.Objectifs
recueil de besoin
prépa affaire
étude biblio
analyse
modélisation
structuration
méthodologie
restitution
Doc spécif. Obj
méthodologie
description process
scénario usage
modèle de données
note Opportunité
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Objectifs
besoins
...
...


Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Phases
Etude de besoin
Conception
Expérimentation
Validation
Doc spécif. Phases
entretien validation
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes phases
...
...
...


Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Type Réunions
Suivi
Technique
Conception livrable
Synthèse
Bilan
Doc spécif. Réunions
Fiche pré-diffusion
Note H
Bilan Annuel
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Réunions
...
...
...


Onglet commun
Activités amont
Activité de production
Suivi de réal
Communication
Administratif
Encadrement
Objectifs
Phases
Réunions
Cycle du produit
Tâche Cycle Produit
préparation
spécification F
conception
développement
maquettage
évaluation
rafinnage
Doc spécif. Cycle Produit
Charte Graphique
loi Proba
Algorithme
extrait note
données
D.E. fortes
CDC
Fichier paramétrage
Maquette
Test Algo
Vérification
Analyse Logs
Intéret du doc
exhaustif
exemple
conclusion
remarque
perspectives
conséquences
réflexions
....
Thèmes Cycle Produit
...
...
...

Orélie DESFRICHES DORIA

La classification à facettes pour la gestion des connaissances métier,

Méthodologie d'élaboration de FolkClassifications à Facettes

Résumé

Tout d'abord, nous abordons les problématiques liées à l'organisation et à la gestion des connaissances ainsi que les principes présidant à l'élaboration des classifications à facettes. Nous proposons ensuite une synthèse des méthodes existantes d'élaboration de classifications à facettes. Nous poursuivons en réalisant une analyse de l'impact des types de Systèmes d'Information sur l'activité de gestion de l'information. Ce faisant, nous produisons une analyse de l'activité inspirée de la Théorie de l'Activité, et influencée par les travaux de la psychologie du travail et de l'ergonomie cognitive. Le prototype Hypertagging développé dans le cadre du projet Miipa-Doc fondé sur le tagging des utilisateurs et structuré par les principes de la classification à facettes est présenté. Notre expérimentation sur l'élaboration de classifications à facettes pour la gestion des documents de travail est détaillée. Enfin, nous exposons notre méthodologie d'élaboration de FolkClassifications à facettes pour la gestion des connaissances métier. Elle vise à appuyer une démarche de gestion des connaissances sur la gestion personnelle de l'information.

Mots-clés : Organisation des connaissances, Système d'Organisation des Connaissances, Gestion des connaissances, Classification, Classification à facettes, Tagging, Indexation, Classement, Méthodologie, Système d'Information, Activité, Métier, Contexte, Gestion de l'information, Système Orienté-utilisateur, Collaboration, Contribution

Titre en Anglais : Faceted Classification for Knowledge Management in specific trades, Method for the development of FolkFaceted Classifications

Résumé en anglais

We address Knowledge Organization and Knowldege Management general issues and we present the basic principles for the development of faceted classifications. Then we introduce a synthesis of existing methods to develop this kind of classification. Next we propose an analysis of effects of Information Systems on information management activities. By doing this, we produce an analysis of the concept of activity inspired by Activity Theory, occupational psychology and french ergonomics. A presentation of Hypertagging prototype developed in the frame of Miipa-Doc research project, based on users tagging and structured by faceted classification principles, is provided. Our experiment about generating faceted classification for document management activities is detailed. Finally, we expose our method for the development of faceted classification for Knowledge Management purposes in the context of specific trades. This Knowledge Management approach is incorporated in Knowledge Organization activities.

Keywords : Knowledge Organization, Knowledge Organization System, Knowldege Management, Classification, Faceted Classification, Tagging, Indexing, Ranking, Methodology, Information System, Activity, Trade, Context, Information Management, User-oriented System, Collaboration, Participation.