

Université Jean Moulin Lyon 3

École doctorale : Sciences économique et de gestion

**Capitalisation d'expériences pour
l'indexation et la recherche
d'information dans le domaine de la
Gestion Electronique de Documents**

par Jean-Luc MARINI

thèse de doctorat en Sciences de l'Information et de la
Communication

sous la direction de Laïd BOUZIDI

présentée et soutenue publiquement le 30 septembre 2010

Membres du jury :

Laïd BOUZIDI, Professeur à l'Université Jean Moulin Lyon 3

Luc QUONIAM, Professeur à l'Université du Sud -Toulon Var

Amos DAVID, Professeur à l'Université de Nancy 2

Widad MUSTAFA EL-HADI, Professeur à l'Université Charles De Gaulle Lille 3

Mohammed SAAD, Professeur à l'Université de Bristol Business School (Angleterre)

Remerciements

Ce travail n'aurait pu être accompli sans le concours de mon directeur de thèse, Monsieur le Professeur Laïd BOUZIDI, de l'Université Jean Moulin - Lyon 3, qui m'a fait partager sa passion pour la recherche et m'a guidé tout au long de ces années.

Qu'il accepte toute ma reconnaissance pour avoir dirigé ce travail de recherche et participé activement à ma formation doctorale.

Je souhaite également exprimer mes profonds remerciements à Monsieur Amos DAVID, Professeur à l'université Nancy 2 et Monsieur Luc QUONIAM Professeur à l'Université du Sud-Toulon Var, pour m'avoir fait l'honneur de rapporter cette thèse et d'avoir accepté d'être membres du jury.

J'exprime aussi mes sincères remerciements à Madame Widad MUSTAFA EL-HADI, Professeur à l'Université Charles De Gaulle Lille 3 ainsi qu'à Monsieur Mohammed SAAD, Professeur à l'Université de Bristol Business School, pour avoir accepté d'être membres du jury de soutenance.

Mes remerciements vont également à tous les membres de l'équipe de recherche SICOMOR (Systèmes d'Information COMMunicants, Management et Organisation) de l'Institut d'Administration des Entreprises, rattachée au Centre de recherche Magellan de l'Université Jean MOULIN - Lyon 3 avec lesquels j'ai pu échanger et confronter des idées.

J'exprime ma profonde gratitude à Mademoiselle Sabrina BOULESNANE, Maître de conférences à l'Université Jean Moulin - Lyon 3 pour son écoute, la lecture des différentes versions de cette thèse et la qualité de ses remarques.

Mes remerciements vont aussi à tous mes collègues enseignants-chercheurs de l'Université Jean Moulin – Lyon 3 pour leurs conseils et leurs soutiens.

Enfin, je remercie plus particulièrement ma famille et mes amis qui m'ont soutenu sans faiblesse durant toutes ces années.

Introduction

Dans un environnement économique hautement concurrentiel, l'entreprise qui veut rester compétitive doit sans cesse repenser son organisation, mobiliser ses compétences et revoir sa stratégie. Cela nécessite une plus grande coopération entre les acteurs, un renforcement de la communication ainsi qu'une capitalisation des connaissances et des processus d'action.

Cette évolution suppose également que l'entreprise ait la capacité à maîtriser l'information, en temps réel, à tout moment et en tous lieux pour construire et faire évoluer sa base de connaissance stratégique.

Etablissant un parallèle entre le management stratégique d'entreprise et l'art de la guerre, Jean-Yves PRAX, Président Directeur Général de POLIA Consulting (société de conseil en management de la connaissance) et auteur de plusieurs ouvrages de référence dans le domaine de la Gestion Electronique de Documents [PRAX JY., 1998] [PRAX JY. et LARCHER S., 2004] et du Knowledge Management [PRAX JY., 2003] a d'ailleurs écrit dans l'un de ses livres que *"Dans l'art de la guerre, un avantage décisif revient à celui qui a la connaissance, du terrain, de l'arme, des plans ou des faiblesses de l'adversaire, sur celui qui ne l'a pas"*.

Cette réflexion sur la connaissance de l'adversaire et de son environnement s'inspire très largement des concepts de stratégies énoncés par SUN TZU, général chinois qui vécut-il y a 2 500 ans et auteur présumé du plus ancien ouvrage connu de stratégie militaire : "L'Art de la Guerre" [SUN TZU, 2001].

Ces concepts anciens et éminemment modernes peuvent s'adresser au management des entreprises emportées dans un environnement économique fluctuant et compétitif qui ont compris que maîtriser l'information sans cesse plus importante est désormais un enjeu vital.

Parmi les technologies de l'information et de la collaboration, la Gestion Electronique de Documents (GED) fournit les outils logiciels nécessaires afin de gérer électroniquement les informations et les documents (supports matériels des informations) qui circulent dans l'entreprise.

Bien que la GED permette de gérer des flux d'informations et d'automatiser des processus de diffusion des documents, une des finalités d'un système de GED est de rechercher rapidement de l'information en vue de la consulter, de la restituer ou de la communiquer à des tiers.

La pertinence et la qualité des informations trouvées dépendent non seulement de la méthode d'indexation et d'acquisition de l'information mais aussi de la technique d'interrogation choisie.

Bien souvent, la recherche d'informations s'opère par l'intermédiaire d'un langage faisant appel à des techniques d'interrogation plus ou moins complexes.

Plusieurs techniques existent et sont mises en œuvre dans les systèmes de GED, soit indépendamment et exclusivement les uns des autres, soit pour certaines d'entre elles de façon combinée.

Mais peu d'entre elles s'intéressent à la façon dont un utilisateur aborde un système de GED et formule sa requête, ni à la "rencontre fortuite" avec l'information.

De plus, rares sont les systèmes de GED qui offrent des outils de capitalisation des expériences de recherche d'information et qui s'intéressent à la formulation et à la représentation du mécanisme de recherche proprement dit.

Par conséquent, notre propos est de s'intéresser plus particulièrement à l'apport du mécanisme de "Sérendipité" (rencontre fortuite) et du raisonnement par cas dans les processus de recherche d'information afin d'élaborer un système d'aide à la recherche d'information dans le domaine de la GED.

Dans un premier temps, nous établirons la problématique de recherche d'information dans les systèmes de GED, puis nous décrirons les processus existants et tenterons de déterminer les limites des techniques de recherche utilisées dans ces processus.

Nous présenterons alors notre approche en matière de sérendipité et d'utilisation du raisonnement par cas pour la recherche d'information dans le domaine de la GED.

Notre approche devrait également donner lieu à une réalisation pratique dans le cadre de la nouvelle version d'un Intranet documentaire qui sera prochainement mise en place au sein d'une Société de Services en Ingénierie Informatique (SSII) de la région Rhône-Alpes.

I. Problématique de recherche

L'accès à une information pertinente, adaptée aux besoins et au contexte de l'utilisateur est un challenge dans une mémoire d'entreprise distribuée via le Web ou un Intranet, caractérisée par une prolifération des ressources hétérogènes (données structurées, documents textuels et images), conduisant à des volumes de données considérables.

Au fur et à mesure que ce volume s'accroît et que les données se diversifient, les systèmes de recherche d'information (moteurs web, systèmes de gestion de bases de données, etc.) délivrent des résultats massifs en réponse aux requêtes des utilisateurs, générant ainsi une surcharge informationnelle dans laquelle il est souvent difficile de distinguer l'information pertinente d'une information secondaire ou même du bruit.

La personnalisation de l'information constitue un enjeu majeur pour toute l'industrie informatique.

Que ce soit dans le contexte des systèmes d'information d'entreprise, du commerce électronique, de l'accès au savoir et aux connaissances, la pertinence de l'information délivrée, son intelligibilité et son adaptation aux usages et préférences des utilisateurs constituent des facteurs clés du succès ou du rejet de ces systèmes.

Aussi, afin d'améliorer la pertinence de l'information délivrée, notre propos a été de s'intéresser à un phénomène bien connu des Internautes qui parcourent le Web à la recherche de données, d'informations et de connaissances, à savoir : l'effet "Sérendipité".

Dans le domaine de la recherche d'information, le mécanisme de "Sérendipité" consiste à transformer en richesse exploitable des données ou des documents glanés au hasard.

Notre approche vise à démontrer que ce mécanisme n'emprunte rien au hasard mais qu'il est en fait une manifestation du mécanisme de contournement du mental conscient d'un acteur humain lors d'une recherche d'information.

Nos travaux de recherche font également appel au raisonnement par cas (RpC) qui est une technique qui se fonde sur des situations antérieures pour en traiter des nouvelles.

C'est également un concept puissant, qui fournit un mode de raisonnement analogique pour résoudre des problèmes.

Le raisonnement par cas consiste à capitaliser les expériences antérieures à une situation donnée et ainsi de générer des connaissances liées à ces dernières.

Le langage de l'inconscient étant analogique, l'utilisation du raisonnement par cas nous permet de capitaliser les expériences de contournement du mental conscient d'un acteur humain en quête d'information pour faciliter le recours à l'effet "Sérendipité" lors d'expériences de recherche d'information ultérieures.

II. Structuration et organisation de la thèse

Cette thèse en Sciences de l'Information et de la Communication s'inscrit plus particulièrement dans le champ des sciences cognitives. Elle a pour objet la création d'un modèle de recherche d'information (RI) utilisant le mécanisme de "Sérendipité" et le raisonnement par cas (RpC) en vue d'améliorer le processus de recherche d'information dans le domaine de la Gestion Electronique de Documents.

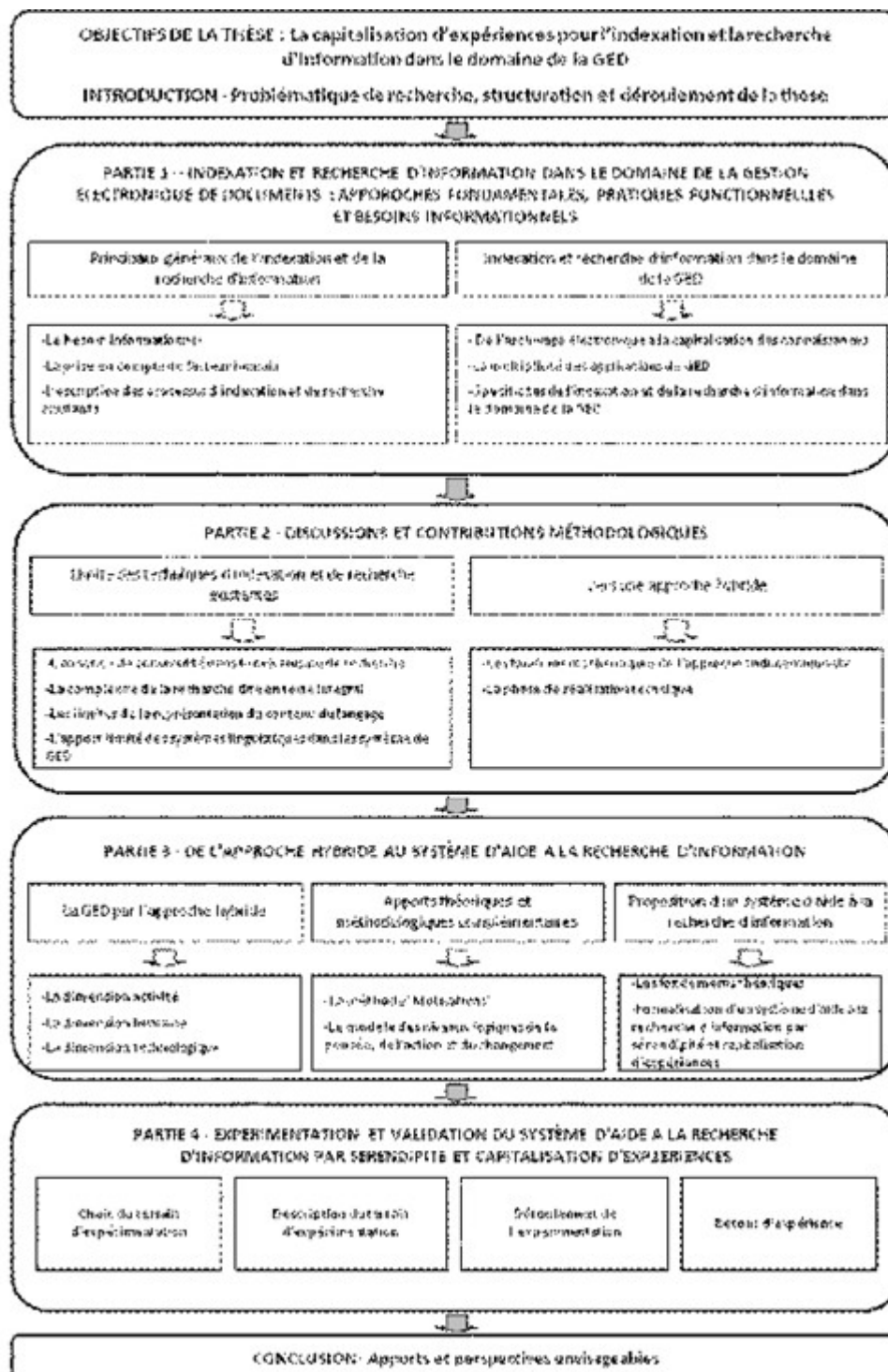
Elle s'organise autour de quatre parties. La première partie, "INDEXATION ET RECHERCHE D'INFORMATION DANS LE DOMAINE DE LA GESTION ELECTRONIQUE DE DOCUMENTS : APPROCHES FONDAMENTALES, PRATIQUES FONCTIONNELLES ET BESOINS INFORMATIONNELS", présente un état de l'art sur l'indexation et la recherche d'information dans le domaine de la Gestion Electronique de Documents.

La deuxième partie, "DISCUSSIONS ET CONTRIBUTIONS METHODOLOGIQUES", présente les différentes hypothèses de recherche envisagées ainsi que les raisons du choix d'un cadre méthodologique qui s'appuie sur une démarche hybride : l'approche tridimensionnelle.

La troisième partie, "DE L'APPROCHE HYBRIDE AU SYSTEME D'AIDE A LA RECHERCHE D'INFORMATION", est consacrée à notre proposition d'amélioration du processus de recherche par l'adoption d'un Système d'Aide à la Recherche d'Information utilisant le mécanisme de "Sérendipité" et le raisonnement par cas (RpC).

La quatrième partie, "EXPERIMENTATION ET VALIDATION DU SYSTEME D'AIDE A LA RECHERCHE D'INFORMATION PAR SERENDIPITE ET CAPITALISATION D'EXPERIENCES", décrit le contexte d'expérimentation du Système d'Aide à la Recherche d'Information proposé ainsi que sa validation par le biais d'une instanciation au sein d'un portail documentaire intégrant des fonctionnalités de Gestion Electronique de Documents et un retour d'expérience mettant en perspective les améliorations possibles.

Figure 1 : Plan de lecture du mémoire de thèse



III. Déroulement et évolution de la thèse

L'observation de ce qui se passe sur et autour d'Internet nous confirme que nous entrons aujourd'hui dans une société du savoir. Cette dénomination n'est certainement pas la bonne et

changera vraisemblablement dans l'avenir mais cela signifie que cette société est d'ores et déjà autre chose que la société de l'information, dont nous sommes en train de sortir.

Dans une telle société, les êtres humains que nous sommes construisent des connaissances, des compétences et des savoir-faire à l'aide notamment des outils numériques disponibles.

Il en résulte que dans les pays industrialisés tout au moins, la plupart des actes de la vie professionnelle mais aussi de la vie privée comprennent des séquences numériques ou sont imprégnés d'une approche numérique. L'exploration documentaire qui permet d'accéder à la connaissance est aujourd'hui transcendée par la puissance des réseaux d'information et de communication ainsi que par le nombre des acteurs qui s'en servent. La recherche d'information sur le Web ou l'accès à des espaces de discussion nous mettent en contact avec des points de vue et des acteurs que l'on ne s'attendait pas à rencontrer et qui sont porteurs de connaissances, de compétences et de savoir-faire qui enrichissent notre vision du monde.

Au niveau des organisations économiques, le passage de l'entreprise industrielle et commerciale à l'entreprise du troisième millénaire qui capitalise sur les connaissances, les compétences et les savoir-faire est désormais initié [SIMON HA., 1977].

Le concept même d'entreprise "intelligente" est considéré par les managers comme une nouvelle voie à explorer face à un environnement économique hautement concurrentiel qui exige plus de performance et d'agilité.

Les dirigeants et actionnaires d'entreprises ont pour la plupart compris la nécessité de favoriser l'émergence d'une performance collective basée sur les coopérations des différents acteurs qui constituent la richesse humaine des organisations.

Mais passer d'une culture de l'information et de la communication à celle de la collaboration ou bien encore de la coopération ne peut pas se faire sans volonté commune, ni s'être assuré au préalable de la capacité de chacun des acteurs à savoir collaborer ou bien encore coopérer.

Les technologies de l'information et de la communication, de la collaboration et de l'intelligence augmentée ne sont alors introduites dans les organisations que pour outiller les processus d'information, de communication, de collaboration et de coopération.

Dans ce contexte, on comprend tout l'intérêt à s'intéresser à certains outils technologiques qui portent ces processus.

La Gestion Electronique de Documents de par sa nature et son appartenance aux technologies de l'information et de la collaboration constitue un formidable terrain d'expérimentation pour

nos travaux de recherche visant à améliorer l'accès à cette information pertinente que certains n'hésitent pas à qualifier de "marchandise des rois".

Car l'enjeu n'est plus l'information en elle-même (bien souvent disponible et accessible au plus grand nombre), ni sa communication (avec Internet, l'information circule très vite à très grande échelle) mais de savoir analyser l'information recherchée en vue d'une exploitation efficace [CHAUMIER J., 1996].

Les entreprises qui demain sauront mobiliser les intelligences internes et externes disponibles pour comprendre la portée des informations auxquelles elles ont accès, leurs enjeux, leur sens seront ainsi en capacité de pouvoir anticiper les crises et faire naître l'innovation au cœur de leur activité.

C'est en prenant conscience de la valeur opérationnelle de l'information que nous nous sommes tout d'abord intéressés à la Gestion Electronique de Documents, véritable "mémoire de l'entreprise".

Puis, c'est seulement après avoir mesuré les limites actuelles des mécanismes d'indexation et de recherche d'information dans le domaine de la GED que s'est constituée notre problématique de recherche.

Afin de mener à bien nos travaux de recherche, il nous a tout d'abord fallu établir un état de l'art de l'indexation et de la recherche d'information dans les organisations et ce, dans l'optique d'en appréhender les principes fondamentaux.

Cet état de l'art traite donc :

- Du besoin informationnel des organisations ;
- Du comportement de l'acteur humain à l'initiative d'une recherche d'information qui agit dans un cadre professionnel donné et en fonction d'un objectif opérationnel ou décisionnel précis ;
- Des processus d'indexation et de recherche d'information existants ;
- Des technologies de l'information, de la communication, de la collaboration et de l'intelligence augmentée ;
- De la multiplicité des applications de Gestion Electronique de Documents ;
- Des spécificités de l'indexation et de la recherche d'information dans le domaine de la GED.

Une fois cet état de l'art établi, le déroulement de cette thèse n'a pas pour autant fait l'objet d'un processus linéaire. En effet, de nombreuses voies ont été explorées afin d'améliorer la pertinence de l'information recherchée dans le domaine de la GED mais aucune n'a retenue de façon significative notre attention si ce n'est la prise en compte du mécanisme de "Sérendipité".

La participation à des manifestations scientifiques et les nombreux échanges entretenus avec les différents membres de l'équipe de recherche SICOMOR n'ont fait que conforter cette orientation. La prise de conscience et la compréhension du mécanisme de contournement du mental conscient n'ont fait que confirmer notre intérêt pour la prise en compte de l'effet "Sérendipité" dans les dispositifs d'aide à la recherche d'information dans le domaine de la GED.

Mais cette démarche visant à concevoir un système d'aide à la recherche d'information par "Sérendipité" dans le domaine de la GED ne peut rester en l'état et demeurerait trop incomplète sans une phase d'expérimentation au sein d'une organisation économique visant à instancier le modèle proposé et à mesurer la validité de notre approche.

Le retour d'expérience issu de cette expérimentation nous fournit alors les éléments nécessaires à l'évolution du modèle proposé et nous ouvre des perspectives de recherche que nous ne manquerons pas d'exploiter.

Partie 1 : Indexation et recherche d'information dans le domaine de la gestion électronique de documents : Approches fondamentales, pratiques fonctionnelles et besoins informationnels

Selon Olivier ZARA, Président fondateur d'Axiopole (société spécialisée dans le conseil et la formation en management de l'intelligence collective ainsi que dans l'édition de logiciels d'aide au management d'équipes et de groupes de travail) et membre du "Collective Intelligence Business Network" [ZARA O., 2008], *"La performance des entreprises dans une société industrielle et commerciale est de savoir produire, de savoir vendre mieux et plus vite que ses concurrents. Aujourd'hui, la plupart des entreprises savent produire et vendre. C'est la raison de leur existence. Si elles n'avaient pas cette capacité, elles auraient déjà disparu. Mais au fil du temps, leur niveau de performance devient de plus en plus homogène. Croître, prendre des parts de marché se fait alors plus facilement en absorbant ses concurrents.*

La performance des entreprises dans une société de l'information est de savoir mobiliser l'intelligence collective et les connaissances de ses parties prenantes (salariés, fournisseurs, clients,...). S'il faut, et s'il faudra toujours, savoir produire et vendre, ce n'est plus aujourd'hui un facteur suffisamment différenciateur dans la compétition internationale. Hier, l'entreprise était industrielle et commerciale. Demain il faudra qu'elle soit de plus en plus une entreprise intelligente.

L'entreprise intelligente repose principalement sur l'intelligence collective (IC) et le Knowledge Management (KM) qui ne peuvent exister et fonctionner efficacement sans les technologies de l'information. Ces outils font partie des technologies de l'intelligence

augmentée (Amplified Intelligence) dont l'objet est d'étendre les capacités intellectuelles humaines, en particulier les capacités cognitives des groupes. Ces technologies ont beaucoup évolué ces dernières années en passant de l'information à la communication, puis aujourd'hui à la collaboration.

Les technologies de l'information et de la communication ont permis de rendre accessible l'information, de stocker et de partager. Les technologies de l'information et de la collaboration (intranet collaboratif) vont beaucoup plus loin. Elles augmentent la performance des interactions humaines et donnent à l'information une valeur opérationnelle".

Mais, force est de constater que les managers d'entreprises sont nombreux à résister aux processus d'intelligence collective parce qu'ils pensent qu'ils vont perdre leur pouvoir. En fait, l'IC n'induit pas une redistribution du pouvoir (chacun reste à sa place, chacun conserve la même quantité de pouvoir) mais un changement dans l'exercice du pouvoir, dans les modes de management. Si l'on se réfère à Pierre LEVY, titulaire d'une chaire en Intelligence Collective à l'université d'Ottawa (CANADA) et auteur de nombreux ouvrages sur les implications culturelles de l'informatisation [LEVY P., 1997], *"l'intelligence collective implique donc une nouvelle gouvernance des organisations qu'on appelle le management de l'intelligence collective"*.

Aussi, c'est dans le cadre de ce nouveau mode de management des organisations et du profond changement que cela impose, que nous allons nous intéresser dans un premier temps aux principes généraux de l'indexation et de la recherche d'information puis à l'indexation et à la recherche d'information dans le domaine de la Gestion Electronique de Documents.

I. Principes généraux de l'indexation et de la recherche d'information

Aborder les principes généraux de l'indexation et de la recherche d'information consiste, dans un premier temps, à tenter de formaliser le besoin informationnel des organisations économiques ainsi qu'à prendre en compte le facteur humain, à l'origine des processus d'indexation et de recherche d'information existants.

1) Le besoin informationnel

Dans l'optique d'appréhender le besoin informationnel (énoncé, réel ou reformulé) des organisations économiques, il convient de définir ce que recouvre exactement le concept d'information et comment il se différencie de la donnée ou bien de la connaissance, autres concepts souvent utilisés en lieu et place de l'information et souvent confondus avec ce dernier.

Une donnée est un fait discret, brut; elle résulte d'une observation, d'une acquisition ou d'une mesure effectuée par un instrument naturel ou artificiel, elle peut être qualitative ou quantitative. Il n'y a pas d'intention, ni de projet dans la donnée, c'est ce qui lui confère son caractère d'objectivité [KORFHAGE R.R., 1997].

Une information est une collection de données organisées dans le but de délivrer un message, le plus souvent sous une forme visible, imagée, écrite ou orale. La façon d'organiser les données résulte d'une intention de l'émetteur, et est donc parfaitement subjective. Pour Gregory BATESON [BATESON G., 1973], *"l'information produit un nouveau point de vue sur des événements ou des objets, qui rend visible ce qui était invisible"*.

On distingue différents types d'information :

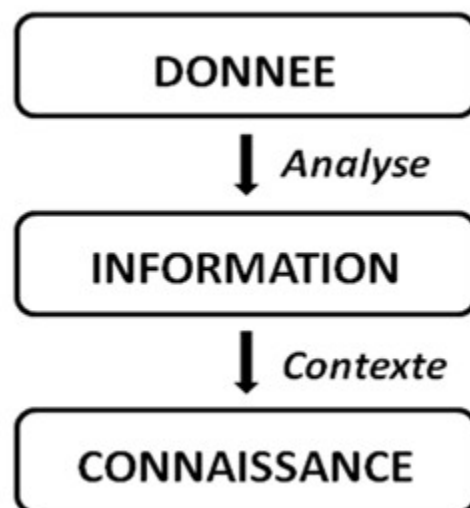
- **L'information physique** : Elle se présente sous la forme d'une agrégation de données quantitatives ou qualitatives décrivant un état ou un fait. L'information est essentiellement déclarative et peut être purement physique (attributs), logique ou instrumentale (par exemple, l'extrait K-bis d'une entreprise).
- **L'information pragmatique** : Elle se réfère à la qualité de l'information, sa pertinence, sa précision, sa validité, sa lisibilité ainsi que sa mise en forme (par exemple, une procédure).
- **L'information rationnelle** : Elle fait référence au raisonnement, à une démonstration reproductible ou déduction logique pour établir une vérité (par exemple, l'absence de résultats d'une société peut résulter d'une baisse de ses recettes et du maintien de ses charges d'exploitation).
- **L'information paradigmatique** : Elle désigne une évidence, une valeur partagée, mais non démontrable. Elle est l'expression d'un consensus plus que de la vérité (par exemple, la démocratie est un système politique qui respecte les droits de l'homme).

- **L'information expressive** : Elle désigne une information destinée à créer une impression, une réaction émotionnelle ou affective pour faire passer un message. Elle peut être non langagière (images, sons, films) ou orale (style narratif, figure rhétorique). Par exemple, le docu-film *Home* du Photographe Yann Arthus-Bertrand.

La connaissance, quant à elle, diffère de l'information par plusieurs aspects :

- Pour qu'une information devienne connaissance, il faut que l'acteur humain puisse construire une représentation qui fasse sens ;
- Pour cela, l'information subit une série d'interprétations (filtres et retraitements), liées aux croyances générales, au milieu socioprofessionnel, au point de vue, à l'intention ainsi qu'au projet de l'acteur humain ;
- Contrairement à l'information, la connaissance, n'est pas seulement mémoire, item figé dans un stock, mais toujours activable selon une finalité, une intention, un projet. En cela, il y a dans la connaissance la notion de processus visant à construire une représentation finalisée d'une situation donnée.

Figure 2 : Donnée, information et connaissance



Pour appréhender le besoin informationnel, il nous faut revenir aux temps anciens, où l'information était réduite à l'échange verbal dont la portée physique était limitée dans l'espace et le temps, comme le savoir était assujéti à la mémoire, donc à une transmission orale.

Pendant longtemps, la science et l'histoire ont été entièrement soumises à l'ordre religieux, autoritaire et hiérarchique qui entendait garder la maîtrise de l'espace et du temps, et qui conservait le privilège de distinguer l'information de la vérité.

Cette rareté de l'information et la précarité des systèmes de communication expliquent l'inexistence de la "Stratégie" au sens moderne du terme (l'art de coordonner l'action collective), et il a fallu attendre la première révolution de l'information, contemporaine des grandes inventions et découvertes (imprimerie, découverte du "nouveau monde", etc.), pour que la stratégie soit enfin considérée comme un système organisé et non plus comme le simple "art du commandement".

La séparation du pouvoir religieux et de l'état ainsi que le rationalisme, vont permettre véritablement la naissance de la stratégie moderne. Dans ce nouveau cadre, l'information a tendance à devenir autonome et à créer sa propre zone de pouvoir. Elle reste évidemment la matière première dont se nourrit la stratégie, mais elle en devient aussi de plus en plus souvent l'enjeu.

Selon Tzvetan TODOROV, auteur de "Mémoire du mal, tentation du bien" [TODOROV T., 2000], "Ayant compris que la conquête des terres et des hommes passe par celle de l'information et de la communication, les tyrannies du XXème siècle ont systématisé leur mainmise sur la mémoire et tenté de la contrôler jusque dans ses recoins les plus secrets".

Il faudra attendre la seconde révolution de l'information, celle de l'informatique, pour que l'information ne soit plus au seul profit des systèmes structurés, autoritaires et hiérarchiques, mais de façon horizontale au niveau des individus, des groupes d'intérêt et des entreprises.

Selon Jean-Yves PRAX, auteur du "Manuel du Knowledge Management" [PRAX JY., 2003], "La diffusion massive de moyens d'information, de relation et de communication enlève aux pouvoirs traditionnels le monopole de l'information. La diffusion du savoir, la démocratisation et l'appropriation de l'information, la banalisation de la communication vont à l'encontre de toutes les organisations sociales et politiques antérieures, fondées sur la verticalité et le pouvoir".

Ces changements s'opèrent également dans les entreprises, où les technologies de l'information, de la communication, de la collaboration et de l'intelligence augmentée ont donnée de la valeur à l'information [ZARA O., 2008].

Dans un environnement économique en perpétuelle évolution, chaque acteur humain au sein d'une organisation peut s'apercevoir à quel point il a besoin d'informations pour agir.

Mais, il doit également se rendre compte que les informations qu'il détient sont partielles et son savoir limité ou spécialisé, ce qui peut impliquer plusieurs interprétations d'une même information.

Face à cette complexité, chaque acteur à la recherche d'informations pertinentes et significatives pour réaliser des opérations ou prendre des décisions peut être amené à ressentir une insatisfaction.

L'expression "besoin informationnel" désigne la situation d'un acteur humain (ou d'un groupe d'acteurs humains) qui ne disposerait pas de toutes les informations qui lui seraient utiles à un moment donné [MEADOW C.T ET AL., 2000].

En d'autres termes, cet acteur humain peut ressentir un manque d'information et, de ce fait, se trouver dans une situation d'insatisfaction. La situation satisfaisante serait celle où les informations sont complètes (on parle alors de complétude des informations) et toutes pertinentes. Cette situation idéale peut à juste titre être considérée comme inatteignable.

Si on introduit une information complète dans un système, ce n'est pas pour autant que l'acteur humain qui accédera à cette information en aura une vision complète.

Il en est de même en matière de pertinence de l'information, il est important de noter qu'elle varie en fonction du besoin informationnel et de l'utilisation que l'acteur humain veut en faire.

Une même information peut être pertinente pour un acteur humain et non pertinente pour un autre.

Cependant, lorsque l'on parle de pertinence de l'information, il convient de distinguer trois notions différentes [DALBIN S., 2002] :

- La "pertinence-système" qui concerne l'indexation des documents, supports matériels de l'information (et qui définit entre autres le bruit et le silence documentaire) ;
- La "pertinence-utilisateur", qui englobe la pertinence de la formulation de la requête, de la présentation des résultats ainsi que celle du document par rapport au besoin informationnel ;
- La "pertinence-thème", qui fait référence au contenu du document par rapport à un thème.

2) La prise en compte du facteur humain

L'homme est le seul être à avoir pleinement conscience qu'il existe et que son existence est limitée dans le temps. Cette certitude fait naître chez lui un étonnement qui contient des degrés de crainte, de surprise, d'ébranlement et de douleur. Il vit tout en sachant que sa vie va

prendre fin et il est terrifié par sa mortalité. Cette peur l'entraîne à se poser des questions sur son existence et à chercher un sens à un monde quelque peu absurde.

Face à cette peur de l'inconnu, l'être humain éprouve alors le désir légitime de modifier le présent pour améliorer son futur. Il utilise pour cela des modèles, qu'ils soient intuitifs ou d'une grande complexité mathématique, pour comprendre la réalité, transformant de simples informations en connaissance et projetant ces dernières dans le futur pour créer de nouvelles réalités.

C'est cette capacité à transformer l'information en connaissance et la connaissance en compétence qui confère à l'être humain toute sa richesse.

Dans une économie fortement concurrentielle et toujours changeante, face aux exigences multiples de clients en perpétuelle évolution, c'est par l'exploitation de ses compétences et de ses connaissances que l'entreprise peut répondre et prospérer. Sa richesse fondamentale est avant tout humaine.

La prise en compte du facteur humain en entreprise, consiste donc à étudier les mécanismes qui existent entre ce qui est perçu par l'homme et les réactions conscientes ou inconscientes qui en découlent.

Seule l'étude de ces mécanismes peut nous permettre de déterminer sur quels éléments se fonde la richesse humaine et ce qui fait qu'une information possède de la valeur aux yeux d'un acteur humain.

Dans le cadre des systèmes d'information, la prise en compte du facteur humain s'appuie sur des notions comme le "Profilage", la prise en compte du ressenti ou bien encore l'analyse du comportement.

3) Description des processus d'indexation et de recherche existants

On peut considérer indexation et recherche comme deux faces d'une même pièce de monnaie dans la mesure où le but de la recherche est de retrouver l'information qui a préalablement été indexée.

Le processus d'indexation fait référence à des opérations intellectuelles plus ou moins complexes dont l'utilisateur final n'a guère idée [LE LOARER P., 1994].

A ce processus, il faut ajouter des composantes techniques liées au mode de stockage du résultat de l'indexation et aux possibilités de recherche d'information à partir de cette indexation.

Dans le domaine de la recherche d'information documentaire, l'objectif du processus d'indexation est de représenter le contenu d'un document sous une forme condensée (ensemble de termes).

Le processus d'indexation est donc l'opération qui consiste à décrire le document afin de permettre une exploitation sans nécessairement recourir à la consultation du document lui-même.

Il doit être la représentation la plus fidèle et la plus exhaustive du contenu et du sens du document afin de faciliter la recherche d'information et de garantir la pertinence de cette information.

Cette représentation peut être réalisée par :

- Une analyse intellectuelle, réalisée par un acteur humain spécialiste du domaine. On parle alors d'indexation manuelle ;
- Un traitement automatique qui se traduit par un ensemble de fonctions et de traitements automatisés. On parle alors d'indexation automatique ;
- Une combinaison des deux approches, où les termes significatifs d'un document sont choisis par un spécialiste utilisant un thésaurus ou une base terminologique. On parle alors d'indexation semi-automatique.

Dans le cas de l'indexation automatique, plusieurs étapes sont à respecter :

- L'analyse lexicale,
- L'élimination des mots vides,
- La lemmatisation,
- La pondération des termes.

L'analyse lexicale consiste à convertir le texte d'un document en une suite de termes. Un terme est une unité lexicale ou un radical. Ce processus permet de reconnaître les espaces de séparation des mots, des chiffres, les ponctuations, etc.

L'élimination des mots vides a pour objectif de supprimer les termes non significatifs (pronoms personnels, prépositions, etc.) ou mots athématiques (mots qui peuvent se retrouver

dans n'importe quel document parce qu'ils exposent le sujet mais ne le traitent pas, comme par exemple contenir, appartenir, etc.).

La lemmatisation provient du fait qu'on trouve des mots dans un texte qui ne sont pas dans leur forme canonique. Ces différentes formes peuvent avoir le même sens ou un sens très proche. Il suffit alors d'indexer les racines, et donc substituer les termes par leur lemme pour éviter d'avoir à indexer tous les mots.

La lemmatisation permet alors d'indexer la forme canonique du terme qui regroupe les différentes variables du mot et ses dérivés. Cette forme est l'infinitif pour les verbes, la forme masculine singulière pour les noms, etc.

La pondération d'un terme indique l'importance du terme dans la caractérisation d'un document. L'objectif est alors de trouver les termes représentant le mieux le contenu du document. Ainsi, une bonne formule de pondération est celle qui assure à la fois un rappel et une précision élevés en prenant bien compte la fréquence d'occurrence d'un terme dans un document et l'importance d'un terme dans toute la collection de documents.

Pour caractériser le processus d'indexation (qu'il s'agisse d'indexation humaine ou automatique), il convient de distinguer :

- Les objets sur lesquels porte l'indexation ;
- Les outils sur lesquels elle peut s'appuyer ;
- Les schémas qu'elle présuppose ;
- Le processus de recherche d'information avec lequel elle est en relation.

Selon Pierre LE LOARER [LE LOARER P., 1994], le processus d'indexation ainsi caractérisé permet de faire le lien entre une offre d'informations (le fonds documentaire) et une demande d'informations (les recherches effectuées par des utilisateurs).

a) Typologie des différentes sources d'information

Le document a été pendant longtemps synonyme de texte, mais depuis plusieurs années il recouvre une réalité bien différente.

Aussi, de par l'interpénétration des médias et des supports, nous nous trouvons face à une typologie documentaire de plus en plus complexe.

Typologie qu'il nous est possible de résumer de la façon suivante :

Documents textuels : Il s'agit essentiellement de documents existants sous une forme papier, qu'ils soient produits ou non par une application informatique. Un document textuel se présente généralement comme un ensemble de termes qui constitue un écrit ou une œuvre originale.

Documents visuels : On entend par documents visuels l'ensemble des sources d'information de nature graphique (plan, carte, schéma, etc.) ou iconographique (dessin, peinture, photographie, etc.) directement visibles par l'œil humain.

Documents sonores : La production de sons intelligibles, par l'homme sans transformation ni utilisation d'une machine, constitue un document sonore qu'il convient d'identifier et de conserver en l'état. Son contenu étant porteur de sens, il représente un témoignage qu'il est nécessaire de stocker le plus fidèlement possible afin de pouvoir le reproduire ultérieurement sans altération ni déformation d'aucune sorte.

Documents audiovisuels : Les documents audiovisuels rassemblent l'ensemble des sources d'information qui utilisent la présentation d'images, de films et d'enregistrements sonores (séquences vidéos, diaporamas, etc.). Il s'agit en fait de documents n'appartenant pas à l'univers de l'écrit, ils apportent une illustration sur des événements ou des objets avec une sensibilité plus importante que la seule description verbale.

Documents structurés : Un document structuré se caractérise par un ensemble d'éléments organisés dont la présentation relève un aspect significatif. Deux niveaux de structuration peuvent être retenus pour définir un document, la structure physique et la structure logique. Cette séparation permet au document électronique d'acquérir sa virtualité et d'être indépendant des matériels de visualisation ou d'impression. Une série de normes internationales permettent de définir cette structuration.

Bien que l'objectif principal d'un système d'information documentaire soit de manipuler différentes sources d'information, il est important de noter que le type de documents à appréhender est parfois très éloigné de la notion de document électronique.

Ce qui sous-entend, qu'un processus de transformation doit être opéré afin de parvenir à stocker sous une forme numérique un document préexistant dont l'aspect dépend essentiellement de son mode de production.

Cette transformation a une répercussion sur le processus d'indexation de l'information et donc par conséquent sur le mode d'interrogation du fonds documentaire via un système d'information documentaire.

b) Le processus d'indexation

Si l'on se réfère aux travaux de Serge CACALY, Yves-François LE COADIC, Eric SUTTER, et Paul-Dominique POMART [CACALY S. ET AL., 2008], l'indexation d'un document (en tant que support d'information) dans un système d'information documentaire peut être définie comme l'opération qui consiste à décrire le document et son contenu en vue de faciliter son exploitation sans nécessairement recourir à la consultation du document lui-même.

Selon Widad MUSTAFA EL HADI [MUSTAFA EL HADI W., 2006], *"cette opération est un préalable indispensable à toute recherche d'information sur le contenu et à d'autres types de traitement des informations. Il est donc difficile de parler de recherche d'information sans parler d'indexation, au sens procédural du terme. Indexation et recherche d'information sont de ce fait deux processus fortement liés. On indexe les documents dans le but de les retrouver."*

Selon Jian-Yun NIE [NIE JY., 2003] de l'Université de Montréal, *"les index (ou termes d'indexation) jouent un rôle important dans la recherche d'information dans la mesure où ils déterminent avec quels mots on peut retrouver un document"*.

Ces index (ou descripteurs) sont généralement stockés dans une base de données sur laquelle se feront les recherches ultérieures.

Selon Jean-Yves PRAX [PRAX JY., 2004], *"l'histoire des langages d'indexation est riche, car elle s'inscrit dans celle de la pensée scientifique, avec deux courants épistémologiques majeurs :*

- *Un courant positiviste qui remonte à la haute Antiquité, avec notamment la classification des connaissances d'Aristote, et qui a engendré l'indexation classificatoire ;*
- *Un courant constructiviste, né vers les années 1960, année d'origine de la description conceptuelle des documents par le thésaurus, puis suivi par des techniques d'indexation en langage naturel et d'hypertextes apportées par l'informatique".*

Ces deux courants ont donné naissance à deux catégories de méthodes de représentation du document :

- **Les langages classificatoires**, utilisés pour représenter le sujet d'un document de manière synthétique à travers la notion de domaines et de sous-domaines (sur plusieurs niveaux) ;

- **Les langages d'indexation**, utilisés pour représenter le contenu d'un document de manière analytique. Ces langages sont parfois nommés **langages combinatoires** pour désigner le fait que les termes et les concepts utilisés pour l'indexation peuvent être combinés entre eux lors de la phase de recherche.

Ces langages dits "d'indexation combinatoire" peuvent être eux-mêmes répartis en deux sous-catégories :

- **Les langages libres**, qui sont construits à posteriori à partir des concepts relevés dans les documents et qui se matérialisent par des mots-clés et/ou des descripteurs libres ;
- **Les langages contrôlés**, qui sont construits à priori, cherchant à représenter les documents par une codification univoque (qui n'a qu'un sens) et que l'on retrouve sous la forme de **listes d'autorité** et de **thésaurus**.

Les langages classificatoires : Ils ont pour but de classer les documents selon une structure hiérarchique (classification par domaine, sous-domaines, etc.). Notons qu'une telle approche est idéaliste et naïve et pose souvent des problèmes lorsqu'il s'agit de classer un document dans deux domaines, de découper des domaines en sous domaines (avec plus ou moins de finesse et sur plusieurs niveaux) et de restructurer la classification existante en tenant compte des documents précédemment classifiés.

Pour s'en convaincre, il suffit de prendre connaissance de la Classification Décimale Universelle (CDU) qui est un système de classification de bibliothèque développé par Paul OTLET et Henri LA FONTAINE [UDC Consortium, 2004], deux juristes belges fondateurs de l'Institut International de Bibliographie en 1895, et qui a été établie à partir de la classification décimale de l'américain Melvil DEWEY [OCLC Inc., 2005] et qui est toujours largement utilisée par les bibliothèques, dont la Bibliothèque Nationale de France.

Les langages libres : Ils ont pour but de respecter la richesse "native" du document et de limiter au maximum la rationalisation (réflexion sur le document et son contenu), d'ailleurs toujours subjective, introduite par les personnes chargées de la phase d'indexation.

Parmi ces langages, les mots-clés se présentent sous la forme d'une liste de termes significatifs extraits du titre, du résumé ou du texte d'un document, supposés être représentatifs du contenu. L'établissement d'une liste de mots-clés soulève bien souvent la question de la personne qui extrait les mots-clés (auteur du document ou personne tierce) et de la manière dont les mots-clés sont extraits (méthode automatique ou manuelle). Mais elle pose

également la question de la constitution et de l'évolution d'une liste préétablie de termes lorsque celle-ci est établie.

Pour ce qui est des descripteurs libres, on peut dire qu'ils sont eux aussi extraits du document mais qu'à la différence des mots-clés, ce sont des mots ou des expressions relevés dans le document sans vérification préalable de leur existence dans une liste. L'utilisation de descripteurs libres permet un traitement automatique (donc plus économique) mais au détriment de l'homogénéité et du contrôle du vocabulaire (ambiguïté sémantique de type synonymie ou polysémie par exemple).

Les langages contrôlés : Ils ont pour but la représentation des documents par une codification univoque que l'on retrouve généralement sous la forme de listes d'autorité et de thésaurus.

Une liste d'autorité se présente généralement sous la forme d'une liste de concepts, mots ou expressions issus d'une liste établie à priori. Cette liste est généralement établie par et pour un groupe d'utilisateurs et fait donc l'objet d'un consensus. Elle convient assez bien aux concepts scientifiques et techniques, où l'on peut supposer qu'un concept recouvre un ensemble défini de phénomènes, voire le résultat d'expériences ou de procédés réguliers et répétables mais elle a du mal à s'appliquer à des concepts linguistiques complexes qui sont susceptibles d'interprétations différentes, d'associations et de représentations subjectives chargées de contextes culturels ou affectifs.

Un thésaurus est un langage contrôlé où les concepts sont liés par des relations de hiérarchie et d'association et replacés dans leur champ sémantique (un champ sémantique est un ensemble de mots qui par leur proximité permettent de replacer un concept dans un domaine bien délimité de signification). Grâce au champ sémantique, il est possible de regrouper les concepts d'une manière plus efficace et moins ambiguë et faire ainsi la différence entre concepts voisins et synonymes.

Un thésaurus va pouvoir comporter jusqu'à quatre unités lexicales :

- Les groupes thématiques de descripteurs ;
- Les descripteurs eux-mêmes, appelés aussi termes ;
- Les non-descripteurs appelés aussi termes équivalents ;
- Les mots outils.

Les groupes thématiques de descripteurs se présentent comme des sous-ensembles servant à faciliter la manipulation de listes comprenant un grand nombre de mots (de quelques dizaines

à plusieurs milliers) et à prendre en compte la notion de champ sémantique en opérant des regroupement de termes par thèmes ou spécialités.

Un autre regroupement est parfois opéré en utilisant une technique inventée par Shiyali Ramamrita RANGANATHAN [RANGANATHAN S.R., 1963] à l'Université de Madras en opérant une classification analytico-synthétique par facette [MANIEZ J., 1999]. Cette approche consiste à prendre en considération les différents angles d'intérêt ou d'éclairage que l'on peut apporter à un domaine. Le regroupement par facettes étant toutefois beaucoup plus abstrait, donc plus difficile à appréhender, que le regroupement par thème (qui correspond d'avantage aux classifications du savoir telles qu'elles nous ont été enseignées depuis l'école primaire), il présente néanmoins un gros intérêt dans certains contextes documentaires où l'on s'intéresse globalement à une activité, à une transformation d'un processus existant ou à un mécanisme de création de valeur.

Selon l'Association Française de NORMalisation (AFNOR), *"un descripteur est un terme (mot ou expression) retenu, parmi un ensemble de synonymes, de quasi-synonymes et de termes apparentés, pour représenter, de manière univoque, un concept susceptible d'intervenir dans des documents et des questions traités dans un système documentaire donné et, dès lors, inclus dans le thésaurus de descripteurs de ce système"* [AFNOR, 1987].

Il est important de souligner que dans un thésaurus, un descripteur doit être univoque, c'est-à-dire qu'un seul descripteur désigne un seul concept.

Les non-descripteurs (ou termes équivalents), sont les synonymes, les quasi-synonymes et les termes voisins, liés par une relation d'équivalence sémantique avec un descripteur du thésaurus.

Les mots outils sont des mots de signification imprécise qu'il est préférable de ne pas utiliser comme descripteur, ou alors seulement à la condition d'être liés à un descripteur.

Un thésaurus peut également comporter jusqu'à cinq types de relations sémantiques :

- L'appartenance des descripteurs à un groupe sémantique ;
- L'équivalence sémantique entre descripteurs et non-descripteurs ;
- La hiérarchie de descripteurs ;
- L'association des descripteurs ;
- L'équivalence linguistique.

L'appartenance de descripteurs à un groupe sémantique nous invite à se poser la question de l'intérêt de créer plusieurs groupes sémantiques au sein d'un même thésaurus plutôt que de procéder à la création de plusieurs thésaurus. Il est souvent plus facile de construire un thésaurus à partir d'un groupe sémantique que de regrouper plusieurs thésaurus.

Selon l'AFNOR [AFNOR, 1987], on parle de "*relation hiérarchique entre deux descripteurs lorsque l'un des deux descripteurs est subordonné à l'autre*".

L'intérêt des relations hiérarchiques est de pouvoir répartir les descripteurs par rapport à des niveaux de spécialisation qui prendront tout leur sens lorsqu'un utilisateur possédant une connaissance plus ou moins approfondie d'un sujet effectuera une recherche d'information et trouvera des documents (relatifs au sujet et à l'information recherchée) qu'il jugera dignes d'intérêt (ni trop généralistes, ni trop spécialisés).

Selon l'AFNOR [AFNOR, 1987], "*une relation d'association est une relation qui indique des analogies ou liens de signification entre les descripteurs, autres que la relation hiérarchique ou d'équivalence*".

Ce type de relation traduit le fait que deux concepts peuvent s'évoquer mutuellement par le biais d'une association d'idées et de points de vue.

Cette mise en relation peut traduire le fait que différents concepts ne peuvent être mesurés qu'indirectement à travers leurs manifestations observables. Elle peut également s'appliquer à des concepts similaires ou ressemblants ou illustrer un lien entre un concept et l'action ou le processus qui l'engendre ou bien encore entre un concept et la discipline auquel il est rattaché.

Dans la pratique, ce type de relation n'est pas forcément facile à appréhender et encore moins à mettre en œuvre.

L'équivalence linguistique fixe en principe un lien biunivoque et symétrique entre des descripteurs de langues différentes. Cette définition suppose qu'un descripteur exprimé dans une langue n'a qu'un seul équivalent (descripteur équivalent) dans une autre langue.

Bien que très restrictive, cette définition ne tient pas compte du problème de traduction d'un concept qui n'a pas de correspondance dans une langue étrangère.

On distingue généralement quatre types d'équivalence linguistique :

- L'équivalence réelle entre des concepts identiques possédant la même étymologie ;
- L'équivalence réelle entre des concepts identiques d'étymologie différentes mais de même construction linguistique ;
- L'équivalence partielle qui résulte de découpages conceptuels différents ;

- L'absence de concept équivalent.

Dans la pratique, l'utilisation de descripteurs de langues différentes ne pose pas de problème insoluble, car la terminologie est en principe connue et peut être précisée dans le cadre du groupe utilisant le thésaurus.

Il existe plusieurs notations normalisées des relations d'un thésaurus (notation par sigles normalisée par l'AFNOR, norme internationale ISO, etc.), mais la méthode la plus simple et la plus utilisée est la présentation alphabétique de listes de mots-clés ou de descripteurs libres.

Il est toutefois important de noter que cette méthode présente deux défauts. Le premier repose sur le fait que la consultation des descripteurs est d'autant plus fastidieuse que le volume des descripteurs est important. Le deuxième traduit une utilisation difficile du thésaurus si les concepts employés dans les listes alphabétiques sont composés de groupes de mots.

Une autre méthode couramment employée permet d'éviter ces défauts, il s'agit de la présentation par liste permutée ou chaque descripteur est trié par chacun des mots qu'il contient et se retrouve présent plusieurs fois.

Cette méthode a été généralisée avec le "permuted kwic (key word in context) index" qui présente une liste de mots-clés contenus dans plusieurs descripteurs en utilisant pour chaque affichage une permutation circulaire, où le mot trié par ordre alphabétique est présenté au milieu d'une ligne contenant ses voisins à gauche et à droite.

Notons également que la présentation des descripteurs peut être graphique. Dans la présentation graphique fléchée, tous les descripteurs d'un même champ sémantique sont regroupés à l'intérieur d'un rectangle. Des flèches relient alors les termes spécifiques aux termes génériques et des lignes non fléchées indiquent les relations associatives. Les mots figurant à l'extérieur du rectangle sont liés à ceux du champ sémantique mais appartiennent à d'autres champs sémantiques.

Le rectangle est parfois quadrillé en cases portant des coordonnées permettant ainsi de localiser chaque descripteur. Le rectangle est lui-même affecté d'un indice et porte le nom du descripteur le plus générique ou du champ sémantique.

Selon Jean-Yves PRAX [PRAX JY., 2004], *"l'indexation, qu'elle soit structurée ou en texte intégral, est le moteur de la recherche documentaire. L'index constitue une série de filtres entre :*

- *L'information réelle avec toute sa richesse ;*
- *Son stockage dans l'ordinateur (indexation) ;*

- *L'information recherchée (requête) par l'utilisateur ;*
- *Les articles extraits par l'ordinateur en réponse à la requête".*

Afin de prendre en compte l'apport d'un mode d'indexation par rapport à un autre, Jean-Yves PRAX [PRAX JY., 2004] introduit également la notion de filtres "a priori" ou "a posteriori".

Dans le cas d'un filtre a priori, un utilisateur chargé de l'indexation de documents renseigne les index ou mots-clés qui serviront une fois pour toutes aux recherches ultérieures (les index ou mots-clés désignent des concepts dont le sens est communément admis); alors que dans le cas d'un filtre a posteriori (texte intégral) les documents sont indexés automatiquement par le système, sans aucune intervention humaine ; tous les mots (hormis les mots vides) deviennent des index possibles de recherche.

C'est cette même notion que l'on retrouve chez Widad MUSTAFA EL HADI qui met également en avant les concepts d'indexation manuelle et d'indexation automatique.

Pour Widad MUSTAFA EL HADI [MUSTAFA EL HADI W., 2006], qui s'appuie sur les travaux de Jacques MANIEZ [MANIEZ J., 2002] et Philippe LEFEVRE [LEFEVRE P., 2000], *"l'indexation manuelle est une opération qui consiste à recenser les concepts dont traite un document et à les représenter à l'aide d'un langage documentaire (classificateur ou combinatoire) souvent par plusieurs indicateurs sémantiques : indice de classification, descripteurs libres, termes autorisés, descripteurs ou mots-clés d'un thesaurus"*.

Cette définition fait ressortir le fait que derrière le concept d'indexation manuelle se cache un important travail de mise au point qui suppose la connaissance du contenu de l'information, le choix des concepts à représenter et la traduction de ces concepts en descripteurs.

Widad MUSTAFA EL HADI précise également que l'utilisateur est totalement absent de cette indexation qui se fait en amont (les listes de concepts et de mots-clés sont constitués en amont et indépendamment des documents à indexer) et que l'évolution du vocabulaire de l'indexation n'évolue pas en fonction de celle des documents indexés pris isolément (cela est essentiellement dû à la difficulté de mettre à jour en temps réel les listes de concepts ou de mots-clés en fonction de l'évolution des connaissances) [MUSTAFA EL HADI W., 2006].

Pour Widad MUSTAFA EL HADI [MUSTAFA EL HADI W., 2006], c'est en raison de ces problèmes que des méthodes d'indexation automatique sont apparues *"consistant à faire reconnaître par l'ordinateur des termes figurant dans le titre, le résumé, le texte complet (s'il est enregistré avec la notice documentaire) et à employer ces termes, soit tels quels soit après*

conversion en d'autres termes équivalents ou conceptuellement voisins, pour en faire des critères incorporés dans le fichier de recherche et utilisables pour retrouver des documents".

Elle souligne également le fait que la première génération des systèmes d'indexation automatique sur le texte intégral était fondée sur la création d'index (fichiers inversés) permettant de localiser un terme ou un ensemble de termes au sein d'une collection de documents.

Par ailleurs, et dans le cadre de la recherche d'information sur le Web, d'autres dispositifs terminologiques viennent aujourd'hui compléter ce panorama d'outils. Il s'agit de réseaux sémantiques, de cartographie conceptuelle et d'ontologies (ensemble structuré de termes et de concepts représentant le sens d'un champ d'informations) qui sont issus des développements en intelligence artificielle et plus récemment en ingénierie des connaissances.

Selon Widad MUSTAFA EL HADI [MUSTAFA EL HADI W., 2006], *"ce sont des structures d'organisation et de représentation des connaissances innovantes qui font appel aux technologies du Web. Conçues sur la base de la modélisation conceptuelle d'un domaine de connaissance, ces structures sont de plus en plus utilisées dans des applications de gestion des connaissances, de veille technologique, d'indexation et de recherche d'information. C'est surtout dans les fonctions d'indexation et de recherche d'information que les systèmes d'organisation des connaissances classiques et innovants convergent vers des points communs"*.

c) Les principales méthodes de recherche

Dans un système d'information documentaire, les documents sont généralement organisés autour de leur descriptif et/ou de leur contenu.

Il est important de noter que cette organisation est directement liée au mode de lecture et à l'approche de recherche d'informations souhaitée.

Elle a pour principal objectif de résoudre efficacement le problème du repérage et de l'accessibilité aux informations.

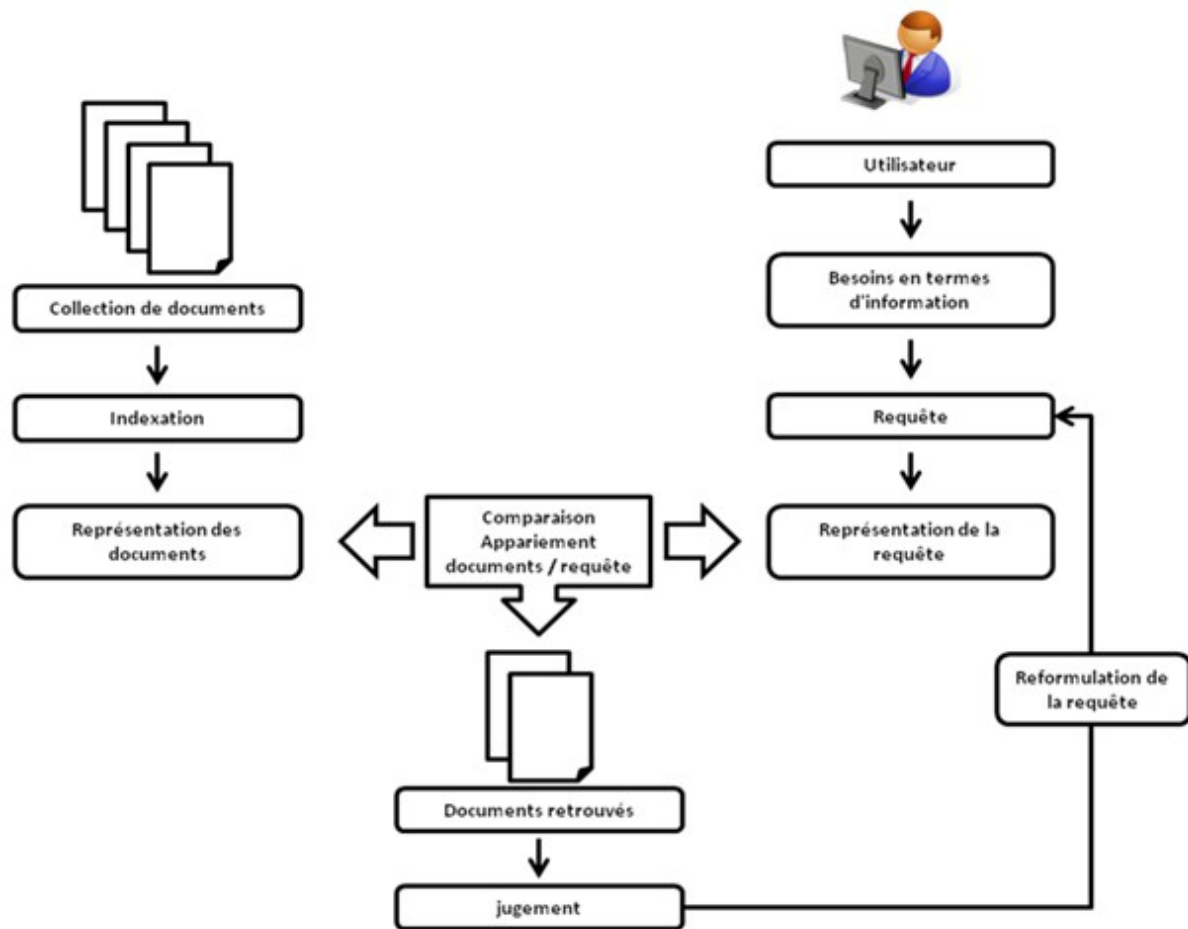
Mais de cette organisation dépend également les outils d'interrogation et de recherche de l'information qui seront mis en œuvre dans le cadre d'un système d'information documentaire.

Pour bien comprendre la mise en relation des besoins utilisateurs et des informations, intéressons-nous dans un premier temps au processus en U de Recherche d'Information [SALTON G. et MC GILL M., 1984].

Selon Gérard SALTON, ce processus se compose des fonctions principales suivantes :

- L'indexation des documents et des requêtes ;
- L'appariement requête-document, qui permet de comparer la requête et le document, et de calculer la similarité entre ces deux éléments. La pertinence du document vis-à-vis de la requête est représentée par un calcul de score. Ce score est calculé à partir d'une fonction ou d'une probabilité de similarité. Cette mesure tient compte du poids des termes dans les documents, déterminé en fonction d'analyses statistiques et probabilistes. La fonction de similarité permet d'ordonner les documents renvoyés à l'utilisateur. Cet ordonnancement joue un rôle primordial puisque l'utilisateur se contente la plupart du temps d'examiner les premiers documents affichés ;
- La fonction de modification, qui intervient en réponse aux résultats obtenus (les modifications pouvant concerner les documents ou la requête). Les modifications sur les documents se manifestent par l'ajout ou la suppression dans la base de données. Tandis que les modifications concernant les requêtes, et qui sont les plus courantes, se révèlent dans ce qu'on appelle la reformulation de la requête.

Figure 3 : Processus en U de recherche d'information



Le processus en U de Recherche d'Information présente trois propriétés importantes :

- L'appariement requête-document se fait entre les index des documents et la requête. Pour cela il faut que les index représentent le plus possible le contenu sémantique du document ;
- Si l'on considère que les représentations des documents et des requêtes sont correctes, il faut encore que la fonction de correspondance qui permet d'évaluer la pertinence des documents soit de bonne qualité ;
- Enfin, pendant la phase de reformulation de la requête, il faut que la traduction de la requête de l'utilisateur (expression des besoins mentaux) soit correctement effectuée par le système de recherche d'information. Ceci en considérant que la requête exprimée par l'utilisateur représente vraiment ses besoins, ce qui est loin d'être toujours le cas.

Pour être en mesure d'offrir aux utilisateurs les informations correspondants à leurs attentes, une solution de recherche d'information doit s'appuyer sur le concept de pertinence (dans le domaine de la recherche d'information, la pertinence est l'inverse du bruit).

Ce concept, évoqué pour la première fois en 1948 par Samuel C. BRADFORD dans son livre intitulé "Documentation" [BRADFORD S.C., 1948], a donné lieu depuis à toute une série d'évaluations expérimentales et de modèles informatiques, faisant de la pertinence un des concepts clé de la problématique de la recherche d'information.

Il n'en demeure pas moins qu'évaluer la pertinence d'un document sur un sujet donné n'est pas chose aisée (ce concept a été plus ou moins précisément défini par la norme ISO 11620) [ISO, 2008].

La notion de taux de pertinence souvent évoquée se résume à un ratio entre le nombre de documents effectivement retrouvés dans une base documentaire lors d'une recherche d'information et le nombre de documents de cette base qui concernent l'information recherchée [RIVIER A., 2007].

De plus, il convient de distinguer, la pertinence objective, mesurée par le système de recherche documentaire et la pertinence subjective qui mesure le point de vue de l'utilisateur.

Selon Jean-Yves PRAX [PRAX JY., 2004], *"la pertinence objective n'est légitime que si l'indexation ne comporte que des descripteurs objectifs"*, c'est-à-dire choisis dans un vocabulaire normé et que les termes employés dans la question (formulation de la recherche d'information) fassent également partie de ce vocabulaire.

Pour ce qui est de la pertinence subjective et selon Alexandre SERRES [SERRES A., 2002-2004], *"elle répond à un besoin subjectif d'information et ne peut donc être évaluée de manière universelle"*.

Ainsi, l'évaluation de la pertinence peut se révéler d'une complexité redoutable, mais dès lors que l'on ne cherche pas à en faire une opération technique, cette complexité peut constituer en retour une grande richesse pour l'utilisateur.

Selon André TRICOT et Jean-François ROUET [TRICOT A. et ROUET JF., 1998], la recherche d'information peut alors se définir comme *"l'ensemble des activités, des processus, des représentations, des transformations et des inférences cognitives qu'élabore et met en œuvre un individu pour trouver des informations utiles et pertinentes par rapport à une tâche qu'il se donne"*.

Cette approche cognitive de la recherche d'information ne prend pas seulement en compte les aspects techniques liés à l'étude des solutions de recherche d'information qui instrumentent au sens de supporter, outiller ces activités de recherche mais aussi les aspects mentaux qui sont au cœur de toute activité de recherche d'information.

Cette définition nous permet également d'introduire une notion importante qui se cache derrière toute activité de recherche d'information. Cette activité peut être assimilée au fait de chercher à "retrouver" une information, ce qui signifie que le domaine de recherche est un ensemble défini à priori, ou bien encore à "trouver" de l'information, auquel cas, la définition du domaine de recherche doit aussi faire l'objet d'une recherche particulière.

L'ensemble des méthodes et techniques adoptées pour rechercher de l'information tiennent compte de ces deux catégories d'activité de recherche.

C'est pourquoi, nous allons consacrer dans ce qui suit un très long développement à ces méthodes et techniques de recherche.

Historiquement, plusieurs approches peuvent être distinguées :

- **L'approche ensembliste** qui considère que l'ensemble des documents s'obtient par une série d'opérations (intersection, union et le passage au complémentaire) ;
- **L'approche algébrique** (ou vectorielle) qui considère que les documents et les requêtes sont représentés par des vecteurs ;
- **L'approche probabiliste** qui essaie de modéliser la notion de pertinence.

Depuis l'introduction du terme "recherche d'information" (information retrieval) par Calvin Northrup MOOERS en 1948 et la première conférence internationale sur le thème en 1958, de nombreux modèles de recherche ont vu le jour se réclamant de ces trois principales approches. Hans Peter LUHN, fin des années 50, décrit un modèle statistique pour la recherche d'information (KWIC - Key Words In Context) [LUHN HP., 1957] puis, en 1960, M.E. MARON & J.L. KUHNS [MARON M.E. et KUHNS J.L., 1960] définissent un modèle de recherche d'information probabiliste. Les modèles booléens et vectoriels pour la recherche d'information connurent également un fort développement dans les années 60 et 70 de même que les années 90 virent l'essor des modèles probabilistes.

Selon Gérard SALTON [SALTON G. et MC GILL M., 1984], le premier et le plus simple de ces modèles est le **modèle booléen** (approche ensembliste). Il est basé sur la théorie des ensembles et l'algèbre de BOOLE. Le modèle booléen traite une requête comme une expression logique. Si D est un document et T un terme, on définit l'expression $R(D,T)$

comme étant vraie si le terme se trouve dans le document et comme fausse, s'il en est autrement. Souvent, en informatique, la valeur "FAUX" est représentée par l'entier "0" et la valeur "VRAI", par l'entier "1".

On définit trois opérateurs ET, OU et SAUF (AND, OR et NOT en anglais) qui influent sur les valeurs booléennes :

- $x \text{ ET } y$: est vrai si et seulement si x et y sont vrais.
- $x \text{ OU } y$: est vrai si et seulement si x ou y , ou les deux, sont vrais.
- $x \text{ SAUF } y$: est vrai si et seulement si x est vrai et y est faux.

Ainsi, l'expression $R(D, \text{Taux ET Inflation})$ sera vraie si et seulement si le document contient à la fois le mot "Taux" et le mot "Inflation". L'expression $R(D, \text{Taux OU Inflation})$ sera vraie si le document contient le mot "Taux", le mot "Inflation" ou les deux. L'expression $R(D, \text{Taux SAUF Inflation})$ sera vraie si et seulement si le document contient le mot "Taux", mais pas le mot "Inflation".

Dans ce type de modèle, le document est pertinent si et seulement si son contenu respecte la formule logique demandée. On parle alors de pertinence binaire dans la mesure où un document est soit pertinent soit non pertinent par rapport à une requête.

Le modèle booléen est transparent et simple à comprendre pour un utilisateur. Il est relativement bien adapté pour des spécialistes qui ont un vocabulaire restreint lié à leur domaine de spécialité.

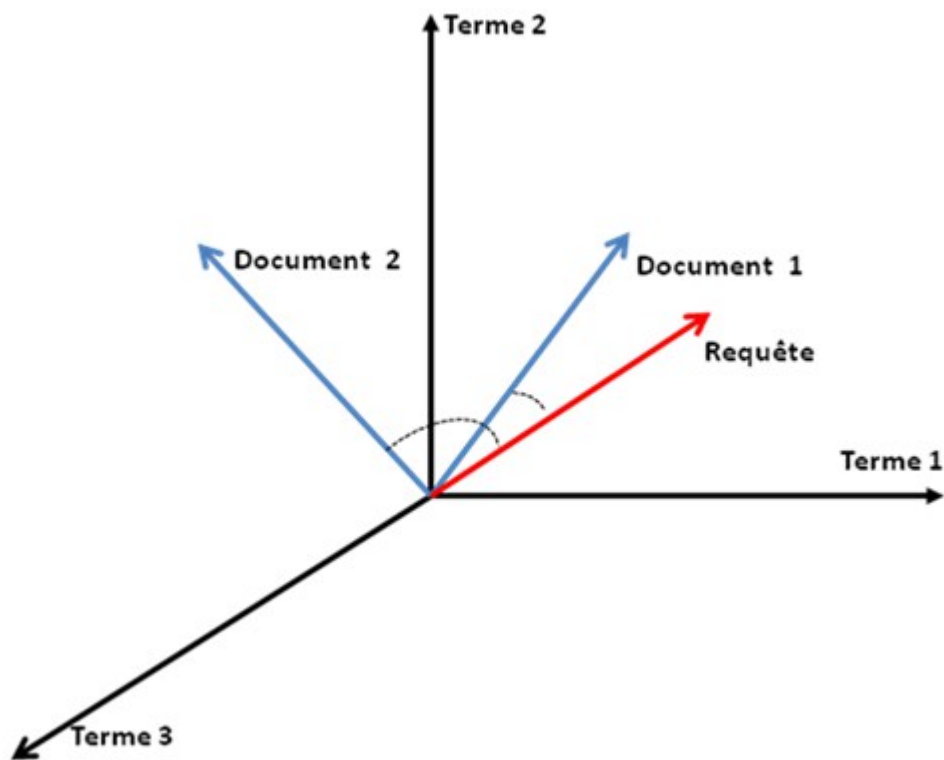
Par contre, ce type de modèle présente aussi un certain nombre d'inconvénients. Notamment, parce qu'il est parfois difficile d'exprimer des requêtes longues sous forme booléenne ou de ne prendre en compte que le seul critère binaire dans une recherche d'information. Enfin, ce type de modèle ne permet pas d'ordonner les résultats ni d'en opérer un classement.

De nos jours, très peu de systèmes de recherche d'information utilisent le modèle booléen standard. C'est plutôt une extension de ce modèle qui est implémentée. Dans le modèle booléen étendu, on préserve la structure d'interrogation inhérente au système booléen, mais en même temps on peut incorporer des termes pondérés, à la fois dans les requêtes et dans les documents stockés.

Il est clair qu'une pondération non binaire des termes de l'index peut amener à des améliorations notables des performances. La pondération de ces termes nous amène donc à introduire le modèle vectoriel.

Parmi les modèles qui relèvent de l'approche algébrique, le **modèle vectoriel** est un modèle statistique où documents (corpus des documents indexés) et requête sont représentés par un vecteur.

Figure 4 : Représentation vectorielle de deux documents et d'une requête dans un espace composé de trois termes



Selon Gérard SALTON [SALTON G., 1970], il tient compte de l'aspect quantitatif des termes et des documents car ses coordonnées sont exprimées dans un espace euclidien à n dimensions (où n correspond au nombre de termes) et sa longueur est proportionnelle au poids des termes.

La pertinence du document dans ce type de modèle correspond au degré de similarité entre le vecteur de la requête et celui du document.

Dans ce type de modèle, le langage de requête est plus simple (de par l'utilisation d'une liste de mots-clés) et les performances sont meilleures grâce à la pondération des termes.

Il est également important de noter que l'évaluation de la pertinence est meilleure (car la pertinence peut-être partielle) et qu'il est possible de trier les documents.

Par contre, le langage de requête est moins expressif et le modèle moins transparent que dans le cadre d'une approche ensembliste et suppose pour être pertinent que tous les termes soient indépendants.

En matière de recherche d'information, le modèle vectoriel est à ce jour le modèle le plus utilisé, notamment parce que les résultats d'une telle approche de recherche d'information peuvent être comparés à de nombreuses nouvelles méthodes d'ordonnement.

Les **modèles probabilistes** en recherche d'information (approche probabiliste) sont importants parce qu'ils représentent une des tentatives les plus significatives pour donner une base théorique solide à la recherche d'information.

Dans les modèles probabilistes, on cherche à estimer la probabilité de pertinence d'un document par rapport à une requête. Selon S.E. ROBERTSON [ROBERTSON S.E., 1977], l'idée est de retrouver des documents qui ont en même temps une forte probabilité d'être pertinents, et une faible probabilité d'être non pertinents.

Dans ce type d'approche, le postulat de départ est que la recherche d'information constitue un processus incertain et imprécis. L'incertitude est dans la représentation des informations tandis que l'imprécision se manifeste dans l'expression des besoins. Ce modèle tend alors à estimer la probabilité qu'un document donné soit pertinent pour une requête donnée, donc de mesurer cette incertitude et cette imprécision.

Les modèles probabilistes retournent de bons résultats par rapport aux modèles booléens et sont indépendants du domaine d'application. Cependant, ils présentent un obstacle majeur dans les méthodes d'estimation des probabilités utilisées pour évaluer la pertinence.

Il existe d'autres modèles pour la recherche d'information issus des travaux en apprentissage dans le domaine de l'intelligence artificielle. Il s'agit notamment du modèle neuronal et du modèle inférentiel. Dans le modèle neuronal, c'est le niveau d'activation final des nœuds documents qui donne la mesure de pertinence utilisée dans le rangement final [KWOK K.L., 1995]. Dans le modèle inférentiel, la pertinence d'un document à une requête correspond au degré de croyance que l'observation du document va satisfaire le besoin de l'utilisateur qui a formulé la requête [TURTLE H. et CROFT W.B., 1990].

Pour échapper au problème du "tout ou rien" qui accompagne les recherches textuelles classiques, de plus en plus de chercheurs s'intéressent aujourd'hui aux technologies avancées de traitement linguistique.

Parmi ces technologies, les clusters sémantiques occupent une place particulière. Un cluster sémantique se présente sous la forme d'une grappe de concepts reliés à des degrés divers en fonction de leur contribution au sujet d'intérêt principal [HAAV H.M. et LUBI T.L., 2001].

La grappe de concepts dépend fortement du domaine sémantique dans lequel l'utilisateur se place (par exemple : un financier qui s'intéresse aux cours du cuivre en dollars n'aura pas le même cluster sémantique qu'un acheteur qui suivra également ces valeurs par pays producteur, par société, etc.).

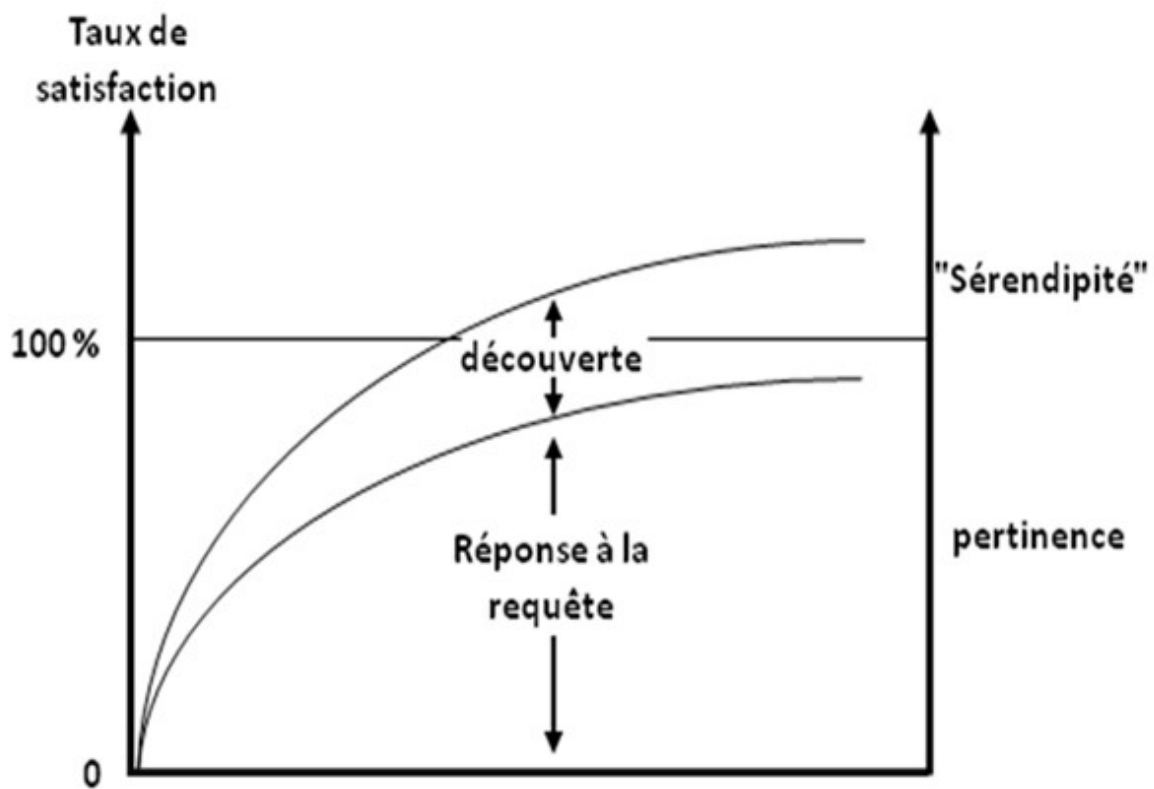
A la différence du thésaurus, les concepts ne sont pas liés par des relations linguistiques, mais par la contribution, le renseignement, qu'ils apportent au sujet d'intérêt principal.

Dans certains cas, il convient également de s'intéresser à l'évolution de la connaissance d'un utilisateur sur un thème donné au cours du temps et à l'évolution du thème lui-même. Sachant qu'à base documentaire constante, un système d'information documentaire est fiable si pour une même requête répétée plusieurs fois, on obtient le même résultat.

Si l'on intègre ces différents éléments dans un système de recherche d'information où un agent de veille documentaire (ce concept s'appuie sur un mécanisme de requête répétée permettant à un profil type d'utilisateur de recevoir des informations sur un ou des sujets donnés) proposerait tous les jours à un utilisateur un certain nombre de documents, le seul critère de pertinence ne suffirait plus et présenterait même le risque d'un appauvrissement progressif du ou des sujets.

Par ailleurs, dans le cadre d'un fonds documentaire important et un mode d'indexation riche, on constate que le taux de pertinence est une courbe asymptotique dont les valeurs en ordonnées sont comprises entre 0 et 1. Le taux de satisfaction de l'utilisateur (ou taux de pertinence subjective) peut cumuler deux variables, le taux de pertinence et le taux de découverte que bon nombre d'auteurs appellent "la rencontre fortuite" ou l'effet "Sérendipité" [VAN ANDEL P. et BOURCIER D., 2009].

Figure 5 : Taux de satisfaction, sérendipité et pertinence



Selon Jean-Yves PRAX [PRAX JY., 2004], *"dans l'activité cognitive d'un professionnel, la rencontre fortuite avec la connaissance est primordiale : elle favorise la découverte de ce qu'on ne cherchait pas mais qui se révèle répondre à des aspirations tacites, elle peut être la clé déclenchant la compréhension soudaine d'une situation complexe ou ambiguë, elle peut être le catalyseur de la créativité et de l'innovation"*.

Dans le domaine de l'analyse statistique et linguistique, certains auteurs [TOUSSAINT Y. et AL., 1997] considèrent les mots-clés comme des indicateurs de connaissance (contenu des documents indexés). Cette méthode permet de mettre en évidence des affinités de sens entre les mots-clés et grâce à un algorithme de classification hiérarchique de construire des clusters (groupes de mots proches les uns des autres). Elle constitue une approche qualitative de la cooccurrence (présence simultanée de deux ou plusieurs mots-clés) mais ne permet pas de mesurer la force des associations. D'autre part, elle a l'inconvénient d'avantager les mots-clés de haute fréquence par rapport à ceux de basse fréquence sans tenir compte de la proximité.

Il est possible de calculer des indicateurs de cooccurrence et de représenter la densité moyenne des associations entre mots-clés au sein d'un cluster. L'objectif est alors de pouvoir positionner ces valeurs sur une carte et de repérer les clusters les mieux structurés du point de vue de leur densité mais aussi de leur centralité.

Cette cartographie de concepts également appelée "diagramme stratégique" permet d'évaluer l'intérêt stratégique des différents clusters présents au sein d'un corpus documentaire et d'en appréhender globalement mais aussi localement le contenu [MIHALCEA R. et MOLDOVAN D.I., 2000].

D'autres approches (modèles connexionnistes) acceptent de traiter des données complexes, incomplètes, bruitées et parfois même partiellement contradictoires. Elles possèdent la capacité d'apprentissage et s'intéressent aux données "floues". C'est le cas notamment de l'approche neuronale [BESSAI F.Z et AL., 1997] qui établie une analogie entre les neurones biologiques et des neurones formels.

Dans les réseaux de neurones (ou réseaux connexionnistes), toutes les connaissances sont représentées par des liaisons entre les unités (neurones) et leurs poids synaptiques (valeurs) associés, par analogie avec la neurobiologie. Un réseau de neurones peut être représenté comme un graphe orienté et pondéré. Les nœuds de ce graphe sont des automates simples nommés neurones formels ou tout simplement unités du réseau. Un neurone formel possède plusieurs entrées d'information, est doté d'un état interne, que l'on appelle état d'activation, et possède une fonction qui permet de calculer une valeur de sortie selon son état d'activation.

En matière de traitement linguistique, [BOURIGAULT D. et JACQUEMIN C., 2000] et [JACQUEMIN C., 2001], il convient de faire appel à plusieurs niveaux d'analyse :

- L'analyse morphologique, qui vise à ramener tous les mots connus dans une phrase à leur forme canonique, en séparant les variations grammaticales (pluriels, conjugaisons, etc.) ;
- L'analyse lexicale qui cherche à identifier le lexème, qui est l'unité élémentaire de signification et qui suppose la prise en compte du domaine ou du contexte ;
- L'analyse syntaxique, qui introduit la notion de syntagme qui analyse l'arrangement des mots, sujets, verbes et compléments selon une unité fonctionnelle dans une phrase. L'analyseur syntaxique doit donc être capable d'utiliser des règles de grammaire ;
- L'analyse sémantique, qui analyse le sens afin de fournir une représentation symbolique du sens des phrases et en se basant sur l'hypothèse que le sens est une configuration particulière de symboles.

D'autres recherches en ingénierie linguistique appliquent avec originalité les outils statistiques et mathématiques à l'analyse textuelle et lexicale [NORMIER B., 2007].

Ces méthodes facilitent l'étude des corpus sémantiques et ouvre la voie de l'analyse lexicale.

Ainsi, à travers un acte de langage, un acteur humain peut dévoiler un certain nombre de postures décelable par l'analyse (engagement, désintérêt, critique, etc.). Encore peu répandus, ces outils commencent à trouver des applications dans différents domaines dont l'analyse de complétude et de non-ambiguïté de la documentation technique.

L'ensemble de ces approches permettent de mieux comprendre le comportement d'un système de recherche d'information et fournissent un cadre théorique pour la modélisation de la mesure de pertinence.

Aussi, et après avoir énoncé les principes généraux de l'indexation et de la recherche d'information, nous allons à présent traiter des spécificités de l'indexation et de la recherche d'information dans le domaine de la Gestion Electronique de Documents.

II. Indexation et recherche d'informations dans le domaine de la Gestion Electronique de Documents

La constitution puis le maintien d'une mémoire d'entreprise reposent de manière indispensable sur la Gestion Electronique de Documents dont l'objectif est de permettre aux différents acteurs d'une organisation de retrouver rapidement l'information pertinente qui se dissimule au sein des documents qui pénètrent, sortent ou circulent à l'intérieur de l'entreprise.

La phase d'indexation représente alors le cœur d'un système de Gestion Electronique de Documents dans la mesure où elle se veut l'opération de représentation la plus fidèle et la plus exhaustive du contenu et du sens d'un document en vue de faciliter la recherche d'information et d'en garantir la pertinence.

C'est pourquoi, dans un premier temps, nous décrirons l'évolution des solutions de GED.

Dans un deuxième temps, nous nous intéresserons à la nature de ces solutions en termes d'applications, avant de présenter les principaux outils d'indexation, leur typologie, leur construction et leur utilisation, en les replaçant dans le contexte d'évolution des technologies de la GED et en s'attardant sur les contraintes spécifiques liées à la multiplicité des applications de Gestion Electronique de Documents.

1) De l'archivage électronique à la capitalisation des connaissances

A l'ère de l'information et de la dématérialisation des échanges, les nouvelles technologies de l'information et de la communication constituent un composant technologique essentiel de la capitalisation de la connaissance et de l'intelligence collective.

Dans ce contexte, la capitalisation de la connaissance qui consiste à identifier, formaliser et conserver la mémoire des activités de l'entreprise, doit permettre de rendre cette mémoire accessible indépendamment des acteurs qui l'ont créée et de manière pertinente par rapport à un contexte d'intérêt donné.

La fonction documentation (mémorisation de l'information) devient alors stratégique dans la phase de mise en place de processus de travail collaboratif, de capitalisation et d'échanges informationnels, dont l'objectif est la recherche de nouveaux gisements de productivité, d'avantages compétitifs mais aussi de facteurs d'innovation et de qualité.

L'outillage de la fonction documentaire et plus particulièrement des phases de recherche d'information et de mémorisation de l'information par le biais du stockage et de l'indexation de son support matériel, nous incite à considérer la GED comme un composant essentiel du système d'information et de communication de l'entreprise.

A partir des années 95 et avec l'expansion des réseaux, les applications de GED ont cessé d'être focalisées sur l'archivage électronique pour devenir des solutions économiques et stratégiques destinées à favoriser l'émergence d'une nouvelle forme d'entreprise : l'entreprise "apprenante".

Cette évolution du concept et des fonctionnalités a été tellement forte que certains éditeurs ont décidé, en commun, de substituer à l'acronyme GED celui de GEIDE (Gestion Electronique d'Informations et de Documents Existants) afin d'étendre le concept en introduisant la notion plus large de gestion d'informations. Le béotien qui cherche à s'informer sur le sujet peut être dérouté par la multiplication des définitions qui apparaissent pour l'acronyme GEIDE. On peut lire parfois "Gestion Electronique de l'Information et des Documents Existants", ou "Gestion Electronique de l'Information Documentaire des Entreprises" ou encore "Gestion Electronique d'Informations et de Documents Existants". Certains détournent même la terminologie pour lui faire épouser le concept très en vogue de "Gouvernance" ce qui donne : "Gouvernance Electronique des Idées et des Décisions en Entreprise".

Mais au-delà des "idées" et des "décisions", ce sont, surtout, les routines et les procédures administratives de l'entreprise qui sont adressées par la GED. Ainsi un Workflow documentaire intégré à une GED permet de créer des procédures de traitement des documents et de garantir leur suivi.

Pour que tous les documents sortants d'une entreprise puissent se retrouver dans la GED, certains éditeurs proposent des modules ou connecteurs permettant de l'alimenter avec les éditions des applicatifs de gestion. Dans ce cas, la copie électronique de chaque document édité est conservée avec les différents index qui permettent de la classer.

Pour que tous les documents entrants puissent être interceptés par la GED, certains éditeurs proposent des logiciels de RAD (Reconnaissance Automatique de Documents)/LAD (Lecture Automatique de Documents) qui sont destinés à automatiser la reconnaissance et l'indexation multicritère des documents.

On trouve sur le marché, différentes familles de GED qui correspondent à des utilisations précises.

2) La multiplicité des applications de GED

Selon Jean-Yves PRAX [PRAX JY., 2004], *"La GED est à vocation universelle. Elle trouve son utilité dans tous les concepts et dans tous les lieux où les documents abondent. Elle permet d'accéder de façon sélective à des documents en quelques secondes là où une recherche manuelle des dossiers aurait nécessité plusieurs minutes, voire plusieurs heures"*.

Les applications de GED sont avant tout des ensembles fédérateurs d'outils qui viennent compléter des applications bureautiques ou autres de fonctions de traitement de documents électroniques [BODIN B., ROUX-FOUILLET JP., 1992].

Grâce aux technologies de réseaux et de télécommunication, elles facilitent également la collaboration, la coordination des acteurs et abolit les contraintes de temps et d'espace en permettant le partage des documents sans nécessairement imposer de déplacements physiques. Si l'on ajoute une application de Workflow à la GED, ce partage peut-être agencé selon des procédures qui permettront la répartition de tâches entre plusieurs acteurs suivant un circuit donné.

Bien que l'offre mondiale en matière d'applications de GED soit particulièrement bien fournie et répond à la plupart des besoins, il est malgré tout possible d'établir une classification des

applications de GED en cinq grands domaines selon leur finalité principale à savoir la conservation, l'archivage et la diffusion d'informations :

La GED administrative et Bureautique : L'intégration et la banalisation des applications de GED comme outils de traitement dans les systèmes bureautiques constituent l'essentiel des applications de cette catégorie. Ces applications s'intéressent aux documents vivants de l'entreprise ainsi qu'aux échanges et aux flux d'information.

Le COLD (Computer Output on Laser Disc) s'intéresse au stockage et à l'archivage des données issues de traitements informatiques. Sous l'appellation COLD (Computer Output Laser Disc) sont regroupés les applicatifs conçus pour gérer des fichiers spools générés par des ordinateurs centraux ou des systèmes transactionnels. Il peut s'agir par exemple de factures, de bons de livraison, de fiches de paie, etc. Le COLD complète ou remplace les traditionnelles éditions de microfiches COM (Computer Output Microfilm) ou les listings.

La GED image : On regroupe essentiellement sous ce terme, l'ensemble des applications de GED qui gèrent des images ou des documents numérisés. Cette segmentation est purement formelle car une même application de GED est bien souvent capable de gérer des images, des documents numérisés et des fichiers issus d'applications bureautiques ou de logiciels métiers.

La GED documentaire : Il s'agit des applications de type bibliothéconomique ou documentaire qui présentent à la fois les aspects les plus variés mais aussi les plus importants de la GED. Ces applications gèrent les documents de référence de l'entreprise. L'objectif est de constituer une mémoire organisationnelle de l'entreprise.

La GED métier : Seules les applications spécifiques propres à un métier et qui demandent des solutions sur mesure sont concernées par cette catégorie.

Par ailleurs, il existe des applications de GED pour la majeure partie des plateformes informatiques et des systèmes d'exploitation du marché que ce soit sur des micro-ordinateurs, des mini-ordinateurs ou en liaison avec des ordinateurs centraux.

3) Spécificités de l'indexation et de la recherche d'information dans le domaine de la GED

Aux vues des éléments précédemment énoncés, il est clair que toutes les applications de GED répondent à un besoin de conservation et de partage des documents générés ou reçus et d'une partie des informations qui s'y rapportent.

Par contre, toutes ces applications ne fournissent pas les mêmes possibilités. Afin de mieux cerner les spécificités de l'indexation et de la recherche d'information dans le domaine de la GED, il nous paraît nécessaire de s'intéresser dans un premier temps aux premiers maillons de la chaîne GED à savoir : l'acquisition numérique et la reconnaissance.

Pour que l'information soit gérable par un ordinateur, elle doit d'abord être disponible sous forme numérique (suite de 0 et de 1). Trois cas sont alors possibles :

- L'information est déjà en mode numérique et stockée dans des fichiers clairement identifiés et localisés dont le format et la forme sont connus (informations textuelles et/ou images au format numérique dont le format et la forme sont connus). On parle alors d'acquisition "directe" ;
- L'information est en mode numérique mais éparpillée et surtout disparate dans son format et dans sa forme (informations contenues dans des documents numériques de nature et de format différents). On parle alors d'acquisition par "conversion de format" ;
- L'information est sur un support analogique qu'il faudra convertir au format numérique (informations contenues dans des documents papier numérisés à l'aide d'un scanner). On parle alors de "capture" et d'acquisition par "conversion".

Les documents traités par les organisations proviennent de différents points d'entrée et nécessitent donc des techniques d'acquisition, de conversion et de capture différentes.

Parmi ces techniques on distingue :

- L'acquisition directe ou par conversion de format de fichiers numériques ;
- Les techniques d'acquisition des télécopies et des courriels envoyés et reçus ;
- Les techniques d'acquisition par numérisation à l'aide de scanners et de cartes d'acquisition (capture d'images, de sons et de vidéos) ;
- L'interprétation directe de sources informatiques comme les spools d'impression ou les fichiers d'éditions provenant des différentes applications métiers.

Généralement, la plupart des applications de GED présentes sur le marché ne couvrent pas l'acquisition de sources d'informations aussi diversifiées. Si bien que les notions d'acquisition/capture se limitent bien souvent dans les organisations à de la numérisation de documents papier.

Parmi les premiers maillons de la chaîne GED, il convient de citer les dispositifs de reconnaissance de caractères, d'images, de documents et de contenu que bon nombre d'auteurs considèrent parfois comme faisant partie des techniques d'acquisition [PRAX JY., 2004], [CHAUMIER J., 1996].

Le type de reconnaissance est fonction de la nature des documents (supports matériels de l'information) :

- Reconnaissance de codes à barres : Un code à barres se présente sous la forme d'une succession de barres claires et foncées d'épaisseur variable, traduisant une suite de caractères numériques ou alphanumériques selon un "alphabet" défini et destinée à être interprétée automatiquement par un lecteur optique.
- Reconnaissance d'images (ou de formes) : Cette technique consiste à identifier des formes pré-décrites dans une image numérique, et par extension dans un flux vidéo numérique.
- Reconnaissance de texte par OCR (Optical Character Recognition). Il s'agit d'un procédé permettant de convertir l'image numérisée d'un texte en un fichier texte.
- Reconnaissance de caractères manuscrits par ICR (Intelligent Character Recognition) : Cette technique peut être considérée comme assez proche de l'OCR au sens où il s'agit d'extraire le contenu textuel d'un document ayant fait l'objet d'une numérisation. Cependant, par rapport à l'OCR, l'ICR ajoute la technique de reconnaissance de caractères manuscrits et intègre également un mécanisme d'apprentissage de nouveaux caractères qui permet au moteur d'ICR d'améliorer les performances lors des reconnaissances qui suivent.
- Reconnaissance et lecture de formulaires : Cette technique se présente sous la forme d'un traitement automatique de questionnaires avec cases à cocher et comportant parfois des champs manuscrits.
- Reconnaissance automatique de documents (RAD) : La technologie RAD permet de reconnaître le type d'un document et d'en analyser son contenu. Après une analyse de la structure d'un document, ce dispositif est capable de le comparer à des modèles stockés dans sa base de données et d'en déterminer la nature. Selon la nature du document traité, une analyse du contenu est alors possible et une extraction de différentes informations pourra alors être opérée par l'utilisation conjointe d'une

application de lecture automatique de documents (LAD) et de reconnaissance optique de caractères (OCR) ou bien encore d'ICR.

Compte tenu des éléments que nous venons d'énoncer, il est clair que l'objectif de l'ensemble de ces dispositifs (d'acquisition, de capture, de conversion et de reconnaissance) est de faciliter la phase d'indexation.

a) L'indexation

Bien que l'indexation soit considérée comme le cœur de la GED au sens où elle doit permettre la représentation la plus fidèle et la plus exhaustive du contenu et du sens de l'ensemble des documents, il ne faut pas perdre de vue que dans la plupart des organisations :

- L'acteur humain qui recherche une information est également un producteur potentiel de documents porteurs d'informations ;
- La politique d'indexation est co-conçue par ces utilisateurs consommateurs et producteurs d'informations ;
- Le choix des documents qui viennent alimenter le fonds documentaire résulte d'une collecte qui implique l'ensemble des acteurs.

Selon Jean-Yves PRAX [PRAX JY., 2004], *"dans le domaine de la GED, il faut considérer deux grandes familles de techniques pour décrire et classer les documents : l'indexation structurée, et l'indexation en texte intégral."*

Avec un système d'indexation structurée de type base de données, l'opérateur doit renseigner les index et mots-clés de chaque document, le document peut alors se présenter sous forme d'image, liée à ces index, quelle que soit la nature du document.

Avec les systèmes d'indexation en texte intégral, il est possible d'auto-indexer les documents par tous les mots du texte qu'ils contiennent."

Au-delà de ces deux catégories, il est possible d'identifier plusieurs méthodes d'indexation :

- L'indexation manuelle avec usage d'un langage documentaire ;
- L'indexation en texte intégral ;
- Les métadonnées ;
- Les moteurs d'indexation.

Pour ce qui est de l'indexation manuelle avec usage d'un langage documentaire et de l'indexation en texte intégral, nous avons déjà traité de ces différentes méthodes dans le cadre

des principes généraux d'indexation documentaire. Par contre, à ce stade de notre réflexion, il nous apparaît important de revenir sur le rôle des métadonnées et des moteurs d'indexation dans les applications de GED.

Les métadonnées sont des informations techniques et descriptives qui sont ajoutées aux documents pour mieux les qualifier. Elles sont surtout utilisées dans les applications de GED pour apporter des informations de nature sémantique sur les documents qu'elles décrivent. Par contre, la conception de ces métadonnées n'a de sens que si des outils de repérage savent les exploiter. On attend donc des applications de GED qu'elles prennent en compte ces informations, pour améliorer la pertinence des recherches qui se fondent en général sur bien d'autres critères.

Les moteurs d'indexation constituent des outils importants de la GED car ils permettent d'indexer automatiquement un certain nombre de documents par rapport à des critères préalablement établis et de créer des dossiers électroniques complets associant de multiples types d'information issues de sources diverses.

Compte tenu des éléments que nous venons d'évoquer, la question qui se pose alors est celle du choix de la meilleure méthode d'indexation. En la matière, plusieurs facteurs sont à prendre en considération :

- La nature du fonds documentaire et sa volumétrie ;
- Le groupe d'utilisateurs chargé de l'indexation des documents ;
- La finalité de l'application de GED ;
- Le cycle de vie de l'information contenue dans les documents.

Toutefois, aux vues de la diversité des applications de GED et des multiples dispositifs de recherche présents dans ces applications, il nous apparaît difficile d'établir une classification des différentes méthodes d'indexation en fonction des critères précédemment énoncés.

Il est par conséquent évident que la meilleure méthode d'indexation est tout simplement celle qui permet à un utilisateur, à travers un ou plusieurs dispositifs de recherche, d'accéder à une information pertinente et adaptée à ses besoins.

b) La recherche d'information

Pour ce qui est des techniques de recherche d'information couramment utilisées dans le domaine de la Gestion Electronique de Documents, on peut citer principalement :

La recherche par mots clés : Dans cette approche, un mot est recherché. Soit ce mot (et ses équivalents) est trouvé soit il ne l'est pas sans autre alternative. Les différents documents répondant à la requête sont rapatriés sans ordre particulier et l'utilisateur doit alors tous les parcourir pour savoir quels sont ceux qui répondent réellement à ses préoccupations.

La recherche par concept : Cette technique est fondée sur une recherche documentaire par thèmes, plutôt que par mots-clés. Dans cette approche, chaque document se voit affecté d'un degré de pertinence par rapport au thème traité.

Les concepts sont au préalable définis selon une arborescence hiérarchique, chaque nœud portant un coefficient de pondération. Les documents sélectionnés par le système de recherche sont alors présentés à l'utilisateur triés par ordre décroissant de coefficient de pondération. Ainsi, les premiers documents de la liste ont ainsi une forte chance d'être les plus pertinents pour l'utilisateur.

Cette technique de recherche est très puissante et assez fiable, mais la fabrication des arborescences de concepts pondérés, en liaison avec les documents, représente un travail considérable qui nécessite une grande expertise.

La recherche plein texte ou en texte intégral ou de texte libre : Il s'agit d'une technique de recherche textuelle qui consiste à examiner tous les mots de chaque document électronique enregistré et à essayer de les faire correspondre à ceux fournis par l'utilisateur dans le cadre de sa requête.

Un des inconvénients majeurs de la recherche plein texte, c'est qu'elle est susceptible de récupérer beaucoup de documents qui ne sont pas pertinents par rapport à la question posée. De tels documents sont appelés faux positifs. La récupération de documents non pertinents est souvent provoquée par l'ambiguïté inhérente au langage naturel (par exemple le mot "avocat" peut aussi bien désigner le fruit que la profession). Aussi, il est possible de lever ces ambiguïtés dues au langage naturel en étiquetant les documents (recherche avec un vocabulaire contrôlé). Par contre, une recherche avec un vocabulaire contrôlé peut omettre de récupérer des documents qui sont vraiment pertinents par rapport à une requête.

Les insuffisances de la recherche plein texte ont été traitées en fournissant aux utilisateurs des outils qui leur permettent d'exprimer leurs requêtes plus précisément, et en développant de nouveaux algorithmes de recherche qui améliorent la précision des récupérations.

Certains systèmes de recherche utilisent des opérateurs booléens ("ET", "OU", "SAUF") pour accroître considérablement la précision d'une recherche plein texte.

La recherche par navigation arborescente : La navigation arborescente est une recherche par circulation dans un plan de classement, une arborescence prédéfinie, en étant guidé par la structuration hiérarchique des informations en domaines et sous-domaines. On parle aussi d'accès hiérarchique à l'information.

La recherche par navigation hypertextuelle : Cette technique suppose une structuration préalable des documents avec la mise en œuvre de liens hypertextuels.

La navigation hypertextuelle est un mode de consultation non-linéaire, selon le principe d'hypertexte. Elle est fondée sur un principe associatif et non hiérarchique, plus proche du fonctionnement intuitif de la pensée.

La recherche en langage naturel : Il s'agit d'un mécanisme d'interrogation reposant sur une analyse linguistique de la requête de l'utilisateur et conduisant à l'élaboration d'une requête "interne" à partir de laquelle la recherche est réalisée. Bien plus riche que le vocabulaire contrôlé, le langage naturel peut toutefois être implicite, redondant et ambiguë.

Une interrogation en langage naturel offre la possibilité à un utilisateur de ne pas avoir à construire une requête en manipulant les opérateurs booléens ni à choisir les bons descripteurs. Par contre, une recherche en langage naturel risque de générer du bruit.

A ce stade de notre réflexion, il est important de rappeler que la plupart des solutions de GED déployées en entreprise se présentent sous la forme de systèmes multi-applications intégrant différentes techniques de recherche d'information.

Cette spécificité provient généralement de l'hétérogénéité des sources d'information et répond aux attentes des différents acteurs et partenaires de l'entreprise qui souhaitent ainsi regrouper l'ensemble des informations qu'ils produisent et utilisent au sein d'un seul et même système.

L'objectif étant de trouver et de sélectionner rapidement la meilleure information, l'utilisateur doit désormais adapter son mode de raisonnement à la logique d'indexation et de classement de l'information électronique.

Il doit également tenir compte de la spécificité du langage et de la structuration de l'information propres à chaque application.

D'autre part, la plupart des solutions de GED actuellement déployées en entreprise fonctionnent comme des applicatifs à part entière, mais force est de constater qu'avec la "Webification" des applications en entreprise (transformation et/ou réécriture des applications en adoptant les technologies et l'aspect dynamique du Web), elles ont tendance à évoluer vers des Intranets documentaires ou des plateformes collaboratives intégrant des fonctionnalités de

GED où les utilisateurs se servent de leur navigateurs Internet pour accéder aux informations de l'entreprise et souhaitent disposer de fonctionnalités en phase avec les pratiques Web actuelles (navigation par hyperliens, utilisation d'un moteur de recherche comme principale fonctionnalité permettant de retrouver des documents, des images, des vidéos et des fichiers, etc.).

Cette "Webification" des applications en entreprise devrait entraîner un recours de plus en plus important aux moteurs de recherche et aux métadonnées dans les applications de GED.

Par conséquent, un des enjeux majeurs de la recherche d'information dans le domaine de la GED consiste à appréhender la façon dont un utilisateur formule une demande en vue de caractériser l'ensemble de cette démarche et de concevoir un modèle d'interrogation généraliste susceptible de répondre aux attentes d'un public très diversifié.

Partie 2 : Discussions et contributions méthodologiques

Si l'on considère que l'entreprise est une organisation vivante et que sa réalité va au-delà de la somme des individus qui la composent, alors il convient de *"mettre l'essentiel au cœur de l'important"* (Citation d'Olivier LECERF, Industriel Français et ancien Président de Lafarge) et d'inscrire l'entreprise dans une perspective dynamique en considérant que sa finalité est de créer de la valeur, à partir des savoirs cumulés et accumulés.

Dans cette perspective, la GED constitue alors une véritable mémoire de l'entreprise, permettant à tous les acteurs d'acquérir, d'indexer, de stocker et de rechercher des informations qui leur permettront de réactiver ou de stimuler des connaissances afin d'opérer des traitements et/ou de prendre des décisions.

Les techniques d'indexation et de recherche d'information que vous venons de présenter dans la première partie de ce document apparaissent alors comme des fonctionnalités essentielles d'un système de GED et peuvent être mises en œuvre soit indépendamment et exclusivement les unes des autres soit pour certaines d'entre elles de manière combinée.

La généralisation du document numérique ainsi que la reconnaissance du fait que la compétitivité économique réside dans la maîtrise des flux d'information ont rendu indispensable l'usage de ces techniques d'indexation et de recherche d'information. Mais, force est de constater, qu'elles connaissent un certain nombre de limites qu'il convient également d'exposer dans la deuxième partie de ce document.

Dans la deuxième partie de ce document, et après une discussion sur les limites des techniques d'indexation et de recherche existantes dans le domaine de la GED, nous tenterons alors d'apporter un éclairage différent sur la manière d'appréhender la réponse aux besoins informationnels des utilisateurs en s'appuyant sur une approche hybride fondée sur trois dimensions : la dimension cadre ou activité, la dimension humaine et la dimension technologique.

Car comme aimait à le rappeler Albert EINSTEIN "aucun problème ne peut être résolu sans changer le niveau de conscience qui l'a engendré".

I. Limites des techniques d'indexation et de recherche existantes

Compte tenu de l'ensemble des éléments qui ont été exposés dans la première partie de ce document, il est clair que les modalités de la recherche d'information dans le domaine de la GED connaissent de profondes transformations et une formidable expansion. Pour autant, l'interrogation obéit toujours à une stratégie de recherche qui emprunte ses principaux modes de fonctionnement aux techniques actuelles, lesquels présentent un certain nombre de limites qu'il convient à présent de préciser et de repositionner dans le contexte spécifique de la GED.

1) L'absence de convivialité dans le mécanisme de recherche

Lorsqu'un utilisateur opère une recherche d'informations dans un système de GED multi-applications, il doit nécessairement passer par une phase de test ou d'apprentissage empirique. Ce phénomène est directement lié au fait qu'à chacune de ces applications correspond des pratiques et des modes de recherche spécifiques (recherche par requête sur des mots-clés dans des champs délimités, navigation arborescente, recherche par requête sur le contenu, navigation hypertextuelle, etc.).

Par conséquent, la logique d'interrogation propre à une application étant parfois très différente de celle rencontrée dans le cadre d'une autre application, les outils de recherche d'informations utilisés dans un système de GED traduisent nettement cette spécificité.

Cette imbrication d'outils, de techniques et de modalités de recherche différents font que l'utilisateur en quête d'informations doit dans un premier temps se familiariser avec cette hybridation des modes de recherche et des outils et analyser la façon dont il doit formuler sa requête pour que le système de GED lui fournisse une réponse plus pertinente.

2) La complexité de la recherche dite en texte intégral

La complexité de la recherche dite en texte intégral est généralement masquée par l'apparente facilité de la forme d'interrogation.

Ainsi, pour un utilisateur averti d'un système de GED, ce mécanisme de recherche est celui qui présente le plus de souplesse de par l'absence de structuration de la requête et l'utilisation d'un vocabulaire simplifié.

Ce qui se traduit généralement par le fait qu'un utilisateur à la recherche d'informations a l'impression de pouvoir s'exprimer beaucoup plus librement et ainsi formuler une demande qui soit plus en rapport avec son objectif initial.

Mais au-delà de l'apparente facilité à utiliser un tel mécanisme, il existe une réelle complexité à exprimer fidèlement l'information recherchée (besoin informationnel).

Cette complexité provient essentiellement de la difficulté à privilégier certaines expressions dans une requête et au fait que pour un utilisateur averti d'un système de GED, certains mots contiennent plus de "sens" que d'autres.

3) Les limites de la représentation du contenu par le langage

La principale limite d'un système de GED repose sur le fait que l'ensemble des mécanismes d'indexation et de recherche ne s'intéressent généralement qu'à l'aspect sémantique du fonds documentaire.

Car même si le multimédia s'est fortement développé ces dernières années, il n'existe pas de moyens fiables pour représenter de manière formelle un document.

Or, la forme d'un document contient autant d'informations dans le mécanisme de validation de l'information que les éléments de contenu qui le caractérisent.

De plus, la façon de décrire un document dépend fortement de l'utilisateur, de son expérience, de ses compétences et de son domaine d'activité.

Les termes utilisés pour indexer puis rechercher de l'information dépendent donc du vécu et de l'environnement de l'utilisateur et ne seraient constituer un référentiel unique.

Par conséquent, une approche de recherche basée essentiellement sur une représentation du contenu par le langage introduit généralement un biais lié à l'absence d'éléments permettant de caractériser de façon formelle un document et sur la spécificité du langage d'indexation et de recherche de l'utilisateur.

4) L'apport limité des systèmes linguistiques dans les systèmes de GED

Actuellement, deux types de modèles sont utilisés dans les approches linguistiques de traitement des documents, le modèle linguistique et le modèle procédural auquel vient s'ajouter dans certains cas une approche probabiliste.

Bien que dans un modèle linguistique, il soit possible de lever les ambiguïtés du langage, la connaissance des concepts et le niveau pragmatique de maîtrise de l'environnement utilisateur sont autant d'éléments qui aujourd'hui ne sont pas suffisamment développés au sein des systèmes linguistiques.

Par conséquent, l'utilisation de ces modèles dans un système de GED imposent au préalable un important travail de développement d'outils linguistiques (dictionnaires spécialisés, réseaux sémantiques, etc.).

Cette étape n'étant pas systématiquement réalisée, l'apport des systèmes linguistiques dans les systèmes de GED est par voie de conséquence extrêmement limité.

Les approches probabilistes permettant quant à elles de tenir compte du poids relatif des mots dans un document, elles offrent la possibilité d'orienter l'utilisateur vers des systèmes de classement des informations trouvées en fonction de leur pertinence.

Cependant, ce type de modèle ne permet pas d'apporter une réponse satisfaisante quant à la recherche d'informations dans un système de GED qui n'offrent pas de fonctions documentaires élaborées.

Le manque de maturité des systèmes de recherche en langage naturel intégrant une aide sémantique et l'absence de dictionnaires de jargon d'entreprise ou de secteur, constituent des éléments qui vont à l'encontre d'une amélioration des performances dans le domaine de la recherche d'information.

De même, l'utilisateur, son expérience et son mode de raisonnement sont des critères bien souvent écartés dans les mécanismes d'interrogation implémentés dans les systèmes de GED multi-applications.

Pourtant ils apportent une dimension supplémentaire aux mécanismes de recherche existants en prenant en compte les compétences, le profil et la logique de formulation d'une requête par un utilisateur.

II. Vers une approche hybride

Notre approche s'appuie sur une démarche méthodologique dite "tridimensionnelle" dont les principes furent énoncés en 2001 par Laïd BOUZIDI [BOUZIDI L., 2001], Responsable scientifique de l'équipe de recherche SICOMOR (Systèmes d'Information COMMunicants, Management et Organisation) de l'Institut d'Administration des Entreprises, rattachée au Centre de recherche Magellan de l'Université Jean MOULIN - Lyon 3.

Cette démarche méthodologique a ensuite été enrichie par les différents travaux des enseignants-chercheurs, des docteurs et doctorants de l'équipe de recherche SICOMOR.

Cette approche générique vise à prendre en considération tous les éléments nécessaires à la conception d'une solution technologique en lien avec le métier et les différents acteurs de l'entreprise et s'articule autour de trois dimensions :

- La dimension cadre ou dimension activité qui s'intéresse au(x) métier(s) de l'organisation ;
- La dimension humaine qui se préoccupe des acteurs humains qui assurent différentes fonctions au sein de l'organisation ;
- La dimension technologique qui tente de comprendre l'apport des outils technologiques au service des différentes fonctions et des acteurs humains de l'organisation.

Elle constitue un cadre théorique permettant d'analyser les problèmes de recherche d'information dans le domaine de la GED, mais aussi pragmatique, en soulignant l'apport de cette problématique de recherche afin d'améliorer l'accès à des informations pertinentes et fiables au sein des organisations.

Dans un premier temps, nous présenterons les fondements théoriques de cette approche puis, nous appliquerons cette méthodologie au contexte spécifique de la recherche d'information dans le domaine de la GED afin de proposer un schéma global de fonctionnement d'un système de recherche d'information dans le domaine de la GED.

1) Les fondements théoriques de l'approche dite "tridimensionnelle"

La méthodologie de conception de système d'information fondée sur l'approche dite "tridimensionnelle" permet de mettre en évidence l'importance des différentes technologies de l'information, de la communication, de la collaboration et de l'intelligence augmentée qui sont susceptibles d'être déployées au sein d'une organisation, les conditions de leur intégration ainsi que la participation de l'ensemble des acteurs humains concernés par l'usage de ces technologies. Elle traite également des interactions entre ces différentes composantes et s'appuie sur trois dimensions : la dimension cadre ou dimension activité, la dimension humaine et la dimension technologique.

Il est important de souligner que l'acteur humain n'intervient pas seulement durant la phase de conception d'un système d'information mais qu'il est au contraire présent à tous les stades de son cycle de vie, de la phase de conception jusqu'à la phase de maintenance et d'évolution en passant par toutes les phases opérationnelles liées à son développement.

La technologie accompagne le développement de l'entreprise ou de l'organisation à laquelle elle est destinée et doit s'adapter à son activité ainsi qu'à l'ensemble des pratiques et usages des acteurs humains qui sont susceptibles de l'exploiter.

Les choix technologiques qui sont alors opérés tiennent compte de l'organisation et de son activité, des différents métiers présents dans l'entreprise et du rôle joué par chaque acteur dont le profil et les besoins informationnels doivent également être pris en considération.

Relevant d'une approche systémique [DONNADIEU G., KARSKY M., 2002] au sens où *"la systémique permet de prendre en compte l'extraordinaire complexité du monde qui nous entoure"*, l'approche dite "tridimensionnelle" dont les principes généraux furent énoncés par Laïd BOUZIDI, permet d'étudier l'activité d'une organisation, ses différents composants et leur interaction dans l'optique de concevoir et de mettre en œuvre un système d'information dont la stratégie est en cohérence avec la stratégie globale de l'entreprise et en lien avec son environnement.

Dans la mesure où l'approche systémique peut être considérée comme une méthodologie de représentation d'un objet actif lui-même constitué d'autres d'éléments actifs en interaction dynamique et où, l'approche dite "tridimensionnelle" prend pour objet d'étude une activité qui

est considérée comme un objet actif, composé de différents niveaux en interaction, on peut alors qualifier l'approche dite "tridimensionnelle" d'approche systémique.

Nous allons à présent décrire cette méthodologie d'un point de vue théorique en nous intéressant successivement aux dimensions activité, humaine et technologique.

a) La dimension activité

La dimension activité ou dimension cadre permet d'analyser le type d'activité dans lequel s'inscrira le système d'information et de mettre en évidence l'ensemble des processus qui la représente. Elle est déclinée à plusieurs niveaux : organisationnel, fonctionnel, relationnel, structurel et temporel.

Le niveau organisationnel : Le niveau organisationnel nous amène à appréhender l'activité étudiée par le biais de l'organisation et à la situer dans son environnement socio-économique. Il s'agit essentiellement de déterminer l'environnement interne (cadre administratif, juridique, financier, etc.) dans lequel l'activité s'exerce mais aussi de prendre en considération l'environnement externe (micro et macro environnement) de l'organisation.

Le niveau fonctionnel : Le niveau fonctionnel permet de recenser les différentes fonctions nécessaires au bon déroulement de l'activité.

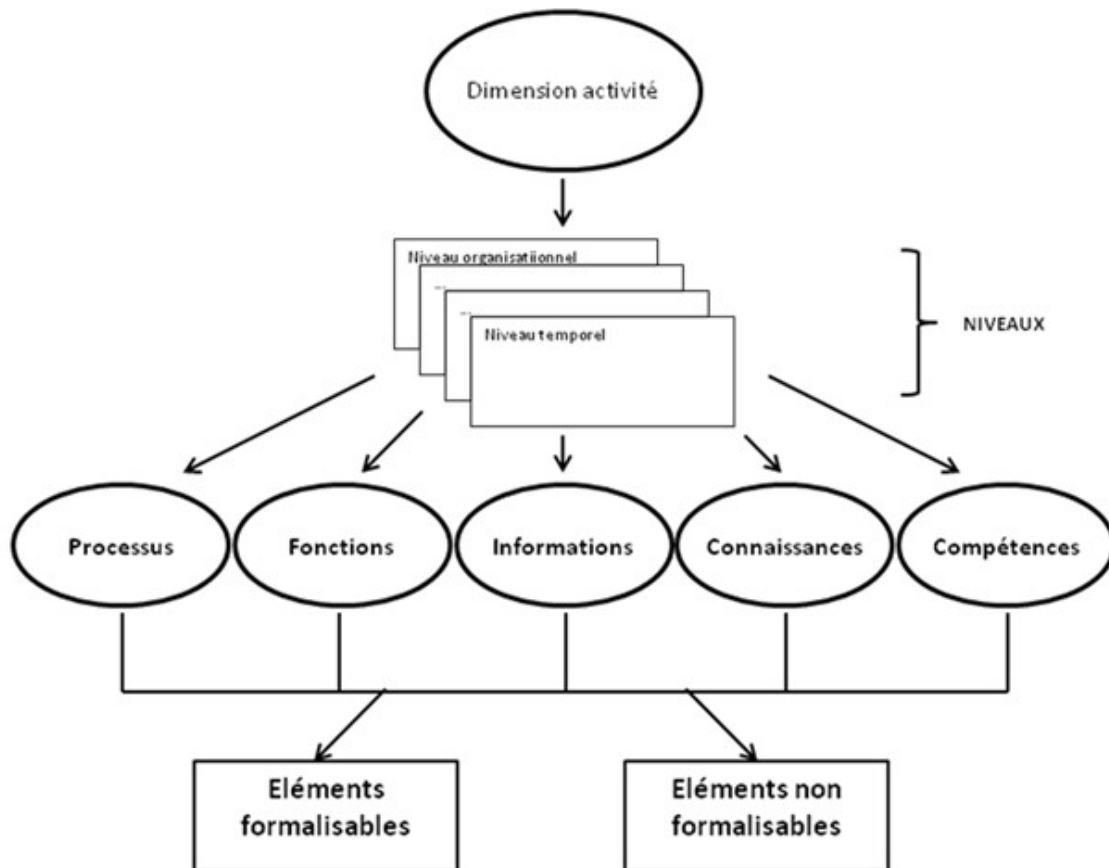
Le niveau relationnel : Il correspond à la prise en compte des interactions entre les différentes fonctions.

Le niveau structurel : Le niveau structurel a pour but de définir la structuration et l'exploitation des informations et de décrire comment sont représentées les connaissances.

Le niveau temporel : L'exercice de toute activité humaine s'inscrivant dans la durée, il en va de même pour toutes les organisations qui pour tenir compte de l'évolution de leur environnement (interne et externe) doivent sans cesse s'adapter au changement qui constitue véritablement la seule permanence que nous ayons au monde.

L'étude de la dimension activité à travers ces différents niveaux nous permet de déterminer les fonctions qui sous-tendent l'activité et de mettre en relief l'ensemble des processus métiers. Elle a aussi pour but de recenser l'ensemble des informations présentes dans l'organisation et de mieux cerner les flux informationnels échangés par les différents acteurs. Cette étude conduit également à identifier les connaissances ainsi que les compétences présentes dans l'organisation.

Figure 6 : La dimension activité de l'approche dite "tridimensionnelle"



b) La dimension humaine

La dimension humaine se préoccupe essentiellement des acteurs humains qui assurent différentes fonctions au sein d'une organisation. Elle est par ailleurs omniprésente dans toutes les phases suggérées par la méthodologie et doit donc être appréciée sous l'angle individuel ou collectif.

Toute organisation pouvant se présenter comme une somme d'individus au service d'un projet commun, c'est grâce à ce facteur humain que des projets aboutissent, que des organisations atteignent leurs objectifs, que des systèmes dits à hauts risques sont fiables.

Dans le cadre de cette méthodologie, la prise en compte de la dimension humaine repose sur plusieurs niveaux qui sont :

- L'identification de tous les acteurs humains ;
- La définition de leur profil ;
- L'identification de leurs besoins informationnels ;
- La définition de leurs attentes en matière d'ergonomie et d'Interface Homme-Machine (IHM).

L'identification des acteurs humains : Elle permet l'établissement d'une classification des différents types d'acteurs humains qu'il est malgré tout possible de regrouper en trois grandes catégories :

- Les acteurs utilisateurs/usagers ;
- Les acteurs informaticiens ou acteurs techniques ;
- Les acteurs experts du ou des domaines liés à l'activité.

La première catégorie d'acteurs (acteurs utilisateurs/usagers) correspond aux utilisateurs finaux et usagers du système d'information. Selon Yves-François LE COADIC [LE COADIC YF., 2001], un utilisateur est *"la personne qui emploie, qui se sert d'un produit, d'un service, d'un système d'information, ces éléments subsistant après leur utilisation"*.

La notion d'usager est un peu différente de celle de l'utilisateur. Selon Yves-François LE COADIC, ce concept fait référence à la notion d'usage qui peut se définir comme *"la reproduction d'utilisations constantes et récurrentes intégrées dans la quotidienneté"*. Compte tenu de ces éléments, la notion d'usager a une portée générale à la différence de l'utilisateur qui se place plus dans la pratique et dans la spontanéité.

La deuxième catégorie d'acteurs correspond aux acteurs techniques présent dans l'organisation qui sont susceptibles d'intervenir tout au long du processus de conception d'un système d'information mais également à tous les stades de son cycle de vie (de la phase de conception jusqu'à la phase de maintenance et d'évolution en passant par toutes les phases opérationnelles liées à son développement).

La troisième catégorie d'acteurs correspond aux experts métiers, qui possèdent la maîtrise d'un ou de plusieurs domaines d'activité et donc, par voie de conséquence de l'ensemble des processus métiers et des fonctions de ce ou ces domaines d'activité.

Cette phase de classification des différents acteurs humains permet de préciser le rôle de chaque catégorie d'acteurs intervenant dans l'organisation ainsi qu'à tous les stades du cycle de vie du système d'information.

La définition de leur profil : La définition du profil des utilisateurs/usagers représente un aspect important de la prise en compte de la dimension humaine. Ce "profilage" constitue un préalable à la phase d'identification des besoins informationnels des différents acteurs mais peut s'avérer une opération particulièrement complexe du fait des nombreux paramètres que comprend un profil et de l'hétérogénéité de la population d'acteurs.

L'identification de leurs besoins informationnels : L'identification des besoins informationnels des différents acteurs constitue le troisième niveau d'analyse de la dimension humaine et s'intéresse tout particulièrement à la manière dont les différents acteurs en charge de fonctions bien spécifiques au sein d'une organisation expriment leurs besoins informationnels à partir de leurs pratiques fonctionnelles et de l'usage qu'ils font des technologies de l'information, de la communication, de la collaboration et de l'intelligence augmentée.

Un parallèle peut être établi entre cette phase d'identification des besoins informationnels et l'approche fonctionnelle menée lors de l'étude de la dimension activité. Toutefois, l'approche menée ici repose plus sur une analyse de l'offre et de la demande d'information avec la volonté d'identifier la nature des informations demandées, les acteurs à l'origine de cette expression de besoins et les usages qui seront faits de l'information auxquels ils auront accès.

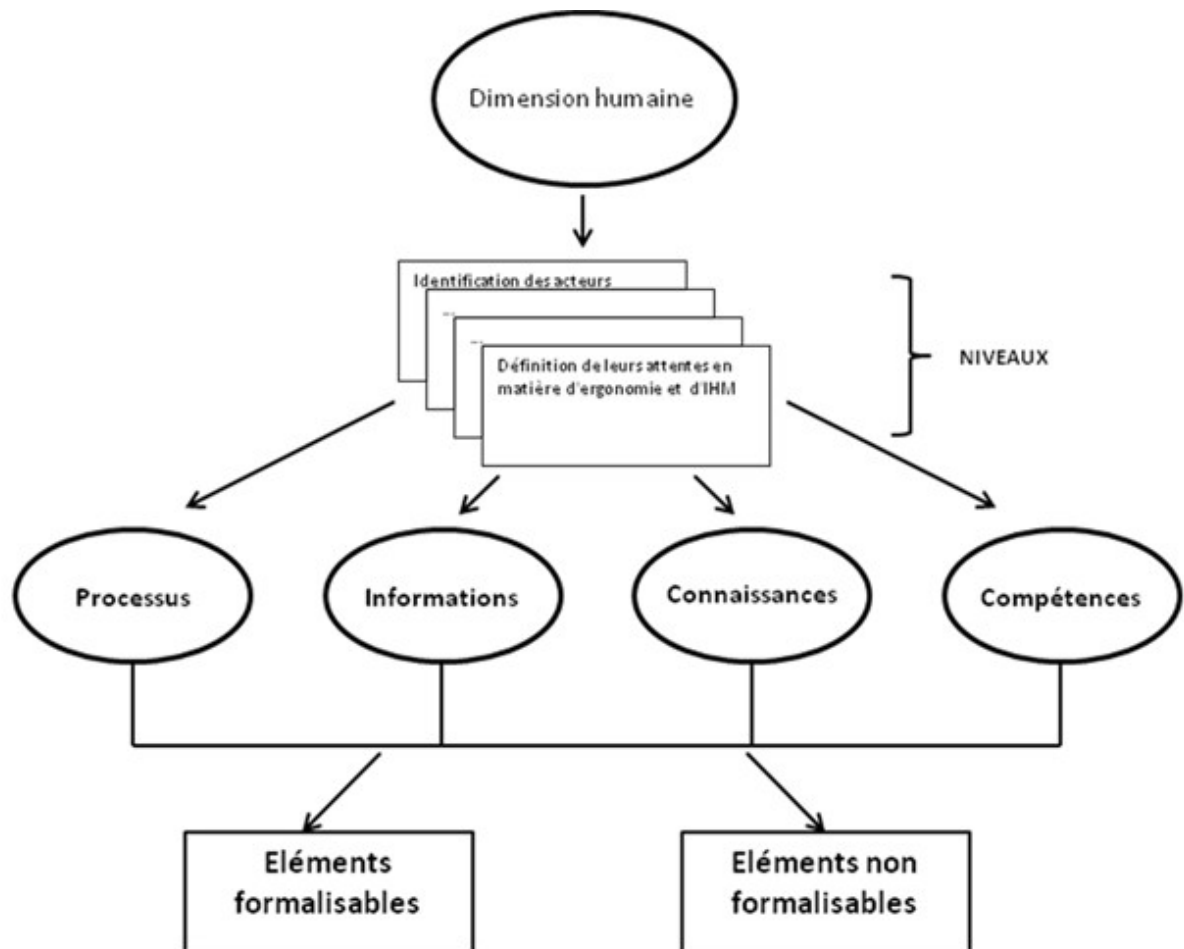
La définition de leurs attentes en matière d'ergonomie et d'IHM : Le système d'information représentant l'ensemble des moyens (organisation, acteurs, procédures, systèmes informatiques) nécessaires au traitement et à l'exploitation d'informations dans le cadre d'objectifs définis au niveau de la stratégie, des métiers et de la réglementation, il est alors nécessaire d'établir le lien entre l'activité réelle d'une organisation dont les principaux acteurs sont des êtres humains et la représentation de cette activité dans le cadre d'un système d'information.

Cet aspect relatif aux attentes formulées par les acteurs et utilisateurs du système d'information en matière d'ergonomie et d'IHM constitue un point important de la démarche visant à appréhender la dimension humaine de l'activité d'une organisation.

Car, c'est en fonction de ces attentes que des outils de représentation et de structuration des informations sont proposés et appréhendés lors des phases de conception et de développement du système d'information. La faisabilité technique relevant des possibilités offertes par la technologie, elle est fortement dépendante des applications et des dispositifs matériels susceptibles d'être mis en place au sein de l'organisation

L'analyse de la dimension humaine, tout comme l'étude de la dimension activité permet de dégager trois éléments caractéristiques : les processus, les informations, les connaissances et les compétences.

Figure 7 : La dimension humaine de l'approche dite "tridimensionnelle"



c) La dimension technique

Le troisième volet d'étude de la méthode dite "tridimensionnelle" est relatif à la prise en compte de la dimension technologique. Cette approche passe tout d'abord par l'analyse des outils susceptibles d'être intégrés (applicatifs et composants d'infrastructure technique) puis par une analyse des différents processus métiers et des logiques fonctionnelles.

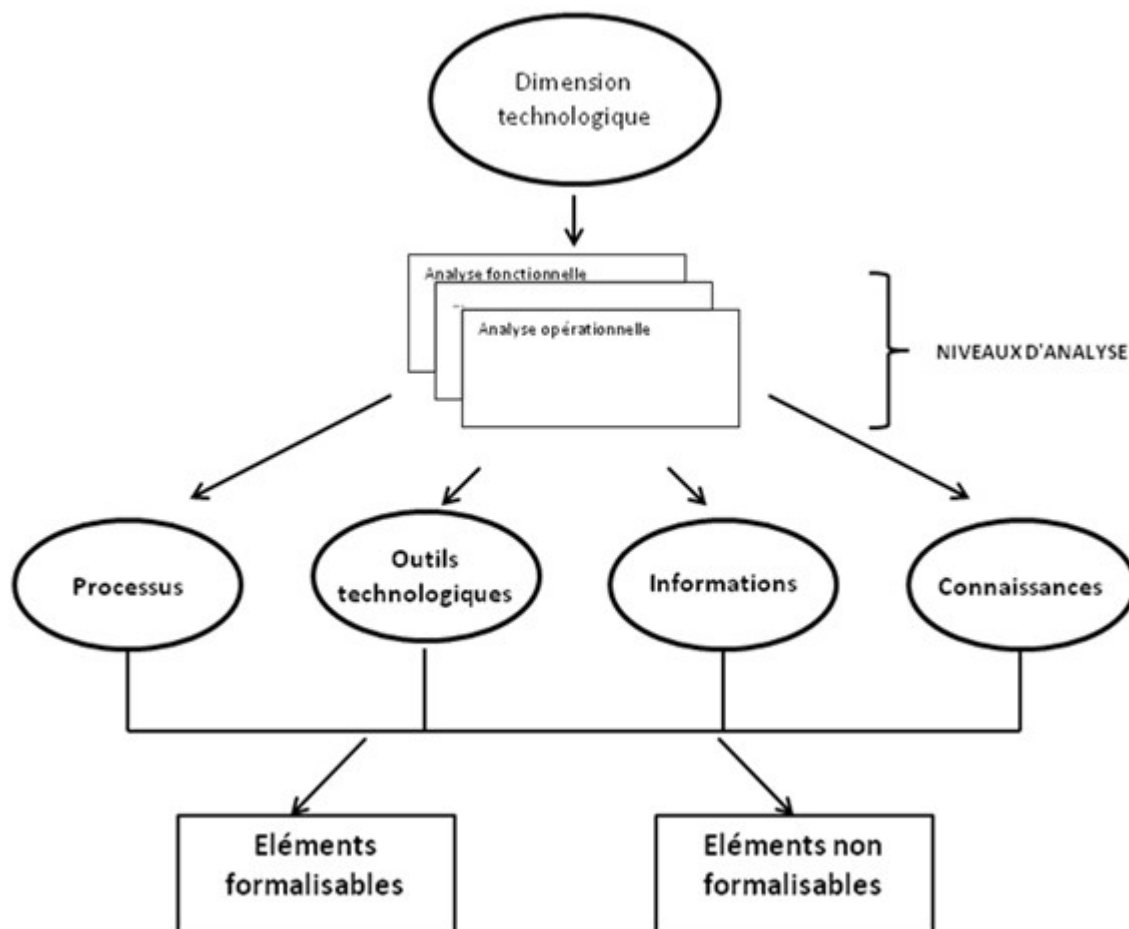
L'analyse fonctionnelle : L'objectif de cette analyse est de décrire l'ensemble des logiques fonctionnelles ainsi que les conditions d'utilisation des différents outils technologiques. Elle a pour vocation de préciser le type d'architecture applicative et de recenser les grandes familles d'applications concernées. Elle comprend également une étude du fonctionnement en réseau et des modes d'accès à l'information. L'ergonomie mais aussi les contraintes imposées par les différents outils technologiques seront également abordées dans le cadre de cette étude.

L'analyse organisationnelle : L'intégration de nouveaux outils technologiques fait généralement apparaître de nouvelles contraintes organisationnelles. L'objectif de cette phase d'analyse est donc de prendre en considération l'ensemble de ces contraintes et d'appréhender

les problématiques de changement générées par l'introduction de ces nouveaux outils technologiques.

L'analyse opérationnelle : Cette phase d'analyse porte sur les différents aspects liés à la mise en œuvre opérationnelle des outils technologiques ainsi qu'à leur exploitation par les acteurs et utilisateurs concernés. Cette étude s'intéresse également à l'ergonomie de ces outils technologiques et aux interfaces homme-machine.

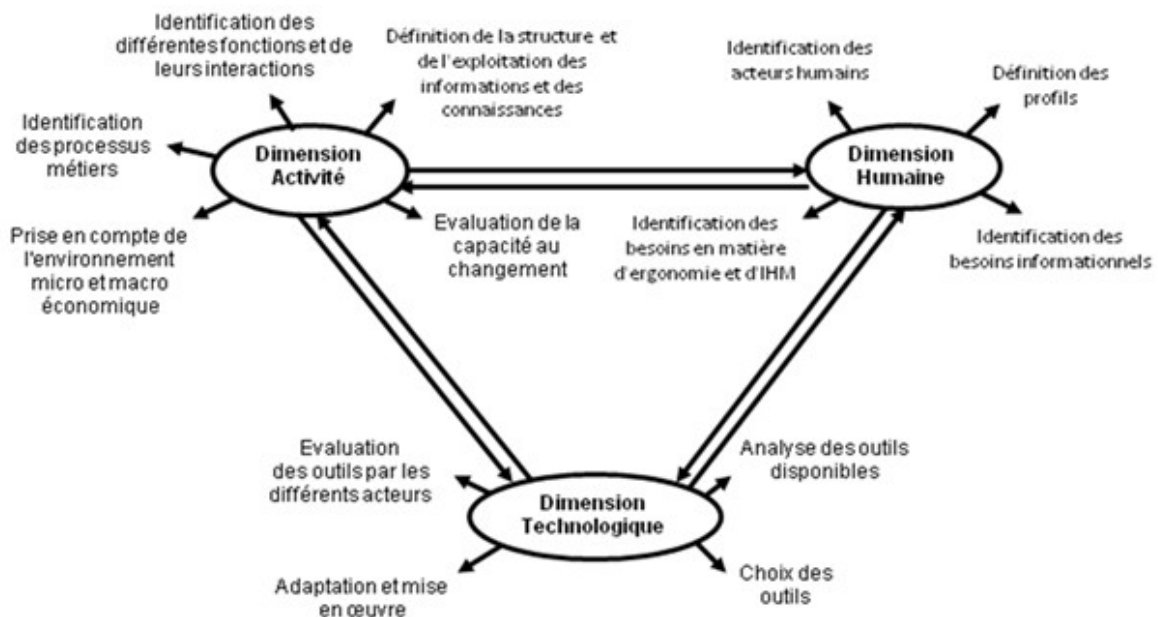
Figure 8 : La dimension technologique de l'approche dite "tridimensionnelle"



d) Les interactions entre les dimensions

La méthode dite "tridimensionnelle" s'appuie sur les trois dimensions et les différents niveaux d'analyse que nous venons d'exposer. Cette approche permet également de mettre en relief les interactions qui existent entre les différents éléments (formalisables) que chaque dimension fait ressortir.

Figure 9 : L'approche dite "tridimensionnelle"



Les interactions entre la dimension activité et la dimension technique : L'étude de la dimension activité nous a permis de mettre en relief l'ensemble des processus métiers, de déterminer les fonctions qui sous-tendent l'activité, de recenser l'ensemble des informations, des connaissances et des compétences présentes dans l'organisation ainsi que de mieux cerner les flux informationnels échangés par les différents acteurs. Une partie des éléments recueillis lors de cette étude sera formalisée à l'aide d'outils d'aide à la représentation qui font partie de la dimension technologique.

Les interactions entre la dimension humaine et la dimension technique : L'étude de la dimension humaine nous a permis d'identifier et de préciser le rôle des différentes catégories d'acteurs ainsi que leurs besoins informationnels qui seront représentés dans la dimension technologique à l'aide d'outils techniques.

Les interactions entre la dimension activité et la dimension humaine : Chaque acteur ayant vu son rôle précisé dans la dimension humaine de l'approche dite "tridimensionnelle", il s'inscrit également dans le cadre d'une activité et de l'exercice d'une ou plusieurs fonctions qui conditionnent et influent ses attentes vis-à-vis des outils technologiques.

2) La phase de réalisation technique

La finalité de l'approche hybride dont les principes généraux furent énoncés par Laïd BOUZIDI, est de concevoir et de mettre en œuvre un système d'information fondé sur les trois dimensions que nous venons d'évoquer.

Cette phase de réalisation technique comprend alors quatre étapes qu'il nous convient à présent d'évoquer.

L'analyse des outils disponibles : Elle permet de recenser l'ensemble des solutions existantes et d'évaluer les fonctionnalités des différents outils technologiques en prenant en compte la dimension technologique de l'organisation (moyens techniques existants) sa dimension humaine (moyens humains et compétences) mais aussi sa dimension activité (notamment en termes de moyens financiers).

Le choix des outils : Il s'effectue à l'aide d'une grille d'évaluation dont les critères sont représentés selon les différentes composantes de la dimension technologique et de la dimension humaine.

L'adaptation et la mise en œuvre : Il s'agit de l'étape de maquettage ou de prototypage de la solution choisie mais aussi d'intégration et d'adaptation des outils technologiques dans l'organisation. Cette étape étant réalisée par les acteurs techniques et les concepteurs du système d'information, elle fait appel aux éléments recueillis par l'analyse des dimensions activité et humaine (contraintes organisationnelles, processus métiers, etc.).

L'évaluation des outils technologiques : Cette évaluation se fait généralement par retour d'expériences et permet de mieux adapter les outils à l'organisation et aux utilisateurs/usagers. Un mécanisme de capitalisation d'expériences peut également permettre d'améliorer les outils mis en œuvre et de susciter de nouvelles évolutions.

Après avoir présenté les fondements théoriques de l'approche dite "tridimensionnelle" ou approche hybride, il convient à présent d'instancier cette méthode dans le cadre de la conception d'un système d'aide à la recherche d'information dans le domaine de la Gestion Electronique de Documents.

Partie 3 : De l'approche hybride au système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences

Cette troisième partie est entièrement consacrée à la conception d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences en vue de contribuer à la résolution de notre problématique de recherche dont la finalité est d'améliorer la pertinence de l'information recherchée dans le domaine de la GED.

Dans un premier temps, nous procéderons à une instanciation de l'approche dite "tridimensionnelle" dans le cadre de la conception d'un système d'aide à la recherche d'information dans le domaine de la Gestion Electronique de Documents.

Nous étudierons également les limites de cette approche notamment en termes de représentation des interactions entre un utilisateur et une application informatique mais aussi en termes de compréhension de la manière dont fonctionne un acteur humain.

Nous présenterons ensuite un certain nombre de concepts qui nous permettront de prendre en compte l'effet "Sérendipité" et de modéliser la démarche que nous avons adoptée.

Ces concepts seront abordés dans le cadre de deux méthodes :

- La méthode "Motivations" qui permettra de mettre en évidence les interactions entre un acteur humain et des applications informatiques ;
- La méthode des niveaux logiques de la pensée, de l'action et du changement qui permettra de mieux comprendre comment fonctionne un être humain.

Par la suite, nous établirons une proposition de conception d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences.

Nous commencerons par exposer les fondements théoriques d'une approche permettant d'améliorer la pertinence de l'information recherchée par la prise en compte du mécanisme de "Sérendipité" et l'apport du raisonnement par cas.

Puis, nous énoncerons le principe de fonctionnement du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences proposé.

Enfin, nous établirons les conditions d'intégration d'un tel système au sein d'une solution de GED et détaillerons les facteurs clés de succès.

I. La Gestion Electronique de Documents par l'approche Hybride

L'approche dite "tridimensionnelle" ou approche hybride dont nous venons d'exposer les fondements, contribue à l'analyse de notre problématique de recherche au sens où elle nous fournit un cadre méthodologique afin de concevoir un système d'aide à la recherche d'information dans le domaine de la GED.

Elle s'articule autour de trois dimensions et de différents niveaux d'analyse qui constituent le socle de cette approche et que nous nous proposons d'appliquer à notre problématique de recherche.

1) La dimension activité

En introduisant dans les organisations des fonctions de gestion ou de traitement des documents vivants, qui viennent en complément des applications métiers et des outils bureautiques, la GED est devenue une composante essentielle des systèmes d'information.

La grande diversité des applications de GED nous montre à quel point elle trouve son utilité dans tous les contextes et dans tous les lieux où les informations et les documents abondent.

Faisant partie intégrante du système d'information, la GED s'est imposée comme une technologie à vocation universelle.

Dans ce contexte, et sachant que la structuration et la mise en œuvre d'un système de GED ne peuvent être effectuées que par l'organisation qui le porte, il nous apparaît important de pouvoir caractériser au mieux l'activité de cette organisation.

La dimension activité nous fournit une approche globale de l'organisation, de ses caractéristiques, contraintes, évolutions, etc.

Adopter une vision globale, dépassant l'organisation elle-même, permet de développer une compréhension fine des problématiques auxquelles elle est confrontée, des grands enjeux

qu'elle doit relever et des dynamiques à mettre en œuvre pour atteindre les objectifs qu'elle s'est fixée.

La prise en compte de l'environnement :

L'environnement des organisations évolue rapidement : globalisation des marchés, concurrence exacerbée, nouvelles réglementations, progrès technologiques, phénomènes conjoncturels, tendances socioculturelles, préservation de l'environnement, etc.

Ces facteurs sont multiples, complexes à appréhender et difficiles à anticiper, notamment de par leur apparition parfois soudaine et leur évolution combinée. Force est de constater qu'ils impactent plus ou moins fortement l'organisation de par la nature de son activité, sa santé financière, sa taille, etc.

Ils sont en effet susceptibles de remettre en cause sa compétitivité, et parfois même sa pérennité.

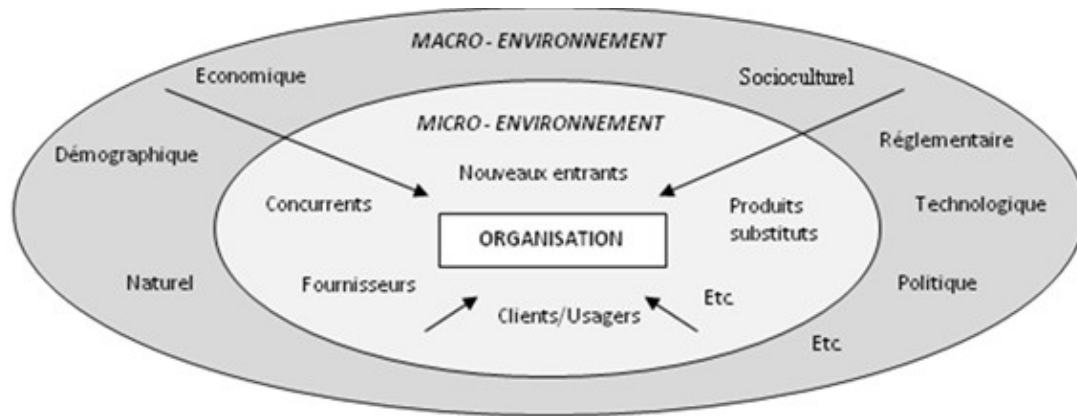
Ne pas suffisamment tenir compte des caractéristiques de cet environnement, de ses règles du jeu et de son évolution, c'est exposer l'organisation à des risques importants, voire fatals. Face à ces phénomènes, l'organisation doit constamment s'adapter.

Il convient d'adopter une méthodologie particulière pour éviter deux écueils préjudiciables :

- Avoir une vision trop réductrice de la réalité : ne pas savoir, ou ne pas accepter de voir, en ne prenant pas en compte des informations stratégiques et déterminantes pour le devenir de l'organisation ;
- Avoir une vision trop large : raisonner au niveau d'une filière en oubliant les forces et les faiblesses de l'organisation, se perdre dans des éléments de détail n'ayant pas d'impact sur l'évolution de l'organisation.

Cette méthodologie relevant d'une approche systémique, il nous est alors possible de représenter simplement par un schéma systémique les relations entre une organisation et son environnement (qui jouent effectivement ou potentiellement un rôle dans son développement) ainsi que les dynamiques externes à l'entreprise lui permettant d'exister et de se développer.

Figure 10 : Vision globale de l'environnement d'une organisation



Le schéma systémique apporte une vision globale de l'environnement et permet de saisir les dynamiques externes qui sont en jeu.

L'identification des processus :

Une organisation peut se définir comme un ensemble d'éléments en interaction, regroupés au sein d'une structure régulée, ayant un système de communication pour faciliter la circulation de l'information, dans le but de répondre à des besoins et d'atteindre des objectifs déterminés.

Compte tenu de cette définition, l'activité d'une organisation peut alors se résumer à un ensemble organisé de tâches qui concourt à la réalisation ou à la transformation d'un bien ou d'un service dans l'optique de satisfaire les besoins d'individus ou d'organisations et d'atteindre des objectifs qui ont été préalablement fixés.

Dans la mesure où toute tâche organisée qui génère un bien ou un service constitue un processus (ou un enchaînement de processus), un des moyens d'appréhender l'activité d'une organisation consiste à recourir à une analyse de l'ensemble de ces processus, mais aussi à étudier ceux qui contribuent à leur bon fonctionnement ainsi qu'à leur pilotage et leur contrôle.

Avant de préciser les différents types de processus présents au sein d'une organisation, il convient de rappeler que la nature de l'ensemble de ces processus est fortement dépendante de l'activité (qui correspond à un ou des métiers) et de l'organisation (marchande ou non marchande).

On distingue généralement trois grandes familles de processus :

- Les processus opérationnels :
Ils contribuent à la réalisation d'un bien ou d'un service. Ils sont dans le "faire" de l'organisation. Plus généralement, les processus opérationnels regroupent toutes les activités dédiées au cycle de vie d'un bien ou d'un service (recherche marketing,

développement de nouveaux produits, avant vente, vente, conception, production, etc.). Ils sont également appelés processus métier.

- Les processus de support :
Également appelés processus de soutien, ils contribuent au bon fonctionnement des autres processus par l'apport de ressources nécessaires (gestion des compétences, formation, gestion financière, achats etc.). Ces processus nourrissent tous les autres.
- Les processus de direction :
Ils participent et contribuent à la détermination et à l'élaboration de la politique ainsi qu'au [déploiement des objectifs](#) dans l'organisation. Ils sont les "fils conducteurs" des processus opérationnels et de soutien. Ils les pilotent et les surveillent.

Dans la mesure, où il est parfois difficile d'appréhender un processus transverse à plusieurs entités d'une organisation, on ne traite bien souvent que des procédures rattachées à diverses entités de cette organisation (direction, service, équipe, etc.). Une procédure représentant la manière de mettre en œuvre tout ou partie d'un processus, on peut alors la définir comme un ensemble de tâches qui s'enchaînent de manière chronologique pour atteindre un objectif au sein d'une organisation.

Il résulte de cette définition qu'à un processus peuvent correspondre plusieurs procédures et qu'une procédure concerne généralement plusieurs fonctions et une ou plusieurs entités au sein de l'organisation.

Une procédure peut et doit être un résultat issu de la modélisation et ne doit pas être confondu avec un mode opératoire qui lui décrit comment réaliser une opération, élément d'une procédure.

Figure 11 : Activité, processus et procédures



L'analyse des processus et procédures nous permet d'appréhender les dynamiques internes qui font que l'organisation existe et se développe.

L'identification des différentes fonctions et de leurs interactions :

L'identification des différentes fonctions présentes dans l'organisation passe généralement par une phase d'entretiens et de description détaillée des missions et des attributions de l'ensemble des membres d'une organisation. La compréhension du "Qui fait Quoi" nous permet alors d'évaluer le niveau de flexibilité de l'organisation et son aptitude à intégrer tous les changements.

La phase d'entretiens nous renseigne également sur la personne qui est en charge d'une fonction (attribution de la fonction), sur les membres de l'organisation qui contribuent à l'exercice de cette fonction (contribution à la fonction) et sur ceux qui assurent le remplacement du titulaire et des contributeurs si ceux-ci sont absents (polyvalence des personnes).

L'utilisation des fiches de définitions de fonctions (appelées également fiches de postes) nous permet alors d'améliorer considérablement la vision que nous avons d'une organisation.

Les fiches de définitions de fonctions, correspondant généralement à des postes types dans l'organisation, décrivent :

- Les missions de la personne qui occupe le poste, les objectifs qui lui sont assignés, les pouvoirs qui lui sont conférés,
- Les attributions, c'est-à-dire ses responsabilités, ses compétences, ses tâches, ses droits de gérer, d'administrer, de connaître,
- Une liste de documents, base de la circulation de l'information entre les différents postes de l'entreprise.

Elles sont la base du projet d'organisation, à travers lequel on peut :

- Identifier les fonctions de manière précise ;
- Définir l'organigramme de l'organisation ;
- Prendre en considération la communication entre les différentes fonctions.

Afin de synthétiser l'ensemble de ces informations, il est alors possible de recourir à un Tableau de Répartition des Responsabilités (TRR) qui a pour objectif :

- De recenser dans le détail l'ensemble des missions, attributions et tâches des collaborateurs de l'organisation et de préciser leurs niveaux de responsabilité ;

- De connaître avec précision le contenu de chaque fonction de l'organisation et de le positionner correctement par rapport aux autres collaborateurs qui assument une responsabilité en amont ou en aval de la fonction considérée.

Un TRR se présente comme un tableau composé de lignes et de colonnes.

Les lignes du TRR correspondent aux opérations, tâches et missions susceptibles d'être réalisées par l'ensemble des collaborateurs de l'entreprise.

Les colonnes du TRR correspondent à l'ensemble des fonctions de l'organisation ainsi qu'aux différents niveaux de responsabilités.

Tableau 1 : Tableau de Répartition des Responsabilités (TRR)

		Direction Générale				DAF				Dir. Client et Marketing				Marketing				ADY				Bureau d'études			
		Elabore	Décide	Applique	Gère	Elabore	Décide	Applique	Gère	Elabore	Décide	Applique	Gère	Elabore	Décide	Applique	Gère	Elabore	Décide	Applique	Gère	Elabore	Décide	Applique	Gère
3	Elabore, propose, définit, participe à (E) Décide, est responsable de (D) Applique, exécute, met sa marque (A) Gère les conséquences, les répercussions (G)																								
4																									
5	Politique commerciale																								
6	Détermination de la politique de prix, conditions de vente																								
7	Détermination des réseaux et méthodes de distribution																								
8	Gestion des modifications de tarifs																								
9	Stratégie de prospection et de vente																								
10	Marketing																								
11	Détermination du marketing mix de la société																								
12	Etablissement du plan marketing																								
13	Adaptation ou abandon de produits existants																								
14	Conception de nouveaux produits																								

L'identification des différentes fonctions présentes dans l'organisation nous permet alors d'appréhender l'ensemble des éléments constitutifs d'une organisation, son niveau de flexibilité et son aptitude à intégrer le changement.

Définition de la structure et de l'exploitation des informations et des connaissances :

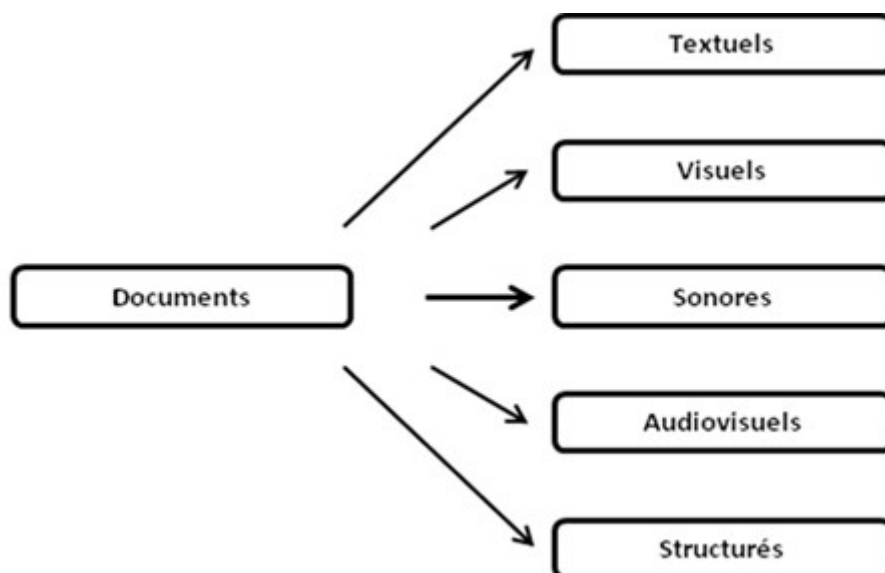
Dans l'optique de mieux comprendre les différentes fonctions présentes dans l'organisation, il est nécessaire d'appréhender les différents flux d'information ainsi que le rôle joué par chacun des acteurs qui est à la fois producteur et consommateur d'information.

Chaque fonction de l'entreprise étant portée par un acteur humain, il convient d'identifier les flux d'information qui lui sont transmis (ou événements entrants), les flux d'information qu'il émet (ou événements sortants), les objectifs de sa mission, les règles/contraintes dont il doit tenir compte et les traitements, c'est-à-dire les actions qu'il doit normalement effectuer.

On distingue essentiellement deux types de traitements opérés par un acteur humain dans le cadre de sa fonction : les opérations (une opération est un ensemble d'actions à réaliser dans le cadre d'un processus métier) et les décisions (une décision est un processus cognitif qui vise à sélectionner un type d'action parmi différentes alternatives). Pour mener à bien ces deux types de traitements, l'acteur humain a besoin d'information. Cette information est alors représentée par le biais d'un objet d'information qui n'est autre que le support matériel de l'information.

Dans le cas de la GED, cet objet matériel (ou support physique de l'information) est un document dont il est possible de préciser la nature.

Figure 12 : Nature des documents



Une typologie des différentes sources d'information peut alors être établie afin de préciser la structure de l'information qui sera susceptible d'être exploitée dans le cadre de la GED.

Tableau 2 : Différents types d'information

<i>Type d'information</i>	<i>Caractéristiques</i>
Physique	Elle se présente sous la forme d'une agrégation de données quantitatives ou qualitatives décrivant un état ou un fait. L'information est essentiellement déclarative et peut être purement physique (attributs), logique ou instrumentale (par exemple, l'extrait K-bis d'une entreprise)
Pragmatique	Elle se réfère à la qualité de l'information, sa pertinence, sa précision, sa validité, sa lisibilité ainsi que sa mise en forme (par exemple, une procédure)
Rationnelle	Elle fait référence au raisonnement, à une démonstration reproductible ou déduction logique pour établir une vérité (par exemple, l'absence de résultats d'une société peut résulter d'une baisse de ses recettes et du maintien de ses charges d'exploitation)
Paradigmatique	Elle désigne une évidence, une valeur partagée, mais non démontrable. Elle est l'expression d'un consensus plus que de la vérité (par exemple, la démocratie est un

<i>Type d'information</i>	<i>Caractéristiques</i>
	système politique qui respecte les droites de l'homme)
Expressive	Elle désigne une information destinée à créer une impression, une réaction émotionnelle ou affective pour faire passer un message. Elle peut être non langagière (images, sons, films) ou orale (style narratif, figure rhétorique). Par exemple, le docu-film Home du Photographe Yann Arthus-Bertrand

La Gestion Electronique de Documents ayant vocation à gérer tous les types de documents et ce, indépendamment de leur contenu et du support original, elle peut également permettre à l'acteur humain de donner à l'information une représentation qui lui fasse sens et ainsi transformer l'information en connaissance.

L'approche systémique que nous avons adoptée nous amène à considérer l'information comme un facteur d'organisation car elle permet de mettre en communication tous les éléments d'une organisation (qui est vue comme un système dynamique).

La compréhension de la structure et du mode d'exploitation des informations et des connaissances peut alors nous aider à améliorer la capacité d'une organisation à évoluer dans le temps tout en restant organisée.

Evaluation de la capacité au changement :

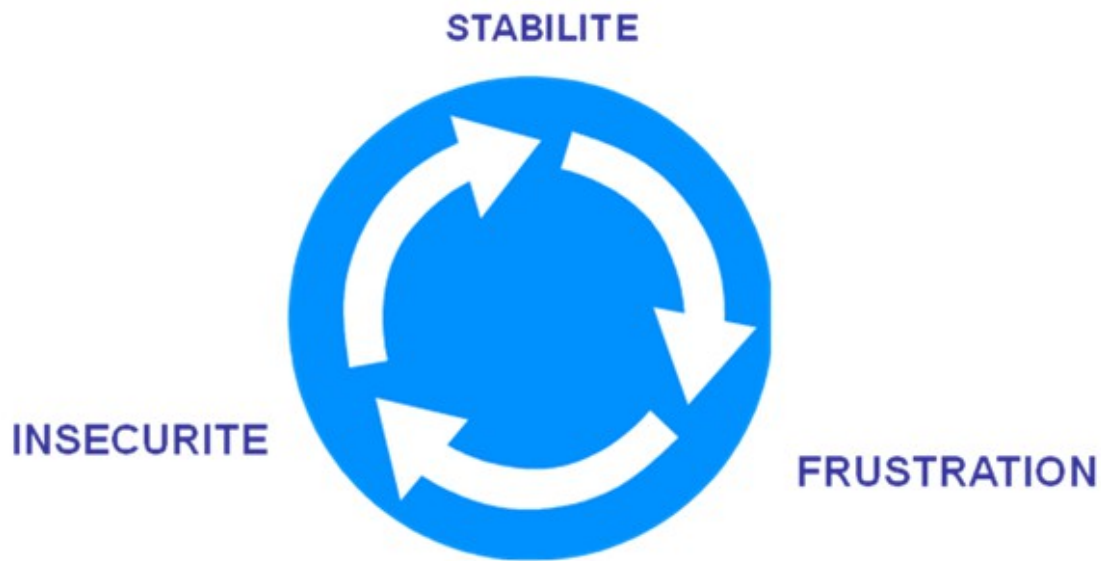
Le changement étant la seule permanence au monde, nulle organisation ne peut exister ou survivre dans l'immobilisme.

Car la stabilité que nous cherchons tous finit toujours par entraîner une forme de frustration.

Cette frustration s'appelle le changement. Mais le changement provoque à son tour l'insécurité et la peur de l'inconnu.

L'insécurité appelant à la stabilité, c'est tout naturellement qu'une organisation qui vit s'inscrit en permanence dans un cycle de perpétuel changement.

Figure 13 : Le cycle du changement



Une fois ce postulat admis, il convient alors de déterminer la capacité de l'organisation à gérer le changement.

Cette capacité repose en grande partie sur l'aptitude au changement des acteurs humains de l'organisation.

Afin d'évaluer cette aptitude collective au changement, il est d'usage de recourir à une analyse de la pratique du changement dans l'organisation.

Cela passe par la réponse à un certain nombre de questions :

- Quels changements majeurs l'organisation a-t-elle connus dans les douze derniers mois ?
- Ces changements ont-ils été voulus ou subis ?
- Comment ces changements ont-ils été gérés ?
- Les différents acteurs de l'organisation ont-ils été impliqués ?
- Avec quels résultats ?
- Etc.

Au niveau de chaque acteur, un questionnement est également opéré :

- Quel changement majeur a-t-il vécu au cours de ces douze derniers mois ?
- Essaie-t-il d'anticiper les changements extérieurs ?
- Le font ils réfléchir à sa stratégie ?
- A-t-il participé à des projets de changement interne ?

- Est-il assez attentif aux suggestions extérieures ?
- A-t-il une réaction positive face aux changements ?
- Etc.

Cet ensemble de questions dont l'objectif principal est de déterminer la capacité des différents acteurs à gérer le changement, nous conduit également à mieux appréhender les moteurs et les freins du changement dans l'organisation.

D'autres éléments sont aussi à prendre en considération si l'on veut évaluer "justement" la capacité de changement d'une organisation. Il s'agit notamment de son histoire, de ses valeurs clés, des identités professionnelles, des pratiques managériales, du dialogue social, des canaux formels et informels d'information et de communication, etc.

La prise en compte de la dimension activité nous a permis d'acquérir une vision globale de l'organisation, de ses caractéristiques, contraintes, évolutions, etc.

Elle nous a également fourni une évaluation de la capacité de l'organisation à gérer le changement occasionné par la mise en œuvre d'un projet de Gestion Electronique de Documents.

Enfin, elle nous laisse entrevoir le rôle considérable de l'acteur humain dans la réussite d'un projet de ce type.

2) La dimension humaine

La dimension humaine s'intéresse au seul paramètre capable de s'auto performer et de performer les autres composantes d'une organisation à savoir : l'acteur humain.

Mais, s'intéresser à l'acteur humain, c'est avant tout prendre en considération ses compétences, son histoire et son rôle au sein de l'organisation.

C'est aussi comprendre, que dans le cadre de sa fonction, il est amené à gérer des flux d'information, à subir des contraintes liées à la nature de sa mission, à opérer un certain nombre de traitements et à exprimer des besoins en termes d'information qui tiennent compte de son profil et de son cadre de référence.

A ce titre, l'acteur humain présente un certain nombre de caractéristiques qui lui permettent ou non de réaliser une activité donnée.

Bien que ces caractéristiques peuvent évoluer dans le temps, la dimension humaine s'attache alors à analyser ces caractéristiques quelque soit le rôle et le profil des différents acteurs humains qui interviennent dans l'activité d'une organisation.

Identification des acteurs humains :

L'objectif est de pouvoir identifier l'ensemble des catégories d'acteurs de l'organisation qui ont recours à la GED dans l'exercice de leur fonction.

La GED étant à vocation universelle, il est parfois difficile de pouvoir déterminer quels acteurs sont véritablement concernés par la GED. D'autant plus, que bon nombre de solutions de GED déployées dans les organisations sont des systèmes multi-applications couvrant divers domaines fonctionnels.

Cependant, une des caractéristiques majeures d'un acteur humain qui utilise un système de GED concerne sa capacité à être producteur et/ou consommateur d'information.

Cette première segmentation permet de déterminer deux grandes catégories d'acteurs :

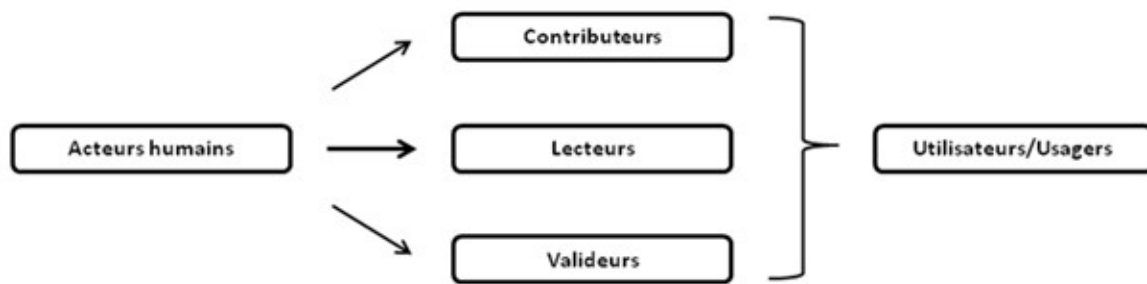
- Les producteurs d'information qui contribuent à faire vivre la GED en émettant des documents, supports matériels de l'information, qui viendront enrichir son fonds documentaire. Ces acteurs sont appelés des contributeurs (ou rédacteurs).
- Les consommateurs d'information qui utilisent régulièrement la GED pour formuler des requêtes, rechercher et consulter des documents et parfois même les transmettre directement depuis la GED à d'autres acteurs internes ou externes à l'organisation. Ces acteurs sont appelés des lecteurs.

Les acteurs qui utilisent régulièrement la GED mais qui contribuent également à la faire vivre et à enrichir son fonds documentaire (contributeur et lecteur) représentent la majeure partie des acteurs humains présents dans les organisations où une GED a été mise en œuvre.

A ces deux catégories d'acteurs, il convient d'en ajouter une troisième :

- Les acteurs humains qui possèdent une connaissance approfondie d'un certain nombre de processus "métier" et qui jouent le rôle d'experts fonctionnels susceptibles d'occuper une place importante dans la phase de caractérisation de l'information et d'indexation des documents. Ces acteurs sont appelés des valideurs.

Figure 14 : Les différentes catégories d'acteurs utilisateurs/usagers de la GED



A côté de ces trois catégories d'acteurs dits "fonctionnels", ayant des compétences en lien avec l'activité et les différentes fonctions de l'organisation, il existe également une quatrième catégorie d'acteurs dits "techniques" qui sont en fait les gestionnaires de la GED déployée au sein de l'organisation.

Chacun de ces individus, mis en situation professionnelle, présente un ensemble de capacités et d'aptitudes (savoirs, savoir faire et savoir être) qui lui permet d'exercer convenablement sa fonction au sein de l'organisation et qui constitue des éléments importants de son profil.

Définition des profils :

La définition des profils des différents acteurs qui seront amenés à utiliser un système de GED constitue un des points importants de notre approche.

La notion de profil fait référence à un ensemble de spécifications qui permet de caractériser un acteur humain et qui est potentiellement un utilisateur d'un système de GED.

Afin de pouvoir déterminer le profil d'un acteur humain, il convient de s'intéresser à son cadre de référence.

Le cadre de référence d'une personne humaine regroupe la culture, la religion, les croyances, les valeurs familiales, l'éducation, la somme des connaissances et des compétences acquises ainsi que le fruit des diverses expériences positives et négatives vécues dans le monde professionnel ainsi que dans la sphère privée.

Parmi tous les éléments constitutifs du cadre de référence d'un acteur humain agissant au sein d'une organisation, certains d'entre eux nous paraissent essentiels afin de comprendre l'émergence des besoins informationnels énoncés par cet acteur.

Il s'agit notamment :

- De la ou les fonctions qu'il exerce aujourd'hui et de son rôle au sein de l'organisation ;
- Des compétences qu'il a su développer ;
- De l'environnement socio-professionnel dans lequel il évolue actuellement.

La ou les fonctions occupées par un acteur humain au sein d'une organisation font référence à des "emplois" ou des "métiers" qu'il est possible d'appréhender à l'aide d'un répertoire professionnel, d'une nomenclature ou de toute autre instance de classification faisant autorité dans un domaine d'activité et servant de référence en matière de ressources humaines.

La notion de compétences fait référence à un ensemble de connaissances, de capacités d'action et de comportements structurés en fonction d'un but à atteindre dans un type de situation donnée.

Ces compétences reposent sur :

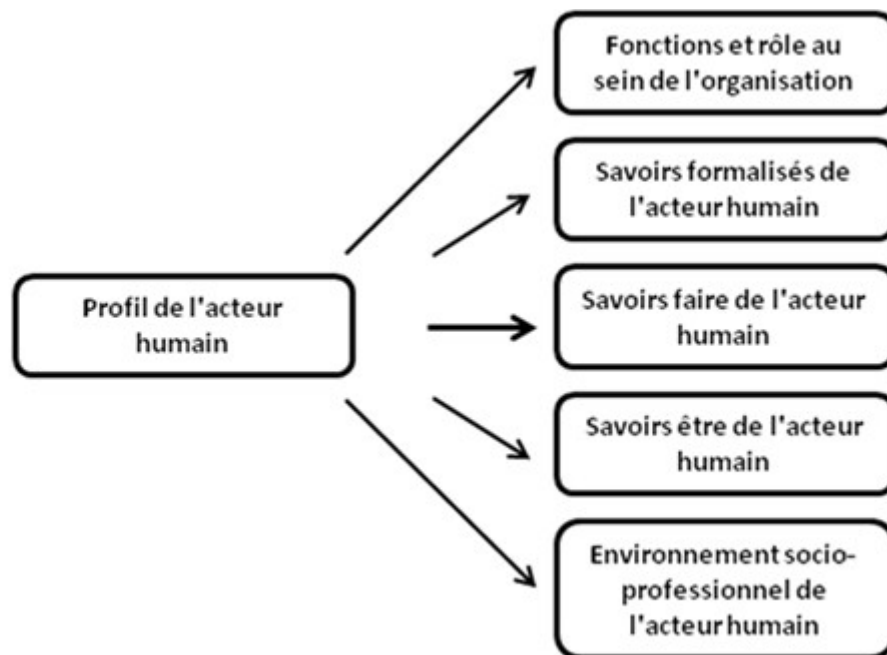
- Des savoirs formalisés : Il s'agit d'un ensemble de connaissances acquises par l'étude. On peut aussi parler de connaissances générales. Ce savoir n'est qu'une partie des connaissances nécessaires pour exercer une activité ou un métier.
- Des savoirs faire : Il s'agit d'un ensemble d'habiletés de différentes natures permettant de résoudre des problèmes plus ou moins complexes d'une situation de travail avec efficacité (qualité, délai, etc.). Le savoir faire se réfère toujours à une situation professionnelle spécifique. Il doit toujours pouvoir être validé à partir d'indicateurs d'efficacité et donc être exprimé en termes de "être capable de". Ce savoir faire s'acquiert par l'expérience et reste sujet à apprentissage.
- Des savoirs être : Il s'agit d'un ensemble de démarches et processus cognitifs (par exemple : capacité d'analyse, de diagnostic et de synthèse), affectifs, relationnels et physiques mobilisés par la mise en œuvre des savoir-faire.

L'environnement socio-professionnel correspond au milieu dans lequel évolue l'ensemble des acteurs humains.

Une partie des éléments qui caractérisent le profil d'un acteur humain au sein d'une organisation sont des paramètres endogènes. Il s'agit notamment de la ou les fonctions que l'acteur humain a pu exercer ainsi que des compétences professionnelles qu'il a pu acquérir par le biais de la formation et de l'expérience. D'autres paramètres, comme l'ensemble des éléments qui caractérisent l'environnement socio-professionnel dans lequel évolue l'acteur humain, constituent des paramètres exogènes qu'il convient également de prendre en considération pour déterminer son profil.

Ces paramètres, loin d'être exhaustifs, permettent de mieux cerner les besoins informationnels émis par un certain nombre d'acteurs que l'on peut identifier par profil et regrouper en grandes catégories.

Figure 15 : Eléments caractéristiques du profil d'un acteur humain



Identification des besoins informationnels :

Lorsqu'un acteur humain utilise un système de GED, c'est généralement parce qu'il recherche de l'information afin de pouvoir réaliser des opérations ou prendre des décisions.

L'ensemble de ces traitements se situant à différents niveaux : niveau stratégique, tactique ou opérationnel; la nature des informations recherchées dépend fortement du rôle joué par l'acteur humain au sein de l'organisation et des objectifs qui lui ont été assignés dans le cadre de sa ou ses fonctions.

La GED ayant pour vocation la représentation, la gestion, l'organisation et l'accès à l'information, elle ne peut apporter des réponses qu'à des requêtes formulées par un utilisateur/usager. Ces requêtes représentent alors le besoin d'information de l'acteur humain, utilisateur/usager d'un système de GED et identifié par un profil.

Compte tenu de ces éléments, il est clair que la prise en compte du besoin d'information de l'acteur humain est au cœur du dispositif de recherche d'information dans un système de GED.

L'analyse de ce besoin exige que soient connues les raisons qui conduisent l'acteur humain à s'engager dans un processus de recherche d'information.

Cette analyse se fait d'abord en aidant l'acteur humain à formuler par questionnaire son besoin d'information. Vient ensuite une phase de négociation de ce besoin par l'instauration d'une interaction informationnelle où alternent questions et réponses.

À l'issue de cette interaction, qu'elle s'établisse entre deux acteurs humains, entre un acteur humain et un ordinateur ou entre deux acteurs humains par l'intermédiaire d'un ordinateur, apparaît tout naturellement un diagnostic du besoin d'information de l'utilisateur/usager.

L'analyse des besoins informationnels des différents acteurs humains fait donc appel à des méthodes et des techniques de questionnement et de diagnostic que bon nombre de concepteurs de systèmes de GED ont besoin de maîtriser afin d'implémenter des outils interactifs permettant de mieux préciser le besoin d'information d'un utilisateur/usager au sein d'un système GED.

Car, plus le besoin d'information sera compris et meilleur sera l'usage finalement fait de l'information et de la GED au sein d'une organisation.

Selon Nicholas J. BELKIN et Alina VICKERY [BELKIN N. J., VICKERY A., 1985], le besoin d'information est créé par la prise de conscience par un individu d'une lacune dans l'état de ses connaissances et peut être vu comme un processus intrinsèquement interactif qui nécessite un dialogue, itératif et dynamique.

La recherche d'information étant rarement une fin en soi (sauf pour les professionnels de l'information), elle participe en général à la réalisation d'une autre tâche. Par conséquent, Il est nécessaire d'analyser ce que l'individu doit faire (ou pense qu'il a à faire) de l'information pour mieux évaluer le besoin d'information.

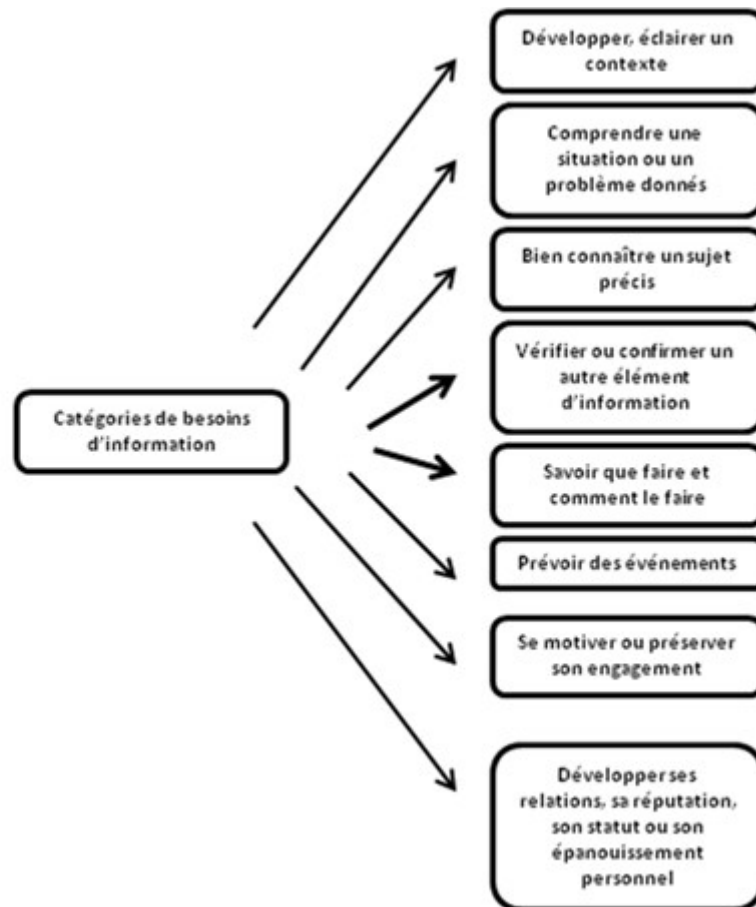
Les travaux de Kalervo JÄRVELIN et Peter INGWERSEN [JÄRVELIN K., INGWERSEN P., 2004] mettent clairement en évidence que pour répondre à un besoin d'information, il est nécessaire d'aider dans certains cas l'individu d'une part à se former une idée assez précise de la tâche qu'il doit effectuer, du but à atteindre et en parallèle, si besoin est, à s'appropriier peu à peu le domaine du sujet de sa recherche : acquisition du vocabulaire, repérage des sources pertinentes, appropriation des modes de raisonnement du domaine, etc.

Ils montrent aussi que le besoin d'information n'est pas ponctuel et de nature unique mais qu'il existe différents types de besoins qui ne demandent pas le même traitement.

Selon Kalervo JÄRVELIN et Peter INGWERSEN, s'appuyer sur des typologies de besoins pourrait permettre de mieux répondre aux demandes des usagers.

Parmi ces typologies, la typologie de Robert S. TAYLOR [TAYLOR R.S., 1986] basée sur le but et l'action permet de mieux appréhender la diversité des situations qui motivent l'acteur humain à rechercher de l'information et distingue huit catégories de besoins d'information qui ne sont pas seulement cognitifs mais également pragmatiques, psychologiques ou sociaux.

Figure 16 : Catégories de besoins d'information [TAYLOR R.S., 1986]



Identification des besoins en matière d'ergonomie et d'IHM :

Dans le domaine de la GED, la représentation de l'information est primordiale. D'où, la nécessité de concevoir des systèmes de GED qui soient ergonomiques, efficaces, faciles à utiliser ou plus généralement adaptés à leur contexte d'utilisation.

Une IHM ou une série d'IHM constitue alors la face visible d'un système de GED, une série d'interfaces qui va permettre aux différents acteurs d'une organisation de l'utiliser. Une IHM d'un système de GED étant liée aux processus d'indexation et de recherche d'information, elle résulte d'une démarche méthodologique précise.

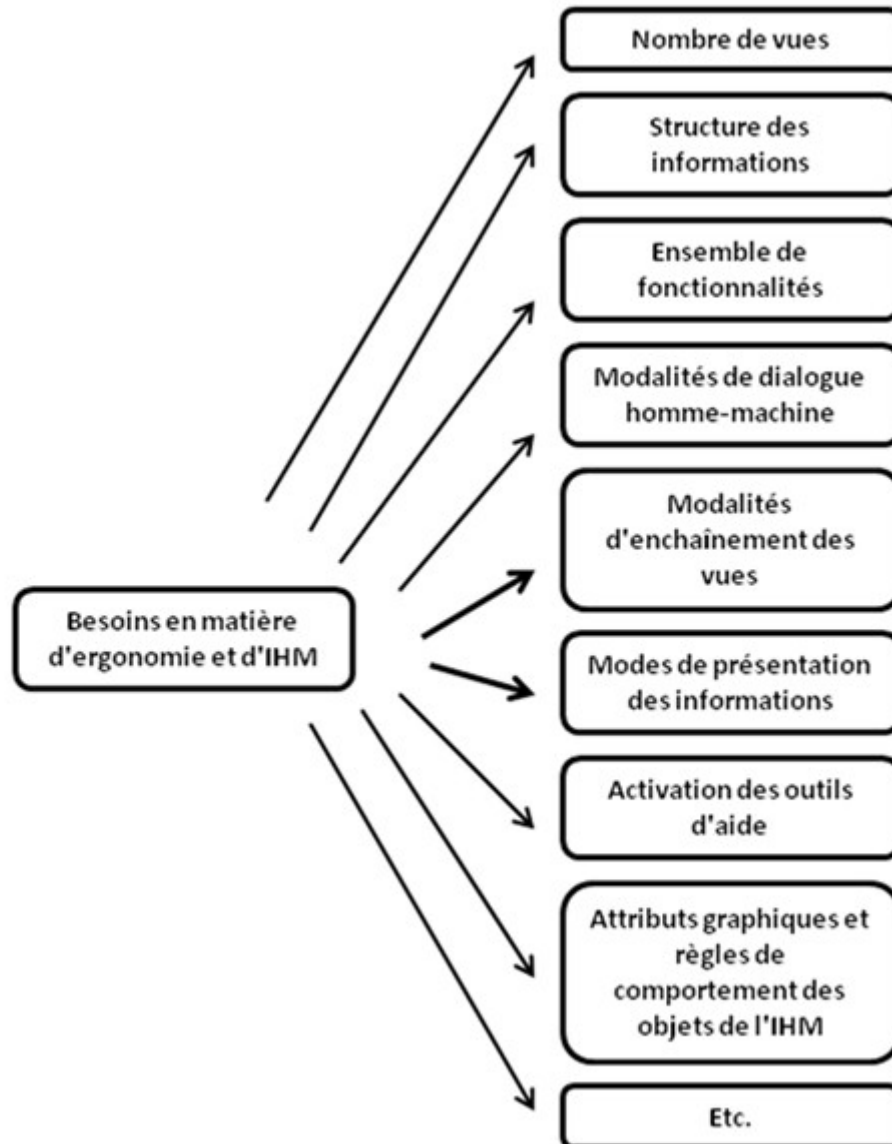
Il s'agit alors de spécifier puis de concevoir le nombre d'écrans à utiliser, certains écrans pouvant être dédiés à des tâches ou des informations particulières, la structure des ensembles d'informations et fonctionnalités à mettre à la disposition des utilisateurs, les modalités de dialogue homme-machine, les modalités d'enchaînement des vues, les modes de présentation des informations, d'activation des différents outils d'aide, les attributs graphiques et règles de comportement des objets de l'IHM, etc.

Pour cela, de nombreuses techniques et méthodes de conception sont disponibles et soulignent à quel point il existe une relation forte entre la conception de l'IHM et la conception interne de l'application de GED.

Cette relation se matérialise par une réflexion de fond sur la façon dont l'outil s'intègre dans son contexte opérationnel qui passe par l'appréhension d'une architecture, d'un mode de navigation et de la définition d'une ergonomie. La logique du concepteur est alors abandonnée au profit de celle de l'utilisateur/usager.

La prise en compte de la dimension humaine nous a permis d'appréhender le profil des différents acteurs qui participent aux processus d'indexation et de recherche d'information qui sont au cœur d'un système de GED. Elle nous a également permis d'identifier les besoins informationnels de ces acteurs ainsi que leurs attentes en matière d'ergonomie et d'IHM.

Figure 17 : Identification des besoins en matière d'ergonomie et d'IHM



3) La dimension technologique

Cette dimension s'intéresse aux composantes technologiques jugées essentielles dans un système de GED et prend en considération les différents éléments que nous avons pu appréhender à travers la dimension activité et la dimension humaine.

Cette approche passe par différents niveaux (fonctionnel, organisationnel et opérationnel) et portent essentiellement sur les éléments suivants :

- Analyse des outils disponibles ;
- Choix des outils ;
- Adaptation et mise en œuvre ;

- Evaluation des outils par les différents acteurs.

Analyse des outils disponibles :

Parmi l'ensemble des applications de GED disponibles sur le marché, il convient de déterminer un type de solutions susceptible de répondre aux différents besoins de l'organisation.

Cette catégorie d'applications relève généralement de l'un des cinq grands domaines suivants :

- La GED administrative et Bureautique : Ces applications s'intéressent aux documents vivants de l'entreprise ainsi qu'aux échanges et aux flux d'information ;
- Le COLD (Computer Output on Laser Disc) : Les applications de type COLD s'intéressent au stockage et à l'archivage des données issues de traitements informatiques ;
- La GED image : Il s'agit de l'ensemble des applications de GED qui gèrent des images ou des documents numérisés ;
- La GED documentaire : Ces applications gèrent les documents de référence d'une organisation avec pour objectif de constituer une mémoire organisationnelle ;
- La GED métier : Seules les applications spécifiques propres à un métier et qui demandent des solutions sur mesure sont concernées par cette catégorie ;

La détermination d'un type de solutions de GED nécessite la prise en compte de plusieurs facteurs qui résultent essentiellement d'une analyse détaillée de l'activité de l'organisation et des besoins informationnels émis par l'ensemble des acteurs humains.

Une fois ce type de solutions déterminé, il convient alors de s'intéresser aux fonctionnalités offertes par les différentes applications de GED relevant de cette catégorie, à savoir :

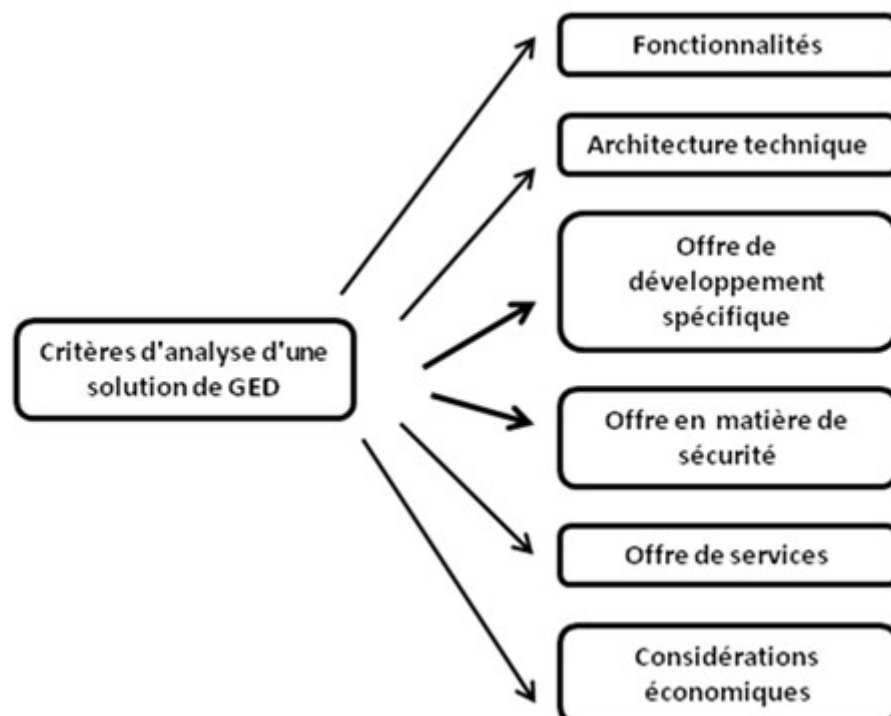
- L'acquisition (capture et indexation) ;
- Le stockage et l'archivage ;
- La recherche ;
- La consultation ;
- La diffusion ;
- Les éventuelles fonctions spécifiques.

D'autres critères doivent également être pris en compte comme :

- L'architecture technique (serveurs, postes de travail, configurations matérielles, systèmes, réseau, protocoles, etc.) ;
- L'offre de développement spécifique (langage de développement, API, interface avec des SGBD, intégration à l'environnement bureautique, etc.) ;
- L'offre de sécurité ;
- L'offre de services (paramétrage, installation, formation, maintenance, support technique personnalisé, services de numérisation, reprise de l'existant, conception d'un thésaurus, etc.) ;
- Des considérations d'ordre économique (prix des licences, des développements, des services, structure de représentation, références et sites installés, politique de mise à jour des produits, etc.).

L'ensemble de ces caractéristiques permettra de pouvoir comparer les différentes applications de GED présentes sur le marché répondant à un objectif bien déterminé. Certaines applications intègrent des fonctions qui peuvent être économiquement pénalisantes si elles ne sont pas utilisées à bon escient. D'autres font appel à des compétences qui, si elles ne sont pas adaptées à la nature des documents et au niveau des utilisateurs conduiront à un rejet de la solution de GED.

Figure 18 : Critères d'analyse d'une solution de GED



Choix des outils :

Après avoir analysé les différentes solutions de GED disponibles sur le marché, il est important de pouvoir rapprocher leurs fonctionnalités de celles requises par l'organisation (couverture fonctionnelle des outils).

Mais le choix d'une solution de GED dépend de bien d'autres facteurs que ceux reposant sur les seuls aspects fonctionnels (critères techniques, ergonomiques, économiques, etc.). C'est pourquoi, il est nécessaire d'établir une matrice de comparaison des différentes solutions susceptibles de répondre aux besoins énoncés par l'organisation qui devra faire ressortir quatre niveaux de critères.

Ces quatre niveaux sont :

- Les critères éliminatoires : Les éléments d'architecture technique et de compatibilité avec les dispositifs matériels et logiciels existants sont primordiaux dans le choix d'une solution de GED. La pérennité, l'expérience et la proximité de la structure de représentation et de support de la solution de GED sont également des facteurs capitaux ;
- Les critères de choix : Ce sont des critères fonctionnels, techniques ou ergonomiques qui vont peser pour le choix d'une solution au détriment d'une autre. Il est alors possible d'opérer des pondérations permettant de donner plus de poids à certains critères par rapport à d'autres ;
- Les critères optionnels : Ce sont généralement des fonctionnalités supplémentaires qui ne correspondent pas à un besoin clairement exprimé par l'organisation mais qui peuvent se révéler, à l'usage, fort utiles et appréciées ;
- Les évolutions annoncées par l'éditeur : La mise en œuvre d'une solution de GED au sein d'une organisation constitue un choix qui engage l'organisation sur plusieurs années. Il est donc important de consacrer une partie de l'analyse des différentes solutions de GED aux évolutions annoncées par les éditeurs dans de prochaines versions (prise en compte des fonctionnalités qui seront offertes et des dates de sortie prévisionnelles) ainsi que de prendre en considération les tendances actuelles et les visions des éditeurs à moyen terme.

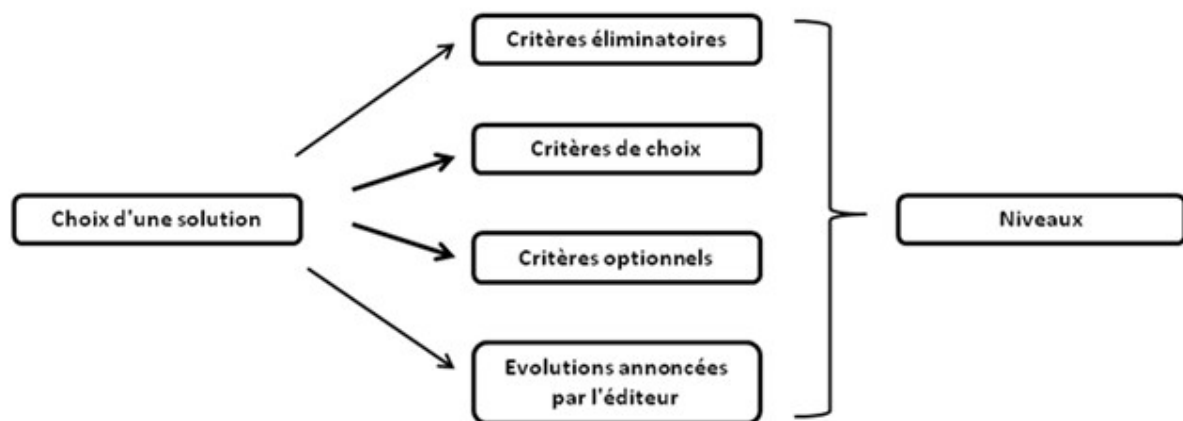
Pour ce qui est de l'analyse des critères éliminatoires, cela peut être opéré en consultant une documentation technico-commerciale ou par le biais d'un échange avec l'éditeur (basé sur un questionnement).

Concernant l'analyse des critères de choix, seule une démonstration de la solution, si possible personnalisée (c'est-à-dire intégrant des documents types de l'organisation), peut permettre une évaluation satisfaisante.

Afin de prendre en considération les critères optionnels, il est possible de recourir à une analyse de la documentation technico-commerciale, d'avoir un échange avec l'éditeur ou d'assister à une démonstration de l'ensemble des fonctionnalités offertes par la solution.

Une grille d'analyse comparative des différentes solutions de GED peut alors être établie afin de choisir l'une d'entre elles sur la base de critères fonctionnels, techniques, ergonomiques et économiques.

Figure 19 : Critères de choix d'une solution de GED



Adaptation et mise en œuvre :

Comme tout projet ayant trait au système d'information, la mise en œuvre d'une solution de GED va susciter des résistances humaines et culturelles, va modifier les modes de travail et par conséquent l'organisation et les rapports hiérarchiques. Elle peut également avoir un impact sur la qualité des services rendus, sur les délais de production ou sur la capacité de l'organisation à réagir face à une situation nouvelle.

La première source de difficultés vient du fait que *"toute organisation est naturellement hostile au changement et développe à son insu, des routines défensives qui sont anti-apprenantes et anti-productives"* [ARGYRIS C., 1995].

La méconnaissance de la solution de GED choisie, et de l'étendue de ses possibilités, par l'ensemble des utilisateurs potentiels, constitue le premier obstacle au déploiement de cette solution.

Dans un contexte innovant où une faible base est installée, le phénomène de rétroaction de l'utilisateur sur le concepteur est minime et ralenti considérablement le processus d'adaptation de l'outil aux besoins des utilisateurs.

L'autre difficulté provient du fait qu'il est très difficile de demander à un utilisateur de s'exprimer sur de nouveaux usages qu'il ne connaît pas encore. L'utilisateur doit alors passer par une phase d'appropriation de l'outil pour pouvoir être en mesure d'opérer un retour vis-à-vis du concepteur.

Lors de l'introduction de nouvelles technologies au sein d'une organisation, existe un risque d'attentisme "contagieux" qui rend difficile tout mécanisme d'adaptation de la solution aux éventuelles attentes des utilisateurs.

Les premières difficultés d'ordre technique rencontrées lors de la phase de mise en œuvre constituent également un facteur de rejet pour certains acteurs qui ne sont pas forcément favorables à la solution de GED.

La deuxième source de difficultés vient du fait que l'outil ne contient pas le progrès, mais que bien utilisé, il peut être un formidable catalyseur du changement.

Selon Chris ARGYRIS [ARGYRIS C., 1995], *"la mise en place de nouveaux modes de fonctionnement suppose une action conjointe à trois niveaux :*

- *Celui de l'outil et de son appropriation ;*
- *Celui du projet, c'est-à-dire des processus d'action qui génèrent les usages ;*
- *Celui des valeurs directrices sous-jacentes, au service d'une stratégie de changement".*

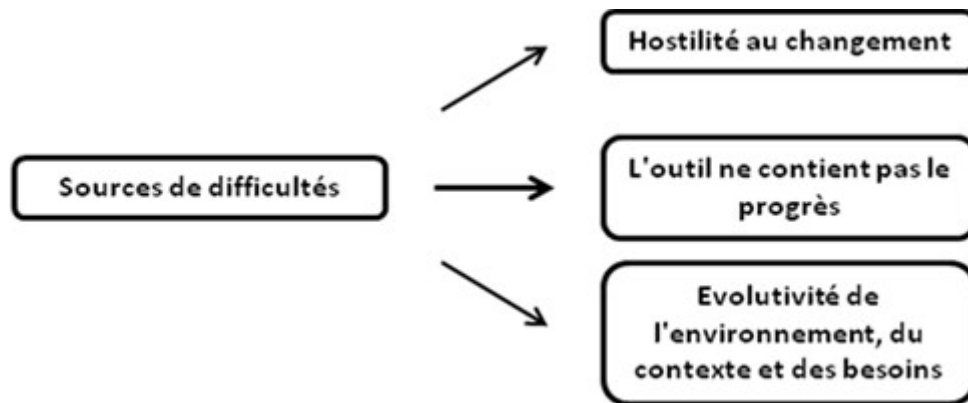
Dans la pratique, ces trois niveaux ne sont pas séparés et se nourrissent l'un l'autre.

La troisième source de difficultés tient à la prise en compte de l'évolutivité de l'environnement, du contexte et des besoins.

Au fur et à mesure de la mise en œuvre de la solution de GED et de son appropriation par les différents acteurs, l'organisation doit continuer à s'adapter à un environnement et une activité en perpétuel changement faisant ainsi naître de nouveaux besoins qui donneront lieu à de nouvelles spécifications qu'il conviendra d'intégrer au niveau de la GED.

L'organisation est alors confrontée au problème d'obsolescence de la solution de GED qui est pourtant conforme aux spécifications d'origine mais qui a du mal à prendre en considération les nouveaux besoins.

Figure 20 : Difficultés rencontrées lors de la mise en œuvre d'une solution de GED



Evaluation des outils par les différents acteurs :

L'objectif est alors d'évaluer le niveau de performance de la solution de GED. Cette évaluation repose sur la détermination de différents indicateurs qui peuvent être :

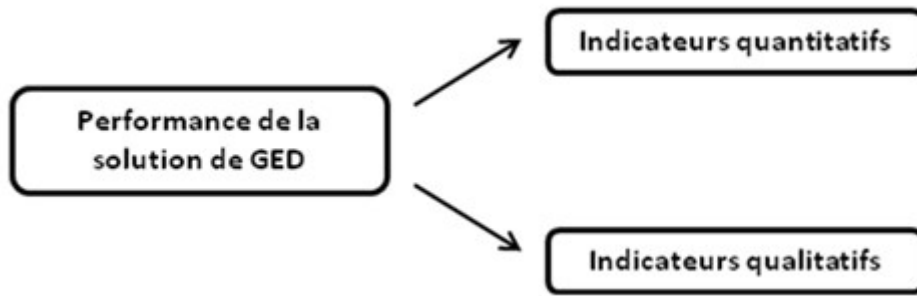
- Quantitatifs : nombre de documents se rapportant à un thème, nombre de consultations par jour, etc.
- Qualitatifs : niveau de satisfaction, facilité d'accès à l'information, etc.

L'implication des différents acteurs dans la définition de ces indicateurs est importante car elle facilitera leur compréhension et leur appropriation et permettra une réelle mobilisation autour d'objectifs partagés.

Cependant la principale difficulté tient à la pertinence de ces indicateurs qui devront :

- Apporter une réponse concrète aux finalités et objectifs exprimés ;
- Etre fiables ;
- Etre sensibles, c'est-à-dire que l'impact d'une action doit faire sensiblement varier l'indicateur ;
- Etre exprimés en termes d'écart ou de position par rapport à une situation souhaitée.

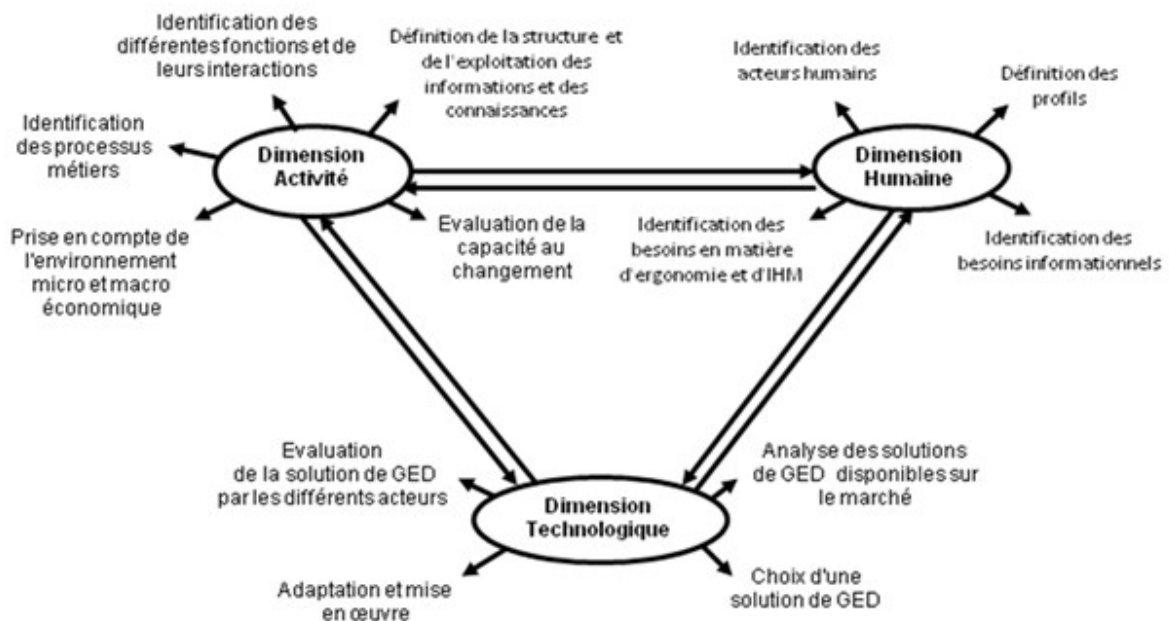
Figure 21 : Evaluation de la solution de GED



La dimension technologique facilite l'analyse des éléments nécessaires à la conception et à la mise en œuvre d'une solution de GED. Elle permet également de traiter de l'adaptation des outils et de leur évaluation par les différents acteurs.

D'un point de vue global, l'approche dite "tridimensionnelle" nous permet de mieux évaluer la démarche de conception et de mise en œuvre d'une solution de GED au sein d'une organisation et de mettre en relief le rôle central joué par l'acteur humain.

Figure 22 : Instanciation de l'approche dite "tridimensionnelle"



II. Apports théoriques et méthodologiques complémentaires

L'approche dite "tridimensionnelle" dont les principes furent énoncés par Laïd BOUZIDI [BOUZIDI L., 2001], constitue un cadre théorique et méthodologique sur lequel nous

pouvons nous appuyer afin de concevoir un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences.

Mais cette approche hybride présente également des limites par rapport à la prise en compte de notre problématique de recherche.

Cela s'exprime notamment en termes de représentation des interactions entre un utilisateur et une application informatique mais aussi en termes de compréhension de la manière dont fonctionne un acteur humain.

C'est pourquoi nous avons souhaité procéder à des apports théoriques et méthodologiques complémentaires afin de pouvoir enrichir notre cadre d'origine.

Ces apports se matérialisent sous la forme de deux méthodes :

- La méthode "Motivations" qui permet de mettre en évidence les interactions entre un acteur humain et des applications informatiques ;
- La méthode des niveaux logiques de la pensée, de l'action et du changement qui permet de mieux comprendre comment fonctionne un être humain.

1) La méthode "Motivations"

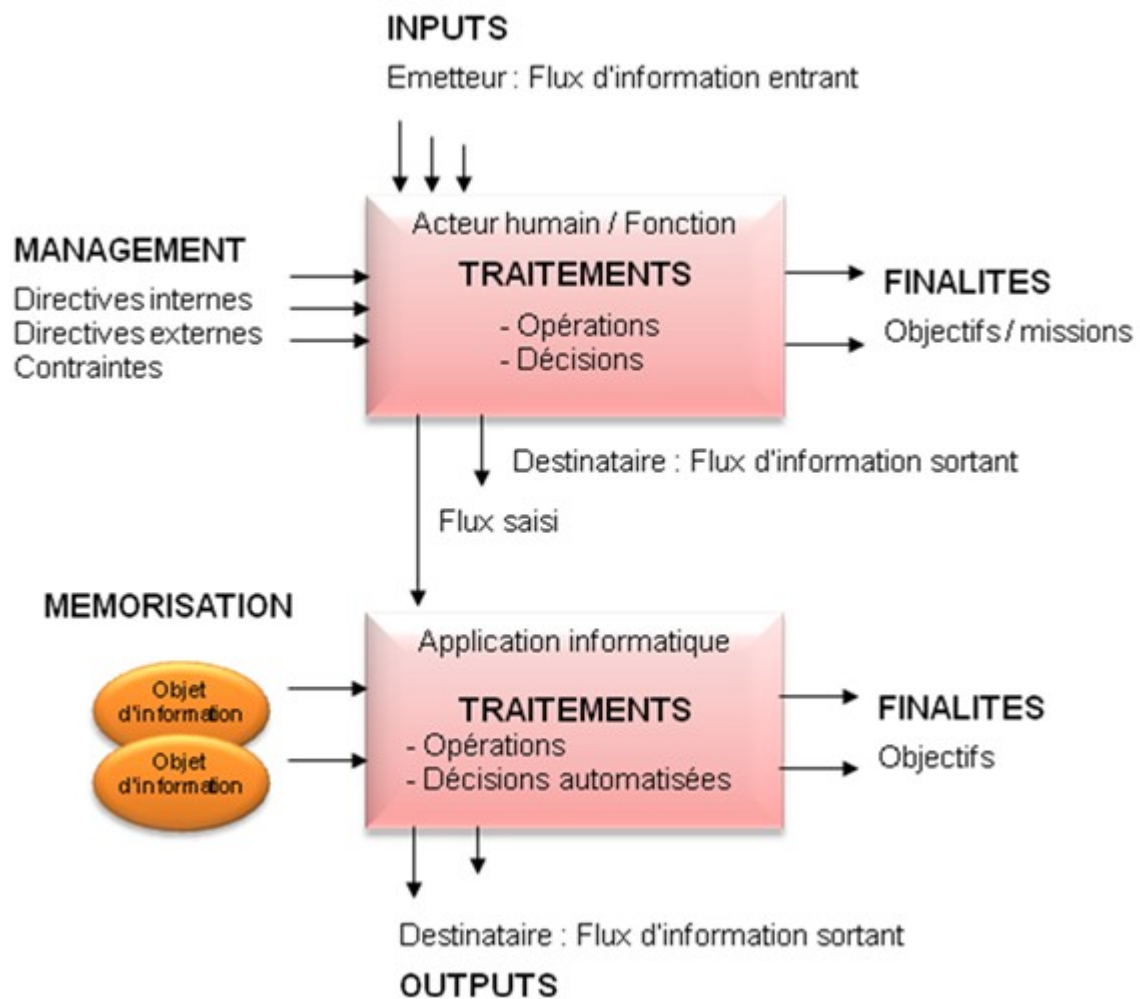
Initiée en 1991 à l'IAE de Lyon par Suzanne ASSIE [ASSIE S., 1999], "Motivations" est une méthode d'audit organisationnel des Systèmes d'Information qui permet d'analyser au sein d'un système d'information le rôle de chaque acteur, son métier ainsi que l'organisation globale de l'entreprise. Il est possible de mettre en évidence la liaison qui relie chaque système utilisé avec un acteur intervenant dans l'organisation.

Cette modélisation met en relief les interactions entre utilisateur et applications informatiques. Elle permet aussi de déterminer le rôle des systèmes informatiques dans la stratégie de l'entreprise. De plus, elle facilite le dialogue entre utilisateur et informaticien grâce à un langage commun. Ceci permet de recueillir des informations pour les étapes de conception.

La vision locale – au niveau de l'individu – relève du "Motif", et permet de comprendre le métier de l'utilisateur. Pour cela, il est nécessaire d'identifier l'acteur puis ses Inputs – flux d'information transmis à l'acteur ou événements entrants –, ses Outputs – flux d'information émis par l'acteur ou événements sortants –, les objectifs de sa mission, les règles / contraintes dont il doit tenir compte et enfin ses traitements, c'est-à-dire les actions qu'il effectue.

La vision globale de la méthode "Motivations" est donnée par une "Fresque". Cette dernière est construite à partir des différents motifs précédemment réalisés. Elle décrit à la fois l'organisation étudiée, les flux d'informations qui la caractérisent et les interactions entre ses acteurs dans le but d'établir un diagnostic de l'entreprise ou de l'organisation.

Figure 23 : Le "Motif", vision locale de la méthode "Motivations"

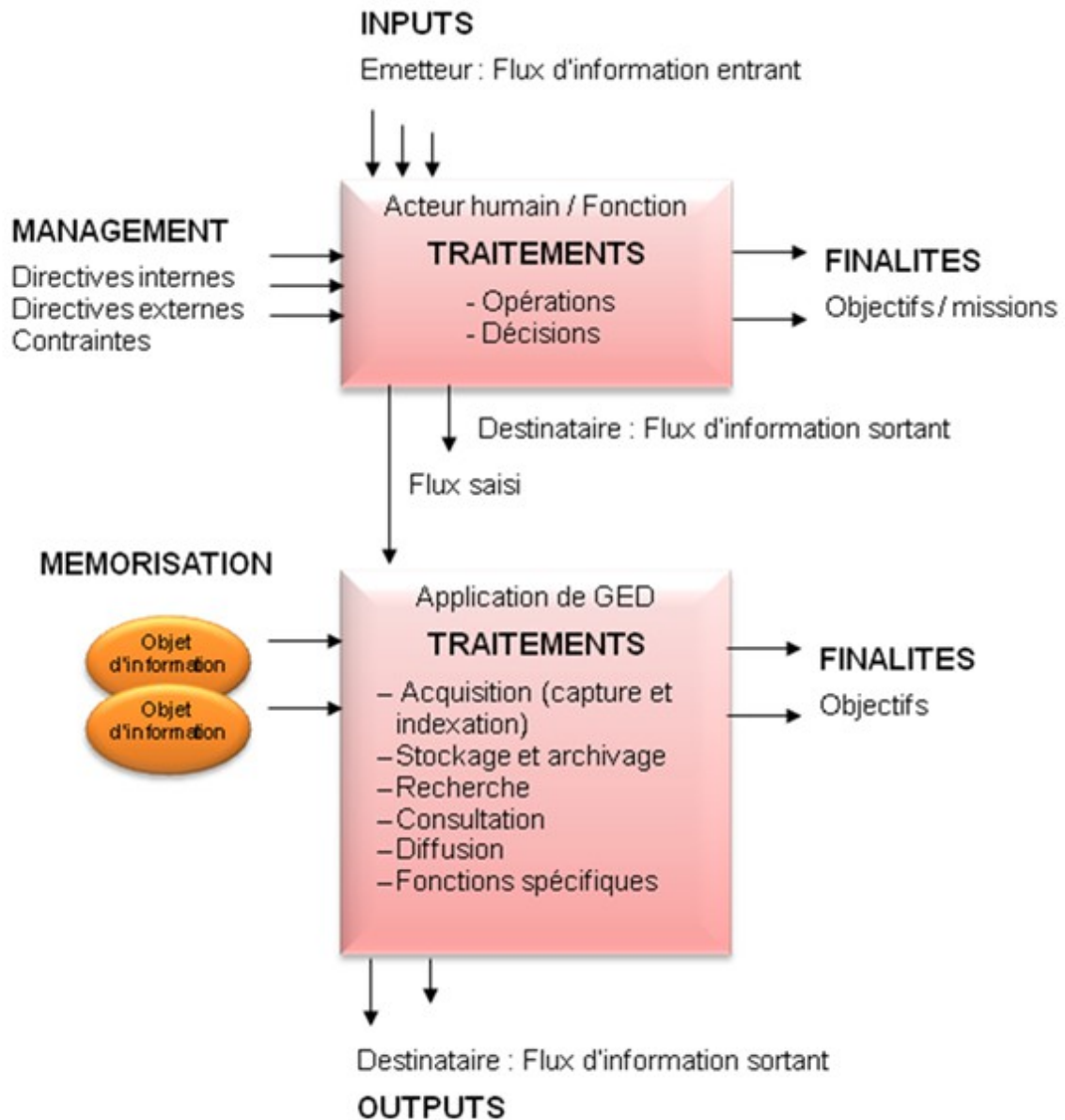


La fresque est une agrégation des motifs. Elle assure la continuité du modèle et représente les composants du système global. Elle met en évidence leurs interactions significatives et répond à une problématique posée tout en facilitant la compréhension de l'architecture globale. Elle est pour l'architecte un instrument de travail et un outil de communication avec les acteurs.

Dans le cadre de notre approche, la vision locale de la méthode "Motivations", à savoir le "Motif", nous permet de représenter l'acteur humain ainsi que l'acteur informatique (solution de GED), et de mettre en lumière l'objectif poursuivi par l'acteur humain dans le cadre de sa

recherche d'information. Il offre également une bonne représentation des interactions qu'entretient l'acteur humain avec la solution de GED.

Figure 24 : Interactions entre un acteur humain et une application de GED



2) Le modèle des niveaux logiques de la pensée, de l'action et du changement

Le modèle des niveaux logiques de la pensée, de l'action et du changement, mis au point par Robert DILTS [DILTS R., 1983] à partir des travaux de Gregory BATESON [BATESON G., 1973], constitue un puissant outil de développement personnel, permettant de mieux

comprendre comment nous fonctionnons, quelles sont nos motivations profondes, et de diriger notre évolution en agissant en cohérence avec ce que nous sommes.

Il donne une représentation structurée de l'être humain au moyen de plusieurs niveaux (ou domaines) de conscience, reliés les uns aux autres.

Les problèmes sont souvent engendrés par la confusion des niveaux logiques. Les changements ne se font pas aux différents niveaux selon les mêmes règles et la résolution d'un problème ne s'effectue pas au niveau où le problème se situe mais au niveau supérieur.

Ces niveaux peuvent être représentés sous forme d'un tableau. En partant du domaine concernant les aspects les plus concrets et les plus extériorisés, situés au bas du tableau, nous pouvons nous élever progressivement vers les autres niveaux, concernant des aspects de plus en plus intérieurs et subtils de notre être ou de l'organisation à laquelle nous appartenons.

Tableau 3 : Niveaux logiques de la pensée selon Robert DILTS

	Sur l'individu	Sur l'organisation
Mission (identité)	La personne que je suis	Mission, vocation/vision que l'on en a
Motivation (croyances)	Comment je perçois mon environnement	Système de culture, objectifs
Orientation (capacités)	Ce dont je suis capable, ce qui est possible pour moi	Politiques, savoir-faire techniques et humains
Action (comportements)	Ce que je fais	Plans d'actions, procédures
Réaction (environnement)	Le monde qui m'entoure	Environnement direct

Comprendre les motivations et le fonctionnement de chacun permet de résoudre certains problèmes qui surviennent bien souvent lors des phases de communication intra et inter personnelle.

Dans le cadre de notre approche, le modèle des niveaux logiques de la pensée, de l'action et du changement, mis au point par Robert DILTS constitue un réel apport méthodologique pour comprendre comment l'être humain fonctionne et comment il convient d'agir pour résoudre certains problèmes de communication non seulement intra et inter personnelle mais également entre un acteur humain et un outil informatique.

Par contre, ce modèle présente une réelle difficulté d'implémentation qui rend sa prise en compte difficile dans un éventuel système de recherche d'information.

Ce modèle a surtout le mérite de nous faire prendre conscience que lorsqu'une recherche d'information échoue, ce n'est pas forcément au niveau des comportements de l'acteur humain qu'il faut en rechercher les causes mais qu'il convient de s'interroger sur les capacités ainsi que sur les objectifs qui sont à l'origine de cette recherche d'information.

III. Proposition d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences

Dans un premier temps, nous présenterons les fondements théoriques d'une approche permettant d'améliorer la pertinence de l'information recherchée par la prise en compte du mécanisme de "Sérendipité" et l'apport du raisonnement par cas.

Nous énoncerons ensuite le principe de fonctionnement du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences proposé.

Enfin, nous établirons les conditions d'intégration d'un tel système au sein d'une solution de GED et détaillerons les facteurs clés de succès.

1) Les fondements théoriques

Dans la deuxième partie de cette thèse, nous avons souligné que malgré de profondes transformations et une formidable expansion, l'interrogation obéit toujours à une stratégie de recherche qui emprunte ses principaux modes de fonctionnement aux techniques actuelles, lesquels présentent un certain nombre de limites que nous avons précisées et repositionnées dans le contexte spécifique de la GED. Afin d'améliorer la performance des systèmes de recherche d'information intégrés aux solutions de GED, nous avons choisi de porter notre attention sur l'acteur humain/utilisateur plutôt que sur l'évolution des outils et techniques de recherche d'information actuels.

a) Une approche centrée sur l'acteur humain

L'approche que nous avons adoptée s'intéresse au seul paramètre capable de s'auto performer et de performer les autres composantes d'une organisation à savoir : l'acteur humain [THOMAS B., MARINI JL., CRETIN R., 2007].

Mais l'acteur humain n'est pas une ressource comme les autres, c'est avant tout un être vivant qui pense, réagit, a des attentes, des stratégies et qui ne peut être réduit à un capital de connaissances et de savoir faire.

S'intéresser à l'acteur humain, c'est prendre en considération sa personnalité et le fait qu'il est soumis à l'imperfection, laquelle s'exprime entre le dit et le non dit, entre le conscient et l'inconscient.

Pour mieux comprendre la structure mentale de l'acteur humain et par voie de conséquence l'ensemble des comportements qu'il est amené à adopter, nous ferons référence aux travaux de Sigmund FREUD [SAURET MJ., 1999], Jacques LACAN [SOLER C., 2009] et Carl Gustav JUNG [JUNG C.G., 1988] sur le conscient et l'inconscient.

b) Conscient et inconscient

Nous pouvons facilement admettre que la personnalité d'un être humain est composée de deux éléments : d'abord du conscient et de tout ce qu'il contient mais aussi d'un arrière-pays infiniment plus vaste nommé l'inconscient. La personnalité consciente peut être délimitée et définie plus ou moins clairement, mais lorsqu'il s'agit de l'ensemble de la personnalité humaine, on est obligé de reconnaître l'impossibilité d'en fournir une description et une définition complète.

En d'autres termes, il existe inéluctablement un élément indéfini et illimité, qui s'ajoute à toute personnalité : cette dernière comprend une partie consciente, susceptible d'être observée, mais elle n'englobe pas certains facteurs, dont cependant nous sommes forcés d'admettre l'existence, si nous voulons expliquer certains faits observés. Ce sont ces facteurs inconnus que nous appelons le secteur inconscient de la personnalité.

Pour tenter de comprendre la dynamique inconsciente d'un être humain ou lui apporter des modèles d'enrichissement de ses dynamiques inconscientes, il est nécessaire d'appréhender son cadre de référence.

c) Le cadre de référence d'un être humain

Le cadre de référence d'un être humain regroupe la culture, la religion, les croyances, les valeurs familiales, l'éducation, la somme des connaissances et des compétences acquises ainsi que le fruit des diverses expériences positives et négatives vécues dans le monde professionnel ainsi que dans la sphère privée.

Selon Jacqui Lee SCHIFF et Aaron Wolfe SCHIFF [SCHIFF JL. et SCHIFF A., 2008], "il fournit à l'individu un ensemble cohérent de perceptions, de concepts, de sentiments et d'actions, qui lui sert à se définir lui-même, les autres et le monde, de façon structurale et dynamique".

Le cadre de référence unique de chacun est donc déterminé par :

- Un formatage ;
- Des perceptions sensorielles ;
- Des limitations liées au langage (communication intra et inter personnelle).

d) La communication intra et inter personnelle

La majorité des problèmes entre êtres humains est générée par des problèmes de communication [PICARD D. et MARC E., 2000].

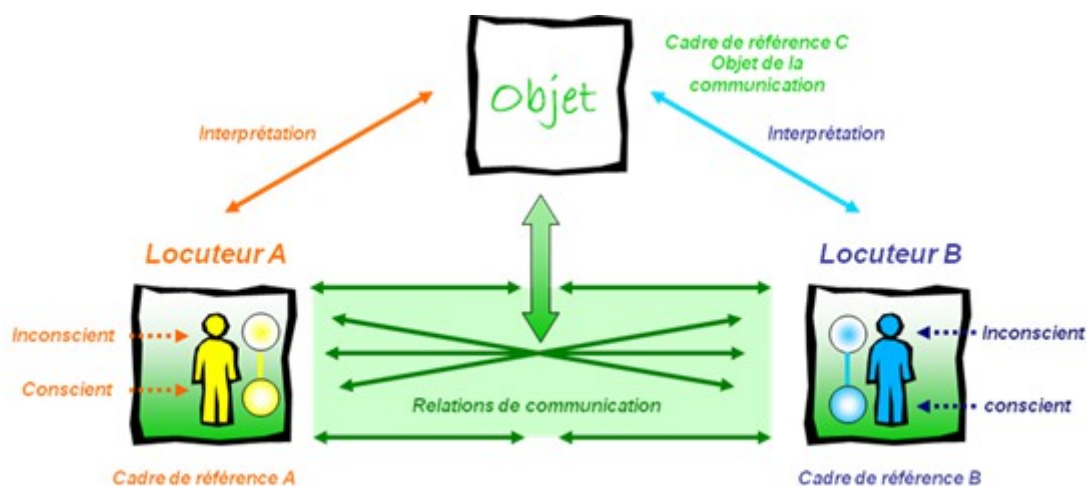
Dans les communications interpersonnelles, les principales difficultés sont dues à des problèmes de langage.

Le langage est une émanation du cadre de référence de chaque personne.

Chaque être humain étant unique, il y a donc autant de cadres de référence, et donc de langages, que d'êtres humains.

Plus les cadres de référence sont différents, plus il y a difficulté à communiquer.

Figure 25 : Cadres de référence et communication interpersonnelle



Dans notre société, ceux qui possèdent l'information et les moyens de la communiquer disposent d'un pouvoir illimité.

Car même à l'ère de l'information, l'information ne suffit pas, le savoir n'est qu'un pouvoir potentiel jusqu'à ce qu'il tombe entre les mains de celui qui sait comment se mettre en condition de prendre des mesures efficaces. La définition littérale du mot "Pouvoir" est "Capacité d'agir", l'action est donc ce qui donne des résultats.

Ce que nous faisons dans la vie est déterminé par la façon dont nous communiquons avec nous-mêmes [CAYROL A. et DE SAINT PAUL J., 2005].

La communication est un pouvoir. Ceux qui en maîtrisent l'emploi peuvent modifier la notion qu'ils ont du monde et la notion que le monde a d'eux. Tous les comportements, toutes les sensations s'enracinent dans une forme quelconque de communication. [ROBBINS A., 2003].

Nous produisons tous deux types de communication qui modèlent notre rapport à l'existence :

- Les communications internes (les choses que nous imaginons et que nous ressentons à l'intérieur de nous-mêmes) ;
- Les communications externes (les mots, les intonations, les expressions du visage, les postures du corps et actions physiques par lesquels nous communiquons avec le monde).

Toute communication que nous émettons est une action, une cause mise en mouvement. Et toutes les communications produisent un effet sur nous-mêmes et sur les autres.

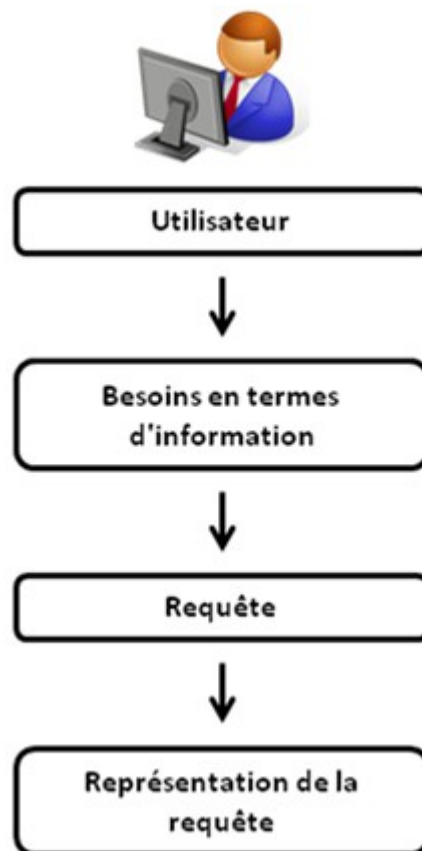
e) La communication entre un acteur humain et un outil technologique

Dans le cadre de notre approche, l'acteur humain qui s'engage dans un processus de recherche d'information communique avec un ordinateur afin de pouvoir utiliser une application de GED.

En utilisant cette application, il souhaite pouvoir accéder à des fonctionnalités de recherche d'information qui lui permettront de satisfaire ses besoins informationnels.

Par l'instauration d'une interaction avec l'application de GED, il formule une requête qu'il soumet au système de recherche d'information (expression des besoins mentaux). Une traduction de cette requête est alors effectuée par le système de recherche d'information (représentation de la requête) dans l'optique de procéder à des comparaisons et d'établir des appariements avec un ensemble de documents issus du fonds documentaire de la solution de GED.

Figure 26 : Représentation du besoin d'information



Ce schéma illustre une situation idéale qui est loin d'être toujours le cas.

Le plus souvent des problèmes de communication apparaissent entre l'acteur humain et le système de recherche d'information (outil technologique). Ils proviennent essentiellement du fait qu'identifier un besoin d'information n'est pas naturel ou inné. On peut ressentir un besoin sans savoir le caractériser, ne pas savoir de quelle information on a besoin ou ne pas avoir conscience qu'il y a problème ou lacune dans ses connaissances [JULIEN H.E., 1999].

Mais il faut également tenir compte du fait qu'un ordinateur n'est pas un acteur humain et que la communication entre un acteur humain et un ordinateur ne relève pas des mêmes principes que ceux évoqués dans le cadre de la communication interpersonnelle (même si celui qui a indexé l'information est un acteur humain et que l'ordinateur et l'application de GED ne sont que des intermédiaires). La différence majeure provient de la capacité d'un ordinateur à traiter et ne comprendre que des chaînes de caractères, c'est-à-dire des mots, alors que l'esprit humain fonctionne avec des concepts [FROCHOT D, 2007].

f) Du besoin informationnel à la recherche d'information

D'après Yves-François LE COADIC [LE COADIC YF., 1998], *"il existe un véritable paradoxe à chercher une chose que l'on ne connaît pas"*. Ce paradoxe a d'ailleurs été mis en évidence par PLATON dans "Menon" et nous amène à nous poser la question de savoir quelle information chercher parmi l'ensemble des informations auxquelles il est possible d'accéder. L'autre question soulevée par la prise en compte de ce paradoxe concerne la manière d'identifier que l'information trouvée est bien celle que l'on cherche.

Selon Jean-François ROUET et André TRICOT [ROUET JF., 2000 et TRICOT A., 2003], pour chercher efficacement de l'information, il faut des connaissances préalables : on ne cherche que si l'on sait que l'on ne sait pas et que l'on peut trouver. Par conséquent, il faut accepter l'incertitude et avoir une motivation pour la réduire ou l'éliminer

Notre approche part du principe que pour mettre en œuvre une stratégie de recherche d'information efficace, il faut avoir une idée assez précise du but à atteindre [ROUET JF., 2000] et pouvoir le décomposer en objectifs intermédiaires. Elle tient compte également du fait qu'à l'origine de tout besoin d'information, il y a une situation perçue comme insatisfaisante, reconnue par un acteur humain qui veut lui apporter une solution.

Dans ce sens, on peut dire que toute recherche d'information est la rencontre de deux réalités : une situation perçue comme insatisfaisante d'une part et un acteur humain d'autre part.

Une recherche d'information est la relation dynamique et créatrice entre ces deux réalités.

Le mécanisme de recherche d'information repose alors sur la prise en compte du besoin informationnel et la définition d'un objectif de recherche qui peut lui-même se décomposer en objectifs intermédiaires (étapes dans le processus de recherche).

g) L'étude du besoin informationnel

Afin de définir un objectif de recherche d'information, il convient au préalable de déterminer l'étendue du besoin informationnel.

A ce niveau une distinction doit être apportée entre deux notions souvent confondues : le besoin et l'attente. Une attente est une projection de l'insatisfaction ressentie par un acteur humain. Elle se formule sous une forme passive et reste générale. Alors qu'un besoin est une sensation d'insatisfaction qui provient de l'interaction entre un individu et son environnement. Il se formule de manière active (avec un verbe d'action). Il est individualisé, précis, conceptualisable, ciblé dans sa formulation et dirigé vers le futur. Il permet de découvrir de nouvelles facettes d'un sujet.

Compte tenu de ces éléments, il apparaît que le besoin d'information d'un acteur humain relève d'un ressenti qu'il est le seul à pouvoir exprimer clairement (pour ce qui est de la partie consciente).

L'acteur humain est alors invité à décrire la situation actuelle qui repose sur :

- Les interactions qu'il entretient avec son environnement ;
- La ou les sources de son insatisfaction.

Par un processus d'abstraction, on tire de cette description, une "représentation" du besoin informationnel ainsi que des "indicateurs" qui permettent de mesurer l'insatisfaction actuelle.

Cette "représentation" et ces "indicateurs" nous serviront ensuite à mesurer le chemin à parcourir afin de revenir à une situation satisfaisante et à déterminer les différentes étapes que suscite cette évolution.

Suite à cette description, l'acteur humain est alors en mesure de fixer des orientations qui tiennent compte :

- De ses attentes en matière d'information ;
- Des divers types et formats de sources potentielles d'information ;
- Des concepts et des termes clés qui décrivent le besoin d'information ;
- De la nature et de l'étendue du besoin d'information.

A partir de la situation actuelle et d'orientations fixées par l'acteur humain, un objectif de recherche d'information peut être alors déterminé.

h) L'objectif de recherche d'information

Si l'on s'intéresse aux acteurs humains qui ont très souvent accès à de l'information pertinente et de qualité, on s'aperçoit alors qu'ils possèdent tous la capacité à se fixer un objectif de recherche qui découle de leur aptitude à bien formuler cet objectif [BANDLER R et AL., 2000].

Pour ce faire, un objectif de recherche d'information doit répondre à un certain nombre de critères :

- Il est formulé en **termes positifs** (sur le plan linguistique) ;
- Il est **précis** (quelle est la nature de l'information recherchée, quelles sont les données que nous possédons pour rechercher cette information, quel est le support matériel de cette information) ;
- Il est **sous le contrôle** de la personne (l'acteur humain a-t-il la maîtrise du système de recherche, possède-t-il les connaissances nécessaires pour valider que l'information trouvée est pertinente et de qualité, quelles sont les contraintes et/ou les directives auxquelles est soumis l'acteur humain et qui pourraient limiter sa capacité d'action, quelles sont les autorisations d'accès à l'information dont dispose l'acteur humain) ;
- Il est **réaliste** (il tient compte des informations présentes au sein de l'organisation et de leur support matériel, des outils et techniques de recherche de ces informations ainsi que de l'état des connaissances de l'acteur humain qui est à l'origine de la recherche d'information) ;
- Il est **contextualisé** (chaque information ayant une durée de vie et une pertinence qui lui sont propres, toute recherche d'information doit s'inscrire dans un délai de réalisation précis) ;
- Il est **testable** (il doit comporter des étapes successives) ;
- Il est **écologique** (pas d'impact sur le système d'information de l'organisation) ;
- Il a un **sens** = méta objectif (apport pour l'acteur humain).

i) Le mécanisme de contournement du mental conscient

Afin de bien comprendre quel est l'apport de notre approche, il est important de préciser que nous vivons dans un monde que nous appelons habituellement la réalité, mais que ce monde n'est en réalité qu'une traduction par notre cerveau des perceptions binaires biochimiques perçues par nos cinq sens [CERI, 2007].

Nous n'avons pas tous la même perception de cet univers puisque c'est une traduction personnelle. Nous avons des codes communs pour communiquer, mais pourtant, nous ne voyons pas la même chose, nous n'entendons pas la même chose, nous ne ressentons pas la même chose, nous ne sentons pas la même chose, nous n'avons pas la même perception des saveurs [BANDLER R. et GRINDER J., 2005].

Dans chaque situation de vie, ce n'est qu'une perception individuelle que nous allons partager avec l'autre. Alors, si nous nous arrêtons uniquement à la perception de nos cinq sens, nous nous limitons dans le spectre de toutes les perceptions possibles de l'univers.

Si nous ouvrons un peu notre esprit, cela va nous permettre d'accéder également à d'autres dimensions qui ne sont pas facilement explicables.

En réalité, tout ce que nous ressentons, ce que nous percevons, ne sont que des échanges biochimiques électriques entre des cellules neuronales.

Si nous faisons un parallèle entre notre cerveau et un ordinateur (utilisation du langage métaphorique), nous pouvons dire que nous créons sans cesse l'équivalent de documents informatiques que nous stockons dans des fichiers, puis dans des dossiers, et enfin dans le disque dur (le cerveau limbique) [CHERICI C., DUPONT JC., 2008].

Notre cerveau comportant trois niveaux : le reptilien, le limbique et le néocortex, on peut dire que le cerveau limbique est semblable au disque dur de l'ordinateur car c'est lui qui emmagasine les informations de notre existence. Il stocke notre "vécu", support de nos "ressentis", eux-mêmes nourris par notre système émotionnel.

A ce stade, il convient de distinguer ce qui est "émotionnel" de ce qui est "ressenti", car ce qui est important, c'est le ressenti, alimenté par les émotions. Les émotions sont les carburants des ressentis. Le "vécu" n'est pas important, bien qu'il porte le ressenti.

Notre cerveau fabrique au travers de ce "vécu", de ce "ressenti", une étonnante équation mathématique qu'il stocke dans les cellules neuronales.

Si l'on admet que le mental d'un être humain est composé de deux éléments : d'abord du conscient et de tout ce qu'il contient mais aussi d'un arrière-pays infiniment plus vaste nommé

l'inconscient; alors il est possible de définir le conscient comme la représentation de la manifestation de notre mental et de notre intellect permettant la conceptualisation du monde.

Tout ce qui est intellectuel et mental participe de notre mécanisme conscient. Au point que si nous ne nous intéressons qu'aux méandres de notre conscient et que nous recherchons des solutions uniquement avec cette partie de notre mental, certes avec l'intention de faire remonter de l'information de l'inconscient vers le conscient, il se peut que cette information ne remonte pas à la surface.

A ce moment-là, notre cerveau qui ne reconnaît que ce qu'il connaît, continue à tourner en rond prisonnier de son ancien schéma. Le but ici est de chercher cette information pour une mise en conscience de l'origine de la problématique afin d'y apporter une solution consciente.

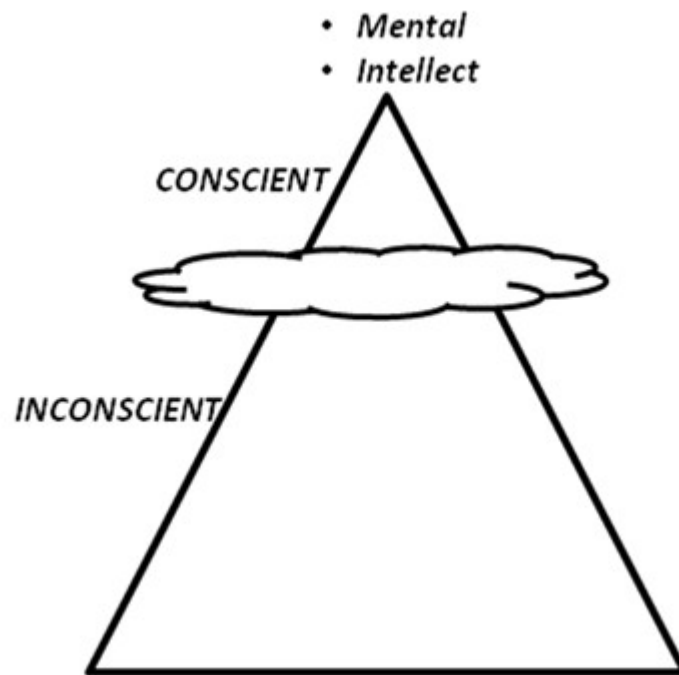
Pour caractériser ce qu'est l'inconscient, on peut dire que l'inconscient représente tout ce à quoi je n'ai pas accès de manière consciente.

Au niveau de notre inconscient, tous les instants sont inscrits dans le même instant. En physique quantique et dans d'autres travaux où nous retrouvons le même concept, nous pouvons percevoir que le passé, le présent et le futur sont inscrits dans le même espace-temps.

Notre cerveau sait tout cela, même si nous n'en sommes pas conscients intellectuellement, et il stocke ces informations dans des chaînes de neurones. Ces données sont entreposées dans notre mémoire cellulaire. Le but est alors d'accéder à cette mémoire, de façon à pouvoir doper la compréhension de certains phénomènes et problèmes auxquels nous sommes confrontés en contournant notre mental conscient pour aller rechercher des informations dans notre inconscient.

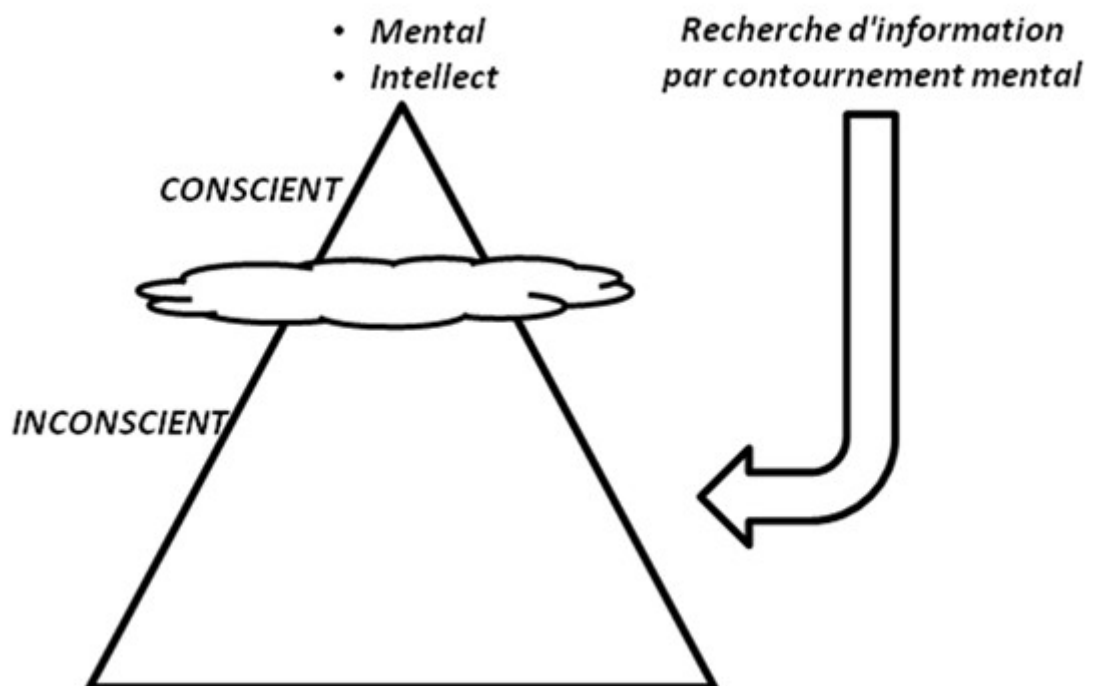
Nous avons l'habitude de représenter le conscient et l'inconscient sous cette forme.

Figure 27 : Représentation du conscient et de l'inconscient



La représentation suivante enrichie la représentation précédente en illustrant le mécanisme de contournement de notre mental conscient.

Figure 28 : Représentation du mécanisme de contournement du mental conscient



j) La sérendipité et la recherche d'information

La "sérendipité" ou l'art de trouver la bonne information par hasard, vient du mot anglais "serendipity" qui a été forgé en 1754 par le philosophe anglais Sir Horace WALPOLE à partir du conte oriental narrant le voyage des trois princes de Serendip (nom médiéval du Sri Lanka).

Ce conte s'inspire d'un des textes d'Amir KHUSRAU, grand poète persan qui vécut à Delhi il y a sept siècles et fait référence aux voyages et aventures des trois princes de Serendip qui après avoir refusé de succéder à leur père furent bannis du royaume et se mirent à voyager à pied à travers le vaste monde afin de découvrir monts et merveilles. Lors de leur périple, ils vécurent de nombreuses mésaventures dont ils se tiraient toujours d'affaire grâce à leur exceptionnel talent pour remarquer, observer et déduire [REMER T.G., 1965].

Le terme de "sérendipité" apparaît dans le domaine scientifique et se conceptualise avec Robert King MERTON [MERTON R.K., 1945] qui le définit comme *"la découverte par chance ou sagacité de résultats que l'on ne cherchait pas"*.

Si l'on se réfère à Jacques PERRIAULT [PERRIAULT J., 2000] qui a participé à introduire le concept de "Sérendipité" dans le champ des sciences de l'information; cet effet *"consiste à trouver par hasard et avec agilité une chose que l'on ne cherche pas. On est alors conduit à pratiquer l'inférence abductive, à construire un cadre théorique qui englobe grâce à un bricolage approprié des informations jusqu'alors disparates"*.

Dans le prolongement des travaux de Jacques PERRIAULT et dans l'optique de comprendre l'effet "Sérendipité", Olivier ERTZSCHEID et Gabriel GALLEZOT [ERTZSCHEID O. et GALLEZOT G., 2003] avancent l'idée que *"dans un effort de maîtrise de l'entropie informationnelle (telle que nous la connaissons actuellement avec le développement d'Internet), l'essor des technologies intellectuelles de représentation et d'accès aux informations fait chaque jour une place plus grande au phénomène de sérendipité comme adjuvant précieux de la recherche d'information"*.

Olivier ERTZSCHEID et Gabriel GALLEZOT précisent également que dans le cadre d'un processus de recherche d'information, "la sérendipité peut être considérée comme un phénomène passager (le temps que des modèles mentaux soient appelés) ou devenir un mode privilégié d'accès à l'information dans le cadre d'un processus de recherche ou de l'une de ses itérations".

Notre approche en matière de recherche d'information s'inscrit dans cette logique, au sens elle repose sur une stratégie de navigation qui vise à prendre en considération le fonctionnement caractéristique de l'esprit humain et à concevoir un système d'aide à la recherche d'information qui fonctionne par association et non plus selon le modèle classique de l'indexation.

Si l'on tient compte des pratiques et usages des Internautes en quête d'information [ERTZSCHEID O. et AL., 2009], on peut constater que l'effet sérendipité se manifeste quand l'acteur humain prend conscience qu'il manque de connaissances, notamment de méta-connaissances (c'est-à-dire des connaissances sur le thème, des compétences documentaires, des savoir-faire, etc.) et qu'il arrive à contourner cette limitation en allant interroger une autre partie de son mental, à savoir : l'inconscient [BADDELEY A., 1999].

Pour bien comprendre à quels mécanismes nous faisons appel, il est nécessaire de préciser que la connaissance résulte de processus cognitifs (perception, apprentissage, raisonnement, mémoire, expérience, etc.) et qu'il est possible d'acquérir de la connaissance sans avoir conscience de la connaissance acquise (connaissance implicite) [SCHLICK M., 2009]. Car, à côté de la connaissance dont nous sommes conscients, il existe une forme de connaissance inconsciente et donc de mémoire inconsciente qui guide nos actions et nos décisions et qu'il est possible d'adresser [JUNG C.G, 1988].

k) La notion d'ancrage

L'ancrage en PNL (Programmation Neuro-Linguistique) est un processus qui associe inconsciemment et automatiquement une réaction interne à un stimulus extérieur. Nous mémorisons ces liens et créons ainsi ce que l'on appelle des "ancres". Dès qu'une ancre est stimulée, la sensation vécue dans le passé revient instantanément. Les ancres peuvent être visuelles, auditives, kinesthésiques, olfactives ou gustatives [BANDLER R. et AL., 2000].

Le mécanisme d'ancrage constitue donc la phase de consolidation cognitive durant laquelle une représentation mentale se lie avec d'autres représentations préalablement inscrites en mémoire.

Stimuler une "ancre" lors d'une recherche d'information peut donc permettre à un acteur humain de faire appel à un mécanisme de contournement de son mental conscient et lui permettre d'améliorer ainsi sa recherche d'information.

2) Formalisation d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences

L'approche hybride que nous avons instanciée précédemment et les différents apports théoriques que nous venons d'opérer nous permettent désormais de pouvoir caractériser le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons.

Cela passe par un énoncé du principe de fonctionnement, une description du mécanisme de contournement du mental conscient (prise en compte de l'effet "sérendipité"), une présentation du processus de capitalisation d'expériences (utilisation du modèle de raisonnement à l'aide de cas) et des facteurs clés de succès.

a) Principe de fonctionnement

Lorsqu'un acteur humain dans une organisation recherche une information dans un système de Gestion Electronique de Documents hétérogènes, il procède souvent par itérations successives.

Si la première action de recherche ne lui apporte pas satisfaction, il procède alors à une deuxième action de recherche et ainsi de suite jusqu'à ce qu'il "ressente" une totale satisfaction.

Le fait que la première action de recherche ne le satisfasse pas totalement peut être imputable à plusieurs facteurs :

- Le premier facteur est que l'objectif de recherche qu'il a formulé n'en est véritablement pas un (pas assez précis par exemple) ;
- Le second facteur provient souvent du fait que l'indexation des informations qu'il recherche (documents et données) n'a pas été opérée sur les mêmes critères que ceux adoptés pour rechercher l'information ;
- Le troisième facteur vient souvent de la mauvaise utilisation des outils et de la non maîtrise des techniques de recherche ;

- Le quatrième facteur provient du fait que l'acteur humain ne possède pas toujours le niveau de connaissances (connaissance limitée du domaine fonctionnel par exemple) nécessaire pour rechercher efficacement l'information ;
- Enfin, il se peut que l'information recherchée ne soit tout simplement pas présente au sein du système de GED.

Lors des actions de recherche suivantes, l'acteur humain cherche alors à comprendre comment faire évoluer son objectif de recherche en essayant de faire varier certains critères de cet objectif notamment ceux qui servent à caractériser l'information ou en utilisant d'autres outils et techniques de recherche.

Il procède alors par identification des similitudes et des différences par rapport à l'action de recherche précédente mais en s'appuyant uniquement sur sa rationalité normée (mental conscient). Seul le recours à des mécanismes comme l'effet "sérendipité" et la prise en compte des conditions qui font, qu'un acteur humain fasse appel à son inconscient pour améliorer son objectif de recherche d'information par le biais de la capitalisation des mécanismes de contournement de son mental conscient, peut constituer une hypothèse de réponse quand à la manière d'améliorer considérablement la recherche d'information dans un système de Gestion Electronique de Documents hétérogènes.

C'est cette hypothèse que nous allons à présent développer et modéliser dans le cadre de notre approche.

Dans un premier temps, nous allons nous intéresser à l'identification des différents acteurs ainsi qu'à la détermination de leurs profils.

Le recours à la méthode "Motivations", dont nous avons présenté les fondements théoriques précédemment, nous permet d'analyser le rôle de chaque acteur ainsi que son métier au sein de l'organisation. Elle nous permet également de mettre en évidence la liaison qui relie le système de GED utilisé avec un acteur intervenant dans l'organisation.

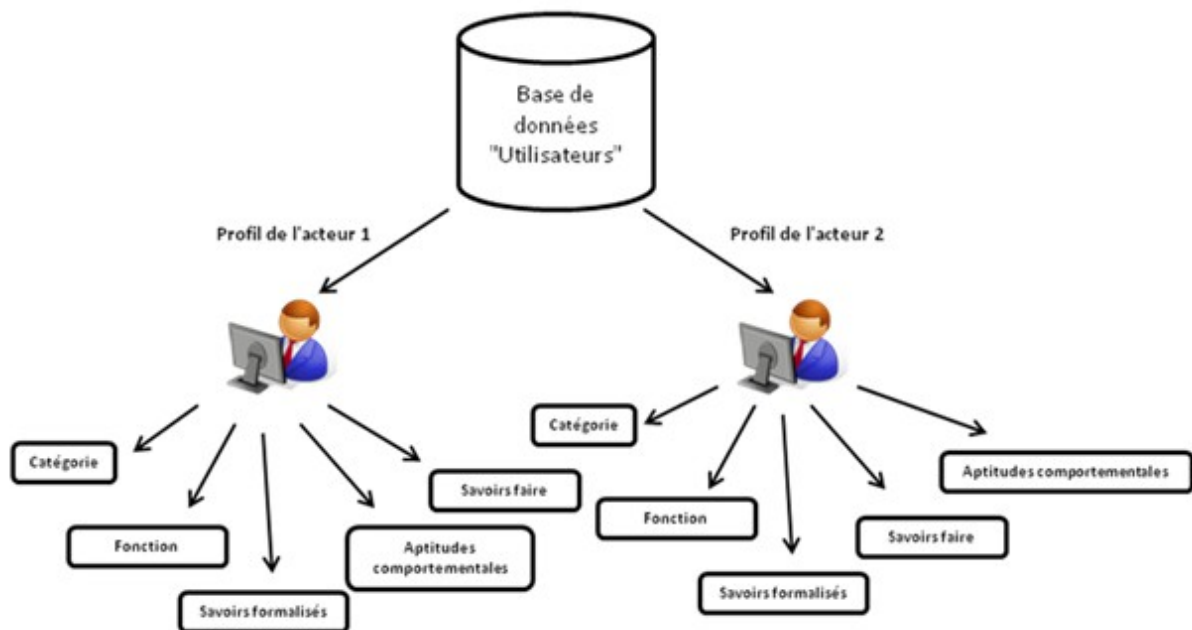
Outre le fait de nous aider à identifier tous les acteurs de l'organisation utilisant la GED, cette méthode (à travers la vision locale qu'est le "Motif") nous permet de comprendre le métier de l'utilisateur, de recenser les flux d'information qui lui sont transmis ou événements entrants, les flux d'information qu'il émet ou événements sortants, les objectifs de sa mission, les règles et contraintes dont il doit tenir compte et enfin ses traitements, c'est-à-dire les actions qu'il effectue.

En fonction des quatre grandes catégories d'acteurs que nous avons déterminées (producteurs, consommateurs et valideurs d'information mais aussi gestionnaire du système de GED), il est possible de construire quatre "Motifs" standards à personnaliser en fonction de chaque acteur/utilisateur du système de GED (cf. figure 24 : Interactions entre un acteur humain et une application de GED).

La construction de la "Fresque" à partir des "Motifs" qui auront été réalisés nous apportera une meilleure compréhension de l'organisation, des flux d'information qui la caractérisent et des interactions entre ses différents acteurs.

Partant de là, il nous est alors possible de déterminer le profil de chaque acteur/utilisateur du système de GED et de constituer une base de données "Utilisateurs" dont la vocation est de stocker les données relatives à chacun des profils d'acteurs.

Figure 29 : Base de données "Utilisateurs"



L'utilisateur d'un système de GED devant généralement s'authentifier par le biais d'un identifiant et d'un mot de passe pour pouvoir accéder aux différentes fonctionnalités d'un système de GED (prise en compte de la sécurité), il dispose alors d'une exploitation personnalisée des différentes stratégies d'indexation et de recherche d'information.

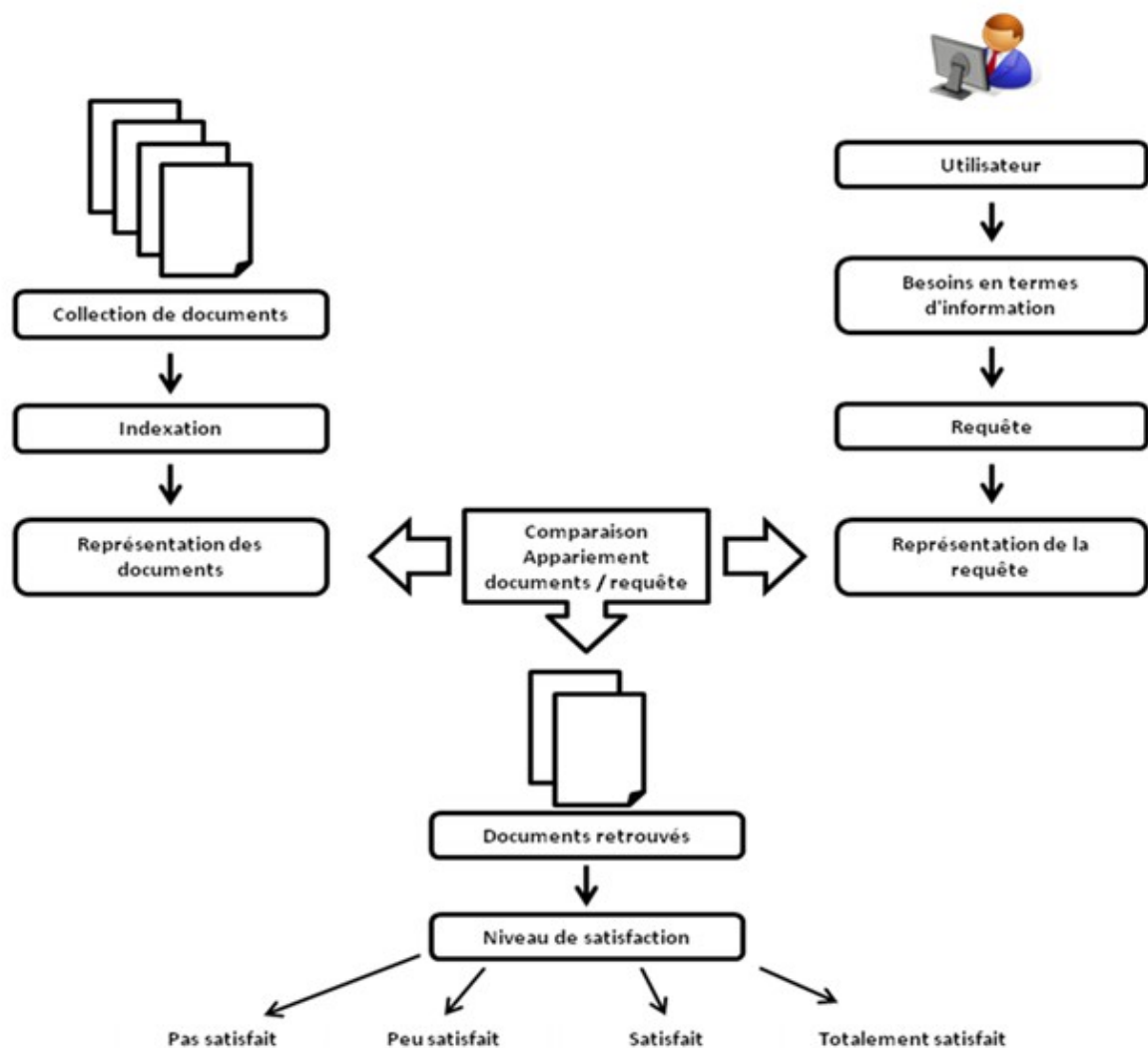
En fonction des autorisations accordées à l'utilisateur (et qui sont liées à son profil), l'acteur/utilisateur peut alors procéder à une première recherche d'information.

Sachant que plusieurs techniques de recherche d'information existent et sont mises en œuvre dans les systèmes de GED, soit indépendamment et exclusivement les unes des autres, soit pour certaines d'entre elles de façon combinée, l'acteur/utilisateur devra alors choisir celle qu'il juge la plus opportune pour interroger le fonds documentaire du système de GED.

A ce stade, il est important de noter que selon le profil de l'utilisateur, le système de GED peut procéder à du filtrage d'information.

Au terme de cette interrogation et lorsque l'utilisateur quitte la fonction de recherche, il est alors amené à se prononcer sur son niveau de satisfaction (pas satisfait, peu satisfait, satisfait, totalement satisfait).

Figure 30 : Expression du niveau de satisfaction



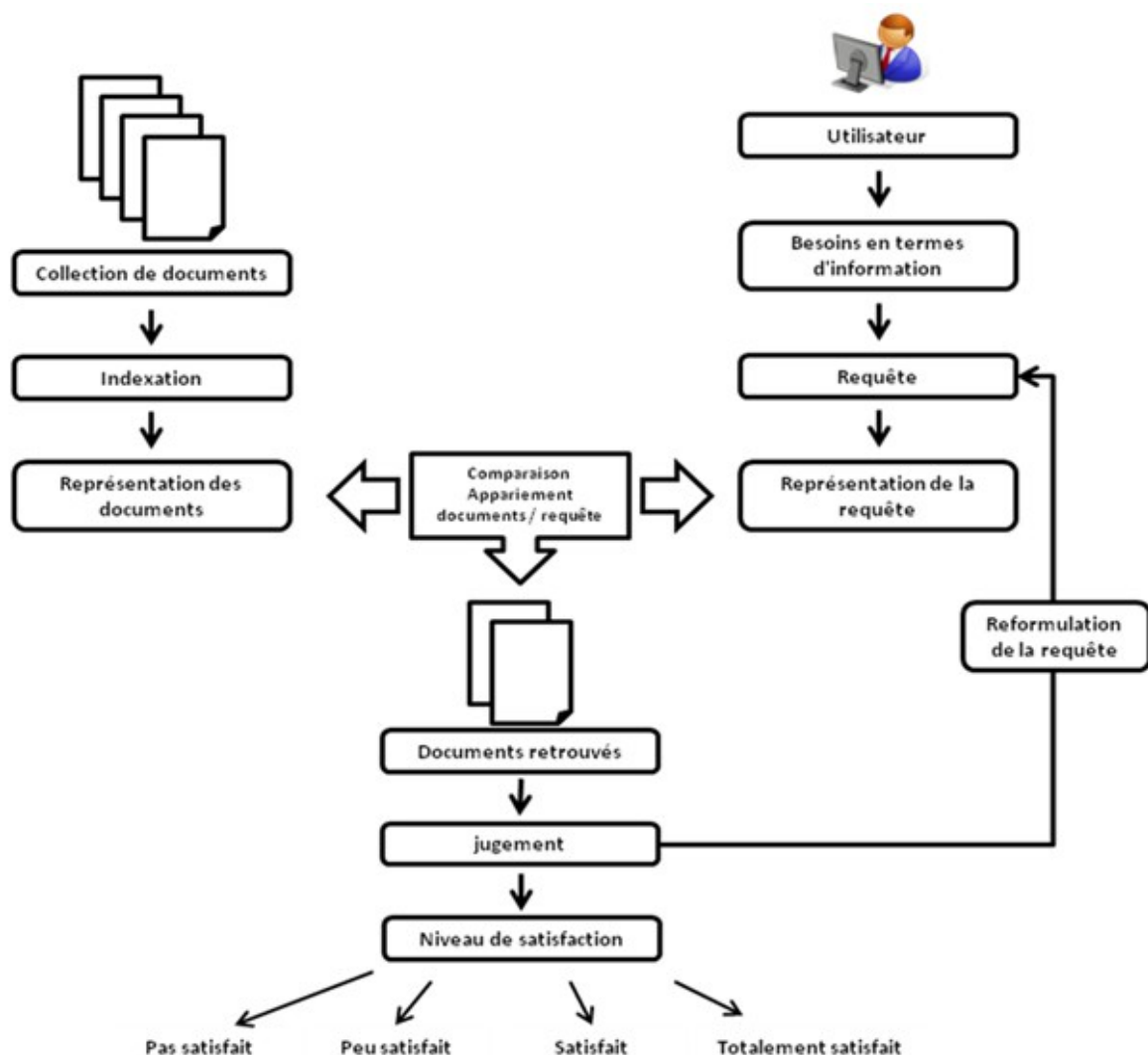
Si la première action de recherche ne l'a pas satisfait et qu'il désire poursuivre sa recherche d'information (pour répondre à l'objectif qu'il s'est fixé), l'acteur/utilisateur peut alors engager

une deuxième action de recherche et ainsi de suite jusqu'à ce qu'il "ressente" une totale satisfaction.

Avant d'engager une autre action de recherche, l'acteur/utilisateur est toutefois amené à se prononcer sur l'action de recherche précédente en précisant son niveau de satisfaction.

Lors des actions de recherche suivantes, l'acteur humain cherche à faire évoluer son objectif de recherche en essayant de faire varier certains critères de cet objectif notamment ceux qui servent à caractériser l'information ou en utilisant d'autres outils et techniques de recherche (sans pour autant quitter la fonction de recherche d'information).

Figure 31 : Nouveau processus de recherche d'information

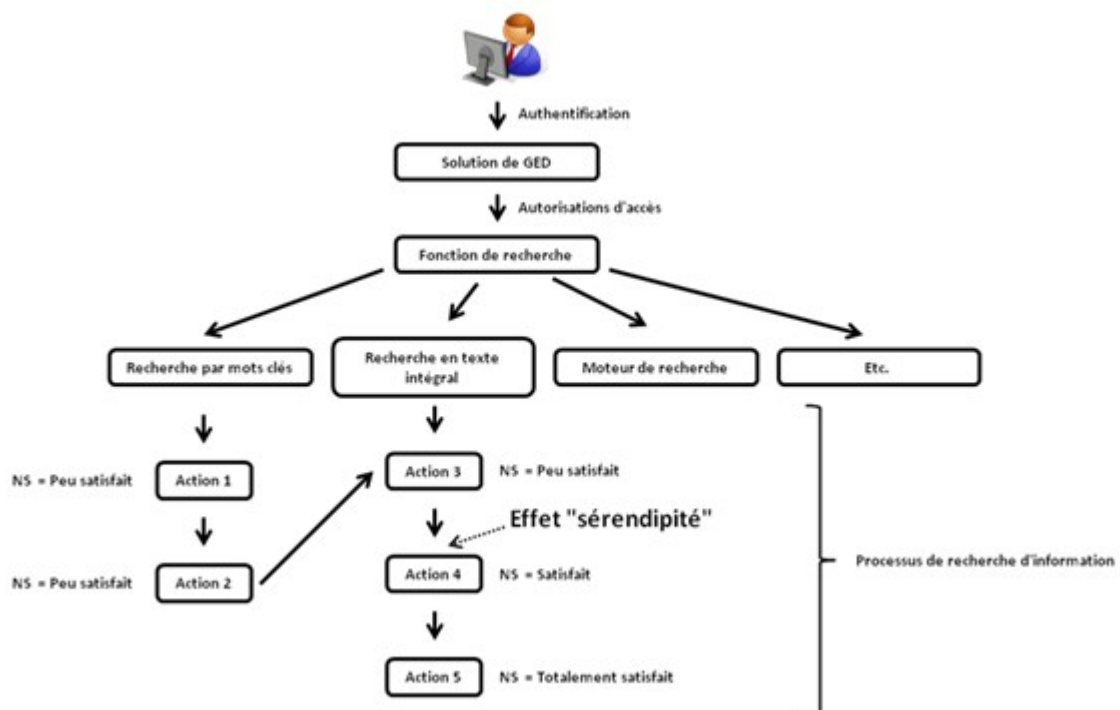


Lorsque l'utilisateur, après plusieurs actions de recherche, ne parvient pas à élever son niveau de satisfaction alors le mécanisme de contournement de son mental conscient (prise en compte de l'effet sérendipité) peut opérer.

Il est important de souligner que la sensibilité du dispositif de contournement du mental conscient dépend du nombre d'actions de recherche permis sans élévation du niveau de satisfaction.

Cette prise en compte de l'effet "sérendipité" se manifeste alors sous la forme d'une "ancree" visuelle apparaissant sur une partie de l'interface utilisateur (d'où le rôle important joué par la conception d'IHM) et qui permet à l'acteur/utilisateur qui utilise une des techniques de recherche d'information intégrées au système de GED de voir apparaître les interrogations qu'il a pu effectuer dans le passé et qui lui ont permis d'élever son niveau de satisfaction.

Figure 32 : Schéma global de fonctionnement



Il est important de noter que le nombre d'expériences positives visualisées par le biais de l'ancree visuelle est lui aussi paramétrable.

Par ailleurs, il convient également de souligner que l'ensemble des actions réalisées par l'acteur/utilisateur dans le cadre du système d'aide à la recherche d'information que nous proposons sont assistées ou supervisées (l'utilisateur est guidé dans la phase d'évaluation et d'interrogation du fonds documentaire et le mécanisme de contournement du mental conscient comme le processus de capitalisation d'expériences est entièrement piloté par le système d'aide à la recherche d'information).

La démarche assistée et supervisée que nous avons adoptée s'appuie sur un processus de capitalisation d'expériences de recherche d'information, un mécanisme de contournement du mental conscient et un dispositif d'évaluation du niveau de satisfaction de l'acteur/utilisateur.

Elle permet ainsi d'améliorer la performance des systèmes de recherche d'information existants en s'adaptant au profil et à l'objectif de recherche d'information de chaque acteur humain.

Pour bien comprendre comment fonctionne le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons, nous allons à présent détailler le mécanisme de contournement du mental conscient tel que l'avons conçu et implémenté.

b) Le mécanisme de contournement du mental conscient (prise en compte de l'effet "sérendipité")

Selon Elaine G.TOMS [TOMS E.G., 2000], il convient de distinguer le raisonnement par analogie (favorisant la sérendipité) de ce qu'elle nomme "blind luck" où seul le hasard est à l'origine d'une découverte informationnelle.

Dans le cadre de notre approche de contournement du mental de conscient à l'aide d'un mécanisme d'ancrage, nous fonctionnons par analogie dans la mesure où nous nous servons de situations antérieures inconsciemment mémorisées pour en comprendre de nouvelles.

Nous fondons notre approche sur le fait que dans la majorité des processus de recherche d'information, l'utilisateur sait déjà, ou en partie, ce qu'il cherche même si l'objet informationnel qu'il recherche n'est pas totalement défini.

Cela fait d'ailleurs référence à la célèbre phrase de Louis PASTEUR sur le fait que *"le hasard ne favorise que les esprits préparés"*, prononcée en 1854 lors de son discours de nomination comme professeur et doyen de la Nouvelle Faculté des Sciences de Lille.

Partant de là, l'acteur/utilisateur va alors interroger le système de GED par le biais d'une série d'actions de recherche utilisant des modèles d'interrogations classiques (recherche par mots clés, par concepts, en texte intégral, etc.). L'acteur/utilisateur est dans une logique de consultation et cherche simplement à accroître son niveau de satisfaction en évaluant la pertinence des résultats obtenus. L'acteur/utilisateur est alors dans un mode de raisonnement de type hypothético-déductif qui peut très vite trouver ses limites (rationalité normée).

Cela se mesure par un niveau de satisfaction constant ou en baisse après différentes actions de recherche où l'acteur/humain a modifié certains critères servant à caractériser l'information ou utilisé d'autres outils et techniques de recherche.

A ce stade, seule une exploration d'une partie de son inconscient peut permettre à l'acteur humain d'aboutir dans sa recherche d'information en lui fournissant l'occasion de prendre conscience de certaines situations qu'il a déjà vécues et mémorisées inconsciemment.

Par un stimulus extérieur, l'acteur/utilisateur accède à des représentations préalablement inscrites en mémoire et qui vont l'aider à faire évoluer son interrogation.

Ce stimulus se manifeste sous la forme d'une zone spécifique de l'interface graphique où figure l'ensemble des requêtes précédemment saisies qui ont permis une augmentation significative du niveau de satisfaction. Cette zone apparaît après un certain nombre d'actions de recherche n'ayant pas abouti à accroître le niveau de satisfaction de l'acteur/utilisateur.

Suivant l'outil et la technique de recherche d'information utilisés, la structuration de cette zone est très différente. La manière de présenter les données relatives aux expériences antérieures et de mettre en évidence les éléments qui ont permis un accroissement du niveau de satisfaction dépend fortement de la nature de l'application de GED et des acteurs/utilisateurs de l'organisation.

Elle se manifeste également lorsque l'acteur/utilisateur commence à saisir sa requête dans le ou les champs de recherche et que s'affichent automatiquement des suggestions pertinentes de termes dans la ou les zones de saisie. Ce phénomène s'accompagne également d'une information complémentaire relative au nombre de résultats qui apparaîtront si l'acteur/utilisateur sélectionne les termes de recherche qui lui sont proposés.

Afin d'illustrer notre propos, prenons l'exemple d'une solution de GED documentaire mono application disposant d'une fonction de recherche en texte intégral et dont l'interface graphique se limite à un seul champ de saisie et un bouton de recherche.

L'acteur/utilisateur qui s'engage dans un processus de recherche identifie au préalable un certain nombre de termes qui, selon lui, caractérisent au mieux l'information qu'il recherche. Après avoir saisi ces termes dans la zone de recherche appropriée, il opère une première action de recherche.

Figure 33 : Première action de recherche

Le diagramme illustre l'interface de recherche. À gauche, un rectangle horizontal contient le texte "Terme 1 Terme 2 Terme 3". À droite, un bouton rectangulaire est étiqueté "RECHERCHER".

Suite à cette interrogation, une liste de documents correspondants aux critères de recherche s'affiche sous la zone de recherche. L'acteur/utilisateur peut alors évaluer la pertinence de chacun des documents sélectionnés par l'algorithme de recherche et se prononcer sur son niveau de satisfaction.

S'il n'est pas satisfait du résultat de sa première action de recherche, il procédera alors à une deuxième action de recherche en faisant varier les critères servant à caractériser l'information recherchée.

Figure 34 : Deuxième action de recherche

Terme 1 Terme 3 Terme 4 Terme 5

RECHERCHER

Au bout de N actions de recherche ne permettant pas à l'acteur/utilisateur d'être au minimum "Satisfait" ou "Totalelement satisfait", l'effet sérendipité se manifeste alors au cours de la $N+1^{i\text{ème}}$ action de recherche sous la forme d'une zone graphique où figurent des expériences de recherche antérieures proches de celle effectuée et ayant permis d'élever de façon significative le niveau de satisfaction de l'acteur/utilisateur.

Figure 35 : Illustration de l'effet sérendipité : $N+1^{i\text{ème}}$ action de recherche

Terme 1 Terme 4 Terme 10 Terme 11

RECHERCHER

Requête X = {Terme 1, Terme 4, Terme 7, Terme 10}	Nbre de documents : 120
Requête X+1 = {Terme 1, Terme 4, Terme 10, Terme 14}	Nbre de documents : 30
Niveau de satisfaction = Peu satisfait → Totalelement satisfait	
Requête y = {Terme 4, Terme 10, Terme 22, Terme 1}	Nbre de documents : 395
Requête y+1 = {Terme 1, Terme 4, Terme 19}	Nbre de documents : 23
Niveau de satisfaction = Peu satisfait → Totalelement satisfait	

Supposons à présent que l'acteur/utilisateur désire procéder à une $N+2^{i\text{ème}}$ action de recherche et qu'il saisisse le premier terme d'une nouvelle requête; une information complémentaire peut lui être fournie par le biais d'un affichage sous la zone de recherche. Cet affichage se présente alors sous la forme d'une liste de requêtes que l'acteur/utilisateur a déjà formulées dans le cadre d'expériences de recherche antérieures.

Figure 36 : $N+2^{i\text{ème}}$ action de recherche

Terme 1 Terme 4 ...		RECHERCHER
Requête X+1 = {Terme 1, Terme 4, Terme 10, Terme 14}	Nbre de documents : 30	
...	...	
Requête y = {Terme 1, Terme 4, Terme 19}	Nbre de documents : 23	
...	...	
...	...	

Au terme du processus de recherche et après avoir atteint un haut niveau de satisfaction, l'acteur/utilisateur quitte la fonction de recherche et/ou l'application de GED non sans avoir signalé au système d'aide à la recherche d'information son niveau de satisfaction (comme il l'aura fait d'ailleurs après chaque action de recherche).

En termes d'assistance supplémentaire aux acteurs/utilisateurs d'une solution de GED intégrant ce type de système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences, il est également possible de signaler la nature (documents textuels, visuels, audiovisuels, sonores, etc.) et le type de fichiers (par le biais de leur extension) qui correspondent aux nombre de documents résultant d'une requête.

Figure 37 : Illustration de l'effet sérendipité : Informations complémentaires

Terme 1 Terme 4 Terme 10 Terme 11		RECHERCHER
Requête X = {Terme 1, Terme 4, Terme 7, Terme 10}	Nbre de documents : 120	
Requête X+1 = {Terme 1, Terme 4, Terme 10, Terme 11}	Nbre de documents : 30	
Niveau de satisfaction = Peu satisfait	Nombre de documents : 120 + Documents textuels → 120 [107 : PDF – 13 : DOC/DOCX]	
Requête y = {Terme 1, Terme 4, Terme 10, Terme 11}	Nbre de documents : 395	
Requête y+1 = {Terme 1, Terme 4, Terme 10, Terme 11, Terme 12}	Nbre de documents : 23	
Niveau de satisfaction = Peu satisfait → Totalemtent satisfait		

De la même manière, et dans le cas de certaines applications de GED, il est possible d'intégrer une zone de "Nouveautés" ou "News" qui au moment de la mise en œuvre de l'effet sérendipité se manifeste sous la forme d'une fenêtre indiquant les documents qui ont été indexés dans les X dernières heures (X étant un paramètre du système à déterminer en fonction de l'usage et des pratiques des différents acteurs/utilisateurs en matière de recherche d'information) et qui contiennent un terme particulier de la requête. Cette aide peut être déclinée en prenant également en compte plusieurs intervalles de temps (par exemple les dernières 24 heures, 48 heures, etc.).

Figure 38 : Illustration de l'effet sérendipité : Zone de nouveautés

Terme 1 Terme 4 Terme 10 Terme 11

RECHERCHER

Requête X = {Terme 1, Terme 4, Terme 7, Terme 10}

Requête X+1 = {Terme 1, Terme 4, Terme 10, Terme 14}

Niveau de satisfaction = Peu satisfait → Totalelement satisfait

Nbre de documents : 120

Nbre de documents : 30

Requête y = {Terme 4, Terme 10, Terme 22, Terme 1}

Requête y+1 = {Terme 1, Terme 4, Terme 19}

Niveau de satisfaction = Peu satisfait → Totalelement satisfait

Nbre de documents : 395

Nbre de documents : 23

Nouveaux Documents

+ Il y a 24 heures

12 Documents contenant le terme *Terme 1* → [Lien 1](#)

1 Document contenant le terme *Terme 4* → [Lien 2](#)

+ Il y a 48 heures

3 Documents contenant le terme *Terme 1* → [Lien 3](#)

1 Document contenant le terme *Terme 4* → [Lien 4](#)

5 Documents contenant le terme *Terme 11* → [Lien 5](#)

Cliquez sur le lien approprié pour accéder à la liste des documents

Après avoir illustrer le mécanisme de contournement du mental conscient (prise en compte de l'effet "sérendipité"), il nous apparaît important de revenir sur le processus de capitalisation des expériences de recherche d'information qui est un point essentiel du système de recherche d'information par sérendipité et capitalisation d'expériences.

Ce mécanisme utilise le modèle de raisonnement à partir de cas dont nous allons à présent expliquer le mode de fonctionnement.

c) Le processus de capitalisation d'expériences (utilisation du modèle de raisonnement à partir de cas)

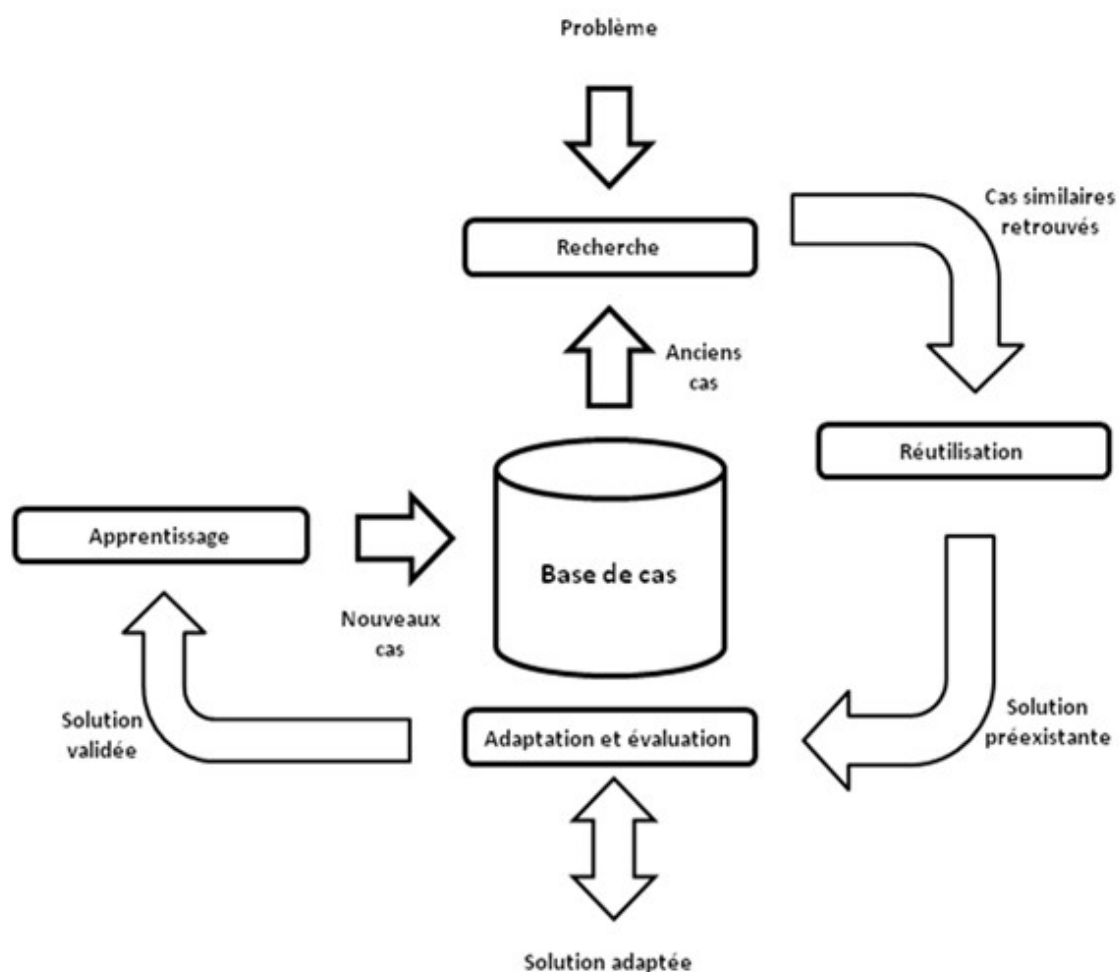
Le modèle de raisonnement à partir de cas :

Il s'agit essentiellement d'une technique qui se fonde sur des situations antérieures pour en traiter des nouvelles. Le Raisonnement par Cas (RpC) ou Case Based Reasoning (CBR), est un "concept puissant qui fournit un mode de raisonnement analogique pour résoudre des problèmes" [AAMODT A. et PLAZA E., 1994]. Cette technique de raisonnement consiste à capitaliser les expériences antérieures à une situation donnée et ainsi générer des

connaissances liées à ces dernières. C'est une approche qui permet d'exprimer la connaissance liée à l'expérience et au savoir-faire. Cette connaissance sera utilisée pour traiter les nouveaux cas jugés similaires.

Le principe général du RpC consiste à traiter un nouveau problème (cas cible) en se remémorant des expériences passées voisines (cas de référence). Ce type de raisonnement repose sur l'hypothèse suivante : si une expérience passée et la nouvelle situation sont suffisamment similaires, alors tout ce qui peut être expliqué ou appliqué à l'expérience passée (base de cas) reste valide si on l'applique à la nouvelle situation qui représente le nouveau problème à résoudre. D'un point de vue très global, le RpC met en œuvre une base d'expériences ou de cas, un mécanisme de recherche et d'extraction des cas similaires et un mécanisme d'adaptation et d'évaluation des solutions des cas extraits pour résoudre le problème spécifié.

Figure 39 : Modèle de raisonnement à partir de cas



Le RpC étant fondé sur la comparaison des nouveaux cas avec des cas indexés et précédemment stockés, la logique du RpC consiste donc à rechercher ceux qui sont similaires et à proposer d'appliquer les éléments de décision et les actions correspondantes en espérant que *"ce qui sera efficace une fois le sera plusieurs fois"* [GUPTA U.G., 1994].

Cette façon d'intégrer l'expérience pour le traitement de cas similaires est, selon le domaine traité, d'un apport important, citons le cas de la médecine, de la documentation et de la recherche d'information sur Internet.

L'approche de RpC est fondamentalement constituée de quatre grandes étapes :

- la représentation et l'indexation de nouveaux cas ;
- la recherche de cas similaires ;
- La réutilisation ou l'adaptation des cas similaires trouvés ;
- L'apprentissage et la mémorisation de nouveaux cas.

L'étape de représentation et d'indexation des nouveaux cas occupe une place importante dans l'approche de RpC. En effet, cette première étape va déterminer l'efficacité et la rapidité de la recherche de cas similaires dans la base de cas. Pour cela, il est nécessaire de bien choisir les informations à stocker pour caractériser chaque nouveau cas et définir des index qui serviront également à la recherche de cas analogues.

La recherche de cas similaires à un nouveau cas utilise les index et les règles d'appariement qui auront été définies pour tenter de retrouver un cas antérieur similaire au problème courant.

Dans certaines approches de RpC, lorsque l'on a retrouvé un cas similaire, on réutilise directement ce cas pour résoudre le problème courant. Dans ce type d'approche, on considère que les similarités sont suffisantes et que l'on peut négliger les différences entre le cas trouvé et le problème courant.

Toutefois, il est rare que l'on trouve un cas identique au problème, il est alors souvent nécessaire d'adapter les solutions préexistantes. L'adaptation consiste donc à construire une nouvelle solution à partir du problème courant et des cas similaires trouvés. Cette phase met l'accent sur les différences entre les cas trouvés et le problème et sur l'information utile à transférer à la nouvelle solution.

La solution adaptée est testée en vue de décider de son adéquation au problème courant. Si la solution proposée est un succès, on incorpore le nouveau cas (avec les caractéristiques de la solution proposée) dans la mémoire de cas (étape d'apprentissage). C'est cette étape d'apprentissage qui permet d'enrichir la base de cas. Si la solution échoue, on tente de

l'expliquer en recherchant les causes de l'échec. Puis, connaissant la ou les causes de l'échec, on corrige la solution avant de réitérer le test.

Le modèle de raisonnement à partir de cas proposé :

Dans le cadre du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons, le processus de RpC que nous avons retenu ne prend pas en considération la phase d'adaptation des cas similaires retrouvés, et propose à l'acteur/utilisateur d'exploiter directement ces cas.

Par conséquent, notre approche en matière de RpC repose essentiellement sur deux grandes étapes :

- La représentation et l'indexation de nouveaux cas ;
- La recherche de cas similaires.

Le problème à l'origine du processus de raisonnement par cas est le suivant : rechercher l'ensemble des actions de recherche précédemment réalisées par un acteur/utilisateur à l'intérieur d'une même session de recherche présentant de fortes similarités avec l'action de recherche que vient de réaliser cet acteur/utilisateur dans le cadre d'une nouvelle session de recherche et qui ont permis une augmentation significative du niveau de satisfaction.

Il est alors possible de définir un "cas" comme deux actions de recherche consécutives à l'intérieur d'une même session de recherche opérée de bout en bout par un acteur/utilisateur donné et qui, au terme de la dernière action de recherche, est au minimum "Satisfait" ou "Totalelement satisfait".

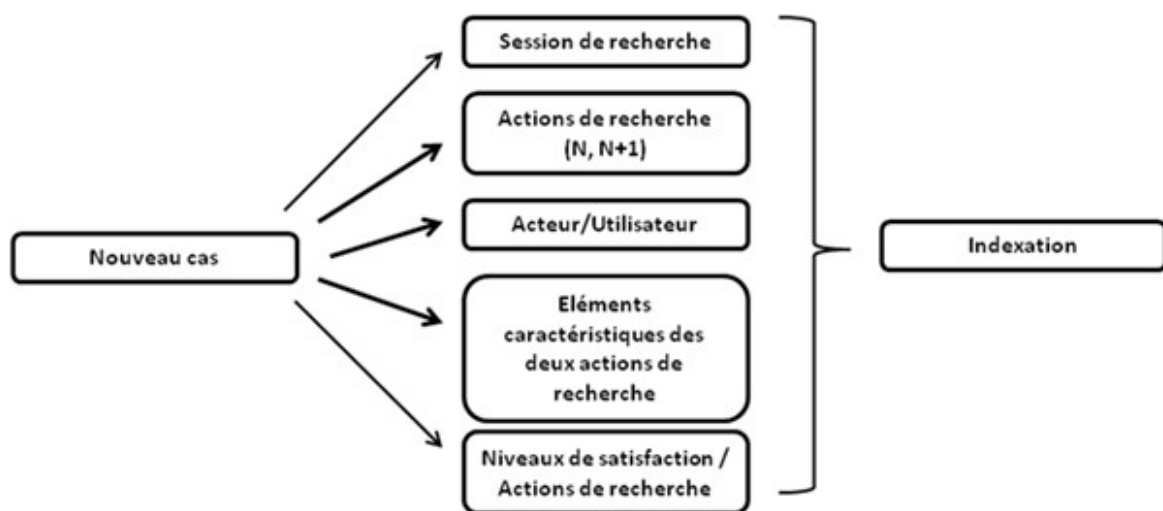
Le problème ainsi posé, il est alors possible de déterminer l'ensemble des informations nécessaires pour caractériser un nouveau cas.

Il s'agit notamment :

- D'un identifiant permettant de faire référence à une session de recherche ;
- D'un identifiant permettant de faire référence à une action de recherche "N" au sein de cette session ;
- Des éléments caractéristiques de l'action de recherche N (requête composée de termes par exemple) ;
- Du niveau de satisfaction de l'acteur/utilisateur après avoir mené l'action de recherche N ;

- D'un identifiant permettant de faire référence à l'action de recherche N+1 au sein de cette session ;
- Des éléments caractéristiques de l'action de recherche N+1 (requête composée de termes par exemple) ;
- Du niveau de satisfaction de l'acteur/utilisateur après avoir mené l'action de recherche N+1 ;
- De l'acteur/utilisateur à l'origine de cette session et de ses actions de recherche.

Figure 40 : Caractérisation d'un nouveau cas



Dans le cadre du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons, il est important de noter que le mécanisme de raisonnement par cas est composé de deux fonctions essentielles mais également complémentaires :

- Une fonction de représentation et d'indexation des nouveaux cas ;
- Une fonction de recherche des cas similaires.

La fonction de représentation et d'indexation des nouveaux cas permet d'enrichir la base de cas.

La fonction de recherche permet, quant à elle, de trouver des cas antérieurs similaires au problème courant.

La base de cas représente ainsi l'expérience capitalisée des différents acteurs/utilisateurs en matière de recherche d'information et d'interrogation du fonds documentaire d'une solution de GED.

Dans le mécanisme de raisonnement par cas que nous avons adopté, la phase de recherche de cas permet de calculer un score de similarité structurale entre les objets composites représentant les cas précédemment stockés et le nouveau cas actuellement considéré.

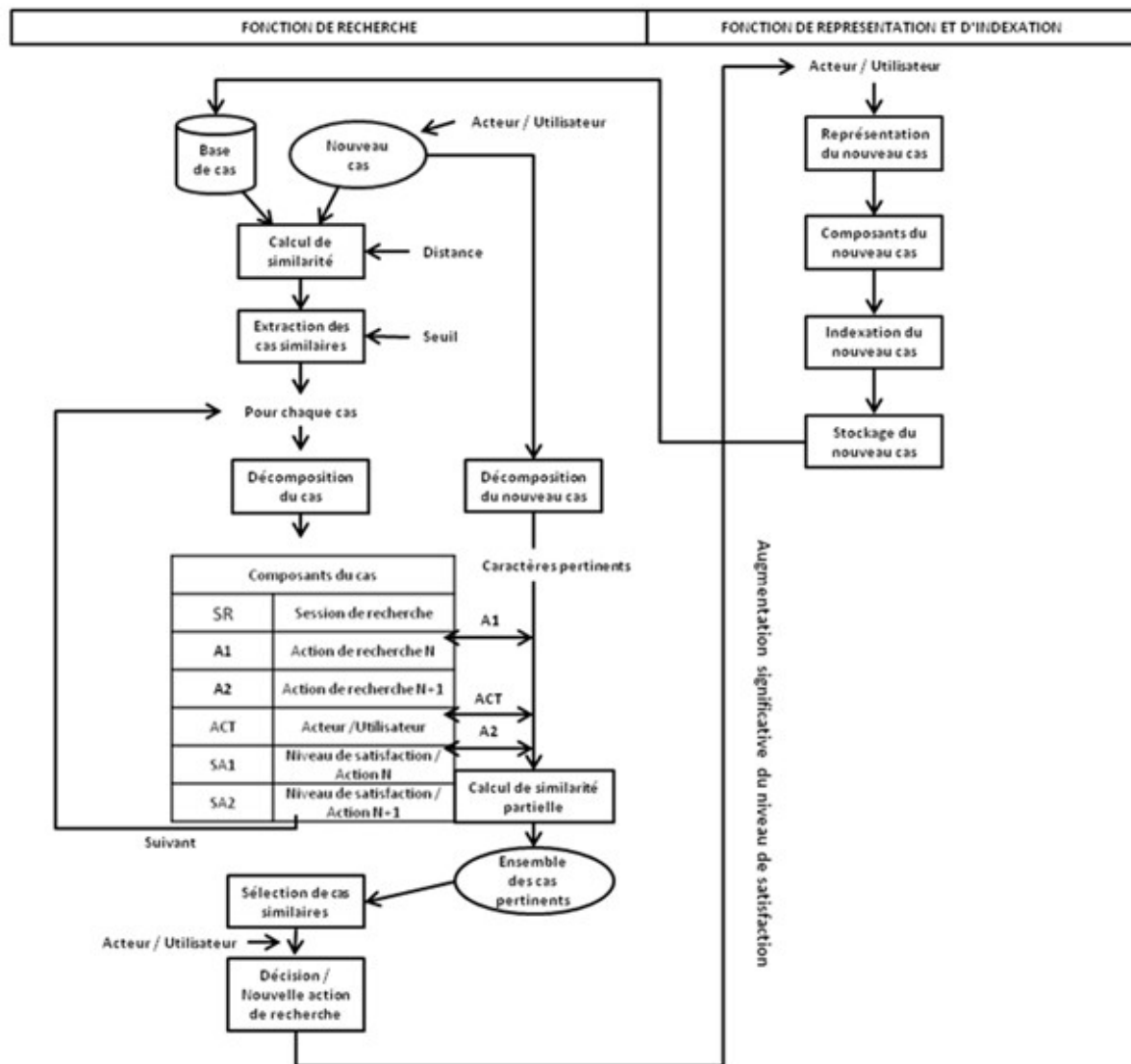
Un processus de décomposition de l'objet composite "cas" produit des sous-objets de types suivants : session de recherche (SR), action de recherche N (A1), action de recherche N+1 consécutive à l'action de recherche N (A2), acteur/utilisateur (ACT), niveau de satisfaction de l'acteur/utilisateur suite à l'action de recherche N (SA1), niveau de satisfaction de l'acteur/utilisateur suite à l'action de recherche N+1 (SA2).

Cette décomposition des objets "cas" en sous-objets permet de déterminer un ensemble de cas pertinents à l'aide d'un calcul de similarité partielle entre des sous-objets des objets composites représentant les cas précédemment stockés et des sous-objets du nouvel objet cas {A1, ACT, SA1}.

Durant la phase d'indexation, le nouvel objet cas est instancié et l'acteur/utilisateur est appelé à fournir des informations complémentaires {SR, A2, SA2} permettant de caractériser complètement le nouvel objet cas {SR, A1, A2, ACT, SA1, SA2}. Enfin, le nouveau cas est indexé et stocké dans la base de cas [COLLOC J. & BOUZIDI L., 2001].

L'indexation des cas repose sur le calcul d'une distance. Différents modèles de distance peuvent être utilisés pour ordonner les cas : la logique floue et la théorie de la preuve [SCHUSTER F.E. et AL., 1997].

Figure 41 : Le cycle du RpC du système d'aide à la recherche d'information proposé



Le processus d'aide à la décision en matière de recherche d'information :

Le processus d'aide à la décision qui permet à un acteur/utilisateur de formuler une nouvelle action de recherche en fonction de cas antérieurs présentant un fort degré de similarité, constitue une phase importante du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons.

Sa prise en compte représente une étape indispensable avant de procéder à l'indexation des différentes actions de recherche consécutives ayant permis un accroissement significatif du niveau de satisfaction d'un acteur/utilisateur au cours d'une session de recherche.

Ce processus repose en grande partie sur :

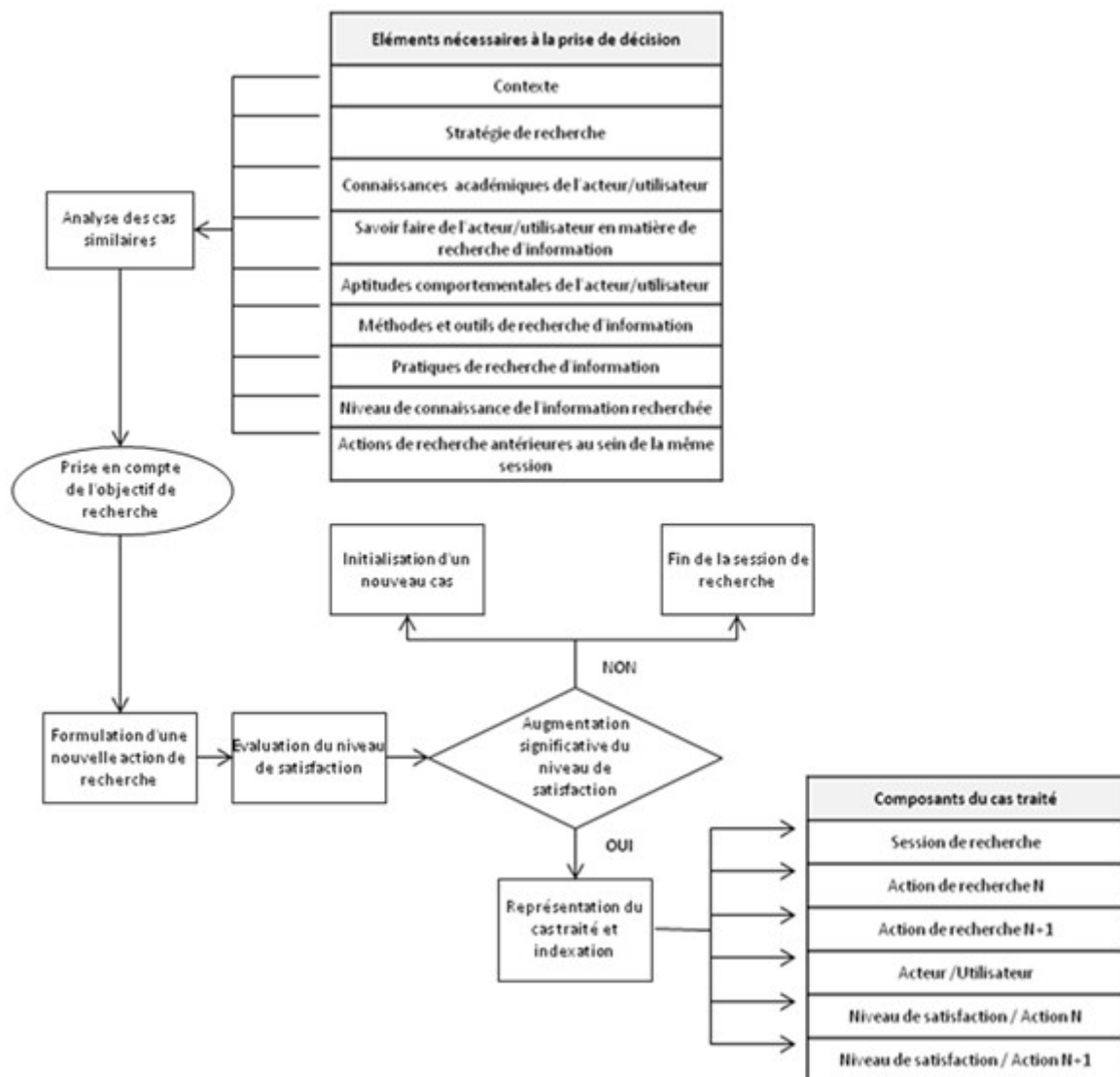
- Le profil de l'acteur humain ;
- Le contexte professionnel dans lequel il évolue ;

- La nature des outils technologiques dont l'acteur humain dispose pour rechercher de l'information ;
- Des expériences de recherche d'information antérieures qu'il a pu réaliser.

Il comprend une phase d'analyse des cas similaires, la prise en compte de l'objectif de recherche de l'acteur/utilisateur et la formulation d'une nouvelle action de recherche.

La figure 42 décrit notamment les éléments de connaissance nécessaires pour parvenir à une prise de décision de l'acteur/utilisateur et à la formulation d'une nouvelle action de recherche.

Figure 42 : Etapes du processus d'aide à la décision



d) les facteurs clés de succès

Identifier l'ensemble des facteurs clés de succès constitue une phase importante de la démarche de mise en œuvre d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences.

Elle nous permet de déterminer les leviers sur lesquels nous pouvons agir afin d'anticiper un certain nombre de problèmes susceptibles d'intervenir dans l'utilisation d'un système de ce type au sein d'une solution de GED.

Mettre en place un tel système peut constituer un risque de dysfonctionnement de la solution de GED et représenter un investissement important pour l'organisation (mobilisation de ressources humaines, financières et technologiques). Mais, il est clair qu'un système de ce type, bien implanté, constitue un facteur de performance non négligeable dans le mécanisme de recherche d'information au sein d'une solution de GED.

L'ensemble des éléments à prendre en considération pour faire du système proposé un facteur de performance supplémentaire dans le processus de recherche d'information se décline essentiellement en termes d'adaptabilité et d'ergonomie.

Une des principales caractéristiques du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons est de pouvoir s'adapter en permanence à l'évolution des besoins des différents acteurs et de ne pas chercher à imposer des stratégies profilées de recherche d'information (notamment par catégories d'acteurs).

Par contre, cette adaptabilité nécessite d'auditer et de contrôler régulièrement les pratiques et les usages liés à ce type de système pour pouvoir définir au mieux l'ensemble des paramètres qui le compose (comme par exemple le nombre d'actions de recherche jugées insatisfaisantes avant déclenchement de l'effet sérendipité, le seuil pris en compte pour le calcul de similarité ou bien encore l'affichage du nombre de documents correspondants à une requête).

Un autre facteur clé de succès repose sur la convivialité des outils d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons. Cela dépend en grande partie de la conception des IHM mais aussi du niveau d'appropriation et de la fréquence d'utilisation du système de recherche d'information par les différents acteurs. Le principe d'ancrage visuel qui n'apparaît qu'après plusieurs actions de recherche où l'acteur/utilisateur du système ne parvient pas à augmenter son niveau de satisfaction contribue fortement à accroître cette convivialité.

Sur un plan opérationnel, il est important d'étudier la façon dont le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons va être techniquement produit. Les performances du système proposé dépendent en grande partie de la façon dont il aura été conçu et techniquement réalisé mais aussi de l'usage qu'il en sera fait.

Du point de vue de l'acteur/utilisateur, cela se traduit par une rapidité du système de recherche à répondre à ses demandes et à enchaîner les différentes phases d'un processus de recherche d'information sans subir de ralentissements.

L'évaluation des résultats (documents trouvés) et de leur degré de pertinence suite aux différentes actions de recherche constitue une opération clé dans un processus de recherche d'information. Cette évaluation est également fonction de l'ergonomie dans la présentation des résultats et nous conduit à porter une attention toute particulière à la façon dont le dialogue entre le système d'aide à recherche d'information et l'acteur humain s'installe.

L'approche dite "tridimensionnelle" que nous avons adoptée nous a permis d'accorder un soin tout particulier à ce dialogue entre l'acteur humain et le système mais aussi de concevoir des dispositifs visuels qui soient le plus adapté possible aux attentes des utilisateurs.

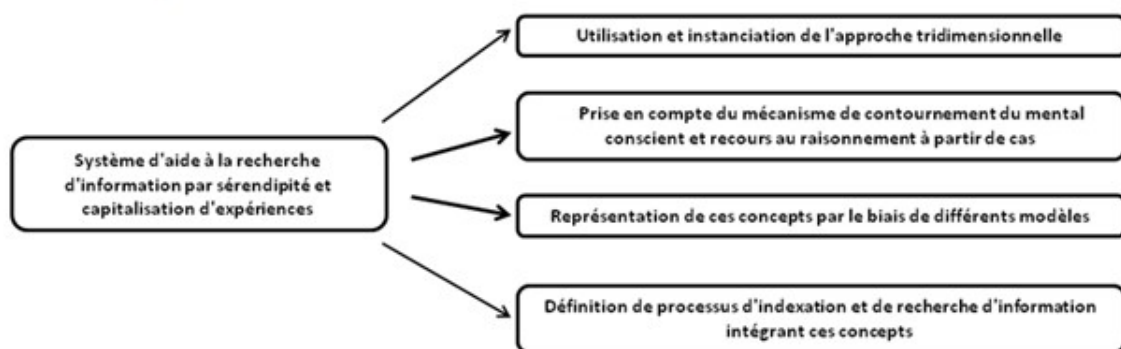
Enfin, le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons se veut auto-adaptatif au sens où le processus de capitalisation des expériences de recherche d'information vise à réutiliser les fruits des expériences de recherche passées tout en tenant compte de l'évolution des savoir faire de l'acteur/utilisateur en matière de recherche d'information. Il permet également de prendre en considération la diversité des organisations, l'évolution des besoins informationnels et le profil des différents acteurs qui utilisent un système de GED dans le cadre de leur activité. On peut ainsi dire qu'il présente différents degrés d'analyse et de représentation du besoin informationnel.

A ce stade de notre réflexion, nous avons pu établir les bases d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences qui se fonde sur les aspects suivants :

- L'utilisation et l'instanciation de l'approche tridimensionnelle ;
- La prise en compte du mécanisme de contournement du mental conscient (prise en compte de l'effet "sérendipité") et le recours à un modèle de raisonnement à partir de cas ;

- La représentation de ces concepts à travers l'utilisation de différents modèles (Motivations, modèle des niveaux logiques de la pensée, de l'action et du changement, modèle de raisonnement à partir de cas, etc.) ;
- La définition de processus d'indexation et de recherche d'information qui intègre ces concepts.

Figure 43 : Fondement d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences



Au terme de cette troisième partie, il apparaît clairement que la mise en œuvre d'un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences peut constituer un facteur de performance non négligeable dans le mécanisme de recherche d'information au sein d'une solution de GED à condition que l'alliance entre l'homme et la machine soit parfaitement réussie (notamment en termes d'adaptabilité et d'ergonomie) et que le système proposé soit en capacité de se rapprocher du mode de fonctionnement de l'esprit humain.

Dans la dernière partie de cette thèse, nous décrirons le contexte d'expérimentation du système d'aide à la recherche d'information proposé ainsi que sa validation par le biais d'une instanciation au sein d'un portail documentaire intégrant des fonctionnalités de Gestion Electronique de Documents et un retour d'expérience mettant en perspective les améliorations possibles.

Partie 4 : Expérimentation et validation du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences

Afin de pouvoir valider notre approche et de voir les limites du modèle proposé, une réalisation pratique a été effectuée et sera prochainement mise en œuvre au sein d'une Société de Services en Ingénierie Informatique (SSII) de la région Rhône-Alpes : la société HEXAGRAM. Dans le cadre de la prochaine version de son Intranet documentaire, qui intègre des fonctionnalités de Gestion Electronique de Documents, une instanciation du modèle proposé sera opérée afin que les différents acteurs humains de l'entreprise puissent utiliser, en complément des outils de recherche traditionnels, cette nouvelle fonctionnalité de recherche d'information.

I. Choix du terrain d'expérimentation

Créée en décembre 2000, HEXAGRAM est une société par actions simplifiée au capital de soixante seize mille euros et réalise un chiffre d'affaires annuel d'environ un million huit cent mille euros.

Premier Eco-infrastructureur de la Région Rhône-Alpes, HEXAGRAM a choisi de s'engager dans la lutte contre le changement climatique et la réduction des émissions de gaz à effet de serre (GES) en proposant à ses clients de concevoir, mettre en œuvre, superviser et maintenir des infrastructures techniques sécurisées à base de solutions (logicielles et/ou matérielles) ayant un moindre impact sur l'environnement

Constituée autour de six associés salariés, HEXAGRAM emploie actuellement 10 personnes dont son Président (occupant également la fonction de consultant AMOA), une Secrétaire Générale, une Assistante de Communication, trois Experts Techniques, un Ingénieur d'Etudes et Développement, deux Techniciens, un Responsable Commercial et un Ingénieur Avant Vente.

La plupart des collaborateurs d'HEXAGRAM sont issus de sociétés de services de la région Rhône-Alpes et possèdent plusieurs années d'expérience dans la mise en œuvre et la maintenance de solutions systèmes et réseaux.

Afin d'asseoir durablement son développement en région Rhône-Alpes, HEXAGRAM peut également compter sur le soutien de partenaires économiques et financiers (à l'image du Groupe Bernard et des investisseurs financiers présents à son capital), de la Chambre de Commerce et d'Industrie et de la région Rhône-Alpes.

Dès sa création, HEXAGRAM a résolument décidé d'orienter son offre commerciale vers un public de grandes et moyennes entreprises sur le grand Rhône-Alpes. Fort d'un important tissu relationnel dans cette cible de clientèle, l'entreprise souhaite conserver cette vocation régionaliste tout en poursuivant sa croissance à un bon rythme.

Habitués à travailler ensemble, les collaborateurs d'HEXAGRAM sont avant tout des professionnels expérimentés, formés en permanence aux nouvelles technologies et qui garantissent aux grandes et moyennes entreprises un haut niveau de réactivité. La complémentarité et l'expérience des collaborateurs font qu'aujourd'hui la société HEXAGRAM connaît une croissance importante et un fort volume d'activité sur un marché hautement concurrentiel.

II. Description du terrain d'expérimentation

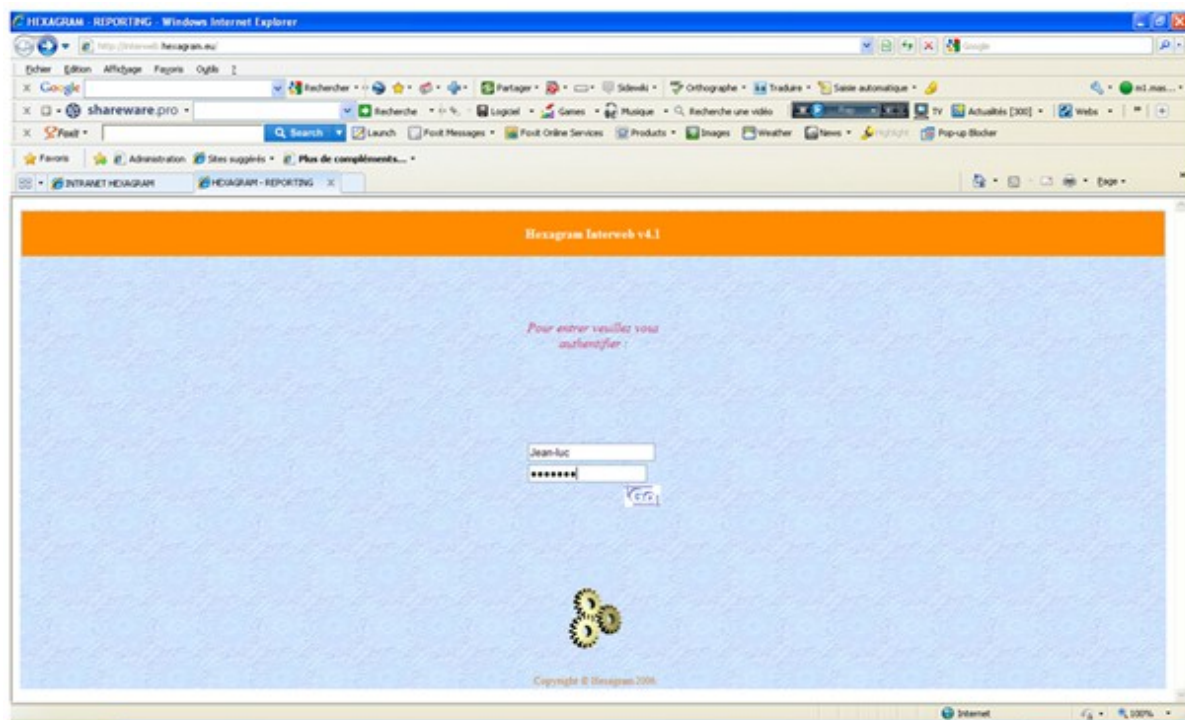
L'Intranet documentaire de la société HEXAGRAM a été développé en 2007 par Alain GALLET, Ingénieur d'Etudes et Développement de la société HEXAGRAM, afin de répondre à l'ensemble des besoins informationnels des différents acteurs humains qui, dans leur fonction et afin d'opérer des traitements opérationnels ou de prendre des décisions, éprouvent le besoin de rechercher de l'information relative à l'activité des experts techniques, de l'ingénieur d'études et développement, des techniciens et du consultant AMOA d'HEXAGRAM.

Les fonctionnalités de Gestion Electronique de Documents implantées dans l'Intranet documentaire de la société HEXAGRAM sont plutôt orientées "archivage électronique" et répondent à l'objectif de constituer une mémoire organisationnelle de l'entreprise.

Cet Intranet est accessible en interne ou en externe par le biais d'un simple navigateur internet et de l'accès à un serveur Web hébergé chez HEXAGRAM.

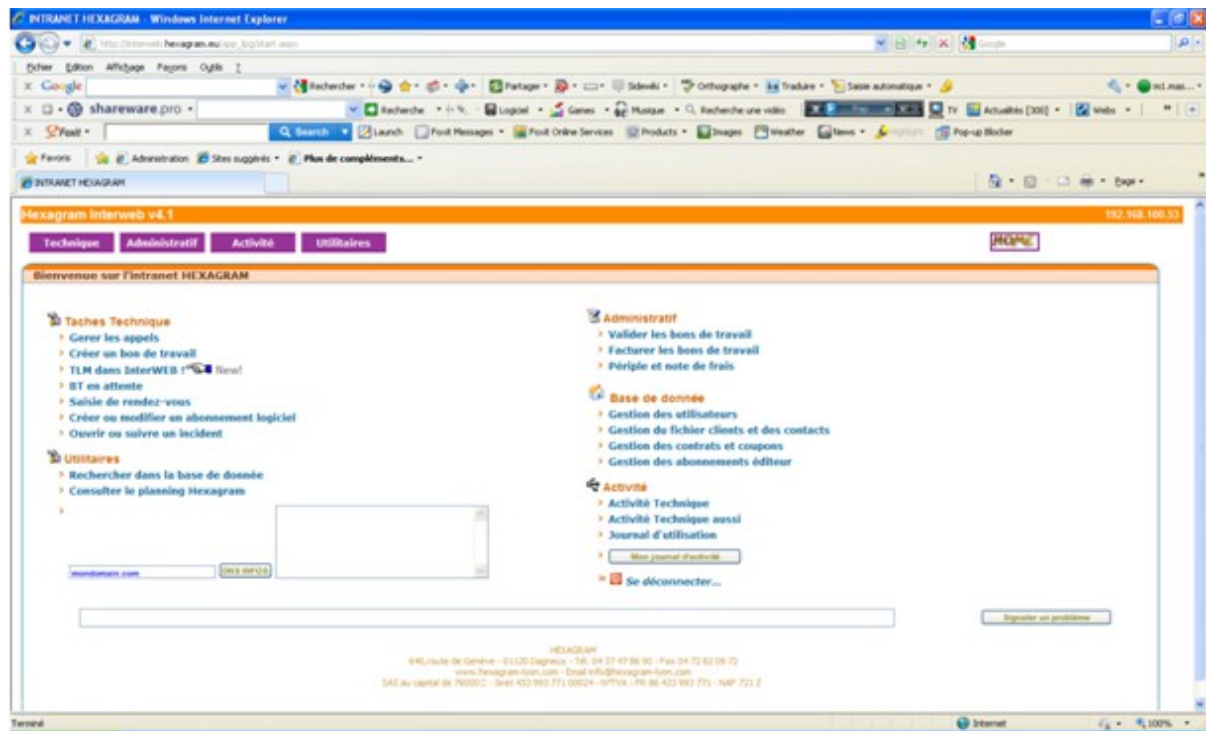
Pour accéder à cet Intranet, un collaborateur doit tout d'abord s'authentifier à l'aide de son compte d'utilisateur présent dans l'annuaire LDAP (pour Lightweight Directory Access Protocol) d'HEXAGRAM et du mot de passe associé (un annuaire LDAP est une bibliothèque électronique mise à jour régulièrement qui regroupe des informations sur l'ensemble des membres d'une organisation).

Figure 44 : Intranet HEXAGRAM – Fenêtre d'authentification



L'Intranet documentaire de la société HEXAGRAM comprend plusieurs fonctionnalités dont une brique fonctionnelle permettant aux experts techniques, à l'ingénieur d'études et développement, aux techniciens et au consultant AMOA d'HEXAGRAM de créer des bons de travaux qui seront par la suite transmis aux clients d'HEXAGRAM où ils sont intervenus pour validation.

Figure 45 : Intranet HEXAGRAM – Page principale



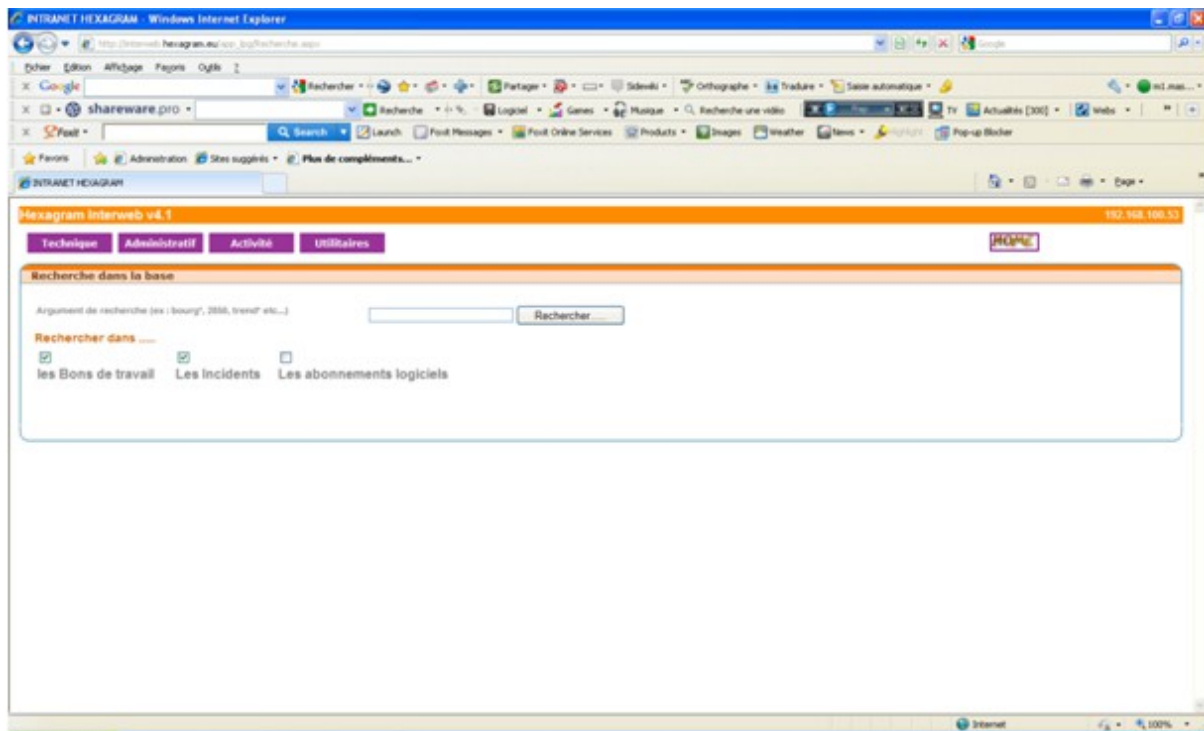
La brique fonctionnelle permettant de valider les bons de travaux associée à un bon de travail créé dans l'intranet et stocké sous la forme d'un fichier au format PDF, une copie numérique de ce même bon de travail validé (c'est-à-dire daté, signé et tamponné) retourné par le client.

Le client peut alors choisir de transmettre ce bon de travail validé soit par le biais d'un document numérique (document papier numérisé) attaché à un courriel, d'une télécopie (document au format numérique) ou d'un document papier qui sera numérisé dans les locaux d'HEXAGRAM.

Une fois la phase de validation opérée, les prestations réalisées par les experts techniques, l'ingénieur d'études et développement, les techniciens ou le consultant AMOA d'HEXAGRAM peuvent alors être facturées.

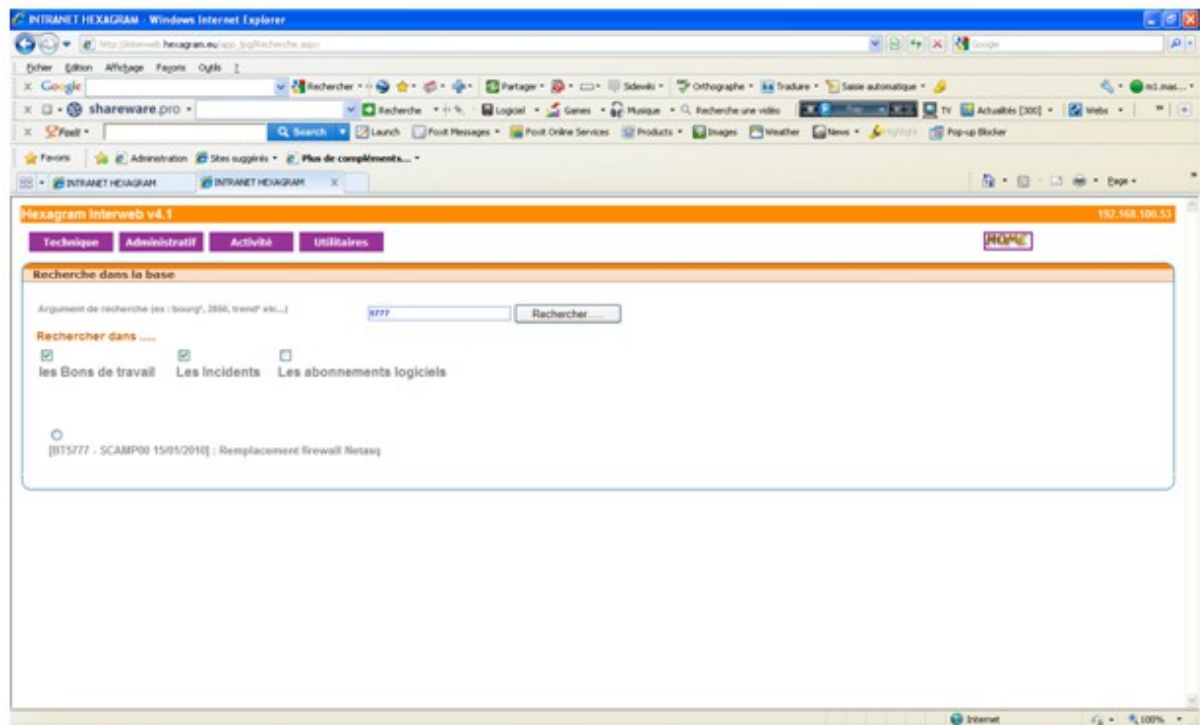
Lorsque qu'un collaborateur d'HEXAGRAM recherche un bon de travail, il utilise la fonctionnalité permettant de rechercher un document dans la base de données.

Figure 46 : Intranet HEXAGRAM – Fonction de recherche



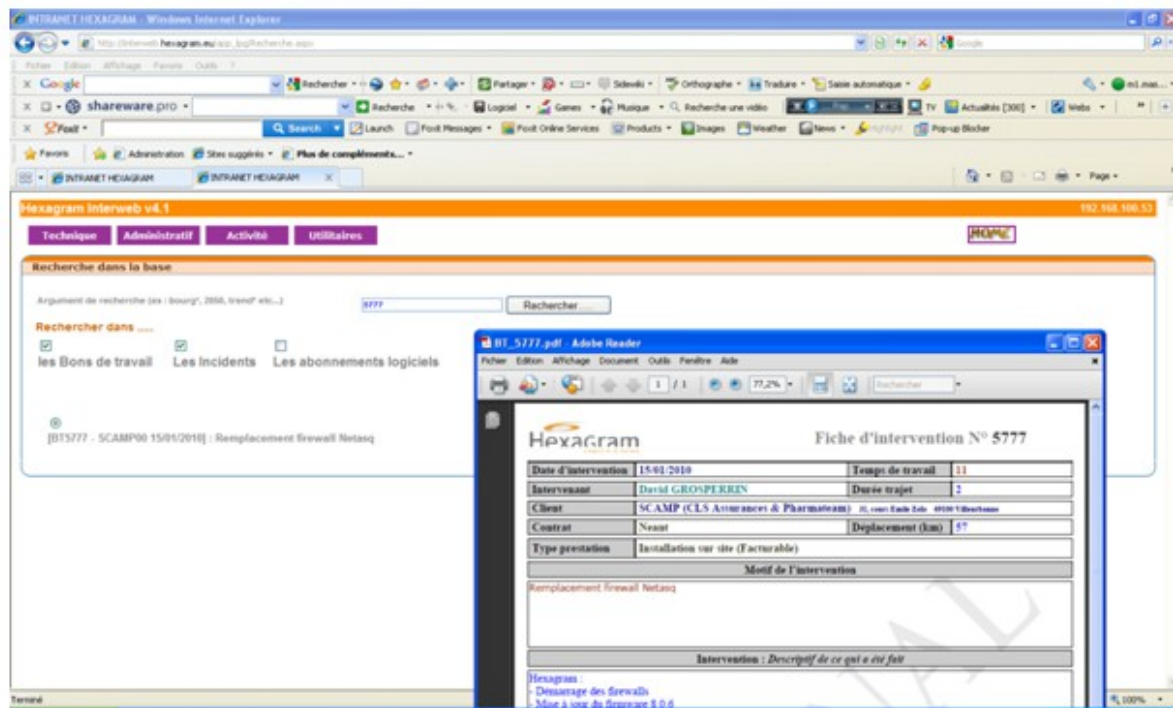
Cette brique fonctionnelle permet de rechercher une information structurée dans la base de données de l'Intranet HEXAGRAM. Si les informations existent, l'Intranet fournit la liste des documents correspondants aux critères de recherche.

Figure 47 : Intranet HEXAGRAM – Recherche d'un bon de travail



Il suffit alors au collaborateur d'HEXAGRAM à l'origine de la recherche d'information de cliquer sur le document qui semble répondre à son besoin informationnel pour voir s'afficher dans une fenêtre, un fichier au format PDF, qui correspond au bon de travail créé dans l'Intranet suivi du document validé par le client.

Figure 48 : Intranet HEXAGRAM – Affichage d'un bon de travail



Outre les bons de travaux, cet Intranet permet également de gérer les incidents techniques mais aussi les abonnements logiciels.

Toutefois, et malgré tout l'intérêt que représente cet Intranet pour l'ensemble des collaborateurs d'HEXAGRAM, la plupart d'entre eux considèrent qu'il n'est pas assez complet et que de nombreux documents relatifs aux opérations techniques d'HEXAGRAM sont absents.

Cette carence informationnelle les amène d'ailleurs à utiliser de manière complémentaire plusieurs plateformes logicielles (logiciel de gestion commerciale WaveSoft, messagerie d'entreprise Microsoft Exchange, dossiers partagés accessibles sur le SAN, etc.) pour rechercher de l'information qu'ils ne trouvent pas toujours dans l'Intranet.

C'est pourquoi, la direction de la société HEXAGRAM a souhaité faire évoluer cet Intranet pour qu'à chaque intervention technique soit associé un certain nombre de documents relatifs à cette intervention (au-delà du bon de travail qui est produit par l'Intranet) et provenant de

sources d'information différentes (environnements de production de documents internes et externes à l'entreprise).

Cette nouvelle version de l'Intranet devra prendre en considération le fait que les interventions réalisées par les experts techniques, l'ingénieur d'études et développement, les techniciens et le consultant AMOA d'HEXAGRAM sont de nature très différente.

Tableau 4 : Nature des interventions et modes d'intervention par catégorie d'acteurs

Acteurs	Nature des interventions	Modes d'intervention
Experts techniques & Techniciens	- Audit - Conception - Installation - Documentation - Transfert de compétences - Dépannage - Assistance technique - Supervision	- Sur site - A distance (depuis une connexion sécurisée au réseau du client) - Dans les locaux d'HEXAGRAM - Par téléphone
Ingénieur d'études et développement	- Audit - Analyse - Réalisation d'interfaces graphiques - Développement - Installation et paramétrage - Documentation - Transfert de compétences - Maintenance applicative - Assistance aux utilisateurs	- Sur site - A distance (depuis une connexion sécurisée au réseau du client) - Dans les locaux d'HEXAGRAM - Par téléphone
Consultant AMOA	- Audit - Conseil - Rédaction de livrables - Transfert de compétences	- Sur site - Dans les locaux d'HEXAGRAM - Par téléphone

Selon le profil des différents acteurs et la nature de leurs interventions, les documents et donc les informations stockées et indexées dans l'Intranet d'HEXAGRAM, ne seront pas forcément du même type.

Tableau 5 : Nature des documents et des informations par catégorie d'acteurs

Acteurs	Nature des documents et des informations
Experts techniques & Techniciens	- Bon de commande client - Bon de commande fournisseur - Bon de commande HEXAGRAM - Bon de livraison fournisseur - Bon de livraison HEXAGRAM - Bon de travail - Cahier des charges - Contrat d'assistance réseau - Contrat de prestation de service OPCA - Contrat de régie - Contrat de service - Contrat de supervision - Contrat d'heures - Convention de formation - Demande d'intervention - Documentation technique - Dossier de conception - Dossier d'installation - Facture fournisseur - Facture HEXAGRAM - Fiche de présence - Guide d'administration - Information de garantie constructeur - Information relative à un abonnement logiciel - Information relative à une demande d'assistance technique - Information relative à un incident technique - Manuel d'utilisation - Procédure d'installation - Rapport d'audit - Rapport d'incident - Rapport journalier ou hebdomadaire de supervision - Recette d'installation
Ingénieur d'études et développement	- Bon de commande client - Bon de commande HEXAGRAM - Bon de travail - Cahier des charges - Contrat d'assistance téléphonique - Contrat de maintenance applicatif - Convention de formation

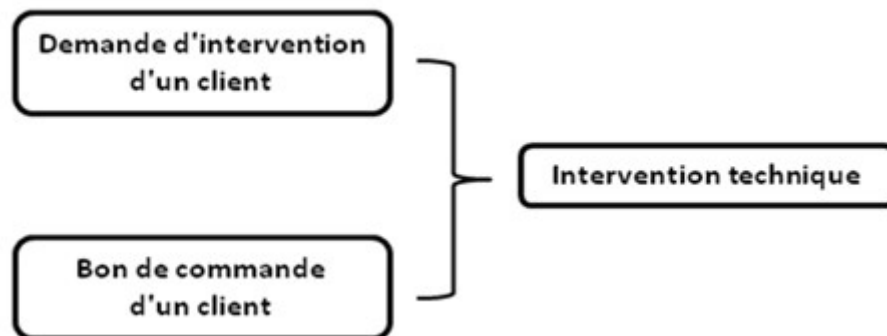
Acteurs	Nature des documents et des informations
	- Demande d'intervention - Dossier de maintenance - Dossier de programmation - Dossier de spécifications fonctionnelles - Dossier de spécifications techniques - Dossier d'installation - Etude d'opportunité - Facture HEXAGRAM - Fiche de présence - Guide d'administration - Information relative à une demande d'assistance téléphonique - Information relative à un dysfonctionnement applicatif - Manuel d'utilisation - Plan de formation - Procédure d'installation - Rapport d'analyse préalable - Rapport d'audit - Rapport de tests - Recette d'installation - Recette des prestations
Consultant AMOA	- Bon de commande client - Bon de commande HEXAGRAM - Bon de travail - Contrat de service - Convention de formation - Demande d'intervention - Documents liés à la mission (schéma directeur, portefeuille de projet, budget de fonctionnement et d'investissement, tableaux de bord, etc.) - Facture HEXAGRAM - Fiche de présence - Lettre de mission - Note de synthèse - Plan de formation - Rapport de mission - Recette de la mission

Pour bien comprendre le processus d'intervention d'un expert technique, d'un ingénieur d'études et développement, d'un technicien et d'un consultant AMOA, il convient dans un premier temps d'identifier tous les événements qui sont à l'origine d'une intervention technique qu'elle soit réalisée sur le site d'un client, dans les locaux d'HEXAGRAM ou depuis un dispositif technique permettant une prise de main à distance.

Ces événements sont essentiellement :

- La réception d'une demande d'intervention (transmise par messagerie électronique ou télécopie et très rarement par courrier) ;
- La réception d'un bon de commande (envoyé par courrier, messagerie électronique ou télécopie).

Figure 49 : Documents à l'origine d'une intervention technique



En cas d'appel téléphonique (notamment des clients sous contrat bénéficiant d'une d'assistance téléphonique) ou de remontée d'alertes de supervision (pour ceux bénéficiant d'un contrat de supervision), une intervention technique de quelque nature que ce soit (prise de main à distance, intervention sur site, etc.) ne peut être initiée sans qu'une demande d'intervention ne soit préalablement adressée à HEXAGRAM.

Par ailleurs, certains des acteurs/utilisateurs de l'Intranet considèrent également que ce dernier n'est pas forcément très performant ni convivial en matière de recherche d'information.

C'est la raison pour laquelle, à l'occasion de l'écriture d'une nouvelle version de cet Intranet, la Direction de la société HEXAGRAM a demandé à son ingénieur d'études et développement de revoir le processus de recherche d'information.

Travaillant en étroite collaboration avec l'équipe de recherche SICOMOR (Systèmes d'Information Communicants, Management et Organisation) et entretenant, depuis sa création, un lien de partenariat avec l'IAE de Lyon, la société HEXAGRAM a bien voulu intégrer à titre d'expérimentation, le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous avons conçu dans sa nouvelle version de son Intranet.

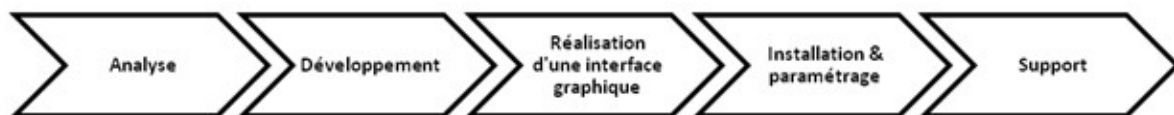
Cette expérimentation au sein d'un Intranet documentaire intégrant des fonctions de GED constitue un terrain tout à fait approprié afin de valider la pertinence et la qualité du système d'aide à la recherche d'information proposé.

III. Déroulement de l'expérimentation

Tout d'abord, il est important de signaler que cette expérimentation s'inscrit dans le cadre d'un projet de refonte de l'Intranet de la société HEXAGRAM qui comprend cinq phases et pour lequel un cahier des charges avait préalablement été constitué.

A ce titre, la prise en compte du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences a nécessité de revoir les spécifications fonctionnelles et techniques de la nouvelle version de l'Intranet.

Figure 50 : Phases du projet de refonte de l'Intranet de la société HEXAGRAM



La phase d'analyse : Il s'agit de l'ensemble des étapes préparatoires nécessaires au lancement du projet, du découpage du projet en tâches élémentaires, de la planification de l'exécution de ces tâches et de la définition des ressources à mobiliser.

La phase de développement : Il s'agit de la réalisation de la nouvelle version de l'Intranet en utilisant la technologie de développement ASP.NET de Microsoft.

La phase de réalisation d'une interface graphique : Il s'agit de la prise en compte de la charte graphique et de l'ergonomie de la nouvelle version de l'Intranet.

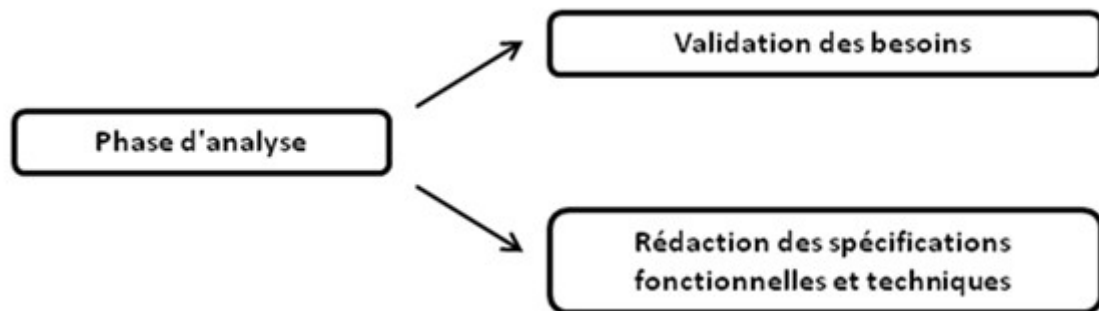
La phase de support : Il s'agit de l'ensemble de la documentation devant accompagner cette nouvelle version mais aussi de la formation des utilisateurs et administrateurs de l'Intranet.

La phase d'installation et de paramétrage : Il s'agit de l'étape d'installation et de paramétrage sur le serveur Web d'HEXAGRAM, de l'ensemble des tests permettant de valider la conformité de la nouvelle version de l'Intranet par rapport au cahier des charges et des corrections éventuelles à apporter à cette nouvelle version.

La phase de support : Il s'agit de l'ensemble de la documentation devant accompagner cette nouvelle version mais aussi de la formation des utilisateurs et administrateurs de l'Intranet.

La phase d'analyse se décompose en deux parties : la validation des besoins et la rédaction des spécifications fonctionnelles et techniques.

Figure 51 : Phase d'analyse du projet de refonte de l'Intranet de la société
HEXAGRAM



Compte tenu du caractère expérimental de la nouvelle fonction de recherche d'information de l'Intranet, le processus de validation des besoins a dû être revu afin de permettre l'intégration du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous souhaitons valider.

La phase de développement a d'ores et déjà été achevée et une pré-version de la nouvelle version de l'Intranet est actuellement en cours de tests. Une fois la nouvelle version de l'Intranet validée, sa mise en production sur le serveur Web d'HEXAGRAM pourra être effectuée.

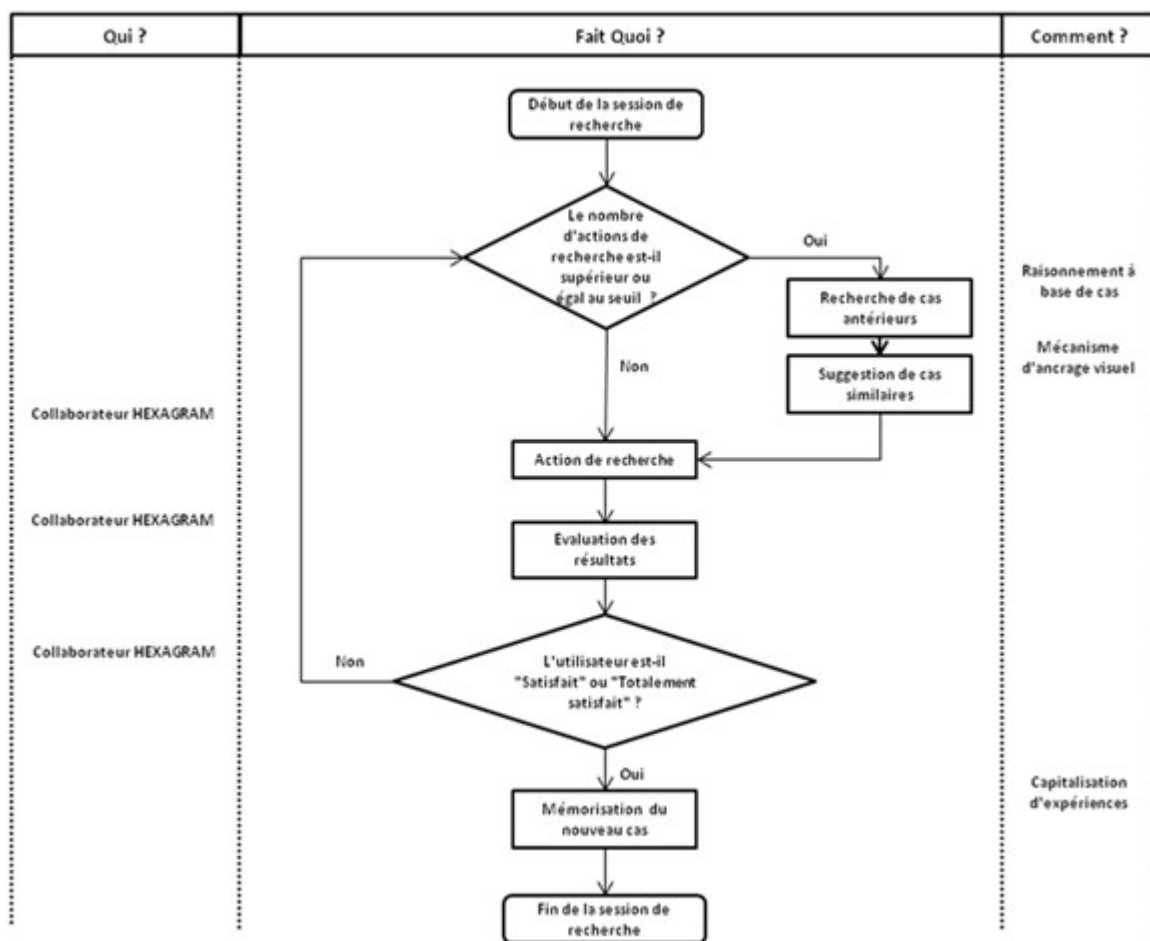
Le processus d'expérimentation que nous avons retenu se décompose en quatre phases :

- La conception d'une nouvelle fonction de recherche d'information au sein de l'Intranet ;
- Le développement de cette nouvelle fonction ;
- L'évaluation de cette fonction par des utilisateurs clés chargés des tests fonctionnels ;
- La détermination des points d'amélioration.

La phase de conception d'une nouvelle fonction de recherche d'information au sein de l'Intranet d'HEXAGRAM a donné lieu à une refonte totale du processus de recherche actuel.

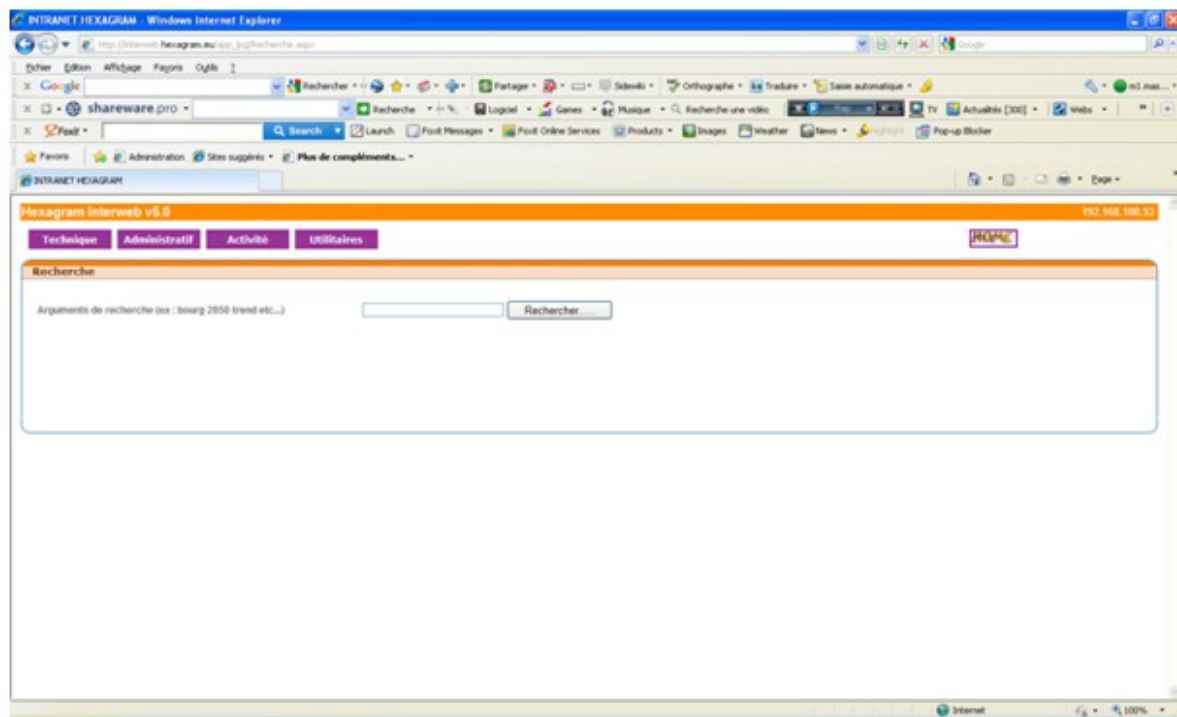
Cette réécriture du processus de recherche s'explique en partie par le fait que la nature des informations susceptibles d'être recherchées a évolué et tient compte de la diversité des documents qui seront désormais stockés et indexés dans l'Intranet d'HEXAGRAM. Elle s'explique également par la prise en compte de l'effet sérendipité et du mécanisme de raisonnement à partir de cas.

Figure 52 : Nouveau processus de recherche d'information



La fonction de recherche d'information qui résulte de ce processus se présente désormais sous la forme d'un moteur de recherche en texte intégral dont l'interface graphique se limite à un seul champ de saisie et un bouton de recherche. Cette fonctionnalité permet de rechercher de l'information structurée (stockée dans une base de données et ayant une architecture interne cohérente et prédéfinie) et non structurée (n'ayant aucune structure fixe et prédéfinie et dont le contenu est variable) et s'appuie sur un processus d'indexation des documents produits et non produits par l'Intranet qui intègre un moteur d'OCR pour la reconnaissance de textes.

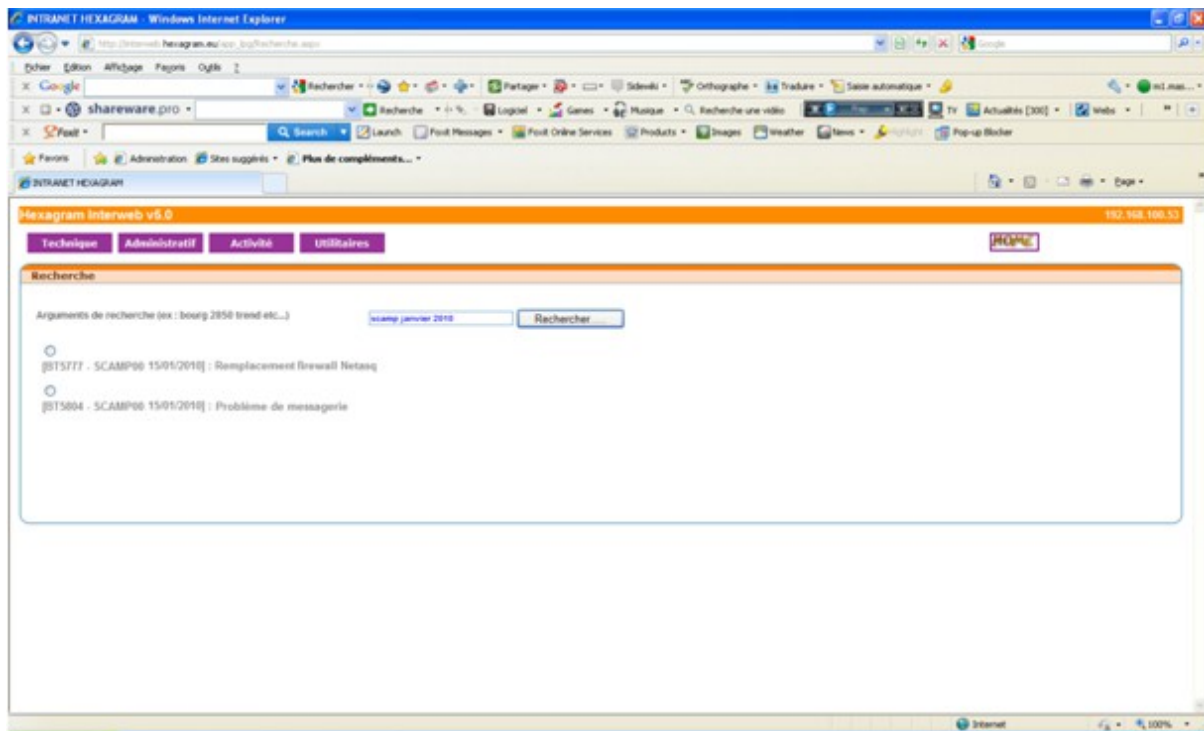
Figure 53 : Intranet HEXAGRAM – Nouvelle fonction de recherche



Afin d'illustrer la manière dont se déroule le processus de recherche d'information dans la nouvelle version de l'Intranet, nous allons procéder à la recherche d'un bon de commande émis par un client de la société HEXAGRAM à l'aide de quelques critères comme la raison sociale du client, le mois et l'année de commande. Le collaborateur à l'origine de ce processus de recherche documentaire a besoin d'accéder à une information précise qui figure sur ce document qui a été transmis par courrier à HEXAGRAM avant d'être numérisé puis stocké et indexé dans l'Intranet.

Pour initier le processus de recherche le collaborateur à l'origine du besoin d'information saisie dans le champ approprié un certain nombre de termes qui, selon lui, caractérisent au mieux le document qu'il recherche (raison sociale du client, mois et année de la commande). Après avoir saisi ces termes dans la zone de recherche appropriée, il opère une première action de recherche.

Figure 54 : Intranet HEXAGRAM – Exemple de recherche

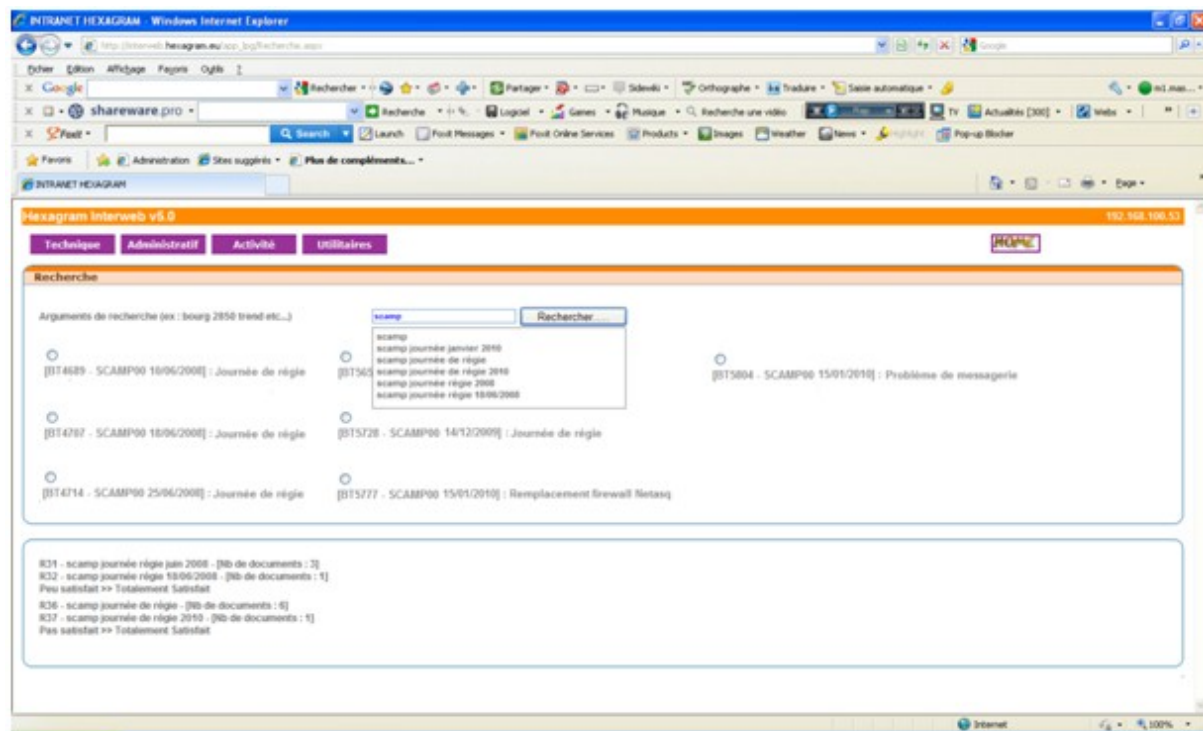


Cette première action de recherche n'étant pas jugée satisfaisante aux vues des documents trouvés et de leur contenu, le collaborateur à l'origine de la première action de recherche est alors amené à procéder à une deuxième action de recherche en faisant évoluer sa requête.

Après une deuxième puis une troisième action de recherche jugée peu satisfaisante par le collaborateur ayant initié le processus de recherche, une quatrième action de recherche est à nouveau menée.

Au premier terme saisi dans le champ approprié, l'effet "sérendipité" se manifeste alors sous la forme d'une zone graphique dans le bas de l'écran où figurent des expériences de recherche antérieures proches de celle dernièrement effectuée et ayant permis d'élever de façon significative le niveau de satisfaction de ce collaborateur.

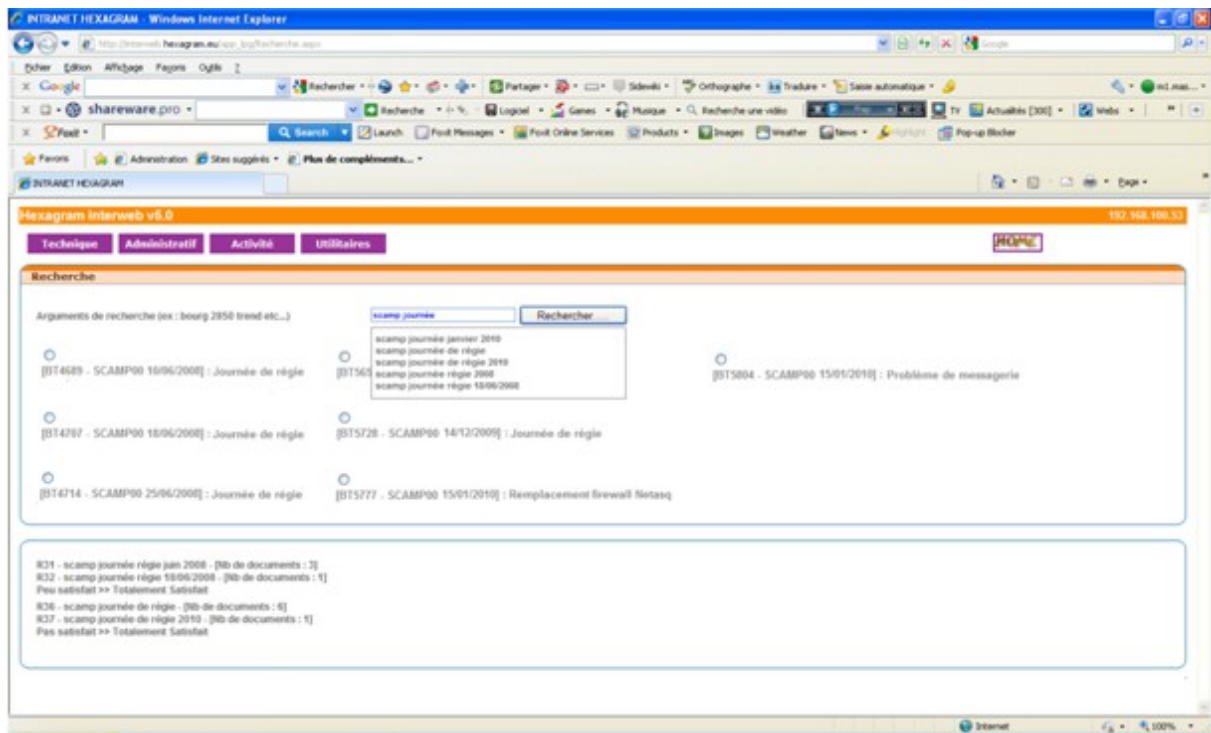
Figure 55 : Intranet HEXAGRAM – Manifestation de l'effet sérendipité



Le fait que cet effet se manifeste à la quatrième action de recherche est l'un des paramètres du système que nous avons à renseigner puis à modifier en fonction du niveau de sensibilité que l'on désire.

Au fur et à mesure que le collaborateur précise les termes de sa requête dans le champ de recherche, des suggestions pertinentes de termes s'affichent en dessous de la zone de saisie. Ce phénomène s'accompagne également d'une information complémentaire relative au nombre de résultats qui apparaîtront si le collaborateur en question sélectionne les termes de la recherche qui lui sont proposés.

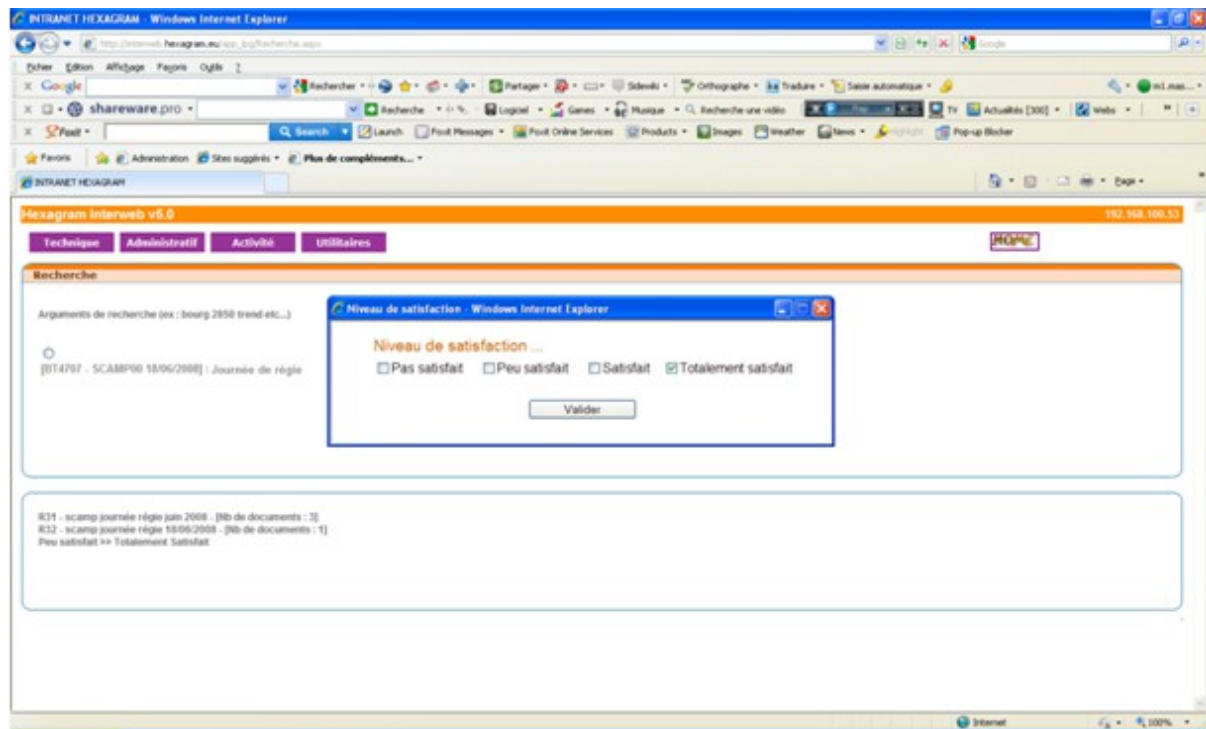
Figure 56 : Intranet HEXAGRAM - Informations complémentaires lors de la saisie d'une nouvelle requête



Au terme du processus de recherche et après avoir atteint un haut niveau de satisfaction, le collaborateur en quête d'une information bien spécifique sur un bon de commande client quitte la fonction de recherche et/ou l'application de GED non sans avoir signalé au système d'aide à la recherche d'information son niveau de satisfaction (comme il l'aura fait d'ailleurs après chaque action de recherche).

Si l'utilisateur veut appeler une autre fonction ou quitter l'application de GED, il ne pourra en aucun cas le faire sans avoir au préalable spécifié son niveau de satisfaction.

Figure 57 : Intranet HEXAGRAM - Evaluation du niveau de satisfaction



Quant au développement de cette nouvelle fonction, il a nécessité un temps de programmation relativement important. La complexité des mécanismes à intégrer (mécanisme d'ancrage et d'affichage dynamique, utilisation du raisonnement à partir de cas, capitalisation des expériences de recherche, etc.) et le mode de fonctionnement de l'Intranet (nouvelles fonctionnalités à intégrer et nouveaux types de documents à gérer) ont généré de nombreux dépassements par rapport aux temps estimés pour les différentes étapes de réalisation de cette fonction.

L'évaluation de cette nouvelle fonction de recherche par des utilisateurs clés chargés des tests fonctionnels n'a pas vraiment donné le résultat escompté pour la simple et bonne raison du manque de documents hétérogènes constituant le fonds documentaire.

La nouvelle version de l'Intranet n'étant pas en production, nous avons donc dû nous adapter à cet état de fait pour essayer d'analyser les usages des différents utilisateurs clés.

Les utilisateurs clés en charge des tests étant au nombre de trois, un par grand domaine fonctionnel de l'entreprise (un utilisateur dans le domaine administratif et financier, un technicien et un commercial) et ayant chacun un emploi du temps relativement chargé, la phase de tests n'a pas été aussi révélatrice que nous l'aurions souhaité. Dans tous les cas, elle ne nous permet pas de valider totalement le système proposé et nécessite de poursuivre après la mise en production de la nouvelle version de l'Intranet.

Toutefois, différents points d'amélioration ont pu d'ores et déjà être dégagés aux vues des problèmes de performance rencontrés, du manque d'ergonomie de l'interface de recherche et de la difficulté pour un utilisateur à évaluer son niveau de satisfaction.

IV. Retour d'expérience

A ce jour, le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous avons projeté de valider dans le cadre de la nouvelle version de l'Intranet de la société HEXAGRAM n'est toujours pas en production.

A cela deux raisons, la première est liée à un contexte économique difficile qui fait que la ressource (ingénieur d'études et développement) affectée au développement de la nouvelle version de l'Intranet est actuellement dédiée à d'autres projets beaucoup plus rémunérateurs pour HEXAGRAM. La deuxième raison tient à un cahier des charges beaucoup trop ambitieux qui nécessite des temps de développement bien supérieurs à ceux qui avaient été estimés et qui font qu'à ce jour certaines fonctionnalités n'ont toujours pas été réalisées ni testées.

Toutefois, l'expérimentation menée sur la version non finalisée de l'Intranet nous laisse entrevoir un certain nombre de problèmes et de facteurs à prendre en considération avant la mise en production de la nouvelle version.

Le premier facteur d'amélioration concerne la performance du système et sa capacité à monter en charge. Lorsque l'effet "sérendipité" se manifeste et que le mécanisme d'interrogation de la base de cas s'opère, on constate des phénomènes de ralentissement important qui s'amplifient lors de l'affichage des cas antérieurs similaires trouvés. La volonté de voir apparaître le nombre de documents répondant aux critères de recherche des cas antérieurs suggérés semblent être pour beaucoup dans ces problèmes d'affichage.

Une étude de caractérisation et de prévision de la charge de même qu'une évaluation prévisionnelle de la sûreté de fonctionnement devraient nous apporter des éléments de compréhension des phénomènes observés. Cette étude nous apparaît indispensable afin de faire évoluer l'architecture applicative existante pour aller vers plus de fiabilité et de disponibilité.

Au niveau des IHM, il semblerait que le mécanisme d'ancrage visuel proposé désoriente un peu trop les utilisateurs, notamment l'affichage de la zone graphique où apparaissent les cas

de recherches antérieures qui présentent une forte similarité avec la dernière action de recherche effectuée.

Un formalisme reste à trouver afin d'attirer l'attention de l'utilisateur sur les zones de l'écran qui sont sensés l'aider à améliorer sa recherche d'information.

Les valeurs des paramètres permettant de régler la sensibilité du système d'aide à la recherche d'information restent difficiles à déterminer et semblent tenir compte des compétences des utilisateurs mais aussi de la nature du fonds documentaire et des pratiques de recherche d'information des différents acteurs.

Enfin, un des points critiques du système proposé réside dans l'auto-évaluation du niveau de satisfaction de l'acteur/utilisateur en quête d'information. Non seulement, chaque acteur éprouve énormément de mal à déterminer le niveau adéquat mais il trouve que ce dispositif vient alourdir le processus d'interrogation lorsqu'il effectue un nombre limité d'actions de recherche et qu'il trouve l'information qu'il désire avant que l'effet "sérendipité" ne se soit manifesté.

Un important travail autour de la question de l'évaluation du niveau de satisfaction d'un acteur reste à faire afin d'améliorer l'efficacité du système proposé.

Au terme de cette phase d'expérimentation, il apparaît clairement qu'elle doit se poursuivre bien au-delà de la mise en production de la nouvelle version de l'Intranet et que des indicateurs de performance, de pertinence et de qualité doivent être déterminés afin de mieux mesurer les critères à prendre en considération afin de faire évoluer le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons.

Pour pouvoir valider le système d'aide à la recherche d'information proposé, il serait bon d'élargir cette expérimentation à d'autres contextes professionnels et à des types de GED différents.

Conclusion

Tout au long de cette thèse qui se veut avant tout le fruit de notre réflexion et le résultat de nos différents travaux de recherche, nous avons tenté de comprendre la façon dont un acteur humain aborde un système de Gestion Electronique de Documents et "rencontre" l'information.

Que cette rencontre résulte d'une démarche consciente faisant appel à des méthodes et des techniques de recherche d'information existantes ou qu'elle soit issue d'un mécanisme de contournement du mental conscient, notre objectif a toujours été d'améliorer le dialogue entre l'homme et la machine afin d'optimiser le processus de recherche d'une information pertinente et de qualité au sein d'un système GED.

Véritable "mémoire organisationnelle", la GED possède une dimension stratégique dans la mesure où elle offre un accès privilégié à l'ensemble des documents de référence d'une organisation. Cependant, l'interrogation d'un système de GED obéit toujours à une stratégie de recherche qui emprunte ses principaux modes de raisonnement aux techniques actuelles et rencontre de ce fait un certain nombre de limites.

La démarche que nous proposons apporte une contribution à la résolution des problèmes de recherche d'information dans les systèmes de GED, et ce quelque soit la nature de ces systèmes, leur finalité et le contexte organisationnel dans lesquels ils sont déployés.

Pour bien comprendre comment évolue la problématique d'indexation et de recherche d'information dans le domaine de la GED, nous avons tout d'abord mené un état de l'art des tendances actuelles.

Puis, dans un deuxième temps, nous nous sommes penchés sur les limites des techniques actuelles pour appréhender à travers une approche dite "tridimensionnelle" la phase de conception et de mise en œuvre d'une solution de GED qui puisse intégrer un système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences.

Cette approche nous a permis d'explorer plusieurs dimensions, notamment la dimension humaine qui est au cœur de la performance des organisations, la dimension activité qui s'intéresse à l'ensemble des processus qu'ils soient opérationnels, de support ou de direction et la dimension technologique qui se préoccupe d'avantage de l'ensemble des outils et systèmes qui sous-tendent l'activité d'une organisation.

Cette approche complétée par d'autres apports théoriques nous a ensuite permis de formuler une proposition de système d'aide à la recherche d'information dans le domaine de la GED qui accorde une place importante à la compréhension du comportement humain.

Durant la phase d'expérimentation, qui malheureusement n'a pas pu aller jusqu'à la mise en production de la nouvelle version de l'Intranet de la société HEXAGRAM, nous nous sommes surtout intéressés aux aspects techniques et à l'opérationnalité des concepts théoriques qui fondent le système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous avons proposé.

Malgré tout, cela nous a quand même permis de mettre en lumière des problèmes de performance qui ont nécessité de revoir le mode de fonctionnement de l'Intranet et de conception de l'interface graphique pour que la notion d'ancrage qui régit le mécanisme de sérendipité puisse jouer pleinement son rôle.

Nous n'avons malheureusement pas pu évaluer les pratiques et usages liés à ce type de système en fonction des différentes catégories d'acteurs dans la mesure où la mise en production de l'Intranet n'a pas encore eu lieu. Toutefois, nous avons quand même pu bénéficier d'un retour de la part des utilisateurs clés qui ont été désignés pour tester le fonctionnement de la nouvelle version de l'Intranet et dégager ainsi deux axes d'amélioration du système proposé.

Le premier axe d'amélioration repose sur la manière d'évaluer le niveau de satisfaction des utilisateurs du système qui ne doit en aucun cas être laissé à l'appréciation de l'utilisateur mais provenir de la prise en compte d'un certain nombre de facteurs qu'il nous faudra définir.

Le deuxième axe repose sur la juste expression des besoins informationnels et la prise en compte des référentiels sémantiques des différents acteurs humains au sein d'une organisation. Car, le simple fait d'accorder aux utilisateurs la responsabilité de l'expression de leurs besoins informationnels sans valider la justesse de cette expression engendre un problème de pertinence du système qu'il nous faudra traiter.

Cette problématique faisant également l'objet de travaux de recherche au sein de l'équipe SICOMOR, un rapprochement avec les enseignants-chercheurs travaillant sur ces problématiques est d'ores et déjà engagé.

Le mécanisme de contournement du mental conscient qui repose sur l'évaluation du niveau de satisfaction d'un utilisateur suite à une action de recherche et l'usage d'une ou plusieurs ancrés se manifestant après plusieurs actions de recherche jugées infructueuses constitue la pierre angulaire du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous proposons.

L'utilisation du raisonnement par cas qui permet de mémoriser des expériences de recherche antérieures ayant permis une augmentation significative du niveau de satisfaction de l'utilisateur constitue un apport certain en matière de recherche d'information dans le domaine de la GED.

Ce mode de raisonnement apporte une plus grande souplesse dans le mécanisme d'interrogation et contribue à améliorer la pertinence du système de recherche d'information.

Cette approche constitue une démarche complémentaire aux différents modèles ayant cours traditionnellement dans le domaine de la GED. Elle souligne également la nécessité de prendre en considération la logique d'interrogation de l'utilisateur et la façon dont il formule sa demande d'information.

Notre approche se voulant avant tout pragmatique, l'expérimentation et la validation du système d'aide à la recherche d'information par sérendipité et capitalisation d'expériences que nous avons conçu devrait se poursuivre au sein de la société HEXAGRAM mais également s'élargir à d'autres contextes professionnels et à des systèmes de GED hétérogènes.

Se trouvant au carrefour des sciences de l'information et de la communication, des sciences cognitives, de l'informatique et des sciences du management, cette thématique nous a permis d'enrichir notre vision des mécanismes d'indexation et de recherche d'information dans le domaine de la GED et nous a ouvert des perspectives de recherche complémentaires qui mettent au centre de leurs préoccupations la dimension humaine des projets d'entreprise.

Bibliographie

- AAMODT A., PLAZA E. – Case-based reasoning: foundational issues, methodological variations, and system approaches – AI Communications – IOS Press – Vol. 7: 1 – 1994.
- ARGYRIS C. – Savoir pour agir – Paris - InterEditions – 1995.
- ASSIE S. - Le diagnostic des systèmes d'information - In Le Diagnostic d'Entreprise, Méthode et processus - ouvrage collectif IAE Lyon - sous la direction d'Alain MARION – ECONOMICA - 1999.
- ASSOCIATION FRANCAISE DE NORMALISATION (AFNOR) - Vocabulaire de la documentation. 2ième édition - Paris La Défense – Afnor – 1987.
- BADDELEY A. - Essential of Human Memory - Psychology Press – Hove - 1999.
- BANDLER R. et GRINDER J. - Les secrets de la communication - Les Editions de l'Homme – Collection Alter ego – 2005.
- BANDLER R., ANDREAS C., ANDREAS S. et DE SAINT PAUL J. - Un cerveau pour changer , La programmation neuro-linguistique – InterEditions – 2000.

- BATESON G. - Step to an Ecology of Mind - Londres – Paladin – 1973.
- BELKIN N. J., VICKERY A. - Interaction in Information Systems: A Review of Research from Document Retrieval to Knowledge Based Systems – London - The British Library Board – 1985.
- BESSAI F.Z., NASRI A., OULAD NAOUI S. - Les réseaux de neurones appliqués à la recherche documentaire – CERIST - vol. 7, no1 – Alger – 1997.
- BODIN B., ROUX-FOUILLET JP. – La gestion électronique de documents – Paris – Dunod – 1992.
- BRADFORD S.C. – Documentation – London - Crosby Lockwood - 1948.
- BOURIGAULT D., JACQUEMIN C. - Construction de ressources terminologiques - Ingénierie des Langues - J.M. Pierrel Ed. – Paris - Hermes - 2000.
- BOUZIDI L. – Systèmes d'aide à l'accès aux connaissances : apprentissage, décision et recherche d'information – HDR – 2001.
- CACALY S., LE COADIC YF., SUTTER E., et POMART PD. – Dictionnaire de l'information – 3^{ième} édition – Collection Dictionnaire - Armand Colin – 2008.
- CAYROL A. et DE SAINT PAUL J. - Derrière la magie : La Programmation Neuro-Linguistique – InterEditions – 2005.
- CERI (Centre pour la Recherche et l'Innovation dans l'enseignement) - Comprendre le cerveau : Naissance d'une science de l'apprentissage – Volume 2007 - Numéro 5 – Editions OCDE – 2007.
- CHAUMIER J. – La gestion électronique de documents - Que sais-je ? – Paris - P.U.F. – 1996.
- CHERICI C., DUPONT JC. - Les Querelles du cerveau. Comment furent inventées les neurosciences – Vuibert - Paris, 2008.
- COLLOC J., BOUZIDI L. – A case based reasoning support system for use in medicine – UPGRADE Vol II, n°1 – 2001.
- DALBIN S. - Instruments de recherche sur le web - In La Recherche d'information sur les réseaux - Cours INRIA - 30 septembre-4 octobre 2002 - Le Bono (Morbihan) – Paris - ADBS Editions – 2002.

- DILTS R. - Roots of Neuro-Linguistic Programming - Meta Publications – Capitola - Etats-Unis - 1983.
- DONNADIEU G., KARSKY M. - La systémique : penser et agir dans la complexité – Liaisons – 2002.
- ERTZSCHEID O. et GALLEZOT G. - Chercher faux et trouver juste : Serendipity et recherche d'information - Bucarest - CIFSIC – 2003.
- ERTZSCHEID O. et GALLEZOT G. et BOUTIN E – Perspectives documentaires sur les moteurs de recherche : Entre Sérénipité et logiques marchandes - In L'Entonnoir - C&F éditions - <http://archivesic.ccsd.cnrs.fr/docs/00/17/21/69/PDF/ertzsgallbout.pdf> – 2009.
- FROCHOT D. - Information, documentation et veille juridique – Paris - Territorial Editions – 2007.
- GUPTA U.G. - How Case-Based Reasoning Solves New Problems – Interfaces - Volume 24 - 1994.
- HAAV H.M., LUBI T.L. - A Survey of Concept-based Information Retrieval Tools on the Web - Proceedings of the 5th East-European Conference ADBIS - Vol 2 - 2001.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO) - Information et documentation - Indicateurs de performance des bibliothèques – Edition 2 - ISO – 2008.
- JACQUEMIN C. - Spotting and Discovering Terms through Natural Language Processing - MIT Press – Cambridge - ÉtatsUnis - 2001.
- JÄRVELIN K., INGWERSEN P. - Information seeking research needs extension towards tasks and technology - Information Research - Volume 10, Paper 212 – 2004.
- JULIEN H.E. - Barrier's to adolescents' information seeking for career decision making - Journal of the American Society for Information Science - Volume 50 - 1999.
- JUNG C.G. – Essai d'exploration de l'inconscient – Gallimard – Collection Folio Essais – Numéro 90 - 1988.

- KORFHAGE R.R. - Information storage et retrieval, Wiley computer publishing - ISBN 978-0-471-14338-3 - 1997.
- KWOK K.L. - A network approach to probabilistic information retrieval - ACM Trans. Inf. Syst. - vol. 13, n° 3 - 1995.
- LE COADIC Y. - Le besoin d'information : formulation, négociation, diagnostic – Paris - ADBS - 1998.
- LE COADIC Y. – Usages et usagers de l'information – Collection 128 - Nathan Université – 2001.
- LEFEVRE P. - La recherche d'information, du texte intégral au thésaurus - Paris - Hermès – 2000.
- LE LOARER P. - Indexation automatique, recherche d'information et évaluation - In Le Traitement électronique du document - Cours INRIA - Aix-en-Provence - 3-7 octobre 1994 - Paris – ADBS – 1994.
- LEVY P. – L'intelligence collective, vers une anthropologie du cyberspace – Collection Poche, numéro 27 – Editions La Découverte – 1997.
- LUHN HP. - A Statistical Approach to Mechanized Encoding and Searching of Literary Information - IBM Journal of Research and Development - vol. 52 n° 12 - 1957.
- MANIEZ J. - Des classifications aux thésaurus : du bon usage des facettes – Documentaliste - Sciences de l'information - vol.36 – 1999.
- MANIEZ J. - Actualité des langages documentaires ; Fondements théoriques de la recherche d'information – ADBS – Paris - 2002.
- MARON M.E. et KUHNS J.L. - On Relevance, Probabilistic Indexing and Information Retrieval - Journal of the ACM (JACM) - Volume 7, Issue 3 – 1960.
- MEADOW C.T., BOYCE B.R. et KRAFT D.H. - Text Information Retrieval Systems - Academic Press - New York City - 2000.
- MERTON R. - The Serendipity Pattern - In Sociological Theory - In American Journal of Sociology - Numéro 50 – 1945.
- MIHALCEA R., MOLDOVAN D.I. - Semantic Indexing using WordNet Senses -Proceedings of ACL Workshop on IR & NLP - 2000.

- MUSTAFA EL HADI W. - Terminologie et accès à l'information – Paris – Lavoisier - 2006.
- NIE JY. - Introduction. Le domaine de la recherche d'information, survol d'une longue histoire dans GAUSSIER E., STEFANINI MH. - Assistance intelligente à la recherche d'information - Collection Traité des sciences et techniques de l'information – Paris – Lavoisier - 2003.
- NORMIER B. - L'apport des technologies linguistiques au traitement et à la valorisation de l'information textuelle - Éditions ADBS – 2007.
- OCLC Inc. - Dewey Decimal Classification – 14ième édition abrégée – ASTED -2005.
- PERRIAULT J. - Effet diligence, effet serendip et autres défis pour les sciences de l'information - <http://www.limsi.fr/Individu/turner/DCP/Paris2000/Perriault.pdf> - 2000.
- PICARD D. et MARC E. - Relations et communications interpersonnelles - Traité scientifique - DUNOD – 2000.
- PRAX JY. - La gestion électronique de documents, manager les flux d'information dans l'entreprise – 2^{ième} édition –Paris – InterEditions – 1998.
- PRAX JY., LARCHER S. - La gestion électronique documentaire – 3^{ième} édition – Paris – Collection InfoPro – DUNOD – 2004.
- PRAX JY. – Le manuel du Knowledge Management, une approche de 2^{ième} génération – Paris – DUNOD – 2003.
- RANGANATHAN S.R. - The five laws of library science - Asia publishing house – 1963.
- REMER T.G. - Serendipity and the Three Princes: From the Peregrinaggio of 1557 - University of Oklahoma Press – 1965.
- RIVIER A. - Aide-memoire d'informatique documentaire – Collection Bibliothèques - Editions du Cercle de La Librairie – 2007.
- ROBERTSON S.E. - The probability ranking principle in IR - Journal of Documentation - 1977.
- ROBBINS A. - Pouvoir illimité – Robert Laffont – 2003.

- ROUET JF. - Les activités documentaires complexes. Aspects cognitifs et développementaux - Mémoire pour l'habilitation à diriger les recherches - Université de Poitiers - France - 2000.
- SALTON G. - The SMART retrieval system: Experiments in automatic document processing - Prentice Hall - 1970.
- SALTON G. et MC GILL M. - Introduction to modern information retrieval – McGraw Hill Int. Book Co - 1984.
- SAURET MJ. – Freud et l'inconscient – Editions Milan – Collection Essentiels – Numéro 125 – 1999.
- SCHIFF JL. et SCHIFF A. - Le cadre de référence - Les Classiques de l'analyse transactionnelle – Volume 2 - Editions d'Analyse Transactionnelle – 2008.
- SCHLICK M. - Théorie générale de la connaissance - Éditions Gallimard – Collection Bibliothèque de philosophie – Paris – 2009.
- SCHUSTER F.E., MORDEN D.L., BAKER T., MC DUNNING K. et HAGAN C. - Management Practice, Organization Climate and Performance : An Exploratory Study - The Journal of Applied Behavioral Science - Volume 33 – Numéro 2 - 1997.
- SERRES A. - Evaluation de l'information sur Internet - Rennes – URFIST - 2002-2004.
- SIMON HA. - The new science of managerial decision – Ed. Prentice-Hall – New Jersey – 1977.
- SOLER C. – Lacan, l'inconscient réinventé – PUF – 2009.
- SUMMONS P., GIBBON G., VILLANUEVA H. and FALCONER J. – Causal reasoning in a medical k.
- SUN TZU – L'art de la guerre – Collection Champs, numéro 58 – Flammarion – 2001.
- TAYLOR R.S. - Value Added Processes in Information Systems – Norwood - NJ : Ablex Publishing. 1986.
- THOMAS B., MARINI JL., CRETIN R., - La dimension humaine : paramètre essentiel de la performance des organisations - Colloque "A la Recherche de la Performance" - ESC Bretagne Brest – 2007.

- TODOROV T. - Mémoire du mal, tentation du bien, Enquête sur le siècle – Paris - Robert Laffont -2000.
- TOMS E.G. - Serendipitous Information Retrieval - http://www.ercim.eu/publication/ws-proceedings/DelNoe01/3_Toms.pdf - 2000.
- TOUSSAINT Y., ROYAUTE J., MULLER C., POLANCO X. - Analyse linguistique et infométrie pour l'acquisition et la structuration de connaissances - Deuxièmes Rencontres Terminologie et Intelligence Artificielle - Université Toulouse-Le Mirail - 3-4 avril 1997.
- TRICOT A. et ROUET JF. - Les hypermédias, approches cognitives et ergonomiques – Paris – Hermès - 1998.
- TRICOT A. - Apprentissage et recherche d'information avec les documents électroniques - Mémoire en vue de l'habilitation à diriger les recherches, Toulouse – 2003.
- TURTLE H., CROFT W.B. - Inference networks for document retrieval - Proceedings of ACM-SIGIR'90 - ACM Press - 1990.
- UDC Consortium - Classification décimale universelle : édition moyenne internationale - Edition du Céfal – Liège – 2004.
- VAN ANDEL P. et BOURCIER D. - De la sérendipité dans la science, la technique, l'art et le droit - Leçons de l'inattendu - Edition L'Act Mem – 2009.
- ZARA O. – Le management de l'intelligence collective, vers une nouvelle gouvernance – 2ième édition - Collection Essais Documents – M21 Editions – 2008.

Annexes

Annexe1 : Glossaire de termes

Base de connaissances :

Base structurée comprenant des connaissances spécifiques à un domaine donné, des documents, données, programmes, procédures, formules, schémas, plans, vidéos, sous une forme exploitable par un ordinateur dont le but est de regrouper thématiquement les connaissances des collaborateurs d'une organisation.

Bruit :

Dans le domaine documentaire, ce terme désigne l'inverse de la pertinence ou du silence, c'est-à-dire des informations qui n'ont aucun rapport avec le sujet d'intérêt ou de requête.

BPM (Business Process Management) :

Le BPM désigne la gestion des processus métier d'une organisation. Afin d'avoir une description précise de celle-ci, les gestionnaires ont la plupart du temps recours à la modélisation des processus, leur objectif étant de les améliorer pour en garantir la qualité et la performance.

Capture :

Processus d'acquisition et de stockage de l'information. Cette information est issue de processus de dématérialisation (numérisation), il peut aussi s'agir de documents électroniques d'origine bureautique, de télécopies, de flux d'impression, de flux EDI, de formulaires électroniques, etc. La capture d'information comporte généralement plusieurs étapes : lecture (capture proprement dite), avec parfois conversion de format, extraction automatique, semi-automatique ou manuelle des métadonnées.

Carte sémantique

La carte sémantique est un outil de construction et de représentation qui sert à élaborer, sous la forme d'un modèle, un réseau d'idées, d'informations ou de connaissances dans un domaine donné. Ce réseau indique les liens qui existent entre ses constituantes.

Classification par facettes :

Ce type de classification permet de pallier les limites des classifications traditionnelles, hiérarchiques ou arborescentes. Ainsi, l'espace de classement est vu comme un espace multidimensionnel, chacune des dimensions (facettes) correspondant à une métadonnée. Pour faciliter le classement et éviter les ambiguïtés, les facettes doivent être mutuellement exclusives.

CMS (Content Management System) :

Système intégré permettant aux entreprises et aux organisations de gérer rapidement et facilement le contenu dynamique et rédactionnel d'un site Web.

Collaboration :

Interactions multiples entre deux ou plusieurs individus. A distinguer de la "communication" (un individu A envoie un message à un individu B) et de "l'interaction" (un individu A envoie un message à un individu B qui lui répond par retour de message).

COLD (Computer Output to Laser Disk) :

Cet acronyme recouvre le stockage des flux d'impression sur des supports de type disque, avec quelques fonctions avancées de recherche, d'extraction et de mise en forme. Aujourd'hui le terme COLD est surtout employé dans le sens "Stockage de spool" avec peu de valeur ajoutée.

Communautés :

Communautés qui regroupent des individus partageant des centres d'intérêts communs, coopérant et échangeant leurs savoirs.

Communication :

Fait d'établir une relation avec quelqu'un ou quelque chose. Cette relation permet à un émetteur de transmettre un message à un récepteur par le biais d'un canal, ou média.

Concept mapping :

Technique de représentation graphique dans laquelle les concepts sont liés entre eux par des liens pour former un réseau. On élabore alors des "cartes conceptuelles" qui permettent de visualiser et de comprendre l'ensemble des interactions entre ces concepts. On formalise ainsi un savoir.

Connaissance / Information / Donnée :

Modèle hiérarchique de la connaissance : avec en bas, le plus basique, les données (éléments bruts en-dehors de tout contexte), puis l'intermédiaire : l'information (données mises en

contexte) ; et enfin, au sommet de la hiérarchie : la connaissance (l'information assimilée pour réaliser une action).

Contenu :

Qui est dans un contenant. Qui est exprimé, signifié par un document. En informatique, terme proche de document, qui désigne l'information non structurée (opposée à données), quel qu'en soit le support ou la source : courriel, message instantané, fichier électronique, document textuel, graphique, image fixe ou animée, plan, flux, etc.

Dématérialisation :

Transformation de supports d'informations matériels (en général papier) en fichiers informatiques.

Donnée :

Voir [Connaissance / Information / Donnée / Métadonnée](#).

ECM (Enterprise Content Management) :

La Gestion des contenus d'entreprise est un terme issu des technologies de l'information, désignant les méthodes et les outils utilisés pour acquérir, gérer, stocker, archiver et diffuser les contenus liés aux processus organisationnels de l'entreprise.

ERM :

L'ERM couvre la capacité d'un système de gestion de contenu à prendre en compte des flux documentaires issus le plus souvent des chaînes éditiques.

Fonds documentaire :

A l'origine, ensemble des documents de toute nature qu'une personne physique ou morale a automatiquement et organiquement réunis dans l'exercice de ses activités et conservés en vue d'une utilisation éventuelle. Aujourd'hui ce terme recouvre l'ensemble du contenu, quel qu'il soit, géré par un système basé sur l'[ECM](#).

Full Web :

Qualifie les interfaces qui ne demandent aucune installation sur le poste de travail, permettant ainsi leur déploiement sans contrainte.

GED (Gestion Electronique de Documents) :

La Gestion Electronique de Document est un terme issu des technologies de l'information, désignant les méthodes et les outils électroniques utilisés pour acquérir, gérer, stocker, archiver et diffuser les documents.

GEIDE (Gestion Électronique de Documents et d'Informations Existants) :

Cette modernisation de l'acronyme GED vise à renforcer l'association entre la notion de gestion d'information et celle de gestion documentaire.

Gestion de la connaissance :

Voir [Knowledge Management](#).

Heuristique :

Du grec "heuriskein" qui signifie "trouver", ce terme de didactique signifie l'art d'inventer, de faire des découvertes. Il a une utilité dans la recherche, notamment la recherche scientifique et épistémologique.

ICR (Intelligent Character Recognition) :

Technologies logicielles permettant la reconnaissance de l'écriture manuscrite, élargissant ainsi les techniques [OCR](#).

ILM (Information Lifecycle Management) :

L'ILM est une méthodologie de gestion de l'information depuis sa création jusqu'à son éventuelle destruction. Le Records Management est aujourd'hui la forme la plus aboutie de l'ILM.

Index :

Valeur fixe (chaîne de caractères) permettant d'obtenir l'adresse de l'objet indexé. En matière d'ECM, les métadonnées sont généralement utilisées en tant qu'index. Voir également "texte intégral", où chaque mot d'un texte devient lui-même une valeur d'index.

Indexation :

L'indexation consiste à saisir une ou plusieurs informations relatives à un document afin de la caractériser et d'en faciliter la recherche et la consultation. Ces informations sont communément appelées index, propriétés, métadonnées. L'indexation peut être une opération manuelle, semi-automatique ou automatique. Elle peut être automatique ou semi-automatique avec l'aide de la reconnaissance de caractères.

KM (Knowledge Management) :

[ECM](#) appliqué au champ de la gestion des connaissances, où les contributions collaboratives enrichissent une base de connaissances.

LAD (Lecture Automatique de Documents) :

Ensemble des moyens logiciels et matériels qui permettent d'extraire, à partir d'informations plus ou moins structurées, les données et métadonnées qui alimenteront un système de gestion, par exemple d'[ECM](#). La LAD s'appuie essentiellement sur les technologies d'[OCR](#) et [ICR](#), elle est souvent couplée à des bases externes pour compenser les erreurs de lecture.

Métadonnée :

Donnée servant à décrire, à définir, voire à compléter une autre donnée. Donnée qui renseigne sur la nature d'une autre donnée et en permet son utilisation pertinente. Dans les technologies de l'information, les métadonnées sont généralement des données structurées qui décrivent des informations peu ou pas structurées et en permettent leur exploitation.

OCR (Optical Character Recognition) :

La reconnaissance optique de caractères est une technologie logicielle permettant la reconnaissance de texte imprimé.

OMR (Optical Mark Recognition) :

Technique mise en œuvre par les périphériques spécialisés de reconnaissance de marques par capteurs optiques. Encore très utilisés, les lecteurs optiques de marques sont extrêmement rapides en acquisition de données mais ne permettent pas la lecture de caractères imprimés ou pré-casés.

Ontologie :

Création d'un vocabulaire unifié pour éviter les ambiguïtés terminologiques.

Plan de classement :

Structure (linéaire, hiérarchique, multidimensionnelle, etc.) qui décrit l'organisation du contenu.

Portail :

Page d'accueil d'un site Web présentant de nombreuses informations et comportant un grand nombre de liens vers différentes parties du site ou vers d'autres sites. Le terme a dérivé en même temps que les apports techniques vers la désignation d'une structuration particulière de site Web permettant d'agréger différents services, contenus et accès de sources variées.

RAD (Reconnaissance Automatique de Documents) :

Souvent couplée à la [LAD](#) (Lecture Automatique de Document). Ensemble de moyens matériels et logiciels qui permettent, par l'analyse du contenu de l'information numérisée, sa reconnaissance et donc sa facilité de traitement ultérieurs. (La génération de métadonnées est un des processus de la RAD).

RM (Records Management) :

Champ de l'organisation et de la gestion en charge d'un contrôle efficace et systématique de la création, de la réception, de la conservation, de l'utilisation et du sort final des documents, y compris des méthodes de fixation et de préservation de la preuve et de l'information liée à la forme des documents. Le RM couvre donc le cycle de vie complet de l'information, quelle qu'elle soit, depuis sa création jusqu'à son archivage.

Text mining :

Technique permettant d'automatiser le traitement de gros volumes de données textuelles, visant à analyser et découvrir des connaissances et des relations à partir des documents disponibles.

Texte intégral :

Technique d'indexation de documents de nature textuelle ou de champs textes de base de données, où chaque mot composant le texte est un index pouvant servir de clé de recherche lors d'une interrogation. On associe généralement un dictionnaire de mots vides, ceux-ci étant des mots qui ne seront jamais indexés car apparaissant de manière trop systématique dans les textes, et donc ne présentant pas de pertinence lors de l'interrogation (les prépositions, les pronoms, etc.). Lors d'une interrogation, les mots utilisés comme clés de recherche peuvent en général être combinés entre eux par des opérateurs logiques de recherche tels que le "ET", le "OU" et le "SAUF", des opérateurs de proximité entre mots (permettant par exemple de rechercher deux mots appartenant à la même phrase), et des opérateurs de troncatures (permettant de rechercher des parties de mots).

Thésaurus :

Dictionnaire de descripteurs linguistiques dans lequel chaque terme est relié sémantiquement.

Web sémantique :

Ensemble de concepts et technologies tendant à rendre le contenu des ressources disponibles sur le web accessible et utilisable par des logiciels, notamment avec des métadonnées formalisées. Web "intelligent" dans lequel les informations, auxquelles on donne une signification bien définie, sont reliées entre elles de façon à ce qu'elles soient comprises par les ordinateurs (métadonnées et liens), dans le but de transformer la masse des pages Web en un index hiérarchisé.

Workflow :

Littéralement traduisible de l'Anglais par "flux de travail". Ce terme désigne la modélisation et le suivi des tâches à accomplir : procédures organisationnelles, historiques des actions et échanges d'informations entre les personnes impliquées (commentaires, notes). Un Workflow intégré à une GED /GEIDE permet de formaliser et de systématiser les procédures de travail autour des documents.

Annexe 2 : Article de recherche - Colloque

International sur l'information numérique et les enjeux de la société de l'information – Tunis - TUNISIE

Titre du document :

Capitalisation d'expériences pour l'indexation et la recherche d'informations numériques dans le domaine de la GED.

Auteur(s) :

Laïd BOUZIDI et Jean-Luc MARINI.

Résumé :

L'intégration de systèmes hétérogènes et le couplage de différentes techniques en matière de stockage et de recherche de l'information numérique devient une nécessité vu les volumes importants qui ne cessent d'augmenter. Parmi les techniques utilisées pour la gestion des flux d'information et l'optimisation des processus de diffusion de documents numériques ou de sources d'information, la Gestion Electronique de Documents (GED) représente une solution adaptée et utilisée dans différents contextes et domaines d'application. Mais en l'absence de techniques fiables de recherche d'information numérique, le problème de la pertinence des informations obtenues et de la qualité des informations sélectionnées dépendent non seulement des méthodes d'indexation et d'acquisition des informations mais aussi des techniques d'interrogation choisies. Le degré de pertinence d'une technique d'indexation et de recherche d'information dépend en premier lieu de l'utilisateur, de son profil, de sa démarche de recherche d'information et du domaine sur lequel il recherche des informations. La démarche que nous proposons intègre, par l'approche GED, les informations sources

intégralement puisque la technique utilisée dans les systèmes de GED consiste à numériser les sources brutes d'informations puis de les indexer, mais aussi le savoir et le savoir-faire des utilisateurs par la capitalisation des expériences de ces derniers en matière d'indexation et de recherche d'information. Pour cela nous proposons l'utilisation du raisonnement par cas RpC qui permet une aide à la recherche d'information en utilisant les recherches antérieures jugées similaires. Les usages et les pratiques en matière d'indexation et de recherche de l'information numérique ne cessent d'évoluer et de se spécialiser par domaine certes mais aussi par l'identification qualitative des usagers. Nous proposons une démarche pouvant aider à maîtriser et à résoudre en partie cette problématique.

Revue / Journal :

Revue maghrébine de documentation.

Editeur :

Université de Tunis - Institut supérieur de documentation – Tunis – TUNISIE (Revue).

Colloque :

L'information numérique et les enjeux de la société de l'information - Tunis - TUNISIE (14/04/2005).

Langue(s) :

Communication : Français.

Revue : Français et Arabe.

Mots-clés :

Recherche d'information, information numérique, GED, aide à la décision, raisonnement par cas, apprentissage.

Annexe 3 : Article de recherche - Colloque International TICE Méditerranée sur l'humain dans la formation à distance et la problématique du changement – Marseille - FRANCE

Titre du document :

Mise en place de l'Environnement Numérique de Travail au sein de l'Université Jean Moulin Lyon 3 : l'Enseignant au cœur d'une dynamique de changement.

Auteur(s) :

Raphaëlle CRETIN, Laïd BOUZIDI et Jean-Luc MARINI.

Résumé :

A l'instar de toutes les organisations, les institutions universitaires ont intégré les TIC dans l'ensemble de l'activité d'enseignement, de recherche et de gouvernance. Nous nous intéressons particulièrement à l'introduction des ENT en nous basant sur une approche tridimensionnelle. La prise en compte de la dimension « humaine » est fortement corrélée au succès ou à l'échec du développement et de la pérennisation d'un ENT. Nous expliciterons la démarche de conduite du changement que nous avons mené afin de favoriser l'adhésion de l'ensemble des acteurs de l'université : enseignants-chercheurs, étudiants, administratifs et décideurs.

Colloque :

L'humain dans la formation à distance et la problématique du changement – TICE Méditerranée – Edition 2007 - Marseille - FRANCE (02/06/2007).

Langue(s) :

Français.

Mots-clés :

Environnement Numérique de Travail, conduite du changement, Enseignement à Distance, espace informationnel.

Annexe 4 : Article de recherche – Colloque sur la recherche de la performance : un concept de gestionnaire essentiel à l'entreprise – Brest - FRANCE

Titre du document :

La dimension humaine : paramètre essentiel de la performance des organisations.

Auteur(s) :

Bernard THOMAS, Jean-Luc MARINI et Raphaëlle CRETIN.

Résumé :

Si on part du principe que la performance globale des organisations repose sur des aspects économique, social et sociétal, il n'en demeure pas moins que son appréciation repose essentiellement sur des indicateurs quantitatifs et qualitatifs. Lesquels traduisent le niveau de performance des principales fonctions des organisations et leur capacité à produire de la valeur ajoutée. Toutefois, seule la dimension humaine présente dans toutes les composantes d'une organisation permet d'évaluer et de potentialiser la notion de performance globale. C'est la raison pour laquelle notre approche repose essentiellement sur la prise en compte de la dimension humaine, seul paramètre capable de s'auto performer et de performer les autres composantes d'une organisation. Cependant, la dimension humaine est constituée d'un paramètre qui est l'imperfection, lequel s'exprime entre le dit et le non dit, entre le conscient et l'inconscient. Alors que la majeure partie des approches qui s'intéressent à la dimension humaine travaillent sur le conscient et le révélé, la méthode que nous nous proposons d'exposer permet d'interroger le non dit et l'inconscient. Nous nommons cette méthode, la méthode d'Amélioration et de Recherche de la Performance Globale en Entreprise (ARPEGE). Cette approche relève d'une approche systémique et se fonde sur l'identification de trois niveaux appelés des états qui sont l'état des lieux, l'état dissocié et l'état désiré. Dans l'article qui suit, nous définissons la notion générale de performance globale ainsi que le concept de dimension humaine à la base de la méthode ARPEGE. Nous présentons les fondements théoriques de notre méthode que nous appliquerons ensuite dans une étude de cas menée au sein de la COMmission Technique d'Orientation et de REclassement Professionnel (COTOREP) d'Arras.

Colloque :

A la recherche de la performance : un concept de gestionnaire essentiel à l'entreprise – ESC BRETAGNE BREST - Brest - FRANCE (01/02/2007).

Langue(s) :

Français.

Mots-clés :

Dimension humaine, performance globale, conscient et inconscient.

Résumé :

Dans un environnement économique hautement concurrentiel, l'entreprise qui veut rester compétitive doit sans cesse repenser son organisation, mobiliser ses compétences et revoir sa stratégie. Cela nécessite une plus grande coopération entre les acteurs, un renforcement de la communication ainsi qu'une capitalisation des connaissances et des processus d'action. Cette évolution suppose également la mise en place d'un dispositif de recherche d'information dans une mémoire d'entreprise hétérogène distribuée via le Web ou un Intranet. La constitution de ce patrimoine s'appuie généralement sur différentes sources d'information et nécessite une prise en compte du niveau d'expertise des différents acteurs. Dans ce contexte, la GED (Gestion Electronique de Documents) constitue un composant technologique essentiel de la capitalisation de la connaissance et de l'intelligence collective. Notre démarche consiste donc à modéliser et à concevoir un système de recherche d'information capable de rendre cette mémoire accessible indépendamment des acteurs qui l'ont créée et de manière pertinente par rapport à un contexte d'intérêt donné.

Mots clés : Recherche d'informations, Capitalisation des connaissances et des processus d'action, GED (Gestion Electronique de Documents), Sérendipité.

Abstract:

In a highly competitive economic environment, the company that wants to remain competitive must continually rethink its organization, mobilize its skills and revise its strategy. This requires greater cooperation among stakeholders, increased communication and an accumulation of knowledge and action processes. This development also implies the establishment of a mechanism for information retrieval in a heterogeneous corporate memory distributed via the Web or an Intranet. The constitution of this heritage is generally based on various sources of information and requires taking into account the level of expertise of different actors. In this context, the EDM (Electronic Document) is an essential technology component of the corporate knowledge and collective intelligence. Our approach is therefore to model and design a system to search for information that can make this memory accessible regardless of actors who have created and so relevant to the context of a given interest.

Keywords: Information Retrieval, Capitalization of knowledge and action processes, EDM (Electronic Document Management), Serendipity.