

BAnQ

BIBLIOTHÈQUE
ET ARCHIVES
NATIONALES
DU QUÉBEC

VIEUX-
MONTREAL

L'indexation automatisée à Bibliothèque et Archives nationales du Québec : un bilan d'expérience

Valérie D'Amour

Directrice de l'ouest du Québec

BAnQ Vieux-Montréal

*Bibliothèque
et Archives
nationales*

Québec 

Plan de conférence

- Les objectifs
- L'appel d'offres et le choix du fournisseur
- Les prémisses nécessaires à l'indexation automatisée
- Les étapes de configuration du logiciel
- Un aperçu
- Résultats atteints à ce jour
- Les limites de l'indexation automatisée
- Les prochaines étapes

Objectifs

- Pourquoi l'indexation automatisée?
 - Rendre plus rapidement accessibles des notices de description jusqu'ici non indexées par l'indexation en lot.
 - Fournir un support à l'indexation humaine: recherche et présentation automatisées d'un choix de termes.

L'appel d'offres et le choix du fournisseur

- Exigences fonctionnelles et non fonctionnelles
- Évaluation qualitative des exigences fonctionnelles : développement de tests pratiques et d'une grille d'évaluation
- Deux fournisseurs ont répondu à l'appel.
- Le fournisseur retenu : Luxid de TEMIS devenu Expert System – Semantic Intelligence

Les prémisses nécessaires à l'indexation automatisée

- Nettoyage du répertoire de termes dans Pistard
- Harmonisation des catégories de termes en fonction des codes du MARC21
- Autorisation pour utiliser d'autres répertoires
- Préparation d'une liste de termes à ne pas indexer
- Détermination d'un échantillon représentatif de notices descriptives

Les étapes de configuration du logiciel

- Intégration dans Luxid du répertoire Pistard et des notices descriptives
- Validation du premier plan d'annotation
- Multiples sessions de travail entre archiviste et programmeur
- Intégration progressive des autres répertoires et établissement des priorités
- Réduction du corpus de validation à 500 notices

Un aperçu



ACCUEIL

PROJET

DOCUMENTS

RESSOURCES

CLASSES

- Evenement
- Lieu
- Nom commun
- Personne morale
- Personne physique

FILTRES

RECHERCHER DANS NOM COMMUN

publicité

1-100 / 746

- publicité
- publicités
- publicité -- campagnes
- publicité -- efficacité
- publicité -- rendement
- publicité -- impression
- publicité artistique -- expositions
- publicité mensongère
- publicité déloyale
- publicité frauduleuse
- publicité immobilière
- publicité lumineuse
- publicité radiophonique
- publicité pharmaceutique
- publicité politique
- publicité sexiste
- publicité touristique
- publicité télévisée
- publicité électorale
- publicité électronique
- publicité ambulante
- publicité commerciale
- publicité fausse
- publicité illégitime

PUBLICITÉ -- CAMPAGNES

GÉNÉRAL

IDENTIFIANT
<http://www.banq.pistard/619070>

LIBELLÉ PRÉFÉRENTIEL
publicité -- campagnes

LIBELLÉS ALTERNATIFS
Publicité

EXTRACTION

NE PAS EXTRAIRE

MÉTHODE D'EXTRACTION

- Même forme
- Même forme et casse
- Étendue




CODE_TERME

Un aperçu

Luxid
WEBSTUDIO

ACCUEIL **PROJET** **DOCUMENTS**





RESSOURCES

CLASSES   

- Evenement
- Lieu
- Nom commun
- Personne morale
- Personne physique

FILTRES




RECHERCHER DANS NOM COMMUN

< 1-100 / 746 >

- publicité
- publicités
- publicité -- campagnes**
- publicité -- efficacité
- publicité -- rendement
- publicité -- impression
- publicité artistique -- expositions
- publicité mensongère
- publicité déloyale
- publicité frauduleuse
- publicité immobilière
- publicité lumineuse
- publicité radiophonique
- publicité pharmaceutique
- publicité politique
- publicité sexiste
- publicité touristique
- publicité télévisée
- publicité électorale
- publicité électronique
- publicité ambulante
- publicité commerciale
- publicité fausse
- publicité illicite

PUBLICITÉ -- CAMPAGNES

▼ **EXTRACTION**

NE PAS EXTRAIRE

MÉTHODE D'EXTRACTION

- Même forme
- Même forme et casse
- Étendue

> **CODE_TERME**

▼ **IDENTIFIANT**

619070

> **IDENTIFIANT_ACCEPTE**

▼ **NOTES_TERME**

NOTES_TERME

Un aperçu



ACCUEIL PROJET DOCUMENTS

RESSOURCES

CLASSES

- Evenement
- Lieu
- Nom commun
- Personne morale
- Personne physique

FILTRES

RECHERCHER DANS NOM COMMUN

publicité

1-100 / 746

- publicité
- publicités
- publicité -- campagnes**
- publicité -- efficacité
- publicité -- rendement
- publicité -- impression
- publicité artistique -- expositions
- publicité mensongère
- publicité déloyale
- publicité frauduleuse
- publicité immobilière
- publicité lumineuse
- publicité radiophonique
- publicité pharmaceutique
- publicité politique
- publicité sexiste
- publicité touristique
- publicité télévisée
- publicité électorale
- publicité électronique
- publicité ambulante
- publicité commerciale
- publicité fausse
- publicité directe

PUBLICITÉ -- CAMPAGNES

TERME ACCEPTE

VOIR AUSSI

VOIR AUSSI

EXTRACTION

TERMES DE CONTEXTE OBLIGATOIRES

Campagnes

PREFLABEL_ACCEPTE

PREFLABEL_ACCEPTE

619066_Campagnes publicitaires

PREFLABEL_VOIR AUSSI

Résultats atteints à ce jour

Taux de précision de 72,5 %

Taux de rappel de 67,5 %

Pour une moyenne de 70 %

Résultats atteints à ce jour

- Bonifications possibles :
 - Fonctionnalité « Ne pas extraire ». Ex. : « Ce dossier contient des coupures de presse et... »
 - Fonctionnalité « Formes interdites ». Ex. : « Ce dossier fait état de dépenses faites par feu Jean Gagné... »
 - Fonctionnalité « Termes de contexte obligatoires ». Ex. : « Cette photographie montre des pêcheurs au Lac Croche dans la municipalité de Colombier sur la Côte-Nord. »
 - Fonctionnalité « Libellés alternatifs ». Ex. : « Cette politique d'immigration... » le terme étant « Émigration et immigration »

Les limites de l'indexation automatisée

■ Non-respect de certains de nos principes d'indexation

Exemple de notice descriptive :

Procès-verbal de bornage, à la requête de Rock Juteau et à la suite du consentement de François Billet, tous deux cultivateurs de la côte Saint-Louis dans la paroisse de Sainte-Thérèse-de-Blainville, pour établir la ligne de division de leurs terrains situés dans la dite paroisse.

Difficile, voire impossible, de limiter l'indexation à l'essentiel, soit :

Bornage (Droit)

Chemin de la Côte-Saint-Louis (Blainville, Québec)

Les limites de l'indexation automatisée

- Marges d'erreur à prendre en compte
 - Indexation de mots vides. Ex. : « Au pied de »
 - Non-extraction de mots pertinents. Ex. : « Les lacs Provost, Chat et Monroe... »
 - Amalgame de concepts. Ex. : « Le dossier porte sur le projet de construction d'un immeuble dans le secteur du boulevard Newman. » Terme d'indexation : Boulevard – Conception et construction.
 - Difficultés avec les cas d'homonymie et de polysémie. Ex. : Pêche (Fruit) et Pêches ou Pierre Beaudry (1774-1848) et Pierre Beaudry (1884-1918).

Les prochaines étapes

- Mise en place d'un suivi avec le fournisseur afin d'assurer la stabilité de la chaîne d'annotation
- Développement du mécanisme de mises à jour des répertoires
- Bonification des règles d'application particulières
- Définition de règles de description pour faciliter l'indexation automatisée
- Intégration du logiciel dans la refonte de Pistard

Région de Montréal :
514 873-1100

Sans frais, d'ailleurs au Québec :
1 800 363-9028

banq.qc.ca

Montréal

Grande Bibliothèque

475, boulevard De Maisonneuve Est
Montréal (Québec) H2L 5C4

BAnQ Rosemont–La Petite-Patrie

2275, rue Holt
Montréal (Québec) H2G 3H1

BAnQ Vieux-Montréal

535, avenue Viger Est
Montréal (Québec) H2L 2P3

Capitale-Nationale

BAnQ Québec

Pavillon Louis-Jacques-Casault
Campus de l'Université Laval
1055, avenue du Séminaire
Québec (Québec) G1V 4N1

BAnQ Gaspé

80, boulevard de Gaspé
Gaspé (Québec) G4X 1A9

BAnQ Gatineau

855, boulevard de la Gappe
Gatineau (Québec) J8T 8H9

BAnQ Rimouski

337, rue Moreault
Rimouski (Québec) G5L 1P4

BAnQ Rouyn-Noranda

27, rue du Terminus Ouest
Rouyn-Noranda (Québec) J9X 2P3

BAnQ Saguenay

930, rue Jacques-Cartier Est, bureau C-103
Saguenay (Québec) G7H 7K9

BAnQ Sept-Îles

700, boulevard Laure, bureau 190
Sept-Îles (Québec) G4R 1Y1

BAnQ Sherbrooke

225, rue Frontenac, bureau 401
Sherbrooke (Québec) J1H 1K1

BAnQ Trois-Rivières

225, rue des Forges, bureau 208
Trois-Rivières (Québec) G9A 2G7

**Bibliothèque
et Archives
nationales**

Québec 