

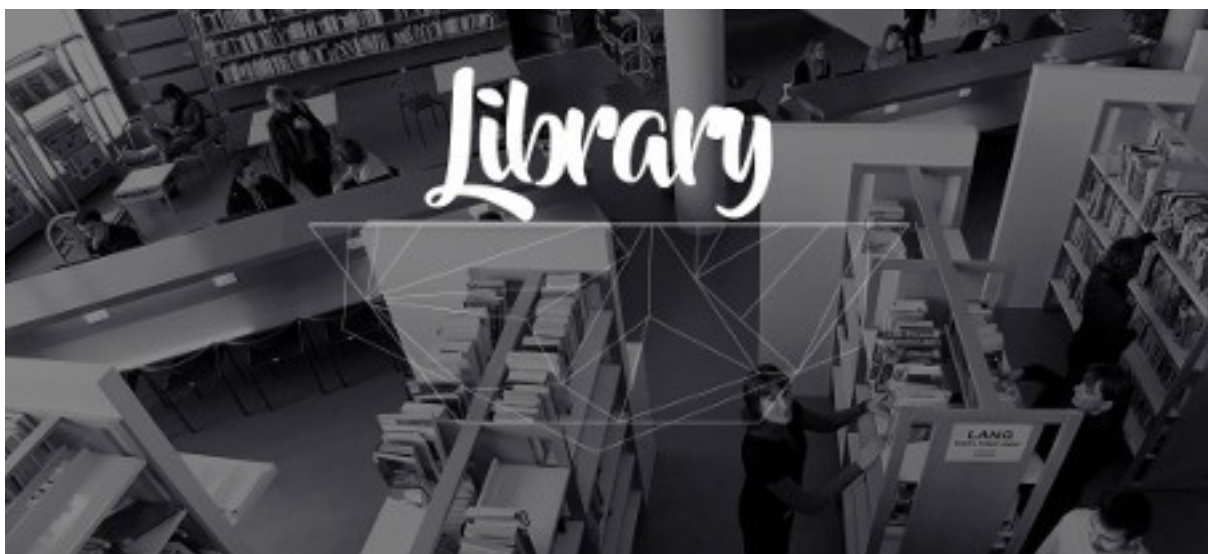
Faculté des Lettres et des Sciences Humaines

Licence Professionnelle

Métiers du livre : documentation et bibliothèques

2017-2018

Mise à jour d'un thésaurus documentaire



Julia POPEK

Stage effectué du 3 janvier au 30 mars 2018

Rapport de stage dirigé par

Anne MÉNER

Responsable des médiathèques

Audencia Business School



We do not realize what tremendous power the structure of an habitual language has. It is not an exaggeration to say that it enslaves us through the mechanism of s[emantic] r[eactions] and that the structure which a language exhibits, and impresses upon us unconsciously, is automatically projected upon the world around us.
– **Sciences and Sanity, Alfred Korzybski, 1933**

Remerciements

Je remercie l'ensemble de l'équipe de la médiathèque d'Audencia Business School qui m'a beaucoup appris sur le métier. Et plus particulièrement Anne Méner, qui m'a donné la chance de me reconverter vers la documentation.

Je remercie aussi toutes les personnes qui ont soutenu mon projet de reconversion : mes proches, l'équipe et les participants de Challenge Emploi ainsi que mes camarades de promotion en licence professionnelle.

Droits d'auteurs

Cette création est mise à disposition selon le Contrat :

« **Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France** »

disponible en ligne : <http://creativecommons.org/licenses/by-nc-nd/3.0/fr/>



Table des matières

| | |
|---|----|
| Introduction..... | 7 |
| 1. Présentation du projet..... | 8 |
| 1.1. Contexte du projet..... | 8 |
| 1.1.1. Audencia Business School..... | 8 |
| 1.1.2. Mise à jour du thésaurus documentaire..... | 8 |
| 1.2. Fonctionnement du thésaurus documentaire..... | 9 |
| 1.2.1. Définition..... | 9 |
| 1.2.2. Les 3 types de relations sémantiques..... | 9 |
| 1.2.2.1. Relation d'équivalence..... | 10 |
| 1.2.2.2. Relation hiérarchique..... | 10 |
| 1.2.2.3. Relation associative..... | 10 |
| 1.3. Étude du projet..... | 11 |
| 1.3.1. Thésaurus Delphes..... | 11 |
| 1.3.2. Thésaurus du Management..... | 11 |
| 1.3.3. Comparaison des deux thésaurus..... | 11 |
| 2. Migration..... | 13 |
| 2.1. Étude de faisabilité..... | 13 |
| 2.1.1. Procédure..... | 13 |
| 2.1.2. Gestion du multilingue..... | 13 |
| 2.2. Mise en œuvre..... | 16 |
| 2.2.1. Prérequis..... | 16 |
| 2.2.2. Conversion..... | 17 |
| 2.2.3. Supprimer l'ancien thésaurus..... | 18 |
| 2.2.4. Import..... | 18 |
| 2.2.4.1. Aciege_terme_domaine..... | 19 |
| 2.2.4.2. Accrocher les domaines à THESAURUS_HEADER..... | 19 |
| 2.2.4.3. Aciege_terme..... | 20 |
| 2.2.4.4. Aciege_terme_interdit..... | 20 |
| 2.2.4.5. Aciege_terme_interdit_anglais..... | 20 |
| 2.2.5. Configuration des vues <i>Cadic</i> | 20 |
| 2.2.5.1. Vue « AUDENCIA : Monographie »..... | 20 |
| 2.2.5.2. Formulaire de recherche..... | 21 |
| 2.2.5.3. Formulaire de résultat..... | 22 |
| 2.2.6. Vue « AUDENCIA : Recherche experte »..... | 22 |
| 2.2.6.1. Formulaire de résultat/Formulaire de saisie..... | 22 |
| 2.2.6.2. Formulaire de recherche..... | 23 |
| 2.3. Vue « Thésaurus affichage »..... | 23 |
| 2.3.1.1. Formulaire de saisie..... | 23 |
| 3. Conception du mini-thésaurus de communication..... | 24 |
| 3.1. Règles d'élaboration du mini-thésaurus..... | 24 |
| 3.1.1. Équipe..... | 24 |
| 3.1.2. Public visé..... | 24 |
| 3.1.3. Domaines d'application..... | 24 |



| | |
|--|----|
| 3.1.4. Périmètre..... | 24 |
| 3.1.5. Base de travail..... | 25 |
| 3.1.6. Règles de construction d'un terme..... | 25 |
| 3.1.6.1. Nombre..... | 25 |
| 3.1.6.2. Noms composés..... | 25 |
| 3.1.6.3. Sigles..... | 26 |
| 3.2. Création du mini-thésaurus..... | 26 |
| 3.2.1. Dégrossissement des termes candidats..... | 26 |
| 3.2.2. Hiérarchisation des termes..... | 26 |
| 3.2.3. Étude intellectuelle des termes..... | 27 |
| Conclusion..... | 28 |
| Références bibliographiques..... | 29 |

Introduction

Informaticienne de formation, en 2016 je décide de me réorienter vers le domaine de la documentation. Pour l'année scolaire 2017-2018 j'ai donc intégré la licence professionnelle « Documentation et Bibliothèque » de l'université de Limoges.

Le présent document est le mémoire de mon stage d'étude. Il prend place dans la médiathèque de l'école de management Audencia Business School à Nantes et porte sur la mise à jour de son thésaurus documentaire.

Ce stage est donc en quelque sorte l'accomplissement de mon choix de carrière, puisqu'il mêle à la fois mes compétences en documentation et celles en informatique, pour respectivement le traitement intellectuel du thésaurus et la mise à jour concrète du thésaurus.

Ce mémoire reflète ainsi cette double coloration, car il décrit aussi bien les notions d'indexation et de conceptualisation derrière le thésaurus que les procédures et scripts informatiques mis en œuvre lors des opérations de mise à jour. Les parties dédiées à l'informatique sont donc principalement destinées à la médiathèque.

Le mémoire est divisé en trois parties : tout d'abord la présentation et description du projet de mise à jour ; secondement la procédure de mise à jour ; et finalement la création d'un mini-thésaurus de complément.

1. Présentation du projet

1.1. Contexte du projet

1.1.1. Audencia Business School

Audencia Business School est un établissement d'enseignement supérieur français fondé à Nantes en 1900. Après plus d'un siècle de développement, l'école compte aujourd'hui près de 4 000 étudiants répartis sur quatre sites à Nantes ainsi qu'un campus à Paris, et depuis 2016, un campus en Chine.

La médiathèque d'Audencia emploie 8 documentalistes et est répartie sur les 3 sites nantais d'enseignement du groupe qui possèdent des orientations différentes :

- **Atlantic Campus** – Le campus principal, spécialisé en management au niveau bac+5, avec des classes internationales ;
- **City Campus** – Campus du centre-ville, spécialisé en management au niveau licence ;
- **Media Campus** – Campus spécialisé dans le secteur de la communication et des médias.

La médiathèque a pour mission de donner accès aux savoirs et connaissances à l'ensemble de la communauté des étudiants et enseignants-chercheurs du groupe. Pour cela, elle constitue une offre documentaire papier et électronique adaptée qui représente aujourd'hui 20 000 livres papier et 18 000 ebooks, ainsi que l'accès à 25 bases de données électroniques. En 2017, le budget annuel d'acquisitions était de 405 k€, dont les 2/3 sont consacrés aux abonnements de ressources numériques.

La médiathèque d'Audencia fait aussi partie du réseau de l'Aciege, une association Loi 1901 qui regroupe les bibliothèques des écoles de management membres de la CGE (Conférence des Grandes Écoles). L'association a pour but de développer l'entraide et la solidarité entre ses membres, d'effectuer les démarches d'intérêt commun auprès des fournisseurs ainsi que de développer et faire évoluer le métier. Elle compte aujourd'hui 37 écoles membres.

1.1.2. Mise à jour du thésaurus documentaire

La médiathèque utilise depuis 2006 le logiciel de gestion documentaire *Cadic Intégrale* de la société Cadic Services.

Cette application métier stocke les données de gestion des notices bibliographiques des 20 000 livres de la médiathèque et des 500 abonnements de revues avec leurs 40 000 exemplaires liés. Il permet aux documentalistes de procéder aux opérations courantes de gestion : commande, réception, catalogage ainsi que les emprunts/retours.

L'indexation des documents se fait à travers : d'un plan de classement élaboré en interne pour l'indexation systématique, ainsi que d'un thésaurus documentaire pour l'indexation matière, qui sera le sujet de ce stage.

Il s'agit du thésaurus matière Delphes élaboré par le groupe IndexPresse. Ce thésaurus est spécialisé dans les thématiques de l'entreprise, ses marchés, ses finances, sa gestion et son environnement.



La médiathèque utilise ce thésaurus dans sa version 2004 depuis l'année 2007. Depuis lors les documentalistes ont fait évoluer le thésaurus en nettoyant plusieurs centaines de termes inutilisés ou obsolètes et en ajoutant une cinquantaine de termes pour suivre les évolutions des besoins d'indexation.

Mais aujourd'hui ce thésaurus est très vieux (presque 15 ans) et son champ lexical ne correspond que très peu aux besoins du fonds du Media Campus. Il apparaît donc nécessaire de le mettre à jour. Les premières réflexions sur le sujet débutent lors de l'année 2015, mais faute de temps le projet est resté dormant jusqu'à aujourd'hui.

Je suis chargée lors de mon stage de m'occuper de cette mise à jour. Dans ce projet, je suis aidée de Fabienne Fouché, documentaliste responsable des ressources numériques, et je travaille aussi avec Anna Guillet, documentaliste du Media Campus, pour que le thésaurus puisse répondre aux besoins de son fonds documentaire.

1.2. Fonctionnement du thésaurus documentaire

Avant d'entrer dans le vif du projet, il convient de définir et de décrire les grands principes du thésaurus documentaire.

1.2.1. Définition

Un thésaurus est un langage artificiel constitué de termes. Ces termes représentent des concepts et servent à indexer un document. On parle ici « d'indexation matière » ou « d'indexation sujet », puisqu'il sert à décrire succinctement le contenu, le « sujet », la « matière » du document.

Cette forme d'indexation est, dans un centre de documentation, généralement associée à une autre forme de classement : l'indexation systématique, comme par exemple l'indexation Dewey, qui catégorise les documents dans de grandes classes.

Une définition concise d'un thésaurus serait :

Un thésaurus est un ensemble de termes appartenant à un domaine précis de la connaissance et représentant chacun un concept ou un ensemble de concepts. Les termes sont reliés entre eux par des relations qui permettent de déterminer leur environnement sémantique. Ces relations sont de trois ordres : relations d'équivalence, relations hiérarchiques et relations associatives. La valeur d'un thésaurus en tant qu'outil documentaire réside moins dans le choix de ses termes - toujours empreint d'une certaine convention - que dans le choix des relations conceptuelles qui en définissent les modalités d'utilisation. (Ménillet 1993)

Concrètement un thésaurus prend la forme d'une liste d'autorités organisée selon des relations sémantiques et dans laquelle sont identifiés des termes. Ceux-ci obéissent à des règles terminologiques propres et sont sélectionnés pour éviter au maximum toute ambiguïté. (Keller 2013)

1.2.2. Les 3 types de relations sémantiques

Comme dit dans la définition d'un thésaurus précédente, il existe trois types de relations :

1.2.2.1. Relation d'équivalence

Un terme peut être « descripteur » (EM) ou « non-descripteur » (EP). Le « descripteur » est le terme retenu pour décrire un concept, quand le « non-descripteur » est un synonyme ou quasi-synonyme d'un descripteur (Keller 2013). Chaque descripteur peut posséder plusieurs non-descripteurs. Dans *Cadic*, un non-descripteur est décrit comme un « terme interdit » (TI).

Par exemple pour un synonyme, le descripteur « PEPINIERE D'ENTREPRISES » peut avoir pour non-descripteur « Incubateur » (les deux mots peuvent être utilisés de façon interchangeable). Ou pour un quasi-synonyme, le descripteur « ENTREPRISE LOCALE » peut avoir pour non-descripteur « Entreprise régionale » (une entreprise régionale est une forme particulière d'entreprise locale).

Par convention et comme illustré dans les exemples précédents, le descripteur est écrit en majuscule lorsque le non-descripteur est écrit en minuscule.

1.2.2.2. Relation hiérarchique

La relation hiérarchique associe un terme générique (TG) à un ou plusieurs termes spécifiques (TS). Cette relation construit ainsi une arborescence de termes qui aide l'utilisateur à naviguer dans le thésaurus et atteindre le niveau de précision qu'il souhaite.

On distingue trois types de relations hiérarchiques (Keller 2013) :

- Relation générique (de divisions à subdivisions) avec par exemple TG=SOCIOLOGIE, TS=SOCIOLOGIE ECONOMIQUE ;
- Relation partitive (de tout à partie) avec par exemple TG=EUROPE, TS=FRANCE ;
- Relation d'instance avec par exemple TG=REGION FRANCAISE, TS=BRETAGNE.

Dans *Cadic* un thésaurus possède ainsi un premier terme de tête, nommé le « thesaurus header » auxquels sont accrochés plusieurs termes spécifiques, qui forment des domaines (DOM). Ces domaines sont le point de départ de plusieurs hiérarchies descendantes de termes, par exemple « ENTREPRISE », « DROIT », « SCIENCES », « GEOGRAPHIE », et cætera.

Le fait qu'un terme soit un quasi-synonyme ou un terme spécifique est défini lors de la création ou de la mise à jour du thésaurus. Par exemple pour le Thésaurus du Management (présenté plus bas), en octobre 2017 le terme EP « Intelligence collective » du terme EM « MANAGEMENT DES CONNAISSANCES » est devenu le TS « INTELLIGENCE COLLECTIVE » du TG « SCIENCE COGNITIVE ». Ce choix est principalement fonction de la pertinence du terme.

1.2.2.3. Relation associative

Une relation associative est une relation non-hiérarchique entre deux termes. Elle relie deux termes qui, bien que décrivant deux concepts distincts, restent liés par une forte connexion sémantique (Keller 2013). Par exemple les termes « INNOVATION » et « START-UP » sont des termes associés (TA).

Il est par ailleurs possible d'associer deux termes qui n'appartiennent pas au même domaine.



Cette association est fondamentalement subjective et donc doit être créée avec attention. Elle sert principalement lors d'une recherche, pour orienter l'utilisateur vers des concepts liés à sa requête initiale, et ainsi enrichir son champ de recherche.

1.3. Étude du projet

Lors de mon arrivée dans le projet, deux options de mise à jour sont alors étudiées :

- Celle de mettre à jour le thésaurus Delphes dans sa dernière version ;
- Celle de changer de thésaurus pour le Thésaurus du Management élaboré par l'Aciege.

Il convient d'étudier les deux options, afin de déterminer le thésaurus qui conviendrait le mieux à nos besoins.

1.3.1. Thésaurus Delphes

Comme présenté précédemment de façon succincte, le thésaurus de Delphes est un thésaurus matière édité par le groupe IndexPresse. Il s'agit d'un langage contrôlé pour indexer et rechercher l'information économique, spécialisé dans les thématiques de l'entreprise, ses marchés, ses finances, sa gestion et son environnement.

Initialement créé en 1980 par la Chambre de Commerce et d'Industrie de Paris, il n'existe qu'en langue française.

Sa dernière version date de 2011. Elle comprend 3 454 descripteurs répartis en 61 domaines, ainsi que 1 784 non-descripteurs. En comparaison de sa version 2004, il met principalement à jour les secteurs des services, du marketing, de la communication, de l'informatique, de la finance, de l'environnement et enfin les évolutions nationales. Il ajoute concrètement 213 nouveaux descripteurs et plus de 100 non-descripteurs.

Après prise de contact avec IndexPresse, il semblerait qu'aucun projet de mise à jour du thésaurus ne soit au programme pour un avenir proche.

1.3.2. Thésaurus du Management

Le Thésaurus du Management est élaboré par l'association de l'Aciege, dont fait partie Audencia. Ce thésaurus est créé à partir des travaux du Bureau Van Dijk et de Georges Sandeau dans les années 1980.

Le thésaurus, disponible sous la licence Creative Commons Attribution, est disponible en trois langues, à savoir le français, l'anglais et l'espagnol. Il est mis à jour chaque année par un groupe de travail composé de représentants des grandes écoles de management membres de l'Aciege.

Sa dernière version date de 2017. Elle comprend 2 447 descripteurs répartis en 69 domaines, ainsi que 1 211 non-descripteurs.

1.3.3. Comparaison des deux thésaurus

Afin de choisir quel thésaurus conviendrait le mieux à nos besoins, nous avons élaboré un tableau recensant les avantages et inconvénients de chaque solution :



| | Avantages | Inconvénients |
|------------------------------|---|---|
| Delphes 2011 | <ul style="list-style-type: none"> • Màj technique Cadic plus facile ; | <ul style="list-style-type: none"> • Pas de màj depuis 2011 et pas de visibilité sur une màj à venir. • Français uniquement ; • Moitié des termes en secteur d'activité peu pertinents ; • Schéma Communication à développer. |
| Thésaurus du Management 2017 | <ul style="list-style-type: none"> • Ciblé sur le management ; • Màj annuelle dans le cadre de l'Aciege ; • Possibilité de contribuer à la màj dans le cadre de l'Aciege ; • Trilingue français / espagnol / anglais. | <ul style="list-style-type: none"> • Intégration Cadic plus compliquée ; • Période de double indexation vraisemblablement plus longue ; • Schéma Communication à développer. |

Comme le rapport avantages/inconvénients penche fortement en faveur du Thésaurus du Management, ce dernier est donc choisi pour la suite du projet de mise à jour du thésaurus documentaire.

2. Migration

2.1. Étude de faisabilité

2.1.1. Procédure

Il existe deux procédures de migration possibles :

1. Écraser l'ancien thésaurus par le nouveau ;
2. Importer le nouveau thésaurus à côté de l'existant, créant une cohabitation des deux thésaurus.

La première solution possède une contrainte majeure : lorsque l'on saisit ou modifie une notice de document dans *Cadic*, au moment de la sauvegarde, il est vérifié si les termes utilisés comme descripteurs existent bel et bien dans le thésaurus ; s'ils n'existent pas la sauvegarde est alors impossible. Cela ne pose pas de problème lors de la saisie d'une nouvelle notice, mais lors de la modification d'une ancienne notice, comme les deux thésaurus ne partagent pas exactement le même vocabulaire, cela forcera le documentaliste à refaire l'indexation matière du document.

La seconde solution ne pose pas ce problème. En outre il s'agit de la solution qui a été sélectionnée à la fois par les écoles (membres de l'ACIEGE) NEOMA Business School et Rennes School of Business qui ont toutes deux migré du thésaurus de Delphes au Thésaurus du Management en 2012. Toutefois cette solution pose un autre problème : s'il y a deux thésaurus, il y a deux champs d'indexation matière pour une même notice. Si l'OPAC de Rennes SB permet de fusionner ces deux champs en un seul, ce n'est pas le cas pour NEOMA BS qui a dû faire afficher les deux indexations sur son OPAC « Mots-clés Management » et « Mots-clés Delphes » peu clairs pour l'utilisateur.

Illustration 1: Exemple de double indexation sur le catalogue de NEOMA BS.

ISSN : 3782-04703-433-0

Prix : 39 EUR

Mots-clés : **Management**
DISTRIBUTION ; COMPORTEMENT DU CONSOMMATEUR ;
MARKETING STRATEGIQUE ; MARQUE ; PUBLICITE ;
PROMOTION DES VENTES ; INTERNET ; ENFANT
Delphes
MARKETING GENERATIONNEL

Classification décimale : 122.53 CONSOMMATEUR

Illustration 1: Exemple de double indexation sur le catalogue de NEOMA BS.

Avec *Cadic*, il semble a priori impossible de fusionner deux champs en un seul pour la partie OPAC, nous plaçant dans la même situation que pour NEOMA BS. Aussi, nous choisissons la première solution, à savoir écraser l'ancien thésaurus par le nouveau.

2.1.2. Gestion du multilingue

Cadic n'offre pas nativement de possibilité d'indexer une notice par mots-clés de manière multilingue. Toutefois, si *Cadic* n'offre pas nativement une indexation multilingue, il est tout de même possible, dans la notice d'un terme du thésaurus, de spécifier sa traduction en anglais, allemand ou espagnol. En conséquence de quoi, l'on peut tout de même associer une équivalence anglaise à un terme français.



Saisie de terme

Décrocher le terme de ORGANISATION DE L'ENTREPRISE ;
 Accrocher le terme

Terme

Termes génériques ORGANISATION DE L'ENTREPRISE ;
 REINGENIERIE ;

Termes spécifiques

Termes associés

RESISTANCE AU CHANGEMENT RESTRUCTURATION DE L'ENTREPRISE

Descripteur oui/non

Traduction Anglaise

Traduction Allemande

Traduction Espagnole

Note d'application

Utilisateur POPEK Julia
 Date de modif. 25/01/2018
 Référence

Illustration 2: Notice *Cadic* du terme « Changement organisationnel » avec ses traductions anglaise et espagnole.

Ces traductions sont traitées comme des termes interdits et permettent donc de rechercher le terme directement dans le thésaurus.

Le terme recherché **organizational change** est un terme interdit ou une traduction.
 Le terme à employer est **CHANGEMENT ORGANISATIONNEL**.

- >> 00 Mots Outils
- >> 01 Relations Economiques Internationales
- >> 02 Organisation de l'Economie
- >> 03 Situation nationale
- >> 04 Aménagement du territoire
- >> 05 Finances publiques
- >> 06 Monnaie
- >> 07 Epargne
- >> 08 Marché des capitaux
- >> 09 Environnement
- >> 10 Science
- >> 11 Droit
- >> 12 Propriété intellectuelle
- >> 13 Entreprise
- >> 14 Gestion de l'entreprise
- >> GESTION DE L'ENTREPRISE
- >> ORGANISATION DE L'ENTREPRISE
- >> **CHANGEMENT ORGANISATIONNEL**
- >> 15 Ressources humaines

Illustration 3: La recherche « Organizational change » trouve bien « Changement organisationnel » comme si c'était un synonyme.

Malheureusement lorsque l'on exécute une recherche d'un document par mots-clés dans *Cadic*, la recherche ne prend en compte que le terme exact utilisé pour indexer le document, et ignore ses synonymes ou traductions.

Ainsi, il apparaît nécessaire de faire apparaître clairement dans la notice une indexation en français et en anglais. Ma première démarche fut donc de regarder si d'autres organismes utilisateurs de *Cadic* indexent en multilingue, et si oui, comment.

C'est par exemple le cas de la Bibliothèque de l'École des Ponts ParisTech. Sur leur catalogue, dans la liste des mots-clés, nous trouvons à la fois les descripteurs en français et les descripteurs en anglais.

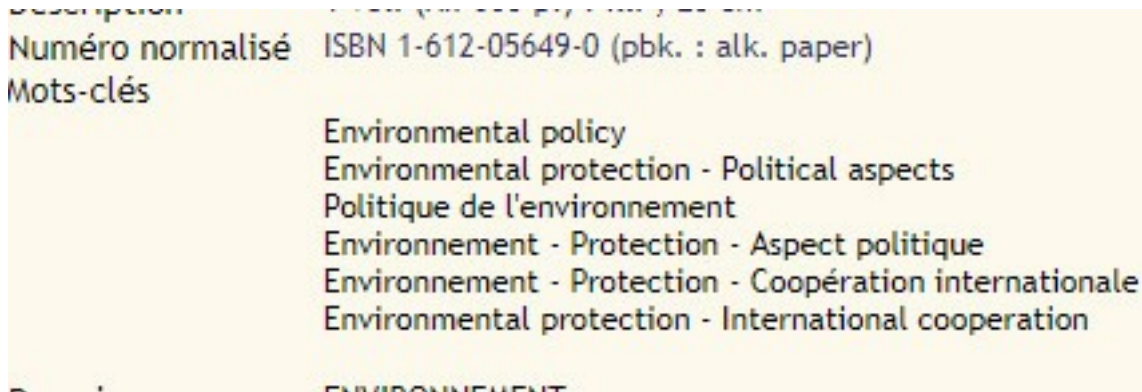


Illustration 4: Exemple d'une indexation bilingue pour l'École des Ponts.

Cette démarche, bien que techniquement simple, présente deux soucis majeurs :

- Une surcharge de la liste des descripteurs avec les deux langues réunies dans un seul champ (--);
- Les documentalistes doivent saisir deux descripteurs pour indexer à la fois en français et anglais (--).

La Vidéothèque Roméo Martinez propose une autre forme d'indexation bilingue, où chaque descripteur est écrit à la fois en français et en anglais.



Illustration 5: Exemple d'une indexation bilingue pour la vidéothèque, chaque descripteur est séparé par un « / ».

Cette démarche présente pour caractéristiques :

- Une surcharge de la liste des descripteurs avec les deux langues réunies dans un seul champ (--);
- Les documentalistes n'ont qu'un seul descripteur à saisir pour indexer à la fois en français et en anglais (++);
- Complexifie la compatibilité avec le Thésaurus du Management, notamment lors du processus de mise à jour (--).

L'idéal serait de n'avoir à indexer qu'une seule fois le document en français, et que l'indexation en anglais se fasse automatiquement par rapport à l'indexation en français (ou inversement). Les mots-clés en français et en anglais se trouveraient dans deux champs différents.

Dans *Cadic*, les notices de documents sont gérées par la table MySQL nommée ILS_DOC. Cette table comprend plusieurs champs pour gérer les mots-clés/descripteurs :

- DOC_DEE : Mots-Clés – Actuellement utilisé ;
- DOC_DE : Mots-clés secondaires – Actuellement inutilisé ;
- DOC_DL : Mots-clés libres – Actuellement inutilisé ;
- DOC_DC : Mots-clés candidats – Actuellement utilisé ;
- AUD_SOCIETE : Mots-Clés sociétés – Actuellement utilisé.

Les champs « DOC_* » sont natifs à *Cadic* cependant que les champs « AUD_* » sont propres à Audencia. Il est donc possible de créer des nouveaux champs afin de stocker les mots-clés en anglais voire en espagnol, ou alors on peut utiliser les deux champs « DOC_DE » ou « DOC_DL », actuellement inutilisés.

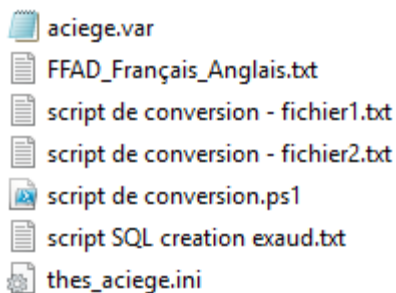
Il est ensuite possible, lors de la saisie d'un mot-clé en français dans le champ « DOC_DEE », d'ajouter en même temps et automatiquement sa traduction anglaise dans le champ dédié. Une telle procédure existe déjà pour la saisie de la cote du document, où l'indice et le libellé de la cote sont saisis simultanément dans deux champs distincts.

2.2. Mise en œuvre

Ce qui suit est la procédure de migration du thésaurus que j'ai rédigé pour Audencia, afin qu'ils puissent répéter l'opération au besoin.

2.2.1. Prérequis

Dans le répertoire de travail, vous devez avoir :



Où :

- **aciege.var** Fichier de configuration pour l'import ;
- **script de conversion***** Fichiers pour la conversion du fichier thesaurus vers un format importable dans Cadic ;
- **script SQL creation exaud.txt** Fichier contenant le script SQL à exécuter pour réinitialiser la table EXAUD (table thésaurus) ;
- **thes_aciege.ini** Fichier de configuration pour l'import ;
- **FFAD_Français_Anglais.txt** Fichier du thesaurus à importer après conversion.

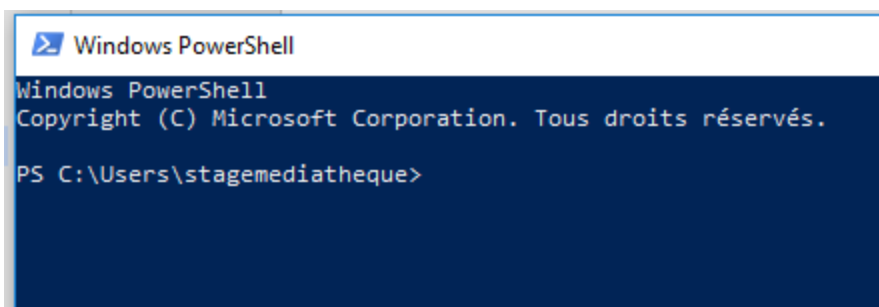


2.2.2. Conversion

Le fichier du thésaurus du management qui nous est fourni, ici « FFAD_Français_Anglais.txt » ne peut pas être importé dans *Cadic* tel quel. Il faut donc le convertir vers un format importable par *Cadic*.

Pour cela, j'ai conçu un script PowerShell pour automatiser l'opération, il s'agit du fichier « script de conversion.ps1 », qui est accompagné de deux autres fichiers « script de conversion – fichier1.txt » et « script de conversion – fichier2.txt » qui contiennent des constantes nécessaires à l'exécution du script.

PowerShell est une interface en ligne de commande intégrée à Windows depuis sa version Windows7. Vous pouvez la lancer directement avec le menu démarrer de Windows, qui vous affichera une fenêtre similaire à celle-ci :



```
Windows PowerShell
Copyright (C) Microsoft Corporation. Tous droits réservés.

PS C:\Users\stagemediatheque>
```

Avec PowerShell, allez dans le répertoire de travail, avec la commande :

```
cd 'P:\chemin\d\acces\du\fichier'
```

Dans mon exemple, ce sera :

```
cd 'P:\Mediatheque\JULIA\Thesaurus\fichiers import'
```

Par défaut, Windows bloque l'exécution de scripts PowerShell, pour désactiver cette protection, il faut utiliser cette commande :

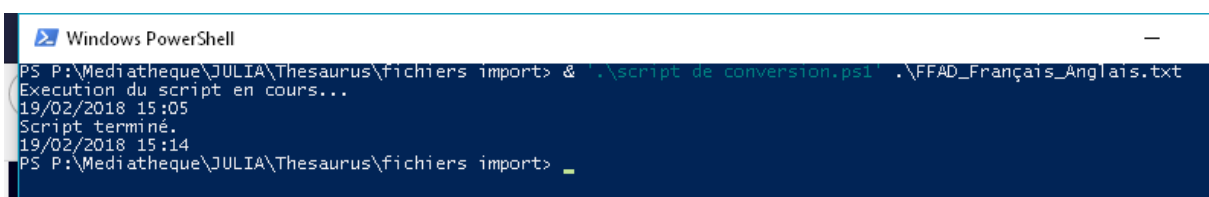
```
Set-ExecutionPolicy -Scope "CurrentUser" -ExecutionPolicy "Unrestricted"
```

Une confirmation vous sera demandé, entrez « o » pour « oui ».

Maintenant il suffit d'exécuter le script pour la conversion, avec cette commande :

```
& '.\script de conversion.ps1' .\FFAD_Français_Anglais.txt
```

L'exécution du script dure une dizaine de minutes. Le script terminé, vous devriez trouver un fichier « out.txt » dans le répertoire de travail : il s'agit du fichier à importer dans *Cadic*.

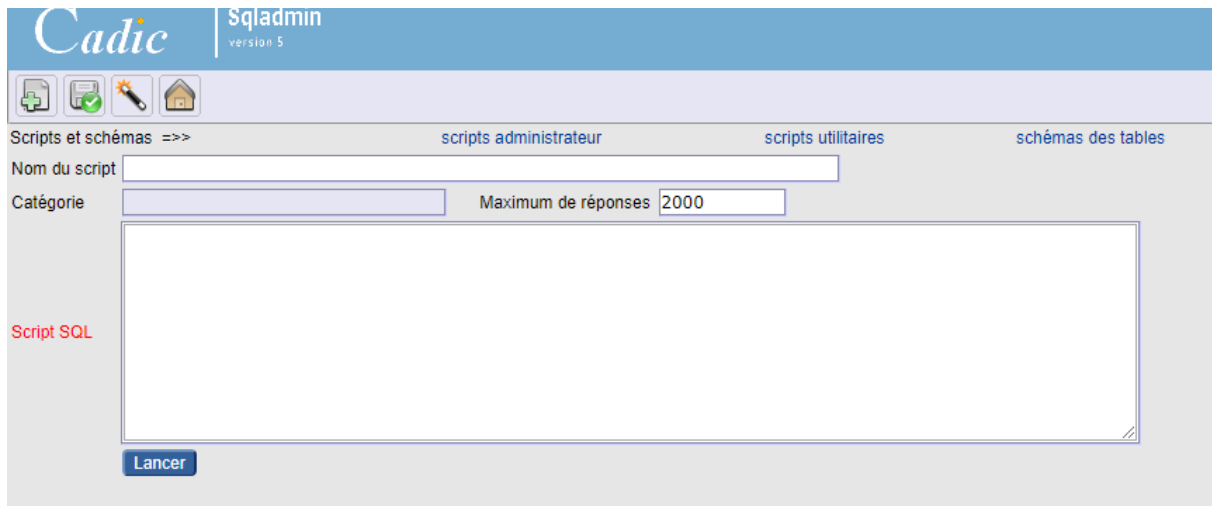


```
Windows PowerShell
PS P:\Mediatheque\JULIA\Thesaurus\fichiers import> & '.\script de conversion.ps1' .\FFAD_Français_Anglais.txt
Execution du script en cours...
19/02/2018 15:05
Script terminé.
19/02/2018 15:14
PS P:\Mediatheque\JULIA\Thesaurus\fichiers import>
```



2.2.3. Supprimer l'ancien thésaurus

Avant d'importer le nouveau thésaurus, il faut tout d'abord supprimer l'ancien. Dans *Cadic*, se rendre dans le menu « Administration SQL », qui vous mènera à cette page :



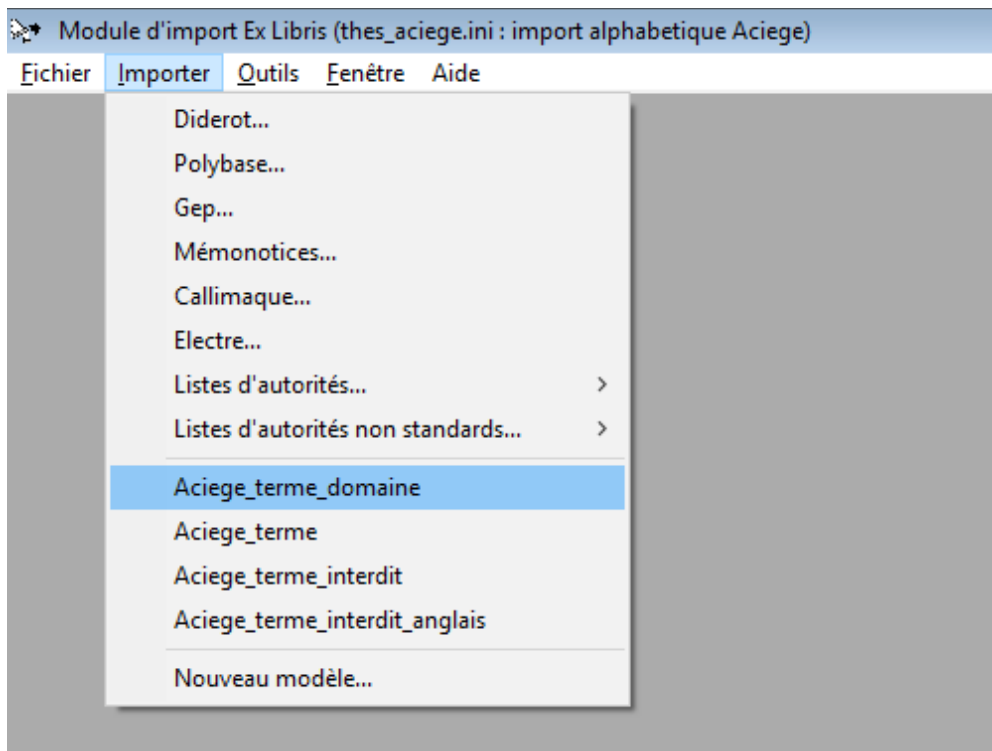
Dans le champ « Script SQL », coller-y le contenu du fichier « script SQL creation exaud.txt » et appuyez sur le bouton « Lancer ».

Cela supprime instantanément l'ancien thésaurus et prépare l'importation du nouveau.

2.2.4. Import

Cadic fournit un module nommé *ExImport* qui permet d'importer un ensemble de données dans *Cadic*. Lancez le module *ExImport*, vérifiez que le fichier de configuration « thes_aciege.ini » pointe bien sur la bonne base, puis ouvrez ce fichier, ce qui ajoute quatre options dans le menu « Importer » :





Exécutez les options dans l'ordre, en sélectionnant à chaque fois le fichier « out.txt ».

2.2.4.1. Aciege_terme_domaine

Importe les domaines du thésaurus.

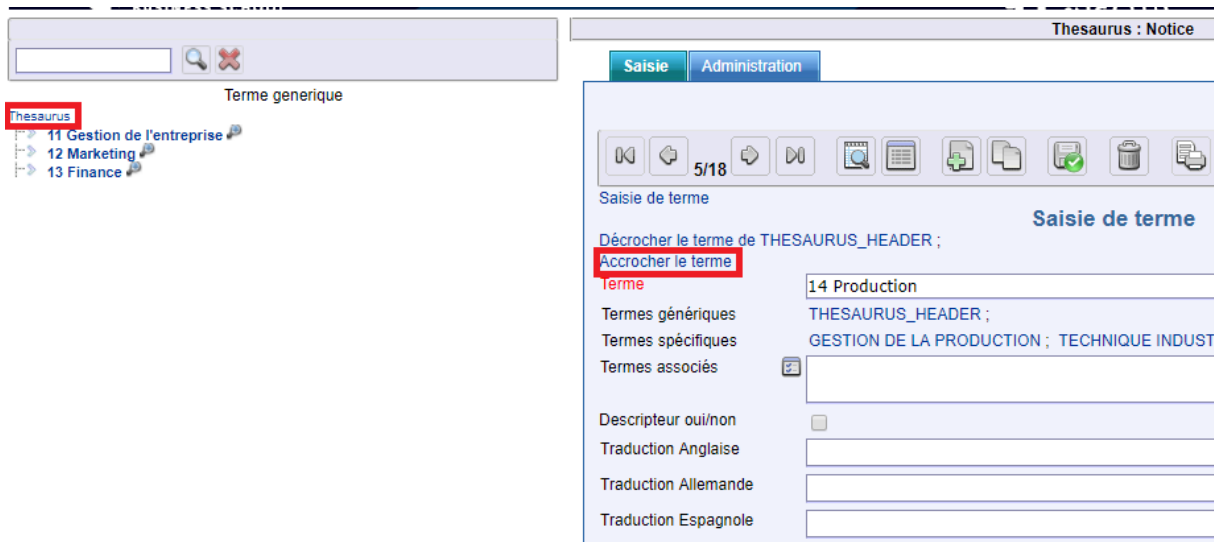
Durée ~10sec.

2.2.4.2. Accrocher les domaines à THESAURUS_HEADER

Lors de l'import, les domaines ne sont pas automatiquement accrochés au header du thésaurus, il faut donc le faire manuellement.

Sur la notice de chaque terme domaine, il faut cliquer sur le bouton « accrocher le terme », ce qui fait apparaître un volet sur la gauche, à partir duquel il faut sélectionner « Thésaurus ». Il n'y a pas besoin de sauvegarder la notice.





2.2.4.3. Aciege_terme

Importe les termes du thésaurus.

Durée ~4min.

2.2.4.4. Aciege_terme_interdit

Importe les termes interdits du thésaurus.

Durée ~2min.

2.2.4.5. Aciege_terme_interdit_anglais

Importe les termes interdits anglais du thésaurus.

Durée ~2min.

2.2.5. Configuration des vues *Cadic*

L'import terminé, il faut maintenant configurer les vues *Cadic* pour prendre en compte les nouveaux éléments, comme la gestion des descripteurs anglais.

Dans *Cadic*, se rendre dans le menu « Administration des vues ».

2.2.5.1. Vue « AUDENCIA : Monographie »

Formulaire de saisie

Pour ajouter le nouveau champ hébergeant les descripteurs anglais :

1. Renommer le champ « Descripteur » en « Descripteur FR ».
2. Dupliquer « Descripteur FR » en « Descripteur EN ».
3. Descripteur EN
 - a. Onglet « Général » : Nom de Champ -> « DOC_DE »
 - b. Onglet « Général » : Valeur du champ -> Option -> Remplacer « GSEP » par « Chr(10) ».
 - c. Onglet « Liste » :



Gestion des champs en SAISIE

LISTE [Général](#) | [Liste](#) | [I](#)

Type de liste: FERMEE [configuration d'une liste d'autorité](#)

Nom de la table: EXAUD

Titre de la liste: Descripteurs anglais Nombre de termes/page

Modèle d'affichage: modeles/liste2.tpl

Fenêtre

Fenêtre externe: Nom Hauteur: Largeur:

Fermeture automatique:

Fenêtre modale ajax: (Attention, il faut que le modèle utilisé soit compatible)

Champs

Champs affichés:

Libellés respectifs:

Champ principal:

Recherche

Champs de recherche:

Restriction Initiale:

Tri Initial: Recherche initiale:

Retour

Ajouter/Remplacer:

Retours des valeurs:

Séparateur de retour: Retour Chariot

- d. Onglet « Thesaurus » : tout supprimer.
- e. Onglet « Validation » : Champ obligatoire -> décocher.

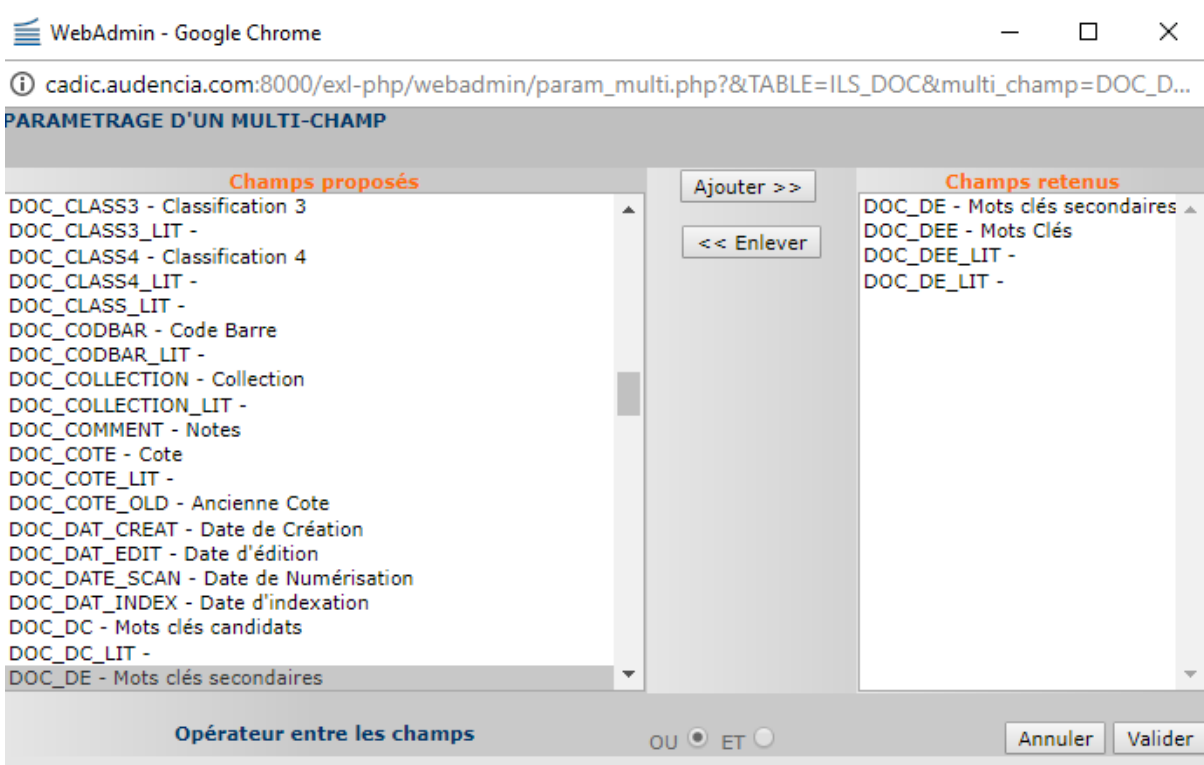
4. Descripteur FR

- a. Onglet « Thesaurus » : Retour des valeurs -> « DOC_DE|TRAD_1 ».

2.2.5.2. Formulaire de recherche

Pour que la recherche par descripteur prenne bien en compte l'anglais et le français :

Champ « Descripteur » : Cliquer sur le lien « multi-champ », cela ouvre une fenêtre, choisir :



Sélectionnez les 4 champs de recherche. Bien vérifier que l'opérateur entre les champs est de type « OU ».

Puis cliquer sur « Valider ». De retour sur la page principale, assurez-vous que ce soit bien le champ « DOC_DEE » qui apparaisse en premier dans la liste :

« DOC_DEE__DOC_DE__DOC_DEE_LIT__DOC_DE_LIT ».

Finalement, re-validez le champ « Lien Hypertexte ».

2.2.5.3. Formulaire de résultat

Pour afficher les descripteurs anglais :

1. Dupliquer le champ « Descripteur » en « Keywords ».
2. Keywords
 - a. Nom de Champ -> « DOC_DE »
 - b. Valeur du champ -> Option -> Convertir les données en lien de recherche (Extraction GSEP) -> Cocher.

2.2.6. Vue « AUDENCIA : Recherche experte »

2.2.6.1. Formulaire de résultat/Formulaire de saisie

Pour afficher les descripteurs anglais :

1. Dupliquer le champ « Mots clés » en « Keywords ».
2. Keywords
 - a. Nom de Champ -> « DOC_DE »
 - b. Valeur du champ -> Option -> Convertir les données en lien de recherche



(Extraction GSEP) -> Cocher.

2.2.6.2. Formulaire de recherche

Pour que la recherche par descripteur prenne bien en compte l'anglais et le français :

Champ « Mots clés » : Cliquer sur le lien « multi-champ », cela ouvre une fenêtre, choisir :

WebAdmin - Google Chrome

cadic.audencia.com:8000/exl-php/webadmin/param_multi.php?&TABLE=ILS_DOC&multi_champ=DOC_D...

PARAMETRAGE D'UN MULTI-CHAMP

| Champs proposés | Opérateur | Champs retenus |
|--------------------------------------|--------------------------|--------------------------------|
| DOC_CLASS3 - Classification 3 | Ajouter >> << Enlever | DOC_DE - Mots clés secondaires |
| DOC_CLASS3_LIT - | | DOC_DEE - Mots Clés |
| DOC_CLASS4 - Classification 4 | | DOC_DEE_LIT - |
| DOC_CLASS4_LIT - | | DOC_DE_LIT - |
| DOC_CLASS_LIT - | | |
| DOC_CODBAR - Code Barre | | |
| DOC_CODBAR_LIT - | | |
| DOC_COLLECTION - Collection | | |
| DOC_COLLECTION_LIT - | | |
| DOC_COMMENT - Notes | | |
| DOC_COTE - Cote | | |
| DOC_COTE_LIT - | | |
| DOC_COTE_OLD - Ancienne Cote | | |
| DOC_DAT_CREAT - Date de Création | | |
| DOC_DAT_EDIT - Date d'édition | | |
| DOC_DATE_SCAN - Date de Numérisation | | |
| DOC_DAT_INDEX - Date d'indexation | | |
| DOC_DC - Mots clés candidats | | |
| DOC_DC_LIT - | | |
| DOC_DE - Mots clés secondaires | | |

Opérateur entre les champs OU ET

Annuler Valider

Sélectionner les 4 champs de recherche. Bien vérifier que l'opérateur entre les champs est de type « OU ».

Puis faire « Valider ». De retour sur la page principale, assurez-vous que ce soit bien le champ « DOC_DEE » qui apparaisse en premier dans la liste :

« DOC_DEE__DOC_DE__DOC_DEE_LIT__DOC_DE_LIT ».

Finalement, re-validez le champ « Lien Hypertexte ».

2.3. Vue « Thesaurus affichage »

2.3.1.1. Formulaire de saisie

Pour que le numéro d'indice du descripteur soit correctement titré comme tel :

Renommer le champ « All. » en « Indice ».



3. Conception du mini-thésaurus de communication

Le contenu de cette section correspond au cahier des charges que j'ai rédigé pour l'élaboration du mini-thésaurus. Pour le rédiger, je me suis basée sur les recommandations de Danièle Degez (Degez 2009).

3.1. Règles d'élaboration du mini-thésaurus

Le présent document fait office de cahier des charges pour l'élaboration du mini-thésaurus visant à compléter le Thésaurus du Management de l'Aciege, afin de couvrir les besoins de l'ensemble de notre offre documentaire.

3.1.1. Équipe

Le mini-thésaurus est principalement élaboré par Anna et Julia, sous la responsabilité de Fabienne. D'autres documentalistes de l'équipe peuvent participer de façon plus sporadique, par exemple à des fins de relecture.

3.1.2. Public visé

Le public cible du thésaurus est principalement constitué :

- D'étudiants (majorité)
- D'enseignants-chercheurs (minorité)

Il s'agit donc d'un public de spécialistes, mais comme la majorité du public est constitué d'apprenants, le thésaurus doit rester accessible pour le néophyte du domaine.

3.1.3. Domaines d'application

Le but principal de ce mini-thésaurus est de couvrir le fonds documentaire du MediaCampus.

Ce campus étant spécialisé dans la communication et les médias, le mini-thésaurus doit couvrir les domaines suivants :

- Sciences de l'information et de la communication
- Industries culturelles, médias et art
- Sociologie

En plus de cela la médiathèque a un projet de développement de son fonds romans et bandes dessinées. Il paraît ainsi pertinent d'ajouter quelques termes descripteurs pour les principaux genres de fiction, qui ne sont pas présents dans le TM.

3.1.4. Périmètre

Le mini-thésaurus possédera différents domaines, dont plusieurs auront probablement une équivalence dans le TM. Comme il s'agit d'un mini-thésaurus pour compléter un existant, seul un petit nombre de domaines sera nécessaire (par exemple 4 ou 5). De même, pour chacun de ces domaines, il est convenu de ne pas avoir plus de deux niveaux de profondeur (en comparaison, le TM possède jusqu'à quatre niveaux de profondeur par domaine).

Nous considérons qu'il est important de préserver l'intégrité du TM. C'est-à-dire que l'on ne va pas ajouter de nouveaux termes dans l'arborescence du TM. De même, nous n'ajoutons



pas de synonyme pour les termes du TM.

En préservant ainsi l'intégrité du TM, il sera plus facile de le tenir à jour vis-à-vis des modifications apportées par l'Aciege.

3.1.5. Base de travail

Pour ce travail, nous nous baserons tout d'abord sur les termes qui ont été validés à l'époque du Thésaurus Delphes et qui n'ont pas trouvé d'équivalence dans le Thésaurus du Management, et sur quelques termes du Thésaurus de Delphes que l'on a souhaité conserver. Ceci forme un total de 19 termes.

Ensuite nous avons la liste des descripteurs candidats élaborée avec le Thésaurus Delphes, qui comprend 227 termes. Après comparaison de cette liste de descripteurs candidats avec le Thésaurus du Management, 136 d'entre eux n'ont pas d'équivalence dans le TM.

Ces 136 termes doivent être étudiés, pour savoir si on les accepte tels quels, après modification, ou si on les rejette.

3.1.6. Règles de construction d'un terme

Pour les règles de constructions des termes, nous nous référerons à celles utilisées par le TM.

3.1.6.1. Nombre

Les termes sont principalement des substantifs singuliers.

Il existe quelques rares exceptions à cela, où la forme plurielle est plus commune dans le langage courant :

- SCIENCES HUMAINES ET SOCIALES
- RELATIONS INTERNATIONALES
- FINANCES PUBLIQUES

3.1.6.2. Noms composés

Le TM utilise trois formes de noms composés : avec tiret ; avec espace ; tout attaché. La plus récurrente semble être la forme avec tiret. C'est donc celle sélectionnée.

Exemples :

- ETAS-UNIS D'AMERIQUE
- E-mail
- Eco-conception
- NOUVELLE-AQUITAINE

Contre-exemples :

- PAYS BAS
- Ecoblanchiment
- E learning
- MOYEN ORIENT
- AFRIQUE DU SUD



3.1.6.3. Sigles

Les sigles sont à privilégier sur les formes détaillées. Ils s'écrivent sans espace ni point. Les formes détaillées, en français et en anglais, doivent être intégrés sous forme de synonymes.

Exemple :

- AELE
 - EP Association Européenne de Libre Echange
 - EP European Free Trade Association

3.2. Création du mini-thésaurus

La création du mini-thésaurus s'est faite en trois étapes principales :

3.2.1. Dégrossissement des termes candidats

Cette étape consiste à prendre la liste des 136 termes candidats + 16 termes déjà validés et de les passer rapidement en revue afin de savoir si *a priori* (sans altération des termes candidats) la ou les notions qu'ils contiennent semblent pertinentes ou non. Cet exercice a été fait conjointement avec Anna, comme elle a une connaissance approfondie du fonds documentaire du campus ScienceCom.

3.2.2. Hiérarchisation des termes

Cette étape consiste à regrouper les différents termes en différents domaines. Ainsi 6 domaines ont été créés :

- **Genres littéraires** : Liste des principaux genres littéraires de fiction pour notre fonds romans et BDs.
- **Société** : Regroupe l'ensemble des sciences humaines, politiques et sociales.
- **Information Communication** : Regroupe les termes des sciences de l'information, de la communication et des médias.
- **Marketing** : Regroupe les termes liés au marketing, à la marque.
- **Culture** : Regroupe les termes liés aux productions culturelles et artistiques, leurs créateurs et leur environnement.
- **Mot outil Audencia** : Liste les quelques termes qui ne peuvent entrer dans les autres domaines.

Ces domaines font plus ou moins redondance avec certains domaines du TM, ce qui est cohérent pour un mini-thésaurus de complément.

Ensuite les termes à l'intérieur de ces domaines peuvent eux aussi se regrouper. Par exemple les deux termes « CENSURE » et « INDEPENDANCE DES MEDIAS » sont devenus des TS du terme « LIBERTE D'EXPRESSION ». En revanche certains termes sont sémantiquement associés, comme « BIG DATA », « OPEN DATA » et « DONNEES PERSONNELLES » alors qu'aucun TG ne semble exister dans la liste des termes candidats. Ainsi le terme « GESTION DES DONNEES » a été créé pour devenir le TG de ces trois termes.

Cette étape terminée, une première relecture a été faite.



3.2.3. Étude intellectuelle des termes

Les termes candidats proposés sont parfois flous, ambigus ou peuvent posséder plusieurs sens. Il s'agit donc d'étudier chaque terme un par un, pour essayer de définir correctement la notion souhaitée et utile pour notre thésaurus.

Par exemple le terme candidat « PUBLIC » fait appel à deux concepts :

- Celui de l'audience d'une œuvre culturelle, artistique, d'un spectacle, et cætera -- c'est le concept désiré pour le thésaurus, afin d'indexer les ouvrages qui étudient le public d'œuvres ;
- Celui du concept d'espace, de bien ou de service public, opposé à ce qui est privé -- ce concept n'est pas désiré pour le thésaurus.

Afin de lever l'ambiguïté, nous avons renommé ce terme en « ETUDE DES PUBLICS ».

Cette étude approfondie des termes permet aussi de se rendre compte que certaines notions se trouvent déjà dans le TM, mais que cela nous avait échappé lors de la première analyse rapide des termes candidats.

Par exemple le terme candidat de « MORT ». Ce terme peut avoir deux usages, celui pour la notion philosophique, inutile pour notre thésaurus, ou celui pour caractériser les services et les problématiques vis-à-vis de la mort, tels des ouvrages à propos du marketing de cercueils écologiques, ou la problématique de la gestion des profils utilisateurs des personnes décédées sur les réseaux sociaux. Après recherche le terme qui semble le mieux convenir est celui de « POMPES FUNEBRES » afin de décrire les entreprises de service autour de la mort. Or, ce terme existe déjà dans le TM. La notion n'a donc pas besoin d'être intégrée dans notre mini-thésaurus.

Pour faire ces analyses, plusieurs outils se sont révélés très pertinents :

- Les encyclopédies **Wikipédia** et **Universalis** : pour définir les notions, parfois trouver les termes voire les traductions.
- Les dictionnaires du **Centre National de Ressources Textuelles et Lexicales** et du **Wiktionnaire** : pour la recherche de synonymes principalement, et plus rarement pour définir des notions.
- Les sites de traduction **WordReference** et **Linguee** : pour aider à la traduction des termes français en anglais.



Conclusion

L'ensemble du projet de mise à jour du thésaurus a été complété lors du stage.

Mais un thésaurus est une entité vivante, qui doit évoluer régulièrement. Pour cela une documentaliste d'Audencia va intégrer le groupe de réflexion de l'Aciege autour du TM, qui se réunit une à deux fois par an.

Le système du thésaurus d'Audencia est désormais structuré en trois parties :

- Le TM, qui évoluera en fonction des avancées du groupe de réflexion de l'Aciege ;
- Notre mini-thésaurus, plus « vivant », il contient des descripteurs propres à notre fonds, mais aussi quelques termes qui seront proposées à l'Aciege pour intégration définitive dans le TM ;
- La liste des « Descripteurs candidats », que l'on remplit lorsque le TM et le mini-thésaurus semblent insuffisants. Cette liste a été vidée à la fin du projet de mise à jour, afin de repartir sur des bases saines. Elle sera réévaluée régulièrement, pour savoir si les descripteurs proposés intègrent ou non le mini-thésaurus, et peut-être pourront être validés à l'Aciege.

Il est aussi apparu une limite, lors de ce processus de mise à jour : seule une petite partie des notices de documents pré-existantes à la mise à jour ont été ré-indexées avec le nouveau thésaurus. Cette procédure de ré-indexation, compte-tenu de la taille du catalogue, ne peut se faire que petit à petit avec le temps et notamment en gré du désherbage du catalogue.

Pour ma part, ce stage et le projet qui m'a été confié m'ont permis de déployer mes compétences en informatique dans un contexte documentaire, d'approfondir mes connaissances sur l'indexation et plus globalement d'apprendre la vie d'une bibliothèque et ses problématiques.

Il confirme ainsi mon projet, d'utiliser mes deux casquettes informatique/documentation.



Références bibliographiques

- Degez, Danièle. 2009. « Construire un thésaurus ». *Archimag*, n° 222 (mars): 44-45.
- Keller, Loraine. 2013. « Encadrer la réingénierie d'un thésaurus : méthode, enjeux et impacts pour l'équipe d'un service de veille et documentation en entreprise ». Institut national des techniques de la documentation du CNAM.
https://memsic.ccsd.cnrs.fr/mem_00945542/document.
- Ménillet, Dominique. 1993. « Thésaurus et indexation ». *BBF*, n° 5 (septembre): 44-46.



Mise à jour d'un thésaurus documentaire

Ce mémoire présente la procédure de migration d'un thésaurus documentaire utilisé dans une école de management, puis la création d'un mini-thésaurus pour compléter ce nouveau thésaurus mis-à-jour.

Il décrit à la fois les aspects documentaires de l'indexation et les aspects informatiques de la mise à jour.

Mots-clés : Indexation, Thésaurus, Informatique.

Update of a thesaurus

This essay presents at first the update of a thesaurus used in a business school, then the conception of a mini-thesaurus designed to complete the updated thesaurus.

It describes both the theory behind the indexation and the computer aspects of a thesaurus' update.

Keywords : Indexing, Thesaurus, Computer.

