



Réingénierie du thésaurus du service InfoDoc

Cédric Rougeaux

► **To cite this version:**

Cédric Rougeaux. Réingénierie du thésaurus du service InfoDoc. Sciences de l'information et de la communication. 2010. <mem_00522567>

HAL Id: mem_00522567

https://memic.ccsd.cnrs.fr/mem_00522567

Submitted on 1 Oct 2010

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

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



Cedric ROUGEAUX

MASTER 1, MENTION ICD
(Option: Sciences de l'Information et du Document)

MÉMOIRE DE STAGE
Mission effectuée du 6 Avril 2010 au 31 Mai 2010

à la
RTBF.be
52, boulevard Auguste Reyers
1044 Bruxelles, Belgique

RÉINGÉNIERIE DU THÉSAURUS DU SERVICE INFODOC

Sous la direction de:
Mr J.S. CARNEL (responsable universitaire)
Mr C. BOLY (tuteur professionnel)

Soutenu le 22 Juin 2010 à l'UFR IDIST
Université Charles de Gaulle, Lille 3 (Campus Pont de Bois)
BP 60149, 59653 Villeneuve d'Ascq Cedex

Année Universitaire 2009/2010

*Je remercie tout particulièrement Claude Boly et ses collaborateurs
pour leur accueil chaleureux et leur disponibilité*

*Je remercie également Monsieur Carnel
pour le temps consacré à mon stage
et pour ses conseils avisés*

*et Monsieur Schöpfel
pour ses cours de préparation de stage qui se sont avérés très utiles*

TABLE DES MATIÈRES

INTRODUCTION.....	4
1. PRÉSENTATION DE LA RTBF : DE L'INFODOC AU THÉSAURUS.....	6
1.1 LA RTBF ET LE SERVICE INFODOC.....	6
1.1.1 La RTBF.....	6
1.1.2 Le service Infodoc.....	7
1.2 LA MISSION DE STAGE.....	10
1.3 LES LIMITES DES THÉSAURUS DE L'INFODOC.....	14
2. LA RÉÉLABORATION D'UN THÉSAURUS.....	16
2.1 LE THÉSAURUS VISÉ.....	16
2.2 QUELQUES CAS DE DIFFICULTÉS RENCONTRÉES.....	18
2.2.1 Le problème de spécificité.....	19
2.2.2 Le problème de l'exhaustivité.....	20
2.2.3 D'autres types de problèmes ?.....	21
2.3 DU CHOIX DES TERMES À LEURS PLACES DANS LE THÉSAURUS.....	23
3. UTILISATION D'UN LOGICIEL DE GESTION DE THÉSAURUS : TEMATRES.....	27
3.1 TEMATRES.....	27
3.2 L'EXPLOITATION DE TEMATRES.....	28
3.2.1 Description du vocabulaire.....	29
3.2.2 Création et gestion du thésaurus.....	30
3.2.3 Visualisation et exportation de données.....	31
3.2.4 Recherche et indexation.....	32
3.3 TEMATRES ÉTAIT-IL UN CHOIX PERTINENT?.....	33
CONCLUSION.....	36
BIBLIOGRAPHIE.....	38
GLOSSAIRE.....	40
ANNEXE.....	42
Exemple de fiche créée sur SharePoint.....	42
Interface de TemaTres.....	43
Mail de prise de contact avec Chalon et Ferreyra.....	44
Document élaboré lors de l'utilisation de TemaTres.....	47

Les termes possédant une astérisque (*) sont renvoyés au glossaire.

INTRODUCTION

Il existe différents types de langages documentaires. Ces « systèmes d'organisation des connaissances »¹ (HODGE, 2000) ont pour vocation de décrire un fonds documentaire afin de servir de guide lors d'une recherche sur ce même fonds. Les taxonomies, les ontologies, et les thésaurus sont des exemples de langages documentaires. Cependant ces langages doivent être en accord avec l'endroit où ils se trouvent et avec le fonds qu'ils recouvrent. D'une part parce que le fonds est différent mais aussi parce que les besoins des usagers sont différents. Par exemple un centre de documentation d'un organisme officiel utilisera un thésaurus différent d'un service de documentation d'un organe de presse.

En effet, pour ce dernier, le thésaurus devra correspondre à un fonds portant sur la presse mais aussi que ses descripteurs correspondent aux termes employés par ses usagers, en l'occurrence les journalistes. La RTBF, qui est un important organe de presse en Belgique, possède un service de documentation. Ce service, appelé service Infodoc, a pour rôle de collecter et de sélectionner l'information de la presse écrite.

Pourtant bien que le fonds soit important dans ce service, les thésaurus utilisés posent des problèmes de cohérence. En effet, ces thésaurus ont été élaborés de façon chaotique. Et à cela s'ajoute un projet de numérisation de ce fonds. Dès lors élaborer un nouveau thésaurus dans le cadre de ce projet d'évolution technologique est l'occasion d'améliorer l'indexation et les recherches futures des documents.

Mais le service Infodoc doit faire face à des contraintes. Élaborer un nouveau thésaurus demande du temps, du personnel et des moyens financiers dont ne dispose pas l'Infodoc. Il n'y a que sept documentalistes œuvrant dans ce service. Ils doivent faire face au travail quotidien de dépouillement de la presse écrite et ont à répondre aux demandes d'informations des journalistes de la RTBF. Malgré ces contraintes, le projet d'élaboration d'un nouveau thésaurus basé sur ceux actuellement utilisés a été lancé. Ce travail de réingénierie documentaire a pour base le thésaurus de l'Eurovoc afin de garantir une sélection efficace et pertinente des descripteurs. Ce nouveau thésaurus devra prendre place dans un outil informatique. En ce sens le logiciel libre de gestion de thésaurus TemaTres fut choisi.

¹ Terme traduit de l'anglais *Knowledge Organization Systems*

Ce projet effectué durant le stage recouvrait la réélaboration du thésaurus mais aussi l'encodage des nouveaux termes dans la base de données de TemaTres. Le thésaurus de l'Eurovoc a été créé dans un autre but que celui de l'Infodoc. Il n'est pas destiné à un organe de presse mais à des organismes officiels. Ces derniers ont des fonds et des besoins différents du service Infodoc, le choix du thésaurus de l'Eurovoc était-il judicieux ? Choisir un thésaurus ne recouvrant pas les mêmes besoins était-il un choix pertinent ? De plus l'utilisation de TemaTres a sans doute eu un impact sur l'encodage des termes. A-t-il rendu la phase d'encodage ardue ? Est-il ergonomique ? D'une façon plus générale, quelles ont été les contraintes auxquelles nous avons dû faire face durant cette réélaboration de thésaurus ? Et quelles ont été les difficultés rencontrées sur cette opération de réingénierie ?

Afin de répondre à ces questions, il nous faudra d'abord présenter le milieu dans lequel a eu lieu cette opération, en l'occurrence la RTBF et plus spécifiquement son service Infodoc. Puis il s'agira d'avoir un aperçu de la mission de stage et des limites que présentent les thésaurus de l'Infodoc. Dans un deuxième temps, nous nous pencherons sur les objectifs que vise le nouveau thésaurus et sur quelques types de difficultés rencontrées lors cette opération. Nous nous questionnerons alors sur la pertinence du choix de l'Eurovoc en regard de ce qui a été accompli durant ce stage. Et dans un dernier temps, nous nous concentrerons sur l'utilisation de TemaTres, sur ce que ce logiciel permet comme outil de gestion de thésaurus et sur la pertinence de son choix. Et enfin nous conclurons.

1. PRÉSENTATION DE LA RTBF : DE L'INFODOC AU THÉSAURUS

1.1 LA RTBF ET LE SERVICE INFODOC

1.1.1 La RTBF

La RTBF (Radio-Télévision Belge de la communauté Francophone) est une entreprise publique à caractère culturel en charge du service public de la radio et de la télévision pour la Communauté française de Belgique (Wallonie et Bruxelles). La VRT (Vlaamse Radio-en Televisieomroep) et la BRF (Belgischer RundFunk) s'occupent respectivement de la communauté flamande et de la communauté germanophone. En outre, elle fait partie de différentes instances internationales comme les Radios francophones publiques, la Communauté des télévisions francophones, le CIRTEF (Conseil International des Radios et Télévisions d'Expression Française) et l'UER (Union Européenne de Radiodiffusion).

Son chiffre d'affaire est de 250 millions d'euros et son siège social est situé au 52, boulevard A.Reyers à Bruxelles. La RTBF est régie par un Conseil d'Administration. Ses membres (de répartitions politiques différentes) sont élus au sein du Conseil culturel de la Communauté française de Belgique qui est l'autorité de tutelle de la RTBF. Cette entreprise fut créée en 1930 par un arrêté royal, l'INR (son ancienne appellation) avait pour but de diffuser des émissions radiophoniques en français et en néerlandais. Elle arrêta d'émettre durant la seconde guerre mondiale et recommença à émettre en 1945. En 1953, elle commence la diffusion d'émissions audiovisuelles. L'INR devient la RTB en 1960. En 1977, suite à l'instauration du fédéralisme séparant les différentes communautés linguistiques, la RTB devient la RTBF, ses émissions ne sont alors diffusées qu'en français.

Depuis l'entreprise n'a eu de cesse d'évoluer. Au milieu des années 1990, la RTBF investit dans la création d'EuroNews (une chaîne d'information) et elle signe un accord de coproduction avec Arte en 1995. Au début des années 2000, elle subit des plans de restructuration, elle cesse alors d'émettre certaines chaînes et sa masse salariale diminue. À partir de 2010, la RTBF devient RTBF.be, montrant par la même occasion son ouverture vers les nouvelles technologies.

1.1.2 Le service Infodoc

Plus spécifiquement, la RTBF possède un service de documentation appelé service Infodoc. Ce service a été créé en novembre 1963. La RTBF avait alors pris conscience suite à la mort du président américain Kennedy qui lui fallait organiser et maîtriser les sources documentaires réunies au sein d'un centre de documentation.

Jusqu'en 2003, chaque centre régional de la RTBF possédait sa propre cellule de documentation spécifique. Chacune de ces cellules était totalement autonome et indépendante du service de documentation Infodoc, même si il existait un esprit évident de collaboration entre ces différentes entités.

Le plan de restructuration Magellan appliqué à l'ensemble de la RTBF dès octobre 2003 a conduit à la suppression de ces cellules de documentation régionales. Suite à ce plan, le service Infodoc a vu le nombre de ses collaborateurs internes et externes réduit de douze à sept personnes. La collecte et la sélection d'informations s'opèrent presque exclusivement à partir des éditions de presse électroniques.

Infodoc gère et alimente un fonds documentaire constitué d'articles de la presse quotidienne et hebdomadaire nationale (francophone et néerlandophone), régionale et étrangère (française) et assure les recherches d'informations pour les journalistes, producteurs, animateurs et autres utilisateurs de la RTBF (radio et télévision). Ces derniers constituent la clientèle de ce service.

Le fonds documentaire est consulté notamment pour les besoins de l'actualité quotidienne et pour des émissions culturelles ou des magazines (« L'Invité de Matin 1^{ère} », « Culture Club », « Tête à tête », « Questions à la une »,...).

Il est important de préciser que le service Infodoc est réservé à l'usage interne de la RTBF. Les utilisateurs (trois cents utilisateurs réguliers recensés et deux milles demandes par an) sont principalement des journalistes et des pigistes qui se rendent sur place pour consulter le fonds. Ils peuvent également adresser leur demande par téléphone ou par mail.

Pour répondre à ces demandes, le service peut numériser des dossiers papiers et les transmettre ensuite à leur destinataire par mail ou via l'intranet basé sur « Sharepoint » de Microsoft.

Celui-ci est utilisé par la RTBF pour assurer une collaboration entre les différents membres du personnel et ses différents services. Il permet au personnel de la RTBF dont les journalistes d'avoir accès aux différents services que propose Infodoc comme des dossiers thématiques, des dossiers « chauds » ou encore une revue de presse.

Les dossiers thématiques ainsi proposés sont mis à jour quotidiennement. Il s'agit par exemple de dossiers sur les élections en Grande-Bretagne, sur le crash de l'avion présidentiel polonais... Ils servent notamment à anticiper des demandes que des journalistes pourraient formuler.

Les « dossiers chauds » également proposés aux usagers sont des dossiers dont les sujets ont une actualité brûlante. Ce sont des sujets qui sont susceptibles d'être traités par les journalistes de par leur récente introduction dans l'actualité.

Cette revue de presse concerne l'actualité quotidienne et permet aux journalistes de leur apporter des idées de sujets originaux. Elle est effectuée tous les matins notamment à partir de « Vers l'Avenir », un quotidien belge.

Le service Infodoc est constitué de sept documentalistes et de Claude Boly le responsable de ce service. Le travail hebdomadaire est réparti autour de trois pôles. Chacune de ces sections nécessite deux à trois documentalistes à temps plein.

Cependant pour gérer ces listes d'autorités, Mr Boly est assisté par Fabian Detienne (un documentaliste gérant l'actualité internationale). Malheureusement la charge de travail qui consiste au dépouillement et à répondre aux demandes de la clientèle étant assez conséquente, peu de temps est consacré à la gestion et à la maintenance des thésaurus.

Le stockage des documents se fait sur support physique. Il est constitué de plus de vingt trois milles dossiers stockés dans cinq cents mètres de rayonnages. Et il possède plus de trente ans d'archives. Le classement des documents se fait selon le sujet du document. Par exemple un article sur un artiste sera classé selon la liste biographique, un pays selon le thésaurus des pays étrangers... Un article traitant d'un thème sera classé suivant deux listes d'autorités selon l'endroit concerné par l'article :

- Thésaurus thématique belge
- Thésaurus thématique étranger

La distinction entre le « belge » et l'« étranger » se fait au niveau du lieu concerné par l'article. Par exemple un article traitant du téléphone portable en Belgique aura sa place dans le thématique belge.

Le service Infodoc gère tout ce qui relève du thématique avec ces deux thésaurus. Mais des descripteurs pour un même sujet sont différents d'un thésaurus à l'autre. À cela s'ajoutent les contraintes liées à l'espace que peut prendre un support physique. Un nouveau descripteur* se traduit physiquement par un dossier. L'espace que peut prendre un nouveau descripteur a une influence sur la gestion des thésaurus. Cette contrainte a un impact sur le nombre de descripteurs autorisés.

De plus le service Infodoc a un projet de numérisation de leurs documents en vue. Ce projet est l'occasion de revoir la classification se faisant selon les listes thématiques notamment en ce qui concerne l'harmonisation des descripteurs. Cela afin de n'avoir qu'un seul thésaurus à gérer et pour classer le thématique.

Cette numérisation se fera sans doute avec Téliio ou Alfresco (le choix n'est pas encore arrêté). La présence d'un outil de GED* facilitera le travail des membres du service Infodoc et leur permettra d'être plus réactif. De plus le gain de temps gagné sur la classification leur permettra de proposer plus de documents à disposition de leurs collègues de la RTBF.

1.2 LA MISSION DE STAGE

Le travail de réélaboration de thésaurus s'est accompagné du travail quotidien au sein du service Infodoc. Il fallait dépouiller des journaux étrangers (« Le Monde », « Libération ») et belges (comme « Le Soir », « La Libre Belgique »,...) et parfois répondre à des demandes de dossier de membres de la RTBF. Ce travail de dépouillement consistait à sélectionner des articles, leurs assigner des mots-clefs servant à les archiver physiquement. Ces mots-clefs étaient choisis en fonction des listes d'autorités utilisées par le service Infodoc.

En plus de ce travail de dépouillement, j'assistais tous les jours aux réunions de rédaction d'information du journal télévisé et de la radio afin d'y apporter un compte rendu au service Infodoc pour que l'équipe de documentalistes puisse anticiper toutes demandes éventuelles de dossiers. Ces réunions m'ont permis de découvrir comment était préparé un journal d'information audiovisuel ou radiophonique. Mais aussi de me rendre compte que la ligne éditoriale d'un organe de presse a une influence sur le service de documentation. La sélection des articles à classer se fait selon l'actualité que la RTBF peut traiter. De plus le choix des termes est influencé par le vocabulaire employé par les journalistes.

Afin de disposer d'un thésaurus efficace, il faut connaître le travail que font les documentalistes au service Infodoc. Ce sont ces documentalistes qui vont indexer et rechercher les documents. Élaborer un langage documentaire répondant aux critères de classification du service Infodoc était primordial.

De fait, ce travail de dépouillement quotidien m'a permis de me familiariser avec les listes d'autorités utilisées par le service Infodoc. Cela me fut utile lors de la réélaboration du nouveau thésaurus à partir des listes thématiques concernant la Belgique et l'étranger. De plus, afin de disposer d'une base claire, cette opération de réingénierie de thésaurus s'est basée sur le thésaurus d'Eurovoc.

Ce thésaurus Eurovoc est un « thésaurus multilingue couvrant tous les domaines de l'activité de l'Union européenne, il permet d'indexer les documents dans les systèmes documentaires des institutions européennes et de leurs utilisateurs. Ce produit

documentaire est actuellement utilisé par le Parlement européen, l'Office des publications des Communautés européennes, les parlements nationaux et régionaux en Europe, des administrations nationales et par certaines organisations européennes »². Il possède 21 domaines et 127 micro-thésaurus. Parmi ces domaines, nous avons par exemple QUESTIONS SOCIALES, AGRO-ALIMENTAIRE ou encore INDUSTRIE.

Il permet d'avoir une base sur laquelle s'appuyer pour concevoir le nouveau thésaurus. Durant le déroulement du projet, il fallait placer les descripteurs issus des anciennes listes d'autorités à ce thésaurus mais en prenant des libertés car celui d'Eurovoc n'a pas été conçu pour le service Infodoc. Il a donc fallu adapter ce thésaurus aux besoins de l'Infodoc.

L'une des contraintes que j'ai rencontré fut de reprendre les descripteurs des listes d'autorités en les insérant dans le thésaurus d'Eurovoc. En effet, il fallait garder un lien entre le nouveau thésaurus et les anciens de sorte qu'il serait possible de ressortir un ancien dossier même avec un descripteur changé, pour cela des termes exclus furent utilisés. De même il était important de conserver les renvois.

Ce travail de réélaboration s'est appuyé sur un travail de tri des descripteurs effectué en amont par un membre de l'équipe Infodoc. Des descripteurs jugés obsolètes furent écartés afin de ne garder que des descripteurs pertinents dans le nouveau thésaurus. Un descripteur qui ne faisait plus l'objet de demande de la part de la clientèle du service Infodoc était écarté.

À partir de ce tri, des fiches³ reprenant les descripteurs des listes d'autorités furent mises sur SharePoint. Ces fiches permettaient de comparer les différents descripteurs et d'avoir accès à leur renvoi, leur terme associé*... Ce projet s'effectuait aussi sur TemaTres. Il s'agit d'un logiciel libre qui permet d'élaborer et de traiter un thésaurus. Au préalable de ce travail, les deux premiers niveaux d'Eurovoc furent insérés dans ce logiciel. Il me restait ensuite à réfléchir à la pertinence des termes et à les incorporer dans ce nouveau thésaurus.

² Définition donnée par le portail en ligne d'Eurovoc
<http://europa.eu/eurovoc/sg/sga_doc/eurovoc_dif!SERVEUR/menu!prod!MENU?langue=FR>

³ Un exemple de fiche SharePoint se trouve en annexe.

Dans ce processus d'élaboration, il était primordial de garder une certaine cohérence. Il fallait garder une trace d'une manière ou d'une autre des anciens descripteurs. Cependant ce nouveau thésaurus ne fait que s'inspirer de celui d'Eurovoc, seuls les deux premiers niveaux ont été conservés, ceci afin de pouvoir garantir une flexibilité dans l'incorporation des mots-clés propres au service Infodoc.

Mon travail sur ce projet ne dura que six semaines (bien que j'ai effectué deux mois de stage) et durant ce laps de temps, le travail de dépouillement quotidien se devait d'être fait. Ainsi je ne passais que deux à trois heures par jour sur ce travail. Ma tâche consistait surtout à élaborer un cadre permettant par la suite à un documentaliste de l'équipe de poursuivre ce travail.

Pendant le déroulement de ce projet, l'inclusion ou non d'un descripteur se faisait en quatre temps :

- **Confrontation des listes d'autorités** utilisées en belge et en étranger

Cette première étape permettait de mettre en exergue les différences entre les descripteurs et d'apporter des solutions aux difficultés d'harmonisation des deux thésaurus.

- **Comparaison avec Eurovoc**

Cette deuxième étape consistait à trouver la place du descripteur des thésaurus thématiques dans l'arborescence de l'Eurovoc.

- **Choix du descripteur retenu**

Lors de cette phase, la formulation du terme était reconsidérée et changée si nécessaire.

- **Réalisation du nouveau thésaurus** avec TemaTres :

Cette dernière étape s'est basée sur un encodage partiel de l'Eurovoc.

- Encodage des listes de domaines et des micro-thésaurus de l'Eurovoc : cet encodage sur les deux premiers niveaux d'Eurovoc a permis d'avoir une base cohérente et solide pour le nouveau thésaurus.

- Intégration des descripteurs dans l'arborescence avec leurs relations : les descripteurs choisis étaient incorporés dans l'arborescence du nouveau thésaurus.

De plus des travaux préparatoires furent effectués avant les quatre étapes citées précédemment afin de se familiariser avec TemaTres :

- Test de fonctionnalité du logiciel TemaTres :

Des essais d'export/Import de la base de données ont été menés, recherche de solution des difficultés rencontrées lors de l'utilisation de ce logiciel et familiarisation avec l'interface de ce logiciel⁴.

- Une recherche approfondie d'information sur Internet pouvant aider à la compréhension de ce logiciel fut entreprise ainsi qu'une prise de contact par mail avec le créateur de ce logiciel⁵.

⁴ À cet effet, un document (présent en annexe) fut créé.

⁵ Les mails de cette prise de contact sont présents en annexe.

1.3 LES LIMITES DES THÉSAURUS DE L'INFODOC

Le service Infodoc se sert encore du support papier pour classer et archiver ses documents. Dans l'optique d'une évolution du service, ces documents seront tôt ou tard numérisés. Mais cette numérisation posera sans doute quelques problèmes. En effet, il faudra absolument pouvoir retrouver les anciens documents dans la base de données ainsi créée. Pour cela reprendre les thésaurus existants serait la solution la plus simple. Mais ces listes d'autorités existantes au sein d'Infodoc ont été créées de façon chaotique.

L'adjonction d'un nouveau terme au thésaurus ne passait pas par une concertation entre les documentalistes. Chacun prenait la liberté d'ajouter un terme quand la nécessité se faisait sentir, souvent à la suite d'une demande de document. Ce manque de collaboration se traduit par des incohérences. Par exemple le terme « AÉROSPATIAL » renvoie au descripteur « INDUSTRIE AÉROSPATIAL » dans le thésaurus thématique étranger et existe sous le terme « AÉROSPACIAL » dans le thésaurus thématique belge. En plus d'un renvoi, on assiste ici à une orthographe différente ce qui montre que ces thésaurus n'ont pas été élaborés conjointement. Des termes furent ajoutés mais sans forcément avoir mené une réflexion quant à leur pertinence⁶.

La numérisation future précédemment évoquée permettra de s'affranchir de la contrainte liée à l'espace dont le service Infodoc dispose pour la classification physique. Cet abandon du support physique pour classer les documents permet de ne pas se limiter dans le nombre de descripteurs. Avec le numérique, la question de la place qu'un dossier peut prendre physiquement perd tout son sens.

L'élaboration de ce nouveau thésaurus ne se fait pas sans difficulté. Harmoniser les listes existantes nécessite de les connaître ou au moins d'en être familier. Ce travail de réflexion se doit d'être approfondi car cet élaboration est importante pour le service et pour la numérisation qui s'annonce. Le but de cette nouvelle liste d'autorités serait surtout de proposer un thésaurus clair et pertinent pour le service Infodoc. Il aura pour rôle l'assignation de mot-clefs permettant de retrouver les documents physiques mais aussi de pouvoir classer de façon numérique les documents à venir.

⁶ Ce dernier point n'a pas pu être approfondi car aucun des membres du service Infodoc d'aujourd'hui n'était présent au moment de la création du service.

Cependant, différents types de problèmes peuvent se poser lors de l'élaboration d'un nouveau thésaurus, notamment en ayant pour point de départ deux autres thésaurus. Dans leur ouvrage, J.Aitchison et A.Gilchrist (AITCHISON, 1992) y exposent les principales sources d'incompatibilité pouvant exister entre deux thésaurus:

- La spécificité: un des deux thésaurus possède une arborescence davantage développée que l'autre. Il possède plus de termes spécifiques pour un même concept présent dans les deux thésaurus.
- L'exhaustivité: un thésaurus a des termes plus généraux qu'un autre, celui auquel il est confronté.
- Les termes composés: un concept peut être représenté par un terme composé dans un thésaurus et non dans l'autre. De plus des normes différentes concernant les termes composés peuvent être appliquées aux deux thésaurus ce qui peut poser des problèmes de compatibilité.
- Les synonymes: les synonymes (Employé pour) entre les différents thésaurus peuvent être différents.
- Les relations entre les termes: les hiérarchies entre les termes peuvent ne pas être similaires (de même que les relations associatives aussi appelé Terme Associé), ce qui poserait des problèmes évidents de compatibilité.

Lors de cette réélaboration, nous avons été confrontés à certains types de problèmes, notamment certains relevant de problèmes de spécificité et d'autres de problèmes d'exhaustivité. Cependant nous verrons que d'autres problèmes se sont posés dûs notamment à la particularité de la tâche effectuée durant ce stage.

2. LA RÉÉLABORATION D'UN THÉSAURUS

2.1 LE THÉSAURUS VISÉ

Selon Maniez (MANIEZ, 1977), un thésaurus est une « liste cohérente de termes d'un domaine particulier choisis en vue de caractériser le contenu des documents [...] et par suite de permettre une sélection rapide et efficace des documents souhaités ». Le thésaurus visé aura donc pour objectif de constituer « un vocabulaire de termes d'indexation contrôlés, structuré de sorte qu'il mette en évidence les relations *à priori* entre les concepts » (AITCHISON, 1987). Il prendra place dans un système de GED que le service Infodoc mettra en place par la suite. Il devra permettre également une recherche efficace d'information du fonds de l'Infodoc. Ce fonds est constitué de documents classés selon des descripteurs, lesquels devront se retrouver dans le nouveau thésaurus. Celui-ci fera en quelque sorte le lien entre les documents papiers et numérique.

Avant de commencer à élaborer un thésaurus, il faut connaître les besoins du lieu où ce thésaurus prendra place, dans le cas présent il s'agit ici du service Infodoc. Ce service, rappelons le, a pour mission de répondre aux demandes d'informations émanant des journalistes. Infodoc possède un fonds assez important. Ce nouveau thésaurus ne concerne que le thématique. L'un des deux thésaurus ne concerne que la Belgique : seuls les documents traitant d'un sujet (ou d'un thème) en Belgique sont classés suivant le thésaurus thématique belge. L'autre thésaurus concerne quant à lui ce qui touche à l'international (du point de vue belge).

Quand la question du choix des descripteurs (et des non-descripteurs*) se posera, il faudra prendre en compte les termes existants (c'est à dire ceux issus des deux anciens thésaurus). Cela permettra entre autre au personnel de l'Infodoc une adaptation plus rapide de ce nouveau thésaurus car il a l'habitude de travailler depuis des années sur les anciens thésaurus. De plus, ce thésaurus visé aura une correspondance avec les domaines thématiques du thésaurus de l'Eurovoc. Il s'agit surtout d'adapter l'Eurovoc aux termes issus des anciens thésaurus.

Dans ce projet de réélaboration de thésaurus, il ne faut pas oublier que bien que ces deux thésaurus soient tous les deux thématiques, ils ne sont pas identiques. Ces deux thésaurus ont été créés au fil du temps mais ils n'ont pas les mêmes termes, ni la même structure. Ils ont été créés indépendamment l'un de l'autre. De fait des termes se trouvant dans l'un ne se retrouvent pas dans l'autre. En effet des termes comme « CONFLIT SENEGALO-MAURITANIEN » ou « O.P.E.P. » n'ont pas leur place dans le thésaurus thématique belge.

Cependant ces deux thésaurus ont été fait de façon chaotique. Un manque de communication régnait au sein du personnel de l'époque. Quand un terme était créé dans l'un, il ne se retrouvait pas dans l'autre. Cette façon de procéder a conduit à une évolution complètement différente de ces thésaurus. Mais lors de la refonte de ces thésaurus en un seul, des termes comme ceux cités précédemment vont se trouver mélangés avec le thématique belge. Le but de ce nouveau thésaurus est de pouvoir mener une recherche efficace mais il y a tout de même un danger à ce que les deux thésaurus (belge et étranger) soient mélangés. Cependant il a été prévu l'utilisation de fiches créées sur SharePoint afin de pouvoir retrouver les dossiers antérieurs à la future numérisation. En effet, il ne faut pas oublier que cette élaboration du nouveau thésaurus est un prélude à la numérisation des documents du service Infodoc.

Le thésaurus visé devra donc permettre une recherche efficiente des descripteurs. De plus il sera utilisé aussi bien par les documentalistes que par les journalistes.

2.2 QUELQUES CAS DE DIFFICULTÉS RENCONTRÉES

Tout au long de ce processus de réélaboration de thésaurus, plusieurs termes nous obligèrent à une réflexion sur leur insertion dans ce nouveau thésaurus. Chacun d'eux doit correspondre à un concept. De plus, des règles furent imposées quant aux choix des termes comme le masculin singulier, ou encore le développement des acronymes, ceux-ci étant présents mais en terme exclu (sous forme de renvoi).

Il fut également décidé que les noms d'entreprises ou de lieux ne devaient pas être traités, ni les affaires judiciaires. Ces termes feront l'objet d'un traitement spécifique à un autre thésaurus ultérieurement. De même que les renvois n'étaient pas traités car le danger était que si le descripteur auquel il renvoyait changeait de nom dans le nouveau thésaurus, ce même descripteur aurait pu être traité une deuxième fois dans la liste des termes. Cela aurait constitué un doublon dans l'arborescence mais aussi faire perdre en précision lors d'une recherche.

Lors de cette réélaboration, nous avons été confronté à certains types de problèmes tels que les définissent J.Aitchison et A.Gilchrist⁷, notamment certains relevant de problèmes de spécificité et d'autres de problèmes d'exhaustivité.

⁷ Voir page 15

2.2.1 Le problème de spécificité

Comme il a été défini auparavant, le problème de spécificité se présente quand un des deux thésaurus possède une arborescence davantage développée que l'autre. C'est à dire qu'il possède plus de termes spécifiques pour un même concept présent dans les deux thésaurus. En d'autres termes, dans notre cas, le thésaurus de l'Eurovoc présente un degré de spécificité supérieur à ceux de l'Infodoc.

Tout d'abord, le terme ARMÉ – SERVICE MILITAIRE ne possédait aucun terme spécifique* dans les thésaurus de l'Infodoc. Par contre celui de l'Eurovoc en possède cinq : INSOUMISSION, OBJECTION DE CONSCIENCE, SERVICE CIVIL, SERVICE FÉMININ et SERVICE VOLONTAIRE. Il fut décidé que d'une part SERVICE MILITAIRE serait changé en SERVICE NATIONAL (terme issu de l'Eurovoc) et que d'autre part seul un des termes spécifiques présent dans l'Eurovoc serait conservé. En effet, le terme OBJECTION DE CONSCIENCE était présent dans le thésaurus belge. Eurovoc présentait dans ce cas plus de termes spécifiques que les thésaurus de l'Infodoc.

Un autre cas de ce type que nous avons rencontré était toujours dans le terme générique* d'ARMÉE. Il s'agissait de l'armée de terre. Les thésaurus de l'Infodoc ne possédaient pas le terme spécifique : ARMÉE DE TERRE. Pourtant le thésaurus de l'Eurovoc le prévoyait en terme spécifique à ARMÉE. Tous les documents ayant trait à l'armée de terre étaient rangés dans le dossier ARMÉE qui regroupait tout ce qui touchait à l'armée en général ainsi qu'à l'armée de terre. Cependant dans une optique de réorganisation de thésaurus et d'augmentation de la précision et du rappel lors d'une recherche, le besoin de créer ce terme se faisait sentir. Il fut créé dans le nouveau thésaurus. Dans ce cas aussi l'Eurovoc présente une spécificité que les thésaurus de l'Infodoc ne possède pas.

D'autres termes ont posé des problèmes de spécificité. Mais parmi les termes traités certains ont posé un autre problème : celui de l'exhaustivité.

2.2.2 Le problème de l'exhaustivité

Rappelons que le problème de l'exhaustivité signifie que le thésaurus de l'Eurovoc possède un degré de spécificité moindre que ceux de l'Infodoc.

Pour le terme ANIMAL par exemple, l'Eurovoc ne possédait pas un degré de spécificité qui se trouvait dans les deux thésaurus de l'Infodoc. Ainsi il a fallu créer des termes spécifiques pour le thésaurus. Ainsi durant cette étape une réorganisation des termes fut effectuée, dans les thésaurus du service Infodoc, les animaux étaient dénommés par leur nom d'espèce (chien, chat, cheval...). Mais dans le nouveau thésaurus les animaux furent renommés par le genre. Ainsi chien devenait canidé, chat devenait félin... Cette distinction, bien qu'elle n'était pas nécessaire à la résolution du problème de l'exhaustivité, fut utile pour différencier les canidés sauvages (chiens sauvages, renard...) des canidés domestiques par exemple.

Parmi les autres termes qui posaient ce problème d'exhaustivité, nous pouvons citer le terme ARME – MINES présent dans les thésaurus de l'Infodoc. Le thésaurus de l'Eurovoc quant à lui possédait le terme ARME ANTIPERSONNEL⁸, lequel regroupait aussi bien les mines que les « armes et engins employés contre les hommes et non contre le matériel » (comme le stipule sa note d'application). C'est à dire tout ce qui relève des mines, des grenades, des bombes à sous-munition... Le choix de faire une distinction entre les différents types d'armes antipersonnel s'offrait à nous mais il fut décidé de regrouper tout ces types d'armes dans un seul dossier, l'ajout de nouveaux documents ne se faisant pas tous les jours. Au final le terme MINES fut remplacé par un terme plus générique à savoir ARME ANTIPERSONNEL.

Le descripteur ARME – STAND DE TIR présent dans le thésaurus belge posa le problème d'exhaustivité. Le thésaurus de l'Eurovoc ne prévoyait pas de place pour ce terme dans son arborescence. Mais afin de l'insérer dans le nouveau thésaurus, il fallait d'abord savoir ce que signifiait ce descripteur : s'agissait-il de stand de tir de professionnel de la sécurité, de police, ou seulement de tir sportif ? Après avoir examiné le dossier physique, il nous est apparu que ce terme regroupait ce qui touchait au tir sportif. Dès lors, son insertion dans le nouveau thésaurus s'en trouva facilitée. Après une phase de réflexion il fut décidé de le créer en tant que terme spécifique de ÉQUIPEMENT SPORTIF.

⁸ ARME ANTIPERSONNEL ou ARME ANTIPERSONNELLE existe tout les deux, le terme fut choisi conformément à l'orthographe de l'Eurovoc.

2.2.3 D'autres types de problèmes ?

Cependant malgré la description des problèmes de J.Aitchison et A.Gilchrist pouvant se présenter durant une réingénierie de thésaurus, d'autres problèmes se sont présentés.

Le terme AGROGLYPHE⁹ posa un problème basé sur la signification même de ce terme. C'est à dire que ce terme peut soit relever du surnaturel ou de la supercherie : tout dépend de la personne. Une solution fut trouvée en considérant les agroglyphes davantage comme des phénomènes inexplicables. Cependant, d'autres concepts comme les superstitions auraient à s'y ajouter. Dès lors il fut considéré de créer un terme générique comme CROYANCE POPULAIRE¹⁰ qui permettrait de regrouper sous un même descripteur tout les phénomènes inexplicables mais aussi ce qui n'était pas considéré comme rationnel. Mais le thésaurus de l'Eurovoc ne prévoyait pas de place dans son arborescence pour une telle notion. La difficulté ici fut de lui trouver une place dans cette arborescence. Ce terme fut mis en terme spécifique de CULTURE ET RELIGION. Le problème relevait ici du fait que l'Eurovoc prévoit son arborescence selon des critères liés à son usage. Il est créé pour des organisations gouvernementales et officielles mais pas pour un centre de documentation comme le service Infodoc. Nous touchons ici à une des limites du thésaurus de l'Eurovoc.

Un problème d'obsolescence des termes est apparu. Ce problème lors de ce travail de réélaboration est dû au fonds du service Infodoc. En effet même si un tri des termes fut effectué en amont de cette réingénierie. Certains termes comme ARMÉE – LANGUES furent jugés obsolètes et les dossiers physiques furent classés en archives. En effet, pour presque chaque terme le dossier physique était consulté afin de vérifier les notions qu'il regroupait. Lors de ces vérifications il apparaissait que certains dossiers n'avaient plus été alimentés depuis plusieurs décennies. En accord avec les service Infodoc il fut décidé que quand un tel cas se présentait il serait classé en archives.

⁹ Il s'agit des cercles de culture aussi appelés des crop circle.

¹⁰ À la fin du stage, la dénomination finale n'était pas encore décidée. Mais le concept de croyance populaire était retenu.

Le problème rencontré avec le terme ARGENT fut un problème d'homophonie. Il concernait surtout les deux thésaurus de l'Infodoc. Ce même terme était présent dans les deux mais il recouvrait des notions différentes. Dans le thésaurus thématique belge, il désignait l'argent dans le sens de la monnaie tandis que dans le thésaurus thématique étranger il recouvrait le métal précieux. Dès le départ il a fallu les différencier et cette différenciation fut ajoutée aux descripteurs des dossiers physiques. Une fois cette distinction faite, trouver leur place dans l'Eurovoc ne posa pas de problème particulier car il prévoyait les deux. L'un était désigné sous le terme générique de MÉTAL PRÉCIEUX et l'autre par ARGENT MONÉTAIRE.

Avant de clore cette partie il est important de préciser que le nombre de termes ayant posé le problème de l'exhaustivité est plus important que les termes ayant posé des problèmes de spécificité. Il nous apparaît à ce stade du traitement des termes des thésaurus de l'Infodoc que le thésaurus de l'Eurovoc n'est pas aussi spécifique que ceux du service Infodoc.

2.3 DU CHOIX DES TERMES À LEURS PLACES DANS LE THÉSAURUS

À partir de cette opération de réingénierie de thésaurus, il nous apparaît que les principales difficultés auxquelles nous avons dû faire face ont été les problèmes d'exhaustivité et de spécificité. Les problèmes de termes composés, de synonymes (employé pour) et de relations associatives ne se sont pas posés. Cependant seuls les termes commençant par la lettre A ont été traités. Par la suite ces problèmes pourraient se présenter. À travers ces difficultés il nous apparaît qu'il faut être rigoureux dans la restructuration des hiérarchies et dans le choix des termes.

D'autres problèmes autres que ceux de l'exhaustivité et de la spécificité ont été rencontrés comme cela a été vu précédemment avec AGROGLYPHE et ARGENT. Pour ce dernier, nous avons vu qu'il recouvrait dans les deux thésaurus de l'Infodoc des notions différentes (problème d'homophonie). Cependant il ne faut pas oublier que les critères définis par J.Aitchison et A.Gilchrist ne furent créés que dans le cadre d'une réélaboration d'un thésaurus par rapport à un autre. Dans le cas que nous avons traité, il s'agissait de partir de deux thésaurus afin d'en élaborer un autre d'où un degré de difficulté supplémentaire.

De plus, il a fallu tout réorganiser aussi bien, tant au niveau des descripteurs qu'au niveau des dossiers physiques afin d'assurer une pleine compatibilité entre le nouveau et les anciens thésaurus. En plus de ces derniers problèmes, le choix des termes demandait également de la réflexion. N'oublions pas qu'un descripteur « est l'unité de représentation d'un contenu qui est de nature conceptuelle [et il] est considéré comme clé d'accès aux documents dont il synthétise le contenu. » (MUSTAFA EL HADI, 2006). Derrière un terme se trouve une notion, un concept. Un terme d'un thésaurus ne doit recouvrir qu'un seul concept. Et ce terme se doit être cohérent par rapport à ce qu'il recouvre et ne doit pas poser d'ambiguïté quant à son sens. Surtout s'il s'agit d'un terme générique, car tous ses termes spécifiques auront à se retrouver dans ce concept représenté par leur terme parent.

Le choix du terme doit aussi se faire par rapport à l'utilisateur potentiel ou futur du thésaurus. C'est ce dernier qui mènera la recherche documentaire. User de mots trop compliqués ou peu utilisés pour les termes risquerait de mener à du bruit ou à du silence. Les termes choisis ne doivent pas être trop restrictifs. Il faut qu'ils soient le plus accessibles possible à l'utilisateur final, c'est à dire dans le cas présent un journaliste de la RTBF. Le but est de lui faciliter la recherche d'information. Nous pourrions toujours utiliser des termes équivalents pour renvoyer à un descripteur, mais l'idéal est d'avoir recours au minimum à un terme équivalent. Lors de l'élaboration d'un thésaurus connaître d'abord le futur utilisateur permet d'éviter un fastidieux travail de recherche de terme équivalent pour un descripteur. Chaque individu sera susceptible d'utiliser un terme différent pour une même notion. Par exemple pour un véhicule l'utilisateur pourrait utiliser le terme VOITURE aussi bien que le terme AUTOMOBILE. Pour le thésaurus de l'Infodoc l'utilisateur futur sera un documentaliste qui avec la pratique se familiarisera avec ce nouveau thésaurus. Ainsi l'accent a été mis sur l'inclusion en terme équivalent des descripteurs utilisés dans les anciens thésaurus afin d'en faciliter la prise en main. Il faut éviter le plus les recherches d'informations infructueuses.

Selon Mustafa El Hadi (MUSTAFA EL HADI, 2006), l'indexation d'un document « consiste en la caractérisation de son contenu à l'aide de descripteurs afin de pouvoir retrouver aussi facilement que possible dans la masse documentaire les informations pertinentes qu'il contient ». Les descripteurs du thésaurus doivent donc correspondre au fonds du service Infodoc. Comme l'indexation et la recherche se feront par rapport à ce thésaurus les termes ne doivent pas poser de problème et doivent être clairs pour les documentalistes mais aussi pour les journalistes. Ce thésaurus doit être en mesure de proposer des termes correspondant à ce que le document concerné recouvre. Les notions et les concepts du document doivent pouvoir se retrouver dans le thésaurus. L'exhaustivité du nouveau thésaurus élaboré sera un facteur déterminant pour une indexation efficace. Partir des anciens thésaurus permet justement d'avoir une ouverture assez large des concepts et d'être en accord avec le fonds existant de l'Infodoc. Cependant le choix du terme sera un jugement subjectif, il ne faut donc pas que le sens de deux descripteurs soit trop proche. Le but d'un thésaurus étant de mener à une recherche de document plus efficace.

À partir de cette réflexion, nous pouvons affirmer que le choix du descripteur lors de l'indexation, quand bien même est-il fait par un documentaliste doit être partagé par ses collègues et autres usagers du service. Ce choix de terme doit être réfléchi et fait en fonction des autres descripteurs du thésaurus. À cette difficulté s'ajoute celle de la réorganisation des arborescences. Il a fallu réorganiser les hiérarchies tout en conservant au maximum les notions et les concepts en accord avec le fonds existant. Mais cette réorganisation a mené à une modification inévitable des descripteurs afin de garantir la cohérence des nouvelles hiérarchies. L'enjeu était donc que cette altération des arborescences d'origine ne modifiait qu'au minimum les termes existants regroupant les concepts.

En plus de la pertinence du choix des termes, se pose aussi la question du choix du thésaurus Eurovoc comme base du nouveau thésaurus. Celui-ci a posé de nombreux problèmes notamment au niveau de sa spécificité. L'exemple le plus flagrant est celui du terme ACCIDENT INDUSTRIEL. Celui-ci existait dans les thésaurus thématiques d'Infodoc mais l'Eurovoc commence par directement par spécifier le type d'industrie. Dès lors il a fallu prendre des libertés face aux anciens thésaurus. Ce terme fut placé dans le nouveau thésaurus mais en ayant d'abord fait un travail de tri entre les différents types d'accidents industriels au niveau de l'ancien thésaurus, ceci afin de garantir un lien entre eux. L'un des objectifs de cette réingénierie était de pouvoir se référer au nouveau thésaurus pour pouvoir retrouver un document classé avec l'ancien thésaurus.

De plus l'Eurovoc possède des termes et une arborescence qui ont parfois prêté sujet à caution. Le terme ANIMAL DOMESTIQUE est dans l'Eurovoc placé dans le domaine 5631 ACTIVITÉ AGRICOLE sous le terme générique de PRODUCTION ANIMALE. Ainsi nous pouvons penser que l'Eurovoc considère les animaux domestiques comme des moyens de production. Or cette place des animaux domestiques ne nous paraissait pas pertinente. Les dossiers traitant des animaux domestiques n'avaient pas de documents les considérant comme des animaux de production. Nous leur avons donc trouvé une autre place qui répondait à la fois à une logique face au fonds documentaire mais aussi par rapport à toute l'arborescence déjà développée.

Cependant le thésaurus de l'Eurovoc a été créé pour des organismes qui ont des buts et qui assurent des missions différentes du service de l'Infodoc. La raison d'être de l'Eurovoc n'est pas la même que le service Infodoc. Cette différence de fonds mais aussi

de contexte d'utilisation rend ce thésaurus incompatible avec l'Infodoc. Le principal enjeu était de remodeler l'Eurovoc afin de le rendre conforme aux besoins du service Infodoc. Même en prenant un thésaurus d'un centre de documentation d'un organe de presse, des problèmes se seraient présentés. Cependant le service Infodoc a cherché un autre thésaurus d'un service de documentation comme lui mais aucun thésaurus satisfaisant à la spécificité de l'Infodoc ne fut trouvé. En créer un permet également de disposer d'un thésaurus en accord avec le fonds existant et répondant aux critères de classification propres à l'Infodoc.

En conclusion de cette partie, nous pouvons nous demander si le thésaurus de l'Eurovoc était un choix pertinent. Il existe d'autres thésaurus comme MOTBIS¹¹ ou UNBIS¹². Chacun a ses spécificités et chacun fut élaboré pour un public différent, et sans nous plonger dans leurs études nous pouvons tout de même affirmer qu'aucun thésaurus n'aurait pu se transposer tel quel au fonds du service Infodoc. Un thésaurus correspond d'abord à un fonds documentaire et à des usagers. Il aurait été impossible de trouver un thésaurus répondant aux attentes du service Infodoc sans passer par une phase de modification. Concevoir un nouveau thésaurus seulement à partir du fonds existant aurait demandé un long travail qui aurait comporté des risques comme des incohérences dans l'arborescence et dans les termes. Partir d'un autre thésaurus comme l'Eurovoc permet d'avoir une base solide à défaut de ne pas être totalement cohérent avec le fonds de l'Infodoc.

¹¹ MOTBIS est un thésaurus centré sur l'éducation. Voir <http://www.motbis.fr/>

¹² UNBIS est un thésaurus recouvrant les domaines d'actions des Nations Unies. Voir <http://unbisnet.un.org/>

3. UTILISATION D'UN LOGICIEL DE GESTION DE THÉSAURUS : TEMATRES

La réélaboration de ce thésaurus s'est accompagnée d'une création d'une base de données réunissant les différents descripteurs, les non-descripteurs, les associations, les niveaux de hiérarchie... Cet encodage des différents termes de ce thésaurus dans un outil informatique s'est basé sur le logiciel TemaTres.

3.1 TEMATRES

TemaTres est un logiciel servant à gérer un langage documentaire comme les thésaurus. Il s'agit d'un logiciel libre élaboré par Diego Ferreyra (expert en bibliothéconomie et en sciences de l'information en Argentine). Ce logiciel possède quelques fonctionnalités comme une suggestion dans la recherche des termes, une classification des termes en terme accepté, terme rejeté et terme candidat ou encore une navigation par domaine ou par ordre alphabétique.

L'installation de TemaTres nécessite un minimum de connaissances en informatique. Il faut l'installer dans une base de données comme Xampp¹³ par exemple. En outre, des connaissances en Php* et SQL* sont tout de même conseillées sans être primordiales. Ajoutons également qu'il est possible de modifier l'interface graphique à condition d'avoir des connaissances en langage CSS*.

Comme il a été dit précédemment, l'encodage des termes dans TemaTres s'est basé sur les deux premiers niveaux du thésaurus de l'Eurovoc qui possède 21 domaines et 127 micro-thésaurus. Cet outil informatique est assez flexible dans l'insertion des termes au sein de sa base de données.

¹³ XAMPP est un ensemble de logiciels permettant de mettre en place facilement un serveur web et FTP. Voir <http://www.apachefriends.org/fr/xampp.html>

3.2 L'EXPLOITATION DE TEMATRES

L'exploitation d'un logiciel de gestion de thésaurus tel que TemaTres ne demande pas de connaissances particulières en informatique. Cependant comme il a été dit précédemment des connaissances en Php et en SQL sont souhaitables pour qui voudrait l'installer sans avoir recours à un service informatique pour peu que la structure où TemaTres prends place en dispose d'un. N'oublions pas qu'il s'agit d'un logiciel libre mais qui dit libre ne dit pas gratuit. En effet pour une petite structure sans service informatique dédié faire le choix de TemaTres impose de disposer de connaissances techniques.

En se basant sur les critères de comparaison de logiciel de thésaurus de Leonard Will¹⁴, Sylvie Dalbin (DALBIN, 2007) définit les fonctionnalités des outils informatiques permettant l'exploitation de thésaurus. En effet, le choix d'un outil informatique doit se faire selon plusieurs facteurs comme les types de relations gérés par le logiciel, l'interface, la gestion des termes, l'ergonomie... mais aussi par rapport au budget, au personnel qui sera dédié à la gestion de ce logiciel. N'oublions pas qu'il n'y a que sept documentalistes à l'Infodoc et que l'activité de ce service est tributaire de l'actualité. Créer un thésaurus demande du temps et du personnel, la création de ce nouveau thésaurus qui a pour base Eurovoc et les anciens thésaurus de l'Infodoc exige beaucoup de temps. Il est tout à fait dans l'intérêt de ce service de disposer d'un outil simple à prendre en main et permettant une gestion de thésaurus sans difficultés particulières.

Les possibilités d'importation et d'exportation sont là aussi très importantes. Pouvoir effectuer des sauvegardes et des exportations permet d'avoir une indépendance face au système utilisé mais aussi de disposer de sauvegarde en cas de panne par exemple.

Confrontons TemaTres à ces critères pour juger de la pertinence du choix de ce logiciel de gestion de thésaurus. À partir des critères proposés par Dalbin, n'en retenons que les quatre plus pertinents que sont la description du vocabulaire, la création et la gestion du thésaurus, la visualisation et l'exportation de données et la recherche et l'indexation.

¹⁴ WILL.L, *Comparison of thesaurus management software for Pcs*, site internet : <http://www.willpowerinfo.co.uk/thessoft.htm>

3.2.1 Description du vocabulaire

Tout d'abord, pour ce qui est **des termes**, TemaTres autorise aux termes d'avoir un statut différent les uns des autres. Les termes peuvent être des termes candidats, des termes rejetés, des descripteurs, des non-descripteurs et des termes libres. L'utilisation de termes candidats et de termes rejetés permet une gestion et une maintenance facilitées du thésaurus. Lors de l'insertion des termes dans la base de données cette distinction entre les différents types de termes s'est avérée utile, notamment lorsque nous hésitions sur l'inclusion ou non d'un descripteur. Par exemple, pour différencier les différents types d'accidents domestiques, des termes libres décrivant le type d'accident ont été créés comme NOYADE, BRÛLURE... En les associant au descripteur ACCIDENT DOMESTIQUE à un document évoquant un accident domestique par noyade, cela permet d'améliorer les résultats d'une recherche portant sur les noyades en milieu domestique.

En outre, ce logiciel autorise l'**élaboration de micro-thésaurus** classés par thèmes. Ce qui s'est avéré utile lors de l'insertion des deux premiers niveaux du thésaurus de l'Eurovoc. De plus les termes composés sont acceptés comme descripteurs. Dans le nouveau thésaurus des groupes de mots ont bien sûr toute leur place comme ACCIDENT PAR TRANSPORT MARITIME. Ces termes composés permettent d'avoir un degré de spécificité supplémentaire ce qui nous donne une précision accrue des concepts et permet de réduire le bruit lors d'une recherche. En considérant le descripteur précédent, il nous apparaît qu'associer ACCIDENT avec TRANSPORT MARITIME élimine tout les autres types d'accidents que ce soit les accidents par transport aérien ou encore les accidents domestiques. La capacité du logiciel à accepter les termes composés est un facteur à prendre en compte pour qui voudrait assurer une recherche efficace à son fonds documentaire.

Pour ce qui est **des relations**, ce logiciel permet bien évidemment les relations hiérarchiques mono-hiérarchiques mais il autorise aussi les relations poly-hiérarchiques, c'est à dire qu'un terme spécifique possède plusieurs termes génériques. Mais dans l'arborescence de ce nouveau thésaurus, aucune relation poly-hiérarchique n'a été créée. Cependant dans le cadre d'une évolution de ce thésaurus, pouvoir créer ce type de relation sans devoir nécessairement changer d'outil technique peut être intéressant.

Il n'y a pas de limites dans le nombre de **niveaux hiérarchiques**. La construction de ce nouveau thésaurus a pu se faire sans aucune limite de niveaux hiérarchiques ce qui nous a donné une certaine liberté dans l'inclusion des descripteurs. Un descripteur dans TemaTres peut aussi avoir une **relation d'équivalence** avec un ou plusieurs non-descripteurs. De plus un descripteur peut aussi avoir une relation associative avec un terme libre*. Ce type de relation associative a notamment été utilisé pour associer les types d'accidents domestiques (noyade, brûlure...) avec le descripteur ACCIDENT DOMESTIQUE¹⁵.

3.2.2 Création et gestion du thésaurus

Pour l'utilisation de TemaTres, la **saisie de plusieurs termes** de même niveau en une seule fois est autorisée. L'import et l'export peut se faire sous plusieurs formats tels que le XML*, le RDF* ou encore le SQL¹⁶. Cette multiplicité de ces différents formats permet de ne pas être dépendant de TemaTres et de pouvoir migrer si le besoin s'en faisait sentir vers un outil de gestion de thésaurus tout en garantissant l'intégrité structurelle du thésaurus.

Nous pouvons également **modifier un terme** même après l'avoir inséré dans une arborescence. Nous pouvons aussi traiter les arborescences par lots, c'est à dire décrocher un terme et ses termes spécifiques pour les adjoindre à un autre terme. De plus en cas de suppression d'un descripteur il faut d'abord supprimer ses liens avec les autres termes (générique, associé et spécifique). Cependant ces termes spécifiques ne seront pas automatiquement reliés au terme parent du descripteur effacé.

Les **termes doublons** peuvent être repérés assez facilement, TemaTres propose une catégorie où les termes répétés sont répertoriés par liste alphabétique. De plus la cohérence des relations associatives est garantie par TemaTres, c'est à dire qu'une relation associative sera toujours réciproque sans qu'il y ait besoin d'intervenir sur les deux termes.

¹⁵ Voir page précédente

¹⁶ Voir <http://tematres.r020.com.ar/wiki/doku.php?id=tematres:features>

Enfin, TemaTres autorise l'**édition de note** accolée à un terme : il peut s'agir d'une note d'application, d'une note historique, bibliographique, de catalogage ou privée. Ces notes ont permis de spécifier l'usage d'un descripteur. Par exemple pour le descripteur MATÉRIEL MILITAIRE, la note d'application : « concerne tout matériel militaire qui n'est pas de l'armement » lui fut accolée. Cela permet d'indiquer à l'indexeur et à l'utilisateur quelle notion se trouve derrière ce descripteur.

3.2.3 Visualisation et exportation de données

En se penchant sur **la visualisation des termes**, nous nous rendons compte que TemaTres permet une visualisation par une liste alphabétique et par une liste thématique. Lors de la visualisation d'un terme, toutes ses relations apparaissent également, que ce soit ses équivalences, ses associations, son terme générique et ses niveaux antérieurs et ses termes spécifiques¹⁷, ainsi que toute note éventuelle qui lui aurait été adjointe. Cette vue d'ensemble permet d'avoir une vision globale de la place du descripteur dans l'arborescence. Visuellement nous pouvons capter toute information utile se rapportant à ce terme.

Ce logiciel possède aussi plusieurs **formats d'exportations** (voir précédemment), pourtant l'exportation est possible directement depuis la visualisation d'un terme particulier mais dans ce cas cette exportation ne concernera que le terme présenté (ainsi que son terme générique et ses termes spécifiques sur un seul niveau). Ce dernier point est intéressant, car il permet l'export d'une seule partie de l'arborescence. Cette flexibilité supplémentaire dans l'exportation de données permet par exemple si le besoin s'en faisait sentir de ne conserver ou de dupliquer qu'une seule partie du thésaurus. Cela peut prendre tout son sens dans le cadre d'un échange avec un autre organisme n'étant intéressé que par une partie de l'arborescence.

¹⁷ Voir image de l'interface de TemaTres en annexe

3.2.4 Recherche et indexation

La recherche de TemaTres autorise les troncatures ainsi que les mots partiels. Mais par contre elle ne prends pas en compte la casse. La recherche s'en trouve facilitée. Lors de la recherche d'un descripteur, taper « accident transport » permet de trouver les termes ayant trait aux accidents par transport maritime, transport routier, transport aérien... Cette possibilité de recherche que présente TemaTres permet d'augmenter le rappel.

De plus, lors de la frappe d'un terme le logiciel nous suggère des termes présents dans sa base de données. Cette **suggestion de terme** qui se fait à l'image de celle du moteur de recherche Google permet de guider la recherche de l'utilisateur mais aussi d'avoir un aperçu de ce que propose la base de données en tapant le terme recherché.

À ce dernier critère de jugement de logiciels de gestion de thésaurus, nous pouvons ajouter que TemaTres propose pour une visualisation d'un terme des liens internet vers Google et certains de ses services associés comme Google books, Google images et Google scholar mais aussi vers Wikipédia. Cette possibilité que présente cet outil permet une ouverture vers le web élargissant ainsi le fonds documentaire vers l'internet.

3.3 TEMATRES ÉTAIT-IL UN CHOIX PERTINENT?

Après avoir considéré ces précédents critères de jugement de fonctionnalité d'un logiciel de gestion de thésaurus, il ne faut pas perdre de vue que le but d'un tel outil est d'obtenir une exploitation satisfaisante du thésaurus pour permettre une indexation et une recherche efficace de documents.

Même si l'outil est efficace, la qualité de l'indexation reste fonction des compétences du documentaliste, l'Infodoc ne disposant pas d'outil d'indexation automatique. Mais si un document est mal indexé, l'utilisateur aura davantage de difficulté à le retrouver voire même qu'il ne pourra pas le retrouver. Connaître les descripteurs du nouveau thésaurus est primordial pour une bonne indexation. Un avantage que présente TemaTres est la visualisation du descripteur au sein de son réseau sémantique d'un affichage sur un seul écran. Cet aspect pouvant paraître anodin prend tout son sens quand il s'agit de ne pas perdre de temps à explorer tout le réseau sémantique d'un descripteur particulier.

De plus un usager non-documentaliste pourrait être amené à effectuer des recherches sur TemaTres. Lors de cette recherche documentaire, la suggestion des termes et la prise en compte de troncature permet d'assister cet usager. TemaTres facilite la recherche et l'exploitation du fonds documentaire. Pour ce nouveau thésaurus, TemaTres reste son support d'exploitation. En ce sens si l'outil ne permet pas une pleine utilisation de ce nouveau thésaurus, il perd de son intérêt. Or TemaTres en ayant la suggestion des termes ou encore en autorisant leur troncature notamment, facilite la recherche de tout usager ne connaissant pas l'orthographe exacte des termes.

L'interface de base de TemaTres reste simple et ergonomique. Elle ne nécessite pas d'apprentissage particulier de la part de l'utilisateur. Cependant il ne faut pas oublier que TemaTres pourrait être à terme incorporé dans une solution logicielle de GED. De fait la modification possible de son interface lui permettrait de s'insérer dans cette solution tout en s'adaptant à sa charte graphique. Ainsi son incorporation dans celui-ci en sera facilitée et son utilisation par les usagers également.

Cette phase d'encodage fut facilitée par des fonctions que présente TemaTres. L'édition des termes par ce logiciel a permis un gain de temps notamment lorsque qu'il s'agissait de modifier des descripteurs déjà insérés dans l'arborescence. Ce dernier point permettra aussi un changement possible des descripteurs durant l'avancement du projet. D'autres termes seront insérés durant le développement de cette arborescence, pouvoir en modifier certains pour éviter par exemple des problèmes de doublons permet un développement moins contraignant. De plus pouvoir classer les termes en termes libres, termes candidats, en termes rejetés... donne une flexibilité accrue à la création et à la maintenance du nouveau thésaurus.

Au regard du service Infodoc et de son fonds documentaire, TemaTres constitue un bon outil. D'une part, sa facilité d'utilisation permet une indexation et une recherche plus efficiente. Car même si c'est le documentaliste de l'Infodoc qui effectue cette phase d'assignation de descripteur à un document, il n'en reste pas moins que c'est le logiciel de par son ergonomie qui facilitera l'accès aux différents descripteurs présents dans la base de données. D'autre part, tout usager comme il a été dit précédemment aura sa phase de recherche facilitée par l'interface ergonomique de TemaTres.

Pour ce qui est de la création du thésaurus, l'avantage de l'utilisation d'un outil informatique n'est plus à démontrer. Que ce soit dans le nombre de descripteurs pouvant être créé, ou dans la maintenance qui en est facilitée. Cependant un logiciel tel que TemaTres présente tout de même quelques aspects qu'il est important de prendre en compte.

Le premier aspect est qu'il s'agit d'un logiciel issu de l'open-source. Il est gratuit mais il ne possède pas de support technique en cas de dysfonctionnement comme peuvent le présenter des logiciels propriétaires. La présence ou non d'une personne ayant des bonnes connaissances informatiques ou la présence d'un service informatique au sein de la structure où le logiciel prendra place est à prendre en compte. Le service Infodoc bénéficie de la présence du service informatique de la RTBF. Cependant durant ce stage il nous est apparu que le service informatique n'intervenait pas sur TemaTres. Cela est dû au fait que ce logiciel en est encore au stade de test et que pour l'instant il ne relève qu'exclusivement du service Infodoc.

Le deuxième aspect concerne la capacité pour les documentalistes de l'Infodoc à s'adapter à ce nouvel outil. Quand bien même l'interface de TemaTres est-elle ergonomique, il faut prendre en compte le délai d'adaptation au nouveau thésaurus. Il ne faut pas oublier que les documentalistes ont acquis des automatismes pour l'indexation des documents au fil des années. Avec l'expérience le questionnement sur le choix d'un descripteur ne se pose plus, le documentaliste les connaît par cœur. Cependant des descripteurs sont créés avec ce nouveau thésaurus. Un délai sera certainement nécessaire pour une utilisation pleine et satisfaisante de ces nouveaux descripteurs.

Nous pouvons terminer cette partie en déclarant que TemaTres répond aux attentes de l'Infodoc ou du moins à l'élaboration du nouveau thésaurus. L'utilisation de cet outil informatique facilite l'indexation. Sa visualisation qui permet de capter le réseau sémantique d'un terme, l'ergonomie de son interface et les possibilités de recherches nous autorise une exploitation efficace du thésaurus. De plus la modification des termes et pouvoir les classer par types (candidat, rejeté ou libre) nous permettent d'obtenir un encodage facilité et une certaine souplesse dans la maintenance future du thésaurus.

CONCLUSION

Ce projet de réélaboration de thésaurus n'a pas été sans difficulté. Différents problèmes ont été rencontrés que ce soit celui de l'exhaustivité ou celui de la spécificité. D'autres se sont présentés à nous comme celui de l'homophonie : deux descripteurs partagent la même orthographe et la même prononciation mais ne partagent pas le même concept.

Le service Infodoc a besoin d'un nouveau thésaurus. La nécessité de disposer d'un thésaurus cohérent et efficace s'est rapidement imposée à nous lors du dépouillement de la presse quotidien et de l'assignation des descripteurs aux documents. Le fonds du service Infodoc se doit d'être valorisé or ce service ne dispose que de peu de moyens humains et financiers. Cela ne l'empêche pourtant pas de se lancer dans ce projet de grande importance tant pour son efficacité et sa visibilité. La numérisation devrait permettre à tout personnel de la RTBF de lancer une recherche depuis son poste de travail et ainsi d'avoir accès au fonds de l'Infodoc. Cependant pour une recherche efficace, encore faut-il que les outils choisis (Eurovoc et TemaTres) soient assez pertinents pour permettre une élaboration de thésaurus efficace.

Comme nous l'avons vu l'Eurovoc a des termes et une arborescence qui peuvent parfois prêter à caution. Cependant l'inclusion de ses deux premiers niveaux dans le nouveau thésaurus permet à celui-ci de disposer d'une base solide et suffisamment large pour englober bon nombre de concepts. Pourtant certains termes comme AGROGLYPHE n'ont pas de place dans l'Eurovoc. Cela est certainement dû au caractère atypique de ce terme. Mais au vu de ce qui a déjà été entrepris dans cette opération de réingénierie, ce type de terme est très peu fréquent.

En ce qui concerne TemaTres, son utilisation n'a souffert d'aucun problème technique. De plus de par son utilisation simple et son interface ergonomique nous en concluons qu'il s'agit d'un outil tout à fait adapté et pertinent à cette réingénierie de thésaurus.

Pourtant, développer un thésaurus demande des moyens financiers, du temps, et des moyens humains. Sur ce premier point le service Infodoc n'en dispose que de très peu. Pour ce qui du temps et des moyens humains l'Infodoc doit composer avec un nombre restreint de documentalistes qui ont à faire le dépouillement et le traitement des demandes des usagers : services de base de l'Infodoc. Pourtant ce thésaurus permettrait une amélioration quantitative de ces services de base : plus de demandes d'usagers pourront être traitées car l'accès aux documents voulus ne se ferait plus que par voie numérique. Cependant nous pouvons tout de même nous questionner sur la capacité du service Infodoc à continuer ce projet et à le finaliser avant la numérisation.

BIBLIOGRAPHIE

AITCHISON.J, *Words in the mind: an introduction to the mental lexicon*. Oxford, Blackwell. 1987

AITCHISON.J, GILCHRIST.A, *Construire un thésaurus – Manuel pratique*, ADBS Éditions, 1992, 222p

CHICHEREAU.D, et al. « Les normes de conception, gestion et maintenance de thésaurus. Évolutions récentes et perspectives » in *Documentaliste-Sciences de l'information*, janvier 2007, numéro 1, volume 44, page 66-74, ADBS Édition.

CYROT.C, PREUSS.C, « Réingénierie de thésaurus : une étude de cas » in *Documentaliste-Sciences de l'information*, janvier 2009, numéro 3, volume 46, page 4-13, ADBS Édition.

DALBIN.S, « Thésaurus et informatique documentaires. Partenaires de toujours ? » in *Documentaliste-Sciences de l'information*, janvier 2007, numéro 1, volume 44, page 42-55, ADBS Édition.

HODGE.G, *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington (DC) : Council on Library and Information Resources, 2000. 43p

MANIEZ.J, « Terminologie et thésaurus: divergences et convergences », *Terminologies 76: Colloque international, Paris-La Défense, 15-18 Juin 1976*, p. IV39-IV50, La Maison du dictionnaire, Paris, 1977.

MUSTAFA EL HADI.W, « Indexation humaine et indexation automatisée: la place du terme et de son environnement ». In : BLAMPAIN.D, THOIRON.P, VAN CAMPENHOUDT.M, « Mots, termes et contextes », Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles 8-10 septembre 2005, coll. « Actualité scientifique », Éditions des Archives Contemporaines et Agence universitaire de la Francophonie

<http://europa.eu/eurovoc/sg/sga_doc/eurovoc_dif!SERVEUR/menu!prod!MENU?langue=FR>, consulté le 08/06/2010

<<http://www.r020.com.ar/tematres/index.en.html>>, consulté le 08/06/2010

GLOSSAIRE

CSS (Cascading Style Sheets) : est un langage utilisé dans la présentation de pages web.

Descripteur : il s'agit d'un mot ou un groupe de mots conceptuels utilisé pour l'emploi d'un thésaurus. L'indexeur se sert du descripteur pour indexer le document.

Format RDF : est un langage formel permettant une représentation standard des thésaurus.

Format XML : est un méta-langage informatique de balise extensible. Il permet notamment de séparer la forme du contenu d'un document.

GED (Gestion Électronique de Document) : il s'agit d'un système de gestion, d'archivage, de tri, et de traitement de documents numériques.

Non-descripteur (renvoi) : est un synonyme, quasi-synonyme ou variante orthographique d'un descripteur. Il n'est pas utilisé pour l'indexation mais sert de renvoi vers un descripteur.

Php (Hypertext Preprocessor) : est un langage de programmation qui permet la création de pages web dynamiques.

SQL : est un langage informatique permet de rechercher, d'ajouter, de modifier ou de supprimer des données dans les bases de données.

Terme associé : c'est un descripteur qui a une relation associative avec un autre descripteur dont il ne partage par nécessairement la même hiérarchie. Cette relation est basée sur un point commun comme un champ sémantique, un concept et son contraire...

Terme générique (terme parent) : il s'agit d'un descripteur qui représente un concept principal et auquel sont subordonnés un ou plusieurs termes spécifiques dans une hiérarchie.

Terme libre : c'est un descripteur qui n'est rattaché à aucune hiérarchie.

Terme spécifique (terme enfant) : il s'agit d'un descripteur qui précise et identifie un concept à l'intérieur d'un concept plus général représenté par son terme générique.

ANNEXE

Exemple de fiche créée sur SharePoint

Bienvenue Cédric ROUGEAX

Collaboratif RTBF InfoDoc Actions du site

> CREATION NOUVEAU THESAURUS > AGENCE INTERNATIONALE DE L'ENERGIE

CREATION NOUVEAU THESAURUS: AGENCE INTERNATIONALE DE L'ENERGIE

Fermer

Nouvel élément | Modifier l'élément | Supprimer l'élément | M'avertir

TERME	AGENCE INTERNATIONALE DE L'ENERGIE
TERME_ISSU_DE	LISTE MOTS CLES BELGE
DESCRIPTEUR THESAURUS BELGE	AGENCE INTERNATIONALE DE L'ENERGIE
DESCRIPTEUR THESAURUS ETRANGER	AGENCE INTERNATIONALE DE L'ENERGIE (A.I.E)
DESCRIPTEUR EUROVOC	AGENCE INTERNATIONALE DE L'ENERGIE
DESCRIPTEUR RETENU	AGENCE INTERNATIONALE DE L'ENERGIE
DESCRIPTEUR CLOTURE LE	03/05/2010
RENVOI	
NOTE_APPLICATION	
REMARQUES	Voir Matières : A.I.E
ETAT_FICHE	TERMINE
TERME SUPPRIME	

Créé le 26/04/2010 14:03 par jmcisadmin
Dernière modification le 03/05/2010 16:32 par Cédric ROUGEAX

Fermer

Ces fiches étaient utilisées pour suivre l'évolution du thésaurus et la modification des termes. Sur une fiche se trouvent le descripteur traité, son thésaurus d'origine (belge ou étranger) et le descripteur correspondant au même concept dans l'autre thésaurus quand ce terme était présent dans les deux thésaurus de l'Infodoc. À cela s'ajoutaient le descripteur de l'Eurovoc et le descripteur retenu pour le nouveau thésaurus. La date de traitement du terme était indiquée et quelques remarques pouvaient y être ajoutées.

Ces fiches avaient pour rôle d'assurer une compatibilité entre les descripteurs d'origine et les nouveaux descripteurs.

Interface de TemaTres

L'interface de TemaTres, lors de la visualisation d'un terme, permet de saisir le réseau sémantique d'un descripteur (ici ARMÉE).

The screenshot displays the TemaTres web interface. At the top, there is a dark red header with the logo 'TemaTres' on the left and a search bar with the text 'Rechercher' on the right. Below the header, the main content area is white. On the left side, there is a vertical navigation menu with links: 'Liste thématique', 'Liste alphabétique', 'A propos de...', 'RTBF, INFODOC', 'Mon compte', and 'Déconnecter'. Below these links is a language dropdown menu set to 'français'. The main content area features a large heading 'armée' in red. Below the heading is a horizontal navigation bar with buttons: 'Ajouter', 'Options', 'Menu', and 'Administration'. To the right of this bar, the text 'Date de création: 27-Mai-2010' and 'Terme accepté: 27-Mai-2010' is displayed. Below the navigation bar is a breadcrumb trail: '► ACCUEIL ► 08 RELATIONS INTERNATIONALES ► 0821 DEFENSE ► armée'. The main content area lists a series of terms with their associated codes and expand/collapse icons: '[x]TG 0821 DEFENSE', 'armée', '[x]TS3 armée de l'air', '[x]TS3 armée de métier [-]', 'TS4 féminisation dans l'armée', 'TS4 recrutement', '[x]TS3 armée de terre [+]', '[x]TS3 discipline militaire [+]', '[x]TS3 force belge à l'étranger', '[x]TS3 marine militaire', '[x]TS3 matériel militaire', '[x]TS3 personnel militaire', and '[x]TS3 service national [+]'.

MADS Zthes SKOS-Core BS8723-5 XTM DC

Auteur: Diego Ferreyra
URI: http://localhost/tematres/
Powered by TemaTres 1.04

Sur cette capture d'écran, nous constatons que pour un descripteur, ses termes spécifiques (TS3) apparaissent, ainsi que ses termes génériques sur plusieurs niveaux. En outre la date d'acceptation du terme, les liens vers des sites web comme Google scholar ou Wikipédia et l'export vers différents formats apparaissent.

Mail de prise de contact avec Chalon et Ferreyra

Dans le cadre du test de fonctionnalité effectué en préambule du travail d'encodage, des échanges de mails avec Patrice Chalon et Diego Ferreyra ont eu lieu.

Patrice Chalon connaît bien cet outil car il a notamment traduit le site de Diego Ferreyra sur TemaTres (<http://www.r020.com.ar/tematres/setup.html>). De plus Monsieur Boly avait déjà pris contact avec Patrice Chalon dans le cadre de l'installation de ce logiciel à l'Infodoc.

Cedric R <cedric.rougeaux@gmail.com>

20 avril 2010 20:11

À: patrice.chalon@gmail.com

Bonjour,

Je suis étudiant actuellement en stage à la RTBF à Bruxelles. Leur service Infodoc souhaite développer un thésaurus à partir de TemaTres.

Je l'ai installé et il fonctionne sans problème. Mais je souhaiterais en savoir plus sur son fonctionnement. Notamment quelles sont les fonctions des différentes tables présentes dans la base de donnée « tematres_101 ».

Pourriez vous m'éclairer à ce sujet et auriez vous éventuellement de la documentation à me fournir?

Très cordialement,
Rougeaux Cedric

Patrice Chalon <patrice.chalon@gmail.com>

20 avril 2010 22:26

À: Cedric R <cedric.rougeaux@gmail.com>

Bonjour,

Je n'ai fait qu'un test d'installation (et transmit mon fichier descriptif à l'auteur) car finalement, nous n'avons pas eu de besoin correspondant à Tematres.

Le logiciel a bien évolué, je pense que le mieux est de contacter directement l'auteur (info@r020.com.ar), en anglais.

Pour quelle raison désirez vous connaitre le détail des tables?

Bien cordialement,

Patrice Chalon

Cependant la discussion avec Monsieur Chalon n'a pu être poursuivie par manque de temps. Ainsi l'échange s'est poursuivi avec Diego Ferreyra, le créateur de TemaTres. Il est Argentin mais n'ayant aucune connaissance solide en espagnol, le mail lui a été adressé en anglais.

Cedric R <cedric.rougeaux@gmail.com>

23 Avril 2010 12:00

À: TemaTres <tematres@r020.com.ar>

Dear Sir,

I am a French student, currently in internship on RTBF in Brussels. The Infodoc department would like to create a thesaurus with TemaTres.

I have installed this software and it seems to work. But I would like to know how it does work. Especially what is the use for every table in "tematres_101"? How to create two distincts tables? And how to mix two tables into one?

Please, would you have any informations or documents to provide?

Cordially,
Rougeaux Cedric.

TemaTres <tematres@r020.com.ar>

23 avril 2010 23:30

À: Cedric Rougeaux <cedric.rougeaux@gmail.com>

Hi dear Rougeaux, In a few days we publish de 1.05 version of TemaTres. This version have easy install script. If you want, you can download it from <http://www.r020.com.ar/tematres/tematres1.05.zip>

What do you mean with "mix two tables in one"?

lc_tema : contain terms data

lc_tabla_rel: contain data about relations between terms

lc_usuario: users data

lc_config: metadata about the vocabualry

lc_notas: term's notes

lc_values: referential values for IDs used in TemaTres

please let me know if I help you :-)

best regards

diego ferreyra

Cedric R <cedric.rougeaux@gmail.com>

26/04/10 14:51

À: TemaTres <tematres@r020.com.ar>

Hi, Sir Ferreyra,

I mean with "mix two tables in one" that I wish import datas from table to another table (no empty) without to lose relations between terms. Every term has a number of identification (ID) but when the process is done, there are terms which have the same number of identification.

So TemaTres can't recognize all relations between terms. How do I realize this operation, please?

Cordially,
Rougeaux Cedric.

Malheureusement cet échange ne s'est pas poursuivi. Diego Ferreyra n'a pas répondu. Il n'a pas été recontacté, le travail devenant très chronophage.

Document élaboré lors de l'utilisation de TemaTres

Ce qui suit a été élaboré, conjointement avec Monsieur Boly, durant les tests de fonctionnalité et l'utilisation de TemaTres.

TEMATRES DESCRIPTION DES FONCTIONS & PROCEDURES D'ENCODAGE
--

1. **Analyse des fonctions.**

Pour afficher le menu complet de la barre des tâches, il faut sélectionner au préalable, soit un élément de la liste thématique, soit un terme de la liste alphabétique.

➤ **AJOUTER**

○ **Terme spécifique**

Cette fonction est utilisée pour créer un terme spécifique et l'intégrer à une arborescence.

○ **Subordonate an Existing Term (TS)**

Cette fonction subordonne un terme existant à un terme situé un niveau hiérarchique au-dessus (créer une relation de terme secondaire).

○ **Terme exclu (EP)**

Cette fonction crée une relation d'équivalence.

C'est-à-dire qu'elle permet d'ajouter un « EP » (employé pour) à un descripteur existant.

○ **Associate an Existing Non-preferred Term**

Cette fonction associe, en tant que terme exclu, un terme existant à un descripteur existant.

○ **Terme associé (TA)**

Cette fonction crée une relation associative entre deux descripteurs existants.

Lorsqu'on crée une relation, elle se fait automatiquement dans les deux sens.

○ **Relation avec un autre vocabulaire**

Cette fonction permet de créer une relation d'équivalence entre des descripteurs de langues différentes.

➤ **OPTIONS**

○ **Editer un terme**

Cette fonction permet de modifier un terme existant (ex : verre -> ver)

○ **Editeur de note**

Cette fonction permet de créer différents types de notes explicatives sur l'usage du descripteur sélectionné.

- ◆ Note d'application (N.A.)
- ◆ Note historique (ex : nouvelle appellation à partir du ...)
- ◆ Note bibliographique
- ◆ Note du catalogueur

◆ Note privée

○ **Subordonner à un terme (TG)**

Cette fonction permet d'ajouter un terme générique au terme sélectionné. Concrètement, on le dote d'un nouveau parent. Cela revient à subordonner le terme sélectionné à un terme considéré du niveau supérieur.

○ **Effacer un terme**

Cette fonction permet d'effacer un terme. Pour pouvoir effacer un terme il faut, le cas échéant, lui enlever au préalable toute relation avec un autre terme. Cette opération se fait en cliquant sur la petite croix [X] qui précède le terme concerné ou son parent.

○ **Accepter le terme**

Cette fonction permet de modifier le statut d'un candidat descripteur en un descripteur accepté. Elle n'apparaît que dans le cas où le terme sélectionné est un candidat descripteur.

○ **Rejeter le terme**

Cette fonction permet de rejeter un terme accepté. Le terme ne peut être rejeté que si on lui a supprimé toute relation avec un terme parent.

➤ **MENU**

○ **Ajouter un terme**

Cette fonction est utilisée pour créer un nouveau terme. On peut spécifier s'il s'agit d'un candidat descripteur. Si on ne précise pas, le terme créé est considéré comme accepté à sa date de création. Comme le terme créé n'est relié à aucune branche hiérarchique, il est considéré comme un terme libre.

○ **Termes libres**

Cette fonction dresse la liste de tous les termes libres. Cela comprend les termes acceptés, les candidats et les rejetés, mais sans lien hiérarchique.

○ **Termes répétés**

Cette fonction repère et dresse la liste des termes répétés.

○ **Termes rejetés**

Cette fonction dresse la liste des termes rejetés.

○ **Termes candidats**

Cette fonction dresse la liste des candidats descripteurs.

➤ **ADMINISTRATION** (pour mémoire) non testé

○ **Configuration du Thésaurus**

Cette fonction semble permettre l'ajout de vocabulaire cible pour des termes de langues différentes.

○ **Utilisateurs**

Cette fonction donne la liste des différents utilisateurs et permet d'en ajouter.

- o **Exportation**
Cette fonction permet d'exporter tout ou partie d'une base de données vers une autre base de données.
- o **Optimiser les tables**
Cette fonction devrait permettre l'actualisation des tables.

2. Procédures d'encodage.

➤ **NOMS DE DOMAINES**

Les NOMS DE DOMAINES sont empruntés au thésaurus Eurovoc et sont encodés en **lettres capitales** précédés de leur numéro de classement de manière identique à Eurovoc.

➤ **MICROTHESAURUS**

Les NOMS DE MATIERES constituant le microthésaurus sont empruntés également au thésaurus Eurovoc et encodés en **lettres capitales** précédés de leur numéro de classement à l'identique d'Eurovoc.

➤ **DESCRIPTEURS & NON-DESCRIPTEURS**

- o Tous les termes descripteurs, libres, exclus, candidats, rejetés ..., sont écrits en **lettres minuscules**.
- o Ils sont au **singulier**, pour autant que cela n'introduise pas d'ambiguïté dans le sens.
- o Les acronymes, sigles et logos sont écrits en **majuscule** mais, sauf exception, ils ne sont pas retenus comme termes descripteurs, exemples :
ABB *USE* *Association Belge des Banques*
OMS *Use* *Organisation mondiale de la Santé*

➤ **CREATION DES RELATIONS ASSOCIEES (TA)**

Les renvois [**Voir aussi**] recensés dans nos listes de descripteurs belges et étrangers doivent être ajoutés en « terme libre ». Cela permet de créer directement la relation associative avec le terme entré. On pourra par la suite, le cas échéant « retravailler » le terme libre et lui donner son appellation définitive tout en étant assuré que la relation associative créée ne disparaîtra pas.

Cette procédure ne s'applique pas aux relations associatives recensées par Eurovoc.

On créera cette relation quand le moment sera venu d'inscrire le descripteur Eurovoc. Comme la relation se fait automatiquement dans les deux sens, cela ne devrait pas poser de problème.

• **DATATION DE CERTAINS DESCRIPTEURS**

Il est nécessaire d'associer une date aux descripteurs relevant d'événements ou faits précis et ponctuels, en respectant le format de date suivant :

AAMMJJ.

Exemples :

- «*Accident de transport par conduite Ghislenghien 2004-07-30* »

➤ **RECENSEMENT DES ENTREPRISES**

Pour l'instant, on se concentre uniquement sur les matières thématiques.
Le recensement des entreprises fait l'objet d'une prochaine phase du projet.

➤ **TERMES CANDIDATS**

Il peut être intéressant de recenser en termes candidats une liste de termes ayant le même objet mais dont la classification n'a pas été prévue par Eurovoc, par exemple : les accidents.

Le but est de repartir de cette liste pour en établir une classification structurée qui tiendra compte de tous les cas de figure et éviter ainsi une classification empirique.

➤ **NOTES DU CATALOGUEUR**

Par précaution, il serait utile d'indiquer en note du catalogueur toute information nécessaire à la complétion future du descripteur encodé.