

DE L'ANALYSE CRITIQUE DES LANGAGES DOCUMENTAIRES : ENJEUX, METHODE ET APPLICATION AU DOMAINE BIOMEDICAL

Isabel SASTRE CANTANO
ULB-Infodoc
Ave, F. Roosevelt, 50
1050 Bruxelles, Belgique
Tél. : 32/ 2/ 650 39 91 Fax : 32/ 2/ 650 39 43
isabel.sastre.cantano@ulb.ac.be

1 INTRODUCTION

Dans le cadre de l'action européenne COST B10 ¹, nous avons conçu un modèle de base de données ² ainsi qu'un thesaurus monolingue en vue de répertoire, dans un environnement Web, les laboratoires de recherche européens actifs dans le domaine du " Brain Damage Repair ". L'étude a montré l'importance d'une analyse critique des langages documentaires ³ concernés. Nous proposons ici

d'étayer cette démarche critique trop peu souvent abordée.

Nous allons, à l'aide de l'exemple suivant, illustrer la problématique qui se pose lors de l'utilisation de ces langages et insister plus particulièrement sur leur structure :

L'exemple évoqué ici est le terme préférentiel " Brain injuries " issu du *Medical Subject Headings (MeSH)* ⁴ [1] :

<p>Nervous System Diseases [C10] Central Nervous System Diseases [C10.228] Brain Diseases [C10.228.140] Brain Injuries [C10.228.140.199]</p> <p>Nervous System Diseases [C10] Trauma, Nervous System [C10.900] Cranocerebral Trauma [C10.900.300] Brain Injuries [C10.900.300.087]</p> <p>Disorders of Environmental Origin [C21] Wounds and Injuries [C21.866] Cranocerebral Trauma [C21.866.260] Brain Injuries [C21.866.260.118]</p> <p>Disorders of Environmental Origin [C21] Wounds and Injuries [C21.866] Trauma, Nervous System [C21.866.915] Cranocerebral Trauma [C21.866.915.300] Brain Injuries [C21.866.915.300.200]</p>
--

Sans entrer dans le détail d'autres relations existantes, nous pouvons constater que le terme médical " Brain injuries " se retrouve à quatre positions différentes au sein des vedettes-matière du *Medical Subject Headings (MeSH)*. Cela peut poser un problème lors de la compréhension de ce terme spécialisé. Une interprétation différente peut être faite selon sa lecture. Ce problème d'interprétation est accentué lors de la consultation d'autres thesaurus

médicaux, car on observe une hétérogénéité structurelle.

Nous pouvons constater qu'en général le monde des thesaurus est un monde peu connu, même par les spécialistes des technologies de l'information. Il est souvent présenté comme figé alors que c'est un domaine extrêmement vivant et riche sur les plans théorique et pratique comme le montrent les différents exemples cités.

La recherche dans des bases de données volumineuses se fait fréquemment à l'aide de plusieurs langages documentaires. Ces langages classificatoires ⁵ ou langages combinatoires ⁶ (thesaurus ou listes d'autorité) sont articulés entre eux, ce qui permet d'accroître les possibilités et la qualité de la recherche. Dans ce cadre-ci, nous nous sommes plus spécialement concentrés sur les thesaurus ⁷ [2,3].

Nous allons au sein de cette analyse décrire en premier lieu les difficultés apparaissant lors de l'utilisation d'un langage documentaire; ensuite nous exposerons les problèmes que nous avons dû résoudre afin de pouvoir proposer une approche cohérente des différents outils évalués dans le cadre de notre étude pour COST B10. Nous terminerons avec des conclusions et un extrait de l'étude.

2 LES DIFFICULTES

Nous avons choisi d'exposer cinq difficultés que l'utilisateur rencontre lors de l'emploi d'un thesaurus. Nous avons voulu insister plus particulièrement sur deux points qui ne sont peut-être pas évidents à première vue : l'entrelacement des langages documentaires et l'évolution des concepts et du réel dans le temps.

Le problème de structure décrit ci-dessus est classique et, même étant identifié, il ne rend pas la question moins complexe.

Cette première difficulté n'est pas spécifique aux thesaurus biomédicaux, mais c'est une difficulté qu'on retrouve dans tout genre de thesaurus et classifications,

qu'ils soient généraux ou appliqués à d'autres domaines.

La deuxième difficulté peut paraître toute simple : l'interprétation d'un concept dans le contexte général du thesaurus ou de la classification. Nous pensons aux relations et plus particulièrement aux relations associatives ou aux renvois. D'une part, les relations élargissent le champ de recherche et représentent une aide concrète lors de la recherche du concept soi-disant "optimal". Mais, d'autre part, elles renvoient vers d'autres concepts qui peuvent, dans leur interprétation, se confondre avec la première notion de recherche. C'est une démarche qui peut perturber l'utilisateur. Celui-ci peut alors avoir une mauvaise compréhension du sujet recherché.

La troisième difficulté se réfère au croisement de plusieurs concepts. Il est possible qu'à première vue un concept exprimé à l'aide de descripteurs et de non-descripteurs recouvre partiellement le contenu d'un autre concept. Même après une étude plus approfondie, il est très difficile de bien cerner le champ couvert par chacun des concepts.

La quatrième difficulté réside dans la combinaison courante des thesaurus et des systèmes de classifications. Elle complique l'interprétation du concept et peut donner lieu à fausse route. Ce type de croisement est à la limite du "tolérable" si on se réfère à la théorie respective. Voici un exemple d'un terme préférentiel du *Grants Classification System Thesaurus* qui illustre ce type de croisement [4]. On observe une grande complexité rien qu'en regardant le nombre et le type de renvois vers d'autres termes.

EMERGENCY ASSISTANCE (food, clothing, cash) (P60)

Use for organizations that provide short-term emergency assistance in the form of food, clothing, cash, etc. to needy individuals and families.

NT : TRAVELERS' AID (P61); VICTIMS' SERVICES (P62)

For organizations that provide only food services, see FOOD SERVICE/FREE FOOD DISTRIBUTION PROGRAMS (K30).

See also DISASTER PREVENTION/RELIEF (M20); HOMELESS-TEMPORARY SHELTER (L41); EMERGENCY MEDICAL SERVICES (E63); HOTLINE-CRISIS INTERVENTION SERVICES (F40); FAMILY VIOLENCE SHELTERS AND SERVICES (P43); HOMELESS PERSONS CENTERS & SERVICES (P85)

La dernière difficulté - et pas la moindre - est celle de l'évolution dans le temps

des concepts et du réel correspondant. Le réel observable est en changement

continu; prenons comme exemple les connaissances sur le virus du Sida qui changent continuellement. Cette notion pourrait être représentée au sein d'un thesaurus à l'aide de descripteurs, de non-descripteurs et de notes explicatives. Lors de grandes découvertes sur ce sujet ou plutôt lors de la prise de conscience d'un changement du réel, l'entrée dans le thesaurus doit être adaptée. On pourrait imaginer la modification de la note explicative lors d'une prochaine mise à jour. Mais le problème du décalage temporel entre l'évolution du réel et l'évolution du concept correspondant au sein de thesaurus demeure intact.

3 LA METHODOLOGIE APPLIQUEE

Nous décrivons en premier lieu les difficultés rencontrées lors de la consultation des sources relatives à l'étude pour l'action " Brain Damage Repair ". Ensuite, nous détaillerons l'approche critique appliquée qui se compose de deux parties :

une grille d'analyse et un tableau récapitulatif.

De nombreux problèmes se sont posés lors de l'élaboration de cette méthodologie : nous avons tout d'abord dû nous frayer un chemin jusqu'aux thesaurus - des outils très chers et rarement accessibles gratuitement au grand public. Nous avons tenu compte des différentes versions d'un même langage documentaire et tenté de travailler sur la dernière version existante. Nous avons également constaté un manque d'information concernant les sources utilisées lors de la conception des thesaurus et classifications. Une dernière difficulté a résidé dans l'élaboration d'éléments communs à tous les outils d'indexation et de recherche.

Après avoir franchi l'ensemble de ces obstacles, nous avons développé une méthodologie qui permet une approche structurée et cohérente des outils d'indexation et de recherche. Nous avons élaboré une grille d'analyse qui a été appliquée aux six langages documentaires ⁸ retenus :

Nom du thesaurus	
Organisation responsable	
Objet	
Source	
Taille	
Genre	<ul style="list-style-type: none"> • Thématique • Général
Profondeur horizontale ⁹	Nombre de domaines ou groupes
Profondeur verticale ¹⁰	Nombre de niveaux hiérarchiques
Langue	<ul style="list-style-type: none"> • Monolingue • Multilingue
Structure	<ul style="list-style-type: none"> • Relation d'inclusion • Relation hiérarchique • Relation d'association • Relation sémantique • Note explicative • Equivalence linguistique
Présentation	<ul style="list-style-type: none"> • Alphabétique • Hiérarchique • Permutée
Commentaire	

L'application de cette dernière a permis une analyse critique de ces outils dont le point 4 propose un extrait. Cette grille d'analyse n'est pas uniquement valable pour les thesaurus biomédicaux : elle est généralisable et peut être utilisée pour d'autres domaines d'application comme ceux de l'éducation, du droit ou de l'économie.

Dans le domaine de l'éducation par exemple, il existe plusieurs thesaurus (*ERIC*, *Thésaurus européen de l'Éducation (TEE)*, *Eurovoc*, *IBE Education Thesaurus* ou *EFICET*) auxquels on pourrait appliquer cette approche [5, 6, 7, 8, 9].

Cette analyse a été synthétisée sous forme de tableau. Le tableau ci-dessous récapitule par exemple les résultats les

plus pertinents au sujet des langages documentaires biomédicaux :

Thesaurus ou Voc. contrôlé	Taille	Genre	Profondeur verticale et horizontale	Mono-/Multilingue	Structure	Présentation	Commentaire
UNESCO Thesaurus	6 600 descripteurs et non-descripteurs (4 261 concepts)	thesaurus général	- 7 facettes - 3 niveaux	Multilingue : anglais français espagnol	relations classiques d'un thesaurus : rel. d'inclusion, rel. hiérarchique (BT/NT) rel. associative, note explicative, non-descripteurs	3 présentations : - alphabétique - structurée - permuée	Le thesaurus couvre en partie la médecine, n'a pas été conçu à cet effet, priorité aux domaines pertinents pour Unesco : l'éducation, l'économie, etc.
EMTREE	42 000 termes préférés + 180 000 synonymes	thesaurus thématique : biomédical	- 15 domaines - 7 niveaux	Monolingue : anglais	relations classiques d'un thesaurus : - rel. d'inclusion, rel. hiérarchique (see under) - rel. associative (see also) - note explicative non-descripteurs - notes historiques code de classification	3 présentations : - alphabétique - structurée - permuée	Le thesaurus est très complet, couvre largement les neurosciences.
BIOSIS	20 000 termes ou phrases (dont 730 mots-clés contrôlés, ~500 Concepts Codes, ~700 Biosystematic Codes)	- 2 systèmes de classification (CC et BC) - une liste de mots-clés : domaines couverts : 'life sciences'	- 110 domaines - CC : 2 niveaux - BC : 3 niveaux - liste des mots-clés : pas de hiérarchisation	Monolingue : anglais	- CC : vedettes-matières + code de classification couvrant les domaines des 'life sciences', notes explicatives aidant à trouver une stratégie de recherche (guide vers d'autres termes) - BC : vedettes-matières + code de classification servant à indexer des organismes - liste des mots-clés : 7 types différents ; structure différente selon le type : définitions, ancienne version en tant que non-descripteur, dates de validité, notes historiques	'Master Index' : alphabétique CC et BC : présentation structurée selon les vedettes matières	Les outils d'indexation sont d'une excellente qualité couvrant un grand nombre de domaines, mais pas spécialement les neurosciences.
INSPEC	15 700 (7 700 termes préférés et 8 000 'lead-in terms')	thesaurus général	- pas explicite - 4 niveaux	Monolingue : anglais	relations classiques d'un thesaurus : - rel. d'inclusion - rel. hiérarchique (BT/NT) - rel. associative - note explicative - non-descripteurs	2 présentations : - alphabétique - structurée - hiérarchique	Le thesaurus couvre un grand nombre de disciplines, mais pas spécialement la médecine.
MeSH	19 000 vedettes-matières principales + 103 500 vedettes-matières	vocabulaire contrôlé couvrant le domaine biomédical	- 16 domaines - 7 niveaux	Monolingue : Anglais	renvois vers - d'autres descripteurs (see related) - d'autres termes (consider also) renvois venant - d'autres descripteurs - d'autres termes notes explicatives : - notes d'indexation/catalogage - notes historiques - notes online (pour chercheurs online) code de classification qualifiants	3 présentations : - alphabétique - structurée - permuée	L'outil d'indexation est très complet et reconnu. Le Vocabulaire contrôlé est le plus utilisé et est apprécié par les experts du domaine.
UMLS Meta-Thesaurus	800 000 concepts (1,9 millions de termes préférés et synonymes)	'méta-thesaurus', structuré autour de la notion du 'concept', couvrant le domaine biomédical	- termes regroupés par concept - 13 niveaux	Monolingue : anglais ; en partie d'autres langues présentes, mais pas de manière systématique	Broader (RB)/Narrower (RN) : rel. hiérarch. Like (RL) : rel. entre 2 concepts similaires Parent (PAR) : relation père Child (CHD) : relation fils Sibling (SIB) : relation 'sibling' (frère) A Q : un qualifiant QB : peut être qualifié Other related (RO) : toute relation qui n'est ni hiérarchique ni de synonymie	accessible via : - Internet - CD-Rom	Outil d'indexation - conçu pour développeurs, pas pour grand public très puissant, très complexe et très complet.

Tableau 01 : Récapitulatif de l'analyse critique de thesaurus et classifications existantes

4. CONCLUSION

Cette approche critique permet une évaluation globale de thesaurus et de classifications. Elle donne la possibilité de comparer des thesaurus similaires couvrant un même domaine comme la médecine. La démarche est généralisable et de ce fait applicable à tout autre domaine empirique.

L'étude permet également de comparer rapidement les critères spécifiques des différents outils. De cette manière, il est possible de voir quels sont les domaines couverts, de comparer par exemple le nombre de concepts intégrés dans chaque langage ou d'avoir une bonne vision de leur structure, ce dernier point étant très important vu sa complexité.

Ces résultats représentent une aide originale en vue du choix d'un outil d'indexation et de recherche pour un système documentaire au sein d'une grande entreprise. Ils offrent également un support intéressant lorsqu'on doit trancher entre l'acquisition (ou l'adaptation) d'un outil déjà existant sur le marché et la conception d'un nouveau langage documentaire dans un domaine spécifique ou générique.

5 LES THESAURUS BIOMEDICAUX

Afin d'étayer la présentation synthétique qui précède, nous évoquerons l'analyse des six langages documentaires suivants : le Thesaurus de l'Unesco *UNESTHES*, *l'EMTREE* utilisé pour la base de données EMBASE, les outils d'indexation employés pour la base de données *BIOSIS*®, les outils d'indexation appliqués dans la base de données INSPEC, le langage classificatoire *MeSH* (National Library of Medicine) et *l'UMLS Metathesaurus* (National Library of Medicine). Pour plus d'informations concernant les thesaurus cités dans cet article, nous renvoyons à l'analyse critique détaillée [10].

5.1 Thesaurus de l'UNESCO

Nom du thesaurus : UNESTHES

Organisation responsable : UNESCO

Objet :

Le thesaurus de l'UNESCO couvre les grands domaines de connaissance qui constituent le champ d'action de l'organisation. Les sujets auxquels l'Unesco s'intéresse plus particulièrement sont traités en détail. Il a été conçu pour indexer et retrouver l'information stockée dans la base de données bibliographiques de l'UNESCO (UNESBIB) et dans diverses autres bases secondaires faisant partie de leur réseau documentaire [11, 12, 13].

Source :

Ce sont en majorité des thesaurus et des systèmes de classification qui ont servi de base lors de la conception du thesaurus. Voici quelques exemples : 16 sources différentes pour le domaine " Science et technologie " dont INSPEC Thesaurus, SPINES Thesaurus, TEST [13, 14, 15, 16]. Les sources utilisées pour les microthesaurus médicaux sont les suivantes : les Medical Subject Headings (MeSH) pour les termes et le BSO : le Système général de classement de l'Unisist pour la structure à facettes [17].

Taille :

Le thesaurus contient 4.261 concepts (6.600 descripteurs et non-descripteurs), 87 microthesaurus, 4.049 relations hiérarchiques (BT/NT) et 5.771 relations associatives (RT).

Genre :

On peut considérer ce thesaurus comme étant général car il couvre des domaines très différents tels, par exemple, l'éducation ou l'économie, mais en gardant toujours la principale mission de l'UNESCO au premier plan.

Profondeur horizontale et verticale :

Ce thesaurus contient 7 domaines divisés en microthesaurus dont le nombre diffère entre dix et dix-sept selon le domaine : 1. Education, 2. Science, 3. Culture, 4. Social and human sciences, 5. Information and communication, 6. Politics, law and economics, 7. Countries and country groupings. Le nombre de niveaux est de 3 [18].

Langue :

Ce thesaurus est multilingue et existe depuis 1984 dans 3 des langues officielles de l'UNESCO : en anglais, français et espagnol.

Structure :

Toutes les relations courantes apparaissent : les relations d'inclusion (MT¹¹), les relations hiérarchiques (BT/NT), les relations d'association (RT). Il existe des poly-hiérarchies (plus d'une relation générique) uniquement en ce qui concerne les noms des pays. Cela a, comme conséquence, une très forte redondance, et donc un gonflement artificiel du thesaurus. Ceci aurait pu être évité en introduisant des relations d'association. Des notes d'application donnant des directives relatives à l'indexation, mais aussi des précisions sur le concept et des relations sémantiques (non-descripteurs), apparaissent également dans ce thesaurus. Comme c'est un thesaurus multilingue, les équivalences linguistiques en français et en anglais sont aussi présentes.

Présentation :

Il existe trois présentations différentes : une liste alphabétique, structurée et permutée, une liste hiérarchique par microthesaurus et un index des descripteurs français/anglais/espagnol (aussi un index espagnol/anglais/français).

Commentaire :

Ce thesaurus couvre en partie le domaine d'application des neurosciences, mais a été développé pour un autre but et

en priorité avec d'autres champs d'application. La médecine et, plus particulièrement, les neurosciences (notamment " brain damage repair ") ne sont pas traitées de manière spécifique. Nous retiendrons que la source principale des termes médicaux sont les Medical Subject Headings (MeSH).

5.2 EMTREE Thesaurus 2001

Nom du thesaurus : EMTREE

Organisation responsable : Elsevier, Excerpta Medica.

Objet :

Le thesaurus EMTREE sert à indexer et retrouver les 8 millions de documents stockés dans la base de données bibliographiques EMBASE [19]. Celle-ci contient essentiellement des articles et des comptes-rendus issus des principaux domaines biomédicaux suivants : Drug Research, Pharmacology, Pharmaceutics, Pharmacy, Side Effects and Interactions, Toxicology, Human Medicine (Clinical and Experimental), Basic Biomedical Sciences, Biotechnology, Biomedical Engineering and Instrumentation, Health Policy and Management, Pharmacoeconomics, Public, Occupational and Environmental Health, Pollution, Substance Dependence and Abuse, Psychiatry, Forensic Medicine [20].

Source :

Ce sont les 16 facettes du MeSH qui ont servi de base pour concevoir la structure du thesaurus EMTREE [21]. Différents vocabulaires contrôlés - MALIMET, EMTAGS et EMCLAS - qui, avant 1991, étaient utilisés pour l'indexation, ont été intégrés dans le EMTREE [22].

Taille :

Le thesaurus contient 42.000 termes préférés et 180.000 synonymes. Parmi ces termes préférés, 10.000 ont un code alphanumérique qui classe l'ensemble des termes sous ses 15 domaines.

Genre :

On peut considérer ce thesaurus comme thématique, car il couvre uniquement des domaines de la médecine.

Profondeur horizontale et verticale :

Ce thesaurus est divisé en 15 facettes (de A à Q) qui sont également divisées en sous-groupes (entre 3 et 30). La profondeur descend jusqu'à sept niveaux [22].

Langue :

Sa conception est anglaise, elle remonte à 1988 et la dernière édition date de 2001.

Structure :

Le thesaurus EMTREE est structuré de manière hiérarchique, incluant les relations courantes suivantes : les relations hiérarchiques (*see under*), les relations d'association (*see also*) et les relations sémantiques (*use*). D'autres champs sont présents par concept comme, par exemple, les champs " history note, " Associated Emtree Code " et autres pour lesquels nous ne rentrerons pas dans le détail. Les termes repris dans le thesaurus n'ont pas tous un code de classification. Les termes sont divisés en " explosion terms " et " non-explosion terms ". Le premier type de terme a un code de classification, le deuxième type de terme n'a pas de code. Les " non-explosion terms " sont toujours mentionnés comme spécifique d'un " explosion term ".

Exemples :

brain **A8.30.20**
1974

Le terme préféré " brain ", a été introduit en 1974 et a comme code A8.30.20.

brain injury **C2.380.380.380.125**
1974 **C2.610.150.10.110**
 C2.610.603.125
 C6.450.380.380.125
 C6.450.610.125
see also :

brain damaged child

Le terme préféré " brain injury ", introduit en 1974 a 5 codes, cela signifie qu'il est représenté cinq fois dans ce " thesaurus " alors que, dans un " vrai " thesaurus, un concept n'est représenté qu'une seule fois. Comme nous l'avons déjà signalé, le thesaurus est divisé en groupes et sous-groupes. A titre d'exemple, nous voyons que dans le groupe C " Physical diseases, disorders and abnormalities " les maladies physiques sont reprises sous plusieurs catégories. Ce groupe C est divisé en 7 sous-groupes : C1 " General aspects of disease ", C2 " Physical disease by anatomical structure ", C3 " Physical disease by body fonction ", C4 " Physical disease by composition of body fluids, excreta and secretions ", C5 " Physical disease by

developmental age ", C6 " Physical disease by etiology and pathogenesis " et C7 " animal disease ". " Le terme " brain injury " se retrouve trois fois dans le sous-groupe C2 " Physical disease by anatomical structure " et deux fois dans le sous-groupe C6 " Physical disease by etiology and pathogenesis ".

La base de données EMBASE contient plusieurs champs pour les descripteurs: Major Medical Descriptors (DEM), Minor Medical Descriptors (DER), Major Drug Descriptors (DRM), et Minor Drug Descriptors (DRR). Chaque enregistrement de la base de données contient ces 4 champs, par conséquent, une indexation à l'aide de ces quatre champs est permise.

La recherche dans la base de données est complétée par les " Links ". Ce sont des qualificatifs (appelés également " qualificatifs ", " link terms " ou " subheadings ") qui spécifient le domaine d'application du descripteur. Il existe trois groupes de " links : drug links (17 qualificatifs), disease links (14 qualificatifs) ajoutés en 1988 " et les " Routes of Drug Administration " (47 qualificatifs) ajoutés en 2000 [23].

Cette approche ressemble fortement à celle utilisée dans Medline [24].

Les règles appliquées lors de la normalisation de la terminologie sont les suivantes [25] :

- Préférer l'ordre naturel des mots (par exemple : acute granulocytic leukemia, et non leukemia, acute granulocytic)
- Utiliser le singulier pour les noms
- Appliquer l'orthographe américaine et non l'anglaise
- Eviter les abréviations
- Ecrire en tous mots les lettres grecques
- Ne pas utiliser de traits d'union, ni d'apostrophes et éviter les virgules (exception : les noms chimiques contenant des chiffres : 4 ethyl 2,5 dimethoxyamphetamine)

Présentation :

Il existe trois présentations différentes : un index alphabétique très complet avec les codes Emtree et les notes d'application, une liste hiérarchique représentant les 15 facettes et un index permuté [21].

Commentaire :

Cet outil d'indexation ressemble plus à une classification qu'à un thesaurus, même s'il est dénommé en tant que tel. Il est très précis, très complet et couvre largement le domaine des neurosciences et celui du " brain damage repair ". Il représente une bonne source pour la conception du thesaurus " brain damage repair ".

5.3 Biosis

Noms des outils d'indexation : le " Master Index ", la liste mots-clés contrôlés, les " Concept Codes " (CC) et les " Biosystematic Codes " (BC) qui seront définis ci-dessous.

Organisation responsable : Biosis®

Objet :

Les différents outils servent à indexer les 8,8 millions de références stockées dans la base de données bibliographiques BIOSIS. Les domaines principaux couverts sont la biologie ainsi que des disciplines jouant un rôle dans une approche interdisciplinaire (" life sciences "); ces domaines sont divisés en 110 groupes tels que Agriculture, Anatomy, Biophysics, Cancer, Conservation, Drugs, Development, Ethnobiology, Forestry, Information Science, Methods, Materials and Apparatus, Nervous System, Pesticide, Pharmacology, Public Health, Soil Science, Taxonomy, Traditional Medicine, Virology, Waste Management ou Zoology. La base de données contient essentiellement des comptes-rendus d'articles, parus dans des revues spécialisées, des monographies, des livres de références et des brevets enregistrés aux Etats-Unis. Quelques 6.400 périodiques de 94 pays sont passés en revue régulièrement; la répartition des sources se fait comme suit : 50 % d'Europe et du Moyen Orient, 31 % d'Amérique du Nord, 14% d'Asie et d'Australie, 3% d'Amérique Centrale et du Sud et 2% d'Afrique [26].

Source :

Ce sont les documents revus par les spécialistes de Biosis qui servent de base pour les trois outils d'indexation utilisés dans la base de données BIOSIS : les " Concept Codes ", les " Biosystematic Codes " et la liste des mots-clés contrôlés. Tous les termes repris dans un des trois vocabulaires contrôlés sont inclus dans le " Master Index ". Les critères d'inclusion dans ce dernier sont la fréquence d'apparition dans la base de données, mais aussi l'importance que peut avoir un terme lors d'une recherche.

Taille :

Le " Master Index " contient 20.000 termes ou phrases (les mots-clés contrôlés, les " CC " et les " BC " sont inclus dans ce comptage). Il existe environ 730 mots-clés contrôlés, 500 " Concept Codes " et 700 " Biosystematic Codes ".

Genre :

La recherche dans cette base de données se fait à l'aide de différents outils de recherche :

Exemple :

BROAD CONCEPT HEADING	CC CONCEPT HEADING	NOTES
DIGESTIVE SYSTEM	CC14001 General Methods CC14002 Anatomy CC14004 Physiology and Biochemistry CC14006 Pathology	

Code " BC " :

La classification des codes " BC " descend jusqu'à trois niveaux.

Mots-clés contrôlés :

Il n'y a pas de hiérarchisation dans les mots-clés contrôlés.

Langue :

L'ensemble de tous les outils de recherche sont accessibles uniquement en langue anglaise alors que les documents stockés dans la base de données sont issus du monde entier et rédigés dans différentes langues.

Structure :

Avant de rentrer dans le détail des trois outils de recherche, nous allons décrire le " Master Index ". C'est un index très complet qui sert à trouver des synonymes et des termes associés, qui guide vers un " CC " ou un " BC " approprié pour une recherche dans la base de données. Il sert également de source pour trouver les dénominations chimiques, les " Enzyme Commission Numbers " et les " Registry Numbers ". L'index est complété par des suggestions de recherche et des explica-

- deux systèmes de classification (" CC " et " BC ")
- une liste de mots-clés contrôlés

Il existe bien évidemment un grand nombre d'autres champs par lesquels une recherche est possible mais nous n'en tiendrons pas compte dans cette analyse.

Profondeur horizontale et verticale :**Code " CC " :**

La classification des codes " CC " a deux niveaux et il existe environ 110 vedettes-matières principales.

tions très précises décrivant l'utilisation des termes. Cet index regroupe toutes les informations existantes sur un concept dans une liste alphabétique avec des renvois vers le(s) CC ou le(s) BC adéquats. La fréquence d'indexation d'un terme (CC, BC et mots-clés contrôlés) est également spécifiée.

Les " Concept Codes " sont des codes en cinq chiffres qui représentent des vedettes-matières dans les domaines couverts par Biosis. Ce code est complété par des notes explicatives très complètes et précises qui guident les utilisateurs vers une autre stratégie de recherche et donnent également des exemples d'application. La fréquence d'indexation en tant que " CC " majeur ou mineur est, de son côté, bien déterminée.

Les " Biosystematic Codes " sont également des codes en cinq chiffres qui, pour se différencier des " CC ", commencent par les lettres BC alors que les " Concept Codes " commencent par CC. Ce code sert à l'indexation d'organismes vivants.

Le troisième outil de recherche est le plus diversifié alors que son nombre total

de mots-clés est le plus petit, si on le compare aux deux autres. Les mots-clés sont classifiés en sept types différents :

1. DRUG AFFILIATIONS (78 mots-clés)
2. OTHER DRUG ACTIONS (6 mots-clés)
3. GEOGRAPHIC INDEXING (550 mots-clés)
4. VIRUS AFFILIATION (20 mots-clés)
5. TAXONOMIC AFFILIATIONS (13 mots-clés)

6. NEW RESEARCH AREAS (3 mots-clés)
7. BIBLIOGRAPHIC DESCRIPTORS

La structure des mots-clés est différente selon le type de mots-clés. En ce qui concerne le premier type, les " DRUG AFFILIATIONS ", c'est une liste composée par un mot-clé principal avec, comme synonymes, les anciennes versions de celui-ci. Le mot-clé est complété par une définition et par les années pendant lesquelles les anciennes versions étaient valables :

DRUG AFFILIATION	DEFINITION	SEARCHABLE YEARS
ANTIHISTAMINE-DRUG use ANTIHISTAMINE use ANTI HIST-DRUG use ANTI HIST	any chemical agent used to neutralize or antagonize the actions of histamine or to inhibit the production of histamine	1985- 1979-1984 1970-78 1969

- Le deuxième type " OTHER DRUG ACTIONS " ne reprend que le mot-clé.
- Le troisième type " GEOGRAPHIC INDEXING " est complété par des notes historiques.
- Le quatrième type " VIRUS AFFILIATION " n'est presque plus utilisé. La plupart de ces mots-clés ont été introduits dans la classification " BC ". Il n'en reste plus qu'une vingtaine.
- Le cinquième type " TAXONOMIC AFFILIATION " permet d'identifier les nouvelles taxa comme par exemple " NEW-COMBINATION, NEW-GENUS, NEW-NAME, NEW-SPECIES, etc.
- Le sixième type " NEW RESEARCH AREAS " regroupe 3 mots-clés (ECOLOGICAL ENERGETICS, GENETIC ENGINEERING et MOLECULAR SEQUENCE DATA) qui décrivent les nouveaux domaines de recherche pour lesquels il n'existe pas encore de CC. Le dernier type de mot-clé, " BIBLIOGRAPHIC DESCRIPTORS " sert à décrire le type de documents source comme par exemple " ABSTRACT ", " ATLAS ", " BIBLIOGRAPHY ", " CORRECTED RECORD ", etc.

Il est également permis d'utiliser des mots-clés libres n'existant dans aucun des outils de recherche, car ils peuvent aider à trouver une référence.

Présentation :

Le " Master Index " est uniquement présenté par ordre alphabétique des concepts. Quant aux " CC " et aux " BC ", ils le sont par l'ordre des vedettes-matières principales.

Commentaire :

L'ensemble des outils conçus pour la base de données BIOSIS est d'une qualité excellente et constitue une grande aide lors de l'établissement de la stratégie de recherche. La base de données et les classifications sont multidisciplinaires et couvrent beaucoup de domaines, mais le domaine des neurosciences n'est pas spécialement couvert. Pour cette raison, ces outils ne seront pas spécialement pris en compte lors de la conception du thesaurus " brain damage repair ".

5.3.1 Thesaurus INSPEC

Nom du thesaurus : INSPEC Thesaurus
Organisation responsable : Institution of Electrical Engineers (IEE)

Objet :

Le thesaurus INSPEC sert à indexer les 6,8 millions de documents stockés dans la base de données bibliographique

INSPEC [27]. Les documents sont issus de différentes disciplines telles que la physique, l'électronique, la communication, l'ingénierie électrique, les technologies de l'information, l'informatique, etc. Plus de 3.500 journaux scientifiques et techniques, 1.500 documents de travail utilisés lors des conférences ainsi que des livres, rapports et dissertations sont analysés par l'équipe d'INSPEC. Leur analyse permet de résumer ces documents qui sont ensuite indexés et introduits dans la base de données [28].

Source :

Pas d'information.

Taille :

L'édition de 1995 contient 15.700 termes dont 7.700 sont des termes préférés et 8.000 sont des " lead-in terms " (terme pointant vers un terme préféré) [27].

Genre :

On peut considérer ce thesaurus comme étant général, car il couvre des domaines très différents.

Profondeur horizontale et verticale :

La présentation hiérarchique n'est pas clairement divisée en domaines comme nous l'avons vu dans les thesaurus ci-dessus. Ici, l'ensemble des top terms (TT)

est repris alphabétiquement dans la présentation hiérarchique. Il est possible qu'un descripteur ait plusieurs pères, cela implique donc qu'il se retrouve plusieurs fois dans la présentation. La profondeur verticale est au moins de quatre niveaux.

Langue :

Le thesaurus, édité pour la première fois en 1973, est élaboré uniquement en anglais.

Structure :

Il existe toujours un terme préféré en gras, suivi des non-descripteurs, lorsqu'ils existent, ensuite sont édités les relations hiérarchiques (BT/NT/TT), des relations associatives (RT) lorsque celles-ci existent, un ou plusieurs code(s) de classification qui relie(nt) tous les termes préférés du thesaurus au code le plus proche issu du système de classification INSPEC. Celui-ci sert également à la recherche des enregistrements bibliographiques de la base de données. Il est divisé en quatre parties : section A : physique, section B : ingénierie électrique et électronique, section C : informatique et contrôle, section D: technologies de l'information qui sont également réparties en sous-groupes. Cette classification contient 2.000 entrées avec mention de la date d'introduction et une note indiquant les méthodes d'indexation antérieures à cette date.

Exemple issu du thesaurus INSPEC :

Preferred Term	medical image processing
Used for	UF biomedical image processing
Broader Term(s)	BT biomedical imaging image processing medical diagnostic computing medical signal processing
Top Term(s)	TT computer applications patient diagnosis signal processing
Related Term	(RT) biomedical measurement computerised tomography medical expert systems PACS
Classification Code	CC A8760; A8770E; B6140C; B7510B; C5260B; C7330
Date of Input	DI January 1993
Prior Term(s)	PT computerised picture processing medical diagnostic computing picture processing

Présentation :

Il existe deux présentations de ce thesaurus : une présentation alphabétique complète avec toutes les relations, les synonymes, les codes et les notes historiques ainsi qu'une présentation hiérarchique qui édite uniquement les liens hiérarchiques.

Commentaire :

Ce thesaurus couvre un grand nombre de disciplines, mais pas spécialement la médecine, ni les neurosciences. Nous ne tiendrons pas compte de ce thesaurus lors de l'élaboration du thesaurus " Brain Damage Repair ", mais il servira à analyser une méthode de suivi historique très complète.

5.4 Thesaurus of Medical Subject Headings (MeSH)

Nom du thesaurus : Thesaurus of Medical Subject Headings (MeSH)

Organisation responsable : U.S. National Library of Medicine

Objet :

Ce vocabulaire contrôlé sert à indexer des articles, à cataloguer des livres et autres documents ainsi qu'à faire des recherches dans des bases de données indexées à l'aide des MeSH, principalement MEDLINE [29]. Cette base de données bibliographique contient plus de 11 millions de références, dans diverses langues du monde, issues de 4.300 journaux biomédicaux [30].

Source :

C'est l'équipe de la section Medical Subject Headings qui révisé et met à jour régulièrement la terminologie du MeSH. En plus de tenir compte des commentaires des indexeurs, ils collectent les nouveaux termes apparaissant dans la littérature scientifique et dans les domaines de recherche. Lors d'importantes restructurations du MeSH, l'équipe est soutenue par

des professionnels des différentes disciplines.

Taille :

Les MeSH contiennent plus de 19.000 vedettes-matières principales qui sont complétées par 103.500 vedettes-matières (Supplementary Concept Records) qu'on retrouve dans un thesaurus destiné aux produits chimiques.

Genre :

Le MeSH est plutôt un vocabulaire contrôlé ou un système de classification et non un thesaurus dans le sens strict du terme. Les termes sont repris plusieurs fois ce qui n'est pas le cas dans un thesaurus; la plupart des termes ont plusieurs pères (polyhiérarchies), ce qui existe dans le thesaurus, mais qui dans la plupart des cas est évité le plus possible. Les thesaurus n'ont en général pas de codes de classification sauf lorsqu'ils sont reliés à un système de classification, ce qui est souvent le cas.

Profondeur horizontale et verticale :

Il existe 16 " facettes " recouvrant toujours les notions générales retrouvées dans les documents de la base de données Medline : 1. Analytical, Diagnostical and Therapeutic Techniques and Equipment, 2. Anatomy, 3. Anthropology, Education, Sociology and Social Phenomena, 4. Biological Sciences, 5. Check Tags, 6. Chemical and Drugs, 7. Diseases, 8. Geographical Locations, 9. Health Care, 10. Humanities, 11. Information Science, 12. Organisms, 13. Persons, 14. Physical Sciences, 15. Psychiatry and Psychology, 16. Technology and Food and Beverages. Le nombre de niveaux s'élève à 7.

Langue :

La langue du MeSH est l'anglais. Mais, il existe un grand nombre de traductions (allemand, espagnol, finnois, italien, portugais, etc.) réalisées par diverses organisations. Cet aspect sera développé sous le point " ULMS Metathesaurus ".

Structure :

Les MeSH ressemblent plus à une classification qu'à un thesaurus. Les termes intégrés dans les MeSH sont très détaillés. Il existe au moins un code de classification pour chaque descripteur; le terme " ACQUIRED IMMUNODEFICIENCY SYNDROME " par exemple a cinq codes de classification ¹² [31].

Trois types de notes sont proposés pour compléter une entrée dans les MeSH: en premier lieu, des " notes d'indexation et de catalogage " très détaillées sur la façon de travailler destinées spécifiquement aux indexeurs et catalogueurs. En deuxième lieu, des " notes historiques " indiquent les anciennes versions d'une entrée avec les dates de validité (cf. exemple ci-dessous) En dernier lieu, des " notes online " sont ajoutées pour les

chercheurs online dans le but de les aider à retrouver des documents indexés avec des anciens termes.

Il existe des renvois - " see related " - vers d'autres descripteurs du MeSH qui sont en relation avec le descripteur. Un autre genre de renvoi - " consider also " - propose d'aller voir des termes proches du descripteur, comme par exemple aller voir tous les termes commençant par le préfixe " ONCO- " pour le descripteur " NEOPLASMS ".

Un élément original est le fait qu'on montre également les renvois faits à partir d'autres termes ou descripteurs vers ce descripteur (cf. exemple ci-dessous). Ils sont signalés à l'aide d'un " X " pour les renvois d'un autre terme et à l'aide d'un " XR " pour les renvois d'un autre descripteur.

Exemple issu des MESH de 1991 [31] :**DESCRIPTOR :** NEOPLASMS**CONSIDER ALSO****CROSS-REFERENCE****TO OTHER TERMS**

consider also terms at CANCER, CARCINO-, ONCO-, and TUMOR

TREE NUMBER : C4+**INDEXING ANNOTATION :**

avoid : too general; prefer specifics; policy: Manual section 24; /chem. ind permitted but consider also CARCINOGENS; /class: consider also NEOPLASM STAGING (see note there) but "grading" = /pathol; /etiol: consider also ONCOGENIC VIRUSES; /vet: Manual 24.32-36 or TN 136; TN 135: MeSH terms for neoplasms classed by tissue; /drug ther: consider also ANTINEOPLASTIC AGENTS & its specifics; /genet: consider also GENES, SUPPRESSOR, TUMOR; /immunol: consider also ANTIGENS, NEOPLASM & ANTIBODIES, NEOPLASM; /microbial: consider also ONCOGENIC VIRUSES; /radiother: consider also BRACHYTHERAPY; do not use /second (= NEOPLASMS METASTASIS); familial: consider also NEOPLASTIC SYNDROMES, HEREDITARY; metastatic cancer of unknown origin: index under NEOPLASM METASTASIS; Tumor Key: TN Suppl

CATALOGING ANNOTATION :

CATALOG: form qualify permitted

HISTORY NOTE :

/diagnosis was NEOPLASM DIAGNOSIS 1964-65; /etiology was NEOPLASM ETIOLOGY 1964-65; /immunology was NEOPLASM IMMUNOLOGY 1964-65; /radiotherapy was NEOPLASM RADIOTHERAPY 1964-65; / therapy was NEOPLASM THERAPY 1964-65; NEOPLASM STATISTICS was heading 1964-65; CARCINOGENESIS 1977 was heading 1977

ONLINE NOTE:

use NEOPLASMS /CI or NEOPLASM /ET to search CARCINOGENESIS 1977

FORWARD SEE RELATED**CROSS-REFERENCE****TO OTHER DESCRIPTORS**

see related

DNA, NEOPLASM
ONCOGENIC VIRUSES
PLEURAL EFFUSION, MALIGNANT
PRECANCEROUS CONDITIONS
RNA, NEOPLASM

BACKWARD SEE**CROSS-REFERENCES****FROM ENTRY TERMS**

X CANCER
X TUMOR

BACKWARD SEE**CROSS-REFERENCE****FROM OTHER DESCRIPTOR**

XR MEDICAL ONCOLOGY

Les descripteurs sont complétés par des qualificatifs qui sont au nombre de 80 et qui ont également une structure hiérarchique comme les vedettes-matières. Ils sont apposés aux descripteurs MeSH afin d'en préciser le sens d'un point de vue conceptuel. La syntaxe est la suivante : terme MeSH/qualificatif. Une indexation telle que " LIVER/drug effects " par exemple, indique que l'article ou le livre ne concerne pas le foie en général, mais l'effet d'un médicament sur celui-ci.

Présentation :

Il existe 3 présentations du MeSH : une liste alphabétique avec des notes complémentaires pour les indexeurs, catalogueurs et les chercheurs (MeSH Annotated Alphabetic List), une liste hiérarchique (MeSH Tree structures) et une liste permutée (Permuted MeSH).

Commentaire :

C'est sans aucun doute une des sources les plus pertinentes dans le domaine biomédical. Ce vocabulaire contrôlé est celui qui est le plus utilisé et plus apprécié par les experts du domaine.

5.4.1 UMLS Knowledge Sources - Metathesaurus

Nom du thesaurus :

UMLS Metathesaurus

Organisation responsable :

U.S. National Library of Medicine

Objet :

Depuis longtemps, le projet UMLS est un projet de recherche et de développement de la NLM. Le but est de faciliter la recherche et l'intégration d'information issue de multiples sources biomédicales, comme les bases de données factuelles, bibliographiques, les systèmes-experts, les dossiers médicaux sur ordinateur, etc. [32].

Le *UMLS Metathesaurus* est un des trois éléments dont sont constitués les UMLS Knowledge Sources. Le "*Semantic*

Network", un réseau des catégories génériques et des types sémantiques auxquels tous les concepts du Metathesaurus ont été assignés, et le "*SPECIALIST Lexicon*", qui contient des informations syntaxiques sur des termes biomédicaux, en constituent les deux autres éléments. Ces derniers ne seront pas décrits dans cette analyse.

Source :

C'est une équipe multidisciplinaire qui dirige le projet UMLS, soutenu par différents spécialistes du domaine. Des experts de l'information ont développé la structure de l'UMLS Metathesaurus et le système pour la réalisation et la maintenance du Metathesaurus.

Une soixantaine de sources a servi de base lors de la conception de ce Metathesaurus. Ce sont des classifications, des listes de vocabulaire contrôlé, des thesaurus, des terminologies établies par les hôpitaux, des services spécialisés en terminologie médicale, des associations, des administrations, etc. issus du monde entier (Etats-Unis, Europe, Amérique du Sud, Australie, etc.). Les MeSH ont été intégrées dans ce Metathesaurus.

Taille :

Le UMLS Metathesaurus contient environ 800.000 concepts (797.359) et 1,9 millions de termes (termes préférés et synonymes).

Genre :

Il est difficile de classer cet outil très particulier. Il a été conçu autour de la notion de " concept " et s'adresse à des développeurs, à des spécialistes et non à des utilisateurs non avertis. Il récolte toutes les informations au sujet d'un concept et les rassemble en un enregistrement. Il tient compte, par exemple, des différentes orthographes, du pluriel et du singulier. De plus, lorsque les concepts sont présents dans différentes sources, toutes les hiérarchies sont reprises.

Profondeur horizontale et verticale :

Etant donné que ce Metathesaurus reprend la structure des différentes sources, il n'est pas possible d'explicitier le niveau de détail de façon précise. Au moins 13 niveaux de profondeur verticale, dépendant de la source, ont été repérés.

Langue :

La langue principale du UMLS Metathesaurus est l'anglais. Mais beaucoup d'autres langues sont également présentes : allemand, basque, danois, espagnol, finnois, français, hébreu, hongrois, italien, néerlandais, norvégien, portugais, etc. Un autre élément fort important est la présence de traductions des MeSH en allemand, espagnol, finnois, français, italien et portugais. Ces traductions ont été réalisées par différentes organisations telles que le Deutsches Institut für Medizinische Dokumentation und Information (DIMDI) (version allemande), Istituto di Sanita Servizio Documentazione (version italienne) ou encore le Latin American and Caribbean Center on Health Sciences Information (version espagnole et portugaise) [32].

Structure :

Le UMLS Metathesaurus est organisé par concepts : l'ensemble des informations concernant un concept est regroupé au sein d'une entrée dans ce Metathesaurus. Il existe toujours un terme préféré en anglais qui est issu d'une ou plusieurs sources. Tous les synonymes, quasi-synonymes, variantes, différences dans l'orthographe (pluriel, singulier, majuscule, minuscule, etc.), sont liés au même concept. Les équivalences linguistiques - si elles existent - sont également présentes, mais sans leurs variantes [32].

Il existe neuf types différents de relations :

- Broader (RB), Narrower (RN) : relations hiérarchiques
- Other related (RO) : toute relation qui n'est ni hiérarchique ni synonymique

- Like (RL) : relations entre deux concepts similaires
- Parent (PAR) : il existe une relation père dans un des vocabulaires source du Metathesaurus
- Child (CHD) : il existe une relation fils dans un des vocabulaires source du Metathesaurus
- Sibling (SIB) : il existe une relation " sibling " (frère) dans un des vocabulaires source du Metathesaurus
- AQ : est un qualifiant dans un des vocabulaires source du Metathesaurus
- QB : peut être qualifié à l'aide d'un des concepts du vocabulaire source du Metathesaurus

L'ensemble de ces types de relation est issu des soixante sources consultées lors de l'élaboration du UMLS Metathesaurus. La documentation sur ce dernier ne donne pas plus de détails sur les différences existantes entre ces relations et vu le nombre de sources, nous n'avons pas pu consulter ces dernières pour spécifier les dites différences.

Les définitions ou notes d'application issues des sources sont également ajoutées au concept. Cela a comme conséquence qu'un concept - donc une même notion - peut avoir plusieurs définitions. Au premier abord, cela peut engendrer de la confusion mais les définitions se complètent et permettent finalement une meilleure compréhension du concept et du champ d'application.

Présentation :

Les UMLS Knowledge Sources sont accessibles via Internet ou via CD-Rom. Il faut savoir que cet outil n'est pas destiné à un large public, mais à des développeurs. Pour cette raison, sa présentation n'est pas facilement abordable. Même si les divers types d'information en intention et en extension sont bien séparés en deux colonnes, il faut tout de même un certain temps avant de bien comprendre les correspondances (p. ex. les codes des sources ou les relations hiérarchiques). Un élément important est le fait de pouvoir choisir la source des termes : on peut res-

treindre sa recherche en définissant la source à laquelle doit appartenir le terme requis.

Exemple :

Le terme " endocrinology " :

Tout d'abord, nous voyons à gauche le terme principal : " Endocrinology "; il est

suivi d'un identifiant unique : UI (Unique identifier), ensuite nous avons des informations de type sémantique, les définitions issues des différentes sources dans lesquelles ce concept est repris, les quasi-synonymes et synonymes en anglais, les sources, les équivalences linguistiques, lorsqu'elles existent, et à gauche les structures hiérarchiques de toutes les sources dans lesquelles est repris le concept.

BASIC CONCEPT INFORMATION

Concept Name: Endocrinology
UI: C0014137
Semantic Type: Biomedical Occupation or Discipline
Definition (CSP2000): for thorough searching see all RTs and check the Permuted Term Index for similar alpha strings.
Definition (MSH2001): A subspecialty of internal medicine concerned with the metabolism, physiology, and disorders of the endocrine system.
Definition (PSY97): Scientific discipline dealing with the study of endocrine glands and internal secretions.
Synonyms : endocrinology (field)
Sources: LCH90 , MSH2001 , PSY97 , CSP2000 , PDQ2000
Other Languages : endokrinologia - Finnish ENDOCRINOLOGIE - French Endokrinologie - German Endocrinologia - Italian ENDOCRINOLOGIA - Portuguese ENDOKRINOLOGIJA - Russian ENDOKRINOLOGIJA I METABOLIZM - Russian METABOLIZM I ENDOKRINOLOGIJA - Russian ENDOCRINOLOGIA – Spanish

ANCESTORS

MSH2001 Biological Sciences (MeSH Category) [G] Health Occupations [G2] Medicine [G2.403] Specialties, Medical [G2.403.776] Internal Medicine [G2.403.776.409] Endocrinology [G2.403.776.409.323]
MSH2001 Biological Sciences (MeSH Category) [G] Biological Sciences [G1] Physiology [G1.782] Endocrinology [G1.782.323]
PSY97 Sciences [45825] Medical Sciences [30410] Endocrinology [17370]
CSP2000 biology [0418-4282] physiology [2324-0184] endocrinology [1025-4867]
AOD99 field, discipline, or occupation [S] medical specialty (field) [SG] medical specialty by body system [SG6] internal medicine (field) [SG6.4] endocrinology (field) [SG6.4.18]

Commentaire :

Le UMLS Metathesaurus est un outil très riche et très complexe qui a été élaboré de manière systématique. Il couvre

l'ensemble des domaines biomédicaux de façon très détaillée. Ce système est idéal pour faire des recherches poussées et représente un vocabulaire d'indexation très puissant.

6 BIBLIOGRAPHIE**Référence principale**

SASTRE CANTANO, Isabel, *Analyse préalable à la conception d'un thesaurus européen dans le domaine du " Brain Damage Repair "*. " Examen des questions sémantiques liées à l'interprétation du domaine d'application et à son évolution dans le temps dans un cadre multilingue ". (Mémoire INFODOC, DES en Sciences de l'Information et de la Documentation, 2000-2001, directeur : Isabelle BOYDENS).

Thesaurus

BIOSIS Search Guide, BIOSIS Previews, Biological Abstracts on CD, Biological Abstracts/RRM on CD, édition de 1995, BIOSIS, 1995.

EMTREE Thesaurus 2000, Amsterdam, Elsevier Science B.V., 2000.

Medical Subject Headings, Annotated Alphabetical List – 1991, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine, Library Operations, Medical Subject Headings, Bethesda, July 1990

INSPEC Thesaurus, The Institution of Electrical Engineers, 1995.

UMLS Knowledge Sources, Metathesaurus, Semantic Network, Specialist Lexicon, U.S. Department of Health and Human Services, National Institutes of Health National Library of Medicine, 12^{ème} édition, Janvier 2001

UNESCO Thesaurus, Thesaurus de l'UNESCO, Tesauro de la UNESCO, 1^{ère} version française, Paris, Unesco Publishing, 1983.

UNESCO Thesaurus, Thesaurus de l'UNESCO, Tesauro de la UNESCO, Introduction, Introduction à la 2^{ème} édition du UNESCO Thesaurus, format PDF, Paris, Unesco Publishing, 1995.

AFNOR-1987 : *Vocabulaire de la documentation*, 2^{ième} édition, Paris, AFNOR, 1987.

ISO 2788-1986 : *Documentation - Guidelines for the establishment and development of monolingual thesauri*, International Organization for Standardization, Ref. No. ISO 2788-1986, 1986.

ISO 5964-1985 : *Documentation - Guidelines for the establishment and development of multilingual thesauri*, International Organization for Standardization, Ref. No. ISO 5964-1985, 1985.

Références spécifiques

- [1] Medical Subject Headings (MeSH), <http://www.nlm.nih.gov/mesh/2002/MBrowser.html>.
- [2] BOYDENS, Isabelle, Documentologie, Section Information et Documentation, 2^{ème} édition 2002-03/1, Bruxelles, Université Libre de Bruxelles (syllabus), 2002.
- [3] ISO 5964-1985 : Documentation – Principes directeurs pour l'établissement et le développement de thesaurus multilingues, International Organization for Standardization, Ref. No. ISO 5964-1985 (F), 1985.

- [4] The Foundation Center, http://fdncenter.org/research/grants_class/thes_e.html.
- [5] ERIC Thesaurus, <http://searcheric.org/>.
- [6] European Education Thesaurus, <http://www.eurydice.org/>.
- [7] EUROVOC, <http://europa.eu.int/celex/eurovoc>.
- [8] IBE, <http://www.ibe.unesco.org/International/Publications/Thesaurus/theshome.htm>.
- [9] European Foundation for the Improvement of Living and Working Conditions, <http://www.eurofound.ie/>.
- [10] SASTRE CANTANO, Isabel, Analyse préalable à la conception d'un thesaurus européen dans le domaine du " Brain Damage Repair ". " Examen des questions sémantiques liées à l'interprétation du domaine d'application et à son évolution dans le temps dans un cadre multilingue ". op. cit.
- [11] University of London Computer Centre (Ulcc), <http://www.ulcc.ac.uk/unesco/index.htm>.
- [12] UNESCO Thesaurus, Thesaurus de l'UNESCO, Tesauo de la UNESCO, Introduction, Introduction à la 2^{ème} édition du UNESCO Thesaurus, format PDF, Paris, Unesco Publishing, 1995.
- [13] UNESCO Thesaurus, Thesaurus de l'UNESCO, Tesauo de la UNESCO, Paris, Unesco Publishing, 1^{ère} version française, 1983.
- [14] INSPEC Thesaurus, a thesaurus of terms in physics, electro-technology, computers and control, B.J. Field, 2^{ème} édition, Londres, Institution of Electrical Engineers, 1975.
- [15] SPINES Thesaurus, a controlled and structured vocabulary of science and technology for policy making, management and development, Paris, Unesco, 1976.
- [16] TEST (Thesaurus of Engineering and Scientific Terms) : a list of scientific and engineering terms and their relationships for use in vocabulary for indexing and retrieving technical information, New York, Engineers Joint Council, 1967.
- [17] BSO , 2^{ème} version révisée, La Haye, FID, 1976.
- [18] UNESCO Thesaurus, Thesaurus de l'UNESCO, Tesauo de la UNESCO, Paris, Unesco Publishing, 2^{ème} version française, 1994.
- [19] Elsevier, <http://www.elsevier.nl/homepage/sah/spd/embase/menu.htm>.
- [20] Embase, <http://www.embase.com/otherdb>.
- [21] Elsevier, <http://www.elsevier.nl/inca/publications/store/6/2/1/5/9/0/index.htm>.
- [22] EMTREE Thesaurus 2000, Amsterdam, Elsevier Science B.V., 2000.
- [23] Quick reference guide and example, Embase on DIMDI, format PDF, DIMDI, Juillet 2000.
- [24] DIMDI, <http://www.dimdi.de/germ/vortraege/wk0399/sld017.htm>.
- [25] Silverplatter, <http://www.silverplatter.com/catalog/out/EMBX/other/EMTREE.html>.
- [26] BIOISIS Search Guide, BIOSIS Previews, Biological Abstracts on CD, Biological Abstracts/RRM on CD, édition de 1995, BIOSIS, 1995.
- [27] INSPEC Thesaurus, The Institution of Electrical Engineers, 1995.
- [28] Institution of Electrical Engineers (IEE), <http://www.iee.org.uk/publish/inspec/about.html>.

- [29] National Library of Medicine, <http://www.nlm.nih.gov/mesh/meshhome.html>.
- [30] National Library of Medicine, <http://www.nlm.nih.gov/pubs/factsheets/medline.html>.
- [31] Medical Subject Headings, Annotated Alphabetical List – 1991, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine, Library Operations, Medical Subject Headings, Bethesda, July 1990.
- [32] UMLS Knowledge Sources, Metathesaurus, Semantic Network, Specialist Lexicon, U.S. Department of Health and Human Services, National Institutes of Health National Library of Medicine, 12^{ème} édition, Janvier 2001.

Notes du texte

- ¹ L'action COST B10 " Brain Damage Repair " est une des actions s'inscrivant dans le cadre de la Coopération européenne dans le domaine de la recherche scientifique et technique (COST) sous l'égide de la Commission européenne. C'est l'initiative de recherche la plus ancienne d'Europe (1971) permettant la coordination de projets entre les centres de recherches nationaux à un niveau européen.
- ² Cette base de données est en phase finale d'élaboration et sera bientôt accessible via le site officiel de COST B10 :
<http://dev.ulb.ac.be/~pochet/cgi/>
- ³ Définition de " Langages documentaires " : " Langages artificiels faits de représentations de notions et de relations entre ces notions, destinés, dans un système documentaire, à formaliser les données contenues dans les documents et les demandes des utilisateurs. "
- ⁴ *Medical Subject Headings* est un des vocabulaires contrôlés le plus utilisé et plus apprécié par les experts du domaine biomédical.
- ⁵ Déf. : Langage documentaire composé de concepts hiérarchisés, selon le principe d'un plan de classification.
- ⁶ Déf. : Ensemble contrôlé de termes sélectionnés à partir du langage naturel et utilisé pour représenter, sous une forme condensée, le contenu des documents.
- ⁷ Déf. : Vocabulaire d'un langage d'indexation contrôlé organisé formellement de façon à expliciter les relations a priori entre les notions (par exemple relations générique-spécifique).
- ⁸ Cf. point 4 pour plus de détail.
- ⁹ Le terme " profondeur horizontale " désigne le nombre de domaines ou de groupes couverts par le thesaurus ou l'outil d'indexation.
- ¹⁰ Le terme " profondeur verticale " désigne le nombre de niveaux hiérarchiques à l'intérieur de chaque domaine ou groupe couvert par le thesaurus ou l'outil d'indexation.
- ¹¹ MT = Microthesaurus.
- ¹² Codes : C2.782.815.483.440.40, C2782.815.616.400.40, C2.800.801.40, C2.839.40, C20.673.483.440.40.