

INDEXATION ET QUALIFICATION DES RESSOURCES

Manuel d'utilisation

BSI, juillet 2010

SOMMAIRE

INTRODUCTION

AVANT-PROPOS : LES LANGAGES DOCUMENTAIRES

I – L'INDEXATION SUJET	P. 6
II – AUTRES ACCES AUX RESSOURCES	P. 8
III – POLITIQUE D'INDEXATION	P. 8
IV – PROCESSUS D'INDEXATION	P. 15
V – GESTION DU VOCABULAIRE	P. 18
VI – EVALUATION DE LA QUALITE DE L'INDEXATION	P. 19
VII – MISE EN PLACE D'UNE PLATEFORME DE RECHERCHE FEDEREE	P. 20
VIII– ACCES A L'INFORMATION DANS LE CONTEXTE DU WEB	P. 21
IX – DU REFERENTIEL THEMATIQUE A LA REALISATION D'ONTOLOGIES ?.....	P. 23

ANNEXES

Annexe I : Présentation du Référentiel thématique	P. 28
Annexe II : Rétroagir sur l'indexation des ressources	P. 39
Annexe III : Fiches de maintenance du Référentiel	P. 41
Annexe IV : Exemples de thésaurus et listes de vedettes matière	P. 49
Annexe V : Glossaires de la documentation	P. 52

INTRODUCTION

Durant ces dernières années, la masse d'informations disponibles sur les portails des bibliothèques, centres de ressources, grandes institutions et sur Internet en général a augmenté de manière fulgurante. Les données y sont très diverses et caractérisées par une grande hétérogénéité : texte, images et son ; description de ressources imprimées, bases de données et sites web ; informations techniques, juridiques ou commerciales.

Pour les producteurs de contenus et les fournisseurs de services, la qualité de l'accès à l'information est devenue un défi majeur à relever.

Si les fonctionnalités de certains moteurs de recherche permettent d'indexer et de classier automatiquement les termes repérés dans les contenus, l'analyse documentaire proprement dite reste encore l'apanage des professionnels de l'information et de la documentation.

Organiser les ressources, les décrire de façon structurée et normalisée, donner une représentation cohérente des contenus en s'appuyant sur des référentiels (langage d'indexation sujet entre autres) optimisent l'accès à l'information.

Ce travail d'analyse, les bibliothécaires et les documentalistes le réalisent depuis longtemps avec les fiches de description des ressources (titre, auteur, sujet), sur support cartonné d'abord, puis en format lisible par les systèmes informatiques, maintenant par le biais de métadonnées (notamment celles relatives à l'indexation sujet).

La qualité de ces métadonnées et leur gestion rigoureuse par les professionnels permettent d'optimiser les tris opérés par les moteurs de recherche et améliorent ainsi la visibilité, sur le Web, d'institutions culturelles comme la nôtre. Elles rendent aussi possible la réutilisation des contenus dans des contextes différents et facilitent l'échange d'informations entre institutions partenaires grâce à l'interopérabilité qu'elles supportent.

Cet effort maintenu dans la description normalisée des contenus préparera Universcience au Web de demain : le Web sémantique.

Avant-propos : les langages documentaires

Un langage documentaire permet de décrire une ressource, qu'elle soit imprimée ou numérique, afin de favoriser l'accès à l'information qu'elle contient.

Les principaux langages documentaires actuellement utilisés dans les centres de ressources - bibliothèques, centres de documentation, archives - et les bases de données accessibles sur le Web - sites d'éditeurs, grands sites commerciaux, bases de données professionnelles - sont les **classifications, les listes de vedettes matière et les thésaurus**.

- **Les classifications** : ce sont les premiers langages documentaires qui, en **répartissant les champs de la connaissance en catégories, elles-mêmes divisées en parties**, ont cherché à classer les concepts dans une structure hiérarchique.
Dans de nombreuses bibliothèques, des classifications encyclopédiques, notamment la Classification décimale universelle, celle de Dewey, celle de la Bibliothèque du Congrès constituent toujours un accès par sujet privilégié en permettant l'organisation spatiale des champs de la connaissance.
Sur le Web, les classifications sont adaptées à la nature et à la masse d'informations qui doivent être rendues accessibles aux utilisateurs.
- **Les listes de vedettes matière** : Library of Congress Subject Headings (LCSH), Répertoire des vedettes matière de l'Université de Laval (RVM), Répertoire d'autorité-matière encyclopédique et alphabétique unifié (Rameau) sont des **listes constituées de termes normalisés, reliés entre eux et d'une syntaxe indiquant les règles de construction** pour décrire le sujet contenu dans les documents. Ces langages sont dits précoordonnés, les vedettes matière se construisant (têtes de vedette + subdivisions) au moment de l'indexation des ressources, et sont encore largement utilisés dans les bibliothèques. Des études consacrées aux listes de vedettes matière pourvues d'une syntaxe avancent toutefois l'hypothèse qu'un excès de complexité peut être un frein à leur bonne utilisation. Des propositions très concrètes d'évolution de Rameau ont été émises en 2005, dont celles portant sur la simplification de la syntaxe.
- **Les thésaurus** : ils peuvent être spécialisés (le thésaurus biomédical MESH, le thésaurus GETTY des noms géographiques, le thésaurus européen des systèmes éducatifs TESE) ou plus généralistes (le thésaurus encyclopédique RASUQAM, le thésaurus EUROVOC de l'Union européenne, le thésaurus de l'UNESCO). Tous ont été conçus pour organiser des corpus de ressources à l'intention de communautés d'utilisateurs bien identifiées.

Leurs trois caractéristiques sont d'être :

- **un langage structuré**, de par la structure sémantique des termes, liés entre eux par des relations de hiérarchie et d'association
- **un langage contrôlé** : on contrôle la définition et l'usage des termes choisis.
- **un langage combinatoire** : l'utilisateur peut combiner les termes lors de sa requête (ce qu'on appelle la post-coordination).

Constitués au départ d'unitermes (comme FLEUR, DOPAGE, MOLECULE), les thésaurus ont évolué et accepté l'utilisation de la précoordination pour exprimer certaines notions correspondant à des expressions fréquemment utilisées (telles que RYTHME BIOLOGIQUE, ILLUSION D'OPTIQUE, COMPLEXE D'ŒDIPE).

Sources :

Les langages documentaires / Bruno Menon, Documentaliste, No1, février 2007

<http://www.adbs.fr/les-langages-documentaires-un-panorama-quelques-remarques-critiques-br-et-un-essai-de-bilan-15471.htm?RH=REVUE>

Intérêt pédagogique des langages documentaires / Marie-France Blanquet, Université Bordeaux III, janvier 2000

<http://www.savoirscdi.cndp.fr/index.php?id=569>

Rameau : bilan, perspectives, BBF 2005, T. 50, No 5

<http://bbf.enssib.fr/consulter/bbf-2005-05-0038-001>

I – L'INDEXATION SUJET

Comment donner **un bon accès par sujet à des ressources** imprimées ou numériques ?

Que l'on soit en contexte de précoordination (avec un langage documentaire comme la liste des vedettes matière Rameau), en contexte de post-coordination (avec un thésaurus uniquement composé d'unitermes) ou dans cette double approche (où des unitermes côtoient des locutions précoordonnées proches du langage naturel comme c'est le cas dans le Référentiel indexation), les **principes de base** de l'indexation sont les mêmes.

On recherchera :

- **La pertinence**

« Ce sont les bonnes réponses obtenues par une requête. Le taux de pertinence est le pourcentage exprimant le rapport entre le nombre de documents pertinents extraits et le nombre total de documents extraits. La pertinence est l'inverse du bruit ».*

En règle générale, on indexera un sujet s'il couvre **au moins 25% de la ressource**. En cas de sujet rare et peu traité ou si l'on cherche à valoriser certaines ressources, ce pourcentage, donné à titre indicatif, peut être de l'ordre de 10% (1).

Dans la majorité des cas, le sujet principal couvre bien plus de 25% de la ressource et un seul descripteur pourrait suffire à l'indexer.

- **La sélectivité par rapport à l'exhaustivité**

On privilégie une indexation sélective (1) qui ne retient que les descripteurs qui décrivent le mieux le document.

Si l'on cherche à indexer de manière exhaustive (c'est-à-dire en indexant des concepts moins importants), l'utilisateur risque d'extraire une grande quantité d'information qui ne l'intéresse pas. **Plus le degré d'exhaustivité dans l'indexation est élevé, moins la proportion de documents extraits qui intéressent réellement l'utilisateur sera élevée** (2).

Les seuls concepts à retenir sont ceux susceptibles d'intéresser l'utilisateur et de lui **apporter une information significative** (3). On évite de ce fait la surindexation.

● La spécificité

De manière générale, les ressources sont indexées **le plus précisément et le plus spécifiquement possible**.

On n'indexe pas à la fois au sujet générique et au sujet spécifique (4).

● L'objectivité

«L'indexation doit refléter le contenu du document, et non émettre de jugement de valeur. Elle **doit exprimer le sujet en toute impartialité**, sans mention d'orientation ou de tendance» (4).

● La cohérence

*« Dans le cas idéal, pour un système donné et dans un environnement donné, l'indexation d'un même document par plusieurs indexeurs devrait être identique, l'indexation de plusieurs documents sur un même sujet devrait être identique, et l'indexation d'un document par un même indexeur à des moments différents devrait être identique »**.*

Pour s'assurer que les descripteurs utilisés conviennent, on peut les comparer avec ceux de ressources traitant du même sujet.

Sources :

* Glossaire des techniques de l'information

<http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/glossaire1.htm>

** Glossaire professionnel (Services techniques des bibliothèques de l'Université du Québec à Montréal)

(1) Politiques d'indexation et de gestion du vocabulaire, Services des bibliothèques de l'UQAM (Université du Québec à Montréal), 2007

<http://www.bibliotheques.uqam.ca/services-techniques/politiques-d%E2%80%99indexation-et-de-gestion-du-vocabulaire>

(2) Présentation du thésaurus TESE : Indexation,

http://eacea.ec.europa.eu/education/eurydice/documents/tese/pdf/tesefr_002_intro.pdf

(3) Guide d'indexation Motbis, www.cndp.fr/motbis/telechargement/guide_d_indexation.pdf

(4) Guide d'indexation Rameau, http://guiderameau.bnf.fr/html/rameau_0078.html

II – AUTRES ACCES AUX RESSOURCES

Les ressources peuvent être qualifiées par leur sujet mais d'autres critères permettent de compléter leur description, notamment :

- les publics visés : tout public, enfants, scolaires, enseignants, etc.
- le type de ressource : encyclopédie, dossier documentaire, conférence, etc.
- le support : ressource imprimée, DVD, site Web, etc.
- l'intention éditoriale : ressource de vulgarisation, ressource humoristique, etc.
- la localisation géographique, la langue, les dates de publication, etc.

La combinaison de tous ces marquages multiplie les accès aux contenus et améliore l'information donnée à l'utilisateur.

III – POLITIQUE D'INDEXATION

La politique d'indexation permet d'harmoniser les pratiques d'indexation et de faciliter l'utilisation du Référentiel thématique.

Elle détermine quelles sont les ressources à indexer. Elle définit le niveau de précision du vocabulaire et le degré d'exhaustivité souhaités dans la description des contenus.

Elle prend en compte les besoins et comportements des usagers.

Elle garantit la cohérence de l'indexation de tous les contenus et ainsi, la pertinence des résultats de recherche.

1 - Quelles sont les ressources que l'on indexe ?

■ 1^{er} critère : l'unité documentaire

Un document fera l'objet d'une indexation du fait de son *unité documentaire*, c'est-à-dire de sa cohérence structurale et de l'homogénéité de son contenu.

Les livres, les périodiques, les cédéroms, les films répondent à ce critère. Ces documents font l'objet d'une présentation matérielle et d'une gestion physique. Les auteurs, les éditeurs et les producteurs les ont désignés en tant qu'unités documentaires.

Pour les documents à grande quantité d'information, on peut toutefois se poser la question d'une indexation au niveau de l'ensemble du document ou à celui de ses parties :

- **Un document à grande quantité d'information et à caractère encyclopédique** sera indexé au sujet principal (et non à ses parties). Il faudra en revanche le repérer en tant que document de référence et donner accès à sa table des matières et à ses index.

L'ouvrage encyclopédique « Biologie » de Neil A. Campbell et Jane B. Reece qui aborde toutes les disciplines des sciences de la vie sera indexé à SCIENCES DE LA VIE et non à chacune des disciplines traitées.

- **Un document à grande quantité d'information** pourra être indexé au niveau de ses parties, à condition que les parties soient bien identifiées, les contenus différents et que la quantité d'information livrée soit jugée suffisante*.

On indexera « Qu'est-ce que l'univers ? », ensemble de contributions organisées en chapitres autonomes sur plusieurs disciplines des sciences exactes à MATHÉMATIQUES, ASTROPHYSIQUE, SCIENCES DE LA TERRE, PHYSIQUE, CHIMIE

** L'information jugée suffisante doit être au moins égale à la quantité d'information livrée dans un article d'une ressource encyclopédique.*

La question de l'unité documentaire à considérer se pose surtout lorsqu'on a affaire à des ressources en ligne. L'émergence de l'information composite sur le web - réutilisation de parties de documents pour réaliser une nouvelle entité, notamment avec l'usage de plus en plus fréquent de la syndication - a fait évoluer la notion même de document. La décomposition d'un document ouvre la porte à la génération d'autres documents.

Les dossiers documentaires et les sélections documentaires, les sites web, les conférences, manipulations et rencontres – tous les contenus en ligne d'Universcience - peuvent être considérés comme des entités en soi et/ou comme les parties d'une unité plus large.

On parle de « **granularité** » pour désigner cette plus ou moins grande finesse de l'unité informationnelle à considérer.

- **S'il y a cohérence, du point de vue du contenu**, entre l'unité principale et les sous-parties, on peut considérer qu'il suffit d'indexer au niveau générique.

Le dossier documentaire en ligne « Déforestation : vers la disparition des forêts naturelles » peut n'être indexé qu'à DEFORESTATION qui en est le sujet principal

- **Si une sous-partie apporte une information différente, originale**, si elle peut être considérée comme une unité cohérente, elle doit être créée en tant qu'objet documentaire à part entière, à qualifier, notamment à indexer.

Dans ce même dossier sur la déforestation, on peut aussi considérer que l'une des parties, le film « Green » de Patrick Roussel (sur les derniers jours de la vie d'un orang-outan victime de la déforestation) donne une information intéressante sur le comportement des orangs-outans. On peut alors décider de faire de cette partie un objet informationnel à qualifier et de l'indexer à ORANG-OUTAN

■ 2ème critère : la quantité d'information contenue dans la ressource

La **quantité d'information** contenue dans la ressource doit être **suffisante**. Seules les ressources ayant un **contenu significatif** sont indexées. Le contenu doit être **exploitable** par les usagers.

Ce critère est néanmoins à nuancer. Certains éléments justifient en effet l'indexation d'une ressource à faible quantité d'information si l'on souhaite la promouvoir auprès des usagers :

- lorsqu'il s'agit d'une production d'Universcience, à **vocation pédagogique**

Dans le « Coin des juniors : Quelques expériences en animation », du Palais de la Découverte, de petites animations ludiques à destination des juniors introduisent des notions scientifiques très importantes, utiles pour toute personne souhaitant s'initier aux disciplines scientifiques

- lorsque la ressource fait preuve d'**originalité**

Dans un programme de la Web TV « C'est beau, une image de science », de petits films de 2 minutes sensibilisent à l'imagerie scientifique. La forme de la ressource (caractère attrayant de l'image) peut justifier qu'on la traite en tant qu'objet documentaire

Ces deux aspects sont particulièrement présents dans les ressources à destination des enfants.

- si le sujet en question est **peu traité** et **les ressources rares**

■ L'indexation de la fiction, des récits, des témoignages

Si, dans une fiction, le propos explicite est de transmettre une information sur un **sujet scientifique ou technique**, il est recommandé d'indexer les ressources concernées. Les récits et les témoignages doivent également obéir à ce critère pour être indexés.

En résumé

Une ressource fera l'objet d'une indexation :

- du fait de son unité documentaire (cohérence structurale, homogénéité de contenu)
- si la quantité d'information est suffisante et exploitable par l'utilisateur
- si le propos est de transmettre une information scientifique ou technique en cas de fiction, récit et témoignage

2 - Quelle profondeur de l'indexation vise-t-on ?

Un des problèmes majeurs rencontrés dans le travail d'indexation est **le choix du niveau d'indexation**, c'est-à-dire la plus ou moins grande finesse de l'indexation, la plus ou moins grande spécificité des descripteurs à utiliser.

En règle générale, les ressources doivent être indexées le plus précisément et le plus spécifiquement possible.

En cas d'absence du descripteur recherché, il faut se poser la question de la marginalité du sujet et de sa spécificité :

■ La marginalité du sujet par rapport au cœur de l'offre d'Universcience

On indexera différemment les ressources qui sont au cœur de l'offre et celles qui sont à la marge.

Lorsqu'aucun descripteur ne permet de traduire un concept marginal ou une thématique périphérique, on indexe de façon plus générique, sachant qu'un accès à la ressource est rendu possible par la recherche « mot » (il est recommandé de faire figurer des mots clés dans les résumés ou les éditos).

Pour indexer une ressource sur la culture de la canne à sucre aux Antilles qui, entre le 16^e et le 19^e siècle, s'appuie sur le système esclavagiste, c'est par HISTOIRE COLONIALE ou HISTOIRE ECONOMIQUE ou HISTOIRE SOCIALE qu'on précisera les descripteurs principaux, le descripteur SYSTEME ESCLAVAGISTE n'existant pas.

■ La grande spécificité du sujet traité par la ressource

Si la règle générale veut que les descripteurs assignés ne doivent pas décrire un contenu plus large que le contenu de la ressource, aboutir à une indexation qui ne pointerait que sur une ou deux ressources, du fait de la grande spécificité du sujet n'est guère pertinent. Si le descripteur recherché n'existe pas, on se contente du niveau de spécificité fourni par le Référentiel et on indexe au niveau de généralité immédiatement supérieur.

« Le lave-vaisselle : son fonctionnement, ses pannes, son entretien ». Le descripteur LAVE-VAISSELLE n'ayant pas été créé en raison de sa grande spécificité et du faible nombre de ressources concernées, le document sera indexé à ELECTROMENAGER. L'accès par mot clé dans le corps de la notice palliera l'absence du descripteur.

En résumé

Les ressources doivent être indexées le plus précisément et le plus spécifiquement possible, sauf :

- **en cas de marginalité du sujet** par rapport au cœur de l'offre d'Universcience : il est possible qu'un sujet marginal ne puisse pas être indexé de façon précise vu l'absence de descripteur adéquat ; il sera indexé au niveau de précision permis par le Référentiel,
- **en raison de la grande spécificité du sujet et du très faible nombre de ressources** concernées : en cas d'absence de descripteur adéquat, on indexe au niveau de généralité immédiatement supérieur.

3 – La pluri-indexation

Il n'est pas souhaitable de chercher à indexer de manière exhaustive, c'est-à-dire d'indexer tous les sujets identifiés, même ceux de moindre importance (se reporter au principe de l'indexation sélective et au critère de la quantité d'information suffisante contenue dans la ressource). **L'objectif de l'indexation est de refléter au mieux le sujet principal d'une ressource.** Surindexer fait perdre de sa pertinence à la description du contenu.

Il est normal en revanche d'avoir parfois recours à un usage croisé de deux, voire trois descripteurs, le Référentiel thématique étant fondé sur une logique de descripteurs qui se combinent a posteriori au moment de la requête (post-coordination).

« Evaluer la prévention de l'infection par le VIH en FRANCE » sera indexé à SIDA et PREVENTION DES MALADIES

De manière générale, la préconisation est de **se limiter à trois descripteurs**, hors précisions géographiques, historiques ou de point de vue (voir plus loin l'utilisation de ces précisions).

L'emploi de descripteurs ayant une relation hiérarchique entre eux (coexistence d'un terme générique et de son/ses spécifique/s) est à proscrire.

Cas de figure pour lesquels se pose la question de la pluri-indexation :

1 - Une ressource a des parties bien identifiées, les contenus sont différents. L'emploi de plusieurs descripteurs peut se justifier, sous réserve que chacune des parties ait une quantité d'information suffisante.

« La nature en France : itinéraires et guide visuel pour tout reconnaître en un clin d'œil » sera indexé à IDENTIFICATION DES ANIMAUX, IDENTIFICATION DES PLANTES, ROCHE et FRANCE plutôt qu'à NATURE et FRANCE

2 - Le contenu de la ressource est très homogène, les sujets qui émergent sont spécifiques d'un ensemble plus générique. La règle générale est d'utiliser le descripteur générique et non plusieurs descripteurs (encore moins la totalité des spécifiques d'un terme générique).

Il suffira d'indexer «Le Recknagel », ouvrage de référence en génie climatique, par GENIE CLIMATIQUE et non à CHAUFFAGE, CLIMATISATION, TECHNIQUE DU FROID qui en sont les spécifiques

3 - Le sujet est transversal et traite de la rencontre de deux champs disciplinaires différents. Avant de pluri-indexer, on cherche si une locution précoordonnée ne rend pas compte de cette rencontre.

*« L'artiste et l'ordinateur ». On adopte le descripteur ART NUMERIQUE
« L'internet des objets ». Indexer à INTERNET et OBJET n'aurait pas de sens. On décide de créer la locution INTERNET DES OBJETS pour rendre compte de cette réalité.*

4 - Descripteurs principaux et secondaires

Le choix a été fait d'attribuer une valeur différente aux descripteurs selon qu'ils sont « principaux » ou « secondaires ». Un descripteur est considéré comme principal s'il décrit le sujet essentiel de la ressource. Un descripteur concernant un aspect moins important (à condition toutefois qu'il couvre au moins 25% de la ressource) ou apportant une précision sur le sujet principal (précision géographique, historique ou de contenu) est considéré comme secondaire.

En cas de pluri-indexation, la pondération définie par la valeur attribuée aux descripteurs est prise en compte par le moteur de recherche. Cela devrait permettre à l'utilisateur de déterminer rapidement si l'information indexée est essentielle ou bien simplement complémentaire.

5 - Utilisation des précisions géographiques, historiques et de point de vue

◆ Descripteurs géographiques

On estime que cette précision est importante.

*« Guide des sites naturels de France », « Science in the new Russia », « Le design au Japon »
« L'avènement de la médecine clinique moderne en Europe »*

◆ Descripteurs historiques

On évite d'utiliser plus de deux descripteurs chronologiques à la suite (15E SIECLE, 16E SIECLE, 17E SIECLE, etc.) lorsqu'une période historique importante est couverte.

« Les grandes Alpes dans la cartographie, 1482-1885 » sera indexé à ALPES et à HISTOIRE DE LA CARTOGRAPHIE

On utilise un descripteur chronologique pour préciser l'histoire d'une discipline.

Pour l'ouvrage « Copernic et le mouvement de la terre », on précisera le descripteur HISTOIRE DE L'ASTRONOMIE à l'aide du descripteur chronologique 16E SIECLE

On différencie les documents qui ont une approche historique (l'histoire d'une discipline), les documents qui parlent d'un objet à une période donnée (on utilise le descripteur qui qualifie l'objet en lui adjoignant un descripteur chronologique) et les sources.

- « La chimie, c'est toute une histoire : idées et conquêtes des origines à nos jours » sera indexé à HISTOIRE DE LA CHIMIE

- On indexera « Le pendule de Foucault » à PENDULE et à 19E SIECLE

- Un document de Copernic sera indexé à ASTRONOMIE et à 16E SIECLE ; on le marquera en tant que « source ».

◆ Précision de point de vue

Elle se traduit par l'emploi de « mots outils ». Elle permet de distinguer ce qui est l'approche d'une discipline, l'angle sous lequel est abordé le sujet, de ce qui est la rencontre de deux champs disciplinaires équilibrés, voire d'un nouveau champ d'étude.

La coexistence des deux descripteurs MATHEMATIQUES et ASPECT PHILOSOPHIQUE ne rend pas compte du même contenu que le descripteur précoordonné PHILOSOPHIE DES MATHEMATIQUES.

La pluri-indexation en résumé

On ne cherche pas à indexer de manière exhaustive. Surindexer fait perdre de sa pertinence à la description du contenu.

Pour rendre compte d'un sujet, on a parfois recours à l'usage croisé de deux, voire trois descripteurs (combinaison au moment de la requête).

La préconisation est de **se limiter à l'emploi de trois descripteurs**, hors précisions géographiques, historiques ou de contenu.

On attribue une **valeur différente aux descripteurs**, selon qu'ils sont principaux ou secondaires.

4 – Le niveau de langue

Le niveau de langue n'est pas un critère retenu. Il a été traité en amont, dans la construction du Référentiel, par les renvois d'exclusion. Le parti a été pris de ne jamais sacrifier la rigueur scientifique au niveau de langue.

ANIMAL DISPARU (EP ANIMAL PREHISTORIQUE)
IMMEUBLE DE GRANDE HAUTEUR (EP GRATTE-CIEL)
OBJET CELESTE (EP ASTRE)
POLYMORPHISME GENETIQUE (EP DIVERSITE GENETIQUE)

La possibilité de faire une indexation « différenciée » existe toutefois.

On n'indexera pas de la même manière une ressource dont le contenu porte sur un objet (VOLCAN), une discipline (VOLCANOLOGIE) ou un processus (ERUPTION VOLCANIQUE).

Les documents de sensibilisation ou à l'intention des enfants ont plus de chance de traiter de l'objet que de la discipline.

IV – PROCESSUS D'INDEXATION

Le processus d'indexation s'effectue en trois étapes et doit aboutir à une phase de contrôle afin de vérifier la cohérence de l'indexation :

1. Analyse du contenu de la ressource
2. Sélection des principaux concepts en langage naturel
3. Traduction de ces concepts en langage documentaire
4. Phase de contrôle

1 – Analyse du contenu de la ressource

Cette analyse s'appuie sur les informations livrées par le titre, le sommaire ou la table des matières, l'introduction et la conclusion, le résumé ou l'éditorial, certaines notions mises en valeur soit dans le contenu même de la ressource soit citées et commentées dans une ressource externe.

Son but est d'identifier et de relever tous les sujets et concepts contenus dans la ressource. L'exhaustivité est ici privilégiée.

2 – Sélection des principaux concepts en langage naturel

Les sujets et concepts trouvés précédemment lors de l'analyse de contenu ne sont pas nécessairement tous retenus.

A cette étape en effet, **ne sont sélectionnés que les éléments utiles**, ceux qui apportent une **information réellement significative et exploitable** par les usagers : on se pose la question de la quantité d'information apportée et de sa pertinence, on étudie l'environnement documentaire.

Sont identifiés :

- le **sujet principal**
- les **sujets secondaires** s'ils constituent une part suffisamment importante de la ressource
- le **lieu, le temps ou la période chronologique couverte, le point de vue** si ce sont des précisions pertinentes apportées au sujet

3 – Traduction de ces concepts en langage documentaire

Les principaux concepts identifiés (on a parfois besoin de plusieurs concepts pour représenter un sujet), on passe à l'opération d'indexation proprement dite, c'est-à-dire au choix des descripteurs qui exprimeront au mieux ces concepts.

Cette étape prend en compte les préconisations de la politique d'indexation :

- précision et spécificité des descripteurs en rapport avec la spécificité des concepts sélectionnés dans le contenu
- choix d'un descripteur plus générique lorsque le descripteur permettant de traduire un concept marginal et/ou extrêmement spécifique n'existe pas (recommandation de faire figurer des mots clés dans les résumés ou les éditoriaux)
- indexation des ressources à forte quantité d'information et à caractère encyclopédique au sujet principal et non à ses parties
- indexation au niveau des parties de la ressource si les parties sont bien identifiées, les contenus différents, l'information délivrée dans chacune des parties jugée suffisante et/ou en cas de rareté des ressources sur le sujet
- recours à l'usage croisé de deux, voire trois descripteurs pour exprimer au mieux le sujet (combinaison lors de la requête en contexte de post-coordination)
- utilisation de locutions précoordonnées lorsqu'elles sont disponibles plutôt qu'emploi de deux descripteurs génériques (DROIT DE L'INFORMATIQUE et non DROIT et INFORMATIQUE)
- utilisation des descripteurs géographiques, historiques ou des mots outils (point de vue sous lequel est abordé le sujet) si ces précisions s'avèrent importantes
- non utilisation simultanée du descripteur générique et de l'un de ses spécifiques

- attribution de la valeur à donner aux descripteurs (descripteur principal ou descripteur secondaire).

Rappel : de manière générale, la préconisation est de **ne pas utiliser plus de trois descripteurs**, hors précisions géographiques, historiques ou de point de vue.

Lorsque le descripteur jugé adéquat n'a pas été trouvé, il est conseillé de naviguer dans l'arborescence du domaine concerné, d'explorer les relations de hiérarchie et d'association et de rechercher le descripteur le plus proche du concept en question ou, à défaut, de choisir le niveau de généralité immédiatement supérieur. En cas de lacune avérée, une proposition de création de descripteur peut être faite. (Voir p. 16)

4- Phase de contrôle

Pour une bonne cohérence de l'indexation, il est recommandé d'interroger la base documentaire avec les descripteurs assignés et d'évaluer la cohérence des réponses.

On peut également s'assurer que les descripteurs utilisés conviennent en les comparant avec ceux de ressources traitant du même sujet.

La prise en compte du nombre de réponses obtenues peut aussi aider à réévaluer l'indexation. Si le descripteur (ou la combinaison des descripteurs) engendre trop de réponses, il peut être judicieux d'indexer plus spécifiquement ou d'adjoindre un nouveau descripteur qui, combiné à ceux déjà assignés, améliorera la précision des réponses. À l'inverse, un trop faible nombre de réponses peut, dans certains cas, alerter sur le fait que le descripteur sélectionné n'est pas adapté.

Sources :

Étapes de l'indexation, Guide d'indexation Rameau

http://guiderameau.bnf.fr/html/rameau_0088.html#d11e2568

Étapes de l'indexation, Guide d'indexation Motbis

http://www.cndp.fr/motbis/telechargement/guide_d_indexation.pdf

L'indexation analytique matière. Dans : Choix de vedettes matières à l'intention des bibliothèques / F. Danset, M. Blanc-Montmayeur, Cercle de la librairie, 2002

Présentation du thésaurus TESE : Indexation

http://eacea.ec.europa.eu/education/eurydice/documents/tese/pdf/tesefr_002_intro.pdf

Processus d'indexation, Services techniques des bibliothèques de l'Université du Québec à Montréal

<http://www.bibliotheques.uqam.ca/sites/default/files/services/services-techniques/politiques-indexation.pdf>

V – GESTION DU VOCABULAIRE

Le Référentiel thématique est amené à évoluer. Comme tout langage d'indexation, c'est un outil qu'il faudra :

- **actualiser, en fonction de l'évolution des domaines de la connaissance**, par l'ajout de descripteurs (nouveaux concepts, nouveaux champs d'étude) et/ou par la suppression de termes obsolètes ou peu pertinents,
- **enrichir au niveau lexical**, par l'apport de relations d'association et de synonymes,
- **modifier ou corriger**, en cas d'erreurs dans les relations hiérarchiques

L'actualisation du Référentiel* n'implique cependant pas qu'il s'accroisse de façon exponentielle. La quantité des descripteurs devra rester à volume constant et il est souhaitable que le niveau de spécificité du vocabulaire n'augmente pas.

Le projet de compléter les descripteurs – tout au moins une partie d'entre eux - de définitions et notes d'application est toujours à l'ordre du jour.

Pour gérer au mieux ce vocabulaire structuré, toute suggestion de création, modification (y compris enrichissement) ou suppression d'un descripteur sera à transcrire par le biais d'une fiche indexation.

Trois fiches correspondent à ces cas de figure :

- **Fiche N° 1 – Création d'un nouveau descripteur**
Chaque descripteur devant être intégré dans une structure hiérarchique, de préférence déjà existante, il est impératif de préciser son supérieur hiérarchique immédiat (**TG : terme générique**). Si ce nouveau descripteur s'accompagne d'un ou plusieurs descripteurs spécifiques (**TS : terme spécifique**), une fiche sera à remplir pour chacun d'eux. Une relation d'association peut également être proposée (**VA : terme associé**). Enfin, en cas de termes sémantiquement très proches (synonymie ou quasi-synonymie), le nouveau descripteur peut être enrichi d'un ou plusieurs termes rejetés (**EP : employé pour**).
- **Fiche N° 2 – Modification (ou enrichissement) d'un descripteur existant**
La modification du descripteur peut concerner son **intitulé** (sans qu'il y ait obligatoirement changement de sa place dans la structure hiérarchique), son ou ses **EP** (Employé pour : terme exclu), son ou ses **VA** (Voir aussi : terme associé), ses **relations hiérarchiques** (terme générique, terme/s spécifique/s).

* Lors des mises à jour, une partie des ressources devrait pouvoir être réindexée automatiquement mais des corrections manuelles seront toujours indispensables.

Fiche N° 3 – Suppression d'un descripteur existant

Lorsqu'on juge qu'un descripteur est inadéquat ou superflu, il est nécessaire de se demander si sa suppression entraîne ou pas la création d'un nouveau descripteur, provoque le changement de place ou la suppression de son/ses spécifique/s, conduit ou pas à le supprimer en tant que terme associé.

Il a été envisagé que ces fiches soient disponibles sur l'intranet et envoyées par Groupwise à indexation@universcience.fr. Lorsque les modalités de leur transmission seront arrêtées, l'information sera largement communiquée.

VI – EVALUATION DE LA QUALITE DE L'INDEXATION

L'indexation sujet a pour objectif de représenter le contenu des ressources et d'en donner le meilleur accès possible. C'est à ce titre qu'elle se doit d'être pertinente et cohérente. De nombreux outils et méthodes permettent d'en évaluer et d'en améliorer la qualité, parmi lesquels :

- **Enquêtes auprès des utilisateurs**
 - des publics pour connaître leur degré de satisfaction (pertinence des résultats de recherche, facilité de navigation dans les relations sémantiques, etc.)
 - des indexeurs pour identifier leurs besoins, leurs difficultés

- **Modules de sensibilisation ou formations à l'indexation, réunions de travail**
 - mises à niveau et formations en fonction des besoins exprimés
 - réunions de travail afin d'harmoniser les pratiques d'indexation
 - réunions pour une analyse concertée des propositions de création, modification ou suppression de descripteurs, en vue de l'enrichissement et de l'actualisation du Référentiel

- **Confrontation avec d'autres systèmes d'indexation**
 - prise en compte de l'indexation automatique du moteur de recherche et du « tagging » des usagers (mots clés attribués aux ressources par les usagers eux-mêmes) si les technologies adéquates sont mises en place

 - comparaison avec l'indexation de notices issues de réservoirs différents (pour la BSI, comparaison avec l'indexation Rameau des notices importées du catalogue du Sudoc - Système universitaire de documentation)

- **Contrôles automatiques réguliers pour connaître le taux d'utilisation du Référentiel et analyse des résultats**

VII – MISE EN PLACE D’UNE PLATEFORME DE RECHERCHE FEDEREE

Plusieurs systèmes d’information coexistent au sein d’Universcience : les systèmes de gestion de bibliothèque du Palais de la Découverte et de la Cité des sciences, le système de gestion de contenu Content Server et les différentes bases de données documentaires (base des audiovisuels, des archives, photothèques, etc.).

Cette diversité rend difficile une vision globale de l’offre.

La mise en place d’une **plateforme de recherche fédérant l’ensemble des contenus d’Universcience quelles que soient leur nature et la source dont ils proviennent** facilitera les recherches des usagers.

Les utilisateurs doivent pouvoir explorer l’ensemble de l’offre au moyen d’une **stratégie de recherche unifiée**, soit en formulant leurs requêtes via un champ de saisie libre avec possibilité d’organiser les résultats selon des filtres de tri, soit en naviguant dans la classification thématique avec possibilité de relancer la recherche sur un thème ou un sous thème. Des requêtes standard par profil pourraient également être proposées.

L’indexation des contenus n’étant opérante que si l’information est inscrite dans les systèmes d’information utilisés, **il est impératif que toutes les zones de qualification des ressources soient renseignées** (sujet, type de publics, support, intention éditoriale, etc.) **au moyen de référentiels communs** (thématique, publics cibles, types de ressources, etc.).

Des informations correctement qualifiées – en particulier celles relatives au contenu des ressources par le biais de descripteurs adéquats - ne peuvent que rendre plus performantes les fonctionnalités d’un moteur de recherche paramétré en conséquence (pondération des champs, création de profils d’usagers, choix des critères de filtrage des résultats, etc.).

Le déploiement du Référentiel thématique dans le moteur de recherche, en permettant au moteur d’effectuer ses tris au moyen d’un vocabulaire contrôlé et structuré, améliorera le taux de pertinence des résultats.

Pour information sur les solutions techniques mises en place ou envisagées, contacter STI-BSI/DSI/DIPN

VIII – ACCES A L'INFORMATION DANS LE CONTEXTE DU WEB

L'information n'a jamais été aussi abondante, les ressources aussi nombreuses. La recherche d'information n'aboutit malheureusement pas toujours à des réponses pertinentes.

L'enjeu est donc de se donner les moyens de trier, sélectionner, vérifier l'information ; de garantir une qualité d'accès à l'offre ; de prendre en compte la diversité des besoins ; de proposer des services innovants.

Deux mouvements coexistent à l'heure actuelle sur le Web :

- **l'un qui va vers plus de normalisation**, plus de structuration dans la représentation des contenus, plus de métadonnées (1) (Dublin Core ou autres formats de description des ressources), plus de relations sémantiques entre les objets documentaires pour une exploration fine de l'information
- **l'autre qui va vers le partage d'une information qui est libre de circuler**, d'être modifiée, commentée, où l'utilisation du Web est en quelque sorte démocratisée (Web 2.0) mais où les contenus produits sont faiblement structurés, indexés de façon personnelle sans réelle normalisation (2).

Pour les professionnels de l'information, se présente l'opportunité de réunir ces deux mondes :

- **en offrant des accès de qualité à des contenus structurés et bien qualifiés**
 - par le biais de systèmes stables qui contrôlent le contenu (langages d'indexation)
 - grâce aux performances accrues des moteurs de recherche (indexation, catégorisation ou clustérisation automatiques du texte intégral (3) - souvent en appui au vocabulaire contrôlé utilisé, techniques d'extension sémantique (4), etc.)
 - à moyen terme, en s'appuyant sur des systèmes de modélisation des connaissances et des usages, avec la réalisation d'ontologies (Voir chapitre suivant)
- **mais en sachant diversifier ces accès** pour mieux répondre aux besoins individuels et aux centres d'intérêt partagés de communautés virtuelles, grâce aux outils de la démarche participative du Web social (5) :
 - hiérarchisation et affichage de « tags » (mots clés attribués aux ressources par les usagers eux-mêmes), possibilité de lancer une requête par « tags »
 - affichage des commentaires, des recommandations, des votes des usagers
 - suggestions personnalisées de ressources proches des requêtes, etc.

L'enjeu serait de réussir à faire coexister la qualité des contenus et la dynamique du Web social pour donner accès à une information plus pertinente.

Sources :

- (1) « Dossier sur les métadonnées », BiblioDoc.Francophonie.org
http://www.bibliodoc.francophonie.org/article.php3?id_article=172

« Documents numériques et métadonnées », BNF
http://www.bnf.fr/fr/professionnels/num_metadonnees/s.num_metadonnees_documents.html

« Dublin Core », BNF
http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_dublin_core.html
- (2) « Folksonomies : les usagers indexent le web », Olivier Le Deuff, BBF, T. 51, No 4, 2006
<http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>
- (3) « Indexation automatique », ADBS
http://www.adbs.fr/indexation-automatique-17377.htm?RH=OUTILS_VOC

« Les techniques de clustérisation », ADBS
http://www.adbs.fr/les-techniques-de-clusterisation-6093.htm?RH=OUTILS_RECHWEB

« Clustering », ADBS
http://www.adbs.fr/clustering--52476.htm?RH=OUTILS_VOC
- (4) « Le sens de la sémantique → L'extension sémantique », Le Journal du Net, 21 février 2010
<http://www.journaldunet.com/solutions/expert/45184/le-sens-de-la-semantique.shtml>
- (5) « Les outils du web 2.0 en bibliothèque », BBF, T. 52, No 6, 2007
<http://bbf.enssib.fr/consulter/bbf-2007-06-0100-011>

« Bibliothèques, tags et folksonomies : l'indexation des bibliothèques à l'ère sociale », Claire Lebreton, Mémoire d'étude, Enssib, mars 2008 (texte intégral consultable en ligne, recherche du titre via Google en cas de lien inactif)
<http://www.enssib.fr/bibliotheque-numerique/document-1750>

Pour une information plus complète sur les nouvelles fonctionnalités des moteurs de recherche, contacter STI-BSI/DSI/DIPN

IX – DU REFERENTIEL THEMATIQUE A LA REALISATION D'ONTOLOGIES ?

Les **ontologies** sont appelées à jouer un rôle essentiel dans le développement du Web par leur capacité à mettre en relation les données qui y sont disponibles, via une formalisation du sens des contenus, des langages de description adaptés et des mécanismes permettant d'effectuer des traitements automatisés.

C'est ce lien entre les données, dans un contexte d'interopérabilité, qui sous-tend les travaux sur le Web sémantique qu'il serait plus juste d'appeler « Web de données ».

Qu'est-ce qu'une ontologie ?

Concept philosophique, le terme, repris dans les domaines de l'Intelligence artificielle et des Sciences de l'information, désigne un **modèle conceptuel d'organisation et de représentation des connaissances**.

« Une ontologie est la conceptualisation d'un domaine, c'est-à-dire **un choix quant à la manière de décrire un domaine**, et la spécification de cette conceptualisation, c'est-à-dire sa description formelle » (Wikipédia, d'après la définition donnée par Thomas Gruber*).

A qui les ontologies s'adressent-elles ?

Bien qu'elles puissent être de niveaux de généralité différents (réalisées avec des concepts généraux ou dites de « domaine », de « tâche », voire « d'application »), leur objectif est de s'adresser à **une communauté d'utilisateurs pour qui cette conceptualisation fait sens**. Le monde ne s'interprétant pas de façon unique pour l'ensemble des individus, il n'y a pas qu'une seule ontologie de référence pour un domaine précis.

Comparaison avec les thésaurus

Les ontologies partagent avec les thésaurus la même structuration hiérarchique des concepts, mais elles s'en écartent du fait de **la plus grande précision et de la richesse des relations d'association** entre les concepts.

Vers la réalisation d'ontologies à Universcience ?

Créer des ontologies à partir du Référentiel serait tout à fait envisageable à condition :

- d'identifier clairement les communautés d'utilisateurs à qui elles sont censées s'adresser (les publics d'Universcience sont variés, les utilisateurs en interne nombreux)

* *Formal ontology in conceptual analysis and knowledge representation*. Kluwer Academic Publishers, 1993.

- d'adapter ces ontologies à la spécificité de leurs besoins, dans des buts d'utilisation bien définis (modélisation des besoins et explicitation des usages)
- de déterminer avec précision les domaines à couvrir et de limiter leur portée aux besoins exprimés (sur un domaine précis, il pourrait y avoir autant d'ontologies que de communautés intéressées par ce domaine mais avec des points de vue et des besoins différents)
- d'anticiper sur leur durée de vie et leur évolution souhaitées, leur réutilisation éventuelle
- d'en rendre l'exploitation informatique possible grâce aux apports des technologies du web sémantique

A titre d'exemples :

- Pour une communauté d'utilisateurs intéressée par la **biodiversité** et pour laquelle la préservation des espèces ne peut s'envisager que dans le cadre d'espaces géographiques, établir des liens « typés » entre une partie de la systématique botanique/zoologique et des lieux géographiques
- Pour ceux qui, dans un contexte médical, s'intéressent aux **molécules chimiques**, les relier avec les médicaments, les maladies ou processus physiologiques impliqués et les organes touchés
- Pour les publics d'Universcience intéressés par **l'évolution des sciences et des techniques**, associer les principaux champs d'étude ou disciplines et les personnalités incontournables (chercheurs, ingénieurs) ou institutions de référence qui les ont illustrés
- Pour les utilisateurs internes, modéliser les relations d'Universcience avec ses partenaires, etc.

Quel que soit le nom qu'on lui donne, « ontologie » ou « extension de thésaurus », c'est à la représentation commune de connaissances dans un domaine donné, à sa compréhension partagée, pour des besoins spécifiques et avec un objectif commun qu'un outil de ce type renvoie.

Concernant le Référentiel indexation, sa structure actuelle devrait pouvoir, dans nombre de cas, convenir pour l'indexation des ressources d'Universcience.

Mais un gros **travail sur les relations d'association** entre les descripteurs reste à faire.

Cet enrichissement, à condition de procéder ultérieurement au « typage » des relations en question, pourrait contribuer à l'élaboration d'ontologies. Ce serait une première étape.

Sources :

http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.elaboration_ontologies.html

http://www.technolangue.net/imprimer.php3?id_article=280

http://interstices.info/jcms/c_17672/ontologies-informatiques?qs=id%3Djalios_5127&part=0

[http://fr.wikipedia.org/wiki/Ontologie_\(informatique\)](http://fr.wikipedia.org/wiki/Ontologie_(informatique))

http://archivesic.ccsd.cnrs.fr/docs/00/20/24/40/PDF/cais-acsi_zacklad_-_avec_ref.pdf

http://liris.cnrs.fr/gom/JFO_2007/11.pdf

ftp://ftp.fao.org/gi/gil/gilws/aims/references/flyers/ontologies_fr.pdf

ANNEXES

Annexe I : Présentation du Référentiel thématique	
1. Structure de base et données chiffrées	P. 28
2. Choix et forme du vocabulaire	P. 35
3. Relations sémantiques	P. 36
4. Elaboration des hiérarchies	P. 37
5. Profondeur des hiérarchies	P. 37
6. Polyhiérarchie	P. 37
7. Arborescence «Histoire des sciences et des techniques»	P. 38
8. Définitions et notes d'application	P. 38
Annexe II : Rétroagir sur l'indexation des ressources	P. 39
Annexe III : Maintenance du Référentiel – Fiches Indexation	P. 41
Annexe IV : Exemples de thésaurus et listes de vedettes matière	P. 49
1. AGROVOC	
2. EUROVOC	
3. MeSH bilingue	
4. MOTBIS	
5. RASUQAM	
6. TESE	
7. Thésaurus de la BDSP	
8. Thésaurus GETTY des noms géographiques	
9. Thésaurus de l'UNESCO	
10. LCSH	
11. Rameau	
12. RVM Laval	
Annexe V : Glossaires de l'indexation	P. 52
1. Glossaire des techniques de l'information (Joconde)	
2. Glossaire professionnel (Services techniques des bibliothèques de l'UQAM)	
3. Vocabulaire de la documentation (ADBS)	

ANNEXE I

PRESENTATION DU REFERENTIEL THEMATIQUE

Le référentiel thématique reflète la richesse de l'offre qu'Universcience propose à des publics très divers et en vue d'usages multiples.

Il est centré sur les sciences et les techniques mais tient compte de l'approche originale de l'établissement, qui est de **relier les sciences et les techniques aux problèmes de société et au monde des arts.**

La partie dédiée aux **métiers, à la formation et à l'emploi** a été développée en écho à l'offre de la Cité des métiers.

1 - Structure de base et données chiffrées

Le Référentiel thématique comprend à l'heure actuelle 8840 descripteurs :

- **7959** descripteurs « noms communs »
- **859** descripteurs « lieux et entités géographiques »
- **22** descripteurs « noms de collectivités »

Ces descripteurs sont répartis en 20 entrées thématiques ou domaines :

- 13 entrées sur les **sciences et les techniques** proprement dites
- 5 liées à une approche « **Sciences humaines** » des sciences et des techniques
- 1 entrée sur l'approche « **Arts et science** »
- 1 entrée consacrée aux **métiers, à la formation et à l'emploi**

L'arborescence principale est complétée par un ensemble de **champs sémantiques transversaux**, une liste structurée de **lieux et entités géographiques** et une liste non hiérarchisée de **mots outils**.

- **Les champs thématiques transversaux** regroupent des concepts qui débordent des domaines scientifiques et techniques ou sont trop génériques en science et en technique pour pouvoir s'insérer dans une arborescence particulière. Ils s'organisent autour de 25 petites unités,
- **L'arborescence géographique** offre une couverture mondiale et est enrichie d'entités administratives et économiques,
- **18 mots outils** apportent des précisions de point de vue.

ARBORESCENCE PRINCIPALE (niveaux 1)

1. MEDECINE – SANTE – PSYCHOLOGIE
2. SCIENCES DE LA VIE
3. SCIENCES DE LA TERRE
4. AGRICULTURE – ALIMENTATION – INDUSTRIE AGROALIMENTAIRE
5. TERRITOIRE – TRANSPORT – VEHICULES
6. GENIE CIVIL – BÂTIMENT ET TRAVAUX PUBLICS – ARCHITECTURE
7. ENERGIE – ENVIRONNEMENT – DEVELOPPEMENT DURABLE
8. MATHEMATIQUES – PHYSIQUE – CHIMIE
9. INGENIERIE – PRODUCTION – GESTION
10. ESPACE ET ASTRONOMIE
11. AUDIOVISUEL – COMMUNICATION – MEDIA
12. SCIENCES ET TECHNOLOGIES DE L'INFORMATION
13. MATERIAUX – MICROTECHNIQUES – NANOTECHNOLOGIES
14. ARTS ET SCIENCE
15. SCIENCES, TECHNIQUES ET SOCIETE
16. HISTOIRE DES SCIENCES ET DES TECHNIQUES
17. PHILOSOPHIE DES SCIENCES ET DES TECHNIQUES
18. SCIENCES DE L'EDUCATION – SYSTEME EDUCATIF
19. MUSEOLOGIE
20. METIERS – FORMATION – EMPLOI

CHAMPS THEMATIQUES TRANSVERSAUX

LIEUX ET ENTITES GEOGRAPHIQUES

MOTS OUTILS

Chaque domaine est organisé en sous domaines (niveaux 2 de la hiérarchie), eux-mêmes dotés d'une structure hiérarchique plus ou moins profonde en fonction des thématiques.

ARBORESCENCE PRINCIPALE ET LISTES ANNEXES (niveaux 1 et 2)

MEDECINE – SANTE – PSYCHOLOGIE

1. Médecine
2. Médecine alternative
3. Pharmacologie
4. Spécialité paramédicale
5. Substances médicamenteuses et drogues
6. Technologie médicale
7. Thérapeutique
8. Pathologie
9. Guérison
10. Mort
11. Santé
12. Psychologie

SCIENCES DE LA VIE

1. Anatomie
2. Biochimie
3. Bioinformatique
4. Biologie cellulaire
5. Biologie du développement
6. Biologie humaine
7. Biologie marine
8. Biométrie
9. Biophysique
10. Biotechnologie
11. Botanique
12. Chronobiologie
13. Corps humain
14. Ecologie
15. Evolution biologique
16. Génétique
17. Histologie
18. Microbiologie
19. Mycologie
20. Physiologie
21. Sciences du comportement
22. Systématique
23. Zoologie

SCIENCES DE LA TERRE

1. Cartographie
2. Climatologie
3. Géochimie
4. Géologie
5. Géologie minière
6. Géophysique
7. Géotechnique
8. Hydrologie continentale
9. Météorologie
10. Milieu naturel
11. Océanographie
12. Risque naturel
13. Topographie

AGRICULTURE – ALIMENTATION – INDUSTRIE AGROALIMENTAIRE

1. Agriculture
2. Alimentation
3. Industrie agroalimentaire

TERRITOIRE – TRANSPORT – VEHICULES

1. Territoire
2. Transport
3. Orientation et navigation
4. Véhicule
5. Construction aéronautique
6. Construction automobile
7. Construction ferroviaire
8. Construction navale

GENIE CIVIL – BÂTIMENT ET TRAVAUX PUBLICS – ARCHITECTURE

1. Génie civil
2. Architecture

ENERGIE – ENVIRONNEMENT – DEVELOPPEMENT DURABLE

1. Energie
2. Ecologie humaine
3. Dégradation de l'environnement
4. Développement durable

MATHEMATIQUES – PHYSIQUE – CHIMIE

1. Mathématiques
2. Physique
3. Matière
4. Chimie

INGENIERIE – PRODUCTION – GESTION

1. Ingénierie industrielle
2. Informatique industrielle
3. Entreprise
4. Normalisation

ESPACE ET ASTRONOMIE

1. Astronautique
2. Astronomie
3. Exobiologie
4. Sciences spatiales

AUDIOVISUEL – COMMUNICATION – MEDIA

1. Audiovisuel
2. Communication
3. Industrie graphique
4. Média

SCIENCES ET TECHNOLOGIE DE L'INFORMATION

1. Sciences de l'information
2. Electronique
3. Informatique
4. Réseau
5. Technique d'imagerie
6. Télécommunication
7. Traitement des données

MATERIAUX – MICROTECHNIQUES – NANOTECHNOLOGIES

1. Matériau
2. Calcul des structures
3. Science des matériaux
4. Textile
5. Microtechnique
6. Nanotechnologie

ARTS ET SCIENCE

1. Médecine et art
2. Médecine et cinéma
3. Médecine et littérature
4. Science et art
5. Science et littérature
6. Science et musique
7. Technique et art

SCIENCES, TECHNIQUES ET SOCIETE

1. Culture scientifique et technique
2. Science et société
3. Sociologie des sciences
4. Sociologie des techniques

HISTOIRE DES SCIENCES ET DES TECHNIQUES

1. Histoire de l'anthropologie
2. Histoire de la biologie et de la médecine
3. Histoire de la psychologie
4. Histoire des relations entre l'homme et l'environnement
5. Histoire des sciences
6. Histoire des sciences cognitives
7. Histoire des techniques
8. Histoire naturelle
9. Méthodologie scientifique - Histoire

PHILOSOPHIE DES SCIENCES ET DES TECHNIQUES

1. Philosophie de la médecine
2. Philosophie des sciences
3. Philosophie des techniques

SCIENCES DE L'EDUCATION – SYSTEME EDUCATIF

1. Sciences de l'éducation
2. Système éducatif
3. Analphabétisme
4. Illettrisme

MUSEOLOGIE

1. Documentation muséale
2. Exposition
3. Formation en muséologie
4. Musée

METIERS – FORMATION – EMPLOI

1. Métier
2. Orientation scolaire et professionnelle
3. Formation professionnelle
4. Recherche d'emploi
5. Travail et emploi

CHAMPS THEMATIQUES TRANSVERSAUX

1. Activité de loisirs
2. Anthropologie
3. Archéologie
4. Art
5. Art et science militaires
6. Cadre institutionnel
7. Culture
8. Design
9. Droit
10. Economie
11. Géographie
12. Gestion du risque – Accidents – Catastrophes – Produits dangereux
13. Histoire
14. Linguistique
15. Méthodologie
16. Organisations
17. Philosophie
18. Politique
19. Préhistoire
20. Recherche
21. Science
22. Sciences cognitives
23. Sciences sociales
24. Sémiologie
25. Technique

LIEUX ET ENTITES GEOGRAPHIQUES

1. Afrique
2. Amérique
3. Antarctique
4. Arctique (Région)
5. Asie
6. Europe
7. Océanie
8. Mers et océans
9. Commonwealth
10. Méditerranée (Région)
11. Pays arabes
12. Pays de l'OCDE
13. Pays développés
14. Pays émergents
15. Pays en développement

16. Pays francophones
17. Pays les moins avancés
18. Pays socialistes

MOTS OUTILS

- | | |
|--------------------------|--------------------------|
| - Aspect économique | - Aspect moral |
| - Aspect écologique | - Aspect philosophique |
| - Aspect éducatif | - Aspect politique |
| - Aspect environnemental | - Aspect psychanalytique |
| - Aspect éthique | - Aspect psychologique |
| - Aspect ethnologique | - Aspect religieux |
| - Aspect génétique | - Aspect sanitaire |
| - Aspect juridique | - Aspect social |
| - Aspect médical | - Aspect symbolique |

2 – Choix et forme du vocabulaire

Les **termes de l'arborescence** ont été pour l'essentiel sélectionnés à partir des ressources de la BSI (données bibliographiques, vedettes matière préexistantes) et de la Cité des sciences (exploration des différents portails du site, lexique, liens vers d'autres institutions).

Le corpus s'est enrichi de concepts émanant de champs d'étude récents ou en émergence, avec prise en compte de l'indice de popularité des termes dans les communautés scientifiques concernées et du nombre d'occurrences sur Internet.

Le vocabulaire se compose d'unitermes (unités sémantiques les plus courtes possible) **et de locutions précoordonnées.**

La majorité des locutions précoordonnées correspondent à des expressions utilisées dans le langage courant : CULTURE IN VITRO, AMENAGEMENT DU TERRITOIRE. D'autres ont été construites pour contourner la polysémie : SOURIS (ANIMAL), SOURIS (INFORMATIQUE) ou pour caractériser un objet : INTESTIN (PATHOLOGIE). Certaines enfin l'ont été pour rendre compte des rapports qu'entretiennent deux champs d'étude : ENVIRONNEMENT ET SANTE, EDUCATION ET MUSEE, SCIENCE ET ART. Leur validité est parfois à remettre en cause et fera l'objet d'un examen ultérieur. (Voir aussi chapitre 7)

En règle générale, l'emploi du singulier a été privilégié, à quelques exceptions près, lorsque les termes ne pouvaient pas être exprimés au singulier (RESEAUX DE PETRI, MŒURS ET COUTUMES, SCIENCES DE LA VIE, METAUX LOURDS, etc.).

3 - Relations sémantiques

Ces relations situent chaque descripteur dans son environnement sémantique. L'arborescence a été construite selon les trois types de relations préconisées par les normes :

- **Relations de hiérarchie** (TG : terme générique et TS : terme spécifique)

Des termes liés à un même sujet, à condition qu'ils appartiennent à la même catégorie fondamentale* (un objet, une action ou un processus, une propriété, etc.), sont organisés en hiérarchies.

La relation hiérarchique comprend la relation générique, la relation hiérarchique tout/partie, et la relation par l'exemple (1). Cette organisation permet d'indexer un contenu à un niveau générique, ou, si le sujet du contenu est plus précis, à un niveau spécifique.

- **Relations d'équivalence** (employé pour : EP)

Au sein d'une liste de termes dont la signification est identique ou très proche - synonymes ou quasi-synonymes - ou en cas de variante lexicale, un terme a été retenu (descripteur), les autres rejetés (non descripteurs). Le descripteur va représenter le concept lors de l'indexation, tandis que le/ les non descripteurs constitueront des termes d'accès qui orienteront l'utilisateur, des termes non sélectionnés vers ceux qui l'ont été (1).

- **Relations d'association** (voir aussi : VA)

Les termes sont étroitement associés mais ils ne sont pas en relation hiérarchique. On peut dire que l'utilisation de l'un suggère fortement l'utilisation de l'autre (il convient d'ajouter qu'une relation associative peut être pertinente pour une communauté particulière d'utilisateurs, dans un domaine spécialisé, et n'avoir aucun sens pour un public différent (1)).

* Nous avons parfois pris la liberté de contourner cette norme, en inscrivant un descripteur spécifique « objet » sous un descripteur générique « processus », les usagers faisant un lien implicite entre le processus et l'objet qui donne réalité à ce processus.

Ex. : TG IMPRIMERIE (processus)
TS ENCRE D'IMPRIMERIE (objet)
TS PRESSE A IMPRIMER (objet)
TS REPROGRAPHIE (processus)
TS TECHNIQUE D'IMPRESSION (processus)

(1) Construire un thésaurus / Jean Aitchison et Alain Gilchrist. ADBS, 1992

4 - Elaboration des hiérarchies

Le travail d'élaboration des hiérarchies s'est toujours appuyé sur la consultation d'autres sources, contrôlées et normalisées, comme les thésaurus, les listes de vedettes matière, les dictionnaires et les encyclopédies, sur support imprimé ou sur le Web.

Lorsqu'un concept très utilisé dans une discipline particulière ou un champ d'application spécifique s'avérait plus générique que ne le laissait supposer son utilisation, le parti a été pris de le « décontextualiser », c'est-à-dire de l'extraire de cette discipline ou de ce champ d'application, et de le classer à son niveau le plus générique en valorisant sa transversalité.

Lorsqu'on évoque l'ÉCHOGRAPHIE ou la RADIOGRAPHIE, on pense immédiatement aux applications médicales alors qu'il y a de plus en plus d'applications industrielles de ces techniques, dans le contrôle des matériaux entre autres. Ce descripteur a donc été classé sous TECHNIQUE D'IMAGERIE

Quelques termes et locutions, non retenus en tant que descripteurs, ont été parfois nécessaires à la cohérence hiérarchique. Ils ont été nommés « Relais virtuels » et ne doivent pas être employés pour indexer des contenus (ils sont au nombre de 27).

5 - Profondeur des hiérarchies

Elle varie en fonction des thématiques traitées :

- **92% des descripteurs s'organisent entre les niveaux 3 et 6**
- 2,5% d'entre eux sont compris entre le niveau 1 et le niveau 2
- 5,5% des descripteurs restants, provenant pour l'essentiel des hiérarchies Zoologie et Botanique systématiques, sont répartis entre les niveaux 7, 8 et 9 (57 descripteurs au niveau 9)

6 - Polyhiérarchie

L'usage de la polyhiérarchie, logique selon laquelle un descripteur peut être inscrit dans plusieurs arborescences, a été limité, les relations sémantiques devenant complexes et la navigation moins aisée. Dans certains cas toutefois (environ 150 descripteurs sont concernés), il a été nécessaire de représenter un concept dans plusieurs de ses dimensions et de faire appel à la polygénéricité (plusieurs TG).

L'ARDOISE, l'AMIANTE et le MARBRE sont des roches métamorphiques mais également des matériaux.

7 - Arborescence «HISTOIRE DES SCIENCES ET DES TECHNIQUES»

La richesse des collections imprimées de la bibliothèque d'Histoire des sciences, la valorisation d'une partie de ses fonds sur le site (Scientifica), l'existence du Centre de recherche en Histoire des sciences et des techniques (CRHST) et les nombreux partenariats noués avec d'autres institutions ont rendu incontournable l'élaboration d'arborescences dédiées à l'histoire des sciences et des techniques.

L'hypothèse d'adopter comme unique logique de requête la post-coordination (un descripteur x coordonné avec le descripteur HISTOIRE) n'a jamais été envisagée, la possibilité pour les usagers de naviguer au sein d'arborescences établies toujours privilégiée.

Cette arborescence n'est pas le double « historique » de l'arborescence générale.

- du fait de l'évolution des sciences et des techniques, l'organisation des disciplines n'était pas la même qu'aujourd'hui
- certains termes utilisés dans le passé sont devenus obsolètes aujourd'hui
- des termes actuels étaient absents autrefois (ou avaient un autre sens)
- il n'y a pas eu lieu de créer l'équivalent « historique » d'une bonne partie des descripteurs de l'arborescence générale, les ressources n'existant pas

Quant aux deux modes de précoordination utilisés dans la construction des descripteurs (par exemple : HISTOIRE DE L'ASTRONOMIE et OBSERVATOIRE ASTRONOMIQUE + HISTOIRE), la raison en est que :

- lorsqu'il existait un réel champ d'étude sur l'histoire d'une discipline ou d'une technique, l'expression « HISTOIRE DE » a été privilégiée
- lorsqu'il s'agissait de descripteurs spécifiques appartenant à ces champs d'étude, l'expression « + HISTOIRE » a été préférée

Ces choix peuvent paraître discutables, la forme des descripteurs est susceptible d'évoluer, le mode de précoordination n'ayant de toute façon aucune influence sur les requêtes faites à l'aide du moteur de recherche ou via la recherche « Mot du sujet » du SIGB Aleph.

8 - Définitions et notes d'application

Toujours à l'état de projet, le chantier n'a pas avancé.

Les définitions doivent émaner d'institutions partenaires fiables. Les notes d'application s'appuieront sur ces définitions.

ANNEXE II

RETROAGIR SUR L'INDEXATION DES RESSOURCES

Ressources de la BSI

Suite aux phases de conversion automatique réalisées par la société Ex-Libris et au nettoyage manuel de ce qui n'avait pu être traité informatiquement, **la base des autorités sujet du SIGB* Aleph (catalogue) a été mise en conformité avec le Référentiel CSI.**

Début 2006, la base des autorités sujet comprenait environ 18 000 vedettes matière dont à peine quelques centaines étaient liées par des relations de hiérarchie ou d'association. Les trois quarts d'entre elles consistaient en vedettes composées d'une ou de plusieurs subdivisions, un quart était constitué d'unitermes ou de locutions précoordonnées.

Début 2010, cette base est constituée des 8840 descripteurs du Référentiel CSI, tous liés entre eux par des relations sémantiques. Cette évolution a permis de créer un univers documentaire cohérent pour les usagers internes et externes.

Le travail consiste actuellement à **remettre à niveau l'indexation de la base bibliographique** qui rassemble **les notices** d'environ 160 000 ressources imprimées (livres, revues) et 5500 ressources numériques (livres et revues électroniques, dossiers documentaires, films de la vidéothèque, DVD, cédéroms).

Cette logique de rétroaction, en lien avec les préconisations émises par la politique documentaire de la BSI et plus largement avec les orientations politiques d'Universcience, a pour objectif de **vérifier la cohérence globale de l'indexation et d'améliorer ainsi la qualité de l'accès aux ressources.** Elle s'accompagne d'outils et d'analyses donnant de premières indications sur l'utilisation des descripteurs du Référentiel.

Les contenus en ligne de la BSI publiés dans Content Server seront indexés dès que les dysfonctionnements du système seront réglés et lorsque la dernière version du Référentiel thématique sera déployée.

* Système intégré de gestion de bibliothèque

Autres ressources d'Universcience

A l'instar de la conversion des anciennes vedettes matière de la BSI en descripteurs CSI, le **travail de confrontation** entre les termes d'indexation utilisés jusque-là par les différents services pour décrire leurs ressources et les descripteurs du Référentiel doit se poursuivre.

Les résultats obtenus permettront d'évaluer le taux de compatibilité entre les différents outils d'indexation et le Référentiel thématique, de réaliser éventuellement une mise en concordance automatique des termes (« mapping » terminologique) et/ou d'envisager une indexation semi-automatique*. Cela permettra d'obtenir in fine une meilleure cohérence de l'indexation des ressources d'Universcience.

* Pour information : Indexation semi-automatique de textes <http://asso-aria.org/coria/2009/151.pdf>

ANNEXE III

FICHES DE MAINTENANCE DU REFERENTIEL

Toute suggestion de **création**, **modification** ou **suppression** d'un descripteur est à transcrire par le biais de fiches Indexation.

Trois fiches correspondent à ces cas de figure :

- Fiche N° 1 pour une proposition de création
- Fiche N° 2 pour une proposition de modification
- Fiche N° 3 pour une proposition de suppression

Les fiches sont à envoyer par Groupwise à indexation@universcience.fr

FICHE N° 1

CREATION D'UN NOUVEAU DESCRIPTEUR

Chaque descripteur doit être intégré dans une hiérarchie, de préférence déjà existante.

Merci de préciser le supérieur hiérarchique immédiat (**TG** : terme générique).

Si ce nouveau descripteur s'accompagne d'un ou plusieurs descripteurs spécifiques (**TS** : terme spécifique), une fiche est à remplir pour chacun d'eux.

Une relation d'association peut également être proposée (**VA** : terme associé).

En cas de termes sémantiquement très proches (synonymie ou quasi-synonymie), le nouveau descripteur peut être enrichi d'un ou plusieurs termes rejetés (**EP** : employé pour).

Les fiches sont à envoyer par Groupwise à indexation@universcience.fr

- **DESCRIPTEUR**

- **TERME GENERIQUE (TG)**

- Descripteur déjà existant
- Création (remplir une fiche)

- **EMPLOYE/S POUR (EP)**

- **TERME/s SPECIFIQUE/s (TS)**

-
- **TERME/s ASSOCIE/s (VA)**

-
- **Commentaires en texte libre** (sources utilisées, note de définition ou d'application, ressources concernées)

-
- **NOM, Prénom**

- **SERVICE**

- **ADRESSE ELECTRONIQUE**

- **TELEPHONE**

- **DATE DE LA SUGGESTION**

FICHE N° 2

MODIFICATION D'UN DESCRIPTEUR EXISTANT

La modification du descripteur concerne :

son **intitulé** (sans qu'il y ait obligatoirement changement de sa place dans la structure hiérarchique)

son/ses **EP** (Employé pour : terme exclu)

son/ses **VA** (Voir aussi : terme associé)

ses **relations hiérarchiques** (terme générique, terme/s spécifique/s)

Les fiches sont à envoyer par Groupwise à indexation@universcience.fr

DESCRIPTEUR

Nouvel intitulé proposé :

-
- **EMPLOYE/s POUR (EP)** (Ne pas remplir de fiche)

Création

Suppression

Création

Suppression

- **TERME/s ASSOCIE/s (VA)**

O Ajout d'une relation d'association

Descripteur/s déjà existant/s (Pas de fiche)

Création (Remplir une fiche)

O Suppression de la relation d'association (Pas de fiche)

-
- **Nouveau TERME GENERIQUE (TG) proposé :**

Descripteur déjà existant

Création (Remplir une fiche)

-
- **Nouveau/x TERME/s SPECIFIQUE/s (TS) proposé/s :**

Descripteur/s déjà existant/s

(Remplir une fiche de modification par descripteur, avec mention de la nouvelle relation sémantique)

Création (Remplir une fiche par nouveau descripteur)

- **Commentaires en texte libre** (justification du choix effectué, précisions à apporter ...)

-
- **NOM, Prénom**

- **SERVICE**

- **ADRESSE ELECTRONIQUE**

- **TELEPHONE**

- **DATE DE LA SUGGESTION**

FICHE N° 3

SUPPRESSION D'UN DESCRIPTEUR EXISTANT

La suppression du descripteur :

- entraîne la création d'un nouveau descripteur
- n'entraîne pas la création d'un nouveau descripteur
- entraîne la suppression du/des terme/s spécifique/s
- entraîne le changement de place du/des terme/s spécifique/s
- entraîne sa suppression en tant que terme associé

Merci de remplir les fiches correspondantes en cas de nouvelle création, autres suppressions ou modifications (Fiches N° 1 Création, N°2 Modification ou N°3 Suppression).

Les fiches sont à envoyer par Groupwise à indexation@universcience.fr

- **DESCRIPTEUR**

qui pour **Relation associative (VA)** :

Cette relation d'association sera supprimée

- **Ressources concernées**

Aucune ressource n'est liée à ce descripteur et on pense qu'il n'y en aura jamais assez

Ces ressources ont une indexation déjà suffisante

Ces ressources doivent être indexées autrement

Titres ou intitulés des ressources :

- **NOM, Prénom**

- **SERVICE**

- **ADRESSE ELECTRONIQUE**

- **TELEPHONE**

- **DATE DE LA SUGGESTION**

ANNEXE IV

EXEMPLES DE THESAURUS ET LISTES DE VEDETTES MATIERE

THESAURUS :

- **AGROVOC** (version française)
 - Thésaurus agricole multilingue de la FAO (Organisation des Nations unies pour l'alimentation et l'agriculture)
 - 28 474 descripteurs (Source : FAO)
 - <http://aims.fao.org/standards/agrovoc/about>

- **EUROVOC** (version française)
 - Thésaurus multilingue couvrant tous les domaines de l'activité de l'Union européenne
 - 6645 descripteurs, répartis en 21 domaines et 127 microthésaurus (Source : Eurovoc)
 - http://europa.eu/eurovoc/sg/sga_doc/eurovoc_dif!SERVEUR/menu!prod!MENU?langue=FR

- **MeSH** (Medical Subject Headings) (version bilingue)
 - Thésaurus de référence dans le domaine biomédical
 - 25 588 descripteurs, répartis en 16 catégories et organisés jusqu'à 11 niveaux de hiérarchie (Source : Mesh bilingue)
 - <http://mesh.inserm.fr/mesh/>

- **MOTBIS**
 - Thésaurus de l'Education nationale
 - 10 160 descripteurs, organisés en 69 microthésaurus et 10 listes annexes (Source : Motbis)
 - <http://www.cndp.fr/thesaurus-motbis/site/>

- **RASUQAM**
 - Répertoire des Autorités sujet de l'UQAM (Université du Québec à Montréal)
 - 40 000 descripteurs, répartis en noms communs, noms d'organismes, titres d'œuvres, noms géographiques, chronologies et descripteurs de forme (source UQAM)
 - Pas d'information sur l'organisation des domaines et sous domaines
 - <http://www.bibliotheques.uqam.ca/aide-la-recherche/rasuqam>
 - Cliquer sur « Thésaurus RASUQAM »

- **TESE (version française)**
 - Thésaurus européen multilingue des systèmes éducatifs
 - 1387 descripteurs, organisés en 6 facettes, 17 microthésaurus et 4 listes additionnelles non hiérarchisées (Source : TESE)
 - http://eacea.ec.europa.eu/education/eurydice/tese_fr.php
 - Télécharger le document par type de présentation (systématique, permutée ou alphabétique) après avoir sélectionné la langue (fr)

- **Thésaurus de la BDSP**
 - Thésaurus de la Santé publique
 - 7144 descripteurs, répartis en 57 sous-domaines (Source : BDSP)
 - <http://www.bdsp.ehesp.fr/Thesaurus/Default.asp>

- **Thésaurus GETTY des noms géographiques**
 - Vocabulaire hiérarchisé d'environ 1,3 million de noms, coordonnées et autres informations pour quelque 911 000 lieux et régions géographiques (Source : Getty)
 - <http://www.getty.edu/vow/TGNHierarchy?find=&place=&nation=&english=Y&subjectid=7029392>

- **Thésaurus de l'UNESCO (version française)**
 - Thésaurus multilingue dont la terminologie reflète les activités de l'Organisation : éducation, culture, sciences naturelles, sociales et humaines, communication et information
 - 8600 descripteurs, répartis en 7 domaines et 88 microthésaurus (Source : Unesco)
 - <http://databases.unesco.org/thesfr/>

LISTES DE VEDETTES MATIERE :

- **LCSH** (Library of Congress Subject Headings)
- 265 000 “subject authority records” (Source : Library of Congress)
<http://authorities.loc.gov/>
→ Subject Authority Headings

- **RAMEAU** (Répertoire d'autorité-matière encyclopédique et alphabétique unifié)
- 160 703 notices autorités sujet, réparties en 97 878 noms communs, 53 584 noms géographiques, 122 subdivisions chronologiques et 9119 noms propres (chiffres 2009, Source : Rameau)
<http://catalogue.bnf.fr>
→ Autorités Rameau

- **RVM Laval** (Répertoire des vedettes matière de l'Université de Laval)
- 203 900 vedettes matière, réparties en 157 200 noms communs et 46 700 noms géographiques ou toponymes (chiffres 2005, Source : IFLA)
<http://rvmweb.bibl.ulaval.ca>
→ *Abonnement exigé pour la consultation du RVM*

ANNEXE V

GLOSSAIRES DE LA DOCUMENTATION

- Glossaire des techniques de l'information (Joconde : catalogue des collections des musées de France)
<http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/glossaire1.htm>
- Vocabulaire de la documentation (ADBS)
<http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm>
- Université de Québec à Montréal (UQAM) : [Glossaire](#)
http://www.bibliotheques.uqam.ca/InfoSphere/fichiers_communs/site/glossaire.html