

La représentation des connaissances en science de l'information-documentation

El-Khansa MKADA-ZGHIDI

Université de La Manouba - Institut supérieur de documentation

mkadakhansa@gmail.com

Résumé : La représentation des connaissances en sciences de l'information-documentation peut être considérée comme une assise à la fois théorique et pratique de cette science. Les systèmes de représentation sont évolutifs, ils sont étudiés en mettant en valeur leur structuration et leur contexte d'apparition notamment à travers les facteurs cognitifs, idéologiques et technologiques qui ont marqué cette évolution.

L'apport de l'interdisciplinarité dans la conception et l'élaboration des systèmes de représentation est également souligné. Cet aspect constitue le thème de la discussion finale qui touche trois axes, à savoir la finalité des systèmes de représentation, les pratiques professionnelles et l'épistémologie de la science de l'information.

Mots clés : Classifications // Ontologie // Représentation des connaissances // Science de l'information-documentation // Thésaurus

المستخلص : يمكن اعتبار تمثيل المعرفة في مجال علم المعلومات والتوثيق كأساس نظري وممارسة في نفس الوقت. فنظم التمثيل قابلة للتطور وقد تمت دراستها خاصة من خلال تسليط الضوء على بنيتها وسياق ظهورها وكذلك العوامل المعرفية والأيدولوجية والتكنولوجية التي تميز هذا التطور..

ولا بد من التأكيد أيضا على مساهمة تعددية الاختصاصات في مجال تصميم وإعداد نظم التمثيل. فهذا الجانب هو موضوع نقاش من ثلاثة محاور رئيسية هي الغرض من نظم التمثيل، والممارسات المهنية ونظرية المعرفة في علم المعلومات..

الكلمات المفتاح : نظم التصنيف // تمثيل المعرفة // الانطولوجيا // علم المعلومات والتوثيق // المكانز

Abstract : The representation of information-documentation of science knowledge can be considered as a foundation for the theory and practice of this science. Representation systems are scalable; they are studied by highlighting their structure and context of appearance especially through cognitive, ideological and technological factors that have marked this evolution.

The contribution of interdisciplinarity in the design and development of systems of representation is also stressed. This aspect is the subject of the final discussion through three key areas, namely the purpose of representation systems, professional practices and the epistemology of information science.

Keywords : Classifications // Ontology // Knowledge representation // Information Science and documentation // Thesaurus

Introduction

La représentation des connaissances en science de l'information-documentation est contemporaine des premières bibliothèques. Elle se fonde sur l'analyse du contenu informationnel des documents. Ces contenus sont indexés soit dans les termes du langage naturel en fonction d'un ordre normalisé, soit dans ceux des langages documentaires. Les langages documentaires fournissent un état de la connaissance à un moment donné à travers des notions plus ou moins complexes et différemment structurées selon les systèmes de représentation. Utilisés comme outils de représentation des contenus informationnels, ils permettent l'accès des usagers aux fonds documentaires mis à leur disposition. Les systèmes de représentation des connaissances n'ont cessé d'évoluer au rythme des avancées de la connaissance elle-même et du progrès technologique.

Cet article étudie le contexte d'apparition, la structuration et les objectifs des quatre grandes étapes caractéristiques de l'évolution de la représentation des connaissances en science de l'information-documentation. L'étape des systèmes de représentation des connaissances contenues dans les documents consacrés aux sciences religieuses et profanes ; celle des systèmes de représentation classificatoires, encyclopédiques et synthétiques basés sur la notation des sujets permettant l'accès au document et à l'information de manière simultanée; celle des systèmes combinatoires, spécialisés et analytiques, privilégiant l'accès à l'information indépendamment de son support, basés sur un réseau de relations entre des descripteurs univoques se prêtant à la post-coordination au moment de la recherche d'informations ; et enfin l'étape

actuelle marquée par les technologies de l'information et les progrès de l'intelligence artificielle, consacrée aux connaissances véhiculées par les ressources électroniques et le web sémantique, et où la représentation des domaines de la connaissance tente de refléter le raisonnement humain à travers les propriétés des concepts et la diversité de leurs relations au moyen d'ontologies construites à l'aide de logiciels dédiés.

1. Problématique et méthodologie

1.1. Problématique

La problématique de la représentation des connaissances en science de l'information-documentation se décline en plusieurs axes. Le premier peut être qualifié de technique, il est relatif à la finalité de la représentation des connaissances dans cette science consacrée à la collecte, au traitement et à la diffusion de l'information documentaire. Comment les systèmes de représentation sont-ils élaborés ? Quel est leur impact sur les contenus informationnels recelés dans un document intégral ou un texte plein lisible sur ordinateur ? Comment donc les divers systèmes de représentation ont-ils évolués, et dans quelle mesure cette évolution a-t-elle suivi celle des supports informationnels ?

Le deuxième axe peut être considéré comme déontologique et porte sur la relation avec l'utilisateur auquel est destiné cette représentation. Ce dernier est en effet confronté à une énorme masse d'informations parmi lesquelles il doit parvenir à celles pouvant satisfaire le mieux ses besoins informationnels. Dans quelle mesure les systèmes de représentation conçus et élaborés par le professionnel de l'information sont-ils des adjuvants pour un accès rapide à l'information pertinente par rapport aux besoins divers d'utilisateurs très variés ?

Enfin le dernier axe de la problématique est épistémologique. Dans ce contexte, il semble légitime de s'interroger sur l'identification des sciences connexes dans la construction des systèmes de représentation des connaissances en science de l'information-documentation. Quel est l'apport de ces sciences connexes et dans quelle mesure contribuent-elles à donner du sens à la représentation des connaissances en elle-même ?

1.2. Méthodologie

Pour traiter les trois axes principaux de la problématique, nous commençons par définir les notions clés de représentation des connaissances et de science de l'information-documentation. Puis nous décrivons l'évolution des systèmes de représentation des connaissances

en essayant d'analyser le contexte qui a favorisé cette évolution. Ce contexte aborde principalement l'avancement des connaissances, la conjoncture politique et économique et le progrès technologique. Ce faisant, nous tentons de cibler les sciences connexes qui ont influencé les diverses mutations des systèmes de représentation.

Chaque système de représentation étudié donne lieu à une synthèse présentant ses principales caractéristiques. A la fin de l'article, les diverses synthèses sont incluses dans une discussion autour des trois axes de la problématique, technique, déontologique et épistémologique, tout en précisant, d'ores et déjà, que notre prétention n'est pas d'apporter des réponses catégoriques mais de contribuer au débat actuel autour des assises théoriques de la science de l'information-documentation.

2. Définitions et objet de la science de l'information

2.1. Définitions

La représentation des connaissances a pour objectif d'organiser les objets du monde. Son objectif est double, d'une part la connaissance de la connaissance en elle-même, sa genèse, ses différents domaines et leurs interrelations et d'autre part sa mise à la disposition de la recherche scientifique afin de construire le savoir d'une société à un moment donné et contribuer à son évolution.

L'information se fonde essentiellement sur la représentation des connaissances. Elle s'est constituée en tant que science au début des années 1960. Elle a été définie comme suit par le psychologue, documentaliste et informaticien américain Harold Boroko lors d'une conférence organisée par l'Institut technologique de Georgie en 1961 : « *la science de l'information est une science qui explore les propriétés et le comportement de l'information, et les moyens de traiter l'information pour une accessibilité et une ergonomie optimales. Ses modalités incluent sa genèse, sa dissémination, son organisation, sa collecte, son archivage, sa recherche, son interprétation et sa réutilisation.* »¹

¹ L'auteur approfondi cette définition dans "Information science: what is it? " American Documentation, 1968. « *Information science is a discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information. This includes the*

Les définitions actuelles focalisent sur l'objet d'étude, « *la science de l'information est l'étude des propriétés générales de l'information (nature, genèse, effets) et l'analyse de ses processus de construction, de communication, d'usage.* »²

2.2. Donnée, information, connaissance et savoir

D'une manière générale, la science de l'information appréhende les notions de donnée, d'information, de connaissance et de savoir comme suit :

- La donnée est la plus petite unité informationnelle, elle est encodée et inscrite sur un support à des fins de stockage et de communication. Les données peuvent être sélectionnées par un sujet ou demeurer à l'état de données ;
- La donnée se transforme en information lorsqu'elle est sélectionnée par un être humain qui lui donne du sens. Cette information est utilisée en fonction de sa pertinence par rapport à un besoin informationnel ou à un projet de recherche ;
- L'information devient connaissance lorsqu'elle est mobilisée par un être humain et lorsqu'elle modifie l'état de ses connaissances antérieures ;
- Le savoir est l'ensemble des connaissances d'une personne ou d'une collectivité, il est reproductible, mesurable et il a une certaine stabilité à un moment donné.

investigation of information representations in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields. It has both a pure science component, which inquires into the subject without regard to its application, and an applied science component, which develops services and products.»

² Yves-François Le Coadic, La science de l'information, PUF, Que sais-je ?, 1994.

2.3. Document et documentation

Un document est tout objet porteur d'informations, et qui est susceptible d'apporter de la connaissance. C'est un support recelant un contenu informationnel sous forme de signes graphiques, iconiques, audio, vidéo, etc. Il ne peut atteindre son objectif que s'il est connu, diffusé ou médiatisé.

L'origine du terme documentation remonte aux environs de 1870 avec le sens de recherche de documents pour répondre à un besoin précis. Ce terme désigne actuellement l'« *ensemble des méthodes et des techniques de traitement systématique de documents ou d'informations, quel que soit leur support, mises en œuvre pour répondre aux besoins des usagers et incluant l'acquisition, le signalement, l'analyse documentaire, l'indexation, le stockage, la recherche, la diffusion de ces documents ou informations.* »³ Par extension, la documentation désigne l'ensemble des documents sur un sujet donné.

2.4. L'objet d'étude de la science de l'information-documentation

L'objet d'étude de la science de l'information-documentation est l'information en tant que « *connaissance inscrite (enregistrée) sous forme écrite (imprimée ou numérisée), orale ou audiovisuelle* »⁴. La consignation de l'information (enregistrement) la distingue nettement de la communication, d'où l'importance de la notion de document ou du caractère documentaire de cette information qui justifie l'appellation que nous adoptons pour cette science.

La représentation des connaissances est l'un des piliers de la science de l'information-documentation. Ses principes, ses objectifs, ses procédés et ses outils peuvent être présentés dans la diachronie suivante où nous étudions les systèmes les plus marquants et les plus répandus.

3. Le système de représentation des connaissances classificatoire

La représentation classificatoire des connaissances a d'abord été mise en œuvre dans les systèmes de classification bibliologiques qui sont

³ Association des bibliothécaires et documentalistes spécialisés. Vocabulaire de la documentation. 2004.

⁴ Le Coadic Yves, La Science de l'information, Puf, (Que sais-je n°2873), 1994

l'un des types du langage documentaire. Ils ont le double objectif de répartir les connaissances selon un ordre bien déterminé et d'établir le classement des documents sur les rayonnages d'une bibliothèque en fonction de cet ordre. Ces systèmes sont encyclopédiques ; mono-hiérarchiques ou poly-hiérarchiques ; synthétiques, analytico-synthétiques ou analytiques.

3.1. Les systèmes de classification mono-hiérarchiques et synthétiques

Les systèmes de classification mono-hiérarchiques et synthétiques sont contemporains des bibliothèques. La première classification connue est celle de la bibliothèque d'Alexandrie (260-240 av. J. C.). Elle permettait l'accès aux documents par catégories d'auteurs comme suit :

Poètes	Historiens
Législateurs	Orateurs
Philosophes	Mélanges

A l'époque où l'enseignement était assuré par les hommes de religion, les bibliothèques théologiques ont adopté des classifications reflétant la confession du lieu. Par exemple, le catalogue de la bibliothèque "El Ebdellia" (de la Mosquée Zitouna) fondée en 1419 dont le fonds était estimé à 30.000 volumes avant l'invasion par les espagnols en 1535, comprend les classes suivantes :

Science du hadith	Guide des prophètes
Genèse du hadith	Vie du prophète Mohamed
Terminologie du Hadith	

En France, certaines bibliothèques monastiques du moyen-âge distinguaient entre œuvres sacrées et œuvres profanes comme suit :

Œuvres sacrées	Œuvres profanes
Écritures	Grammaire
Bible	Rhétorique
Ouvrages ecclésiastiques (Pères)	Logique
Homilétique	Arithmétique
	Géométrie
	Musique
	Astronomie

Le développement de ces systèmes de classification fut prolifique et varié jusqu'en **1876** où un jeune mathématicien âgé de 21 ans, Melvil Dewey, recruté comme assistant-étudiant dans un collège de New-York publia la classification qui porte son nom, et qui est aujourd'hui la plus utilisée dans le monde, traduite en arabe, elle est aussi la plus utilisée dans ces pays.

L'objectif de la Classification Décimale de Dewey (CDD) demeure identique à tous les systèmes de classification bibliologiques (accès au document et en même temps accès à l'information), mais elle se distingue par :

- une nette influence des systèmes philosophiques, notamment la classification de F. Bacon⁵ en ordre inversé (Philosophie et Sciences / Arts et Littérature / Histoire) ;
- la représentation de tous les domaines de la connaissance qui reflète l'état du savoir à cette époque ;
- la répartition des connaissances humaines en 10 grandes classes, 100 divisions, 1000 sections et une multitude de sous-sections.
- et surtout, du fait que les classes se subdivisent de 10 en 10, par la numérotation des sujets dans une structure arborescente décimale. Cette numérotation est utilisée comme adresse physique ou cote de classement des documents sur les rayonnages d'une bibliothèque et permet de les repérer rapidement. De plus, la numérotation extensible à l'infini offre des capacités d'aménagement du système de classification en fonction du progrès scientifique et de l'évolution du savoir.

⁵ Classification tripartite de Platon, classification « idéaliste » d'Aristote, classification « matérialiste » de Francis Bacon, jusqu'à l'ordre de filiation des sciences d'Auguste Comte et l'ébauche d'Engels, en passant par les classifications de musulmans comme El Kindi et El Farabi, El Ghazeli, Ibn Khaldoun. Pour en savoir plus sur ces dernières, voir Al-Rabe, Ahmad Abdulla. *Muslim Philosopher's Classifications of the Sciences : Al-Kindi, Al-Farabi, Al-Ghazali, Ibn Khaldun*. Ann Arbor, Michigan: UMI, Dissertation Information Service, 1987. et le Groupe Cyclope, *Bibliographie sélective : encyclopédisme et organisation du savoir en Islam*, disponible sur <http://cyclopes.fltr.ucl.ac.be/materiaux/biblioarabe.htm> 22 décembre 2011.

<p>000 Généralités</p> <p>010 Bibliographie</p> <p>020 Bibliothéconomie et sciences de l'information</p> <p>030 Encyclopédies générales</p> <p>040 <i>Non attribué</i></p> <p>050 Publications en série d'ordre général</p> <p>060 Organisations générales et muséologie</p> <p>070 Médias d'information, journalisme, édition</p> <p>080 Recueils généraux</p> <p>090 Manuscrits et livres rares</p> <p>100 Philosophie et psychologie</p> <p>110 Métaphysique</p> <p>120 Théorie de la connaissance, causalité, genre humain</p> <p>130 Phénomènes paranormaux</p> <p>140 Écoles philosophiques particulières</p> <p>150 Psychologie</p> <p>160 Logique</p> <p>170 Morale</p> <p>180 Philosophie ancienne, médiévale, orientale</p> <p>190 Philosophie occidentale moderne</p> <p>200 Religion</p> <p>210 Philosophie et théorie de la religion</p>	<p>500 Sciences naturelles et mathématiques</p> <p>510 Mathématiques</p> <p>520 Astronomie et sciences connexes</p> <p>530 Physique</p> <p>540 Chimie et sciences connexes</p> <p>550 Sciences de la terre</p> <p>560 Paléontologie Paléozoologie</p> <p>570 Sciences de la vie Biologie</p> <p>580 Plantes</p> <p>590 Animaux</p> <p>600 Technologie (Sciences appliquées)</p> <p>620 Ingénierie et techniques connexes</p> <p>630 Agriculture</p> <p>640 Économie domestique et vie familiale</p> <p>650 Gestion et services auxiliaires</p> <p>660 Génie chimique</p> <p>670 Fabrication industrielle</p> <p>680 Fabrication de produits à usages particuliers</p> <p>690 Bâtiments</p> <p>700 Arts Beaux-arts et arts décoratifs</p> <p>710 Urbanisme et art du paysage</p> <p>720 Architecture</p> <p>730 Arts plastiques Sculpture</p>
--	---

220 Bible	740 Dessin et arts décoratifs
230 Christianisme Théologie chrétienne	750 Peinture et peintures
240 Théologie morale et spirituelle chrétiennes	760 Arts graphiques Gravures
250 Églises locales et ordres religieux chrétiens	770 Photographie et photographies
260 Théologie chrétienne et société et ecclésiologie	780 Musique
270 Histoire du christianisme et de l'Église	790 Loisirs et arts du spectacle
280 Confessions et sectes chrétiennes	800 Littérature et techniques d'écriture
290 Religions comparées et autres religions	810 Littérature américaine en anglais
300 Sciences sociales	820 Littératures anglaise et du vieil anglais
310 Statistiques générales	830 Littératures des langues germaniques
320 Science politique	840 Littératures des langues romanes
330 Économie politique	850 Littératures italienne, roumaine, rhéto-romane
340 Droit	860 Littératures espagnole et portugaise
350 Administration publique et science militaire	870 Littératures des langues italiques Littérature latine
360 Problèmes et services sociaux; associations	880 Littératures helléniques Littérature grecque classique
370 Éducation	890 Littératures des autres langues
380 Commerce, communication, transports	900 Géographie et histoire
390 Coutumes, étiquette, folklore	910 Géographie et voyages
400 Langues	920 Biographies, généalogie, emblèmes, insignes
410 Linguistique	930 Histoire du monde antique jusque vers 499
420 Anglais et vieil anglais	940 Histoire générale de l'Europe

430 Langues germaniques Allemand	950 Histoire générale de l'Asie Extrême-Orient
440 Langues romanes Français	960 Histoire générale de l'Afrique
450 Italien, roumain, rhéto-roman	970 Histoire générale de l'Amérique du Nord
460 Espagnol et portugais	980 Histoire générale de l'Amérique du Sud
470 Langues italiques Latin	990 Histoire générale des autres aires géographiques
480 Langues helléniques Grec classique	
490 Autres langues	

Les 100 divisions de la CDD- 21^{ème} éd. 2005

Etant donné l'objectif pratique de la CDD, plusieurs procédés mnémotechniques sont exploités, par exemple le 0 garde sa valeur de généralité dans toutes les classes et toutes les divisions, ex. : 500 = Généralités sur les Sciences Pures, 530 = Généralités sur la physique.

Un indice a toujours au moins 3 chiffres. Un « point » sépare le 3ème du 4ème chiffre (641.5); un espace sépare le 6ème du 7ème chiffre (704. 039 7).

Le contenu d'un document est toujours représenté par un seul indice de classification quel que soient le nombre des éléments informatifs qu'il recèle. Ce procédé est contraint par le fait que le document ne peut avoir qu'une seule adresse physique dans une bibliothèque correspondant à son indice dans la classification. L'exemple suivant montre l'extension d'un indice et le caractère dit « synthétique » du système



Dewey decimal system -

http://cultureandcommunication.org/deadmedia/index.php/Dewey_Decimal_System

- A la fin des années **1880**, la bibliothèque du congrès américain commence à développer sa propre classification, la Library of Congress Classification⁶ (LCC) à partir du traitement des livres conservés dans son propre fonds et en s'inspirant de l'Expansive classification du bibliothécaire Charles Ammi Cutter⁷, inventeur des nombres Cutter pour coder l'auteur sur la cote des livres. Dans la LCC, les classes sont désignées par des majuscules.

⁶ Pour la consulter en ligne, voir <http://www.loc.gov/catdir/cpsol/lcco/>

⁷ Charles Ammi Cutter, Expansive Classification, Harvard University, 1893.

Les classes de la LCC

A : Généralités	K : Droit
B : Philosophie. Psychologie. Religion	L : Éducation
C : Sciences auxiliaires de l'histoire	M : Musique et livres sur la musique
D : Histoire du monde et histoire de l'Europe, l'Asie, l'Afrique, l'Australie, la Nouvelle- Zélande, etc.	N : Beaux arts
E : Histoire des Amériques (généralités et États-Unis)	P : Langue et littérature
F : Histoire des Amériques (autres pays d'Amérique)	Q : Sciences
G : Géographie. Anthropologie. Loisir	R : Médecine
H : Sciences sociales	S : Agriculture
J : Sciences politiques	T : Technologie
	U : Science militaire
	V : Science navale
	Z : Bibliographie. Bibliothéconomie. Ressources d'information

Les sous classes de la LCC comportent des minuscules et des chiffres pouvant se décimaliser. Les classes E et F sont consacrées à l'histoire des États-Unis et du continent américain. Les lettres I, O, W, X et Y sont réservées pour les développements futurs. Depuis 1930 l'indice CDD est ajouté aux indices de la LCC.

Exemple d'indexation :

- C : Sciences auxiliaires de l'histoire
 - CJ : Numismatique
 - CJ 1-4625 : Pièces de monnaie

- En **1904**, deux juristes belges, Paul Otlet (Président du Premier *Congrès des nationalités*, prélude à l'ouverture de la Société Des Nations) et Henri La Fontaine (Prix Nobel de la Paix en 1913), fondateurs de l'Institut International de Bibliographie (actuellement Fédération Internationale d'Information et de Documentation) obtiennent l'autorisation de Dewey pour adapter et traduire en français la CDD. Fidèles à l'humanisme qui prévalait à leur époque, ces deux pacifistes

ambitionnent de construire un système de classification universel, valable en tous lieux et en tous temps, et élaborent la Classification Décimale Universelle (CDU) en y introduisant les modifications suivantes:

- Simplification de la numérotation (ex. : 0 à 9 au lieu de 000 à 900)
- Création des classes principales (les 10 grands domaines de la connaissance) avec déplacement des Langues avec la Littérature dans la classe 8 afin de libérer la classe 4 en prévision d'un nouveau domaine de la connaissance nécessitant une classe entière
- Ajout de 5 classes auxiliaires, appelées subdivisions communes, dont les indices de classification peuvent être ajoutés à chaque indice de la table principale. Il s'agit des subdivisions communes de lieu (indice entre parenthèses), de temps (indice entre guillemets), de langue (indice précédé du signe =), de forme (indice commençant par (0) et de point de vue (indice précédé par un tiret).
- Ont été aussi ajoutés le signe d'agrégation (+) pour indiquer la présence de sujets différents mais non reliés entre eux dans un document, le signe d'extension (/) pour indiquer la relation entre des sujets différents mais se trouvant dans une même classe, le signe de relation (:) pour indiquer un rapport entre des sujets n'appartenant pas à la même classe.

La CDU est actuellement en ligne et mise à jour par un consortium CDU international⁸.

- En **1959** la bibliothèque de Lénine, la bibliothèque d'Etat de l'URSS, la bibliothèque publique de Sitykov-Scedrin, la bibliothèque de l'académie des sciences de l'URSS et la chambre du Livre de l'Union, se réunissent pour développer la Classification Bibliothéco-Bibliographique (CBB) à partir de la CDU. Plus de 800 scientifiques y ont contribué sous la direction de O. P. Teslenko, Docteur en philosophie⁹.

⁸ Pour la consulter en ligne, voir <http://www.udc-hub.com/>

⁹ <http://www.enssib.fr/bibliotheque-numerique/documents/62943-etude-sur-la-classification-bibliotheconomique-bibliographique-sovietiquenote-de-synthese.pdf>

La CBB comporte 6 catégories et 28 classes auxquelles sont affectées les 28 lettres de l'alphabet cyrillique. Son système est alpha-décimal, les lettres sont suivies de fractions décimales inspirées de la CDU. Elle est encore en vigueur dans plusieurs bibliothèques de l'ex-bloc soviétique, dans une version parfois adaptée.

Catégories de la CBB

Marxisme Léninisme	Société et les Sciences de la société
Nature et les Sciences de la nature	Pensée et les sciences de la pensée
Technique Agronomie Médecine	Littérature de contenu international

3.2. Système de classification poly-hiérarchique et analytico-synthétique.

Tout comme Melvil Dewey, le concepteur de ce système, Shiyali Ramamrita Ranganathan¹⁰ est un mathématicien devenu bibliothécaire suite à son recrutement en 1924 à la bibliothèque de l'Université de Madras en Inde. Ne connaissant rien à ce domaine, il bénéficie d'un séjour de quelques mois à Londres pour étudier la bibliothéconomie. Il passa près de huit années à réfléchir sur les lois de la pensée, la philosophie, notamment hindoue, et sur l'organisation des connaissances. Il étudiait les classifications bibliologiques existantes. Il était critique par rapport à la CDD, un système préfabriqué, rigide et à tiroirs, par contre il partageait le principe de Charles Ammi Cutter pour qui l'ordre des sciences est l'ordre des choses, et l'ordre des choses est celui de leur complexité. Il avait la conviction que, dans une certaine mesure, chaque système de classification est fondé sur une idéologie et qu'aucun système ne peut être évalué en dehors de son époque et de la culture qui est à son origine.¹¹

En 1933 il développe un système de classification poly-hiérarchique et analytico-synthétique. qui reflète la pensée hindoue, c'est la Colon Classification (CC) ou Classification à facettes. Au début la CC

¹⁰ Shiyali Ramamrita Ranganathan par Marie-France Blanquet, [mai 2007] <http://www.cndp.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-presse/histoire-du-livre-et-de-la-documentation/biographies/Shiyali-ramamrita-ranganathan.html>

¹¹ Il eut plusieurs échanges à ce sujet avec M. Dewey. Pour en savoir plus, consulter *Ranganathan's Monologue on Melvil Dewey* (1964). Voir http://www.hyperorg.com/misc/ranganathan_on_dewey_transcript.html

comportait 33 classes principales (actuellement il y en a 108), 10 classes généralisées (largement divisées entre les sciences humaines et les sciences exactes) et quatre tables flottantes qui correspondent aux subdivisions de forme, de lieu géographique, de temps et de langue du document et non pas du sujet traité dans le document.

Selon Ranganathan, les sujets les plus complexes peuvent toujours être réduits par l'analyse à cinq catégories fondamentales ou facettes à concrétivité décroissante (PMEST) :

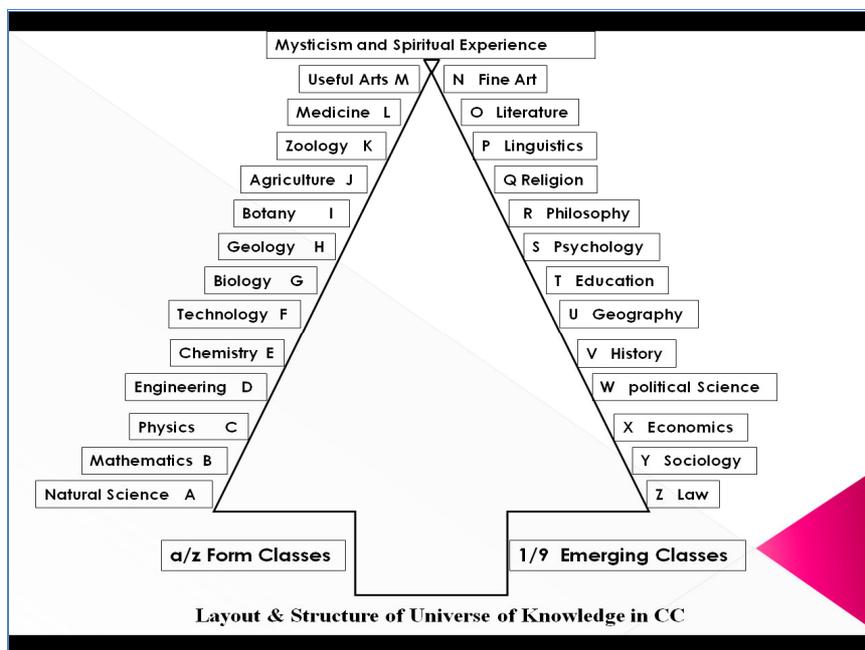
1. **P**ersonnalité : le concept principal du document
2. **M**atière : une substance ou une propriété
3. **E**nergie : l'opération ou action subie par l'objet
4. **e**Space : localisation géographique du sujet
5. **T**emps : localisation chronologique et temporelle du sujet

La facette P est présente dans toutes les classes de la CC. Les autres facettes sont réparties en fonction des concepts, par exemple **E**nergie se trouve en Economie (où elle désigne le commerce, la distribution, etc.) mais pas en Droit ou en Littérature, **M**atière existe dans la classe Arts, mais pas dans la classe Mathématiques.

L'obligation d'enchaîner les facettes toujours dans le même ordre permet d'aboutir à une notation homogène afin de classer les livres. La notation utilise des lettres majuscules pour indiquer les classes, des chiffres décimalisés pour indiquer les facettes, des symboles alphanumériques pour les subdivisions communes, des signes de ponctuation pour les relations entre sujets et enfin des lettres minuscules pour les relations entre les classes. Elle comprend aussi des caractères grecs et différents signes (P ; M : E . S ' T)

Ce système de classification permet de complexifier la séquence PMEST, autrement dit de répéter chaque facette en faisant intervenir des cycles (rounds) et des niveaux (level).

La structure des classes principales peut être schématisée comme suit :



Dans <http://fr.slideshare.net/ARUNVR2467/mapping-of-knowledge-in-colon-classification>

Exemple de « montage » des indices de la CC sans cycle ¹² :

La chirurgie dentaire est symbolisée par L 214: 4: 7

Ce code est créé en combinant la lettre L pour la médecine, le nombre 214 pour les dents, le nombre 4 pour les maladies, et le nombre 7 pour la chirurgie.

Ranganathan veillait lui-même à la mise à jour de la CC (1987, 7^{ème} éd. Posthume).

Le principe des facettes a été repris dans les Thésaurus et surtout dans les annuaires des moteurs de recherche, le Web 2.0 et les logiciels libres. Une étude a démontré que, par rapport à la classification monohiérarchique « *la classification à facettes est souple et ouverte à de nouvelles catégories. Elle est aussi plus appropriée pour représenter des idées ou des objets complexes. Elle est plus particulièrement utile pour*

¹² Pour voir des exemples comportant des cycles et des niveaux, voir Colon classification by S. R. Ranganathan : an outline with examples. Disponible sur <http://www.iskoi.org/doc/colon.htm>

organiser des pièces ou des documents complexes dans un environnement multidisciplinaire »¹³.

Synthèse

L'objectif de repérage simultané de l'information et du document a conditionné la représentation des connaissances basée sur les systèmes classificatoires. Ceci n'a pas empêché ces dernières d'être idéologiquement et culturellement marquées par l'environnement dans lequel elles ont été créées. Mais aussi, dès lors que toutes les disciplines produisent leurs résultats sous forme de documents, ces documents confèrent aux classifications un caractère encyclopédique car ils touchent à toutes les choses, et cela confère à la représentation de la connaissance un caractère universel. L'« être documenté » est pour la bibliologie ce que l'« être de la raison » est pour la logique¹⁴. Au plan pratique, les codes obtenus pour représenter le contenu d'un document sont le résultat de la pré-coordination de ses divers éléments informatifs au moment de l'indexation.

Le fait que la CDD soit le système de classification le plus utilisé dans le monde, plus de 135 pays en tout, en particulier dans les pays arabes¹⁵ et traduite dans 37 langues, peut s'expliquer par le fait que c'est d'abord la seule qui ait été traduite en arabe (au moins 3 fois selon les éditions successives), que le système mono-hiérarchique est simple et que les indices sont mnémotechniques. Les professionnels qui l'appliquent continuent toutefois de lui reprocher d'une part un « christiano-

¹³ KYUNG6SUN, Kim et alii. Analyse des facettes des catégories utilisées dans les annuaires du Web : étude comparative. Séoul : IFLA 2006. Disponible sur http://www.researchgate.net/publication/266272680_Analyse_des_facettes_des_catgories_utilises_dans_les_annuaires_du_Web_tude_comparative.

¹⁴ Paul Otlet. Traité de documentation. Editions Mundaneum, 1934. Disponible en ligne http://fr.wikisource.org/wiki/Trait%C3%A9_de_documentation

¹⁵ Pour en savoir plus, consulter OCLC, Services Dewey, disponible sur <http://www.oclc.org/fr-CA/dewey.html> et en particulier le Projet WEB Dewey en arabe de la Bibliothèque d'Alexandrie disponible sur https://www.google.tn/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0CEAQFjAF&url=http%3A%2F%2Fwww.melcominternational.org%2Fwp-content%2Fcontent%2Fpast_conf%2F2007%2F2007_papers%2Fiman.doc&ei=F5YSVYaaHsPSaI_1gaAJ&usg=AFQjCNEqyyotxZzMOoCirlcIXj5kXwiFww&bvm=bv.89184060,d.d24

centrisme », que Dewey reconnaissait lui-même, et d'autre part le fait que, pour les ouvrages ne relevant pas d'un domaine du savoir occidental, les cotes sont nettement plus longues comme le montre l'exemple suivant¹⁶ :

André Miquel - <i>L'Islam et sa civilisation</i>		
	Version française	Traduction en arabe
Indice Dewey	297.09	909.49 27

4. Le système de représentation des connaissances combinatoire et analytique

4.1. Origine du système

Le système de représentation des connaissances combinatoire et analytique se démarque du système classificatoire, il a été développé pendant la 2ème guerre mondiale. A la faveur de cet événement, l'information a en effet acquis une valeur stratégique et l'objectif d'y accéder rapidement et indépendamment de son support s'est imposé de manière cruciale. De plus, l'accroissement exponentiel de la production documentaire, la spécialisation de plus en plus poussée et surtout l'interdisciplinarité ont révélé les limites des systèmes de classifications dont le double objectif conditionne pratiquement une indexation du document au "sujet principal" afin qu'il soit localisé dans une seule adresse physique (sa cote de classement).

De nombreuses recherches, dans lesquelles l'informatique naissante aura un rôle important, jalonnèrent le passage des systèmes classificatoires vers les systèmes combinatoires. Les résultats des travaux menés au cours de ces étapes n'ont été rendus publics qu'après la guerre grâce à Vannevar Bush dont il sera question ci-dessous. Nous citons les étapes les plus significatives de ce passage et ses acteurs les plus décisifs.

- Les systèmes de fiches perforées à sélection visuelle : en réalité, ce type de fiche dont l'objectif est la représentation et le repérage rapide de l'information connaît plusieurs prototypes et applications

¹⁶ Pour avoir plus de témoignages de professionnels à ce sujet, consulter Savoir et science dans les bibliothèques (François Dingremont) 2012. Disponible sur <http://lieuxdesavoir.hypotheses.org/1128>

depuis les années 1920¹⁷. Nous décrivons en particulier le système Uniterm (substantif simple, plus petit élément significatif du langage) du bibliothécaire documentaliste Mortimer Taube (1950)¹⁸ qui a travaillé avant à la Bibliothèque du Congrès, au Département de la Défense, et au sein de la Commission de l'énergie atomique. L'uniterme représente le contenu d'un document en langage naturel, c'est-à-dire dans les termes de l'auteur du document et non pas en ayant recours à un système de classification pré-établi. Tous les sujets traités dans un document sont représentés, et pour chaque sujet, il est établi une seule fiche. Des perforations sont faites dans les colonnes de la fiche correspondant aux numéros d'inventaire de tous les documents traitant le même sujet, celles-ci apparaîtront sous forme de trous lumineux lors de la recherche. Ce système a été testé à la Bibliothèque du Congrès américain où plus de 60.000 fiches bibliographiques ont pu être réduites à 3000 fiches perforées à sélection visuelle, soit respectivement autant de sujets construits avec autant d'unitermes. Lors de la recherche d'un document traitant de deux sujets déterminés on confronte les fiches de leurs unitermes pour déceler les numéros d'inventaire qui sont communs à ces deux sujets. Le chercheur procède ainsi à une post-coordination entre les sujets de sa recherche. Mais ce système fut générateur de beaucoup de bruit lors de la recherche d'information (le chercheur obtient plus d'informations qu'il n'en a besoin). Il s'avéra que ce bruit est dû à l'uniterme lui-même et non pas au principe de la post-coordination qui fut retenu et qui exploita les opérateurs de la logique booléenne, à savoir l'intersection (ET), la réunion (OU) et l'exclusion (SAUF) entre les sujets d'une recherche.

- La critique des systèmes de représentation existants : cette critique est notamment due à l'ingénieur électronicien américain, responsable de l'organisation du National Defense Research Committee (NDRC) et de l'Office of Scientific Research and Development (OSRD) de 1940 à 1947, Vannevar Bush. Il leur reproche le fait que « *chaque item n'y est classé que sous une seule*

¹⁷ Pour en savoir plus, consulter *Les systèmes de sélection*, Paule Salvan et Paul Piondron, 1957, disponible en ligne <http://bbf.enssib.fr/consulter/bbf-1957-06-0455-001>

¹⁸ Taube (Mortimer) et alii. - *Studies in coordinate indexing*. - Washington, 1953-1956. - 3 vol.

rubrique, et le rangement y est purement hiérarchique (classes, sous-classes,...) », alors que selon lui « l'esprit humain fonctionne par associations. Il saute d'une représentation à l'autre le long d'un réseau enchevêtré. » V. Bush recommande au groupe de recherche sur le traitement de l'information (dont M. Taube et A. Turing, faisaient partie) de mettre au point des outils méthodologiques qui permettraient au lecteur de nouer des liens indépendamment de toute classification hiérarchique, entre une information quelconque et n'importe quelle autre.¹⁹ Les idées de l'indexation associative et du Memex (considéré comme l'origine des liens hypertextes) dont le fonctionnement a été décrit en 1945²⁰ ne sont que des étapes s'ajoutant aux précédentes afin d'aboutir aux langages documentaires combinatoires connus sous le nom de thésaurus²¹.

- La notion de descripteur : il est quasi-unanime d'attribuer l'invention de cette notion à Calvin Mooers, mathématicien et informaticien, qu'il a utilisé dans son mémoire de maîtrise (M.I.T. 1948) et dans ses écrits à partir de 1950. Il mit également au point un système de sélection en 1952 fondé sur un sélecteur automatique de fiches pré-perforées, les Zatocards, alliage des codes et des descripteurs pour analyser l'information, la coder et la retrouver. Là où M. Taube utilise des unitermes, C. Mooers utilise ce qu'il appelle les descripteurs ou mots conceptuels, que

¹⁹ Marie-Madeleine Varet-Pietri L'ingénierie de la connaissance: la nouvelle épistémologie appliquée, Volume 696. Presses Universitaires Franc-Comtoises. 2000.

²⁰ « Un memex, c'est un appareil dans lequel une personne stocke tous ses livres, ses archives, ses idées et sa correspondance, et qui est mécanisé de façon à permettre la consultation à une vitesse énorme et avec une grande souplesse. Il s'agit d'un supplément agrandi et intime de sa mémoire » dans : Vannevar Bush. « As we may think », The Atlantic Monthly ; July, 1945 ; Volume 176, No. 1 ; pages 101-108. Trad. fr. Charles Monatte. En ligne sur http://mediateur.free.fr/web/hist_aswemaythink_fr.htm,

²¹ Le mot Thésaurus est un emprunt au grec ancien et au latin (répertoire). Il existe des thesaurie (1592), utilisés dans le sens de dictionnaire en 1592. En 1840, thésaurus a le sens « d'encyclopédie avec informations ». Le terme thésaurus utilisé en science de l'information-documentation serait emprunté au Thésaurus de Peter Mark Rodget, Thésaurus of English Words and Phrases Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary publié en 1852.

l'AFNOR définira plus tard (en 1987) comme des : « *termes ou groupe de mots retenu dans un thésaurus et choisi parmi un ensemble de termes équivalents pour représenter sans ambiguïté une notion apparaissant dans un document ou dans une demande de recherche documentaire* ». Plusieurs critères morphosyntaxiques et sémantiques, dont le plus fondamental est la bi-univocité (un descripteur = un et un concept / un concept = un et un seul descripteur) doivent satisfaire la notion de descripteur.

4.2. Premiers systèmes de représentation combinatoire et analytique (Thésaurus)

Si la secrétaire de V. Bush, H. -L. Brownson, a utilisé publiquement le terme de thésaurus documentaire en 1958 lors de la Conférence internationale sur l'information scientifique « Application of a mechanized thesaurus based of neetworks of related meanings », la paternité de ce système demeure cependant incertaine. Elle est notamment revendiquée par les deux associés américains Bernier et Crâne du Chemical abstracts service, qui menèrent, en 1947, des travaux sur des alternatives aux cartes perforées consistant essentiellement à indexer les termes de recherche employés avec succès de telle sorte que, si la même question ou une question similaire est répétée, alors la consultation du fichier révélerait l'ensemble antérieurement réussi des termes de l'index caractérisés par des relations de subordination, de coordination et d'équivalence. Cette paternité est aussi revendiquée par l'ingénieur allemand Peter Luhn, qui fut responsable du service d'information chez IBM après avoir servi l'armée allemande pendant la guerre en tant que responsable de la communication. Il fut le concepteur, en 1947, du premier prototype de la « Luhn Scanner » qui sera 'utilisé par la CIA dans la recherche d'information. Luhn revendique la paternité du thésaurus en se basant sur un article publié en janvier 1953, intitulé « Nouvelle méthode d'enregistrer et de rechercher une information » dont le contenu a été soumis à l'American chemical society -à laquelle était rattachés Bernier et Crâne- le 11 septembre 1951.

Les premiers thésaurus publiés furent pratiquement tous consacrés aux domaines de l'armement et des sciences exactes. Ce fut le cas en mai 1960 aux Etats Unis avec le Thesaurus of ASTIA descriptors du Département de la Défense, avec des concepts selon la méthode de Calvin Mooers. En 1961 paraît le Chemical Engineering Thesaurus, élaboré par l'American Institute of Chemical Engineers (AIChE), rapidement suivi par le premier thésaurus européen, le Thesaurus EURATOM (Communauté européenne de l'énergie atomique), et le

premier thésaurus français, celui du Centre de documentation de l'armement (Thésaurus CEDOCAR).

4.3. Normalisation des thésaurus

Dès 1971, l'UNESCO a organisé plusieurs rencontres avec l'UNISIST et l'ISO pour mettre au point des principes directeurs pour l'établissement et le développement des thésaurus monolingues²² et multilingues²³. En 1972, leur construction a été normalisée au plan national et international,²⁴ pratiquement pour entériner un état de fait, en fusionnant au fur et à mesure des versions, les principes des thésaurus mono et multilingue.

Le thésaurus a été défini en 1981 par l'Association Française de Normalisation (NF Z47-100:1981), comme « *un langage documentaire fondé sur une structuration hiérarchisée d'un ou plusieurs domaines de la connaissance et dans lequel les notions sont représentées par des termes d'une ou plusieurs langues naturelles et les relations entre les notions par des signes conventionnels.* ». La définition actuelle (norme ISO 25964-1:2011) met en valeur quant à elle la notion de concept et le définit comme « *un vocabulaire contrôlé et structuré dans lequel les concepts sont représentés par des termes, organisés de façon à ce que des relations entre les concepts soient explicitées, et dont les termes préférentiels sont accompagnés par des entrées vers leurs synonymes ou quasi-synonymes.* »

Dans un thésaurus, les termes sont appelés descripteurs et obéissent au principe de la bi-univocité (un descripteur= un concept / un concept = un descripteur). La notation existe mais n'a qu'une valeur indicative pour situer le descripteur dans la facette.

²² <http://unesdoc.unesco.org/images/0000/000059/005951FB.pdf>

²³ <http://unesdoc.unesco.org/images/0003/000392/039209Fb.pdf>

²⁴ Dernière version de la norme ISO 25964-1:2011 Information et documentation -- Thésaurus et interopérabilité avec d'autres vocabulaires -- Partie 1: Thésaurus pour la recherche documentaire. Elle remplace les normes ISO 2788:1986 (Thésaurus monolingue) ; ISO 5964:1985 (Thésaurus multilingue) • NF Z47-100:1981 (Thésaurus monolingue) ; NF Z47-101:1990 (Thésaurus multilingue). La Partie 2 de la norme ISO 25964-1:2011 concerne l'interopérabilité entre différents thésaurus (à schéma fléché, à arborescence ; graphique, circulaire, micro et macrothésaurus etc.) et d'autres types de vocabulaires structurés (classifications, taxonomies, listes d'autorité, ontologie, etc.)

- Relation hiérarchique : héritée des systèmes de classification, relation d'inclusion et de supériorité entre terme spécifique (TS) et terme générique (TG).
- Relation d'équivalence intra-linguistique entre des synonymes ou des quasi-synonymes. On utilise la notation EM (Employer) et le EP (Employé pour). Pour l'utilisateur cette relation est signalée par *Voir*.
- Relation d'association : elle est établie entre descripteurs ayant une relation sémantique mais se trouvant des facettes différentes. On utilise la notation TA (terme associé) pour signaler diverses associations possibles (cause/effet ; tout/partie ; action /agent ; objet/application ; etc.). Cette relation est utilisée pour élargir la recherche. Pour l'utilisateur elle est signalée par *Voir aussi*.
- Relation inter-linguistique : relation symétrique entre deux descripteurs de deux ou plusieurs langues. On utilise les sigles des différentes langues : An/ FR/ ES
- Notes d'application (NA) : définitions qui expliquent les modalités d'emploi d'un descripteur dans le but d'éviter les ambiguïtés, en particulier la polysémie.

Exemples extraits du Thésaurus des Nations Unies (UNBIS' Thesaurus):

MANDATS (SDN)

01.05.00 - DECOLONISATION, TUTELLE ET APARTHEID

Employé pour

TERRITOIRES SOUS MANDAT

Note d'application: Mandats d'administration de certains territoires confiés par la SDN à ses Etats membres. Pour la mission confiée à un organe ou organisation, utiliser **MANDAT**

Termes associés

CONDOMINIUM

PROTECTORATS

QUESTION DE NAMIBIE

SPHERES D'INFLUENCE

TERRITOIRES SOUS TUTELLE

Synthèse

Actuellement, les thésaurus se comptent par centaines²⁵, ils sont quasiment tous en ligne et gérés par des logiciels spécifiques. Ils sont surtout utilisés dans les centres de documentation et les centres de recherche spécialisés préoccupés du repérage et de l'accès à l'information courante. Ils n'ont pas supplanté les systèmes de classification qui les ont précédés dans les bibliothèques conservant des fonds encyclopédiques à l'attention d'usagers nombreux et diversifiés.

A l'instar des systèmes de classification, ils ont une nette empreinte idéologique notamment dans le domaine des sciences humaines et sociales. Par exemple, dans l'UNBIS thésaurus, le descripteur FEDDAYIN est rattaché à la facette DESARMEMENT ET QUESTIONS MILITAIRES alors que son terme générique GUERILLEROS est associé au descripteur MOUVEMENTS DE LIBERATION NATIONALE qui dépend de la facette DECOLONISATION, TUTELLE ET APARTHEID.

La construction des thésaurus a proliféré jusqu'aux années 1990, lorsque le web sémantique a offert de nouvelles possibilités de représentation des connaissances adaptées aux supports électroniques diffusés par des flots informationnels, en l'occurrence les ontologies.

5. Les Ontologies

5.1. Origine des ontologies

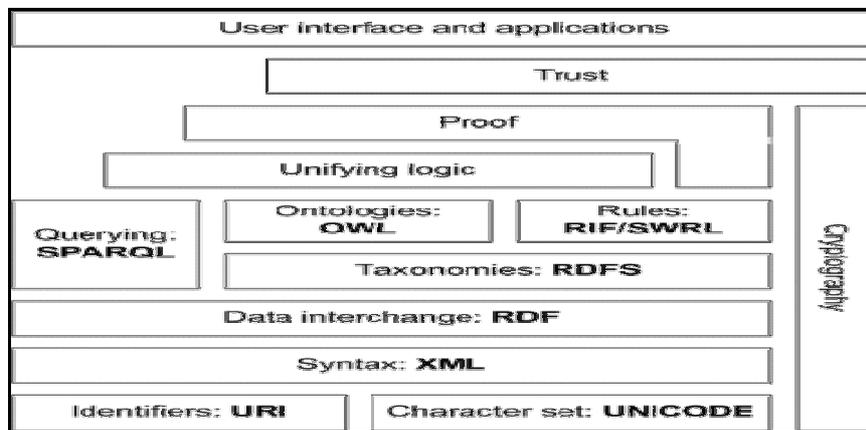
Les ontologies sont apparues suite à une réflexion qui commença en 1994, au moment de la première conférence World Wide Web à Genève, et qui s'acheva en 1998 lorsque Tim Berners-Lee, physicien britannique inventeur du WWW, publia le projet de configuration du Web sémantique. L'objectif du WWW est que les ordinateurs du réseau Internet (dont la vision prémonitoire est due à P. Otlet dans son *Traité de documentation*, 1934) puissent prendre en charge les tâches de recherche et d'association d'information à la place des usagers au moyen des liens

²⁵ Pour en savoir plus, voir la liste des thésaurus francophones en ligne sur <http://www.dmoz.org/World/Fran%C3%A7ais/R%C3%A9férences/Th%C3%A9saurus>

hypertextes²⁶ qui sont une invention du sociologue et informaticien américain Theodor Holm Nelson en 1965.²⁷

Nous sommes en mesure d'avancer que les ontologies sont le fruit de la rencontre entre, d'une part l'objet de la science de l'information-documentation qui est l'étude des propriétés générales de l'information et l'analyse de ses processus de construction, de communication et d'usage, et d'autre part l'objet de la technologie de l'information qui consiste en la conception de produits, de systèmes et de services qui permettent la construction, la communication, le stockage et l'usage de l'information.

L'emplacement des ontologies dans les briques technologiques du Web sémantique (*layer cake*) est représenté comme suit par Berners-Lee :



5.2. Définition des ontologies

Cette rencontre entre la science et la technologie a également favorisé l'ingénierie des connaissances (IC), historiquement issue de l'intelligence artificielle (IA), et dont le projet fondamental est le

²⁶ Dans ses réponses aux jeunes T. Berners-Lee déclare «*Je n'ai fait que prendre le principe d'hypertexte et le relier au principe du TCP et du DNS et alors – boum ! – ce fut le World Wide Web !* » Disponible sur <http://www.w3.org/People/Berners-Lee/Kids.html>.

²⁷ «*Let me introduce the word 'hypertext' to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper* Pour en savoir plus, consulter Jean Clément. *Du texte à l'hypertexte: vers une épistémologie de la discursivité hypertextuelle*. 1995. Disponible sur <http://hypermedia.univ-paris8.fr/jean/articles/discursivite.htm#donnees>

développement des systèmes à base de connaissance (SBC) dans lequel les ontologies fournissent des représentations formalisées et structurées du vocabulaire spécifique à un domaine de la connaissance.

Comparées à l'ontologie philosophique qui est la science de l'*être* (l'étude de l'être en tant qu'être selon Aristote), les ontologies représentent l'ensemble des objets / concepts renvoyant à ce que l'on considère exister dans un domaine, autrement dit c'est l'étude de l'*étant*, d'où la relation *is-a* entre un concept et ses propriétés dans les ontologies.

Plus précisément, une ontologie est une « *description formelle explicite des concepts dans un domaine du discours (classes (appelées parfois concepts)), des propriétés de chaque concept décrivant des caractéristiques et attributs du concept (attributs (appelés parfois rôles ou propriétés)) et des restrictions sur les attributs (facettes (appelées parfois restrictions de rôles))*.²⁸ » Une ontologie ainsi que l'ensemble des instances individuelles des classes constituent une base de connaissances.

Les classes sont en général organisées hiérarchiquement par la relation «est-un» (*is-a*), elles génèrent des individus ou instances qui ont des valeurs spécifiques. Les propriétés peuvent être de deux types : les propriétés d'objets utilisées pour les attributs dont la valeur correspond à une instance d'une autre classe, et les propriétés de types de données qui sont les propriétés simples exprimées par des chiffres ou des notation alphabétiques, les propriétés peuvent aussi être classées hiérarchiquement.

5.3. Typologie et construction des ontologies

Il existe divers types d'ontologies qui correspondent à divers niveaux d'abstraction, les plus globales sont les ontologies génériques (ex. : MIKROKOSMOS), les ontologies de domaine sont spécifiques (ex. MENELAS, FAO), et enfin les ontologies de tâches sont destinées à une application dans un domaine. (ex. : ONTOLINGUA).

²⁸ Natalya F. Noy et Deborah L. McGuinness. Développement d'une ontologie 101 : Guide pour la création de votre première ontologie. Université de Stanford, Stanford, CA, 94305

Traduit de l'anglais par Anila Angjeli, BnF, Bureau de normalisation documentaire. (2002) Disponible sur <http://www.bnf.fr/documents/no-DevOnto.pdf>

Les ontologies sont construites en quatre étapes²⁹, les deux premières étant souvent confondues :

- Conceptualisation : s'effectue à partir de l'analyse et de la description du corpus caractérisant le domaine de l'ontologie.
Corpus = Signifié
- Normalisation sémantique : a pour objectif l'identification des objets du domaine et de leurs types. Signifié = Signifié normé (ou Concept linguistique)
- Ontologisation : engagement ontologique. Concept linguistique = Concept formel
- Opérationnalisation : dans un langage de représentation des connaissances (OWL : Ontology Web Language) à partir d'un logiciel de construction d'ontologies.

La conceptualisation peut se faire à partir d'un corpus de texte et d'un thésaurus et elle est souvent précédée d'une analyse des besoins auprès des futurs usagers de l'ontologie.³⁰

L'ontologisation ou la construction du modèle formel est variable et dépend du contexte d'usage de l'ontologie. Par exemple, la notion d'EMPLOYE peut être modélisée en mettant en relation employé/employeur, elle peut aussi être modélisée en considérant l'action EMPLOYER qui fait intervenir les deux relations casuelles AGENT (un employeur) et PATIENT (un employé), elle peut encore être modélisée en considérant l'action EMPLOYER en rapport avec le concept d'EMPLOI, etc.

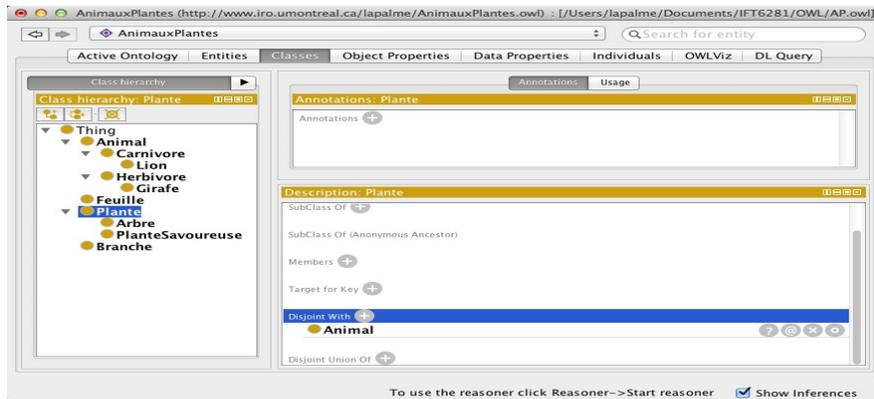
Pour illustrer l'étape de l'opérationnalisation, nous nous basons sur un exemple simple employant l'éditeur d'ontologies Protégé 2000 (version 4.3)³¹. Cet éditeur n'est pas le seul, mais c'est l'un des plus

²⁹ Bruno Bachimont. (2000). Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en Ingénierie des connaissances. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds.), Ingénierie des connaissances, évolutions récentes et nouveaux défis. Paris: Eyrolles.

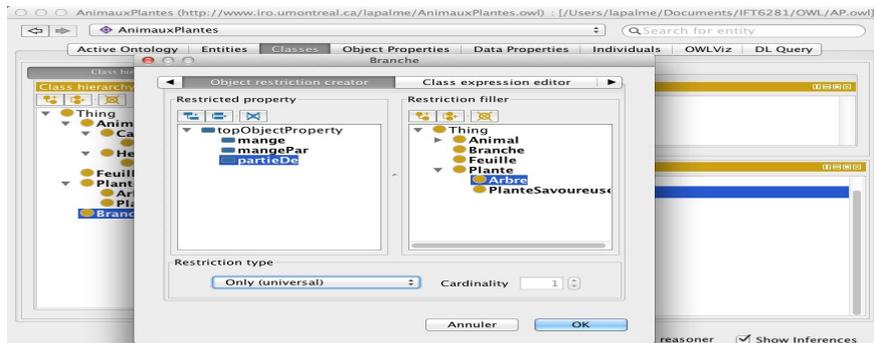
³⁰ Par exemple la FAO a transformé son Thésaurus AGROVOC en Serveur de concepts puis en Service d'ontologie agricole.

³¹ Guy Lapalme Création d'une petite ontologie avec Protégé.. Disponible sur <http://www.iro.umontreal.ca/~lapalme/ift6282/OWL/EtapesCreationOntologie.html>

utilisés³². L'image ci-dessous montre une hiérarchie de deux classes disjointes « Animal » et « Plante » avec leurs sous-classes respectives.



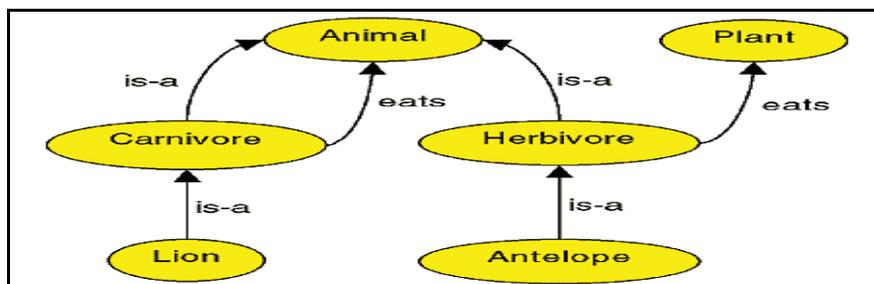
Des fonctionnalités sont disponibles pour créer les propriétés mange (son domaine est un Animal), mangePar (inverse de mange) et partieDe (propriété transitive) comme l'illustre l'image suivante :



L'éditeur d'ontologie permet aussi de créer des classes complexes (comme par ajouter une sous-classe « feuille » subordonnée à « branche » ou bien indiquer qu'un Carnivore est un Animal qui mange aussi un Animal ou encore qu'une Girafe est un Herbivore qui ne mange que des Feuilles, etc.). Il est également possible d'ajouter des individus ou instances (dans notre exemple cela consiste par exemple à donner des noms aux lions.)

³² L'éditeur est open source et téléchargeable sur <http://protege.stanford.edu/>

Le résultat de l'ontologie à un certain moment de sa création est le suivant :



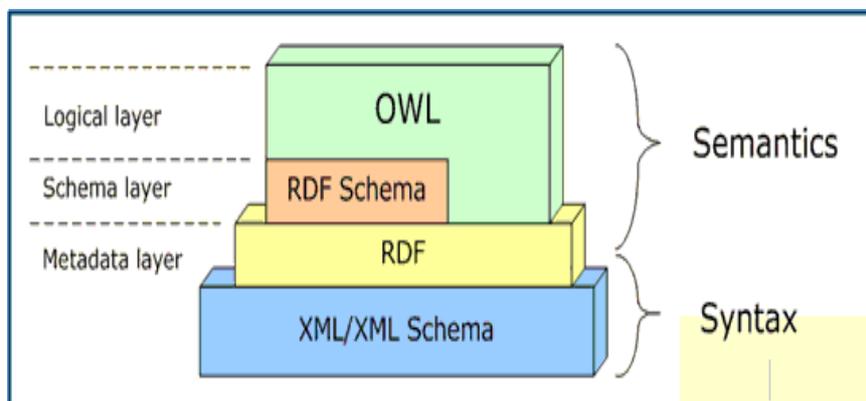
La dernière étape de l'opérationnalisation est la vérification automatique de la cohérence de l'ontologie qui détecte les définitions contradictoires, les classes inconsistantes, etc. grâce à la fonctionnalité *Reasoner*.

Synthèse

Selon nous, comparées aux thésaurus, les principaux apports des ontologies sont des relations associatives plus explicites, l'intégration d'un moteur d'inférences (déductive, inductives et abductive) qui simule le raisonnement humain (*reasoner/start reasoner* dans le menu de l'éditeur d'ontologie), la souplesse et le dynamisme de la structuration lors de l'opérationnalisation (de simples clics) et enfin le résultat final sous divers graphismes au choix, (*ontograpf*) représentant les objets, leurs propriétés, et la nature de leurs relations sur un simple passage du curseur.

Le problème de l'accès ne se pose pas pour les ressources électroniques car l'ontologie (représentation du contenu) et les métadonnées (*metadata* - description du document) sont associés dans des briques complémentaires.³³

³³ Pour consulter un exemple concret, voir Groupe de travail québécois sur les normes et standards en TI pour l'apprentissage, l'éducation et la formation . Ontologie de description et vocabulaire de métadonnées pour les scénarios pédagogiques Michel Léonard, Gilbert Paquette. 18 Juin 2014. Disponible sur <http://www.gtn-quebec.org/rapport/ontologie-de-description-et-vocabulaire-de-metadonnees-pour-les-scenarios-pedagogiques>.



6. Discussion

6.1. Au plan de la finalité de la représentation

Au terme de cette présentation des différents systèmes de représentation de la connaissance en science de l'information-documentation³⁴, il est important de préciser que la représentation des contenus documentaires n'a jamais eut la prétention de se substituer au document lui-même – ou à la ressource- produit par un auteur ou par une collectivité. En fait, avant d'être élaborés, tous les systèmes de représentation reconnaissent les concepts d'un domaine et collectent les termes à partir des documents de base en citant généralement leurs sources dès l'introduction et dans une bibliographie. Les experts d'un domaine sont consultés et le système est testé auprès d'un certain nombre d'utilisateurs.

Nonobstant l'orientation idéologique de certains systèmes, la représentation des connaissances en science de l'information-documentation est un moyen offert au lecteur ou à l'utilisateur, confrontés soit à une énorme masse documentaire, soit à un flux continu de

³⁴ Dans laquelle ont été omis d'autres types de représentation de la connaissance typiquement bibliothéconomiques comme les listes de vedettes-matières ou d'autorités (ex. RAMEAU) dont la construction et l'application sont normalisées (norme NF Z 47-100, 1981) car leur structure alphabétique ne pose pas de problématique particulière en terme de représentation, et comme les cartes mentales ou les graphes conceptuels dont le formalisme a été introduit par John F. Sowa en 1984 (Freemind, etc.) car ce ne sont pas des outils conçus pour la représentation du contenu des documents bien qu'ils peuvent être utilisés à cette fin.

ressources, pour repérer le document qui répond le mieux, ou qui se rapproche le plus de ses centres d'intérêts et le consulter dans son intégralité. Ce faisant, la représentation de la connaissance s'inscrit dans l'éthique et la déontologie du domaine de la science de l'information-documentation, à savoir l'accès de tous au savoir, et même davantage le libre accès et la libre circulation de l'information³⁵.

6.2. Au plan des pratiques professionnelles

Au plan des pratiques professionnelles, la principale remarque est que, plus les modes de représentations évoluent et plus la représentation s'imbrique dans le document dont elle est censée représenter le contenu. En effet, les ontologies sont « avec » la ressource qu'elles représentent, tandis que les index des sujets, issus des systèmes de représentation tels les classifications et les thésaurus, sont « à part » dans des fichiers matières, des catalogues ou des répertoires alors que les documents sont entreposés ailleurs. Actuellement, dans les OPAC (Online Public Access Catalogue) les références sont en ligne et pointent vers les documents numérisés et accessibles en texte plein grâce aux liens hypertexte.

6.3 Au plan épistémologique

Au plan épistémologique, il est à présent opportun de déclarer que notre domaine d'application, justement en raison du fait que la pratique y a souvent précédé la théorie, vient à peine de se démarquer du domaine purement technique pour se revendiquer pleinement du domaine des sciences humaines et sociales. Il convient aussi de signaler que même l'appellation de la science dont nous parlons n'est pas encore établie, entre science au singulier et au pluriel, science de l'info-com., l'info-doc., ou de l'information tout court, etc.

Toutes les sciences étant évolutives, et toutes se basant sur des documents, la notion de représentation de la connaissance en science de l'information-documentation constitue l'un des repères théoriques de cette science qui fait régulièrement l'objet de recherches et de questionnements. Cela est également valable pour les autres aspects en

³⁵ Voir à ce propos l'affaire entre une société de production de bases de données et un journal renommé qui a duré près de ans (épilogue en 1987) et qui a fait jurisprudence « Les conséquences juridiques de l'affaire Microfor/Le Monde ». Disponible sur file:///C:/Users/isd/Downloads/les-infostrateges_31_16-03-1988.pdf

relation avec son objet d'étude tels que la gestion ou le droit de l'information.

Pour asseoir ces repères et forger ses concepts, la science de l'information-documentation fait autant appel à l'interdisciplinarité (linguistique, philosophie, informatique) qu'à l'intradisciplinarité (les théories de tous les domaines de la connaissance traités, par exemple l'algèbre de Boole pour construire les équations de recherche de l'information, la théorie de la qualité pour améliorer les services ou les produits documentaires, la psycho-sociologie pour analyser les usages et les usagers, etc.).

Nombreux sont les colloques consacrés à complexité de l'épistème de cette science, mais d'une manière ou d'une autre, la majorité des théoriciens s'accordent sur le fait qu'elle « *demande encore des recherches, afin de développer lois, concepts et méthodes. Elle doit accumuler et interpréter observations et expériences pour aboutir à une théorie générale.* »³⁶

Notre opinion s'inscrit dans la tendance interdisciplinaire de l'Association internationale de l'organisation des connaissances³⁷. Nous penchons en effet vers la nécessité d'un regard anthropologique qui pourrait analyser, et pas simplement décrire, les symboliques et les situations vécues à différents moments de l'évolution des systèmes de représentation ainsi que la nature des rapports entre les différents groupes sociaux impliqués dans la même entreprise de construction du savoir humain, depuis l'auteur/producteur jusqu'à l'utilisateur ou reproducteur, en passant par l'informatiste/médiateur. Cela contribuera sans nul doute à approfondir la relation entre le document et les différents

³⁶ Yves-François Le Coadic. La science de l'information. Paris : PUF, 1994. (Que sais-je ? 2873).

Citons aussi les travaux de Hubert Fondin, entre autres La science de l'information ou le poids de l'histoire. Article inédit. Mise en ligne: 24 mars 2006, disponible sur <http://lesenjeux.u-grenoble3.fr/2005/Fondin/index.php>.

Et aussi l'un des derniers articles car la liste serait longue, Stefanie Averbeck-Lietz, Fabien Bonnet et Jacques Bonnet, « Le discours épistémologique des Sciences de l'information et de la communication », *Revue française des sciences de l'information et de la communication* [En ligne], 4 | 2014, mis en ligne le 15 janvier 2014, consulté le 06 avril 2015. URL : <http://rfsic.revues.org/823>.

³⁷ Pour en savoir plus, voir les conférences de l'ISKO (International Society for Knowledge Organisation) en ligne sur <http://www.isko.org/events.html#int>

regards portés sur lui et entre l'information et les différents sens qu'on lui attribue.

Dans la notion de document, nous intégrons bien évidemment les différents systèmes de représentation des connaissances qui ne se réduisent pas à l'usage utilitaire mais qui sont aussi porteurs de sens.

Bibliographie

Al-Rabe, A. A. (1987). *Muslim Philosopher's Classifications of the Sciences : Al-Kindi, Al-Farabi, Al-Ghazali, Ibn Khaldun*. Ann Arbor, Michigan: UMI, Dissertation Information Service.

Averbeck-Lietz S. Bonnet F. et Bonnet, J. (15 janvier 2014). Le discours épistémologique des Sciences de l'information et de la communication, *Revue française des sciences de l'information et de la communication*. 4 - 2014. Repéré sur <http://rfsic.revues.org/823>.

Bachimont B. (2000). *Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en Ingénierie des connaissances*. J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*. Paris: Eyrolles.

Barthe E. (Avril 2004). *Le lien hypertexte et la loi*. Repéré sur http://www.precisement.org/panor_presse/resume_archimag.htm

Bessagnet M.-N. ,Kergeosien E. ,Gaio. M. (2000). *Extraction de termes, reconnaissance et labellisation de relations dans un thésaurus – Vers une ontologie*. Repéré sur <https://hal.archives-ouvertes.fr/hal-00451960/document>

Blanquet M.-F. (2007). *Shiyali Ramamrita Ranganathan*. Repéré sur <http://www.cndp.fr/savoirscdi/societe-de-linformation/le-monde-du-livre-et-de-la-presse/histoire-du-livre-et-de-la-documentation/biographies/shiyali-ramamrita-ranganathan.html>

Blanquet M.-F. (2002). *Les origines des sciences de l'information et de la communication : regards croisés*. Robert Boure, éd. Villeneuve d'Ascq : Presses Universitaires du Septentrion

Borko H. (1968). *Information science: what is it?* American Documentation.

Bush V. (Juin 1945). *As we may think*. The Atlantic Monthly. Volume 176, No. 1 .p. 101-108. (traduit par Monatte C.) Repéré sur http://mediateur.free.fr/web/hist_aswemaythink_fr.htm

Clément Jean. (1995). *Du texte à l'hypertexte: vers une épistémologie de la discursivité hypertextuelle*. Repéré sur <http://hypermedia.univ-paris8.fr/jean/articles/discursivite.htm#donnees>

Fondin H. (24 mars 2006). *La science de l'information ou le poids de l'histoire*. Article inédit. Repéré sur <http://lesenjeux.u-grenoble3.fr/2005/Fondin/index.php>

Groupe Cyclope (22 décembre 2011). *Bibliographie sélective : encyclopédisme et organisation du savoir en Islam*, Repéré sur <http://cyclopes.fltr.ucl.ac.be/materiaux/biblioarabe.htm>

Hudon, M. (2005). Le passage au XXI^e siècle des grandes classifications documentaires. *Documentation et bibliothèques*, p.85-97

ISO (1986) Norme ISO 25964-1:2011 *Information et documentation -- Thésaurus et interopérabilité avec d'autres vocabulaires -- Partie 1: Thésaurus pour la recherche documentaire*

Champavère J. (10 mars 2010). De la représentation des connaissances au Web sémantique : Un survol. Repéré sur <http://www.grappa.univ-lille3.fr/~champavere/Enseignement/0809/13miashs/ia/rc-ws.pdf>

Kyungsun, K. et alii. (2006). *Analyse des facettes des catégories utilisées dans les annuaires du Web : étude comparative*. Séoul : IFLA
Repéré sur :
http://www.researchgate.net/publication/266272680_Analyse_des_facettes_des_catgories_utilises_dans_les_annuaires_du_Web_tude_comparative.

Lapalme G. Création d'une petite ontologie avec Protégé. Repéré sur <http://www.iro.umontreal.ca/~lapalme/ift6282/OWL/EtapesCreationOntologie.html>

Le Coadic Y.- F. (1994). *La science de l'information*. PUF, Que sais-je ?

Léonard, M. Paquette G. (18 Juin 2014). Ontologie de description et vocabulaire de métadonnées pour les scénarios pédagogiques. Groupe de travail québécois sur les normes et standards en TI pour l'apprentissage, l'éducation et la formation. Repéré sur <http://www.gtn->

quebec.org/rapport/ontologie-de-description-et-vocabulaire-de-metadonnees-pour-les-scenarios-pedagogiques.

Maniez, J. (1999). Des classifications aux thésaurus : du bon usage des facettes. *Documentaliste-Sciences de l'information*, juillet-septembre, vol.36, no 4-5, p. 249-262

Noy, N. F. McGuinness. D. L. (2002). Développement d'une ontologie 101 : Guide pour la création de votre première ontologie. Université de Stanford, Stanford, CA, 94305 (traduit par A. Angieli). BnF, Bureau de normalisation documentaire. Repéré sur <http://www.bnf.fr/documents/no-DevOnto.pdf>

Otlet. P. (1934). *Traité de documentation*. Editiones Mundaneum,. Repéré sur http://fr.wikisource.org/wiki/Trait%C3%A9_de_documentation

Ranganathan's Monologue on Melvil Dewey (1964). Repéré sur http://www.hyperorg.com/misc/ranganathan_on_dewey_transcript.html

ISKO. (2006). Ranganathan. S. R. Colon classification: an outline with examples. (Italia. Documenti). Repéré Disponible sur <http://www.iskoi.org/doc/colon.htm>

Salvan P. et Piondron P. (1957). *Les systèmes de sélection*. Repéré sur <http://bbf.enssib.fr/consulter/bbf-1957-06-0455-001>

UNESCO, UNISIST (1973). Principes directeurs pour l'établissement et le développement de thesaurus monolingues. <http://unesdoc.unesco.org/images/0000/000059/005951FB.pdf>

UNESCO, UNISIST (1980). Principes directeurs pour l'établissement et le développement de thesaurus multilingues <http://unesdoc.unesco.org/images/0003/000392/039209Fb.pdf>

Varet-Pietri M.-M. (2000). *L'ingénierie de la connaissance: la nouvelle épistémologie appliquée*, Volume 696. Presses Universitaires Franc-Comtoises.