



HAL
open science

L'avenir des langages documentaires dans le cadre du Web sémantique : conception d'un thésaurus iconographique pour le Petit Palais

Béatrice Pierre

► **To cite this version:**

Béatrice Pierre. L'avenir des langages documentaires dans le cadre du Web sémantique : conception d'un thésaurus iconographique pour le Petit Palais. domain_shs.info.docu. 2010. mem_00575053

HAL Id: mem_00575053

https://memsic.ccsd.cnrs.fr/mem_00575053v1

Submitted on 9 Mar 2011

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

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



CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE pour obtenir le
Titre professionnel "Chef de projet en ingénierie documentaire" INTD
Niveau I

Présenté et soutenu par
Béatrice PIERRE

Le 18 octobre 2010

L'avenir des langages documentaires dans le cadre du
Web sémantique : conception d'un thésaurus
iconographique pour le Petit Palais

Jury
Cécile Kattnig
Claire Boisseroles
Valérie Fours

Cycle supérieur Promotion XL

Remerciements

Je tiens à remercier l'équipe de documentation du Petit Palais, et tout particulièrement Claire et Valérie qui m'ont très bien accueillie et qui m'ont permis de réaliser mon stage dans d'excellentes conditions.

Je remercie Elizabeth Gava pour ses précieux conseils, sa patience et son optimisme constant.

Merci également à Cécile Kattnig pour ses conseils et ses recherches personnelles lors de la rédaction de ce mémoire.

Merci à Sylvie Dalbin et Gérard Régimbeau qui ont accepté d'apporter leur aide dans mes recherches.

Merci à Nicolas James de m'avoir présenté une partie de son travail et donné l'autorisation d'en publier une partie dans ce mémoire.

Merci aux personnels des musées et de la RMN de m'avoir ouvert leurs portes et partagé avec moi une partie de leur travail sur l'indexation des images.

Enfin, un grand merci à Corinne Welger-Barbosa de m'avoir accordé une entrevue malgré son emploi du temps surchargé. Ce fut un grand plaisir et privilège de pouvoir la rencontrer.

Notice

Résumé :

Ce mémoire est une étude sur les langages documentaires et leur utilité pour structurer les informations dans l'idée d'un Web sémantique. Les nouveaux modes d'indexation collaborative, folksonomies et tags, qui ont vu le jour avec le Web 2.0 sont également abordés, notamment pour illustrer une partie consacrée aux musées et leur usage du (des) réseau(x). La dernière division de ce travail est consacrée à l'utilisation des langages documentaires dans les institutions patrimoniales et la réalisation d'un thésaurus iconographique pour le musée du Petit Palais.

Mots-clé :

Thésaurus, image, langage documentaire, web, web sémantique, folksonomie, indexation, musée, thésaurus iconographique.

Table des matières

Remerciements.....	2
Notice	3
Résumé :	3
Table des matières.....	4
Introduction.....	8
Première partie : Définition et évolution des langages documentaires	11
1 Les langages de type classificatoire	12
1.1 Classifications.....	12
1.1.1 La classification décimale Dewey.....	Erreur ! Signet non défini.
1.1.2 La classification décimale universelle.....	12
1.1.3 La classification de la librairie du Congrès.....	13
1.2 Les taxonomies ou taxinomies	15
1.3 Les ontologies.....	17
2 Les langages de type combinatoire.....	20
2.1 Le répertoire de vedettes-matière.....	20
2.1.1 Définition.....	20
2.1.2 Le RVM de la bibliothèque de l'université de Laval.....	20
2.1.3 RAMEAU.....	22
2.2 La classification à facettes.....	23
2.3 Le thésaurus.....	25
2.3.1 Origine et définition.....	25
2.3.2 Méthodologie d'élaboration d'un thésaurus (monolingue).....	28
2.3.3 Les normes.....	32
3 Théorie de l'indexation	35

3.1	Principes généraux de l'indexation	35
3.2	Indexation de l'image fixe.....	36
	Deuxième partie : Vers un Web sémantique	38
1.	Les enjeux du Web sémantique.....	39
1.1	Présentation.....	38
1.2	L'usage des ontologies dans le web sémantique.....	39
1.3	Les applications du Web sémantique.....	41
1.3.1	Projet SKOS.....	41
1.3.2	Topics Maps ou cartes conceptuelles.....	42
1.3.3	Le TAL.....	44
2.	Adaptation des langages documentaires au Web 3.0 et interopérabilité	48
2.1	Les technologies du Web sémantique.....	47
2.2	Le multilinguisme comme facteur d'interopérabilité.....	49
3.	Emergence d'un nouveau type d'indexation	53
3.1	Les différents grades de l'indexation collaborative.....	52
3.2	L'intérêt et l'usage des folksonomies.....	52
3.3	L'indexation collaborative sur le site Flickr.....	54

3.4 Limites de folksonomies.....	55
3.5 Bilan sur les folksonomies.....	56
4. Les musées et leur usage du Web	58
4.1 Qu'est-ce que la muséologie ?.....	57
4.2 L'usage du web par les musées.....	58
Troisième partie : Conception d'un thésaurus iconographique pour le compte du Petit Palais.....	63
1 Présentation du contexte.....	65
1.1 Bref historique sur la création du musée	65
1.2 Contexte du projet.....	66
2 Image et représentation d'œuvres d'art	68
2.1 Qu'est-ce qu'une image ?.....	68
2.2 Spécificité de l'image comme représentation d'œuvre d'art.....	69
3 L'utilisation des langages documentaires au sein des musées.....	71
3.1 Point sur la situation.....	71
3.2 Les principaux langages documentaires utilisés par les musées.....	71
3.3 Compte-rendu de visites.....	75
4 Processus d'élaboration du thésaurus iconographique.....	77
4.1 Indexation des représentations d'œuvre	77
4.2 Structure et relations entre les termes.....	78
4.3 Contraintes et difficultés	79
4.4 Complément et mise à jour.....	81
Conclusion.....	82

Bibliographie	84
ANNEXES	98
ANNEXE 1- AGENDA.....	98
ANNEXE 2- TRAVAUX EXPERIMENTAUX DE N.JAMES SUR LES ONTOLOGIES ..	100
ANNEXE 3- INTERVIEW DE CORINNE WELGER-BARBOSA	103
ANNEXE 4- INDEXATION SUR FLICKR	111
ANNEXE 5 – INDEXATION SUR LE SITE STEVE MUSEUM.....	113
ANNEXE 6 INDEX ET THESAURUS ICONOGRAPHIQUE	114

Introduction

« Aucune œuvre d'art ne doit être décrite ni expliquée sous les catégories de la communication », déclara Théodor W. Adorno. Il exprime ainsi le fait qu'une œuvre existe par elle-même, vouloir l'analyser, la classer est par essence un acte réducteur.

Cette théorie, bien que tout à fait légitime sur le plan philosophique, va à l'encontre de l'idée même du métier de documentaliste, qui s'articule autour de trois axes : la collecte d'informations, leur traitement et leur communication. Le professionnel des Sciences de l'information s'inscrit dans une logique plus large de l'organisation et de la diffusion du savoir. Cette discipline, qui est apparue dans les années 40 avec la naissance de la cybernétique, n'a cessé d'être disputée par les scientifiques qui peinaient à en circonscrire les contours institutionnels.

De fait, les objets d'étude des documentalistes et autres professionnels ont souffert, et c'est encore le cas, de cette confusion générale. Les langages documentaires, dans le jargon francophone, ou « controlled vocabularies » dans la version anglaise, en sont l'exemple parfait. Pour cette raison, j'ai choisi dans ce mémoire d'effectuer un travail de définition pour démêler l'enchevêtrement des définitions et acceptions de ces outils. En effet, il semblait opportun d'établir leur véritable nature et objectif avant d'aborder la question de leur avenir et éventuelle transposition sur le Web.

Dans quelle mesure cette adaptation quelque peu forcée des langages documentaires est-elle un signe de leur obsolescence ? A contrario, doit-on l'envisager comme une reconversion raisonnable, moderniste pour leur permettre d'exister sur la toile ?

La complexité technique et conceptuelle de l'interopérabilité a conduit les spécialistes à envisager de nouveaux moyens pour l'organisation du savoir sur Internet. Le paradoxe de cette nouvelle technologie, qui multiplie les informations sans pour autant que les utilisateurs en aient un réel accès, oblige les intellectuels (nous entendons ici un collègue de scientifiques, linguistes, philosophes ...) à penser de nouveaux systèmes de représentation et de structuration, tels que les ontologies ou les Topics Maps. Elles constituent un tremplin pour assurer l'avenir (virtuel) des langages documentaires, en particulier dans l'optique du Web sémantique (ou web 3.0).

En même temps, l'avènement récent d'un web 2.0, appelé plus communément web collaboratif, a conduit à l'apparition de nouveaux modes d'indexation, des textes comme des images. Chaque internaute peut devenir auteur et se doit de référencer lui-même sa production d'information ou celle d'autrui via les sites de partage, notamment grâce aux tags et aux folksonomies. Ces nouveaux modes de communication libre laissent poindre quelques interrogations : cette nouvelle typologie de langages vient-elle concurrencer les thésaurus et classifications ou les complètent-elles ? Quel langage pour quel usage ? Internet peut-il être à la fois libre et structuré ?

Ces bouleversements technologiques ont amené le milieu culturel à se remettre en question et à revoir ses pratiques. Les musées et autres institutions patrimoniales ont peu à peu tissé leur toile sur le Net. Certes, l'on peut encore observer une certaine réserve sur le fait de rendre accessible à tous des œuvres et un savoir prestigieux mais on ne peut nier un réel progrès et bond en avant d'un milieu longtemps considéré comme assez discret et peu porté sur les considérations techniques et technologiques.

Cela m'a amenée à nous intéresser au cas des musées, en particulier le Petit Palais qui, mû par une quantité de projets, a l'ambition de valoriser sa collection d'œuvres. Ainsi, la nécessité de réaliser un thésaurus iconographique, destiné en premier lieu pour le musée lui-même, mais à terme pour tous ceux de la Ville de Paris, s'avère utile dans ce contexte. Si, de nos jours, la question de l'indexation de l'œuvre d'art est moins confidentielle, il a toutefois été difficile de constituer une bibliographie conséquente à ce sujet. C'est pourquoi un panorama des principaux langages documentaires utilisés dans les musées a été ainsi brossé et donne lieu ensuite à une méthodologie personnelle, en tout point perfectible il est sûr, pour élaborer ce thésaurus iconographique.

**Première partie :
Définition et évolution des
langages documentaires**

1 Les langages de type classificatoire

1.1 Classifications

1.1.1 La classification décimale Dewey

La CDD doit son nom à son créateur, Melvil Dewey, qui l'a mise au point en 1873. Il fut également à l'origine du développement de la bibliothèque de l'université Columbia à New York, de la fondation de la première école de bibliothéconomie et de l'American Library association. Tout comme la CDU, la classification Dewey comporte dix classes et chaque classe est elle-même subdivisée en dix autres classes. Sa notation est donc très simple, ce qui en fait un outil facile à manipuler et à adapter. Elle est de loin la plus connue et la plus répandue à travers le monde : environ 135 pays sur les cinq continents l'ont adoptée. On compte à ce jour une trentaine de traductions, ce qui en fait un système vivant et ouvert sur le monde. Cependant, du fait de sa popularité, la classification Dewey est également la plus critiquée. Ses détracteurs lui reprochent d'adopter une vision du monde très anglo-saxonne et donc un certain manque d'objectivité. Sa macrostructure constitue également un handicap car elle ne peut pas convenir à bon nombre d'institutions qui possèdent une collection très spécialisée ou pointue dans un domaine. Enfin, les notions d'interdisciplinarité et multidisciplinarité sont peu présentes voire inexistantes et il serait de bon ton d'examiner ces manques pour faire de la classification Dewey un outil complet et efficient sur le web.

1.1.2 La classification décimale universelle (CDU)

La CDU date de 1905. Deux avocats belges, Paul Otlet et Henri Lafontaine, également fondateurs de l'Institut International de Bibliographie, mirent au point ce système de classification des bibliothèques. Inspiré de la Dewey, Il est composé de neuf classes, comme suit :

- 0 Sciences et connaissances
- 1 Philosophie et psychologie
- 2 Religion et théologie
- 3 Sciences sociales, statistiques et économie
- 4 inoccupée
- 5 Mathématiques, sciences exactes et naturelles
- 6 Sciences appliquées, médecine et technologies
- 7 Art, divertissement et sport

- 8 Langue, linguistique et littérature
- 9 Géographie, biographie et histoire

Le chiffre 0 désigne toujours les généralités, donc le code 20 renvoie à des généralités sur la religion.

Cette classification est très peu connue en Amérique du Nord et d'une manière générale on peut dire qu'elle est exclusivement utilisée dans les pays européens non anglophones (Belgique, France, Espagne¹ ...), bien qu'à l'origine, elle avait pour vocation d'être multilingue. La CDU a pour avantage d'avoir un système de notation assez simple et flexible et peut donc être employée tant dans des domaines très spécialisés que dans des bibliothèques publiques ou universitaires. Elle convient à la fois pour classer des collections d'objets ou de documents. Bien qu'elle soit dotée d'un niveau de précision plus élevé que la Dewey et la LCC, la CDU connaît depuis quelques années une certaine régression compte tenu du fait l'irrégularité de ses mises à jour. Par ailleurs, elle ne bénéficie pas de caution particulière, comme le fait d'être l'outil de référence d'une grande bibliothèque, ce qui lui permettrait d'asseoir sa légitimité. Les moyens financiers investis pour cette classification ne sont pas non plus suffisants pour envisager des études lui assurant une certaine pérennité. Cependant, il est nécessaire d'indiquer une collaboration entre l'équipe de la CDU et celle de la Dewey

1.1.3 La classification de la librairie du Congrès

Elaborée entre 1890 et 1901, cette classification a pour objectif d'organiser le gros volume de documents détenus par la bibliothèque du congrès. Elle est répartie sur 21 classes représentée chacune par une lettre de l'alphabet :

¹ Etude réalisée en 2004 par UDC Consortium (qui détient les droits d'auteur de la CDU et assure son développement)

- A : Généralités
- B : Philosophie, psychologie, religion
- C : Sciences auxiliaires de l'histoire
- D : Histoire générale et histoire de l'Europe
- E : Histoire de l'Amérique (généralités et États-Unis)
- F : Histoire de l'Amérique (autres pays d'Amérique)
- G : Géographie, anthropologie, loisirs
- H : Sciences sociales
- J : Sciences politiques
- K : Droit
- L : Éducation
- M : Musique et musicologie
- N : Beaux arts
- P : Langage et littérature
- Q : Sciences
- R : Médecine
- S : Agriculture
- T : Technologie
- U : Science militaire
- V : Science navale
- Z : Bibliographie, Sciences de l'information et des bibliothèques

La LCC est une classification alphanumérique car pour les subdivisions, les lettres sont accompagnées de chiffres. Cela en fait un système très précis et plutôt convivial. Elle remporte un certain succès aux États-Unis mais est très peu utilisée en dehors de ses frontières. Cela peut s'expliquer par le fait que la classification ne s'appuie sur aucune source comme une encyclopédie ou autre outil de connaissance fiable contrairement à la Dewey qui avait bénéficié de toute la richesse culturelle de son auteur. Il peut également lui être reproché un certain américano-centrisme car elle reflète avant tout les préoccupations des États-Unis. Par exemple, la notion d'indiens d'Amérique du Nord évoque la situation des indiens sur le sol américain mais n'inclut pas forcément les populations indigènes du Canada, les Inuits, alors que ces derniers sont pourtant des indiens d'Amérique du Nord.

Outre l'absence de contrôle terminologique et syntaxique, certains pans de la classification sont totalement obsolètes et mériteraient une révision totale. Il n'y a pas non plus de notation hiérarchique et par conséquent l'organisation du savoir n'en est que partielle. L'index, qui aurait été un outil de navigation supplémentaire, est totalement inexistant. Par ailleurs, son évolution s'avère très lente mais toutefois il est possible de rajouter des termes par le biais de chiffres cutter et la classification reste stable.

Au début des années 1980, des travaux d'automatisation avaient été entrepris mais l'arrivée d'Internet a obligé l'équipe de maintenance à tout revoir et à s'interroger sur les processus de compatibilité et d'interopérabilité, à commencer par celle entre la classification elle-même et le répertoire de vedettes-matière LCSH qui provient de la même institution. Pour le moment, la classification de la librairie du congrès n'est disponible qu'en anglais.

1.2 Les taxonomies ou taxinomies

Le terme taxonomie fut créé en 1813 par le botaniste Suisse Augustin Pyrame de Candolle dans son ouvrage « *Théorie élémentaire de la botanique ou exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux* ». Si l'on s'intéresse à l'étymologie latine, taxonomie signifie « règle de classement ».

Emile Littré, lexicographe et philosophe français, a contesté la graphie proposée par Candolle et proposa le terme de « taxinomie ». Néanmoins, la première acceptation est la plus répandue, notamment parce qu'elle a été validée par la littérature anglophone.

La taxonomie peut également être définie comme ceci : « Classification systématique et hiérarchisée des taxons dans diverses catégories selon les caractères qu'ils ont en commun, des plus généraux aux plus particuliers ² ». La taxonomie est régie par des relations de subsomption, c'est-à-dire qu'un élément fait forcément partie d'un autre. On peut les représenter sous forme d'un arbre et chaque nœud est un taxon. Par exemple, les chats font partie de la famille des félins qui eux-mêmes sont apparentés à l'espèce animale.

Cependant, nous ne pouvons nous contenter d'une seule définition du terme de la taxonomie, tout dépend en effet de son contexte d'utilisation. D'un point de vue documentaire, on lui reconnaît déjà plusieurs sens mais elle peut tout à fait être envisagée dans d'autres disciplines. Nous n'en ferons le cas ici et nous concentrerons essentiellement sur son utilisation dans le champ des Sciences de l'information. Pour Danièle Dégez et Dominique Ménillet³, la taxonomie est la classification des formes vivantes. Alan Gilchrist⁴ propose cinq acceptations qui peuvent toutefois se chevaucher :

1) Taxonomie comme répertoire de site Web : Elle est souvent utilisée sur Internet et en augmentation constante sur les Intranets. Le site Dmoz est un très bon exemple de cet usage fait des taxonomies. Il n'y a pas de niveau hiérarchique et les termes peuvent être répétés indéfiniment d'une classe à une autre, ce qui offre à l'utilisateur plusieurs voies de navigation.

2) Taxonomie pour appuyer l'indexation automatique : Elle peut être cachée derrière une classification supportant des algorithmes comprenant eux-mêmes des ensembles de mots, des syntagmes, des synonymes ou encore des instructions. Cette base de règles est utilisée pour extraire automatiquement les termes d'indexation qui peuvent ou

² Définition de la taxonomie par l'URFIST

³ Danièle Dégez et Dominique Ménillet, *Thésauriglossaire des langages documentaires : un outil de contrôle sémantique* », p 41

⁴ Alan Gilchrist, *Thesauri, taxonomies and ontologies*, *Journal of documentation*, p 10 à 12

non être présents dans les documents. Cette approche est intéressante dans la mesure où la prolifération des informations rend impossible une indexation 100% manuelle. On peut supposer que le site du BHV cache une taxonomie pour proposer sa très vaste quantité de produits aux usagers.

3) Taxonomie créée par catégorisation automatique : Cette définition rejoint la première dans la mesure où il est question de générer automatiquement des catégories pour classer de grandes quantités de documents et la présentation ressemble à celle des répertoires de site web cité auparavant. Les logiciels utilisés pour aboutir à ce résultat sont globalement les mêmes, bien que dans ce cas présent, l'intervention humaine permet de les augmenter.

4) Taxonomie comme filtre de recherche : Ici, une taxonomie est aussi bien créée, importée ou utilisée dans une formule de requête. Les homographes peuvent être débarrassés de leur ambiguïté et les synonymes regroupés entre eux. L'utilisateur peut naviguer dans les hiérarchies et passer d'un terme à un autre. Cette taxonomie est finalement un thésaurus qui a été adapté aux besoins des utilisateurs et leur permet une navigation fonctionnelle.

5) Taxonomie dans l'entreprise : La taxonomie est une des solutions envisageables pour les entreprises de permettre l'accès de leurs informations à leur personnel. Dans un contexte professionnel où les informations proviennent de sources variées, externes ou internes, le besoin d'ordonner pour en faciliter l'accès devient pressant. Ces taxonomies peuvent être perçues comme un système hybride de thésaurus et d'ontologies, un nouveau type de taxonomies dont l'appellation anglophone exacte est « corporate taxonomies ». Ce nouveau procédé pourrait être adapté au fonctionnement des moteurs de recherche et servir de modèle pour l'indexation.

La taxonomie utilise à la fois les techniques des classifications et des thésaurus mais pour autant on ne doit pas confondre ces trois notions. Il y a des similarités entre chaque technique dans la manière d'indexer mais les degrés de granularité font la différence.

Selon la société LINGWAY⁵, la taxonomie est « un réseau sémantique dans lequel une seule relation est la relation hiérarchique ». Cette définition nous montre bien que si la

⁵ Société spécialisée dans les moteurs de recherche sémantique

taxonomie peut parfois être apparentée au thésaurus, elle n'en possède que certaines caractéristiques, l'amalgame n'a donc pas lieu d'être.

Enfin, d'après l'ADBS⁶, la taxonomie est « un cadre d'organisation pour des ressources numériques de toutes natures (et pas seulement documentaires – en cela les taxonomies ne sont pas toujours des langages documentaires) destiné à en permettre une présentation ordonnée et y donnant accès par navigation hypertextuelle. »

1.3 Les ontologies

S'il y a un terme qui prête particulièrement le flanc à la controverse, c'est bien le terme « ontologie(s) ». Il est parfois utilisé à tort et à travers et il n'est pas toujours aisé de comprendre son sens véritable et l'utilisation qui en est faite dans le champ des sciences de l'information et dans le domaine informatique. Cela tient au fait que des acteurs issus de ces disciplines se sont approprié ce concept en s'éloignant de son origine, la philosophie.

Dans le domaine philosophique, l'ontologie étudie l'être en tant que tel, ses propriétés intrinsèques. Elle doit sa paternité à Parménide, philosophe grec ayant vécu au Vème siècle avant JC. Il y a une dimension métaphysique dans cette étude. L'intérêt pour les ontologies comme dispositif de gestion de connaissances ne date pas d'hier. Déjà, dans les années 80, quelques chercheurs tels que Barry Smith avait conclu à ce rapprochement. Vickery, en 1997, a également œuvré dans ce sens et a souligné l'émergence des ontologies dans l'ingénierie des connaissances. Depuis, la littérature spécialisée ne cesse de s'y intéresser et les articles évoquant les ontologies sont en constante augmentation.

Pour couper court aux contestations au sujet du sens du terme « ontologie », deux variations du terme ont été acceptées. D'un côté, il y a l'ontologie, au singulier, qui désigne toujours la discipline philosophique, et de l'autre, les « ontologies », au pluriel, et qui sont considérées comme des objets techniques et opérant dans divers univers tels que l'intelligence artificielle, la linguistique ou encore l'ingénierie des connaissances.

⁶ Association des professionnels de l'information et de la documentation. L'équipe chargée de l'étude des langages documentaires avait réalisé en 2007 (Vol 44 N°1) un dossier spécial sur ce sujet. Voir page 23

Dans le champ même de l'ingénierie des connaissances, le terme « ontologie » peut revêtir plusieurs significations. Tantôt les ontologies sont définies⁷ comme « spécification formelle, explicite, d'une conceptualisation partagée », ou « une ontologie est parfois appelée ontologie formelle uniquement pour nous référer à un champ de recherche philosophique ».

Le rapprochement avec les langages documentaires s'explique par le fait que les ontologies servent à normaliser un corpus de termes et à expliciter les liens sémantiques qui les lient.

Voici un exemple :

Représentant de :		Se joue avec :	
football	sport	football	ballon
tennis	sport	tennis	balle
badminton	Sport	badminton	volant

☞ Le football est un sport et on y joue avec un ballon

Les ontologies seraient donc un cadre conceptuel ou un mode de représentation dans lequel les thésaurus ou les taxonomies pourraient être englobés mais ne sauraient être des langages documentaires à eux seuls, comme semble l'affirmer Michèle Hudon⁸ : « Au fil des ans se sont développés des langages documentaires de type catégoriel ou classificatoire (schémas de classification, taxinomies, ontologies) et des langages de type combinatoire (répertoires de vedettes-matière, thésaurus, liste de mots-clés contrôlés) ».

Selon Jacques Chaumier⁹, « une ontologie fournit le vocabulaire spécifique à un domaine de la connaissance et, selon un degré de formalisation variable, fixe le sens des concepts et des relations qui les unissent ». Il complète cette première définition en stipulant qu'une ontologie comprend une ou plusieurs taxonomies, elles-mêmes subdivisées en classes et sous-classes composées d'instance représentant les individus ou les objets.

Si les ontologies ne sont pas à proprement parler des langages documentaires, en revanche, on peut constater que leur processus d'élaboration est proche de celui des thésaurus. Danièle Dégez¹⁰ en conviendra : « De même, quelle différence entre ontologie et thésaurus ? Lors d'une conférence, une personne affirmait que le Mesh qui est un thésaurus

⁷ Michèle Hudon In : Article de Bruno Menon, Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan », p 24

⁸ Bruno Menon , Les langages documentaires : un panorama, quelques remarques critiques et un essai de bilan », p 25

⁹ Jacques Chaumier, Ontologies, antécédents, aspects techniques et limites, p 81

¹⁰ Danièle Dégez, le thésaurus est un outil de connaissances, Archimag, novembre 2004, n°179, p 15 à 17

médical, était une ontologie ! Cela fait plus chic ! Pour moi, une ontologie est toujours construite sur les bases d'un thésaurus, c'est un outil pour structurer l'information. Mais l'ontologie est davantage liée à un métier, à une activité, et elle comprend davantage de relations sémantiques identifiées, disant que tel concept renvoie à tel autre ou lui est opposé. [...] J'avoue que je suis très mal à l'aise devant ces concepts car c'est souvent le snobisme qui recouvre des concepts anciens par un mot nouveau ». En effet, dans les deux cas, il s'agit de normaliser les termes d'un fonds lexical, d'analyser d'un point de vue linguistique ce fonds, de l'uniformiser sémantiquement et établir les relations entre les mots ou les syntagmes.

Enfin, la principale difficulté qui entache le travail fait sur les ontologies est que les concepts ne sont pas les mêmes pour tout le monde, même au sein d'une même culture. L'exemple qui me semble le mieux éclairer cette théorie est celui de l'argent (la monnaie). Je pense que chaque personne interrogée aura une vision du concept « argent » différente de son voisin et qu'il sera très difficile de les faire s'entendre à ce sujet.

2 Les langages de type combinatoire

2.1 Le répertoire de vedettes-matière

2.1.1 Définition

Un répertoire de vedettes-matière est une liste encyclopédique dont les termes sont liés les uns aux autres par une syntaxe particulière. Il est avant tout conçu pour les bibliothèques dans un objectif de catalogage de leur fonds. Danièle Dégez et Dominique Ménillet¹¹ le définissent ainsi : « Suite ordonnée de termes représentant systématiquement le sujet ou l'un des sujets d'un document, plus particulièrement dans un catalogue de bibliothèque ».

Les répertoires de vedettes-matières les plus connus sont le RDV de la bibliothèque de l'université de Laval, les LCSH¹² de la bibliothèque du congrès et le RAMEAU¹³ de la Bibliothèque Nationale de France.

2.1.2 Le RVM de la bibliothèque de l'université de Laval

La première édition du RVM de la bibliothèque de Laval au Canada, entreprise dès 1946, a vu le jour en 1962. Il est en partie une traduction du LCSH mais nous verrons qu'il n'en est absolument pas une pâle copie mais un outil distinct et personnalisé. Il est d'ailleurs utilisé dans d'autres bibliothèques que celle de l'université, tel que Bibliothèque et Archives Canada qui a érigé le RDV au statut de norme en matière d'indexation en langue française en 1974 et lui verse une allocation non négligeable en compensation. La bibliothèque nationale du Luxembourg a elle aussi adopté ce répertoire de vedettes-matière.

Le succès du RVM s'explique en partie par la grande rigueur et la cohérence mobilisée pour créer et mettre à jour cet outil. Une équipe de six bibliothécaires assure son développement et travaille indifféremment sur chaque vedette, à l'exception faite de la biologie. Tout est passé en revue chaque semaine et par chaque membre de l'équipe, ce qui confère au RVM homogénéité et précision.

¹¹ Danièle Dégez et Dominique Ménillet, thésauroglossaire des langages documentaires : un outil de contrôle sémantique », p 44

¹² Library of Congress Subject Headings

¹³ Répertoire d'autorité matière encyclopédique et alphabétique unifié

D'abord en raison des coûts documentaires importants mais aussi dans un souci de partage des ressources, ils ont rapidement entrepris de traduire les vedettes des LCSH, du moins au début. En effet, il s'est avéré bien plus intéressant de les adapter aux besoins de la bibliothèque et de la culture canadienne que de les traduire stricto sensu. Ainsi, environ quatre mille vedettes n'ont pas d'équivalents en anglais, soit 20% du RVM, afin de répondre aux besoins d'indexation spécifique de la documentation en français. Le Canada reproche aux LCSH un certain américano-centrisme, déjà palpable dans la classification de la bibliothèque du congrès, et prend donc une distance critique par rapport à son modèle.

Afin de diminuer la charge humaine, les responsables ont mis en place un système de traduction automatique de l'anglais au français, 75% des vedettes sont ainsi transcrites et rendent le travail moins fastidieux.

Le répertoire de vedettes-matière de l'université de Laval est un outil moderne et en perpétuelle évolution. Cela se traduit de diverses manières :

- N'importe quel abonné peut suggérer de nouvelles entrées de vedettes-matière
- Les termes sont réexaminés régulièrement et adaptés aux connaissances actuelles. Par exemple, le mot « élasticité » sur le plan de la personnalité est devenu « résilience » car des publications francophones ont validé ce nouveau vocable.
- 70% des vedettes de la liste hebdomadaire sont traitées
- Activité de veille importante : consultation du journal RAMEAU, la rubrique « livres de la semaine » du périodique français « Livres Hebdo », suggestion des indexeurs de la bibliothèque ... Il suffit d'un document pour justifier la création d'une nouvelle vedette.
- Communication directe entre les responsables du RVM et des LCSH depuis 2004
- Longueur d'avance : le RVM respectait déjà les principes édictées par la norme ISO 2788, parue environ trente ans après sa création : forme des vedettes-matière, traitement des synonymes, utilisation des qualificatifs pour distinguer les homonymes, notes explicatives.

Concernant l'emploi du singulier et du pluriel, le RVM reflète davantage les normes anglo-saxonnes. Denise Dolbec l'exprime très explicitement dans son article¹⁴ : « la présence du pluriel ou du singulier dans les vedettes de nom commun, selon qu'elles représentent des

¹⁴ Denise Dolbec, Le répertoire de vedettes-matière : outil du XXIème siècle, p 103

entités dénombrables ou non, demeure sans doute la caractéristique la plus frappante qui distingue le RVM des thésaurus de langue française ». Justement, l'évolution du RVM l'a conduit à ressembler davantage à un thésaurus qu'au moment de sa création. Ses relations sémantiques sont pour ainsi dire similaires : générique, d'instance, partitive, associative et même polyhiérarchique.

Par ailleurs, le répertoire de vedettes-matières de l'université de Laval a adopté un langage plus naturel et pré-coordonné dans un souci de précision, notamment au niveau de la recherche. Il s'est également enrichi d'autres thésaurus tels que l'AAT, le MeSH¹⁵ et le National Library of Medicine.

2.1.3 RAMEAU

Ce répertoire de vedettes-matières a été réalisé en 1980 par la bibliothèque nationale de France. Il est en partie décliné du répertoire du RDV mais connaît toutefois un développement autonome. D'autres structures, telles que les bibliothèques universitaires ou des organismes privés l'utilisent également.

Cependant, l'instauration de ce répertoire ne s'est pas faite sans difficulté. Bien qu'il fut très intéressant pour la BNF de partir d'un langage existant, issu des LCSH qui jouissent d'une certaine notoriété et traduit par le Québec, il y avait encore un travail de formatage à effectuer pour tenir compte des spécificités linguistiques et culturelles françaises.

Une infrastructure institutionnelle a été créée pour assurer au langage RAMEAU un développement cohérent, le centre national RAMEAU, composé de deux entités : comité d'orientation et comité opérationnel. Il existe aussi un *guide d'utilisation RAMEAU* qui assure un usage correct de ce langage et deux fois par an, le journal *des créations et des modifications* informe les utilisateurs des évolutions du répertoire.

RAMEAU est un langage d'indexation pré-coordonné, composé de termes reliés entre eux et d'une syntaxe qui définit les règles de création de vedettes-matières. Il est structuré sur trois niveaux : terminologique, sémantique et syntaxique. Il y a les termes retenus, nommés vedettes-matières, et les termes exclus. Cet ensemble constitue une « grappe terminologique¹⁶ » qui fait la richesse du langage RAMEAU.

Néanmoins, sa précoordination assez sophistiquée, instituée pour exprimer des notions assez complexes, a posé quelques problèmes pour l'indexation et a conduit le centre

¹⁵ Medical Subject Headings

¹⁶ Michel Mingam : <http://bbf.enssib.fr/consulter/bbf-2005-05-0038-001>

national RAMEAU à réviser sa politique syntaxique afin de la simplifier, du moins en surface, car le volume sans cesse croissant de documents obligeait à revoir l'indexation de milliers de documents, ce qui aurait occasionné un travail titanesque.

Le projet phare en ce début de XXIème siècle pour RAMEAU est de modéliser des ontologies entre domaines et sous-domaines, nourries de relations sémantiques qui existent déjà entre les termes.

Pour mener à bien cette entreprise, il sera nécessaire de s'éloigner des modèles de base, le RVM et les LCSH, sans toutefois en renier l'héritage.

2.2 La classification à facettes

Ranganathan, bibliothécaire indien, a finalisé la première classification à facettes, nommée la Colon Classification (CC), en 1933. « Colon » désigne en français le signe « : », utilisé pour concaténer les différents indices servant à indexer le document.

L'objectif était de trouver une alternative aux classifications classiques hiérarchisées, telles que la Dewey ou la LCC. Dans le cas de la CC, les documents sont représentés en quelque sorte par une combinaison de concepts. Le bibliothécaire a créé cinq facettes, fondatrices et emblématiques de son système, que l'on appelle aussi le PMEST :

- Personnalité ☞ sujet ou concept principal du document
- Energie ☞ opération ou action subie par le sujet
- Matière ☞ substance ou une propriété
- Espace ☞ localisation géographique
- Temps ☞ localisation chronologique ou temporelle

Cette classification comporte de nombreux avantages¹⁷. Premièrement, elle établit des liens fonctionnels entre les sujets et cette vision pluridimensionnelle semble apporter plus de cohérence que les schémas classiques qui reposent sur la hiérarchie et la synonymie. Justement, cette absence de hiérarchie induit que les concepts sont indépendants les uns des autres. Il n'y a pas non plus de sujet qui ne puisse pas être envisagé grâce à ce système, la CC possède l'atout d'être universelle. Grâce à la concaténation des indices, tout sujet, aussi complexe soit-il, peut être exprimé de manière simple et explicite. Par ailleurs, ce langage, à la base en mode post-coordonné, peut aussi bien fonctionner en pré-coordonné. La flexibilité pour ordonnancer les termes, les facettes ou les indices de classement en font un outil très complet pour représenter les sujets.

¹⁷ Michèle Hudon, Le passage au XXIème siècle des grandes classifications documentaires, p 94

Il n'empêche, malgré tous ses atouts, que la Colon classification intéresse assez peu les professionnels de la bibliothéconomie. Le système de notation mixte et les procédures d'application les rebutent, même s'ils jugent que c'est un outil plutôt stimulant sur le plan intellectuel. Selon Jacques Maniez¹⁸, « la formule PMEST est le point faible de ce système et que le statut hybride qu'il réserve aux facettes a contribué dès le départ à brouiller la perception de ce concept en documentation ». Même en Inde, le pays qui l'a vue naître, une infime quantité de structures en fait usage et les rares à s'y oser le font sans discernement ni à bon escient. La septième et dernière révision date de 1987. Elle n'est quasiment pas remise à jour et de manière très sporadique. En conséquence, elle ne bénéficie pas de soutien institutionnel qui lui permettrait de vivre, ou plutôt de survivre.

Dans les pays anglophones, dont l'Inde, beaucoup de chercheurs ont repris le travail de Ranganathan en tentant de le simplifier. Ainsi, Bhattacharya, théoricien indien, a inventé en 1979 le système POPSI¹⁹ inspiré de la CC mais dont il a corrigé les anomalies. Les facettes sont conservées mais complétées d'une syntaxe. Le chercheur a retenu quatre grandes familles :

- Discipline : facette de base
- Entité : remplace « personnalité » dans la CC
- Action : remplace « énergie » dans la CC et comprend deux branches : l'action réflexive (Ex : respirer) et l'action externe tel que boire ou chanter.
- Propriété : remplace « matière » de la CC et symbolise les propriétés des entités ou des actions

Dans le système POSPI, des règles syntaxiques ont été édictées. Par exemple, lorsqu'un sujet est subordonné à un autre, il doit le suivre directement dans la formule d'indexation. En ce qui concerne les facettes, il convient de respecter l'ordre suivant : discipline, entité, propriété, action.

Les classifications à facettes ont fait l'objet de nombreuses recherches, notamment par le Classification Research Group. Brian Vickery²⁰, a d'ailleurs publié un ouvrage à ce sujet, intitulé « Les classifications à facettes », publié en France en 1963.

¹⁸ Jacques Maniez, Du bon usage des facettes, p 253

¹⁹ Postulate-Based Permuted Subject Index

²⁰ Professeur anglais émérite qui a longtemps travaillé sur les classifications mais aussi sur le web sémantique. Il a fait partie du CRG, d'ISKO en 2008 puis fut également le directeur de l'école de bibliothèque et d'études archivistiques en Angleterre de 1973 à 1983. Il est décédé le 17 octobre 2009.

D'autres auteurs, tels que Jacques Maniez, font le rapprochement entre les facettes et les thésaurus et Jean Aitchison²¹ est connue pour avoir créé le néologisme « thesaurofacet » pour symboliser l'union de ces deux langages d'indexation.

2.3 Le thésaurus

2.3.1 Origine et définition

A l'origine, le terme thésaurus vient du grec ancien et signifie « trésor ». Sa forme latine est *thesaurus*, sans accent. Si les deux orthographes sont admises, la première semble être la plus répandue dans la littérature francophone. Les anglo-saxons privilégient la version latine et son corollaire au pluriel : *thesaurus/thesauri*. Le thésaurus est réellement apparu au XVI^{ème} siècle grâce à Robert Estienne²² et son *thesaurus lingua latinae* paru en 1531. Cet ouvrage avait pour objectif de dresser un panorama de la latinité depuis ses débuts jusqu'à Isidore de Séville. Ensuite, Au XVII^{ème} siècle, l'évêque John Wilkins créa un type de dictionnaire amélioré constitué de relations entre les mots et destiné à organiser les objets du référentiel d'histoire naturelle de la royal Society. En 1852, Peter Mark Roget²³ élaborait aussi un thésaurus qu'il qualifia de « classification verbale », basé sur l'œuvre philosophique de Leibniz. Les thésaurus ne furent l'objet de normes qu'à partir des années 70 et connurent un certain succès dans les années 80.

Il existe plusieurs définitions du mot « thésaurus », que ce soit dans la littérature francophone ou anglophone, mais d'une manière générale toutes concordent vers le fait qu'il s'agit d'un vocabulaire contrôlé, outil d'indexation et d'aide à la recherche, comprenant des relations sémantiques entre les termes. Voici ce que nous proposent Danièle Dégez et Dominique Ménillet dans leur thésauroglossaire²⁴, l'ouvrage que j'ai choisi comme référence en matière de définition : « Liste organisée de termes contrôlés et normalisés (descripteurs et non-descripteurs) servant à l'indexation des documents et des questions dans un système documentaire. Les descripteurs sont reliés par des relations sémantiques (génériques, associatives et d'équivalence), exprimés par des signes ou symboles conventionnels. Les synonymes (non-descripteurs ou termes interdits) sont reliés aux descripteurs par une seule

²¹ Professeur de langue et communication à la faculté de langue anglaise et littérature à l'université d'Oxford.

²² Michèle Hudon, *Le thésaurus, conception, élaboration, gestion*, p 16

²³ Physicien, théologien et lexicologue britannique (1779/1869)

²⁴ Danièle Dégez et Dominique Ménillet, *Thésauroglossaire des langages documentaires*,

relation d'équivalence. L'OED²⁵ définit le thésaurus comme suit : « Une liste classifiée de termes, plus spécialement des mots-clés, dans un champ particulier, pour un usage d'indexation et de recherche d'information ».

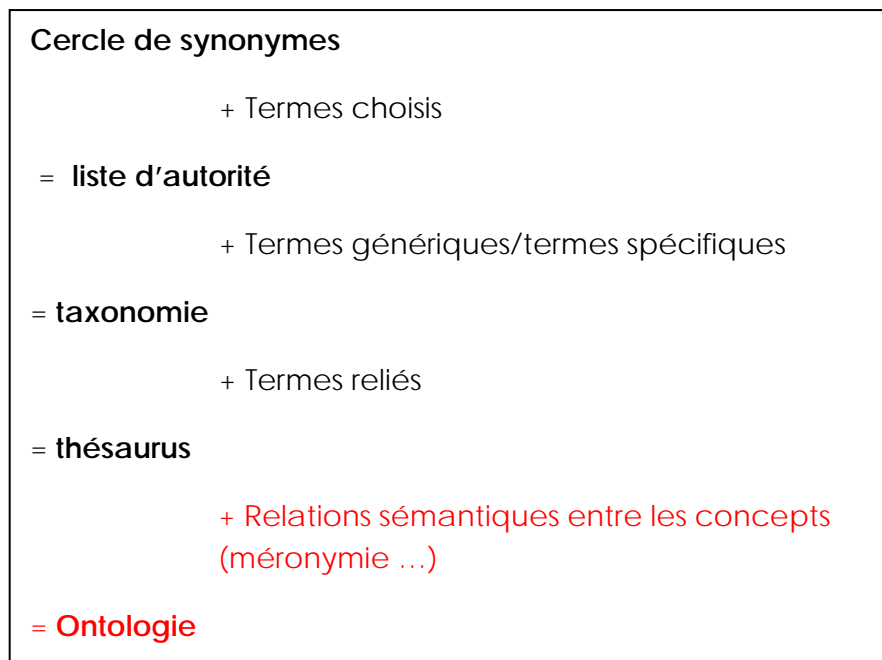
Malgré l'abondance des définitions données par les uns et les autres, le terme de thésaurus demeure assez vague. Les frontières entre les différents types de langages documentaires ont tendance à s'estomper et certains professionnels des sciences de l'information ont parfois tendance à faire l'amalgame entre plusieurs notions :

- **Thésaurus et dictionnaire** : le thésaurus est souvent perçu par les non-spécialistes comme un dictionnaire. La différence est assez simple, le thésaurus n'a pas pour vocation d'être aussi exhaustif et le dictionnaire ne comporte pas de relations entre les termes excepté la synonymie. Dans un dictionnaire, le mot est le point de départ et on en dégage plusieurs concepts, dans le cas du thésaurus, c'est l'inverse, le concept est décrit par plusieurs termes.
- **Thésaurus et glossaire** : le but d'un glossaire est d'apporter des définitions à des termes dans un domaine donné. Même si un thésaurus est lui aussi un ensemble de vocables rattaché à une spécialité, il n'y a pas d'obligation de spécifier le sens de certains mots par une définition, même si les auteurs sont formellement invités à le faire.
- **Thésaurus et liste de vedettes-matière** : la distinction principale entre ces deux notions réside dans le fait qu'une liste de vedettes-matière a été conçue pour cataloguer une collection documentaire et qu'un thésaurus a pour objectif d'indexer un corpus de documents. L'unité de base de la liste est la vedette, soit un sujet qui regroupe plusieurs concepts tandis que celle du thésaurus est le descripteur qui ne renvoie qu'à un seul concept. En outre, la liste de vedettes-matière est essentiellement destinée au bibliothécaire tandis que le thésaurus offre un double usage documentaliste/usager.
- **Thésaurus et taxonomie** : la différence est bien plus ténue. L'on peut dire que le degré de précision sera l'élément clé pour distinguer les deux langages documentaires. Un thésaurus est supposé être bien plus riche qu'une taxonomie qui n'est à la base qu'une classification.
- **Thésaurus et ontologie** : La littérature grise traite souvent les similarités entre les deux objets, ce qui est tout à fait censé, mais rarement des différences. Les ontologies sont d'abord présentées comme un prolongement du thésaurus et une manière de le rendre viable et fonctionnel sur le web. Dès lors, une ontologie serait une

²⁵ Oxford English Dictionary

version de thésaurus bien plus sophistiquée, avec une vision pluridimensionnelle des termes.

Finalement, à y regarder de plus près, il n'y a qu'une notion de granularité entre chaque langage documentaire. Le schéma récapitulatif de Fred Leise²⁶ m'a paru assez simple pour synthétiser les nuances. Toutefois, il m'a paru possible de le compléter (partie en rouge) car il me paraissait inachevé :



A l'aide à ce schéma, on réalise pourquoi le thésaurus est le langage documentaire le plus complexe à réaliser, compte tenu de toutes les relations qui le composent.

²⁶ Fred Leise, Controlled vocabularies : an introduction, p 124

2.3.2 Méthodologie d'élaboration d'un thésaurus (monolingue)

La réalisation d'un thésaurus étant donc complexe et fastidieuse, la rédaction d'un cahier des charges²⁷ peut s'avérer utile voire même indispensable. Voici la liste des éléments à prendre en compte :

- Le type et le nombre d'éléments à collecter (livres, articles de revues, images fixes ou animées ...)
- La/les spécialité(s) concernée(s) par le thésaurus
- Le public et ses besoins
- Les objectifs
- Les logiciels
- Les ressources disponibles (financières, humaines ...) : il est fortement conseillé de confier cette mission à plusieurs personnes plutôt qu'une dans un souci d'objectivité.
- Le planning (il est important d'assigner à ce projet un délai sous peine de s'enliser et de ne jamais vraiment boucler le projet, même si on considère qu'un thésaurus n'est jamais terminé. Il faut compter environ entre six mois et un an)

Une fois le cahier des charges établi, il existe un certain nombre d'étapes²⁸, celles-ci varient en fonction du type de thésaurus (monolingue, bilingue ou multilingue), à suivre dans un ordre précis par souci de cohérence et de rigueur.

1^{ère} étape - collecte du vocabulaire ou lexique

Au préalable, les concepteurs du thésaurus devront se prémunir d'un certain nombre de documents, voués à les aider dans la réalisation d'un thésaurus. Des thésaurus existants dans le domaine d'expertise ou cas échéant des glossaires, dictionnaires ou banques de données terminologiques constituent une bonne base pour construire cet outil. Certains pourront toutefois objecter qu'un thésaurus trop pointu peut s'avérer inapproprié, la question de cerner le ou les publics est donc ici cruciale. Le degré de précision des termes n'en sera que meilleur. Ensuite, il est indispensable de consulter les dernières normes²⁹ en vigueur dans le domaine des thésaurus. Répondre aux normes offre un cadre rassurant et limite déjà un certain nombre de contestations.

2^{ème} étape – listage des catégories et plan de classement

Une fois le vocabulaire collecté, il convient de classer les termes en plusieurs catégories. Il existe une règle tacite qui veut que le nombre de descripteurs soit divisé par

²⁷ Danièle Dégez, Construire un thésaurus, Archimag N°222 mars 2009, p 44

²⁸ Van Slype, les langages d'indexation, 1986, p 103 à 133

²⁹ Normes AFNOR et ISO qui feront l'objet de la sous-partie suivante.

soixante pour déterminer le nombre de microdisciplines. Ces dernières seront ensuite regroupées autour de six ou douze branches plus générales.

3^{ème} étape – établissement des règles d'écriture et relation entre les mots

- **Structure d'équivalence sémantique :**

Il est ici question de dresser la classe d'équivalence, c'est-à-dire attribuer à chaque concept, une liste de synonymes et de quasi-synonymes. Parmi les termes de cette liste, il faut choisir ceux qui seront descripteurs et non-descripteurs et les lier par un rapport d'équivalence. Cette étape est la plus délicate car le choix d'un terme descripteur pose certaines difficultés. Il doit être le plus « neutre » possible donc le moins ambigu, le plus utilisé dans la profession ou dans l'entreprise. Enfin, on termine en éliminant les mots vides et en rédigeant des notes d'application.

- **Structure hiérarchique :**

Elle établit un rapport de subordination entre les termes et organise la lecture du thésaurus : des termes les plus généraux vers les plus spécialisés. On compte trois sortes de relations hiérarchiques : génériques (genre-espèce), partitives (tout-partie) ou d'instance (qui renvoie un nom propre à une catégorie. Ex : Chabrol TGI réalisateur). Dans le cas où les non-descripteurs seraient très nombreux, il est possible d'instaurer la polyhiérarchie.

Le choix du nombre de niveaux hiérarchiques pose souvent problème. Un thésaurus à trois niveaux est assez avantageux sur le plan de la réalisation mais s'avère peu satisfaisant pour la recherche. A contrario, un thésaurus de dix niveaux peut être intéressant car très spécialisé mais sa conception et sa lecture en seront nettement compliquées. Le compromis se trouve entre six et sept niveaux.

- **Structure associative :**

Elle met en relation des termes ayant des significations proches sur le plan sémantique mais qui ne sont pas liés entre eux par des rapports d'équivalence ou de hiérarchie. On ne peut pas créer d'association entre descripteur et non-descripteur ainsi qu'entre descripteurs de la même chaîne hiérarchique.

4^{ème} étape – test

Cette phase est essentielle pour tester la cohérence du thésaurus et la pertinence des termes choisis. Pour ce faire, il est conseillé de se munir d'un corpus de documents (une centaine environ) que l'on indexera avec les termes du thésaurus. C'est l'occasion de

rajouter des descripteurs ou non-descripteurs, de modifier les relations hiérarchiques, associatives ou d'équivalence. Pour une efficacité optimale du test, il sera demandé aux indexeurs de remplir une fiche indicative sur laquelle ils consigneront les descripteurs les plus fréquemment utilisés et calculeront le ratio entre le nombre de descripteurs utilisés au moins une fois et le nombre total de descripteurs du thésaurus. Ce ratio se situe idéalement entre 0,5 et 2. Plus le ratio est faible, moins le thésaurus est structuré. Par contre, un taux trop élevé est signe que les relations sont trop nombreuses et nuisent à la compréhension. Un trimestre est généralement considéré comme suffisant pour tester le thésaurus.

5^{ème} étape : présentation du thésaurus

Selon les normes en vigueur, le thésaurus doit comporter une introduction présentant les éléments suivants :

- Le nom des auteurs (personne, organisme ...)
- Le contexte et les objectifs du thésaurus
- Le domaine couvert
- Le nombre de descripteurs et non-descripteurs
- La méthode d'élaboration du thésaurus (sources des termes, choix des descripteurs ...)
- Le mode de présentation, la signification des abréviations employées ...
- Le mode d'emploi du thésaurus et de ses différentes parties (table des matières)

Le thésaurus peut être présenté de différentes manières :

- Présentation alphabétique
- Présentation alphabétique permutée
- Présentation thématique
- Présentation hiérarchique
- Présentation graphique

Un thésaurus peut cumuler plusieurs de ces formes, comme c'est généralement le cas.

6^{ème} étape - Validation

Suite à l'évaluation, en fonction des remarques et des changements, l'équipe de réalisation procède à la validation du thésaurus et lance la première édition.

7^{ème} étape- Mise à jour

Un thésaurus doit être régulièrement mis à jour sous peine de devenir rapidement obsolète. Les descripteurs et non-descripteurs feront l'objet d'un examen critique et avant d'ajouter un nouveau terme, il est vivement conseillé de vérifier qu'il n'existe pas déjà en tant que synonyme ou terme associé. Une personne devrait être désignée pour effectuer ce travail car il est souvent observé que si chacun peut ajouter lui-même des termes, le thésaurus comporte alors des doublons ou des triplons, notamment à cause de l'orthographe qui n'est pas toujours respectée. Il arrive également que la hiérarchie ne soit pas prise en compte et les termes sont placés à des niveaux inappropriés. Une trop grande liberté a tendance à « corrompre » l'intégrité du thésaurus. L'acquisition de nouveaux documents, l'évolution du langage dans le domaine étudié sont autant de prétextes pour réviser le thésaurus. Cependant, la mise à jour d'un thésaurus appelle à la prudence. En effet, un seul changement peut en induire d'autres en cascade et ainsi mettre à mal la colonne vertébrale du thésaurus. Par ailleurs, la suppression d'un descripteur ne doit être envisagée que si le terme n'a jamais été utilisé pour l'indexation.

Si nous avons dit auparavant qu'une seule personne devait être en charge pour la mise à jour du thésaurus, rien ne l'empêche, bien au contraire, de mettre à contribution les utilisateurs, le plus régulièrement possible, en leur envoyant un formulaire à compléter. Plus le thésaurus est spécialisé, plus sa mise à jour sera évidemment fréquente.

Les étapes de réalisation du thésaurus sont assez communes aux divers auteurs que j'ai pu lire. Je me suis appuyée sur les écrits de Georges Van Slype³⁰, Michèle Hudon³¹, Danièle Degez³², Michel Lubkov³³, Jean Aitchison et Alan Gilchrist³⁴ pour retracer les différentes phases. On constate toutefois quelques fois des différences notables. Par exemple, M. Lubkov suggère d'établir les relations sémantiques entre les termes avant la catégorisation, j'ai choisi de m'écarter de cette façon de faire car il me paraissait logique que les termes soient classés avant d'être liés entre eux.

³⁰ G. Van Slype, les langages d'indexation, p 103 à 133

³¹ M. Hudon, le thésaurus, conception, élaboration, gestion, p 76 à 80,

³² D. Dégez, construire un thésaurus, p44-45

³³ M. Lubkov, Comment réaliser un thésaurus ?, p 30 à 33

³⁴ J. Aitchison et A. Gilchrist, Construire un thésaurus

2.3.3 Les normes

La première norme en matière de thésaurus, la NF 47-100, a été instaurée en France en 1973, révisée en 1981. Un an après, l'organisme ISO publia la norme internationale 2788, actualisée en 1986. L'Angleterre a, elle aussi, contribué à l'établissement d'un cortège de règles sur les thésaurus avec la BS8723 (anciennement BS 5723 datant de 1979 et revue en 1987). Les Etats-Unis est le pays qui a le plus fait évoluer ses normes. En effet, entre 1974 et 2005, on compte quatre révisions : 1980, 1993, 2003,2005.

Nous étudierons ces normes en nous basant sur sept points :

- **Périmètre d'application** : La norme américaine Z39.19 a étendu son champ d'étude aux *synonym rings*, aux taxonomies et aux lexiques. Ces vocabulaires contrôlés ont en commun avec le thésaurus le même système de choix des termes et des relations qui les lient. La norme anglaise BS 8723 a un spectre un peu plus large car elle comprend les thésaurus bien sûr mais aussi les classifications, les taxonomies, les répertoires de vedettes-matière et les ontologies.

- **Choix des termes** : les normes sont quasiment identiques à ce niveau mais il a été constaté que les normes les plus récentes, telle que ANSI/ISO, présentent plus de cas pratiques, notamment quant à l'emploi des adjectifs ou les notes d'application. Les différents types de concepts sont également bien développés dans les normes anglaises et américaines : chose, matériau, activité etc. En ce qui concerne l'emploi du singulier ou pluriel, tout dépend de la langue utilisée : les anglophones privilégient le pluriel pour tout ce qui n'est pas dénombrable. Ainsi, dans les récentes normes, les éléments du corps humain qui existent par deux ont été mis au singulier et non pas au pluriel comme c'était le cas auparavant. Les objets de musée sont eux-aussi au singulier.

La norme AFNOR stipule que tous les termes doivent être au singulier sauf ceux qui sont naturellement au pluriel dans la langue (ex : travaux publics).

Les noms propres, plutôt ignorés dans les versions précédentes, sont reconsidérés La norme ANSI/ISO s'est inspiré des règles de catalogage relatives aux noms géographiques pour traiter le cas de ces nouveaux concepts. D'une manière générale, c'est cette norme qui s'intéresse le plus aux utilisateurs, en étant plus accessible par le choix du vocabulaire et la variété d'exemples pour chaque situation.

- **Pré et postcoordination** : toutes les normes s'accordent à dire que les termes pré-coordonnés (ou composés) sont difficiles à traiter. Il est tout à fait possible de faire des recherches avec des termes post-coordonnés et des opérateurs booléens. La norme américaine développe davantage cette problématique que les autres. Elle propose de conserver les formes pré-coordonnées que sous certaines conditions : vocabulaire usuel (l'expression existe en tant que tel dans la langue), perte de sens si on décompose l'ensemble de mots, si cet ensemble comporte un nom propre, si le terme distinctif a perdu de son sens ou qu'il y a ambiguïté avec un autre concept, si l'adjectif modificateur ne définit pas une sous-classe ou si le terme principal est un mot vide. La norme ANSI/ISO complète la réflexion en abordant le sujet de la présentation du thésaurus, qu'il soit imprimé ou informatisé, cela change la donne pour la précoordination car le mode de lecture est différent. Par exemple, dans le cas d'un thésaurus sur Internet, l'emploi des termes pré-coordonnés se justifie pleinement car il aide l'utilisateur à naviguer dans la hiérarchie.

- **Relations entre les termes** : elles sont détaillées de manière identique dans les quatre normes. Pour ce qui est de la relation d'équivalence, la norme américaine se distingue par l'évocation des *synonym rings*, ce qui revient à utiliser plusieurs termes semblables mais sans en choisir un préférentiel. Pour ce qui est des relations hiérarchiques, l'Angleterre et les Etats-Unis ont décrit plus finement les relations d'instance et partitives en les illustrant d'exemples. Par ailleurs, ces normes évoquent également la notion de *top term* (terme à la tête de la hiérarchie), absente des normes AFNOR et ISO. En outre, elles innovent en proposant d'autres types de relation, du type cause à effet (partitive) et dénomination/sigle (équivalence).

- **Présentation du thésaurus** : une fois de plus, la norme américaine se démarque des autres normes en proposant un mode de représentation du thésaurus en fonction de l'utilisateur : gestionnaire de langage, indexeur et utilisateur final. Plus exigeante, la norme BS 8723 prévoit de renseigner un certain nombre d'éléments, en sus des relations classiques : code de classification, le terme de tête, la définition et la note historique. Globalement, les dernières normes font cas des nouvelles potentialités offertes par Internet pour présenter les thésaurus, en particulier les liens hypertextes qui permettent une lecture interactive.

- **Construction du thésaurus** : toutes les normes s'entendent sur la méthode d'élaboration mais chacune a ses spécificités. Les logiciels d'aide à la construction de thésaurus sont évoquées, excepté dans les normes AFNOR et ISO, compte tenu de leurs dates d'édition. Il est conseillé aux chefs de projet de consulter les quatre normes pour avoir

- **Interopérabilité des langages contrôlés** : les normes BS 6723 ou ISO 5964 abordent la question de l'interopérabilité bien plus largement que les normes AFNOR ou NISO car plus récentes, elles tiennent compte d'un univers technologique bien plus ouvert qu'il y a une vingtaine d'années. Cependant, les protocoles et les formats n'étant pas encore bien définis en 2005, les normes ne sont pas encore au point en matière d'interopérabilité.

L'évolution galopante des nouvelles technologies, la perspective d'un nouvel Internet, le Web sémantique, a obligé les organisations à revoir leur copie en matière de normes. Jusqu'à présent, elles n'évoquaient que superficiellement les questions d'interopérabilité et de recherche d'information.

En 2007, un projet de norme a vu le jour, l'ISO 25964 reposant sur la norme anglaise BS 8723. Sans doute parce que celle-ci englobait déjà les autres types de vocabulaires et ne s'arrêtait pas purement et simplement à la notion de thésaurus. La première partie de la norme ISO 25964 sera publiée en, mai 2011, la deuxième en août 2013. Elle pose véritablement la question de la recherche d'information et de l'interopérabilité avec les autres langages contrôlés que sont les classifications, les taxonomies, les vedettes-matières et ontologies. Cette nouvelle norme prendra en considération le langage SKOS et OWL, validés par le W3C, instance de référence dans le domaine du web.

Elle concerne à la fois les thésaurus mono et multilingue, à contrario de la France qui avait érigé deux normes distinctes : NF Z47-100 et Z47-101.

Cette nouvelle norme internationale fit l'objet d'un travail commun considérable, mené par deux comités, l'un international et l'autre français. Dans le premier nous retrouvons des figures importantes telles qu'Alan Gilchrist (E.U) ou Michèle Hudon (Canada). Dans le groupe français figurent Sylvie Dalbin (également présente dans le premier groupe), Danièle Dégez, Dominique Ménillet et Michel Mingam.

3 Théorie de l'indexation

3.1 Principes généraux de l'indexation

Les termes « indexer » et « indexation » apparurent pour la première fois vers 1948, soit cinq siècles après l'invention de l'imprimerie par Gutenberg. Pendant très longtemps, l'indexation n'a fait l'objet d'aucune théorie, seul le nombre exponentiel de documents a fini par justifier que l'on s'y intéresse de plus près. L'avènement d'internet et la possibilité d'y trouver des documents en texte intégral a rendu délicate cette question de l'indexation. L'indexation manuelle s'avère être une charge incommensurable sur le plan humain. De nombreux logiciels de traitement automatique virent le jour mais le résultat était dans le meilleur des cas moins bon que ceux réalisés par une main humaine. Comme le dit si justement Muriel Amar³⁵, ce n'est pas tant le mot qui compte que la chose, le concept auquel il renvoie. A l'heure actuelle, les techniques d'indexation automatique se sont améliorés mais comportent encore quelques lacunes.

La notion de lexique pour indexer un document a été unanimement rejetée par les professionnels des sciences de l'information et de la linguistique. La seule valeur admise est celle des langages documentaires dont l'unité de base est le descripteur. Le descripteur renvoie à un concept et donne accès au document ou à un corpus documentaire.

L'indexation a deux fonctions reconnues :

- Indiquer de manière concise le contenu d'un document ☞ descripteur devant exprimer un concept
- Permettre de rechercher efficacement l'information contenue dans un document ☞ descripteur comme point d'accès au document

La littérature spécialisée reconnaît plusieurs typologies d'indexation, marquées notamment par des oppositions.

- **Indexation manuelle (humaine)/indexation automatique** : ce qui fait la différence entre les deux indexations, c'est l'intervention d'un système informatisé. Seulement, selon Muriel Amar, la discordance entre les deux notions n'est pas réellement fondée, l'indexation automatique tentant de reproduire l'indexation manuelle, elles évoluent en parallèle mais sans se faire profiter mutuellement de leurs avancées.

³⁵ Muriel Amar, les fondements théoriques de l'indexation, p 8

- **Indexation manuelle/indexation en texte intégral** : cette opposition ressemble à la précédente car on assimile l'indexation en texte intégral à l'indexation automatique. L'indexeur humain n'est pas obligé de lire tout le texte pour en comprendre le sens et choisir les descripteurs, contrairement à la machine qui doit analyser, décortiquer mot à mot pour en extraire le thème grâce aux liens sémantiques existant entre les termes.

- **Indexation libre/indexation contrôlée** : la présence d'un langage documentaire est la caractéristique majeure de l'indexation contrôlée. Ce langage a été construit sur l'indexation de textes antérieurs alors que dans le cas de l'indexation libre, l'indexeur choisit lui-même les descripteurs sur la base d'un seul document.

L'indexation libre a l'avantage de proposer un vocabulaire assez large et plus proche des utilisateurs mais en revanche on peut aussi lui reprocher une trop grande synonymie. De plus, un même mot clé peut être utilisé par différents indexeurs et recouvrir des réalités différentes. Le vocabulaire contrôlé, s'il peut s'avérer trop spécialisé pour un certain type de public, a l'avantage d'être cohérent sur le plan terminologique. Il est conseillé de combiner les deux techniques afin d'indexer au plus juste, même s'il est vrai que sur le plan pratique, le processus est assez compliqué à mettre en place.

- **Indexation post-coordonnée/indexation précoordonnée** : ce cas a déjà été traité précédemment au niveau des normes, il n'est donc pas utile de revenir dessus puisque la problématique est identique.

3.2 Indexation de l'image fixe

Pendant longtemps, l'image a souffert de préjugés. Il n'était pas admissible qu'elle puisse être un vecteur de connaissance. D'après Michel Melot³⁶, elle fut même accusée de divulguer un savoir « sauvage, superficiel et fragmenté », de tromper celui qui la regarde. Pourtant, de tout temps, elle a existé, les peintures pariétales du paléolithique ne sont-elle pas des images ? Déjà à cette époque, les hommes ornaient leur intérieur de représentations figuratives ou symboliques.

L'image est partout, les techniques photographiques, le numérique en particulier, et les nouvelles technologies l'ont rendue omniprésente. Les personnels des organismes culturels furent donc obligés d'appivoiser ce nouveau support d'érudition et de lui accorder un traitement qui lui soit propre. En effet, une image n'est pas un texte et ne s'indexe pas de la même façon. Il ne faut pas croire non plus qu'elle se suffit à elle-même. La légende et

³⁶ Michel Melot, l'image dans les bibliothèques : trente ans après, p 67

tout autre commentaire accompagnent l'image et la rendent explicite. De même, elle fait souvent partie d'une série, comme un reportage par exemple. L'analyse de l'image est difficile et demande un travail intellectuel important. Par ailleurs, un autre élément vient s'ajouter et complexifier le tout : la reproduction d'image. Une photo de la Joconde est une autre image que le célèbre chef d'œuvre de Léonard de Vinci, d'autant plus qu'elles n'ont pas le même auteur.

Deux écueils sont à éviter dans le cadre de l'analyse de l'image. D'une part, contrairement à ce que l'on pourrait penser, l'image ne s'oppose pas à l'écriture mais au langage³⁷, au son. L'écriture est déjà une représentation graphique des choses, des idées, des concepts. Ensuite, l'image n'est pas véritablement calquée sur la réalité, comme on aimerait le croire. D'autant plus que les techniques de traitement de l'image permettent de la modifier à l'infini.

Le documentaliste ou le bibliothécaire qui aura pour mission d'indexer des images pourra s'appuyer sur les standards développés par l'IPTC. Cette instance a mis au point l'IMM (Information Interchange Model), ensemble de métadonnées qui peuvent être appliquées à des documents textuels, des images ou des vidéos. La norme AFNOR Z44-077 (1997) pour le catalogage de l'image fixe peut également être une bonne base sur laquelle s'appuyer. Par ailleurs, Cécile Kattnig³⁸, chargée de mission à la BNF, a publié un ouvrage abordant entre autre le traitement intellectuel et la valorisation de l'image.

L'avènement des nouvelles technologies ont donné lieu à deux techniques d'indexation de l'image fixe. L'indexation humaine s'attache à analyser l'image selon le contexte et la description de ce qui y est représenté tandis que l'indexation automatique³⁹ exploite les caractéristiques physique de l'image, c'est-à-dire les formes, les couleurs ou encore les contrastes. Pour l'instant, ces outils informatiques ne sont pas encore au point et l'indexation manuelle a toujours la préférence des spécialistes.

Les nouveaux types d'indexation tels que les folksonomies, les tags, seront abordés dans la deuxième partie du mémoire puisqu'ils sont liés à l'essor du Web 2.0 dit collaboratif et feront certainement l'objet de structuration dans le cadre du web 3.0 ou le Web sémantique. La question de l'image comme représentation d'œuvre d'art sera donc également abordée dans la deuxième partie et aussi dans la troisième liée à la réalisation d'un thésaurus iconographique de musée.

³⁷ Ibid.

³⁸ Cécile Kattnig, gestion et diffusion d'un fonds d'image, 2002

³⁹ Voir les travaux de N.James en annexe 2 page 99-101

Deuxième partie : Vers un Web sémantique

1. Les enjeux du Web sémantique

1.1. Présentation

L'expression Web sémantique⁴⁰ fut lancée par Tim Berners Lee, inventeur d'Internet et fondateur du consortium W3C⁴¹, organisme voué à établir des standards pour la compatibilité des différents systèmes technologiques. Cette formulation ne date pas d'hier, dans les années 90, au moment où Internet est apparu, les informaticiens pensaient déjà à mettre en place une relation sémantique entre les contenus publiés sur le web. Bruno Menon l'a d'ailleurs qualifié de « vin vieux dans des outres neuves⁴² ». On pourrait relativiser ce propos en disant que dans les années 90, même si l'idée était là, les technologies n'étaient pas suffisamment avancées pour la concrétiser. Le web sémantique ne remplace pas l'Internet actuel, il joue plutôt le rôle d'une brique supplémentaire permettant de structurer l'information contenue sur le Web. Voici comment T. Berners-Lee définit le lui-même : « Le web sémantique n'est pas un Web séparé, mais une extension du Web actuel, dans lequel l'information est munie d'une signification bien définie, permettant aux ordinateurs et aux personnes de mieux travailler en coopération ». En effet, il existe un réel paradoxe lié au net : il y a sur la toile des milliards de documents, l'utilisateur est submergé par l'information mais ne peut y accéder véritablement faute de structuration. L'enjeu principal du web sémantique est l'aide à la recherche. Le but est de faire en sorte que l'internaute trouve facilement et rapidement l'information qu'il recherche. Outre cet objectif majeur, le W3C prévoit de rendre le web encore plus collaboratif. En filigrane, le dessein du Web sémantique avec son accès à l'information se veut démocratique. Or, en plus des contraintes technologiques, on peut redouter un refus de transparence de certaines entreprises, organisations ou même nations, comme cela a été le cas à la naissance du Web (outil militaire avant de devenir « civil »).

Le web actuel est plutôt syntaxique. Si l'on regarde la définition du terme syntaxique dans un dictionnaire, nous retenons deux acceptions : d'une part, la syntaxe est la règle qui régit les relations entre les mots à l'intérieur d'une phrase, d'autre part elle se rattache à la forme, et donc s'oppose au sens, la sémantique. Il est interprétable par les humains mais très peu par les machines. Le web sémantique vise donc un traitement automatique des données. Le système mis en œuvre pour rendre effectif ce web sémantique s'appuie néanmoins sur le fonctionnement cérébral humain. A l'heure actuelle, les moteurs de recherche éprouvent encore des difficultés à désambigüiser un terme et à appréhender les

⁴⁰ Philippe Laublet, Jean Charlet et Chantal Reynaud, Introduction au Web sémantique, web

⁴¹ Fondé en octobre 1994, source : Wikipedia

⁴² Asuncion Valderrama et Arlette Boulogne, Web 3.0 et web sémantique : du vin vieux, dans des outres neuves, Documentaliste-Sciences de l'information, 2009, Vol 46 n°3, p 20

notions de polysémie et homonymie. Les métadonnées⁴³ des documents, qui pourtant pourraient justement aider à limiter le bruit et l'extraction de documents non pertinents, ne sont pas toujours fiables (quand elles existent !) et surtout prises en compte par les moteurs de recherche. Il s'agit donc de modéliser les données des documents et de les relier entre elles pour créer du sens.

Toutefois, il est important de souligner que le Web sémantique a été conçu pour organiser un contenu à valeur ajoutée, émanant d'entreprises, d'organisations, de musées, dans un souci de valorisation du capital informationnel et qui a notamment trait au Knowledge management.

1.2. L'usage des ontologies dans le Web sémantique

Dans la première partie du mémoire, nous avons défini ce qu'était une ontologie en partant de son étymologie. Il est question maintenant d'examiner son usage dans le cadre du web sémantique. Les ontologies jouent un rôle très important. Selon B. Le Grand et M. Soto⁴⁴ : « la structuration des données ne suffit pas à les rendre intelligibles par un utilisateur ou une machine. Quelle que soit la pertinence des regroupements de topics que cette structure permet d'obtenir, ces regroupements doivent être complétés par du sens pour être pleinement exploitables. Un moyen de fournir ce sens est l'utilisation d'ontologies ». Les premières expérimentations dans le domaine faisaient déjà cas des ces outils conceptuels. Ainsi, les projets SHOE (université du Maryland) et Ontobroker (Université de Karlsruhe) nés dans les années 90 étaient basés sur les ontologies. Quelques années plus tard, des chercheurs européens ont bâti le réseau Ontoweb dans le but de rassembler les travaux dans ce domaine et de les faire aboutir.

Une ontologie, dans le contexte du web, serait un ensemble de concepts liés à un univers. Dans la littérature spécialisée, les auteurs apparentent souvent les ontologies au thésaurus, leur point commun étant l'organisation et la structuration d'un ensemble de connaissances. Certains⁴⁵ les considèrent même comme des « concurrentes » des thésaurus. Le thésaurus version web serait donc une ontologie, une version améliorée et adaptée à la nature du réseau. On peut distinguer plusieurs types d'ontologie⁴⁶ :

- Les ontologies de haut niveau (ontologies formelles) expriment des concepts assez abstraits tels que le temps, l'espace, les matières ou les événements.

⁴³ Véronique Mesguich, Le Web sémantique : utopie ou réalité ?, Netsources, p 2

⁴⁴ Béatrice Le Grand et Michel Soto, Topics Maps et navigation intelligente sur le Web sémantique », p 5

⁴⁵ François Feyler, de la compatibilité à l'interopérabilité en matière de repérage d'information pertinente, la problématique et l'exemple d'OTAREN, Documentaliste-Sciences de l'information, 2007, Vol 44, n°1, p 86

⁴⁶ Manuel Zacklad, Classifications, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'informations (ROI), p 6

- Les ontologies par domaine : la médecine ou l'architecture.
- Les ontologies par activité : enseigner, plaider

Il existe au moins trois domaines pour lesquels l'usage des ontologies s'avère particulièrement intéressant. Le premier étant le E-commerce. Le système ONTOSEEK, mis au point entre 1996 et 1998, est exploité pour les services de pages jaunes ou de catalogues de produits. Il permet notamment à l'utilisateur de rechercher des marchandises en utilisant ses propres mots, alors que ceux-ci ne correspondent pas au langage employé par les professionnels. La technologie sous-jacente va justement mettre en correspondance ou traduire le vocabulaire pour faire jaillir directement la bonne information. Le deuxième champ d'application des ontologies est la médecine. Le domaine médical fait l'objet chaque jour d'une littérature toujours plus abondante, et le besoin de structurer l'information se fait encore plus pressant que d'ordinaire. Ainsi, la base de données bibliographiques Medline a mis au point un thésaurus qui est toutefois perçu comme une ontologie, MeSH. Il rend la recherche d'article plus aisée et pertinente. Enfin, l'efficacité des ontologies s'illustre aussi dans le domaine touristique, notamment pour le personnel des agences de voyage. Né en 2002, Hi-Touch, un programme de recherche et de développement de web sémantique appliqué au milieu touristique, est en voie de révolutionner l'organisation des voyages. Le but est de créer un environnement sémantique qui permettrait de lier des activités différentes en fonction du goût et des spécificités des individus. Par exemple, il sera facile d'organiser un voyage pour une personne végétarienne, qui apprécie les musées d'art moderne, qui voyage avec son chien et qui doit se rendre une fois par semaine à l'hôpital pour effectuer une dialyse. Ce qui était un casse-tête il y a encore quelque temps, voire même maintenant pour un agent de voyage, deviendra une simple formalité grâce à l'intégration des ontologies dans le web sémantique.

Enfin, WordNet, sans doute l'ontologie la plus représentative et la plus efficiente du Web sémantique. WordNet⁴⁷ est une base de données lexicale conçue par des linguistes de l'Université de Princeton (E.U). Elle se veut comme un répertoire de syntagmes ou éléments lexicaux, les uns liés aux autres par de multiples relations, afin de rendre compte de la richesse de la langue anglaise. WordNet a été traduite en d'autres langues mais la version anglaise est la plus complète à ce jour. L'élément de base du système est le « synset », abréviation de *synonym* et *set* (ensemble). La base comprend environ 200 000 mots et 115 000 synsets. Il peut désigner des concepts concrets comme *voiture* ou *maison*, mais aussi des notions plus abstraites comme *entité* ou *quelque chose*. C'est dans ce sens qu'on peut légitimement la considérer comme une ontologie. Evidemment, ce système connaît quelques imperfections : on peut déplorer que la richesse des liens sémantiques entre les

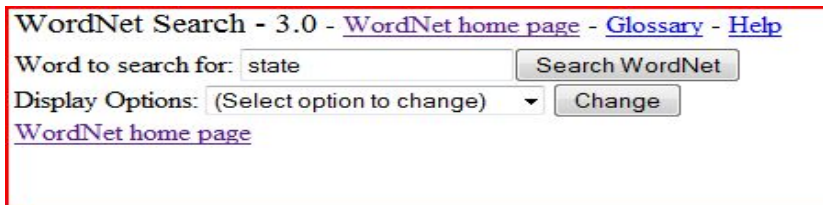
⁴⁷ Source de la définition : Wikipedia

syntagmes ne soit pas équilibrée selon les classes de mots, tels que les verbes par exemple. WordNet est un outil téléchargeable et fait l'objet d'une licence.

Voici un exemple illustré de copies d'écran sur ce que produit cette ontologie :



Page d'accueil du site de WordNet



Formulaire de recherche : J'ai effectué une requête sur le terme « State » (Etat en français) qui est polysémique pour tester l'efficacité de

Noun

- [S: \(n\) state, province](#) (the territory occupied by one of the constituent administrative districts of a nation) *"his state is in the deep south"*
- [S: \(n\) state](#) (the way something is with respect to its main attributes) *"the current state of knowledge"; "his state of health"; "in a weak financial state"*
- [S: \(n\) state](#) (the group of people comprising the government of a sovereign state) *"the state has lowered its income tax"*
- [S: \(n\) state, nation, country, land, commonwealth, res publica, body politic](#) (a politically organized body of people under a single government) *"the state has elected a new president"; "African nations"; "students who had come to the nation's capitol"; "the country's largest manufacturer"; "an industrialized land"*
- [S: \(n\) state of matter, state](#) ((chemistry) the three traditional states of matter are solids (fixed shape and volume) and liquids (fixed volume and shaped by the container) and gases (filling the container)) *"the solid state of water is called ice"*
- [S: \(n\) state](#) (a state of depression or agitation) *"he was in such a state you just couldn't reason with him"*
- [S: \(n\) country, state, land](#) (the territory occupied by a nation) *"he returned to the land of his birth"; "he visited several European countries"*
- [S: \(n\) Department of State, United States Department of State, State Department, State, DoS](#) (the federal department in the United States that sets and maintains foreign policies) *"the Department of State was created in 1789"*

Verb

- [S: \(v\) state, say, tell](#) (express in words) *"He said that he wanted to marry her"; "tell me what is bothering you"; "state your opinion"; "state your name"*
- [S: \(v\) submit, state, put forward, posit](#) (put before) *"I submit to you that the accused is guilty"*
- [S: \(v\) express, state](#) (indicate through a symbol, formula, etc.) *"Can you express this distance in kilometers?"*

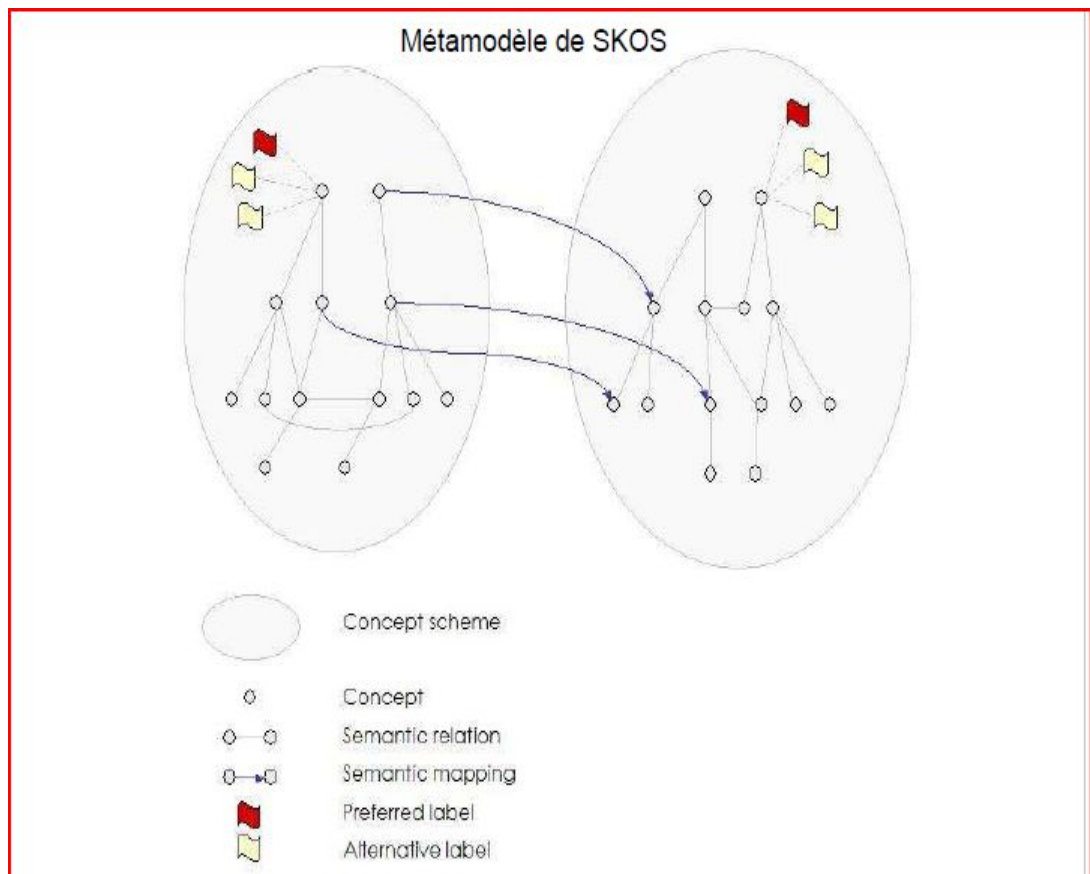
1.3. Les applications du web sémantique

1.3.1. Le projet SKOS

Des ingénieurs ont imaginé des dispositifs permettent de représenter des schémas de concepts, tels que les thésaurus, les taxonomies ou les classifications. Le plus connu est SKOS, développé à partir de 2002 par le projet SWAD-Europe et financé par la communauté

européenne dans le but de développer la recherche dans le domaine des technologies de la société de l'information. Il a ensuite été repris par le W3C pour être amélioré et associé aux travaux sur les standards du Web sémantique (RDF, OWL, etc.). SKOS est composé de plusieurs éléments. Le « SKOS Core » qui est le noyau du système et qui renvoie aux propriétés de base. Ce noyau est complété par les « SKOS extensions » permettant :

- de clarifier la nature des relations (relation hiérarchique partitive ou d'instance)
- de spécifier un concept par une note d'application



Source : <http://www.u-picardie.fr/~furst/4-SKOS.pdf>

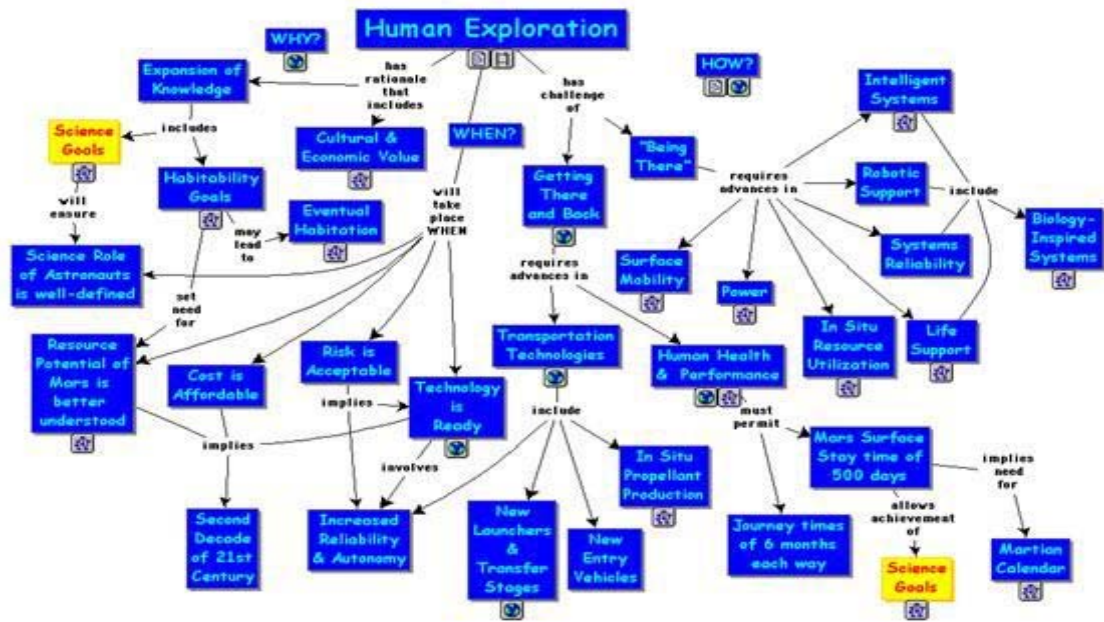
Un certain nombre de langages contrôlés ont été transposés au Web par le biais de SKOS, tels que le répertoire RAMEAU, les thésaurus GEMET et AGROVOC et les LCSH de la bibliothèque du congrès.

1.3.2. Les Topics Maps ou cartes conceptuelles

Apparus dans les années 90, les Topics Maps sont également des outils de représentation des connaissances, au même titre que SKOS. Elles font d'ailleurs l'objet d'une norme ISO/CEI 13250. Le principe est de partir d'un concept (une idée, un objet et même une personne) et d'y agréger tout un ensemble d'autres concepts, nommés « occurrences », le tout sous forme graphique comme l'expression le laisse entendre. La différence entre SKOS

et les cartes conceptuelles résident dans le fait que ces dernières peuvent aller beaucoup plus loin dans la représentation. On peut les considérer comme une extension de SKOS. Lorsque les concepts deviennent trop nombreux et que la carte devient trop complexe, il convient de procéder au « clustering », c'est-à-dire au regroupement de concepts.

Elles n'utilisent pas le langage RDF mais des travaux sont entrepris pour les rendre compatibles.



Source : <http://www.knowledgemapping.net/images/stories/illustrations/Carthe%20conceptuelle%20de%20l%20exploration%20humaine%20%5BNovak,%202001%5D.jpg>

1.3.3. Le TAL

Les premiers travaux sur le traitement automatique de la langue ont débuté dans les années 50 et n'avaient donc rien à voir avec le web sémantique. Le lien entre les disciplines peut s'envisager de deux façons⁴⁸ : Qu'est-ce le Web sémantique apporte au TAL ? En quoi le TAL peut-il favoriser le développement du Web sémantique ?

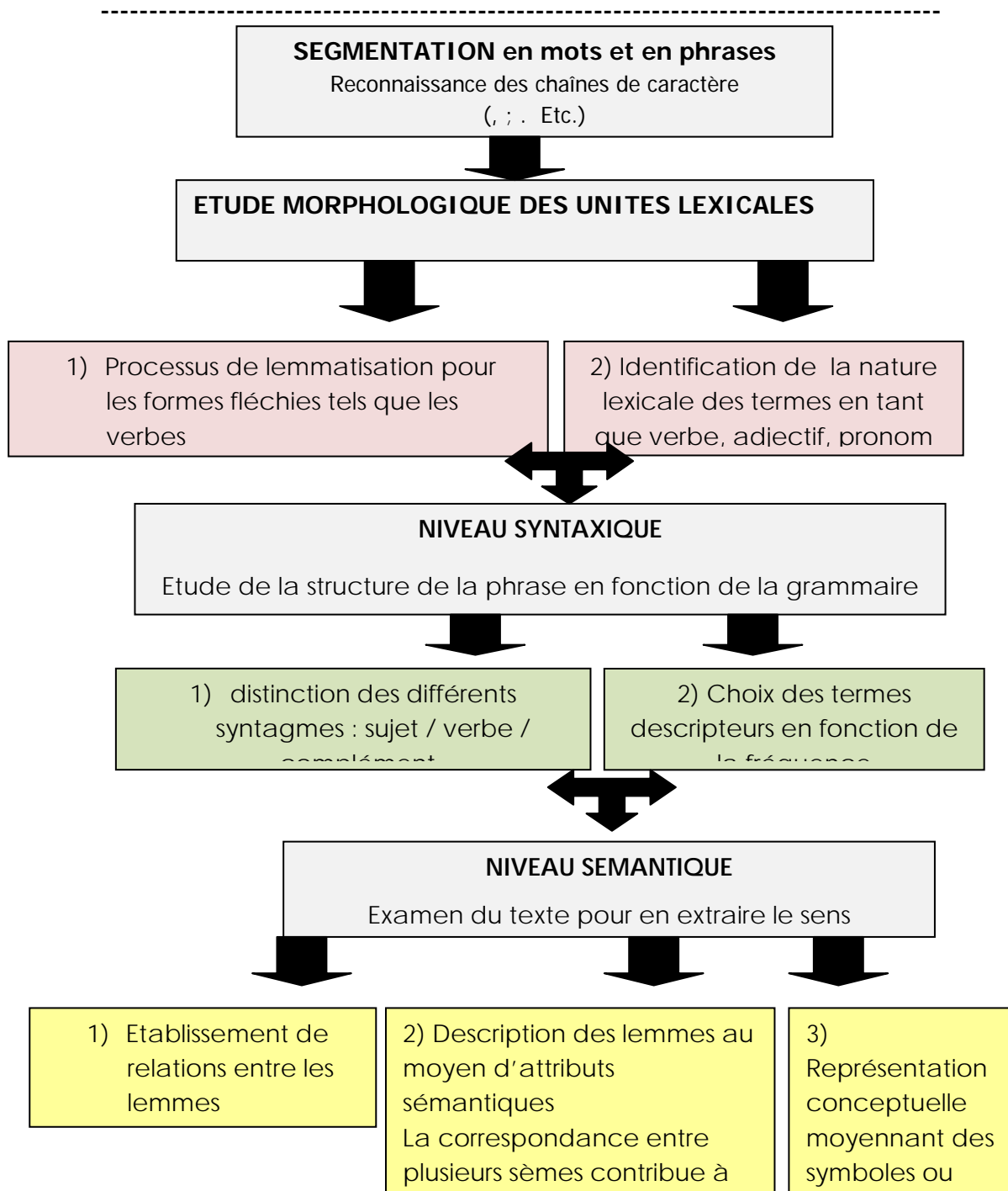
Dans le traitement automatique de la langue, on étudie la relation des mots entre eux au sein d'une phrase pour en saisir le sens. La sémantique a permis d'étendre cette analyse au discours pour en faire émerger le sens, le thème principal. Certains logiciels de TAL, exploités dans le domaine de la veille, parviennent à déterminer si un texte est plutôt positif ou négatif, très intéressant dans le domaine de l'E-réputation par exemple.

A l'ère du tout technologique, qui a induit une croissance exponentielle de documents, l'indexation humaine devient une lourde charge et il permet d'espérer que les avancées dans le domaine du traitement de la langue permettront d'impulser le Web sémantique.

Plus qu'un long discours, un schéma synthétique et récapitulatif m'a paru plus pertinent pour rendre compte du fonctionnement du TAL.

⁴⁸ Alain léger, Jean Charlet, Applications du Web sémantique, Information, Interaction Intelligence, HS, 2004 p 182

I- TRAITEMENT DE LA LANGUE



II – MODE DE REPRESENTATION

- Indexation libre ou contrôlée
- Classification automatique (ou clustering)
- Catégorisation des documents (notamment grâce à un thésaurus)
- Résumé automatique

1.3.4. Le projet Scriptures⁴⁹

Les ontologies sur le Web n'ont pas été conçues uniquement pour les documents textuels mais également d'autres supports tels que les images, les sons ou les vidéos. Le département informatique de Télécom Bretagne⁵⁰, basé à Brest, a mis au point le projet Scriptures il y a de cela quelques années. Ce projet a pour objectif l'indexation des contenus multimédias d'un point de vue sémantique. Véronique Mesguich⁵¹ explicite la méthodologie ainsi : « ce projet propose une méthodologie qui couple les processus d'indexation et de recherche d'information multimédia, dans un espace pré-structuré par une analyse d'expert. L'utilisateur se sert de ces schémas, et peut également indexer des images par du texte, qu'il écrit ou récupère par ailleurs ».

L'expérience Scriptures se situe au confluent de plusieurs thématiques :

EN AMONT	EN AVAL
Problématique du Web sémantique	La recherche et l'indexation sémantique de documents
Traitement automatique de la langue	L'analyse interprétative des données
Sémiologie des contenus et des connaissances	La composition adaptative et contextuelle de documents et de systèmes d'information

D'autres projets similaires à Scriptures, notamment portant sur l'utilisation des ontologies pour le traitement des images. Citons par exemple les travaux de N.JAMES⁵², doctorant au MAS, qui travaille sur l'annotation automatique pour l'indexation des images.

⁴⁹ Semantic Composition and Retrieval of Information Pictures or Services

⁵⁰ Ecole d'ingénieurs et centre de recherche international dans les sciences et technologies de l'information

⁵¹ Véronique Mesguich, Web sémantique : utopie ou réalité ?, Netsources n°71, p 4

⁵² Voir annexe 2 p98-100

2. Adaptation des langages documentaires au Web 3.0 et interopérabilité

Dans un premier temps, il serait intéressant de spécifier la notion d'« interopérabilité » que beaucoup confondent encore avec compatibilité. Voici comment Patrice Landry⁵³ conçoit ce principe : « On définit l'interopérabilité comme la capacité de plusieurs systèmes à communiquer entre eux sans ambiguïté et à échanger de l'information sans difficulté ». Toutefois, la nuance entre les deux idées est mince : la compatibilité désigne la communication entre deux systèmes A et B tandis que dans le cas de l'interopérabilité, il n'y a pas de restriction du nombre.

2.1. Les technologies du Web sémantique

Comme nous l'avons vu précédemment, l'adaptation des langages documentaires ou ontologies à l'environnement du Web nécessite un panel de dispositifs dont SKOS. Ces dispositifs dépendent eux-aussi de protocoles établis par le W3C pour rendre les liens sémantiques possibles. Cependant, la multiplicité des « langages » qui constituent les fondations du web sémantique tend à complexifier cette infrastructure. Je m'efforcerai donc de proposer une démonstration assez brève des standards du W3C les plus représentatifs en définissant chaque élément séparément :

- **RDF (Resources Description Framework)** : c'est un modèle de graphes censé représenter les diverses ressources du web et leurs métadonnées de manière à en permettre le traitement automatique. En outre, elle favorise l'interopérabilité en annotant des documents ou applications non structurés sur le Web. Un document modélisé par RDF est composé de trois éléments :
 - Le sujet = la ressource à décrire
 - Le prédicat = un type de propriété correspondant à cette ressource
 - L'objet = il représente une autre donnée ou une autre ressource

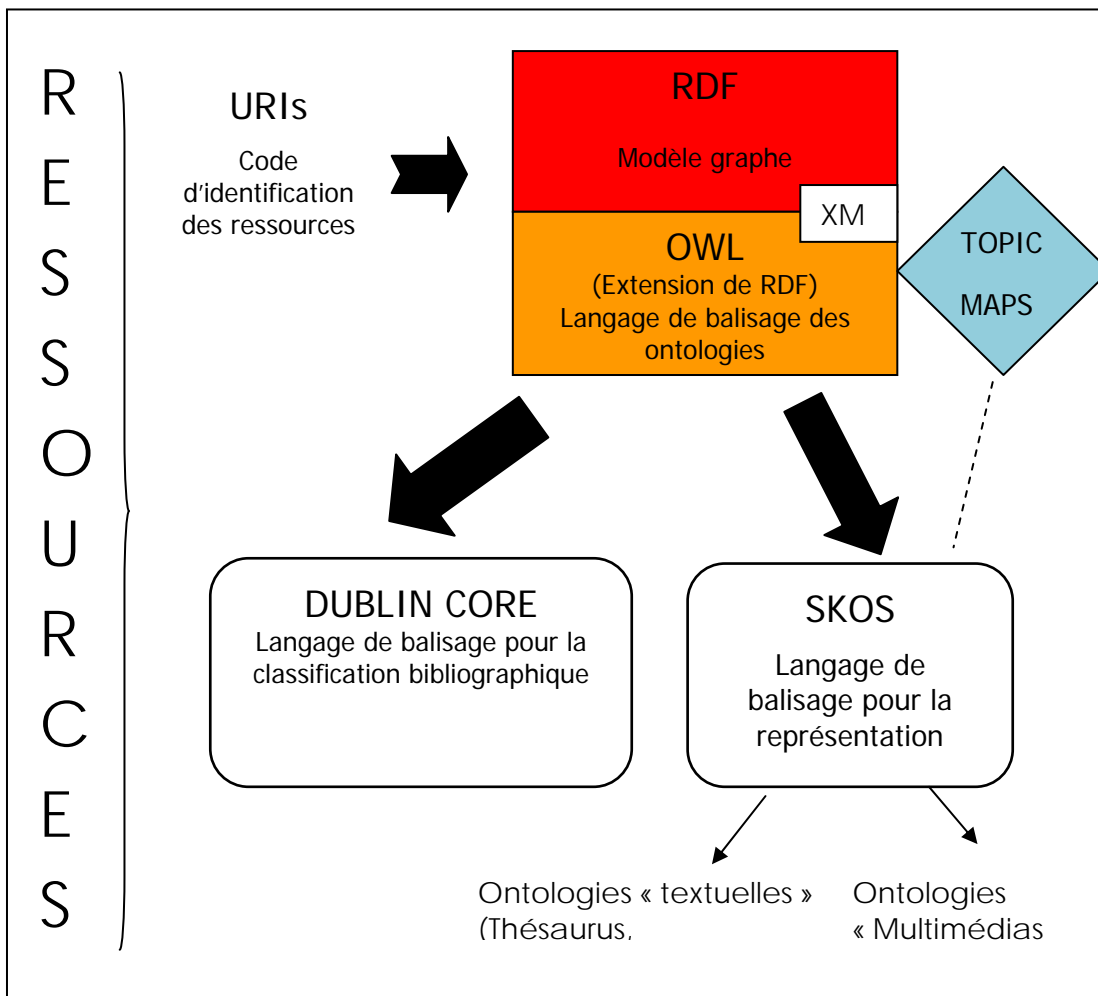
RDF sert de base à de nombreux langages édifiés par le W3C, tels que RDFS ou OWL pour les ontologies, SKOS pour la représentation des thésaurus, Dublin Core pour le classement bibliographique ou encore FOAF pour la description des personnes (celui-ci est nettement moins connu)

- **RDFS (Resources Description Framework Schema)** : il complète le langage RDF en précisant les données définies par celui-ci.

⁵³ Patrice Landry, Multilinguisme et langages documentaires : le projet MACS en contexte européen, Documentation et Bibliothèques, avril/juin 2006, p 121

- **OWL (Ontology Web language)** : Langage conçu pour représenter et modéliser des ontologies, notamment des thésaurus. Il est basé sur une structure RDF traduite en langage XML. Il est considéré par les spécialistes comme une extension de RDF. Il est décliné en trois variantes : OWL-Lite, la version la plus simple, OWL-DL et OWL-Full.
- **URIs (Uniform Resource Identifiers)** : Ce sont des composants du web actuel et qui deviennent fondamentaux dans le contexte du Web sémantique. Ils rendent non seulement possible l'identification des ressources sur Internet mais également de créer du lien entre elles.
- **XML (Extensible Markup language)** : Il sert à stocker et à transférer des données structurées en arborescence. C'est un langage de mise en forme, contrairement à l'HTML qui se contente d'ordonner les données avec des hyperliens. Les deux langages se juxtaposent et on utilise aujourd'hui l'appellation XHTML.

Le schéma ci-dessous a pour fonction de synthétiser de la manière la plus simple possible la structuration du Web sémantique.



2.2. Le multilinguisme comme facteur d'interopérabilité

L'interopérabilité passe aussi par le multilinguisme. L'accès « universel » à l'information étant l'objectif ultime du web sémantique, l'internaute doit pouvoir, en s'exprimant dans sa langue maternelle, trouver le renseignement qu'il lui faut, fût-il en japonais, russe ou en anglais. Malheureusement, il faut encore de nos jours recourir à ses propres connaissances linguistiques ou recourir à des dictionnaires pour traduire les documents qui nous intéressent. Les applications Web telles que *Google translate* peuvent dépanner mais les traductions ne sont pas toujours très pertinentes (quand ce n'est pas à la limite du comique !).

Avant même l'ère d'Internet, des tentatives pour rendre compatibles plusieurs thésaurus de langues variées avaient vu le jour et avaient été couronnées de succès. On

peut par exemple citer l'*UMLS*⁵⁴, le *ROOT thesaurus* ou bien le thésaurus de l'UNESCO, *l'integrated Multilingual thesaurus*, établis au format de communication MARC. La bibliothèque du congrès a également converti son répertoire de vedettes-matière, les LCSH, au multilinguisme. Le Canada en a fait autant avec le RVM Laval mais seulement en version anglaise et française.

Le développement du Web et la variété de ressources qu'il contient ont impulsé de nouvelles initiatives. En effet, des classifications, telles que Dewey et sa transposition sur Internet, WebDewey, se sont vues insufflées et projetées sur la toile. Plusieurs thésaurus ont été mixés et ont donné lieu à des métathésaurus grâce à un système de traduction manuel ou automatique. TermSciences en est un exemple assez parlant.

Préoccupée par cette problématique du multilinguisme, la CENL⁵⁵ a mis au point en 1997 le projet MACS, en concertation avec quatre bibliothèques : la BN de Suisse en tant que pilote du projet, la BNF François Mitterrand, la DNB d'Allemagne et la British Library. Le but était de favoriser l'accès aux catalogues de bibliothèques en ligne dans un contexte multilingue. Pour ce faire, elles ont procédé à une mise en concordance des termes dans les langages d'indexation, sans passer par une traduction automatique des vedettes. L'objectif n'était pas de créer un métathésaurus multilingue mais de créer des équivalences entre des « grappes conceptuelles » d'une langue à l'autre. Les expérimentations du projet MACS ont été validées par la CENL en janvier 2005. Un des inconvénient de MACS est sa mise en pratique sur le Web, des études sur l'infrastructure seraient toujours en cours pour rendre efficient cet outil.

Ainsi, comme nous venons de le voir, les langages documentaires ont un avenir au sein du Web sémantique. Ils constituent une aide à la structuration et modélisation des données. Les professionnels de l'AFNOR ou de l'IFLA les ont d'ailleurs regroupés sous le terme de « référentiels »⁵⁶. La question de l'interopérabilité ou de l'harmonisation des outils documentaires prend tout son sens dans cette optique de développement de Web sémantique. Cela a également l'avantage de faire collaborer des professionnels qui n'avaient jusqu'ici que peu d'occasions de se rencontrer, de partager des connaissances et de trouver des consensus. En effet, d'aucuns⁵⁷ reprocheront au Web de déformer, d'« affaiblir » les langages documentaires. Cependant, n'est-ce pas le prix à payer pour pouvoir les adapter au Net, sans lequel ils seraient sans doute voués à une mort certaine ? Il me paraît légitime de se poser la question. D'autant qu'au XXI^{ème} siècle, la flexibilité est de rigueur, on demande de plus en plus aux individus une certaine plasticité cérébrale pour assimiler sans cesse de

⁵⁴ Unified Medical Language System

⁵⁵ Conférence des bibliothèques nationales européennes

⁵⁶ Muriel Amar, Référentiels, données d'autorité, thésaurus, ontologies, taxonomies ... pour en savoir plus !, Documentaliste-Sciences de l'information, 2007, Vol 45, n°3, p 14

⁵⁷ Bruno Menon, Optimiser l'accès à l'information, une opportunité pour les langages documentaires ?, Journée d'étude ADBS, Documentaliste-Sciences de l'information, 2007/6, Vol 44, p 387

nouvelles informations ou concepts, il paraît donc normal que les objets ou les outils subissent eux-aussi cet état de fait pour s'assurer une place dans ce monde virtuel.

Cependant, la mise en place du Web sémantique et de l'interopérabilité rencontre quelques obstacles. Dans un premier temps, on observe qu'il est assez difficile finalement de construire des ontologies⁵⁸, surtout si l'on considère le fait qu'elles sont elles-mêmes des versions améliorées des thésaurus. Les techniciens veilleront à établir une méthodologie qui prendra en compte le maintien et l'interopérabilité avec d'autres types de ressources du Web. Ensuite, vient la question des métadonnées, le réflexe d'enregistrer des informations décrivant des ressources n'est pas encore acquis et elles doivent respecter un certain formalisme. Les langages qui permettent cette interopérabilité sont également multiples et complexes. Enfin, le Web sémantique est un projet très ambitieux et sa mise en œuvre occasionne des coûts très élevés, que certaines organisations pourront supporter et d'autres pas.

⁵⁸ Bruno Menon, le web sémantique : de nouveaux enjeux documentaires ?, Journée d'étude ADBS/IUT Paris V , Documentaliste-Sciences de l'information, 2003, vol 40, n°6 p 387

3. Emergence d'un nouveau type d'indexation

La vocation des langages documentaires étant d'indexer, il était sans doute logique que leur adaptation au web allait bouleverser la pratique même de l'indexation. A l'heure actuelle où nous parlons Web 2.0 ou Web collaboratif, les internautes se sont emparés de cet usage jusqu'ici réservé au milieu documentaire. Cette nouvelle tendance, apparue vers 2003, porte le nom de « folksonomies ». Selon Wikipedia, « une folksonomie est un néologisme désignant un système de classification collaborative décentralisée spontanée ». C'est un terme anglo-saxon, combinaison de Folk qui veut dire *peuple* et *taxonomy*, qui a été francisé mais qui n'a pas encore trouvé sa forme définitive dans la langue française. Les prêcheurs de la francophonie lui préfèrent en général les vocables « potonomie » ou « peplomonie ».

3.1. Les différents grades de l'indexation collaborative

Si le concept peut paraître assez floue aux yeux des novices, on peut toutefois distinguer plusieurs niveaux d'indexation collaborative⁵⁹ qui empiètent les uns sur les autres. Le premier échelon serait l'indexation par l'auteur. Quand un individu ou une organisation met en ligne un site web, il choisit lui-même les descripteurs qui caractérisent le contenu de son site pour le référencer sur les sites Web. C'est l'indexation comme autopromotion. Ensuite, les internautes qui vont consulter ce site pourront eux-aussi le référencer avec leurs propres termes, qui ne seront probablement pas les mêmes que ceux de l'auteur. C'est la deuxième couche de l'indexation collaborative. Les individus peuvent à l'aide de sites comme delicious (ou del.icio.us) ou yoono classer leur sites préférés et les faire partager au reste du monde grâce aux folksonomies. Elles se présentent sous forme de nuages de mots, lesquels apparaissent de taille variable : les plus gros étant les plus usités par la communauté. Chacun peut décider d'utiliser les mots proposés par d'autres ou d'en incorporer de nouveaux. Il y a une logique de partage, de consensus.

3.2. L'intérêt et l'usage des folksonomies

Le nombre sans cesse croissant de documents sur le Web oblige incontestablement tout un chacun à contribuer à l'indexation, au moins celle qui concerne ses propres contenus. Comme le dit Olivier Ertzscheid⁶⁰, « l'indexation à l'ère industrielle est nécessairement sociale ». L'indexation automatique et humaine (on entend ici

⁵⁹ Elie Francis et Odile Quesnel, Indexation collaborative et folksonomies, Documentaliste-Sciences de l'information, 2007, Vol 44, n°1, p 59-60

⁶⁰ Olivier Ertzscheid, Culture documentaire et folksonomie : l'indexation à l'ère industrielle et collaborative », Documentaliste-Sciences de l'information, 2007, vol 47, n°1

professionnelle, c'est-à-dire la pratique exercée par des documentalistes) ne suffisent pas pour indexer l'ensemble du Web. On peut aussi supposer que les internautes, un peu perdus et perplexes face à l'immensité du Web, cherchent à le contrôler en s'y investissant, en y créant des points de repère et se rassurent par l'établissement de relations avec autrui. Les liens ainsi tissés lui permettront également de remonter jusqu'aux ressources les plus pertinentes du web. Les termes proposés dans les folksonomies correspondent aux besoins des utilisateurs et cette pseudo-hiérarchie entre les mots est censée filtrer les mauvais choix et les abus des utilisateurs. En effet, un processus de sélection « naturelle » se construit, les termes les moins populaires finissant par disparaître. Par ailleurs, on constate une triple économie du fait de l'usage des folksonomies : il ya un gain financier lié à l'absence de prestation documentaire, un faible coût cognitif car les internautes proposent en général les premiers mots qui leur viennent à l'esprit. L'effort à fournir est minimal. Il y a aussi un bénéfice sur le plan informatique car toute ressource décrite par les internautes, c'est autant de travail en moins pour l'indexation automatique. Qui plus est, l'homonymie et l'utilisation erronée de certains termes donnent lieu à une culture de la sérendipité, principe qui consiste à découvrir de manière totalement fortuite une information. Les unités de base des folksonomies sont les tags, parfois aussi appelés « hashtags⁶¹ », notamment pour les sites comme Twitter. Des spécialistes⁶² ont retenu six types de tags portant sur : le contenu de la ressource, le contributeur ou la source du document, une évaluation de la qualité de la ressource, une description des caractéristiques de la ressource, la relation avec le tagueur ou l'activité associée à la ressource (recherche d'emploi, indexation photo ...). Il a été aussi observé que la plupart des internautes indexent de façon différente selon qu'ils le font pour eux-mêmes ou pour la communauté. Dans le deuxième cas, ils font un « effort » de précision dans la description. S'ils indexent pour eux-mêmes, ils emploient davantage d'abréviations ou un langage que parfois eux seuls pourront comprendre. Dans le cas de twitter, site sur lequel apparaissent des brèves souvent exprimées dans un langage semi-codé, il existe des dictionnaires appropriés, tels que « whatTheHashtag », pour déchiffrer les informations.

Outre les fonctionnalités pratiques des folksonomies, d'indexation et de recherche d'informations, il est intéressant de prendre en compte la dimension psychologique de ce nouveau procédé. En intercédant sur le web, l'individu se plaît à penser qu'il « joue un rôle », comme au cinéma, sauf qu'il s'agit d'un rôle plus accessible. Il peut publier et/ou donner son avis et si son opinion est validée par d'autres, cela lui confère une certaine confiance en lui-même. Cela se vérifie avec des sites comme Flickr ou Dailymotion où les internautes se fixent comme objectif de faire figurer leurs « œuvres » au Hit parade. Enfin,

⁶¹ Terme d'indexation précédé du signe #

⁶² Maxime Crépel, les folksonomies comme support émergent de navigation sociale et de structuration de l'information sur le Web, Lavoisier-Réseaux, 2008, n°152, p180

par le simple fait d'indexer, l'internaute se sent utile à lui-même et aux autres. Le succès des folksonomies est donc attesté.

L'indexation par tag et la recherche par les folksonomies semblent donc plus évidente, plus décomplexée que sur les autres banques d'images. Une expérience⁶³ réalisée sur des utilisateurs cherchant des images sur la base de l'agence photo Gamma a révélé que ces derniers sont très peu familiarisés avec les langages documentaires employés par les banques d'images en ligne, et ce pour plusieurs raisons : ces langages sont assez complexes, très hétérogènes selon les sources et ne facilitent donc pas l'assimilation, et les usagers ne prennent pas non plus la peine de consulter les guides d'indexation qui leur permettraient d'effectuer des recherches efficaces (utiliser ou non les guillemets etc.). Enfin, ils privilégient en grande majorité des moteurs de recherche simples car ils considèrent le module de recherche avancée trop difficile à utiliser.

3.3. L'indexation collaborative des images sur le site Flickr

Flickr est un site de partage de photos et vidéos⁶⁴, plébiscités autant par le grand public que par les professionnels de la photo. Conçu en 2002 par la société canadienne Ludicorp, le projet initial de Flickr n'avait pas pour objectif d'être une plateforme d'échange mais plutôt un jeu vidéo, finalement abandonné. La société fut rachetée par Yahoo ! en 2005. Chacun peut exposer ses images en ligne, décider de les diffuser à grande échelle ou plutôt de les réserver à un cercle intime tels que les amis ou la famille. Le site est gratuit mais certaines fonctionnalités sont payantes, telles que la suppression de la publicité, un espace de stockage illimité et une application statistique. Les utilisateurs peuvent consulter le blog Flickr pour être avertis des nouveautés.

Lorsque l'internaute possède déjà un certain nombre de photos indexées, le site Flickr lui propose des tags déjà utilisés afin de lui faire gagner du temps. Ils peuvent donc tagger en série un corpus d'images sur un même thème. Ces tags servent avant tout chose aux utilisateurs à faire des recherches dans leur propre collection. Lorsqu'ils décident de rendre publiques leurs images, les photographes amateurs ou professionnels utilisent leurs pseudo dans le système de tagging afin d'en garder le contrôle mais aussi pour distinguer leurs photos de celles d'autrui qui sont identiques et décrites avec les mêmes tags.

Les tags des photos Flickr comportent aussi des dates, notamment celles publiés par les membres faisant partie d'un groupe. Les amoureux de Paris qui se sont donnés rendez-vous le 25 juin 2010 pour faire des photos du jardin Luxembourg pourront par exemple

⁶³ Coralie Picault, Usages et pratiques de recherche des utilisateurs d'une banque d'images, *Documentaliste-Sciences de l'information*, 2007, vol 44, n°6, p 376

⁶⁴ Voir Annexe 4 p 110-111

indexer leurs clichés de la façon suivante : 25062010jardluxparis. Lors de la création de groupe, certaines règles de « tagage » sont clairement établies et les membres se doivent de les respecter sous peine d'être rappelés à l'ordre par les modérateurs. L'objectif premier étant de se démarquer du flot d'utilisateur et de filtrer les photos. En bref, la forme d'indexation signerait en quelque sorte l'identité du groupe.

Par ailleurs, l'instinct communautaire des « flickr addicts » se traduit par l'envie de faire partager ses photos au monde entier, de favoriser l'accessibilité de l'information. Or, on peut se dire que cet « altruisme » déguise en réalité une envie d'imposer son identité, sa vision du monde à la face des autres. Les plus passionnés d'entre eux cherchent à ce que leurs photos figurent dans « l'explore », un classement des meilleures images. Pour cela, ils vont jusqu'à y accoler plusieurs dizaines de tags (le maximum autorisé étant 75). Cela se rapproche plus de la publicité (mensongère ?), du matraquage que de la pure indexation.

Cependant, Flickr prévoit un système de « nettoyage » des clichés jugés de mauvaise qualité par les internautes. Chaque groupe décide qu'à partir d'un certain nombre de « viremoi », ou « deleteme », la photo doit être supprimée du site par son auteur. Les tags sont pris en compte pour évaluer les contenus et décréter qu'une photo mérite d'être conservée ou non. Tout dépend de la culture du groupe. Certains tags, formulés en phrases complètes, permettent aux internautes de dialoguer entre eux sans être compris des autres. L'humour et la taquinerie rentrent souvent en jeu dans ce système de tags codés. La notion de web collaboratif (et web sémantique ?) prend ici tout son sens. Par contre, certains apprécient peu que leurs images soient taguées par des inconnus. En effet, ils préfèrent que leurs images soient commentées car ils considèrent les tags d'autrui comme une intrusion dans leur univers personnel. Indexer une image, c'est comme y apposer sa marque, l'estampiller de son identité. Cela est d'autant plus vrai que les photos peuvent être facilement détournées de leur contexte original et cela s'avère particulièrement vexant pour les auteurs des images.

3.4. Limites des folksonomies

Malgré tout l'intérêt qu'elles suscitent, les folksonomies ne sont pas exemptes de défauts. Le premier est qu'elles génèrent beaucoup de bruit, notamment à cause des homonymes, et donc ne facilitent pas du tout l'accès à l'information. Devant la masse de réponses, l'utilisateur aura tendance à abandonner plutôt qu'à trier. L'indexation collaborative n'est pas non plus exhaustive car l'individu va décrire la source de manière subjective, en fonction de l'intérêt qu'il lui porte, par conséquent certains sujets seront sciemment ou non évacués. La manière d'orthographier les mots interfère aussi dans la qualité de l'indexation. Par exemple, un même mot exprimé au pluriel ou au singulier donnera lieu à deux mots, pareil pour les termes composés pour lesquels on décidera

arbitrairement d'y mettre un trait d'union ou non. Les internautes, qui ne sont pas des linguistes, loin s'en faut, passent largement outre l'exactitude et la pertinence des termes proposés. Quand bien même ils s'en préoccuperaient, la question de l'ambiguïté se posera toujours car un concept est différemment appréhendé selon la culture, la mentalité et le niveau de connaissance de celui qui cherche l'information. Enfin, il est important de souligner certains actes délibérés de malveillance, que l'on peut qualifier « d'infopollution⁶⁵ » et qui consiste à indexer des ressources « douteuses » avec des termes (« spam taggings ») qui y sont totalement étrangers pour inciter les internautes à consulter ces sites.

3.5. Bilan sur les folksonomies

Les folksonomies suscitent une vraie polémique au sein de la profession. D'un côté, les spécialistes les plus enthousiastes envisagent l'indexation collaborative comme une alternative aux langages documentaires, assez complexes aux yeux des utilisateurs et qui éprouvent des difficultés à s'imposer sur le Net. De l'autre côté, certains experts des Sciences de l'information opposent radicalement indexation collaborative et indexation normée. Les documentalistes tiennent à faire valoir leur connaissance en la matière et estiment qu'indexation professionnelle et indexation collaborative ne se valent pas. Certes, ils n'ont pas tout à fait tort mais reconnaître l'intérêt des folksonomies ne remet pas en cause leur savoir-faire. Il faudrait plutôt envisager ce nouveau mode comme complémentaire et voué au domaine public. En effet, les organisations ont d'autres besoins que les folksonomies ne peuvent pas satisfaire et pour lesquelles l'emploi d'un thésaurus, d'une taxonomie ou d'une classification s'avère plus judicieux.

Selon O. Ertzscheid, les différentes typologies de l'indexation, professionnelle, industrielle (moteur de recherche) et collaborative sont amenées à interagir et à faire bénéficier les unes aux autres de leurs divers atouts dans le but de produire une indexation de qualité. Google le fait déjà en se servant des tags des internautes pour référencer des sites.

⁶⁵ Maxime Crépel, les folksonomies comme support émergent de navigation sociale et de structuration de l'information sur le Web, Lavoisier-Réseaux, 2008, n°152, p 177

4. Les musées et leur usage du Web

Pendant très longtemps, les musées détenaient un savoir assez confidentiel, destiné à une élite. L'apparition du Web a déclenché quelques vagues d'austérité et de méfiance de la part des conservateurs qui craignaient la mauvaise vulgarisation de cette culture. L'avènement d'Internet a entraîné une mutation du domaine culturel. Les musées et les autres organisations patrimoniales ont su prendre le virage du XXIème avec habileté et s'emparer des outils sociaux pour se forger une nouvelle identité, à tel point qu'on parle aujourd'hui de musée⁶⁶. 2.0. En effet, les musées sont présents sur tous les fronts : Facebook, Twitter, Flickr, Youtube ... Chaque application leur permet de s'adresser à différents publics, et de se mettre en valeur de différentes façons. L'union de ces deux disciplines a donné naissance à une nouvelle discipline, la muséologie⁶⁷.

4.1. Qu'est-ce que la muséologie ?

Voici la définition donnée par Wikipédia : « La muséologie est une « méta discipline » qui étudie, réfléchit, questionne l'institution muséale, le musée, et travaille intellectuellement l'ensemble des activités liés au champ muséal, les sciences et techniques du musée que sont la gestion, la recherche, la conservation, le classement, la mise en valeur des objets, œuvres ou patrimoine (les collections), la médiation [...]. Elle étudie notamment l'histoire, les fonctions philosophiques, sociales, culturelles, des musées, tout comme les attentes, les besoins, et les pratiques des publics. Elle interroge également l'avenir du muséal dans l'espace public, et s'interroge sur les orientations et sur les nouvelles formes que le musée pourrait éventuellement prendre ». Cette vaste définition prend en compte les différentes préoccupations des musées mais reste assez vague sur l'influence des nouvelles technologies dans la sphère culturelle. Elle ne précise d'ailleurs pas que la muséologie est une discipline rattachée au champ des sciences de l'information⁶⁸.

Afin d'officialiser cette union, il convenait de créer des cursus universitaires qui intégraient à la fois des modules de gestion muséale dans les formations de bibliothéconomie et des programmes de muséologie dans les départements d'études culturelles. Le Canada (Université de Laval, Projet DOCAM ...), les Etats-Unis et l'Australie, pour ne citer que ces pays, sont bien plus avancés en ce domaine que la France qui

⁶⁶ Guillaume Nuttin, « Il est absurde d'avoir un site web 2.0 et un musée 1.0 », Archimag, n°234, p42

⁶⁷ Interview de Corinne Welger-Barbosa, voir question 5 page 104-105

⁶⁸ Voir interview de Corinne Welger-Barbosa, page 102-109 et article de James Turner, *Airs de famille : DOCAM, entre la muséologie et les sciences de l'information*, Documentation et bibliothèques, Oct/déc 2009, p 154

commence à s'y intéresser. L'école du Louvre a justement créé un Master en muséologie et nombres de conservateurs se sont formés auprès d'instances plus techniques pour développer leurs compétences.

On peut envisager cette convergence des deux domaines comme une solution, voire une aide au développement du Web sémantique. En effet, les musées, comme toute autre organisation possédant un capital d'informations à valeur ajoutée, se doivent de maîtriser ce qu'ils diffusent et faire un effort de structuration des données⁶⁹ pour les rendre accessibles à un public profane comme un public d'initiés. La mise en forme de ces informations compte tout autant, l'interface doit être claire et attractive, notamment pour séduire les jeunes internautes.

4.2. L'usage du Web par les musées

Au début des années web, les responsables de ces institutions culturelles⁷⁰ craignaient que la création d'un site Internet, ou « musée virtuel » nuise à la fréquentation et que les usagers se contentent d'une version factice du musée. Dès lors, ils s'investissaient assez peu dans la création et la mise à jour de leur site. Comme le dit si justement Corine Welger-Barbosa⁷¹ : « la notion de musée virtuel se répand à la façon d'un concept mou ». Du moins était-ce le cas il y a une dizaine d'années. Depuis, le dynamisme des musées américains et anglais, par exemple, a bousculé les musées français dans leur inertie et leur a permis d'entrer réellement en contact avec le public par le biais des outils sociaux. Les sites se sont aussi nettement améliorés : les visiteurs peuvent créer leur espace personnel et constituer leur propre collection, les fiches des œuvres sont enrichies et sont valables dans plusieurs langues, il y a possibilité d'agrandir les photos et de zoomer sur certaines parties. Les internautes peuvent s'abonner aux flux RSS pour obtenir les dernières actualités.

Facebook sert à nouer de nouvelles relations avec le public, mais surtout à approfondir celles déjà existantes avec des internautes qui fréquentent à la fois le site et les musées. Il permet de lancer des événements, tels que des vernissages, et de publier des photos ou vidéos, même si elles n'ont pas pour vocation d'être vraiment diffusées sur ce support. La page Facebook peut être perçue comme un « produit d'appel » ou une vitrine miniature du musée qui donnerait l'envie à l'internaute d'en savoir plus. Il peut d'ailleurs devenir « fan de » grâce à l'application pourvue à cet effet. Si à la base, les musées comptaient sur Facebook pour attirer une jeune clientèle, cela n'a pas vraiment fonctionné. En

⁶⁹ Geneviève Vidal, Pour en finir avec le Web 2.0 : vers le Web 3.0 dans les musées », Documentation et bibliothèques, octobre/décembre 2009, p 205

⁷⁰ Il s'agit principalement du cas français, nous verrons par la suite que les musées anglophones étaient moins farouches dans ce domaine.

⁷¹ Corinne Welger-Barbosa, Du musée virtuel au musée médiathèque : le patrimoine à l'ère du numérique, p 119

effet, la tranche d'âge des membres⁷² qui s'intéressent aux musées sur Facebook se situe entre 35 à 49 ans, un quart d'entre eux ayant plus de 50 ans.

Les musées tiennent aussi des blogues leur permettant d'établir une extension de leur site sur le Web. Il a longtemps été reproché aux musées de ne pas proposer sur leur site des liens avec d'autres ressources numériques et de cultiver cette réputation d'être hors du temps, figé et dépassé. Grâce à ce nouvel espace d'expression, les institutions culturelles font leur autopromotion en informant sur les expositions, les projets développés sur un ton journalistique. Il s'en suit des discussions entre les internautes sur ces sujets. Le blogue est un outil assez souple et offre divers usages (diffusion, communication, veille ...) en fonction de la personnalité et des besoins de la structure. A l'inverse, Twitter devient un canal de communication privilégié entre les musées et les journalistes. Le format des informations (brèves) convient parfaitement à ce type de public qui doit se nourrir d'informations en permanence et dispose de peu de temps pour fouiller les sites des musées.

Pour les novices, Youtube (ou Dailymotion) est un site de diffusion et de partage de vidéos. Beaucoup de musées disposent d'une page personnelle sur cette plate-forme exogène et communiquent grâce à l'image animée, censée capter davantage l'attention de l'auditoire. Il peut s'agir de visites virtuelles, d'expositions sur une thématique donnée ou de faire découvrir un métier en présentant un membre du personnel. Ce sont en quelque sorte des magazines culturels⁷³ (ils sont numérotés) diffusés mensuellement. L'internaute a ainsi l'illusion de pénétrer dans le cercle intime des musées. Ces spots, voulus assez courts (moins de cinq minutes), correspondent davantage aux goûts d'un public assez jeune même si les musées visent la communauté en général. Les musées ont tendance à reproduire leur identité visuelle sur les différents supports qu'ils exploitent. Ainsi, en cliquant sur les liens du site, l'utilisateur peut avoir l'impression d'y être encore. On peut supposer qu'ils le font pour plusieurs raisons : pour asseoir leur identité, pour montrer qu'ils ont réussi à maîtriser ces nouveaux outils et pour permettre aux internautes de les reconnaître instinctivement.

Un autre outil, un peu moins cité dans la presse spécialisée mais tout aussi utilisé, est le podcast. Il peut être assimilé aux vidéos mais il existe aussi en simple format sonore, tout comme le principe des guides audio qui existent déjà depuis longtemps dans les musées. Le Tate, musée britannique, a nommé sa compilation de podcasts « Tateshots ». Cet ensemble est composé de plusieurs rubriques telles que : « Story of a Masterpiece », « Work in focus » ou bien « Current Exhibition ». Certaines conférences filmées sont également transformées en podcast sonore et sont téléchargeables gratuitement. Le contenu sur ces sites est très riche et n'a rien à envier aux radios dites culturelles. Le Walker

⁷² Sheila Carey, Musées et médias sociaux, Documentation et bibliothèques, octobre/décembre 2009, p 182

⁷³ Aurélie Chossegros, les musées au risque du web 2.0, article publié en 2008 sur le site web de l'observatoire critique (voir bibliographie en fin de mémoire pour l'URL)

Art Center a d'ailleurs intitulé sa rubrique de podcast « Walker Channel » comme s'il s'agissait d'une chaîne de télévision.

Flickr est de loin la plateforme d'échange la plus populaire. La grande majorité des musées y ont un compte et diffusent des photographies d'œuvre. Le droit de l'image, qui était un frein à la diffusion de contenus auparavant, devient une moindre préoccupation grâce à la licence « Creative Commons ». Ce nouvel arsenal juridique est une alternative au Copyright, jugé trop rigide. L'auteur peut décider d'abandonner une partie ou tous ses droits (à condition qu'il en possède la totalité) sur l'œuvre pour favoriser la libre circulation et le partage. Flickr s'avère être un terrain d'expérimentation particulièrement apprécié, autant par les institutions culturelles que par les chercheurs. En effet, des musées comme le Brooklyn Museum⁷⁴ ou le MET (Metropolitan museum of New York) incitent les visiteurs à prendre des photos des œuvres, se prendre en photo avec et poster les photos sur le groupe. D'ailleurs, on retrouve cette relation entre le sujet (le visiteur) et l'objet (l'œuvre d'art) dans la manière d'indexer l'image. Ce n'est pas tant l'œuvre en elle-même qui compte mais le lien tissé avec elle. Par ailleurs, les marques et les références de l'appareil photo sont souvent mentionnées dans les tags, ce qui est assez surprenant.

Certains musées ou institutions (Library of Congress par exemple) ont exposé leurs propres images en ligne pour observer le comportement des internautes. Ils ont analysé les annotations, et lorsque celles-ci étaient jugées pertinentes, elles étaient reprises pour enrichir l'indexation faite par le musée. Ensuite, ils ont essayé de retracer le circuit des images pour voir comment elles étaient réutilisées : copiées sur des sites personnels, sur Facebook ou sur des blogues. Cette initiative fut couronnée de succès puisque la bibliothèque du Congrès a constaté un boom d'activité sur son propre site grâce à Flickr. La plate-forme de partage a rebondi sur cette réussite en créant « The Commons » pour encourager d'autres organisations culturelles à diffuser leurs images. Il est toutefois important de préciser que les œuvres les plus prestigieuses, régies par des droits d'auteur bien plus stricts, ne figurent évidemment pas sur un site tel que Flickr. Une autre expérience de plus grande envergure a été entreprise par le STEVE Museum⁷⁵ (que l'on peut traduire comme « Le musée de Mr Tout le monde »). L'équipe de chercheurs de cette association prend très au sérieux cette nouvelle forme de marquage social et estime que le tagging est en voie de révolutionner la manière d'indexer les collections d'œuvres des musées. Entre 2006 et 2008, ils ont sollicité des utilisateurs et leur ont demandé de décrire rigoureusement une série d'images, les termes devant ensuite être exploités par la galerie d'art. Il en est ressorti deux constatations : d'une part, le vocabulaire du grand public et celui des

⁷⁴ Très souvent cité dans la littérature spécialisée, ce musée est un pionnier dans la maîtrise des nouvelles technologies, il semble avoir toujours une longueur d'avance sur les autres.

⁷⁵ Sheila Carey, Musées et médias sociaux, Documentation et bibliothèques, octobre/décembre 2009, p 184 + voir annexe 5 page 112.

professionnels est réellement différent, d'autre part les folksonomies et les tags améliorent notablement l'accessibilité aux collections des musées. Dans un style plus ludique et plus léger, le musée Mc Cord (Montréal, Canada) a lancé un jeu d'annotations : Un joueur, jumelé à un autre, doit décrire le maximum d'images en trois minutes. Dès que les cyber-joueurs ont noté le même mot, l'image disparaît laissant place à une autre, et ainsi de suite jusqu'au temps écoulé. Cécile Tardy⁷⁶ a analysé le traitement de la photographie par le musée Mc Cord : « Ce qui est intéressant avec le musée Mc Cord, c'est le renouvellement de l'encadrement visuel de la photographie à partir du jeu, forme très représentative du rapport entretenu par le site à la photographie : ludique (manipulation de l'image), individuelle (prise en charge de l'internaute dans son programme d'activité), éditoriale (privilège la transformation de la photographie en objet éditorial de nature différente : carte postale, affiche, duo, photo ...) ».

Dans un premier temps, nous pouvons dresser un double bilan suite à ces propos. D'une part, nous observons que ce sont surtout les pays anglophones (lesquels n'ont pas le même rapport à la culture que les français) qui sont à l'origine de ce type d'initiatives. La France se fait encore assez discrète dans l'investissement des nouvelles technologies, même si des musées comme celui du Quai Branly offrent des prestations intéressantes. Il demeure une réticence de la part des représentants de la culture à partager un savoir et à lâcher du lest quant aux droits des images. Comme le dit Corinne Welger-Barbosa dans l'interview, nous ne sommes pas encore très éloignés de la situation qu'elle a dépeinte dans son ouvrage paru en 2001. Les directeurs de musée pratiquent encore massivement la surtarification des images, comportement qui n'est pas vraiment en phase avec le contexte actuel du tout numérique. D'autre part, les musées ou autres institutions culturelles (bibliothèques, Archives) ont pour responsabilité de gérer et de (faire) respecter le patrimoine, ils doivent donc contrôler et gérer la diffusion de leurs contenus. Le musée est considéré par le public comme une source d'information fiable et doit absolument le rester. Le public est également hétérogène et les responsables doivent en tenir compte. Les historiens d'art, entre autre, ont des attentes spécifiques et n'apprécient guère les annotations et commentaires parfois (ou souvent ?) futiles des autres usagers. Utiliser des plates-formes exogènes, telles que Facebook ou Flickr, permet justement aux musées d'assurer une séparation nette entre un contenu rationnel et la production d'informations assez disparate réalisée par les internautes. Le site reste donc une source fiable et contrôlée tandis que les réseaux sociaux invitent à une plus grande liberté. L'arrivée du Web

⁷⁶ Cécile Tardy, la photographie documentaire et l'épreuve du dessin informatisé : mise en contexte sur les sites web de musée, Documentation et bibliothèque, janvier/mars 2010, p 21

sémantique constitue une belle opportunité pour les organisations culturelles qui sont plus que jamais concernées par la structuration des informations et l'accès raisonné au savoir.

Enfin, on peut émettre quelques réserves quant à l'influence des réseaux sociaux ou des sites web de musée. Toutes les informations délivrées par les musées, surtout sur le plan événementiel, profitent à un cercle assez restreint et sur une zone géographique assez limitée. En effet, d'une manière générale, ceux qui fréquentent les musées consultent les sites et vice versa. On admettra également que les internautes de la région parisienne se sentent plus concernés par les actualités du musée d'Orsay qu'un américain de l'Arkansas, cela va sans dire. De plus, il est difficile de mesurer le rapport entre visite virtuelle et réelle, il semblerait que la fréquentation des musées⁷⁷ stagne bien qu'elle ait augmenté de 4% entre 2006 et 2007.

⁷⁷ Chiffres du Ministère de la Culture et de la Communication

3^{ème} partie : Conception d'un
thésaurus iconographique pour le
compte du Petit Palais

1 Présentation du contexte

1.1 Bref historique sur la création du musée

Construit par l'architecte Charles Giraud pour l'exposition universelle de 1900, le Petit Palais est devenu un musée en 1902. Pour comprendre un peu mieux sa raison d'être, il est nécessaire de rappeler que jusqu'à la fin du XIX^{ème} siècle, la capitale ne possédait aucun musée. La majorité des collections était conservée au Museum, l'actuel musée du Louvre, et le reste dans les musées de province. Le musée Carnavalet, consacré à l'histoire de Paris, fut inauguré en 1881 et devint officiellement le premier musée détenu par la ville. Auparavant, tableaux, sculptures et autres œuvres d'art étaient exposés à l'hôtel de Ville, dans des salons, mais n'étaient pas accessibles au public. Grâce à l'exposition universelle, les œuvres purent être enfin exposées et admirées et le Petit Palais devint officiellement le musée des Beaux-arts de la ville de Paris.

La collection du musée provient en grande partie de legs et de donations. Le legs le plus important est celui des frères Dutuit en 1902, qui a notamment donné son nom à la bibliothèque du musée. Il y eut aussi celui de Girardin en 1953. Les donations Tuck en 1930 et Zoubaloff en 1935 ont également permis d'enrichir la collection.

De 2001 à 2005, le Petit Palais a connu une grande phase de rénovation. Les collections permanentes et temporaires ont chacune leur zone, et de nouveaux espaces tel que l'auditorium ont été créés. Ces travaux ont été engagés afin de satisfaire les multiples ambitions des services scientifiques et pédagogiques : une surface d'exposition permanente plus vaste, rendre le parcours des visiteurs plus pertinent, faciliter l'accès à tous les publics (handicapés moteurs et sensoriels) et installer des lieux de convivialité comme le café sur deux étages et le jardin réaménagé. La Réunion des Musées Nationaux tient une petite librairie où les visiteurs peuvent acheter ouvrages et souvenirs.

Le musée est dirigé depuis 1998 par Gilles Chazal, conservateur général du patrimoine. Les conservateurs forment une équipe d'une dizaine de personnes environ. Au service documentation, une bibliothécaire gère la collection d'ouvrages et du plan de classement (actuellement en révision : la classification Dewey devrait être conservée mais sera plus détaillée), une documentaliste s'occupe de créer et d'enrichir les dossiers d'œuvres et d'artistes, l'administratrice de la base de données vient de clore le dossier de la base de données Adlib, outil commun à l'ensemble des musées de la ville. En temps que membre de l'équipe « Projet informatisation des collections », elle travaille sur le suivi de la mise en œuvre de l'expression des besoins des utilisateurs (tests, rédaction d'anomalie ...) et forme

les utilisateurs de tous les musées de la ville de Paris ainsi que les services centraux de la Direction des Affaires Culturelles puis rédige les notices sur la base.

Enfin la chef de service supervise les différents membres et est en train de revoir progressivement le système des archives. Elle est également à l'étude sur la documentation numérique en ligne et a réalisé une enquête (en association avec la personne chargée de la collaboration scientifique) auprès d'historiens de l'art de l'INHA et d'autres institutions. Le fruit de ce travail fera l'objet d'un article à paraître à la fin de l'année dans le livre blanc de l'AGCPF⁷⁸.

1.2 Contexte du projet

La réalisation d'un thésaurus iconographique vient se greffer sur d'autres projets actuellement en cours. Un nouveau site Internet, très attractif, a été mis en ligne au mois de mai. Il offre une documentation claire et précise sur les œuvres (Ex : n° d'inventaire, mode d'acquisition) ainsi que plusieurs photographies pour les illustrer. L'internaute peut zoomer sur les clichés pour apprécier certains détails que la photographie de plain-pied ne permet pas de visualiser. Tous les événements du musée sont judicieusement mis en valeur et le visiteur peut s'abonner aux flux RSS pour recevoir les dernières actualités.

Depuis ce mois de septembre, le Petit Palais est doté d'une toute nouvelle base de données Adlib, système de gestion des collections commun à tous les musées de la Ville de Paris qui permet de gérer à la fois les collections, les archives mais aussi la bibliothèque et la documentation. Jusqu'à présent, il existait plusieurs applications complémentaires mais séparées pour documenter, gérer et consulter les informations liées aux collections (Gcoll et Navig'art créés par Videomuseum) et pour cataloguer les ouvrages de la bibliothèque (Alexandrie). Dorénavant, la nouvelle base devrait permettre de tout concentrer grâce à différents modules. Un thésaurus général permet de naviguer dans le logiciel mais aussi d'indexer et de retrouver des œuvres. Il est divisé en trois segments : thème représenté, sujet, lieu. Il y a une entrée spécifique pour les personnages.

Le service documentation m'a confié la mission de réaliser un thésaurus iconographique destiné à être intégré à cette base de données, et par extension au site Internet (puisque la base est déjà connectée au site). Il doit permettre de répondre à des requêtes générales ou spécifiques de la part du public, telles que « le sport dans l'art », ou « le suicide dans l'art ». La collection d'œuvres étant très riche (environ 45000) et ma mission assez courte, je devais me concentrer sur la peinture ancienne (plus de 200 œuvres) pour construire le thésaurus. La peinture ancienne est assez figurative, les symboles

⁷⁸ Association générale des conservateurs des collections publiques de France

religieux et mythologiques sont nombreux, et présente un avantage pour l'indexation. A terme, ce thésaurus sera enrichi pour couvrir toutes les œuvres du Petit Palais et les collections des musées de la ville de Paris (Carnavalet, Cernuschi, Bourdelle etc.). Voici quelques données chiffrées au sujet de l'image au Petit Palais :

- Nombre d'œuvres avec au moins une image : 4160
- Nombre d'images sur la base de données Adlib : 4561

2 Image et représentation d'œuvres d'art

2.1 Qu'est-ce qu'une image ?

Sur le plan étymologique, nous savons que le terme image vient du latin *imago* et désigne les masques mortuaires. D'après Wikipédia, « une image est « une représentation visuelle voire mentale de quelque chose (objet, être vivant et/ou concept). Elle peut être naturelle (ombre, reflet), ou artificielle (peinture, photographie), visuelle ou non, tangible ou conceptuelle (métaphore), elle peut entretenir un rapport de ressemblance directe avec son modèle ou au contraire y être liée par un rapport plus symbolique ». D'une autre façon, nous pouvons distinguer deux familles : « image mentale » et « image perceptible ». Elles ne s'opposent pas dans la mesure où la première précède souvent la seconde. L'œuvre d'un peintre ou d'un photographe est généralement le fruit d'une représentation mentale d'un paysage, d'un personnage ou d'une situation, qu'il s'appuie sur un modèle ou non. Cécile Kattnig⁷⁹ évalue les différentes formes que l'image peut revêtir : « le terme d'image est très largement utilisé mais il prend des significations bien différentes lorsqu'il désigne une affiche, un dessin, une bande-dessinée, une peinture préhistorique, une image de cinéma ou un clip vidéo. L'image recouvre aussi bien les représentations s'inspirant des modèles du monde visible, la sculpture grecque ou celle du Moyen Age, les compositions plus abstraites comme les peintures de Kandinsky mais également l'imagerie scientifique (ondes, rayons X, images macroscopiques ...) ».

La reproduction d'une image est une autre image, beaucoup en conviendront, surtout si nous les considérons en tant qu'objet. Au demeurant, une question se pose : les objets d'art ayant une valeur relative, qu'en est-il de leur représentation photographique ? Quel traitement lui réserver ? Cette citation de Cécile Tardy⁸⁰ peut constituer un début de réponse : « Pour comprendre la photographie documentaire, il est nécessaire de se placer du point de vue de sa pratique en tant qu'outil. Cette pratique ne vise pas à dire la réalité par la photographie, mais à construire une représentation du réel dans le cadre de programmes de travail et de l'exercice de métiers qui requièrent un type d'image descriptive plutôt qu'expressive ».

⁷⁹ Cécile Kattnig, *Gestion et diffusion d'un fonds d'image*, p 5

⁸⁰ Cécile Tardy, *la photographie documentaire et l'épreuve du dessin informatisé : mise en contexte sur les sites web de musée, Documentation et bibliothèque*, janvier/mars 2010p 17

2.2 Spécificité de l'image comme représentation d'œuvre d'art

Pendant longtemps, la reproduction d'œuvre d'art a fait les frais d'une part d'une iconophobie⁸¹ latente et de l'autre de la condescendance de l'élite. On allait même jusqu'à considérer que la photographie était un « faux » et l'utilisateur un être superficiel. Sa fonction de transmission de la connaissance n'était certainement pas acceptée, voire méprisée. Que l'œuvre d'art fut un sujet d'analyse documentaire fut également difficilement admis. Il semblerait que l'ère numérique ait permis de revoir ces positions dépassées et de reconsidérer la photographie d'œuvre d'art: « la photographie documentaire n'est plus considérée seulement un élément technique de description d'un objet, mais comme un élément explicatif d'un objet par rapport à une thématique plus vaste. Ainsi la photographie d'inventaire porte une orientation de lecture qui la dépasse, articulant un savoir ethnologique large autour de l'objet qu'elle représente et un sujet thématique recouvrant d'autres objets⁸² ».

Outre le statut de l'image, l'indexation de la représentation d'œuvre d'art soulève elle aussi quelques interrogations :

- **La polysémie de l'image** : cet aspect met fortement mal à l'aise les personnes qui doivent indexer l'image. Comment ne pas se tromper ? Effectivement, Matisse disait lui-même : « Une œuvre d'art a une signification différente selon l'époque dans laquelle on l'examine ». Concernant une peinture abstraite, comment indexer sans trahir l'artiste ? En même temps, un peintre qui réalise un tel type d'œuvre ne laisse-t-il pas à chacun la possibilité de l'interpréter comme il le souhaite ? Ainsi, pour faire face à cette difficulté, le documentaliste peut indexer le texte qui accompagne l'image. Par ailleurs, comme le dit G.Régimbeau dans sa thèse, le texte a toujours prévalu à l'image. Il lui est naturellement supérieur, et quand bien même le texte serait en proportion inférieur à l'image, on continue de parler de « texte illustré ».

- **L'indexation par unité ou par lot** : tout dépend de la logique de la structure mais il s'avère que les bibliothèques qui doivent cataloguer une masse de documents ont tendance à privilégier le traitement par lot, par gain de temps mais aussi parce que les coûts documentaires sont moindres. Par contre, il est peu concevable d'imaginer que cela puisse convenir aux musées. Une reproduction d'œuvre a beau ne pas avoir la valeur de l'œuvre, elle en est le reflet, le messager. Une indexation erronée de cette photographie, par

⁸¹ Gérard Régimbeau, *Thématique des œuvres plastiques contemporaines et indexation documentaire*, p 346

⁸² Cécile Tardy, *la photographie documentaire et l'épreuve du dessin informatisé : mise en contexte sur les sites web de musée, Documentation et bibliothèque*, janvier/mars 2010, p 20

ricochet, ne porte t'elle pas atteinte à l'œuvre ? Si des spécialistes (muséologues ...) concèdent une valeur nulle à la reproduction d'œuvre, comment expliquer alors que les photographies d'œuvres les plus prestigieuses soient totalement absentes des sites collaboratifs comme Flickr ? Pourquoi ne pas admettre que les responsables de patrimoine ne veulent pas prendre le risque d'une insulte à l'œuvre, à l'artiste et au patrimoine tout court ?

- **Informations connotatives/dénotatives** : Dans quelle mesure l'indexeur doit-il intégrer des informations dénotatives dans la description de l'image ? N'oublions pas que l'indexation est directement liée à la recherche. Le documentaliste doit se préoccuper des besoins du public qu'il vise, non seulement avec le vocabulaire utilisé mais aussi avec le choix des informations divulguées. Les initiés seront demandeurs de ce type de renseignements, le grand public, pas forcément. Gérard Régimbeau cite René Berger qui proposa une méthodologie pour analyser une œuvre d'art :

- 1^{ère} étape : extraire les informations de base : auteur, date, technique, dimensions ...
- 2^{ème} étape : catégoriser l'œuvre en lui attribuant un genre, un sujet
- 3^{ème} étape : Expliquer l'œuvre en mettant en relation la forme et le contenu (couleur, valeur, composition etc.

3 L'utilisation des langages documentaires au sein des musées

3.1 Point sur la situation

L'arrivée d'Internet et la publication en ligne des collections a quelque peu obligé les musées à utiliser des outils adéquats pour structurer les informations liées aux œuvres pour les exposer sur la toile, surtout dans l'éventualité d'un Web sémantique. Comme le dit Madeleine Lafaille⁸³, « Les langages documentaires, en tant qu'outils efficaces d'organisation et de représentation, peuvent contribuer à cette mise en valeur des collections de musée ». Que ce soit sur le plan de l'indexation ou la recherche, du documentaliste à l'utilisateur final, le langage documentaire s'avère l'outil ad hoc pour présenter une information structurée, et à fortiori dans les musées qui détiennent un capital de savoir très important. A l'heure actuelle, il existe en muséologie un certain nombre de vocabulaires contrôlés mais on notera toutefois que les listes d'autorité et les thésaurus y sont peu nombreux, et très rarement en langue française.

3.2 Les principaux langages documentaires utilisés par les musées

Nous allons dresser ici un panorama des langages documentaires les plus populaires dans les musées (à l'échelle internationale).

- **The Getty (AAT et TGN)**

The Art and architecture Thesaurus est un thésaurus à facettes mis au point par le Getty information Institut. Amorcé dans les années 70, il a été rendu effectif au début des années 80. Il a pour objectif d'aider au catalogage et à la visibilité des collections dans les domaines de l'art, de l'architecture, des artefacts et des archives. Il a été conçu pour répondre aux besoins des bibliothèques d'art et de la presse spécialisée qui commençaient à automatiser leurs procédures de catalogage et d'indexation. La description du contenu de l'œuvre se fait donc au travers de sept facettes :

⁸³ Madeleine Lafaille, l'utilisation des langages documentaires dans les musées, Documentation et bibliothèques, avril/juin 2006, p 132

- Concepts associés ☞ Toutes les idées connexes qui ont trait à l'air : la beauté, l'équilibre, les différents mouvements artistiques, les idéologies ...
- Attributs physiques ☞ Les caractéristiques visibles et mesurables des matériaux et des objets (propriétés chimiques, qualité de texture, forme ...)
- Style et époque ☞ Informations stylistiques sur les objets et en rapport avec les époques (Style Louis XIV, Bonaparte ou style bayadère ...)
- Agents ☞ Les personnes ou organisations liées aux métiers du domaine étudié. Les animaux et les plantes sont également identifiés sous cette facette.
- Activités ☞ Les activités, les événements, les actions et les processus en relation avec les objets d'art et d'architecture
- Matériaux ☞ Toutes les substances (matière, colorant ...) utilisées dans la création d'œuvre, qu'elles soient naturelles (bois ...), obtenues par mélange ou créées artificiellement.
- Objets ☞ C'est la facette la plus importante. Elle désigne toutes ces choses inanimées et créées par l'Homme. Elles sont de l'ordre de l'utilitaire et de l'esthétique. Cela va de la chaise au jardin, paysage artificiel imaginé par l'humain.

L'AAT répond aux normes ISO et NISO, il a notamment été construit sur les bases du thésaurus MeSH. Il comprend environ 131 000 termes et 34000 concepts.

C'est un thésaurus très bien construit mais assez lourd d'utilisation⁸⁴. Par exemple, pour le concept « entreprise dont une femme est propriétaire », huit expressions ont été retenues. L'indexation à l'aide de cet outil paraît assez fastidieuse et réclame environ quarante minutes pour un documentaliste chevronné. On peut également lui reprocher, tout comme aux LCSH, de s'intéresser principalement à la culture américaine. Cela dit, même si le thésaurus n'existe pour le moment que dans la version anglaise, des collaborations avec d'autres institutions sont en cours et le thésaurus pourrait être traduit dans d'autres langues. L'AAT est disponible en ligne sur Internet et ne fait plus l'objet d'édition papier depuis environ 1997.

En 1987, le Getty Information Institute a également lancé un autre thésaurus, le TGN (Thesaurus of Geographic Names). Il contient autour d'un million de lieux et couvre l'ensemble des continents et des nations du monde entier. Il est complémentaire de l'AAT dans la mesure où il recense notamment les lieux importants dans le domaine des arts et de l'architecture. Le TGN est construit sur des relations hiérarchiques, associations et d'équivalence. Il n'existe également qu'en version anglophone.

⁸⁴ Marie Després-Lonnet, thésaurus iconographiques et modèles culturels, Document numérique, Vol 4, 2000, n°1-2, p 161

- **Iconclass**

Cette nomenclature a été créée par Henri Van de Waal, vers 1950, dans l'objectif de classer la documentation iconographique d'histoire de l'art des Pays-Bas. Suite à son décès en 1972, les travaux de recherche ont continué et la version finale d'Iconclass a été publiée à la fin des années 70, notamment grâce au soutien de la KNAW (Académie néerlandaise des arts et des sciences) et dont le professeur avait été membre.

Ce système comprend neuf classes :

- | | |
|------------------------------------|-----------------------------------|
| 1. Abstrait/Art non représentatif | 5. Société, Civilisation, Culture |
| 2. Religion et Magie | 6. Idées abstraites et Concepts |
| 3. Nature | 7. Histoire |
| 4. Etre humain et Homme en général | 8. Bible |
| | 9. Littérature |

La notation alphanumérique permet de rendre compte de manière très réaliste certaines activités ou certaines scènes, notamment celles décrites dans la Bible. Ce système a l'avantage de limiter l'ambiguïté et les termes vides puisque ceux-ci sont contextualisés. Il a été conçu pour répondre à des besoins très précis et s'avère exhaustif pour recenser les personnages bibliques et la mythologie.

Sa richesse et sa précision le rendent toutefois difficile à appréhender et à adapter pour une collection assez étendue dans le domaine de l'art.

- **Ethnophoto**

Ce thésaurus, qui date de 1977, a été créé spécifiquement pour indexer les photographies du Musée national des Arts et traditions populaires (MNATP). Environ 500 000 images sont indexées aujourd'hui grâce à cet outil. Sa constitution est hybride dans la mesure où des termes spécifiques ont été choisis pour décrire de manière factuelle la vie des hommes sur un plan ethnologique (métier, psychologie, histoire ...) mais il a aussi emprunté au Thésaurus Garnier la partie consacrée à l'iconographie religieuse.

L'analyse comparative que Marie Després-Lonnet a effectuée a révélé que le mot « femme » n'existait pas en tant que tel dans Ethnophoto mais figure dans une quinzaine d'expressions : « chaussure de femme », « costume de femme », « travail des femmes » etc. L'utilisateur qui lance une requête sur le terme « femme » peut donc voir s'afficher des images qui ont trait aux femmes mais qui n'en représentent aucune. Par ailleurs, le fait d'avoir incorporé un descripteur directement lié à une exposition (ATP mari et femme) est

certes novateur mais en même temps le rend peu utilisable car le préfixe ATP n'est connu que du gestionnaire ou de l'indexeur.

- **TGM (Thesaurus for Graphic materials)**

Le TGM a été produit par la bibliothèque du Congrès. Il se veut assez généraliste, pour couvrir l'ensemble des sujets potentiels et les différents supports de l'image (bande-dessinées, plans, dessins ...) de la collection du département « Prints and photographs ». Il est possible de rajouter des facettes à ce thésaurus et les termes peuvent être pré-coordonnés. Le descripteur « femme » fait l'objet de 29 termes spécifiques et neuf termes associés. Il peut être combiné à des catégories telles que « ethnie », « personne handicapée » ou encore « indigène », ce qui assouplit son utilisation. Il a ensuite été proposé à d'autres institutions, son amplitude lexicale et sa structure en faisant un outil adaptable.

- **Le thésaurus Garnier (thésaurus iconographique : système descriptif des représentations)**

Ce thésaurus a été élaboré par François Garnier à la demande du ministère de la Culture. Il fut publié pour la première fois en 1984. Il devait servir de support pour indexer les œuvres et les objets d'art dans leur ensemble. La base de données Joconde se sert du thésaurus Garnier pour décrire les collections des musées français. Il est divisé en trois catégories : la principale étant la description de la représentation (elle-même subdivisée en deux parties : thème et sujet), viennent ensuite les sources écrites et la datation. La sous-rubrique thème comprend quinze classes qui sont liées aux attributs généraux de l'image : genre iconographique, angle de vue etc. La sous-partie « sujet » se compose de huit classes qui définissent les lieux, les sujets, les personnages.

Pour reprendre l'exemple cité précédemment pour d'autres thésaurus, le terme « femme », qu'il soit isolé ou combiné à d'autres termes génère beaucoup de bruit, environ 26000 notices, ce qui en fait un descripteur peu pertinent pour la recherche.

- **Le thésaurus des images médiévales**

Il a été conçu par le Groupe d'anthropologie historique de l'Occident pour assister les médiévistes dans la réalisation de bases de données iconographiques. Il est considéré avant tout comme un outil d'aide à la recherche. Ainsi, au moment de l'indexation, certains termes, comme « femme », sont d'office éliminés car peu pertinents lors de la requête. Les personnages féminins sont donc décrits par leur nom, leur fonction ou leur aspect physique.

Comme son nom l'indique, c'est un thésaurus très spécialisé, et à priori seuls les spécialistes de la période médiévale sauront l'exploiter de manière optimale.

Voici la liste non exhaustive des thésaurus les plus populaires dans le monde muséal. Toutefois, d'autres outils tels que le thésaurus des objets religieux ou celui du British Museum peuvent être intéressants pour inspirer ou enrichir un thésaurus existant mais incomplet dans un domaine particulier.

3.3 Compte-rendu de visites

Dans le cadre de ma mission, j'ai effectué quelques visites de musées et j'ai pu observer les différents outils utilisés pour l'indexation⁸⁵. Au musée du Louvre, pendant très longtemps, les professionnels se sont servis d'index iconographiques pour repérer les œuvres représentant un thème précis. Un numéro d'inventaire était associé à un descripteur. Ils disposent maintenant d'un thésaurus directement intégré à leur base de données. Je me suis également rendu aux Archives départementales du Morbihan, (qui possède un fonds notable de photographies et qui organise régulièrement des expositions), je n'ai pas pu obtenir d'informations précises mais il m'a semblé que l'organisation de la collection reposait avant tout sur un simple index et la mémoire du personnel dévoué à cette fonction. Il serait intéressant pour cette structure d'envisager un système plus construit, vecteur de connaissance pour les potentiels salariés nouvellement recrutés.

D'une manière générale, j'ai constaté que la plupart des musées visités s'étaient inspiré du thésaurus Garnier et qu'ils l'avaient adapté en fonction de la collection. Celui du musée d'Orsay en est un exemple mais ne possède pas de relations hiérarchiques. Non pas qu'ils en soient spécialement satisfaits mais ils ont estimé que c'était un outil de base et que rien de plus pertinent n'avait vu le jour depuis, du moins en français. J'aimerais attribuer une mention spéciale au musée Cluny qui a réalisé un très bon thésaurus accompagné d'un guide d'utilisation.

La RMN⁸⁶, qui gère une collection d'images très importante (elle gère les collections de 35 musées nationaux), possède un métathésaurus composé de six branches dont le thésaurus commun qu'elle m'a présenté. Cette organisation a un statut spécifique car depuis 1990, elle est devenue un institut à caractère industriel et commercial. Elle vend donc des photos et sa politique d'indexation doit refléter cette réalité. Par exemple, pour satisfaire des

⁸⁵ Voir annexe 6 p 113

⁸⁶ Réunion des Musées Nationaux

demandes publicitaires, les œuvres sont classées par couleur : dans les tons de bleu, de rouge etc. C'est un outil très précis et une seule personne a la responsabilité de valider ou non les termes candidats. Cela assure au thésaurus son caractère homogène et pertinent. Dans certaines structures visitées, j'ai pu vérifier que lorsque chaque documentaliste pouvait enregistrer des termes dans la base, les homonymes et les erreurs étaient plus nombreuses (fautes d'orthographe, de frappe, termes rentrés au pluriel sans raison ...).

4 Processus d'élaboration du thésaurus iconographique

4.1 Indexation des représentations d'œuvre

Après avoir balayé la littérature spécialisée, il m'a été permis d'en déduire que c'était un sujet assez peu traité d'une manière générale. Comme le dit Corinne Welger-Barbosa⁸⁷, l'indexation d'image était une pratique assez intimiste et qui se transmettait tacitement d'un professionnel à l'autre. J'ai donc élaboré une méthodologie basée sur ma propre réflexion, construite à partir des lectures que j'ai faites. J'ai analysé image par image les représentations d'œuvre et j'ai dégagé de manière instinctive les principaux thèmes qui ressortaient de la collection : paysages, Nature morte, portrait ... Je me suis aidée d'ouvrages spécialisés en art et d'Internet pour traduire ces thèmes avec la terminologie adaptée. Cette étape a donné lieu à une première réunion au Bureau des Musées⁸⁸ (où ma tutrice travaille en alternance pour le projet de la base de données) pour valider cette liste qui constituera le premier niveau du thésaurus. Ce brainstorming nous a permis d'arrêter les termes génériques suivants : Allégorie, Architecture, Genre iconographique, Lieu⁸⁹, Ornementation, Paysage, Psychologie et Morale, Religion, Représentation Humaine, Représentation de la Nature, Représentations d'objets et de concepts, Représentation scientifique et Sujet de société.

Pour décrire chaque image, je m'étais fixé une fourchette de quatre à dix descripteurs. En dessous de quatre, l'indexation est assez vague et à la recherche, il risque d'y avoir trop de bruit. Au dessus de dix termes, cela alourdit le thésaurus et le rend inopérant. J'ai constaté qu'en moyenne, sur les deux cents peintures décrites, le nombre de descripteurs était de six ou sept, et cela correspondait aux préconisations des experts dans ce domaine. Les peintures représentant des scènes mythologiques ou des natures mortes étant souvent celles qui donnaient lieu à une indexation plus riche. Je me suis également aidée des dossiers d'œuvre pour compléter l'indexation d'une œuvre qui me posait problème et aussi par simple souci d'objectivité et d'exactitude. J'avais toutefois conscience que ne pouvais pas être aussi exhaustive que possible et que ce n'est pas l'objectif d'un thésaurus.

⁸⁷ Voir interview en annexe page 102-109

⁸⁸ Voir agenda du stage en annexe 1 page 98

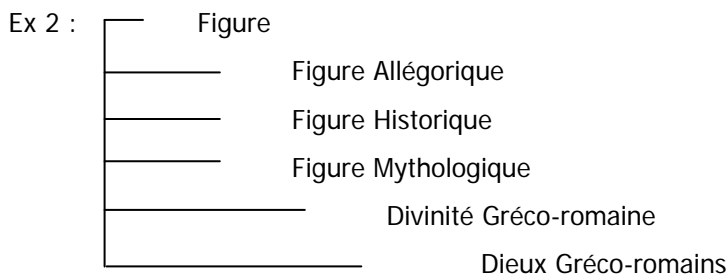
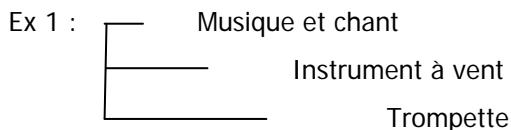
⁸⁹ Les lieux feront l'objet d'une autre entrée dans le thésaurus

4.2 Structure et relations entre les termes

Le thésaurus du Petit Palais est structuré sur deux axes : hiérarchique et alphabétique. Je me suis basée sur les cours que nous avons eus à l'INTD et les normes en vigueur. Le thésaurus comporte donc des relations hiérarchiques, d'équivalence et associatives. Dans la mesure du possible, et conformément à la norme AFNOR, les termes sont au singulier (contrairement aux musées canadiens qui utilisent le pluriel pour décrire les objets de musée). Afin de démontrer les différentes relations qui lient les termes entre eux, il m'a paru opportun de les illustrer par des exemples concrets :

- **Relation hiérarchique :**

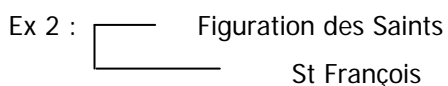
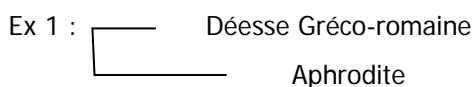
- Générique :



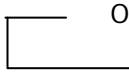
- Partitive

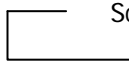
Cette relation n'est pas encore très développée actuellement car le thésaurus est encore à l'état d'ébauche. Cela dit, le terme spécifique « Corps humain » appartenant à la catégorie « représentation humaine » donnera lieu à une relation hiérarchique partitive.

- d'instance



- **Relation associative**

Ex 1 :  Ornementation
Animal fabuleux (**TA représentation animalière**)

Ex 2 :  Scène
Scène de genre (**TA Sujet de société**)

- **Relation d'équivalence**

Ex 1 : Virginal = clavecin

Ex 2 : Marie de Nazareth = Madone, Sainte Vierge, Vierge, Marie

Le thésaurus de la base de données Adlib permet d'enregistrer un nombre illimité de termes équivalents ou associatifs, ce qui offre la possibilité d'un champ lexical assez riche et d'une recherche fructueuse.

4.3 Contraintes et difficultés

En réalisant ce thésaurus, je me suis heurtée à un certain nombre d'obstacles, liés en partie au fait qu'il était créé pour la base de données et que je le réalisais sur « papier ». Je ne pouvais pas rendre compte des liens aussi facilement que l'informatique le permet. Par ailleurs, « thème représenté » et « sujet » sont deux entrées différentes sur la base et il m'était difficile de calquer cette réalité dans mon thésaurus qui devait être homogène. Le terme « Sujet » m'a au tout début paru ambigu et je ne savais pas exactement ce qu'il recouvrait. Les sujets étant la plupart du temps les termes spécifiques des thèmes représentés, je ne voyais pas comment les séparer.

Avec l'équipe du Petit Palais, nous nous sommes longtemps posés la question de l'intérêt d'intégrer une catégorie « psychologie et morale ». Elle est certes présente dans la majorité des thésaurus iconographiques que j'ai pu consulter, dont Garnier, mais attribuer des sentiments ou des attitudes aux personnages induit toujours un risque de fausse interprétation. De plus certains sont plus difficiles à faire ressortir sur une œuvre que d'autres. Comment peut-on être sûr de lire la lassitude ou le sentiment d'échec sur le visage d'un personnage, à moins que le titre l'indique explicitement ? D'un autre côté, des attitudes

ou émotions telles que la contemplation ou la souffrance sont bien plus souvent exprimées, notamment dans la peinture religieuse. Il serait donc dommage voire même inconséquent de se priver de ce thème. Il en va de même pour la rubrique « représentation d'objets et concepts » qui mériterait d'être intégrée à la base, même si ce n'est pas une priorité actuellement.

La question de la religion a sans doute été la plus difficile à résoudre. De mon point de vue, l'application stricte du thésaurus Garnier pour cette partie me paraissait trop complexe et le but n'était pas de copier à l'identique un outil existant. En même temps, ce thème étant assez riche en symboles, en personnages et en pratiques, il n'était pas non plus possible de le simplifier à outrance. Les sujets bibliques et les pratiques religieuses étant deux entités bien différentes, nous avons pris le parti de les séparer. Le thème « Religion » a donc été scindé en deux classes :

- Religion-sujet religieux ☞ Sainte Famille, Evangiles, Saints ...
- Religion-Vie religieuse ☞ Organisation, rites, sacrements ...

Cela nous a également permis d'alléger la notation car avec un thème aussi développé, un dixième niveau de hiérarchie est vite atteint et nuit à la lisibilité et compréhension du thésaurus. Grâce à cette division, le dernier niveau de chaque catégorie atteint au maximum cinq chiffres, ce qui demeure tout à fait raisonnable. Le thème « Religion », tout comme le thésaurus d'une manière générale, devra être étoffé voire modifié car il n'est pas certain, le temps nous ayant manqué, que la structure soit tout à fait satisfaisante. Les attributs des Saints ou les Evangiles apocryphes devront notamment être notifiés.

La base de données Adlib trie directement tous les termes par ordre alphabétique, ce qui constitue l'avantage de réduire le travail de l'indexeur. En revanche, cette opération systématique ne respecte pas toujours l'ordre logique des concepts. Ainsi, dans le thème « Représentation de la Nature » et sous classe « Cycle de vie », le terme « mort » vient avant le terme « Vie ». Cela n'est pas choquant en soi pour les individus qui croient en la réincarnation, mais le principe de neutralité est fortement biaisé dans le cas présent.

4.4 Complément et mise à jour

Le langage documentaire que j'ai réalisé pendant ma mission est un micro-thésaurus. Il n'est censé représenter que le corpus de peintures anciennes, bien qu'un certain nombre de termes aideront à l'indexation d'autres œuvres.

Un thésaurus est un outil vivant qui n'est jamais véritablement achevé. Il nécessite un entretien permanent, en particulier s'il est très spécifique à une discipline. Dans un domaine très créatif et en perpétuelle mouvance, comme celui de l'art, il est vivement conseillé d'effectuer une veille régulière (littérature spécialisée, colloques ...) pour repérer les nouvelles tendances, les néologismes. Je nuancerai toutefois mon propos concernant le Petit Palais car celui-ci expose avant tout des collections datant de l'Antiquité jusqu'au début du XXème siècle et que par conséquent la terminologie de base évoluera sans doute moins rapidement.

Un thésaurus informatisé est moins fastidieux à maintenir qu'un support papier. Il suffit d'être vigilant et de vérifier que le terme nouveau n'ait pas déjà été entré sous cette forme ou sous une autre. Généralement, on observe qu'un thésaurus devient dysfonctionnel parce que d'une part le choix et l'intégration des termes sont effectués avec moins de rigueur et d'autre part, les velléités pour le mettre à jour s'affaiblissent au fil du temps.

Conclusion

A travers cette étude sur les langages documentaires, leur transposition sur le Web et leur utilisation dans le contexte culturel permettent de dégager plusieurs remarques. D'une part, ces outils constituent le pivot de la fonction documentaire dans le sens où ils servent à classer et diffuser un savoir organisé. Pour cette raison, il convient de leur apporter une définition claire et définitive. Comme le disait B.MENON, peu de professionnels des sciences de l'information ont remis en cause ce terme de langage documentaire. Est-ce pour cette raison que le flou qui entoure les notions de classifications, ontologies, taxonomies et thésaurus est plus que jamais avéré ?

L'autre raison, qui peut s'avérer complémentaire à la première, voudrait que ce soit l'adaptation de ces langages documentaires au contexte du web, à fortiori au web sémantique, qui soit la cause de cet imbroglio. En effet, transposer des thésaurus a conduit parfois à les simplifier à outrance, quitte à les faire ressembler à de simples classifications. En outre, l'apparition de nouveaux concepts telles que les ontologies et les topics Maps est venu complexifier la situation. A cela se conjugue la multitude des protocoles édictés par le consortium du W3C pour formaliser ces nouveaux modes de représentation. Cependant, on pourra objecter à ces griefs que pour assurer un certain avenir aux langages documentaires et notamment au thésaurus, il fallait envisager quelques réaménagements, au risque effectif de constater un affaiblissement de leur structure. Le Web sémantique en est encore au stade expérimental et il serait de bon augure d'être optimiste quant au développement de cette nouvelle version du web.

La problématique de l'indexation de l'image, et en particulier des œuvres d'art devrait prendre d'ici à quelques années une autre tournure. L'influence des folksonomies et des tags, notamment sur les sites tels que Flickr, est à prendre au sérieux dans la mesure où il a été prouvé que ces outils d'indexation collaborative pouvaient venir enrichir la description de collections d'œuvres. Les professionnels reconnaissent d'ailleurs que la meilleure alternative pour garantir une recherche d'information efficace est de mixer l'indexation en vocabulaire libre et contrôlé. En même temps, il y a une certaine urgence à normaliser la pratique d'indexation de représentation d'œuvre pour faire face aux dérives que l'indexation participative peut engendrer. Peu de personne à l'exception de G.REGIMBEAU et quelques autres s'y sont véritablement intéressées. On peut espérer que cette fonction viendra élargir les perspectives de la muséologie.

Bibliographie

Cette bibliographie analytique a été arrêtée le 15 septembre. Elle se compose de 78 références, réparties selon le plan de classement suivant :

- Langages et outils documentaires :
 - indexation
 - Langages documentaires
 - Thésaurus
- Internet et Web sémantique
- Interopérabilité
- Iconographie
- Muséologie
- Normes
- Ontologies

Ce corpus de documents assez important m'a permis de développer mes connaissances et de réaliser ce mémoire mais ne saurait couvrir de manière exhaustive la problématique et les divers domaines abordés.

LANGAGES ET OUTILS DOCUMENTAIRES

Indexation

[01] Les fondements théoriques de l'indexation : une approche linguistique. AMAR, Muriel. Paris : ADBS, 2000. 355 p. ISBN 2-84365-042-9.

☞ Cet ouvrage, qui est une publication de la thèse de doctorat de M.AMAR, pose une fois pour toutes les fondations de la pratique documentaire de base : l'indexation. Elle constitue une méthodologie de référence pour les professionnels de la documentation et n'hésite pas à s'affranchir des mythes établis pour imposer son point de vue.

[02] Indexation collaborative et folksonomies. FRANCIS, Elie ; QUESNEL, Odile. Documentaliste-Sciences de l'information, 2007, Editions ADBS, Vol. 44, n°1, p 58-63.

☞ Les auteurs recensent dans ce document quatre types d'indexation : la classification personnelle, indexation par l'auteur, indexation par l'utilisateur, et classification globale. Outre cette énumération, l'article repose sur une réflexion sur le risque d'abus et de désinformation que ces nouveaux modes peuvent générer.

[03] L'indexation à l'heure du numérique. MENON, Bruno. Documentaliste-Sciences de l'information, 2004, Editions ADBS, Vol. 41, n° 6, p 340 à 342.

☞ B.MENON revisite à travers cet article la notion d'indexation, notamment dans l'optique des changements qu'occasionne le Web. Il évoque l'opposition entre indexation libre et indexation contrôlée puis conclut l'article en amenant la question des nouvelles compétences et responsabilités professionnelles pour faire face à cette nouvelle donne.

[04] Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires. VAN SLYPE, Georges. Paris : Les Editions d'Organisations, 1986. 277 p. ISBN 2-7081-0760-7. 0

☞ Cette monographie a l'inconvénient de dater un petit peu mais propose une démarche claire et efficace pour élaborer un thésaurus. La lecture d'ouvrages similaires et plus récents ont montré que les processus intellectuels n'avaient pas tant évolué que cela. Qui plus est, G. VAN SLYPE est une figure de référence chez les anglo-saxons.

[05] Indexation des images et des sons : le numérique a-t-il changé les pratiques ? : Journée d'étude ADBS audiovisuel. BOULOGNE, Arlette ; JADOUL, M.B ; MASSE, Claudine ; MELLET D ; PAPPANOT, H. Documentaliste-Sciences de l'information, Editions ADBS, 2006, Vol. 43, n°2, p 136 à 142.

☞ L'article est un condensé de ce qui a été dit lors de la journée d'étude consacrée à l'indexation des images et des sons. Il en est ressorti que cette pratique a connu des bouleversements dû à l'avènement des technologies du numérique.

[06] Description et indexation automatique des documents multimédias : du fantasme à la réalité. GROS, Patrick. Documentaliste-Sciences de l'information, Editions ADBS, 2005, Vol 42, n°6, p 383 à 391.

☞ Ce document traite de l'indexation automatique des images et des vidéos et souligne en particulier les lacunes de ces systèmes à l'heure actuelle.

[07] Culture documentaire et folksonomies : l'indexation à l'ère industrielle et collaborative. ERTZCHEID, Olivier. Documentaliste-Sciences de l'information, Editions ADBS, 2010, Vol. 47, n°1, p 45 à 47.

☞ L'auteur présente les folksonomies, nouvel outil d'indexation sur le mode collaboratif et opérant sur Internet. Il relate notamment les différentes raisons qui expliquent le succès des folksonomies auprès des internautes.

[08] Technologies linguistiques et mode de représentation de l'information textuelle. CHAUDIRON, Stéphane. Documentaliste-Sciences de l'information, 2007, Vol 44, n°1, p 30-39.

☞ S.CHAUDIRON explore à travers ce document les différentes techniques pour représenter l'information, en particulier le TAL (traitement automatique de la langue) qui a pour but de faciliter les processus d'indexation de document. Il traite également de l'évolution des techniques dans le cadre du développement du web sémantique.

Langages documentaires

[09] Les langages documentaires : un panorama, quelques remarques et un essai de bilan. MENON, Bruno. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol 44, n°1, p 18 à 28.

☞ B.MENON dresse le constat d'un passage peu convaincant des langages documentaires sur Internet. Il évoque également l'implication des ontologies comme support des classifications sur Internet et met en balance les notions de langage libre et contrôlé.

[10] Optimiser l'accès à l'information : une opportunité pour les langages documentaires ? ; Journée d'étude ADBS. MENON, Bruno. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°6, p 385 à 388.

☞ Ce compte-rendu examine le rôle que peuvent jouer les langages documentaires dans la recherche d'informations. Il est notamment dit que les thésaurus, s'ils jouent un rôle réel, sont quelque peu malmenés et affaiblis dans leur adaptation au web.

[11] Langages documentaires et outils linguistiques : principes, usages, perspectives. Rupture ou continuité ? MANIEZ, Jacques. Documentaliste-Sciences de l'information, 2007, Vol.44, n°1, p 12-16.

☞ J.MANIEZ tente d'éclaircir la notion de langage documentaire. Il reproche aux spécialistes de ne pas avoir remis en cause ni leur contenu, ni leur appellation. Il compare notamment les différences de conceptualisation entre les pays anglo-saxons et la France.

[12] Le répertoire de vedettes-matière : outil du XXIème siècle. DOLBEC, Denise. Documentation et Bibliothèques, 2006, Vol. 52, n°2, p 99-108.

☞ Depuis sa création en 1946, le répertoire de vedettes-matière a connu une certaine évolution, notamment au gré des innovations technologiques. D.DOLBEC expose les efforts constants de l'équipe du RVM pour maintenir à flot cet outil documentaire.

[13] Actualité des langages documentaires : les fondements théoriques de la recherche d'information. MANIEZ, Jacques. Paris : ADBS, 2002.395 p. ISBN 2-84365-060-7.

☞ J.MANIEZ s'interroge sur le rôle que peuvent encore jouer les langages documentaires à l'aube du XXIème siècle, sachant que l'informatique prend de plus en plus en charge les fonctions basiques de la recherche documentaire : classement alphabétique, text mining ...

[14] Le passage au XXIème siècle des grandes classifications documentaires. HUDON, Michèle. Documentation et Bibliothèques, Montréal : ACBLF, 2006, Vol. 32, n°2, p 85 à 97.

☞ Cet article de M.HUDON est une vue d'ensemble sur les grandes classifications et de leur adaptation sur le Web. Certaines, en fonction de leur structure, ont mieux amorcé le virage du siècle nouveau.

[15] Classifications, thésaurus, ontologies, Folksonomies : comparaison du point de vue de la recherche ouverte d'information (ROI). ZACKLAD, Manuel. In : CAIS/ACSI 2007, 35e Congrès annuel de l'Association Canadienne des Sciences de l'Information. Partage de l'information dans un monde fragmenté : Franchir les frontières, sous la dir. de C. Arsenault et K. Dalkir. Montréal : CAIS/ACSI, 2007. (Consulté le 3/10/2010)

Accès : http://www.cais-acsi.ca/proceedings/2007/zacklad_2007.pdf

☞ M.ZACKALD effectue une comparaison des systèmes d'organisation des connaissances (classifications, thésaurus, ontologies ...) basée sur divers critères pour évaluer le bien-fondé et l'efficacité de ces outils dans le cadre de la recherche d'information.

[16] Les nouvelles tendances des langages documentaires. DAVIES, Ron. Cahiers de la documentation, Mars 2006, p 4-10.

☞ Cet article évoque l'apparition des taxonomies et des ontologies, comme toiles de fond d'un web sémantique en devenir. Il fait également cas de la révision des normes, à l'échelle nationale et internationale.

[17] Thésauroglossaire des langages documentaires : un outil de contrôle sémantique. DEGEZ, Danièle ; MENILLET Dominique. Paris : ADBS, 2001. 181 p. ISBN 2-84365-051-8.

☞ Cet ouvrage m'a avant tout servi pour cerner clairement les notions de langages documentaires grâce aux définitions édictées par ces spécialistes

[18] Controlled Vocabularies: an introduction. LEISE, Fred. The indexer, Society of indexer: Londres, 2008, Vol. 26, n°3, p 121 à 126.

☞ F.LEISE introduit les différences de base entre les différents langages documentaires et de leur usage dans le cadre de la recherche d'information.

[19] L'utilisation des langages documentaires pour la recherche d'information. ARSENAULT, Clément. Documentation et Bibliothèques, Montréal : ACBLF, 2006, Vol. 52, n°2, p 139-148.

☞ L'article présente l'implication des langages documentaires dans le processus de recherche d'information. L'auteur établit tour à tour les forces et les faiblesses des vocabulaires libres et contrôlés.

[20] Thesauri, taxonomies and ontologies : an etymological note. GILCHRIST, Alan. Journal of Documentation, 2003, Vol. 59, n°1, p7-18.

☞ A.GILCHRIST effectue un exercice de définition entre les différentes notions de thésaurus, de taxonomies et d'ontologies en se basant sur l'étymologie des mots. Il était intéressant d'avoir un point de vue anglo-saxon sur ces systèmes de classification, surtout du fait qu'il s'agisse de Gilchrist qui est une référence en ce domaine.

[21] Taxonomies, ontologies, Folksonomies [en ligne]. AMAR, Muriel. URFIST PARIS. Mars 2009, [consulté le 5 octobre 2010].

Accès :

<http://urfist.enc.sorbonne.fr/sites/default/files/file/traitementdoc/TaxonomiesAndCie-s-support-de-formation.pdf>

☞ Ce document Power Point est un support de cours réalisé par M.AMAR pour l'URFIST. Il aborde la question des nouveaux langages documentaires: définitions, enjeux, principes et application dans le cadre du Web sémantique

Thésaurus

[22] Guide pratique pour l'élaboration d'un thésaurus documentaire. HUDON, Michèle ; MENILLET, Dominique. Montréal : Ed. ASTED ; Paris : ADBS, 2009.274 p. ISBN 978-2-923563-17-6.

☞ Cet ouvrage est une réédition revue et enrichie de celui paru en 1995. M.HUDON insiste davantage sur les processus intellectuels qui jalonnent la construction d'un thésaurus et les innovations technologiques qui font évoluer la notion de thésaurus.

[23] Construire un thésaurus : manuel pratique. AITCHISON, Jean ; GILCHRIST, Alan. Paris : ADBS Ed. Cop, 1992. 217 p. ISBN 2-901046-43/ISSN 9782901046431.

☞ Autre méthodologie de référence réalisée par deux spécialistes des thésaurus outre-Atlantique.

[24] Construire un thésaurus. DEGEZ, Danièle. Archimag, Groupe Serda, 2009, n°222, p 44-45.

☞ D.DEGEZ propose à son tour une procédure de base pour concevoir un thésaurus. Le paragraphe sur l'intérêt d'un cahier des charges et les divers éléments qu'il doit contenir a retenu mon attention.

[25] Comment réaliser un thésaurus ? LUBKOV, Michel. Archimag, Groupe Serda, 1994, n°77, p 30-33.

☞ M.LUBKOV propose dans cet article une méthodologie de construction de thésaurus sur dix étapes. En fin d'article, le lecteur trouvera un focus bref mais intéressant sur le nombre de niveaux hiérarchiques.

[26] Thésaurus, l'outil des documentalistes gardiens du trésor. NUTTIN, Guillaume. Archimag, Groupe Serda, 2008, n°219, p 30-32.

☞ Le thésaurus est ici présenté comme l'outil ad hoc pour un double usage : l'indexation et la recherche d'information. L'auteur expose les fonctionnalités essentielles d'un thésaurus informatisé efficient.

[27] Des classifications aux thésaurus : du bon usage des facettes. MANIEZ, Jacques. Documentaliste-Sciences de l'information, Editions ADBS, 1999, Vol. 36, n° 4-5, p 249-262.

☞ Cet article a pour objectif d'éclaircir la notion de facettes. J.MANIEZ reprend le travail de Ranganathan et de ses successeurs pour en établir une analyse fine et constructive.

[28] Thésaurus et informatique documentaire : des noces d'or. DALBIN, Sylvie. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°1, p76-80.

☞ S.DALBIN fait cas de l'adaptation des thésaurus aux logiciels documentaires. Elle dresse un bref panorama de l'évolution de l'informatique documentaire et conclut l'article en évoquant l'avenir des thésaurus sur le Web.

[29] Thésaurus et informatique documentaire : partenaires de toujours ? DALBIN, Sylvie. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°1, p 42-55.

✍ L'ergonomie des logiciels documentaires n'a, selon l'auteur, pas été assez exploitée pour venir en aide à la construction des logiciels d'indexation ou des moteurs de recherche. La question de la transposition des thésaurus sur Internet est clairement posée.

[30] Le thésaurus est un outil de gestion des connaissances. REMIZE, Michel ; DEGEZ, Danielle. Archimag, Groupe Serda, 2004, n°179, p 14-16.

✍ D.DEGEZ, spécialiste des thésaurus, exprime dans cet article l'idée que le thésaurus véhicule un savoir et trouve parfaitement sa place dans le cadre du Web sémantique. Elle déplore toutefois l'engouement excessif de certains professionnels qui se sont emparés de ces nouveaux concepts que sont les ontologies. Cet article datant de 2004, et depuis les travaux sur les ontologies ont donné des résultats assez probants, il peut être conseillé aux lecteurs de ce document de prendre du recul par rapport au discours tenu.

[31] True and tested products: thesauri on the web. HUDON, Michèle. The indexer, Society of indexer, 2003, Vol. 23, n° 3, p115-119.

✍ Dans ce document, M.HUDON fait le point sur les thésaurus qui ont réussi leur adaptation sur Internet et montre en quoi ils sont efficaces en matière de recherche d'informations.

[32] EUROVOC, thésaurus multilingue : maintenance et aspects pratiques, exemple à bibliothèque du parlement européen. GAUTIER, Isabelle. Cahier de la documentation, janvier 2006, p 25-34.

✍ Cet article, présentant le thésaurus Eurovoc (réalisé pour la bibliothèque du parlement européen), a été rédigé suite à la conférence « Systèmes et langages documentaires » donné par I.GAUTIER en 2005 à Bruxelles.

[33] Thésaurus et systèmes de traitement automatique de la langue. DA SILVA, Lyne. Documentation et Bibliothèques, Ed. ASTED, avril-juin 2006, p 149-155.

✍ Ce texte expose le rôle que le thésaurus documentaire classique peut jouer dans le traitement automatique de la langue. En effet, c'est un outil intéressant car les relations sémantiques entre les termes peuvent être exploitées pour faciliter l'extraction de l'information tout en respectant le sens du discours.

[34] thésaurus, texte intégral et Web sémantique : la guerre des trois n'aura pas lieu. ROUMIEUX, Olivier. Archimag, Groupe Serda, 2002, n°154, p 48-49.

✍ O.ROUMIEUX contredit ici la thèse du déclin des thésaurus dans le cadre du web sémantique. Il défend l'idée de plusieurs modèles qui se complètent plus les uns aux autres qu'ils ne se concurrencent.

[35] Utilisation du thésaurus MeSH dans le site CisMef. KERDELHUE, Gaetan. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°1, p 29.

✍ L'auteur de cet article évalue la pertinence de l'usage du thésaurus MeSH sur le site Cismef. Ce document est intéressant dans la mesure où l'adaptation de ce thésaurus s'avère

une expérience couronnée de succès et contredit la thèse de la disparition des langages documentaires causée par les nouvelles technologies

INTERNET ET WEB SEMANTIQUE

[36] Les folksonomies comme support émergent de navigation sociale et de structuration sur le web. CREPEL, Maxime. Réseaux, 2008, n°152, p 169-204.

Etude très intéressante sur ce nouveau mode d'indexation collaborative que sont les folksonomies et leur usage par les internautes. L'auteur fait allusion à de nombreuses expériences sur l'indexation sur Flickr qui m'ont servi dans la deuxième partie du mémoire.

[37] Le Web sémantique : utopie ou réalité ? MESGUICH, Véronique. Netsources, 2007, n°71, p 1-5.

V.Mesguich se pose la question de l'existence effective du web sémantique. Elle évoque les différentes potentialités de cette nouvelle version du web et mentionne au passage l'intérêt des ontologies, leur utilisation et les expérimentations lancées par plusieurs structures.

[38] Introduction au Web sémantique. LAUBLET, Philippe ; CHARLET Jean ; REYNAUD Chantal. Information Interaction Intelligence, Hors série 2004, p 7-20.

Ce document se veut comme une synthèse des études faites sur le Web sémantique. Il s'agit d'une approche assez technique toutefois mais quand même assez compréhensible pour un public non initié.

[39] Application du Web sémantique. LEGER, Alain ; CHARLET Jean. Information Interaction Intelligence, Hors série 2004, p 169-185.

Les auteurs font ici mention des différents domaines d'application du Web sémantique : médical, E-Commerce, portails et Intranets d'entreprises.

[40] Le Web sémantique : de nouveaux enjeux documentaires ? Journée d'étude ADBS : IUT Paris V. Documentaliste-Sciences de l'information, 2003, Vol. 40, n°46, p 387-391.

Le groupe d'étude de l'ADBS s'est penché sur la question des éventuels apports du Web sémantique aux professions documentaires. Le principe et les applications du Web sémantique sont clairement présentés à travers cet article.

[41] Accès à l'information : du Web antique au Web sémantique. EUZENAT, Jérôme ; REMIZE, Michel ; OCHANINE, Hélène. Archimag, Groupe Serda, 2003, n°165, p 21 à 29.

Réflexion autour du concept du Web sémantique. J.EUZENAT fait cas des premières expérimentations dans le domaine et des différentes techniques qui facilitent l'accès à l'information.

[42] Web 3.0 et recherche sémantique : du vin vieux dans des outres neuves ? VALDERRAMA, Asuncion. Documentaliste-Sciences de l'information, 2009, Vol. 46, n°3, p 20-21.

☞ L'auteure se fait ici l'écho du propos de B.MENON, comme quoi la recherche sémantique n'a rien de totalement nouveau. Il a toujours été question de structurer les informations via les métadonnées pour les rendre accessibles à l'utilisateur.

[43] An overview of W3C Semantic Web activity. MILLER, Eric ; SWICK Ralph. Bulletin of assist, avril 2003. Vol. 29, p 8-11.

☞ Les auteurs définissent le principe de web sémantique comme une extension du web actuel. Ils en détaillent les différentes fonctions et les activités du Consortium W3C pour développer le web sémantique.

[44] The semantic Web : more than a vision. GREENBERG, Jane. Bulletin of assist, avril 2003, Vol. 29, p 6-7.

☞ Cet article introduit l'idée que le Web sémantique n'est pas un concept irréaliste d'une poignée de scientifiques visionnaires mais une extension authentique du web actuel. L'auteur admet que cette avancée n'a rien de révolutionnaire mais que ce sont les technologies développées actuellement qui rendent le Web sémantique possible.

INTEROPERABILITE

[45] SKOS : un langage de représentation de schémas de concept. LENART, Michèle. Documentaliste Sciences de l'information, Editions ADBS, 2007, Vol 44, n°1, p75.

☞ M.LENART décrit avec des détails techniques le projet SKOS qui est un système de représentation de concept sur le web.

[46] De la compatibilité à l'interopérabilité en matière de repérage d'informations importante : la problématique et l'exemple d'OTAREN. FEYLER, François. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°1, p 84-92.

☞ Il est ici question de l'avenir des thésaurus dans le cadre de la recherche sur Internet. Ces derniers sont concurrencés par de nouveaux modes d'indexation et l'auteur se base sur l'exemple d'OTAREN, dispositif expérimental pour tester l'interopérabilité entre les ressources.

[47] Multilinguisme et langages documentaires : le projet MACS en contexte européen. LANDRY, Patrice. Documentation et Bibliothèques, Ed. ASTED, avril-juin 2006, p 121-129.

☞ Le multilinguisme est ici abordé comme facteur d'interopérabilité entre les langages documentaires. L'internaute doit pouvoir accéder à l'information quelque soit sa langue naturelle. Le projet MACS, développé par plusieurs bibliothèques, dont la BNF, est un des pionniers en matière de multilinguisme.

[48] Référentiels, données d'autorité, thésaurus, ontologie, taxonomies ... pour en savoir plus ! AMAR, Muriel ; BOURDON Françoise. Documentaliste-Sciences de l'information, Editions ADBS, 2008, Vol. 45, n°3, p14-15.

✍ M. Amar et F. Bourdon font ici référence à l'interopérabilité entre les différents langages documentaires dans le cadre du Web sémantique. Elles font notamment mention d'une double convergence, d'une part entre les travaux de l'AFNOR et ceux du W3C mais aussi entre les professions.

ICONOGRAPHIE

[49] Thésaurus iconographique : système descriptif des représentations. GARNIER, François. Paris : Le léopard d'or, 1984, 239 p. ISBN 2-86377-032-2.

✍ Ouvrage de référence en matière de thésaurus iconographique. Réalisé à la demande du ministère de la culture et la direction des musées nationaux par F.GARNIER, ce thésaurus possède une hiérarchie assez forte et n'est pas forcément accessible à un public non averti. On peut également reprocher au thésaurus Garnier de ne pas être suffisamment généraliste pour être adopté par tous les musées. Les exemples choisis pour l'indexation représentent avant tout des œuvres à caractère médiéval.

[50] Gestion et diffusion d'un fonds d'image. KATTNIG, Cécile. Paris, Nathan Université, ADBS, 2005, 127 p. Collection 128. ISBN 2-09-191150-X.

✍ Cette monographie a pour objet le traitement documentaire de l'image. C.KATTNIG englobe toutes les dimensions de l'image : sa nature multiple, son traitement intellectuel et matériel, sans oublier le cortège juridique qui l'encadre.

[51] Thésaurus iconographiques et modèles culturels. DESPRES-LONNET, Marie. Document numérique, Ed. Hermès, 2000, Vol 4, n°1/2, p 153-165. ISBN 2-762-0224-7/ISSN 1279-5127.

✍ Ce document s'impose comme un état des lieux dans le domaine des thésaurus iconographiques. M.DESPRES-LONNET se base sur la recherche du terme « femme » pour tester l'efficacité de chacun des systèmes.

[52] Usages et pratiques de recherche des utilisateurs d'une banque d'images : l'exemple de l'agence de photographie de presse Gama. PICAULT, Coralie. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol. 44, n°6, p.74-381.

✍ Cet article est le résultat d'une enquête lancée par le service d'indexation de l'agence Gamma pour évaluer l'usage à distance des images par les utilisateurs.

[53] L'image n'est plus ce qu'elle était. MELOT, Michel. Documentaliste-Sciences de l'information, Editions ADBS, 2005, Vol. 42, p 361-365.

✍ M.MELOT s'interroge sur la nature même de l'image et de son rapport au texte. Par ricochet, il s'intéresse brièvement au sort de l'image comme représentation d'œuvre d'art.

[54] Image et audiovisuel : la documentation entre technique et interprétation, critiques et perspectives. BACHIMONT, Bruno. Documentaliste-Sciences de l'information, Editions ADBS, 2005, Vol. 42, p348-353.

✍ *B.BACHIMONT, spécialiste en ingénierie documentaire audiovisuelle fait référence ici aux méthodes d'indexation des images fixes et animées. Il fait également allusion à l'usage des ontologies comme référentiel pour représenter les thésaurus, outil de structuration des contenus.*

[55] Les éléments formels de la fiche descriptive des banques d'images et leur nécessaire normalisation. DOUCET, Anne-Vinciane. Documentation et Bibliothèques, Editions ASTED, juillet/septembre 2008, p 209-219.

✍ *L'auteure de ce document présente en tout premier lieu un cadre conceptuel de l'image et ensuite préconise l'utilisation des métadonnées pour en assurer la normalisation.*

[56] Traitement documentaire de l'image ordinaire : analyse de deux approches d'indexation. MENARD, Elaine. Documentation et Bibliothèques, Editions ASTED, janvier/mars 2009, p 29-39.

✍ *Cet article est le fruit d'une recherche sur la comparaison entre indexation libre et indexation contrôlée à l'aide d'un langage documentaire tel que le thésaurus.*

[57] L'image dans les bibliothèques, trente ans après. MELOT, Michel. Bulletin des bibliothèques de France, 2007, Vol. 52, n°2, p 67-69

✍ *L'image n'a pas toujours eu bonne presse dans les milieux intellectuels et son entrée dans le monde des bibliothèques ne s'est pas faite sans difficulté. L'indexation de l'image a suscité une certaine défiance de la part des bibliothécaires et a bouleversé les pratiques professionnelles.*

MUSEOLOGIE

[58] Le patrimoine à l'ère du document numérique : du musée virtuel au musée médiathèque. WELGER-BARBOSA, Corinne. Ed. L'Harmattan, 2001, 313 p, ISBN 2-7475-1725-X.

✍ *C.WELGER-BARBOSA dépeint la situation des musées français et de leur présence (fictive ?) sur Internet. Cet ouvrage, paru il y a dix ans, affiche un postulat considéré encore d'actualité par son auteure (voir Interview en annexe).*

[59] Musées 2.0. CRENN, Gaëlle ; VIDAL, Geneviève. Cultures et Recherches, 2008/2009, n°118/119, p 39.

✍ *Article rédigé à quatre mains sur la construction d'une relation interactive entre les musées et le(s) public(s).*

[60] Il est absurde d'avoir un site web 2.0 et un musée 1.0. NUTTIN, Guillaume. Archimag, Groupe Serda, 2010, n°234, p 42-43.

✍ *G.NUTTIN se propose de recueillir les propos de S.BAUSSON (webmestre et animateur de communauté) au sujet de l'expérience du musée de Toulouse avec les nouvelles technologies, notamment les réseaux sociaux.*

[61] Pour en finir avec le Web 2.0 : vers le Web 3.0 dans les musées. VIDAL, Geneviève. Documentation et Bibliothèques, Editions ASTED, octobre/décembre 2009, Vol. 55, n°4, p 201-207.

La responsabilité culturelle des musées les engage à diffuser de manière rationnelle un capital informationnel de qualité. Alors qu'ils viennent juste de maîtriser les tenants et les aboutissants du Web 2.0, les musées sont invités à envisager très rapidement leur ascension vers un Web 3.0 pour offrir aux internautes un contenu structuré et accessible.

[62] L'utilisation des langages documentaires dans les musées. LAFAILLE, Madeleine. Documentation et Bibliothèques, Editions ASTED, 2006, Vol. 52, n°2, p 131-137.

Rétrospective sur les langages documentaires dans les musées d'une manière générale et étude comparative des outils les plus connus, tel qu'Iconclass ou l'AAT.

[63] La photographie documentaire, l'épreuve du dess(e)in informatisé : mise en contexte sur les sites web de musée. TARDY, Cécile. Documentation et Bibliothèques, Janvier/Mars 2010, p 15-23.

Voici une réflexion très riche sur le statut de la photographie documentaire dans les musées. Qu'il s'agisse de son intégration à des bases de données ou la mise en ligne d'images sur le site web des musées, la notion de photographie documentaire est pluridimensionnelle.

[64] Quelques enjeux culturels du Web 2.0. CHANTEPIE, Philippe ; PIERRE, Jocelyn. Culture et Recherche, 2007, n°112, p35-36.

Article sur l'impact des nouvelles technologies dans le milieu culturel et qui se veut le compte-rendu du séminaire organisé en 2006 et 2007 par le DEPS et l'IRI.

[65] Le portail documentaire du musée quai Branly : la diffusion virtuelle des collections. DALEX, Françoise. Culture et Recherche, 2007, n°112, p 29.

F.DALEX, responsable du portail documentaire du musée du quai Branly, expose les diverses fonctionnalités du site Internet : l'entière collection d'œuvre est mise en ligne, 150 000 images, une mise à jour quotidienne, plusieurs modes de consultation ...

[66] Internet et les musées : de nouvelles offres pour de nouveaux usages. AILLAGON Thomas. Culture et Recherche, 2008/2009, n°118/119, p 35.

T.AILLAGON étudie à travers plusieurs exemples de musées (français et étrangers) les usages faits par ces derniers du web et des médias sociaux.

[67] Airs de famille : DOCAM, entre la muséologie et les sciences de l'information. TURNER James. Documentation et Bibliothèques, octobre/décembre 2009, p 153-158.

Document très pertinent sur le rapprochement entre la muséologie et les sciences de l'information à travers du projet de recherche DOCAM, destiné à faire cohabiter l'enseignement des deux disciplines dans un même cursus.

[68] Muséologie et sciences de l'information : regards croisés, destins croisés. MENARD Elaine. Documentation et Bibliothèques, octobre-décembre 2009, p151-152.

E.MENARD fait état de la rencontre entre deux champs d'étude : les musées et les sciences de l'information. La convergence a donné lieu à une tierce discipline : la muséologie.

[69] Musées et Médias sociaux. CAREY Sheila. Documentation et Bibliothèques, octobre-décembre 2009, p 177-187.

Cet article de fond explique comment les musées se sont appropriés les NTIC, et notamment les réseaux sociaux, tels que Facebook ou Flickr, pour rencontrer et attirer leur public. L'auteur cite un par un ces sites participatifs pour en démontrer les différents usages par les internautes dans le cadre culturel.

[70] Les musées au risque du Web 2.0. [En ligne]. CHOSSEGROS, Aurélia. Observatoire critique, 12 juillet 2008, [consulté le 5 octobre 2010].

Accès : http://www.observatoire-critique.org/dossier_web2.php3

Ce texte est le fruit d'une recherche assez poussée sur la présence des musées sur les réseaux sociaux, l'exacte nature et la qualité des liens tissés entre eux et les publics. La rédactrice a truffé sa réflexion de multiples exemples concrets qui m'ont inspiré pour la rédaction de mon mémoire (2^{ème} partie). Ce site est une mine d'or pour tous ceux qui sont intéressés, entre autre, par l'usage des technologies dans le monde muséal.

[71] Thématique des œuvres plastiques contemporaines et indexation documentaire. REGIMBEAU, Gérard, 1996, Lille, thèse, Edition ANRT. [Accessible à la bibliothèque de l'INHA sur microfiche].

Cette thèse rédigée par G.REBIMBEAU a le mérite de s'intéresser à l'indexation des œuvres d'art (contemporaines) et propose une approche plutôt philosophique que technique. Elle est disponible à l'INHA sous forme de microfiches. J'aurai aimé pouvoir la lire dans son ensemble mais je dois avouer que la qualité du matériel, le coût des reproductions ainsi que le temps, ont constitué des freins non négligeables pour me permettre de livrer davantage la pensée de l'auteur à travers mon mémoire.

[72] L'image d'art entre analyse critique et analyse documentaire. REGIMBEAU Gérard. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol 44, n°2, p 130-137.

Dans cet article, l'auteur reprend le sujet de sa thèse et propose un cadre d'analyse de la représentation d'œuvre d'art. Il illustre son propos par une analyse en trois dimensions d'une toile de Kandinsky.

NORMES

[73] Les normes de conception, gestion et maintenance de thésaurus. Evolution récente et perspective. CHICHEREAU, Dominique ; CONTAT, Odile ; DEGEZ, Danièle, DENIAU, Alina ; LENART Michèle, MASSE Claudine, MENILLET, Dominique. Groupe Langage documentaire de l'ADBS. Documentaliste-Sciences de l'information, Editions ADBS ? 2007, Vol. 44, n°1, p 66-74.

Comme son titre l'indique, les auteurs font le point sur les normes en matière de thésaurus. En se basant sur plusieurs aspects, tels que le choix des termes, la précoordination ou la présentation des thésaurus, ils comparent les normes AFNOR, ISO, NISO (USA) et BS (GB).

[74] NF Z47-100, Documentation, Règles d'établissement des thésaurus monolingues. AFNOR, 1981. 20 p. ISBN 2-00-258358-6.

☞ *Norme française qui a pour objet d'édifier les principes des thésaurus monolingues. Cette norme a été revue dans le cadre du projet ISO 25964.*

[75] Le projet de norme : ISO DIS 25964-1 [en ligne]. DALBIN, Sylvie. Support de réunion AFNOR TC46/CN357/GE10, 7 novembre 2008. [Consulté le 5 octobre 2010].

Accès : <http://www.slideshare.net/Dalb/np-25964-future-norme-sur-les-thsaurus-presentation-737506>

☞ *Cette présentation Power Point visible en ligne présente les divers points abordés lors de la réunion AFNOR pour la création de la norme NP25964.*

ONTOLOGIES

[76] Les ontologies : antécédents, aspects techniques, limites. CHAUMIER Jacques. Documentaliste-Sciences de l'information, Editions ADBS, 2007, Vol.44, n°1, p 81-83.

☞ *A travers cet article, J.MANIEZ retrace les origines des ontologies. Il définit cette notion et son application dans le contexte du Web sémantique. Il évoque également les différents protocoles informatiques qui permettent le développement des ontologies. Ce panorama très bien dressé sert de base pour cerner le concept d'ontologie.*

[77] Un glossaire peut cacher une ontologie : l'expérience de Lafarge. GICQUEL, Florence. Documentaliste-Sciences de l'information, 2007, Vol.44, n°1, p 81-83.

☞ *Cet article repose sur un retour d'expérience. La firme Lafarge, spécialisée dans les matériaux pour les travaux publics, a commandité la réalisation d'un outil documentaire pour rendre compte d'un vocabulaire professionnel assez pointu. Suite à l'appel d'offres, la société Mondeca, dont l'activité gravite autour du web sémantique et des ontologies, a été sélectionnée pour monter le projet.*

[78] Ontologies on the Semantic Web. JACOB, Elin.K. Bulletin of assist, avril/mai 2003, Vol. 29, n°4.

☞ *Il s'agit d'un texte assez précis sur la nature des ontologies, leur usage et quelques exemples précis d'application. Les langages informatiques pour formaliser les ontologies, tels que RDF, sont également mentionnés. L'auteur conclut l'article en disant que ces nouveaux outils d'organisation du savoir constituent un vocabulaire contrôlé, à condition que la définition de ce terme soit reconsidérée.*

ANNEXES

ANNEXE 1 - AGENDA

15/28 février : 1^{ère} cession de stage

24 mars : RDV Musée Carnavalet avec Valéry JOBBE-DUVAL et Christelle COURTOIS-thésaurus Adlib

11 juin : début de la 2^{ème} cession de stage

16 juin : RDV avec Valérie FOURS, Isabelle SIBIRIL et Isabelle JOLFRE au BDM pour formation au thésaurus Adlib

25 juin : Réunion avec Valérie FOURS, Claire BOISSEROLE, Claire MARTIN sur le thésaurus- 1^{ère} présentation de mon travail

29 juin : RDV RMN avec Diana SEGURINI – consultation du thésaurus commun sur la base Cortex

9 juillet : RDV Musée du Louvre avec Michèle PERNY- chargé d'étude documentaire – consultation et photocopie des index iconographiques

13 juillet : Réunion avec Valérie FOURS, Claire BOISSEROLE et Isabelle SIBIRIL au BDM pour déterminer les grandes thématiques du thésaurus et le nombre de hiérarchies.

20 juillet (15h) : RDV avec Annie CARON au Musée des arts décoratifs

21 juillet (14h30) : RDV avec Sophie LAGABRIELLE au Musée Cluny

5 août (10h) : RDV avec isabelle GAETAN au musée d'Orsay

10 août (14h30) : RDV au Bureau Des Musées pour rentrer thésaurus sur BDD Adlib

12 août (10h) : RDV avec Cécile DUPRE (COARC) pour travailler sur la thématique « RELIGION »

20 août : Fin du stage

27 août (14h) : Visite aux Archives départementales du Morbihan

ANNEXE 2- TRAVAUX EXPERIMENTAUX DE N.JAMES SUR LES ONTOLOGIES

Nicolas James est doctorant en informatique au MAS (Mathematics and Systems laboratory) de l'école centrale Paris situé à Châtenay-Malabry.

Son principal intérêt de recherche est centré autour de la vision par ordinateur, de l'apprentissage automatique, du text-mining et d'ontologies pour une compréhension de haut niveau de documents multimédias. Plus précisément, il travaille sur l'intégration d'une connaissance a priori du contenu basé sur des systèmes d'annotation sur l'image pour l'annotation sémantique des images, le sens de marquage de l'apprentissage machine ensembles de données pour améliorer la récupération de l'image basée sur le contenu, la modélisation de l'annotation sur l'image en utilisant des ontologies dans un cadre d'indexation.

Voici comment il définit son sujet d'étude :

« Ma thématique principale n'est pas l'indexation des images, mais l'annotation automatique pour l'indexation (i.e. produire automatiquement des descriptions d'images, descriptions textuelles, qui sont par la suite indexées, mais ce que je veux dire c'est que la problématique est la production automatique de la description des images). Et c'est là qu'on utilise des ontologies, pour introduire de la connaissance dans le processus de calcul des descriptions. Par exemple, en utilisant l'ontologie wordnet je sais qu'il existe une relation de type tout à partie entre car et wheel, (une roue est une partie d'une voiture), en détectant car sur une image, si je peux détecter aussi wheel, j'augmente la probabilité que ce soit réellement pertinent pour décrire l'image (nous sommes dans un domaine probabiliste). Ça c'est une partie de mon travail.

La seconde grande partie des mes travaux est du matching d'ontologie, le but est de décrire un concept d'une ontologie en utilisant les concepts d'une autre ontologie. En gros, disons que comme il existe des centaines d'ontologies destinées à des usages différentes, mise au point avec des techniques différentes, nous faisons face au problèmes de la gestion de plusieurs ontologies dans les systèmes d'indexation, mais ces ontologies possèdent de la connaissance commune. (Par exemple, nous utilisons une ontologie nommée LSCOM qui contient un concept dont l'étiquette est « car », un concept avec une étiquette identique est présent dans l'ontologie WordNet, mais est-ce le même concept?). »

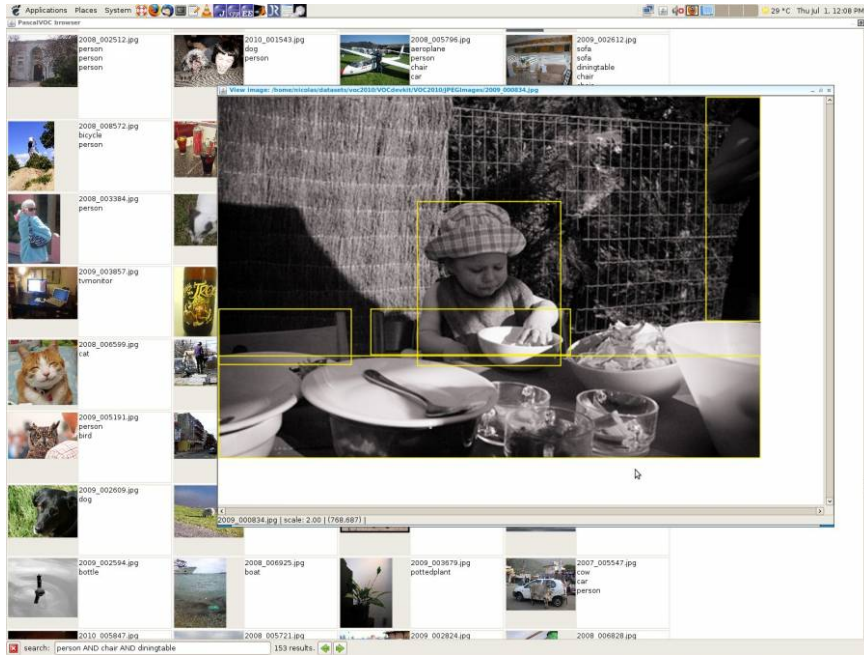


Image 1 :

Galatée est une bibliothèque Java pour l'exploration et la recherche dans la collection de l'image où les images annotées. C'est un logiciel libre.

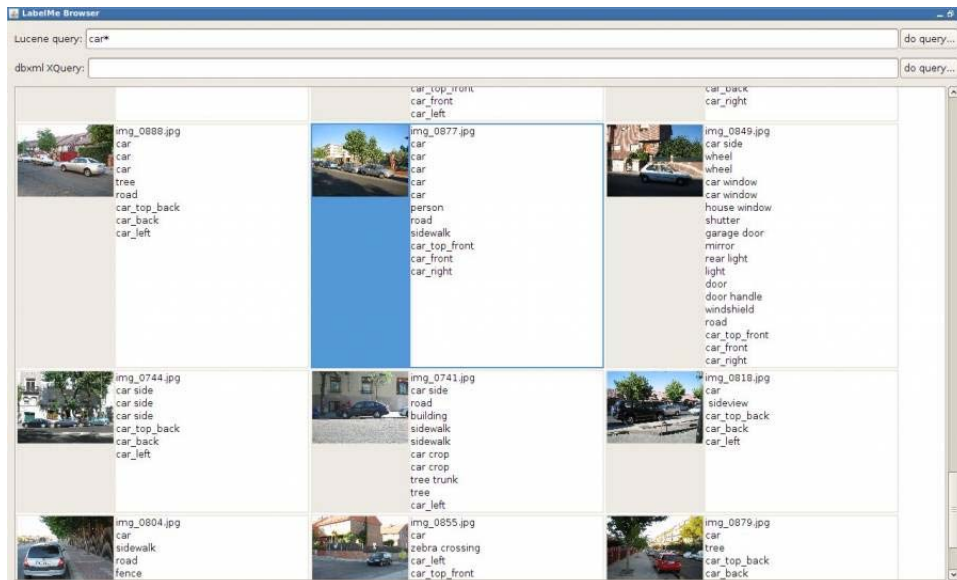
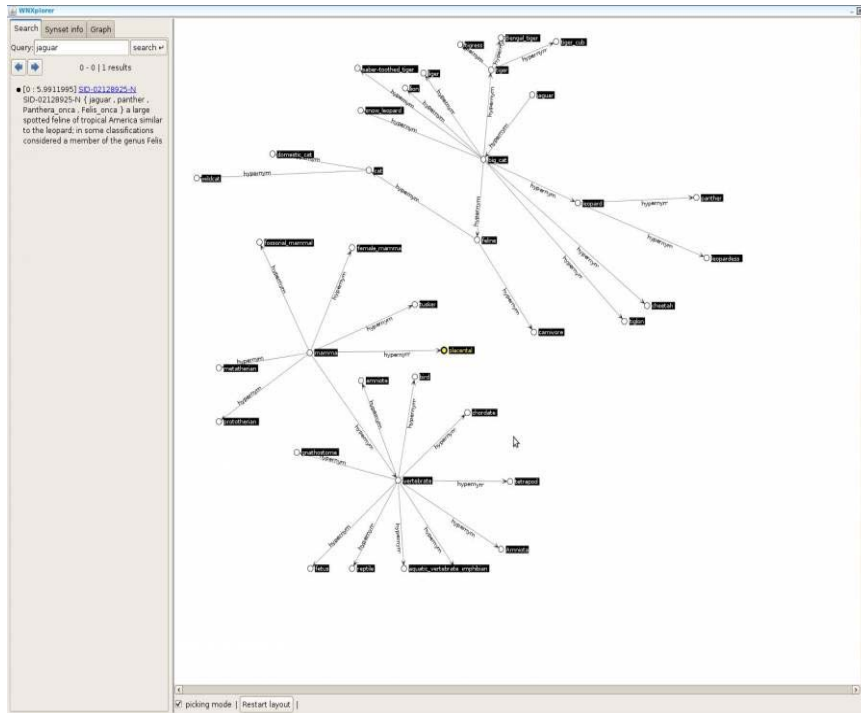


Image 2 :

Extraction de métadonnées

Image 3 :

Matching d'ontologies sur le principe Wordnet



Ces informations sont issues d'un échange de mail avec N.JAMES mais il est possible d'obtenir plus d'informations en allant sur son site et sur son blog.

ANNEXE 3- INTERVIEW DE CORINNE WELGER- BARBOSA

(Maître de conférences en Sciences de l'information et de la Communication ainsi qu'en Histoire de l'Art à l'institut Michelet-Université Paris I)

Auteur de la monographie : « Du musée virtuel au musée médiathèque : le patrimoine à l'ère du numérique : »

9/09/2010 à l'INHA

1- **Dans votre ouvrage « Le patrimoine à l'ère du numérique : du musée virtuel au musée médiathèque », vous écrivez que les conservateurs de musée craignaient de ne plus tenir le monopole du discours sur les œuvres d'art (p 37). A votre avis, cela a-t-il changé depuis 2001 ?**

Réponse : C'est une très bonne question car elle se posait en des termes différents et en même temps il y avait la perception d'un enjeu fondamental. En effet, à l'époque (fin des années 90 et début des années 2000), le fait de mettre en ligne de la documentation faisait craindre en tous cas qu'il y ait cette remise en question d'un monopole par le simple fait de la diffusion dans un environnement non contrôlé. C'était l'Internet de cette époque là de toute façon. Aujourd'hui, cette même question a pris un tour différent, plus concret pourrais-je dire, plus inattendu et inenvisagé à l'époque, et dont à mon avis l'expression la plus critique (critique dans le sens du mot « crise) du point de vue des mœurs de la profession est bien la possibilité de nommer, d'identifier et de documenter d'une certaine façon les œuvres, tel que cela a été apparu et mis en œuvre sur le Web. Alors bien évidemment, je pourrai faire quelques détours plus tard sur la Folksonomie, on pourra y revenir, il y a plusieurs cas de figure qu'on peut développer sur ce sujet qui ne peut pas être pris comme un bloc homogène.

En fait, on ne peut pas vraiment parler d'une façon générale de tous les conservateurs. Quand j'ai fait ce livre, je me suis beaucoup appuyée sur la situation française. Aujourd'hui, si on veut répondre à votre question, on est obligé de constater qu'il y a des attitudes différentes qui ne sont pas totalement homogènes par pays ou régions culturelles mais quand même on peut noter qu'en France nous ne sommes pas particulièrement à la pointe pour ce qui est des changements de la culture professionnelle des conservateurs, et que le web est encore envisagé par la plupart des musées comme un environnement qui fait encore question. C'est-à-dire que de nombreux musées, des musées d'art en régions, se posent encore la question de l'opportunité, la pertinence de mettre des informations en ligne. Les musées qui le font toujours un certain problème à franchir la barrière du web. C'est-à-dire que le web est bien évidemment davantage accepté comme un environnement de diffusion mais par nécessité. Il y a plus d'éléments qui sont mis en ligne, si on prend par exemple Orsay, le Louvre ... mais on n'est encore susceptible d'être porté par cette problématique de diffusion de savoir au(x) public(s). Dans d'autres contextes, d'autres pays, il y a une approche de l'environnement du web qui est davantage marqué maintenant par d'autres types d'échanges avec ce qu'on considérait avant comme des publics à qui on s'adresse, là on est davantage dans une forme d'échange, de dialogue, d'ouverture ... On est

dans une autre problématique mais en France nous ne sommes pas encore dans ce cas de figure. Notez qu'il y a un blogue sur le site du Louvre tenu par la Vénus de Milo et qui par le fait même que ce soit un blog est bien sûr intéressant et puis aussi parce que justement il y a une parole qui circule, qui va au-delà de ce qui est tenu d'être délivré au public. Il y a plus de détails sur des opérations internes du musée. Il y a donc des secteurs au sein même du musée qui sont plus perméables aux tendances à l'œuvre aujourd'hui, qui sont moins sur la défensive mais il ne faut pas oublier que dans les très grandes institutions, cela peut se réduire à quelques personnes ou à des services qui sont certes en relation avec la conservation mais qui ont des objectifs propres, leur propre développement.

En conclusion, si on prend le corps en tant que tel des conservateurs en France, on ne peut pas dire qu'on soit très éloigné de la situation que j'ai pu décrire dans mon livre publié en 2001 mais qui était le fruit d'études antérieures.

**2 - Vous évoquez également la surtarification des images par les musées (p 38).
Ce procédé a-t-il toujours cours aujourd'hui ?**

Réponse : Malheureusement oui ! On n'a pas beaucoup changé d'état d'esprit. On fait vaguement passer sur le registre d'un droit d'auteur ce qui n'est pas vraiment de l'ordre du droit d'auteur mais qui était depuis 1920 si je me rappelle bien une redevance qui avait été créée pour les services rendus pour l'image. On est un peu resté sur ce schéma là avec une attitude très défensive par rapport au fait que le numérique pénètre dans le musée par l'appareil photo numérique tout simplement et que les liaisons du numérique, que ce soit la photo, le web ... circulent sans rupture. Tout est dans la même chaîne de circulation numérique. Il n'est pas rare qu'en France il y ait encore des interdictions de photographier des œuvres du domaine public alors que l'on sait qu'il n'y a aucune raison pour cela. Légalement on a le droit de le faire. Bien sûr on peut demander au public de ne pas utiliser les flashs mais de toute façon malgré les meilleures mesures hygrométriques etc., cela porte atteinte aux œuvres. Il y a donc un certain nombre de rationalisations qui ne sont pas forcément rationnelles et qui sont plutôt mises au service de la préservation d'une possession, d'une privatisation des œuvres. Les musées sont non seulement garants de l'entretien et de la mise à disposition au public du patrimoine mais en même temps d'une certaine façon ils se le sont accaparé. Avec le numérique, on ne peut pas aller à l'encontre d'une circulation des images des œuvres qui soit forcément libre. Quoi qu'il en soit, on ne peut rarement atteindre une qualité de l'image numérique telle qu'elle puisse donner lieu, comme on peut le redouter, à des ouvrages imprimés sans que les droits ne soient payés (qui ne sont pas des droits d'ailleurs ...). C'est irrationnel mais il y a toute une « rationalisation » qui va à l'encontre de la diffusion possible de toutes les images. Là encore il y a des différences importantes au sein des pays européens, je reviens il y a quelques mois de ça d'un séjour à Berlin, je suis beaucoup allée dans les musées et j'étais très surprise de voir que dans un musée, c'est le cas dans tous les musées, mais un musée d'art contemporain où là le problème de droit d'auteur les œuvres se pose (les auteurs sont ou vivants ou morts depuis peu), on a même une incitation à photographier les œuvres. Toutes les rationalisations qui sont mises en œuvre ici : pour dire qu'on n'a pas le droit, que ça va abîmer les œuvres ... ce n'est pas vrai partout, ce n'est pas vrai de façon universelle ! En Angleterre, le musée Victoria and Albert fut parmi les premiers à considérer, il y a trois ou quatre ans, que cela ne servait à rien de restreindre l'accès aux images sur le web, que c'est une bataille perdue, ça n'a pas grand sens. Plusieurs musées ont suivi, en France on n'a pas suivi. Je pense que l'on ne peut rien faire contre ça, à terme, c'est la logique même

du numérique. Il faut savoir qu'aux Etats-Unis par exemple, le MET⁹⁰ qui est un musée encyclopédique, encourage les visiteurs à photographier les œuvres comme ils le souhaitent, à se photographier avec aussi et à mettre toutes les images en ligne sur un compte Flickr créé par le MET. Comme vous le voyez, on est dans des attitudes très diverses et moi ma conviction c'est que ce sont les plus attitudes les plus ouvertes et les plus en phase avec la culture numérique qui s'invente sur le web qui vont l'emporter. C'est ma conviction !

3 - Le fait que les institutions se protègent sous la licence Creative Commons a-t-il réellement changé la donne ?

Réponse : Oui, c'est intéressant parce que déjà c'est la reconnaissance d'une autre façon de voir les choses, c'est-à-dire qu'on n'est plus exactement dans la logique du copyright, le copyright s'inscrit davantage dans une logique de protection mais qui est souvent interprétée comme une logique de l'interdit. Les Creative Commons introduisent d'abord la logique de bien commun et énoncent différentes façons, ou choix pour ce qui s'en servent, d'ouvrir justement la reprise de ce qui est mis en ligne par les autres avec une sorte de contrat de confiance qui est établi et qui change complètement la problématique. Nous ne sommes pas dans une problématique, excusez-moi le terme, paranoïaque, et privative, mais on est du coup dans une problématique contractuelle et d'intersection où la question de la culture de bien commun qui a été posée dans des termes nouveaux, avec la neutralité du web, les logiciels libre etc. on est plus en phase, cohérent avec un discours public qui veut que le patrimoine appartienne à tout le monde.

4 - Le clivage entre les musées anglo-saxons plutôt ouverts sur la « marchandisation » de la culture et le retranchement des musées européens s'est-il estompé ? Je vous pose quand même cette question bien que vous y avez en partie répondu précédemment.

Réponse : Je ne poserai pas la question dans ces termes là. La question de la marchandisation, elle se pose pour tout le monde. Tout dépend de savoir à quoi on l'applique. Si les musées peuvent avoir des rentrées supplémentaires avec l'offre de services ou produits dérivés ... tout le monde le fait, ça. La question des contenus, des documents, des informations diffusés sur l'internet, à mon avis, ne s'incluent pas là-dedans, bien que, au départ, parfois de temps en temps on voit encore des discours dans ce sens là, s'exprime la conviction qu'à terme on va pouvoir faire payer tout ça.

5- Vous évoquez un rapprochement entre les Sciences de l'information et l'histoire de l'art, j'ai pu observer cette évolution à travers mes lectures. Quelles sont à l'heure actuelle les dispositions prises dans l'enseignement de l'histoire de l'art pour rendre effective cette union des deux disciplines ?

Réponse : Là, je vais avoir du mal à vous répondre, et la réponse est malheureusement très simple, à ma connaissance, en France, en termes de formation ou d'enseignement, je suis la seule à avoir proposé des cursus sur ces questions. Je suis gênée pour le dire, c'est toujours gênant de dire qu'on est la seule à ...

⁹⁰ Metropolitan Museum à New York

Alors, il y a sans doute des formations en infocom qui proposent des cours sur la muséologie.

Ailleurs, évidemment, ça existe. Mais ce ne sont pas forcément des rapprochements ... en fait, il y a deux choses dans votre question : ici on voit que les Sciences de l'information et de la communication sont une discipline en tant que tel puisqu'elle est reconnue au CNU, c'est toujours ça notre repère en France, elle a une section du Conseil National des Universités qui en fait une discipline qui s'enseigne à l'université. Pour l'histoire de l'art, c'est un petit peu plus compliqué parce que l'histoire de l'art appartient à une section plus large qui est l'histoire de l'art moderne et contemporaine puis une autre section qui est l'histoire des mondes anciens. Donc, il se trouve que j'ai un cursus double, j'ai été qualifiée dans les deux disciplines. Mais ce qui est plus intéressant me semble t'il pour cette situation française un peu particulière, c'est qu'il existe des formations en infocom où les gens travaillent sur la question des musées. La muséologie a été plutôt rapatriée du côté Sciences de l'information et de la communication et pas côté histoire de l'art. C'est comme ça !

Dans d'autres pays, en revanche, on voit très bien, au Canada, vous faites très bien d'en parler, que des départements d'histoire de l'art, en Angleterre aussi, ont ouvert la formation à l'impact du numérique et des nouvelles technologies sur l'histoire de l'art elle-même. C'est tout l'intérêt, c'est ça qui me semble être l'urgence aujourd'hui. Il y a plusieurs lieux en Angleterre, le premier étant le Birkbeck College à Londres, il y en a d'autres, à l'UCAM de Montréal aussi, il y a déjà un certain temps. Il y a presque une tradition, pourrait-on dire, qui n'est pas seulement liée à la muséologie mais à la prise de conscience de l'impact du numérique ou des sciences de l'information et de la communication dans la discipline elle-même. C'est aussi le cas aux Etats-Unis, en Australie. Nous sommes comme toujours un petit peu en retard ...

6- Au sujet des publics, vous dites : « il nous importe de savoir comment les nouvelles productions mettent au jour la façon de penser des publics et la vision du patrimoine qui se profile à partir de ces offres ... ». Selon vous, la vision du public grâce aux NTIC a-t-elle réellement évolué et si oui de quelle manière ?

Réponse : Ca, on en déjà un peu parlé auparavant puisque je vous disais que la façon de s'adresser au public peut suivre un fil, une logique traditionnelle, didactique en gros, d'éducation des publics.

Aujourd'hui, les gens qui réfléchissent, à ce qui convient d'appeler la médiation culturelle, à des modes de participation du public. C'est un thème central désormais, du fait du web 2.0. on a parlé de la Folksonomie mais il y a aussi d'autres dimensions d'échange, comme Facebook ... il ya quand même toute cette dimension et puis il y a aussi du coup une réflexion de la part des musées les plus avancés en ces domaines sur le fait d'aller jusqu'au bout, c'est-à-dire d'introduire au sein même des musées physiques le fait que la visite du public ne soit plus conçue comme elle l'a toujours été. Là encore, on réfléchit à des dispositifs plus participatifs de la part des publics. C'est une question for discutée. Je ne sais pas si vous connaissez le blog qui s'appelle « Museum 2.0 », le cœur des discussions porte là-dessus.

7 - Pour reprendre vos propres termes, vous affirmez que « la notion de musée virtuel se répand à la façon d'un concept mou » (p 119). Considérez-vous qu'il y avait alors un abus de langage lorsque l'on parlait de musée virtuel en 2001 ?

Est-ce encore le cas ou au contraire les musées ont su s'approprier tous les outils technologiques pour valoriser de manière optimale leurs collections ?

Réponse : Je n'ai jamais accordé beaucoup d'intérêt à cette polémique, cette notion de musée virtuel me paraît déphasée à plus d'un titre, c'est-à-dire que, je pense l'avoir déjà exprimé dans cet ouvrage, c'est presque une façon de se faire peur, de la part de l'institution. Parler de musée virtuel comme si le public allait désertier le musée pour fréquenter virtuellement ... il me semble que c'est là une conception fautive et pas très intéressante. Les développements justement du web montrent très clairement que d'une part ce qui est intéressant et qui n'est pas virtuel ou alors si on prend le terme virtuel dans sa véritable acception, c'est-à-dire que virtuel c'est avant tout un fichier qui peut être actualisé, ce qui est virtuel n'existe véritablement que quand il est affiché sur votre écran. C'est de l'ordre du substitutif au musée. Cela me semble un peu hypocrite aussi, c'est pour cela que je me suis un peu battue contre ce terme parce qu'agiter cette menace empêchait justement de penser le développement documentaire, c'est-à-dire la valorisation des sources ou des savoirs construits partir de sources détenues par le musée qui peuvent prendre effectivement une autre dimension dans l'environnement du web et être livrée au public, aux chercheurs, aux scientifiques aussi. Quand le musée parle de public, il ne pense jamais aux scientifiques, or la documentation en ligne elle concerne essentiellement les scientifiques ou les autres conservateurs et ça ce n'est pas au centre de la préoccupation des musées. Je continue de penser que c'est dommageable.

8 - Que sait-on aujourd'hui du rapport de cause à effet entre visite virtuelle et visite réelle ?

Réponse : Pour l'instant, ce que j'en sais, il y a quelques études mais il n'y en pas énormément. Il y a des publications suite aux colloques, aux conférences qui ont lieu chaque année, aux Etats-Unis ou au Canada. Dans ces articles en ligne, il y en a certains qui réfèrent à des études de publics. Les études sur le public en ligne sont très difficiles à faire. Ces indications sont à manipuler avec des pincettes. Il n'y a pas de désertion des musées. Pas du tout ! Est-ce que ça va jusqu'à attirer de nouveaux publics ? Je voudrai voir l'effet des réseaux sociaux, sur Facebook ... de la présence des musées mais on constate plutôt que les gens qui ont l'habitude de fréquenter les musées vont aussi sur internet consulter de la documentation en ligne et échanger. C'est logique, on ne voit pas tellement pourquoi les gens qui ne vont pas au musée iraient sur les sites des musées. Il y a des considérations très surprenantes là-dessus. On voit qu'en France, par exemple, les visites augmentent depuis qu'on a instauré qu'un dimanche par mois la gratuité des musées. Là, effectivement, on peut mesurer facilement l'accroissement du public. Est-ce que les gens qui vont sur internet vont davantage au musée ? Personne n'arrive à répondre à cette question. Ceux qui fréquentaient les musées avant Internet continuent de les fréquenter. Les études en France montrent qu'il y a une faible progression, il y a même une certaine stagnation dans la fréquentation globale des musées. C'est assez compliqué de répondre à votre question. En revanche, ce que j'en vois sur le web, il y a une certaine façon de faire vivre le musée beaucoup plus ouverte mais dans des cultures différentes. Ce qu'on voit aux Etats-Unis, par exemple, c'est que la multiplication des événements autour des musées a peut-être un peu d'impact localement.

9 - Y a-t-il un moyen de mesurer de manière qualitative et quantitative l'usage des sites de musées ? Par le biais de sondages, d'analyser les pages visitées ou les photos copiées ... ?

Réponse : On peut analyser les clics, mesurer le temps passé sur le site. L'origine des consultations ne peut se faire qu'à travers des formulaires d'enquête dont on sait qu'il y a un faible taux de réponse. Cela donne des valeurs indicatives mais pas plus. La fréquentation des jeunes, on sait qu'elle est particulièrement drainée par les enseignants et que dans certains milieux le cadre familial en est à l'origine.

Sur le site du ministère de la culture, il y a des chiffres clés par secteur. Chaque année, il y a des statistiques qui sortent sur la billetterie, les entrées non payantes pour un certain nombre de musées, les plus importants, le Petit Palais en fait partie je crois. Sur les sites des grands musées, vous avez souvent les rapports d'activité annuels qui contiennent des informations sur la fréquentation. S'ils ont lancé une étude sur les publics, ça y figurera.

10 – Il y a assez peu de littérature sur l'indexation des œuvres d'art (G.REGIMBEAU- M.LAFAILLE). Cela vous paraît-il exact ? Comment l'expliquez-vous ?

Réponse : C'est quelque chose qui concerne une activité professionnelle propre aux documentalistes au sein des institutions et qu'on commence à s'y intéresser autrement parce que le phénomène du Net étant finalement la logique d'indexation. Cela n'a jamais été le cas auparavant. C'était vraiment un domaine fermé de spécialistes et la demande d'information était une demande de transmission de savoir-faire, de références, de vocabulaires contrôlés ... de savoir les manier dans des petits milieux où il n'était question que de transmission professionnelle. Aujourd'hui, vous comprenez que l'indexation est un problème qui concerne tout le monde ! On ne peut pas se servir un tant soit peu du web si on ne comprend pas la chose la plus élémentaire qu'est l'indexation, grâce notamment aux moteurs de recherche qui fonctionnent sur ce principe. Il y a d'autres ingrédients bien sûr pour faire un moteur de recherche. Mais on est dans un monde, sur le web où la question de l'indexation est une question clé pour comprendre le numérique. Alors oui, il n'y a pas grand-chose, surtout une littérature grise sur le sujet, notamment l'élaboration de thésaurus, parce qu'il suffisait d'apprendre, de transmettre un savoir-faire finalement, tandis que maintenant on réfléchit à un autre degré sur ces outils intellectuels. Cela devient des outils intellectuels qu'on interroge. On se demande si on peut les utiliser au-delà du monde professionnel, c'est ce qu'on est en train de faire, enfin c'est mon interprétation. C'est pour ça que, quand j'ai eu l'occasion de parler de cela dans un séminaire ou un colloque, en février dernier, j'ai été très étonnée qu'il y ait beaucoup de monde qui vienne des musées. Je me suis dit voilà des gens qui font avec cela mais ils ont besoin d'y réfléchir. Ils ont besoin, parce qu'ils n'ont pas le temps de le faire, d'avoir une vision qui excède leur propre usage et donc de savoir comment cela se passe en ligne et ailleurs. C'est pour cela que j'ai montré beaucoup d'exemple de bases de collection en ligne, de musées très différents. C'est justement ça l'intérêt, le fait que l'on n'a pas le temps dans sa vie réelle, professionnelle de s'informer au-delà de ce qu'on a besoin de faire. Disons, que ceux qui en savent plus que moi sont venus m'écouter, j'estime qu'ils en savent plus que moi mais j'ai un autre regard et ils ont envie d'en parler en fait, je crois, c'est comme ça que je l'ai interprété.

11 – Le fait que sur Flickr on trouve un petit peu à boire et à manger en matière d’indexation, les professionnels se posent-ils la question de normaliser davantage les pratiques ?

Réponse : Alors cela me paraît une question intéressante parce que d’une part le mouvement d’indexation « sauvage », je ne dis pas sauvage dans le sens péjoratif mais disons que c’est un petit peu ça, je trouve cela passionnant ! C’est passionnant parce que pour arriver à décrire une image, il faut la regarder, et avant on n’a jamais été sûr, quoi que l’on fasse, qu’on la regardait bien. Qu’est-ce qu’on regarde ? On ne sait jamais ce qui est vraiment regardé, on ne peut pas le savoir. Là, ceux qui se prennent un peu au jeu de l’indexation, au moins regardent quelque chose dans l’image, pour la décrire, l’identifier. Alors, en même temps, il y a Folksonomie et Folksonomie ! Il y a des mises en situation de public non professionnel qui se mettent à tagger, pour dire les choses autrement, et ça va être différent. Ces situations sont différentes et il faut les prendre en considération. Je vous parlais des groupes formés par plusieurs musées, celui du MET mais il y en a d’autres, sur Flickr. On invite les gens à faire des photos dans le musée, les mettre en ligne et à les tagger. Dans ces conditions là, ce qui m’a beaucoup frappé, c’est que ceux qui prennent les photos et les mettent en ligne dans ce cadre là n’identifient pas vraiment l’image mais donnent quelques mots clés qui ont plus à voir avec le nom de l’œuvre ou la situation qu’ils ont vécue avec cette œuvre. En revanche, dans le cadre de ce long processus expérimental qu’est le « STEVE Museum » que vous connaissez sans doute, des chercheurs ont incité des publics à tagger des représentations d’œuvre. Ils ont déjà à une deuxième phase d’expérimentation et ils ont mis en ligne les résultats, c’est passionnant ! On a donc accès aux résultats de tagging des œuvres. Là on voit vraiment qu’il y a une recherche de la part du public qui s’est prêté au jeu pour être précis dans la description. Parce qu’il y a un cadre, un objectif ! Il n’est pas question de laisser jusque quelques tags mais de vraiment décrire l’œuvre. Les résultats de cette étude sont absolument passionnants. D’ailleurs je vais en faire un article court, une traduction parce que je trouve que ça n’a pas été traité, c’est dommage, c’est important. Il en ressort des choses très intéressantes. Un des résultats qui m’a frappé à la lecture de ce rapport c’est que d’une manière générale il y a 30% des descripteurs qui sont utiles à la recherche des images, que ce soit dans les tags créés par les professionnels ou que ce soit par les tags des non-professionnels. Il y a toute sorte d’indications quantitatives qui sont données dans cette étude et qui sont vraiment intéressantes. Deuxième élément que j’ai trouvé très passionnant, c’est que les questions que se posaient deux qui étaient à l’initiative de cette expérimentation, c’était de savoir dans quelle mesure ces tags seraient utiles aux professionnels des musées pour qu’on puisse circuler mieux dans les bases de collection mises en ligne. C’était déjà un à priori, une hypothèse quand même très ouverte. Et d’après les résultats de l’étude, ils conviennent du fait que les apports du public, mais d’un public, je le répète, qui fait l’effort de décrire véritablement, sont réels et permettent vraiment de circuler mieux dans les bases de données. Je vais faire un article bientôt là-dessus, il faut lire ce rapport. C’est en anglais mais ça vaut le coup.

12 - Pour terminer, quelles sont vos attentes ou vos espoirs concernant l’entente entre ces deux domaines que sont les Sciences de l’information et de l’histoire de l’Art ?


Réponse : je me sens très partie prenante d’un mouvement international qui s’appelle les « Digital Humanities » qui fait en sorte que les Sciences Humaines d’une façon générale se donnent les moyens de réaliser les choses avec la technique et en même temps réfléchissent

aux nouvelles possibilités que leur apportent les NTIC. C'est mon orientation principale si vous voulez. Il se trouve qu'il y a beaucoup de laboratoires, de centres qui se créent, en Europe ou aux Etats-Unis, et qui font que dans les universités, les enseignants, les étudiants et les chercheurs peuvent expérimenter la mise en œuvre de certains projets qui ont à voir ce qu'ils étudient : histoire de la littérature, l'histoire de l'art, la philosophie ... j'espère que l'on va aller vers cela.


Fin de l'interview (durée : 1h)

ANNEXE 4- INDEXATION SUR FLICKR

Voici un échantillon de ce qu'on l'on peut trouver comme type de tag sur le site Flickr.




Carnegie Hill, NY, États-Unis est le lieu où cette photo a été prise le 15 janvier 2010.



190 vues 1 commentaire

Cette photo fait partie de

Galerie de miguelino (1317)



Cette photo est également visible dans

Metropolitan Museum New York (album: 28)

Tags

Nueva York • New York • NYC • NY •
Metropolitan • Museum • art • arte •
Olympus • E-30 • Zuiko • ZD14-54mmll

Licence

TAGS : Nueva York – New York – NYC- NY- Metropolitan – museum – art- arte – Olympus – E-30 – Zulko - ZD14-54mmll



Cette photo a été prise vers le 1947.

989 vues 15 commentaires 92 favoris

Cette photo fait partie de

Galerie de The Library of Congress (10314)

photo la plus récente

Cette photo est également visible dans

Gottlieb Jazz Photos (album: 518)

Tags

Library of Congress • Ella Fitzgerald • single band • stage • microphone • hat • woman • Dizzy Gillespie • jazz

Afficher les tags de programmation (5)

Licence

Aucune restriction de copyright connue [?]

Confidentialité

TAGS : Library of Congress- Ella Fitzgerald – singer – band – stage – microphone – hat – woman – Dizzy Gillespie - jazz



70 vues

Cette photo fait partie de

Galerie de mbell1975 (12529)



Cette photo est également visible dans

New York Metropolitan Museum ... (album)

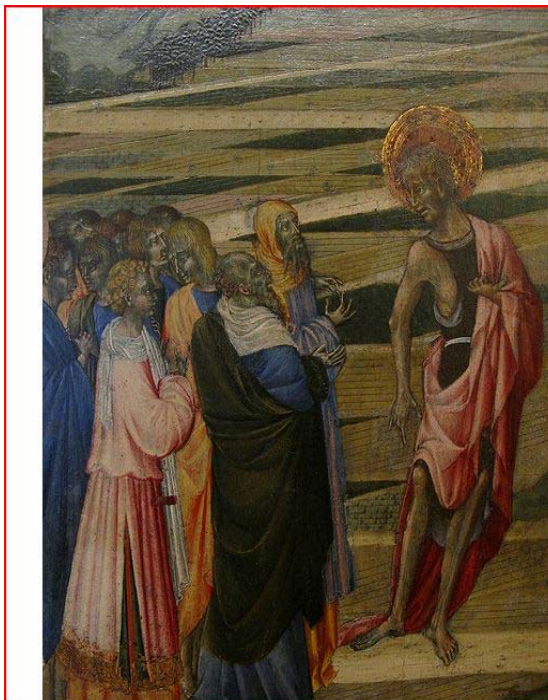
Paintings and Tapestry (album)

Tags

New • York • Metropolitan • Museum • Art • Egyptian • Collection • egypt • drawing • writing • city • New York • New York City • US • USA • Met • Antiquities • Antiquity • Ancient



TAGS : New – York – Metropolitan – Museum – Art – Egyptian – Collection – Egypt – drawing – writing – city – New York – New York City – US – USA – Met – Antiquities – Antiquity – Ancient



Par Thierry Bouts
Thierry Bouts

Cette photo a été prise le 8 février 2010 avec une Canon PowerShot G10.

16 vues

Cette photo fait partie de

Galerie de Thierry Bouts (305)



Cette photo est également visible dans

Geometrie, Geometry (groupe)

Tags

Petit Palais • Avignon • Saint Jean-Baptiste • géométrie • géométrisation du réel • perspective • renaissance

Licence

© Tous droits réservés

Confidentialité



TAGS : Petit Palais – Avignon – Saint Jean-Baptiste – géométrie- géométrisation du réel – perspective - Renaissance

ANNEXE 5 – INDEXATION SUR LE SITE STEVE MUSEUM


Steve est une collaboration de professionnels de musées et d'autres qui croient que le marquage social peut fournir de nouvelles façons de décrire et d'accéder aux collections du patrimoine culturel et favoriser la participation des visiteurs avec des objets de collection.



Fashion Plate (Paris Promenade Dress)

Top 100 Terms

abstract ancient asian beautiful beige **black blue** boat branches bridge **brown**
 building buildings chair cloth clouds color dark darkgray darkgreen darkorange darkred darkslategray
 design dress exhibitions eyes fabric face fashion flower flowers forest forestgreen gold good grass
 gray green grey hair hat hills horse house lady landscape lavender leaves lightblue lightgray
 lightpink lines man maroon men metal modern mountain mountains nature nice old olive
 orange orangered pattern pencil pink portrait purple **red river** road rocks royalblue sculpture
 sea seagreen sienna silver simple sitting sketch sky slategray stone **tan tree trees**
 village wall water **white** window windows woman women wood yellow



Williamsport, 1867 - 1933

Created

early 1920s

Institution

Los Angeles County Museum of Art



▼ **Tags** ► **All Metadata** ► **More Like This**

american,darkbackground americandarkbackground **bluebowtie** chair,sitting,portrait,man
 chairsittingportraitman cigarette early1920s,unitedstates early1920sunitedstates greysuit guitar guitar,musician,pedro
 guitarmusicianpedro hat kneecrossedleaningtable kneecrossedleaningtable man mexican painting,oiloncanvas,h
 paintingoiloncanvas hat sitting smoking whiteshirt,graypants whiteshirtgraypants

ANNEXE 6 INDEX ET THESAURUS ICONOGRAPHIQUE

Fi Collection Guéguen, photographies

Index

Légende

Descripteur matière libre (desc. libre)
Nom géographique (nom géo.)

- abbaye : 8 Fi 365-367.
allégorie : 8 Fi 122.
animal d'élevage : 8 Fi 2-3, 21-25, 36-41, 54-55, 67-68, 109, 112, 140-146, 188, 199, 206-210, 316, 416-419.
animal domestique : 8 Fi 106-108, 110, 231, 234, 273-274, 330, 380.
apiculture : 8 Fi 1.
artisanat : 8 Fi 167.
automobile : 8 Fi 66, 211, 214, 321, 370-371.
barrage : 8 Fi 118-119, 163.
bas-relief : 8 Fi 122.
basilique : 8 Fi 373, 384.
bateau : 8 Fi 302.
bateau à vapeur : 8 Fi 12, 127, 305.
bateau à voiles : 8 Fi 97, 127-128, 307, 332, 363.
bateau à voiles : 8 Fi 13, 294.
bateau de guerre : 8 Fi 14-15.
bateau fluvial : 8 Fi 12, 74, 77-78.
Belle-Ile-en-Mer : 8 Fi 332-334, 339.
bicyclette : 8 Fi 88, 341.
Blavet (rivière) : 8 Fi 74, 76-78, 270.
Boloré (Saint-Aignan ; lieu-dit) : 8 Fi 179.
Bon-Repos, Abbaye de (Saint-Gelven) : 8 Fi 365-367.
Bot-Pléven, Château de (Saint-Aignan) : 8 Fi 1, 4-5, 17-18, 20, 29, 43, 60, 68-69, 80, 82-83, 86-87, 91, 105-117, 130-146, 168-171, 184-200, 203-213, 218, 221-222, 224-229, 232-236, 239, 242, 245-257, 261-265, 269, 273-279, 281-291, 316-319, 322, 324-331, 335-338, 344-359, 362, 368-371, 380, 388, 390, 392-393, 409, 412-415, 418-419.
bovin : 8 Fi 145-146, 197-200, 206-207, 209-210, 236, 316, 318-319, 418-419.
café : 8 Fi 340, 397-398.
canal : 8 Fi 270.
Carnac : 8 Fi 92-96, 98-100.
caserne : 8 Fi 76, 381.
cathédrale : 8 Fi 266.
cavalier : 8 Fi 2-3, 6-7, 21-25, 27, 36-38, 41, 49-53, 67, 70, 80, 239, 245, 264, 324-327, 416-417.
cérémonie militaire : 8 Fi 147-155.
cérémonie religieuse : 8 Fi 156-159.
chapeau : 8 Fi 248.
chapelote : 8 Fi 95, 158-159.
charrète : 8 Fi 200.
- Calliope** : Le Sueur 535.
Callisto (Coréus et) : Fragonard 268.
Callisto (Jupiter et) : Marot 554.
Calypso (Télémaque et) : Raoux 701.
Camille et le maître d'école de Faléris : Poussin 660.
Camondo (comte I. de) : Perronneau (attr.) 638.
Campagne romaine (Paysage de la) : Dughet 245.
Campaspe (Apelle peignant) : Wleughels 931.
Campo Vaccino : Gellée 300.
Cananéenne (Le Christ et la) : Drouais (G.J.) 237.
Canaque (Saint Philippe baptisant l'eunuque de la reine) : Berrin 18.
Cangoxima (Saint François et la fille d'un habitant de) : Poussin 665.
Cany (marquis de) : Bourdon (?) 80.
Capitole (Vue d'un port avec le) : Gellée 299.
Caprarola (Escalier tournant à) : Robert 731.
Caracalla (Homme tenant un buste de) : Bourdon (?) 79.
- (Sévère et) : Greuze 319.
Carlos (Guérison miraculeuse de Don) : Licherie 546.
Caron : Subleyras 784.
Carrasses (Les deux) : Gillot 312.
Carthage (Didon faisant bâtir) : Sève (G. de) 768.
Carthaginois (Bataille entre Romains et) : Dieu 218.
Casscelles de Fivoll : Fragonard 266.
Cassagnade (Mme) : Licherie 546.
Castagnier (F. de) : Rigaud 727.
Castor : Renou 704.
Catherine II : Greuze 324.
Cavalerie (Combat de) : Courtois 160, 162.
Cavaliers (Rencontre de) : Courtois 159.
Cavedone : Vouet (d'après) 914.
Caylus (comte de) : Le Brun 434.
Cazes : Perrot 639.
Cène : Champaigne 103, 104, Le Nain 475.
Centtaures (Hercule combat les) : Boullongne (B.) 59.
Céphale et l'Aurore : Boucher 50.
Cérés implorant le soleil : Durameau 259.
- **protégeant Triptolème** : Dumont 249.
- (L'Amour se réfugie dans les bras de) : Le Sueur 526.
- (Vénus, Bacchus et) : Boullongne (B.) 62.
Cervantes : Coypel (C.A.) 181 à 183, Natoire 587.
Châalis. Abbaye : Restout 709.
Challe (Noël) : Anonyme 1015, 1016.
Châlons (duc de) : Lépicié 493.
Chambré (A. de) : Le Sueur 497.
Chamillart (M. de) : Bourdon (?) 80.
Champaigne (J.B.) : Champaigne 99.
Champaigne (Ph. de) : Champaigne (J.B.) 90.
Champaigne (Ph. de) : Anonyme 945, 950.
Champaigne (Sœur Catherine de Sainte Suzanne) : Cham-
- Chemise enlevée** : Fragonard 276.
Chéron (E.S.) : Chéron 152.
Chesneau (C.) : Tocqué 799.
Cheval blanc : Van der Meulen 841.
- **isabelle** : Desportes 215.
Chevalier (A.) : Debucourt 200.
Chien gardant du gibier : Desportes 212.
- (Bator et perdrix gardés par un) : Oudry 607.
- (Gibier, fleurs, fruits et un) : Desportes 210.
- (Gibier gardé par un) : Anonyme 1019.
Chiens et lapin : Desportes (N.) 217.
- (Gibier, légumes, fruits et deux) : Desportes 209.
Chimiste dans son laboratoire : Chardin 128.
Choiseul (duc de) : Van Loo (L.M.) 858, 859.
Choisy. Château : Chardin 147, 148, Vernet 870 à 873.
Christ à la colonne : Vouet (attr.) 915.
- **aux outrages** : Valentin (d'après) 826.
- **chez Marthe et Marie** : Jouvenet 361.
- **en croix** : Champaigne 112, Champaigne (atel.) 113, Dorigny 220, Rivalz 730, Tournier 807, Vouet 905.
- **et la Cananéenne** : Drouais (G.J.) 237.
- **et les enfants** : Bourdon 77.
- **mort** : Champaigne 99, Le Brun 433.
- **servi par les anges** : Le Brun 437.
- (Baptême du) : Restout 708.
- (Ensevelissement du) : Varin 860.
- (Entrée du - à Jérusalem) : Wampe 916.
- (Tentation du) : Verdier 861.
- **Voir aussi Jésus**.
Christiani (baron de) : Deshayes (?) 205, Drouais (F.H.) 235, Hilair 342.
Chrysis (Ulysse remet - à son père) : Gellée 308.
Cible d'amour : Boucher 47.
Cimon et Ephigène : Blanchard (Jacques) 21.
Cimon (Libéralité de) : Hallé (N.) 338.
Cielle passant le Fibre : Stella 777.
Cléopâtre avalant le poison : Coypel (C.A.) 179.
- (Débarquement de) : Gellée 305.
Clermont d'Amboise : Dufresnoy 243.
Clio, Euterpe et Thalie : Le Sueur 531.
Clotilde (Madame) : Drouais (F.H.) 233.
Clotilde (Madame) : Drouais (F.H.) 231.
Codazzi (N.V.) : Houasse 344, 345.
Colbert : Le Brun (atel.) 461.
Coligny (G. de) : Anonyme 937.
Collin-maillard : Fragonard 281.
Colisée (Intérieur du) : Robert 736.
Collé (Piron à table avec Vadé et) : Jeaurat 360.
Collection (Ancienne) : Allegrain 1, Baugin 13, Blin 25, 26, Boucher 29, 48, Boullongne (attr. à B.) 64, Cazes 89, Champaigne 115, Courtois 159, 161, Coypel (N.N.) 195,

1) Index
iconographique des
Archives
départementales du
Morbihan

2) Index
Iconographique
du Louvre

