

Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI)

Manuel Zacklad

▶ **To cite this version:**

Manuel Zacklad. Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI). 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, May 2007, Montréal, Canada. sic_00202440

HAL Id: sic_00202440

https://archivesic.ccsd.cnrs.fr/sic_00202440

Submitted on 6 Jan 2008

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.

M. Zacklad, Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI), 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. Disponible sur http://www.cais-acsi.ca/proceedings/2007/zacklad_2007.pdf

Manuel Zacklad
Université de Technologie de Troyes, ICD, Equipe Tech-CICO, FRE CNRS 2848
12, rue Marie Curie, BP 2060, 10 010 Troyes Cedex, France
manuel.zacklad@utt.fr

Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information (ROI)[1]

Résumé : Cet article compare différents systèmes d'organisation des connaissances (classifications, thésaurus, ontologies formelles, ontologies sémiotiques, folksonomies) selon différents critères pour évaluer leur pertinence en regard de la Recherche Ouverte d'Information.

Abstract: This paper compares several Knowledge Organisations Systems (classifications, thesaurus, formal ontology, semiotic ontology, folksonomy) according to different criteria in order to assess their relevance according to Open Information Research.

1. Introduction

En matière de recherche d'information, la généralisation du web évoque surtout les moteurs de recherche et les systèmes d'indexation automatiques sur lesquels ils sont basés. Mais elle a aussi vu récemment apparaitre de nouvelles formes d'indexation manuelles, que nous appellerons ici *indexation interprétatives*, réalisées par des groupes de professionnels ou d'utilisateurs volontaires qui conçoivent des *sémiologies externes* (pour utiliser l'expression de Hjelmslev, Badir 2000) plus ou moins structurées. Aux thésaurus utilisés par les bibliothécaires sont venus s'ajouter les ontologies dans le champ du web sémantique et plus récemment les folksomies dans les portails documentaires relevant du Web 2.0. L'ensemble de ces dispositifs produisent une classification plus ou moins systématique de l'information disponible sur le web et peuvent être considérés comme relevant de la catégorie générale des Systèmes d'Organisation des Connaissances (en anglais Knowledge Organisation Systems, p.e. Hodge 2000).

Au sein de l'équipe Tech-CICO de l'Université de Technologie de Troyes, nous défendons depuis plusieurs années un courant original à l'intérieur du Web Sémantique que nous avons nommé le *web socio-sémantique*. Nous considérons en particulier que la constitution des ressources terminologiques appelées à constituer des sémiologies externes pour les documents du web dans un contexte intranet ou internet dépend toujours des transactions coopératives entre les communautés d'usagers impliquées dans l'exploitation, le classement et la recherche de ces documents. Dans l'approche du web socio sémantique, la sémantique est donc pour nous intrinsèquement dépendante des interactions entre des communautés d'acteurs, des finalités de ces interactions et des problèmes de coordination qu'ils sont susceptibles de rencontrer. Dans ce contexte, nous

[1] Nous remercions chaleureusement l'équipe du Web Socio Sémantique du laboratoire Tech-CICO à l'UTT pour les discussions stimulantes ayant permis la rédaction de cet article: JP Cahier, A. Bénel, H. Zaher. & C. Lejeune (CEMAD Liège). Nous remercions également T. Fournier conservateur au SCD de l'UTT.

définissons des ontologies sémiotiques (Zacklad 2005a) en les différenciant des ontologies formelles (ou ontologies computationnelle) du Web Sémantique, mais pas encore des thésaurus (Caussanel et al. 2003).

Le web socio sémantique est lui-même associé à une vision originale de la recherche d'information dite *Recherche Ouverte d'Information* (ROI, Zacklad 2005b, Zaher et al. 2006), dans laquelle la finalité principale de la recherche n'est pas la récupération ou l'accès à un document ou à un ensemble de documents pertinents par rapport à une interrogation déjà formulée, mais l'assistance à un utilisateur engagé dans une démarche d'enquête (au sens de John Dewey 1938) à travers laquelle il sera amené à préciser *simultanément les termes de sa demande et les ressources documentaires ou non documentaires susceptibles de contribuer à y répondre*. Cette vision de la ROI est basée sur le développement de plateformes multifonctionnelles qui associent la consultation des systèmes d'organisation des connaissances utilisant des sémiologies (annuaires, cartes de thèmes, thésaurus…), l'utilisation de moteurs de recherche, l'interrogation de base de données structurées, l'accès à des ressources en ligne et des échanges communautaires[1].

Sur la base de cette vision de la recherche d'information, nous proposerons une comparaison des systèmes d'organisation des connaissances. Selon nous, la comparaison des avantages respectifs et des complémentarités entre ces systèmes (vocabulaire contrôlé, thésaurus, ontologies formelle et sémiotique, folksonomie…) ne doit pas se faire uniquement en considérant les propriétés structurelles des métasémiotiques ou des métalangages sur lesquelles ils s'appuient, mais en tenant compte de leur rôle dans la conduite des activités de ROI et des modalités de leur constitution et de leur maintenance. Cela implique de tenir compte des efforts entrepris pour constituer et maintenir les sémiologies de manière interprétative en fonction des caractéristiques acteurs impliqués et de la dynamique sociale associée (Caussanel et al. 2003, Lejeune 2006).

2. Recherche Ouverte d'Information et démarches d'enquête

Dans les domaines de l'informatique et des sciences de l'information, la recherche d'information (« information retrieval » en anglais), désigne essentiellement la problématique de l'accès à un document ou à un fragment documentaire numérisé dont la localisation n'est pas connue mais dont l'existence ou la pertinence est acquise. Mais dans de nombreux cas, la recherche d'information s'inscrit dans la problématique de l'analyse d'une situation complexe dont on cherche à percevoir les contours sans savoir à l'avance s'il existe des ressources documentaires susceptibles de répondre au besoin. Pour différencier ces cas de la recherche d'information classique nous avons proposé de les caractériser comme relevant d'une *Recherche Ouverte d'Information* (ROI, ou en anglais Open Information Research, Zacklad 2005, Zaher et al. 2006).

La Recherche Ouverte d'Information (ROI), correspond ainsi à la phase d'une enquête permettant aux personnes de construire une représentation de la situation problématique en rassemblant les ressources informationnelles permettant d'instruire les questions posées. La ROI ne consiste pas seulement à accéder à des documents ou à des fragments documentaires pertinents mais à délimiter les contours du problème en identifiant les informations existantes et celles qui ne le sont pas : c'est un processus de découverte et d'apprentissage permettant de poser un problème dans le cadre d'une démarche d'enquête. Cette définition rejoint celle de chercheurs portant une attention particulière aux phénomènes de sérendipité (découverte fortuite) associés à la RI, comme Toms

(2000), qui parle de « serendipitous information retrieval » ou Ertzcheid et Gazzelot (2003) qui mettent en avant les « processus d'apprentissage dynamique » associés à la RI qu'ils souhaitent, comme nous, plus ouverte aux phénomènes de complexité et de créativité.

Pour mieux appréhender les phénomènes de ROI et l'aide que peuvent fournir les systèmes informatiques, il nous faut resituer plus globalement celle-ci dans les démarches d'enquête. Nous considérerons que l'enquête donne lieu au déploiement de deux phases principales. La première est une phase de *recherche d'information*, ou de compréhension, permettant l'analyse de la situation, des options, des ressources disponibles. La seconde est une phase *de mise en œuvre*, de traitement, de décision, de synthèse. Ces deux phases ne sont pas strictement ordonnées et on peut les voir également comme deux dimensions de l'enquête qui s'entremêlent de manière étroite. Ainsi, les phases de mise en œuvre sont également une manière d'acquérir de l'information qui contribue à la RI. La phase de recherche d'information, qui nous intéresse principalement dans le cadre de cet article, peut elle-même être de nature essentiellement intra-documentaire ou de nature extra-documentaire. Dans le premier cas, l'analyse de la situation se fait en exploitant des documents qu'il importe d'identifier, de localiser, d'acquérir. Dans le second, elle se fait par observation directe de la situation et de ses ingrédients ou en interagissant avec des personnes susceptibles d'apporter des réponses aux interrogations rencontrées.

Première typologie des enquêtes	**Mise en œuvre exprimée de manière non documentaire**	**Mise en œuvre exprimée de manière documentaire**
Recherche d'information extra-documentaire	*Extra-documentaire pur* Les sources primaires sont extra-documentaires et la mise en œuvre est extra-documentaire : p.e. choisir un ami pour l'inviter à une sortie	*Extra /Intra-documentaire* Les sources primaires sont extra-documentaires mais la mise en œuvre est de nature documentaire : p.e. rédiger le compte rendu de plusieurs visites d'appartements
Recherche d'information intra-documentaire	*Intra /Extra-documentaire* Les sources primaires sont documentaires mais la mise en œuvre est extra-documentaire : p.e. faire un voyage lointain après s'être documenté	*Intra-documentaire pur* Les sources primaires sont documentaires comme la finalité qui se traduit par une mise en œuvre documentaire : p.e. rédaction d'un article à partir d'autres articles

Tableau 1 - Caractère extra-documentaire ou intra-documentaire de l'enquête

Les documents mobilisés dans les phases de recherche d'information sont de nature diverses. Dans la tradition de la documentologie on établit ainsi une distinction entre des informations ou documents primaire, secondaire voire tertiaire. Les documents primaires sont disponibles quand la situation a été au préalable en partie documentée. Ils constituent des sources plus ou moins anciennes, pertinentes et fiables qui contribuent à alimenter le processus de RIO. Les documents secondaires facilitent l'accès au document primaire soit en présentant une synthèse de leur contenu soit en contenant des annotations-index (Zacklad 2007b) renvoyant de manière plus ou moins systématique aux documents secondaires. Nous proposons de nommer ce deuxième type de documents, document secondaires articulatoires, à la fois parce que les annotations-index qu'ils contiennent articulent de manière externe les documents primaires (Zacklad 2007a) et parce qu'ils sont à la fois codifiés et mis à jour pour permettre la coordination entre les professionnels de la documentation et/ou entre ces professionnels et les usagers (Zacklad 2007c).

Ces documents nous intéresseront bien sur au premier chef puisque ce sont eux qui contiennent les différentes classifications, thésaurus et autres ontologies. Enfin, il est également possible de définir des documents tertiaires qui contiennent les résultats consignés de la recherche d'information, voire des résultats de la mise en œuvre de l'enquête, quand celle-ci est également de nature documentaire. Ces documents tertiaires ont vocation à enrichir le fonds des documents primaires à la fin de l'enquête. Le tableau 1. présente une première typologie des enquêtes selon que la recherche d'information et la mise en œuvre sont exprimées de manière intra-documentaire ou extra-documentaire. Il met en évidence quatre types allant des enquêtes extra-documentaires pures dans lesquelles ni la recherche d'information ni la mise en œuvre ne sont documentaire à des enquête intra-documentaire pures dans lesquelles la recherche d'information comme la mise en œuvre sont documentaire (cf. tab. 1).

Phase de l'enquête / Type d'enquête	**RI documentaire**	**RI extra-documentaire**	**Phase de mise en œuvre documentaire**	**Phase de mise en œuvre non documentaire**
Localisation-accès	Chercher dans des annuaires la localisation d'un magasin de disques	Demander à des personnes la localisation d'un magasin de disque	Télécharger le disque en format numérique (document numérique)	Acheter le disque et l'emmener avec soi
Sélection-décision	Rassembler les différents catalogues rassemblant les voyages existants	Faire le tour de ses connaissances pertinentes en les interrogeant sur les voyages existants	Faire un tableau comparant les différents types de voyages possibles et permettant de conduire à la décision	Prendre la décision organisant ses arguments dans un dialogue intérieur ou avec des tiers
Synthèse-conception	Rechercher des documents permettant d'imaginer un nouveau voyage	Discuter et faire des visites pour imaginer un nouveau voyage	Rédiger une proposition de voyage inédite en documentant les étapes et les options prévues	Réaliser le voyage selon les étapes et options prévues

Tableau 2. Nature des enquêtes

Le deuxième tableau présente une typologie des enquêtes inspirée des typologies utilisées en psychologie cognitive pour classer les méthodes de résolution de problème. On y distingue trois types d'enquêtes selon leur finalité principale :

1. Enquête de localisation et d'accès : dans ces enquêtes la situation problématique est générée par l'absence d'un objet connu. La démarche d'enquête vise à permettre de trouver la localisation de cet objet et la mise en œuvre consiste à y accéder ou à l'acquérir.
2. Enquête visant une sélection ou une décision : dans ces enquêtes la situation problématique est générée par l'absence d'un objet dont on connait l'existence parmi d'autres mais qui n'a pas encore été identifié. La démarche d'enquête vise à permettre de sélectionner l'objet souhaité au sein d'un ensemble en décidant d'y accéder ou de l'acquérir.
3. Enquête visant une synthèse ou une conception : dans ces enquêtes la situation problématique est générée par l'absence d'un objet n'ayant pas encore d'existence. La démarche d'enquête vise à permettre de synthétiser ou de concevoir cet objet.

Après cette présentation des types d'enquêtes, nous allons maintenant passer en revue les principaux systèmes d'organisation des connaissances existant afin de pouvoir les évaluer en tenant compte de leur contribution à la ROI.

3. Les classifications épistémiques universelles de la bibliothéconomie

Les schémas de classification utilisés dans le domaine de la bibliothéconomie (la classification décimale de M. Dewey et la classification décimale universelle de P. Otlet et H. La Fontaine) sont les outils aujourd'hui les plus répandus pour le classement documentaire au sein des bibliothèques. Ils sont basés sur une approche épistémologique visant à classer l'ensemble du savoir humain selon une division hiérarchique. Ils visent à la fois à faciliter le rangement des exemplaires physiques et à fournir une organisation systématique des ouvrages permettant « *au chercheur de repérer des documents pertinents qu'il ne connaît pas encore* » (Hudon 2001). Alors que la CDD requiert en principe une classification unique de chaque ouvrage, la CDU permet une combinaison d'indices qui, bien qu'accroissant la précision, n'en facilite pas toujours l'usage.

Dans un effort similaire de couverture du sens du sujet d'un livre par combinaison d'indices, le bibliothécaire indien S. R. Ranganathan propose en 1924 la « Colon Classification » (CC) basée sur le principe de l'addition de facettes classificatoires (Maniez 1999). Chaque sujet doit être qualifié de cinq manières, selon la personnalité, la matière, l'énergie, l'espace et le temps. La CC à donné lieu à plusieurs améliorations (Maniez 1999) concernant notamment le caractère plus ou moins universel des facettes retenues dans chaque domaine de connaissance. Cependant, pour chaque domaine, le principe est toujours de définir un vocabulaire universellement accepté pour faciliter le rangement des livres.

4. Les thésaurus

Selon la norme internationale ISO 2788 (1986), les thésaurus nés dans les années 50, sont le « *vocabulaire d'un langage d'indexation contrôlé organisé formellement de façon à expliciter les relations a priori entre les notions (par exemple relation générique-spécifique)* ». Selon la même norme un langage d'indexation est un « *ensemble contrôlé de termes choisis dans une langue naturelle et utilisés pour représenter sous forme condensée, le contenu des documents* » (Saadani L. & Bertrand-Gastaldy S. 2000). Contrairement aux usages des classifications, simple (CDD) ou analytico-synthétique (CC), qui visent à définir un indice représentant au mieux le contenu du document, indice pouvant résulter dans le cas de la CC d'une combinaison d'indices élémentaires assemblés selon une syntaxe précise représentant différentes facettes, l'usage d'un thésaurus autorise l'indexeur à utiliser autant de descripteurs que bon lui-semble.

Le thésaurus contient un lexique (l'ensemble des termes d'une langue de spécialité) définissant les descripteurs et les non descripteurs (termes interdits), un sous-ensemble de définitions et de notes d'application pratique et une structure classificatoire exprimée par des relations sémantiques entre les termes du lexique : relation d'équivalence intra linguistique (synonymie), relation d'équivalence inter linguistique (traduction), relation hiérarchique, relation d'association. Alors que les classifications organisent les sujets des documents, les termes des thésaurus visent à décrire des concepts. Selon Maniez (1999), la distinction entre sujet et concept est assimilable à la distinction entre parole et langue. Alors que les sujets sont en nombre potentiellement infini, les concepts correspondent à

un ensemble restreint de notions associées aux ressources cognitives d'une collectivité et dépendant notamment de sa langue (« *Ce qui distingue le concept du sujet est son statut sociolinguistique et son statut cognitif* » Maniez 1999). C'est la raison pour laquelle les concepteurs d'un thésaurus vont se fixer sur une expression linguistique, le descripteur, et le considérer « toute choses étant égales par ailleurs » comme le meilleur représentant du concept visé. C'est à cette condition qu'ils seront en mesure d'établir des équivalences inter-linguistiques (traduction), « l'opérateur d'équivalence » étant précisément le concept. Si les concepteurs de thésaurus ont intérêt à travailler avec les terminologues, c'est pour identifier les variations en discours du concept et ne retenir qu'une forme canonique.

Soulignons à ce stade deux caractéristiques. D'une part, comme le rappelle Maniez, les concepts du thésaurus sont spécialement définis à fin d'indexation à partir d'un fonds documentaire donné pour en faciliter l'interrogation ultérieure. C'est ce critère qui justifie la sélection du descripteur parmi d'autres possibles. D'autre part, si l'on suit J. Maniez, les thésaurus s'appuient sur une caractérisation des concepts qui les font au moins pour partie dépendre des langues et des mises en discours. Or, cette dépendance du concept vis à du système de la langue et la variabilité intrinsèque induite par cette dépendance, n'est pas une vision partagée par tous les concepteurs de Système d'Organisation des Connaissances, en particulier dans le domaine des ontologies.

5. Les ontologies formelles et le Web-Sémantique

Comme le souligne J. Charlet (2002), l'apparition des ontologies entendues comme une nouvelle approche de la modélisation des connaissances, s'est fait clairement dans le contexte de l'ingénierie informatique. Les ontologies s'inscrivent dans la continuité de nombreux travaux sur la représentation des connaissances, réseaux sémantiques, cartes conceptuelles, graphes conceptuels et leur popularité à principalement bénéficiée du développement du Web Sémantique, une vision prospective et normative du web proposée par Tim Berners Lee (2001) en 1994, sans avoir connu, à ce jour, le succès escompté. Comme les principaux langages de représentation des connaissances issus des sciences cognitives et de l'Intelligence Artificielle, les ontologies visent également à décrire des concepts, qui sont appréhendées comme des représentations mentales plus ou moins universelles ou comme des catégories a priori largement partagées dans la droite ligne de la philosophie de la connaissance (Guarino 1998).

On distingue des ontologies de différents niveaux de généricité : des ontologies dites de haut niveau qui contiennent « *des concepts très généraux comme l'espace, le temps, la matière, les objets, les événements, les actions, etc. [qui] ne dépendent pas d'un problème ou d'un domaine particulier* » (Lando 2006) ; des ontologies de domaine (médecine, architecture, mécanique..) ; de tâche (diagnostiquer, enseigner) ; voire d'application, dans lesquelles les concepts appartiennent à un domaine et à une tâche particulière (enseigner la médecine). Les ontologies ont plusieurs caractéristiques importantes :

- Comme d'autres langages de représentation des connaissances, elles n'ont pas une vocation exclusivement documentaire au sens de l'indexation et de la recherche d'information mais *elles visent aussi à participer de l'ingénierie des connaissances d'un domaine* et en particulier à « spécifier explicitement une conceptualisation » pour reprendre les termes de T. Gruber (1993).

- Point corolaire du précédent elles n'ont pas à être conçues strictement à partir d'un fonds documentaire qu'elles viseraient à indexer. Même si les ingénieurs de la connaissance responsables de leur conception utilisent souvent des outils terminologiques appliqués à des textes de référence du domaine, *ils peuvent également s'appuyer sur d'autres sources d'information* comme des entretiens auprès d'experts, l'analyse de bases de données, ou des conceptualisations ad hoc issues de leur propre synthèse des connaissances du domaine considéré.
- Bien que la dimension documentaire ne soit pas, comme on l'a dit, leur justification intrinsèque, leur ancrage au sein du web sémantique tendrait néanmoins (pour autant qu'elles se diffusent effectivement) à leur faire jouer un rôle essentiel dans la recherche et la mise en relation d'information. Mais *l'information dont elles visent à faciliter l'accès est d'abord celle du web invisible*, celui constitué par les multiples bases de données qui consignent l'information structurée des processus d'affaires et des références techniques. Cette vocation première est aujourd'hui concurrencée par l'usage des ontologies pour annoter des documents plus classiques, mais cette tendance reste minoritaire.
- En conséquence, les ontologies formelles ne sont pas faites pour être directement exploitées par des usagers humains engagés dans une navigation hypertextuelle comme cela pourrait être le cas pour une classification documentaire ou un thésaurus. Au contraire, elles sont le plus souvent conçues pour être *exploitée par des programmes informatiques* (des agents de recherche automatique sur le web), l'utilisateur interagissant avec l'agent à l'aide d'un formulaire ou d'un autre type de langage de requête.
- De ce fait, les ontologies gagnent à être *représentées à l'aide de langages formels*, le standard proposé par le W3C (World Wide Web Consortium) étant aujourd'hui OWL (Ontology Web Language), qui s'exprime à partir du langage RDF (Ressource Description Framework), proche des réseaux sémantiques, lui-même exprimé à l'aide de balises XML comme tous les langages du web sémantique. Les classifications exprimées en OWL s'appuient sur une stricte séparation classe/instance, l'héritage de propriétés, l'expression de contraintes de cardinalité et de contraintes logiques sur les relations entres propriétés, etc. Cette formalisation extrême, vise à répondre aux objectifs ultimes du web sémantique tel que définis par T. B. Lee qui était de fournir des réponses logiquement fondées (« vraies ») aux requêtes des utilisateurs.
- Enfin, la raison d'être première des ontologies formelles, liée à la manipulation des données structurées du web invisible en réponse à des requêtes complexes sur la base d'une sémantique formelle, à des conséquences sur le type de signification associé à ces langages. La *sémantique des ontologies est une sémantique référentielle* au sens du positivisme logique, les termes recevant une valeur de vérité ancrée sur des référents externes objectivables. Les concepts décrits par les termes de l'ontologie ont donc également principalement une valeur référentielle comme dans la tradition artistotélicienne où « *le sens d'un signe est conçu comme représentation mentale (concept), et défini par ce à quoi il renvoie dans le monde (le mot « chien » « signifie » ce quadrupède à poils ras).* » (Lacour 2004, citant Rastier 1997). Or, cette vision du concept est largement incompatible avec les épistémologies de la philosophie pragmatique (J. Dewey) ou de la tradition herméneutique (cette dernière étant largement répandue dans les sciences humaines et sociales), épistémologies que nous défendons dans le cadre du web socio sémantique.

Pour conclure cette partie, insistons sur le fait que le terme d'ontologie est aujourd'hui utilisé de manière bien plus large et imprécise pour désigner toute classification aisément partageable sur le web grâce à l'usage des standards du W3C que sont le langage XML et le protocole http, sans qu'il soit nécessaire de faire référence au web sémantique formel.

6. Les ontologies sémiotiques et le web socio-sémantique

Proposé au sein de l'équipe Tech-CICO de l'Université de Technologie de Troyes et issus de préoccupations conjointes à l'ingénierie des connaissances, au CSCW (Computer Supported Cooperative Work) et au Social Informatics (Turner 2007), le web socio sémantique s'oppose à la vision logiciste du web sémantique initialement proposés T. Berners-Lee tout en s'inscrivant dans le projet d'une structuration des ressources documentaires partagées sur le web facilitant la réalisation d'activités coopératives distribuées (Zacklad 2005, Zacklad et al. 2007, Cahier 2005). Pour instrumenter cette vision du web, nous proposons un format de représentation de l'information, la métasémiotique HyperTopic, permettant construire et de partager aisément des ontologies sémiotiques de type cartes de thèmes (Cahier 2005) ou réseaux de description (Bénel 2003)[2]. L'application de la métasémiotique à un domaine de connaissance produit une sémiotique tâche-domaine pouvant servir d'index pour caractériser des « items » qui peuvent eux-mêmes être associés à des ressources (voir Zacklad et al. 2007)[3].

Les ontologies sémiotiques partagent avec les thésaurus leur structure hiérarchique rassemblant des expressions significatives du domaine selon une relation général / spécifique, sans imposer un formalisme logique ou « orienté objet » (pas de relation d'héritage au sens strict, par exemple). Elles s'en différencient également de part les caractéristiques suivantes :

- Elles sont considérées comme des ontologies dans la mesure où elles visent à classifier des situations, des personnes ou des artefacts qui ne sont pas nécessairement entièrement documentés. Si elles ne sont donc pas produites exclusivement par l'analyse d'un corpus de documents primaires, elles donnent cependant toujours lieu à la réalisation d'un document secondaire articulatoire (un document pour l'action, Zacklad 2007) dans lequel la sémiotique tâche-domaine est exprimée. Comme les ontologies formelles, ou plus largement les modèles dit de « représentation des connaissances » (p.e. cartes conceptuelles), elles sont construites par des ingénieurs de la connaissance, sur la base de l'analyse de documents primaires mais également à l'aide d'entretiens avec des experts ou par l'analyse ad hoc d'un domaine.
- Elles sont toujours multiples et visent à caractériser un ensemble d'items selon différents points de vue. Chaque point de vue est en principe défendu par un ou plusieurs acteurs et peut être socialement et/ou cognitivement conflictuel avec un autre. L'organisation de la diversité des points de vue permet de médiatiser la coopération entre des acteurs ou des communautés d'acteurs hétérogènes. Les domaines traités sont toujours en partie controversés ou sujets à des interprétations ou des formes d'expériences contrastées. De ce fait, les ontologies sémiotiques doivent être évolutives et pouvoir être facilement actualisées par les représentants des points de vue en présence ou par l'intermédiaire de l'ingénieur de la connaissance qui les accompagne. Dans certaines applications actuelles basées sur Hypertopic (Agorae, Cahier 2005), l'ingénieur de la connaissance à été le médiateur humain responsable de la construction des différentes ontologies sémiotiques. Cependant, l'objectif à terme, est que les plateformes informatiques

permettent l'élaboration dynamique, par les acteurs eux-mêmes, de ces ontologies, ce qui n'exclut d'ailleurs pas la fonction régulatrice de l'ingénieur de la connaissance.

- Les ontologies sémiotiques, cartes de thèmes ou réseau de descripteurs selon les applications, contiennent des thèmes plutôt que des concepts ou des sujets. Sous un certain angle, les thèmes pourraient être assimilés à des descripteurs visant à caractériser des « concepts sémiotiques » (voir Cahier 2005) représentés par un ou plusieurs items selon la généralité du concept considéré. Nous sommes néanmoins toujours prudents dans l'utilisation du terme de concept à cause de l'acception référentielle dominante qu'il possède dans les ontologies formelles. Notre approche du sens peut-être dite rhétorico-herméneutique dans la lignée de Rastier (2003)[4], l'appréhension du sens d'un terme dépendant au moins de trois facteurs : la proximité de ce terme vis-à-vis de termes voisins mais différents, le parcours de lecture permettant de l'appréhender, dépendant ici des associations hypertextuelles définies par l'ontologie sémiotique et la situation d'énonciation[5] (ou contrat de lecture), correspondant souvent en ingénierie des connaissances à la « tâche ».

7. Les annuaires de ressources internet collaboratifs et les folksonomies

Nous avons regroupé dans cette partie les annuaires internet et les folksonomies parce qu'ils s'appuient sur des applications qui mettent en synergie une communauté d'indexeurs coopérant à travers le web dans une logique proche de celle du web socio sémantique. Les annuaires de ressources s'inscrivent dans la logique des schémas de classification de la bibliothéconomie présentés plus haut, sans toutefois revendiquer l'organisation disciplinaire à laquelle ceux-ci se conforment (Hudon 2001). Les annuaires de ressources internet qui nous intéressent ici sont notamment ceux pris en charge par des communautés de bénévoles comme l'annuaire « libre » dmoz (http://dmoz.org/), étudié par C. Lejeune (2006, 2004). A la différence d'un annuaire comme celui de Yahoo, par exemple, construit par des salariés placés sous l'autorité de managers définissant une politique de référencement explicite, le processus social d'indexation mis en œuvre dans ces communautés relève d'un type de contrôle social qui se situe entre celui de Wikipédia et celui, plus hiérarchique, des communautés Open Source. Selon leur niveau de réputation, les membres de la communauté virtuelle peuvent contrôler des niveaux plus ou moins haut de la classification, rajouter de nouvelle branches dans les domaines au sein desquels ils font « autorité », supprimer les descriptions effectuées dans la « notice » par des participant moins renommés, etc. L'accroissement de réputation qui permet à un membre de la communauté de se voir octroyer des droits supplémentaires se fait notamment en traçant ses contributions déjà effectuées.

Mais le dispositif qui connait aujourd'hui le plus grand succès est celui constitué par les folksonomies dans le contexte des business model du Web 2.0. Les folksonomies s'appuient sur des dispositifs informatiques donnant « *la possibilité à l'usager d'indexer des documents afin qu'il puisse plus aisément les retrouver grâce à un système de mots-clés* » (Le Deuff 2007). Chaque utilisateur des plateformes telles que Del.icio.us (http://del.icio.us) ou Flickr (http://www.flickr.com), peut déposer des ressources, marque-page ou photo personnelle, et leur associer des mots-clefs qu'ils peut ensuite partager avec les autres utilisateurs. Malgré les défauts liés à la faible cohérence des descripteurs (synonymie, polysémie, non explicitation des facettes prises en compte, absence de relation sémantique…) les folksonomies semblent connaitre un réel succès.

Comme le soulignent O. Ertzscheid et G. Gallezot (2006), les folksonomies, qu'ils assimilent à des pratiques d'indexation sociales, tirent leur force de deux phénomènes. D'une part, du faible effort cognitif requis par leur utilisation en comparaisons des classifications épistémiques de la bibliothéconomie et, d'autre part, de la fonction de régulation offerte par la mise en visibilité des mots-clefs déposés par l'ensemble des utilisateurs qui permet d'avoir un effet de feed-back rapide sur leur popularité et leur degré de couverture (Ertzscheid & Gallezot 2006). Cet effet est renforcé par la possibilité d'accéder directement au site identifié par le marque-page (Del.icio.us) ou à la photo indexée (Flickr), ce qui réduit le cout potentiel de l'erreur d'indexation et permet de désambigüiser rapidement certains mots-clefs. Ces propriétés sont d'autant plus essentielles qu'à la différence des annuaires de ressources internet qui sont contrôlés par des indexeurs sélectionnés selon leur réputation, les folksonomies sont totalement ouverte au public inscrit sur les sites.

Les annuaires de ressources internet collaboratifs, comme les folksonomies, ne représentent pas des innovations majeures du point de vue de l'organisation conceptuelle des descripteurs. Le premier dispositif offre des schémas de classifications bien moins sophistiqués que ceux développés par la bibliothéconomie et les folksonomies correspondent à des listes de vocabulaire dont la cohérence apparait comme bien faible eu égard de celles proposées par les professionnels de la documentation. Dans les deux cas, l'innovation majeure se situe dans le processus collaboratif de construction des schémas de classification ou des listes de descripteurs et dans le processus d'indexation associé à cette construction à partir d'un flux de documents primaires très hétérogènes et dont le volume s'accroit très rapidement. Dans la perspective du développement du web socio-sémantique, mais également dans la perspective du développement de nouveaux outils de gestion des bibliothèques numériques, l'hybridation entre des dispositifs professionnels de type schémas de classification, thésaurus et ontologies et ces nouveaux dispositifs de gestion collaborative de l'information numérique via le web, nous semble très prometteuse.

8. Evaluation des Systèmes d'Organisation des Connaissances du point de vue de la RIO

Après ce tour d'horizon des différents Systèmes d'Organisation des Connaissances (SOC) nous sommes maintenant en mesure, dans cette dernière partie, d'esquisser une comparaison qui ne prenne pas seulement en compte leur structure syntaxique et sémantique, mais qui intègre la logique de construction et d'usage de ces dispositifs pour les évaluer du point de vue de la RIO. On se souvient que les principaux critères ayant permis de faire une typologie des processus de recherche ouverte d'information portent sur le caractère intra-documentaire ou extra-documentaire des sources et sur le type d'enquête : enquêtes de localisation-accès, de sélection-décision et de synthèse-conception. Les critères que nous passerons en revue sont les suivants : le degré de formalité de la métasémiotique ou de la métalangue utilisée, le caractère théorique ou a-théorique de la classification sous-jacente, le type de communauté responsable de la conception, la nature des sources utilisées pour fonder la classification, les théories de la signification sous-jacentes aux termes, les modalités de mise à jour et enfin les systèmes de consultation offerts.

Le premier critère qui est classiquement évoqué est celui du degré de formalité de la métasémiotique ou de la métalangue utilisée (tab. 3). Ce critère est corrélé à de nombreux

autres mais il ne véhicule pas, en tant que tel, une fonction d'usage spécifique. Le critère suivant est lié au caractère théorique ou a-théorique des classifications sous-jacentes (tab. 4). L'absence de théorie du domaine peut être vue comme un avantage ou un inconvénient. Quand cette théorie est alignée sur les préoccupations de l'enquêteur, elle offre un avantage considérable. Dans le cas contraire, elle constitue plutôt un handicap. Pour une recherche de sources documentaires associées à des disciplines scientifiques installées le recours à des théories disciplinaire « positivistes » est donc une bonne chose. Quand les sources documentaires sortent du cadre académique, sont innovantes ou interdisciplinaires, des théories ad hoc sont souvent préférables même si elles requièrent un effort d'assimilation en cours de consultation.

Pas de syntaxe pour la gestion combinée des termes	Méta sémiotique explicite	Méta langage formel
• Classification décimale de Dewey • Folksonomie	• CDU • Thésaurus • Ontologie sémiotique • Annuaire collaboratif internet • Carte conceptuelle	• Ontologie formelle

Tableau 3 - Degré de formalisation du langage et des combinaisons syntaxiques

Pas de théorie explicite du domaine	Théorie faisant l'objet d'une controverse explicite	Théorie disciplinaire (souvent positiviste)
• Folksonomie • Annuaire de ressources internet • Cartes conceptuelles (point de vue subjectif revendiqué)	• Ontologies sémiotiques (points de vue complémentaires mais potentiellement conflictuels)	• Classifications universelles (CDD, CDU…) • Thésaurus • Ontologies formelles

Tableau 4 - Présence ou non de justifications à caractère théorique

Conception par des professionnels des bibliothèques ou de la documentation	Conception par des ingénieurs de la connaissance (avec des experts du domaine)	Conception participative régulée (expertise multiple basée p.e. sur la réputation)	Conception participative grand public non contrôlée (sagesse des foules ou effet de mode)
• Classifications universelles • Thésaurus	• Ontologies formelles • Cartes conceptuelles	• Annuaire internet • Ontologies sémiotiques (éventuellement avec la médiation d'un ingénieur de la connaissance)	• Folksonomies

Tableau 5 - Type de communauté impliquée dans la conception des SOC

La meilleure approche serait, selon nous, de pouvoir confronter des théories du domaine concurrentes et explicites comme dans le projet du web socio sémantique. Mais les formes de mise en œuvre à grande échelle de ce principe restent à inventer. Le troisième critère est lié aux communautés responsables de la conception des SOC (tab. 5). On note principalement une distinction entre les deux premières et les deux dernières colonnes du tableau 5, qui correspondent à des formes de participation les usagers-experts à la conception des outils d'indexation et de recherche. Il est possible que ces formes de participation à l'élaboration des SOC facilitent la recherche d'information associée à des enquêtes créatives (synthèse-conception) en stimulant des effets de sérendipité.

Cependant, comme le souligne O. Le Deuff (2007) au sujet des folksonomies, la pertinence et la popularité ne vont pas toujours de pair.

Les critères suivants sont liés à la nature des sources utilisées pour construire les classifications (tab. 6) et aux théories sous-jacentes de la signification des termes (tab. 7). L'avantage des approches ontologiques dans le cadre d'une recherche d'information qui n'est pas uniquement documentaire est de ne pas s'en tenir strictement à un fonds de documents primaires déjà constitué mais de pointer également sur d'autres composantes de la situation d'enquête. Il existe ainsi des ontologies sémiotiques utilisées pour la réalisation d'annuaires de compétences au sein d'une communauté métier sur une base qui n'est pas strictement liée à une documentation préexistante mais qui fait aussi appel à des informations déposées par les intéressés eux-mêmes ou par leurs collègues (Cahier et al. 2005).

L'opposition entre sémantique référentielle et sémantique rhétorico-herméneutique est très discriminante du point de vue des fonctions d'usage. Quand les enquêtes sont liées à des problématiques de localisation et d'accès, la dimension référentielle est essentielle. C'est la raison pour laquelle la CDD continue de jouer un rôle si important pour le classement à l'intérieur des bibliothèques. C'est également la raison pour laquelle les ontologies formelles jouent un rôle sans doute plus important pour faciliter l'interopérabilité entre bases de données que pour assister la recherche ouverte d'information. De même, dans les enquêtes de type sélection-décision essentiellement basées sur des critères objectifs et mesurables, la dimension référentielle peut-être un atout important. Inversement, quand dans ces mêmes enquêtes des critères plus subjectifs et heuristiques entrent en jeu, ce sont les approches rhétorico-herméneutiques du sens qui sont préférables. C'est également le cas dans les enquêtes de type synthèse-conception.

Basé sur un fonds bibliographique	Structuration du domaine empruntant à d'autres sources d'information
• Classifications universelles • Thésaurus • Folksonomies • Annuaire internet	• Cartes conceptuelles • Ontologies sémiotiques • Ontologies formelles

Tableau 6 - Nature des sources utilisées

Sens basé sur une sémantique référentielle et possibilité d'accès selon cette sémantique	Sens rhétorico-herméneutique et accès selon les associations « heuristiques »
• Classification universelle (CDD) • Ontologies formelles	• CDU • Cartes conceptuelles • Annuaire internet • Thésaurus • Ontologie sémiotique • Folksonomie

Tableau 7 - Théorie de la signification sous-jacente.

Les deux derniers critères que nous passerons en revue dans le cadre de cet article ont trait à la fréquence et à la facilité de mise à jour des SOC (tab. 8) et au système de consultation proposé (tab. 9). Trois caractéristiques restreignent la facilité de mise à jour : le respect de la cohérence conceptuelle d'ensemble de la classification, la complexité du formalisme syntaxique et l'existence ou non de systèmes informatisés en permettant l'actualisation « en ligne ». Ces caractéristiques expliquent pour l'essentiel la répartition à l'intérieur du tableau 8. Le projet des ontologies sémiotiques nous semble offrir un

compromis entre cohérence et diversité. En explicitant les points de vue concurrents en présence, elles autorisent une certaine diversité facilitant les mises à jour. Elles sont également plus structurées que les folksonomies dont les mots-clefs relèvent de facettes ou de points de vue potentiellement très hétérogènes.

La variété des systèmes de consultation associés aux SOC (tab. 9) tient à la fois à la sémantique des termes, aux modalités de mise à jour et au caractère plus ou moins largement numérisé des ressources. La vision référentielle du sens et la formalisation extrême des ontologies formelles les destinent plutôt à des programmes informatiques automatisés, le dialogue avec l'utilisateur se faisant par l'intermédiaire des interfaces d'applications spécifiques souvent liées à des bases de données non documentaires. Les classifications universelles et les thésaurus doivent offrir à leurs utilisateurs une « vue dépliée » leur permettant d'interpréter le sens respectif des termes. Même si celui-ci est référentiel dans le cas de la CDD, la sélection s'appuie sur une comparaison entre les intitulés des indices.

Mise à jour rare et complexe (maintien de la cohérence)	Mise à jour systématique régulée par l'évolution du fonds	Mise à jour progressive et négociée	Mise à jour fréquente, facile, immédiate
• Classification universelles • Ontologies formelles	• Thésaurus	• Annuaire internet • Ontologie sémiotique • Cartes conceptuelles	• Folksonomie

Tableau 8 - Fréquence et facilité de mise à jour

De type formulaire ou lié à l'interopérabilité entre programmes	Vue dépliée permettant l'interprétation des descripteurs	Systèmes de navigation hypertexte ergonomiques
• Ontologies formelles	• Classifications universelles • Thésaurus	• Ontologies sémiotiques • Cartes conceptuelles • Folksonomie • Annuaire internet

Tableau 9 - Systèmes de consultation

Dans le cas des thésaurus, la nécessité de pouvoir comparer les termes est encore plus grande du fait de l'approche résolument différentielle de la signification (le sens d'un descripteur est précisé par celui des termes équivalents ou interdits). Dans les systèmes qui combinent des points de vue multiples, l'interface d'exploration du réseau de thèmes est essentielle pour préciser l'interrogation. La possibilité d'accéder directement à la ressource numérique est également précieuse pour être en mesure de désambiguïser les descripteurs. Le principe des logiciels associés aux folksonomies qui offrent une rétroaction directe sur les d'indexation effectuées par les usagers (nombre de ressources indexées, termes préférés dans la base, associations entre tag, etc.) semble très efficace et est surement amené à se généraliser dans d'autres dispositifs.

En guise de conclusion sommaire, il nous semble important d'insister sur le fait qu'une bonne compréhension des différents systèmes d'organisation des connaissances existants doit permettre d'imaginer, pour les prochaines années, de nouvelles solutions hybrides d'aide à la recherche ouverte d'information dans le contexte de la numérisation accélérée des ressources documentaires et de la participation de plus en plus aisée des experts et des utilisateurs. Ces solutions doivent être développées en pensant la complémentarité

avec l'usage des moteurs de recherche, une de nos préoccupations majeure dans le cadre des travaux que nous conduisons sur la recherche ouverte d'information. Selon nous, grâce à ces développements, les potentialités de l'indexation interprétative des ressources documentaires sont encore loin d'être épuisées et doivent constituer une voie de recherche majeure à l'intersection des sciences de l'information, des sciences de la communication, de l'ingénierie des connaissances et du CSCW.

9. Bibliographie

Badir, S. (2000), *Hjelmslev*, Paris, Belles-Lettres.

Bénel, A. (2003). *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications à l'archéologie*, Thèse de doctorat en informatique, INSA de Lyon, décembre 2003.

Berners-Lee, T., Hendler J., Lassila O. (2001) The Semantic Web, *Scientific American*, http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

Cahier, J.P., Zacklad, M., Monceaux, A. (2004) Une application du Web socio-sémantique à la définition d'un annuaire métier en ingénierie, *Actes des journées Ingénierie des Connaissances 2004*, Mai 2004, Lyon.

Cahier, J.-P. (2005), *Ontologies sémiotique pour le Web socio sémantique: étude de la gestion coopérative des connaissances avec des cartes hypertopiques*, Thèse en informatique, Université de Technologie de Troyes, déc. 2005.

Caussanel, J., Cahier J.-P., Zacklad M., Charlet J. (2002), Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique ? *Conférence Ingénierie des Connaissances IC2002*, Rouen.

Charlet, J. (2002). *L'Ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*. Habilitation à diriger des recherches, Université Paris 6.

Dewey, J. (1938). *Logique, la théorie de l'enquête*, trad. Gerard Deledalle, col. L'interrogation philosophique, PUF, 1938 (re-éd.1993).

Ertzscheid, O., Gallezot, G. (2003), Chercher faux et trouver juste, *Actes du X° Colloque bilatéral franco-roumain, CIFSIC Université de Bucarest*, 28 juin – 3 juillet 2003, Bucarest SFSIC - Université de Bucarest (Ed.) en ligne sur http://archivesic.ccsd.cnrs.fr/

Ertzscheid, O., Gallezot, G. (2006), Etude exploratoire des pratiques d'indexation sociale comme une renégociation des espaces documentaires. Vers un nouveau big bang documentaire ? in Chartron G. & Broudoux, E. (Eds). *Document numérique et société*, ADBS Éditions, 2006 en ligne sur http://archivesic.ccsd.cnrs.fr/

Guarino, N. (1998), Guarino, N., "Formal Ontology and Information Systems," in *Formal Ontology in Information Systems*, N. Guarino, Ed. Amsterdam, Netherlands: IOS Press, 1998

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199-220.

Hodge, G. (2000) Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files, *The Digital Library Federation Council on Library and Information Resources* http://www.clir.org/pubs/abstract/pub91abst.html

Hudon, M. (2001), Structuration du savoir et organisation des collections dans les répertoires du Web, *BBF*, Paris, T. 46, n°1, 2001, pp. 57-62.

ISO, (1986). Organisation internationale de normalisation. Documentation - Principes directeurs pour l'établissement et le développement de thésaurus monolingues : norme internationale ISO 2788. [Géneve]: ISO, 1986.

Lacour, P. (2004). L'oubli de la sémantique dans le programme cognitiviste : réflexions sur l'oeuvre de François Rastier. *Texto !* décembre 2004 [en ligne]. Disponible sur : http://www.revue-texto.net/Inedits/Lacour_LOubli.html

Lando, P. (2006), Conception et développement d'applications informatiques utilisant des ontologies : application aux EIAH, *actes des Premières rencontres jeunes chercheurs en EIAH* (RJC-EIAH'06), Évry, France, 11-12 mai 2006.

Le Deuff, O. (2007). Folksonomies : Les usagers indexent le web, *BBF*, 2006, n° 4, p. 66-70 [en ligne] <http://bbf.enssib.fr>

Christophe Lejeune (2004). *_Sociologie d'un annuaire de sites Internet*. Les sciences documentaires saisies par l'informatique libre, Thèse de Doctorat, Université de Liège, Liège.

Lejeune, C. (2006), Ce que l'annuaire fait à internet : Sociologie des épreuves documentaires; *Cahiers de la documentation*, vol. 60, no3, pp. 12-22.

Maniez, J. (1999) Des classifications aux thésaurus : du bon usage des facettes. *Documentaliste-Sciences de l'information*, juillet-octobre 1999, vol. 36, n° 4-5, p. 249-262.

Toms, E. 2000. Serendipitous Information Retrieval. In *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, Zurich, Switzerland: European Research Consortium for Informatics and Mathematics. http://citeseer.csail.mit.edu/toms00serendipitous.html

Turner, W. (2007), Eléments pour une socio-informatique, in Reber, B., Brossaud, C., Humanités numériques. *Nouvelles technologies cognitives et concepts des sciences sociales*, Hermes Publishing, Londres-Paris (à paraître).

Rastier, F. (1997) Herméneutique matérielle et sémantique des textes, in Salanskis et al. (éds.), *Herméneutique : textes, sciences*, Paris, PUF.

Rastier, F. (2003). De la signification au sens. Pour une sémiotique sans ontologie. *Texto!* juin-sept. 2003 [en ligne]. Disponible sur : <http://www.revue-texto.net/Inedits/Rastier/ Rastier_Semiotique-ontologie.html>

Saadani, L., Bertrand-Gastaldy, S., (2000). Cartes conceptuelles et thésaurus : essai de comparaison entre deux modèles de représentation issus de différentes traditions disciplinaires. In : Kublik, A., ed. *Les dimensions d'une science de l'information globale*, ACSI-CAIS http://www.cais-acsi.ca/2007call_fr.htm

Zacklad, M. (2005a), Introduction aux ontologies sémiotiques dans le Web Socio Sémantique. In: Jaulent, M.-C. *16èmes journées francophones d'Ingénierie des Connaissances*, 30-03 Avril 2005, Nice. Grenoble: PUG, 12 p.

Zacklad, M. (2005b), Ergonomie de la Recherche d'Information (RI) sur le Web : quel modèle de l'activité ? Présentation à l'école d'été du GDR TICS, septembre, 2005, http://gdrtics.u-paris10.fr/pdf/ecoles/sept2005/transparents_Zacklad.ppt

Zacklad, M. (2007a). Processus de documentarisation dans les Documents pour l'Action (DopA). Babel - edit -, *Le numérique : impact sur le cycle de vie du document*. ENSSIB - février 2007 [en ligne] http://babel.enssib.fr/document.php?id=727

Zacklad, M. (2007b). Annotation : attention, association, contribution. In P. Salembier et M. Zacklad eds, *Annotations dans les Documents pour l'Action*. Lavoisier, Paris : 29-46.

Zacklad, M. (2007c), Une théorisation communicationnelle et documentaire des TIC, in Reber, B., Brossaud, C., *Humanités numériques. Nouvelles technologies cognitives et concepts des sciences sociales*, Hermes Publishing, Londres-Paris (à paraître).

Zacklad, M., Cahier, J.-P., Bénel, A., Zaher, H., Lejeune, C., Zhou, C., (2007), Hypertopic : une métasémiotique et un protocole pour le Web socio-sémantique, in *Actes des 18èmes journées francophones d'Ingénierie des Connaissances*, 2-6 juillet 2007, Grenoble (à paraitre).

Zaher, H., Cahier, J.-P., Zacklad, M., (2006), Information Retrieval and E-Service : Towards Open Information Retrieval, In Proceedings of *International Conference on Service Systems and Service Management IC SSSM'2006*, IEEE, ISBN 1-4244-0451-7, France, october 2006, pp. 41-46.

[1] On peut penser aux plateformes commerciales de type Amazon mais également aux développements actuels de « moteurs de recherche » qui combinent différentes formes d'indexation.

[2] Près d'une quinzaine d'applications dans différents domaines ont été réalisées à ce jour, mais d'avantage dans un contexte intranet qu'internet (Zacklad & al. 2007).

[3] Dans une base documentaire décrivant des projets de Recherche & Développement, les projets sont considérés comme des items caractérisés par des ressources documentaires (des fichiers), des attributs standards (dates, participants…) et des attributs heuristiques, c'est-à-dire des cartes de thèmes représentants les ontologies sémiotiques.

[4] Pierce est également un inspirateur important pour certain d'entre nous (Cahier 2005).

[5] Moins présent à notre connaissance chez Rastier.