

De la recherche d'information à une recherche ouverte d'information

L'Hédi Zaher, Jean-Pierre Cahier et Manuel Zacklad

*Tech-CICO - Institut Charles Delaunay
Université de Technologie de Troyes / CNRS (FRE 2848)
12 rue Marie Curie, 10010, Troyes Cedex, France
zaher@utt.fr*

Résumé : Nous proposons d'élargir la notion de Recherche d'informations (RI) à celle de recherche ouverte d'information (ROI), dans un concept basé sur la notion « d'entité », permettant l'accès à des formes de recherche riches, mieux reliées aux usages métier spécifiques. Pour cela nous montrons sur un exemple (choisi dans le domaine des places de marché électroniques) comment ces formes riches peuvent s'appliquer. Le modèle de représentation de connaissances Hypertopic, proposé dans le cadre du Web socio-sémantique, est ici mis à profit pour organiser la « mémoire externe de construction » sur laquelle prend appui la pratique de cette ROI. Nous explorons ainsi une voie vers des outils de ROI mieux adaptés à l'activité, permettant à l'utilisateur d'investir ses connaissances personnelles et d'être plus actif dans la construction de la signification en rapport avec des usages précis.

Mots-clés : Ingénierie des connaissances, recherche d'informations, Place de marché virtuelle, Web socio sémantique.

INTRODUCTION

Dans un contexte professionnel ou non, de nombreuses activités de résolution de problème (ou plus généralement d'enquête au sens de Dewey [DEW 38]) s'appuient aujourd'hui sur des ressources documentaires et informationnelles plus ou moins formalisées, accessibles dans des environnements Internet ou intranet.

Le premier mode d'accès aux ressources du Web est sans conteste constitué par les moteurs de recherche qui s'appuient sur une indexation de chaque page Web rencontrée et par la mise en correspondance de mots-clés exprimés dans une requête avec ceux contenus dans l'index du document.

L'autre modalité d'accès aux ressources s'applique quand les caractéristiques de celles-ci ont fait l'objet d'une « formalisation » poussée de leurs caractéristiques. Cette formalisation permet leur stockage dans un entrepôt de données (e.g. une base de données ou sous forme de fichiers XML, etc) et l'accès à celui-ci par le biais de requêtes logiques. Dans l'optique, notamment, d'une interopérabilité entre ces bases de données, le programme du Web

sémantique prolonge la vision d'un accès généralisé à des ressources formalisées sur une partie du Web dont le Web sémantique est une extension.

Enfin, les annuaires de ressources proposent un classement de ces ressources, souvent réalisé par une communauté d'indexeurs, qui met en relief leur pertinence en regard de différents registres de préoccupations. L'organisation de ces annuaires s'appuie sur ce que nous appelons, dans le contexte du Web Socio Sémantique, des ontologies sémiotiques ([ZAC 05], [CAH 05]).

Pour comprendre et définir les infrastructures informationnelles qui permettent aux usagers des environnements « webisés » de progresser dans leurs enquêtes, il est nécessaire de s'interroger sur les méthodes leur permettant de découvrir les faits pertinents (au sens de faits relevant de l'enquête), souvent associés de manière plus ou moins directe à une ou plusieurs ressources numérisées.

A première vue, cette problématique semblerait relever du domaine de la Recherche d'Information (RI). Mais un premier examen des principaux travaux réalisés dans ce domaine nous a conduits au constat suivant : la manière dont les questions de recherche scientifique sont posées ne permet pas d'aborder

globalement la question des méthodes de recherche adaptées à la conduite d'une enquête basée sur des ressources numérisées. En effet, la question principale qui semble être abordée dans ce courant est celle de la « récupération d'un document » à partir de sa description plus ou moins stricte, et non celle de la résolution d'un problème (cf. § 1.2.), à partir d'une multiplicité de « documents » disponibles permettant à chacun d'apporter un éclairage particulier sur le problème posé.

En d'autres termes, comme nous l'expliciterons plus bas, alors que de nombreux travaux de RI abordent surtout la question de l'accès à des documents dont la localisation n'est pas connue mais dont l'existence ou la pertinence est acquise (ou de l'accès à des fragments de textes dans un document considéré comme pertinent), notre problématique, relevant de la Recherche Ouverte d'Information (ROI) [ZAC 05] [ZAH 06], vise à permettre la *découverte et la définition originale d'une entité complexe* à partir d'un ensemble de ressources numérisées dont le contour n'est pas défini *a priori* et dont la définition s'appuie à la fois sur une variété de types de contenus (documents textuels ou imagés, bases de données...) et sur une variété de systèmes d'accès à ces types de contenus (moteur de recherche, système de requête, ontologies formelles ou sémiotiques...).

Dans cet article, nous commencerons à esquisser le modèle conceptuel de la ROI et nous évoquerons quelques pistes permettant la conception d'une infrastructure assistant les chercheurs d'information dans leurs activités. Nous pensons que les hypothèses du Web socio sémantique et le projet de standard HyperTopic [CAH 05] [CAH 04] permettent d'en guider utilement la conception. C'est ce que nous tenterons de montrer en nous appuyant sur un exemple mis en œuvre avec la plateforme Agoræ.

1. Recherche d'information

1.1. Champs d'étude de la RI

L'acception que nous adoptons de la notion de recherche d'information (RI) se distingue de l'usage du terme privilégiant l'étude de la représentation et des inscriptions formelles des informations, de l'organisation des entrepôts de ressources numériques, des formalismes d'interrogation et de requête ainsi que les algorithmes de recherche automatique dans ces entrepôts (bases de données, bases documentaires, voire le Web, etc.) dont quelques représentants seraient [MEL 99] [PRI 00].

Notre conception de la RI comme une activité cérébrale d'utilisateurs humains en interaction avec des artefacts TIC se rapproche du sens donné par [ROU 98] et s'apparente à une approche « cognitive » de la RI. Ainsi, là RI serait l'ensemble « des activités, des processus, des représentations, des transformations et des inférences cognitives qu'élabore et met en œuvre un individu pour trouver des informations utiles et pertinentes par rapport à une tâche qu'il se donne »

[ZAH 05].

Cette recherche s'effectue alors en ayant recours à des artefacts, et ce de manière plus ou moins interactive. Ces artefacts peuvent se différencier par rapport à leurs supports (Documents papiers, Hypertextes numériques, système informatique à base de langage de requêtes) et leurs conceptions différentes en matière d'usages (pages jaunes, annuaires, cartographies, moteurs de recherche plein texte, etc.). Éventuellement, ils peuvent être conçus pour la recherche dans un corpus fermé relatif à un métier ou un domaine donné (e.g. catalogue électronique des pièces de rechange d'un constructeur automobile, catalogues papier des produits d'un vendeur par correspondance, moteur de recherche de films et de séances d'une chaîne de cinémas, etc.).

La RI est alors non seulement l'étude des aspects mentaux d'une tâche de recherche (comme dans [ZAH 05]) mais aussi l'étude des systèmes d'information qui instrumentent au sens de supporter, outiller ces tâches de recherche, et l'étude de leurs interfaces homme-machine et des représentations qui leurs sont sous-jacentes.

1.2. Quelques approches cognitives de la RI

Une première revue des approches cognitives de la RI permet de distinguer – non exhaustivement et non exclusivement – trois approches.

Selon la première approche, **(1) la RI est un processus non linéaire**, et est une activité **exploratoire** de type « cueillette de baies » (berry picking). Elle met l'accent la flexibilité du besoin d'information, évoluant tout au long de la recherche [BAT 89].

Quant à la seconde approche, elle considère que dans une tâche de RI, un sujet met en œuvre une construction progressive de buts, de méthodes pour les atteindre et une représentation (de contenu) [ROU 98]. Ces auteurs proposent un modèle de la RI qui l'analyse en tant que **(2) activité de résolution de problème**.

Selon ce modèle, la RI correspond à un cycle de trois processus :

- le processus d'évaluation a pour fonction d'identifier les informations qui manquent pour effectuer la tâche ; le processus de sélection de l'information pertinente correspond à une prise de décision en référence aux critères élaborés dans la phase évaluative ;
- le processus de traitement de l'information recherchée met en jeu principalement une activité de lecture ;
- à la compréhension de texte viennent s'ajouter d'autres tâches (décision quand à la pertinence du contenu, etc.). Les cycles se succèdent jusqu'à l'atteinte du but. Des mécanismes de gestion cognitive de l'activité accompagnent ces processus (planification,

contrôle, régulation).

La troisième approche tente d'intégrer les contraintes liées au système dans son analyse de la tâche de RI dans les bases documentaires [KOL 98]. Elle la traite comme un **(3) problème de conception**. La RI consisterait en un changement de représentation du problème prenant en compte les contraintes qui se traduit par le passage d'un but flou (besoin d'information) à un but opérationnel (plan d'action sur le système).

L'activité du sujet consiste à élaborer des plans déclaratifs (explicitation du but) et procéduraux (mise en œuvre de l'atteinte du but). Les contraintes sont liées au but (quel domaine ?...) mais également au système d'information (système d'interrogation et fonctionnalités). [KOL 98] considère qu'un système hypertexte propose une représentation du domaine à l'utilisateur lui permettant d'élaborer des plans déclaratifs contribuant à la focalisation progressive.

1.3. Dispositifs de RI par requêtes vs dispositifs par navigation

Nous nous contenterons ici de rappeler quelques éléments d'état de l'art qui nous semblent significatifs et qui nous seront utiles, à travers un certain nombre de travaux qui se sont intéressés spécifiquement aux deux grandes classes d'environnements de RI : Les dispositifs par requête et les dispositifs par exploration / navigation.

Dans un dispositif par requête, comme un moteur de recherche, la requête est comparée avec les représentations des documents dans la collection [DUN 98]. Le système, en réponse à la requête, produit une liste de documents (ou sites...) qui correspond aux documents de la collection compatibles avec la requête (contenant le mot ou des combinaisons de mots, etc.). Pour être utilisés pleinement ces outils impliquent des connaissances sur la façon dont le système opère.

Quand une requête n'apporte pas de résultat intéressant, l'utilisateur doit diagnostiquer pourquoi la réponse du système est inattendue et résoudre le problème en proposant une nouvelle requête. Ce système permet de n'avoir accès qu'à une seule partie des ressources (des documents de la collection), ceux correspondant au résultat de la requête. Par conséquent, l'utilisateur doit évaluer l'écart possible entre cette « fenêtre » sur la collection que renvoie le système et son besoin d'information [CHI 97].

C'est pourquoi un des enjeux pour ce type d'outils est de proposer des feedbacks (retours) pertinents permettant d'apprendre le fonctionnement de l'outil et d'interpréter cette « fenêtre ». Ainsi, certains travaux expérimentent des systèmes aidant l'utilisateur dans la formulation et reformulation de la requête, que ce soit par la sélection de champs préalables [MAN 97] ou par la réalisation d'inférences à partir des requêtes précédentes. D'autres travaux s'intéressent à l'effet des techniques de visualisation des réponses [CIA05].

Dans les dispositifs par exploration / navigation (ou encore hypertexte) la collection de documents est accessible en naviguant de lien en lien. Les liens peuvent se situer entre les documents (système hypertexte classique) et/ou dans une structure décrivant l'organisation des informations (carte conceptuelle, index, etc.). La variété de ces outils est assez large, allant d'index peu structurés à des index fortement structurés comme divers types de systèmes de navigation. La structuration des informations guide l'utilisateur tout le long de sa recherche (choix d'un lien ou d'un autre). L'utilisation d'un tel outil ne nécessite pas de connaissances sur le fonctionnement informatique.

En revanche, la navigation conduit souvent à une désorientation des utilisateurs caractérisée par la consultation de nombreux documents et une baisse des performances dans le temps [MAR 95] [McD 98]. L'utilisateur n'obtient pas une vue d'ensemble de la structure des informations et éprouve des difficultés à établir un but et sa planification [KIM 95].

Nous pouvons à cette étape retenir que, pour être utilisés pleinement, les systèmes par requête requièrent de l'utilisateur une bonne connaissance des modalités de fonctionnement du système et en particulier de la manière de formuler les requêtes pour parvenir à des résultats pertinents. De leur côté, les systèmes par navigation sont susceptibles d'entraîner une forme de désorientation des utilisateurs.

Pour y pallier, il est bénéfique de proposer des aides à la navigation sous la forme d'une structuration des informations utiles à la navigation. C'est à ce niveau de la structuration que vont intervenir les schémas de classification que nous chercherons à évaluer en leur faisant jouer un rôle d'aide à la navigation, sachant qu'ainsi ils sont susceptibles de fournir un guidage palliant aux risques de désorientation cognitive des utilisateurs. Mais on notera que lorsqu'un schéma de classification est mémorisé et partagé par les membres d'une organisation, il peut tout aussi bien aider ces acteurs qui s'appuient sur la mémorisation de ces schémas lorsqu'ils utilisent un système par requête, en les aidant à trouver les mots clés pertinents.

1.4. Enseignements pour la RI ouverte

De cette première revue non exhaustive des approches cognitives par lesquelles la question de la RI est abordée, nous retenons trois enseignements qui nous semblent essentiels pour la question de systèmes informatiques de support à la RI : (1) la RI est une activité progressive dans laquelle les représentations internes se construisent au fil de l'eau de la recherche ; (2) la RI est une activité cyclique qui enrichit progressivement la représentation du but de l'utilisateur [BAT 89] ; (3) la RI est une activité impliquant des changements de systèmes de représentation impliquant notamment des plans déclaratifs du domaine.

D'une part cette analyse nous permet de distinguer

que selon les connaissances préalables de l'utilisateur, on peut considérer que l'on a affaire à :

- Une RI limitée lorsque l'utilisateur connaît l'entité singulière recherchée et souhaite essentiellement la « localiser ». C'est notamment le cas lorsque la recherche est considérée comme s'effectuant selon un arbre de décision, sans que des résultats intermédiaires de la recherche puissent venir remettre en cause ou affiner le but (absence d'imprévu et de sérendipité¹). Notons que l'on retrouve ici la distinction profondément développée sur un exemple « d'organisation de soirée » par [HAT 02], entre le type de décisions nécessaire pour choisir un film (quel film aller voir dans une liste finie) et pour organiser une fête (nécessité de définir d'abord ce qu'on entend par ce terme, puis les caractéristiques, les choix sémantiques ayant une influence sur la suite de la recherche, qui comporte beaucoup plus d'imprévu).

- Une RI ouverte, lorsque l'utilisateur ne sait pas quelle entité singulière peut répondre à son besoin – (voire ne sait pas si elle existe) et va donc travailler à la découverte de cette entité, y compris en faisant valoir sa grille d'interprétation sémantique. Il peut alors être guidé par un artefact (l'espace de découverte) basé sur un modèle de connaissances organisant dans cet espace les différents types de « prises cognitives » disponibles sur l'entité.

Aussi, une première comparaison de l'utilisation de différents dispositifs de recherche d'information – dont un moteur de recherche et une carte de thèmes – par des sujets non-experts du domaine essayant de répondre à des consignes à partir d'un corpus documentaire fermé a montré que même si les utilisateurs d'un moteur de recherche répondent davantage aux consignes, le nombre de réponses correctes penche légèrement en faveur de la carte de thèmes (sans que le nombre des sujets de l'expérience ne permette d'en tirer des conclusions statistiquement valides) [ZAH 05]. L'idée qui s'en dégage est que la qualité de réponse – *e.g.* la certitude mesurable par le taux de réponses correctes par réponses incorrectes – serait meilleure avec une carte de thèmes.

Une hypothèse assez crédible qui reste à vérifier serait que la supériorité (en termes d'efficacité) des différents dispositifs (modes d'accès) pourrait varier en fonction du corpus (taille, fermé ou ouvert, mono ou multi domaine, degré de formalisation) aussi bien qu'en fonction du sujet (maîtrise du domaine, expertise de l'usage des dispositifs de RI...).

D'autre part l'espace problème qui va s'appuyer sur l'artefact gagne à être lié à l'activité de l'utilisateur, à laquelle correspond une classe de

situations problèmes. L'espace de découverte n'a donc pas vocation à prendre en compte « l'ensemble des objets reconnus comme existants » dans une visée universalisante (comme le proposent certaines ontologies de domaine) mais au contraire à ne sélectionner que certains objets utiles à l'usage considéré et à mettre alors davantage en valeur ces objets dans la recherche, notamment en utilisant des attributs « heuristiques ». Dans la RI ouverte, « l'espace problème » personnalisé est plus restreint et pertinent que « l'espace des possibles » résultant de la combinatoire des attributs objectifs.

2. Les systèmes d'aide à la recherche d'information : modèles sous-jacents

Nous présentons ici trois approches qui coexistent et qui mènent à des conceptions visiblement différenciées des systèmes informatisés de RI².

Une **approche orientée données** qui part d'une hypothèse forte selon laquelle tout objet « informatique ou informatisable » peut être décrit et représenté sinon « strictement et exhaustivement » du moins suffisamment par un certain nombre de qualificatifs très souvent (suffisamment) prévisibles, dénombrables, finis et figés pour être listés et précisés très tôt³. Ces qualificatifs que nous appelons plus tard des *attributs standards*, prennent des valeurs dans des types de données courants.

Si une telle hypothèse reste implicite assez fréquemment, elle est sous-jacente à bon nombre de systèmes informatiques en général, et de systèmes à base de connaissances et de RI en particulier.

La classification à facettes⁴ et les systèmes de facettes fixes adoptés en bibliothéconomie ne sont pas étrangères à cette approche, qu'elles illustrent bien. Pour des raisons épistémologiques, historiques et pratiques liées au développement de l'informatique, cette approche rencontre un certain succès.

Le statut des attributs standards leur permet d'être adoptés fréquemment pour exprimer des données ou des métadonnées, et résiste bien aux phénomènes de mode (programmation structurée, orienté objet, système multi-agents, etc.).

L'énumération des couples de qualificatifs / types⁵ est un fondement usuel pour définir une classe

¹ « La sérendipité est le don ou la faculté de trouver quelque chose d'imprévu et d'utile en cherchant autre chose, ou encore, l'art de trouver ce qu'on ne cherche pas » (Wikipédia)..

² Dans le reste de cette contribution, nous nous traiterons que les aspects liés à l'informatisation -- au sens d'outiller par l'informatique, et non forcément d'automatiser -- des tâches de RI bien que nous illustrerons par quelques exemples et métaphores non « assisté par ordinateur ».

³ Par rapport au cycle de développement d'un projet informatique en génie logiciel.

⁴ Introduit dans la théorie de la classification en 1930 par S.R. Ranganathan dans « Colon Classification ». Ranganathan est un mathématicien, documenta-liste et «classificationniste» indien.

⁵ Sans oublier l'énumération des traitements possibles en orienté objet, et des relations entre classes et instances d'objet et règles d'inférence dans le cas des ontologies dites « formelles » ou « lourdes ».

d'objets. Les différentes instances d'une même classe sont obtenues et se distinguent mutuellement par l'affectation de valeurs spécifiques à ces attributs standards.

Cette approche transcende généralement aussi bien les bases de données, la programmation et la conception orientées objet, les systèmes multi-agents et les systèmes à bases de connaissance ou le Web sémantique [CHR 03] [BAG 03]. Les avatars les plus courants en RI sont les moteurs de recherches « non plein texte »⁶ mais aussi des places de marché numériques, ou pour retrouver une ressource, on sélectionne par une combinaison logique d'attributs standards tels que l'auteur, une date, un fabricant, etc (recherche dite « paramétrique »).

Une **approche orientée documents** [APA 04] [BEN 01] postulant que la source et le vecteur le plus important et le plus riche de la connaissance sont – et restent – le document (*cf.* le courant scientifique « document numérique »).

Étant un support peu (voire non-) structuré, des agents logiciels, des experts du domaine traité au sein du document et des experts documentalistes construisent des structures et des index *ad hoc* et externes décrivant ces documents - structures et index dont l'existence et la légitimité sont intimement liées aux documents en question. Ces descriptions traitent aussi bien les contenus des documents (à différents niveaux : lexique, sémantiques et sens, connaissances opératoires, etc.) que les contenants-documents (sous forme d'attributs standards du contenant-document mais aussi de plus en plus la relation entre documents, citations, références, etc.).

Cette tendance est liée non seulement à la nature des tâches métiers faisant appel à des documents, mais aussi au malaise consécutif au traitement automatique des documents sans méta-données ni contextualisation. Malaise dû à des structures, des sens et des descriptions des modes opératoires non explicités sinon absents dans le document, en tout cas non appréhendables par des agents logiciels « ordinaires ».

Toute description est alors facilitatrice sans être substitutive ou suffisante en soi, ce qui est l'une des distinctions par rapport à la première approche « orientée données ».

L'approche orienté document est prédominante dans le cas des bibliothèques numériques, des CMS (Content Management Systems) et du Web, et ses avatars les plus répandus en RI sont les moteurs de recherche plein texte de type Google (qui marie une recherche plein texte lexicale et pointant sur des ressources documentaires Web à une recherche sur des index construits automatiquement et des index construits par des relecteurs humains).

Enfin, une **approche orientée cartographie** amène à faire jouer un rôle important à des artefacts utilisés dans un mode de navigation : classification [PIE 87] [WYN 92], annuaires, cartographies [STU 04], Hypertexte, ontologies sémiotiques [CAH 05] [ZAC 05a] [ZAC 05b]. Dans ce dernier cas, les attributs heuristiques ou « thèmes » reflètent un point de vue qui est un *point de vue sur l'entité*, par exemple un point de vue métier ou une vision d'expert pour considérer cette entité. Le modèle Hypertopic est précisément conçu pour permettre de bâtir les relations nécessaires entre points de vue et thèmes et les constituer en carte de thèmes multi-points de vue ou ontologies sémiotiques. [ZAC 05b] montre que l'aide apportée par une ontologie sémiotique dans la recherche et la sélection d'information comme dans la navigation sémantique, est toujours relative à une classe de situations problèmes.

Pour l'approche orientée cartographie, les avatars les plus connus en ingénierie des connaissances et en recherche d'information sont les graphes conceptuels, les réseaux sémantiques, les taxonomies et les ontologies « légères »⁷ et les cartographies de thèmes.

L'exemple le plus parlant est celui des Topic Maps, dérivés de Hytime (longtemps utilisés en sciences de la documentation) et normalisés par l'ISO⁸ avant d'avoir une déclinaison XML appelée XTM⁹. Les Topic Maps privilégient la représentation des associations abstraites liant les différents objets décrits (appelés sujets).

En recherche d'information, les Topic Maps débouchent sur des apports en matière de conception d'index et des cartes permettant la recherche par navigation et exploration ainsi que des modes de recherche dits augmentés ou sémantiques. Les Topic Maps ont inspiré nos travaux menant à HyperTopic.

Aux origines de ce troisième mode d'accès « cartographique » aux entités, nous faisons l'hypothèse fondamentale qu'un certain sens est véhiculé par la manière d'associer entre eux un ensemble d'indices (attributs heuristiques). L'importance des associations, éventuellement de leurs types, des distances, de la disposition, des similarités, des proximités, de la disposition au sein du rendu graphique à l'utilisateur, sont aussi importantes que les termes qualifiant les indices.

Bien que ces trois modes correspondent à des courants et à des paradigmes bien distincts dans la littérature scientifique (sans être vraiment

⁶ Souvent appelés paradoxalement recherche avancée, ou encore recherche multi-critères.

⁷ À distinguer des ontologies dites lourdes, qui associent à une taxonomie, des classes (et de l'héritage), un réseau de relations et un ensemble de règles d'inférences.

⁸ ISO / IEC FCD 13250 : 1999, norme publiée début 2000.

⁹ XML Topic Maps (XTM) 1.0, spécification non normative de TopicMaps.org.

contradictoires), l'étude empirique (§4) montre que les systèmes les plus répondus et rencontrant le plus de succès sont assez hybrides et marient avec plus ou moins de nuances ces trois modes. Il s'agit d'une tendance qui s'accroît.

3. Recherche ouverte ou fermée d'information dans le contexte du Web Socio Sémantique

La RI s'inscrit toujours dans le contexte d'une activité finalisée. Dans l'exemple que nous utiliserons plus bas, l'objectif du chercheur d'information est de renouveler son téléphone portable. Dans de très nombreux cas, cette situation génère une activité d'enquête qui implique à la fois des considérations liées à l'équipement (e.g. « quel appareil téléphonique serait compatible avec mon usage et mon budget ? »), et aux services associés offerts par les opérateurs (pour bénéficier des réductions et d'une partie des fonctionnalités de l'équipement).

Dans cette situation, la recherche fermée correspond au cas où le chercheur d'information sait déjà quel appareil et quelle offre associée répondent à son besoin. Il souhaite accéder à cette offre et préciser son besoin sans que celui-ci ne soit *a priori* susceptible d'être remis en cause. Cette situation ne nécessite pas de faire appel à des capacités de sérendipité. Au contraire, toute distraction dans cette recherche focalisée sera considérée comme malvenue.

Au contraire, dans la *ROI*, l'utilisateur ne sait pas *a priori* quelle entité complexe est en mesure de répondre à son besoin, voire ne sait pas si cette entité existe. Son objectif est donc tout à la fois de découvrir [BAT 89] [KOL 98] [ZAC 05a][ZAH 05] et de construire cette entité complexe. Pour reprendre les termes de Chateauraynaud [CHT05] dans le contexte des enquêtes sociologiques, son usage des ressources de la toile vise à lui permettre de « confirmer ou infirmer des thèses, authentifier des sources, recouper ou évaluer des informations ».

Pour découvrir/construire l'entité complexe, le chercheur d'information recourt alternativement aux différents modes d'accès à l'information que nous avons évoqués plus haut. Conformément aux hypothèses du Web Socio Sémantique, nous considérons que ces différents *modes d'accès* à l'entité correspondent eux-mêmes à divers *modes de définitions de l'entité* plus ou moins contraints, formalisés, contextualisés, consensuels.

Ces modes de définition de l'entité renvoient eux-mêmes aux situations sociales dans lesquelles celle-ci est produite et/ou échangée c'est-à-dire au plus ou moins grand degré de standardisation des transactions communicationnelles par lesquelles elle est mise en

forme sur un plan symbolique (cf. la discussion sur l'opposition entre formalité sémiotique et machinale dans [ZAC 05b]).

Dans un contexte de vente de produit, par exemple, on pourra ainsi opposer des critères tangibles liés à la mesure physique des caractéristiques de l'équipement aux critères plus immatériels liés à ses modalités d'appropriation et d'usage. Ces critères sont eux-mêmes mobilisés dans les négociations plus ou moins conflictuelles entre différentes séries d'acteurs (fabriquant, distributeur, détaillant, client, etc.) et de métiers ou rôles (ingénieur, marketeur, vendeur, acheteur, technicien du SAV, etc.).

Pour identifier les différents modes de définition et d'accès aux entités complexes, nous utilisons le langage de représentation HyperTopic que nous avons proposé dans le contexte du Web Socio-Sémantique (figure 1, 2) et que nous allons maintenant commenter.

3.1. Attributs standards, données formatées, ontologies formelles

Le premier mode de définition de l'entité utilise un certain nombre de qualificatifs dénombrables, finis et figés. Ils sont suffisamment stables pour permettre la conception de structures de données pérennes sur lesquelles pourront être mis en place des systèmes de requête et d'inférence. Ces qualificatifs que nous appelons des *attributs standards*, prennent des valeurs dans des types de données courants.

Ce mode de description des entités est le plus utilisé en informatique dans des contextes aussi divers que les bases de données, la programmation et la conception orientées objet, les systèmes multi-agents et le web sémantique. Dans le domaine de la RI cette modalité renvoie aux systèmes de requêtes utilisant une combinaison logique d'attributs standards tels que l'auteur, une date, un fabricant, etc. dans le cadre de recherches dites « paramétriques ».

3.2. Ressources documentaires

Le second mode de définition de l'entité est le document basé sur une forme d'expression linguistique ou imagée. Il peut s'agir d'un texte plus ou moins narratif, d'une photo, d'un extrait de musique, d'une séquence filmée, ou de toute combinaison possible de ces modalités d'expression. On établira ici une distinction entre deux types de situations :

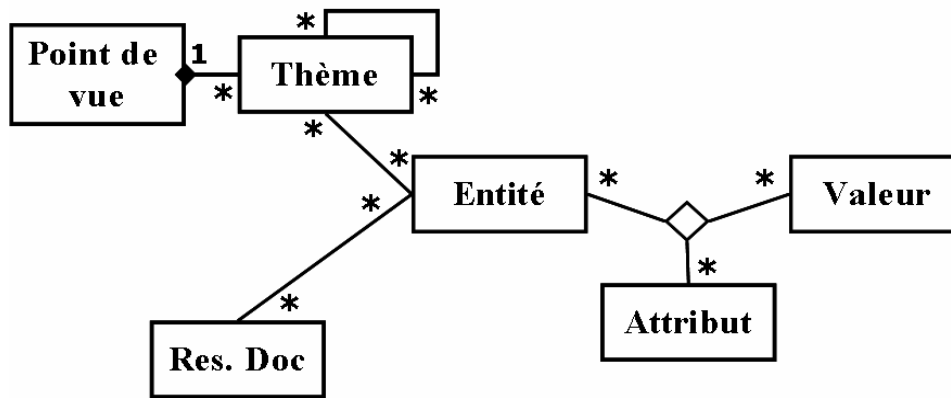


Fig.1 Le modèle Hypertopic (UML - diagramme de classes)

- celles dans lesquelles des composantes de l'entité complexe sont directement accessibles parce que celles-ci sont pour une large part intrinsèquement sémiotiques et que leur numérisation a permis de les rendre accessibles *via* la toile (livre, film, photo, musique...).

- celles dans lesquelles les composantes de l'entité complexe sont elles-mêmes largement de nature matérielle, instrumentale, relationnelle et ne sont pas en tant que telles numérisables (appareil, véhicule, voyage, formation, etc.). Dans ce cas, les documents fournissent un accès indirect aux composantes de l'entité comme c'est le cas pour les autres modes d'accès (attributs standards ou heuristiques).

Il existe deux principaux types d'accès à ces ressources. Par le biais de leur notice quand celle-ci est disponible ou par intermédiaire de l'indexation de la totalité du document, comme c'est possible dans le cas des documents « textuels ». La recherche par le biais de la notice, nous ramène à la première modalité. Le second type de recherche, de loin le plus populaire aujourd'hui, correspond à l'utilisation des moteurs qui combinent accès aux notices et accès aux contenus des documents, et ce de manière en partie indifférenciée.

3.3. Annuaire, ontologie sémiotique, cartographie

Le troisième mode de définition et d'accès repose sur la navigation hypertextuelle dans des annuaires, cartographies, ontologies sémiotiques [CAH 05]. Ceux-ci décrivent les entités par des thèmes selon différentes approches : pour partie liées à leur « définition logique », mais également contextuelle – les corpus d'où les thèmes sont extraits – situationnelle ou pragmatique – les situations d'usages

des termes – [ZAC 05]. La signification des associations dépend ainsi de leurs types, des distances qui les séparent, de leur disposition respective notamment d'un point de vue graphique.

Dans le cas des ontologies sémiotiques [CAH 05], les attributs heuristiques ou « thèmes » sont organisés selon les différents *points de vue sur l'entité* qui peuvent être adoptés par les acteurs qui produisent ou échangent les entités. Au sein du modèle Hypertopic cette caractéristique reçoit une attention particulière pour permettre la construction des cartes de thèmes ou ontologies sémiotiques.

4. Places de marché Web : étude empirique des fonctionnalités de RI

En nous appuyant sur la définition des différents modes d'accès aux entités complexes définis dans le modèle Hypertopic nous avons cherché à spécifier une infrastructure informatique pour la ROI. Le terme d'infrastructure renvoie pour nous à l'existence de différents dispositifs de stockage et composants logiciels associés rendus interopérables par la standardisation. Dans le contexte de notre étude nous avons du nous contenter de la réalisation d'un « environnement informatique » unique réalisé à l'aide du système Agoræ développé au sein du laboratoire Tech-CICO.

Nous avons décidé de travailler dans le contexte des places de marché en considérant que ces environnements offraient aujourd'hui un ensemble de fonctionnalités pour partie interopérables mettant bien en évidence la diversité des exigences associées à la ROI. Le prototype Agoræ avait d'ailleurs lui-même été considéré comme constituant une « place de marché à base de connaissances » ou KBM (Knowledge Based Marketplace, [CAH 04]).

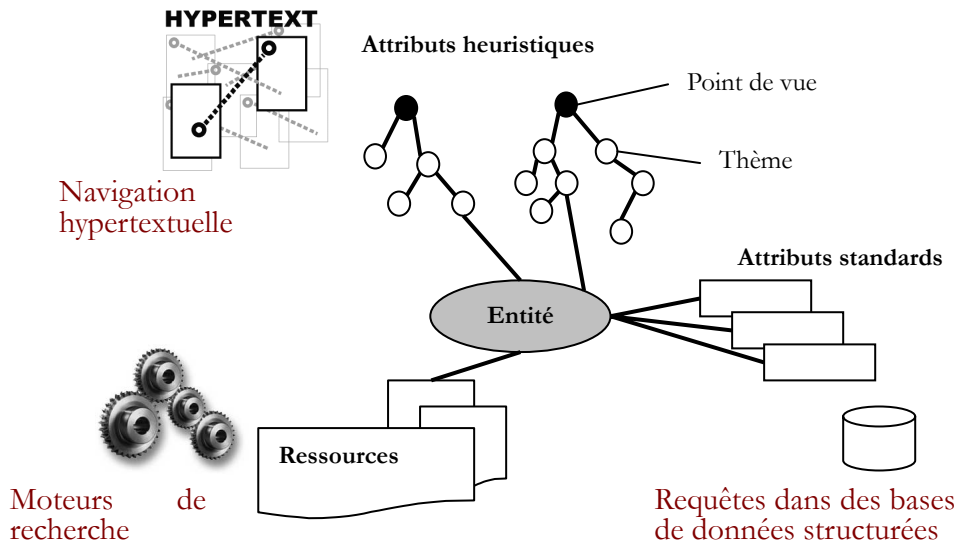


Fig.2 Le modèle Hypertopic appliqué à la ROI

Nous avons commencé à passer en revue un certain nombre de places de marché marchandes françaises et internationales présentes sur le Web dans le domaine des biens matériels et culturels « high-tech »¹⁰. Deux de ces exemples fédèrent plusieurs vendeurs, les autres correspondent à un modèle à vendeur unique. Les interfaces proposées par les places de marché étudiées associent – et cela semble être une tendance grandissante – un moteur de recherche plein texte et une cartographie plus ou moins sommaire (par produit puis par marque le plus souvent) mais aussi par usage pour certaines d'entres-elles. On trouve également souvent des moteurs de recherche multicritères. Parmi les dispositifs originaux que nous avons noté figure un moteur de recherche « situé », qui élague une partie de la cartographie pour ne traiter que les branches sélectionnées. On trouve également des moteurs de recherche qui explorent le contenu des livres.

Pour rechercher un produit dans ces places de marché, après avoir défini le type du produit (ici un téléphone mobile), l'utilisateur se voit souvent offrir trois techniques de recherche qui sont une particularisation des trois modes de définition et d'accès génériques aux entités complexes que nous avons présentées :

(1) **Un moteur de recherche plein texte** – qui s'apparente à Google dans son interface homme machine, bien qu'assez différent quant à la panoplie des technologies utilisées –,

(2) **Un moteur de recherche multicritère** en fonction des valeurs associées aux attributs standards du téléphone (poids, réseaux compatibles, gamme de prix, type d'abonnement, etc.),

(3) **Une arborescence présentant l'article selon différents points de vue** (même si la distinction entre les points de vue n'est pas aussi riche que dans les annuaires de type « Yahoo »).

A ces techniques de recherche correspondent différentes modalités de description de l'entité complexe :

(1) **Une fiche technique** fournissant les attributs standards stabilisés accessibles par le moteur de recherche multicritères.

(2) **Plusieurs ressources** documentaires sous forme d'image, de sons, de fichiers attachés et de descriptions textuelles.

(3) **Un ensemble de thèmes** extraits de l'annuaire et associés à l'entité notamment sous la forme du chemin retraçant les étapes de la navigation dans la hiérarchie ayant conduit à cette entité.

D'autres éléments peuvent être également disponibles : lien vers d'autres produits complémentaires ou concurrents, vers d'autres sites Web, commentaires d'internautes le plus souvent évaluatifs, etc.

5. Expliciter la représentation hybride des entités complexes

Bien que divers modes de représentation et différentes techniques de recherches soient offertes sur les places de marché, ceux-ci ne sont généralement pas considérés comme correspondant à des modes d'accès complémentaires au même type d'entité complexe. De ce fait, la plupart du temps, les modalités de passage d'une représentation à l'autre de l'entité ne sont que peu développées, à la fois pour assister la conception du système documentaire et son indexation et pour faciliter les activités des chercheurs d'information.

¹⁰ dont Fnac.fr, Surcouf.fr, intercity.fr, rue du commerce.fr, rue-montgallet, Amazon, E-Bay, worldgsm.com, gsmarena.com et France-gsm.fr

L'invitation à une mise en relation des différentes formes de définition et d'accès à l'entité complexe nous semble être la principale vertu du standard Hypertopic. Les objets matériels et immatériels qui font l'objet des transactions sont représentés dans les environnements informatiques numérisés selon les trois modalités génériques que nous avons présentées. La prise en compte des interdépendances entre ces formes de représentation est susceptible de faciliter les activités de ROI en multipliant les prises sur les entités que les chercheurs d'information devront tout à la fois découvrir et construire.

Une des manières d'organiser les relations entre les différentes facettes des entités complexes serait d'établir une ontologie formelle exhaustive. L'équipement téléphonique serait relié aux différents services offerts par les opérateurs et ceux-ci seraient associés aux différents usages possibles imaginés par les services marketing. Chaque image représentant le téléphone et ses usages serait précisément analysée et l'ensemble des ses « objets » formalisés logiquement.

Nous avons déjà expliqué ailleurs [CAH 05] [CAU 02] [ZAC 05] la raison pour laquelle cette voie n'était pas adaptée dans le cadre d'objets complexes en évolution rapide, souvent objets de controverses techniques et commerciales. Dans le contexte du Web Socio Sémantique, certains représentants de l'entité complexe, dont la dénomination est déjà en soi un enjeu complexe, sont rattachés par des liens faibles à ses différentes descriptions. C'est toujours l'utilisateur qui, *in fine*, à l'issue de ses activités d'usage, définira l'entité réelle dont les systèmes d'information ne peuvent lui fournir qu'une représentation approximative.

Les différentes modalités de représentation de l'entité complexe correspondent, dans Agoræ, aux trois composantes du modèle HypeTopic (Fig.1, Fig2) :

- Un **ensemble d'attributs standards** pour la description des valeurs des différentes instances. Dans l'exemple où l'objet représenté (et donc potentiellement recherché dans une ROI) est un téléphone mobile associé à des services (forfait, garantie, reprise, renouvellement...), il est nécessaire d'identifier une entité <téléphone> dotée d'une marque, d'un modèle, d'un opérateur, d'un poids :

```
<attribute uri="http://tech-ada.utt.fr/attribute/
name/"> marque</attribute>
```

La valeur de cet attribut pour la première instance de téléphone est par exemple Xinëø :

```
<entity uri="http://phone.zaher.tech-
cico.fr/phone/1/">
<attribute name="marque" value="Xinëø" />
</entity>
```

- Un certain nombre de **ressources documentaires** décrivant des caractéristiques moins standardisées des

instances destinées à être interprétées par les clients. Dans la plateforme Agoræ, il s'agit de ressources numérisées (textes mis en forme, images, séquences audio et vidéo...). Les ressources documentaires ou les fragments documentaires sont liés en Hypertopic *via* leurs url et uri¹¹, aux instances d'entités qu'ils décrivent. Par exemple, pour associer une image à notre téléphone :

```
<resource uri="http://www.xineo.com.tw/
cellphone/img/321.png">
<entity uri="http://phone.zaher.tech-
cico.fr/phone/1/">
</resource>
```

- Une **ontologie sémiotique** qui regroupe des thèmes correspondant à des attributs heuristiques organisés dans différents points de vue. La définition de ces points de vue ainsi que la hiérarchisation des thèmes correspondent aux intérêts d'un groupe d'acteurs. Pour les téléphones mobiles de notre exemple, un des points de vue pourrait être celui des « usages » (a), dans lequel le thème « Photo » (c) serait un sous-thème (cb) de « Multimédia »(b) :

```
<points-of-view>
<point-of-view uri="http://tech-ada.utt.fr/pov/1/">
Par usage </point-of-view>
</points-of-view> (a)
```

```
<topic name="Multimédia" uri="http://tech-
ada.utt.fr/viewpoint/1/topic/2/">
<point-of-view uri="http://tec...">
<related-topic rel-type="includes" uri="http://
tech-ada.utt.fr/pov/1/topic/3/">Photo
</relatedTopic> (cb)
</topic>(b)
```

```
<topic name="Photo" uri="http://tech-
ada.utt.fr/viewpoint/1/topic/3/">(c)
```

Le même thème « photo » indexe notre instance de téléphone (d) :

```
<entity uri="http://phone.zaher.tech-
cico.fr/phone/1/"> ...
<topic uri="http://tech-ada.utt.fr/viewpoint/1/
topic/3/"> (d)
</entity>
```

La figure 3 présente l'interface d'Agoræ au démarrage de la recherche dans la branche des appareils multimédia possédant une photo. La figure 4 présente l'interface au moment de l'exploration d'une des facettes de l'entité complexe par le biais de documents multimédias.

¹¹ Et dans les versions prochaines via Web-DAV.

6. Perspectives : Vers la définition d'un module de « mémoire externe de construction »

Sur la base de la mise en relation des différents modes d'accès et représentations offerts dans l'environnement Agoræ, notre objectif est aujourd'hui de faciliter les activités de découverte/construction de l'entité complexe par le chercheur d'information. L'objectif de ce module, que nous nommerons « mémoire externe de construction », est de faciliter la mise en relation des différents éléments de description hybride de l'entité en assurant une trace des recherches déjà effectuées et en contraignant les recherches ultérieures soit en fournissant une aide mnésique consultable soit en propageant les contraintes issues des décisions antérieures. Aujourd'hui, cette possibilité n'est offerte dans aucun des sites que nous avons étudiés chaque modalité d'accès à l'information étant indépendante de l'autre (à l'exception du site restreignant le champ de recherche du moteur à la branche de l'ontologie sémiotique en cours d'exploration, mais sans mémorisation des actions antérieures).

Nous n'en sommes aujourd'hui qu'à la phase de spécification de ce module qui nous semble devoir offrir une aide essentielle aux activités de ROI. Tout en s'apparentant au « panier » très utilisé dans les sites de commerce électronique, ce module doit également

permettre une mémorisation structurée non seulement des représentants des entités complexes candidates mais encore des critères de choix et des indices découverts lors de l'exploration. Ces indices peuvent provenir des différents modes d'accès à l'entité (attributs standards, thèmes et points de vue, documents textuels et multimédia) que l'utilisateur a mobilisé dans ses dernières activités de ROI pour lui permettre de les réutiliser lors de recherches futures.

Dans notre exemple, le chercheur d'information devrait non seulement pouvoir retenir certains critères représentés par des attributs standards mais également pouvoir annoter telle ou telle partie d'un document de présentation, d'une démo en ligne, d'une photo, ou le son d'un choix de sonneries proposées, etc. Il devrait pouvoir compléter les ontologies sémiotiques pour son usage propre en rajoutant à la volée des thèmes personnels liés aux thèmes offerts par défaut.

Dans les contextes où les entités changent rapidement et où les besoins des différents usagers sont susceptibles d'évoluer en permanence le recours aux attributs standards permettant de construire des tableaux comparatifs paramétrique n'est pas toujours généralisable. L'approche de la ROI hybridant données, documents et ontologies sémiotiques est une solution permettant de ne pas exiger des informaticiens de rallonger indéfiniment les listes d'attributs gérés sous formes de données. De ce point



Fig. 3 - Maquette de navigation augmentée par des moteurs de recherche de type Agoræ pour la ROI


Attributs heuristiques

Catalogue > MOBILES > Par usage > Photo > APN 1 méga > Motorola V3i mais aussi dans [Marques > Motorola et Nouveautés](#)

MOBILES


- ▶ [Bons plans](#)
- ▶ [Nouveautés](#)
- ▶ [Kits](#)
- ▶ [Par usage](#)
 - ▶ [3G](#)
 - ▶ [Bandes](#)
 - ▶ [Couleur](#)
 - ▶ [Photo](#)
 - ▶ [Format VGA](#)
 - ▶ [Format 1 méga](#)
 - ▶ [Format 2 mégas](#)
 - ▶ [Bluetooth](#)
 - ▶ [MP3](#)
- ▶ [Marque](#)



Motorola V3i Razzr



Le V3i de Motorola a été conçu pour fournir des performances exceptionnelles.

Sous son design ultra-mince se dissimulent des fonctionnalités avancées telles que la lecture de clips vidéo MPEG4, la technologie



Bluetooth®, un appareil photo numérique 1,3 mégapixels, jusqu'à 512 Mo de mémoire, et bien plus encore...

Son clavier de conception raffinée, son style minimaliste et sa finition métallisée confèrent au V3i un aspect aussi sophistiqué que ses performances.

Poids & dimensions

Dimensions	93 x 46,6 x 21,7 mm
Poids	92 g

Affichage

Résolution écran interne	TFT 262 144 couleurs, 128 x 160 pixels
Résolution écran externe	-

Appareil photo

Appareil photo intégré	Oui, avec flash et enregistrement vidéo
Résolution	SXGA (1280 x 960 pixels), 1,3 Mégapixels

Rechercher :

Clips vidéo mot

Rechercher par attribut :

Prix

Poids

Ecran

Autonomie

[rechercher juste dans cette branche](#)

Attributs standards

Ressources documentaires

Fig. 4 - Aperçu pour évaluation d'une entité candidate dans une interface de type Agoræ

de vue la ROI offre une piste permettant de faciliter la gestion des grands systèmes de gestion de la relation client (CRM) en intégrant données structurés et données semi-structurées.

7. Conclusion

Bien qu'encore programmatique la conceptualisation en termes de ROI nous semble ouvrir de multiples voies de recherche à la fois théoriques et pratiques pour l'amélioration des méthodes de recherche sur le Web.

L'étude des sites marchands et la maquette réalisée à l'aide d'Agoræ a déjà enrichi nos questionnements de manière pertinente. De nombreuses techniques documentaires pourraient contribuer au développement du module de mémoire externe de construction (métadonnées, annotations, syndication...), renforcées ou non par des techniques d'intelligence artificielle (e.g. CBR).

Sur un plan théorique, la définition conceptuelle de

l'entité complexe et de ses relations avec les trois représentations génériques que nous avons mises en évidence doit faire également l'objet de recherches prochaines.

REFERENCES

R. Apascal et al., Document à structures multiples, *In actes de Sciences of Electronic, Technologies of Information and Telecommunications (Setit'04)*, Sousse, 15-20 mars 2004.

J.-F. Baget, É. Caunaud, J. Euzénat et M. H. Hacid Les langages du Web sémantique, In J.Charlet, Ph. Laublet, Ch. Reynaud (eds) *Web Sémantique : Rapport final de l'action spécifique 32 CNRS/STIC*, CNRS, octobre 2003.

M.-J. Bates, The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13(5):407-431, 1989.
url : <http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>

- A. Bénel et al., Truth in the digital library, from ontological to hermeneutical systems, In *proceedings of the European Conference on Digital Libraries (ECDL 01)*, pp 366-377, Darmstadt, 2001.
- J.-P. Cahier, M. Zacklad, Towards a Knowledge-Based Marketplace model (KBM) for cooperation between agents. In *proceedings of COOP'2002*, pages 226-238, St Raphael : IOS Press, 2002
- J.-P. Cahier, M. Zacklad, A. Monceaux, Une application du Web socio-sémantique à la définition d'un annuaire métier en ingénierie. In *actes des 15^{ème} journées francophones d'Ingénierie des Connaissances IC 2004*, pages 29-40, Lyon : PUG, 2004.
- J.-P. Cahier Ontologies sémiotiques pour le Web socio sémantique: Etude de la gestion coopérative des connaissances avec des cartes hypertextuelles, Thèse en Informatique de l'Université de Technologie de Troyes, soutenue le 14 dec. 2005.
- J. Caussanel , J.-P. Cahier, M. Zacklad, J. Charlet, Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web Sémantique?. In *actes des journées francophones d'ingénierie des connaissances IC2002*, Rouen : 2002.
- J. Charlet, Ph. Laublet, Ch. Reynaud, Web Sémantique : Rapport final de l'action spécifique 32 CNRS/STIC , CNRS, octobre 2003.
- F. Chateauraynaud, Moteurs de la recherche et pragmatique de l'enquête, les sciences sociales face au Web connexionniste, Colloque Classification et Histoire, Nanterre, 5 octobre 2005.
- Y. Chiaromonte, Browsing and querying: two complementary approaches for multimedia information retrieval, Hypertext-Information Retrieval- Multimedia, *HIM'97*, 1997.
- A. Ciaccia, et D. Martins, « Recherche d'informations sur le Web : Etude de l'influence de facteurs liés à l'interface, à l'utilisateur et à la tâche », *Revue d'intelligence artificielle*, 19, 1/2, p.159-177, 2005.
- J. Dewey, *Logique, la théorie de l'enquête*, (trad. Gerard Deledalle re-éd. 1993), collection L'interrogation philosophique, PUF, 1938.
- M. D. Dunlop, C.W. Johnson, Reid, Exploring the layers of information retrieval evaluation, *Interacting with computers*, 10, p. 225-236, 1998.
- A. Hatchuel, Towards Design Theory and expandable rationality : The unfinished program of Herbert Simon, *Journal of Management and Governance*, 5:3-4, 2002.
- H. Kim, and S.C. Hirtle, Spatial metaphors and disorientation in hypertext browsing, *Behaviour and Information technology*, 14, pp 239-250, 1995.
- E. Kolmayer, Démarche d'interrogation documentaire et navigation , In J.F. Rouet et B. De La Passardière (eds.), Actes quatrième colloque Hypermédiat et Apprentissage, p.121-134, Poitiers, 1998.
- G. Marchionini, *Information seeking in electronic environment*, Cambridge University Press, 1995.
- M. Magennis, C. J. Van Rijsbergen, The potential and actual effectiveness of interactive query expansion, In *Proceedings of ACM Conference on Research and Development in Information Retrieval (SIGIR'97)*, Philadelphia, 1997.
- S. McDonald and S., R. Stevenson, Navigation in hyperspace: an evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext, *Interacting with computers*, 10, p. 129-142, 1998.
- M. Melucci, An Evaluation of Automatically Constructed Hypertexts for Information Retrieval, In *Information Retrieval*, ISSN: 1386-4564, Volume 1, Numbers 1-2, pp 91-114, Springer Netherlands, 1999.
- R. Pietro-Diaz and P. Freeman Classifying software for reusability, *IEEE Software*, 18(1), 1987.
- Y. Prié, T. Limane et A. Milles, Isomorphisme de sous-graphe pour la recherche d'information audiovisuelle contextuelle. In *Actes du 12ème congrès Reconnaissance de Formes et Intelligence Artificielle, (RFIA'2000)*, v 1, pp 277-286, Paris, 2000.
- J.-F., Rouet, A. Tricot, Chercher de l'information dans un hypertexte : vers un modèle des processus cognitifs. In *Hypertextes et Hypermédiat*, n° hors série, p. 57-74, 1998.
- H. Stuckenschmidt et al. , A Topic-Based Browser for Large Online Resources, In E. Motta and N. Shadbolt (eds) *Proceedings of the 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW04)*, Lecture Notes in Artificial Intelligence, Springer-Verlag, 2004.
- B.S Wynar and A. G. Taylor, *Introduction to Cataloging and Classification*, Libraries Unlimited, 1992.
- M. Zacklad, et al., Ergonomie de la recherche d'information sur le Web : quel modèle de l'activité, *Modèles et modélisation de l'Internet*, Ecole d'été du GDR TICS, Batz sur mer : 2005.
- M. Zacklad, Introduction aux Ontologies sémiotiques dans le Web Socio Sémantique, in M.-Ch. Jaulent (eds) *Actes des 16ème journées francophones d'Ingénierie des Connaissances (IC'2005)*, pp. 241-253, PUG, ISBN 2-7061-1284-0, Nice, 2005.
- L'H. Zaher, J.-P. Cahier, N. Gauducheau et M. Zacklad, Vers une évaluation des schémas de classification pour la gestion de l'information métier , In I. Saleh et J. Clément (eds) *Actes de la conférence Hypermédiat, Hypertexts, Products, Tools and Methods (H2PTM'05)*, pp. 365-379, Hermès Science, ISBN 2-7462-1244-7, Paris, 2005.
- L'H. Zaher, J.-P. Cahier, M. Zacklad, Towards Open Information Retrieval, In *Proceedings of IEEE International Conference on Service Systems and Service Management (IC SSSM'2006)*, pp. 41-46, IEEE, ISBN 1-4244-0451-7, Troyes, 2006.