

# Intelligence qualitative dans les situations à risque complexes et apports du web socio-sémantique

Manuel Zacklad<sup>1</sup>, Jean-Pierre Cahier<sup>1</sup>, Aurélien Bénel<sup>1</sup>, L'Hédi Zaher<sup>1</sup>, Christophe Lejeune<sup>2,1</sup>, Chao Zhou<sup>1</sup>

<sup>1</sup>Tech-CICO - Institut Charles Delaunay  
Université de Technologie de Troyes / CNRS (FRE 2848)  
12 rue Marie Curie, 10010 - Troyes Cedex, France  
<http://www.tech-cico.fr>  
{zacklad, cahier, benel, zaher, chao.zhou}@utt.fr  
<sup>2</sup>CEMAD, 4000 Liège, Belgique  
Christophe.Lejeune@ulg.ac.be

**Résumé** – La prise en compte des informations qualitatives dans la gestion des situations à risque complexes doit contribuer au développement d'une branche, selon nous essentielle, des activités associées à la veille stratégique, au renseignement et à la recherche d'information, que nous suggérons de nommer « intelligence qualitative ». Celle-ci correspond à la mise en œuvre d'activités relevant d'une rationalité documentaire qui dépendent étroitement des environnements informationnels de plus en plus numérisés à l'intérieur desquels s'exercent l'activité des acteurs. Nous contribuons à la fois à l'analyse de ces environnements et à la conception de solutions d'assistance au service de la rationalité documentaire et de l'intelligence qualitative dans le cadre de nos recherches sur le web socio-sémantique opérationnalisées dans la méta-sémiotique et le protocole HyperTopic.

**Abstract** – The consideration of qualitative information in the management of complex risk situations must contribute to the development of a research path, according to us essential, bound to strategic intelligence and information seeking that we suggest naming "qualitative intelligence". It corresponds to the implementation of activities being a matter for a *documentary rationality* that depends closely on the informational environments more and more digitized in which the activities of actors unfold. We contribute to the analysis of these environments and to the conception of support systems to the service of documentary rationality and qualitative intelligence in our researches on socio-semantic web implemented in the metasemiotic and HyperTopic protocol.

## 1. Gestion de l'information qualitative dans les situations à risque complexes (SIRIX)

Dans cette communication nous aborderons l'apport des méthodes de gestion et de recherche d'information relevant du web socio-sémantique dans les Situations à Risque complexes (SIRIX). Cette expression est d'ailleurs en partie pléonastique : toute situation complexe présente des risques potentiels mais toutes les situations à risque ne sont pas complexes. Si en reprenant la formulation de Lemoigne, on considère qu'une situation complexe implique de modéliser et raisonner « dans une interaction récursive et projective permanente » (Lemoigne 1997) on définira une situation à risque complexe comme *une situation dans laquelle la nature du danger est en permanence requalifiée dans l'interaction récursive et projective avec la situation (prise d'information, interactions sociales, manipulation physique d'artefacts...) et sa représentation*.

Ces situations impliquent une gestion et un traitement en continu de l'information permettant, notamment, de diminuer la surcharge cognitive face à l'afflux de données associé au risque pris en charge (cf. les principes de rationalité limitée mis en évidence par H. A. Simon, 1969). Parmi l'ensemble des informations qui circulent dans les SIRIX, nous considérerons essentiellement ici les

informations de nature qualitative (informations issues de bases données, messages, dépêches, images, film, enregistrements divers, articles, textes de référence, manuels de procédure, documents de référence...) qui soit, sont elles-mêmes issues de l'agrégation ou de la synthèse d'informations provenant de capteurs, soit ont été directement saisies, rédigées, enregistrées, par des acteurs humains. Quand ces informations sont inscrites ou enregistrées sur des supports pérennes, nous considérons, sous certaines conditions, qu'elles relèvent de la catégorie générale des documents et donc qu'elles constituent des informations documentaires.

La prise en compte des informations qualitatives dans la gestion des situations à risque complexes doit contribuer au développement d'une branche, selon nous essentielle, des activités associées à la veille stratégique, au renseignement et à la recherche d'information, que nous suggérons de nommer « intelligence qualitative ». Celle-ci correspond à la mise en œuvre d'activités relevant d'une *rationalité documentaire*<sup>1</sup> qui dépendent étroitement des environnements informationnels de plus en plus numérisés à l'intérieur desquels s'exercent l'activité des acteurs.

<sup>1</sup> Dans un contexte différent, J. Goody (1979), parle de « raison graphique ». La généralisation de la numérisation documentaire a encore accentué le phénomène pour produire ce que Pédaque (2006) nomme la « redocumentarisation du monde ».

Nous contribuons à la fois à l'analyse de ces environnements et à la conception de solutions d'assistance au service de la rationalité documentaire et de l'intelligence qualitative dans le cadre de nos recherches sur le web socio-sémantique opérationnalisées dans la méta-sémiotique et le protocole HyperTopic.

## 1.1 De la rationalité instrumentale à la rationalité documentaire

Les progrès récents des sciences cognitives, notamment quand elles prennent en compte la dimension collective de l'activité et les problématiques de conception d'environnements de support à ces activités comme dans le courant du CSCW (Computer Supported Cooperative Work), accordent une place de plus en plus notable aux artefacts instrumentaux dans la compréhension des activités cognitives (par exemple, Salembier 2007). Plus encore que d'une rationalité limitée, qui met principalement l'accent sur les capacités limitées du système de traitement de l'information humaine, notamment en terme de mémoire et de contraintes temporelles associées à la prise de décision, il s'agit de la mise en évidence de la dimension profondément externalisée de la cognition humaine dans ce qui relève d'une *rationalité instrumentée* (par exemple, Rabardel, 1995, Bachimont 2004).

Quand cette rationalité instrumentée mobilise principalement des productions sémiotiques, elle débouche sur une *rationalité sémiotique* qui peut être définie comme une conduite finalisée qui tout à la fois (a) s'alimente de productions langagières (au sens de système de signes produits par l'homme), (b) s'appuie dans son déroulement sur la production continue de ces signes et (c) dont la finalité relève elle-même de la réalisation de nouvelles productions sémiotiques ou intègre de manière consubstantielle, dans des actions à dominante physique, leur justification langagière. Quand cette rationalité sémiotique s'exerce dans le cadre de transactions coopératives spatio-temporellement distribuées (Zacklad 2006, 2007a), les productions sémiotiques sont documentarisées et la rationalité sémiotique est essentiellement une *rationalité documentaire* à l'intérieur de laquelle les conduites humaines individuelles et collectives s'alimentent à partir de fragments documentaires en constante évolution et produisent en continu d'autres fragments en soutien de l'activité et comme résultat principal de celle-ci.

Cette prise en compte de la rationalité documentaire s'inscrit elle-même dans le cadre des réflexions actuelles en sciences de l'information sur la notion de document (R. T. Pédaque 2006) auxquelles nous contribuons par la théorie de la documentarisation et des Documents pour l'Action (Zacklad 2006, 2007a), qui nous permet de décrire l'usage des Technologies de l'Information comme relevant de pratiques documentaires et les Systèmes d'Information comme étant constitués d'un ensemble de documents inter-reliés et dynamiques (Zacklad 2007c). Notons bien ici que l'emploi du terme d'information

diffère du sens extrêmement large qu'il possède chez Simon, dans la modélisation du système humain de traitement de l'information, comme chez la plupart des psychologues cognitivistes et des économistes, pour ne désigner que des artefacts sémiotiques externes au sujet. En économie, par exemple, le terme d'information est le plus souvent utilisé de manière imprécise sans distinction entre les stimuli que l'agent perçoit et les connaissances qu'il construit à partir d'eux. Notre emploi du terme n'est pas non plus lié à sa signification mathématique basée sur des considérations probabilistes.

## 1.2 Activités d'enquêtes suscitées par la gestion des SIRIX

La rationalité instrumentée s'exerce dans le cadre d'activités d'enquête (Dewey, 1938) qui comportent deux phases principales (Zacklad 2007d) auxquelles il est possible d'en adjoindre une troisième (Fig 1) : une phase d'analyse et de recueil d'information permettant de poser le problème, que nous appellerons dans notre contexte phase de Recherche Ouverte d'Information (Zaher et al. 2006, 2007), une phase d'exploitation permettant de poursuivre l'interprétation et d'élaborer des solutions pour transformer la situation et une phase de capitalisation. Ces trois phases se retrouvent dans tous les types d'enquête que nous proposons de regrouper selon trois catégories (Zacklad 2007d) : enquête de localisationaccès (par exemple, localisation physique d'un objet recherché), enquête d'analyse-décision (par exemple, choix d'un emplacement, identification d'une cause, sélection d'un moyen...), enquête de synthèse-conception (par exemple, élaboration d'un plan d'intervention, élaboration d'une solution technique, etc.). Localisation, analyse et synthèse correspondent plutôt à la phase dite de Recherche Ouverte d'Information (Zaher et al. 2006, 2007) tandis qu'accès, décision et conception correspondent à la phase de mise en œuvre ou d'exploitation.

- |  |
|--|
| <ol style="list-style-type: none"><li>1. Recherche Ouverte d'Information - ROI<br/>(compréhension du problème)<ol style="list-style-type: none"><li>1. Recherche (pull) / Filtrage (push)</li><li>2. Interprétation / Codage</li><li>3. Organisation / Classement -&gt; Archives courantes</li></ol></li><li>2. Exploitation (résolution du problème)<ol style="list-style-type: none"><li>1. Interprétation / Codage</li><li>2. Rédaction / Mise en forme</li><li>3. Edition / Partage (pull) / Diffusion (push)</li></ol></li><li>3. Capitalisation (mémorisation)<ol style="list-style-type: none"><li>1. Classement / Stockage -&gt; Archives intermédiaires</li></ol></li></ol> |
|--|

Fig. 1. Trois phases et sept étapes correspondant aux manipulations documentaires associées aux enquêtes

Pour bien cerner le rôle que l'on peut attendre des concepts, méthodes et outils issus des sciences de l'information, de l'ingénierie des connaissances, de l'informatique documentaire et du CSCW qui sont dans le focus de notre problématique, nous avons également

proposé de distinguer, à l'intérieur de chacun de ces types d'enquête, le caractère documentaire ou non documentaire des activités de recherche d'information et d'exploitation (Zacklad 2007d). On peut ainsi distinguer des phases de recherche d'information documentaire (utiliser un moteur de recherche et lire des extraits de texte) ou non documentaire (interroger des acteurs ou réaliser des observations) et des phases de mise en œuvre documentaire (impliquant une rédaction) et non

documentaire (impliquant communication verbale et actions physiques, par exemple). Les méthodes que nous examinerons ici sont bien sur essentiellement pertinentes quand les phases de Recherche Ouverte d'Information (ROI) et d'exploitation s'exercent dans le cadre d'une rationalité documentaire. Quand les activités d'enquêtes sont suscitées par une SIRIX dans le cadre d'une rationalité documentaire elles correspondent au déploiement de l'intelligence qualitative.

TAB 1. Exemple de type d'enquête prototypique selon la rationalité instrumentée

| Phase de l'enquête / Type d'enquête   | Recherche ouverte d'information documentaire   | Recherche ouverte d'information extra-documentaire   | Phase d'exploitation documentaire (mise en œuvre)   | Phase d'exploitation non documentaire (mise en œuvre)                                     |
|---|--|--|---|---|
| <b>Localisation/ accès</b> (items = des zones à analyser/ explorer)                                       | Trouver la localisation d'un objet/personne à l'aide d'un document (p.e .carte)  | Localiser des objets ou individus en se déplaçant  | Organiser l'accès à l'objet choisi à l'aide de documents  | Récupérer les objets matériels localisés  |
| <b>Analyse / décision</b> (items = artefacts, personnes, situations à sélectionner puis choisir)          | Analyser des objets/ personnes/ situation pertinents à partir de leur description documentaire   | Manipuler /explorer les objets / personnes / situations pour obtenir l'information caractérisant l'item        | Faire un tableau comparant les différents types de solutions selon différents critères                        | Débattre ou comparer matériellement les items déjà sélectionnés et partiellement analysés |
| <b>Synthèse/ conception</b> (items = des composants, savoir-faire, contraintes à assembler puis combiner) | Rassembler les connaissances, composants, contraintes nécessaires à la conception d'une action / situation / artefact à partir de leur documentation | Recueillir des composants et connaissances en interagissant avec d'autres acteurs ou par manipulation physique | Concevoir une solution sous une forme langagière et documentaire (forme finale ou préparant la mise en œuvre) | Réaliser matériellement la solution par des manipulations physiques                       |

### 1.3 Corpus primaire, secondaire et tertiaire

L'enquête exploite et produit différents types de documents au sens élargi que nous donnons à ce terme qui englobe à la fois des productions sémiotiques multimédia sous forme papier ou numérique (Zacklad 2006, 2007a, Pédaque 2006) et la plupart des supports informatiques allant des bases de données, aux messageries électroniques en passant par les forums et les blogues (Zacklad 2007c). Ces documents jouent un rôle selon les phases de l'enquête. En nous inspirant d'une analyse effectuée dans le contexte de la définition des principes de fonctionnement des collaboratoires<sup>2</sup> dans le domaine des

sciences humaines et sociales (Zacklad 2007e) nous ferons la distinction entre trois ensembles documentaires constituant des corpus documentaires définis comme des « archives structurées de documents sélectionnés et/ou produits de manière systématique pour permettre le déroulement d'une enquête » (Zacklad 2007e) :

- Les corpus primaires sont ceux qui constituent la base « empirique » de la recherche et à partir desquels seront élaborées les conclusions de l'enquête.
- Les corpus secondaires sont constitués par des documents qui contribuent à l'étude des sources primaires : documents produits à l'issue d'autres enquêtes similaires, ouvrages méthodologiques, plan de codage. état de l'art, document méthodologiques.
- Les corpus tertiaires sont constitués par les productions originales de l'enquête, dans un état intermédiaire ou final (Fig. 2).

<sup>2</sup> Les collaboratoires sont des laboratoires collaboratifs permettant le travail à distance entre scientifiques.

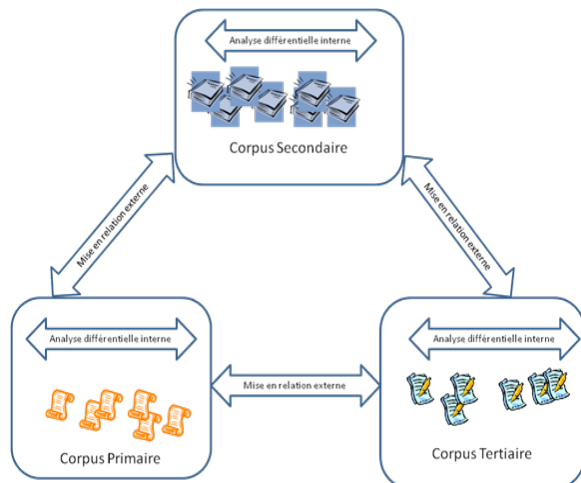


Fig. 2. Les phases de ROI produisent essentiellement les corpus primaire et secondaire et la phase d'exploitation plutôt le corpus tertiaire

On établit une distinction entre le corpus effectif, qui contient des documents dont la pertinence a été attestée et le corpus virtuel, documents dont la sélection n'a pas été justifiée « par une analyse comparative explicitant les formes d'association entre les documents et les fragments documentaires rassemblés » (Zacklad 2007e). L'explicitation qui produira le **corpus effectif**, sera réalisée à partir de systèmes d'organisation des connaissances (SOC, cf. infra) permettant l'annotation ou l'étiquetage, dans le cadre d'opérations de codage.

### 1.4 Environnements physiques et réseaux d'acteurs associés aux activités d'enquête

Les différentes phases de l'enquête impliquent une manipulation des documents et leur circulation entre plusieurs environnements physiques adaptés aux formes de traitements qui les visent, traitements qui associent conjointement informatique et action humaine (Fig 3). On distinguera trois principaux environnements :

- L'environnement externe d'entrée est constitué par des dossiers, des documents, des fragments documentaires issus du web public, de bases de données générales ou préparées par des informateurs spécifiques. C'est l'environnement où se déploie l'activité des réseaux d'informateurs potentiels.
- L'environnement de travail des gestionnaires de risque est celui dans lequel se déroulent les activités d'enquête (activité de ROI, d'exploitation et de capitalisation). Il est alimenté par les activités de ROI à partir de l'environnement externe, activités de ROI qui interagissent elles-mêmes avec les activités d'exploitation. Les activités d'exploitation s'appuient sur trois espaces interdépendants permettant la réalisation des phases d'activités correspondantes : interprétation/codage, rédaction/mise en forme, diffusion/partage (Fig 3). Ces activités et les espaces correspondant interagissent entre eux comme avec les

activités de ROI. Les activités de capitalisation alimentent l'espace d'archivage intermédiaire.

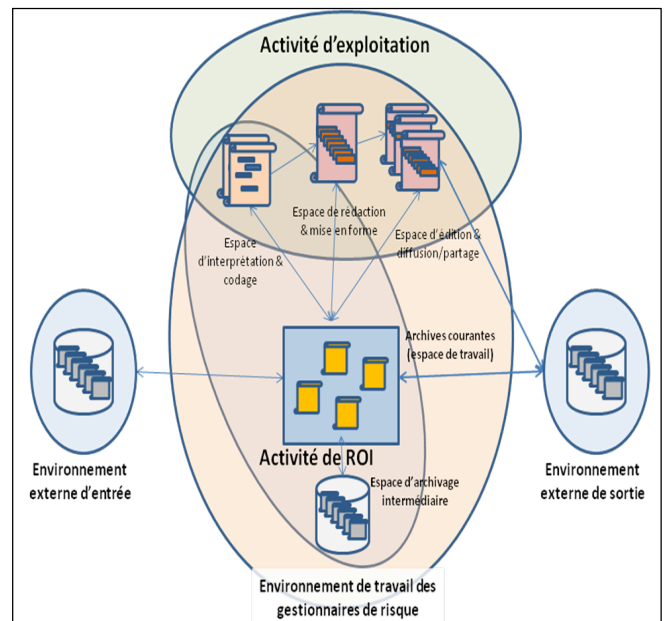


Fig. 3. Les environnements physiques dans lesquels se déploie la rationalité documentaire.

- L'environnement externe de sortie est alimenté par les activités d'édition. Selon une modalité « push » (adressage) on peut considérer que les documents circulent effectivement vers un environnement physique différent pour être distribués à leurs destinataires. Selon une modalité « pull » (mise à disposition pour l'accès) l'environnement externe de sortie est constitué par un portail d'accès. C'est l'environnement où se déploie l'activité des destinataires potentiels des résultats partiels ou définitifs de l'enquête. Dans certains cas, ces destinataires peuvent être les mêmes personnes que les informateurs potentiels qui tout à la fois véhiculent les informations primaires et sont les bénéficiaires des activités de gestion de risque productrices des informations tertiaires.

### 1.5 Séquencement et division du travail d'intelligence qualitative

Chacune des étapes de gestion de l'information qualitative peut être réalisée isolément ou de manière coopérative. Dans le second cas, plusieurs acteurs défendant chacun leur point de vue peuvent intervenir dans l'interprétation ou la rédaction. Cette caractéristique a des conséquences très importantes sur les propriétés des Systèmes d'Organisation des Connaissances qu'il faudra mettre à disposition du collectif. Par ailleurs, le séquencement des différentes phases dépend des contraintes et des stratégies adoptées. Il est possible de paralléliser la recherche et l'exploitation des informations et à l'intérieur de la phase d'exploitation et de paralléliser en partie les étapes d'interprétation, de rédaction et de

diffusion/partage. Quand il y a simultanément travail coopératif et parallélisation on peut imaginer des situations dans lesquelles les destinataires des résultats de l'enquête prennent connaissance de certains éléments d'information en cours d'élaboration (corpus tertiaire) et viennent contribuer à l'interprétation à partir des informations dont ils disposent, voire viennent enrichir les corpus primaires et secondaires.

## **1.6 Prise en compte de la vitesse d'évolution des SIRIX**

Une autre caractéristique discriminante des situations à risque complexe est liée à la vitesse d'évolution de la situation, vitesse qui a elle-même des conséquences sur la vitesse à laquelle les enquêtes doivent être conduites. Nous proposons d'établir une distinction entre trois familles de processus évolutifs en fonction de leur dynamique : des processus à délais de réponse rapide (quelques heures), des processus à délai de réponse intermédiaire (quelques jours) et des processus à délai de réponse lent (situation de crise progressive pouvant mettre plusieurs mois à se déclarer avant d'atteindre son climax). La dynamique du processus a un impact très important sur la nature des solutions à envisager pour gérer les phases de recherche d'information et d'exploitation documentaire.

Notons que, quand les situations à délais de réponse rapide ont un degré de récurrence important et que les responsables de ces situations souhaitent développer à leur propos des processus d'apprentissage, ils peuvent parfois être en partie ramenés à des processus à délai de réponse lent contenant plusieurs sous-processus à délais d'intervention brefs. Selon l'état de développement de la crise, risque latent, risque avéré mais n'impliquant pas d'action urgente, risque impliquant une réaction rapide, les modalités de gestion de l'information et d'organisation des acteurs responsables diffèrent.

## **2. Modélisation de l'environnement documentaire du web socio-sémantique avec HyperTopic dans le cadre de la rationalité documentaire**

L'environnement physique dans lequel s'exerce la rationalité documentaire est aujourd'hui de plus en plus numérisé et inscrit dans les infrastructures du web. Les frontières entre le web public, les web d'entreprise et l'environnement de travail personnel des acteurs sont de plus ténues et il est nécessaire de disposer d'outils de recherche et de gestion de l'information permettant de maintenir une certaine cohérence lors du franchissement de ces frontières. Pour conceptualiser les nouvelles pratiques documentaires de plus en plus numérisées, nous développons au sein de l'équipe Tech-CICO l'approche du Web socio-sémantique (W2S).

Selon cette vision, le web est d'abord appréhendé comme un instrument de gestion documentaire facilitant des transactions coopératives interpersonnelles éventuellement très asynchrones et distribuées entre des

acteurs individuels et collectifs. Pour faciliter ces transactions coopératives, il est nécessaire de s'appuyer sur des ressources terminologiques organisées dans des Systèmes d'Organisation des Connaissances (SOC) : classifications, thésaurus, ontologies formelles, ontologies sémiotiques, folksonomies, etc. Dans l'approche du web socio-sémantique, la sémantique de ces SOC est 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 (Zacklad 2005, 2007d, Zacklad et al. 2007).

A la fois pour modéliser les activités de recherche et d'exploitation de l'information et pour concevoir des environnements d'assistance, nous avons développé la métasémiotique HyperTopic. Celle-ci est directement conçue pour faciliter des activités d'enquête mises en œuvre par les membres d'une communauté ou d'une organisation impliqués dans des tâches récurrentes et pour partie communes médiatisées par le Web socio-sémantique (Cahier, 2005). Un modèle Hypertopic résulte de l'application de la métasémiotique à un domaine et à une tâche pour documentariser des contenus sémiotiques afin de mettre en œuvre des raisonnements relevant de la classification heuristique (Clancey, 1985).

La classification heuristique correspond pour nous à un raisonnement d'un très haut niveau de généralité applicable aux différents types d'enquêtes de localisation/accès, analyse/décision et synthèse/conception. Vue à travers le prisme de la classification heuristique, l'enquête consiste à caractériser une situation ou un artefact qu'il s'agit d'analyser ou de concevoir en référence à d'autres situations ou artefacts analogues. Ces situations ou artefacts, dont la caractérisation est le résultat de l'enquête, correspondent à la catégorie d'item. L'expression d'item vise notamment à souligner que la situation ou l'artefact étudié fait lui-même partie d'une collection et que l'analyse des similarités entre les éléments de cette collection est essentielle au raisonnement.

Il existe deux grandes étapes itératives dans le cycle de vie d'un modèle Hypertopic. La première consiste à enrichir le modèle en caractérisant de nouveaux items qui ne figurent pas encore dans le modèle. La seconde consiste à rechercher des items déjà en partie caractérisés pour résoudre de nouveaux problèmes. L'enrichissement du modèle implique trois activités correspondant aux trois formes de description de l'item (Fig. 4).

### **2.1.1 Spécification référentielle via les attributs standards**

La première est la spécification référentielle de l'item qui est réalisée à l'aide de la catégorie d'attribut standard. Les attributs standards sont des descriptions conventionnelles et standardisées des situations et des artefacts qui s'appuient souvent sur leurs caractéristiques mesurables. Ils correspondent aux mesures, aux coordonnées topographiques, aux données administratives. La sémantique est référentielle et les outils principalement

utilisés sont des bases de données. Quand ils sont organisés, les attributs standards constituent des SOC formels : schéma entité-association, modèle objet, graphe conceptuel, ontologie formelle (cf. Zacklad 2007d). Les requêtes utilisées sont elles-mêmes de nature formelle.

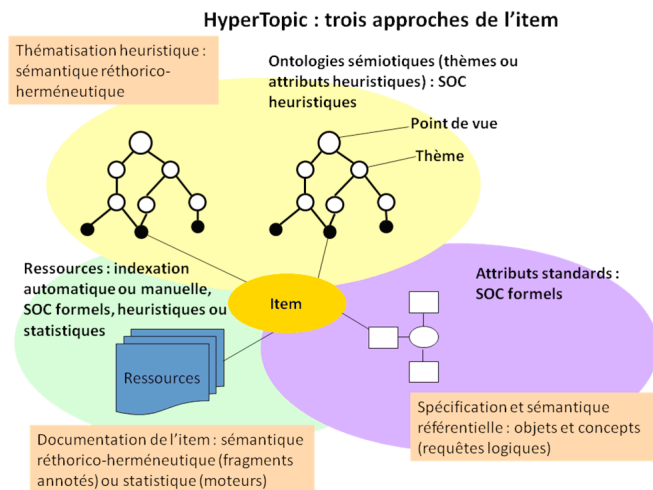


Fig. 4. Les différentes caractérisations de l'item selon HyperTopic

### 2.1.2 Thématisation heuristique via les thèmes

La thématization heuristique de l'item, réalisée à l'aide de la catégorie thème (ou topic), correspond à la caractérisation de l'artefact ou de la situation considérée en fonction de compétences métiers, d'hypothèses théoriques ou d'intérêts particuliers à l'aide d'une langue opérative (Falzon 1989) et d'expressions opératives issues de celle-ci. Les attributs standards sont également issus d'une langue opérative mais celle-ci est entièrement standardisée. Dans la thématization heuristique, les expressions opératives sont plus idiosyncrasiques et peuvent faire l'objet de connotations différentes selon les points de vue. Les thèmes sont organisés selon une arborescence qui traduit des relations de plus ou moins grande généralité (versus spécialisation) selon un point de vue donné. Cette organisation définit une ontologie sémiotique (Zacklad 2005).

La sémantique des thèmes, de nature rhétorico-herméneutique (Rastier 2003), s'appuie sur des relations de proximité sémantique basées sur des relations qui peuvent être proches de celles utilisées pour les attributs standards : relation de généralisation, partie-de, fonctionnelle, de causalité, d'association simple... Mais en l'absence d'une sémantique référentielle forte, ces relations ne permettent pas d'inférences formelles automatiques. Par exemple, les relations de généralisation peuvent ne pas supporter l'héritage des attributs. Pour exploiter les ontologies sémiotiques, on s'appuie le plus souvent sur une combinaison de représentations visuelles représentant les associations et de classement alphabétique. Quand les SOC thématiques s'appuient sur des supports numériques, on

privilégie la navigation de type hypertextuelle et/ou l'utilisation de techniques de visualisation.

### 2.1.3 Documentation via les ressources documentaires

La troisième activité est la documentation de l'item, réalisée à l'aide de la catégorie ressource, qui consiste, quand cela est possible, à associer à l'item des documents ressources ou des fragments de ces documents qui en fournissent une description sous la forme de monographie<sup>3</sup> : fichier texte, image, audio, vidéo, etc. Parce que ces productions sémiotiques sont le plus souvent déposées sur un support pérenne, nous les décrivons en HyperTopic comme relevant de la catégorie des « ressources documentaires ». En informatique, ces productions sémiotiques sont considérées comme constituant de l'information « non structurée », mais cette appellation n'est pas très heureuse, car les textes ou les images contiennent bien sur une structure. Il s'agit en réalité de descriptions qualitatives relevant de la narration ou de formes d'argumentation s'appuyant sur le déploiement d'un raisonnement articulé, basé sur des textes, des images, des graphiques ou la combinaison de ces moyens d'expression.

A la différence des attributs standards ou heuristiques (thèmes) les monographies sont des descriptions porteuses d'un potentiel d'interprétation ouvert dont l'exploitation à fin d'analyse pose deux problèmes. D'une part celui de la sélection de la bonne ressource documentaire qui caractérise un item parmi une collection éventuellement vaste. D'autre part, celui de la sélection, à l'intérieur de la ressource documentaire d'un ou plusieurs fragments fournissant une caractérisation jugée pertinente de l'item considéré en constituant une mosaïque de fragments annotés. Pour résoudre ces deux problèmes il est possible de recourir à une indexation manuelle (ontologies sémiotiques ou basées sur des attributs standards), ou à des index automatiques constitués par les moteurs de recherche.

En ce qui concerne le problème de la sélection des ressources documentaires pertinentes, les spécialistes des sciences de l'information (dans notre contexte les analystes du risque), développent eux-mêmes des SOC spécialisés qui facilitent le classement et l'accès aux documents ou ils utilisent des moteurs de recherche, particulièrement quand les monographies sont textuelles. Il existe deux types de SOC utilisés en documentation : basés sur des attributs standards (nom de l'auteur, dates de parution, etc.) ou heuristiques (thésaurus décrivant les thèmes des

<sup>3</sup> Il peut arriver que l'item soit lui-même un document, un livre, un film, un rapport. Dans ce cas, les monographies documentaires doivent être considérées comme des documents secondaires décrivant des documents primaires, comme c'est le cas de certaines notices accompagnant les ouvrages dans un fond bibliographique (la notice pouvant regrouper attributs standards, thèmes et petits textes descriptifs)

documents, par exemple). Quand des moteurs de recherche sont utilisés, ceux-ci indexent de manière automatique les ressources documentaires et constituent des SOC automatiques, le plus souvent cachés aux utilisateurs, qui permettent, à partir des termes d'une requête adressée au moteur, d'associer certaines monographies considérées comme pertinentes. La sémantique de ces SOC automatiques dépend la technologie utilisée par le moteur (Zaher et al. 2006, 2007) mais est le plus souvent de nature statistique.

En ce qui concerne le problème de la sélection des fragments documentaires pertinents, il faut également les sélectionner puis les annoter pour mémoriser la justification de leur sélection. L'indexation peut se faire de manière manuelle, à l'issue d'un processus de lecture et d'interprétation, en s'appuyant sur des thèmes organisés dans des ontologies sémiotiques représentant différents points de vue. Mais l'indexation peut également bénéficier des associations statistiques effectuées entre des termes par un moteur de recherche. On est dans un cas fréquent où le raisonnement statistique vient alimenter un raisonnement qualitatif. Selon les cas, l'analyste pourra effectuer lui-même une lecture et une sélection rigoureuse des fragments issus des propositions du moteur et la sémantique sera alors pleinement rhétorico-herméneutique, ou il se contentera de catégoriser les propositions qui lui sont faites en les survolant rapidement et la sémantique restera de nature statistique.

#### 2.1.4 Points de vue complémentaires et/ou concurrents

Ces trois activités et plus particulièrement celle de la thématisation heuristique, s'appuient sur une cinquième catégorie essentielle de la métasémiotique Hypertopic, celle de point de vue<sup>4</sup>. C'est la diversité des formes de caractérisation de l'item selon les communautés professionnelles, d'intérêt, de pratiques ou les hypothèses en cours d'examen qui justifie le recours à plusieurs thématisations potentiellement concurrentes de l'item. C'est également la multiplicité des points de vue qui permet de structurer le raisonnement heuristique en permettant, selon les tâches, de faire varier les caractérisations de l'item<sup>5</sup>. Dans les implémentations actuelles d'Hypertopic, les points de vue s'appliquent surtout à des thématisations différentes de l'item. Il pourrait être pertinent qu'ils sélectionnent également certaines formes de documentation, voire des spécifications référentielles différentes.

<sup>4</sup> Nous renvoyons à Bénel (2003) et Cahier (2005) pour une discussion approfondie de cette catégorie.

<sup>5</sup> Dans les exemples canoniques du raisonnement diagnostique en médecine, un problème médical peut être décrit selon le point de vue des symptômes (point de vue des patients) ou selon celui des processus physiopathologiques en cause (point de vue des médecins).

## 2.2 Définition d'un espace de travail numérique pour la gestion de risque combinant différentes sémantiques

La combinaison appropriée des différents types de sémantique est de la première importance dans l'intelligence qualitative associée à la gestion des SIRIX. Dans les architectures développées au sein de l'équipe Tech-CICO et ses partenaires, les différents types de SOC permettent aussi bien de travailler de manière différentielle sur les documents que de travailler à l'intérieur d'un document en exploitant ses différents fragments. La métasémiotique HyperTopic a également donné lieu à un protocole du même nom permettant de faire collaborer différents serveurs et clients comme le montre la figure 5 (Zhou et al.2006, Zacklad et al.2007) : le serveur Argos stocke les items et leur description, le serveur Cassandra permet le traitement de monographies textuelles, les clients Agorae (Fig. 6) et Porphyry permettent la création des attributs et des thèmes et l'organisation et la recherche des items ainsi caractérisés. Agorae caractérise globalement un document ou un fragment documentaire tandis que Porphyry permet à l'utilisateur de délimiter des zones précises à l'intérieur d'une image ou d'un texte selon sa visée.

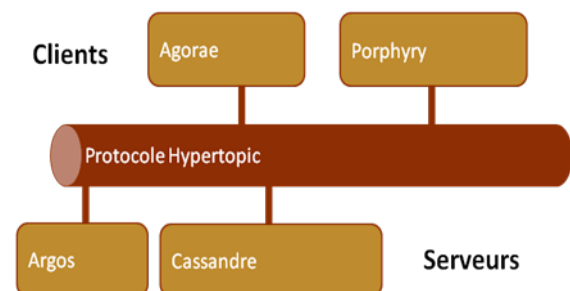


Fig. 5. Serveurs et clients reliés par le protocole HyperTopic.

Ces outils contribuent principalement à la Recherche Ouverte d'Information où à l'étape d'interprétation dans la phase d'exploitation. L'assistance informatique à la phase d'exploitation (rédaction et mise en forme) dépend des contraintes propres à chaque type d'enquête et à chaque domaine de connaissance : cartographie, architecture, art graphiques, administration, recherche en sciences humaines, mécanique, électronique, programmation, médecine, etc. Cela est particulièrement vrai dans la phase de « synthèse rédaction » mais cela peut également intervenir dans l'étape d'annotation où les formats de fichiers peuvent être très divers et nécessiter une adaptation de notre outillage.

Pour fixer les idées, prenons l'exemple d'une crise sanitaire dans laquelle il faut estimer la gravité d'une épidémie potentielle et la nature des mesures prophylactiques à prendre. Les items, documentés dans des petits dossiers informatisés, correspondent aux différents

*événements sanitaires suspects* qui doivent être analysés par les gestionnaires de risque. Ceux-ci sont décrits par des expressions opératives formelles (formalité machinale) correspondant aux standards promus par les autorités sanitaires et recensés dans un fichier (date, localisation, caractéristiques des victimes, analyses médicales disponibles) parfois complétés par des documents monographiques (photo du contexte, des patients, images issues des analyses médicales, comptes-rendus textuels du médecin). Certains d'entre eux ont fait l'objet d'une première analyse et ont été annotés par un médecin épidémiologiste à l'aide d'expressions opératives heuristiques (formalité sémiotique) qui véhiculent les hypothèses sur les causes possibles des événements constatés. Les différents événements peuvent avoir été situés avec plus ou moins de précision sur une carte. Le choix de la carte et les conventions de codage des événements sont une autre manière de caractériser de manière heuristique les événements.

Dans les termes de la métasémiotique HyperTopic, les items sont les événements sanitaires suspects, les expressions opératives formelles correspondent aux attributs standards de l'item, les monographies aux ressources documentaires et les expressions opératives heuristiques aux thèmes de l'item, organisés dans des « ontologies sémiotiques » plus ou moins systématiques selon les investissements définitionnels consentis (Zacklad 2005). A partir de nos environnements informatiques en cours de développement, les gestionnaires de risque peuvent réaliser plusieurs traitements complémentaires.

- Navigation libre entre les items à partir des attributs standards ou des thèmes permettant d'enrichir la description de l'item par le biais de la création de nouveaux thèmes inscrits dans un point de vue existant ou relevant d'un nouveau point de vue à créer (les points de vue sont souvent associés à des compétences, des hypothèses, des communautés professionnelles) : client Agorae/serveur Argos.
- Exploitation des monographies constituées par des images et/ou des photos en annotant certaines zones de l'image jugées significatives en regard d'une hypothèse en cours d'examen à l'aide d'un thème nouveau ou issu d'une ontologie sémiotique existante : client Porphyry/serveur Argos.
- Exploitation des monographies textuelles en constituant certaines expressions récurrentes en nouveaux thèmes venant à l'appui d'une hypothèse en cours d'examen. Possibilité de rechercher automatiquement les occurrences correspondant à ces thèmes dans l'ensemble du corpus associé aux items (dans le cadre d'un raisonnement statistique) et de mettre en évidence les fragments textuels d'où ils sont issus : serveurs Cassandre et Argos/clients Agorae et Porphyry.
- Mise en relation des items caractérisés par certains thèmes ou attributs standards à partir de leur description en langue opérative formelle ou heuristique (attributs standards et thèmes des

ontologies sémiotiques) ou issue des monographies imagées ou textuelles associées : fonctionnalités du client Agorae à développer.

- Recherche sur internet ou des bases de données de cas similaires sur la base des cooccurrences statistiques entre les thèmes heuristiques et d'autres documents : interface et paramétrage d'un moteur de recherche à développer.

Ces différents traitements, qui correspondent à la Recherche Ouverte d'Information, permettent d'aboutir à un premier classement des *événements sanitaires suspects* (items) et de la documentation associée. Ils correspondent également à la l'étape d'interprétation de la phase d'exploitation. C'est sur la base de ces analyses qu'une synthèse argumentée pourra être réalisée donnant lieu à une diffusion ou à une mise à disposition raisonnée aux acteurs concernés (résolution du problème). La figure 6 présente une copie d'écran du client Agorae issue d'un autre contexte dans laquelle on voit apparaître la fiche descriptive d'un item, en l'occurrence une situation de risque de pollution, définie par des thèmes heuristiques issus de points de vue prédéfinis ou d'une caractérisation spontanée correspondant à des « associations d'idées » inédites (représentation folksonomique).

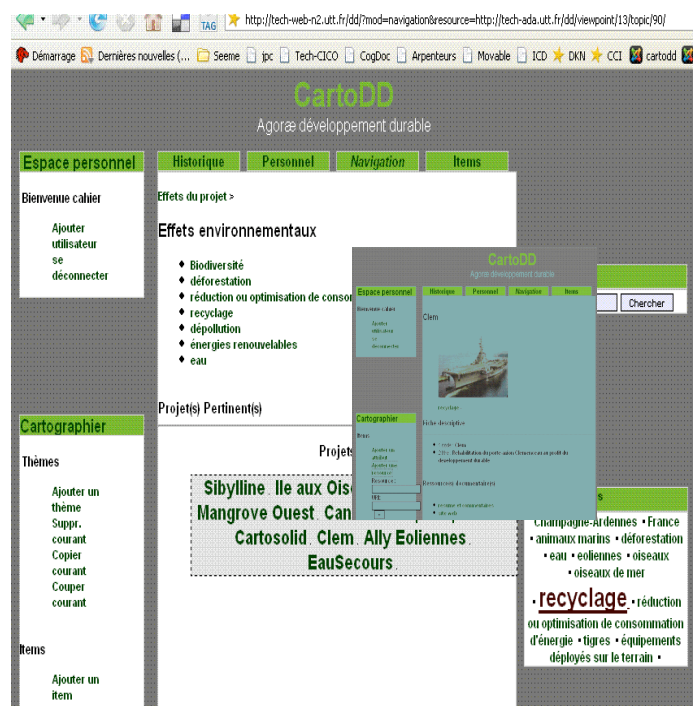


Fig. 6. Copie d'écran du client Agorae

### 3. Conclusion

Contrairement à d'autres formes de rationalité qui équipent la décision à partir de calculs de nature statistiques nombreux et sophistiqués, la rationalité documentaire privilégie la constitution d'environnements sémiotiques facilitant la conception continue et répétitive de solutions nouvelles à partir de différents

environnements de gestion de corpus où les fragments documentaires servent à la fois de ressource pour l'évaluation et pour la conception. Bien sur, la rationalité documentaire met à profit les calculs statistiques effectués par les moteurs de recherche pour fournir de nouveaux documents candidats. Mais elle vient surtout fournir des moyens permettant l'évaluation de la pertinence des résultats, souvent en très grand nombre, fournis par le moteur.

Pour être efficaces, les SOC doivent être enrichis au fur et à mesure des sessions par les actions répétées du même utilisateur ou par le cumul des usages distribués d'un collectif. Les SOC heuristiques peuvent bénéficier de différentes méthodes de visualisation, tels que les représentations arborescentes, les cartes folksonomiques, les représentations de graphes. La rationalité documentaire s'inscrit également naturellement dans les environnements de travail coopératif dans la mesure où, dans de nombreux cas, les interactions entre acteurs sont médiatisées par des messages documentarisés. Elle bénéficie également des technologies de gestion de réseaux sociaux qui facilitent la documentation des caractéristiques des acteurs en présence. Dans de nombreuses SIRIX, l'intelligence qualitative vient ainsi apporter un complément indispensable à d'autres techniques de traitement de données, voire peut s'avérer suffisante pour assister la gestion de bout en bout des crises. Une analyse des complémentarités entre les différentes formes d'intelligence offertes par les technologies actuelles de traitement de l'information reste à réaliser pour aider responsables et décideurs à sélectionner les environnements les plus pertinents selon les SIRIX.

## Références

- [1] Bachimont, B. (2004). Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle. Habilitation à diriger des recherches de l'Université Technologique de Compiègne.
- [2] 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.
- [3] Cahier, J.-P. (2005), Ontologies sémiotique pour le Web socio sémantique: étude de la gestion coopérative des connaissances avec des cartes hypertextuelles, Thèse en informatique, Université de Technologie de Troyes, déc. 2005.
- [4] Clancey W., Heuristic Classification. Artificial Intelligence Journal, 27, pp. 289-350, 1985.
- [5] Dewey, J. (1938). Logique, la théorie de l'enquête, trad. Gerard Deledalle, col. L'interrogation philosophique, PUF, 1938 (re-éd.1993).
- [6] Falzon, P. (1989). Ergonomie cognitive du dialogue, PUG, Grenoble.
- [7] Goody, J. (1979). La raison graphique, la domestication de la pensée sauvage. Les Editions de Minuit, Paris.
- [8] Lemoigne, J.L. (1997), La décision est modélisation dynamique, in résumé des Rencontres MCX 6 - Poitiers, Futuroscope, 9 et 10 Juin 1997, <http://www.mcxapc.org/docs/dossiermcx/poitiers/poitiers1.htm> (consulté le 04/01/2007).
- [9] Pédaque, R. T. (2006) Le document à la lumière du numérique, Caen : C&F Editions.
- [10] Rabardel, P. (1995). Les hommes et les technologies. Approche cognitive des instruments contemporains. Paris: Colin.
- [11] 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](http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html)
- [12] Salembier, P. (2007), Analyse, modélisation et instrumentation des activités coopératives situées, Habilitation à diriger des recherches de l'Université de Nancy.
- [13] Simon, H. A. (1969). The sciences of the artificial. Cambridge, MA: MIT Press (seconde édition, 1981).
- [14] 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.
- [15] Manuel Zacklad: Documentarisation Processes in Documents for Action (DofA): The Status of Annotations and Associated Cooperation Technologies. Computer Supported Cooperative Work 15(2-3): 205-228 (2006)
- [16] 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>
- [17] 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.
- [18] 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 : 20-35.
- [19] Zacklad, M. (2007d), Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information, in actes de la Conférence CAIS/ACSI 2007, Montréal, [http://www.cais-acsi.ca/search\\_fr.asp?year=2007](http://www.cais-acsi.ca/search_fr.asp?year=2007)
- [20] Zacklad, M. (2007e), Démarche de caractérisation des collaboratoires en SHS et apports du web socio-sémantique, présentation au séminaire « Economie politique de l'édition numérique » de l'ENS Lyon

animé par G. Wormser et P. Mounier (prochainement en ligne).

- [21]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.
- [22]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.
- [23]Zaher, H., Cahier, J.-P., Zacklad, M. (2007) Une plateforme de recherche ouverte d'information : Agoræ et Hypertopic, In Imed Saleh, Khaled Ghedira, Balhassen Badreddine, Nasresddine Bouhai et Bernhard Rieder (eds), actes de la conférence Hypertexte et Hypermédia, «Collaborer, échanger, inventer : expériences de réseaux» H2PTM'2007, France, Hermès Science, pp. 27-38, ISBN 978-2-7462-1891-8.
- [24]Zhou C., Lejeune C. & Bénel A. (2006). Towards a standard protocol for community-driven organizations of knowledge, Proceedings of the thirteenth international conference on Concurrent Engineering, Frontiers in Artificial Intelligence and Applications, vol. 143. Amsterdam: IOS Press, 2006. p.438-449. Disponible sur : [http://benel.techcico.fr/publi/zhou\\_CE\\_06.pdf](http://benel.techcico.fr/publi/zhou_CE_06.pdf)