
Utilisation des ontologies pour le partage de l'information dans le domaine de l'eau

Lylia Abrouk* — **Abdelkader Gouaïch****

* *LE2I, Laboratoire Electronique Informatique et Image
UMR CNRS 5158, Université de Bourgogne
21000 Dijon, France*

lylia.abrouk@u-bourgogne.fr

** *LIRMM, Laboratoire d'Informatique, de Robotique
et de Microelectronique de Montpellier
161 rue Ada, 34392 Montpellier Cedex 5*

gouaich@lirmm.fr

RÉSUMÉ. La description des ressources dans une communauté doit suivre un vocabulaire contrôlé. Il s'agit d'un ensemble de termes définis par un groupe afin de pouvoir labelliser des contenus ou de décrire un document. Dans cet article nous présentons l'utilité des ontologies dans un système d'information dédié au domaine de l'eau, le SEMIDE (Système Euro-Méditerranéen d'Information sur les savoir-faire dans le Domaine de l'Eau) est un instrument d'échange d'information et de connaissances entre tous les pays du Partenariat Euro Méditerranéen.

ABSTRACT. The description of the resources in a community must follow a controlled vocabulary. It's a set of terms defined by a group in order to be able to stamp the contents or to describe a document. In this article we present the utility of ontologies in an information system dedicated to the water domain, the EMWIS (Euro-Mediterranean Information System on the know-how in the Water sector) is an instrument of exchange of information and knowledge between all the countries of the Euro-Mediterranean Partnership.

MOTS-CLÉS : ontologies, système d'informations, enrichissement.

KEYWORDS: ontologies, information system, enrichment

1. Introduction

La rapidité de l'évolution de la masse d'informations dans tous les domaines a généré un besoin d'organisation et de structuration des contenus. Les ontologies servent à la représentation des données échangées dans un domaine particulier afin de faciliter la communication interne au système informatique et externe entre les différents acteurs du domaine. Leur utilisation peut varier de la représentation des données à la recherche d'informations.

Les informations sur l'eau sont nombreuses, d'ordre qualitatif ou quantitatif et sont dispersées chez de très nombreux acteurs. Que ce soit au niveau local, national ou européen, le citoyen, le gestionnaire, le chercheur ou le décideur a besoin d'accéder aux informations disponibles. L'information disponible dans ce domaine n'existe que de façon fragmentaire, dispersée et hétérogène, il est apparu nécessaire d'engager un effort de rationalisation et de lisibilité pour la rendre facilement accessible et utilisable. C'est pourquoi, il a été décidé d'étudier les modalités de mise en œuvre d'un système d'information qui, à travers l'utilisation des moyens modernes de communication, permettrait de mettre en réseau les sources existantes : le SEMIDE¹ (Système Euro Méditerranéen d'Information sur les savoir-faire dans le Domaine de l'Eau).

Dans tous les domaines de connaissances, le Web sémantique vise à fournir un médium d'échange d'information et de savoir en partageant les ressources. Dans notre travail, nous avons cherché à appliquer les technologies du Web sémantique afin d'offrir une plateforme technique réalisant les objectifs du SEMIDE.

Le SEMIDE étant un système euro méditerranéen, les acteurs sont de pays différents et les ressources sont spécialisées et multilingues. Par conséquent, pour pouvoir répondre au problème du partage d'informations, nous devons pouvoir utiliser un vocabulaire commun du domaine (spécialisé au domaine de l'eau), indépendant de la langue afin de partager la connaissance entre les différents acteurs de la communauté. Les ontologies sont un des concepts de base du Web sémantique, elles représentent un vocabulaire défini par une communauté afin de représenter un domaine. Elles servent pour la structuration et l'exploitation des métadonnées.

La suite de cet article est structurée de la façon suivante : la section 2 décrit des travaux qui traitent de la construction et l'enrichissement des ontologies, l'ontologie définie pour le Semide et l'approche d'enrichissement proposée sont décrites dans la section 3, enfin nous concluons et présentons quelques perspectives dans la section 4.

2. Utilisation des ontologies : techniques de construction d'enrichissement

Plusieurs travaux se sont intéressés à la construction de ces ontologies dans différents plans : *(i)* extraction de termes représentatifs dans un domaine spécialisé, *(ii)* identification de relations lexicales entre les termes, *(iii)* placement des nouveaux

1. www.semide.org

termes dans une ontologie existante. On verra que dans ces travaux, le terme ontologie prend plusieurs sens, que l'on parle de thésaurus, taxonomies ou plus généralement de vocabulaire contrôlé.

2.1. Méthodes de construction

Les approches de construction d'ontologies sont basées généralement sur un corpus de documents, et font intervenir des experts pour le placement des termes dans l'ontologie ou bien afin de valider le résultat.

Le système *Adaptiva* [BRE 02] est l'implémentation d'une méthode centrée sur les utilisateurs, ces derniers doivent être capables de : créer des ontologies, valider des phrases qui expriment une relation entre deux termes et nommer la relation. la construction de l'ontologie consiste en l'apprentissage de la relation d'hyponymie (*is-a*) par le système et la validation de l'ontologie par l'utilisateur. D'autres approches basées essentiellement sur l'analyse linguistique, Sanderson et Croft [SAN 99] utilisent la co-occurrence de deux termes x, y dans les documents du corpus afin de générer une relation de spécialisation entre les deux termes. Ferret [FER 02] aborde la construction de hiérarchies à partir d'un corpus en se basant sur le contexte des termes. Le système trouve des relations sémantiques entre les termes et plus particulièrement les relations d'hyperonymie afin de construire une hiérarchie. L'approche se base sur le sens d'un terme qui se traduit par un ensemble de contextes associés au terme dans le corpus, le paragraphe étant le délimiteur du contexte ou bien la délimitation se fait en utilisant un outil.

2.2. Méthodes d'enrichissement

L'approche de Faatz et Steinmetz [FAA 02] s'applique sur une ontologie déjà existante afin de l'enrichir en se focalisant uniquement sur la relation généralisation/spécialisation. Les auteurs posent le problème de l'enrichissement d'ontologie comme un problème d'optimisation de paramètres. Ces paramètres indiquent combien un concept co-occure avec d'autres concepts candidats par rapport à des règles prédéfinies dans une collection de textes. L'idée est de comparer les mots qui co-occurrent dans le corpus avec l'hypothèse qu'il existe une mesure calculant la distance sémantique entre deux concepts dans l'ontologie. Amardeilh [AMA 05b] enrichit une ontologie à partir de documents textuels en s'appuyant sur l'analyseur linguistique Insight Discoverer Extractor (IDE) [AMA 05a]. L'analyseur produit en sortie un arbre conceptuel étiqueté où chaque nœud porte une étiquette sémantique attribuée à l'unité textuelle extraite en fonction du domaine traité.

Une autre approche d'enrichissement consiste à ne s'intéresser qu'au contexte d'un concept. Harabagiu [HAR 99] a noté que les réseaux de type Wordnet présentaient un manque, dans la mesure où il existe une absence de relation sémantique (appartenance de mots à un même thème). Afin de combler ce manque, Harabagiu [HAR 02] propose

une approche afin d'extraire des relations thématiques de Wordnet en se basant sur les définitions associées aux concepts. Ceci consiste à construire pour chaque concept la liste de mots qui lui sont relatifs. Cette approche permet à un utilisateur de manipuler tout le contexte d'un concept, ce qui peut être utile à la recherche d'informations. Par contre cette méthode ne donne pas une relation explicite entre deux concepts.

La hiérarchisation de concepts par Cimiano [CIM 03] basée sur l'analyse formelle des concepts (AFC) vise à créer une taxonomie de manière automatique qui peut être appliquée sur des domaines différents. L'idée est d'extraire automatiquement un ordre partiel entre les concepts. Les connaissances sont représentées par une matrice objets-verbs (par exemple un appartement peut être acheté). Ensuite ces informations sont représentées par un treillis, où l'on retrouve un ordre partiel entre les nœuds verbes-concepts. La hiérarchie entre les concepts est extraite directement du treillis. Latiri [LAT 05] présente une approche utilisant aussi l'AFC afin de construire une ontologie. Ceci se fait en passant des relations du treillis aux relations ontologiques (relation père/fils, relation d'héritage).

Enfin, il existe des travaux utilisant des ressources externes. Simon et al [SIM 03] proposent une approche se basant sur un thésaurus et sur des techniques du *Text Mining* afin d'enrichir l'ontologie de manière semi-automatique en laissant l'utilisateur sélectionner tout ou une partie des concepts proposés. D'autres ressources utilisant un vocabulaire contrôlé peuvent être utilisées pour l'enrichissement d'une ontologie. Parekh [PAR 04] décrit comment enrichir une ontologie en se basant sur des textes, glossaires et dictionnaires du domaine afin de trouver des groupes de concepts/termes reliés entre eux.

L'analyse des textes pour la construction d'une ontologie passe par une phase déterminante qui est la constitution d'un corpus. Or cette tâche n'est pas facile dans le sens où on ne retrouve pas forcément des textes très représentatifs pour n'extraire que les termes importants du domaine. Cette analyse génère généralement trop de résultats (bruit) ou pas assez (silence), ce qui nécessite un travail fastidieux et complémentaire par l'expert.

Évidemment, il n'est pas réaliste de penser qu'on puisse se passer du travail humain, car il est impossible de créer une ontologie de manière complètement automatique ou du moins jusqu'à présent. Bien que la méthodologie dépende principalement du besoin, on a déjà pu tirer les constats suivants :

- 1) le traitement linguistique est essentiel, même s'il ne représente qu'une petite étape dans le processus ;
- 2) le choix d'un corpus représentatif est une tâche difficile qui ne dépendra que d'un petit nombre d'experts, si ce n'est d'un seul ;
- 3) enfin, l'ontologie concerne un domaine et par ce fait les acteurs du domaine. Une collaboration de ces derniers pourrait être une approche intéressante.

Dans la section suivante, nous présentons l'ontologie du Semide et l'approche d'enrichissement proposée.

3. Les ontologies dans le domaine de l'eau

L'ingénierie ontologique consiste en la recherche de concepts généraux, réutilisables, partageables et durables pour construire un modèle de connaissances capable d'aider des personnes à résoudre des problèmes [MIZ 04]. L'utilisation des ontologies dans différents secteurs (biomédical, automobile...) a connu une expansion due essentiellement à un souhait de s'orienter vers une vision Web sémantique. Le but commun, dans ces différents secteurs, est d'optimiser la représentation des connaissances et leurs exploitation.

Le partage de l'information au niveau du SEMIDE ne peut se faire que si les acteurs d'un tel système partagent des concepts communs, d'un côté pour décrire le domaine et d'un autre côté pour décrire l'information, et ainsi faciliter son utilisation.

3.1. L'ontologie du Semide

Nous nous sommes intéressés dans le cadre de notre travail à mettre en œuvre une ontologie pour le SEMIDE. Nous avons rapproché notre ontologie d'un thésaurus dans les relations existantes, cependant, ces relations peuvent évoluer dans le temps. L'ontologie existante est basée sur les termes et les relations du thésaurus de l'office Internationale de l'Eau (OIEau²). Ceci a été réalisé principalement afin de :

- 1) partager le savoir commun entre la communauté ou les agents logiciels : par exemple chaque *PFN*³ possède des informations sur l'eau et nous souhaitons partager et publier les mêmes termes. Le système sera capable d'extraire et de fusionner les informations des différents *PFN* fournisseurs, et l'utiliser pour répondre à la requête ;
- 2) pouvoir utiliser des ontologies d'autres domaines afin d'enrichir notre ontologie (notre vocabulaire) : par exemple si on souhaite ajouter le terme *informatique* dans notre ontologie, on peut réutiliser l'ontologie de ce domaine pour enrichir la nôtre ;
- 3) faire en sorte de rendre le domaine explicite.

L'ontologie du SEMIDE est décrite par les concepts du domaine et les relations qui les relient. La figure 1 illustre un extrait cette ontologie :

- un nœud représente un concept, représenté par un cercle dans la figure (par exemple le concept C_1) ;
- les concepts sont reliés par des arcs orientés définissant la relation de spécialisation/généralisation, ici le concept C_2 est une spécialisation du concept C_1 ;
- à chaque concept sont reliés des termes, un terme pouvant avoir des termes synonymes ;
- les termes sont dans différentes langues, actuellement arabe, anglais et français et bientôt espagnol. le terme *Hydraulique agricole* est associé au concept C_1 .

2. <http://www.oieau.fr/>

3. Point Focal National

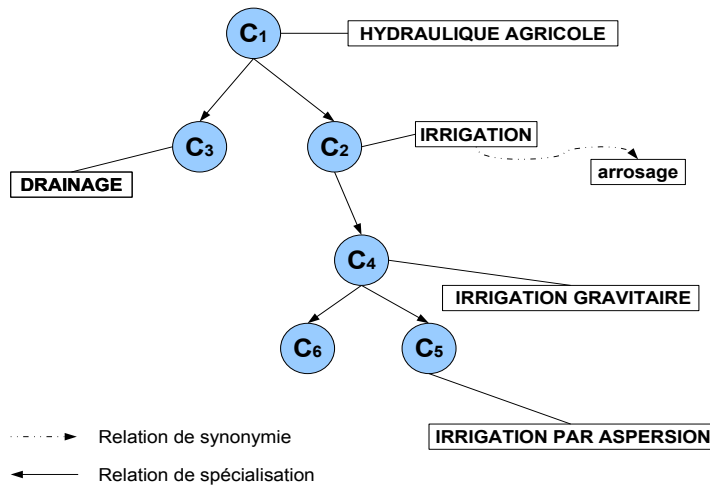


Figure 1. Extrait de l'ontologie du Semide

3.2. Une approche pour l'enrichissement de l'ontologie

Le traitement des ontologies est un domaine très actif où on cherche à automatiser au mieux ce processus car il n'existe pas d'approche unique. La construction/enrichissement des ontologies dépend dans un premier lieu des besoins des acteurs ainsi que de l'existant. L'approche que nous proposons est adaptée au contexte du SEMIDE, elle se base essentiellement sur la phase de recherche effectuée par les utilisateurs. Le SEMIDE souhaitant faire participer les différents acteurs du système pour la mise en œuvre de l'ontologie du domaine, nous avons traduit ceci par la saisie des utilisateurs de termes n'appartenant pas à l'ontologie afin d'effectuer une recherche sur la base documentaire. Cette recherche est guidée par l'ontologie du domaine, tout en laissant la liberté à l'utilisateur de saisir d'autres mots clés s'il considère que les termes existants ne suffisent pas pour constituer la requête. Notre approche consiste à exploiter ces termes, trouver les relations lorsqu'elles existent avec les termes de l'ontologie.

L'idée de notre approche est basée sur l'exploitation des requêtes des utilisateurs afin d'enrichir l'ontologie du SEMIDE. La figure 2 donne un aperçu des différentes étapes pour l'enrichissement de l'ontologie.

A partir du moteur de recherche, les utilisateurs soumettent leurs requêtes. Le moteur de recherche renvoie des réponses et l'utilisateur peut modifier sa requête (ici la requête est un ensemble de termes issus ou non de l'ontologie). Toutes ces requêtes sont sauvegardées dans un entrepôt de requêtes. Le processus complet pour chaque utilisateur constitue une session de recherche.

Les différentes étapes sont les suivantes :

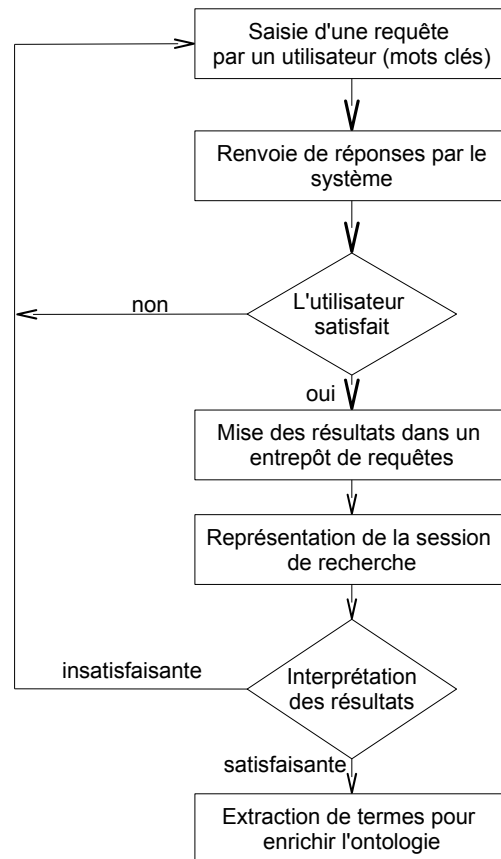


Figure 2. *Enrichissement à partir de la phase d'interrogation*

1) identification des sessions de recherche par utilisateur, c'est-à-dire des ensembles de requêtes qui sont affinées par les utilisateurs qui y rajoutent ou suppriment des termes. Nous utilisons le moteur de recherche du SEMIDE ;

2) représentation des sessions de recherche ;

3) ajout de nouveaux termes apparaissant dans les requêtes. Ces termes sont utilisés par les experts dans leur recherche et n'existent pas au préalable dans l'ontologie ;

4) enrichissement de l'ontologie avec ces termes.

3.2.1. *représentation des sessions de recherche*

Afin de représenter l'association (requête/documents), nous utilisons le treillis de concepts à partir d'un entrepôt de requêtes par utilisateur. La représentation des couples (termes de requêtes/documents réponses) sert à extraire les termes non présents dans l'ontologie, mais répondant à la recherche (besoin) des utilisateurs. Les treillis de concepts sont utilisés dans la recherche d'informations pour affiner ou généraliser la requête de l'utilisateur [MES 04]. L'utilisation des treillis est ici différente, l'utilisateur pose sa requête, l'affine ou la généralise en fonction des réponses et de sa recherche. Les différentes requêtes constituent une session de recherche représentée par un treillis de concepts.

Nous nous basons sur l'algorithme de Bordat, non incrémental afin de construire le treillis de Galois et ceci pour chaque session de recherche. Cet algorithme est adapté dans le cas où le diagramme de Hasse⁴ est à générer [NGU 05].

3.2.2. *Construction de la hiérarchie*

La structure du treillis est utilisée afin d'extraire les relations candidates entre les termes. nous ne nous intéresserons dans le treillis qu'aux termes. Nous construisons la hiérarchie des termes afin de ne garder qu'une seule occurrence des termes. Nous partons de l'ensemble des termes du plus haut niveau et nous éliminons les occurrences de chaque élément dans les niveaux inférieurs.

On définit l'ordre partiel P_t comme étant la restriction du treillis aux nœuds qui contiennent les éléments maximaux de t .

La première approche que nous avons réalisée afin d'exploiter ce treillis rejoignait celle de Sanderson et Croft [SAN 99], laquelle se base sur les co-occurrences des termes pour déduire que x et un terme générique de y si les documents où y apparaît sont un sous-ensemble de ceux où x apparaît.

Cependant, dans la majorité des corpus que nous avons étudiés, cette inclusion ensembliste n'est jamais vérifiée, car le terme y apparaît régulièrement dans des documents où x n'apparaît pas, ce qui nous empêche d'utiliser cette propriété pour conclure sur le type de la relation entre x et y .

Cette hypothèse peut toutefois s'avérer pour certains corpus à un instant donné, mais les corpus de documents à exploiter sont en continuelle évolution, et il s'avère que la propriété ne reste jamais longtemps vérifiée. D'autre part, cette hypothèse ne peut pas parfaitement s'appliquer sur notre approche car le treillis ne prend pas en compte les documents qui répondent uniquement au terme y .

4. Le diagramme de Hasse de (G, \leq) est une représentation graphique qui contient toutes les informations concernant la relation d'ordre représentée.

Le treillis construit à partir des sessions de recherche définit bien une relation thématique entre les deux termes mais qui ne se traduit pas forcément en une relation de généralisation. Pour cette raison, la relation de hiérarchie entre un terme x et un terme y dans le treillis est transformé en relation ontologique *related to*.

Nous définissons la relation "related to" comme la relation déduite à partir du treillis décrivant deux termes qui sont liés par inclusion thématique. La thématique de y est incluse dans celle de x si x est ancêtre de y dans la restriction du treillis.

3.2.3. Exemple

Session de recherche

Illustrons la constitution d'une session de recherche par un exemple. Un utilisateur U effectue ses recherches selon les étapes suivantes :

1) dans un premier temps, U saisie la requête *hydrogéologie*, il obtient comme réponse les documents D_1, D_2, \dots, D_6 ;

2) U essaie d'affiner sa requête en recherchant différentes combinaisons (*hydrogéologie, forage*), (*hydrogéologie, alimentation de nappe*), ou bien (*hydrogéologie, aquifère*) ;

3) l'utilisateur peut ne pas être satisfait et ajouter un terme qui n'est pas dans l'ontologie, comme par exemple la combinaison (*hydrogéologie, aquifère, barrage souterrain*) ;

4) l'utilisateur termine sa recherche quand il est satisfait du résultat, ou bien il refait une autre recherche s'il combine un ensemble de termes disjoints des précédents.

Représentation de requête

Notre approche consiste à construire un entrepôt de requêtes à partir des recherches des utilisateurs. La représentation des réponses se fera par utilisateur. Le tableau 1 représente l'exemple précédent de correspondance entre 6 documents réponses des 5 termes $\{t_1, t_2, t_3, t_4, t_5\}$.

Les termes sont les suivants : (t_1 , hydrogéologie), (t_2 , aquifère), (t_3 , forage), (t_4 , alimentation de nappe), (t_5 , barrage souterrain).

Nous avons ici, par exemple, la correspondance de Galois :

$$- O_1 = \{D_3, D_4\} \Rightarrow f(O_1) = \{t_1, t_3\}$$

$$- A_1 = \{t_1, t_3\} \Rightarrow g(A_1) = \{D_1, D_3, D_4\}$$

Dans cet exemple, on a le couple ($\{t_3\}, \{D_1, D_3, D_4\}$) qui signifie que le résultat de la requête avec le terme t_3 donne pour réponse les documents D_1, D_3 et D_4 .

Construction de la hiérarchie

| | t_1 | t_2 | t_3 | t_4 | t_5 |
|-------|-------|-------|-------|-------|-------|
| D_1 | X | | X | | X |
| D_2 | X | X | | X | |
| D_3 | X | | X | X | |
| D_4 | X | | X | | |
| D_5 | X | | | X | |
| D_6 | X | X | | | |

Tableau 1. Exemple de correspondance

Nous illustrons le résultat de l'algorithme sur l'exemple du tableau 1

$$\begin{aligned}
c &= (\emptyset, \{t_1, t_2, t_3, t_4, t_5\}). \\
\delta_c &= \max\{f_c(D_1), f_c(D_2), f_c(D_3), f_c(D_4), f_c(D_5), f_c(D_6)\} \\
&= \max\{\{t_1, t_3, t_5\}, \{t_1, t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_3\}, \{t_1, t_4\}, \{t_1, t_2\}\} \\
&= \max\{\{t_1, t_3, t_5\}, \{t_1, t_2, t_4\}, \{t_1, t_3, t_4\}\}
\end{aligned}$$

Dans ce cas les successeurs directs de c sont :

$$\begin{aligned}
c_1 &= (\{D_1\}, \{t_1, t_3, t_5\}) \\
c_2 &= (\{D_2\}, \{t_1, t_2, t_4\}) \\
c_3 &= (\{D_3\}, \{t_1, t_3, t_4\})
\end{aligned}$$

De la même façon, on calcule les successeurs directs de c_1 :

$$\begin{aligned}
\delta_{c_1} &= \max\{f_{c_1}(D_2), f_{c_1}(D_3), f_{c_1}(D_4), f_{c_1}(D_5), f_{c_1}(D_6)\} = \max\{\{t_1\}, \{t_1, t_3\}, \{t_1, t_3\}, \{t_1\}, \{t_1\}\} \\
&= \max\{\{t_1, t_3\}, \{t_1, t_3\}\}
\end{aligned}$$

Les successeurs directs de c_1 sont :

$$c_4 = (\{D_1, D_4, D_3\}, \{t_1, t_3\})$$

La suite du calcul des successeurs directs se fait de la même manière. Le résultat de l'exemple du tableau 1 est le suivant (figure 3)

Le résultat de la hiérarchisation des concepts de l'exemple est illustré dans la figure 4.

4. Conclusion et perspectives

Cet article a décrit l'utilisation des ontologies dans le milieu industriel et plus particulièrement au sein d'une communauté multilingue traitant du partage de l'information dans le domaine de l'eau. La volume des ressources étant en continuelle évolution, l'ontologie doit s'enrichir tout en répondant au besoins des acteurs du domaine.

L'extraction de nouveaux termes issus de la recherche utilisateur consiste à combler le manque de l'ontologie, représentée par le besoin des utilisateurs. La structure des treillis de concepts est utilisée afin de représenter les sessions de recherche

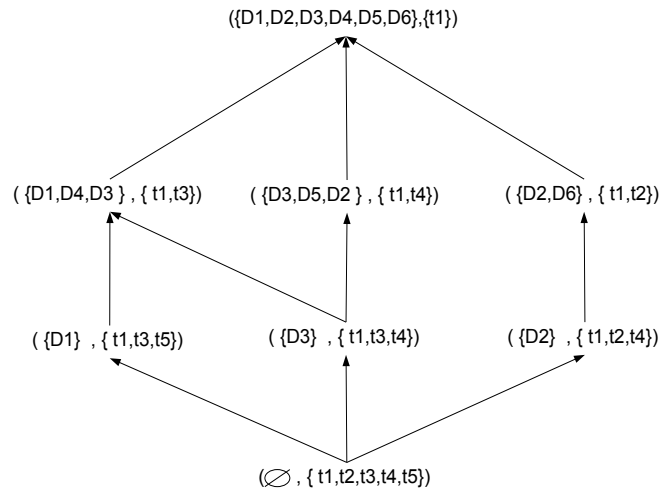


Figure 3. Treillis de Galois

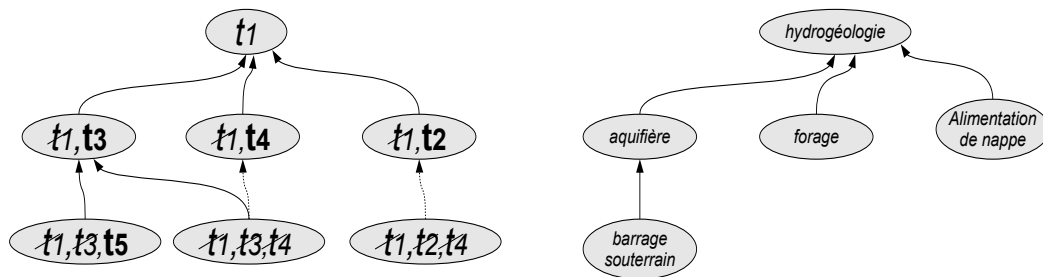


Figure 4. La hiérarchie des concepts

par utilisateur et extraire la hiérarchie des concepts. Une perspective intéressante à cette représentation serait de se baser sur les treillis construits à partir des sessions de recherche, et ceci afin de construire des classes d'utilisateurs par rapport à leur recherche. Effectivement, une ontologie ne peut pas satisfaire tous les utilisateurs, surtout s'ils ont des niveaux de connaissance différents. D'un autre côté, plus une ontologie sera grande et plus elle perdra du sens.

La classification de treillis (session de recherche) permettrait d'enrichir l'ontologie par classe d'utilisateur et ainsi répondre aux besoins des acteurs en fonction de leur niveau.

5. Bibliographie

- [AMA 05a] AMARDEILH F., LAUBLET P., MINEL J. L., « Document Annotation and Ontology Population from Linguistic Extraction », *Proceedings of Third International Conference on Knowledge Capture*, 2005.
- [AMA 05b] AMARDEILH F., LAUBLET P., MINEL J., « Annotation documentaire et peuplement d'ontologies à partir d'extractions linguistiques », *Proceedings of IC 2005*, Août 2005.
- [BRE 02] BREWSTER C., CIRAVEGNA F., WILKS Y., « User-Centred Ontology Learning for Knowledge Management », *Proceedings of NLDB '02 : 6th International Conference on Applications of Natural Language to Information Systems-Revised Papers*, London, UK, 2002, Springer-Verlag, p. 203–207.
- [CIM 03] CIMIANO P., STAAB S., TANE J., « Automatic Acquisition of Taxonomies : FCA meets NLP », *Proceedings of the ECML/PKDD Workshop on Adaptive Text Extraction and Mining*, 2003.
- [FAA 02] FAATZ A., STEINMETZ R., « Ontology enrichment with texts from the WWW », *Proceedings of Semantic Web Mining 2nd Workshop at ECML/PKDD-2002*, Août 2002.
- [FER 02] FERRET O., FLUHR C., HANS F., SIMONI J., « Building domain specific lexical hierarchies from corpora », *Proceedings of LREC 2002*, Las Palmas, Espagne, May 29-31 2002.
- [HAR 99] HARABAGIU S., MILLER G., MOLDOVAN D., « WordNet 2 - A Morphologically and Semantically Enhanced Resource », *Proceedings of SIGLEX Workshop*, 1999, p. 1–8.
- [HAR 02] HARABAGIU S., MAIORANO S., « Multi-Document Summarization with GISTexter », *Proceedings of Third LREC Conference 2002 (LREC 2002)*, 2002.
- [LAT 05] LATIRI C., MTIR M., BENYAHIA S., « Méthode de construction d'ontologie de termes à partir du treillis de l'iceberg de Galois », *Proceedings of Extraction et Gestion des Connaissances (EGC2005)*, Paris, 19-21 Janvier 2005, p. 365-376.
- [MES 04] MESSAI N., « Treillis de Galois et ontologies de domaine pour la classification et la recherche de sources de données génomiques », Juin 2004.
- [MIZ 04] MIZOGUCHI R., « Le rôle de l'ingénierie ontologique dans le domaine des EIAH. », *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, vol. 11, 2004, p. 231–246.
- [NGU 05] NGUIFO E. M., NJIWOUA P., « Treillis de concepts et classification supervisée », *Technique et Science Informatiques*, vol. 24, n° 4, 2005, p. 449-488.
- [PAR 04] PAREKH V., GWO J., FININ T., « Mining Domain Specific Texts and Glossaries to Evaluate and Enrich Domain Ontologies », *Proceedings of International Conference of Information and Knowledge Engineering*, 2004.
- [SAN 99] SANDERSON M., CROFT W. B., « Deriving Concept Hierarchies from Text », *Research and Development in Information Retrieval*, 1999, p. 206-213.
- [SIM 03] SIMON L., DESMONTILS E., JACQUIN C., « Utilisation de techniques d'enrichissement d'ontologies pour améliorer le processus d'indexation structurée », *Proceedings of AFIA 2003*, France, juillet 2003.