

Les treillis de Galois pour l'organisation et la gestion des connaissances

Laszlo Szathmary, Amedeo Napoli

LORIA/Inria-Lorraine, 615 rue du Jardin Botanique, BP 101
F-54600, Vandœuvre-lès-Nancy, France
{szathmar,napoli}@loria.fr

Résumé. Dans cet article, nous étudions la construction de treillis de Galois à partir de différentes sources d'information (par exemple des documents du Web ou des notices bibliographiques) afin d'organiser les connaissances qui peuvent en être extraites. L'organisation en treillis qui en résulte peut alors être utilisée pour satisfaire un certain nombre de buts, comme par exemple la gestion de la connaissance dans une organisation, la recherche documentaire sur le Web, etc. En outre, la classification par treillis peut également tirer parti d'une ontologie des propriétés du domaine considéré. Notre objectif global est de mettre en place un processus de classification par treillis pour enrichir une ontologie qui à son tour permet de guider le processus de découverte de connaissances dans les données.

1 Introduction

Dans cet article¹, nous cherchons à analyser d'un point de vue global le travail effectué au sein d'une organisation ou d'une entreprise. Dans notre cas, nous avons considéré une équipe de recherche comme une petite entreprise. L'expérience montre que fréquemment les chercheurs, pouvant appartenir à la même équipe ou au même laboratoire, ne savent pas toujours exactement sur quoi travaillent exactement les autres chercheurs. Notre but est de faire émerger des relations entre les travaux des différents chercheurs, pour comprendre les orientations de recherche principales et/ou marginales d'une équipe de recherche et ainsi fournir des explications sur le travail de recherche.

Dans une équipe de recherche, les publications sont une bonne façon de décrire les centres d'intérêts d'un chercheur. C'est pourquoi nous avons choisi d'analyser les notices bibliographiques de l'équipe. Nous avons travaillé sur les descriptions BibTeX des articles, qui nous fournissent les méta-données d'un article, comme le titre, les auteurs, les mots-clés, le résumé, l'année de publication, etc. Une entrée BibTeX s'appuie globalement sur le standard du Dublin Core². La norme Dublin Core comprend un ensemble d'éléments simples et représentatifs décrivant les caractéristiques des ressources sur le réseau, articles scientifiques en particulier. Cette norme est généralement utilisée pour la gestion des méta-données dans les pages HTML. Les descriptions BibTeX ayant un

1. Ce travail de recherche est réalisé en partie dans le cadre d'un programme de recherche franco-hongrois Balaton (Balaton F-23/03).

2. <http://www.dublincore.org>

“vocabulaire contrôlé”, c’est-à-dire un ensemble limité et consistant de termes bien définis, peuvent être considérées comme étant alignées sur le Dublin Core.

Pour analyser les publications, nous avons utilisé la classification par treillis comme technique d’extraction de connaissances à partir de données. Le processus d’extraction peut s’appuyer sur une ontologie, qui fournit un modèle des connaissances du domaine considéré. Dans notre cas, nous avons construit une ontologie représentant certains éléments de connaissances sur les membres et les publications de l’équipe. En utilisant cette ontologie, nous pouvons affiner le processus de fouille et d’analyse de données sur les publications de l’équipe.

Dans cet article, nous avons étudié plus précisément les relations entre les personnes, les publications et les thématiques. L’ontologie construite sert à guider la classification, en corrigeant certaines erreurs et une certaine redondance dans les données. Nous montrons et discutons les premiers résultats obtenus.

L’article est organisé comme suit. Dans la section 2, nous donnons une vue d’ensemble des ontologies et nous décrivons notre ontologie de domaine. La section 3 définit l’extraction de connaissance, présente le rôle des ontologies dans le processus d’extraction, et introduit les bases des treillis de concepts. La section 4 décrit le processus de classification proprement dit avec quelques exemples. La section 5 propose une synthèse et nous discutons les perspectives du travail de recherche présenté dans cet article.

2 Ontologies

Une ontologie peut se voir comme une spécification formelle et explicite d’une conceptualisation (c’est-à-dire de l’ensemble des concepts relatifs à un domaine) [Gruber, 1993]. Une ontologie s’appuie donc sur une représentation des connaissances d’un domaine, et peut être partagée entre personnes et systèmes. Une ontologie facilite le partage et la réutilisation des connaissances d’un domaine [McGuinness, 2002]. Une ontologie identifie:

- les concepts principaux dans un domaine
- les relations principales entre ces concepts
- un vocabulaire pour les concepts et les relations.

Plusieurs ontologies sont disponibles sur le Web, comme celles de la bibliothèque DAML Ontology³ ou le projet DMOZ⁴ (Directory Mozilla). À notre connaissance, il n’existe pas d’ontologie des thématiques de recherche et nous avons décidé de construire une ontologie prototype pour les besoins de notre travail de recherche. Cette ontologie va servir de base de connaissances de référence pour le domaine considéré. Elle va également servir à guider le processus de classification pour la correction d’erreurs et la gestion des synonymes.

Un extrait de notre ontologie est donné à la figure 1. Un noeud est un concept, et un arc orienté représente la relation sous-classe/classe. Tous les concepts sont des sous-classes d’un concept racine, appelé “Top”. Les mots-clés des publications sont

3. <http://www.daml.org/ontologies/>

4. <http://www.dmoz.org>

représentés par les concepts de cette ontologie. Pour plus de détails consultez la section 4.2.

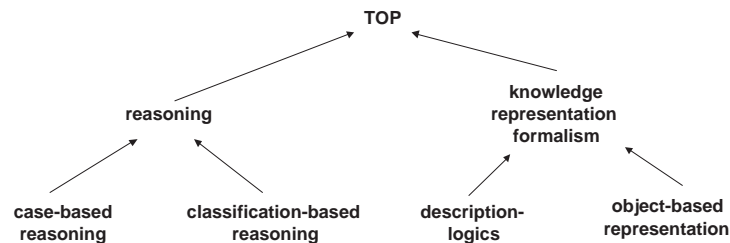


FIG. 1 – *Extrait de l'ontologie des mots-clés.*

3 L'extraction de connaissances dans les bases de données

Dans cette section nous introduisons l'extraction de connaissances dans les bases de données (ECBD). Puis, nous détaillons comment une ontologie peut être utilisée dans le processus d'ECBD. Ensuite, nous présentons les grands principes de la classification par treillis.

3.1 Les ontologies dans le processus d'ECBD

L'ECBD est un processus qui cherche à extraire des unités de connaissances nouvelles et réutilisables dans les grands volumes de données [Fayyad *et al.*, 1996], [Goebel et Gruenwald, 1999]. Ce processus peut être contrôlé par un analyste, qui est un spécialiste du domaine, et guidé par une ontologie portant sur le domaine des données. En retour, les résultats du processus d'ECBD peuvent venir enrichir l'ontologie considérée.

Notre travail de recherche porte plus spécialement sur l'organisation en treillis de connaissances relatives à une équipe de recherche. Pour mettre en oeuvre la classification par treillis, nous exploitons une ontologie qui porte sur les thématiques de recherche étudiées dans l'équipe, et qui nous permet de:

- corriger des erreurs d'orthographe et de traiter des problèmes de synonymie en regroupant les diverses étiquettes attribuées à une même information;
- guider le processus de classification en permettant de considérer les connaissances d'un domaine à différents niveaux de granularité, selon qu'un élément de connaissance est plus ou moins spécifique dans l'ontologie. La prise en compte du niveau de généralité permet de construire une famille de treillis sur le même ensemble de données.

De plus, la classification par treillis peut être utilisée en parallèle comme un module de recherche d'informations selon certains critères, comme cela est introduit dans [Carpineto et Romano, 2000]. Nous avons également exploité cette approche, mais il n'en sera pas question dans cet article.

3.2 Treillis de Galois

Les treillis de Galois (ou treillis de concepts) fournissent un processus de classification formel et efficace pour découvrir et représenter des hiérarchies de concepts [Guénoche et Van Mechelen, 1993]. Ils sont à la base de ce qui est aussi appelée "analyse de concepts formels", principalement utilisée en analyse des données et en extraction de connaissances [Ganter et Wille, 1998], [Ganter, 1999], [Ganter et Wille, 1999].

Comme donnée d'entrée, nous considérons une relation binaire entre un ensemble d'individus et un ensemble de propriétés, par exemple un ensemble de documents D et un ensemble des mots-clés W . Un contexte est un triplet (D, W, R) , où $R \subseteq D \times W$. $R(d, w)$ signifie que le document d est étiqueté par le mot-clé w . De cette relation binaire, on peut dériver pour chaque document son motif de mots-clés, c'est-à-dire, l'ensemble de tous les mots-clés qui s'appliquent à lui. De même, on peut dériver pour chaque mot-clé son modèle de document, c'est-à-dire, l'ensemble de tous les documents auxquels il s'applique.

Un *concept* C est déterminé par son extension et son intension: $(Extension(C), Intension(C))$, où l'extension se compose de tous les objets qui partagent les attributs de l'intension, et vice versa, l'intension se compose de tous les attributs qui sont communs aux objets de l'extension. Ces concepts sont *formels*, ce qui signifie qu'ils sont des entités mathématiques qui ne correspondent pas nécessairement à un concept du monde réel.

La *relation de subsumption* est la relation d'ordre partiel \leq qui est définie entre les concepts: $C_1 \leq C_2$ ce qui se lit C_1 est subsumé par C_2 , ou C_2 subsume C_1 , ssi: $Extension(C_1) \subseteq Extension(C_2)$, et de façon duale $Intension(C_2) \subseteq Intension(C_1)$. Le treillis de Galois se définit à partir d'un contexte (D, W, R) en considérant la table de la relation binaire R et les couples maximaux *individus* \times *propriétés* (nous ne détaillons pas ici cette construction, voir [Guénoche et Grandcolas, 1999], [Barbut et Monjardet, 1970]).

Parallèlement aux treillis, nous nous sommes intéressés aussi aux motifs et aux règles d'association [Bastide *et al.*, 2002]. Dans le contexte (D, W, R) soient X et $Y \subset W$, où $X \cap Y = \emptyset$. Une règle d'association est une implication de la forme $X \Rightarrow Y$. La règle $X \Rightarrow Y$ dans le contexte (D, W, R) s'interprète de la façon suivante: chaque document dans D qui contient les mots-clés X a tendance à contenir les mots-clés Y .

Considérons par exemple un contexte (D, W, R) , où $D = \{d_1, d_2, d_3\}$ et $W = \{w_1, w_2, w_3, w_4\}$. Dans le tableau suivant, 1 signifie qu'un document contient le mot-clé donné.

Les documents d_1 et d_3 sont les seuls documents qui ont le mot-clé w_3 , et réciproquement, le mot-clé w_3 est possédé seulement par les documents d_1 et d_3 . Ainsi $\{d_1, d_3\} \times \{w_3\}$ est un concept du treillis de Galois. La figure 2 montre le diagramme de Hasse du treillis de Galois obtenu pour ce tableau. Dans ce diagramme, chaque noeud est un concept et les arcs représentent la relation de subsumption \leq entre concepts.

	w_1	w_2	w_3	w_4
d_1	1		1	
d_2	1	1		
d_3			1	1

TAB. 1 – Un exemple simple documents \times mots-clés.

Deux concepts sont *voisins* s'il existe un arc direct entre eux. Deux règles d'association peuvent être extraites du treillis: $w_2 \Rightarrow w_1$, et $w_4 \Rightarrow w_3$.

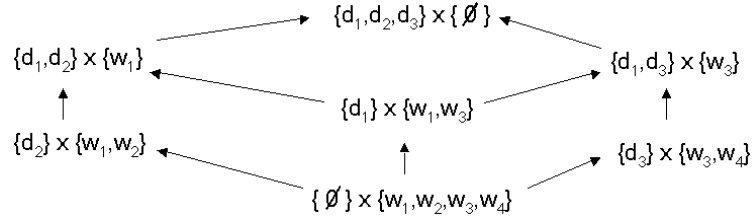


FIG. 2 – Diagramme de Hasse du treillis.

Le concept *Top* ($\{d_1, d_2, d_3\} \times \{\emptyset\}$) a une extension qui contient tous les objets et le concept *Bottom* ($\{\emptyset\} \times \{w_1, w_2, w_3, w_4\}$) a une intension qui contient tous les attributs.

Pour construire le treillis, nous avons employé l'algorithme incrémental décrit dans [Godin *et al.*, 1995], qui construit un treillis en ajoutant un nouvel objet à un treillis de Galois déjà existant.

4 Le treillis de Galois connecté à une ontologie

4.1 L'architecture générale du système

Dans la suite nous décrivons notre travail de recherche qui s'appuie sur la classification par treillis combinée avec une ontologie, pour l'organisation de connaissances dans un domaine donné. L'architecture générale du système est décrite à la figure 3.

Nous avons choisi d'étudier les corrélations existant entre les personnes, les documents et les thématiques dans notre équipe de recherche. Une personne a écrit un article, probablement avec d'autres personnes, sur une thématique donnée définie par une liste de mots-clés. Pour cette expérience, nous avons employé la bibliographie de l'équipe qui est stockée dans une base de données enregistrant la bibliographie entière du laboratoire. L'archive globale de bibliographie peut être consultée par une interface Web. Nous avons conçu un wrapper pour extraire les descriptions BibTeX des articles bibliographiques de l'équipe à partir des pages HTML. Puis, selon notre intérêt, nous construisons un treillis de Galois s'appuyant sur la relation *personnes* \times *articles*, ou *articles* \times *mots-clés*, ou *personnes* \times *mots-clés*.

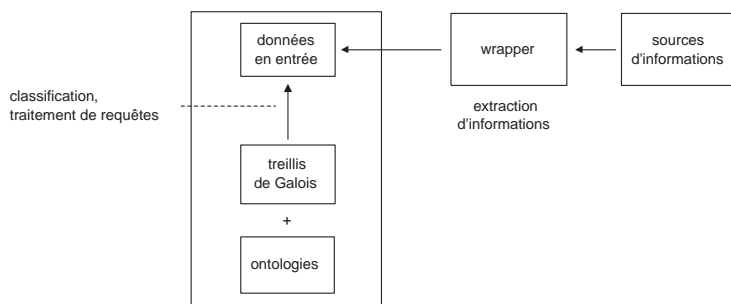


FIG. 3 – *L'architecture générale du système.*

4.2 La correction d'erreurs et la mise en facteur d'information avec une ontologie

Dans cette section, nous détaillons pourquoi et comment une ontologie peut être utilisée pour guider le processus d'extraction et d'organisation de connaissances à partir d'une base de données.

Divers problèmes ont pu être traités grâce aux connaissances intégrées à l'ontologie:

- Les publications sont indexées par des mots-clés qui sont fournis manuellement. Ces données sont très souvent entachées d'erreurs d'orthographe. Si un mot-clé est syntaxiquement incorrect, alors une recherche sur la base du même mot-clé écrit correctement ne peut pas aboutir.
- Synonymes: les mots-clés peuvent avoir divers synonymes. Comme l'association d'un mot-clé à un document ne s'appuie pas sur des règles bien définies ou une grammaire précise, plus d'un mot-clé peut être attaché à un document pour la même thématique. Dans ce cas, une recherche de documents doit pouvoir aboutir en utilisant n'importe lequel des synonymes.
- Langues: comme nous travaillons avec une bibliographie où il y a au moins deux langues, français et anglais, des mots-clés peuvent être employés dans les deux langues. Ceci est pris en considération dans l'ontologie de la même façon que les synonymes: les mots-clés sont donc fournis dans les deux langues étudiées.

Nous avons exploité l'ontologie des thématiques de la façon suivante. Tous les mots-clés qui servent à indexer les publications de l'équipe ont été répertoriés et filtrés à l'aide de l'ontologie: le filtrage permet de regrouper en une même classe les synonymes, les mots-clés de même nature et de langues différentes, les variantes syntaxiques. Cependant, si ce processus fonctionne bien pour un nombre limité de documents, il est difficile de l'étendre à l'échelle du Web car la liste des variantes pour les thématiques de recherche est alors impossible à construire de façon préalable. Un processus de classification approximative dans ce cas reste à construire.

Pratiquement, 147 publications auxquelles sont attachés 335 mots-clés différents ont été analysées. L'ensemble des 335 mots-clés s'est réduit à 89 mots-clés après filtrage,

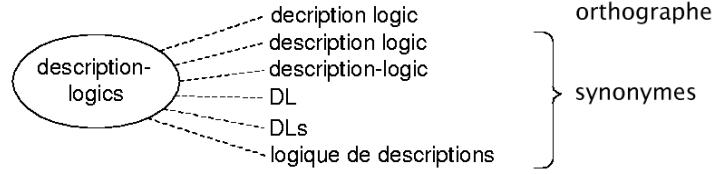


FIG. 4 – Groupement des mots-clés.

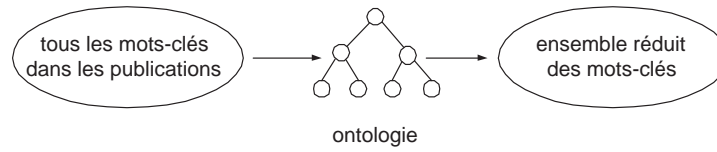


FIG. 5 – Filtrage des mots-clés.

ce qui montre, en particulier, le bruit existant dans les données brutes. Par exemple, l'ensemble de mots-clés suivants ('DL', 'DLs', 'case-based problem solving', 'CBR', 'galois connection') se transforme après le filtrage en ('description logics', 'case-based reasoning', 'Galois lattices').

4.3 L'organisation de documents avec des treillis de Galois

Dans cette section, nous montrons deux exemples d'organisation de documents s'appuyant sur les treillis de Galois, où la donnée d'entrée est la liste des publications de l'équipe. Dans les deux cas, nous avons besoin des mots-clés des publications. Nous avons employé l'ensemble réduit de mots-clés comme mentionné dans la section précédente.

4.3.1 “Quels sont les documents écrits sur une thématique commune?”

Dans cet exemple, nous nous intéressons plus particulièrement à la classification par treillis de documents ayant une thématique commune. Le treillis résultant donne des indications sur la proximité ou l'éloignement des thématiques pour les membres de l'équipe, mais aussi sur les groupes d'auteurs publiant ensemble sur un sujet donné.

Ainsi, pour étudier une interaction entre les documents x qui traitent d'une thématique y , il est possible de construire le treillis de Galois de la relation “le document x traite de la thématique y ”. Pour traiter cet exemple, nous avons pris un tableau de départ volontairement simplifié (voir tableau 2) où figurent 5 individus et 6 thématiques (ontologies, web sémantique, raisonnement à partir de cas, règles d'association, bioinfor-

matique, adaptation). Le treillis de Galois associé à ce tableau booléen est donné à la figure 6, où les concepts sont donnés sous forme de couples $\{\text{extension}\} \times \{\text{intension}\}$. Par exemple, il est possible de voir pour un concept quels auteurs travaillent sur un sujet donné, sachant qu'une clé comme *cadot03b* permet d'accéder à la notice bibliographique correspondante et au champ "Auteur" associé.

	ont.	sw	cbr	assoc. rules	bioinfo.	adapt.
cadot03b				x		
cherfi03c				x	x	
daquin02a						x
daquin03a	x	x	x			
lieber02a			x			x

TAB. 2 – Tableau d'entrée d'articles \times mots-clés.

En outre, il est encore possible d'extraire à partir du treillis des règles d'association, comme par exemple: *bioinformatique* \Rightarrow *règles d'association*. Les règles donnent un point de vue alternatif sur les données étudiées.

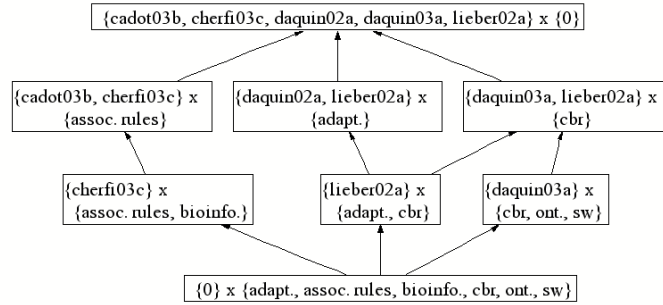


FIG. 6 – Treillis de Galois regroupant les documents selon leurs thématiques.

4.3.2 “Quelles sont les personnes travaillant sur une thématique commune?”

Dans une deuxième étape, nous nous sommes intéressés aux personnes travaillant sur une thématique commune dans l'équipe. La réponse à une telle question peut être très utile à l'échelle d'un laboratoire ou d'une région. Dans ce cas, nous devons rechercher des interactions entre les personnes et les thématiques (voir le tableau 3). Nous avons construit un autre treillis de Galois, qui diffère du précédent dans le sens où l'ensemble des extensions des concepts contient des auteurs. L'ensemble des intensions se compose des mots-clés qui sont communs aux publications de toutes les personnes présentes dans l'ensemble des extensions.

Encore une fois, l'exemple contient un nombre limité de mots-clés pour garder un treillis de taille modérée. Du diagramme de Hasse (figure 7) nous pouvons lire les réponses qui nous intéressent, c.-à-d. $\{Brachais, S., Lieber, J., d'Aquin, M.\}$ travaillent

	repr.	bioinfo.	classification	cbr
Lieber, J.	x	x	x	x
Brachais, S.	x			
Le Ber, F.			x	x
d'Aquin, M.	x	x		x

TAB. 3 – Tableau d'entrée d'auteurs \times mots-clés.

sur la *représentation*; $\{Le\ Ber, F., Lieber, J., d'Aquin, M.\}$ travaillent sur le *raisonnement à partir de cas*; $\{Lieber, J., d'Aquin, M.\}$ travaillent sur $\{bioinformatique, raisonnement à partir de cas, représentation\}$; et $\{Le\ Ber, F., Lieber, J.\}$ travaillent sur $\{raisonnement à partir de cas, classification\}$. Dans notre expérience nous avons pris tous les auteurs et coauteurs de nos publications (96) et l'ensemble réduit des mots-clés (89 mots-clés). Le diagramme de Hasse résultant compte 287 concepts.

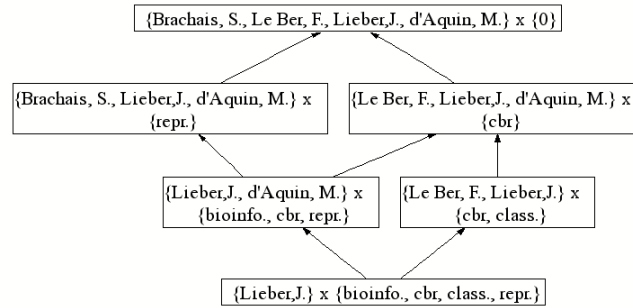


FIG. 7 – Les personnes travaillant sur la même thématique.

4.4 État du système

Pour l'implantation, nous avons employé la technologie Java Servlet/JSP. En Java il y a plusieurs API d'ontologie disponibles pour RDF(S), OWL, etc., rendant la manipulation des ontologies assez facile. Nous avons utilisé l'API d'ontologie Jena2⁵. Pour créer les ontologies nous avons utilisé l'éditeur Protégé-2000, et comme langage d'ontologie nous avons employé RDF(S). RDF(S) est un modèle pour décrire des relations et des classes pour représenter des ressources du Web. Actuellement, c'est plutôt le langage OWL (une extension de RDF(S)) qui permet de représenter des ontologies dans le cadre du Web sémantique. Nous envisageons d'utiliser OWL dans nos prochaines investigations.

Pour visualiser les treillis nous avons utilisé la boîte à outils Graphviz⁶ disponible librement qui permet de dessiner le treillis entier. Toutefois, la navigation dans un grand treillis comptant plus de cent concepts est difficile. Nous projetons d'essayer d'autres

5. <http://jena.sourceforge.net>

6. <http://www.graphviz.org>

méthodes de visualisation comme les techniques de “fisheye” ou d’arbres hyperboliques [Lamping et Rao, 1994].

5 Synthèse et conclusion

Dans cet article nous avons présenté une première expérience de l’application des treillis de Galois sur les données bibliographiques d’une équipe de recherche. Nous avons également montré comment utiliser une ontologie pour améliorer la construction des treillis, en particulier pour la correction des erreurs et le filtrage des données. Ce filtrage permet d’avoir une méthode beaucoup plus souple pour ensuite interroger les informations disponibles, autorisant par exemple une recherche approchée. Nous avons également montré comment extraire des relations entre les personnes, les documents, et les thématiques dans notre équipe, et nous avons proposé quelques exemples.

Pour notre travail, nous nous sommes appuyés sur trois projets comme travaux de référence. Galicia⁷ est un outil pour la construction et la visualisation de treillis. Galicia implante plusieurs algorithmes de classification par treillis, y compris des treillis d’icebergs. Le système Weka regroupe une collection d’algorithmes d’apprentissage automatique pour la fouille de données comme l’extraction de motifs fréquents et l’extraction des règles d’association [Witten et Frank, 1999]. Les systèmes Galicia et Weka sont écrits en Java et disponibles sur le Web. SEAL (SEmantic portAL) est un système permettant de construire des sites Web communautaires. Il utilise des ontologies comme éléments principaux pour gérer des sites Web communautaires et des portails Web sémantiques. SEAL est utilisé dans le portail OntoWeb⁸.

Nous avons choisi ces projets comme références pour notre travail de recherche parce qu’un de nos objectifs est de créer un portail sémantique pour l’équipe intégrant des possibilités d’extraction de connaissances. La construction d’un portail sémantique est aujourd’hui une étape importante dans la gestion des connaissances. Il est possible avec un tel portail de réunir en un même point d’accès toute l’information importante aussi bien de l’intérieur que de l’extérieur d’une organisation [Maedche *et al.*, 2002].

Notre approche peut également être utilisée sur différentes sources de données pour indiquer les relations entre les ensembles d’objets et d’attributs. Dans une première étape, une ontologie du domaine doit être construite (à partir de zéro ou mieux en réutilisant des ontologies déjà existantes) qui puisse contrôler l’ensemble des attributs. Dans une deuxième étape, l’ensemble réduit des attributs peut être employé pour construire les treillis de Galois souhaités.

Dans la suite de nos travaux, nous voulons tout d’abord étudier l’extraction automatique de règles à partir des treillis de Galois comme “dans 68% de cas, l’auteur A a publié avec l’auteur B”. Une autre piste de recherche intéressante est l’utilisation de treillis des icebergs. Le treillis des icebergs montre seulement la partie la plus élevée d’un treillis de concepts. Un tel treillis peut être utilisé en tant qu’outil de regroupement conceptuel, comme une méthode de visualisation, comme base pour des règles d’association et comme outil de visualisation pour ces règles d’association [Stumme *et al.*, 2002].

7. <http://www.iro.umontreal.ca/~galicia/>

8. <http://www.ontoweb.org>

Enfin, ce travail demande à être approfondi sur l'organisation multi-dimensionnelle des connaissances, ainsi que sur la façon de gérer plusieurs treillis “parallèles” ou “orthogonaux”, en prenant en compte les connaissances du domaine à différents niveaux de granularité.

Les questions qui ont été posées dans cet article peuvent être généralisées à des sources externes. Il doit par exemple devenir possible de traiter des questions comme: “quelles sont les conférences qui peuvent avoir un intérêt pour l'équipe et qui vont se tenir dans la prochaine année?”. La réalisation d'un tel système soulève plusieurs questions mais l'idée principale est ici que des machines effectuent des recherches sur le Web pour trouver l'information importante. Un tel projet s'appuie sur les principes du Web sémantique.

Références

- [Barbut et Monjardet, 1970] M. Barbut et B. Monjardet. *Ordre et classification*. Hachette, 1970.
- [Bastide *et al.*, 2002] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, et L. Lakhal. Pascal : un algorithme d'extraction des motifs fréquents. *Technique et science informatiques*, 21(1):65–95, 2002.
- [Carpineto et Romano, 2000] C. Carpineto et G. Romano. Order-Theoretical Ranking. *Journal of the American Society for Information Science*, 51(7):587–601, May 2000.
- [Fayyad *et al.*, 1996] U. Fayyad, G. Piatetsky-Shapiro, et P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17:37–54, 1996.
- [Ganter et Wille, 1998] B. Ganter et R. Wille. Applied lattice theory: formal concept analysis. In G. Grätzer, editor, *General lattice theory*, pages 591–605. Springer Verlag, 2 edition, 1998.
- [Ganter et Wille, 1999] B. Ganter et R. Wille. *Formal concept analysis: mathematical foundations*. Springer, Berlin/Heidelberg, 1999.
- [Ganter, 1999] B. Ganter. Attribute exploration with background knowledge. *Theoretical Computer Science*, 217(2):215–233, 1999.
- [Godin *et al.*, 1995] R. Godin, R. Missaoui, et H. Alaoui. Incremental Concept Formation Algorithms Based on Galois (Concept) Lattices. *Computational Intelligence*, 11(2):246–267, 1995.
- [Goebel et Gruenwald, 1999] M. Goebel et L. Gruenwald. A survey of data mining and knowledge discovery software tools. *SIGKDD Explorations Newsletter*, 1(1):20–33, 1999.
- [Gruber, 1993] T. R. Gruber. A Translation Approach to Portable Ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [Guénoche et Grandcolas, 1999] A. Guénoche et S. Grandcolas. Tree adjustments for partial distance. *Math. Inform. Sci. Humaines*, (146):51–64, 1999.
- [Guénoche et Van Mechelen, 1993] A. Guénoche et I. Van Mechelen. Galois Approach to the Induction of Concepts. In I. Van Mechelen, J. Hampton, R. S. Michalski, et P. Theuns, editors, *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. Academic Press Ltd, 1993.

- [Lamping et Rao, 1994] J. Lamping et R. Rao. Laying out and visualizing large trees using a hyperbolic space. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 13–14. ACM Press, 1994.
- [Maedche *et al.*, 2002] A. Maedche, S. Staab, N. Stojanovic, R. Studer, et Y. Sure. SEMantic PortAL - The SEAL Approach. In D. Fensel, J. Hendler, H. Lieberman, et W. Wahlster, editors, *Spinning the Semantic Web*. MIT Press, Cambridge, MA, 2002.
- [McGuinness, 2002] D. L. McGuinness. Ontologies Come of Age. In D. Fensel, J. Hendler, H. Lieberman, et W. Wahlster, editors, *Spinning the Semantic Web*. MIT Press, 2002.
- [Stumme *et al.*, 2002] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, et L. Lakhal. Computing Iceberg Concept Lattices with TITANIC. *Data and Knowledge Engineering*, 42(2):189–222, 2002.
- [Witten et Frank, 1999] I. H. Witten et E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 1999.