

Sémantique et contextes conceptuels pour la recherche d'information

Marie-Aude Aufaure^{1,2}, Bénédicte Le Grand³, Michel Soto³

¹Supélec – Computer Science department – plateau du Moulon – 3, rue Joliot Curie – F-91192 Gif sur Yvette Cedex

Marie-Aude.Aufaure@supelec.fr

²Projet Axis – INRIA Rocquencourt – Domaine de Voluceau – Rocquencourt
B.P. 105 78153 Le Chesnay Cedex

Marie-Aude.Aufaure@inria.fr

³Laboratoire d'Informatique de Paris 6 – 8, rue du Capitaine Scott – F-75015 Paris

{Benedicte.Le-Grand, Michel.Soto}@lip6.fr

Résumé. Cet article propose une méthodologie de recherche d'information qui utilise l'analyse conceptuelle conjointement avec la sémantique dans le but de fournir des réponses contextuelles à des requêtes sur le web. Le *contexte conceptuel* défini dans cet article peut être *global* – c'est-à-dire stable – ou *instantané* – c'est-à-dire borné par le contexte global. Notre méthodologie consiste en une première phase de pré traitement permettant de construire le contexte global, et une seconde phase de traitement en ligne des requêtes des utilisateurs, associées au contexte instantané. Notre processus de recherche d'information est illustré à travers une expérimentation dans le domaine du tourisme.

1 Introduction

Cet article présente un outil de recherche d'information associant sémantique et contexte conceptuel. Notre objectif est d'utiliser conjointement l'**analyse conceptuelle** et la **sémantique** afin de fournir des réponses contextuelles aux requêtes des utilisateurs sur le web.

Dans cet article, nous présentons notre méthodologie et nous l'illustrons par une recherche d'information effectuée sur un ensemble de pages web relatives au domaine du tourisme. Le processus de recherche d'information est divisé en deux étapes :

- traitement hors ligne de pages web ;
- traitement contextuel en ligne de requêtes utilisateurs.

Le prétraitement consiste à construire un treillis conceptuel à partir de pages web, par exemple dans le domaine du tourisme, de manière à obtenir un contexte conceptuel global ; cette notion est définie dans la section 3.2. Chaque concept du treillis correspond à un groupe de pages web ayant des propriétés communes. Un appariement sémantique est effectué entre les termes décrivant chaque page et un thésaurus du domaine du tourisme (thésaurus de

l'Organisation Mondiale du Tourisme), permettant de labelliser chaque concept de façon standardisée.

Tandis que le traitement des pages web est effectué hors ligne, la recherche d'information se fait en temps réel : les utilisateurs formulent leurs requêtes à l'aide des termes du thesaurus. Cette classe de termes est alors comparée avec les labels des concepts et les concepts les plus pertinents sont délivrés à l'utilisateur. Celui-ci peut alors naviguer à travers le treillis de manière à généraliser ou, au contraire, à spécialiser sa requête.

Cette méthode présente plusieurs avantages :

- les résultats sont fournis à la fois en fonction du contexte de la requête et du contexte des données disponibles. Par exemple, seuls les raffinements de requêtes correspondant à des pages touristiques existantes sont proposés ;
- l'ajout de sémantique peut dépendre de l'utilisateur cible ;
- une sémantique plus puissante, comme les ontologies, peut être ajoutée. Ceci permet d'améliorer la formulation des requêtes et la pertinence des résultats.

Cet article est organisé de la manière suivante : la section 2 introduit la notion de *contexte*, dans le sens général et dans le domaine de l'informatique. La section 3 décrit brièvement l'analyse formelle de concepts et les treillis de Galois, puis définit notre notion de *contexte conceptuel global* et *instantané*. Enfin, nous concluons et donnons quelques perspectives concernant la poursuite et l'application de ces travaux.

2 Notion de contexte

Un contexte est une notion abstraite et ne peut pas être défini de manière précise puisqu'il est lié à une situation particulière. Nous avons tendance à associer un contexte de manière implicite à un ensemble d'actions, une attitude, etc. dans des situations courantes. Des définitions de la notion de contexte ont émergé en psychologie cognitive, philosophie, ainsi que dans des domaines de l'informatique comme le traitement du langage naturel.

Le concept de contexte formel a été introduit par McCarthy dans (McCarthy 1968, 1987). Selon Giunchiglia, qui a également effectué des travaux de recherche sur la formalisation de contexte, « un contexte est une théorie sur le monde qui englobe les perspectives subjectives des individus ». Cette théorie est partielle – incomplète – et approximative du fait que le monde n'est jamais décrit dans tous ses détails (Giunchiglia, 1993).

La notion de contexte est importante pour beaucoup de communautés de recherche comme l'intelligence artificielle, la résolution de problèmes, etc. (Brezillon 1999a, 1999b), (Theodorakis et Spyrtos, 2002). En ce qui concerne l'intelligence artificielle, l'interaction entre contextes se fait au moyen de règles, qui permettent de naviguer d'un contexte à un autre (Guha et McCarthy, 2003). Les contextes peuvent être représentés par des graphes contextuels, des topic maps, les logiques de description avec notamment les extensions OWL, etc.

Comme dans le cas du web sémantique, le contexte est utilisé soit en tant que filtre dans un but de désambiguïsation pour la recherche d'information (Dolog et al, 2006), soit pour définir des services web contextuels (Mrissa et al, 2006), ou enfin comme un moyen d'intégrer ou de fusionner des ontologies, (Bouquet et al, 2004), (Doan et al, 2002). Un contexte peut être spécifié à différents niveaux de granularité (document, page web, etc.). Cette information additionnelle peut être liée à chaque ressource.

3 Contextes conceptuels - Relation avec les ontologies

Dans la section précédente, nous avons présenté diverses définitions de la notion de *contexte*. Dans cet article, nous définissons des *contextes conceptuels*, basés sur l'analyse formelle de concepts, en particulier les treillis de Galois. Beaucoup de travaux de recherche ont appliqué les treillis de concepts à la recherche d'information (Priss, 2000). Les concepts formels peuvent être vus comme des documents pertinents pour une requête donnée. L'introduction d'une ontologie de domaine, combinée avec les treillis de concepts pour améliorer la recherche d'information est plus récente. (Messai et al, 2005) proposent une approche basée sur l'analyse formelle de concepts pour classifier et rechercher des sources de données pertinentes pour une requête donnée. Ces travaux ont été appliqués à des données de bioinformatique. Un treillis de concepts est construit en fonction des métadonnées associées aux sources de données. Puis, un concept construit à partir d'une requête donnée est fusionné à ce treillis de concepts. Dans cette approche, le raffinement de requête s'effectue en utilisant une ontologie de domaine. Le processus de raffinement d'*OntoRefiner*, outil dédié aux portails web sémantiques (Safar et al, 2004), est basé sur l'utilisation d'une ontologie de domaine pour construire un treillis de Galois pour le processus de raffinement de requête. L'ontologie de domaine évite de construire complètement le treillis ; ce travail vise à améliorer la construction du treillis, ce qui n'est pas l'objectif de notre travail. Enfin, le système *CREDO* (Carpineto et Romano, 2004) permet à l'utilisateur d'interroger des documents web et de voir les résultats à travers la navigation dans un treillis de concepts (<http://credo.fub.it>).

(Dolog et al, 2006) ont proposé une méthode pour relâcher automatiquement des requêtes trop contraintes en se basant sur la connaissance du domaine et les préférences utilisateur. Leur approche combine raffinement et relaxation de manière à permettre un accès personnalisé à des données RDF hétérogènes. Contrairement à cette approche, notre méthode est dédiée à des requêtes imprécises et centrées utilisateurs.

Dans notre proposition, les treillis de Galois sont construits pour représenter le contenu de pages web. L'utilisateur peut alors naviguer dans ces treillis de manière à raffiner ou généraliser sa requête. Comparativement aux approches décrites ci-dessus, notre méthode n'est pas seulement dédiée à la recherche d'information mais peut être utilisée pour d'autres objectifs comme le peuplement d'ontologies, la comparaison de sites web à travers leurs treillis respectifs, une aide au concepteur de site web pour vérifier que le contenu du site reflète bien le message qu'il a voulu faire passer, etc.

Cette section est organisée de la manière suivante : après une brève introduction aux treillis de Galois, nous proposons notre définition de contextes conceptuels global et instantané.

3.1 Introduction à l'Analyse Formelle de Concepts et aux Treillis de Galois

L'Analyse Formelle de Concepts est une approche mathématique de l'analyse de données qui permet de fournir une structure à l'information. Cette approche peut être utilisée pour le clustering conceptuel, comme montré dans (Carpineto et Romano, 1993) et dans (Wille, 1984).

Sémantique et contextes conceptuels pour la recherche d'information

La notion de treillis de Galois établissant une relation entre deux ensembles est à la base d'un ensemble de méthodes de classification conceptuelles. Cette notion fut introduite par (Birkoff, 1940) et (Barbut et Monjardet, 1970). Les treillis de Galois consistent à regrouper des objets en classes qui vont matérialiser les concepts du domaine d'étude. Les objets individuels sont discriminés en fonction de leurs propriétés communes, ce qui permet d'effectuer une classification sémantique. L'algorithme que nous avons implémenté est basé sur celui proposé dans (Godin, 1998).

Nous introduisons tout d'abord les principaux concepts des treillis de Galois.

Soient deux ensembles finis E et E' (E est un ensemble d'objets et E' est l'ensemble de leurs propriétés), et une relation binaire $R \subseteq E \times E'$ entre ces deux ensembles. La figure 1 montre un exemple de relation binaire entre deux ensembles. Selon la terminologie de Wille (Wille, 1992), le triplet (E, E', R) est un contexte formel correspondant à un unique treillis de Galois. Il représente des regroupements naturels d'éléments de E et E'.

Soient P(E) une partition de E et P(E') une partition de E'. Chaque élément du treillis est un couple, appelé aussi concept, noté (X, X'). Un concept est composé de deux ensembles $X \in P(E)$ et $X' \in P(E')$ satisfaisant les deux propriétés suivantes :

$$X' = f(X) \text{ où } f(X) = \{ x' \in E' \mid \forall x \in X, xRx' \} \tag{1}$$

$$X = f'(X') \text{ où } f'(X') = \{ x \in E \mid \forall x' \in X', xRx' \}$$

Un ordre partiel sur les concepts est défini de la manière suivante :
Soit $C_1=(X_1, X'_1)$ et $C_2=(X_2, X'_2)$,

$$C_1 < C_2 \Leftrightarrow X'_1 \subseteq X'_2 \Leftrightarrow X_2 \subseteq X_1 \tag{2}$$

Cet ordre partiel est utilisé pour dessiner un graphe appelé *diagramme de Hasse*, illustré par la figure 1. Il existe un arc entre deux concepts C_1 et C_2 si $C_1 < C_2$ et s'il n'y a pas d'autre élément C_3 dans le treillis tel que $C_1 < C_3 < C_2$. Dans un diagramme de Hasse, la direction des arcs est par convention toujours dans le même sens, soit vers le haut. Le graphe peut être interprété comme une représentation de la relation de généralisation / spécialisation entre les couples, où $C_1 < C_2$ signifie que C_1 est plus général que C_2 (et C_1 est situé au-dessus de C_2 dans le diagramme).

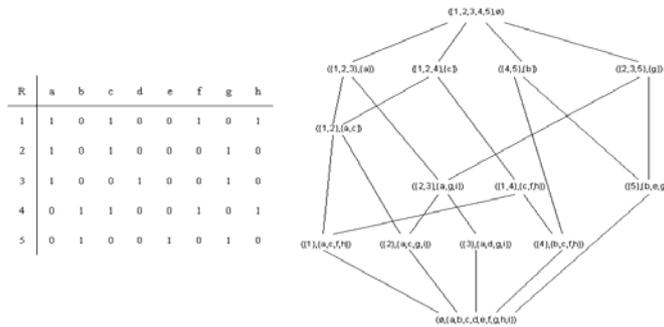


Fig. 1. Relation binaire et treillis de Galois associé (Diagramme de Hasse).

Le treillis de concepts montre les points communs entre les objets du domaine. La première partie d'un concept est l'ensemble des objets et est appelée l'extension. La seconde partie – l'intension – révèle les propriétés communes des objets de l'extension. La partie droite de la figure 1 montre un exemple de treillis de concepts. Par exemple, dans ce treillis, le concept $\{(1, 4), (c, f, h)\}$ contient dans son extension les objets 1 et 4, qui ont les propriétés c, f et h en commun.

3.2 Définition d'un contexte conceptuel

Dans notre définition de contexte conceptuel, nous distinguons le contexte conceptuel *global* et le contexte conceptuel *instantané* :

- Le contexte conceptuel *global* consiste à construire un treillis de Galois à partir d'une sélection de sites web du tourisme où chaque concept est un ensemble de pages web labellisées en utilisant les termes d'un thesaurus ou d'une ontologie. Cette étape est détaillée et illustrée dans la section 4.1.

Propriétés :

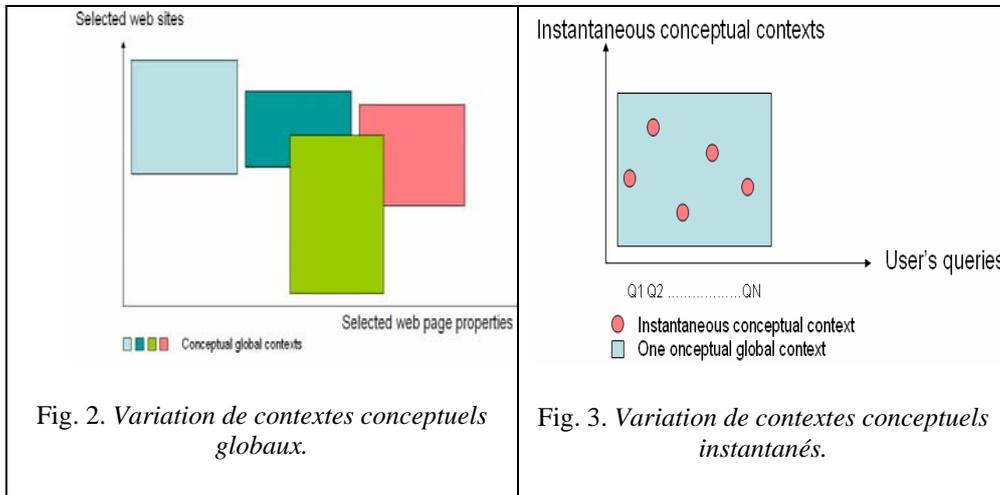
1. La principale propriété de ce contexte conceptuel global est qu'il est stable. La seule façon de le faire évoluer est soit de modifier la sélection de sites web de tourisme et/ou de changer l'ensemble des propriétés sélectionnées pour une page web comme illustré sur la figure 2.
2. Une page web donnée peut appartenir à plusieurs contextes conceptuels globaux. Par exemple, si nous considérons une page web sur le musée du *Louvre*, son contexte conceptuel global sera fonction du site web d'où est extraite cette page (site dédié au tourisme, au musées nationaux, site web personnel, etc.). La similarité avec d'autres pages web dépendra de son contexte conceptuel global.

- Le contexte conceptuel *instantané* représente à la fois la requête exprimée par l'utilisateur et le résultat de cette requête. Cette étape est détaillée et illustrée dans la section 4.2.

Propriétés :

1. Du fait des aspects fortement exploratoires et itératifs du processus de recherche d'information, la principale propriété de ce contexte conceptuel instantané est qu'il évolue à chaque fois que l'utilisateur modifie, raffine ou généralise sa requête.
2. La variation de ce contexte conceptuel instantané est bornée par le contexte conceptuel global, comme illustré dans la figure 3.
3. Une page web donnée peut appartenir à différents contextes conceptuels instantanés du fait des relations de généralisation/spécialisation inhérentes aux treillis de Galois.

Enfin, l'information fournie par le contexte conceptuel (global ou instantané) est complémentaire à l'information intrinsèque des pages web (les propriétés, dans notre cas, représentent les termes les plus significatifs de la page).



4 Méthodologie pour une coordination sémantique de contextes conceptuels et d'ontologies

Notre méthodologie de recherche d'information sémantique et contextuelle peut être divisée en deux étapes : un pré traitement hors ligne et un traitement contextuel en ligne des requêtes des utilisateurs. Dans cette section, nous décrivons ces opérations et les illustrons avec un exemple simple dans le domaine du tourisme.

4.1 Etape 1: prétraitement hors ligne

4.1.1 Sélection de sites web de tourisme

La première étape consiste à construire un contexte conceptuel global dans un domaine donné et sur un ensemble de pages web pertinentes. Ces pages peuvent appartenir à plusieurs sites web. Pour des besoins d'illustration, nous avons sélectionné 5 pages web à partir d'un site touristique français (le site de la mairie de Metz) : <http://tourisme.mairie-metz.fr/>. Cet ensemble de pages web constitue l'information de base à partir de laquelle la recherche d'information sémantique et contextuelle est réalisée.

4.1.2 Analyse des pages Web : génération des données d'entrée pour le treillis de Galois (objets et propriétés)

Le contexte conceptuel global est représenté par le treillis de Galois construit à partir de l'ensemble des pages web sélectionnées ; nous devons générer des données en entrée appropriées pour le calcul du treillis conceptuel :

- Chaque page web correspond à un *objet*;

- Les *propriétés* d'une page web sont les noms significatifs les plus fréquents de la page (ils sont extraits à l'aide de notre outil basé sur Tree-Tagger comme l'illustre la figure 4).

La liste des objets et de leurs propriétés est stockée dans une base de données Mysql, comme le montre la figure 4. Par exemple, l'*objet* 3 (correspondant à une page web spécifique) est décrit par les propriétés *spectacle* et *réservation*.

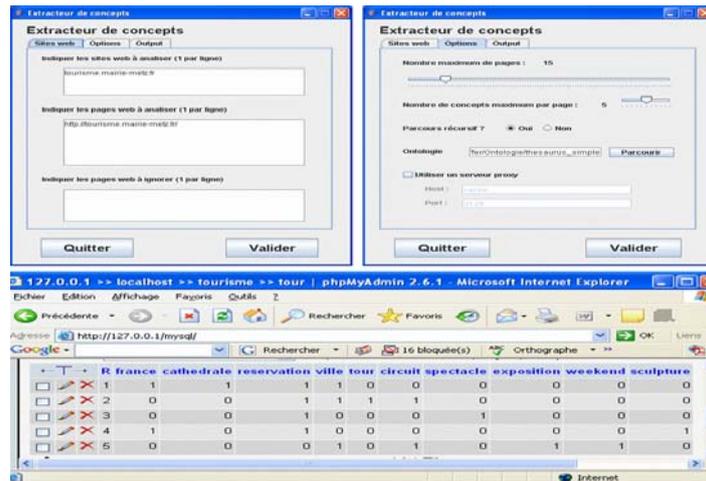


Fig. 4. Extraction d'objets/propriétés et génération de la base de données.

4.1.3 Construction du treillis de Galois

La base de données décrite ci-dessus –contenant des objets et propriétés – est utilisée comme entrée pour la construction d'un treillis de Galois. Nous utilisons un algorithme existant pour la construction incrémentale du treillis de Galois (Godin, 1998), car notre contribution concerne l'exploitation et l'interprétation de ces treillis et non l'optimisation de leur construction. La construction du treillis consistant à identifier et structurer tous les regroupements d'objets en fonction de leurs propriétés communes, le nombre de concepts du treillis devient rapidement élevé si le nombre d'objets et/ou de propriétés est important. Dans le cas de sites possédant de nombreuses pages, cela aura donc un impact sur le temps de calcul. Ceci n'est pas trop pénalisant dans notre approche puisque cette étape est effectuée lors d'une phase de prétraitement.

En sortie, nous obtenons un treillis de concepts où chaque concept consiste en un ensemble d'objets ayant des propriétés communes. La liste des objets d'un concept est appelée l'*extension* d'un concept et les propriétés partagées correspondantes constituent l'*intension* du concept. Le treillis généré à partir de la base de données de la figure 4 est illustré sur la figure 5.

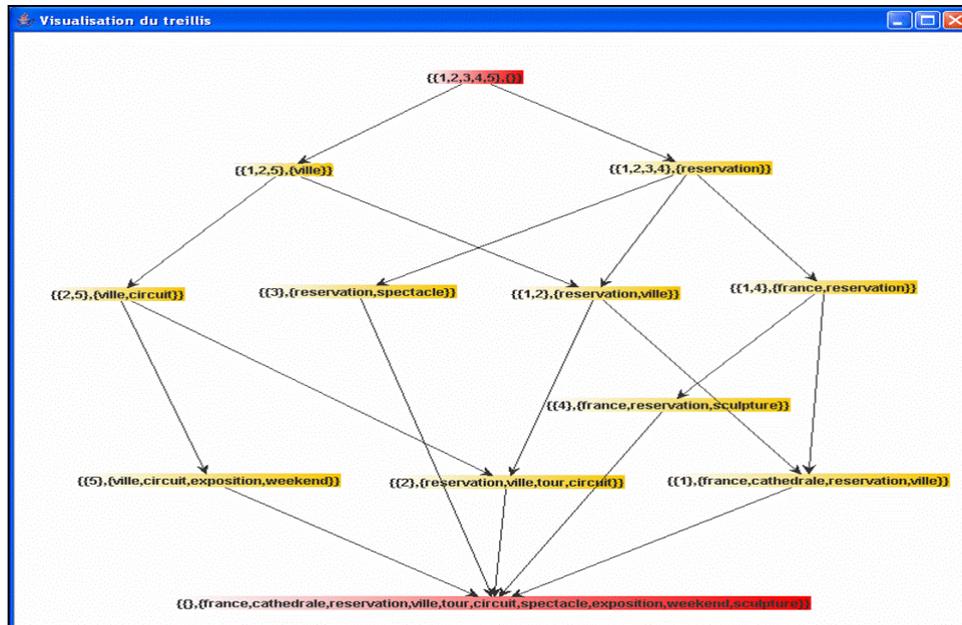


Fig. 5. Le contexte conceptuel global : le treillis de Galois.

Le treillis de la figure 5 contient 12 noeuds (concepts). Chaque concept est caractérisé par son extension – la liste des objets qu’il contient – et son intension – la liste des propriétés communes des objets de l’extension. Le nœud en haut du treillis est le plus général et contient tous les objets, qui n’ont pas de propriétés communes. Plus nous descendons dans le treillis, plus les noeuds deviennent spécifiques. Lors de l’interprétation du treillis, il est intéressant d’étudier quels objets de l’extension sont similaires à d’autres.

Ce treillis nous montre que l’*objet 3* apparaît seulement dans deux concepts car il possède seulement deux propriétés et ne partage que la propriété *réservation* avec les *objets 1, 2 et 4*. Nous pouvons également remarquer qu’aucun objet ne partage la propriété *spectacle* avec l’*objet 3*. Au contraire, l’*objet 2* apparaît dans cinq concepts car il possède plus de propriétés (4) que l’*objet 3* et aussi parce que ses propriétés sont partagées par plus d’objets (4) que l’*objet 3*.

L’interprétation que nous pouvons faire de ce treillis est que :

- Le contexte conceptuel global du site touristique de la ville de Metz indique que celui-ci est moins ciblé vers le spectacle et plus dirigé vers les centres d’intérêts de la ville elle-même.
- Les requêtes des utilisateurs sur les spectacles ne pourront pas être raffinées mais généralisées sur le concept réservation, et de ce fait, sont fortement bornées par le contexte conceptuel global.

En d’autres termes, si un utilisateur est plus intéressé par un tourisme basé sur le spectacle, la ville de Metz n’est pas l’endroit le mieux adapté (au vu des informations disponibles sur le Web). L’analyse des requêtes des utilisateurs va permettre de déterminer si l’utilisateur est plus intéressé par les spectacles ou par une autre forme de tourisme. Dans ce

cas, il sera possible de lui proposer un autre contexte conceptuel global plus orienté “spectacles”. Ce nouveau contexte peut être relié au contexte conceptuel global du tourisme de la ville de Metz en réalisant un appariement avec des ontologies comme nous le proposons dans notre conclusion.

4.1.4 Labellisation des concepts du treillis

La dernière étape du traitement hors ligne consiste à affecter un label aux concepts du treillis de manière normalisée, c’est à dire en utilisant les termes d’un thesaurus ; nous avons utilisé le thesaurus de l’Organisation Mondiale du Tourisme qui décrit le tourisme et les activités de loisirs. La même opération peut être réalisée avec une ontologie plutôt qu’un thesaurus.

L’analyse des pages web consiste à extraire les mots les plus fréquents d’une page (voir section 4.1.2). Notre analyse va plus loin en réalisant un appariement syntaxique entre ces mots fréquents et les termes du thesaurus de l’OMT. Par exemple, la propriété *spectacle* est liée à l’entrée *loisirs* du thesaurus. Le label normalisé d’un concept est constitué de l’ensemble des labels normalisés associés aux objets de son extension.

Le lien entre les données textuelles – les pages web de tourisme – et une structure sémantique – le thesaurus de l’OMT – nous permet d’obtenir un contexte conceptuel global plus riche. En effet, ce contexte reflète aussi bien l’information issue des données d’origine que les connaissances générales du domaine.

4.2 Etape 2: traitement en ligne des requêtes des utilisateurs

Une fois le contexte conceptuel global construit, le contexte conceptuel instantané est calculé pour chaque requête utilisateur.

4.2.1 Formulation d’une requête avec des mots-clés

Les utilisateurs formulent leurs requêtes avec des mots-clés (nous pouvons restreindre ces mots-clés aux entrées du thesaurus ou de l’ontologie). L’intérêt d’avoir normalisé les labels des objets et concepts est la possibilité d’utiliser un vocabulaire contrôlé.

4.2.2 Identification des concepts les plus pertinents du treillis

La réponse à une requête correspond au concept dans le treillis dont les propriétés de l’intension s’apparient le mieux avec les mots-clés de cette requête. S’il n’existe pas de concept correspondant parfaitement, les concepts les plus pertinents sont proposés à l’utilisateur à travers un raffinement ou une généralisation, qui s’effectue en fonction des données disponibles (dans les pages web de tourisme), c’est-à-dire correspondant au contexte des données.

La figure 6 montre une autre représentation visuelle du treillis de Galois. Cette interface est plus riche du fait qu’elle montre également les liens avec les données sémantiques – les entrées du thesaurus de l’OMT. Supposons que l’utilisateur entre une requête avec le mot-clé *réservation*. Le contexte conceptuel instantané correspondant à cette requête est illustré dans la fenêtre de gauche avec une focalisation sur le concept du treillis (concept numéro 5), dont l’extension est $\{page1, page2, page3, page4\}$ et la propriété commune entre ces objets est

Sémantique et contextes conceptuels pour la recherche d'information

{réservation}. Nous pouvons également noter que ce concept est labellisé avec trois concepts du thesaurus: *monument*, *loisirs*, *europe*.

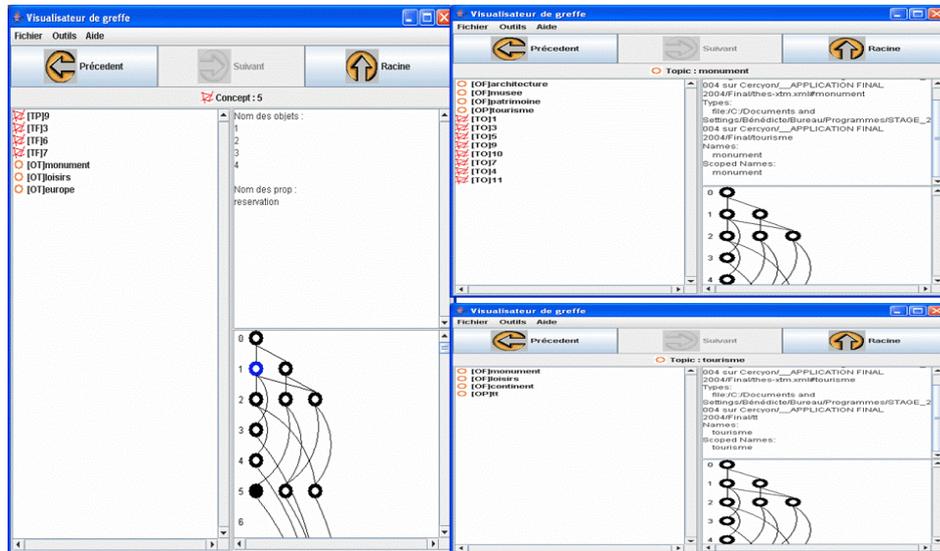


Fig. 6. Navigation à travers le contexte global enrichi avec de l'information sémantique

A partir de ce contexte instantané (noeud numéro 5), l'utilisateur est libre de naviguer soit dans d'autres noeuds du treillis – à travers des concepts plus généraux et plus spécifiques – soit dans le thesaurus.

La partie en haut à droite de la figure 6 donne un aperçu de l'entrée *monument* du thesaurus, et la partie gauche de la fenêtre montre la hiérarchie du thesaurus –*monument* est une sous-classe de *tourisme* et une superclasse de *architecture*, *musée* et *patrimoine*– de même que les autres concepts du treillis labellisés avec l'entrée *monument*–soient les concepts 1, 3, 5, 9, 10, 7, 4 et 11. Encore une fois, la navigation peut se poursuivre soit à travers le thesaurus ou en revenant sur le treillis conceptuel.

L'avantage de cette interface de navigation est que l'utilisateur peut explorer facilement les données via le thesaurus – ou une ontologie – et le contexte conceptuel global, ceci de manière totalement transparente. L'élargissement du contexte des données avec la connaissance du domaine véhiculée par un thesaurus ou une ontologie permet de fournir des réponses plus riches et plus pertinentes aux requêtes des utilisateurs. Cette interface n'est qu'un exemple de représentation et de navigation illustrant l'utilisation des contextes conceptuels. Dans l'avenir, nous nous pencherons plus en détail sur la question des interfaces afin qu'elles permettent d'optimiser l'utilisation de ces contextes globaux et instantanés.

5 Conclusion et perspectives

Dans ce papier, nous avons présenté une méthodologie combinant l'analyse conceptuelle avec la sémantique. Notre méthodologie est divisée en deux étapes ; la première consiste en

un prétraitement hors ligne d'un ensemble de pages web à partir desquelles un treillis conceptuel est construit. Chaque concept correspond à un cluster de pages web partageant un ensemble de propriétés (un ensemble de termes communs). Puis, les termes pertinents des pages web sont mis en correspondance avec un thesaurus du domaine du tourisme, permettant ainsi d'affecter des étiquettes (labels) à chaque concept du treillis de manière normalisée. La deuxième étape est une phase de traitement contextuel des requêtes des utilisateurs. Les termes constituant la requête sont alors comparés avec les labels des concepts. L'utilisateur peut ensuite naviguer comme il le souhaite dans le treillis pour raffiner ou bien généraliser sa requête ; il peut également naviguer à travers la structure sémantique – thesaurus ou ontologie – s'il a besoin de la connaissance du domaine pour affiner sa recherche.

Nous avons illustré cette méthodologie dans le domaine du tourisme. Un thesaurus a été utilisé, mais les travaux futurs prévoient d'étendre cette méthodologie aux ontologies. L'avantage des ontologies ou des thesaurus pour les contextes est qu'il est possible de lier divers contextes à travers ces structures sémantiques. Le contexte peut être élargi par l'ontologie. Inversement, l'avantage des contextes pour les ontologies est que le fait de les relier à différents contextes revient à les instancier avec ces différents contextes. Nous pouvons donc voir ces contextes comme un moyen de peupler les ontologies.

Dans ce papier, nous avons montré que l'utilisation conjointe de treillis conceptuels et de la sémantique pouvait permettre d'obtenir des résultats intéressants pour la recherche d'information sémantique et contextuelle. Mais cette combinaison de l'analyse formelle de concepts et de la sémantique peut également être utilisée dans le cadre d'objectifs différents, comme par exemple le peuplement d'ontologies. Cette méthodologie peut également permettre de comparer différents sites web à travers leurs treillis respectifs. Un autre aspect intéressant est que, les termes correspondants à des points d'entrée d'un ou plusieurs treillis, les utilisateurs vont pouvoir naviguer d'un treillis à l'autre pour reformuler leurs requêtes. Enfin, une autre application possible est d'aider les concepteurs d'un site web du fait que le treillis va refléter le contenu du site. Il sera alors aisé de comparer le site web/treillis résultant avec les objectifs initiaux du concepteur de site, et ainsi de vérifier si le site est bien conforme à ces objectifs.

Références

- Barbut, M., Monjardet, B., *Ordre et classification*, Algebre et combinatoire, Tome 2, Hachette, 1970.
- Birkhoff, G., *Lattice Theory*, First Edition, Amer. Math. Soc. Pub. 25, Providence, R. I., 1940.
- Bouquet P., Giunchiglia F., Van Harmelen F., Serafini L., Stuckenschmidt H.: *Contextualizing Ontologies*. Journal of Web Semantics, 1(4):1-19, 2004
- Brézillon, P., Context in Artificial Intelligence: I. A survey of the literature. *Computer and Artificial Intelligence*. 18(18): 321-340. 1999.
- Carpineto, C., Romano, G., *Galois: An order-theoretic approach to conceptual clustering*, Proc. Of the 10th Conference on Machine Learning, Amherst, MA, Kaufmann, pp. 33-40, 1993.

- Carpineto, C., Romano, G., Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. *Journal of Universal Computer Science*, vol. 10, n. 8, pp. 985-1013, 2004.
- Doan, A., Madhavan, J., Domingos, P., Learning to Map between Ontologies on the Semantic Web. In the 11th International World Wide Web Conference (WWW'2002), May 7-11, Hawaii, 2002.
- Dolog, P., Stuckenschmidt, H., Wache, H., Robust Query Processing for Personalized Information Access on the Semantic Web. *FQAS 2006*: 343-355
- Godin, R, Chau, T.-T., *Incremental concept formation algorithms based on Galois Lattices*, *Computational intelligence*, 11, n° 2, p246 –267, 1998.
- Guha, R., McCarthy, J., Varieties of contexts. 4th International and Interdisciplinary Conference, CONTEXT 2003. *Lecture Notes in Computer Science*, vol 2680, pp. 164-177, 2003.
- Messai, N., Devignes, M-D., Napoli, A. and Smail-Tabbone, M., Querying a Bioinformatic Data Sources Registry with Concept Lattices. *13th International Conference on Conceptual Structures - ICCS 2005*. (Kassel, Germany). Springer, 2005. *Lecture Notes in Computer Science*. vol 3596. pp.323--336.
- Mrissa, M., Ghedira, C., Benslimane, D., Maamar, Z., A Context Model for Semantic Mediation in Web Services Composition. 25th International Conference on Conceptual Modeling (ER2006) November 6-9 2006, Tucson, Arizona, USA. 2006.
- Priss, U., "Lattice-based Information Retrieval." *Knowledge Organization*, Vol. 27, 3, 2000, p. 132-142.
- Theodorakis M. & Spyratos N. Context in artificial intelligence and information modelling. *Proceedings of the second Hellenic Conference on Artificial Intelligence (SETN'02)*, Thessalonique, 2002.
- Wille, R., *Line diagrams of hierarchical concept systems*, *Int. Classif.* 11, pp. 77-86, 1984.
- Wille, R., Concept lattices and conceptual knowledge systems, *Computers & Mathematics Applications*, 23, n° 6-9, pp. 493-515, 1992.

Summary

This paper presents an information retrieval methodology which uses Formal Concept Analysis in conjunction with semantics to provide contextual answers to Web queries The *conceptual* context defined in this article can be *global* - i.e. stable- or *instantaneous*- i.e. bounded by the global context. Our methodology consists first in a pre-treatment providing the global conceptual context and then in an online contextual processing of users' requests, associated to an instantaneous context. Our information retrieval process is illustrated through experimentation results in the tourism domain. One interest of our approach is to perform a more relevant and refined information retrieval, closer to the users' expectation.