

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