

Classement des fragments de documents XML par une méthode d'aide à la décision

Faiza Abbaci*, Pascal Francq**

*

Departement de sciences de l'information et de la communication,
Université Libre de Bruxelles.
50, Av. F. D. Roosevelt, CP 123,B-1050 Bruxelles, Belgique
fabbaci@ulb.ac.be <http://homepages.ulb.ac.be/fabbaci>

**

Departement de sciences de l'information et de la communication,
Université Libre de Bruxelles.
50, Av. F. D. Roosevelt, CP 123,B-1050 Bruxelles, Belgique
pfrancq@ulb.ac.be

Résumé. Vu l'accroissement constant du volume d'information accessible en ligne sous format XML, il devient primordial de proposer des modèles adaptés à la recherche d'information dans les documents XML. Tandis que la recherche d'information classique repose sur l'indexation du contenu des documents, la recherche d'information dans les documents XML tente d'améliorer la qualité des résultats en tirant profit de la sémantique véhiculée par la structure des documents. Dans cet article, nous présentons une méthode de classement des items (éléments XML) retournés lors d'une recherche dans une collection de documents XML. Le classement repose sur la prise en compte d'un ensemble de critères discriminants. La particularité de notre approche réside dans la façon dont nous les utilisons : Nous employons une méthode décisionnelle pour classer les items en les comparant deux-à-deux là où en général une fonction de scoring globale est utilisée.

1 Introduction

L'une des conséquences de la prolifération de l'information en ligne de nos jours est la diversité des données. XML se distingue comme le format par excellence pour la représentation, le stockage et l'échange de données sur Internet.

Les systèmes de recherche d'information dans les documents XML (RI-XML) utilisent soit le paradigme de l'*appariement exact* soit celui de l'*appariement approximatif* (ou *appariement par classement*). Dans le premier cas, la requête doit vérifier les contraintes sur le contenu et la structure spécifiées dans la requête, ainsi chaque item (document, fragment de document ou élément XML) sur lesquels la recherche est effectuée et jugé pertinent ou non. Dans le second cas, les items sont classés selon leur pertinence à la requête. Dans le contexte du Web, l'appariement approximatif est plus approprié. En effet, l'appariement exact nécessite un langage