

Recherche d'Information Agrégée dans des documents XML basée sur les Réseaux Bayésiens

Najeh Naffakhi^{*,**}, Mohand Boughanem^{*}, Rim Faiz^{***}

^{*}IRIT, Université Paul Sabatier
118, route de Narbonne, 31062 Toulouse Cedex 9
{naffakhi, bougha}@irit.fr
<http://www.irit.fr>

^{**}LARODEC, Université de Tunis, ISG
41, rue de la liberté, cité Bouchoucha, 2000 Le Bardo, Tunis
najeh.naffakhi@isg.rnu.tn
<http://www.isg.rnu.tn/larodec>

^{***}LARODEC, Université de Carthage, IHEC
2016 Carthage Présidence, Tunis
rim.faiz@ihec.rnu.tn
<http://www.isg.rnu.tn/larodec>

Résumé

Dans cet article, nous nous intéressons à la recherche agrégée dans des documents XML. Pour cela, nous proposons un modèle basé sur les réseaux bayésiens. Les relations de dépendances entre requête-termes d'indexation et termes d'indexation-éléments sont quantifiées par des mesures de probabilité. Dans ce modèle, la requête de l'utilisateur déclenche un processus de propagation pour trouver des éléments. Ainsi, au lieu de récupérer une liste des éléments qui sont susceptibles de répondre à la requête, notre objectif est d'agréger dans un agrégat des éléments pertinents, non-redondants et complémentaires. Nous avons évalué notre approche dans le cadre de la campagne d'évaluation INEX 2009 et avons présenté quelques résultats expérimentaux mettant en évidence l'impact de l'agrégation de tels éléments.

Summary

In this paper, we are interested in aggregated search in structured XML documents. We present a structured information retrieval model based on the Bayesian networks theory. Query-terms and terms-elements relations are modeled through probability. In this model, the user's query starts a propagation process to recover the XML elements. Thus, instead of retrieving a whole document or a list of disjoint elements that are likely to answer partially the query, we attempt to build a virtual document that aggregates a set of elements, that are relevant, non-redundant and complementary. We evaluated our approach using the INEX 2009 collection and presented some empirical results for evaluating the impact of the aggregation approach.