

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@irit.fr
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.

1 Introduction

Typiquement, un système de Recherche d'Information (RI) retourne une liste ordonnée des documents en réponse à une requête. Ensuite, l'utilisateur doit les examiner linéairement. Un problème principal dans la RI structurée est comment sélectionner l'unité d'information qui répond le mieux à une requête composée de mots clés (CO : Content Only) (Kamps et al., 2003) (Fuhr et al., 2004). La plupart des approches en RI structurée (Sigurbjornsson et al., 2003), (Ogilvie et Callan, 2003), (Lalmas, 1997), (Lalmas et Vannoorenberghe, 2004), (Piwowski et al., 2002) considèrent que les unités retournées sont sous forme d'une liste d'éléments dis-joints. Nous supposons que cette unité pertinente n'est pas nécessairement des éléments adjacents ou un document, elle pourrait aussi être une agrégation d'éléments de ce document. Afin