N° d'ordre: 10/2012-D/INF

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE
FACULTE D'ELECTRONIQUE ET D'INFORMATIQUE

THESE

Présentée pour l'obtention du grade de DOCTEUR
En : INFORMATIQUE
Spécialité : Intelligence artificielle et Bases de Données avancées

Par : ***Bessai Fatma Zohra née Mechmache***

Sujet

Modèle possibiliste pour la recherche d'informations agrégée dans des corpus de documents semi structurés

Soutenue publiquement, le 12/12/2012, devant le jury composé de :

Mme. AISSANI MOKHTARI Aicha	*Professeur à l'USTHB*	*Présidente*
Mme. ALIMAZIGHI Zaia	*Professeur à l'USTHB*	*Directrice de thèse*
M. BOUGHANEM Mohand	*Professeur à l'IRIT Toulouse*	*Co-Directeur de thèse*
Mme. ABED Hafida	*Professeur à l'U. Blida*	*Examinatrice*
M. KECHID Samir	*Maître de conférences, USTHB*	*Examinateur*
M. AHMED OUAMER Rachid	*Maître de conférences, U. Tizi Ouzou*	*Examinateur*

Résumé

Nous nous sommes intéressés, dans ce travail, à la recherche d'information XML qui vise à récupérer non pas un ensemble de documents pertinents, mais un ensemble d'éléments (parties du document) pertinents par rapport à une requête.
Grâce à la théorie des possibilités et plus particulièrement au réseaux possibilistes, nous proposons un nouveau modèle de recherche d'information XML qui permet une sélection automatique d'un ensemble d'éléments pertinents provenant de différentes parties du document XML. Les relations termes-éléments et éléments-document sont modélisées par des mesures de possibilité et de nécessité. Dans ce modèle, la requête de l'utilisateur déclenche un processus de propagation pour retrouver des portions de documents nécessairement ou au moins possiblement pertinents par rapport à la requête.

Ce modèle permet de revisiter la granularité de l'unité d'information retournée. En effet, au lieu de retourner une liste d'éléments séparément, comme il se fait habituellement dans la plupart des systèmes de recherche d'information XML, notre modèle essaye de générer la meilleure agrégation des éléments pertinents susceptible de répondre au mieux au besoin de l'utilisateur formulé à travers une liste de mots clés.

Mots clés : Recherche d'information XML, Document XML, Théorie des possibilités, Réseaux possibilistes, agrégation des éléments, Recherche d'informations agrégée.

Abstract

We were interested, in this work, in XML information retrieval which aims to retrieve not a set of relevant documents, but a set of elements (parts of document) relevant to a query.
Thanks to possibilistic theory and more especially to possibilistic networks, we propose a new XML information retrieval model, which allows an automatic selection of a set of relevant elements from various parts of XML document. Relations terms-elements and elements-document are modeled through possibility and necessity. In this model, the user's query starts a process of propagation to recover parts of document necessarily or at least possibly relevant.

This model revisits the granularity of the unit of information returned. Indeed, instead of returning a list of elements separately, as it is usually done in most XML information retrieval systems, our model tries to build the best aggregation of relevant elements which is likely to be relevant to a query composed of key words.

Keywords: XML Information Retrieval, XML document, Possibilistic Theory, Possibilistic Network, Elements Aggregation, Aggregated search.

A mes parents, mon mari, mes enfants,
mes sœurs et mon frère.

Cherchons comme cherchent ceux qui doivent trouver et trouvons comme trouvent ceux qui doivent chercher encore. Car il est écrit, celui qui est arrivé au terme ne fait que commencer.

Saint Augustin

Remerciements

Je suis heureuse de pouvoir exprimer mes vifs remerciements au Professeur Nadjib Badache, Directeur du CERIST, pour m'avoir encouragé et soutenu dans la réalisation de cette thèse. Qu'il trouve ici l'expression de mon profond respect.

Je tiens tout particulièrement à exprimer toute ma gratitude et mes vifs remerciements à mes Directeurs de thèse Madame Zaia Alimazighi et Monsieur Mohand Boughanem pour m'avoir encadré durant ces années de thèse et m'avoir aidé à mener à bout ce travail, mais aussi pour leurs précieux conseils et réflexions critiques sur mes propositions qui ont été d'un grand apport pour la finalisation de ce travail. Qu'ils soient assurés de mon très grand respect.

Je remercie également les membres du jury :
Mme Aissani Mokhtari Aicha, professeur à l'université des Sciences et de la Technologie Houari Boumediene pour m'avoir fait l'honneur de présider ce jury; Mme Bouarfa Abed Hafida, professeur à l'université Saad Dahlab de Blida ; Mr Kechid Samir, maître de conférences à l'université Houari Boumediene et Mr Ahmed Ouamer Rachid, maître de conférences à l'université Mouloud Mammeri de Tizi Ouzou pour avoir accepté d'évaluer mon travail.

Pour finir, je remercie mes parents, toute ma famille, mon mari, mes enfants chéris Selma, Mohamed Cherif, Zahir et Nazim ainsi que ma chère amie et sœur Bensefia Hassina pour leur amour et leur soutien.

J'adresse aussi mes remerciements à tous mes amis et collègues du Centre de Recherche sur l'Information Scientifique et Technique (CERIST) pour m'avoir encouragé durant toutes ces années.

TABLES DES MATIERES

Première Partie: Recherche d'Information et Structure

Deuxième Parie: Un Modèle Possibiliste pour la Recherche d'Information Structurée

LISTE DES FIGURES

LISTE DES TABLEAUX