République algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université M'hamed BOUGARA de BOUMERDES
Faculté des sciences
Département Informatique

MEMOIRE

Présenté pour l'obtention du diplôme de Magister
Spécialité:
INFORMATIQUE
Option:
Spécification de logiciels et traitement de l'information

Par : **BOUCHAM Souhila**

Thème

Une approche basée Ontologies pour l'indexation automatique et la Recherche d'Information Multilingue (RIM)

Soutenu devant le jury:

Pr. M. Mezghiche	Professeur, Université de Boumerdes	Président
Dr. O. Nouali	Maître de recherche, CERIST	Examinateur
Pr. Alimzighi Zaia	Professeur, USTHB	Rapporteur
Mme Aliane Hassina	Chargée de recherche / CERIST	Invitée

Année universitaire 2008-2009

Résumé

Notre travail se situe dans le contexte de la *recherche d'information (RI)*, plus particulièrement la recherche d'information mulilingue (RIM).
L'objectif de ce projet est de proposer une solution pour la recherche d'information multilingue afin d'explorer l'apport des approches Web Sémantique en particulier l'utilisation des ontologies pour améliorer la description sémantique des documents et des requêtes.

Nous proposons dans ce travail une approche pour l'indexation et la recherche d'information pour un corpus trilingue : arabe, français et anglais. Le système proposé est fondé sur un formalisme de représentation de connaissances, plus précisément les graphes sémantiques qui supportent une ontologie de domaine. Les documents et les requêtes sont aussi représentés dans ce formalisme.
L'ontologie du domaine constitue le noyau du système et est utilisée aussi bien pour l'indexation que pour la recherche. Le système d'indexation utilise une méthode d'extraction qui est basée sur le calcul de segments répétés en utilisant des filtres linguistiques. Le système de recherche consiste en une comparaison de graphes pour trouver les documents qui répondent à la requête étendue de l'utilisateur.

Mots-clés : *recherche d'information multilingue, Indexation automatique, Ontologie, Extraction de connaissances, expansion de la requête, Graphes sémantiques.*

Abstract

This work deals with information retrieval (IR), more particulary multilingual information retrieval (MIR).
The aim our project is to propose a solution to MIR and explore the contribution semantic web approachs, particulary the the use of ontologies to improve the semantic description of documents and queries.

We propose an approach to indexing and retrieval in a trilingual corpus: arabic, french and english. The proposed system is founded on a knowledge representation formalism, namely semantic graphs which supports a domain ontology. Documents and queries are also represented in this formalism. The domain ontology constitutes the kernel of the system and is used both for indexing and retrieval.
The indexing system uses an extraction method based on repeated segments calculation. When identified, repeated segments are submitted to a filtering procedure using linguistic filters.
The retrieval system consists of a graph comparison to find relevant documents for the extended user query.

Keywords: multilingual information retrieval, automatic indexing, ontology, Extracting information, query expansion, semantic graphs.

Table des matières

Partie II