UNIVERSITÉ MOHAMMED V – AGDAL
FACULTÉ DES SCIENCES
Rabat

N° d'ordre: 2418

TF
20

THÈSE DE DOCTORAT

Présentée par

Nom et Prénom : **BOULAKNADEL Siham**

Discipline : Sciences de l'ingénieur
Spécialité : Informatique et Télécommunications

Titre : «**Traitement Automatique des Langues et Recherche d'Information en langue arabe dans un domaine de spécialité : Apport des connaissances morphologiques et syntaxiques pour l'indexation**»

Soutenue le : 18 Octobre 2008

Devant le jury

Président :

J. Martinez	PES à l'Ecole Polytechnique de l'Université de Nantes, France.

Examinateurs :

A. Hamdani	PES à l'Institut des Etudes et Recherches pour l'Arabisation, Rabat.
J. Mothe	PES à l'Université de Toulouse, France.
B. Daille	PES à l'Université de Nantes, France.
D. Aboutajdine	PES à la Faculté des Sciences de Rabat.
A. Elqadi	PA à l'Ecole Supérieure de Technologie de Meknès

UNIVERSITÉ MOHAMMED V – AGDAL
FACULTÉ DES SCIENCES
Rabat

DOCTORAT

Résumé de la Thèse

Discipline : Sciences de l'ingénieur
Spécialité : Informatique et Télécommunications
UFR : Informatique et Télécommunications
Responsable de l'UFR : Prof .Driss ABOUTAJDINE
Période d'accréditation : 2005-2008

Titre de la thèse : « **Traitement Automatique des Langues et Recherche d'Information en langue arabe dans un domaine de spécialité : apport des connaissances morphologiques et syntaxiques pour l'indexation**»

Prénom, Nom : **Siham Boulaknadel**

Résumé :

La Recherche d'Information a pour objectif de fournir à un utilisateur un accès facile à l'information qui l'intéresse, cette information étant située dans une masse de documents textuels. Afin d'atteindre cet objectif, un système de recherche d'information doit représenter, stocker et organiser l'information, puis fournir à l'utilisateur les éléments correspondant au besoin d'information exprimé par sa requête. La plupart des systèmes de recherche d'information (SRI) utilisent des termes simples pour indexer et retrouver des documents. Cependant, cette représentation n'est pas assez précise pour représenter le contenu des documents et des requêtes, du fait de l'ambiguïté des termes isolés de leur contexte. Une solution à ce problème consiste à utiliser des termes complexes à la place de termes simples isolés. Cette approche se fonde sur l'hypothèse qu'un terme complexe est moins ambigu qu'un terme simple isolé. Notre thèse s'inscrit dans le cadre de la recherche d'information dans un domaine de spécialité en langue arabe. L'objectif de notre travail a été d'une part, d'identifier les termes complexes présents dans les requêtes et les documents. D'autre part, d'exploiter pleinement la richesse de la langue en combinant plusieurs connaissances linguistiques appartenant aux niveaux morphologique et syntaxique, et de montrer comment l'apport de connaissances morphologiques et syntaxiques permet d'améliorer l'accès à l'information. Ainsi, nous avons proposé une plate-forme intégrant divers composants dans le domaine public ; elle conduit à montrer l'apport significatif et tranché de plusieurs de ces composants. En outre, nous avons défini linguistiquement les termes complexes en langue arabe et nous avons développé un système d'identification de termes complexes sur corpus qui produit des résultats de bonne qualité en terme de précision, en s'appuyant sur une approche mixte qui combine modèle statistique et données linguistiques.

Mots-clefs (5): Langue Arabe, Recherche d'information, Terme complexe, Terme simple, Extraction de termes, Traitement automatique de la langue arabe.

Sommaire