

# THÈSE DE DOCTORAT

Présentée par

**Boulaknadel Siham**

**Titre : TAL et RI en langue arabe dans un domaine de spécialité :  
apport des connaissances morphologiques et syntaxiques pour  
l'indexation**

Discipline : Sciences de l'ingénieur

Spécialité : Informatique et Télécommunications

UFR : Informatique et Télécommunications

Période d'accréditation : 2005-2008

Responsable de l'UFR : Prof .Driss ABOUTAJDINE

Directeurs de thèse : Prof. Driss ABOUTAJDINE/ Prof. Béatrice Daille

## **Soutenance :**

Date : **17 Octobre 2008**

Heure : **15h**

Lieu : **Amphi El birouni**

## **Devant le jury :**

### **Président :**

J. Martinez , Professeur à l'Ecole Polytechnique de Nantes.

### **Examineurs :**

A. Hamdani , Professeur à l'Institut des Etudes et de Recherche pour  
l'arabisation, Rabat.

J. Mothe , Professeur à l'Université de Toulouse, Toulouse.

D. Aboutajdine, Professeur à la Faculté des Sciences, Rabat.

B.Daille , Professeur à la faculté des Sciences et Techniques, Nantes.

A. Elqadi, Professeur assistant à l'EST, Meknès

**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 (6):** Langue Arabe, Recherche d'information, Extraction terminologique, Domaine de spécialité, Indexation, Traitement automatique de la langue arabe.

---