

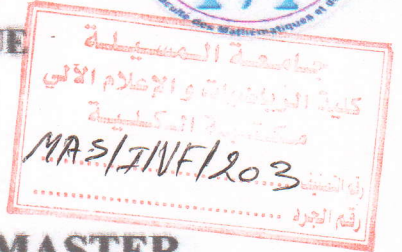
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



**UNIVERSITE MOHAMED BOUDIAF - M'SILA**  
**FACULTE DES MATHÉMATIQUES ET**  
**DE L'INFORMATIQUE**



**DEPARTEMENT D'INFORMATIQUE**



**MEMOIRE de fin d'étude**

**Présenté pour l'obtention du diplôme de MASTER**

**Domaine : Mathématiques et Informatique**

**Filière : Informatique**

**Spécialité : Technologie de l'Information et de Communication**

**Par: KHAOUI Dahbia**

**SUJET**

**AIDE A LA DECISION TEXTUELLE BASEE LOIS**

**Soutenu publiquement le : 31/ 05 /2016 devant le jury composé de :**

**M.BOURAHLA**  
**S.HAMMANI**  
**L.NOUIWA**

**Université de M'sila**  
**Université de M'sila**  
**Université de M'sila**

**Président**  
**Rapporteur**  
**Examineur**

**Promotion : 2015 /2016**

# Table des matières

DEDICACE.....	I
REMERCIEMENTS .....	II
TABLE DES MATIERES .....	III
LISTE DES TABLEAUX ET FIGURES .....	IV
INTRODUCTION GENERAL .....	1

## CHAPITRE 1 : RECHERCHE D'INFORMATION

1 Introduction .....	2
2 La recherche d'information .....	2
3 Processus de recherche d'information .....	2
3.1 Collection de documents (corpus) .....	4
3.2 Besoin en information .....	4
3.3 Indexation .....	4
3.3.1 Les Modes d'indexation .....	4
3.3.2 Prétraitements .....	4
3.3.2.1 Segmentation .....	4
3.3.2.2 Normalisation .....	5
3.3.2.3 Etiquetage morphosyntaxique (optionnel) .....	6
3.3.2.4 Lemmatisation .....	6
3.3.2.5 Dés-suffixation « stemming » .....	6
3.3.2.6 Suppression des mots vides .....	7
3.3.3 Fonction de pondération .....	7
3.3.4 Création de l'index .....	8
3.4 L'appariement requête-document .....	8
4.3.1 Mesure d'appariement .....	8

3.4.1.1 la distance euclidienne .....	8
3.4.1.2 la similarité cosine.....	9
3.5 La fonction de modification de requête.....	9
4 Représentation « les modèles de RI » .....	9
4.1 Le modèle booléen .....	9
4.2 Le modèle probabiliste .....	10
4.3 Le modèle vectoriel.....	10
5 Critères d'évaluation des SRI.....	11
5.1 Rappel.....	11
5.2 Précision .....	11
5.3 F-mesure.....	11
6 Conclusion.....	11

## CHAPITRE 2 : LA LANGUE ARABE

1 Introduction .....	12
2 Morphologie de l'arabe .....	12
2.1. Les lettres .....	12
2.2 Les voyelles.....	13
2.3 Structure d'un mot.....	14
2.3.1 Les antéfixes.....	14
2.3.2 Les préfixes .....	14
2.3.3 Les suffixes .....	14
2.3.4 Les post fixes.....	14
2.4. Catégories d'un mot .....	15
2.4.1 Les verbes.....	15
2.4.2 Les noms .....	16
2.4.3 Les particules.....	17
3 Complexité du traitement automatique de l'arabe .....	17
3.1 L'ambiguïté .....	17
3.2 Problème de la forme et capitalisation .....	18

4 Conclusion.....	18
-------------------	----

## CHAPITRE 3 : CONCEPTION ET IMPLEMENTATION

1 Introduction .....	19
2 La Recherche d'information dans la langue arabe .....	19
2.1 Le corpus utilisé .....	19
3 Système de recherche d'information SRI.....	19
3.1 Formulation de la requête (besoin en information) .....	20
3.2 Indexation.....	20
3.2.1 Prétraitements nécessaire .....	21
3.2.1.1 Segmentation.....	21
3.2.1.2 Normalisation.....	22
3.2.1.3 Suppression des mots vides.....	22
3.2.1.4 Stemming .....	22
3.2.2 Pondération des termes.....	23
3.2.3 L'index .....	23
3.3 L'appariement requête-document.....	23
4 Implémentation.....	23
4.1 NetBeans .....	23
5 Conclusion.....	24
CONCLUSION GENERAL .....	25
BIBLIOGRAPHIE .....	26

## Introduction général

L'homme ne peut pas vivre seul, depuis les temps anciens l'individu vivant dans la communauté avec ses pairs. Afin de construire une société base sur la stabilité et l'ordre il doit exiger des règles juridiques générales contraignantes pour les membres de trouver un équilibre entre les intérêts contradictoires et des libertés de position opposée. [27]

Plusieurs lois dans différent formes (livres, journaux officiels, sites web, etc.) fondée pour réalise ce but et sujet a étés changeant et ça produire un taille énorme d'information et de documents rendre la recherche d'une loi quelques très difficiles. C'est-à-dire la grande quantité d'informations accessibles et les documents disponibles rendent les utilisateurs submergés. En effet, ces utilisateurs ne sont plus capables de choisir l'information ou trouver les documents pertinents qui satisfont ces besoins.

Donc le problème n'est plus d'atteindre aux informations mais de déterminer l'information utile [10]. L'élément clés pour répondre aux besoins de l'utilisateur est d'être capable d'offrir rapidement l'information utile à partir des documents existants. Dans ce contexte il y a des outils pour faire ça comme les systèmes de recherche d'information SRI.

L'objectif principal des Systèmes de Recherche d'Information (SRI) est de répondre au besoin en information des utilisateurs. Les utilisateurs interrogent, au moyen d'une requête, SRI renvoie une liste de documents susceptibles de répondre à leur besoin. [10].

Dans le cadre de ce mémoire, nous nous intéressons à des collections des documents représentent un ensemble des articles extrait de la loi algérienne. Notre problématique est d'aider l'utilisateur à prendre ces décision c'est-à-dire choisir l'article ou les articles d'une loi qui corresponde son problème.

### Organisation du mémoire :

Cette mémoire s'articule autour de trois chapitres :

- Le chapitre 1, présente la recherche d'informations, leur processus général et on va rappeler les différents méthodes utilise dans chaque étapes de ce processus.
- Le chapitre 2, traiter la morphologie et ambigüité de la langue arabe.
- Le chapitre 3, décrit la conception et l'implémentation de notre système d'aide à la décision.

Enfin, une conclusion générale et perspectives futures pour ce travail.

## Conclusion générale

Plusieurs lois fondée pour organier notre vie et ça produire un taille énorme d'information et de documents rendre la recherche d'une loi quelques très difficiles donc les utilisateurs ne sont plus capables de choisir l'information ou trouver les documents pertinents qui satisfont leurs besoins. Notre travail s'inscrit dans le cadre de la recherche d'information dans la langue arabe. L'objectif général de ce mémoire est d'implémenter un système de recherche d'information pour les documents textes écrit en langue arabe. Ce système comporte trois étapes principales (Indexation, La formulation de la requête, l'appariement requête-document). Ainsi que ce système doit prendre en considération la complexité de la morphologie de la langue arabe.

## Perspectives futures

Pour les perspectives de ce travail nous proposons pour l'amélioration de notre système les points suivants :

- L'utilisation d'une ontologie pour les termes juridiques afin de facilite la recherche.
- Utilise la similarité sémantique entre documents-requête.
- Utilise lemmatisation à la place de stemming dans le processus de prétraitement

## BIBLIOGRAPHIE :

- [1] ABBASSI MEFTAH, Un modèle de reformulation des requêtes pour la recherche d'information sur le Web. mémoire de master.
- [2] Abdenour Mokrane. Représentation de collections de documents textuels : application à la caractéristique thématique. Human-Computer Interaction [cs.HC]. Université Montpellier II -Sciences et Techniques du Languedoc, 2006. French.
- [3] Ahmed H. Aliwy, Tokenization as Preprocessing for Arabic Tagging System, International Journal of Information and Education Technology, Vol. 2, No. 4,2012.
- [4] B. Habert, A. Nazarenko and A. Salem. Les linguistiques de corpus. Armand Colin/Masson (edts.). 1997. ISBN : 2200017758. 240p.
- [5] C. Fay-Varnier, C. Fouqueré, G. Prigent et P. Zweingenbaum. Modules syntaxiques des systèmes d'analyse du français. TSI –Techniques et Science Informatiques. Editions AFCET-Bordas, 1991. Volume 10, N°6, pp 403-425.
- [6] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, “An Introduction to Information Retrieval”, Cambridge UP, Cambridge, England, 2009.
- [7] Dhaou Ghoul. Outils génériques pour l''étiquetage morphosyntaxique de la langue arabe : segmentation et corpus d'entraînement. Linguistics. 2011. P 13.
- [8] Dr. S. Vijayarani, Ms. J. Ilamathi, Ms. Nithya ;Preprocessing Techniques for Text Mining - An Overview ; International Journal of Computer Science & Communication Networks, Vol 5(1),7-16,page 9
- [9] Fouad Soufiane Douzidia ; Résumé automatique de texte arabe, Mémoire de M.Sc en informatique ; Université de Montréal Septembre, 2004.

- [10] Fouzi Harrag ; Une approche de fouille des textes basée sur la classification et la segmentation thématique : Application au corpus des Traditions Prophétiques "Hadith", THÈSE Pour obtenir le titre de Docteur en Sciences en INFORMATIQUE, Université Ferhat ABBAS- Sétif;2010-2011
- [11] Gilles Souvay, Jean-Marie Pierrel. LGeRM Lemmatisation des mots en Moyen Français. Traitement Automatique des Langues, ATALA, 2009, 50 (2), pp.21.
- [12] Jason Wexbridge et Walter Nyland, NetBeans Platform for Beginners, Lean Publishing, 31-08-2014
- [13] Kadri.Y and Benyamina.A. (1992). Système d'analyse syntaxicosémantique du langage arabe. mémoire d'ingénieur univercité d'Oran.
- [14] Majed SANAN, Mahmoud RAMMAL et Khaldoun ZREIK, L'accès Multilingue à l'information scientifique et technologique : limitations des moteurs de recherche en langue Arabe, cide [En ligne], CIDE 10, Session Recherche d'information.
- [15] Miloš Radovanović, Mirjana Ivanović, TEXT MINING: APPROACHES AND APPLICATION, Novi Sad J. Math.Vol. 38, No. 3, 2008, 227-234, page 228
- [16] Mohamed Zied Maala, Alexandre Delteil, Amedeo Napoli. Distance sémantique entre concepts définis en ALE. Conférence Francophone Langages et modèles à objets - LMO 07, 2007, Toulouse, France. Hermès/Lavoisier, pp.117–130, 2007.
- [17] Peter D. Turney, Patrick Pantel, From Frequency to Meaning: Vector Space Models of Semantics, Journal of Artificial Intelligence Research, 2010.
- [18] R. Baeza-yates and B. Ribeiro-Neto. Modern Information Retrieval, Addison-Wesley and ACM Press, 1999, ISBN: 0-201-39829-X.

- [19] Rachid Arezki, Pascal Poncelet, Gérard Dray, David William Pearson ; Représentation Dynamique de Documents pour une Recherche Documentaire Intelligente, Centre LGI2P EMA, Site EERIE Parc Scientifique Georges Besse, EURISE, Université Jean Monnet de Saint-Etienne, CORIA05.
- [20] Romaric Besançon, Martin Rajman, Filtrages syntaxiques de co-occurrences pour la représentation vectorielle de documents, Ecublens 1015 Lausanne, 24–27 juin 2002.
- [21] Romaric Besançon, Martin Rajman, Validation de la notion de similarité textuelle dans un cadre multilingue, JADT 2002 : 6es Journées internationales d'Analyse statistique des Données Textuelles, CH-1015 Lausanne – Suisse.
- [22] Salma Jamoussi ; Une nouvelle représentation vectorielle pour la classification sémantique, Multimedia, Information systems and Advanced Computing Laboratory, MIRACL. TAL. Volume 50 – n° 3/2009.
- [23] SALTON G. & MCGILL M. Introduction to Modern Information Retrieval. McGraw Hill. (1983).
- [24] Salton, G. The SMART retrieval system: Experiments in automatic document processing. PRENTICE.
- [25] Tahar DILEKH, Implémentation d'un outil d'indexation et de recherche des textes en arabe, Mémoire de Magister, Système d'Information et de Connaissance (SIC) Université Hadj Lakhdar – Batna, 28 /09 /2011.
- [26] Vincent Claveau, Vectorisation, Okapi et calcul de similarité pour le TAL : pour oublier enfin le TF-IDF, Campus de Beaulieu, 35042 Rennes, France, Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2 : TALN, pages 85–98, Grenoble.
- [27] دكتور عمرو طه بدوي محمد، المدخل لدراسة القانون ، نظرية القانون، جامعة القاهرة-كلية الحقوق، 2007.

## ملخص:

تندرج هذه المذكرة ضمن مجال البحث عن المعلومات في الوثائق النصية المكتوبة باللغة العربية. عملنا يهدف إلى إنجاز نظام بحث عن المعلومات لاسترجاع جميع الوثائق التي لها صلة باستعلام المستخدم، وفي هذا السياق قدمنا نظام البحث عن المعلومات الذي طورناه، والذي يهدف إلى مساعدة المستخدم على اتخاذ قراره أي اختيار بند أو بنود من القانون الذي يناسب مشكلته.

الكلمات المفتاحية: البحث عن المعلومات، اللغة العربية، نظام بحث عن المعلومات.

## Abstract:

This memory comes in the field of Information Retrieval (IR) in textual documents written in Arabic. Our work aims to achieve an information retrieval system for retrieving all documents that are relevant to a user query, In this context we present the information retrieval system that we develop, for the purpose to help the user to make his decisions which means choose article or articles of a law that fits his problem.

**Key words:** Information Retrieval IR, Arabic, Information Retrieval System IRS

## Résumé :

Ce mémoire s'inscrit dans le domaine de la recherche d'information (RI) dans les documents textuels écrits en langue arabe. Notre travail s'attache à réaliser un système de recherche d'information permettant de récupérer tous les documents qui sont pertinents à une requête de l'utilisateur, dans ce contexte nous avons présente le système de recherche d'information qui nous avons développé, dont l'objectif est d'aider l'utilisateur à prendre ses décisions, c'est-à-dire choisir l'article ou les articles d'une loi qui corresponde son problème.

**Mot clés :** Recherche d'information RI, la langue arabe, Système de Recherche d'information SRI.