



**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 la Communication**

**Par : Behloui Selma**

**SUJET**

**Une approche sémantique pour la classification  
des traditions prophétiques**

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

.....	Université de M'sila	Président
Mm. HELASSA Madiha	Université de M'sila	Rapporteur
.....	Université de M'sila	Examineur
.....	Université de M'sila	Examineur

**Promotion : 2015 /2016**

## Table des matières

<b>INTRODUCTION GENERALE</b> .....	1
<b>CHAPITRE 1 : Les caractéristiques de la langue arabe</b>	
1 Introduction .....	3
2 Particularité de la langue arabe .....	4
2.1 a structure d'un mot arabe .....	5
2.2 Les catégories des mots .....	5
2.2.1 Verbe .....	5
2.2.2 Particule .....	5
2.2.3 Nom .....	6
2.3 Morphologie arabe .....	7
2.3.1 Principe .....	7
2.3.2 Objectifs .....	7
• Les schèmes .....	7
• Les racines .....	8
• Les stems .....	8
• Les signes diacritiques .....	9
• Mots dérivés .....	9
• Mot isolés .....	9
• Les affixes .....	9
3 Quelques Problèmes du traitement automatique de la langue arabe	10
2.4 Le problème de la voyellation .....	10
2.5 Le problème de l'agglutination .....	10
2.6 L'extraction de la racine .....	11
4 Conclusion .....	11

<b>CHAPITRE 2 : Fouille de données et Fouille de textes</b>	<b>30</b>
1 Introduction .....	12
2 Fouille de données .....	12
2.1 Définition d'ECD (KDD) .....	12
2.2 processus d'extraction de connaissances à partir des données..	13
2.3 Définitions de la fouille de données .....	14
2.4 Pourquoi la fouille de données ? .....	14
2.5 Les domaines d'application de la fouille de données .....	14
2.6 Difficultés de la fouille de données .....	15
2.7 Tâches de fouille de données .....	16
3 La fouille de textes .....	17
3.1 Définitions de la fouille de textes .....	17
3.2 Les objectifs de fouille de textes .....	18
3.3 Tâches de la fouille de textes .....	18
3.4 Processus de fouille de textes .....	20
3.5 Domaine d'application de la fouille de textes .....	21
4 Conclusion .....	22
<b>CHAPITRE 3 : Corpus des hadiths</b>	<b>35</b>
1 Introduction .....	23
2 Le Hadith .....	23
2.1 La Sunna .....	24
2.2 Composants du hadith .....	24
2.3 Classification des Hadiths.....	26
2.4 Collections importants de hadith .....	28
2.5 Imâm Al-Bukhârî .....	29
2.6 Sahih Al-Bukhârî, le livre le plus juste après le Coran .....	30

3	Les corpus .....	30
3.1	Définition .....	30
3.2	Le corpus en littérature .....	31
3.3	Le corpus en linguistique .....	31
3.4	Le corpus dans la science .....	31
3.5	Domaine d'Utilisation du corpus .....	31
3.6	Conditions pour construire un corpus textuel .....	32
3.7	Méthodologie d'utilisation d'un corpus .....	32
3.8	Les avantages d'une analyse de corpus .....	32
4	Conclusion .....	33
	<b>CHAPITRE 4 : Classification des documents textuels : Etat de</b>	<b>45</b>
	<b>l'art</b>	<b>46</b>
1	Introduction .....	34
2	Définition de la Classification .....	34
2.1	Pourquoi automatiser la classification ? .....	34
2.2	Domaines d'application de la classification .....	35
2.3	Classification bi-classe et multi-classes .....	35
2.4	Classification de textes et Text Mining .....	35
2.5	Hiérarchie des méthodes de classification .....	36
2.6	Les types de classification .....	37
2.6.1	Classification Non supervisé (clustering) .....	37
2.6.2	Classification Supervisé (catégorisation) .....	37
3	Définition de la Catégorisation des textes .....	37
3.1	Comment catégoriser un texte ? .....	38
3.2	Domaine d'application de la Catégorisation de textes .....	39
3.3	Problèmes de la catégorisation de textes .....	40

4	Choix des termes .....	42
4.1.	Représentation en « sac de mots » « bag of words » .....	42
4.2.	Représentation des textes par des phrases .....	43
4.3.	Représentation des textes avec des racines lexicales (stemming) .....	43
4.4.	Représentation des textes avec des lemmes (lemmatisation)	43
4.5.	Représentation des textes avec la méthode des n-grammes	44
5	Calcul de poids .....	44
5.1	Codage TF-IDF .....	45
5.2	TFC .....	45
6	Réduction de la dimension .....	46
6.1	Sélection des termes .....	46
6.2	Extraction d'attributs .....	46
7	Algorithmes d'Apprentissage .....	47
7.1	Algorithme des k-voisins les plus proches .....	47
7.2	Arbre d décision .....	47
7.3	Classificateur bayésien naïf .....	47
7.4	Réseaux de neurones .....	48
7.5	Machines à support vectoriel .....	48
8	Critères d'évaluation d'algorithmes d'Apprentissage .....	48
9	Etat de l'art .....	50
10	Conclusion .....	53
<b>CHAPITRE 5 : Réalisation et expérimentation</b>		
1	Introduction .....	54
2	Le langage et l'environnement de programmation .....	54
2.1	L'environnement de programmation (Visual Studio 2013) .....	54

2.2	Le langage de programmation (C#)	54
3	Présentation de corpus utilisé	55
4	Processus de catégorisation suivi par notre application	56
5	Présentation de l'application réalisée	57
5.1	Interface principale	57
5.2	Prétraitement sur le « Matn »	57
	• Tokenization	59
	• Normalisation	59
	• Mot vide	60
	• Lemmatisation et stemming	61
5.3	L'approche utilisée pour la représentation de corpus	62
5.4	Représentation de texte par tf*Idf et apprentissage	62
5.5	Choix de l'algorithme d'apprentissage	62
5.6	Evaluation des résultats du classifieur	63
6	Conclusion	63
	<b>CONCLUSION GENERALE</b>	<b>64</b>

## INTRODUCTION GENERALE

La recherche en informatique accorde ces dernières années, beaucoup d'importance au traitement des données textuelles. Ceci pour plusieurs raisons : un nombre croissant de collections mises en réseau et distribuées sur le plan international, le développement des infrastructures de communication et d'Internet. Les traitements manuels de ces données s'avèrent très coûteux en temps et en personnel, ils sont peu flexibles et leur généralisation à d'autres domaines est presque impossible ; Pour cela, plusieurs méthodes et techniques de gestion et de traitement d'information ont été développées.

Comme nous le savons, tout le monde a besoin d'informations dans sa vie quotidienne. Nous avons besoin de l'information pour prendre les meilleures décisions possibles. Dans chacune de nos activités personnelles, les décisions sont requises et l'information est nécessaire pour soutenir ces décisions. L'information donc est nécessaire presque dans tous les domaines de la pensée et de l'action humaine.

Le domaine de la fouille de textes (text mining) s'est développé pour répondre à volonté à la gestion par contenu des sources volumineuses de textes. A l'heure actuelle, de nombreux logiciels de classification de textes sont disponibles, ils ont fait l'objet de publications et leurs champs d'application s'élargissent de jour en jour. En général, ces systèmes sont basés sur des algorithmes d'apprentissage automatique.

Avec le développement de l'informatique et l'apparition de la fouille de textes, plusieurs applications autour des traditions du Prophète Muhammad (paix et salut sur lui) ont été développées ; malgré cela il y a absence d'une référence standard utilisable.

Notre travail se situe dans le contexte de l'extraction des connaissances intéressantes et non-triviales à partir des corpus de textes prophétiques en utilisant des techniques issues du domaine de la fouille de texte. Nous nous intéressons principalement au problème de la classification thématique des traditions prophétiques.

### Organisation du mémoire

Pour l'organisation de notre mémoire, nous proposons le plan suivant :

- Une introduction générale donne un aperçu sur le sujet de notre mémoire.
- Le premier chapitre est consacré à la présentation de la langue arabe et ses principales caractéristiques influant sur son traitement automatique.

- Le deuxième chapitre permettra de présenter quelques notions sur la fouille de données et la fouille de textes. Ce chapitre est très didactique et agréable à lire car Il offre un aperçu très large du domaine et les références sont nombreuses et utiles pour le lecteur.
- Dans le troisième chapitre, nous présentons une étude sur les traditions prophétiques et quelques aspects issus des sciences de hadith, aussi nous donnons une généralité sur les corpus textuelle
- le quatrième chapitre donne une présentation générale sur l'apprentissage automatique et la catégorisation des textes, ainsi que les différentes méthodes utilisées dans la littérature.
- Le cinquième chapitre décrit le fonctionnement de l'application suivi par une phase d'évaluation des résultats obtenus à l'aide du classifieur implémenté afin de mesurer sa performance.
- En fin, nous terminerons avec une conclusion qui dégage quelques perspectives importantes qui font suite à ce travail.

Les formes		Des lettres	
Fin	milieu	début	isolée
ب	با	بـ	ب
ت	تا	تـ	ت
ث	ثا	ثـ	ث
ن	نا	نـ	ن

Tableau 1.1 Etat de transcription des lettres arabes.

Une unité lexicale arabe s'écrit avec des consonnes et des voyelles. Les voyelles sont ajoutées au-dessus ou au-dessous des lettres. Elles sont nécessaires à la lecture et à la compréhension correcte d'un texte et elles permettent de différencier des graphèmes ayant la même représentation.

Pour mieux comprendre prenons l'exemple de «*ب*» du (Tableau 1.2) [23].

- *ب* - Il a écrit
- *با* - il a été écrit
- *بـ* - Livres

## CONCLUSION GENERALE

La classification de textes s'est avérée au cours des dernières années comme un domaine majeur de recherche pour les entreprises comme pour les particuliers. La langue arabe est une langue riche morphologiquement ce qui ouvre les portes à plusieurs recherches dans le domaine de fouille de textes et de TALN.

Cette vue d'ensemble du domaine de la fouille de données et particulièrement la fouille de données textuelles nous permet d'affirmer que ce domaine de recherche est mûr, il est maintenant possible, en utilisant différents outils, de faire ressortir des informations de manière automatique à partir d'un ensemble de textes, sans même les lire.

Malgré toutes les difficultés qu'on les a rencontrés telle que : le manque de références et des travaux concerne la classification des textes arabes et spécialement la classification des traditions prophétiques et d'outils spécialisés, la non suffisance du temps, aussi la difficulté liées au traitement automatique de la langue Arabe. Mais, nous pensons qu'on a quand même pu relever le défi. Et par la même occasion, apprendre beaucoup de nouvelles connaissances tout au long de la réalisation de ce travail telles que :

- ✓ La programmation objet sous le langage C #.
- ✓ Des nouveaux concepts dans des domaines d'actualité tels que : le Data Mining, le Text Mining, et l'apprentissage automatique,
- ✓ La classification automatique supervisée de documents qui se situe dans l'intersection des domaines précités.
- ✓ Le traitement automatique du langage naturel TALN (en particulier la langue arabe).

Toutefois, le sujet abordé étant très vaste, il reste beaucoup à faire pour améliorer notre système. A cet effet, nous proposons comme perspectives :

- ✓ Utiliser des corpus d'apprentissage et de test de grandes tailles pour donner plus de crédibilité aux résultats obtenus.
- ✓ Utiliser d'autres formes de documents (HTML, XML, ...).
- ✓ Etudier le cas de la classification multi-classe et la classification non supervisée (clustering).
- ✓ Nous envisageons de partitionner les traditions prophétiques non plus en hadith Sahih El bukhari mais en autres Sahih ou en Charh de Hadith, cela pourrait résoudre le problème de l'homogénéité des documents.

Pour conclure, ce travail peut être amélioré au future par l'ajout des ontologies puisque ils sont devenues des outils importants pour structurer la connaissance pour des résultats plus sûr, ou bien l'élaboration d'une ontologie pour les documents prophétique ce qui va permettre de d'établir des liens entre les différentes branches de la culture islamique comme le Hadith, le Coran, la Jurisprudence et les sciences de la langue arabe.

[3] Amel KHEMAKHEM, *ArabiNet 2008 : une base lexicale normalisée pour la langue arabe*, mémoire de MASTER en Systèmes d'Information et Nouvelles Technologies, Université de Sfax Faculté des Sciences Economique et de Gestion, 2008.

[4] Akila DJEBBAR-ZAIDI, *Optimisation de la recherche d'articles Bayésien*, Thèse de doctorat en Informatique, université Hadji Mokhtar Annabi, 2013.

[5] Ali El Akadi, *contribution à la sélection de variables pertinentes en classification supervisée : Application à la sélection des gènes pour les puces à ADN et des caractéristiques faciales*, mémoire de doctorat d'informatique, Université Mohammed V - Agdal, Rabat, 2012.

[6] Amroui Karima, *de couverte des dépendances fonctionnelles floues dans des modèles relationnels sous imprécision*, mémoire de magister d'informatique, Université Mohamed Bouguet de Boumerdes, 2009.

[7] Basil Meriem, *optimisation de la technique des n-gramme dans l'extraction des racines en langue arabe*, mémoire de master d'informatique, université de m'ilia, 2013.

[8] DAHOUMI Fares, *Identification de la langue et catégorisation thématique des textes d'un corpus multilingue en utilisant les signatures : KJ, SVM*, mémoire de master d'informatique, Université De M'ilia, 2013.

[9] Fouad Soufiane Derzidie, *Récupération automatique de texte arabe*, Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de M.Sc en Informatique Septembre, 2004.

[10] Harug Fouzi, *Modélisation d'expertise dans les bases de données : application à l'authentification du " oudith "*, Université Fehat Abbas Saïf, 2005.

## BIBLIOGRAPHIE

- [1] Abd El Salam AL HAJJAR, Extraction et gestion de l'information à partir des documents arabes, mémoire de Doctorat d'informatique, Université paris – saint dénis, Décembre 2010.
- [2] ABIDI Karima, la catégorisation de texte multilingue, mémoire de magistère d'informatique, Ministère De L'enseignement Supérieur Et De La Recherche Scientifique Ecole Supérieur D'informatique, 2011.
- [3] Aïda KHEMAKHEM, ArabicLDB : une base lexicale normalisée pour la langue arabe mémoire de MASTER en Systèmes d'Information et Nouvelles Technologies, Université de Sfax Faculté des Sciences Economique et de Gestion, 2006.
- [4] Akila DJEBBAR-ZAIDI, Optimisation de la recherche d'un cas Bayésien, Thèse de doctorat en Informatique, université Badji Mokhtar Annaba, 2013.
- [5] Ali El Akadi, contribution à la sélection de variables pertinentes en classification supervisé : Application à la sélection des gènes pour les puces à ADN et des caractéristiques faciales, mémoire de doctorat d'informatique, Université Mohammed V – Agdal, Rabat, 2012.
- [6] Amroun Karima, découverte des dépendances fonctionnelles floues dans des modèles relationnels sous imprécision, mémoire de magister d'informatique, Université M'hamed Bougara de Boumardas, 2008.
- [7] Baali Meriem, utilisation de la technique des n-gramme dans l'extraction des racines en langue arabe, mémoire de master d'informatique, université de m'sila, 2015.
- [8] DAHOUMI Fares, Identification de la langue et catégorisation thématique des textes d'un corpus multilingue en utilisant les algorithmes : NB, SVM, mémoire de master d'informatique, Université De M'sila, 2013.
- [9] Fouad Soufiane Douzidia, Résumé automatique de texte arabe, Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de M.Sc en informatique Septembre, 2004.
- [10] Harrag fouzi, Modélisation d'expertise dans les bases de données : application à l'authentification du ' hadith '. Université Ferhat Abbes Sétif, 2005.

- [11] Harrag Fouzi, Une approche de fouille des textes basée sur la classification et la segmentation thématique : Application au corpus des Traditions Prophétiques "Hadith", mémoire de doctorat d'informatique, Université Ferhat ABBAS, Sétif, 2011.
- [12] HELASSA MADIHA, Extraction de connaissances à partir de données : application au hadith, mémoire de Magister Université Ferhat Abbas-Sétif, 2012.
- [13] LAMICHE Chaabane, fusion et fouille de données guidées par les Connaissances : application à l'analyse d'image, mémoire de doctorat d'informatique, Université Mohamed khider – Biskra, 2013.
- [14] Lazhar FAREK, Identification d'opinions dans les journaux arabes, mémoire de magister d'informatique, Université Badji Mokhtar–Annaba, 2009.
- [15] MATALLAH Hocine, Classification Automatique de Textes Approche Orientée Agent, mémoire de Magister en informatique, Université Aboubekr Belkaid-tlemcen, 2011.
- [16] Nicolas Béchet, Extraction et regroupement de descripteurs morphosyntaxiques pour des processus de Fouille de Textes, thèse de doctorat, Université des Sciences et Techniques du Languedoc, 2008.
- [17] Radwan JALAM, Apprentissage automatique et catégorisation de textes multilingues, thèse de doctorat, université Lumière Lyon2, Année 2003.
- [18] Rahmani Rabah, découvert d'associations sémantique dans les bases de données relationnelles par des méthodes de data mining, mémoire de magister d'informatique, Université mouloud mammeri tizi-ouzou, 2001
- [19] SASSI Amina, Une approche basée agent pour la fouille de données, mémoire de Magister en Informatique, Université HADJ LAKHDAR – BATNA, 2013.
- [20] Siham Boulaknadel, 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, mémoire de doctorat d'informatique, Université de Nantes, octobre 2008.

- [21] SIMON RÉHEL, catégorisation automatique de textes et cooccurrence de mots provenant de documents non étiquetés, Mémoire présenté à la Faculté des études supérieures de Université Laval dans le cadre du programme de maîtrise en informatique pour l'obtention du grade de maître ès sciences (M.Sc.), 2005.
- [22] Slim MESFAR, Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard, mémoire de doctorat d'informatique, Université de Franche-Comté, novembre 2008.
- [23] Sofiane baloul, Développement d'un système automatique de synthèse de la parole à partir de texte arabe standard voyelle, mémoire de doctorat d'informatique, Université du Marie-France, 2003.
- [24] Soraya Zaidi-Ayad, Une plateforme pour la construction d'ontologie en arabe : Extraction des termes et des relations à partir de textes (Application sur le Saint Coran), mémoire de Doctorat en Informatique, Université. Badji Mokhtar, Annaba, 2013.
- [25] Tahar DILEKH, Implémentation d'un outil d'indexation et de recherche des textes en arabe, mémoire de Magister en Système d'Information et de Connaissance (SIC), Université Hadj Lakhdar – Batna, 2011.
- [26] Taoufik Guernine, classification hiérarchique floue basée sur le SVM et son application pour la catégorisation des documents, Mémoire présente au Département d'informatique En vue de l'obtention du grade de maître ès sciences (M.sc.), Université de Sherbrooke ,sherbrooke, québec,Canada,14 avril 2010.
- [27] Yannick Toussaint, Fouille de textes : des méthodes symboliques pour la construction d'ontologies et l'annotation sémantique guidée par les connaissances, Mémoire de doctorale en informatique, Université Henry Poincaré, Nancy 1, 21 novembre 2011

## Article

- [28] Marti.A.Hearst. Automatic acquisition of hyponyms from large text corpora. In Proceedings of the Fourteenth International Conference on Computational Linguistics, Nantes, France, juillet 1992, pages 539-545.
- [29] Moath Najeeb, Abdelkarim Abdelkader, Musab Al-Zghoul, Abdelrahman Osman, A Lexicon for Hadith Science Based on a Corpus, Moath Najeeb et al, / (IJCSIT) International

Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, pages 1336-1340

[30] Mohammed N. Al-Kabi, Heider A. Wahsheh, Izzat M. Alsmadi, Abdallah Moh'd Ali Al-Akhras, Extended Topical Classification of Hadith Arabic Text, International Journal on Islamic Applications in Computer Science And Technology, Vol. 3, Issue 3, September 2015, pages 13-23.

## Webographie

[31] VOLLE Michel. "Intranet et Datamining". Article en ligne 2001. Disponible sur [www.volle.com/lectures/ACM1.htm](http://www.volle.com/lectures/ACM1.htm). Consulté le : 07.05.2016.

[32] <https://www.visualstudio.com/products/visual-studio-community-vs> Consulté le : 15.05.2016.

[33] Mohammad Arshi Saloot· Norisma Idris· Rohana Mahmud· Salinah Ja'afar· Dirk Thorleuchter· Abdullah Gani, Hadith data mining and classification : a comparative analysis, Article en ligne 2016.

[http://www.researchgate.net/publication/290222515\\_Hadith\\_data\\_mining\\_and\\_classification\\_a\\_comparative\\_analysis](http://www.researchgate.net/publication/290222515_Hadith_data_mining_and_classification_a_comparative_analysis) Consulté le : 15.05.2016

[34] Mohamed Amine, Chérargui Youssef Hceini et Moncef Abbas, Une approche multicritère pour lever l'ambiguïté morphologique dans le texte arabe pages 356-367, Disponible sur : [http://dspace.univ-ouargla.dz:8080/jspui/bitstream/123456789/2820/1/Mohamed% 20Amine.pdf](http://dspace.univ-ouargla.dz:8080/jspui/bitstream/123456789/2820/1/Mohamed%20Amine.pdf) Consulté le : 15.03.2016

[35] Fathi DEBILI, Voyellation automatique de l'arabe. 42-49, Disponible sur : <https://www.aclweb.org/anthology/W/W98/W98-1006.pdf> Consulté le : 15.03.2016

[36] Djamel Abdelkader ZIGHED & Ricco RAKOTOMALALA, Extraction des Connaissances à partir des Données (ECD), disponible sur :

<http://www.techniques-ingenieur.fr/base-documentaire/technologies-de-l-information-th9/bases-de-donnees-42309210/extraction-de-connaissances-a-partir-de-donnees-eed-h3744/>  
Consulté le : 18.04.2016

[37] [http://www.lattice.cnrs.fr/sites/itellier/poly\\_fouille\\_textes/fouille-textes.pdf](http://www.lattice.cnrs.fr/sites/itellier/poly_fouille_textes/fouille-textes.pdf)  
Consulté le : 18.04.2016

- [38] Shafi Mohammad, The HADITH - How it was Collected and Compiled, Disponible sur: <http://www.daralislam.org/portals/0/Publications/TheHADITHHowitwasCollectedandCompiled.pdf> Consulté le : 22.04.2016
- [39] Julien Élie, Éléments de réflexion sur le saint Coran, Version du 24 octobre 2010. Disponible sur : <http://www.trigofacile.com/divers/spiritualite/pdf/reflexions-coran.pdf> Consulté le : 16.04.2016
- [40] Xavier Polanco, text mining et intelligence économique, Disponible sur : <http://lpvist.free.fr/IE/Text%20mining%20et%20IE.pdf> Consulté le : 15.03.2016
- [41] <http://scialert.net/fulltext/?doi=jas.2005.584.587&org=11> Consulté le : 15.03.2016
- [42] [https://www.researchgate.net/profile/Mohammed\\_Al-Kabi/publication/241109392\\_A\\_COMPARATIVE\\_STUDY\\_OF\\_THE\\_EFFICIENCY\\_OF\\_DIFFERENT MEASURES TO CLASSIFY ARABIC TEXT/links/00b495345aa866e000000.pdf](https://www.researchgate.net/profile/Mohammed_Al-Kabi/publication/241109392_A_COMPARATIVE_STUDY_OF_THE_EFFICIENCY_OF_DIFFERENT_MEASURES_TO_CLASSIFY_ARABIC_TEXT/links/00b495345aa866e000000.pdf) Consulté le: 22.05.2016.
- [43] [http://link.springer.com/chapter/10.1007%2F978-3-642-22191-0\\_26#page-1](http://link.springer.com/chapter/10.1007%2F978-3-642-22191-0_26#page-1) Consulté le : 22.05.2016.
- [44] [https://www.researchgate.net/publication/276854961\\_Towards\\_Innovative\\_System\\_for\\_Hadith\\_Isnad\\_Processing](https://www.researchgate.net/publication/276854961_Towards_Innovative_System_for_Hadith_Isnad_Processing) Consulté le: 22.05.2016
- [45] <http://www.ijitcs.com/2ndicekmt/Kawther+AAldhaln.php> Consulté le : 22.05.2016.
- [46] <http://sunnah.com/bukhari/about>. Consulté le : 20.04.2016.
- [47] [http://www.memoireonline.com/12/09/2917/m\\_Algorithmes-dapprentissage-pour-la-classification-de-documents0.html](http://www.memoireonline.com/12/09/2917/m_Algorithmes-dapprentissage-pour-la-classification-de-documents0.html) Consulté le : 20.04.2016.

## ملخص

على الرغم من العدد الهائل للمتكلمين باللغة العربية إلا أنها تحوي ظواهر جَدَّ خاصة في المورفولوجيا والنحو، هذه الخصوصية ترجع بصفة خاصة إلى تركيبها الإعرابية وتعدّد أشكالها وهذا أدى إلى صعوبة المعالجة الآلية لها. الهدف من هذا البحث هو دراسة إحدى الطرق لتصنيف المعارف في قواعد المعطيات النصّية مثل تلك الخاصة بالأحاديث النبوية الشريفة هذه الطريقة تمثل عملية لمعالجة مدوّنة الأحاديث النبوية الشريفة.

في هذه الدّراسة استعملنا المصنّف الآلي (Naïve Bayes) المعروف بكفاءته في تصنيف الأحاديث النبوية الشريفة، معتمدين في ذلك على تصنيف صحيح البخاري  
الكلمات المفتاحية: اكتشاف المعارف، التقيب عن البيانات، التقيب عن النصوص، التلقين الذاتي، اللغة العربية، مدوّنة الأحاديث النبوية الشريفة

## Abstract

Even though, The Arabic language has a very important number of speakers, it present special morpho-syntactic phenomena. This particularity is mainly related to the inflectional and agglutinative morphology, and the multiplicity of its forms, this leads to a difficulties in the automatic processing. This research aims to study methods of classification in textual databases such as those of Prophet Sayings (Hadiths). This method provide a process of treatment for the corpora of prophetic traditions.

In this study, we use Naïve Bayes classifieur which known by its effectiveness to classify the prophetic sayings based on Sahih Al-Bukhârî classification.

**Keywords:** Knowledge discovery, Data mining, Text mining, Machine Learning, Texts classification, Arabic language, Prophetic sayings corpora.

## Résumé

La langue arabe, bien que très importante par son nombre de locuteurs, elle présente des phénomènes morpho-syntaxiques très particulières. Cette particularité est liée principalement à sa morphologie flexionnelle et agglutinante, et à la multiplicité de ses formes ; cela induit des difficultés de traitement automatique. Cette recherche a pour but d'étudier une méthode de classification des connaissances dans les bases de données textuelles telles que celles des Traditions Prophétiques (Hadiths). Cette méthode représente un processus de traitement du corpus des traditions prophétiques.

Dans cette étude, nous avons utilisé le classifieur Naïve Bayes qui est connu par son efficacité pour classifier les traditions prophétiques en se basant sur la classification de Sahih Al-Bukhârî.

**Mots clés:** Découverte de connaissances, Fouille de données, Fouille de textes, Apprentissage Automatique, Classification de textes, Langue arabe, Corpus des traditions prophétiques,