

LISTE DES FIGURES

Fig. 1.1. Composants du Hadith.	11
Fig. 1.2. Classification du Hadith.	12
Fig. 1.3. Résultat du processus d'authentification.	14
Fig. 1.4. Interface du logiciel «Encyclopédie des Neufs Livres».....	17
Fig. 1.5. Interface du logiciel «l'Encyclopédie Dorée des Traditions Prophétiques».	18
Fig. 1.6. Interface du logiciel «Bibliothèque Moderne des Traditions Prophétiques».	19
Fig. 1.7. Interface du logiciel «Encyclopédie Aisée de la Tradition Prophétique et ses Sciences».....	20
Fig. 1.8. Interface du logiciel «Al-Muhaddith».	21
Fig. 1.9. Interface du Site web du projet «Encyclopédie de Hadith».	22
Fig. 2.1. Processus de la fouille de données textuelles.....	28
Fig. 2.2. Un système de recherche d'information typique.	30
Fig. 2.3. Représentation d'un texte en fichier inverse.....	33
Fig. 2.4. Partition de la collection pour une requête.	39
Fig. 2.5. Courbes de rappel-précision pour les systèmes S_1 et S_2	40
Fig. 3.1. Structure typique d'un Hadith du « <i>Sahîh Al-Boukhârî</i> ».....	43
Fig. 3.2. Modèle d'extraction d'informations pertinentes dans le corpus « <i>Sahîh Al-Boukhârî</i> ».....	44
Fig. 3.3. Automate de concepts pour la production des séquences dans le modèle d'extraction d'information dans le corpus « <i>Sahîh Al-Boukhârî</i> ».....	45
Fig. 3.4. Indexation et recherche de Hadiths dans le modèle vectoriel	47
Fig. 3.5. Schéma du processus d'indexation.....	50
Fig. 3.6. Création de la matrice Doc-Terme	51
Fig. 3.7. Le cosinus comme mesure de similarité	52
Fig. 3.8. Vérification expérimentale de la loi de Zipf sur le corpus « <i>Sahîh Al-Boukhari</i> ».....	55
Fig. 3.9. Informations du Hadith H_{33} (Requête R_1).....	56
Fig. 3.10. Informations du Hadith H_{129} (Requête R_2).....	57
Fig. 3.11. Informations du Hadith H_{5834} (Requête R_3).	57
Fig. 3.12. Courbe Précision pour les trois requêtes R_1 , R_2 et R_3	60
Fig. 3.13. Courbe Rappel pour les trois requêtes R_1 , R_2 et R_3	60
Fig. 3.14. Courbe Rappel/Précision pour les deux requêtes R_1 , R_2 et R_3	61
Fig. 4.1. Processus de catégorisation de texte.....	65
Fig. 4.2. Maximisation de la marge avec les SVM.	72
Fig. 5.1. Un exemple de l'arbre de décision résultant de l'opération de division des attributs.....	78
Fig. 5.2. Choix du meilleur seuil de sélection d'attributs pour le corpus Hadith.....	81
Fig. 5.3. Choix du meilleur seuil de sélection d'attributs pour le corpus Scientifique.....	83
Fig. 5.4. Influence du critère «sélection d'attributs» sur les performances du classificateur pour le corpus Hadith.....	83
Fig. 5.5. Influence des critères de sélection d'attributs sur les performances du classificateur pour un corpus scientifique de 280 documents	84

Fig. 5.6. Les résultats de classification du corpus Hadith pour un ensemble d'apprentissage de "150" documents.....	85
Fig. 5.7. Les résultats de classification du corpus Hadith pour un ensemble d'apprentissage de "191" documents.....	85
Fig. 5.8. Les résultats de classification du corpus Hadith pour un ensemble d'apprentissage de "239" documents.....	86
Fig. 5.9. Les résultats de classification du corpus Hadith pour un ensemble d'apprentissage de "266" documents.....	86
Fig. 5.10. Les résultats de classification du corpus Hadith pour un ensemble d'apprentissage de "340" documents.....	87
Fig. 5.11. Les résultats de classification du corpus Scientifique pour un ensemble d'apprentissage de "100" documents.....	87
Fig. 5.12. Les résultats de classification du corpus Scientifique pour un ensemble d'apprentissage de "140" documents.....	88
Fig. 5.13. Les résultats de classification du corpus Scientifique pour un ensemble d'apprentissage de "180" documents.....	88
Fig. 5.14. Les résultats de classification du corpus Scientifique pour un ensemble d'apprentissage de "200" documents.....	89
Fig. 5.15. Les résultats de classification du corpus Scientifique pour un ensemble d'apprentissage de "280" documents.....	89
Fig. 5.16. Influence du critère de la taille d'ensemble d'apprentissage sur les performances du classificateur pour le corpus Hadith.....	90
Fig. 5.17. Influence du critère de la taille d'ensemble d'apprentissage sur les performances du classificateur pour le corpus Scientifique.....	90
Fig. 5.18. Les résultats moyens des mesures d'évaluation pour les classes du corpus Hadith.....	91
Fig. 5.19. Les résultats moyens des mesures d'évaluation pour les classes du corpus scientifique.....	92
Fig. 5.20. Comparaison des résultats d'évaluation pour les deux corpus.....	92
Fig. 6.1. Un réseau de neurones (MLP) typique de trois couches à rétro-propagation du gradient de l'erreur.....	98
Fig. 6.2. Structure typique d'un réseau de neurones RBF.....	98
Fig. 6.3. Vue d'ensemble du classificateur de texte neuronale basé sur la technique SVD.....	100
Fig. 6.4. Projection des 200 premiers facteurs du l'ensemble de 1065 attributs.....	104
Fig. 6.5. Proportion accumulée des facteurs singuliers produite par SVD.....	104
Fig. 6.6. ECM en fonction d'époques pour la classification des textes prophétiques en utilisant le modèle original du réseau de neurones MLP.....	105
Fig. 6.7. ECM en fonction d'époques pour la classification des textes prophétiques en utilisant le modèle original du réseau de neurones RBF.....	106
Fig. 6.8. ECM en fonction d'époques pour la classification des textes prophétiques en utilisant le modèle à base SVD du réseau de neurones MLP.....	106
Fig. 6.9. ECM en fonction d'époques pour la classification des textes prophétiques en utilisant le modèle à base SVD du réseau de neurones RBF.....	107
Fig. 6.10. Macro-Rappel en fonction du nombre de dimension pour les deux modèles du réseau de neurones MLP.....	108

Fig. 6.11. Macro-Rappel en fonction du nombre de dimension pour les deux modèles du réseau de neurones RBF.	108
Fig. 6.12. Macro-Précision en fonction du nombre de dimension pour les deux modèles du réseau de neurones MLP.....	109
Fig. 6.13. Macro-Précision en fonction du nombre de dimension pour les deux modèles du réseau de neurones RBF.	109
Fig. 6.14. Macro-F1 en fonction de dimension pour MLP.....	110
Fig. 6.15. Macro-F1 en fonction de dimension pour RBF.	110
Fig. 6.16. Comparaison des résultats de classification pour MLP.	111
Fig. 6.17. Comparaison des résultats de classification pour RBF.....	111
Fig. 7.1. Principe de la segmentation thématique.	116
Fig. 7.2. Principe de Dotplotting.....	119
Fig. 7.3. Architecture de segmenter.....	121
Fig. 7.4. Construction des chaînes lexicales	122
Fig. 7.5. Sélection et pondération des mots du réseau de collocations.....	128
Fig. 8.1. La phase de pré-traitement	131
Fig. 8.2. Le principe de la méthode TextTiling.....	132
Fig. 8.3. Calcul du score de cohésion (profondeur).....	134
Fig. 8.4. Le score de cohésion dans 3 situations critiques.....	134
Fig. 8.5. Application de la convolution.....	135
Fig. 8.6. Principe générale de la méthode C99	136
Fig. 8.7. Exemple d'une matrice de similarité.....	137
Fig. 8.8. Représentation graphique de la matrice de similarité	137
Fig. 8.9. Calcul de la matrice de rang à partir de la matrice de similarité.....	138
Fig. 8.10. Représentation visuelle : à gauche de la matrice de similarité ; à droite de la matrice de rang	138
Fig. 8.11. Les ruptures des lecteurs contre les ruptures des algorithmes.....	141
Fig. 8.12. Les résultats d'évaluation des jugements des lecteurs	142
Fig. 8.13. Les résultats d'évaluation pour l'algorithme TextTiling.....	143
Fig. 8.14. Les résultats d'évaluation pour l'algorithme C99.....	144
Fig. 9.1. Niveau de Rappel pour la requête R3 avec et sans expansion	150
Fig. 9.2. Niveau Précision pour la requête R3 avec et sans expansion.....	151
Fig. 9.3. Courbes Rappel/Précision pour la requêtes R3 avec et sans expansion.....	151
Fig. 9.4. Segmentation thématique du document D1 en réponse à la Requête R6 du premier jeu de requêtes.....	153
Fig. 9.5. Segmentation thématique d'un document D2 en réponse à la Requête R4 du deuxième jeu de requêtes.....	153
Fig. 9.6. Niveau de Rappel pour le corpus C1 avec et sans segmentation	155
Fig. 9.7. Niveau de Rappel pour le corpus C2 avec et sans segmentation	155
Fig. 9.8. Niveau de Précision pour le corpus C1 avec et sans segmentation.....	156
Fig. 9.9. Niveau de Précision pour le corpus C2 avec et sans segmentation.....	156
Fig. 9.10. Comparaison des deux corpus	157
Fig. 9.11. Comparaison des valeurs F1 pour la Requête $n=4$ du corpus C1.....	157
Fig. 9.12. Comparaison des courbe Rappel/précision pour la Requête $n=4$ du corpus C1.....	158
Fig. 9.13. Modèle proposé pour l'organisation de l'information pertinente	160
Fig. 9.14. répartition des segments pertinents par catégorie thématique.....	162

Fig. 9.15. l'évaluation des scores de similarité de chaque catégorie thématique avec la requête R3 (H5834)	163
Fig. 9.16. Niveau de Rappel pour la requête R3 avec et sans classification.....	164
Fig. 9.17. Niveau de Rappel pour la requête R3 avec et sans classification.....	164
Fig. 9.18. Courbes Rappel/Précision pour la requêtes R3 avec et sans classification.....	164