



## Utilisation des ressources externes pour la reformulation des requêtes dans un système de recherche d'information

Mohammed El Amine Abderrahim

Laboratoire de Traitement Automatique de la Langue Arabe  
Faculté de Technologie, Université de Tlemcen  
Algérie

---

### Abstract

Dans un Système de Recherche d'Information (SRI), les démarches pour la reformulation de la requête sont nombreuses. Elles peuvent être classées selon les ressources utilisées en trois grandes approches : l'utilisation des ressources externes, l'analyse globale et l'analyse locale. Dans ce contexte et dans le cadre des SRI pour les textes Arabes, nous nous intéressons à l'évaluation des performances de la première approche. A cet effet deux ressources différentes ont été utilisées à savoir : WordNet Arabe et le Dictionnaire (thésaurus) des Sens Arabe. Les expérimentations réalisées sur un corpus de texte Arabe nous ont permis de mesurer l'apport de cette approche de reformulation de requête dans un SRI arabe.

---

### 1. Introduction

Afin de réduire la distance entre la pertinence système et la pertinence utilisateur, un SRI peut guider l'utilisateur vers une bonne formulation de ses besoins. Les solutions proposées tournent autour de trois approches à savoir : la Reformulation de la requête (RQ), le ré-ordonnancement des documents, la combinaison des résultats issus de différents SRI ou l'intégration du profil utilisateur dans le processus de recherche d'information. Dans cet article on s'intéresse particulièrement à la RQ. Les démarches pour cette dernière sont nombreuses et peuvent être classées selon les ressources utilisées en trois grandes classes (voir figure 1) :

1. L'analyse globale : cette approche consiste à analyser tout l'ensemble des documents de la collection pour extraire les termes pertinents à ajouter à la re-

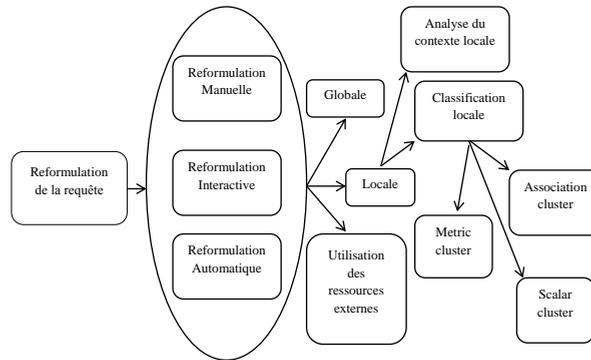


Fig. 1 : Les approches pour la reformulation de la requête

quête initiale. Deux techniques sont alors développées (Baeza-Yates and Berthier, 1999) : le thesaurus de similarité (similarity thesaurus) et le thesaurus statistique (statistical thesaurus).

2. L'analyse locale : les documents retournés en réponse à une requête sont analysés pour extraire des termes pertinents qui serviront à étendre la requête. Les études effectuées dans (Baeza-Yates and Berthier, 1999 ; Carpineto and Romano, 2012 ; Aalbersberg, 1992 ; Xu and Croft, 2000 ; Lee et al., 2008 ; Salton and Buckley, 1990) montrent que contrairement à l'analyse globale, l'analyse locale est plus simple à réaliser et permet d'améliorer les performances d'un SRI. Deux techniques pour l'analyse locale sont proposées dans la littérature (Baeza-Yates and Berthier, 1999) :
  - La classification locale (local clustering) : consiste à construire une matrice d'association qui quantifie les relations de corrélation entre les termes issus de l'ensemble des documents retournés en réponse à la requête initiale. Selon la méthode de construction des relations de corrélation on distingue trois types de clusters : association clusters, metric clusters et scalar clusters.
  - L'analyse du contexte local : consiste à utiliser les concepts à la place de mots-clés pour représenter les documents (Xu and Croft, 2000).
3. L'utilisation des ressources externes : consiste à utiliser les ressources externes comme les ontologies ou les thésaurus pour trouver des termes similaires à la requête initiale (Efthimiadis, 1996 ; Harb et al., 2011). Nous nous intéressons dans la suite de cet article à cette technique de reformulation.

En plus de la reformulation manuelle, nous distinguons deux manières pour la RQ, la première est basée sur un processus automatique, elle se déroule sans l'intervention de l'utilisateur, on parle de reformulation automatique de la requête, alors que

la seconde est basée sur un processus interactif entre le SRI et l'utilisateur, on parle de reformulation interactive de la requête. Les expérimentations effectuées dans le cadre de cette dernière ont montré qu'elle permet d'améliorer la précision des résultats, néanmoins son efficacité reste fortement lié à la disposition des utilisateurs et leurs aptitudes à juger la pertinence des documents (Baeza-Yates and Berthier, 1999 ; Bodo, 2005 ; Hlaoua, 2007 ; Bruande and Chevallet, 2003 ; Salton and Buckley, 1990 ; Black et al., 2006).

## 2. La reformulation de la requête par utilisation d'une ressource externe

La RQ par utilisation d'une ressource externe consiste à analyser premièrement la requête pour détecter les termes qui renvoient à des concepts de l'ontologie ou du thésaurus. Ces termes seront donc remplacés par des concepts proches en utilisant les relations sémantiques de l'ontologie.

Dans notre cas nous avons exploité le contenu de WordNet Arabe (AWN)<sup>1</sup> ou du Dictionnaire (thésaurus) des Sens Arabe (DSA)<sup>2</sup> pour reformuler et étendre des requêtes (par expansion) de manière à retrouver plus précisément les bons documents (voir figure 2). Dans le cadre de notre expérimentation nous avons testé la relation de synonymie.

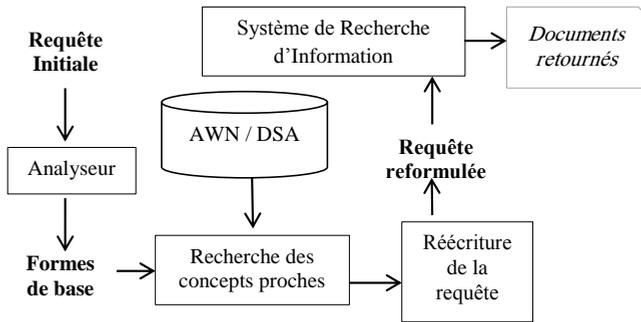


FIG. 2 : Les étapes de la reformulation de la requête

Il faut remarquer que le processus de modification de requêtes est indépendant du SRI, autrement dit l'enrichissement ne modifie pas la représentation interne des documents et des requêtes du SRI. Le but de l'enrichissement étant la formulation d'une

1. <http://www.globalwordnet.org/AWN/AWNBrowser.html>, voir aussi (Elkateb et al., 2006)

2. <http://www.almaany.com>

requête plus riche et donc plus précise. Une conséquence directe étant l'amélioration des performances du SRI en renvoyant des résultats plus pertinents.

Les travaux sur l'évaluation des techniques de la RQ pour les textes en arabe ne sont pas nombreux. Nous trouvons à titre d'exemples :

- Les travaux effectués dans Kanaan et al. (2005) ont montré que la reformulation manuelle par repondération des termes de la requête permet une amélioration des performances (rappel et précision) du SRI Arabe. Pour leur expérimentation, Kanaan et al. (2005) ont utilisé un corpus de 242 documents et un jeu de neuf (9) requêtes.
- Les travaux de Hammo et al. (2007) portant sur l'expansion de la requête par des termes issus d'un thésaurus montrent une amélioration dans le rappel du SRI arabe. Nous notons que le corpus utilisé était le coran.
- Les travaux de Xu et al. (2002) montrent que l'utilisation d'un thésaurus améliore considérablement (18%) les performances d'un SRI arabe. Xu et al. (2002) ont montré aussi que l'utilisation d'une indexation basée sur les racines est plus performante que l'utilisation des schèmes pour les textes arabes.
- Les travaux de Zaidi and Laskri (2007) sur l'expansion de la requête en utilisant une ontologie du domaine juridique et WordNet ont permis d'obtenir des améliorations considérables dans les performances du SRI.
- Le système de Ahmed and Nürnberger (2008) propose d'assister l'utilisateur dans la reformulation de sa requête, par l'ajout des formes proches morphologiquement des formes de la requête initiale, en se basant sur le calcul de similarité des n-grams entre les mots de la requête initiale et ceux enregistrés dans un lexique. Etant basée sur la similarité des chaînes de caractères, cette approche ne peut résoudre le problème des variations lexicales ou sémantiques. Pour les opérations d'indexation et de recherche, Ahmed and Nürnberger (2008) ont utilisé les services du moteur de recherche Google.
- Les travaux de Abderrahim and A (2010) qui se résument à l'utilisation d'une ressource lexicale (WordNet Arabe) et un analyseur morphologique pour reformuler, par expansion, la requête de l'utilisateur permettent d'améliorer le rappel, mais pas la précision du SRI.
- Les travaux de Abderrahim and A (2012) sur l'évaluation de la stratégie de la classification locale pour les textes arabes montrent une amélioration dans les performances du SRI.

Cet article constitue une extension des travaux de Abderrahim and A (2010) en proposant d'évaluer l'apport réel de la reformulation de la requête guidée par une ontologie dans un SRI Arabe. Dans ce qui suit nous allons décrire notre expérimentation et discuter les résultats obtenus.

### 3. Expérimentation et discussion

Pour notre expérimentation nous avons utilisé un corpus de plus de 22 000 documents arabes (environ 180 Mo) de différents domaines (santé, sport, politique, science, religion, ...). Ce corpus compte environ 17 000 000 mots dont 612 650 mots différents. Un ensemble de 50 requêtes de type mots clés et de différents thèmes sont retenues pour l'expérimentation.

Pour les opérations d'indexation et de recherche nous avons utilisé l'API Lucene<sup>3</sup> qui est librement disponible sur le net. Par ailleurs, pour le processus de reformulation nous avons codé en java trois stratégies différentes que nous décrirons dans la suite de cette section. Nous avons aussi utilisé deux ressources externes différentes pour la recherche des concepts proches dans la phase de la reformulation de la requête. La première est WordNet arabe qui est une des rares ressources librement disponibles pour la langue arabe. Il compte actuellement 11 269 synsets et 23 481 mots (Alkhalifa, 2006; Elkateb et al., 2006; Black et al., 2006). Par ailleurs la seconde ressource est le DSA qui est aussi librement consultable en ligne. Dans sa version actuelle, il compte 20 500 synonymes et 35 000 mots. Du point de vue du contenu, le DSA est plus riche en informations comparé à WordNet arabe.

Dans un premier temps nous avons indexé l'ensemble des documents de notre corpus. Le résultat de cette opération nous a permis d'obtenir un index d'une taille d'environ 37 Mo (environ 1/5 de la taille du corpus utilisé). Cet index sera utilisé pour faire la recherche des informations en réponse aux 50 requêtes reformulées selon différents protocoles. La stratégie de reformulation est basée sur une expansion aveugle, contrôlée (sélective) et pondérée des concepts présents dans les requêtes. Pour chaque stratégie de reformulation nous avons fait correspondre un type de recherche que nous allons étudier séparément pour évaluer son apport dans l'amélioration des performances de la recherche. Il faut noter toutefois que chaque stratégie a été testée deux fois : une fois en utilisant WordNet arabe, et une autre fois en utilisant DSA.

Nous avons examinés quatre différents types de recherche :

1. Recherche simple ou recherche avant enrichissement (RS) : nous avons utilisé la liste des 50 requêtes simples de type mots clés, par exemple les requêtes :

صبيغ الشيب, صداقة جادة

2. Recherche Aveugle (RA) : nous avons utilisé une liste de 50 requêtes déduites des requêtes simples de départ par un enrichissement aveugle (nous ajoutons à la requête initiale l'ensemble des synonymes trouvé dans WordNet arabe ou DSA). Le tableau 1 présente des exemples de ce type de recherche selon la ressource externe utilisée.

---

3. <http://lucene.apache.org/>

Num	Requête simple (RS)	Nouvelle requête enrichie (RP) A partir de WordNet Arabe	Nouvelle requête enrichie (RP) A partir de DSA
1	صبيغ الشيب	صبيغ الشيب ، طلى ، دهن ، لون ، صبغة ، تلوين ، صباغ ، خضب	صبيغ الشيب ، لون ، خضب ، تلوين ، طلى ، لون ، ضمخ
2	صداقة جادة	صداقة جادة ، ألفة ، عشرة ، صحية ، خلة	صداقة جادة ، صحية ، رفقة ، مخادنة ، ألفة ، مصاحبة ، إخاء ، مؤانسة

TAB. 1 : Exemples de requêtes en recherche Aveugle selon la ressource utilisée

3. Recherche Contrôlée (RC) : nous avons utilisé une liste de 50 requêtes déduites des requêtes simples par un enrichissement manuel (nous ajoutons à la requête seulement les synonymes sélectionnés manuellement). La construction de la requête enrichie se fait interactivement entre le système et l'utilisateur, le scénario est le suivant : Après analyse de la requête, le système recherche les synonymes des termes de la requête et ensuite, il propose à l'utilisateur de valider les termes suggérés (par suppression des termes non appropriés). Le tableau 2 présente des exemples de ce type de recherche selon la ressource externe utilisée.

Num	Requête simple (RS)	Nouvelle requête enrichie (RP) A partir de WordNet Arabe	Nouvelle requête enrichie (RP) A partir de DSA
1	صبيغ الشيب	صبيغ الشيب ، لون ، خضب ، تلوين ، صباغ ، صبغة	صبيغ الشيب ، خضب ، تلوين ، طلى ، لون ،
2	صداقة جادة	صداقة جادة ، ألفة ، عشرة ، صحية ، خلة	صداقة جادة ، صحية ، رفقة ، ألفة ، مصاحبة ، إخاء ، مؤانسة

TAB. 2 : Exemples de requêtes en recherche contrôlée selon la ressource utilisée

4. Recherche Pondérée (RP) : nous avons utilisé une liste de 50 requêtes déduites des requêtes simples par un enrichissement automatique. Nous avons ajouté à la requête seulement un seul synonyme choisi automatiquement. La procédure de choix se base sur la pondération affectée à chaque terme. Le terme ayant la pondération la plus élevée est retenu, intuitivement, c'est le sens le plus utilisé.

Comme nous ne disposons pas de cette information de pondération des termes dans WordNet arabe ou DSA, nous avons calculé cette valeur à partir d'une analyse statistique de tous les documents de notre corpus. Dans la littérature il existe plusieurs méthodes de pondération, cependant, elles présentent toutes une variation de la formule « TF.IDF ». Dans cette expérimentation, nous avons choisi comme pondération la valeur TF (nombre d'occurrence d'un terme dans le corpus). Il est clair que la valeur de discrimination « IDF » ne présente aucun intérêt pour notre cas puisqu'elle ne permet de mettre en valeur que les termes qui apparaissent dans peu de documents. Le tableau 3 présente des exemples de ce type de recherche selon la ressource externe utilisée.

Num	Requête simple (RS)	Nouvelle requête enrichie (RP) A partir de WordNet Arabe	Nouvelle requête enrichie (RP) A partir de DSA
1	صغ الشيب	صغ الشيب ، لون	صغ الشيب ، لون
2	صداقة جادة	صداقة جادة ، عشرة	صداقة جادة ، صحبة

TAB. 3 : Exemples de requêtes en recherche pondérée selon la ressource utilisée

Les résultats obtenus par ces quatre différents types de recherche sont consignés dans différents fichiers et pour chaque type de recherche et chaque requête nous avons calculé les différents rappels et précisions du système.

#### 4. Analyse des résultats et discussion

Le tableau 4 présente les précisions à 11 points de rappels des différents systèmes associés à chaque type de recherche.

La figure 3 présente les différentes courbes rappel/précision obtenues à partir du tableau 4.

L'examen de la figure 3 nous permet de constater que :

- La reformulation testée avec ses différents types permet d'avoir une amélioration des performances du SRI dans l'intervalle de rappel  $[0; 0,1]$  de la RP utilisant DSA.
- Quelle que soit la ressource utilisée la RP donne les meilleures performances.
- La RA présente la technique la plus pauvre en performance.
- L'utilisation de DSA est plus bénéfique que WordNet Arabe.

Pour comprendre l'effet des différents types de recherche sur chaque requête, nous avons établi différentes mesures qui sont principalement basées sur la comparaison

Rappel	RS	Utilisation de WordNet Arabe			Utilisation de DSA		
		RA	RC	RP	RA	RC	RP
0	0,620	0,484	0,543	0,623	0,419	0,517	0,646
0,1	0,616	0,453	0,516	0,577	0,394	0,473	0,584
0,2	0,587	0,427	0,487	0,522	0,386	0,456	0,532
0,3	0,557	0,392	0,452	0,461	0,362	0,428	0,471
0,4	0,543	0,359	0,430	0,424	0,336	0,399	0,436
0,5	0,508	0,330	0,398	0,396	0,315	0,373	0,401
0,6	0,464	0,305	0,362	0,364	0,291	0,347	0,373
0,7	0,431	0,293	0,340	0,340	0,269	0,328	0,336
0,8	0,402	0,281	0,323	0,319	0,259	0,314	0,312
0,9	0,368	0,257	0,298	0,287	0,243	0,290	0,284
1	0,304	0,207	0,242	0,226	0,201	0,231	0,233

TAB. 4 : Les précisions à 11 points de rappels selon la ressource utilisée

des résultats avant et après enrichissement. Dès lors, pour une requête donnée, trois cas peuvent se présenter :

- Amélioration : toutes les précisions à 11 points de rappel avant sont inférieures à ceux d'après (i.e. : la courbe rappel / précision après est au-dessus de avant).
- Pas d'amélioration : c'est le cas inverse du précédent (i.e. : la courbe rappel / précision après est au-dessous de avant).
- Sans décision : pour certaines précisions à 11 points il y a amélioration et pour d'autres il n'y a pas d'amélioration (i.e. : intersection des deux courbes rappel / précision avant et après).

Le tableau 5 présente pour chaque requête utilisée dans l'expérimentation l'indicateur Amélioration(+), Pas d'amélioration(-) ou Sans décision(X) des différents types de recherche.

L'examen des résultats obtenus dans le tableau 5 nous permet de déduire les faits suivants :

1. Quelle que soit la méthode de reformulation testée, l'utilisation de WordNet Arabe permet d'avoir :
  - Une amélioration (+) dans 4 requêtes (les requêtes numéro : 12, 27, 28, 38) soit 8%.
  - Pas d'amélioration (-) dans 21 requêtes (42%).
  - Une indécision (X) dans 6 requêtes (12%).
  - Un ensemble de 19 requêtes (39%) pour lesquelles il existe au moins une amélioration dans 6 requêtes (12%) dans l'une des méthodes de reformulation

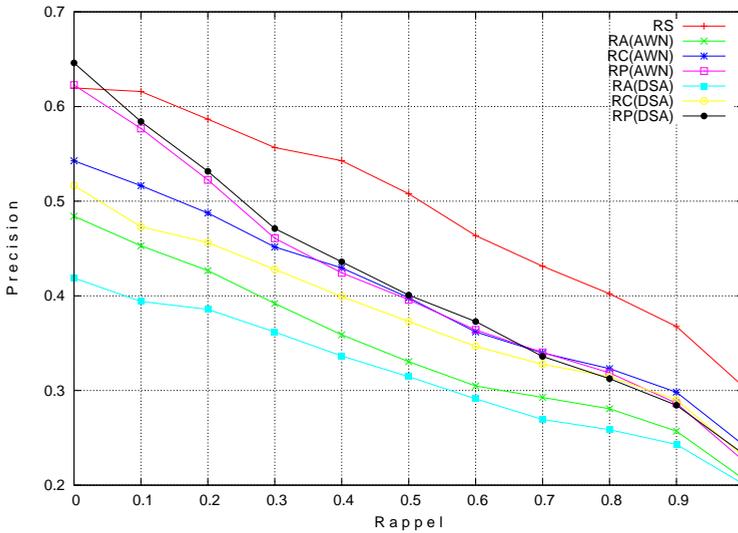


FIG. 3 : Les courbes rappel/précision à partir du tableau 4

testée. Pour les 13 requêtes restantes (26%) soit il n’y a pas d’amélioration soit une indécision.

2. Quelle que soit la méthode de reformulation testée, l’utilisation de DSA permet d’avoir :
  - Une amélioration (+) dans 2 requêtes (les requêtes numéro : 11, 36) soit 4%.
  - Pas d’amélioration (-) dans 16 requêtes (32%).
  - Une indécision (X) dans 6 requêtes (12%).
  - Un ensemble de 26 requêtes (52%) pour lesquelles il existe au moins une amélioration dans 6 requêtes (12%) dans l’une des méthodes de reformulation testée. Pour les 20 requêtes restantes (40%) soit il n’y a pas d’amélioration soit une indécision.

Du point de vue de l’amélioration, l’analyse des faits précédents (1 et 2) nous permet d’annoncer que la reformulation par utilisation d’une ressource externe permet d’améliorer les performances d’un SRI Arabe d’environ 6% ; Par ailleurs, quelle que soit la méthode préconisée pour la reformulation, l’utilisation de WordNet Arabe est bien meilleure que DSA. Pour déterminer la meilleure technique de recherche entre RA, RC et RP nous avons comptabilisé le nombre des requêtes pour chaque type de recherche dans le tableau 6.

Les résultats obtenus dans le tableau 6 ne permettent pas de faire un choix entre RA, RS et RP dans le cas de l’utilisation de WordNet Arabe, car nous avons le même

Num Re- quête	Utilisation de WordNet Arabe			Utilisation de DSA			Num Re- quête	Utilisation de WordNet Arabe			Utilisation de DSA		
	RA	RC	RP	RA	RC	RP		RA	RC	RP	RA	RC	RP
1	-	-	-	-	-	-	26	-	X	-	-	X	X
2	X	X	X	-	-	-	27	+	+	+	X	X	X
3	-	X	X	-	-	X	28	+	+	+	-	-	+
4	X	-	X	-	-	-	29	-	-	-	-	-	-
5	X	X	X	-	X	X	30	-	-	X	-	-	-
6	-	-	-	-	-	-	31	-	-	-	-	-	X
7	-	-	X	-	X	-	32	+	+	X	X	+	X
8	-	-	-	-	X	-	33	-	X	X	-	-	X
9	+	X	+	X	-	X	34	-	-	-	-	-	X
10	-	-	X	-	-	-	35	-	-	-	X	X	X
11	X	-	-	+	+	+	36	X	X	-	+	+	+
12	+	+	+	+	+	X	37	-	-	-	-	-	-
13	X	X	-	-	X	X	38	+	+	+	X	X	+
14	X	X	X	X	X	X	39	-	-	-	-	X	-
15	-	-	-	-	-	X	40	-	-	-	-	-	-
16	-	-	-	-	-	-	41	-	-	-	-	-	X
17	-	-	-	-	-	-	42	-	-	-	X	X	+
18	-	-	-	-	-	-	43	X	-	-	-	-	-
19	X	X	X	X	X	X	44	X	X	-	X	X	-
20	-	-	-	-	-	-	45	-	-	X	-	X	X
21	X	X	X	X	X	X	46	+	+	X	-	X	X
22	X	+	X	+	X	X	47	-	-	-	-	X	-
23	X	X	+	X	X	X	48	-	-	-	-	-	X
24	X	X	X	X	X	-	49	-	-	-	-	-	-
25	-	-	-	-	-	-	50	-	-	+	-	-	-

Tab. 5 : L'indicateur Amélioration(+), Pas d'amélioration(-) ou Sans décision(X) des différents types de recherche selon la ressource utilisée.

taux d'amélioration (14%) pour les trois cas. En revanche, il est clair que la technique RP (10%) présente le meilleur taux d'amélioration dans le cas de l'utilisation de DSA, d'ailleurs l'analyse des courbes rappel/précision de la figure 3 ne fait que confirmer ce résultat. Tout compte fait, nous pouvons conclure que l'apport de l'utilisation d'une ressource externe dans un SRI Arabe est d'environ 6%. Par ailleurs, il apparaît que WordNet Arabe est meilleur que DSA et que RP offre le taux d'amélioration le plus élevé.

	Nombre de requêtes					
	Utilisation de WordNet Arabe			Utilisation de DSA		
	RA	RC	RP	RA	RC	RP
Amélioration	7 (14%)	7 (14%)	7 (14%)	4 (8%)	4 (8%)	5 (10%)
Pas d'amélioration	29 (58%)	29 (58%)	27 (54%)	34 (68%)	26 (52%)	23 (46%)
Sans décision	14 (28%)	14 (28%)	16 (32%)	12 (24%)	20 (40%)	22 (44%)

TABLEAU 6 : Le nombre de requêtes vérifiant les conditions : Amélioration, Pas d'amélioration et Sans décision selon la ressource utilisée.

## 5. Conclusion

Dans cet article, nous avons examiné différentes manières pour faire la reformulation d'une requête dans un SRI Arabe. Cette reformulation étant basée sur une ressource externe, nous avons particulièrement expérimentée deux ressources à savoir WordNet Arabe et DSA. Les résultats obtenus nous ont permis de mesurer l'apport (6%) d'une telle approche dans l'amélioration des performances globales d'un SRI Arabe. Du point de vue du nombre de requêtes ayant effectivement conduit à une amélioration, les résultats de comparaison entre l'utilisation de WordNet Arabe et DSA sont en faveur du premier. En revanche, du point de vue "rappel/précision", l'utilisation de DSA est plus bénéfique que celle de WordNet Arabe. Par ailleurs, nous avons conclu que la technique de RP offre un meilleur taux d'amélioration des performances d'un SRI Arabe.

Cette étude nous a permis d'ouvrir la voie pour tester et comparer d'autres méthodes de reformulation avec les mêmes données de cette expérimentation afin de déterminer la technique la plus appropriée à intégrer dans un SRI Arabe.

## 6. Summary

In information retrieval systems (IRS), approaches to query reformulation are numerous. They can be classified according to the used resources in three main approaches: those using external resources, global analysis, and local analysis. In this context and as part of an IRS for Arabic texts, we are interested in evaluating performance of the first approach. For this purpose, two different resources are used, namely Arabic WordNet and the Arabic Dictionary (thesaurus) of Meaning. The ex-

periments conducted on a corpus of Arabic text allowed us to measure the contribution of this query reformulation approach applied to an Arabic IRS.

## Références

- Aalbersberg, Ijsbrand Jan. Incremental Relevance Feedback. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 11–22. ACM, 1992.
- Abderrahim, Med-El-Amine and Med-Alaeddine A. Using Arabic Wordnet for Query Expansion in Information Retrieval System. In *IEEE The Third International Conference on Web and Information Technologies*, Marrakech, Morocco, June 2010.
- Abderrahim, Med-El-Amine and Med-Alaeddine A. Réinjection automatique de la pertinence pour la recherche d'informations dans les textes arabes. In *IEEE 4th International Conference on Arabic Language Processing (CITALA)*, pages 77–81, Rabat, Morocco, May 2012.
- Ahmed, Farag and Andreas Nürnberger. Arasearch: Improving Arabic Text Retrieval via Detection of Word Form Variations. In *SIIE 2008*, pages 309–323, Hammamet, Tunisie, February 2008.
- Alkhalifa, Musa. Arabic Wordnet and Arabic NLP. In *Journées d'Etudes sur le Traitement Automatique de la Langue Arabe (JETALA)*, Rabat, June 2006.
- Baeza-Yates, Ricardo and Ribeiro-Neto Berthier. *Modern Information Retrieval*. Addison-Wesley, New York City, NY ACM Press, 1999.
- Black, William, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. Introducing the Arabic WordNet Project. In *Proceedings of the Third International WordNet Conference*, pages 295–300, 2006.
- Bodo, Billerbeck. *Efficient Query Expansion*. PhD thesis, RMIT University, Melbourne, Australia, 2005.
- Bruande, Marie-France and Jean-Pierre Chevallet. Assistance intelligente à la recherche d'information. *Lavoisier*, pages 99–129, 2003.
- Carpineto, Claudio and Giovanni Romano. A survey of Automatic Query Expansion in Information Retrieval. *ACM Computing Surveys (CSUR)*, 44(1):1, 2012.
- Efthimiadis, Efthimis N. Query Expansion. *Annual review of information science and technology*, 31:121–187, 1996.
- Elkateb, Sabry, William Black, Piek Vossen, David Farwell, H Rodríguez, A Pease, and M Alkhalifa. Arabic WordNet and the Challenges of Arabic. In *Proceedings of Arabic NLP/MT Conference, London, UK*, 2006.
- Hammo, Bassam, Azzam Sleit, and Mahmoud El-Haj. Effectiveness of Query Expansion in Searching the Holy Quran. In *Proceedings of L'institut organise le 2ème Colloque International sur le Traitement Automatique de la Langue Arabe (CITALA)*, volume 7, pages 18–19, Morocco, 2007.
- Harb, Hany M, Khaled M Fouad, and Nagdy M Nagdy. Semantic Retrieval Approach for Web Documents. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2(9):11–75, 2011.

- Hlaoua, Lobna. *Reformulation de requêtes par réinjection de pertinence dans les documents semi-structurés*. PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2007.
- Kanaan, G, R Al-Shalabi, M Abu-Alrub, and M Rawashdeh. Relevance Feedback: Experimenting with a Simple Arabic Information Retrieval System with Evaluation. *International Journal of Applied Science and Computations*, 12(2), 2005.
- Lee, Kyung Soon, W Bruce Croft, and James Allan. A Cluster-based Resampling Method for Pseudo-relevance Feedback. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 235–242. ACM, 2008.
- Salton, Gerard and Chris Buckley. Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41(4):288–97, 1990.
- Xu, Jinxi and W Bruce Croft. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Transactions on Information Systems (TOIS)*, 18(1):79–112, 2000.
- Xu, Jinxi, Alexander Fraser, and Ralph Weischedel. Empirical Studies in Strategies for Arabic Retrieval. In *Annual ACM Conference on Research and Development in Information Retrieval: Proceedings of the 25 th annual international ACM SIGIR conference on Research and development in information retrieval*, volume 11, pages 269–274, 2002.
- Zaidi, S and M Laskri. Expansion de la requête Arabe sur le réseau internet. In *Barmajiat (CSLA): Les applications logicielles en arabe: Pas vers le e-gouvernement*, Alger, December 2007.

**Address for correspondence:**

Mohammed El Amine Abderrahim  
medamineabd@yahoo.fr  
Université de Tlemcen  
Faculté de Technologie  
BP 230 Chetouane,  
Tlemcen (13000) Algérie