

# Un nouveau dictionnaire électronique structuré et évolutif de la langue arabe : DESELA

**Abd El Salam AL HAJJAR (1)(2), Mohammad HAJJAR (1),  
Khalidoun ZREIK (2)**

*(1)Institut Universitaire de Technologie, Université Libanaise, Liban*

*(2)Laboratoire Paragraphe, Université de Paris 8 - Vincennes - Saint-Denis, France*

**Mots-clés :** Langue arabe, Corpus, Dictionnaire, Extraction d'information, Racine.

**Keywords:** Arabic Language, Corpus, Dictionary, Information Extraction, Root.

**Résumé :** Dans cet article, nous proposons un nouveau dictionnaire électronique structuré et évolutif de la langue arabe (DESELA) qui peut être présenté sous la forme d'une base de données relationnelle ou d'un document XML et qui est facilement exploitable à l'aide des langages de requêtes appropriés. En effet, on trouve beaucoup de dictionnaires arabes mais qui ne sont pas directement exploitables puisqu'ils sont sous forme des fichiers texte plats. DESELA contient essentiellement les racines, les préfixes, les suffixes, les infixes, les modèles et les mots dérivés. De plus, pour un mot donné, il fournit les liens avec sa racine, avec les affixes associés et avec son modèle éventuel. DESELA est alimenté automatiquement à partir d'un ou de plusieurs dictionnaires textuels classiques et est enrichi en permanence avec des corpus textuels arabe quelconques grâce à un système que nous avons construit. Ce système est composé d'un parseur, d'un classifieur, d'un comparateur et d'un analyseur. Le parseur permet de transformer une source textuelle (dictionnaire ou corpus textuel) en un ensemble de mots. Le classifieur permet de classer un mot donné et de l'ajouter au DESELA en tant qu'une racine ou en tant qu'un mot dérivé. L'analyseur permet d'extraire les affixes et le modèle à partir d'un mot dérivé et de sa racine. Le comparateur permet d'éviter d'avoir des doublons, à tous les niveaux, dans DESELA. Ce dictionnaire peut être utilisé pour évaluer ces méthodes d'extraction d'information à partir d'un document arabe. Étant donné que le vocabulaire de la langue arabe est essentiellement construit à partir des racines, un mot arabe est construit à partir de sa racine en y ajoutant des affixes (préfixe, infixé, ou suffixe) selon un modèle précis. La plupart des méthodes d'extraction d'information à partir d'un document arabe procèdent inversement en extrayant la racine à partir du mot.

**Abstract:** In this article, we propose a new structured and progressive electronic dictionary for the Arab language (DESELA) which can be presented in the form of a relational database or in the form of an XML document which can be easily exploitable using suitable query languages. Indeed, many Arab dictionaries are found but are not directly exploitable since they are in flat textual files form. DESELA contains the roots, the prefixes, the suffixes, the infixes, the patterns and the derived words. Moreover, for a given word, it provides links to its root, to their associated affixes, and to its possible pattern. DESELA is supplied automatically from one or several traditional textual dictionaries and is enriched permanently with any Arab textual corpus using system that we built. This system is composed of a parser, a classifier, a comparator and an analyzer. The parser allows transforming a textual source (dictionary or textual corpus) into a set of words. The classifier allows to classify a given word and to add it to DESELA as a root or a derived word. The analyzer allows extracting the affixes and the model from a derived word and of its root. The comparator permits to avoid duplication of roots, affixes or patterns in DESELA. This dictionary can be used to evaluate the information extraction methods from an Arab document, given that; the vocabulary of the Arab language is essentially built from the roots. In general, an Arab word is built from its root while adding to it the affixes (prefix, infix, or suffix) according to a precise pattern. Most methods of information extraction starting from an Arab document proceed conversely by extracting the root from the mot.

## 1 Introduction

Les performances des systèmes d'extraction d'information en langue arabe restent très problématiques et ceci pour plusieurs raisons [1], [3], [6]. L'une des raisons principales est due au fait que le vocabulaire de la langue arabe est essentiellement construit à partir des racines. En effet, la langue arabe possède cinq à sept milles racines distincts. Un mot arabe est construit à partir de sa racine en y ajoutant des affixes (préfixe, infix, ou suffixe) selon un modèle précis [2], [3], [8]. Ces modèles sont au nombre de cent vingt, environ. Les méthodes d'extraction d'information à partir d'un document arabe procèdent inversement en extrayant la racine à partir du mot. Dans ce domaine, plusieurs méthodes ont été proposées [1], [4], [6], [8], [11], [14], [15], [17], [21], [23], [24]. Ces méthodes sont soit basées sur les caractéristiques morphologiques de la langue arabe soit sur des calculs statistiques. Pour évaluer ces méthodes, nous avons développé un système d'évaluation et nous avons construit un corpus limité à vingt racines et à deux milles mots. Pour valider ces résultats, il faut bien sur un corpus plus important, un dictionnaire par exemple.

En effet, on trouve beaucoup de dictionnaires arabes comme Lisan Al-Arab, Al Qamous Al Mouhit, Al Wasit, Al Mouhit, Mouhit Al Mouhit, Al Ghani et d'autres [26], [27], [28], [29]. Bien que ces dictionnaires indiquent la racine, la définition, l'orthographe, les sens et les modes d'utilisation d'un mot donné, ils ne sont pas directement exploitables informatiquement puisqu'ils sont aux formats textuels non structurés (fichiers texte plats). Donc, l'absence d'un tel dictionnaire nous a poussé à construire un dictionnaire électronique structuré et informatiquement exploitable pour l'utiliser dans l'évaluation des méthodes d'extraction d'information à partir des documents arabes.

Dans cet article, nous proposons un dictionnaire électronique structuré et évolutif de la langue arabe (DESELA). Ce nouveau dictionnaire peut être présenté sous la forme d'une base de données relationnelle [18], [19] ou d'un document XML [22] facilement exploitable à l'aide des langages de requêtes appropriés. Ce nouveau dictionnaire contient les racines, les préfixes, les suffixes, les infixes et les modèles, en plus des informations fournies par un dictionnaire classique. De plus, il fournit les liens d'un mot donné avec sa racine, avec les affixes associés et avec son modèle éventuel. Pour atteindre cet objectif, nous avons construit un système automatique qui permet d'alimenter DESELA à partir d'un ou de plusieurs dictionnaires textuels classiques. Ce système permet aussi d'enrichir DESELA, en permanence, à partir d'un corpus textuel arabe quelconque d'où l'évolutivité de notre dictionnaire.

## **2 Architecture**

La figure 1 présente l'architecture générale de notre système qui permet d'alimenter et d'enrichir automatiquement DESELA à partir de plusieurs dictionnaires textuels classiques et des corpus textuels arabe quelconques. Ce système est composé de plusieurs modules qui sont le parseur, le classifieur, le comparateur et l'analyseur.

- Le parseur permet de transformer une source textuelle (dictionnaire ou corpus textuel) en un ensemble de mots.
- Le classifieur permet de classer un mot donné et de l'ajouter au DESELA en tant qu'une racine ou en tant qu'un mot dérivé. L'analyseur permet d'extraire les affixes et le modèle à partir d'un mot dérivé et de sa racine.
- Le comparateur permet d'éviter d'avoir des doublons dans DESELA, à tous les niveaux. Une remarque d'ordre générale pour cette partie est que les transcriptions de tous les mots arabes utilisés (racine, affixes, modèles, mots) dans ce document sont données dans la table 1.

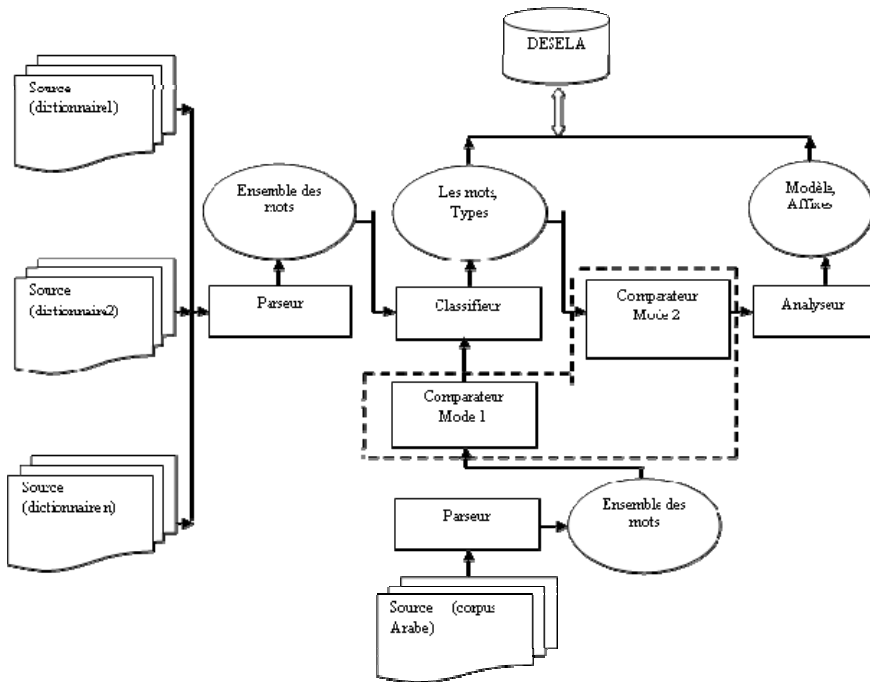


Figure 1 : Architecture générale du système d'alimentation et d'enrichissement automatique de DESELA.

## 1.1 Le parseur

Le parseur constitue le point d'entrée de notre système. L'objectif de ce composant est de transformer une source textuelle en un ensemble de mots. Il peut bien s'agir d'un dictionnaire sous format des fichiers texte plats d'où d'un corpus textuel quelconque. Le parsing d'une source textuel est effectué en plusieurs étapes. La première sert à déterminer les délimiteurs qui séparent les mots. Ces délimiteurs peuvent être des espaces, des symboles particuliers ou d'autres selon le document à traiter. La deuxième étape consiste à fournir un premier ensemble des mots bruts à partir du document source. La dernière étape dans ce module sert à nettoyer l'ensemble de mots bruts ainsi obtenus. Cette étape consiste en plusieurs phases. La première sert à éliminer les non-caractères, les chiffres et les symboles de l'ensemble des mots bruts. La deuxième sert à en supprimer les mots parasites, ou des mots courts (من, إلى, ...) pour les ajouter à DESELA. Donc, la sortie de cette dernière étape, et du cout du parseur, est un ensemble de mots qui sont, soient des racines, soient des mots dérivés des racines.

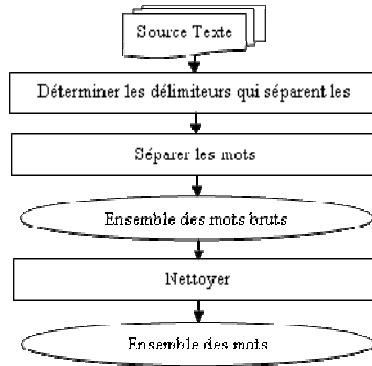


Figure 2 : Le parseur.

## 1.2 Le classifieur

Les entrées de ce composant peuvent être fournies soit par le parseur, soit par le comparateur. L'objectif de ce composant est de décider si un mot est une racine ou non. Le classifieur permet de classer un mot et de l'ajouter au DESELA. Trois classes sont possibles : racines, mots dérivé d'une racine, mot isolé. S'il est une racine, il l'ajoute au DESELA en tant qu'une racine. Dans le cas contraire, il détermine la racine de laquelle il dérive, l'ajoute au DESELA en tant que mot dérivé et établit le lien avec sa racine. S'il n'a pas de racine, il est isolé, dans ce cas aucun lien n'est établi avec les racines.

La question primordiale à résoudre dans ce composant est : comment déterminer si un mot est une racine ? Pour répondre à cette question, plusieurs cas se présentent (Figure 3).

Dans le cas d'un dictionnaire, les racines sont, en générale, encadrées par des séparateurs spéciaux et les mots, qui sont situés après cette racine et avant la racine suivante, dérivent de la première. Le fait de valider un mot avec sa racine est dû au fait que certains mots qui se trouvent après une racine peuvent ne pas dériver d'elle. Ce type des mots est à ne pas considérer dans DESELA. Pour déterminer ce type des mots, nous utilisons l'une des méthodes d'extraction de la racine d'un mot arabe [24]. Par exemple, dans le cas du dictionnaire Lissan Al Arabe [26], [27]. Chaque racine est précédée par le symbole « @ » et suivit par le symbole « : » (Figure 4), la plupart des mots qui sont situés après une racine et avant la racine suivante dérivent de la première. Considérons l'exemple de la racine أكل donné dans la figure 4. Dans cet exemple, tous les mots qui sont situés entre les deux racines أكل et غرب sont validés par une méthode d'extraction de la racine arabe, en l'occurrence « Arabic Stemming without a root dictionary » [8]. Par contre le mot تقول, qui situe entre les deux racines أكل et غرب ne dérive pas de la première racine أكل.

Dans ce cas, le mot تقول, qui n'est pas validé par rapport à la racine أكل, est à ne pas considérer dans DESELA.

La méthode « Arabic Stemming without a root dictionary » est basée sur l'élimination de plusieurs ensembles de diacritiques et d'affixes et sur l'application de plusieurs modèles qui ont déjà défini [8]. Nous avons choisis cette méthode pour déterminer si un mot est une racine ou pour valider les mots par rapport à une racine parce qu'elle n'utilise aucun dictionnaire pour extraire la racine arabe.

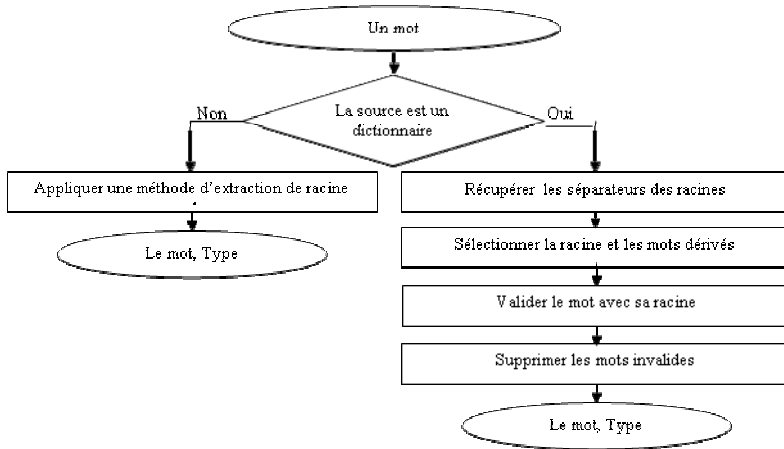


Figure 3 : Le classifieur.

Par contre, dans le cas d'un corpus quelconque, nous utilisons l'une des méthodes d'extraction de la racine d'un mot arabe [24] pour décider si un mot est une racine ou pour déterminer la racine de laquelle elle dérive.

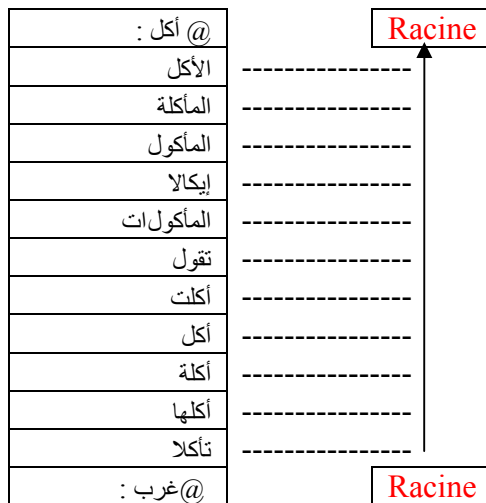


Figure 4 : Cas du dictionnaire Lisan Al Arabe

## 2.1 1.3 L'analyseur

En générale, un mot arabe est dérivé à partir de sa racine en y ajoutant des affixes (préfixe, infixe, ou suffixe) selon un modèle précis. L'analyseur permet d'extraire les affixes et le modèle à partir d'un mot dérivé et de sa racine. Ce composant prend en entrée un couple {Mot dérivé, Racine} et produit en sortie les préfixes, les suffixes et les infixes éventuels ainsi que le modèle selon lequel le mot est dérivé. Pour ce faire, Nous commençons par repérer les positions des lettres constituant la racine dans le mot dérivé. L'étape suivante consiste à déterminer les lettres appartenant au mot dérivé et ne faisant pas partie de la racine. Ainsi, les lettres qui précèdent la première lettre de la racine, si elles existent, dans le mot dérivé constituent les préfixes. De même, les lettres qui suivent la dernière lettre de la racine, si elles existent, dans le mot dérivé constituent les suffixes. En suite, les lettres qui sont situés entre la première lettre et la dernière lettre de la racine, si elles existent, dans le mot dérivé et qui ne font pas parties de la racine constituent les infixes. L'étape suivante consiste à déduire le modèle dans le mot dérivé, le modèle est déduit, selon les positions des lettres constituant la racine dans le mot dérivé. La première étape consiste à supprimer les suffixes, la deuxième consiste à supprimer les préfixes s'ils n'appartiennent pas à l'ensemble {س, ل, م, ت}, la troisième étape consiste à transformer les lettres après les préfixes de l'ensemble {س, ل, م, ت} s'ils y existent de la racine dans l'ordre où la première lettre permute en "ف", la deuxième permute en "ع" et la troisième permute en "ل". Les infixes sont repris tel qu'ils sont.

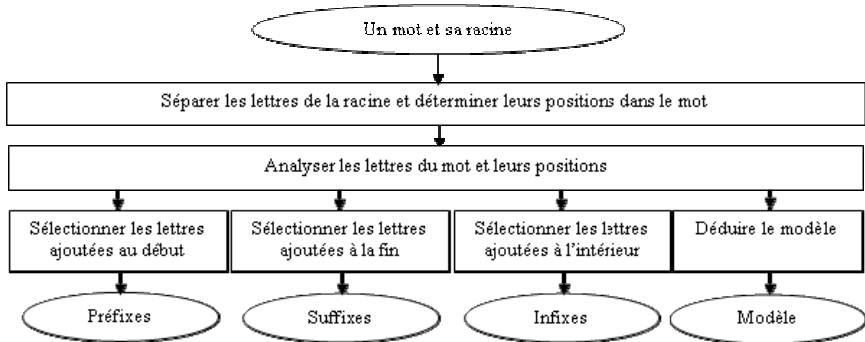


Figure 5 : L'analyseur

Considérons l'exemple du couple {Mot dérivé = المأكولات, Racine = أكل}, la phase du repérage des lettres de la racine (en rouge) dans le mot dérivé

donne المأكولات. En suite, les lettres qui précèdent la première lettre de la racine الم constituent les préfixes (en vert). Les lettres qui suivent la dernière lettre de la racine ات constituent les suffixes (en jaune). Les lettres qui sont situés entre la première lettre et la dernière lettre de la racine et qui n'en font pas parties و constituent les infixes (en bleu). Le modèle est déduit comme لومفع, en partant des lettres du mot المأكولات, nous supprimons le suffixe ات et le préfixe الم, car م appartient à {س, ت, م, ه, ن} nous obtiendrons le mot المأكولات, ensuite nous permutons respectivement المأكولات en لومفع. L'infixe و est repris tel qu'il est pour obtenir le modèle لومفع.

## 1.4 Les comparateurs

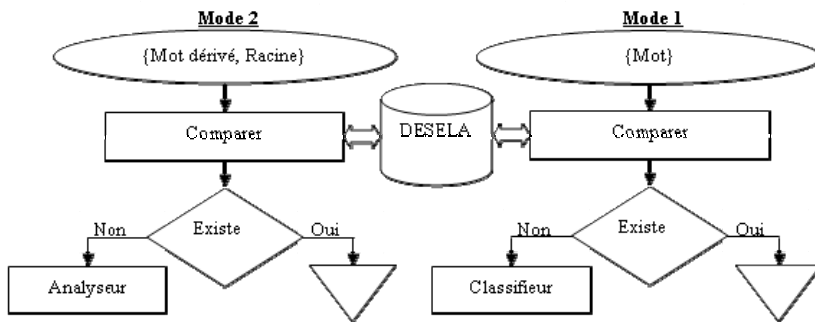


Figure 6 : Le comparateur

Le comparateur permet d'éviter d'avoir des doublons dans DESELA aux niveaux des mots, des racines, des préfixes, des infixes, des suffixes et des modèles. Ce composant est sollicité seulement en phase d'enrichissement. Cette phase d'enrichissement présente le problème suivant : comment enrichir vraiment notre dictionnaire et non pas ajouter des doublons à tous les niveaux. Donc, le rôle du comparateur est de filtrer les mots avant de les ajouter au DESELA. Ce comparateur possède deux modes de fonctionnement. Le premier quand il reçoit en entrée un ensemble de mots fournis par le parseur. Dans ce cas, pour chaque mot il va vérifier s'il existe dans DESELA, s'il n'y est pas il le passe au classifieur pour le traiter. Le second mode de fonctionnement du comparateur est quand il reçoit en entrée un couple {Mot dérivé, Racine} fournis par le classifieur. Ce couple n'est ajouté au DESELA avec le lien entre la racine et le mot dérivé que s'il n'y était pas. Si seule la racine y était, ce couple est passé à l'analyseur pour extraire les affixes et le modèle et si ceux-là n'existent pas dans DESELA, ce mot est ajouté au DESELA avec le lien avec la racine, ils y sont ajoutés.



Mot/Lettre	Transcription
إ	Alef avec Hamza au dessus
أ	Alef avec Hamza on dessous
آ	Alef avec Maada
ب	Baa
ة	Taa Marbouta
ت	Taa
ث	Tha
ج	Jeem
ح	H'a
خ	Khaa
ر	Raa
ز	Thal
س	Seen
ش	Cheen
ص	Saad
ض	Daad
ط	T'aa
ظ	Zha
ع	Ain
غ	Ghain
ف	Faa
ق	Qaf
ك	Kaf
ل	Lam
م	Meem
ن	Noon
ه	Haa
و	Waw
ؤ	Waw avec Hamza
ى	Alif Makzora
ي	Yaa
ئ	Hamza avec Hamza
من	Min
إلى	Ila
أكل	Akala
الأكل	Alakil
المأكلة	Almaekala
المأكول	Almaekol
إيكالا	ikalan
المأكولات	Almaekolate
تقول	takole

أكلت	Akalte
أكل	Akale
أكلا	Aklan
أكلة	Aklah
أكلها	Akalaha
تأكلا	Taeakolan
غرب	Garaba

*Table 1 : Les transcriptions des lettres et des mots arabes utilisés dans ce document.*

### 3 Résultat

Le résultat principal de ce travail est le nouveau dictionnaire électronique structuré et évolutif de la langue arabe (DESELA) qui peut être présenté sous la forme d'une base de données relationnelle ou d'un document XML. DESELA contient essentiellement les racines, les préfixes, les suffixes, les infixes, les modèles et les mots dérivés. De plus, pour un mot donné, il fournit les liens avec sa racine, avec les affixes associés et avec son modèle éventuel.

Un deuxième résultat est le système qui permet d'alimenter DESELA automatiquement à partir d'un ou de plusieurs dictionnaires textuels classiques et de l'enrichir en permanence avec des corpus textuels arabe quelconques. Ce système est composé d'un parseur, d'un classifieur, d'un comparateur et d'un analyseur. Le parseur permet de transformer une source textuelle (dictionnaire ou corpus textuel) en un ensemble de mots bruts. Le classifieur permet de classer un mot donné et de l'ajouter au DESELA en tant qu'une racine ou en tant qu'un mot dérivé. L'analyseur permet d'extraire les affixes et le modèle à partir d'un mot dérivé et de sa racine. Le comparateur permet d'éviter d'avoir des doublons, à tous les niveaux, dans DESELA.

### 4 Conclusion

Dans cet article, nous avons présenté DESELA le nouveau dictionnaire électronique structuré et évolutif de la langue arabe. Ce nouveau dictionnaire peut être présenté sous la forme d'une base de données relationnelle ou d'un document XML facilement exploitable à l'aide des langages de requêtes appropriés. Ce nouveau dictionnaire contient les racines, les préfixes, les suffixes, les infixes et les modèles, en plus des informations fournies par un dictionnaire classique. De plus, il fournit les liens d'un mot donné avec sa racine, avec les affixes associés et avec son

modèle éventuel [2], [3], [6]. Nous avons présenté aussi le système automatique qui permet d'alimenter et d'enrichir DESELA à partir d'un ou de plusieurs dictionnaires textuels classiques et des corpus textuels arabe quelconque.

Notre dictionnaire électronique évolutif et structuré comble un besoin au niveau du patrimoine électronique arabe. Ce dictionnaire peut être utilisé pour évaluer les méthodes d'extraction d'information à partir d'un document arabe cette évaluation contribue sans doute à améliorer les méthodes existantes d'extraction d'information à partir des documents arabes. L'originalité de notre dictionnaire réside dans le fait qu'il s'agit d'un dictionnaire évolutif qui contribue aussi à l'évolution de la langue arabe.

## **5 Perspective**

La prochaine étape est de doter DESELA d'une dimension sémantique en ajoutant des relations sémantiques entre les mots. Pour établir les relations sémantiques entre les mots il faut que nous exploitons les caractéristiques des dictionnaires classiques. En général, un dictionnaire classique fournit les mots avec leurs synonymes. Ces synonymes peuvent être des mots ou des racines. Donc, nous utiliserons le classifieur et l'analyseur pour établir les relations sémantiques aux niveaux des mots et des racines. Ainsi, pour déterminer les relations sémantiques entre deux mots on pourra passer par leurs racines.

**Remerciements :** Ce travail est effectué dans le cadre des projets "Arabic Web Intelligence" financé par le Centre National de Recherche Scientifique Libanais (CNRSL) et « Recherche d'information Multimedia Multilingue Arabe » financé par le comité Franco-Libanaise (CEDRE).

## **6 Références bibliographique**

- [1] W. Adamson George, J. Boreham, *The use of an association measure based on character structure to identify semantically related pairs of words and document titles*, Information Storage and Retrieval, Vol. 10, pp 253-260, 1974.
- [2] I. Al Kharashi, *A Web Search Engine for Indexing, Searching and Publishing Arabic Bibliographic Databases*, 1999.
- [3] A. Chen, F. Gey, *Building an Arabic stemmer for information retrieval*.TREC-11 conference 2002.

- [4] K. Darwish, *Building a Shallow Arabic Morphological Analyzer in One Day*. The ACL-02 Workshop on Computational Approaches to Semitic Languages, Philadelphia, USA, 2002.
- [5] L. S. Larkey, L. Ballesteros, M. E. Connel, *Improving Stemming for Arabic Information Retrieval: Light Stemming and Co-occurrence Analysis*, Proc. of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 275 – 282, 2002.
- [6] H. Suleiman Mustafa, *Character contiguity in N-gram based word matching: the case for Arabic text searching*. Information Processing and Management.41 (4), 819-827, 2004.
- [7] G. Kanaan, R. Al-Shalabi, J. Jaarn, M. Al-Kabi, A. Hasnah, *A New Stemming Algorithm to Extract Quadri-Literal Arabic Roots*, 2004.
- [8] K. Taghva, R. Elkoury, J. Coombs, *Arabic Stemming without a root dictionary*, International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume I pp. 152-157, 2005.
- [9] H. Al Ameen, S. Al Ketbi, A. Al Kaabi, K. Al Shebli, N. Al Shamsi, N. Al Nuaimi, S. Al Muhairi, *Arabic Light Stemmer: A new Enhanced Approach*, The Second International Conference on Innovations in Information Technology (IIT'05), 2005.
- [10] L. Larkey, L. Ballesteros, M. Connell, *Light Stemming for Arabic IR*, Arabic Computational Morphology: Knowledge-based and Empirical Methods, A. Souidi, A. Van Bosch, and G. Neumann Editors. Kluwer/Springer's series on Text, Speech, and Language Technology, 2005.
- [11] F .Douzidia, G. Lapalme, *Un système de résumé de textes en arabe*, 2ème Congrès International sur l'Ingénierie de l'Arabe et l'Ingénierie de la langue, Alger, 2005.
- [12] Y. Kadri, J. Nie, *Effective Stemming for Arabic Information Retrieval*, proceedings of the Challenge of Arabic for NLP/ MT Conference, Londres, Royaume-Uni, 2006.
- [13] L. Khreisat, *Arabic Text Classification Using N-gram Frequency Statistics A Comparative Study*, The 2006 International Conference on Data Mining Part of the 2006 World Congress in Computer Sciences DMIN: 78-82, 2006.
- [14] F. Ahmed, A. Nürnberger, *N-grams Conflation Approach for Arabic*, ACM SIGIR Conference, Amsterdam, 27 Juillet 2007.
- [15] A. M. El-Halees, *Arabic Text Classification Using Maximum Entropy*, The Islamic University Journal (Series of Natural Studies and Engineering) Vol. 15, No.1, pp 157-167, ISSN 1726-6807, <http://www.iugzaza.edu.ps/ara/research/>, 2007.
- [16] A. Khemakhem, B. Gargouri, A. Abdelwahed, G. Francopoulo, *Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF - ISO 24613*, Traitement Automatique des Langues Naturelles, Toulouse, France, 5- 8 Juin 2007.

- [17] M. Ben Abderrahmen, B. Gargouri, M Jmaiel, *LMF-QL: A graphical Tool to Query LMF databases*, Third Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland, 2007.
- [18] E. Norbert, *Arabic Language Support in SQL Server*, Microsoft corporation, *SQL Server Technical Article*, [http://msdn.microsoft.com/en-us/library/cc295829\(SQL.90\).aspx](http://msdn.microsoft.com/en-us/library/cc295829(SQL.90).aspx), 2008.
- [19] C-A. Comes, L-D. Savu, I-O Spatacean, B. Stefan, A. Avram, *Universal Symbolic Translator for Procedural Language over SQL*, 7th WSEAS Int. Conf. on Applied Computer & Applied Computational Science (ACACOS '08), Hangzhou, Chine, 6-8 Avril, 2008
- [20] J. Micher, C.Voss, *Buckwalter-based Lookup Tool as Language Resource for Arabic Language Learners Software Engineering, Testing, and Quality Assurance for Natural Language Processing*, pages 66–67, Columbus, Ohio, USA, June 2008.
- [21] M. Sinane, M. Rammal, K. Zreik, *Arabic documents classification using N-gram*, Conférence ICHSL6, Toulouse, 2008.
- [22] F. Baccar, A. Khemakhem, B. Gargouri, K. Haddar, A. Ben Hamadou, *Modélisation normalisée LMF des dictionnaires électroniques éditoriaux de l'arabe*, TALN 2008, Avignon, France, 9-13 juin 2008.
- [23] G. Francopoulo, M. George, *Language resource management – Lexical markup framework (LMF)*, ISO/TC 37/SC 4 Rev.15, 2008.
- [24] A. Al Hajjar, M. Hajjar, K. Zreik, *Classification of Arabic Information Extraction methods*, 2nd International Conference on Arabic Language Resources and Tools, Le Caire, Egypte, 21-23 Avril 2009.
- [25] A. Al Hajjar, M. Hajjar, K. Zreik, *Un nouveau système d'évaluation des méthodes d'extraction de la racine des mots arabes*, (soumis), 2009.
- [26] Ibn Manzour, *Lisan Al-Arab*. [www.muhammadith.org](http://www.muhammadith.org), 2009.
- [27] Sakher, *Lexicons: Lisan Al-Arab, Al Qamous Al Mouhit, Al Wasit, Al Mouhit, Mouhit Al Mouhit, Al Ghani, Taj Al Arous, Najaat Al Raed*, <http://lexicons.sakhr.com>, 2009.
- [28] Academy of the Arabic Language, <http://lexicons.sakhr.com/intro/intro.aspx?fileurl=intro01.asp>, 2009.
- [29] Islamic Library, *Arabic Dictionaries: Al Misbah Al Mounir, Al Qamous Al Mouhit, Moujam Makayys Al Lougha, Moukhtar Al Sihah*, <http://www.islamweb.net/newlibrary/bookslst.php?subject=كتب اللغة العربية>, 2009.