P- ISSN 1991-8941 E-ISSN 2706-6703 *Journal of University of Anbar for Pure Science (JUAPS)* Open Access
2016,10 (3) :41-49

Root-Stem Approach in General Analyzer System for Arabic Language (RSGAS)

Abeer K.Al-Mashhsdany[*] **Abdulwadood K. Al-Mashhadany**[**] **Waleed K. Al-Mashhadany**[***]

[*] Al-Nahrain University - College of Science
[**] Iraqi Ministry of Health, Baghdad.
[***] Iraqi Ministry of Education, Baghdad

ARTICLE INFO

Received: 19 / 5 /2022
Accepted: 28 / 5 /2022
Available online: 19/7/2022
DOI: 10.37652/juaps.2016.135154

Keywords:
Root-Stem-Based Approach;
Shell Systems;
Word Analysis;
Arabic Morphological Rules;
Synonyms and antonyms words.

ABSTRACT

Many application programs require knowledge from the user. In such programs, number of problems may occur. Problems related to natural language understanding such as; typos, duplication, similarity, and inconsistent knowledge. So, the analysis of text is very important in such programs. A previous work was done to solve problems of natural language understanding. That work build Arabic analyzer and merge it with a shell hybrid system. This work attempts to build a general Arabic Analyzer system, so that it could be merged with other programs to solve the natural language understanding problems. This work attempts to correct the previous Arabic analyzer, so its dictionary will be smaller. Morphology in this work merges root-based with stem-based approach. This merging forces this work because it gains the advantages of the two approaches. Root-based approach gives the advantage of smaller size. Stem-based approach gives the advantage of passing the problem of irregular cases. This work is compared with other two valued techniques. Results of the comparison show that this work has a lot of advantages over the previous techniques.

Introduction

It becomes necessary to provide text analyzer system, which has ability to be merged with application programs that need knowledge from the user. Because of problems that may occur during receiving of knowledge. Knowledge may be wrong due to; typos, inconsistent with another received knowledge, or it may be duplicated knowledge [1, 2].

This work attempts to build an Arabic analyzer system, which has the ability to be merged with systems that acquire knowledge from external user. One example of such systems is shell systems. Shell system acquires knowledge from external user, and it acquires problems from another external user. In an Arabic expert system shell, the main components of shell system are merged with technique for Arabic analyzer. Shell system has the same components of an expert system except the knowledge base [1, 2, 3].

Arabic language is a highly inflected language and has a complex morphological structure. Stemming is one of many tools used for solving Arabic morphological problems [4, 5, 6]. Previous researches proved that light stemming is effective for Arabic language [1, 7].

There are few attempts to build an Arabic analyzer. Haidar Moukdad (2006)[8], he used rules of Arabic language in the field of information retrieval. His paper compares the effects of stemming and root retrieval in Arabic information retrieval through an exploratory study of the handling of Arabic words by an English-language search engine (ELSE). The results of the experiments show that more effective retrieval can be accomplished through stemming, and that it is

[*] Corresponding author at:Al-Nahrain University - College of Science.E-mail address: aabeeeeraa@yahoo.com

possible to adapt an ELSE to be used with Arabic without the need to develop root-retrieval features.

In two previous works, attempts to build Arabic analyzer system had been made. The two previous works used Arabic morphology in two different methods. Each one of the two previous works embeds different strategy of Arabic morphology. ADESS is a root-based Arabic analyzer system (2010) [9]. It applies the classical Arabic morphological rules. It uses the grammar of deriving verbs in different formulas, and the grammar of deriving nouns from verbs. KISB is a stem-based Arabic analyzer system (2012) [1]. The result of comparison between the two Arabic analyzers shows that KISB Arabic analyzer is professional in passing irregular cases in Arabic language, but its dictionary needs more and more size in comparison with ADESS.

This paper uses strategy merges between root-based and stem based approaches to build the Arabic analyzer system. This work (RSGAS) forces Arabic analyzer system. It provides the advantages of both stem-based and root-based approaches. The results of comparison between the three Arabic analyzer systems ADESS, KISB, and RSGAS will be shown later in this paper.

The Proposed System

Root-Stem approach in General Analyzer System for Arabic language (RSGAS). It consists of five main modules (as illustrated in figure 1). Each module performs a specific task; more explanation about each module is given.

1. System Interface Module

System interface module performs the task of communication between RSGAS Arabic analyzer and application program. It receives Arabic phrases (sentences) from the user of the application program.

System interface module interacts with application program via both external user and database. The received Arabic phrase must be analyzed. Then the analyzed phrase posted to database. During analyzing time, problems may occur. Typos cause problems. When there is any unknown word, this module has to tell the user, and it gives him list of suggestions to choose the best suited action. Other problems occur because of inconsistency. An Arabic phrase may be; matched, closed in meaning, or inconsistent to another existing one. This module discovers that by its interaction with the database. So it must tell the user to decide the best suited action.

Inside the analyzer, system interface module interacts with the word analysis, dictionary, and meaning modules. Interface module performs the task of phrase (sentence) analysis. The following steps describe briefly “phrase analysis process” and the tasks of system interface module.

Step1/ decomposing and filtering: system interface module decomposes the input phrase into words. Then removes all stop words as well as noise words, via its interaction with the dictionary module. But a checking is required before removing stop words. If there is a stop word refers to negation then turn on the negation flag of this phrase. Then it checks the synonym file to replace the remained words with their synonyms if found.

Step2/ Word analysis: Now, interface module sends each word to the word analysis module then receives the root of word as a feedback. In case of error, feedback will contain list of suggestions to solve the problem.

Step3/ Phrase meaning: When the input phrase is ready with its roots, it will be send to the meaning module, in which any similarity between phrases will be discovered.

Step4/ Learning antonymous phrases: Now check if there is word stored in the antonym file. If so, add a pointer to this phrase.

2. Dictionary module

RSGAS design of dictionary agrees with KISB dictionary in its main structure. It has a separate dictionary for each domain; variable part. And all domains associated with a part of dictionary called constant part.

As found in KISB, general files are in the constant part (stop-word negation file, other stop-word file, affixes file and other files needed for analysis). The idea of the constant part is to isolate the words that are necessary for all domains of knowledge bases. It means storing them only one time instead of repeating them for all domains. The variable part at RSGAS dictionary is decomposed into domains. Each domain has dictionary. The idea of decomposing the dictionary is to limit the range of search during word analysis, because of the huge amount of words in Arabic language. Each dictionary contains; root file, stem file, noun file, synonym file, antonym file and noise file. RSGAS dictionary differs from KISB dictionary in few points. These points ensure enhancement of system performance. In constant part; it decomposes the stop-words file into two files; negation stop-word file, others stop-word file. In variable part; it decides to use the property of antonym words. This property helps in forcing the work of meaning module. So the antonym words file is used to store each word with its antonym. Each one of the two words is linked to list. The list contains pointers to all phrases that contain this word. At first this list must be empty. During building the knowledge base, RSGAS will learn and fill the contents of the list. This will be done by meaning module.

Arabic, as a member of the Semitic Languages Family, suffers from huge number of compound words, which result in high out-of-vocabulary (OOV) when using finite-sized lexicons. It contains words that are highly infected and derived and therefore needs special techniques and algorithms for solving its morphological problems [9]. It is difficult or impossible to provide technique cover all cases of morphological problems. Because there are many irregular cases, RSGAS dictionary provides stems only for solving problem of irregular cases. It does not save stems that could be analyzed by morphology technique. So it saves roots for all domain words, and saves stems of irregular cases. This will decrease dictionary size in comparison with KISB.

When a database has to interact with RSGAS; the knowledge base is desired to be decomposed into domains. RSGAS provides the ability of ordering knowledge bases in two levels of domains. First there are domains, and inside each domain there are specific domains. For each specific domain, a number of knowledge bases could belong.

3. Morphological Rules Module

RSGAS morphological rules module provides classical Arabic morphological rules (grammar). There are many morphological rules such as: the rules that is used in the derivation of the verbs in different formulas (verbs tenses), and rules that is used in the derivation of the nouns from verbs. In other word, rules of derivations means, rules of adding infixes for roots, in order to generate measures. Figure 2 shows RSGAS measures of derivations. In figure 2 verbs are showed in past tense. Each verb could be formulated in different tenses. Figure 3 describe the rule of verbs tenses.

In addition to rules of derivations (infixes), there are rules of affixes (prefixes & suffixes). Adding infixes to root generates stem. Prefixes and/or suffixes may be attached to stems. So the form of word will be; prefix, followed by stem, then followed by suffix. Figure 2 and 3 show measures of all stems in Arabic language. The stem may be noun or verb. List of prefixes and suffixes that attached to verbs differs from others that attached to nouns. RSGAS depends on lists of prefixes and suffixes that are mentioned in details in KISB [1].

4. Word Analysis Module

Word analysis module merges between Arabic morphological analysis (root-based approach) in ADESS and stem-based approach in KISB. It controls the task of word analysis in RSGAS. It interacts with interface module, from which it receives the source word to be analyzed. Then the root of the analyzed word will be send back. If there is error in source word, then word analysis module will send back the source word with list of suggestions. That is because word analysis module includes good spellchecker routine in order to solve the problem of typos.

Word analysis module interacts with the two modules; dictionary and morphological rules. This interaction gives it the ability of analyzing Arabic words. Each source word is converted into root depending on the morphological rules then it must be found in the dictionary, to be sure that the source word is right word according to the specified domain. Word analysis module has the ability of learning new words and adding them to the dictionary.

The steps of word analysis are the following:

Step1/ Affix-1: This step removes prefixes and suffixes from the source word. This process depends on options of affixes provided by morphological rules module. The result of this step is the stem.

Step2/ Stemming: Now, the word may be one of the irregular stems that have been saved in the dictionary. If it is so, take its root from the dictionary.

Step3/ Affix-2: This step removes infixes. It applies rules of measures that provided by morphological rules module. It determines infixes depending on measures. Now the root word is ready. It means that the source word is right word belong to the specified domain dictionary.

Step4/ Error checking: If the root could not be found, in this case two options are possible: First is typos, and second if it is a right word but it is not found in the dictionary. Typos problem is solved by spellchecker routine, list of right words will be generated. The problem of missed word is solved by the ability of learning new words. List of right words will be sent to external user, to choose one of them. Or, he can say that the source word is right, at this time word analysis module knows that this word must be added to the dictionary.

5. Meaning Module:

RSGAS meaning module is necessary for discovering inconsistency. It interacts with system interface module, which represents the intermediate module between it and the knowledge bases that stored in the database (at the application program).

Knowledge bases that are acquired from external user may contain many mistakes. When knowledge base is huge, more than one user may be needed for entering rules. Rules may be duplicated, or one rule may be inconsistent with another rule. So, meaning module is needed to perform the task of checking similarity ratio between two Arabic phrases (sentences). To perform this task, meaning module needs interaction with the dictionary module. This interaction is connected via system interface module.

Meaning module works in two directions. First direction studies data on dictionary. It performs dictionary refinement process. While the second direction studies the received phrase. Dictionary refinement process includes considering data stored in dictionary. Then try to inference new facts. Figure 4 explain refinement process. At the second direction; when it works on the received phrase, new data will be generated and added to dictionary. It generates links from each word in the antonym file to any phrase that contains it.

This module receives Arabic phrase (sentence) as a list of root words. It provides exchanger routine to replace words; depending on meaning, then check whether the new phrase is similar to another existing one. The following steps describe what will be happen to any received phrase.

Step1/ Synonym: This step checks the received root words to replace synonyms; depending on synonym file that is saved in the dictionary module.

Step2/ Antonymous: This step checks if the received words include antonym. Then turn on the antonymous flag, and point to the antonymous words.

Step3/ Duplication: Now, compute the similarity rate (Sr) between this phrase and others that are stored. When Sr is more than 60%, the two phrases may be the same.

Sr= N2/N1 *100, where N1 is number of words (roots) in phrase, N2 is number of matched words.

Meaning module has to prevent the duplication of a phrase in one rule. Also it has to prevent the duplication of rules.

Meaning module recognizes the antonym words in the phrase. Now, RSGAS can discover any two antonymous phrases. So it can make inferences about the following:

- If there are two antonymous phrases with different negation flags then they are the same.
- Prevent any two antonymous phrases at the same rule.
- When any phrase is proved, the antonymous phrase must be impossible.

Results and Discussions

This work needs real KB and real dictionary, because it is necessary to test its behavior and checking its rightness. It is good idea to use the Abdominal Pain KB, which is provided by ADESS [9]. First, Medicine-Domain Dictionary has been built and provided to RSGAS (Appendix A). The older versions of Medicine-Domain Dictionary were provided to ADESS and KISB. Each new version is different from the previous in a number of points. That is because each system has its own structure of dictionary which is different from others.

Now, we have three versions of Medicine-Domain Dictionary, and two previous versions of Abdominal Pain KB, if we know that the knowledge base was constructed using ADESS [9] and KISB [1]. This work could not be used to construct any KB unless merging it with external application program to provide the KB and hold it. So, that will be next work at future time. Real Arabic sentences and phrases from the Abdominal Pain KB were analyzed in this system. To study the behavior of RSGAS, a comparison among the three systems will be stated here.

The three systems were built by Visual Prolog. The results showed that RSGAS likes ADESS and KISB in its ability to learn new words, discovering typos, and discovering the Arabic phrases that are similar in meaning. But RSGAS has more tools and facilities. RSGAS can discover the antonymous sentences (phrases). It links each antonym word with its appearances in the stored sentences (phrases), this facilitate the task of meaning module, when it is necessary to discover the antonymous sentences. RSGAS has the ability of refining its dictionary. It makes inferences to learn new data.

The merging of Root-Stem-based approaches in this work gives the insurance for passing the problem of wideness Arabic morphological rules. RSGAS succeeded in covering irregular cases which could not be covered completely in ADESS. This work solves the problems of many cases, such as: “جمع التكسير”، “اللفيف”، “المفروق”، “اللفيف المقرون”، “الاجوف”، “المثال”, and “الناقص”. For example, the stems “أعضاء”، “عضو” and “عضوي” (Table.1 in appendix A) are derived from root “عضا” which could not be covered in ADESS. Also KISB succeeded in covering irregular cases in Arabic language rules. But KISB needs more size than RSGAS. The merging of Root-Stem-based approaches in this work affects in decreasing dictionary size. KISB stores all needed stems in its dictionary, while RSGAS use stems only for irregular cases. It uses root-based approach for regular cases. For example, the words (“انقباض”، “ينقبض”، “انقبض”، “منقبض”، “تنقبض”” and “يقبض”) are derived from the root “قبض”. In RSGAS, the dictionary contains only the root “قبض” (Table.1 in appendix A). Other derivation words are found by word-analysis-module. In irregular cases, RSGAS stores all stems with root in the dictionary. For example the stems (“تفيق”، “فوق”، “تفوق”، “فوقان”، “يفيق”” and “فاقة”) with root “فاق” (Table.1 in appendix A).

The three systems succeeded in discovering typos and inconsistent rules with high accuracy rate 98%-100%. But each one of the two previous systems has its limitations. ADESS and KISB did not study antonym property, so at such text they may fail in discovering similar or inconsistent phrases. Another limitation point at ADESS, it could not covering irregular cases in Arabic language rules, so it may fail too. But at RSGAS, there is no failing, as long as all words in tested phrases are studied in the dictionary. If any word not found, RSGAS can learn it, then analyses of meaning will not fail.

Differences among ADESS, KISB and RSGAS are seamed clearly in their dictionaries. Table 1 and figure 5 illustrate some important points in their dictionaries variable parts. ADESS has the smallest size, but it failed with irregular cases. The size was big at KISB then go down at RSGAS. Size of main file was increased 3.9 times at KISB then decreased 19% at RSGAS. The antonyms file that is used by RSGAS increases its size 11 K.byte, but it provides more abilities. It forces performance of meaning module. To discuss the advantages of meaning module follow the following points:

1. In RSGAS, dictionary refinement process discovers new facts after considering data stored in dictionary. For example; it discovers the synonyms (“قوا”=“ثقل” and “كثر”=“زاد”). It also discovers the antonyms (“شدّ”X”خفّ” and “قسا”X”خفّ”).
2. In RSGAS, meaning module validates knowledge, see the following examples:
   - It discovers that the two sentences (“ الألم يزول عند النوم” and “لا يظهر الألم أثناء النوم”) are the same. So they must not be duplicated.
   - It discovers that the two sentences (“سرعة دقّات القلب” and “تباطئ بدقّات القلب”) are antonymous. So must not be at the same rule.
   - Once the phrase “الألم مفاجيء” is proved then it discovers that the phrase “يظهر الألم بالتدريج” is impossible.

Conclusions

1-Merging stem-based with root-based reduces the dictionary size. So RSGAS dictionary saves space in comparison with KISB dictionary, in which, stem-based approach is used only.

2-Although RSGAS saves space, it succeeded in passing the problem of wideness Arabic morphological rules and irregular cases.

3-Adding antonym words gives RSGAS more advantages. Studying more properties of words such as synonym and antonyms makes refinement process more flexible and makes RSGAS close to natural language understanding.

References:

[1]Al-Mashhadany A.(2012). Knowlwdge Acquisition & Hybrid Inference with A Stem-Based Approach (KISB). Journal of Al Nahrain University-Science Vol.15: No.2 169-183.

[2]Al-Mashhadany A. (2012). General Arabic Diagnosing Expert System Shell (GADESS). LAP LAMBERT Academic Publishing, Germany.

[3]Al-Khateeb B., Samawi V. and Bashaga T.(2005). Visual Arabic Expert System Shell (VAES). Journal of Science and Engineering Vol.3: 75-83.

[4]Goweder A., Alhami H., Rashed T. and Al-Musrati A. (2008). A Hybrid Method For Stemming Arabic Text. Proceedings of the 9th. International Arab Conference on Information Technology, Zarqa Private University, Jordan. http://eref.uqu.edu.sa/files/eref2/folder6/f181.pdf

[5]Al Ameed H., Al Ketbi S., Kaabi A. Al, Al Shebli K., Al Shamsi N., Al Nuaimi N. and Al Muhairi S.(2005). Arabic Light Stemmer: Anew Enhanced Approach. The Second International Conference on Innovations in Information Technology (IIT'05), Dubai, UAE. http://www.it-innovations.ae/iit005/proceedings/articles/g_1_iit05_hayder.pdf

[6]Al-shalabi R. and Kanaan G.(2004). Constructing An Automatic Lexicon For Arabic Language. International Journal of /computing & Information Science.Vol.2:N0.2. http://www.ijcis.info/Vol2N2/114-128OKS.pdf

[7]Duwairi R., Al-Refai M. and Khasawneh N.(2007). Stemming Versus Light Stemming as Feature Selection Techniques for Arabic Text Categorization. Innovations in Information Technology 2007.IIT'07. 4th International Conference on P 446-450, 18-20 Nov. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4430403&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4430403

[8]Moukdad H.(2006). Stemming and root-based approaches to the retrieval of Arabic documents on the Web. Webology Vol.3:No.1. http://www.webology.org/2006/v3n1/a22.html

[9]Al-Mashhadany A., Bashaga T. and Samawi V.(2010). Arabic Diagnosing Expert System Shell (ADESS). The Proceedings of 2nd Conference for Information Technology: Applications and Horizons, University of Technology, Baghdad, Iraq P172-193.

[10]السعيد عبداللطيف .(2006)قواعد اللغة العربية المبسّطة. شبكة مشكاة الإسلامية.almeshkat.net http://www.almeshkat.net/books/open.php?cat=16&book=2437

[11](١٩٩٩)الرازي محمد .مختار الصحاح. المكتبة العصرية- الدار النموذجية.

[12]Colledge N., Walker B. and Ralston S.(2010). Davidson's Principles & Practice of Medicine. 21St Edition, Robert Britton, Churchill Livingston Elsevier.

[13]Goldman L., Ausiello D., Arend W. and Armitage J.(2007). Cecil MEDICINE. Saunders.

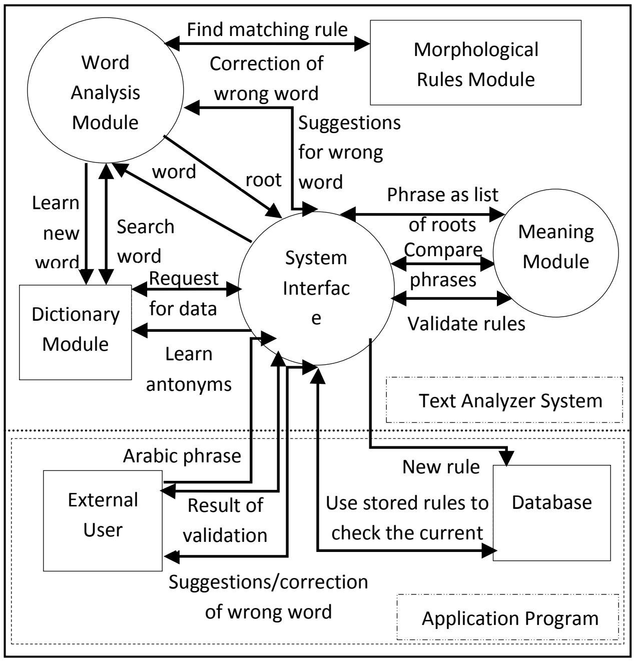

Figure 1. RSGAS architecture.

Roots of words in Arabic language are either “فعل” or “فعلل”

Measures of verbs derived from “فعل”:

افعل، فاعل، فعّل، انفعل، افتعل، افعلّ، تفعّل، تفاعل، استفعل، افعوعل، افعالّ.

Measures of verbs derived from “فعلل”:

تفعلل، افعنلل، افعللّ.

Measures of nouns derived from verbs:

فعالة، فعال، فعيل، فعلان، فعول، فعلة، فعلل، فعللة، فعلال، افعال، افعلة، تفعيل، تفعلة، مفاعلة، انفعال، افتعال، افعلال، استفعال، تفاعل، تفعلل، فاعل، افعل، مفعول، فعّال، مفعيل، فعّيل، فاعول، فاعلة، فعّالة، فعلى، مفعل، مفعلة، مفعال، متفاعل، مفعّل، فعلي، تفعّل، متفعّل، فعلاء، مفاعيل، مستفعل.

Figure 2. Arabic Morphological Grammar.

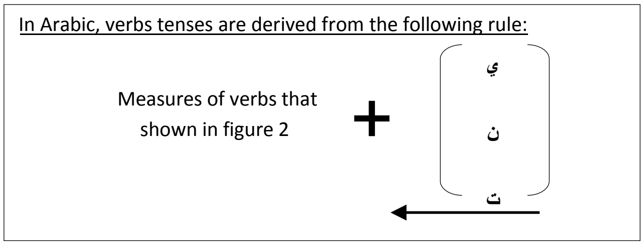
In Arabic, verbs tenses are derived from the following rule:
Measures of verbs that shown in figure 2
+
ي
ن
ت

Figure 3. Arabic Verbs Tenses.

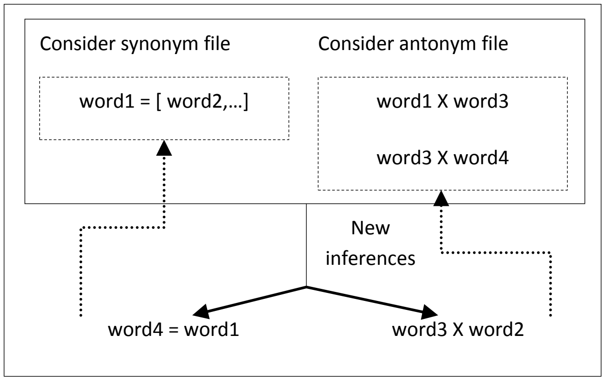

Figure 4. Dictionary refinement process.

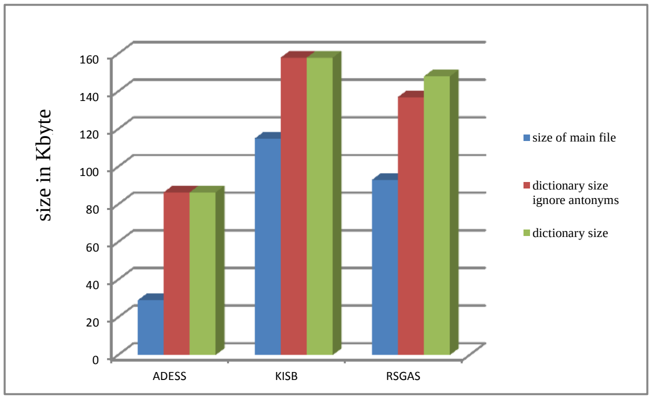

Figure 5. Dictionary size of the three works.

Table 1. Main differences between the three works' dictionaries (ADESS, KISB, and RSGAS).

Points of differences Works	Dictionary files at variable part	The main file & size	Dictionary size
ADESS	Root, Noun, Noise and Synonym	Root file 29 K.byte	86.4 K.byte
KISB	Stem, Noun, Noise and Synonym	Stem file 115 K.byte	158 K.byte
RSGAS	Root, Stem, Noun, Noise and Synonym, antonyms	Root & Stem files 93 K.byte	Without antonyms file 137 K.byte With antonyms file 148 K.byte

Appendix A

Table.1 Medicine-Domain Dictionary (Roots, Stems)

Root	Verb-Stems	Noun-Stems	Root	Verb-Stems	Noun-Stems	Root	Verb-Stems	Noun-Stems
أصل			أكل	آكل، تآكل، يتآكل، تتآكل	تآكل، آكل	ألم	آلم	مؤلم، آلام
أنث		أنثى، إناث، مؤنّث	امّ			انف		
بدا	تبدّى، استبدا، تبادى، يبتدي، يبتدّى،يستبدي، يتبادى		بدأ		بداية، ابتداء	برز		
بشر			بطن			بطأ	يتباطأ، تتباطأ، تباطأ	تباطأ، بطيء
بعد			بقي	تبقى، يبقى، تبقّى، يتبقّى، تتبقّى		بهر		
بوب		باب	بول	با ل		بيض		
تبل		توابل	تعب			تمّ		
تبع			ثبت			ثقل		
ثلث			ثني		اثنان،أثن اء	جبا		جيب، جيوب
جزأ		جزئي، أجزاء	جاع	يجوع، تجوع	جائع، جوع، جوعان	جدر		
جرح			جسم			جفّ		
جلد			جلس			جمع		
جنب			جنس			جهد		
جهز			جوف			حج ب		
حجم			حدّ			حرّ		
حرق			حرك			حس ب		
حسّ			حشا		حاش ية	حص ى		حصوة
حكّ			حلّ			حمر		
حمض			حمل			حمّ		حمّى ، حمية
حوا	احتوى، يحوي، يحتوي، تحتوي	محتو ى	حوض			حوط	يحيط ، تحيط	محيط
حول		حال، حالا	حيض	حـ ا ض	حائض، إستحاض ة	خاف		خوف
خبث			خثر			خدم		
خرج			خصر			خطر		
خفّ			خفق			خفي	اختف ى	إخفاء، اختفاء
خلل			خلف			خلق		
خلّ			خمر			داخ	يدوخ ، تدوخ	دوخة
دام	يدوم، تدوم	دائم، دوام	دخل			دخن		
درج			درّ			دقّ		دقائق
دكن			دلا		دوال ي	دما		دم، دموي، دماء

دمن			دهن			دوى		دواء، أدوية
دود		ديدان	دور			ذكر		
ذهب			رأس			رأى		رئة
راح	يستريح، تستريح		ربع			رجف		
رجل			رحم			ردّ		
رعف			رفع			رقّ		
رقب			زاد	يزيد، تزيد، يزداد، تزداد	زيادة، إزدياد	زال	يزول، تزول	زوال
زحر			زمن			زوى		زاوية
سال	يسيل، تسيل	سيلان، سائل	سبّ			سبق		
سحب			سدّ			سرع		
سرق		مساريق	سعف			سعل		
سفل			سقى		إستسقاء	سمر		
سمن			سهل			سوء	ساء	سيّئ
سوع		ساعتين ، [illegible]	شبع			شحب		
شدّ			شرب			شرى		شريان، شرايين
شعب			شمّ			شنج		
شها	اشتهى، يشتهي، تشتهي	شهية، شهوة	شيم		شامة	صاب	يصيب، تصيب	
صار	يصير، تصير		صبح			صدر		
صدع			صدم			صرّ		
صعب			صغر			صفر		صفراوي
صفق			صفي			صلب		
صمّ			صوت			ضاق		
ضحل		مضمحل	ضخم			ضعف		
ضغط			ضلع			ضمر		
طحل			طرف			طعم		
طفّ			ظهر			عجز		
عدّ			عدل			عدم		
عرض			عرق			عسر		
عشر			عضا		عضو، أعضاء، عضوي	عضل		
عقد			عكب		عنكبوت	علا		علوي، أعلى، عالي
علق			عمر			عمق		
عمّ			عند			عنق		
غثا		غثيان	غدّ			غدر		
غشا		غشاء، مغشيّ، غشاوة	غضب			غمق		
غمي		إغماء	غوط			غير		
فاق	يفوق، تفوق، يفيق، تفيق	فوق، فاقة، فوقان	فتح			فتق		
فجأ		مفاجئ	فجر			فحص		
فرز			فرط			فرغ		
فصل			فعل			فقد		
فقر			فم			قام	يقوم، تقوم	قيام
قبض			قبل			قدم		
قرب			قرص			قرح		
قسا		قاسي قسوة	قصر			قلب		
قلص			قلّ			قنا		قنوات، أقنية

قوا	تقوّى، يقوّي، تقوّي، تقوّى، تتقوّى	قوي، قوّة، مقوّي، مقويّات، تقوية	قوم		قيام، إقامة، تقويم	قيأ		
قيح			كأب		كآبة	كبد		
كبر			كتف			كتل		
كثر			كرّ			كسب		
كسل			كلا		كلى، كلية، كلوي	كلّ		
كمل			كيس			لوى		إلتواء، ملتوي
لحّ			لسن			لمس		
لهب			لوث			لوح		
ليل			ماء		مائي	مال	يميل، تميل	مائل
مدّ			مرا		مريء ، مريئ	مرّ		
مرض			مزق			مسك		
معى		معوي، أمعاء، معي	معد			مغض		
ملّ						موج		
ناب	تناوب، يتناوب، تتناوب	نوبة، نوبات، تناوب، مناوبة، متناوب	نام		نوم، نائم	نبذ		
نتج			نتن			نحف		
نزف			نسا	نسي، ينسى، تنسى	نسيان، نساء	نشط		
نصب			نصف			نظر		
نظف			نظم			نفخ		
نفس			نقر			نقص		
نقط			نقل			نما	ينمو، تنمو	نمو، نامي
نهي			هاج	هيّج، يهيّج، تهيّج، يتهيّج	هيجان، تهيّج، متهيّج	هبل		
هجم			هضم			وجع		
وذم			ورث			ورد		
ورم			وسط			وسع		سعة، متسّع
وصل		متّصل	وعى	يعي، تعي	وعاء، أوعية	وقع		
وضع			وعك			وقت		مؤقّت
ولد			يرق			يسر		
يمن								

Table.2 Medicine-Domain Dictionary (Nouns, Noise, Synonyms and Antonyms)

Nouns	Noise				Synonyms	Antonyms
اسبيرين	مدّة	بشكل	مثل	وخاصة	أكل = طعم	سرع X بطئ
إمرأة	لمدّة	وبشكل	شعور	لون	ألم = وجع = عسر	سرع X درج
بنكرياس	ولمدّة	شكل	وشعور		أنث =إمرأة = نساء	سمن X نحف

فايروس	غازات	سو	سونار	ستيرويد	ردبة	حرقف	جير	ثدي	تحت
توجد	ويوجد	وموجودة	وموجود	موجودة	موجود	يوجد	ولفترة	لفترة	فترة
ويتناول	يتناول	وتناول	تناول	والكشف	يكشف	الكشف	كشف	الشكل	وشكل
منطقة	المنطقة	الجهات	جهات	والجهة	الجهة	جهة	حاجة	ويشعر	يشعر
فجأ = سرع	غمق = دكن	صبح = صار	خمر = كحول	كلّ = تامّ = جميع	زال = ذهب = خفي	عند = أثناء = خلال	شدّ = قسا = قوا = فرط = كبر	وصل = مرّ = دام = بقي	مباشر = حال = حالا
X جاع شبع	بعد X قبل	بدا X نهي	X مسك سهل	غمق X فتح	وسع X ضاق	كمل X نقص	X كسل نشط	كثر X قلّ	X سهل صعب

ويس	هيموغلوبين	مالوري	لولب	كرون	كحول	كارنت	قولون
مرض	مريض	المريض	المرض	جنس	بوجود	وجود	وتوجد
مثلاً	العلامات	العلامة	علامات	علامة	ومادّة	مادّة	بتناول
خاصة	يكون	أحد	ودرجة	درجة	بالمنطقة	بمنطقة	ومنطقة
		نحف= ضعف	خبث = خطر	رجل = ذكر	مسك = قبض	جفّ = إستسقاء	خرج = برز
		يسر X يمن	X زاد قلّ	X خفي ظهر	X خفّ ثقل	X خفّ قوا	X ذكر أنث

طريقة الجذر – الجذع في نظام المحلل العام للغة العربية

عبير خالد المشهداني عبدالودود خالد المشهداني وليد خالد المشهداني

aabeeeeraa@yahoo.com , Wadod_73@yahoo.com . Weeed9691@yahoo.com

الخلاصة

تتطلب العديد من البرامج التطبيقية المعرفة من المستخدم. في مثل هذه البرامج، قد تحدث عدد من المشاكل. المشاكل ذات الصلة مع فهم اللغة الطبيعية مثل؛ الأخطاء المطبعية، والتكرار ، والتشابه، والمعرفة غير المتناسقة. لذا، فإن تحليل النص مهم جدا في مثل هذه البرامج. تم انجاز عمل سابق لحل مشاكل فهم اللغة الطبيعية. ذلك العمل بنى محلل عربي ودمجه مع نظام قشرة هجين. هذا العمل يحاول بناء نظام محلل عربي عام، بحيث يمكن دمجه مع برامج أخرى لحل مشاكل فهم اللغة الطبيعية. هذا العمل يحاول تصحيح محلل العربية السابق، لذلك سوف يكون له قاموس أصغر. الصرف في هذا العمل يدمج المستندة إلى الجذر مع النهج القائم على الجذعية. هذا الدمج يقوي هذا العمل لأنه ربح مزايا النهجين. النهج القائم على الجذر يعطي ميزة الحجم الأصغر. النهج القائم على الجذعية يعطي ميزة اجتياز مشكلة الحالات غير النظامية. تم مقارنة هذا العمل مع اثنين من التقنيات الأخرى المقيمة سابقا. نتائج المقارنة تبين أن هذا العمل لديه الكثير من المزايا على التقنيات السابقة.