Journal of Theoretical and Applied Information Technology
15th December 2020. Vol.98. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING ARABIC QUESTION ANSWERING SYSTEM BY MERGING ANER TECHNIQUE, UPDATED QUESTION CLASSIFICATION TECHNIQUE AND STOP WORDS TECHNIQUE

BELAL ABUATA[1], IYAD ALAZZAM[2] & FADI ALKHAZALEH[3]
[1]Department of Information Technology, Yarmouk University, Irbid, Jordan
[2]Department of Information Systems, Yarmouk University, Irbid, Jordan
[3]Department of Computer Science, Yarmouk University, Irbid, Jordan
E-mail: [1]belalabuata@yu.edu.jo*, [2]eyadh@yu.edu.jo, [3]fadialkhazaleh@yahoo.com

ABSTRACT

Question answering systems are basically of three types: open domain, closed domain and restricted domain. In this paper we build a question answering system that tries to take users' questions, and provide them the answer in Arabic language. The proposed system is called AMAU (Arabic question answering system by merging Aner technique, updated question classification technique and stop words technique). AMAU focuses on question processing and document processing. After building the AMAU system, it was evaluated based on precision, recall and f-measure by ANER QA and ANER data set. The AMAU system is also compared with Arabic NER system using precision, recall and f-measure. The evaluation results show a better performance by AMAU. AMAU produced an average precision of up to 80% compared to Arabic NER system that produced an average precision of up to 63%. The proposed system performed best when stemming and stop words were used and resulted in the highest values of Recall, Precision and F-measure compared to NER system.

Keywords: *Information retrieval, Question Answering, Hybrid Approach*

1- INTRODUCTION

Because of the vast increase of information in internet, users need to obtain answers from internet, and they like them short and simple. Users also prefer the answers to be in their native languages without involving another language. Question answering (QA) system is the system which takes natural language questions that appear in natural language and it try to provide answer also in natural language. As for Arabic natural language processing, researches and especially in QA is still in its primary stages compared with other languages such as English language and European languages. This lack of research is mainly due to the complexity of Arabic language. Arabic language complexity main factors are [1]:

Arabic is highly inflectional and derivational.

The absence of diacritics (which represent most vowels) in the written text creates ambiguity.

The writing direction is from right-to-left and some of the characters change their shapes based on their location in the word.

Capitalization is not used in Arabic, which makes it hard to identify proper names,

Lack of Arabic corpora, lexicons, and machine readable dictionaries, which are essential to advance research in different areas.

In addition to the above factors, there is also a lack of Arabic corpora, lexicons, and machine-readable dictionaries, which are necessary to advance research in the field.

There are many proposed approaches for question classification such as ruled-based and learning-based approaches [2]. Ruled-based uses hand-crafted rules and needs to generate large number of rules. Ruled-based approaches perform very well on specific dataset but they don't perform the same with updated or new datasets. Unlike ruled-based approaches, learning-based approaches can handle the changes in data set, and can learn with a new data set [2] and hence we used the learning-based approaches to build the proposed Arabic question answering system.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

QA system includes three Phases. Each Phase is divided into three modules as shown in figure 1 [1].

There are many researches on these phases where researchers try to build their own QA system, and try to improve one of these phases. After that the researcher evaluates his QA system based on precision ,recall, f-measure, lenient, strict, MRR (mean reciprocal rank) and CWS (confidence weighted score). Each of these measures evaluates QA system based on specific criteria such as relevancy, correctness, conciseness, completeness, justification.

In this paper, we propose a QA hybrid system that integrates three methods that will help in getting better and faster answers for Arabic questions. The evaluation implemented and the comparisons carried in this paper can be used by future researches as a testing corpus for hybrid Arabic QA systems.

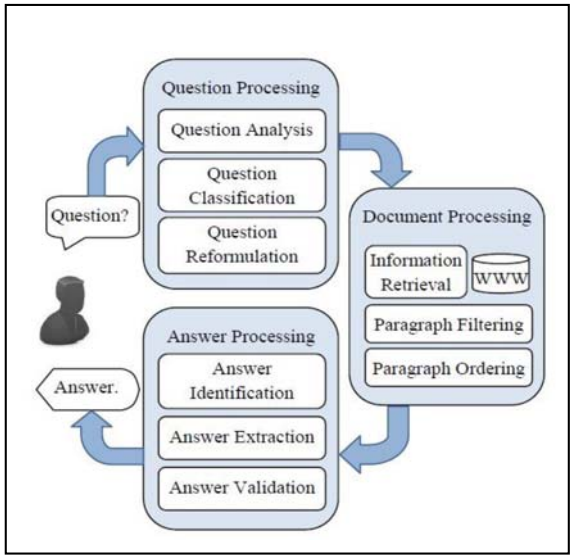

Figure 1. QA system Phases and modules.

2- RELATED WORK

It is evident from the escalating number of systems taking part in TREC-9 and the overall keenness for this research field that QA is the most improving framework for finding answers to natural language inquiries from a tremendous measure of literary information. Researchers in [3] pointed attention to that building "open-ended question answering systems that allow users to pose questions of any type and in any language, without domain restrictions, is still beyond the scope of any QA system today". Researchers in [4] demonstrated that advanced tools, (for example, such as dialog understanding and text mining) are fundamental for the accomplishment of future QA frameworks. Until the propelled devices are actualized, she proposed that we continue approximating the complication of QA with NLP upgrades of IR and IE strategies [4].

Examples of systems that combine IR and NLP processes are: QARAB [5], QA-LaSIE [6] and [7]. These types of systems perform linguistic analysis. They use IR system that deals with the question as a query and returned a set of ranked documents or paragraphs. While the NLP process parses the question and analyzes the returned documents or paragraphs or uses keywords to extract the answer.. A query having semantic representation is instantiated against semantic representation of analyzed document to find the answer.

Arabic NER system (ANERsys) [8] adopted the machine learning approach to build Arabic QA system, and it performs extraction for the user question by taking the user question and removes the stop words and performs question analysis to return the answer. The researchers created a corpus that contains large data set that includes 200 questions and keyword corpus to test their Arabic QA system.

Question classification for Arabic QA system is based on question classification technique that takes the user question, and extracts the pattern based on interrogative particle [2]. The pattern indicates the expected answer. Question classification technique determines a set of patterns for each interrogative particle. The result for Question classification for Arabic QAS is: The recall was (recall 0.93) by using 200 Arabic questions ,test questions 200 Arabic questions, and 186 questions out of 200 Questions.

Multi domain systems such as Start are obtainable on the internet (http://start.csail.mit.edu/) [9]. It answers user's query expressed in natural language. It offers important contribution in QA system area and document processing and IR. . IT answers questions in high precision, and it uses natural language annotations as a mechanism to match the questions to candidate answers. It uses natural language annotations [10] to decrease the gap between sentence level text analysis abilities and full complication of unrestricted natural language. Other systems proposed by [11] and [10] rely on collaborative question answering architecture on how to collect knowledge for QA system from different sources.

QA systems that rely on NLP techniques to detect answers from big collection of documents

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

process questions by merging syntactic and semantic information that distinguish the question [12]. Such systems defined eight heuristic rules for the extraction of keywords used to find the answer. They later developed another QA system known as FALCON [13] which uses WordNet for semantic processing of questions. It removed erroneous answers to obtain one answer by performing justification option.

Other systems such as [14], [15], [16], [17], [18] use different frameworks such as: questions Treebank, pattern learning from online data using seed questions and answer anchors, machine learning algorithms, deep linguistic analysis and surface pattern learning and POS taggers, named entity recognition, stop-word list to extract phrases.

IBM research developed a QA system called Watson [19]. Watson performs at human champion level in terms of precision, confidence and response speed. It processes the relevant content so deeply and precisely, and also justifies answers to user's natural language questions. Watson handles query in natural language analyze it, identify sources, finding and generating hypothesis, finding and scoring evidence and merging and ranking hypothesis.

The three techniques we merged in our system are briefly explained as in the following:

Aner Technique

A Named Entities Recognition (NER) is required as a tool for most of the QA system components. Those NER systems allow extracting proper nouns as well as temporal and numeric expressions from raw text. Usually every QA system will have a NER system for it. Paolo Rosso, et al. developed An Arabic NER system called Aner [20], [21]. Aner technique is composed of three stages:

Question processing**:** Question analysis model performs extraction for the user question by taking the user question and removes the stop words. In the proposed system, Aner takes the user question without stemming?

Document processing: Aner technique uses the Arabic-JIRS (*Information Retrieval (IR)*). IR system takes *Bag of words*, and performs elementary search to obtain the candidate documents. The candidate documents contain the passages that have relevance with the user's questions.

Answer processing: Answer processing model takes the passages that are relevance to the user's questions and extracts the answer from these passages based on keywords. Every interrogative particle has a set of keywords.

Question Classification Technique

Question classification technique takes the user question, and extracts the pattern based on interrogative particle. The pattern indicates the expected answer.

For example:

The question is: من هو محافظ المفرق والذي قال وجه؟ ان الاجهزة المعنية قامت بدورها على اكمل

"Who is the Governor of Mafraq, who said that the agencies concerned have done their best?"

The pattern is: محافظ المفرق والذي قال ان الاجهزة المعنية قامت بدورها على اكمل وجه

The Governor *of Al-Mafraq, who said that the concerned agencies have done their best*

The question classification technique determines a set of patterns for each interrogative particle. Examples of patterns are:

Who ,من, has one pattern , IPW NP WF ,
For example : من هو محافظ المفرق والذي قال .....؟
Ipw : من
Np: محافظ المفرق
Wf: the others والذي قال.....
The pattern, NP WF , indicates the expected answer . i.e. the nouns ,محافظ المفرق , indicate the expected answer.

How ,كم, has two patterns
Pattern1 IPW NP WF for example كم كتاب قرأت؟
Pattern2 IPW VP WF for example كم يبلغ ارتفاع جبل ...؟
The pattern, NP WF and VP WF , indicates the expected answer . i.e the noun and the verb , و كتاب يبلغ, indicate the expected answer.

Where , اين, has one pattern
Pattern1 IPW VP WF for example اين تقع الاردن...؟
The pattern, vp WF, indicates the expected answer, i.e. the verb ,تقع, indicate the expected answer.

when , متى, has one pattern
Pattern1 IPW VP WF for example متى تاسست الاردن...؟

The pattern, vp WF, indicates the expected answer, i.e. the verb, تاسست, indicate the expected answer.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In the proposed system, the question classification technique is updated as follows:

A new pattern is added to who ,من, because the verb can come after ,who, من. Hence,

Who, من, can have two patterns.

Pattern 1: IPW NP WF for example من هو محافظ المفرق والذي قال ...؟

Pattern 2: IPW VP WF for example من هو الذي قتل الامير قطز بعد المعركة؟

Bag of words will be changed based on interrogative particle.

If interrogative particle is (who ,من, or how,كم,) then the Bag of words will contains the (VP WF and NP WF) because the (VP WF and NP WF) indicate the expected answer.

If interrogative particle is (where, اين, or when, متى) , Bag of words will contain the (VP WF) because the (VP WF) indicate the expected answer.

For example:

انتصار المقاتلين بعد من هو الذي قتل الامير قطز بعد المعركة؟..

The interrogative particle is who, من, so Bag of words will contain the nouns and verbs and wf ,

هو الذي قتل الامير قطز بعد انتصار المقاتلين بعد المعركة .

because the nouns and verbs indicate the expected answer.

Stop words such as, بعد الذي هو, will be eliminated.

Suffixes and prefixes will be also eliminated.

As a result, the bag of words will be قتل امير قطز انتصار مقاتل معركة

Stopwords Technique

The stopwords don't indicate the expected answer, and it is not useful in retrieval, and removing the stopwords reduces the size of index. On the other hand, removing stop words may reduce the recall as shown in figure 2.

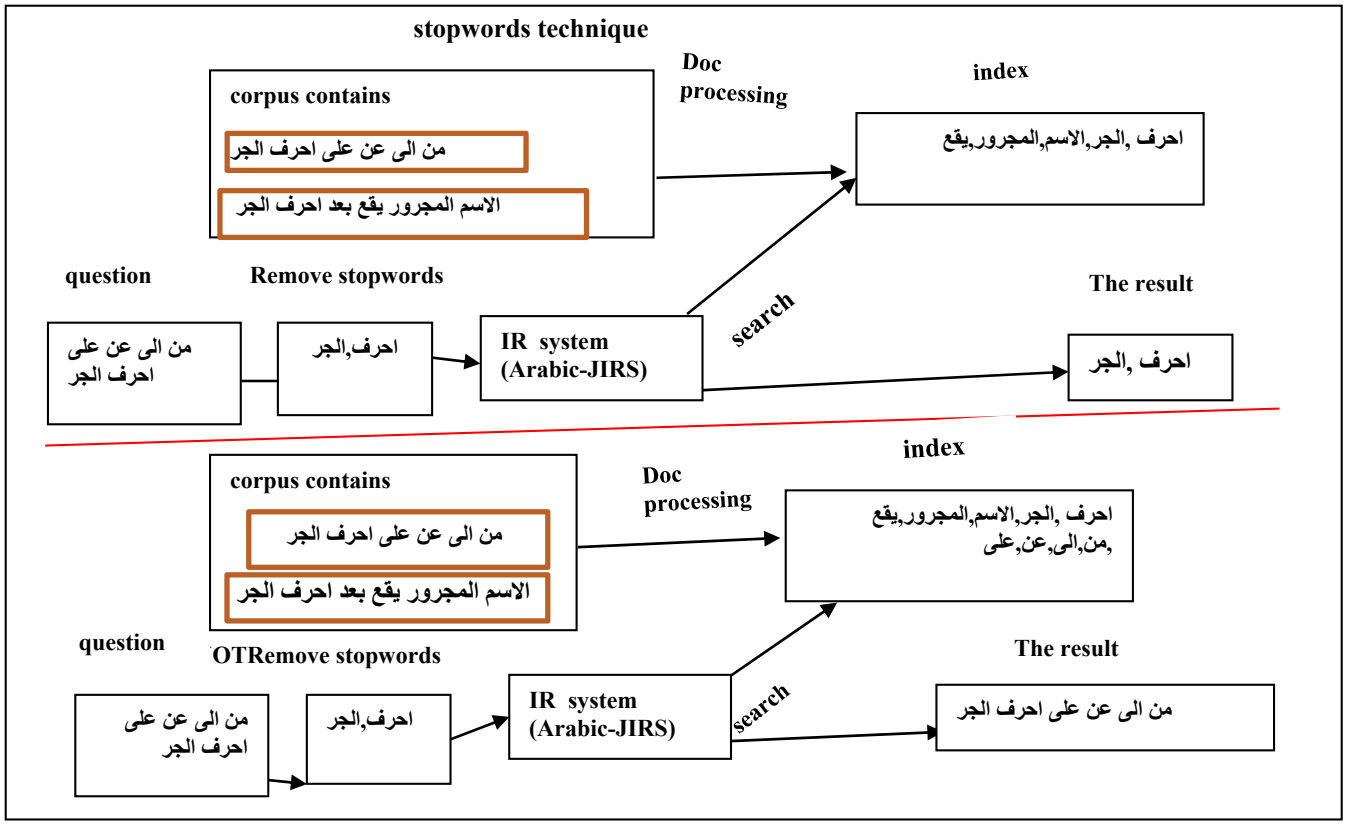

Figure 2: Recall reduction due to stop word removal.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3- METHODOLOGY

The proposed Arabic QA system AMAU is an open domain QA. AMAU is composed of three stages:

1) Question processing.

2) Document processing.

3) Answer processing.

The methodology used in building AMAU is explained based on the following example:

Suppose there are set of documents from tasweer news website (http://www.tasweernews.net/ تصوير نيوز), and the user wants to answer a question from these documents.

The Question is:

من هو محافظ المفرق والذي قال ان الاجهزة المعنية قامت بدورها على اكمل وجه اثناء الظروف الجوية الحالية بكل دقة وسرعة في الاستجابة لملاحظات المواطنين؟

Suppose that there are four files, and each file contains several paragraphs as shown in figure 3.

- **Question Processing Stage**

The question processing stage consists of six steps:

AMAU will take the user's question as token and stem the token using Khoja stemmer [22]. AMAU will then takes the token and removes the pronoun, beginnings, surplus, suffix and prefix and so on as shown in table 1.

In this step, AMAU defines the part of speech for each stem by using Stanford Log-linear Part-Of-Speech Tagger [23]. The output of this step is shown in table 2.

From step 2, AMAU will take the nouns and verbs, as bag of word called Bag of word 1 as shown in table 3.

AMAU takes the nouns and verbs and it doesn't take the stopwords because the nouns and verbs indicate the aon't indicate the answer.

Who, من, can have two patterns.

Pattern 1: IPW NP WF

Pattern 2: IPW VP WF

AMAU generates another bag of word called bag of words 2. It is similar to bag of words but without removing the pronoun, beginnings, surplus, suffix and prefix.

AMAU generates a third bag of word called bag of words 3. It is composed of the question words including the stopwords shown in table 5.

Document Processing stage

In this stage AMAU uses the Lucin IR system [5]. The IR the system searches in data set (sets of files), and retrieve the files that are closest to bag of word 1. This stage consists of the following seven steps:

AMAU sends Bag of words 1 to IR system IR takes Bag of words 1, and searches in the data set (sets of files). Then it will give the candidate files. The candidate files contain the paragraphs that are relevant to bag of word 1. Suppose that bag of words 1 is the one in table 3. The data set that contains the set of files (f1 f2 f3 f4) and the candidate file is f1. F1 content is shown in figure 3.

In this step, the file f1 includes two paragraphs and the system selects the paragraph which is closet to Bag of word 1, and contains more keywords. Each interrogative particle, such as who has its keywords. For example the interrogative particle, who, has the keywords such as الدكتور, الشيخ,المحامي,عبد

In our example the paragraph which is closet to Bag of word 1, and contains more keywords is p1 as shown in figure 4.

AMAU sends Bag of words 2 (shown in table 4) to the IR system and the IR searches in the data set. The system will then return the candidate files that contain the paragraphs that have relevance with bag of word 2. In our example the returned candidate file is f4 shown in figure 3.

File 4 includes two paragraphs, the system selects the paragraph which is closet to Bag of word 2, and contains keywords. Each interrogative particle, such as who, has its keywords, for example the interrogative particle, who, has the keyword such as..... عبد ,المحامي ,الشيخ ,الدكتور

In our example the paragraph which is closet to Bag of word 2, and contains keywords is p3 shown in figure 5.

5) AMAU offers Bag of words 3 to the IR system. The IR system will performs elementary search and returns the candidate files that contain the passages that have relevance with bag of word 3. In our example, the file f2 will be returned. F2 includes two paragraphs (P5 and P6), the system selects the paragraph which is closet to Bag of word 3, and contains keywords. Each interrogative

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

particle, such as who, has its keywords. For example the interrogative particle, who, has the keywords such as ..... عبد,المحامي,الشيخ, الدكتور

In our example the paragraph which is closet to Bag of word 3, and contains keywords is the p 6 shown in figure 6.

In our example we obtained p1 from bag of word 1 and p3 from bag of word 2. After that the system compares between these two paragraphs and selects the paragraph that has highest relevance with user's question based on cosine similarity measure. In our example the paragraph that has highest relevance with user's questions is p1 as shown in figure 7.

Then the result from step 6 (P1) is compared with the paragraph returned from bag of word 3 (P6). The system will then selects the paragraph that has highest relevance with user's question based on cosine similarity measure. The result will be paragraph P6 found in file f2.

Answer Processing Stage

The system uses interrogative particles, such as who, when, what ...., from questions to define the Query Type As it is shown in the table 6 [1], [5], [24].

Such as: who(من) person ⟶

Then the system uses the keywords in table 7 to extract the answer from the candidate paragraph p6 which was retrieved in step 7 of document processing stage. The keyword الدكتور (doctor) is used to mark an Arabic personal name [13]. Finally AMAU extracts the answer that is الدكتور أحمد عبد الله

The architecture of AMAU is shown in Figure 8. All the stages and steps explained previously are summarized in this architecture.

4- EVALUATION AND ANALYSIS OF RESULTS

After building the proposed AMAU system, it was evaluated based on precision, recall and f-measure. AMAU effectiveness was also compared Arabic NER system. The complete results, comparisons and their analysis are explained in the following paragraphs. The equations for the three measures used are shown in equations 1, 2, and 3.

Precision = (number of correct answers) / (number of questions answered) (1)

Recall = (number of correct answers) / (number of questions to be answered) (2).

F-measure = (2*(precision*recall)) / (precision + recall) (3).

The three used measures are most adequate and powerful measures used by QA and IR systems.

Four experiments were carried. Aner corpus was used for the evaluation purpose during all experiments as a corpus test [8]. These experiments varied according the use of stemming and stop words and they are defined as shown in table 8:

The recall values for the four experiments are shown in table 9. The values were calculated for four query types with different number of questions for each type as shown in table 9. AMAU2 performed better than AMAU1 and AMAU3 for query type person. This means that when stop words are not eliminated and no stemming used give better results compared to using or not using stemming and eliminating stop words, i.e. stop words elimination will decrease the recall values. As for the other three query types, non-elimination of stop words will give same results as when elimination of stop words is used except for the query type “date”, it will give better recall values. Also, using similarity between the three techniques will give better results for all query types.

AS for the precision values shown in table 10, using stemming will lead to less precision values in mostly all query types. However, the effect of stop words elimination depends on the query types as can be seen from the values for the query types “person” and “date” in AMAU1 and AMAU2 in table 10. For AMAU4, it always gives the best precision values compared to the other three techniques.

As for the f-measure results as shown in table 11, it follows the same trends in precision effects of stemming and stop words. Also, for AMAU4, it always gives the best f-measure values compared to the other three techniques.

Arabic NER system was evaluated using four query types as shown in table 12. Two of these types are common with AMAU evaluation. The comparison summary between the two systems is found in table 13.

AMAU performed better than Arabic NER based on the three evaluation measures used. It scored higher precision, recall and f-measure

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

values for the query type person. It also performed better for the query type location except for the precision measure as it slightly performed less.

AMAU Implementation and Corpus

The following tools were used in the implementation:

Java language

Part-Of-Speech *Tagger* (POS *Tagger*): Standford log –linear pos tager is used. It is developed by java at Stanford University [25].

Khoja Stemmer: It is one of the famous Arabic stemmer that is available to use in the web. It removes suffix, prefix and stop words [22].

Lucen, Information Retrieval (IR)): Lucen is an open source search engine written in java. It contains several classes to handle Arabic text, and takes set of words called bag of words, and searches in documents by creating index and rank for each document, and returns the closer documents to bag of words [25].

Cosine similarity measure: It is the most popular similarity measure that is used to find the similarity between two vectors when two documents are represented as term vector. The angle between two vectors is called cosine similarity and when the angle decreases, the similarity will increase.

ANER corpus was developed as part of the researchers extensive research on machine learning-based Arabic NER. It contains test sets for Arabic QA:

5- CONCLUSION

In this paper, we proposed an Arabic QA system called AMAU. This system works by merging between Aner technique, updated question classification technique and stopwords technique to improve Arabic question answering system. This Merging produced good results where the overall precision was 80%. AMAU has ambiguity in extracting the answer when there are many keywords in the same paragraph that is returned by IR lucen system:

For example:

Suppose the question is:

من هو محافظ المفرق والذي قال ان الاجهزة المعنية قامت بدورها على اكمل وجه اثناء الظروف الجوية الحالية بكل دقة وسرعة في الاستجابة لملاحظات المواطنين؟

and the IR system return this paragraph :

اخبر محافظ المفرق الدكتور أحمد عبد الله الزعبي رئيس صحيفة الدستور ان الاجهزة المعنية في جميع مناطق المحافظة قامت بدورها على اكمل وجه اثناء الظروف الجوية الحالية بكل دقة وسرعة في الاستجابة لملاحظات المواطنين.

Then AMAU will mark the keywords for who, من, and hence AMAU will mark the answer, محافظ رئيس, الدكتور. This will be analyzed and solved as future work.

REFERENCES

[1]. Khoufi, N., Aloulou, C., & H. Belguith, L. (2016). Parsing Arabic using induced probabilistic context free grammar, International Journal of Speech Technology, Int. J. Speech Technol. 19(2), 313–323.

[2]. Al-Chalabi, H., Ray, S., & Shaalan, K. (2015). Question Classification for Arabic Question Answering Systems. Proceedings of International Conference on Information and Communication Technology Research, 307-310.

[3]. Cardie, C., Ng, V., Pierce, D., & Buckley, C. (2000). Examining the role of statistical and linguistic knowledge sources in a general-knowledge question-answering system. Proceedings of the sixth conference on applied natural language processing, 180-187.

[4]. Harabagiu, S., Paşca, M., & Maiorano, S. (2000). Experiments with open-domain textual Question Answering. Proceedings of the 18th conference on Computational linguistics, 292-298.

[5]. Hammo, B., Abu-Salem, H., Lytinen, S., & Evens, M. (2002). QARAB: A Question Answering System to Support the Arabic Language. Workshop on Computational Approaches to Semitic Languages (ACL 2002), 55-65.

[6]. Gaizauskas, R., & Humphreys, K. (2000). A combined IR/NLP approach to question answering against large text collections. In Proceedings of the 6th Content-Based Multimedia Information Access Conference (RIAO-2000), 1288-1304

[7]. Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a Question Answering system. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 41-47.

[8]. Benajiba, R., & Lyhyaoui, A. (2006). Towards an Arabic Question Answering System. In Proceedings of 4th Conference on Scientific Research Outlook & Technology

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Development in the Arab world (SRO4), 11–14.

[9]. Katz, B., Borchardt, G., Felshin, S. (2006). Semantic an syntactic de-composition strategies for question answering from multiple resources. In proceedings of the AAAI workshop on inference for Textual Question Answering, 35-41.

[10]. Katz, B., Borchardt, G., Felshin, S. (2006). Natural Language Annotations for Question Answering. In Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, 303–306.

[11]. Katz, B., Lin, J., & Felshin, S. (2001). Gathering knowledge for a question answering system from heterogeneous information sources. Proceedings of the workshop on Human Language Technology and Knowledge Management, 1-8.

[12]. Moldovan, D., Harabagiu, S., Pasca, M., Mihalcea, R., Goodrum, R., Girju, R., & Rus, V. (1999). LASSO: A Tool for surfing the Answer Net. In Proceedings of the Eighth Text REtrieval Conference (TREC-8), 175-185.

[13]. Harabagiu, S., Moldovan, D., Pasca, M., Mihalcea, R., Surdeanu, M., Buneascu, R., Gîrju, R., Rus, V., & Morarescu, P. (2001). FALCON: Boosting Knowledge for Answer Engines. Proceedings of the 9th Text Retrieval Conference (TREC-9), 479--488.

[14]. Hermjakob, U. (2001). Parsing and question classification for question answering. Proceedings of the workshop on Open-domain question answering, 1-6.

[15]. Zhang, D., & Lee, W. (2003). A web-based question answering system. In Proceedings of the SMA Annual Symposium, 56-65.

[16]. Zhang, D., & Lee, W. (2003). Question classification using support vector machines. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 26-32.

[17]. Peng F., Weischedel, R., Licuanan, A., & Xu, J. (2005). Combining deep linguistic analysis and surface pattern learning: A hybrid approach to Chinese definitional question answering. In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language, (HLT/EMNLP), 307-314.

[18]. Stoyanchev, S., Chol Song, Y., & Lahti, W. (2008). Exact Phrases in Information Retrieval for Question Answering. Proceedings of the 2nd workshop on Information Retrieval for Question Answering (IR4QA), 9-16.

[19]. Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J., Nyberg, E., Prager, J., Schlaefer, N., & Welty, C. (2013). Building Watson: An overview of the DeepQA project. AI Magazine, 31(3):59–79.

[20]. Benajiba,Y., & Rosso, P. (2007). ANERsys 2.0: Conquering the NER task for the Arabic language by combining the maximum entropy with POS-tag information. In Proceedings of Workshop on Natural Language-Independent Engineering, 3rd Indian International Conference on Artificial Intelligence (IICAI-2007), 1814–1823.

[21]. Benajiba, Y., Rosso, P., & Ruiz, J. (2007). ANERsys: An Arabic named entity recognition system based on maximum entropy. In Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2007), 143–153.

[22]. Khoja, S., & Garside, R. (1999). Stemming Arabic text. Lancaster, UK, Computing Department, Lancaster University

[23]. Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In Conference of the North American Chapter of the Association for Computational Linguistics & Human Language Technologies (NAACL-HLT), 252-259.

[24]. Li, X., & Roth, D. (2002). Learning question classifiers. Proceedings of the 19th international conference on Computational linguistics, 1-7.

[25]. Ray, S., Shaalan, K. (2016). A review and future perspectives of Arabic question answering systems. IEEE Trans. Knowl. Data Eng. 28(12), 3169–3190.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

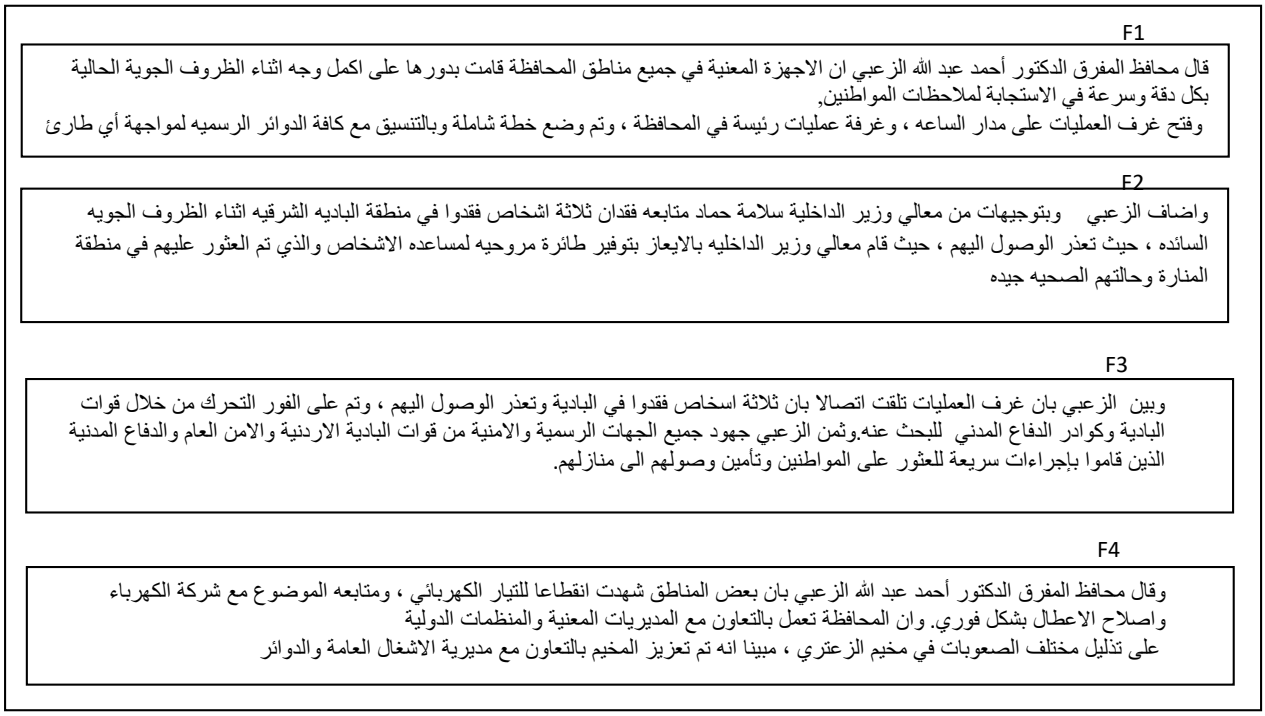

Figure 3: The files 1, 2, 3 and 4.

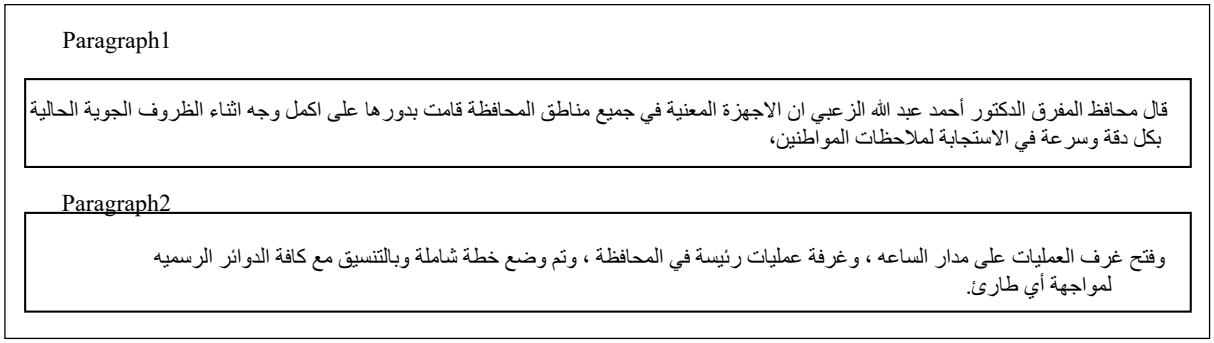

Figure 4: The File F1 Two Paragraphs.

وقال محافظ المفرق الدكتور أحمد عبد الله الزعبي بان بعض المناطق شهدت انقطاعا للتيار الكهربائي ، ومتابعه الموضوع مع شركة الكهرباء

وان المحافظة تعمل بالتعاون مع المديريات المعنية والمنظمات الدولية .واصلاح الاعطال بشكل فوري

Figure 5: F4, P3

قال محافظ المفرق الدكتور أحمد عبد الله الزعبي ان الاجهزة المعنية في جميع مناطق المحافظة قامت بدورها على اكمل وجه اثناء الظروف الجوية الحالية بكل دقة وسرعة في الاستجابة لملاحظات المواطنين الاردنيين.

.

Figure 6. F2, P6.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

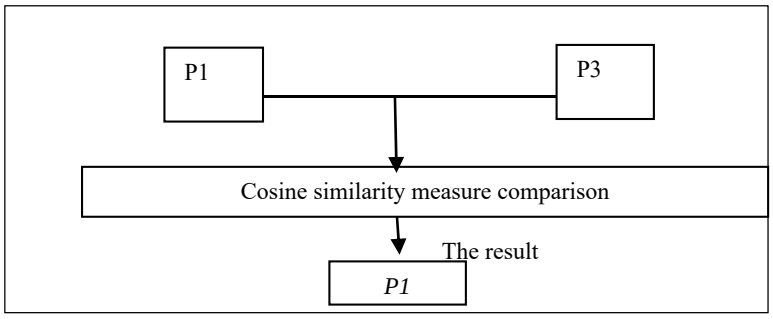

Figure 7: Comparison between bag of words 1 and 2

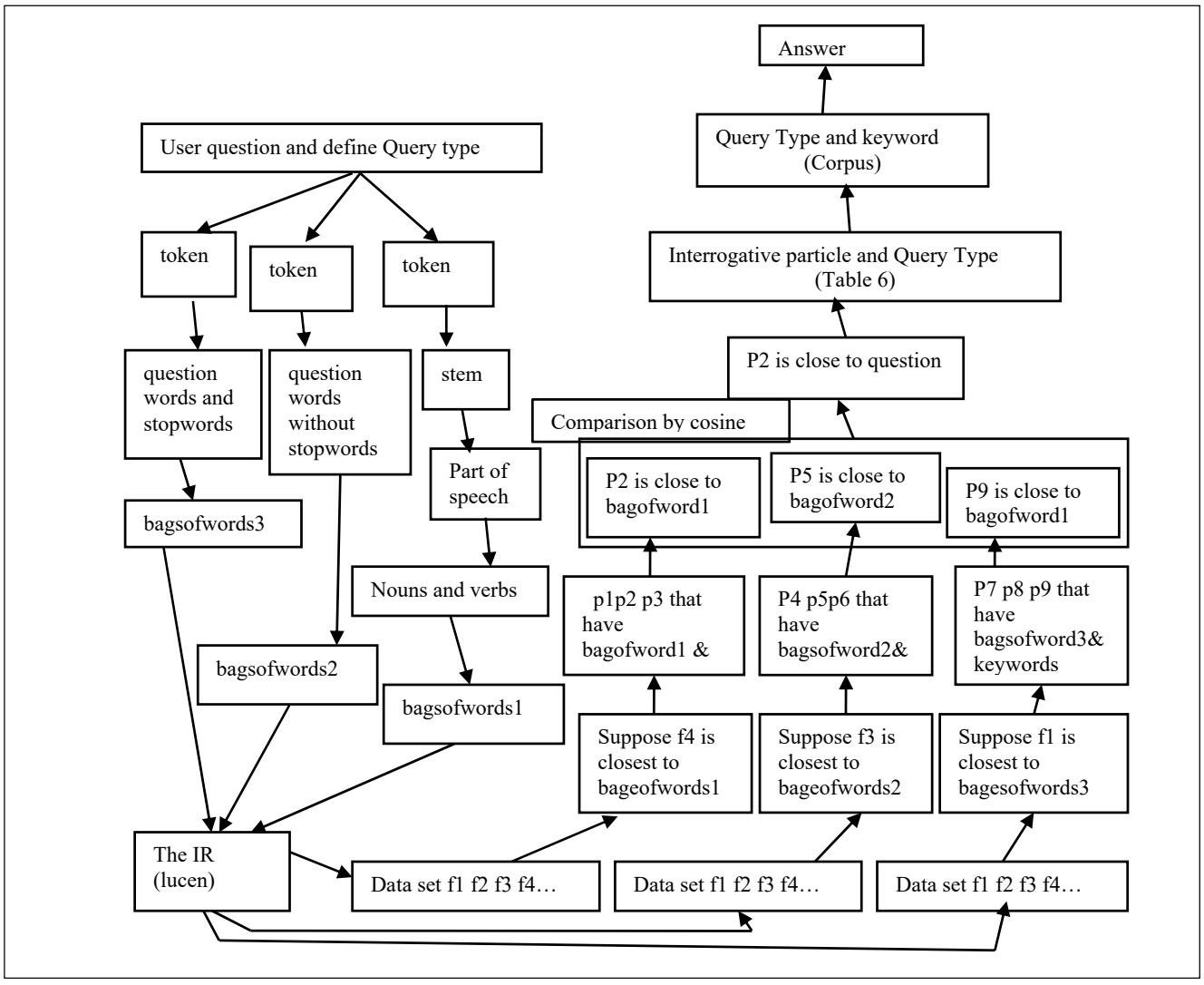

Figure 8: AMAU Architecture

Table 1: AMAU Removal of the Pronoun, Beginnings, Surplus, Suffix and Prefix

Toke	هو	محافظ	المفرق	و	الذي	قال	ان	الاجهزة	المعنية	قامت	بدورها	على	اكمل	وجه	اثناء	الظروف	الجوية	الحالية
Stem	هو	محافظ	مفرق	و	الذي	قال	ان	اجهزة	معنية	قام	دور	على	اكمل	وجه	اثناء	ظروف	جوية	حالية

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Part of Speech Out.

Token	هو	محافظ	المفرق	و	الذي	قال	ان	الاجهزة	المعنية	قامت	بدورها	على	اكمل
Stem	هو	محافظ	مفرق	و	الذي	قال	ان	اجهزة	معنية	قام	دور	على	اكمل
Part of	Pronoun	noun	noun	Conjunction	Pronoun	verb	Particle	noun	noun	verb	noun	Preposition	Adj
Bag of		yes	yes			yes		yes	yes	yes	yes		

Table 3 Bag of Words 1

محافظ	مفرق	قال	اجهزة	معنية	قام	دور	وجه	ظروف	جو	حال	دقة	سرعة	استجابة	ملاحظة	مواطن

Table 4: Bag of word2

محافظ	المفرق	قال	الاجهزة	المعنية	قامت	بدورها	اكمل	وجه	الظروف	الجوية	الحالية	دقة	سرعة	الاستجابة	لملاحظات	المواطنين

Table 5: Bag of words 3

هو	محافظ	المفرق	و	الذي	قال	ان	الاجهزة	المعنية	قامت	بدورها	على	اكمل	وجه	أثناء	الظروف	الجوية	الحالية	بكل

Table 6: Interrogative particles and corresponding query's type

Interrogative particle	Query type
من	person
متى	date
اين	place
كم	number

Table 7 Keywords that might identify personal names

Query type	Keywords
person	,المحامي ,الشيخ,الدكتور,.....
Date	,عام ,سنة,.....
place	الأردن,الامارات,المفرق.....
number	دينار,دولار,100,200.....

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 8: Experiments Definition

Exp. Num	Exp. Name	Exp. description
1	AMAU1	Without stemming, without stop words
2	AMAU2	Without stemming, with stop words
3	AMAU3	With stemming, without stop words
4	AMAU4	Using similarity between the three techniques and the question

Table 9: Recall results for the four experiments

No. of questions	AMAU4	AMAU3	AMAU2	AMAU1	
53	0.8	0.6	0.7	0.6	من
18	0.8	0.7	0.7	0.7	كم
17	0.9	0.7	0.8	0.8	متى
10	0.8	0.8	0.8	0.8	أين
	0.8	**0.7**	**0.7**	**0.7**	**overall**

Table 10: Precision results for the four experiments

No. of questions	AMAU4	AMAU3	AMAU2	AMAU1	
53	0.8	0.6	0.7	0.6	من
18	0.8	0.7	0.7	0.7	كم
17	0.9	0.7	0.8	0.9	متى
10	0.8	0.8	0.8	0.8	أين
	0.8	**0.7**	**0.7**	**0.7**	**overall**

Table 11: F-measure results for the four experiments

No. of questions	AMAU4	AMAU3	AMAU2	AMAU1	
53	0.8	0.6	0.7	0.6	من
18	0.8	0.7	0.7	0.7	كم
17	0.9	0.7	0.8	0.9	متى
10	0.8	0.8	0.8	0.8	أين
	0.8	**0.7**	**0.7**	**0.7**	**overall**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 12: Arabic NER System Recall, Precision and F-measure Results.

Query type	Precision	Recall	F-Measure
Location	0.82	0.78	0.80
Misc	0.61	0.32	0.43
Organization	0.45	0.31	0.37
Person	0.54	0.41	0.47
Overall	**0.63**	**0.49**	**0.55**

Table 13: Results comparison between AMAU4 and Arabic NER

	Arabic NER system			AMAU4 system		
	Precision	Recall	F-measure	Precision	Recall	F-measure
(person)	0.54	0.41	0.47	0.80	0.80	0.80
(locatio	0.82	0.78	0.80	0.80	0.80	0.80