

Three-level approach for Passage Retrieval in Arabic Question/Answering Systems

Lahsen Abouenour*, Karim Bouzoubaa* and Paolo Rosso**

* Mohammadia School of Engineers, Med Vth University, Rabat, Morocco.

Email : abouenour@yahoo.fr Email : karim.bouzoubaa@emi.ac.ma

** Natural Language Engineering Lab., ELiRF, Dpto. Sistemas Informáticos y Computación, Universidad Politécnica Valencia, Spain. Email : proso@dsic.upv.es

Abstract—This paper describes a three-level approach for improving the Passage Retrieval (PR) stage in the context of the Arabic Question/Answering (Q/A) task. This approach proposes to use a semantic reasoning on top of keyword-based and structure-based levels. Results of experiments conducted with a set of CLEF and TREC questions show an improvement of the Accuracy and the Mean Reciprocal Rank measures. An example shows also how the re-ranking based on the semantic score and Conceptual Graphs (CG) operations is even more relevant.

Keywords—Question/Answering; Query Expansion; Arabic WordNet; Semantic Reasoning;

I. INTRODUCTION

Having a great amount of available information, especially from Internet, today's user needs in terms of tools have changed and are becoming very challenging. The Natural Language Processing (NLP) domain has provided (and is still providing) various systems that contribute to overcome such needs. Search Engines (SE) and Information Retrieval (IR) systems remain among the most popular ones. Although they are currently widely used, SE and IR systems are no longer adequate for specific user expectations. Indeed, the large lists of snippets returned by SE are to be browsed by users. This is a time consuming task, in particular in the case of a specific search such as questions.

The goal of Question/Answering (Q/A) systems is to extract precise answers to user questions. Questions, as well as answers, are formulated in natural language.

Unlike languages such as English, Spanish, French or Italian [9], the Arabic Q/A task presents additional challenges to researchers in this field [4] [7]. This is mainly due to the particularities of the Arabic language (short vowels, absence of capital letters, complex morphology, etc.). However, the Q/A process stages are generally the same, namely: (i) question analysis, (ii) Passage Retrieval (PR) and (iii) answer extraction. Concretely, a user question expressed in natural language is to be analyzed. Many processes could be integrated at this step: question classification, Named Entity Recognizing, keyword extraction, etc. After that, the role of a PR is to retrieve and re-rank candidate passages that may contain the expected answer. The latter is to be

extracted and returned to the user as a final step of the Q/A process.

Relying on the sequence of tasks described above, an answer returned to the user could be irrelevant if candidate passages are too. Actually, if documents containing the right answer do not include keywords of the question they will not appear among candidate passages. In order to overcome such problem a Query Expansion (QE) module could be used. The module extends the list of keywords to be used at the PR stage to a larger list containing keywords related to the question ones. This process has been classically developed on the basis of morphological relations. We have proposed in [3] a more advanced QE processes using semantic relations between question keywords and document keywords. For example, the keyword اليابسة (land) has a hyponymy relation with the keyword الأرض (earth) which means that the former is a specialization of the latter. Therefore, if the keyword الأرض is among the user request, documents containing the word اليابسة could be also relevant for answering the question. However, the tasks related to the PR stage are even more challenging. Indeed, Q/A systems are particular in that the relevance required at the PR stage depends not only on the existence of the keywords related to the question but also on the existence of the related structure. For instance, the answer of the question “من الذي أسس مدينة مراكش؟” (Who did build the city of Marrakech?) could be expected in a passage whose structure is similar to “...أسس مدينة مراكش...” (... has built the city of Marrakech...). Therefore, the relevance of candidate passages should be higher for the passages that contain the whole or the longest part of the question structure.

We have integrated a structure-based process among our QE approach for the Arabic Q/A task [2]. The evaluation of this approach has showed that the candidate passages are re-ranked better with respect to the user question. This means that the first ranked passages are more likely containing the expected answer structure using the original keywords or the extended ones.

Yet even with our structure-based process, we do not guarantee a complete success of our Arabic Q/A

task. For example, while processing the question “ما هي عاصمة المغرب؟”, we have the two following passages: passage 1 “عاصمة المغرب العلمية هي مدينة فاس” and passage 2 “عاصمة المغرب تقع على الساحل الأطلسي و تسمى مدينة الرباط”. From the structure point of view passage 1 is ranked better than passage 2 because it contains a high density of question-related keywords. However, from the semantic point of view passage 2 is the one containing the expected answer. The system may better re-rank the returned passages if somewhere it integrates a semantic reasoning deciding that “عاصمة علمية” (cultural capital) is different from “عاصمة إدارية” (administrative capital).

In the present paper, we describe our three-level approach for PR in the context of Arabic Q/A. Hence, the next section of the paper is divided into four sub sections: Sub Section A presents a general overview of our approach. Sub Sections B, C and D are devoted to describe the keyword-based level, the structure-based level and the semantic-based level. Finally, we draw some conclusions and present future works.

II. THE THREE-LEVEL APPROACH

A. General overview

As we have mentioned previously, the PR process is a key issue of the Q/A system. We have seen also that the structure and the meaning of the user question are even important as well as its related keywords. Therefore, efforts aiming to reach an efficient retrieving and ranking of the candidate passages must focus on how to cover these problem sides.

Thus, we propose for Arabic Q/A systems a three-level approach for PR. Figure 1 shows its general overview:

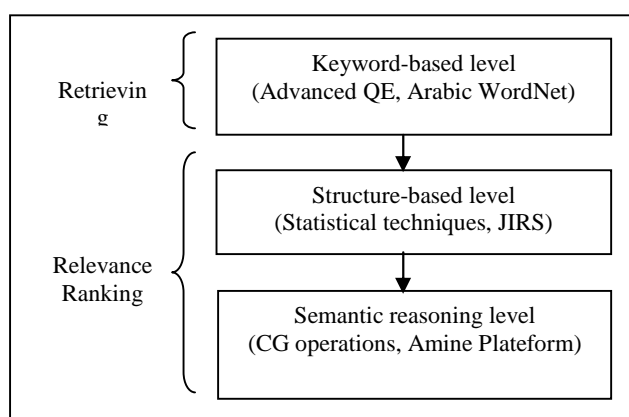


Figure 1. The three-level approach overview

Figure 1 illustrates that the first level role is to efficiently retrieve passages with respect to the question keywords and the extended ones as well according to semantic relations that exist between them.

A statistical method is used then in order to bring up the passages with the highest density of the considered keywords. The distance between keywords represents another criterion of ranking. Indeed, the passages are to be better ranked when they maintain similar distances between keywords as in the user question.

The final level is dedicated to the semantic reasoning. It uses as input the filtered passages from the second level and semantically re-ranks them. In this level we consider the representation in terms of Conceptual Graphs (CG) of the passages and the question.

In order to evaluate our levels added value deeply, we have carried out a few experiments. These experiments use a set of 82 CLEF¹ and 82 TREC² [22] questions manually translated into the Arabic language³. Actually, we have translated 400 TREC questions. However, since only 82 of the available translated CLEF questions are extensible using our QE process, and in order to make a fair comparison, we have considered the same number of TREC questions. They are classified into different domains (sport, geography, politic, etc.) and different types (questions seeking for time answers, persons, places, etc.).

The Yahoo and Google SE have been used at the PR stage. The answer of each question is then checked in the first five returned passages.

Two measures of the Q/A field have been considered:

- The Accuracy which is the average of the questions where we find the right answer in the first snippet;
- The Mean Reciprocal Rank (MRR). The reciprocal rank of a query response is the multiplicative inverse of the rank of the correct answer: MRR is the average of the reciprocal ranks of results for a sample of queries⁴ [22].

In the next sections we give more details about the components of each level together with the results of its evaluation.

B. First level: keyword-based level

The first level of our approach is keyword-based. We have showed with examples [3] that advanced QE could improve the number of relevant passages retrieved in the context of Arabic Q/A systems. Our

¹ Cross Language Evaluation Forum, <http://www.clef-campaign.org>

² Text REtrieval Conference, <http://trec.nist.gov>

³ List of questions available at <http://www.dsic.upv.es/grupos/nle/downloads.html>

⁴ http://en.wikipedia.org/wiki/Mean_reciprocal_rank

advanced QE uses the Arabic WordNet semantic resource.

The Arabic WordNet⁵ ontology (AWN) [10] is among the most interesting Arabic resources available for Arabic NLP researchers. The AWN ontology is a free lexical resource for modern standard Arabic [18]. It is based on the design and contents of Princeton WordNet [11] and it can be mapped onto it as well as a number of other WordNets, enabling translation on the lexical level to and from dozens of other languages. AWN is also connected to SUMO (Supper Upper Merged Ontology) [15] [16]. The SUMO is an upper level ontology which provides definitions for general-purpose terms and acts as a foundation for more specific domain ontologies. It contains about 2000 concepts.

The advanced QE process tries to take benefit from the advantages provided by AWN, namely: its openness, its interesting structure as ontology, the research carried out using other WordNets, etc. The global QE process uses, in addition to the morphological QE, four of AWN semantic relations: synonymy, definition, hyponymy and hypernymy. Figure 2 illustrates this process.

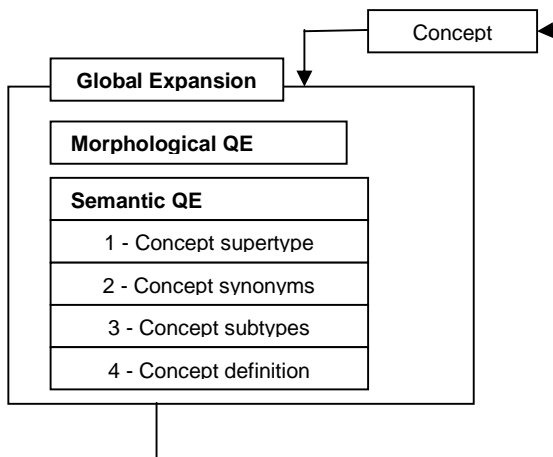


Figure 2: Expansion model to be integrated with a Q/A system

The usefulness of our QE process has been proofed by experiments. In these experiments different Google and Yahoo SE have been used and sets of CLEF and TREC questions have been tested. Table 1 shows the results.

TABLE 1. KEYWORD-BASED EVALUATION RESULTS

	Google (CLEF)		Yahoo (CLEF)		Yahoo (TREC)	
	no QE*	QE	no QE	QE	no QE	QE
Accuracy	29,26%	32,92%	1,22%	7,32%	5,02 %	6,95 %
MRR	10,15	11,25	0,99	3,25	2,04	2,88

* no QE: without using any QE process

As we can see, using different kind of questions and different SE, the accuracy and the MRR have both been improved when using our QE process. For instance, using QE with CLEF question we have obtained 7,32% (1,22% without QE) as accuracy and 3,25 (0,99 without QE) as MRR. The high performances of the preliminary experiments (using Google) is due to the fact the evaluation process was manual unlike the other experiments (using the Yahoo API which allows a larger number of requests per day than the Google one). For example, in the manual process we can identify answers composed of more than one word. For example, the answer of the question “كم تبلغ القيمة المادية لجائزة المغرب للكتاب؟” (What is the value of the Moroccan book award?) is “7000 دولار”. If a snippet contains for instance the expression “سبعة آلاف دولار” the answer is considered correct. More details of these experiments could be found in [1] [2].

C. Second level: structure-based level

The first level improves the accuracy and the MRR. The ranking of the passages is almost based on the number of question (or related) keywords they contain. This could be useful for simple IR needs. However, the context of Q/A is different. Indeed, at the answer extraction level, the structure of the question is to be considered.

To do so, we have chosen the Java Information Retrieval System (JIRS⁶) as PR module. JIRS is a language-independent PR system which has been already adapted to a few non-agglutinative European languages (such as English, Spanish and French) as well as to the Arabic language [5].

The re-ranking of the retrieved passages is based on a distance density n-gram model. In [5] authors explain the idea of this model which gives more weight to the passages where the most relevant question structures appear nearer to each other. In [12] some experiments were carried out to re-rank snippets obtained with Yahoo in order to return the most relevant ones containing the answer.

We have evaluated the second level using JIRS on the top of our QE and Yahoo. Table 2 shows the obtained results.

TABLE 2. STRUCTURE-BASED EVALUATION RESULTS

	Yahoo (CLEF)		Yahoo (TREC)	
	no QE	QE	no QE	QE
Accuracy	15,85%	19,51%	2,7 %	10,81 %
MRR	5,46	7,85	0,67	4,53

As we can see, with respect to the automatic evaluation the use of our QE based on AWN

⁵ <http://www.globalwordnet.org/AWN/>

⁶ <http://sourceforge.net/projects/jirs/>

together with the structure-based re-ranking using JIRS gives the best performances in terms of accuracy (19,51% and 10,81%) and MRR (7,85 and 4,53).

The obtained results are not the same for the CLEF and TREC questions. This may be due to the longer structure of the considered TREC questions (e.g. كم عدد الكروموسومات (التي تحنوي عليها الخلية البشرية الملقحة؟) compared to the CLEF questions ones (e.g. من بنى هرم خوفو؟). Nevertheless, the gain in terms of accuracy and MRR is highest for the set of TREC questions. JIRS may improve PR even more with questions whose structure is longer. The difference in performance between TREC and CLEF questions should be analyzed more in depth.

In [12] authors have showed that the performances provided by JIRS are more likely better when the number of processed passages overpass 1000. In our experiments, we have just considered the first 1000 passages. Therefore, the performances of the structure-based level may be even higher if the number of considered passages is greater than 1000.

So far, the accuracy and the MRR have been improved in the keyword-based level and even more in the structure-based level of our approach regardless the questions nature. Even though, the relevance ranking of the candidate passages should take into account not only the structure side but also the semantic one. Indeed, our aim is to ensure that the expected answer could be extracted efficiently as much as possible in the first passage.

In the next subsection we describe how the semantic reasoning level could allow us to reach this aim.

D. Third level: semantic-based level

The idea to use semantic reasoning process in NLP tasks is not new. In [8] author has adopted it for Machine Translation. Nowadays, various domains of NLP are concerned by semantic reasoning: Information Retrieval (IR), Word Sense Disambiguation [17], automatic acquisition of knowledge [21], etc.

The semantic reasoning to be used in our third level requires the choice of a good technical platform which allows such reasoning for the Arabic language [6]. We have adopted the Amine Platform. Amine⁷ is a Java open source multi-layer platform dedicated to the development of intelligent systems and multi-agents systems [13]. In addition to these characteristics, Amine has been chosen due to its features, namely:

- The choice of the right knowledge representation is one of the key elements of a semantic reasoning process. The Conceptual Graphs are among the formalisms which could be used for representing the meaning of a text formulated in natural language. A Conceptual Graph is a directed graph of nodes that correspond to concepts, connected by labeled and

oriented arcs that represent conceptual relations [19]. In Amine, the CG formalism is used for knowledge representation. The CG formalism has the advantage to be close to natural language and can be manipulated by computers.

- Amine is also a modular environment which provides: (i) an Ontology layer: we use this layer for manipulating the AWN ontology; (ii) an Algebraic layer: in addition to the elementary and the structured data types, this layer provides also various matching-based operations (like match, equal, unify, subsume, compare, maximalJoin, generalize, analogy, etc.); (iii) dynamic and basic ontology processes and (iv) Knowledge Base (KB) support.

The idea is to re-rank the filtered passages at the semantic level, i.e. try to find among them which one is expecting to represent the answer of the original question. In order to do so, we perform the steps showed in Figure 3:

- Since we consider only factoid questions at this level, the question is mapped onto an expected answer which in turn is translated into a corresponding CG: CG-EA (CG Expected Answer). For instance, for the question أين تقع أعلى نقطة على سطح الأرض؟ we replace "أين" with "في" and add it at the end. The expected answer is: « تقع أعلى نقطة على سطح الأرض في X ». For other simple types of factoid questions the following transformations are to be performed:
 - ؟ Q من هو → the expected answer is « X هو Q »
(Who is Q ? → Q is X)
 - ؟ Q ما هي → the expected answer is « X هي Q »
(What is Q ? → Q is X)
 - ؟ Q في أي → the expected answer is « X في Q »
(In which Q ? → Q in X)
 - ؟ Q ماذا → the expected answer is « Q »
(What did Q ? → Q)

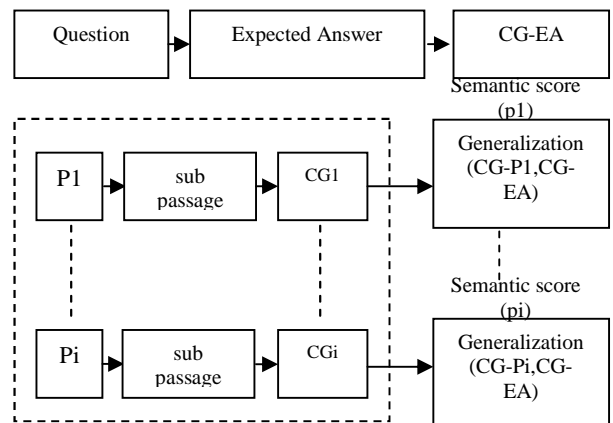


Figure 3. The semantic reasoning level

⁷ <http://amine-platform.sourceforge.net>

(ii) For all filtered passages:

- We start by selecting the text using delimiters and we translate that text to a corresponding CG-Pi (CG Passage i).
- We calculate the semantic score for the passage Pi by performing the generalization operation (provided in Amine Platform) between CG-EA and CG-Pi (the method is explained later).

(iii) We re-rank the passages by sorting their semantic scores.

This approach is similar to the one presented in [20]. In that approach, since no structure-based level is used, the number of passages to consider, and hence to translate into CG is very high. Fortunately, in our approach the use of JIRS allows to downsize the number of passages to be translated.

Let us show with an example how the described approach could improve the relevance of returned passages in the context of Arabic Q/A systems.

Let us consider the TREC question: أين تقع أعلى نقطة على الأرض؟ which means "Where is the highest point on the surface of the earth?". Using the Google search engine (which provides longer passages) we have the first six passages below:

TABLE 3. THE SNIPPETS RETURNED BY GOOGLE FOR THE CONSIDERED TREC QUESTION

ID	Rank	Passage
P1	1	بالإضافة إلى أعلى وأخفض نقطتين في الأرض، حيث قمة افرست في جبال الهملايا ترتفع 8848 متر عن سطح البحر وبذلك تكون أعلى نقطة في سطح الأرض، بينما أخفض نقطة على سطح الأرض تكون قرب البحر .. لا تقع جزر المحيط الهادي في أوقيانوسيا ضمن أي قارات
P2	2	كل نقطة على الأرض عدا القطبين تقع على تقاطع خط طول ومواز، وهكذا يمكن أن ننسب إليها ... المهم أن نتذكر أنه من أجل وصف نقطة على سطح الأرض نحتاج إلى عديدين، ولهذا ... على المستوي المماس للكرة في "أسفل" نقطة منها انطلاقاً من "أعلى" نقطة فيها. ...
P3	3	ولتمثيل ابعاد سطح الأرض على الخريطة، ترسم هذه الأبعاد بنسب خاصة تتناسب مع ... خفيف ثم يزداد الإنحدار مع الارتفاع حتى يصبح الإنحدار حاد عند أعلى نقطة ...
P4	4	هواء المريخ لا يتمتع بنفس كثافة هواء الأرض إذ يبلغ الضغط الجوي على سطح ... فهي تقع مقابل الشمس بالضبط في سماننا؛ إذ تظهر في الغروب، وتصل إلى أعلى نقطة لها ...
P5	5	إذا مثلنا الأرض بصورة كرة صغيرة قطرها 457 ميلي متر فيكون ارتفاع أعلى ... كان المفروض فيه أن الأرض تقع في مركز العالم والشمس (وغيرها) تدور حول الأرض. ... حركة الأرض حول الشمس باختلاف مواضعها على مدارها فسرعتها في نقطة الأوج هي 29.3 كم ... 5 ظهور الأنوار القطبية فيما يرتفع إلى ألف كيلومتر من سطح الأرض يدل على أن ...
P6	6	أعلى نقطة في اليابسة هي قمة جبل إفرست لأنها ترتفع 8846 متراً عن سطح البحر وتقع في جبال الهملايا، والواقع أن أطول جبل على سطح الأرض هو موناكي في هاواي، إذ ...

The ranking of the passages above is provided by Google. As we can see, the passage "p6" contains the answer which is جبل إفرست (the Everest mountain). Therefore, it is the most relevant passage in the context of a Q/A task. Let us see if the use of JIRS PR system could improve the re-ranking of returned passages.

Table 4 shows the new ranking of passages with the structure score between each passage and the question (this measure is calculated by JIRS).

TABLE 4. THE JIRS RE-RANKING RESULT

ID	Structure Score	Passage
P1	0,67	بالإضافة إلى أعلى وأخفض نقطتين في الأرض، حيث قمة افرست في جبال الهملايا ترتفع 8848 متر عن سطح البحر وبذلك تكون أعلى نقطة في سطح الأرض، بينما أخفض نقطة على سطح الأرض تكون قرب البحر .. لا تقع جزر المحيط الهادي في أوقيانوسيا ضمن أي قارات
P6	0,63	أعلى نقطة في اليابسة هي قمة جبل إفرست لأنها ترتفع 8846 متراً عن سطح البحر وتقع في جبال الهملايا، والواقع أن أطول جبل على سطح الأرض هو موناكي في هاواي، إذ ...
P2	0,54	كل نقطة على الأرض عدا القطبين تقع على تقاطع خط طول ومواز، وهكذا يمكن أن ننسب إليها .. المهم أن نتذكر أنه من أجل وصف نقطة على سطح الأرض نحتاج إلى عديدين، ولهذا ...
P4	0,49	هواء المريخ لا يتمتع بنفس كثافة هواء الأرض إذ يبلغ الضغط الجوي على سطح ... فهي تقع مقابل الشمس بالضبط في سماننا؛ إذ تظهر في الغروب، وتصل إلى أعلى نقطة لها ...
P3	0,39	ولتمثيل ابعاد سطح الأرض على الخريطة، ترسم هذه الأبعاد بنسب خاصة تتناسب مع ... خفيف ثم يزداد الإنحدار مع الارتفاع حتى يصبح الإنحدار حاد عند أعلى نقطة ...
P5	0,29	إذا مثلنا الأرض بصورة كرة صغيرة قطرها 457 ميلي متر فيكون ارتفاع أعلى ... كان المفروض فيه أن الأرض تقع في مركز العالم والشمس (وغيرها) تدور حول الأرض. ... حركة الأرض حول الشمس باختلاف مواضعها على مدارها فسرعتها في نقطة الأوج هي 29.3 كم ... 5 ظهور الأنوار القطبية فيما يرتفع إلى ألف كيلومتر من سطح الأرض يدل على أن ...

The passage which contains the expected answer is now in the second rank after using JIRS as PR system. The first passage contains also the answer but the second passage is the easiest to be used at the answer extraction stage. Our aim now is to investigate whether this passage will be ranked as the most relevant passage using the semantic reasoning

based on the CG representation.

Let us now represent the expected answer in term of CG. The following is the Linear Form of the considered passage:

CG-EA : [نقطة]-
 -attr->[أعلى],
 -ala->[الأرض],
 <-agnt-[مفهوم عام]-fi->[نقطة]

Note that Universal concept represent the answer (referred by X previously). Table 5 shows the CG representation of each (or part of) passage returned by JIRS.

TABLE 5. CG REPRESENTATION OF RETURNED PASSAGES

ID	Passage	CG
P2	المهم أن نتذكر أنه من أجل وصف نقطة على سطح الأرض نحتاج إلى عددين	[نحتاج]- -agnt->[عددين], -inOrder->[Proposition = = -وصف]- <-attr-[نقطة]-ala->[سطح الأرض]]
P3	ولتمثيل ابعاد سطح الأرض على الخريطة ، ترسم هذه الأبعاد بنسب خاصة تتناسب مع	[تمثيل]- -agnt->[أبعاد : X]<-attr-[سطح الأرض], [الخريطة]-ala->[ترسم]- -agnt->[أبعاد : X], -attr->[نسب]-attr->[خاصة]
P4	هواء المريخ لا يتمتع بنفس كثافة هواء الأرض إذ يبلغ الضغط الجوي على سطح	[هواء]- <-agnt-[يتمتع]- -bi->[كثافة : X]; <-attr-[Universal : [المريخ [هواء]- <-agnt-[يتمتع]- -bi->[كثافة : Y]; <-attr-[الأرض] [كثافة : X]<-differentOf-[كثافة : Y]
P5	ظهور الأنوار القطبية فيما يرتفع إلى ألف كيلومتر من سطح الأرض يدل على أن	[يدل]- -agnt->[Proposition = [ظهور]- -agnt->[الأنوار القطبية], -fi->[Proposition = [يرتفع]- -ila->[Integer : 1000]- attr->[كيلومتر], -min->[سطح الأرض]]

P6	أعلى نقطة في اليابسة هي قمة جبل افرست لأنها ترتفع 8846 مترا عن سطح البحر وتقع في جبال الهملايا	[نقطة]- -attr->[أعلى], -attr->[قمة]- <-attr-[جبل : افرست]; -ala->[اليابسة]
P1	بالإضافة إلى أعلى وأخفض نقطتين في الأرض، حيث قمة افرست في جبال الهملايا ترتفع 8848 متر عن سطح البحر وبذلك تكون أعلى نقطة في سطح الأرض	[ترتفع]- -agnt->[قمة]- <-attr-[افرست]- [الهمالايا]-in->[Integer = 8848] -aan->[البحر]

In order to use the CG operations supported by the Amine Platform, we have considered the part of AWN ontology that covers the concepts of the returned passages and the expected answer as well. The concepts which do not exist in AWN are added manually (for instance the concept type الأنوار القطبية is manually added). Figure 3 is an extract of the ontology.

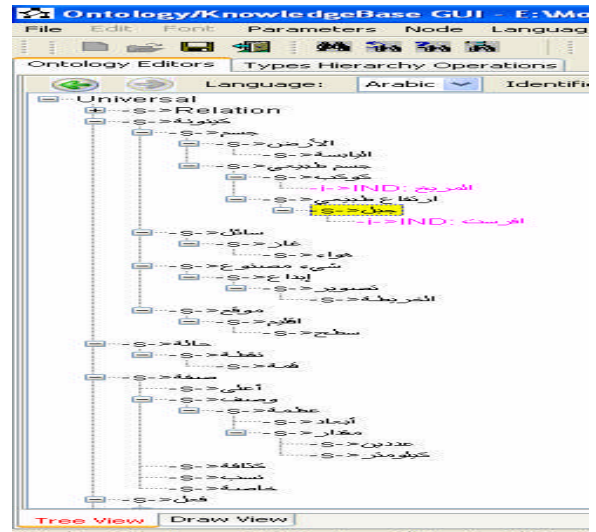


Figure 4. Extract of Amine AWN ontology

Using the Amine Platform we have calculated the generalization between the CG of the expected answer and the CG of each passage. Table 6 shows the result of this operation.

TABLE 6. RESULTS OF THE GENERALIZATION OPERATION

Passage	Generalization (CG-Q, CG-P)
1	[صفة]
2	[صفة]
3	[صفة]
4	[فعل]
5	- [نقطة #0] -attr->[أعلى], -ala->[الأرض]
6	[حدث]-[agnt]-[نقطة]

Two measures for semantic comparison could be considered: semantic score which is concept-based and relational similarity. In the current work we consider only the first measure. Indeed, relational similarity requires a relevant relation hierarchy in the considered ontology. The current release of AWN does not provide such pre-request.

For calculating the semantic score we use the following formula [14]:

$$\text{SemanticScore}(P) = \frac{\sum_{ci \in C} (\text{weight}(ci) * \beta(ci, \pi(ci)))}{\sum (\text{weight}(ci))}$$

Where C is the collection of the concepts of the passage P, weight(ci) is a weight assigned to the word related to the concept ci of the graph CG-P and $\beta(ci, \pi(ci))$ is the distance between ci and its projection in the generalization graph between CG-EA and CG-P, it is defined as follows:

$$\beta(ci, \pi(ci)) = 1 \text{ [if type}(ci) < \text{type}(\pi(ci))]$$

$$\beta(ci, \pi(ci)) = 1 - \min(\delta(\text{type}(ci), \text{type}(\pi(ci))), 5) / 5$$

$$\text{[if type}(ci) \geq \text{type}(\pi(ci))]$$

Where $\delta(\text{type}(ci), \text{type}(\pi(ci)))$ is the number of nodes between type(ci) and type($\pi(ci)$) in the ontology.

We use also the same weights proposed in [20] namely: 1 for verbs, 0.8 for nouns and 0.16 for adjectives and adverbs.

Let us calculate the semantic score for the passage p6: the CG of this passage contains the following concept types (weights are between parenthesis): أعلى (0,8), نقطة (0,16), قمة (0,8), جبل (0,8), اليابسة (0,8). Then, we identify the projection of each concept type in the generalization CG. For instance, $\pi(\text{اليابسة})$ is الأرض. In AWN the latter concept type has اليابسة as a direct subtype. Thus, in this case $\text{type}(c_i) \geq \text{type}(\pi(c_i))$ and $\delta(\text{type}(c_i), \text{type}(\pi(c_i)))=1$. As a result, $\beta(c_i, \pi(c_i)) = 1 - \min(1,5)/5 = 0,8$. Applying the same process to the other concept types, we get: Semantic Score(P6) = 0,66.

Doing the same for the five other passages, the semantic scores are:

- Semantic Score(P1) = 0,29
- Semantic Score(P2) = 0,04
- Semantic Score(P3) = 0,20
- Semantic Score(P4) = 0,19
- Semantic Score(P5) = 0,26

Table 7 below shows the new ranking with respect to the semantic score obtained.

TABLE 7. RESULTS OF THE SEMANTIC RE-RANKING

ID	Semantic Score	Passage
P6	0,66	أعلى نقطة في اليابسة هي قمة جبل إفرست لأنها ترتفع 8846 مترا عن سطح البحر وتقع في جبال الهملايا ، والواقع أن أطول جبل على سطح الأرض هو موناكي .. في هاواي ، إذ
P1	0,29	بالإضافة إلى أعلى وأخفض نقطتين في الأرض، حيث قمة إفرست في جبال الهملايا ترتفع 8848 متر عن سطح البحر وبذلك تكون أعلى نقطة في سطح الأرض، بينما أخفض نقطة على سطح الأرض تكون قرب البحر ... لا تقع جزر المحيط الهادي في أوقيانوسيا ضمن أي قارات
P5	0,26	إذا مثلنا الأرض بصورة كرة صغيرة قطر ها 457 ميلي متر فيكون ارتفاع أعلى ... كان المفروض فيه أن الأرض تقع في مركز العالم والشمس (وغيرها) تدور حول الأرض. ... حركة الأرض حول الشمس باختلاف مواضعها على مدارها فسرعتها في نقطة الأوج هي 29.3 كم ... 5 ظهور الأنوار القطبية فيما يرتفع إلى ألف كيلومتر من سطح الأرض يدل على أن ...
P3	0,20	ولتمثيل ابعاد سطح الأرض على الخريطة ، ترسم هذه الأبعاد بنسب خاصة تتناسب مع .. خفيف ثم يزداد الانحدار مع الارتفاع حتى يصبح الارتفاع حاد عند ... أعلى نقطة
P4	0,19	هواء المريخ لا يتمتع بنفس كثافة هواء الأرض إذ يبلغ الضغط الجوي على سطح ... فهي تقع مقابل الشمس بالضبط في سمائنا؛ إذ تظهر في الغروب، وتصل إلى .. أعلى نقطة لها
P2	0,04	كل نقطة على الأرض عدا القطبين تقع على تقاطع خط طول ومواز، وهكذا يمكن أن ننسب إليها .. المهم أن نتذكر أنه من أجل وصف نقطة على سطح الأرض .. نحتاج إلى عددين، ولهذا

As we can see, according to the semantic score measure the passage P6 is re-ranked as the most relevant. This passage is the one containing the expected answer. Moreover, the passage P1 which also contains a part of the expected answer is now ranked in the second position.

III. CONCLUSIONS

In this paper, we exposed and evaluated a three-level approach dedicated to the PR task in the context of Arabic Q/A systems. The semantic reasoning level integrated on top of the keyword-based and the structure-based one tries to improve even more the performances of the PR module. Conducted experiments show that our approach is adequate in order to overcome challenges related to the PR stage of Arabic Q/A systems.

As future works we keep on enhancing our approach by:

- Automating the semantic reasoning level module. This step concerns also the automatic extraction of CG representation from pattern structures to be identified from the passages. The use of a syntactic tool could be intended.
- Conducting corresponding experiments
- Covering all CLEF and TREC questions
- Integrating more enriched releases of Arabic WordNet. In this context, and from our experiments, we have noticed that:
 - Many concepts are not yet included in AWN. We had to manually include them in the ontology.
 - More efforts have to be considered to propose a more refined hierarchy for "Relation" types. That is why, we were unable to consider the "Relational similarity" in the calculation of the semantic score.

AKNOWLEDGMENTS

The research of the third author was made possible thanks to the AECID PCI B017961/08 and CICYT TIN2006-15265-C06 projects. We would like also to thank Dr. Adil KABBAB (INSEA, Rabat, Morocco) to help us with the use of CG.

REFERENCES

- [1] Abouenour L., Bouzoubaa K., Rosso P. (2009) Structure-based evaluation of an Arabic semantic Query Expansion using the JIRS Passage Retrieval system . In: Proc. Workshop on Computational Approaches to Semitic Languages, E-ACL-2009, Athens, Greece.
- [2] Abouenour L., Bouzoubaa K., Rosso P. (2008) Improving Q/A using Arabic WordNet. In: Proc. Int. Arab Conf. on Information Technology, ACIT-2008, Hammamet, Tunisia, December, 16-18.
- [3] Abouenour L., Bouzoubaa K., Rosso P. (2008) Système de Question/Réponse dans le cadre d'une plateforme intégrée: cas de l'Arabe. (in French) In: Proc. Rencontre Nationale en Informatique : Outils et Applications, RINO-2008, Errachidia, Morocco, June 5-7
- [4] Benajiba Y., Rosso P., Lyhyaoui A. (2007). Implementation of the ArabiQA Question Answering System's components, In: Proc. Workshop on Arabic Natural Language Processing, 2nd Information Communication Technologies Int. Symposium, ICTIS-2007, Fez, Morocco, April 3-5.
- [5] Benajiba Y., Rosso P., Gómez J.M. (2007). Adapting JIRS Passage Retrieval System to the Arabic. In: Proc. 8th Int. Conf. on Comput. Linguistics and Intelligent Text Processing, CILing-2007, Springer-Verlag, LNCS(4394), pp. 530-541.
- [6] Bouzouba K., Kabbaj A. (2007). An integrated development platform for Arabic language processing. ISCAL-07.
- [7] Brini W., Ellouze M., Hadrich Belguith L. (2009). QASAL : Un système de question-réponse dédié pour les questions factuelles en langue Arabe. In: 9ème Journées Scientifiques des Jeunes Chercheurs en Génie Electrique et Informatique, Tunisia. (To appear in March; in French).
- [8] Brown C., T. pattabhiraman, M. Boyer, D. Massam V. Dahi. (1986) Tailoring Conceptual Graphs for Use in NL Translation. Technial Report LCCR 86-14, Simon Fraser University.
- [9] Buscaldi D., Rosso P., Gómez J.M., Sanchis E. Answering Questions with an n-gram based Passage Retrieval Engine. In: Journal of Intelligent Information Systems (82) (in press) DOI: 10.1007/s10844-009-0082-y.
- [10] Elkateb, S., Black W., Vossen P., Farwell D., Rodríguez H., Pease A., Alkhalifa M. (2006). Arabic WordNet and the Challenges of Arabic. In Proceedings of Arabic NLP/MT Conference, London, U.K.
- [11] Fellbaum C. (2000). WordNet: An Electronic Lexical Database. MIT Press, cogsci.princeton.edu/~wn, September 7.
- [12] Gómez J. M., Rosso P., Sanchis E. (2007). Re-ranking of Yahoo snippets with the JIRS Passage Retrieval system. In: Proc. Workshop on Cross Lingual Information Access, CLIA-2007, 20th Int. Joint Conf. on Artificial Intelligence, IJCAI-07, Hyderabad, India, January 6-12.
- [13] Kabbaj A., (2006). Development of Intelligent Systems and Multi-Agents Systems with Amine Platform, in 15th Int. Conf. on Conceptual Structures, ICCS'2006, Springer-Verlag.
- [14] Montes-y Gómez, M., Gelbukh, A., López López, A., and Baeza-Yates, R. (2001). Flexible Comparison of Conceptual Graphs. In Proc. of the 12th International Conference on Database and Expert Systems Applications (DEXA 2001), volume N 2113 of LNCS, Munich, Germany. Springer-Verlag. <http://cseg.inaoep.mx/~mmontesg/papers/FlexibleCompariso nCG-dexa01.pdf>.
- [15] Niles I., Pease A. (2001). Towards a Standard Upper Ontology. In: Proceedings of FOIS 2001, Ogunquit, Maine, pp. 2-9.
- [16] Niles I., Pease A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In Proceedings of the 2003 International Conference on Information and Knowledge Engineering, Las Vegas, Nevada.
- [17] Rifat O. Y., Aslandogan A. (2004). Concept Based Information Access Using Ontologies and Latent Semantic Analysis. CSE Publications, available at <http://www.cse.uta.edu/research/publications/Downloads/CS E-2004-8.pdf>.
- [18] Rodríguez H., Farwell D., Farreres J., Bertran M., Alkhalifa M., Antonia Martí M., Black W., Elkateb S., Kirk J., Pease A., Vossen P., and Fellbaum C. (2008). Arabic WordNet: Current State and Future Extensions in: Proceedings of the Fourth International GlobalWordNet Conference - GWC 2008, Szeged, Hungary, January 22-25, 2008. <http://www.lsi.upc.edu/~nlp/papers/rodriguez08a.pdf>.
- [19] Sowa John F. (1984). Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley Company.
- [20] Stephane N. (2003). Sesei : un filtre sémantique pour les moteurs de recherche conventionnels par comparaison de structures de connaissance extraites depuis des textes en langage naturel. Mémoire présenté à la Faculté des études supérieures de l'Université Laval pour l'obtention du grade de maître ès sciences (M.Sc.), Septembre.
- [21] Svetlana H. (2004). Construction of Conceptual Graph Representation of Texts. In Proc. HLT-NAACL 2004: Student Research Workshop pp 49-54, Boston, Massachusetts, USA.
- [22] Voorhees E.M., The TREC-8 question answering track report. 1999. In Proceedings of the 8th Text Retrieval Conference, Gaithersburg, Maryland, USA, pp. 77-82.