ALQASIM: Arabic Language Question Answer Selection in Machines

Ahmed Magdy Ezzeldin, Mohamed Hamed Kholief, and Yasser El-Sonbaty

College of Computing and Information Technology, AASTMT
Alexandria, Egypt
a.magdy@a1works.com, {kholief,yasser}@aast.edu

Abstract. This paper presents "ALQASIM", a question answering system that focuses on answer selection and validation. Our experiments have been conducted in the framework of the main task of QA4MRE @ CLEF 2013. ALQASIM uses a novel technique by analyzing the reading test documents instead of the questions, which leads to a promising performance of 0.31 accuracy and 0.36 C@1, without using the test-set background collections.

Keywords: Question Answering, QA4MRE, Machine Reading Evaluation, Answer Selection, Answer Validation.

1 Introduction

"ALQASIM" is a Question Answer (QA) selection and validation system that aims at answering the multiple choice questions of QA4MRE @ CLEF 2013 test-set. It could also be used as a part of the answer validation module of any ordinary Arabic QA system. In the upcoming sections, the related works, system architecture, evaluation and discussion and the future work of ALQASIM are demonstrated.

2 Related Works

In CLEF 2012, Arabic QA4MRE was introduced for the first time. Two Arabic systems participated in this campaign. The first system, IDRAAQ [1], achieved a 0.13 accuracy and a 0.21 c@1. It used the Distance Density N-gram Model and semantic expansion using Arabic WordNet, and did not use the CLEF background collections.

The second system by Trigui et al. [6] achieved the accuracy and c@1 of 0.19 with their system. They used semantic expansion using inference rules on the background collection. They also determined the question focus and aligned the retrieved passages with the multiple answer choices of the question. However, these systems do not compare to the system created by Bhaskar et al. [2] for English QA4MRE @ CLEF 2012 that has an accuracy of 0.53 and c@1 of 0.65.

P. Forner et al. (Eds.): CLEF 2013, LNCS 8138, pp. 100–103, 2013.

3 ALQASIM Architecture

Most QA systems are composed of three main phases, which are: Question Analysis, Passage Retrieval and Answer Extraction. However, these systems are mainly targeted at searching for answers in a large collection of documents or on the Internet, which makes passage retrieval efficient [3]. QA4MRE is different in that aspect because the answer to a question is found in only one document, so there is not enough information redundancy to help the IR statistical approaches of passage retrieval. Thus, the ordinary QA pipeline is not the best approach to QA4MRE; the best approach is the one used by human beings in reading tests. A person would normally read and understand a document thoroughly, and then begins to tackle the questions. So, the suggested approach divides the QA4MRE process into three phases: (i) Document Analysis, (ii) Locating Questions & Answers, and (iii) Answer Selection.

3.1 Document Analysis

In the Document Analysis phase, the reading test documents are analyzed using MADA+TOKAN [3] morphological analyzer to stem each word in the documents and get its Part-of-Speech (PoS). Then, stop words are removed, and an inverted index of the remaining words stems is created, which contains the locations of each stem and its weight. Arabic WordNet (AWN) is then used to expand the words semantically by adding the synonyms of each word to the inverted index of that document. The weight of each word in the inverted index is assigned according to its PoS and repetition. So, nouns, verbs, adjectives, adverbs, proper nouns and the other parts of speech are assigned different weights. These weights mark word importance and are assigned according to our experiments with QA4MRE @ CLEF 2013 questions, by assigning the weights that yield the best results. Then the weight of a word is divided by its count in the document, thus, the more a word is repeated the less its weight will be. Thus, if the word is repeated many times in the target document, it is less likely to mark a question/answer snippet, because it appears in many sentences.

$$K_i = \frac{W_i \times S}{C_i} \tag{1}$$

K_i : the Weight of the word (i) saved in the inverted index
W_i : the weight of the word (i)
S : the synonym multiplier if the word is semantically expanded using AWN
C_i : the number of repetitions of the word (i) in the document

3.2 Locating Questions and Answers

In the second phase, every question and answer choice is handled as follows. Keywords are identified by stemming and removing stop words. The inverted index is then searched to find the best scoring three snippets locations for each question and

answer choice keywords. This score is calculated according to: (i) the number of keywords found within a distance threshold, (ii) the weights of all found keywords and (iii) the distance between these keywords. The impact of keywords count and weights is positive while the impact of distance is negative which means that snippets locations scores are penalized for higher distance among its keywords.

$$S_n = \left(\sum_{i=1}^{N} K_i \right) + N - \left(\sum_{i=2}^{N} d_i \right) \tag{2}$$

S_n : the score of snippet (n) which is found keywords for a question or answer choice.
N : the number of found keywords for the snippet
K_i : the weight of the keyword (i) as found in the inverted index. See equation 1.
d_i : the distance between the found keywords (i) and (i-1)

3.3 Answer Selection

By now, the question and its five answer choices have three scored snippets locations each. In this phase, answer choices snippets locations are scored by summing the scores of one question location and one answer choice location and subtracting the distance between them. The maximum of these scores is selected as the answer choice score. The best scoring answer choice is then selected as the question answer. If there is more than one best scoring answer choice, the question is marked as unanswered.

$$\text{Score}_{\,n}A_i = \text{QS}_n + A_iS_n - D_{QS_n A_i S_n} \tag{3}$$

$Score_nA_i$: the score of one question snippet with one answer choice snippet.
$D_{QSnAiSn}$: the distance between the question snippet and the answer choice location.

$$ScoreA_{\,i} = \max\left(Score_{\,n}A_i\right) \tag{4}$$

$ScoreA_i$: the maximum score of all Answer Choice (i) snippets.
$Score_nA_i$: the score of one question snippet with one answer choice snippet.

4 Evaluation and Discussion

The test-set, used by ALQASIM, is the set of questions and answers provided by CLEF 2011 [5] and translated to Arabic in 2012. ALQASIM uses Accuracy and C@1 [4] as evaluation metrics. It performs at an Accuracy of 0.31 and a C@1 of 0.36, which is considered promising, as it did not use any background collections.

Our system performs better than the other two Arabic QA4MRE systems from CLEF 2012 mainly because it analyses the reading test documents instead of the questions and answers. Documents have much more words than questions and answers, which gives context for morphological analyzers to produce more accurate analyses.

This explains why ALQASIM performs better on questions and answers with more keywords. On the other hand, many incorrectly answered questions are causative and list questions and questions that were incorrectly translated due to erroneous automatic translation. It is also noticed that sometimes the correct answer choice has fewer keywords than the other choices, which misleads the system into selecting an incorrect answer choice with more keywords, thus higher weight.

Table 1. Performance of ALQASIM and QA4MRE systems

	Accuracy	**C@1**
IDRAAQ [1]	0.13	0.21
Trigui et al. [6]	0.19	0.19
Bhaskar et al. [2]	0.53	0.65
ALQASIM	**0.31**	**0.36**

5 Conclusion and Future Work

This paper presents "ALQASIM" a Question Answer Selection and Validation system that can answer the multiple choice questions of QA4MRE @ CLEF 2013 test-set with an accuracy of 0.31 and a C@1 of 0.36. We are currently working on integrating Named Entity Recognition (NER), anaphora resolution, and temporal inference. We are also working on handling cause/effect relationship, and building an ontology from the background collections to expand questions and answers keywords.

References

1. Abouenour, L., Bouzoubaa, K., Rosso, P.: IDRAAQ: New Arabic Question Answering System Based on Query Expansion and Passage Retrieval. In: CLEF 2012 Workshop on Question Answering For Machine Reading Evaluation, QA4MRE (September 2012)
2. Bhaskar, P., Pakray, P., Banerjee, S., Banerjee, S., Bandyopadhyay, S., Gelbukh, A.: Question Answering System for QA4MRE@CLEF 2012. In: CLEF 2012 Workshop on Question Answering For Machine Reading Evaluation, QA4MRE (September 2012)
3. Habash, N., Rambow, O., Roth, R.: MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, PoS Tagging, Stemming and Lemmatization. In: Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, pp. 102–109 (2009)
4. Peñas, A., Rodrigo, A., del Rosal, J.: A simple Measure to Assess Non-response. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 1415–1424 (2011)
5. Peñas, A., Hovy, E., Forner, P., Rodrigo, Á., Sutcliffe, R., Forascu, C., Sporleder, C.: Overview of QA4MRE at CLEF 2011: Question answering for machine reading evaluation. Working Notes of CLEF (2011)
6. Trigui, O., Belguith, L.H., Rosso, P., Amor, H.B., Gafsaoui, B.: Arabic QA4MRE at CLEF 2012: Arabic Question Answering for Machine Reading Evaluation. In: CLEF 2012 Workshop on Question Answering For Machine Reading Evaluation, QA4MRE (September 2012)