Using an Arabic ontology to improve the Q/A task

Lahsen Abouenour
Mohammadia School of Engineers - Med V University - Agdal,
Rabat – Morocco, abouenour@yahoo.fr
Karim Bouzoubaa
Mohammadia School of Engineers - Med V University - Agdal,
Rabat – Morocco, karim.bouzoubaa@emi.ac.ma

Abstract

Users in an open environment such as the Web need advanced tools which make easier the access to the information. Question/Answering systems are new types of Information Retrieval (IR) systems. They reduce the effort to be done by users when searching for a precise answer to a question instead of a set of documents. In this paper, we describe our project of improving the Arabic Q/A task by using ontologies in the semantic Query Expansion (QE) approach that we propose. We present the steps that we followed in the preparation of our semantic resource and the experiments related to the evaluation stage as well.

Keywords: Question/Answering, Query Expansion, Ontology, Arabic WordNet, Semantics, Morphology

1. Introduction

The Web has grown larger and has become a main resource where users look for information or services. Therefore, the Natural Language Processing (NLP) domain is important since it provides many kinds of applications (Search Engines (SEs), Information Retrieval (IR) systems, etc.) which make the access to this information easier. One of the most advanced of these tools are Question/Answering systems (Q/A) which processes a user question in natural language and provides an answer, in natural language too, instead of listing a set of documents that the user has to browse.

The basic Q/A processing cycle is composed of three major stages [7]: (i) Processing the input question, (ii) Retrieving, with an IR system, the candidate documents (paragraphs) containing the answer, (iii) Processing each one of the candidate documents (paragraphs) in the same way as the question is processed and returning those sentences that may contain the answer.

The candidate documents retrieved in the second stage are important since the answer extraction depends on them. In the context of Q/A, processing a document is relevant by containing passages which have a structure (or sub structure) similar to the one of the user question. Query Expansion (QE) is one of the NLP techniques which can improve this stage by expanding the list of keywords used in the second stage.

Our project concentrates on an Arabic Q/A system. Currently only a few implementations of Arabic Q/A systems exist [7, 9, 3]. Classically, keywords are expanded using a morphological process [2]. For example, the keyword وظف can be expanded with its morphological nearly forms like توظيف - وظيفة – موظف. In this paper, we present a new approach where keywords are expanded using not only morphological but semantic relations as well and we show that this new approach improves results for searching documents from the web.

In Section 2, we present Amine Arabic WordNet (AAWN) as the semantic resource to be used at the expansion stage. In Section 3, we detail our QE cycle, and present, in Section 4, the evaluation steps. Finally, we conclude the paper in section 5.

2. Building the Amine AWN ontology

The use of semantic relations in a QE module has already been tested in similar approach [13] where authors are using of the concept of ontology in a Q/A system for the English language.

Building an Arabic ontology is not a simple task. For doing so, we need semantic resources. In comparison with other languages, not many are the NLP tools and resources in general (corpora, gazetteers, etc) which are available for Arabic [12]. This is especially true for semantic resources. Recently, this picture is about to change with the new release of Arabic WordNet (AWN).

AWN is a free lexical resource for modern standard Arabic [5]. It is based on the design and contents of Princeton WordNet (PWN) [6] and can be mapped onto PWN as well as a number of other WordNets, enabling translation on the lexical level to and from dozens of other languages. Moreover, the mapping of WordNet to the Suggested Upper Merged Ontology (SUMO) [10, 11] provides opportunities to use the semantic side in some Arabic NLP applications. To take advantage of using intelligent inferences later, we imported AWN ontology in Amine Platform which is a software for developing many kinds of intelligent systems [8]. We are calling this ontology Amine Arabic WordNet ontology (AAWN). Figure 1 shows a part of AAWN ontology.

Fig. 1: Part of Amine AWN ontology

AAWN is an ontology containing a hierarchy of concepts with their synonyms and definitions. For instance, the concept التحاق بتنظيم (joining an organization) has a subtype which is the concept عملية تنظيمية توظيف (hiring) and a supertype which is (Organizational Process). The concept توظيف has a synonym which is موظف and is defined in our ontology as "&%OrganizationalProcesses where someone is made an employee of an &%Organization". Note that the symbol &% shows that the concept exists in our ontology. Our semantic resource integrates around 10 000 synsets1 linked to 2000 concepts of the SUMO.

3. Semantic QE cycle

According to the structure of our ontology we can expand an original keyword to another one using the following semantic links: (i) Concept synonyms; (ii) Concept supertype; (iii) Conceptual structure definition; (iv) Concept subtypes. Hence, the algorithm used to expand keywords in general is as follows (figure 2): In addition to the first morphological expansion, we look for the supertypes of the current keyword which are semantically more general; we get, then, the equivalent keywords (synonyms); at the third step we get the list of more specialized keywords (subtypes), and we, finally, look for the concepts involved in the definition of the equivalent concept of the given keyword. This cycle is repeated recursively.

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

Figure 3 shows an example of the results reached when applying the above process with the original keyword منظمة (organization). In this figure boxes with labels 1, 2, 3 and 4 refer to the first, second, third and fourth semantic expansion. In [1], we present other examples and detail how these new reached keywords (... حركة، لجنة) are helpful for getting the right documents.

Fig.3: Semantic query expansion processing based on Amine AWN ontology

---

1 A given concept with all its synonyms is called a synset.

4. Evaluation steps

To confirm the effectiveness of our approach, we considered two kinds of experiments, one manual and a second one automatic. In the first one, we took a set of Arabic CLEF2 questions3 and the experiment consisted of looking for the correct answer (whitout using AAWN and using it) in the first five snippets returned by Google4. By using AAWN the accuracy has been improved from 29,26% to 32,92%, and the Mean Reciprocal Rank5 (MRR) has reached 11,25 against 10,15. The results show also that for 19,5% of the questions we success to get the expected answer in the first five snippets using our semantic expansion after failing to get it without any QE.

In the second experiment and in order not to require any manual checking of the snippets and to evaluate our approach with a larger number of CLEF questions, the Java Information Retrieval System (JIRS) in combination with the Yahoo search engine will be used instead of Google. JIRS would be used to re-rank the snippets provided by Yahoo. JIRS is a language-independent Passage Retrieval (PR) system which has been already adapted to a few non-agglutinative European languages (such as English and French) as well as to the Arabic language [4]. The re-ranking stage uses the Distance Density n-gram model. JIRS showed to be able to return the most probable snippets containing the answer. In this experiment step, our QE process will be combined with the JIRS process to investigate the improvement reached.

5. Conclusion

In this project we try to improve the stage of retrieving the candidate documents in an Arabic Q/A process. Our approach is to expand the list of keywords of the user question not only using morphological relations but semantic ones too. Our QE module uses Amine AWN which is an Arabic ontology containing a hierarchy of concepts with their synonyms and definitions as well. The preliminary experiments show that the accuracy and the MRR have been improved. In order to confirm these results we are looking for another experiment

---

² Cross Language Evaluation Forum, http://www.clef-campaign.org
³
http://www.dsic.upv.es/~ybenajiba/downloads.html
⁴ www.google.com
⁵ The Mean Reciprocal Rank (MRR): The reciprocal rank of a query response is the multiplicative inverse of the rank of the correct answer and MRR is the average of the reciprocal ranks of results for a sample of queries (http://en.wikipedia.org/wiki/Mean_reciprocal_rank)

based on the JIRS which is a PR dedicated to the Q/A task.

6. References

[1]   Abouenour L., Bouzoubaa K., Rosso P. 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, RINOA-2008, Errachidia, Morocco, June 5-7.

[2]   Abouenour L., El Hassani S., Yazidy T., Bouzoubaa K.,   Hamdani A. Building an Arabic Morphological Analyzer as part of an Open Arabic NLP Platform. In the Language Resources and Evaluation Conference (LREC), Marrakech, Morocco, 31st May, 2008.

[3]   Benajiba Y., Rosso P., Lyhyaoui A. "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, Morroco, April 3-5, 2007.

[4] Benajiba Y., Rosso P., Gómez J.M. "Adapting JIRS Passage Retrieval System to the Arabic". In: Proc. 8th Int. Conf. on Comput. Linguistics and Intelligent Text Processing, CICLing-2007, Springer-Verlag, LNCS(4394), pp. 530-541, 2007.

[5] Elkateb, S., Black W., Vossen P., Farwell D., Rodríguez H., Pease A., Alkhalifa M. "Arabic WordNet and the Challenges of Arabic". In proceedings of Arabic NLP/MT Conference, London, U.K, 2006.

[6] Fellbaum C. 1998. "WordNet: An Electronic Lexical Database". MIT Press, cogsci.princeton.edu/~wn, September 7, 2000.

[7] Hammou B., Abu-salem H., Lytinen S., Evens M. "QARAB: A Question answering system to support the ARABic language". In: Proc. of the workshop on Computational approaches to Semitic languages, ACL, pages 55-65, Philadelphia, 2002.

[8] Kabbaj A., "Development of Intelligent Systems and Multi-Agents Systems with Amine Platform", In 15th Int. Conf. on Conceptual Structures, ICCS'2006, Springer-Verlag, 2006.

[9] Mohammed F.A., Nasser K., Harb H.M., "A knowledge-based Arabic Question Answering

System (AQAS)". In: ACM SIGART Bulletin, pp. 21-33, 1993.

[10] Niles I., Pease A., "Towards a Standard Upper Ontology". In: Proceedings of FOIS 2001, Ogunquit, Maine, pp. 2-9, 2001.

[11] Niles I., Pease A., "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, 2003.

[12] Rosso P., Benajiba Y., Lyhyaoui A., "Towards an Arabic Question Answering system (in Arabic)". In: Proc. 4th Conf. on Scientific Research Outlook & Technology Development in the Arab world, SROIV, Damascus, Syria, 11-14 December, 2006.

[13] Yarmohammadi Mahsa A., Shamsfard M., Yarmohammadi Mahshid. A., Rouhizadeh M., "Using WordNet in Extracting the Final Answer from Retrieved Documents in a Question Answering System". In : Proc. The Fourth Global WordNet Conference, Szeged, Hungary, January 22-25, 2008.