A Lexical Approach for Spanish Question Answering

Alberto Téllez, Antonio Juárez, Gustavo Hernández, Claudia Denicia,
Esaú Villatoro, Manuel Montes, and Luis Villaseñor

Instituto Nacional de Astrofísica, Óptica y Electrónica
Laboratorio de Tecnologías del Lenguaje
Luis Enrrique Erro no. 1, Sta. María Tonantzintla, Pue., Mexico
{albertotellezv, antjug, ghernandez, cdenicia,
villatoroe, mmontesg, villasen}@inaoep.mx

Abstract. This paper discusses our system's results at the Spanish Question Answering task of CLEF 2007. Our system is centered in a full data-driven approach that combines information retrieval and machine learning techniques. It mainly relies on the use of lexical information and avoids any complex language processing procedure. Evaluation results indicate that this approach is very effective for answering definition questions from Wikipedia. In contrast, they also reveal that it is very difficult to respond factoid questions from this resource solely based on the use of lexical overlaps and redundancy.

1 Introduction

Question Answering (QA) has become a promising research field whose aim is to provide more natural access to information than traditional document retrieval techniques. In essence, a QA system is a kind of search engine that allows users to pose questions using natural language instead of an artificial query language, and that returns exact answers to the questions instead of a list of documents.

Current developments in QA tend to use a variety of linguistic resources to help in understanding the questions and the documents. The most common linguistic resources include: part-of-speech taggers, parsers, named entity extractors, dictionaries, and WordNet [1–3]. In contrast to these developments that point to knowledge rich methods, we have proposed a straightforward QA system that avoids using any kind of linguistic resource, and therefore, that can be easily adapted to different domains and languages. This system is supported by two simple ideas. First, that questions and answers are commonly expressed using the same set of words, and second, that different kind of questions requires different kind of methods for their treatment.

A complete description of the proposed system can be found in [4]. This paper, on the contrary, focuses on discussing the system's evaluation results at the QA task of CLEF 2007. In particular, it gives some insights on the usefulness of lexical information for QA and also on the appropriateness of our approach for dealing with semi-structured collections such as Wikipedia.

2 Our System at a Glance

Our QA system is based on a full data-driven approach that exclusively uses lexical information to determine relevant passages as well as candidate answers. The system is divided in two basic components.

The first component focuses on answering *definition questions*. It determines the target term by a regular expression analysis, then it retrieves the most relevant page from Wikipedia using a traditional information retrieval technique, and finally, it extracts the target definition from the first paragraph of the selected page.

The second component focuses on answering *factoid questions*. It applies a passage retrieval process in order to find relevant passages from the EFE collection and Wikipedia. After that, it determines a set of candidate answers by a regular expression analysis. Finally, it uses a machine-learning strategy (a Naïve Bayes classifier) to calculate the confidence value for each candidate answer. In this case, the answer having the highest value is selected as final answer.

On the other hand, our system also contemplates the treatment of *linked questions*, where the first question indicates the focus of the group and the rest are somehow dependent from it. This treatment is quite simple: it basically considers the enrichment of dependent questions by adding some keywords (and the answer) from the self-contained question.

It is important to mention that this system continues our previous year work [5], but incorporates some new elements. Mainly, it takes advantage of the structure of Wikipedia to easily locate definition phrases, and applies a technique for query expansion based on association rule mining to enhance the passage retrieval (refer to [4] for more details).

3 Evaluation Results

This section presents the experimental results corresponding to our participation in the monolingual Spanish QA track at CLEF 2007. This evaluation exercise considered two basic types of questions, definition and factoid. However, this year also were included some groups of linked questions (where the first one –the self-contained question– indicates the focus of the group and the rest of them –the linked questions– are somehow dependent on it).

From the given set of 200 test question, our QA system treated 34 as definition questions and 166 as factoid. Table 1 details our general accuracy results. It is very interesting to notice that our method for answering definition questions was very precise. It could answer almost 90% of the questions; moreover, it never supplied wrong or unsupported answers. In addition, given that all these questions were answered from Wikipedia, this result evidenced that our approach could effectively take advantage of its inherent structure.

On the other hand, Table 1 also shows that our method for answering factoid questions was not completely adequate (it only could answer 23% of this kind

Table 1. System's general evaluation

Questions	Right	Wrong	Inexact	Unsupported	Accuracy
Definition	30	–	4	–	0.88
Factoid	39	118	3	6	0.23
TOTAL	69	118	7	6	0.34

of questions). Taking into account that 82% of the factoid questions were answered from Wikipedia, we presumed that the poor performance was caused by the Wikipedia's structure. Two characteristics of Wikipedia damaged our system's behavior. First, it is much less redundant than general news collections; and second, its style and structure favor the presence of anaphoric and ellipsis phenomena, and thus make lexical contexts of candidate answers less significant than those extracted from other free-text collections.

In order to illustrate the last problem consider the question *"How old was Alfred Hitchcock when he died?"*. A correct answer for this question is located at the Wikipedia's document called *"Alfred Hitchcock"*, in the text fragment *"One year later, the April 29 of 1980, he died in his home located at The Angels when he was 80 years old ..."*. As can be noticed, the ellipsis in the text fragment (i.e., the ommision of the name Alfred Hitchcook) produces a poor lexical overlap between the question and the answer's context, and therefore, complicates the extraction of the given answer.

Finally, Table 2 shows some results from the treatment of groups of linked questions. It is clear that our approach was not useful for dealing with this kind of questions. The reason for this poor performance was that only 38% of the self-contained questions were correctly answered, and therefore, in the majority of the cases, the linked questions were enriched with erroneous information.

Table 2. Evaluation details about answering groups of linked questions

Questions	Right	Wrong	Inexact	Unsupported	Accuracy	NIL	
						Right	Wrong
Self-contained	64	95	6	5	0.38	3	35
Linked	5	23	1	1	0.17	0	5

4 Conclusions

This paper presented a QA system that allows answering factoid and definition questions. This system is based on a lexical approach. Its main idea is that questions and their answers are commonly expressed using almost the same set of words, and therefore, it simply uses lexical information to identify relevant passages as well as candidate answers.

The proposed method for answering definition questions is quite simple; nevertheless it allowed achieving very high precision rates. We consider that its success is mainly attributable to its capability to take advantage from the style and structure of Wikipedia (the used target document collection). On the contrary, our method for answering factoid questions was not equally successful. Paradoxically, the style and structure of Wikipedia, which favor the presence of anaphoric and ellipsis phenomena, caused a detriment in the lexical overlaps and in the answer redundancies, and consequently in the answer extraction process.

About the treatment of groups of linked questions, our conclusion is that the achieved poor performance (17%) was consequence of a cascade error. Only 38% of self-contained questions were correctly answered, and thus, most linked questions were expanded using incorrect information.

Acknowledgments

This work was done under partial support of CONACYT (project grant 43990 and scholarship 171610). We also thank the CLEF organizers.

References

1. de Pablo-Sánchez, C., González-Ledesma, A., Martínez-Fernández, J.L., Guirao, J.M., Martínez, P., Moreno-Sandoval, A.: MIRACLE's cross-lingual question answering experiments with Spanish as a target language. [6] 488–491
2. Ferrés, D., Kanaan, S., Ageno, A., González, E., Rodríguez, H., Turmo, J.: The TALP-QA system for Spanish at CLEF 2005. [6] 400–409
3. Roger, S., Ferrández, S., Ferrández, A., Peral, J., Llopis, F., Aguilar, A., Tomás, D.: AliQAn, Spanish QA system at CLEF-2005. [6] 457–466
4. Téllez-Valero, A., Montes-y-Gómez, M., Villaseñor-Pineda, L.: INAOE's participation at QA@CLEF 2007. In: Working notes for the 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary (September 2007)
5. Juárez-González, A., Téllez-Valero, A., Denicia-Carral, C., Montes-y-Gómez, M., Villaseñor-Pineda, L.: Using machine learning and text mining in question answering. In Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M., eds.: CLEF. Volume 4730 of Lecture Notes in Computer Science., Springer (2006) 415–423
6. Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., eds.: Accessing Multilingual Information Repositories, 6th Workshop of the Cross-Language Evalution Forum, CLEF 2005, Vienna, Austria, 21-23 September, 2005, Revised Selected Papers. In Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., eds.: CLEF. Volume 4022 of Lecture Notes in Computer Science., Springer (2006)