Experiments Adapting an Open-Domain Question Answering System to the Geographical Domain Using Scope-Based Resources

Daniel Ferrés and **Horacio Rodríguez**

TALP Research Center
Software Department
Universitat Politècnica de Catalunya
{dferres, horacio}@lsi.upc.edu

Abstract

This paper describes an approach to adapt an existing multilingual Open-Domain Question Answering (ODQA) system for factoid questions to a Restricted Domain, the Geographical Domain. The adaptation of this ODQA system involved the modification of some components of our system such as: Question Processing, Passage Retrieval and Answer Extraction. The new system uses external resources like GNS Gazetteer for Named Entity (NE) Classification and Wikipedia or Google in order to obtain relevant documents for this domain. The system focuses on a Geographical Scope: given a region, or country, and a language we can semi-automatically obtain multilingual geographical resources (e.g. gazetteers, trigger words, groups of place names, etc.) of this scope. The system has been trained and evaluated for Spanish in the scope of the Spanish Geography. The evaluation reveals that the use of scope-based Geographical resources is a good approach to deal with multilingual Geographical Domain Question Answering.

1 Introduction

Question Answering (QA) is the task of, given a query expressed in Natural Language (NL), retrieving its correct answer (a single item, a text snippet,...). QA has become a popular task in the NL Processing (NLP) research community in the framework of different international ODQA evaluation contests such as: Text Retrieval Conference (TREC) for English, Cross-Lingual Evalua-tion Forum (CLEF) for European languages, and NTCIR for Asian languages.

In this paper we describe our experiments in the adaptation and evaluation of an ODQA system to a Restricted Domain, the Geographical Domain.

GeoTALP-QA is a multilingual Geographical Domain Question Answering (GDQA) system. This Restricted Domain Question Answering (RDQA) system has been built over an existing ODQA system, TALP-QA, a multilingual ODQA system that processes both factoid and definition questions (see (Ferrés et al., 2005) and (Ferrés et al., 2004)). The system was evaluated for Spanish and English in the context of our participation in the conferences TREC and CLEF in 2005 and has been adapted to a multilingual GDQA system for factoid questions.

As pointed out in (Benamara, 2004), the Geographical Domain (GD) can be considered a middle way between real Restricted Domains and open ones because many open domain texts contain a high density of geographical terms.

Although the basic architecture of TALP-QA has remained unchanged, a set of QA components were redesigned and modified and we had to add some specific components for the GD to our QA system. The basic approach in TALP-QA consists of applying language-dependent processes on both question and passages for getting a language independent semantic representation, and then extracting a set of Semantic Constraints (SC) for each question. Then, an answer extraction algorithm extracts and ranks sentences that satisfy the SCs of the question. Finally, an answer selection module chooses the most appropriate answer.

We outline below the organization of the paper. In the next section we present some characteristics of RDQA systems. In Section 3, we present

the overall architecture of GeoTALP-QA and describe briefly its main components, focusing on those components that have been adapted from an ODQA to a GDQA. Then, the Scope-Based Resources needed for the experimentation and the experiments are presented in Sections 4 and 5. In section 6 we present the results obtained over a GD corpus. Finally, in Section 7 and 8 we describe our conclusions and the future work.

2 Restricted Domain QA Systems

RDQAs present some characteristics that prevent us from a direct use of ODQA systems. The most important differences are:

- Usually, for RDQA, the answers are searched in relatively small domain specific collections, so methods based on exploiting the redundancy of answers in several documents are not useful. Furthermore, a highly accurate Passage Retrieval module is required because frequently the answer occurs in a very small set of passages.

- RDQAs are frequently task-based. So, the repertory of question patterns is limited allowing a good accuracy in Question Processing with limited effort.

- User requirements regarding the quality of the answer tend to be higher in RDQA. As (Chung et al., 2004) pointed out, no answer is preferred to a wrong answer.

- In RDQA not only NEs but also domain specific terminology plays a central role. This fact usually implies that domain specific lexicons and gazetteers have to be used.

- In some cases, as in GD, many documents included in the collections are far to be standard NL texts but contain tables, lists, ill-formed sentences, etc. sometimes following a more or less defined structure. Thus, extraction systems based, as our, on the linguistic structure of the sentences have to be relaxed in some way to deal with this kind of texts.

More information about RDQA systems can be found in the ACL 2004 Workshop on QA in Restricted Domains[1] and the AAAI 2005 Worshop on Question Answering in Restricted Domains (Molla and Vicedo, 2005) .

---

[1] http://acl.ldc.upenn.edu/acl2004/qarestricteddomain/

3 System Description

GeoTALP-QA has been developed within the framework of ALIADO[2] project. The system architecture uses a common schema with three phases that are performed sequentially without feedback: Question Processing (QP), Passage Retrieval (PR) and Answer Extraction (AE). More details about this architecture can be found in (Ferrés et al., 2005) and (Ferrés et al., 2004).

Before describing these subsystems, we introduce some additional knowledge sources that have been added to our system for dealing with the geographic domain and some language-dependent NLP tools for English and Spanish. Our aim is to develop a language independent system (at least able to work with English and Spanish). Language dependent components are only included in the Question Pre-processing and Passage Pre-processing components, and can be easily substituted by components for other languages.

3.1 Additional Knowledge Sources

One of the most important task to deal with the problem of GDQA is to detect and classify NEs with its correct Geographical Subclass (see classes in Section 3.3). We use Geographical scope based Knowledge Bases (KB) to solve this problem. These KBs can be built using these resources:

- **GEOnet Names Server (GNS[3])**. A worldwide gazetteer, excluding the USA and Antarctica, with 5.3 million entries.

- **Geographic Names Information System (GNIS[4])**. A gazetteer with 2.0 million entries about geographic features of the USA.

- **Grammars for creating NE aliases**. Geographic NEs tend to occur in a great variety of forms. It is important to take this into account to avoid losing occurrences. A set of patterns for expanding have been created. (e.g. <toponym>_Mountains, <toponym>_Range, <toponym>_Chain).

- **Trigger Words Lexicon**. A lexicon containing trigger words (including multi-word terms) is used for allowing local disambiguation of ambiguous NE, both in the questions and in the retrieved passages.

---

[2] **ALIADO**. http://gps-tsc.upc.es/veu/aliado
[3] **GNS**. http://earth-info.nga.mil/gns/html
[4] **GNIS**. http://geonames.usgs.gov/geonames/stategaz

Working with geographical scopes avoids many ambiguity problems, but even in a scope these problems occur:

- **Referent ambiguity problem.** This problem occurs when the same name is used for several locations (of the same or different class). In a question, sometimes it is impossible to solve this ambiguity, and, in this case, we have to accept as correct all of the possible interpretations (or a superclass of them). Otherwise, a trigger phrase pattern can be used to resolve the ambiguity (e.g. "Madrid" is an ambiguous NE, but in the phrase, "comunidad de Madrid" (State of Madrid), ambiguity is solved). Given a scope, we automatically obtain the most common trigger phrase patterns of the scope from the GNS gazetteer.

- **Reference ambiguity problem.** This problem occurs when the same location can have more than one name (in Spanish texts this frequently occurs as many place names occur in languages other than Spanish, as Basque, Catalan or Galician). Our approach to solve this problem is to group together all the geographical names that refer to the same location. All the occurrences of the geographical NEs in both questions and passages are substituted by the identifier of the group they belong to.

We used the geographical knowledge available in the GNS gazetteer to obtain this geographical NEs groups. First, for each place name in the scope-based GNS gazetteer we obtained all the NEs that have the same feature designation code, latitude and longitude. For each group, we then selected an identifier choosing one of the NE included in it using the following heuristics: the information of the GNS field "native" tells if a place name is native, conventional, a variant, or, is not verified. So we decided the group representative assigning the following order of priorities to the names: native, conventional name, variant name, unverified name. If there is more than one place name in the group with the same name type we decide that the additional length gives more priority to be cluster representative. It is necessary to establish a set of priorities among the different place names of the group because in some retrieval engines

(e.g. web search engines) is not possible to do long queries.

3.2 Language-Dependent Processing Tools

A set of general purpose NLP tools are used for Spanish and English. The same tools are used for the linguistic processing of both the questions and the passages (see (Ferrés et al., 2005) and (Ferrés et al., 2004) for a more detailed description of these tools). The tools used for Spanish are:

- *FreeLing*, which performs tokenization, morphological analysis, POS tagging, lemmatization, and partial parsing.

- *ABIONET*, a NE Recognizer and Classifier (NERC) on basic categories.

- *EuroWordNet*, used to obtain a list of synsets, a list of hypernyms of each synset, and the Top Concept Ontology class.

The following tools are used to process English:

- *TnT*, a statistical POS tagger.

- *WordNet lemmatizer 2.0*.

- *ABIONET*.

- *WordNet 1.5*.

- *A modified version of the Collins parser*.

- *Alembic*, a NERC with MUC classes.

3.3 Question Processing

The main goal of this subsystem is to detect the Question Type (QT), the Expected Answer Type (EAT), the question logic predicates, and the question analysis. This information is needed for the other subsystems. We use a language-independent formalism to represent this information.

We apply the processes described above to the the question and passages to obtain the following information:

- Lexical and semantic information for each word: form, lemma, POS tag (Eagles or PTB tag-set), semantic class and subclass of NE, and a list of EWN synsets.

- Syntactic information: syntactic constituent structure of the sentence and the information of dependencies and other relations between these components.

Once this information is obtained we can find the information relevant to the following tasks:

- **Environment Building.** The semantic process starts with the extraction of the semantic relations that hold between the different components identified in the question text. These relations are organized into an ontology of about 100 semantic classes and 25 relations (mostly binary) between them. Both classes and relations are related by taxonomic links. The ontology tries to reflect what is needed for an appropriate representation of the semantic environment of the question (and the expected answer). A set of about 150 rules was built to perform this task. The ontology has been extended for the GD (see below the classes related with this domain).

```
ENTITY
    ENTITY_PROPER_PLACE
        GEOLOGICAL_REGION
            ARCHIPELAGO
            ISLAND
            LAND_FORM
                MOUNTAIN
            SEA_FORM
                CAPE
                GULF
                SEA
            WATER_FORM
                RIVER
        POLITICAL_REGION
            CITY
            CONTINENT
            COUNTY
            COUNTRY
            STATE
ENTITY_QUANTITY
        NUMERIC
MAGNITUDE
        AREA
        LENGTH
        FLOW
        WEIGHT
```

- **Question Classification.** Our ODQA system uses 25 QTs. For the GD we only used 10 Question Types (see Table 1). Only 5 QTs are common with the ODQA QTs, 5 QTs have been specially created for this domain.

Question Type	Expected Answer Type
Count_objects	NUMBER
How_many_people	NUMBER
What_area	MEASURE_AREA
What_flow	MEASURE_FLOW
What_height	MEASURE_HEIGHT
What_length	MEASURE_LENGTH
Where_action	LOCATION_SUBCLASS
Where_location	LOCATION_SUBCLASS
Where_quality	LOCATION_SUBCLASS
Default_class	LOCATION

Table 1: QTs and Expected Answer Types.

In order to determine the QT our system uses a Prolog DCG Parser. This parser uses the following features: word form, word position in the question, lemma and part-of-speech (POS). A set of DCG rules was manually configured in order to ensure a sufficient coverage.

The parser uses external information: geographical NE subclasses, trigger words for each Geographical subclass (e.g. "poblado" (*ville*)), semantically related words of each subclass (e.g. "water" related with *sea* and *river*), and introductory phrases for each Question Type (e.g. "which extension" is a phrase of the QT *What_area*).

- **Semantic Constraints Extraction.** Depending on the QT, a subset of useful items of the environment has to be selected in order to extract the answer. Accordingly, we define the set of relations (the semantic constraints) that are supposed to be found in the answer. These relations are classified as mandatory, (MC), (i.e. they have to be satisfied in the passage) or optional, (OC), (if satisfied the score of the answer is higher). In order to build the semantic constraints for each question, a set of rules has been manually built. A set of 88 rules is used. An example of the constraints extracted from an environment is shown in Table 2. This example shows the question type predicted, the initial predicates extracted from the question, the Environment predicates, the MCs and the OCs. MCs are *entity(4)* and *i_en_city(6)*. The first predicate refers to token number 4 ("autonomia" (*state*)) and the last predicate refers to token number 6 ("Barcelona").

Question	¿ A qué autonomía pertenece Barcelona? (At which state pertains Barcelona?)
Q. Type	*where_location*
Predicates	*city('Barcelona'),state(X), pertains('Barcelona',X)*
Environment	*action(5), participant_in_event(5,4), theme_of_event(5,6),prep(4,2),entity(4), i_en_proper_place(6),det(4,3),qu(3)*
Mandatory Constraints	*entity(4),i_en_city(6)*
Optional Constraints	*action(5),theme_of_event(5,6), participant_in_event(5,4),prep(4,2), type_of_location(5,5,i_en_state), property(5,5,pertenecer,3,6)*

Table 2: Question Analysis example.

3.4 Passage Retrieval

We use two different approaches for Passage Retrieval. The first one uses a pre-processed corpus as a document collection. The second one uses the web as document collection.

3.4.1 Off-line Corpus Retrieval

This approach uses a pre-processed and indexed corpus with Scope-related Geographical Information as a document collection for Passage Retrieval. The processed information was used for indexing the documents. Storing this information allows us to avoid the pre-processing step after retrieval. The Passage Retrieval algorithm used is the same of our ODQA system: a data-driven query relaxation technique with dynamic passages implemented using Lucene IR engine API (See (Ferrés et al., 2005) for more details).

3.4.2 Online Web Snippet Retrieval

The other approach uses a search-engine to get snippets with relevant information. We expect to get a high recall with few snippets. In our experiments, we chose Google as the search-engine using a boolean retrieval schema that takes advantage of its phrase search option and the Geographical KB to create queries that can retrieve highly relevant snippets. We try to maximize the number of relevant sentences with only one query per question.

The algorithm used to build the queries is simple. First, some expansion methods described below can be applied over the keywords. Then, stopwords (including normal stop-words and some trigger words) are removed. Finally, only the Nouns and Verbs are extracted from the keywords list. The expansion methods used are:

- **Trigger Words Joining (TWJ).** Uses the trigger words list and the trigger phrase pattern list (automatically generated from GNS) to join trigger phrases (e.g. "isla Conejera" o "Sierra de los Pirineos").

- **Trigger Words Expansion (TWE).** This expansion is applied to the NEs that were not detected as a trigger phrase. The expansion uses its location subclass to create a keyword with the pattern: *TRIGGER + NE* (e.g. "Conejera" is expanded to: ("isla Conejera" OR "Conejera")).

- **GNS Grouping Expansion (CE).** Noun Phrase expansion based on the groups generated from GNS Gazetteer.

- **Question-based Expansion (QBE).** This method appends keywords or expands the query depending on the question type. As an example, in the case of a question classified as *What_length*, trigger words and units associated to the question class like "*longitud*" (*length*) and "*kilómetros*" (*kilometers*) are appended to the query.

3.5 Answer Extraction

We used two systems for Answer Extraction: our ODQA system (adapted for the GD) and a frequency based system.

3.5.1 ODQA Extraction

The linguistic process of analyzing passages is similar to the process carried out on questions and leads to the construction of the environment of each sentence. Then, a set of extraction rules are applied following an iterative approach. In the first iteration all the MC have to be satisfied by at least one of the candidate sentences. Then, the iteration proceeds until a threshold is reached by relaxing the MC. The relaxation process of the set of semantic constraints is performed by means of structural or semantic relaxation rules, using the semantic ontology. The extraction process consists on the application of a set of extraction rules on the set of sentences that have satisfied the MC. The Knowledge Source used for this process is a set of extraction rules owning a credibility score. Each QT has its own subset of extraction rules that leads to the selection of the answer.

In order to select the answer from the set of candidates, the following scores are computed and accumulated for each candidate sentence: i) the rule score (which uses factors such as the confidence of the rule used, the relevance of the OC satisfied in the matching, and the similarity between NEs occurring in the candidate sentence and the question), ii) the passage score, iii) a semantic score (see (Ferrés et al., 2005)) , iv) the extraction rule relaxation level score. The answer to the question is the candidate with the best global score.

3.5.2 Frequency-Based Extraction

This extraction algorithm is quite simple. First, all snippets are pre-processed. Then, we make a ranked list of all the tokens satisfying the expected

answer type of the question. The score of each token in the snippets is computed using the following formula:

$$Score(tk_i) = \sum_{o \in Occurrence(tk_i)} \frac{1}{snippet\_rank(o)}$$

Finally, the top-ranked token is extracted.

4 Resources for Scope-Based Experiments

In this section we describe how we obtained the resources needed to do experiments in the Spanish Geography domain using Spanish. These resources were: the question corpus (validation and test), the document collection required by the offline ODQA Passage Retrieval, and the geographical scope-based resources. Finally, we describe the experiments performed.

4.1 Language and Scope based Geographical Question Corpus

We obtained a corpus of Geographical questions from Albayzin, a speech corpus (Diaz et al., 1998) that contains a geographical subcorpus with utterances of questions about the geography of Spain in Spanish. We obtained from Albayzin a set of 6887 question patterns. We analyzed this corpus and we extracted the following type of questions: Partial Direct, Partial Indirect, and Imperative Interrogative factoid questions with a simple level of difficulty (e.g. questions without nested questions). We selected a set of 2287 question patterns. As a question corpus we randomly selected a set of 177 question patterns from the previous selection (see Table 3). These patterns have been randomly instantiated with Geographical NEs of the Albayzin corpus. Then, we searched the answers in the Web and the Spanish Wikipedia (SW). The results of this process were: 123 questions with answer in the SW and the Web, 33 questions without answer in the SW but with answer using the Web, and finally, 21 questions without answer (due to the fact that some questions when instantiated cannot be answered (e.g. which sea bathes the coast of Madrid?)). We divided the 123 questions with answer in the SW in two sets: 61 questions for development (setting thresholds and other parameters) and 62 for test.

¿A qué comunidad autónoma pertenece el <PICO>?
At which state pertains <PEAK>?
¿Cuál es el capital de <COMUNIDAD>?
Which is the capital of <STATE>?
¿Cuál es la comunidad en la que desemboca el <RÍO>?
What is the state in which <RIVER> flows into?
¿Cuál es la extensión de <COMUNIDAD>?
Which is the extension of <STATE>?
Longitud del río <RÍO>.
Length of river <RIVER>.
¿Cuántos habitantes tiene la <COMUNIDAD>?
How many people does <STATE> has?

Table 3: Some question patterns from Albayzin.

4.2 Document Collection for ODQA Passage Retrieval

In order to test our ODQA Passage Retrieval system we need a document collection with sufficient geographical information to resolve the questions of Albayzin corpus. We used the filtered Spanish Wikipedia[5]. First, we obtained the original set of documents (26235 files). Then, we selected two sets of 120 documents about the Spanish geography domain and the non-Spanish geography domain. Using these sets we obtained a set of Topic Signatures (TS) (Lin and Hovy, 2000) for the Spanish geography domain and another set of TS for the non-Spanish geography domain. Then, we used these TS to filter the documents from Wikipedia, and we obtained a set of 8851 documents pertaining to the Spanish geography domain. These documents were pre-processed and indexed.

4.3 Geographical Scope-Based Resources

A Knowledge Base (KB) of Spanish Geography has been built using four resources:

- GNS: We obtained a set of 32222 non-ambiguous place names of Spain.

- Albayzin Gazetteer: a set of 758 places.

- A Grammar for creating NE aliases. We created patterns for the summit and state classes (the ones with more variety of forms), and we expanded this patterns using the entries of Albayzin.

- A lexicon of 462 trigger words.

We obtained a set of 7632 groups of place names using the grouping process over GNS. These groups contain a total of 17617 place

---

[5]**Spanish Wikipedia**. http://es.wikipedia.org

names, with an average of 2.51 place names per group. See in Figure 1 an example of a group where the canonical term appears underlined.

{*Cordillera Pirenaica*, *Pireneus*, *Pirineos*, *Pyrenaei Montes*, *Pyrénées*, *Pyrene*, *Pyrenees*}

Figure 1: Example of a group obtained from GNS.

In addition, a set of the most common trigger phrases in the domain has been obtained from the GNS gazetteer (see Table 4).

	Geographical Scope	
	Spain	UK
Top-ranked Trigger Phrases	*TRIGGER* de *NE* *TRIGGER NE* *TRIGGER* del *NE* *TRIGGER* de la *NE* *TRIGGER* de las *NE*	*NE TRIGGER* *TRIGGER NE* *TRIGGER* of *NE* *TRIGGER* a' *NE* *TRIGGER* na *NE*

Table 4: Sample of the top-ranked trigger phrases automatically obtained from GNS gazetteer for the geography of Spain and UK.

5   Experiments

We have designed some experiments in order to evaluate the accuracy of the GDQA system and its subsystems (QP, PR, and AE). For PR, we evaluated the web-based snippet retrieval using Google with some variants of expansions, versus our ODQA Passage Retrieval with the corpus of the SW. Then, the passages (or snippets) retrieved by the best PR approach were used by the two different Answer Extraction algorithms. The ODQA Answer Extractor has been evaluated taking into account the answers that have a supported context in the set of passages (or snippets). Finally, we evaluated the global results of the complete QA process with the different Answer Extractors: ODQA and Frequency-Based.

6   Results

This section evaluates the behavior of our GDQA system over a test corpus of 62 questions and reports the errors detected on the best run. We evaluated the three main components of our system and the global results.

- **Question Processing**. The Question Classification task has been manually evaluated. This subsystem has an accuracy of 96.77%.

- **Passage Retrieval**. The evaluation of this subsystem was performed using a set of correct answers (see Table 5). We computed the *answer accuracy*: it takes into account the number of questions that have a correct answer in its set of passages.

Retrieval Mode	Accuracy at N passages/snippets			
	N=10	N=20	N=50	N=100
Google	0.6612	0.6935	0.7903	0.8225
+TWJ	0.6612	0.6774	0.7419	0.7580
+TWJ+TWE	0.6612	0.6774	0.7419	0.7580
+CE	0.6612	0.6774	0.7741	0.8064
+QBE	0.8064	0.8387	0.9032	0.9354
+TWJ+QB+CE	0.7903	0.8064	0.8548	0.8870
Google+All	0.7903	0.8064	0.8548	0.8870
ODQA+Wiki	0.4354	0.4516	0.4677	0.5000

Table 5: Passage Retrieval results (refer to section 3.4.2 for detailed information of the different query expansion acronyms).

- **Answer Extraction**. The evaluation of the ODQA Answer Extractor subsystem is shown in Table 6. We evaluated the accuracy taking into account the number of correct and supported answers by the passages divided by the total number of questions that have a supported answer in its set of passages. This evaluation has been done using the results of the top-ranked retrieval configuration over the development set: the *Google+TWJ+QB+CE* configuration of the snippet retriever.

Accuracy at N Snippets		
N=10	N=20	N=50
0.2439 (10/41)	0.3255 (14/43)	0.3333 (16/48)

Table 6: Results of the ODQA Answer Extraction subsystem (accuracy).

In Table 7 are shown the global results of the two QA Answer Extractors used (ODQA and Frequency-Based). The passages retrieved by the *Google+TWJ+QB+CE* configuration of the snippet retriever were used.

	Accuracy	
Num. Snippets	ODQA	Freq-based
10	0.1774 (11/62)	0.5645 (35/62)
20	0.2580 (16/62)	0.5967 (37/62)
50	0.3387 (21/62)	0.6290 (39/62)

Table 7: QA results over the test set.

We analyzed the 23 questions that fail in our best run. The analysis detected that 10 questions had no answer in its set of passages. In 5 of these questions it is due to have a non common question or location. The other 5 questions have problems with ambiguous trigger words (e.g. *capital*) that confuse the web-search engine. On the other hand, 13 questions had the answer in its set of passages, but were incorrectly answered. The reasons are mainly due to the lack of passages with the answer (8), answer validation and spatial-reasoning (3), multilabel Geographical NERC (1), and more context in the snippets (1).

7   Evaluation and Conclusions

This paper summarizes our experiments adapting an ODQA to the GD and its evaluation in Spanish in the scope of the Spanish Geography. Out of 62 questions, our system provided the correct answer to 39 questions in the experiment with the best results.

Our Passage Retrieval for ODQA offers less attractive results when using the SW corpus. The problem of using SW to extract the answers is that it gives few documents with the correct answer, and, it is difficult to extract the answer because the documents contain tables, lists, ill-formed sentences, etc. Our ODQA AE needs a grammatically well-structured text to extract correctly the answers. The QA system offers a low performance (33% of accuracy) when using this AE over the web-based retrieved passages. In some cases, the snippets are cut and we could expect a better performance retrieving the whole documents from Google.

On the other hand, web-based snippet retrieval, with only one query per question, gives good results in Passage Retrieval. The QA system with the Frequency-Based AE obtained better results than with the ODQA AE (62.9% of accuracy).

Finally, we conclude that our approach with Geographical scope-based resources are notably helpful to deal with multilingual Geographical Domain Question Answering.

8   Future Work

As a future work we plan to improve the AE module using a semantic analysis with extended contexts (i.e. more than one sentence) and adding some spatial reasoning. We also want to improve the retrieval by crawling relevant documents from web search-engines instead of using snippets. This could be a good method to find more sentences with supported answers. Finally, we expect to do tests with English in another scope.

Acknowledgements

References

F. Benamara. 2004. Cooperative Question Answering in Restricted Domains: the WEBCOOP Experiment. In *Proceedings of the Workshop Question Answering in Restricted Domains, within ACL-2004*.

H. Chung, Y. Song, K. Han, D. Yoon, J. Lee, H. Rim, and S. Kim. 2004. A Practical QA System in Restricted Domains. In *Proceedings of the Workshop Question Answering in Restricted Domains, within ACL-2004*.

J. Diaz, A. Rubio, A. Peinado, E. Segarra, N. Prieto, and F. Casacuberta. 1998. Development of Task-Oriented Spanish Speech Corpora. In *Procceedings of the First International Conference on Language Resources and Evaluation*, pages 497–501, Granada, Spain, May. ELDA.

D. Ferrés, S. Kanaan, A. Ageno, E. González, H. Rodríguez, M. Surdeanu, and J. Turmo. 2004. The TALP-QA System for Spanish at CLEF 2004: Structural and Hierarchical Relaxing of Semantic Constraints. In C. Peters, P. Clough, J. Gonzalo, G. J. F. Jones, M. Kluck, and B. Magnini, editors, *CLEF*, volume 3491 of *Lecture Notes in Computer Science*, pages 557–568. Springer.

D. Ferrés, S. Kanaan, E. González, A. Ageno, H. Rodríguez, M. Surdeanu, and J. Turmo. 2005. TALP-QA System at TREC 2004: Structural and Hierarchical Relaxation Over Semantic Constraints. In *Proceedings of the Text Retrieval Conference (TREC-2004)*.

C-Y. Lin and E. Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *COLING*, pages 495–501. Morgan Kaufmann.

D. Molla and J.L. Vicedo. 2005. *AAAI-05 Workshop on Question Answering in Restricted Domains*. AAAI Press. to appear.