

AliQAn, Spanish QA System at CLEF-2005

S. Roger^{1,2}, S. Ferrández¹, A. Ferrández¹, J. Peral¹, F. Llopis¹, A. Aguilar¹ and D. Tomás¹

¹ Grupo de Investigación en Procesamiento del Lenguaje y Sistemas de Información

Departamento de Lenguajes y Sistemas Informáticos

University of Alicante, Spain

² Departamento de Computación

University of Comahue, Argentine

{sroger,sferrandez,antonio,jperal,llopis,dtomas}@dlsi.ua.es

Abstract

This paper describes AliQAn, a monolingual open-domain Question Answering (QA) System developed in the Department of Language Processing and Information Systems at the University of Alicante for CLEF-2005 Spanish monolingual QA evaluation task. Our approach is based fundamentally on the use of syntactic pattern recognition in order to identify possible answers. Beside, Word Sense Disambiguation (WSD) is applied to improve the system. The results achieved (overall accuracy of 33.00%) are shown and discussed in the paper.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval

General Terms

Algorithms, Measurement, Performance, Experimentation

Keywords

Question Answering for Spanish, Syntactic Patterns, Natural Language Processing, Word Sense Disambiguation

1 Introduction

Question Answering (QA) is not a simple task of Information Retrieval (IR). A QA system must provide concise answers to questions stated by the user in natural language.

The research in open domain QA has mainly focused around English due to the advances in IR and Natural Language Processing (NLP). However, the Cross-Language Evaluation Forum

Campaigns (CLEF) provide a multilingual forum for evaluation of QA systems in languages other than English. Multilingual open domain QA systems have been recognized as an important issue for the future of information search.

Nowadays, there are several types of implementations of Spanish QA systems. Generally, most of the systems are based on NLP tools [2, 5, 7, 8], such as Part of Speech (PoS) taggers, syntactic parsers, etcetera. On the other hand, some other approaches use machine learning and statistical models [3] like Hidden Markov Models in order to find the answer. Also, there are systems that combine NLP tools with statistical data redundancy techniques [9, 10].

The systems based on NLP tools are complex because of the number of different NLP tools that they use. Moreover, a good integration between them is needed. Our system have been developed during the last two years in the Department of Language Processing and Information Systems at the University of Alicante. It is based on complex pattern matching using NLP tools. Beside, Word Sense Disambiguation (WSD) is applied to improve the system.

As usual, in our approach, three task have been defined: question analysis, selection of relevant passages and extraction of the answer.

The rest of this paper is organized as follows: section two describes the structure and functionality of the system. Afterwards, the achieved results are shown and discussed in section three and finally, section four details our conclusions and future work.

2 System description

2.1 Overview

In this section, the structure and functionality of our approach to open domain Spanish QA system are detailed. The next paragraph presents the phases of the QA process.

Our approach is based fundamentally on syntactic analysis of the questions and the Spanish documents (the EFE collection in this case), where the system tries to localize the answer. In order to make the syntactic analysis, SUPAR [4] system is used, which works in the output of a PoS tagger [1]. SUPAR performs partial syntactic analysis that lets us to identify the different grammatical structures of the sentence. Syntactic blocks (SB) are extracted, and they are our basic syntactic unit to define patterns.

Using the output of SUPAR we are going to identify three types of SB: verb phrase (VP), simple nominal phrase (NP) and simple prepositional phrase (PP). For example in the sentence: *Hillary Clinton was in Jerusalen*, the obtained list of SB is: [NP, hillary*clinton] [VP, to be] [PP, in: jerusalen].

The overall architecture of our system (Figure 1) is divided in two main phases: Indexation phase and Search phase.

- **Indexation phase.** Indexation phase consists of arranging the data where the system tries to find the answer of the questions. This process is a main step to accelerate the process. Two different indexation are carried out: IR-n and QA indexation. The first one is carried out by IR-n system and it is independent from the second one, in which more syntactic and semantic information is stored. For example, the QA indexation stores the NP, VP and PP obtained from the parsing, and it also stores the results of the WSD process.
- **Search phase.** This phase follows the most commonly used schema. The three main modules of our approach are:

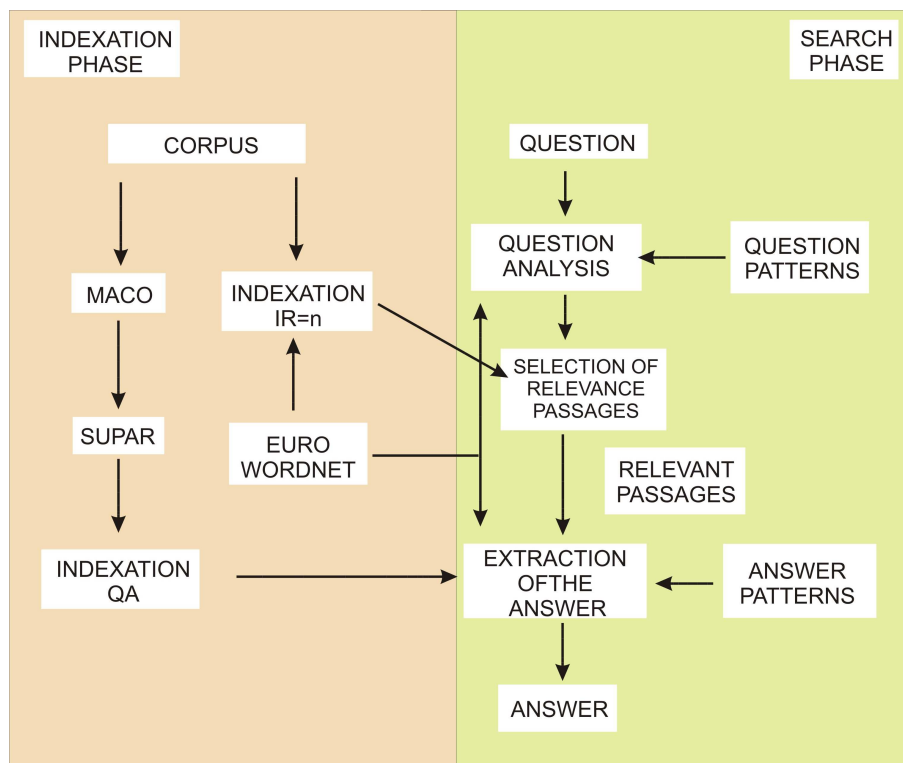


Figure 1: System architecture

- Question analysis
- Selection of relevant passages
- Extraction of the answer

These modules are described below. Previously, the used annotation is commented. The symbols “[]” delimit a SB (NP, VP and PP), “*sp*” is a preposition of a PP, the term “*ap*” indicates that PP is an apposition of the previous nominal head, *SOL* is the place where the answer can be found and the symbols “[...]” indicate some irrelevant SB for the search.

2.2 Question analysis

In this step the system carries out two tasks:

- To detect the type of information that the answer has to satisfy to be a candidate of answer (proper name, quantity, date ...).
- To select the question terms (keywords) that make possible to locate those documents that can contain the answer.

We have based on WordNet Based-Types and EuroWordNet Top-Concepts in order to develop our taxonomy that consists of the next categories: person, group, object, place, place city, place

capital, place country, abbreviation, event, numerical quantity, numerical economic, numerical age, numerical measure, numerical period, numerical percentage, temporary year, temporary month, temporary date and definition.

The expected answer type is achieved using a set of syntactic patterns. The question posed to the system is compared with all the patterns of all the categories. For each category a score is assigned that measures its probability of being the correct type. We choose the category having the highest probability.

We have 173 syntactic patterns for the determination of the different semantic category of our ontology. The system compares the SB of the patterns with the SB of the question, the result of the comparison determines the category of the question.

The next example shows the behavior of question analysis:

- Question:
 - *Quién es el Secretario General de la ONU? (Who is the General Secretary of the ONU?)*
- Syntactic Block
 - [IP *quién*](who)
 - [VP *ser*](to be)
 - [NP *secretario_general* [PP, *de: onu*]] (General Secretary of the ONU)

We have a pronoun or interrogative particle *quién* (*who*) followed by two syntactic blocks: a verb phrase and a nominal phrase. This example matches with the next pattern:

[IP, *quién* | *quiénes*] (*who*) [VP, *ser*] (*to be*) [NP, *hipónimo persona*] (*hyponim person*)

therefore, the category of the question is *person*.

For each SB of the pattern, we keep a flag in order to determine whether the SB of the question is considered for the next stage of the QA process or not.

2.3 Selection of relevant passages

This second module of the QA process creates and retrieves passages using IR-n system [6]. The goal of IR-n system is to extract a set of passages, where at least one passage contains the answer for the input question.

The inputs of IR-n are the detected keywords in question analysis, IR-n returns a list of passages where we apply the extraction of the answer process. Beside, the objective of this task is reducing complexity of the process of searching the solution by means of reducing the amount of text in which the system searches for the answer.

2.4 Extraction of the answer

The final step of QA is the extraction of the answer. In this module, the system takes the set of retrieved passages by IR-n and tries to extract a concise answer to the question.

Moreover, the type of question, SB of the question and a set of syntactic patterns with lexical, syntactic and semantic information are used in order to find a possible answer.

As shown in the next list, the system use the following NLP techniques.

- *Lexical level.* Grammatical category of answer must be checked according to the type of the question. For example, if we are searching for a *person*, the proposed SB as possible answer has to contain at least a noun.
- *Syntactic level.* Syntactic patterns have been defined. Those let us to look for the answer inside the recovered passages.
- *Semantic level.* Semantic restrictions must be checked. For example, if the type of the question is *city* the possible answer must contain a hyponym of *city* in EuroWordNet. Semantic restrictions are applied according to the type of the questions. Some types are not associated with semantic restrictions, such as *quantity*.

In order to design and group the patterns in several sets, the cases of the question are used. The patterns are classified in the followings three cases:

- **Case 1.** In the question, one SB of type NP or PP is only detected. For example:
 - Question: *Who is the president of Yugoslavia?*
We only have a SB, the verb *to be* that is not used to find the answer because it is a copulative verb.
 - SB: *[NP, president [PP, of: Yugoslavia]]*
- **Case 2.** A VP is detected in the question. This verb expresses an action that must be used in order to search the answer. For example:
 - Question: *Who did write Star Trek?*
 - SB: *[VP, to write] [NP, star*trek]*
- **Case 3.** VP is preceded by a NP or PP. In this case we used three sections to find out the possible answer.
 - Question: *Which team did win the NBA tournament?*
 - SB: *[NP, team] [VP, to win] [NP, NBA*tournament]*

When the system tries to find a possible answer in a sentence, first, the SB of the question are localized in the text, secondly the system attempts to match the pattern in the sentence. If this has been possible, then a possible answer has been founded that must be appraised using lexical and semantic restrictions according to the type of the question. Spanish QA system has about 60 patterns, the number of patterns that is processed in each sentence depends on the type of the question. Therefore, a question of case 1 and type “*person*” processes different patterns than a question of case 1 and type “*place_city*” .

The next example shows the used pattern and the behavior the extraction of the answer:

- *[SOL[PP, sp: NP1]] [...] [VP][...] [NP2]*

First, NP2 (or PP2) and VP are searched by the system, afterward the NP1 with the answer must be found. Next example shows the process:

- **Question:** *¿Qué presidente de Corea del Norte murió a los 80 años de edad? (What North Korea’s president died at the age of 80?)*
- **Type:** *person*
- **Case:** *3*

- **List of SB:** [NP, north*korea*president] [VP, to death] [PP, at: age [PP, of: 80]
- **Text:** [...] Kim Il Sung, presidente de Corea del Norte, murió ayer a los 82 años [...] ([...] Kim Il Sung, president of North Korea, died yesterday at the age of 80 [...])
- **List of SB of sentence:** [...] [NP, kim*il* sung [PP, apposition: president [PP, of: north*korea]]] [VP, to death] [PP, at: age [PP, of: 80] [...]
- **Answer:** Kim Il Sung

2.4.1 Value of the solution

In order to select the answer from a set of candidates, each possible answer is scored. The calculation of the value of the solution on each pattern is described in this subsection of the paper.

The score of a candidate is structured in three phases: comparison of the terms inside a nominal head of a SB with the terms of the nominal head of another SB, comparison of a SB of the question with a SB of the text and weighting of a pattern according to the different SB.

Comparison of the terms of a nominal head. When the system is comparing two terms, the system does not only contemplate the literal value of terms, also checks the relations between these terms in EuroWordNet. So, weighting of terms is calculated using the equation 1, where N is the number of terms inside nominal head and pt_i is the value of the terms that is calculated using EuroWordNet (*1 same lemma, 0.8 synonym and 0.6 hyponim*).

$$Pt = \frac{\sum_{i=1}^N pt_i}{N} \quad (1)$$

The equation 2 shows the process of comparison between the terms “*old Bosnian leader*” and “*senior Bosnian sailor*” obtaining the following results:

$$Pt = \frac{0.8 + 1 + 0}{3} = 0.6 \quad (2)$$

where *old* and *senior* have a synonym relation and both SB contain the lemma Bosnian. If the number of terms is different, the system divides using the great number.

Comparison of the SB. In our approach, the comparison of the SB occurs in two kinds of circumstances. When the SB of the question is localized in the text in order to apply a pattern and when the system is analyzing a SB to find the answer.

The first type of comparison is called “*value of terms*”, this measure can be affected by fixed circumstances, such as:

- *Depth of appearance.* The terms of the SB of the question may not appear as nominal heads in a SB of the text.
- *Excess or missing of modifiers.* If the nominal head of the question has more or less modifiers its value is penalized.
- *Appearance of terms, but some complements are missing.* When the system detects only the term of the question in the text, then it continues the searching until it is able to find the complements.

Second type of comparison of SB is the calculation of the *value of solution*, this value is calculated when it is searching for a possible answer. It takes into account a set of evaluation rules according to the type of the question, such as:

- *Lexical restrictions.* Grammatical category of the answer depends on the type of the question. For example, a question of type “persona (person)” the answer must have at least a proper noun or common noun.
- *Semantic restrictions.* The system leaks the answer according to semantic relations such as hyponymy. For example, a question of type “ciudad (city)” the answer must be a hyponim of “ciudad (city)” in EuroWordNet.
- *Ad-hoc restrictions.* An example of this kind of restrictions is founded in the questions of type “fecha (date)”, when the system penalizes the value of solution if the answer does not contain day, month and year.

Comparison of the patterns. When the system is evaluating a pattern in the text, a set of circumstances are considerate in order to provide the value of solution. The total value of an answer is defined by the equation 3, where N is the number of retrieved SB of the question, vt_i is the value of terms of each SB, d is the distance between the localized SB in the text and vs is the value of solution. As shown in the equation 3, vs is 30% of total and the remaining ones is the 70%.

$$Vr = \left(\frac{\sum_{i=1}^N vt_i}{N} - d * 0.1 \right) * 0.7 + vs * 0.3 \quad (3)$$

Final evaluation of patterns. The system generates a list of candidate solutions, where each solution has been obtained in a passage. If two solutions have the same value for a question, the system chooses one considering the proposed order by IR-n.

Spanish QA system must determine when a question has answer or not. In order to do that we suggest an umbral that indicates if an answer is solution or not. A question has answer if its Vr is higher than 0.5.

Next, an example (question 114, *In Workshop of Cross-Language Evaluation Forum (CLEF 2003)*) of resolution of one question, where system chooses the correct solution since the Vr is higher than 0.5.

- **Question:** A qué primer ministro abrió la Fiscalía de Milán un sumario por corrupción? (To whom prime minister the Office of the public prosecutor of Milan opened a summary for corruption?)
- **Type:** person
- **Case:** 3
- **List of BS:**
 - **NP1:** ([NP, primer*ministro])
 - **VP:** ([VP, abrir])
 - **NP2:** ([NP, fiscalia [PP, de: milan]])([NP, sumario [PP, por: corrupcion]])
- **Text where is a correct solution:** “[...] la **Fiscalía de Milán** abrió, hoy martes, un **sumario** al **primer ministro**, *Silvio Berlusconi*, por un supuesto delito de **corrupción** [...]”
- **Value of the solution:**

$$Vr = \left(\frac{1([NP,fiscalia[PP,de:milan]])+0.65([NP,sumario[PP,por:corrupcion]])+1([NP,primer*ministro])+1([VP,abrir])}{4} - 0 * 0.1 \right) * 0.7 + 1(Silvio - Berlusconi) * 0.3 = 0.93$$

Run	alia051eses-assessed	alia052eses-assessed
Right	66	60
Inexact	24	26
Unsupported	0	0
Wrong	110	114
Presicion(NIL)	0.25	0.24
Recall(NIL)	0.45	0.45
Accuracy over:		
overall questions	33.00%	30.00%
Factoid questions	29.66%	26.27%
Definition questions	40.00%	36.00%
Temporal Restricted Factoid questions	34.38%	34.38%

Table 1: General results obtained for each runs

Type	Right	Inexact	Number Questions
Factoid	35	10	118
Definition	20	13	50
Temporal	11	1	32
Total	66	24	200

Table 2: Accuracy over questions

- **Text where is an incorrect solution:** “[. . .] **primer ministro** y líder socialista, *Bettino Craxi*, al que el pasado 21 de septiembre Paraggio **abrió un sumario** relacionado con el proyecto Limen por supuestos delitos de **corrupción** [. . .]”

- **Value of the solution:**

$$V_r = \frac{0([NP,fiscalia[PP,de:milan]])+0.45([NP,sumario[PP,por:corrupcion]])+1([NP,primer*ministro])+0([VP,abrir])}{4}$$

$$0 * 0.1) * 0.7 + 1(Bettino - Craxi) * 0.3 = 0.52$$

- **Answer:** Silvio Berlusconi

As the previous example shows, the system chooses the correct answer among several possible solutions. The correct answer has been chosen due to the *value of terms*.

In the sentence, with the right answer, the value of terms is higher than in other sentences. Although, the VP and NP1 are in both sentences, the NP2 is just completely in the first sentence.

3 Results

This section describes some tables related with the results and the evaluation of our system in CLEF-2005. The proposed system was applied to the set of 200 questions, all of them was supported by our system.

For the development of our system we used as training set the questions developed for CLEF-2003 and CLEF-2004 questions.

During this test process many faults were detected in the tools used in the lexical and morphological phases. The analysis of question 145 of CLEF-2003 shows one of these errors:

- *Quién es el ministro de economía alemán? (Who's the German Minister of Finance?)*

The term *Alemán* is not in the prepositional phrase where the term *economía* is, because of *economía* is tagged as feminine and *aleman* is tagged as masculine. So, when searching for SB in the corpus to find an answer for the questions, it gives wrong answers.

We submitted two runs. The first run was obtained applying the system after repairing the lexical and morphological errors that we have detected (*alia051eses*) while the second run (*alia52eses*) performed QA process without repairing these faults. Table 1 shows the results for each run and how these errors lowered our system performance giving wrong answers.

Inexact answers also lower the system performance and in our system, these are due to errors in parsing process. An answer was judged inexact when the answer string contained more or less than just the correct answer, ie. the system finds this answer in the text but it does not extract the part of the information needed to return it as an answer. Our system returned 24 inexact answers (see Table 1). We may obtain a higher level of performance (45%) if we take into account that these inexact answers include the expected answer.

Finally, Table 2 shows that the accuracy over temporal questions was 34.38%, ie. we have obtained 11 right answers over 32. This is considered a good score because no special mechanism was developed.

4 Conclusion and future work

For our first participation in the QA@CLEF track, we proposed a QA system designed to search Spanish documents in response to Spanish queries. To do so we used a Spanish syntactic analyzer in order to assist in identifying the expected answers and the solution of the question.

All questions given in this track have been supported by our system. The results showed overall accuracy levels of 33%.

As previously mentioned, the used tools reduce exactitude to our system (*alia52eses*). These are encountering results that show the potential of the proposed approach, taking into account that the use of patterns is a less expensive recourse compared with other proposals.

Ongoing work on the system is focused on multilingual task, temporal question treatment and the incorporation of knowledge to those phases that can be useful to increase the our system performance.

5 Acknowledgments

This research has been partially funded by the Spanish Government under project CICYT number TIC2003-07158-C04-01 and by the Valencia Government under project number GV04B-268.

References

- [1] S. Acebo, A. Ageno, S. Climent, J. Farreres, L. Padró, R. Placer, H. Rodriguez, M. Taulé, and J. Turno. MACO: Morphological Analyzer Corpus-Oriented. *ESPRIT BRA-7315 Aquilex II, Working Paper 31*, 1994.

- [2] A. Ageno, D. Ferrés, E. González, S. Kanaan, H. Rodríguez, M. Surdeanu, and J. Turmo. TALP-QA System for Spanish at CLEF-2004. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 425 – 434, 2004.
- [3] C. de Pablo, J.L. Martínez-Fernández, P. Martínez, J. Villena, A.M. García-Serrano, J.M. Goñi, and J.C. González. miraQA: Initial experiments in Question Answering. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 371 – 376, 2004.
- [4] A. Ferrández, M. Palomar, and L. Moreno. An Empirical Approach to Spanish Anaphora Resolution. *Machine Translation. Special Issue on Anaphora Resolution In Machine Translation*, 14(3/4):191–216, December 1999.
- [5] J. Herrera, A. Peñas, and F. Verdejo. Question Answering Pilot Task at CLEF 2004. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 445 – 452, 2004.
- [6] F. Llopis and J.L. Vicedo. Ir-n, a passage retrieval system. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, 2001.
- [7] E. Méndez-Díaz, J. Vilares-Ferro, and D. Cabrero-Souto. COLE at CLEF 2004: Rapid prototyping of a QA system for Spanish. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 413 – 418, 2004.
- [8] M. Pérez-Coutiño, T. Solorio, M. Montes y Gómez, A. López-López, and L. Villaseñor-Pineda. The Use of Lexical Context in Question Answering for Spanish. *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 377 – 384, 2004.
- [9] J.L. Vicedo, R. Izquierdo, F. Llopis, and R. Muñoz. Question Answering in Spanish. *Comparative Evaluation of Multilingual Information Access Systems: 4th Workshop of the Cross-Language Evaluation Forum, CLEF, Trondheim, Norway, August 21-22, 2003, Revised Selected Papers. Lecture Notes in Computer Science*, 3237/2004:541, 2003.
- [10] J.L. Vicedo, M. Saiz, and R. Izquierdo. Does English helps Question Answering in Spanish? *In Workshop of Cross-Language Evaluation Forum (CLEF)*, pages 419 – 424, 2004.