XisQuê: An Online QA Service for Portuguese

António Branco, Lino Rodrigues, João Silva, and Sara Silveira

University of Lisbon, Portugal
{antonio.branco,lino.rodrigues,jsilva,sara.silveira}@di.fc.ul.pt

Abstract. This paper describes XisQuê (`http://xisque.di.fc.ul.pt`) an online service for real-time, open-domain question answering (QA) on the Portuguese Web.

Key words: QA, question answering, real-time QA, open-domain QA, web-based QA, factoids

1 Introduction

In this paper we present XisQuê a real-time, on-line service for open-domain Question Answering (QA) over the Portuguese Web.

Paper structure. Section 2 presents the architecture adopted for the QA system and in Section 3, the performance of the system is described in terms of its speed and ability to deliver appropriate answers.

2 The Underlying QA System

XisQuê is supported by a QA system developed to comply with the following major design features:

Portuguese input: the admissible input are well-formed questions from Portuguese (e.g. *Quem assassinou John Kennedy?*).

Real-time: the system provides the output in real-time.

Web-based: the answers are searched in documents retrieved on the fly from the Web.

Portuguese Web: the documents are obtained in the Portuguese web, that is the collection of documents written in Portuguese and available on-line.

Open-domain: the questions may address issues from any subject domain.

Extraction-based: the answers returned are excerts of the retrieved documents without additional processing.

At the system's heart lies the QA infrastructure described in [1], which is responsible for handling the basic non-linguistic functionality. Its architecture follows what has become a quite standard configuration that has been explored and perfected in similar QA systems for other natural languages [2]:

Question Processing. This phase involves three tasks: (i) detection of the expected semantic type of the admissible answers; (ii) gathering of relevant keywords; (iii) extraction of the main verb and major supporting NP of the input question.

Document Retrieval. In this phase, the system acts as a client of search engines (viz. Ask, Google, MSN Live and Yahoo!), submitting the list of keywords obtained in previous phase and retrieving relevant documents.

Answer Extraction. The last phase includes two tasks performed over the retrieved documents: (i) the sentences most likely containing an admissible answer are selected; (ii) candidate answers are extracted from the selected sentences. XisQuê delivers up to 5 candidate answers (termed "short answers" below) together with the sentences from which they were extracted ("long answers"). It may happen that for some answers only "long answers" are provided. See the example of an outcome in the Annex.

On top of this infrastructure, the natural language driven modules were implemented by using state-of-the-art shallow processing tools developed at our group. They include tools for sentence and token segmentation, POS annotation, morphological analysis, lemmatization and named entity recognition, specifically designed to cope with the Portuguese language [3–5].

3   Performance

The online service was evaluated along two simensions: (i) timeliness, or the speed at which answers are returned; and (ii) appropriateness, or the ability of the system to answer appropriately. A total of 60 test question were randomly picked from Trivial Pursuit® cards, by selecting 15 questions for each of the four interrogative pronouns the system handles (viz. `Quem`, `Quando`, `Onde` and `Que`) This test set is at `http://xisque.di.fc.ul.pt/features.html`.

	Question type				Overall
	Quem	Quando	Onde	Que	(average)
Total time (msec.)	18896	20026	22706	25093	21680
"Outside" time	11569	12058	12421	17488	13839
Core QA system time	7327	7968	8465	7605	7841
Answers returned (short)	60.00%	66.67%	46.67%	53.33%	56.67%
Answers returned (all)	100.00%	100.00%	100.00%	100.00%	100.00%
Accuracy (short)	60.00%	66.67%	40.00%	53.33%	55.00%
Accuracy (all)	93.00%	100.00%	100.00%	100.00%	98.33%
MRR (short)	0.5167	0.4778	0.4333	0.5000	0.4819
MRR (all)	0.6489	0.6667	0.7444	0.8889	0.7372

Table 1. Timeliness and Appropriateness scores obtained March 3-5, 2008

3.1   Timeliness

The service was assessed with respect to time it takes on average to return answers to the input questions. From a development point of view, it is instructive

to also determine how much of that time is spent searching for and downloading documents, since those tasks are contingent on third-party search engines that lie outside the QA system proper.

Table 1 shows the average running time in miliseconds. There are some variations when we consider different questions types, but it is mostly caused by fluctuations in the retrieval time (2 696 std. dev.) since the variations for system time are much smaller (492 std. dev). Overall, the system takes an average of 22 seconds to display the page with the results, with 14 (ca. 64%) of those being spent "outside" the system.

3.2   Appropriateness

Evaluating the appropriateness of a QA system that runs over the Web poses specific problems since the Web is mutable and the results that are obtained for the same set of test questions under different evaluation runs may vary due to external factors, such as website availability of the relevant documents. As a consequence, there is no fixed gold standard against which the output of the system can be automatically compared. Nevertheless, it is possible to obtain an indicative measure of the system's performance through sampling, by manually evaluating the answers to the set of questions.

Table 1 summarizes the scores for a few evaluation metrics: **Answers returned** is the proportion of questions for which the system provided at least a candidate answer — regardless of its rank in the five answer list or even regardless its being a correct answer. Overall, the system provides candidate answers (short- or long-) to 98.33% of the questions in the test set. In turn, it provides short candidate answers to 58.33% of the test set questions. **Accuracy** is the proportion of questions for which a correct answer was provided — regardless its rank in the five returned answer list. In the "all" line, a long-answer is counted in the lot of the correct ones in case it is correct and no short-answer (correct or not) was extracted from it. The system provides a correct short-answer to 45.00% of the test set questions and a correct answer (short- or long-) to 98.33% of that same set. **MRR** stands for mean reciprocal rank: it is a measure commonly adopted in QA evaluation of how highly, on average, the first correct answer is ranked in the answer list [6]. For instance, if all questions have a correct answer and these all appear in position 1, the MRR scores 1; in case they would all appear in position 2, the MRR would score 0.5. The overall value obtained for the QA system is 0.7539 when short- and long-answers are considered, and is 0.4819 when only short-answers are taken into account (a value of 0 was assigned for questions without any short-answer).

4   Conclusion

In this paper we presented the first QA service that complies with all of the following desgin features: it is a real-time, open-domain, freely accessible on-line factoid QA service for the Portuguese Web.

References

1. Rodrigues, L.: Infra-estrutura de um serviço online de resposta a perguntas com base na web portuguesa. Master's thesis, Universidade de Lisboa, Portugal (2007)
2. Paşca, M.: Open-Domain Question Answering from Large Text Collections. CSLI Studies in Computational Linguistics. CSLI Publications (2003)
3. Silva, J.R.: Shallow processing of Portuguese: From sentence chunking to nominal lemmatization. Master's thesis, Universidade de Lisboa, Portugal (2007)
4. Nunes, F.: Verbal lemmatization and featurization of Portuguese with ambiguity resolution in context. Master's thesis, Universidade de Lisboa, Portugal (2007)
5. Ferreira, E., Balsa, J., Branco, A.: Combining rule-based and statistical methods for named entity recognition in Portuguese. In: Actas da 5ª Workshop em Tecnologias da Informação e da Linguagem Humana. (2007)
6. Voorhees, E.: The TREC8 question answering track report. In: Proceedings of the 8th Text REtrieval Conference (TREC). (1999)

Annex – System Output Example