

## Developing a Question Answering System

Victoria Maxim, Victoria Lazu, Victoria Bobicev

**Abstract:** The paper presents a question-answering system created in the project „Research in the field of Information Retrieval for question answering system creation”. The system consists of question analysis and answer retrieval and extraction module. At the first stage the system is working on the basis of documents provided by IDSI (Institutul de Dezvoltare a Societății Informaționale - Information Society Development Institute). The system is available online in order to collect questions from different users and to improve it by using collected information.

**Keywords:** automatic text processing, question answering system, automatic analysis of questions, information retrieval.

### 1 Introduction

At present, the most effective method of finding and acquiring information represents search engines, whose goal is to provide a list of links to websites for the user where he can find the necessary information. Often, the links proposed by the search engines do not meet the user's desire to get an adequate answer. In addition, they do not provide a specific response to user's queries, but only a set of links to web pages and the user is forced to spend time extracting the necessary information. Thus, the next step in the information acquisition is development of systems able to answer user's questions in natural language.

A Question Answering (QA) system requires a more complex natural language processing than a simple search engine. Most question answering systems use a variety of linguistic resources to help in understanding the user's query and matching sections in documents.

Our system is created within the project "Research in information retrieval to create electronic public information system" and operates with documents provided by IDSI (Institutul de Dezvoltare a Sicietății Informaționale - Information Society Development Institute) at the first

stage of the project. It is available online on <http://lilu.fcim.utm.md/QASystem/> in order to collect questions from different users and to improve it by using collected information.

## 2 Question Processing

Almost all QA systems use question type to judge the answer. They classify the question types on the basis of types of answers users are expecting. For example, “Who is ...?” will be assigned to “PERSON/ORGANIZATION” type; while “When did ... happen?” will be classified as “DATE/TIME” question [2].

The classical methodology of question’s semantic analysis proposed in [1] requires dividing the interrogative sentence into the following items: interrogative pronouns (eg, who, what, where, when) or interrogative prepositional group (eg, in what, to whom, from where) and the inverted sentence. Inverted sentence begins with verbal group followed by the noun group. At the end of the sentence there is the so-called "gap" - free space provided for the answer. Such methodology is developed for questions in English and it was adapted to analyze the interrogative sentences in Romanian.

Table 1. The pattern of question analysis

Nr.	Asking point	Verb part	Key words
1	care	este	domeniul de reglementare al prezentei legi
2	în ce	constă	scopul prezentei legi
3	ce	reprezintă	un proiect de inovare și transfer tehnologic
4	la propunerea cui	se formează	parcul științifico-tehnologic sau incubatorul de inovare
5	cum poate	activa	incubatorul de inovare
6	cine	își desfășoară	activitatea în parcul științifico-tehnologic

The noun group of the inverted sentence is considered the key phrase for searching response. Verb and noun groups in the reversed sentence change their places and if this question begins with a preposition, it is attached to the end. The free space for the answer is reserved after these elements.

Table 1 presents the types of questions and methodology of their analysis. After processing, the question is transformed into a query phrase for the search system.

### 3 Evaluation

The answer extraction module selects the sentence or paragraph that is top ranked by the weight calculated on the basis of the question's keywords. Figure 1 presents an answer of the system in the current form with some additional details in order to evaluate their performance.

***Analiza întrebării și vizualizarea răspunsului***

***Întrebarea pusă: Ce este cercetare***

***Răspunsul la întrebarea pusă s-a căutat în Codul cu privire la știință și inovare al Republicii Moldova***

***Răspunsul cu scorul maxim este #7 din 22 alese:***

***Titlu: I Capitol: II Articol: 6***

***Cercetare fundamentala - activitate orientata spre dobindirea de noi cunostinte stiintifice, spre formularea si verificarea de noi ipoteze si teorii.***

alte răspunsuri selectate  
la documentul în care s-a căutat răspunsul  
la pagina precedenta

\*\*\*

Figure 1. Web page with the answer and choices offered to the user.

We performed the preliminary testing of the question-answering system developed within the project, and highlighted the strong and weak points of the system

We concentrated on the reasons of the incorrect answers extracted by the system and the most important are:

- First of all, the questions wrong formulated from grammatical point of view. Most of people are posing the questions as they speak and it's difficult to analyze a question that looks like a simple sentence.

- Incorrect parsing of the question by the system.

- The above mentioned reasons lead to wrong keywords extraction, and incorrect answers.

While we cannot influence the first reason, we focused on the question parsing and key words formation. We should improve the analysis of the question and keywords extraction in order to obtain the adequate answer.

After the complex analysis of answers we calculated that the system correctly answers only 35.8% of questions. It's a small range but it gives us new challenge to improve the system.

## 4 Conclusion

This article describes analysis of question answering system that is created in the project "Research in the field of Information Retrieval for question-answering system creation" The system is available online as a web application. The system was evaluated with a set of questions and the causes of wrong answers are analyzed. The next step of the project is to improve the question analysis on the base of presented evaluation.

## References

- [1] R. Baeza-Yates, B. Ribeiro-Nieto, *Modern Information Retrieval*, Addison Wesley, 2000.
- [2] D. Tufiş, D. Ştefănescu, R. Ion, A. Ceauşu. RACAI's *Question Answering System at QA@CLEF 2007*. In Alessandro Nardi and Carol Peters, editors, Working Notes for the CLEF 2007 Workshop, pages 15–21, 2007.
- [3] V. Bobicev *Preprocesarea textelor în sistemele de tip „întrebare-răspuns*. 7th International Conference on Microelectronics and Computer Science, Chişinău, 2011.
- [4] V. Lazu, T. Prodan *Procesarea și analiza întrebărilor într-un sistem de „ÎNTREBARE-RĂSPUNS”*, In International Conference ICTEI 2012, Chişinău. Vol. I, pp.376-380, 2012.

Victoria Maxim<sup>1</sup>, Victoria Lazu<sup>2</sup>, Victoria Bobicev<sup>3</sup>

<sup>1,2,3</sup>Technical University of Moldova

<sup>1</sup>E-mail: maxivica@yahoo.com

<sup>2</sup>E-mail: lazu.vica@mail.utm.md

<sup>3</sup>E-mail: vika@rol.md