IJRIT International Journal of Research in Information Technology, Volume 1, Issue 6, June 2013, Pg. 46-51

International Journal of Research in Information Technology (IJRIT)

www.ijrit.com ISSN 2001-5569

# A Turkish Automatic Question Answering System with Question Multiplexing:Ben Bilirim

[1] Göksel Biricik, [2] Sercan Solmaz, [3] Ertan Özdemir, [4] M. Fatih Amasyalı

[1] Research assistant, Ph.D., Computer Engineering Department, Yildiz Technical University, Istanbul, Turkey
[2] B.Sc. pursuing, Computer Engineering Department, Yildiz Technical University, Istanbul, Turkey
[3] B.Sc. pursuing, Computer Engineering Department, Yildiz Technical University, Istanbul, Turkey
[4]Assistant professor, Ph.D., Computer Engineering Department, Yildiz Technical University, Istanbul, Turkey

[1] goksel@ce.yildiz.edu.tr, [2] info@sercansolmaz.net,
[3] ertanozdemir88@gmail.com, [4] mfatih@ce.yildiz.edu.tr

**Abstract**

Most people query search engines as if they were talking to another person. Those who know how these engines work use special keywords and escape characters in queries. Since most of the query results include the desired answer, there is a considerable amount of unnecessary results. In fact, it is possible to give the desired answers to questions that are asked in certain forms. There are automatic question answering implementations for English that take the question in natural language and give the correct answer. In this study, we developed an automatic question answering system for Turkish. Our system primarily reshapes the user's question to a search engine query string with a unique contributive method and then chooses possible answers from the result set of the search engine. Possible answers are then scored according to a variety of categories and ten highest scored result sentences are shown to the user. Our system found the correct answer in first three sentences on 82% of the evaluation questions, when only the search result summaries are used.

**Keywords:** *Automatic question answering, Verb multiplexing, Query processing, Natural language processing.*

## 1. Introduction

One of the most common and easiest ways to retrieve information is making use of the Internet, and using search engines is the most straightforward and fastest way of reaching information on it. Internet search engines are web sites that index other sites according to their contents and interconnections. Thus, they allow users to reach to the sites which contain information they are looking for.

In the recent years, search engines are the most visited sites on the web. Furthermore, they rapidly improve and usage rates constantly increase. However, the problem of organizing information on the net is not yet efficiently solved. Search engines crawl the net using bots to build their indexes. These search bots collect, store and index all kinds of data in the sources they visit, as long as they continue their crawl. The continuity of the process is achieved

by incrementally adding all of the discovered links to the crawl visiting queue. These addresses are connected to index keywords in order to reach the content in that Internet address. The crawled Internet resources are weighted either by the originality of the content and meta-data, or number and quality of inbound and outbound links, or both. When a user enters a query, the search engine parses the query and extracts keywords, searches in the index and retrieves the results, where the most relevant weighted resources are shown in top spots [1].

As discussed above, search engines allow us to retrieve information on the web. However, there are lacks of the indexing and weighting techniques that they use. For example, answering questions given in natural language is still a challenging problem. There are studies [2, 3, 4] which can give answers to questions in English [2, 3]. The studies in other languages are still inadequate.

In this study, we developed a system that searches the Internet vie to answer the questions asked by a user in Turkish language. Because of the difficulties of Turkish language processing, the types of questions that our system can find a solution to are limited. Our system can give answer to real life questions like "Who discovered electricity?", "When was Mahatma Gandhi born?", and "Which city is famous with the Golden Temple?". Even if it is possible to find the answers to these questions through search engines by conventional ways, the proposed system can present the answers without strolling among tens of query results.

This paper is organized as follows. In second section, previous related work is summarized. Third section explains the principles of how the designed system works. In the last section we evaluate the outcomes of our system.

## 2. Related Work

Automatic question answering systems that work on natural language processing principles get more popular as more information pile up in the Internet. Previous studies in this field are designed to work on specific topics. Each solution is tailored for a specific language. Since there are examples of academic research on Turkish automatic question answering [5], they did not turn into usable systems. On the other hand, it is easy to find many examples in English. This section briefly introduces well-known systems.

### 2.1 Answerbus

Answerbus [2] can accept queries in English, German, French, Italian and Portuguese. In the first step, the system determines the language of the query with a simple module. If the query is not given in English, it is translated via Altavista Babelfish, which is now migrated to Bing translation service [6]. After this step, the system determines the question type and generates queries suitable for different search engines. Each question type is associated with a different search engine. After getting the results from the search engine, it selects and scores the sentences in the result set. In order to compare the performances of search engines on question types, 2000 TREC questions [7] are used. Every word in these questions are indexed and scored by counts of correct answers to questions, on search engine basis. Answerbus offers that YahooNews performs better than Google for current affairs.

### 2.2 Quantum

Quantum also works on TREC database [3]. The system analyzes user question and selects an extraction function that shows how the answer ranks among the result set, by using labels and synonyms. Definition, quantity, feature, person and place are the examples of extraction functions. The sum of two ranking methods is used for scoring the possible answers. The first method ranks by the extraction function's relevancy score. The second method identifies the score given to the paragraph which contains the possible answer.

### 2.3 AskMSR

AskMSR reforms and weights the user's question based on manually defined rules and posts it to Google [4]. Rather than searching the whole page contents in the result set, 1-gram, 2-gram and 3-grams in the result summaries generated by Google are used. These n-grams are weighted by their frequencies and the weight of the query. This approach makes AskMSR to process queries very fast.

## 3. Proposed System

We aim to find the most relevant answers to the user's question. Typically, our proposed system carries same characteristic system architecture with other automatic question answering systems. Our main contribution is that we determine the verb of the question sentence given in natural language and multiplex the query with its synonyms. Our system works for Turkish. Figure 1 shows the outline and operation processes of our proposed system. Our system accepts questions in natural language. It is important to obey the grammatical and spelling rules in order to find more precise answers.

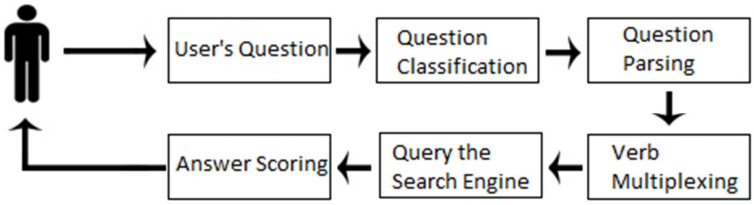


**Fig. 1 Operation steps of the proposed system**

In the first step, our system classifies the user's question. We choose five types of questions for evaluation. These types are "Date/Time", "Person", "Definition", "Quantity" and "Geography". The questions which cannot be classified into one of the predefined classes are labeled as "General". A keyword dictionary is utilized in order to find the corresponding class of the question. The question is parsed into words and word groups, which are then matched with the keyword dictionary. The question is categorized if there is a match, or labeled as "General" otherwise.

All of the words in user's question may not be helpful for search engine querying. We eliminate prepositions and conjunctions as they do not have a solitary meaning. We also eliminate stopwords. Remaining words are stemmed. We used Zemberek [8], a Turkish open source natural language processing library for stemming. Zemberek also helps tagging the part of speech, as verb or noun.

We find the synonyms of the verb tagged words by looking up in the thesaurus dictionary. We use these synonyms to multiplex the user's query. This process enhances the probability of retrieving more precise results. We send the produced queries to the search engine and retrieve results. Since it is possible to process the full contents of the result set, we only used the summary snippets of the search engine query results for ease of processing and speed considerations.

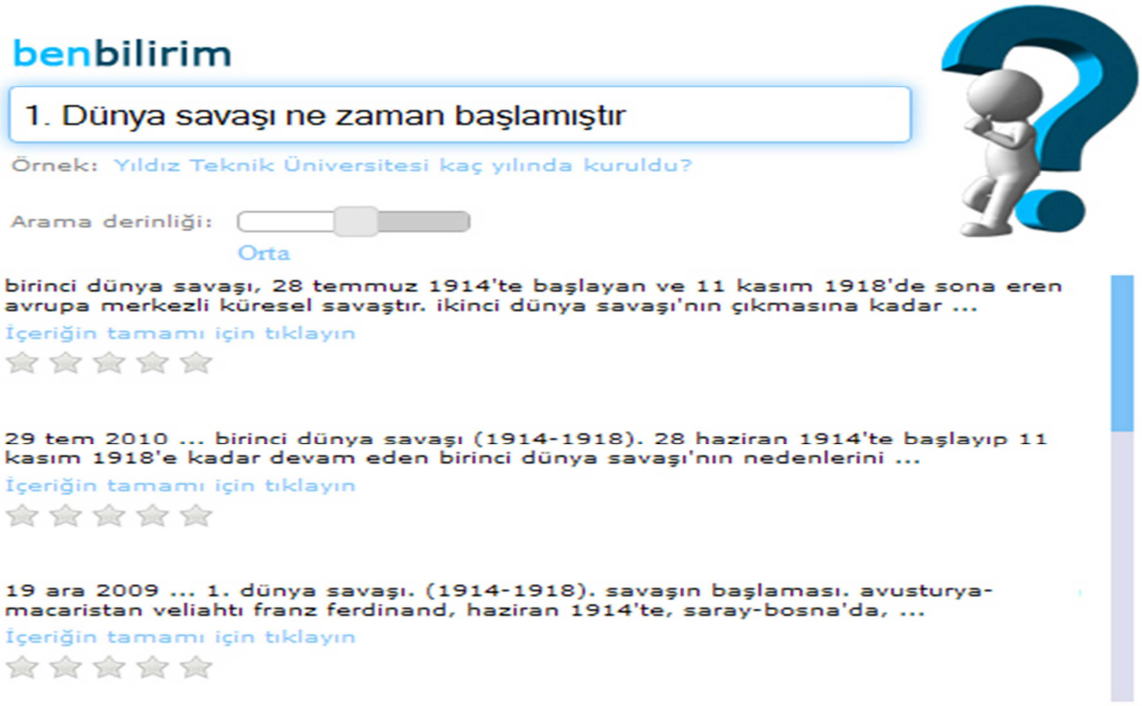


**Fig. 2 An example query screenshot from the system**

After retrieval, every item in the result set is scored in order to identify their relevance to the user's query. For scoring, term frequencies of the keywords are utilized. After scoring each item on the result set, top ten scoring discrete results are given to the user. A screenshot of the working system is given in Figure 2. Since our system works for Turkish, the question and answers in Figure 2 are in Turkish. The question is "When did the World War I begin?"

## 4. Experimental Results

We designed a feedback mechanism for testing the performance of our system. Users can rate the answers that the system presents, in a scale ranging from 1 to 5. 1-2 points are given to unsuccessful or wrong answers. 3 points means that user has no idea about the correctness of the results. Correct answers worth 4-5 points. A result with 1 point means that it is definitely wrong. Contrarily, result with 5 points is the correct answer for the user's question. Since there is no publicly available tool or study exists for automatic question answering in Turkish, we did not have the opportunity to achieve comparison tests for our system.

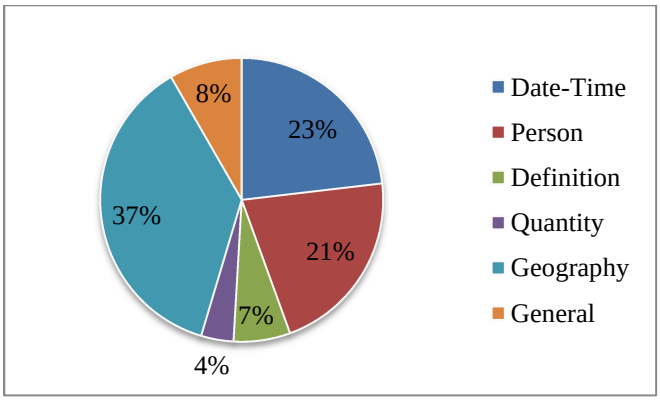


**Fig. 3 Distribution of question types in the evaluation question set**

We evaluated our system with 108 sample questions. Distribution of questions among question types is shown in Figure 3. Answers of 89 questions are rated as correct by the users. Our system retrieved wrong answers for 9 questions. Users were uncertain for the answers of 10 questions in the evaluation set. Overall performance graphic is given in Table 1 and visualized for easier comparison in Figure 4.

**Table 1 Evaluation of answering performance based on question types**

| Question Type | Correct | Uncertain | Incorrect | Total |
|---|---|---|---|---|
| **Date-Time** | 18 | 5 | 2 | **25** |
| **Person** | 15 | 3 | 5 | **23** |
| **Definition** | 7 | 0 | 0 | **7** |
| **Quantity** | 4 | 0 | 0 | **4** |
| **Geography** | 36 | 2 | 2 | **40** |
| **General** | 9 | 0 | 0 | **9** |
| **TOTAL** | **89** | **10** | **9** | **108** |

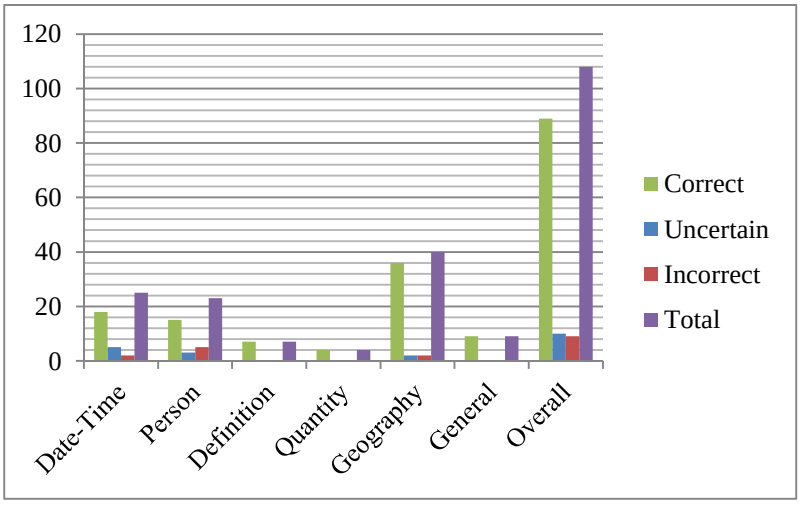


**Fig. 4 Overall performance comparison based on question types**

When we analyze the evaluation results, we see that our approach works best on “definition” and “quantity” question types. Most of the questions in the evaluation are in “geography” type, where the system also shows a great performance of 90% correct answer ratio. “Person” question type produced the worst correct answer ratio, with 15 correct answers out of 23 questions. Overall correct answer ratio performance of the system is very convincing, with 89 correct answers to 108 questions. Incorrect answer ratio ignoring uncertain labeled answers is 9 out of 108.

## 4. Conclusions

In this study we developed an online Turkish automatic question answering system based on question multiplexing by using the synonyms of the verbs in the questions. In the first step, our system determines the verb of the given question and finds its synonyms, which is our contribution. The question is then reshaped into multiplexed search engine query strings using the synonymous verbs. The result set is processed for scoring, where term frequencies of the keywords in the query are used. Scoring is enhanced by using a predefined dictionary, built upon previous questions. Users can get answers in low, middle or high accuracy levels, which are determined by the score scales. The system chooses possible answers from the result set of the search engine, which is ordered by their relevancy rankings based on the scores and selected accuracy level. After retrieval, every item in the result set is evaluated by the user to identify their relevance to the user’s query. This step also helps us to improve the system and measure its performance.

Our evaluation tests based on the user feedbacks show that our system achieved 82.4% correct answer hit ratio. Incorrect answer ratio is 8.3%, excluding uncertain results. If all answers that are not correct are considered, then the error rate is calculated as 17.6%. Since we do not limit the question types that the user can give to the system, this performance is very convincing. The overall operation time for a query directly depends on the Internet connection speed. In our evaluation tests, our system responded averagely in 3 seconds.

Our future work includes extending the question types that our system can identify and respond. Our ultimate aim is to enhance the system to give one exact right answer for any given question, rather than listing ten answers and responding the first three answers.

## 5. References

[1] W.S. Jawadekar, Knowledge Management: Text and Cases, New Delhi, Tata McGraw Hill, 2011.

[2] Z.Zheng, “AnswerBus question answering system", in 2nd Int. Conf. on Human Language Technology Research, 2002, pp. 399–404.

[3] L. Plamondon, and L. Kosseim, "QUANTUM: A Function-Based Question Answering System", in 15th Conference of the Canadian Society for Computational Studies of Intelligence, 2002.

[4] M. Banko, E. Brill, S. Dumais, and J. Lin, "AskMSR: Question Answering Using the Worldwide Web", in 2002 AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases, 2002.

[5] M. F. Amasyalı, and B. Diri, "Bir Soru Cevaplama Sistemi: Baybilmiş", Computer Science and Engineering Journal of Turkish Informatics Foundation, Vol.1, 2005, pp.37–51.

[6] J. Callaham, "Yahoo’s Babel Fish replaced with Bing Translator", http://www.neowin.net, 2012.

[7] E. M. Voorhees, and D. M. Tice, "The TREC-8 Question Answering Track Evaluation", in 8th Text REtrieval Conference (TREC-8), 1999.

[8] Zemberek – A Turkish NLP Library, https://code.google.com/p/zemberek/