Search by Question

Mehmet Şerefoğlu– Koray Danışma– Ali Dorukhan Karakaya
Advisor: Dr. Engin Demir

Çankaya University, Department of Computer Engineering

Abstract

Finding correct answers to queries in search engines is still a problem as most of the search engines do search by keyword based interfaces. In this project, a question and answer type of intelligent search engine prototype is developed by exploring the data collection using interactive filtering to learn the actual intent of the user. Then, search by question system finds results of asked 5w1h question in the document collection and list them. Here, IBM Watson tool is used to process the data collection.

Introduction

Search by Question is basis of semantic analysis that try to answer 5w1h questions. The main point here is natural language processing, NLP for short, is a method for system to analyze, understand, and make meaning from human language in a smart way. To produce a solution, NLP should be used efficiency and develop an useful algorithm to extract the possible answer(s) in returned data(s) from IBM Watson.

Solution

First step of solution process is getting the entered query from user interface. If it could be pass the question check algorithm (then it is ensured that query is a 5w1h question), IBM Watson Discovery returns related passage that includes possible answers from data collection. This step is passage retrieval step. As a question term, search by question system understand the type of answer is called entity type. For instance, if a question term is "when", then the question type should be "date". In returned passages from IBM Watson Discovery tool, first sentence(s) that contain(s) subject (subject is found by IBM Watson Discovery) is/are taken. After that, we move on in each sentence from left of the subject and right of the subject to find related entity type at the same time. Thanks to IBM Watson Natural Language Understanding tool that is trained by Search by Question development team, system could understand and find the related entity type. However, it is not enough to return the found word(s) which is trained as a related entity type. Because in same sentence there could be more than one word with related entity type. For example, for "date" entity type again, if user ask the born date of subject, we get the smallest date and return it as the answer. In addition to, IBM Watson Discovery could find the verb in related passages. If there is "born" in sentence, we return the first date as the answer to user.
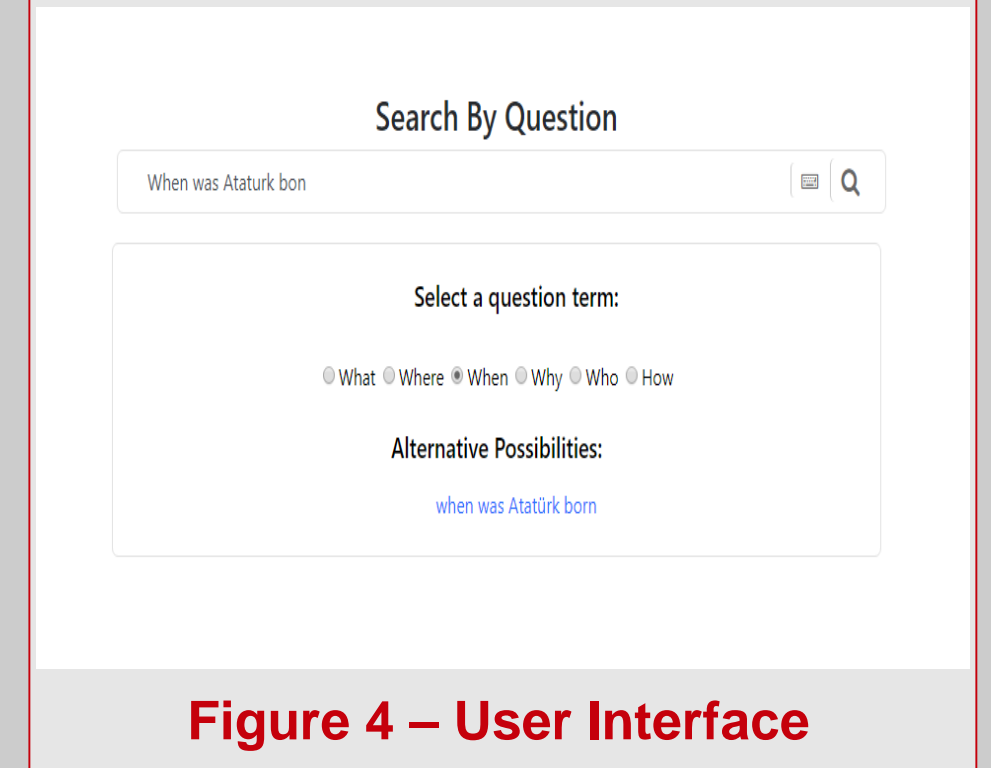

Figure 1 - Flowchart

PLACE FOR
TEAM AND ADVISOR
PHOTO

Figure 2 – Finished Product

Figure 3 – Train the NLU

Figure 4 – User Interface

Results & Conclusion

The last section describing your work should be an experimental results section showing textual or graphical output obtained from running your programs. You can include some facets of the implementation here but do not swamp the reader with code.

State what you learned from your work. Summarize what you did. State what you learned. State future work and directions, and then list any open problems.
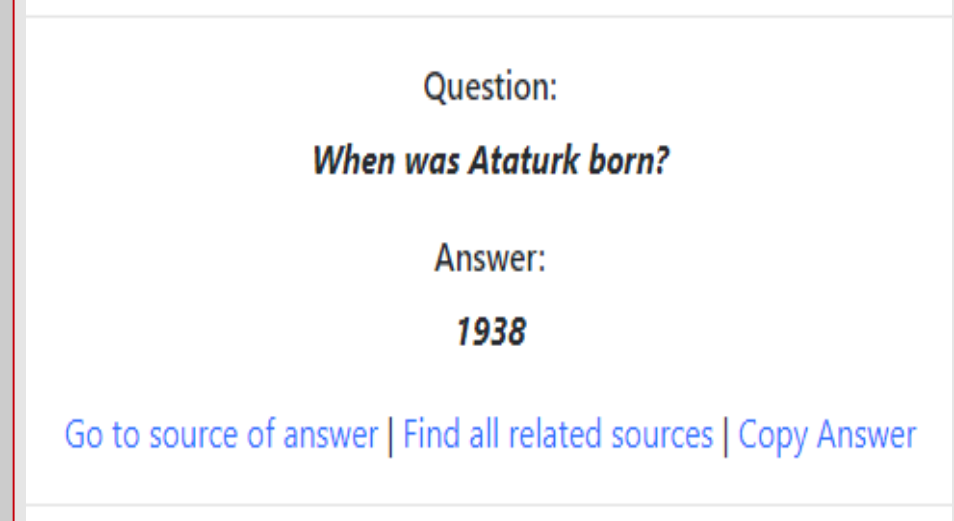
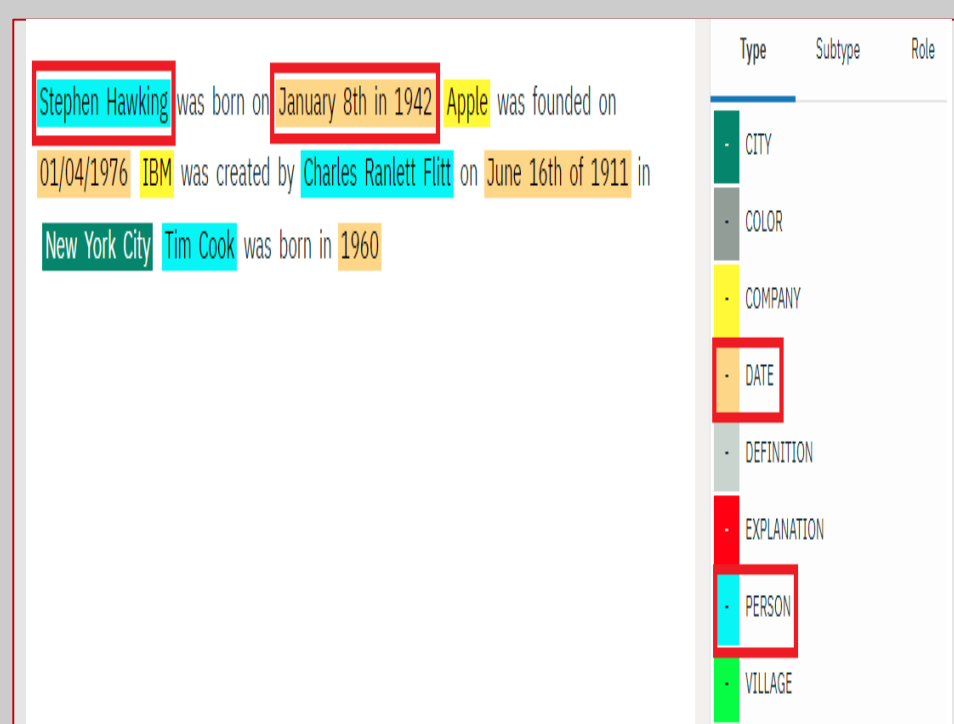## Acknowledgement

It has been a great opportunity to gain lots of experience in real time projects, followed by the knowledge of how to actually design and analyze real projects. For that we want to thank all the people who made it possible for students like us. Thanks to also our advisor Dr. Engin Demir for the efforts he did to provide us with all useful information and making the path clear for the team to implement all the education periods in real time project design and analysis.