



# Search by Question (The SbQ)

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 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

The SbQ 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 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 from IBM Watson.

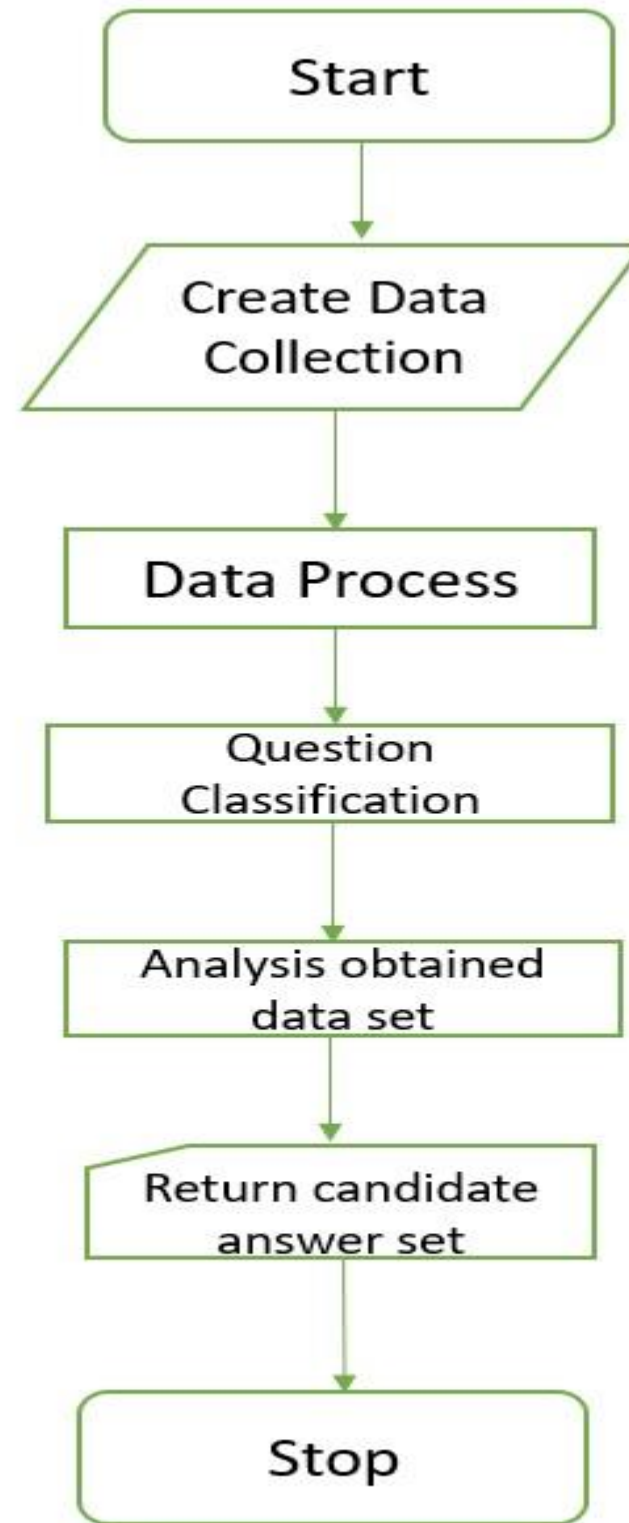


Figure 2 - Flowchart

## Solution

Firstly, Data collection (set of documents) is created in IBM Watson. Then, step 1 is data processing. Data that are in the created document set are tagged by their entity type such as person entity type for Ataturk or company entity type for Microsoft. The data model is trained by these tagged entity types.

Step 2 is question classification. Question is parsed to realize which entity type(s) could be contain the possible answer(s). The relevant information about question is proceed by IBM Watson and then we obtain the result data set.

Step 3 is analysis the obtained result data set. Part(s) which contains subject of the question are found in the result data set. Then, developed question answering algorithm that, move on in each place from left of the subject and right of the subject at the same time to find related entity type, is run. All found data related with searched entity type are checked for suitability to define it as possible answer. For instance, if we search “date” entity type since the question is asked the born date of subject with “when”, and also there are 2 same entity type in the same part, we define the smallest date as possible answer. Finally, candidate answer set is returned to user.

### Search By Question

When was Ataturk born

Select a question term:

What Where When Why Who How

Alternative Possibilities:

when was Ataturk born

Figure 1 – User Interface

### Search By Question

Question:

**When was Ataturk born?**

Answer:

**1938**

[Go to source of answer](#) | [Find all related sources](#) | [Copy Answer](#)

Figure 3 – Finished Product

## Contributions

Over the past few years, massive quantity of world knowledge have been collected in openly available knowledge bases. In addition to, these knowledge bases are seriously incomplete. In this project, we follow a way of using document collections to fill in the gaps in knowledge bases.

In industry, when companies are in research process, they want to find informations that directly related with their area despite all documents that contain their research key words. The SbQ provides that opportunity for companies.

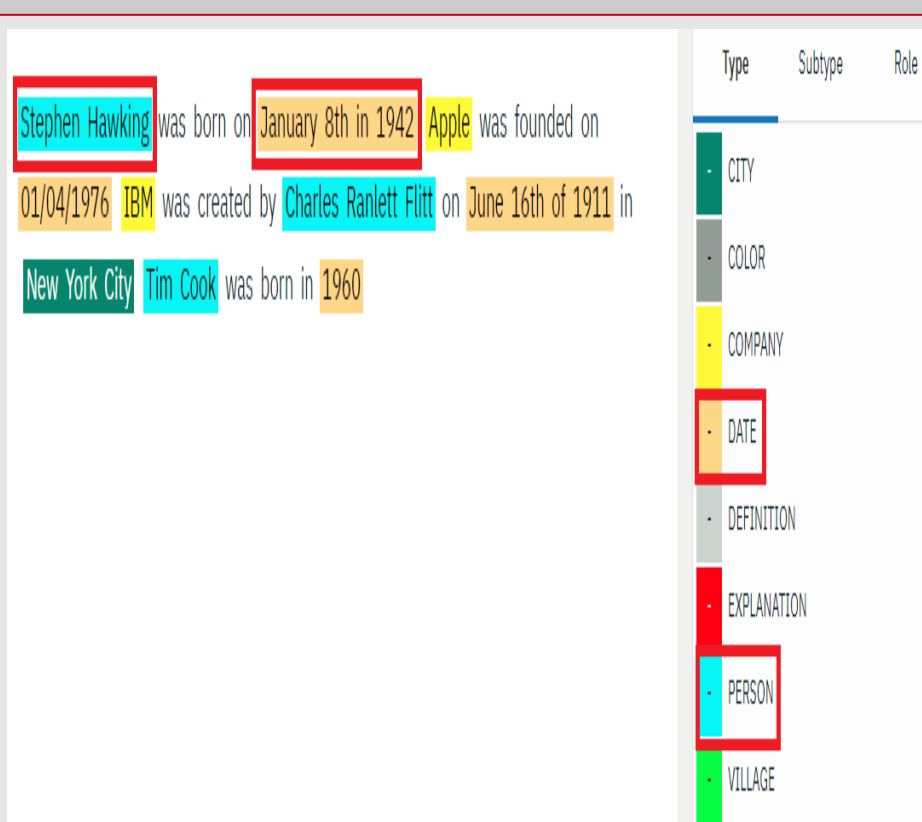


Figure 4 – Tagged data by their entity type

## Results & Conclusion

Search by Question makes searching process easier especially for industry. It provides making filter for researching area that is defined by user with creating a data collection. Moreover, it tries to find directly answer of asked question instead of show all document.

This project is seriously helpful for us to learn how to work with unstructured datas, analyze the machine learning algorithm and decide which one is appropriate to use in project, how to work in Natural Language Understanding process, how computers do semantic analysis and how we should train them.

In addition to, we learned how to work with web based program with using technologies that are php and nodejs.