First, a document retrieval hashmap is created to map a non-stop word in a document to the index of the document in the corpus. When given a keyword from a question, only the relevant documents containing the keywords are examined. One issue with this is that the tokenization of words may not be entirely accurate as the articles are not formatted perfectly. However, almost all words are captured correctly.

When given a question, it is classified into its various types. For example, if a question begins with “Who” or “Whom” then the answer should contain a person. However, this is not robust as not all question types can be captured. The accuracy of this question type classification method is sensitive to the quantity of situations captured.

Keywords are also extracted from the question. The keywords extracted are all non-stop words of a quoted expression, name entities, and nouns and their adjectival modifiers. A part-of-speech tagger is used on the sentence before extracting keywords so that only nouns and adjectives are extracted. Sometimes, verbs are incorrectly classified as nouns and adjectives are incorrectly classified as verbs.

With documents selected based on question keywords, the okapi BM25 algorithm can be applied to determine how relevant a document is to the given question based on a determined score. The top 5 documents with the highest scores are then chosen to extract answers from. 5 is not too many so that irrelevant documents are selected and not too few so that answers are limited. However, a perfect number of documents to use can not be determined as that depends on how general or specific a given question is.

Among the 5 documents, the sentence with the highest cosine similarity to the question is chosen to extract an answer from. The answers extracted are the determined entities based on the question type.

Overall, all the issues explained above lead to inaccuracies in answering. For questions such as “Which companies went bankrupt?”, organizations extracted from a sentence are not always companies. This can be solved by including an accurate company classifier. More weight also needs to be placed on the word “bankrupt” when scoring sentences. Questions related to “What affects GDP?” are also very general and can not have a specific question type classification. This can be solved by having “affects GDP” as a more heavily weighted bigram feature in determining similarity of sentences. However, the QA system can obtain accurate answers for “Who is the CEO of company X?” as that question type has a small number of keywords and requires a person to be in the answer sentence.