Sample Output:

Q1: Which company went bankrupt in 2009 January?

Answer: Lehman Brothers Holdings Inc.

Q1: Which company went bankrupt in 2008 June?

Answer: First Republic

Q1: Which company went bankrupt in 2008 December?

Answer: GITIC

Q2: What affects GDP? What percentage of drop or increase is associated with this property?

Answer1: Stock market earnings, 3%

Answer2: Tax, 1%

Answer3: Fiscal tightening, 1%

Q3: Who is the CEO of company Gap Inc.?

Answer: Glenn Murphy

Q3: Who is the CEO of company Fisker?

Answer: Tony Posawatz

Q3: Who is the CEO of company Sears Holdings?

Answer: Eddie Lampert

Instructions: Type in the Question line manually and the answer in the sentence will pop out after several commands, then extract the keywords from answer sentence.

Firstly determine the question type by figuring out what is it exactly requiring, for example if asking what companies, then we extract the company list and trace them back to the processed text (tokenization, stemmed, removing the stop words and so on) and export the whole sentence within the company name and scoring them separately.

The score process should be 1) count the adjacent words appearing in both question and candidate answer. Add the tf-idf of the words shared in both, while subtract the tf-idf of the words not the same in both. The sum should be the total score. By selecting the top score, the candidate should be high likely the best answer.

By using the regular expression and extract the key answer to be the final answer. During which the corpus should be processed and being indexed thus to create a document term matrix for the convenience of search, by selecting the answer the similarity measures should also be taken into consideration.

The question about the GDP is very frequently brought up to the articles in the dataset, so I made a chart to find out the most influencer to the GDP.

Here list the five highest influencer for GDP according to the articles. To sum up, GDP can be easily affected by Industrial production, retail and fixed asset investment, and most of which is based on the policy published by the government.