

Executive Summary

Yuxiang (Alvin) Chen

This Q&A system developed could take in 4 types of questions to answer about company bankruptcy, GDP factors and how they affected GDP and a company's CEO name using 730 business articles from 2013 and 2014. The Q&A system could take in a question and classify it as 1 of the 4 types or identify it as an unidentifiable question and notify the user. Through experiments and observations during development, I found that accuracy is important and it's more important for a good Q&A system to provide accurate answers than to be able to answer all questions. Thus, this Q&A system pays more attention to give qualified answers. Some questions that may have answers in the database may lead to an answer of 'not found in current database' to indicate the system is not capable of answering the question. It is also found that the distance between the keyword of question and the actual answer entity plays a big role in the confidence of finding the answer so distance is used extensively in this Q&A system algorithm development. Also, it is found that sometimes, keyword appearance does not mean the sentence has valid information to answer the question. For example, 'bankrupt' could appear in a sentence about a plan for filing bankruptcy or in a 'if then' statement. Thus, the past tense of words such as 'declare', 'file' is an important indicator of whether the sentence is giving information about a happened bankruptcy. Discoveries and findings like these help the development of this Q&A system. Although the final Q&A system may not be able to discover all bankrupted companies, find a CEO even if theoretically it is available from the article dataset used, the final Q&A system developed is ensured to provide answers with high confidence so the users could get a qualified answer if an exact entity is found and presented.