Steffany Bahamon

IEMS 308, F16

Question and Answering System Homework

<u>Executive Summary</u>

My objective was to create a question and answer system that would answer simple factoid questions about GDP, CEO's, and companies. I attempted to use a purely python implementation of an indexing engine to create a querying system, but was unable to implement it properly on my computer.

<u>Problem Statement</u>

I aim to create a simple text based question and answering system for Business Insider articles from 2013 and 2014

<u>Methodology</u>

The crux of this system revolves around the Whoosh python package, which is a search engine package purely in Python meant to compete with Elasticsearch, Solr, and Lucune. In addition,it has a querying system based on the Okapi modeled we discussed in class. However, I could not set up it or any of the either search engines listed above functioning properly on my computer, so I continues to construct the system almost as a theoretical exercise since I wouldn't be able to get the most basic part of it (indexing the corpus) working. I extracted keywords from the queries by stripping them of stop words, then running the stripped phrases through the query parser provided with Whoosh. Whoosh then proceeds to rank the results by article using an Okapi score. Using a provided fragmenter, I contextually remove the phrase inside the document with the highest score. This QA system doesn't take into account type of question or respond with a complex answer, but should function appropriately.