Brian Gee  
IEMS 308 HW 4: QA Systems Report

**Executive Summary (Findings)**

My QA model distinguishes between people, CEOs within people, nouns, and organizations. The quality of outputs could be improved as the system still behaves in ways consistent with low IQs. It is not intelligent enough to distinguish between two names in one sentence, often it returns both along with the sentence. For nouns, it returns every noun in sentences within documents that TF-IDF scores highly – but there are still many extraneous/unimportant entities extracted. For organizations, the model has trouble applying qualifiers (such as “bankrupt” or “profitable”) in its search and will often prioritize just one or two of the keywords in the input question; this results in outputs that sometimes just contain general organizations while missing the specific types of organizations queried

Overall, this model could be improved in these areas before future deployment. After these improvements are made, it can be used to identify organizations and CEOs as they pertain to different strategic topics.

**Instructions on Using the QA system**

Start by downloading the whole folder of files. The smaller essential files and folders for running the QA system are “QA.py”, “HW 4 QA Systems (Build).py”, and “Saved Objects” are inside the “Homework 4” folder. The “Data” folder should be empty, place all 730 text files from Business Insider into that folder. In order to be able to run everything you will first need to generate the required pickle files by running the “HW 4 QA Systems (Build).py” script. This will give you will have the necessary files needed for the “QA.py” script.

Prior to running the python file, ensure that your working directory is set to the root of the “Homework 4” folder.

To use the This QA system can be run from the command line. Assuming you already have Python 3 installed, and your working directory is set to the root of “Homework 4”, type in “python QA.py” and wait for the input prompt, “What is your question?”, to pop up. Once you see that prompt, you may type in your question and hit enter. The **system takes approximately 4 minutes to run** and outputs answers to your question (drawn from a database containing 2 years’ worth of Business Insider articles).

**Answers are outputted in the form:**

**“Answer”  
“Sentence containing the answer”**

I chose to keep the answers in this form to address the inaccuracy resulting from the nature of TF-IDF and scoring functions relying on keywords and their proximity to each other. For example, in the question “Who is the CEO of Facebook?” there is a sentence that contains both “Facebook CEO” and “Microsoft CEO” in the same sentence, which the system can’t distinguish between the two CEOs, so it outputs them both. Including sentences in the output provides context for each answer so the user won’t be misinformed.

For your reference there are also iPython Jupyter notebooks included in the submission. The notebook labeled “HW 4 QA Systems (Build)” was used to create pickle files for TF-IDF pandas DataFrames to avoid recreating them every time the code ran. The notebook labeled “HW 4 QA Systems (Run)” is everything essential for running the model.

**Sample Outputs:**

Type 1: Who is the CEO of \_\_\_\_\_?

* Question: **“**Who is the CEO of Facebook?”  
  Answers:
  + Steve Ballmer  
    Determined to keep Facebook away from Google, Microsoft CEO Steve Ballmer offered to buy the company.
  + Mark Zuckerberg

As it turns out, CEO Mark Zuckerberg had other plans for continuing Facebook's outstanding growth that didn't require more ads, and based on some recent data, it appears those plans are working even better than expected.

* Question: “Who is the CEO of JP Morgan?  
  Answers:
  + Jamie Dimon

(Photo by Mario Tama/Getty Images) JP Morgan is sometimes called 'The House of Dimon' for the all encompassing presence of its CEO, Jamie Dimon.

* Question: “Who is the CEO of Walmart?”   
  Answers:
  + Mike Duke

Walmart President and CEO Mike Duke tried to assure investors, however, that "no matter what environment we are in, Walmart will win."

* + Bill Simon

Lagging demand for electronics, such as big screen televisions, has been a drag on sales, Walmart U.S. CEO Bill Simon said at the investors' meeting Tuesday.Â  Meat, produce, home and apparel continue to be Walmart's best performing categories.Â  That has led to a problem among Walmart's discount stores, which don't offer the grocery options of its supercenters, Simon said.

Type 2: Which companies \_\_\_\_\_?

* Question: “Which companies went bankrupt in March of 2013?”  
  Answers:
  + Sunlot

"The company negotiated the settlement as part of its effort to take over the bankrupt exchange before a court-ordered liquidation," Sunlot's statement said.

* Question: “Which companies went bankrupt in 2008?”

Answers:

* + Lehman Brothers

Most will remember September 2008, which was when the credit crisis got really ugly as Lehman Brothers went bankrupt and interest rates surged.

* Question: Which companies were acquired in 2013?  
  Answers:
  + Chiquita

In pre-market trade on Thursday, shares of Chiquita were up nearly 8% after Cutrale-Safra offered to acquire the company for $14.50 per share.

Chiquita

* + Earlier this week, Chiquita issued a lengthy statement responding to what it called "inaccurate and misleading statements," made by Cutrale and Safra in their efforts to acquire the company.
  + AbbVie

2) In mergers &amp; acquisitions news, AbbVie agreed to acquire Shire in a deal worth $53 billion that will allow the company to move its tax base to Jersey, an island in the English Channel.

* + Shire

2) In mergers &amp; acquisitions news, AbbVie agreed to acquire Shire in a deal worth $53 billion that will allow the company to move its tax base to Jersey, an island in the English Channel.

* Question: Which companies’ stocks grew in 2013?

Answers:

* + Romero

However , spite denying had even heard CYNK prior the firestorm its stock price surged , Romero indicated was `` proposed transaction '' would involved purchasing shares the company 's prior CEO , Marlon Luis Sanchez .

* + Regeneron &

The other companies with expensive valuations read like a Who's Who of so-called momentum stocks, including Regeneron &lt;REGN.O&gt;, Alexion, TripAdvisor and Vertex Pharmaceuticals &lt;VRTX.O&gt;.

* + TripAdvisor

The other companies with expensive valuations read like a Who's Who of so-called momentum stocks, including Regeneron &lt;REGN.O&gt;, Alexion, TripAdvisor and Vertex Pharmaceuticals &lt;VRTX.O&gt;.

* + CYNK.Some

site.introbiz.comA screenshot of the homepage of IntroBiz, the social network associated with CYNK.Some of the people associated with CYNK, the mysterious social networking company that generated headlines last week when its stock surged 25,000% in just 16 trading days, have worked together before.

* + Romero

Interestingly, the June 18 date Romero identified as Berkowitz's start with the company coincides almost exactly with the surge in CYNK's stock price.

* + CYNK

Interestingly, the June 18 date Romero identified as Berkowitz's start with the company coincides almost exactly with the surge in CYNK's stock price.

* + SEC

In 2000, he was hit with a $11,000 federal court judgment after the SEC filed a complaint accusing him of "passing himself off as an independent and impartial stock picker when, in fact, he is nothing more than a paid pitch man for the companies he hypes."

* + Romero

However, in spite of denying he had even heard of CYNK prior to the firestorm after its stock price surged, Romero indicated there was a "proposed transaction" that would have involved him purchasing shares from the company's prior CEO, Marlon Luis Sanchez.

Type 3: What \_\_\_\_?

* Question: “What affects GDP?”

Answers:

* + stocks

Historically , stocks posted most consistent gains GDP been around 3 % .

* + Democratic presidents

It found that the average real GDP growth rate under Democratic presidents was 4.35%, a marked increase from a 2.54% growth under Republicans.

* + Republicans

It found that the average real GDP growth rate under Democratic presidents was 4.35%, a marked increase from a 2.54% growth under Republicans.

* + solid double-digit stock market gains

Contrary to conventional wisdom, and what may be a surprise to those who see low single-digit rates of gross domestic product (GDP) growth as incompatible with solid double-digit stock market gains, GDP does not have to be booming to produce solid gains in the stock market — as 2013 can attest.

* Question: “What percentage drop or increase is associated with GDP?”

Answers:

* + 2.3%

— Analysts Underestimate Earnings Around Turning Points: Economists forecast a 2.3% YOY increase in GDP in 4Q.

* Question: “What is the projected growth of the US economy?”

Answers:

* + Even South Africa

Even South Africa , Credit Suisse dubbed most troubled emerging market economy a year ago , see currency stabilize , fewer labor strikes , higher industrial output , a doubling GDP growth 1.2 percent 2.4 percent .

* Question: “What percent has unemployment changed by from 2008 to 2014?”

Answers:

* + 6 percent

Bernanke also stated that 2016 was a reasonable estimate for how long it would take to reach a 6 percent unemployment rate.

* + 7 percent

Hiring has picked up, helping to push the jobless rate to a five-year low of 7 percent in November, but some 11 million Americans are still unemployed, and many are earning less than before the recession.

* + 7 percent

Many of them look at signs of economic growth and an unemployment rate now down to 7 percent and expected to drop further as evidence the additional weeks of benefits are no longer necessary.