# Findings & Business Values

**Yuxiang (Alvin) Chen**

## Development Findings

I will separate my development findings into the 4 main question types.

### Which companies went bankrupt in month X of year Y?

1.  Keywords that are related with bankruptcy include ['bankrupt', 'bankruptcy', 'Chapter', 'liquidation', 'liquidate']

2.  Many sentences include >= 1 of these keywords. However, the sentence may be about planned bankruptcy or the bankruptcy keyword may be in an 'if… then…' statement (may not happen at the specified time in the end) so I sued past tense detection to further confirm it is a bankruptcy that happened.

3.  There are three main ways to identify the time of a mentioned bankruptcy that actually happened.
    a.  The sentence specified a year and month of the happening (such as for Layman Brothers bankruptcy.
    b.  The sentence specified a past month (compared to the article date) but no year information. This means the year probably is the year of the article
    c.  The sentence specified no month or year. This means the bankruptcy happened recently.

4.  Two company names such as X & X Holdings may indicate two different companies. Usually X Holdings has multiple sub-companies and one of them is X.

### What affects GDP?

1.  The factors of GDP are usually nouns. However, it's hard to automatically identify most of them without including nouns that are around GDP but not an affecting factor of GDP.

2.  This question is a fixed question without variables and also it is a question that will lead to further questions. Thus, it is a good approach to provide a fixed answer with each factor also a good valid input for the follow-up question.

**What percentage of drop or increase is associated with Z**

1. Words that usually indicate 'increase in GDP' - ['up', 'increase', 'rise', 'add']

2. Words that usually indicate 'decrease in GDP' - ['down', 'decrease', 'drop']

3. Words that does not give a change percentage of GDP but confusing so should be excluded - ['up to', 'down to']

4. A common structure that should be excluded because it does not indicate change percentage - '{percentage} of GDP'

5. The distance between the percentage found and the GDP word is very important because many percentages could appear in the sentence.

**Who is the CEO of company X**

1. To make sure the answer (if the answer is not 'not found') provided by the system is correct and accurate with high confidence and quality, I only use 'CEO' or 'chief executive officer' as keyword instead of using my algorithm in Assignment 3.

2. In most cases, the CEO's name is very close to the keyword so the distance between a person's name and the CEO keyword is important. I try to look for the closest over all articles in the dataset.

3. CEOs change for a company. Thus, it's important to provide the source article's time. Some companies I used in my experiment do have a CEO change in early 2014.

## Business Insights and Value

This Q&A system has three main functions:

1. Month + Year -> Companies bankrupted in the month of the year
2. Factor -> How the factor affected GDP
3. Company -> CEO of the company

This system and the algorithm could be useful for 2 main groups of user:

1. Companies which need to get a large amount of this sort of information from large amounts of text data sources such as a market research company.

2. Individual users who would like to search the answer to these types of questions. This system could be used in an online knowledge database that uses article dataset.