ITCS 6100 – Big Data for Competitive Advantage
Data Extraction and Data Wrangling
Executive Summary Report

The Big 3 patent office's i.e. US, Europe and Japan,grants and publishes patent filed by a lot of organizations out of which the fortune 500 company patents are a key interest in our project. These companies will fund US Presidential Elections in favor of both the Democrats and Republicans as they will need the support of the ruling government in publishing their patent. The patent publications and funding the election campaigns directly impact the annual revenue generated by the fortune 500.

Data Extraction

We collected US and European patent grants of fortune 500 companies published from 2005 to 2015 downloaded from different sources and archives. The annual sales i.e. revenue of fortune 500 companies is extracted for the year 2005 to 2014 and US Presidential Election outcomes from 1940 to 2012. We also concentrated on what happened post-election like the list of bills, amendments passed by a ruling party in their period. Moreover, the monetary funding statistics for the election campaigns by the top 50 companies has been extracted from 2000 to 2015.

Insights from Data

Now, we have the data and we need to make use of it in such a way to deduce some knowledge and information from the data and make predictions. The information retrieved includes number of patents published by the fortune 500 companies in a year, total revenue generated and so on which could help to answer some of the following questions: -

- Is there any relationship between the annual revenue generated and its total number of patent publication in each year?
- How much the organizations are spending on the election campaigns?
- Why do organizations need to fund them and what's the correlation between the patent publications and the ruling government?
- Who benefits more – the elected government or the organizations funding them?
- How a company can affect its competitors in the same industry by favoring the government?

These are some of the critical questions that the data can answer when analyzed deeply and can bring more insights.

Knowledge Extraction from Data

Advanced tools and techniques are used to do high level calculations and extract relevant information from raw data. This information can help to predict the correlation of patent publications and annual revenue with the government that will be elected in 2016 and can answer many of the questions based on the data we have.