****

**Rapid Integrated Assessment (RIA) of National Planning Documents**

Phase I Report

Team: Brahms

Date: 02/01/2016

**Phase I Summary**

1. Successfully transformed 46 PDF documents of four countries into text files with automated program in Python;
2. Used Python packages, i.e. NLTK, Gensim, Sklearn, to preprocessing the Bhutan’s Five Year National Plans
   1. Extracted complete sentences
   2. Tokenized each sentences
   3. Removing “stop words”
   4. Stemming
3. Researched the most recent semantic similarity academic papers to find best solutions.
   1. Traditional machine learning: word2vec, feature selections, etc.
   2. Deep learning: Recurrent Neural Networks, Long Short-term Memory
4. Build dictionaries for each SDG goals to enrich the training dataset
5. Data output and core codes:

<https://github.com/ICT4SD/Rapid_Assessment_Tools/tree/Fordham-Brahms>

**Rapid Integrated Assessment (RIA) Of National Planning Documents**

**Project Objectives:**

The project is going to analyze the impact of companies’ earning calls on their stock prices. We will mainly focus on the stock price fluctuation before and after the earning calls. The time intervals that we are going to test are 10 mins, 30 mins, and 60mins.

**Project Data Source:**

Earning Call Transcripts: **LexisNexis Database**

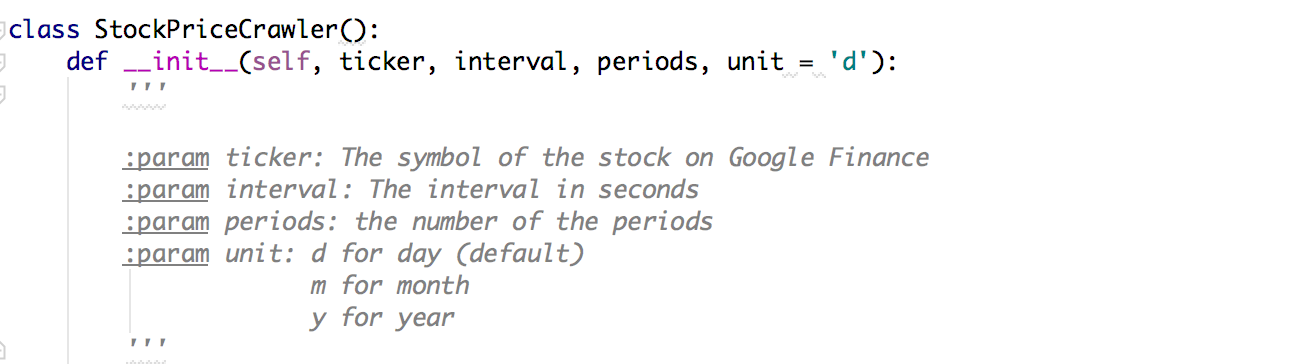
Stock Price: **Google Finance**

Tools: IBM Watson Tone Analyzer, Python

**Project Workflow**:

**Current Status:**

* Building Python crawling codes, which enable users to choose ticker, intervals and period of time:



* Sample Data:

