

---

## ### Supplementary Materials ###

This README is designed to layout the contents of the supplementary materials folder and give instructions on running the demonstration code. Completed by Lily Garfinkel, Joe Scheufele, Ben Staw, and Wayne Whichard.

### # Passive\_Data\_Practices\_Literature\_Review.pdf

This file contains the team's literature review of passive data collection practices.

### # Current\_System\_SWOT.png

This file is a graphical summarization of the team's SWOT analysis conducted with USPTO employees.

### # Detailed\_SWOT\_Analysis.xlsx

This file adds details to the SWOT analysis that were left out of the synthesis.

### # Evaluation\_Matrix.png

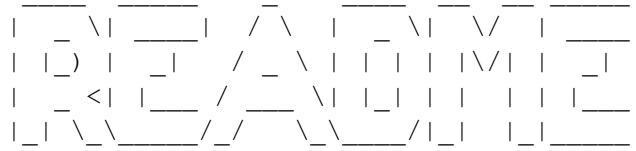
This file ranks each of the analysis strategies against each other on a variety of criteria, with 1 being the highest ranked and 6 being the lowest.

### # Detailed\_Eval\_Matrix.pdf

This file goes in more depth about why each strategy was ranked how it was.

### # Implementation\_Flowchart.pdf

This file shows the thought process for deciding the best analysis strategy to use for a given context.



---

# USPTO\_Interview\_Questions.pdf

This file contains the blank interview questions given to the USPTO employees.

# Non\_USPTO\_Interview\_Questions.pdf

This file contains the blank interview questions given to the interviewees not associated with the USPTO.

# Patent\_Application\_Process\_Flowchart.jpg

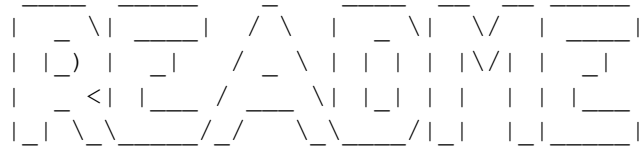
This file shows the process of obtaining a patent from start to finish.

# Presentation.pptx

This file contains the group's final presentation slides.

# Glossary.pdf

This file contains careful descriptions of words chosen to be beyond the scope of the project.



---

## Demonstration Code ##

## CONTENTS

Web\_Scraper\_Demo.zip

Web\_Scraper\_Demo/backup.txt

    /BackUpWordCloud.png

    /example.txt

    /scrapy.cfg

Web\_Scraper\_Demo/demo/items.py

    /lisaExample.py

    /middleware.py

    /pipelines.py

    /settings.py

## REQUIREMENTS

Python version 3.6+

    To download python follow the instructions from:

<https://www.python.org/downloads>

Scrapy version 2.4+

    Once python is installed scrapy can be installed with:

    >> pip install scrapy

PLSA version 0.6+

    >> pip install plsa

Matplotlib version 3.3+

```
>> pip install matplotlib
```

## Instructions

Unzip the folder, then open a command prompt at the folder location: Web\_Scraper\_Demo

To run the web crawler simply enter:

```
>> Scrapy crawl example
```

This will run the web scraper, print some diagnostic information and will write all the found text into the text file 'example.txt'.

To run the analysis program, enter:

```
>> python demo/lsaExample.py
```

This will print the first list of words with their importance within each topic. It will also print the most important words from each of the 6 sources. Lastly it displays a matplotlib figure with 6 word clouds. The size of the words is relative to the importance with each topic. To end the program, close the matplotlib figure.

To make sure the program is running correctly, 'example.txt' and the resulting word cloud can be compared with 'backup.txt' and 'BackUpWordCloud.png' respectively.