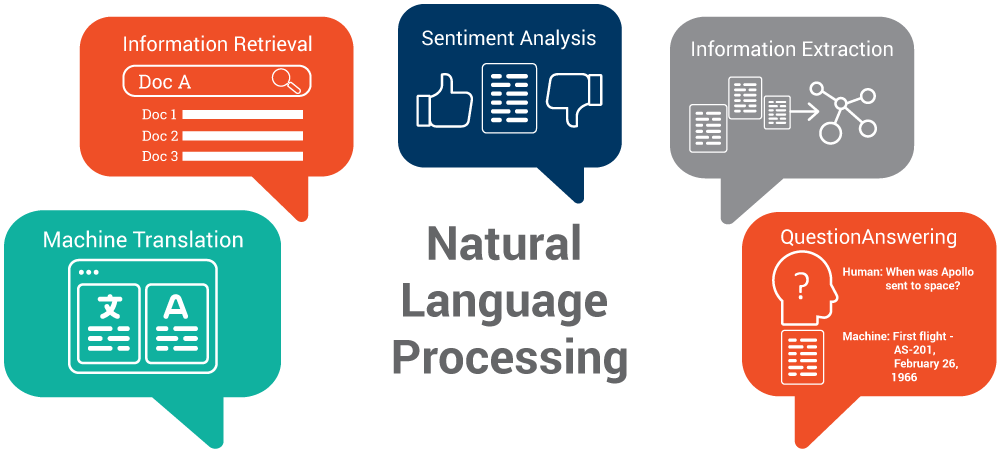
## horizontal line

**Data Extraction and NLP**

09.04.20XX

**─**

Your Name

Your Company

123 Your Street

Your City, ST 12345

# The problem

Extracting data from websites and performing text analysis and sentiment analysis

# Our solution

1. First i started exploring the objective.txt which problem statement and solution was explained
2. After the exploration started crawling the data from the given web urls using beautifulsoup library and stored all the text files in the txt format in a data/processed directory.To execute the operation of crawling data from websites run the src/scraper.py file.
3. Before running the .py file install all the dependencies mentioned in requirements.txt file.

CLI : pip install -r requirements.txt

1. Stored all my extracted text files in a folder with the similar url\_id’s name for each . Next I started text and sentiment analysis.
2. Text analysis started with removing stop words and tokenization of words from the text files using the regular expression ‘re’ module.
3. Approached text analysis.docs file to perform analysis mentioned in the Text analysis file.
4. Followed all the instructions given in the Text analysis.doc and performed one by one all the contents using python programming language.

# Generating output from my assessment project

Step-1. Install all the dependencies from requirements.txt

Cli : pip install -r requirements.txt.

Step2. run scraper.py file to generate the text from web urls

Cli : python src/scraper.py

Step-3. After generating the text files from web urls run analysis.py file to perform sentimental analysis according to the content provided in the text analysis document.

Cli : python src/analysis.py

Step-4. You can see all the logs in log directory log/scraper.logs and analysis.logs and all the output scores are stored in output directory output data structure.xlsx file.

## Utilized python3 to extract the data from URLS and TEXT ANALYSIS, beautifulsoup python library for data crawling.

.

# Thank you.