**Tech for Good - Brief**

**Overall Goal**

To build a web application which can scan content online (e.g. news articles) and provide insights into the intent/sentiment behind them. In the world of mass media, it can be difficult to remain aware of the validity or intent behind the content you are consuming.

**Personas**

User – Wants to get a better understanding of the content they are consuming online and whether it comes from reliable sources and/or has underlying biases.

[ANA] Are we going to define a TYPE of content? E.g news articles, hotel reviews, etc?

**Features**

As a User I want to be able to:

• Input a website URL or upload a document to be reviewed

[ANA] OK, Python library

• View the output analysis

[ANA] Some Ideas:

1. Prediction???
2. Search for a specific word, or set of words (OK – function “context\_search”). User enters what he wants to analyse (e.g. opinion about “Brexit”) and the APP will return all snippets of text in that website contains information about it.
3. Filter all Adjectives, or all Adverbs, or all Nouns (OK – function “preprocess\_pos”)
4. Sort the word frequency map by word counts (OK – function “collect\_word\_map”) (???)

Analysis may include (but not limited to):

• Highlighting “Interesting” words or statements

• Use of hyperbole *(***Hyperbole***, from a Greek word meaning "excess," is a figure of speech that uses extreme exaggeration to make a point or show emphasis. It is the opposite of understatement. ... Hyperboles are not comparisons, like similes and metaphors, but extravagant and even ridiculous overstatements, not meant to be taken literally.)*

• Tone of article (e.g. left-wing/right-wing)

**Stretch Goal Features**

• Use of Machine Learning [OK]

• Cross-analysis of multiple documents

[ANA] Use of document similarity: You can find reviews similar in content and sentiment to yours by applying the similarity function to the whole review and finding the other review with the overall highest similarity score to the input one [REVIEWS???]

• Use of RSS

[ANA] Some notes:

<https://rss.softwaregarden.com/aboutrss.html>

**What information does RSS provide?**

RSS provides very basic information to do its notification. It is made up of a list of items presented in order from newest to oldest. Each item usually consists of a simple title describing the item along with a more complete description and a link to a web page with the actual information being described. Sometimes this description is the full information you want to read (such as the content of a weblog post) and sometimes it is just a summary.

For example, the RSS information for headlines on a local news website could contain the following information:

|  |  |
| --- | --- |
| **Item 1:** |  |
| *Title:* | Sidewalk contract awarded |
| *Description:* | The city awarded the sidewalk contract to Smith Associates. This hotly contested deal is worth $1.2 million. |
| *Link:* | http://www.gardencitynews.com/contractawards/sidewalk.htm |
| **Item 2:** |  |
| *Title:* | Governor to visit |
| *Description:* | The governor is scheduled to visit the city on July 1st. This is the first visit since the election two years ago. The mayor is planning a big reception. |
| *Link:* | http://www.gardencitynews.com/news/2004/06/gov-visit.htm |

The RSS information is placed into a single file on a website in a manner similar to normal web pages. However, the information is coded in the XML computer language for use by a program (the RSS aggregator) and not by a person like a normal web page.

• Sentiment Analysis

Points of Contact

David Hills – Senior Software Engineer – [david.hills@eigentech.com](mailto:david.hills@eigentech.com)

Rebeka Marton – Engineering Manager – [rebeka.marton@eigentech.com](mailto:rebeka.marton@eigentech.com)

Document Analysis:

1. Sentiment Analysis

The sentiment property returns a namedtuple of the form Sentiment(polarity, subjectivity).

The polarity score is a float within the range [-1.0, 1.0].

The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.

from textblob import TextBlob

text = '''test 123'''

testimonial = TextBlob(text)

testimonial.sentiment

**Sentiment(polarity=0.39166666666666666, subjectivity=0.4357142857142857)**

1. Definition of a word

from textblob import Word

Word("python").definitions

**['large Old World boas',**

**'a soothsaying spirit or a person who is possessed by such a spirit',**

**'(Greek mythology) dragon killed by Apollo at Delphi']**

1. Spelling correction

b = TextBlob("I havv goood speling!")

print(b.correct())

**I have good spelling!**

1. Word and Noun Phrase Frequencies

Using the word\_counts dictionary.

monty = TextBlob("We are no longer the Knights who say Ni. "

"We are now the Knights who say Ekki ekki ekki PTANG.")

monty.word\_counts['ekki']

**3**

1. Translate a text

b = TextBlob("Simple is better than complex")

b.translate(to="es")

**TextBlob("Lo simple es mejor que lo complejo")**

1. Language Detection

b2 = TextBlob("Simple is better than complex")

b2.detect\_language()

**'en'**

1. X

Analysis may include (but not limited to):

- Highlighting “Interesting” words or statements

- Use of hyperbole

- Tone of article (e.g. left-wing/right-wing)

Issues:

1. How to get an Unformatted URL?