Specification:

Rate My Tweet: Understanding Comparative Judgement in the Wild

Andy Gray: 445348

## Overview

The main focus of the MSc project will be to create a human centric design focusing on comparative judgement and Twitter tweets. A web interface will be required to be able to conduct the actual comparative judgement and provide information about what comparative judgement is and how it works. Additionally, to provide the user with feedback in terms of the results. Additionally, NLP techniques will be used to be able to extract information about the tweets and see if there are any patterns within the tweets and the order of their popularity.

Aims

* To get an understanding of a few Key Technologies required for the PhD Project.
  + Comparative Judgement
  + NLP and Information Extraction
* To create a web app to be able to conduct the Comparative Judgement.
* Extract the key words from the tweets to see if there are any patterns in the tweets and their popularity.
* Provide some form of feedback and insights to the user on the results.

Potential Tools

* Python
* Python Web App
  + Flask: Flask is a micro [web framework](https://en.wikipedia.org/wiki/Web_framework) written in [Python](https://en.wikipedia.org/wiki/Python_(programming_language)). It is classified as a [microframework](https://en.wikipedia.org/wiki/Microframework) because it does not require particular tools or libraries.
  + Django: Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.
* NLP Libraries:
  + SpaCy: excels at large-scale information extraction tasks. It's written from the ground up in carefully memory-managed Cython.
  + NLTK: is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to [over 50 corpora and lexical resources](http://nltk.org/nltk_data/) such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries
* Visualisation Tools
  + Bokeh: Bokeh is a data visualization library in Python that provides high-performance interactive charts and plots. Bokeh output can be obtained in various mediums like notebook, html and server. It is possible to embed bokeh plots in Django and flask apps.
  + Plotly: is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases. Plotly allows users to import, copy and paste, or stream data to be analysed and visualized

Additional Tool

* IDE:
  + VS Code: Visual Studio Code is a [source-code editor](https://en.wikipedia.org/wiki/Source-code_editor) made by [Microsoft](https://en.wikipedia.org/wiki/Microsoft) for [Windows](https://en.wikipedia.org/wiki/Windows), [Linux](https://en.wikipedia.org/wiki/Linux) and [macOS](https://en.wikipedia.org/wiki/MacOS)
* AWS
  + Hosting
  + Mechanical Turk
* Task Tracking
  + Trello
  + https://toggl.com/

Evaluation

The project supervisory team comprising of Swansea University and CDSM representatives will evaluate the project regarding the user-centric design of the web application for comparative judgement and the information extracted with NLP of the tweets. The web application will be a prototype in regard to a greater system to seek the potential capabilities of comparative judgement. The aim will be to allow users to effectively compare all unique combinations of tweets to see which one they think is better. In result creating the comparative judgement score, which can then be used against other tweet metrics to see if the comparative judgement results match up to the tweets metrics, and to see if there are any patterns within the tweets which will be extracted through NLP. The results and decision-making output will be evaluated to determine insights has been successful by including human in the loop in its design.

Timeline

Application

Description automatically generated with medium confidence

Graphical user interface, application, table, timeline, Excel

Description automatically generated

Timeline

Description automatically generated