CS6109
Compiler Design

Project Proposal

Feature extraction on comments to analyze emotion using NLP

Team Members:

1. V. S. Suryaa - 2018103610
2. Srihari S. - 2018103601
3. Reuel Samuel Sam - 2018103053

Problem Addressed:

In this day and age, where the internet has taken control of most people's lives, social media exists as one of the most powerful platforms to share and voice one's opinions. People put up content looking for appreciation and constructive criticism from the general public to help improve their skills in their art. However these responses, usually in the form of comments, are at times more negative and positive and thereby may cause some form of emotional pain to the original content poster.

This project tries to tackle this situation head on by segregating the comments left on posts into 3 main categories - positive, neutral and negative. The comments are analyzed using a Lexicon based sentiment analysis approach that breaks down the comment to its constituent parts and provides the user an output with desirable filtered comments.

Abstract Idea

This project focuses on comments on social media posts and segregates them based on certain features. A comment mainly is made up of 2 parts - content and tags.

The content mainly refers to the reaction left to a post in forms of sentences or short phrases. For easier processing of an output, our lexical analyzer will remove the usage of proper nouns that do not contribute to the overall emotion of the post.

Furthermore, a comment can include tags. The main 2 types of tags are

1. Hashtags
    a. Depicted by '#' followed by a trending topic
2. Mentions
    a. Depicted by '@' followed by a username

We plan on segregating comments based on tags used and keep track of the usage of these tags along with a few banned tags to make sure none of the comments violate the guidelines.

Procedure

The program is planned to be divided into 3 main sections:

(i) Comment feature extraction
(ii) Database Population
(iii) Categorization

Step 1:

The comments are provided as inputs (most likely through a file) and are then analysed by the lexical component. The tags are placed in separate files and provided to the database section for the population of the database.

The comments, after removal of these tags, are analysed once again and are further refined by removing proper nouns.

Step 2:

The files containing the tags are read and the database is populated. The database will consist of tables that maintain a count for each of the tags and a mapping for the comment under which the tag is mentioned.
This will permit the user to be able to sort the comments based on tags and will allow the user to understand the response to his post in a better light.

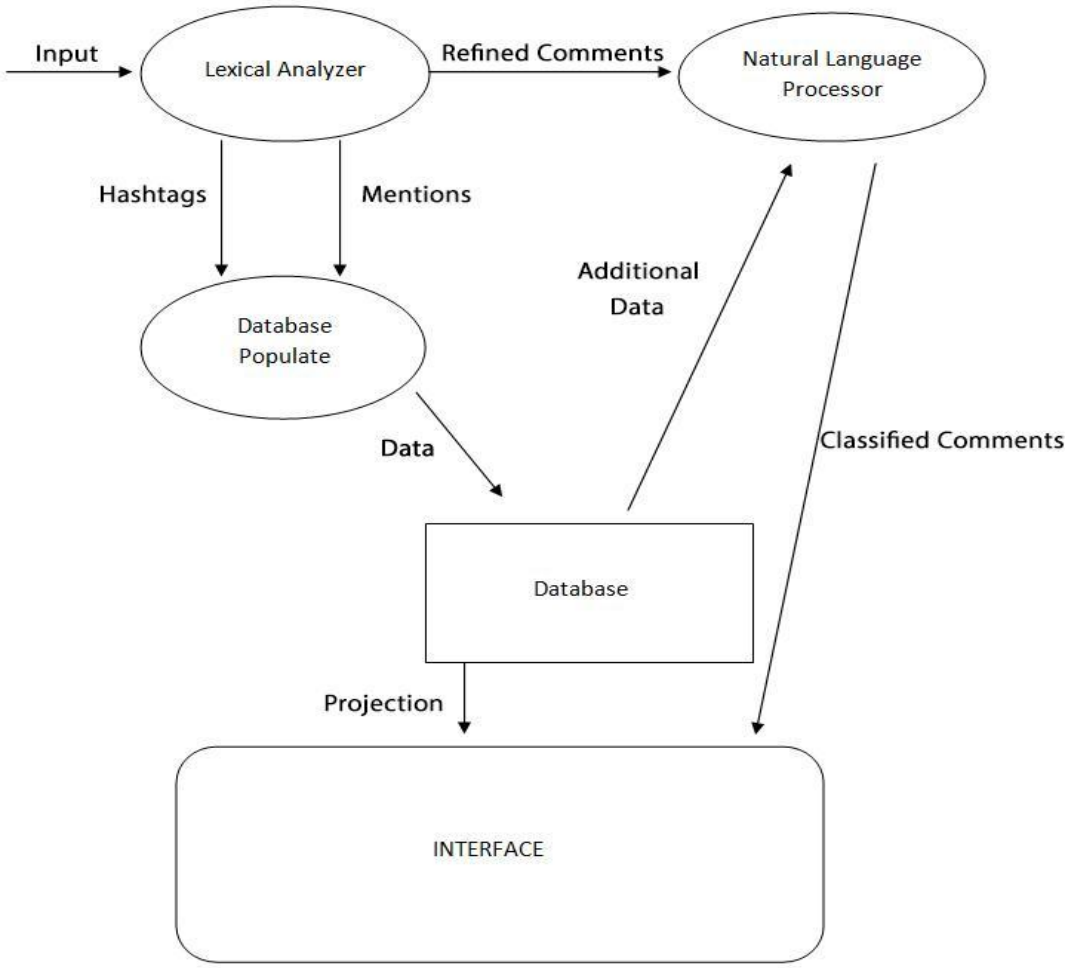
Step 3:

The file with the refined comments (formatted comment without tags and proper nouns) is then sent to the categorization section. Here, Machine Learning techniques are implemented to classify the comments based on sentiment and emotion. NLP and neural network concepts are utilized in this step.

Output:

Users will be able to view comments classified as positive, negative and neutral. Users will also be able to view comments categorized according to the tags used. Furthermore, a tagged user will be notified about being tagged via an email.

Flow of Events



References

1. R. Singh, R. Bagla and H. Kaur, "Text analytics of web posts' comments using sentiment analysis," 2015 International Conference and Workshop on Computing and Communication (IEMCON), Vancouver, BC, 2015, pp. 1-5, doi: 10.1109/IEMCON.2015.7344534.

2. K. M. Anitha, V. Saraswathy, K. Gayathri and S. S. Priya, "An Approach To Comment Analysis In Online Social Media," 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2019, pp. 332-335, doi: 10.1109/ICCCT2.2019.8824949.