**Assignment 11**

Case study of any one text mining / text analytics tool

**Author**

Name: Sakshi Tantak

Class: TE-9

Batch: M-9

Roll No: 33156

**Detailed description of the tool**

Tool used: Google NLP API Tool

The Google Natural Language API is an easy to use tool to a set of powerful NLP models which have been pre-trained by Google to perform various tasks. As these models have been trained on enormously large document corpuses, their performance is usually quite good as long as they are used on datasets that do not make use of a very idiosyncratic language.

The biggest advantage of using these pre-trained models via the API is, that no training dataset is needed. The API allows the user to immediately start making predictions, which can be very valuable in situations where little labeled data is available.

The Natural Language API comprises five different services:

* Syntax Analysis
* Sentiment Analysis
* Entity Analysis
* Entity Sentiment Analysis
* Text Classification

**Syntax Analysis**

For a given text, Google’s syntax analysis will return a breakdown of all words with a rich set of linguistic information for each token. The information can be divided into two parts:

1. **Part of speech:** This part contains information about the morphology of each token. For each word, a fine-grained analysis is returned containing its type (noun, verb, etc.), gender, grammatical case, tense, grammatical mood, grammatical voice, and much more.
2. **Dependency trees:** The second part of the return is called a dependency tree,

which describes the syntactic structure of each sentence.

Both the above parts have been depicted in the attached screenshots

Apart from English, the syntactic analysis supports ten additional languages: Chinese (Simplified), Chinese (Traditional), French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish.

**Sentiment Analysis**

Google’s sentiment analysis will provide the prevailing emotional opinion within a provided text. The API returns two values: The “score” describes the emotional leaning of the text from -1 (negative) to +1 (positive), with 0 being neutral.

The “magnitude” measures the strength of the emotion.

Google’s sentiment analysis model is trained on a very large dataset. Unfortunately, there is no information about its detailed structure available.

**Entity Analysis**

The Google Natural Language API provides some basic information about each detected entity and even provides a link to the respective Wikipedia article if it exists. Also, a salience score is calculated. This score for an entity provides information about the importance or centrality of that entity to the entire document text. Scores closer to 0 are less salient, while scores closer to 1.0 are highly salient.

**Entity Sentiment Analysis**

If there are models for entity detection and sentiment analysis, it’s only natural to go a step further and combine them to detect the prevailing emotions towards the different entities in a text.

While the Sentiment Analysis API finds all displays of emotion in the document and aggregates them, the Entity Sentiment Analysis tries to find the dependencies between different parts of the document and the identified entities and then attributes the emotions in these text segments to the respective entities.

**Text Classification**

Lastly, the Google Natural language API comes with a plug-and-play text classification model.

The model is trained to classify the input documents into a large set of categories. The categories are structured hierarchical, e.g. the Category *“Hobbies & Leisure”* has several sub-categories, one of which would be *“Hobbies & Leisure/Outdoors”* which itself has sub-categories like *“Hobbies & Leisure/Outdoors/Fishing.”*

I found the results of the classification model meaningful in most cases. Still, as all other models from the Google Natural Language API, the classifier comes as a black-box solution which cannot be modified or even fine-tuned by the API user. Especially in the case of text classification, the vast majority of companies will have their own text-categories that differ from the categories of the Google model and therefore, the Natural Language API text classification service might not be applicable for the majority of the users.

**Overall Cost of Google NLP API Tool**



The Google Natural Language API is a very convenient option for quick, out-of-the-box solutions. Very little technical knowledge and no understanding of the underlying machine learning models is required.

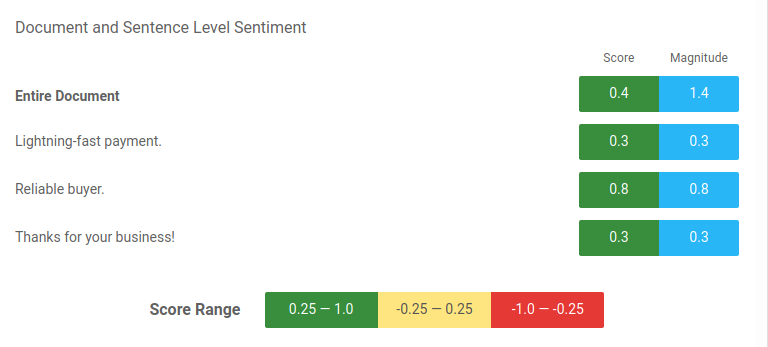
**Discussion about output/outcome**

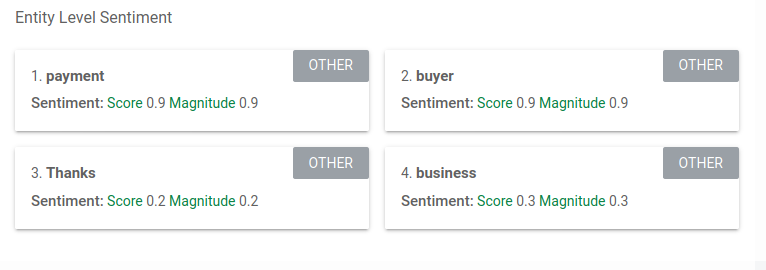
We implemented sentiment analysis on two feedbacks given by users: “Seller sent damaged item; completely uncooperative about refund” and “Lightning-fast payment. Reliable buyer. Thanks for your business.”.

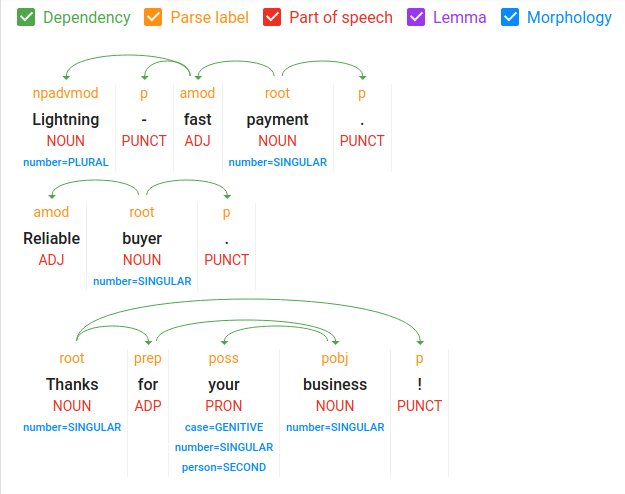
The first feedback is positive while the second is neutral.

Output given by the tool for these two feedbacks is depicted in the snapshots below:

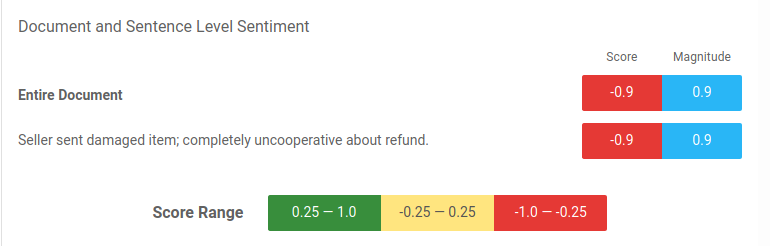
Positive Feedback output:

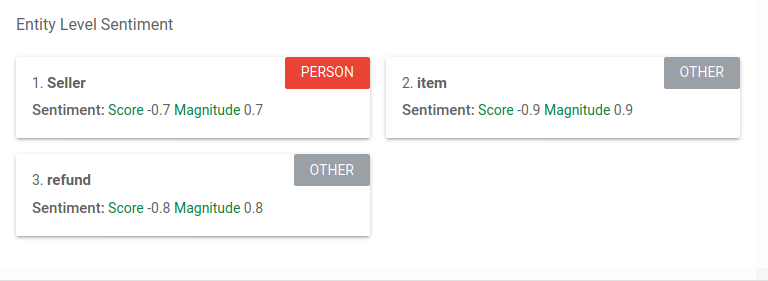


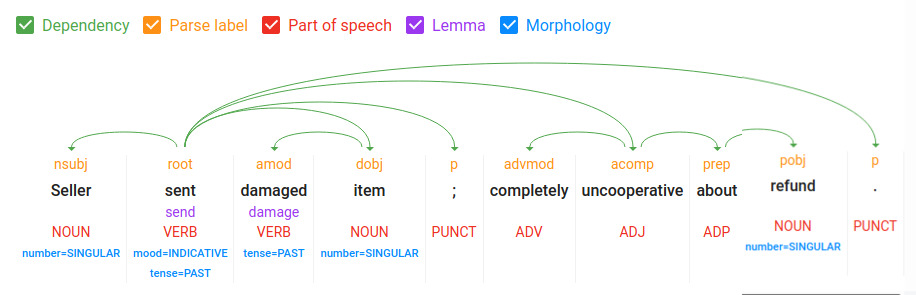




Negative Feedback Output:







**Conclusion**: Thus, the required text analytics by using a text analytics tool has been successfully carried out.