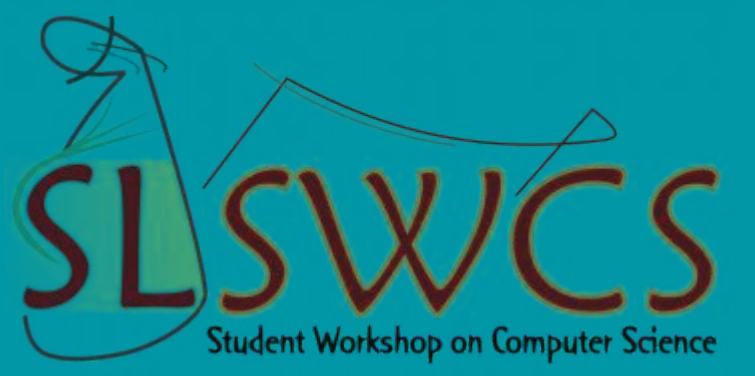


A psychological based analysis of Email Subject Lines

R. Miller, E.Y.A. Charles

Department of Computer Science, Faculty of Science, University of Jaffna

miller@senzmate.com



Abstract

The present email advertisement revenue has reached \$19,353 million, it has not reached to a significant proportion. A marketing email becomes a success only if the email is opened and read by the receiver. The subject line of an email and the email address of the sender are the main deciding factors for one to open an email or leave it. Sentiment analysis is a popular research area which uses natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. This research work analyses the email subject lines in a psychological point of view, for their effect in a person when he/she read it and the decision he/she makes to open that email or neglect it.

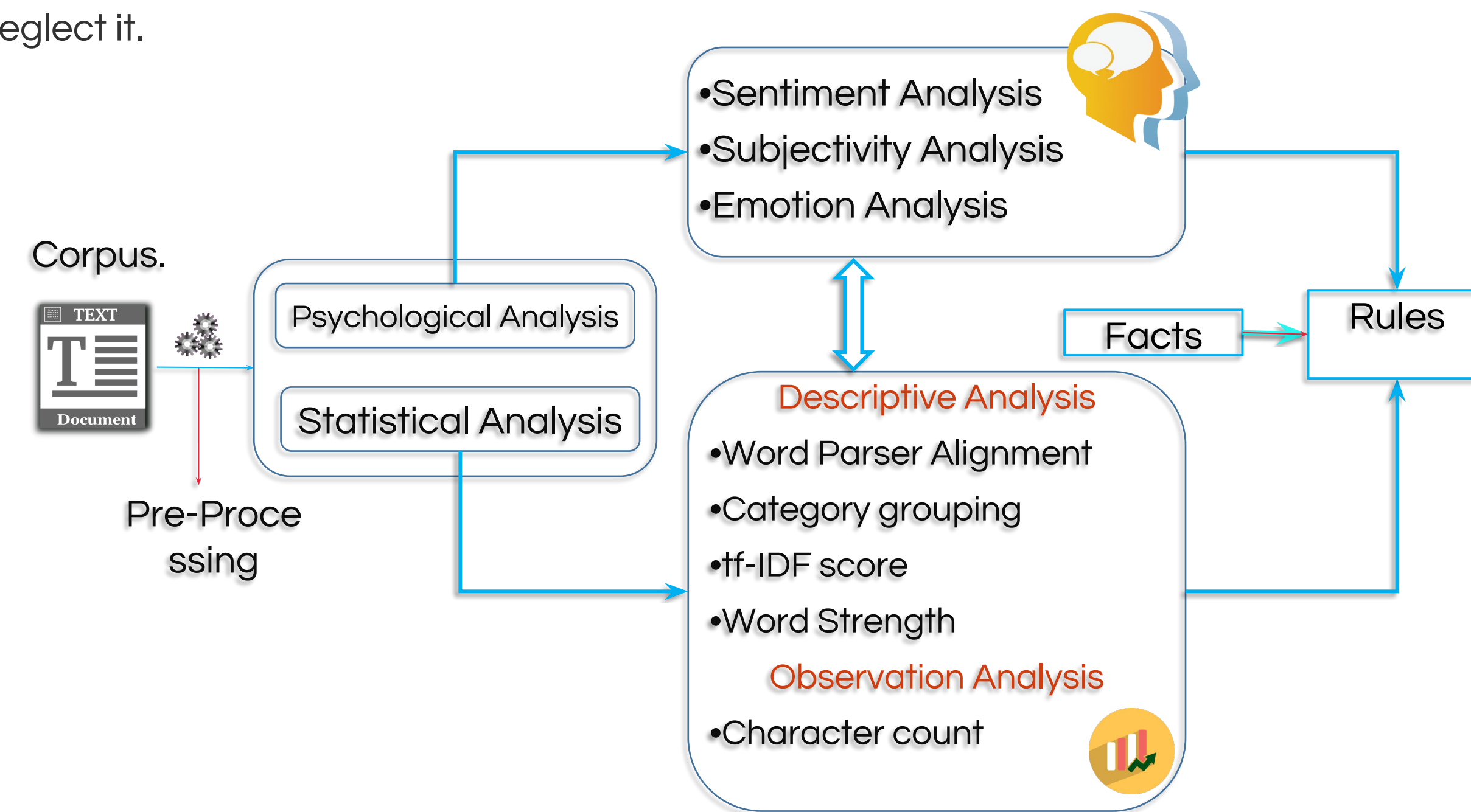
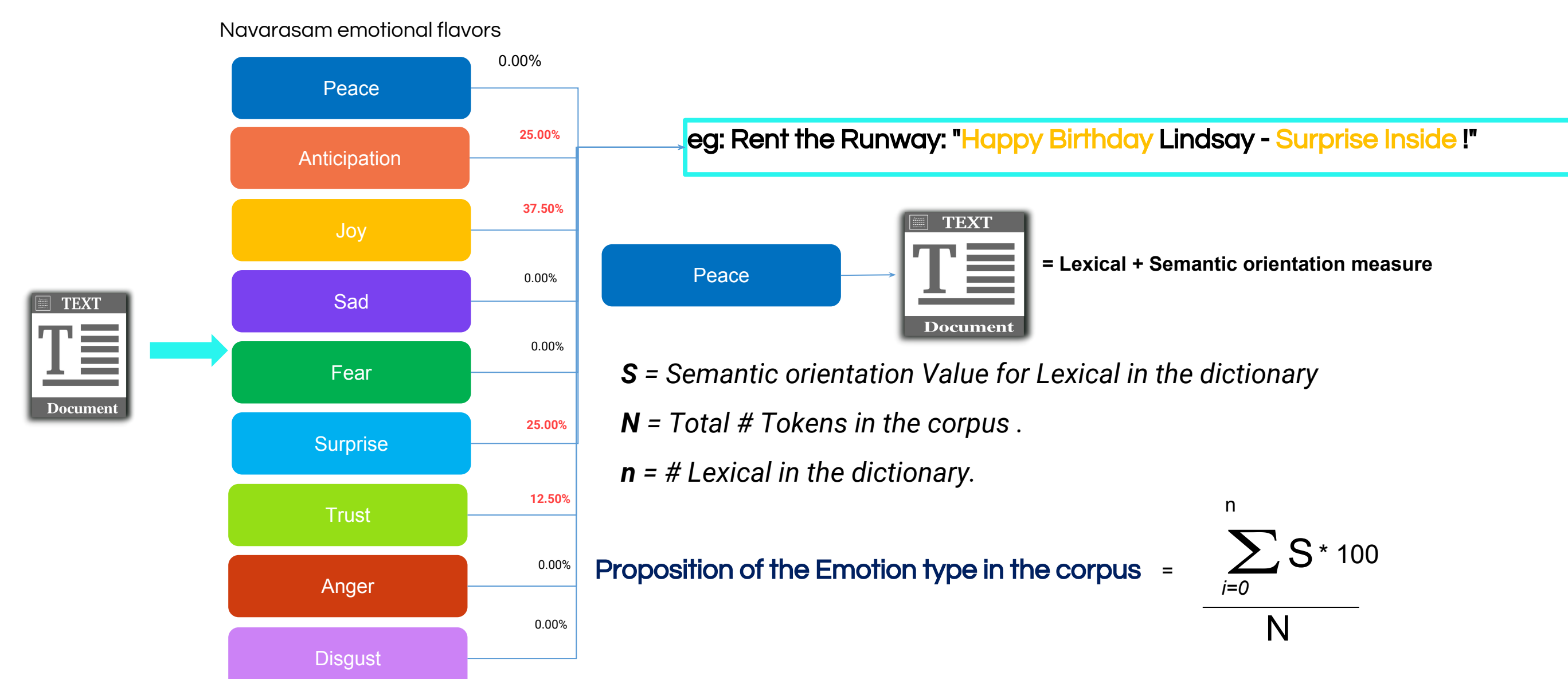


Figure 1: Flow diagram of the proposed method.

Methodology

Email activities represent precious human social and organizational relations, and the objective of this research is to develop methods that can quantify the psychological emotions induced by an email subject line which is correlated with human social and corporate relationships. To achieve the aim of this research, psychological effect analysis was carried on email subject line text in three different approaches, namely:

- Semantic related analysis
- Descriptive analysis
- Observation analysis



Real World Problems

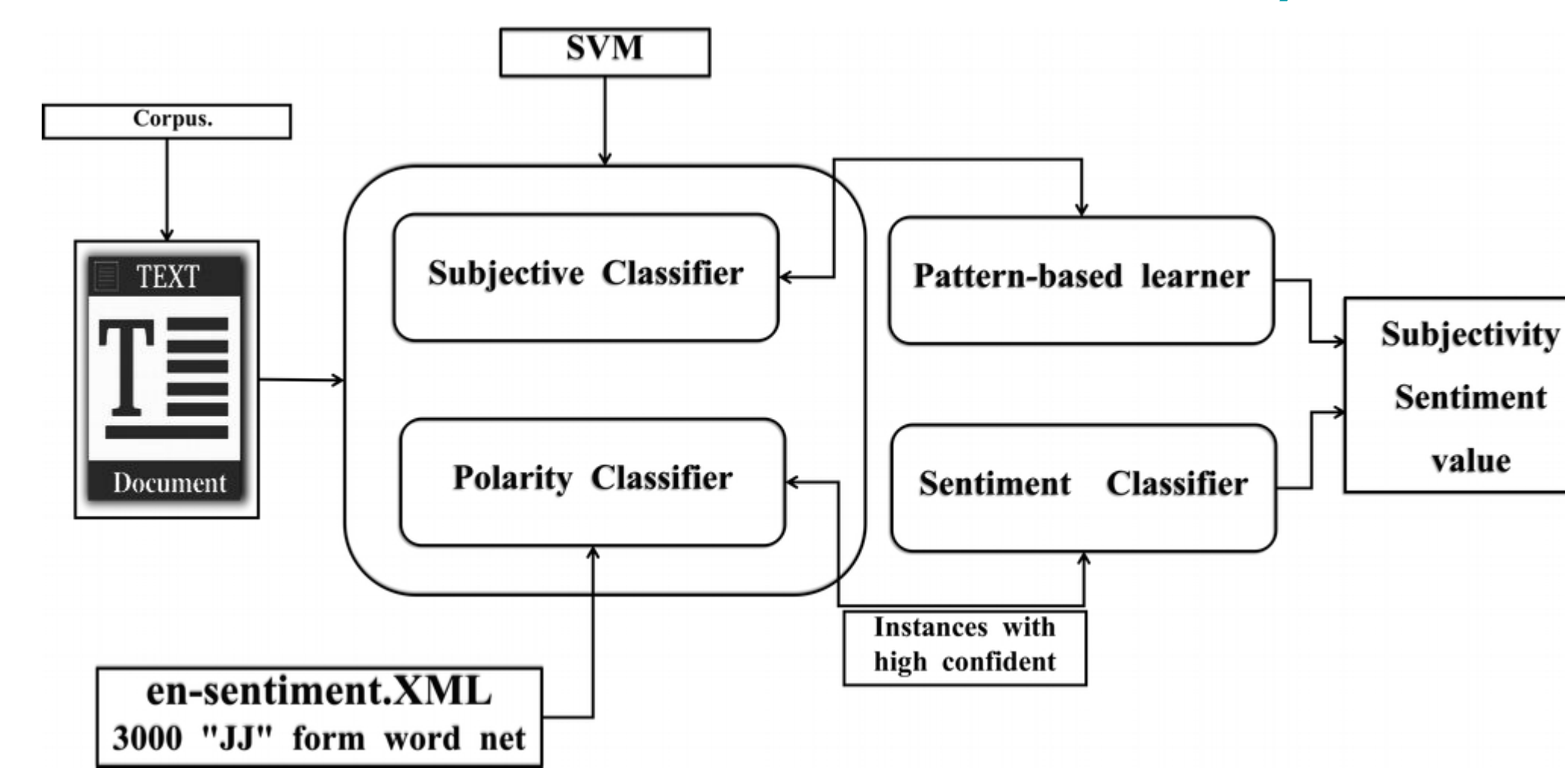
The objective of this research is to develop methods that can quantify the psychological effects induced by an email subject line; where this quantification is correlated with the action performed by the email receiver. This research work was carried out through Sentiment analysis and data mining on three large data sets, namely Enron e-mail, Spamdex Digital Spam Archive and Google Open Adwords. Based on studies in the area of sentiment analysis seven types of analysis techniques were identified and utilised in this study. These seven type of analysis are categorised into three, namely Semantic analysis, Descriptive Analysis and Observational Analysis. Semantic analysis considered the meaning of words in the email subject lines and their psychological effects in three different aspects, namely, Emotional Analysis, Subjectivity Analysis and Sentiment Analysis.

Semantic related analysis

1. Emotional Analysis

Emotional analysis is a field of study which tries to identify the emotion expressed by the writer in text. This research attempts to analyse the effect of the email subject lines based on the possible inducible emotions by the adjectives in the subject line. Here the adjectives are classified into nine predefined groups instead of the six emotions. This nine groups are based on "Navarasa", the nine mental state or primary feelings defined in the Indian aesthetics.

2. Sentiment Analysis



Sentiment Analysis also known as polarity analysis is the process of determining whether a piece of writing is positive, negative or neutral regarding its attitude, opinion or feeling towards something,

such as a person, organisation, product or location. Here the objective is to quantify the sentiment such that it can be utilised along with the subjectivity analysis to determine the effectiveness of the email subject lines.

Descriptive Analysis

3. Word Parser Alignment



Parser alignment assigns meaningful tags to words and groups of words in a sentence. In natural language processing these tags are part-of-speech tags and are assigned to an individual word according to its role in the sentence.

Resources and Pre - Processing Of Data

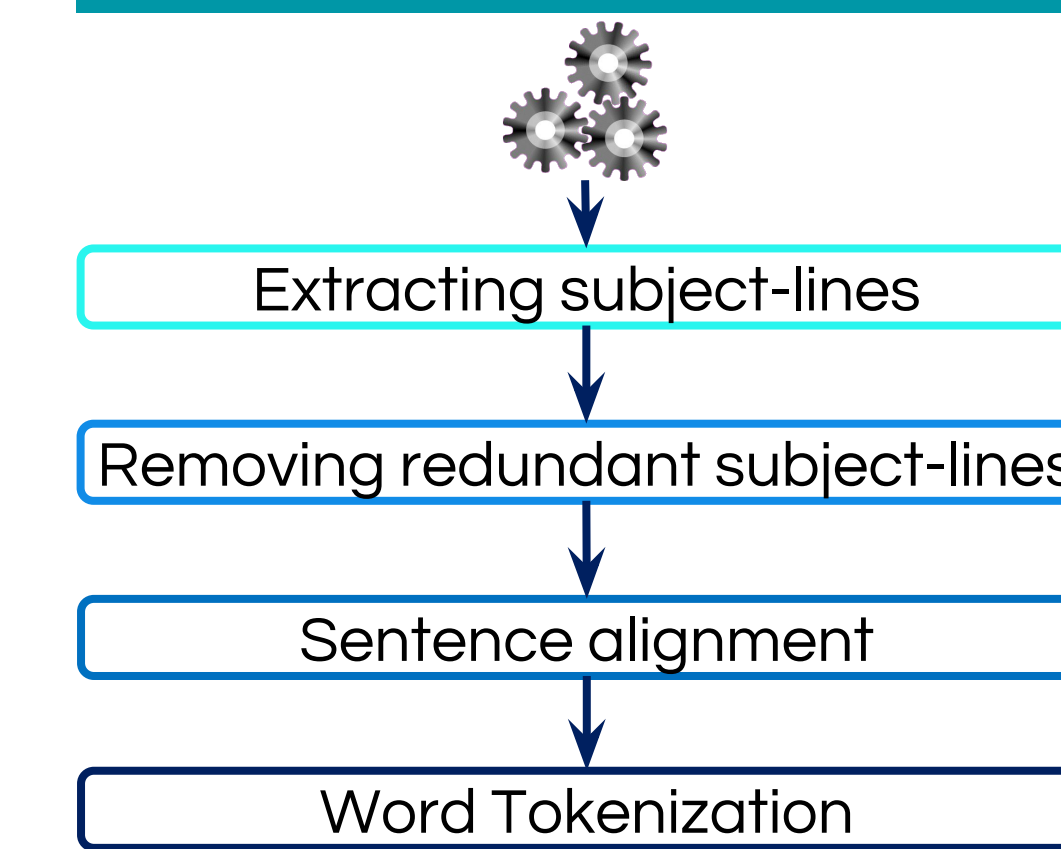


Figure 2: Stages of pre-processing

From the 0.6 million messages in the Enron email data set, approximately 165,000 unique subject lines were extracted. Out of this, replies were sent for 23,000 emails. Hence these were considered as successful emails since the receiver had opened these emails. Apart from Enron data set, 23,000 spam email subject lines were obtained from Spamdex and these were considered as not opened by the receiver.

Word Strength Analysis

TF-IDF score based Analysis

Term Frequency - Inverse Document Frequency, is a numerical statistics which reflects the importance of a word in a document in a collection or corpus.

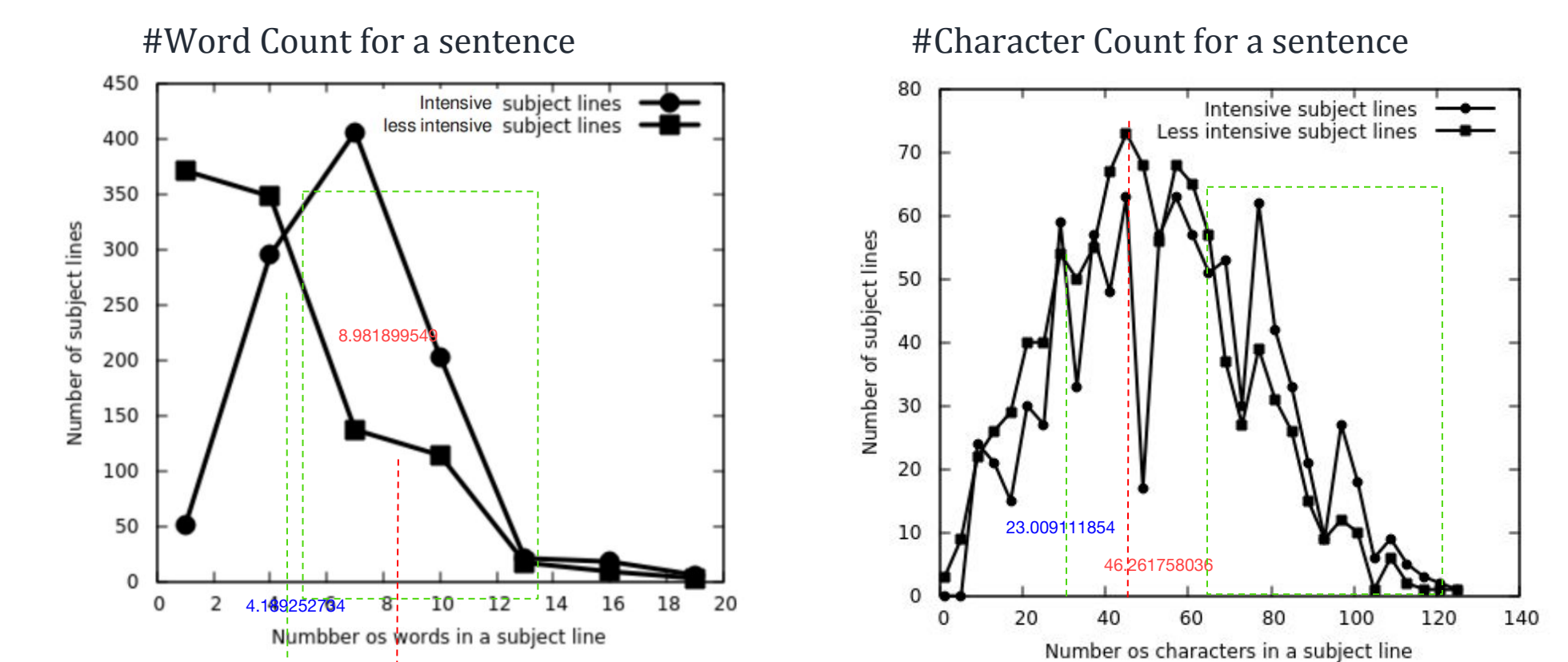
$f(t)$ = (Number of times term t appears in a document) / (Total number of terms in the document).

$idf(t)$ = $\log e$ (Total number of documents / Number of documents with term t in it)

$$tf - idf = f(t) * idf(t)$$

Word and character count based analysis

This analysis was performed to determine correlation tween the number of characters and number of words in a subject line with the email subject lines psychological effect.



Conclusion

From this this study 40 unique rules were proposed for creating the best reader intensive marketing email subject lines. The results found on this study complies with a neurological study which mapped the brain activity through visual stimulation related to emotions. Learn more at [Electronic ISSN: 2472-7598](https://doi.org/10.2472/7598)

- Rule 1: IF "the Subject line have positive emotion words" THEN it deduce more impression.
- Rule 5: IF "the Subject line delivers a blend of negative emotions with surprise, anticipation" THEN it can give more response.
- Rule 15: IF "the Subject line starts with an adjective " THEN it conclude poor impression.
- Rule 38: IF "the Subject line contains capitalized first letters" THEN it deduce more impression.