INTRODUCTION

Overview
To classify reviews of web-service to a negative or a neutral sentiment. A classification model could help identify good and bad features in a web-service.

Hypothesis
Lexical methods may not work due to amount of technicality in the reviews of web-services.
Thus relying on the labels assigned to there views itself might help us to classify the documents to an appropriate sentiment polarity.
The use of delta-tf.idf as a scoring scheme can be realized to verify this hypothesis

DATASET
The corpus of data is obtained by crawling questions posted on Stackoverflow on social media APIs.
The documents either have a neutral sentiment or a negative sentiment, thus classification is done on these two sentiments.
This is because, question usually do not have a positive sentiments associated with them.

ARCHITECTURE

Scoring scheme (Delta tf.idf)
Calculate the term frequency(tf) for a bigram of a document.
Calculate the inverse document frequency(df) of the bigram in the neutral sentiment and in the negative sentiment.
The equation to calculate the weight: $w = tf \times \log \frac{N1\cdot df_2}{N2\cdot df_1}$

RESULTS
Comparisons made lexicon based methods such as Aspect Sentiment Unification Model(ASUM) and SentiStrength.
Lexical based models do not work well as the reviews contain abundant technical content and emotions are expressed subtly.
Since the matrix obtained is very large, experiments of Delta-tf.idf with SVM were performed using linear kernel and 10-fold CV.
SVM performs better since SVM is modelled to work better with sparse data, whereas ANN works best with dense and non-linear data.

REFERENCES
- Georghi Paltoglou and Mike Thelwall. 2010. A study of information retrieval weighting schemes for sentiment analysis.
- Y. Kim, “Convolutional neural networks for sentence classification,”

FUTURE WORK
Extraction of topics in a document using topic modelling technique.
Use GPUs for faster training of the neural net.