Introduction:

SentiWordNet is built via a semi supervised method and could be a valuable resource for performing opinion mining tasks: it provides a readily available database of term sentiment information for the English language, and could be used as a replacement to the process of manually deriving ad-hoc opinion lexicons. In addition, SentiWordNet is built upon a semi-automated process, and could easily be updated for future versions of WordNet, and for other languages where similar lexicons are available. Thus, an interesting research question is to assess how effective is SentiWordNet in the task of detecting sentiment in comparison to other methods, and what are the potential advantages that could be obtained from this approach.

Technique to use SentiWordNet in review classification

The aim of this work is to improve the efficiency of using the SentiWordNet in the task of sentiment classification for the reviews. The proposed technique consists of two phases. A word "fun" average positive and negative scores Sentiment Polarity Phase: As the Term Counting method did not take into consideration the magnitude value of positive and negative scores for words which have a significant effect on the level of word sentiment, therefore we consider the effect of magnitude value in the following calculations methods. 'Term Score Summation' method: in which the summation of positive and negative scores for each term found in a review, is calculated to get the positive and negative scores for all review words. Then, the review sentiment is determined based on which score has the highest value. This method has an advantage over Term Counting method that it takes into consideration the magnitude scores for words.

'Average on Sentence and Average on Review' method: for each sentence in the review, positive and negative scores determined by calculating the average of positive and negative scores for each term found in it. Then calculate the average of positive and negative scores for these sentences to get the positive and negative scores for the review and determining sentiment polarity based on which score has the highest value. Since each word in the SentiWordNet lexicon has three values for its positive, negative and objective sentiment and since the sum of these scores equal to 1, therefore the words which have low positive and negative scores have high objective scores. So, these words most probably have a neutral sentiment. Different thresholds applied to neglect the words which more objective score than positive or negative ones. For example, threshold 0 means neglect words which have positive and negative scores equals to 0.

SentiWordNet can be imported like this:

```
>>> from nltk.corpus import sentiwordnet as swn
```

SentiSynsets

```
>>> breakdown = swn.senti_synset('breakdown.n.03')
>>> print(breakdown)
<breakdown.n.03: PosScore=0.0 NegScore=0.25>
>>> breakdown.pos_score()
0.0
>>> breakdown.neg_score()
0.25
>>> breakdown.obj_score()
0.75
```

Lookup

```
>>> list(swn.senti_synsets('slow')) # doctest: +NORMALIZE_WHITESPACE
[SentiSynset('decelerate.v.01'), SentiSynset('slow.v.02'),
SentiSynset('slow.v.03'), SentiSynset('slow.a.01'),
SentiSynset('slow.a.02'), SentiSynset('slow.a.04'),
SentiSynset('slowly.r.01'), SentiSynset('behind.r.03')]

>>> happy = swn.senti_synsets('happy', 'a')

>>> all = swn.all_senti_synsets()
```

Term Counting

SentiWordNet scores were calculated as positive and negative terms were found on each document and used to determine sentiment orientation by assigning the document to the class with the highest score. This method yielded an overall accuracy of 65.85%, with results detailed in the table below.

Class	Positive	Negative
Predicted Positive	576	259
Predicted Negative	424	741
Total	1000	1000
Class Recall	57.6%	74.1%
Class Precision	68.98%	63.76%

SentiWordNet Features

Experiment	Accuracy
SentiWordNet Features (no refinement).	67.40%
- Including Linear Weight Adjustment to Scores.	68.00%
- Including Negation Detection and Linear Weight Scoring.	68.50%
SentiWordNet, Negation Detection, Linear Scoring and Feature Selection.	69.35%