Predicting News Article Stance

Prajwal Rao (5176504) & Julian Blair (3463793)  
COMP9417 2018s1 – Assignment 3

# Features

## Paraphrasing

The basic idea of using this feature is to determine if words from the headline appear in the body of the article as paraphrased words. This is to try and obtain a more accurate count of binary co-occurrence of tokens from the headline in the body.

The database used for determining paraphrases for words in the headline is WordNet [(link for bib)](https://wordnet.princeton.edu/) provided through the Natural Language Toolkit (nltk) package for python [(link for bib).](http://www.nltk.org/howto/wordnet.html)

Each headline is tokenized into words and each word is tagged into categories – noun, verb etc. Following this, words which belong to either noun or verb categories are queried into the wordnet database to obtains synonym sets for each. Finally, building each word’s synonym set, new words are added to the paraphrased list and returned. The binary co-occurrence function then checks for these words in the article body and updates the counts.