Davies Omale Odu

ECE 590.14

Homework #3

Task: Generate code implementing the algorithm described in the paper "Banerjee and Pederson, 2003. Extended Gloss Overlaps as a Measure of Semantic Relatedness."

# Overview

Method summary: two concepts/senses are similar if their glosses contain overlapping words.

Tools used: I implemented this using Java Api for Wordnet Searching(JAWS).

Algorithm summary:

- Input two words A and B from standard input.
- Find phrasal gloss overlaps between A and B.
- Next, find phrasal gloss overlaps between
    - every synset connected to A, and
        - every synset connected to B
- Compute phrasal scores for all such overlaps.
- Add phrasal scores to get relatedness of A and B.
- A and B can be from different parts of speech.

INPUT: Two strings to be typed into standard input.

HOW I RAN IT: The program runs as a normal Java project in eclipse IDE. In addition to installing WordNet through terminal, it includes the following dependency:

- jaws -bin-1.2.jar

OUTPUT: This program outputs a numerical value which summarizes the relatedness of these two words A and B. The code is appropriately commented to highlight specific use of functions.

Sample output.

```
Enter first concept:
pen
Enter second concept:
pencil
Synsets to be compared are "pen" and "pencil"
----------------------------------------
Relatedness(pen,pencil) = 39
It seems the two words are quite related.
```

References:
1) http://www.cs.cmu.edu/~banerjee/Publications/ijcai03.ppt
2)Banerjee and Pederson, 2003. Extended Gloss Overlaps as a Measure of Semantic Relatedness