**IS620 Project 4**

**Student: Neil Acampa**

6.10.3 Word sense disambiguation

This program prompts for a keyword in Alice and Wonderland (The default keyword is **look)**. Then reads the Alice in Wonderland corpus and selects sentences which contain that key word. The sentences are stored in the array: **sentword.**

Then for each keyword sentence, a new sentence is created containing 2 words prior to the keyword, the keyword and 2 words after the keyword. The adjusted sentence is stored in the array: **sentwordadj.**

Using the keyword and Word Net all synset instances are found and stored in arrays: **syns, synsdef and synsname (Synset, the sentence definition and the extracted synset name i.e “look.n.02”)**

**Some examples of Synsets for Keyword: look**

Synset('look.n.02') : the act of directing the eyes toward something and perceiving it visually

Synset('look.n.03') : physical appearance

Synset('spirit.n.02') : the general atmosphere of a place or situation and the effect that it has on people.

Synset('look.v.09') : accord in appearance with

Features are then created using **Lesk’s algorithm:**

For a given word and sentence: Lesk’s algorithm returns a Synset with the highest number of overlapping words between the context sentence and the different definitions from each Synset.

Example: **Corpus: Alice in Wonderland by Lewis Carroll 27 instances of Keyword: look**

**Sentence Sense Tag Sense Tag Description**

{"(['down', 'to', 'look', 'about', 'her'] : Synset('expect.v.03'))": 'look forward to the probable occurrence of'}

{"(['tried', 'to', 'look', 'down', 'and'] : Synset('expect.v.03'))": 'look forward to the probable occurrence of'}

{"(['made', 'her', 'look', 'up', 'in'] : Synset('look.v.09'))": 'accord in appearance with'}

Then the sentiment value for each trimmed sentence is found using **senti\_synet**. The positive/negative scores for each instance of the keyword in the trimmed sentence stored in featuresets1.

Some examples are:

**Corpus: Alice in Wonderland by Lewis Carroll 27 instances of Key Word: look**

**Sense Tag Positive score Negative score**

Synset('expect.v.03') 0.00 0.25

Synset('look.v.09') 0.38 0.00