

HW #6 Notes

February 26, 2014

WordNet-based WSD

- Perform word sense disambiguation of probe word
 - In context of word set
 - Line news,lot,joke,half,show,cast,brainstorm
 - Tie jacket, suit
- An answer key is provided
 - Don't expect to get them all right!

Implementation

- Implement a simplified version of Resnik's
 - "Associating Word Senses with Noun Groupings"
 - Select a sense for the probe word, given group
 - Rather than all words as in the algorithm in the paper
 - For each pair (probe, noun_i)
 - Loop over sense pairs to find MIS, similarity value (v)
 - Update each sense of probe, descended from MIS with v
 - Also accumulate normalization term
 - Select highest scoring sense of probe

Components

- Similarity measure:
 - IC:
 - /corpora/nltk/nltk-data/corpora/wordnet_ic/ic-brown-resnik-add1.dat
 - NLTK accessor:
 - `wnic = nltk.corpus.wordnet_ic.ic('ic-brown-resnik-add1.dat')`
 - Note: Uses WordNet 3.0
- You can compute your own for extra credit!!

Components

- >>> from nltk.corpus import *
>>> brown_ic = wordnet_ic.ic('ic-brown-resnik-add1.dat')
>>> wordnet.synsets('artifact')
[Synset('artifact.n.01')]
>>> artifact = wordnet.synset('artifact.n.01')
>>> artifact.offset
21939
>>> information_content(artifact, brown_ic)
2.4369607933293391

Components

- WordNet API
 - NLTK
 - Others exists, but no warranty
- <http://nltk.googlecode.com/svn/trunk/doc/howto/wordnet.html>
 - <http://nltk.googlecode.com/svn/trunk/doc/api/nltk.corpus.reader.wordnet-module.html>
 - <http://nltk.googlecode.com/svn/trunk/doc/api/nltk.corpus.reader.wordnet.WordNetCorpusReader-class.html>

Note

- You can use supporting functionality, e.g.:
 - Common_hyponyms, full_hyponyms, etc
- You can NOT just use the built-in resnik_similarity, etc
 - If you're unsure about acceptability, just ask...