**2.7 WordNet**

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the [browser](http://wordnetweb.princeton.edu/perl/webwn" \t "_self). WordNet is also freely and publicly available for [download](http://wordnet.princeton.edu/wordnet/download/" \t "_self). WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus do not follow any explicit pattern other than meaning similarity.

**2.7.1 WordNet Similarity**

WordNet::Similarity is a Perl module that implements the semantic relatedness measures described by [7], [8], [9], [10], [12], [13], the extended gloss overlap measure by [14], and two measures based on context vectors by [15]. The details of the vector measure are described in the Master's thesis work described in [15], and the vector\_pairs measure is derived from that.

The Perl modules are designed as objects with methods that take as input two word senses. The semantic relatedness of these word senses is returned by these methods. A quantitative measure of the degree to which two word **senses are related has wide ranging applications in numerous areas, such as word sense disambiguation, information retrieval, etc. For example, in order to determine which sense of a given word is being used in a particular context, the sense having the highest relatedness with its context word senses is most likely to be the sense being used. Similarly, in information retrieval, retrieving** documents containing highly related concepts are more likely to have higher precision and recall values.