ABSTRACT

Measuring semantic relatedness is a task to estimate the closeness of two different word meanings. This task can be easily done by human using their experience and knowledge, however computers can only do the task by using helpful lexical resources such as dictionary, thesaurus, and encyclopedia. One of the most used lexical resources to help with this task is WordNet which shall be used in this final project. This task will make other Natural Language Processing tasks easier, such as detecting plagiarism between two different text data.

In this final project, measuring the semantic relatedness is done in a graph model which will be built by using WordNet. Each word and their meanings inside WordNet will be used as a group of nodes which are connected with several types of edges based on the relation that connected them. Every type of edges has different weighting measures. The weighting measure will highly affecting the outcome of measuring semantic relatedness.

The graph that has been built will be used to measure semantic relatedness using the method called random walks. This graph will be the place for the particle from random walks to walk freely. This particle will walk from node to another node by using its transition probabilities to another node. The transition probabilities only dependent on the node which the particle is currently in, this is why the random walk is called a Markov chain.

The result of the system's evaluation using Rubenstein and Goodenough dataset has the highest score using Pearson correlation coefficient of 0.477.

Keywords: *semantic relatedness, WordNet, Natural Language Processing, random walks, graph, Markov chain, Pearson correlation coefficient*