

Rohan Nyati

500075940

R177219148

B5-Ai & MI

Experiment -6

Study the 'WordNet' ontology and how can be used for NLP purposes.

WordNet is a [lexical database](#) of [semantic relations](#) between [words](#) in more than 200 languages.^[2] WordNet links [words](#) into [semantic relations](#) including [synonyms](#), [hyponyms](#), and [meronyms](#). The synonyms are grouped into [synsets](#) with short definitions and usage examples. WordNet can thus be seen as a combination and extension of a [dictionary](#) and [thesaurus](#). While it is accessible to human users via a [web browser](#),^[3] its primary use is in automatic [text analysis](#) and [artificial intelligence](#) applications. WordNet was first created in the [English language](#)^[4] and the English WordNet [database](#) and [software](#) tools have been released under a [BSD style license](#) and are freely available for download from that WordNet website.

For the past several decades, fields of studies such as computational linguistics, NLP, machine learning (ML), and AI have developed methods and algorithms for information retrieval and extraction from free-text knowledge resources. Some of these methods have been used and tested for ontology learning from text and have shown promising results. In general, these methods can be categorized into symbolic, statistical, and hybrid approaches ([Table 1](#)). The symbolic approach utilizes linguistic information to extract information from text. For example, noun phrases are considered to be lexicalized concepts and are often used to represent concepts in an ontology. Linguistic rules describing the relationships between terms in the text can also be used to identify conceptual relationships within an ontology. The most common symbolic approach is to use lexico-syntactic pattern (LSP) matching, which was first explored by Hearst [\[37\]](#). LSPs are surface relational markers that exist in a natural language. For example, in the phrase “systemic [granulomatous diseases](#) such as Crohn’s disease or sarcoidosis” the words “such as” can help us infer that “systemic granulomatous diseases” is a hypernym of “Crohn’s disease” and “sarcoidosis”. Another symbolic approach is to use the internal [syntactic structure](#) of component terms. Concepts are often represented

using compound, multi-word terms. In general, a compound term is more specific than a single compositional term. The basis of this method is the assumption that a compound term is likely a hyponym of a single term. For example, using this approach the term “prostatic carcinoma” can be considered to be a hyponym of “carcinoma”. It is also possible to use multiple symbolic approaches at the same time, for example the LSP method can be used with information from compound terms.

Among 308 word returned from WordNet, human expert judged that 252 of the extracted words were correct. Of the 56 incorrect words returned, there were no senses in WordNet relevant to the amphibian domain. Since each of above 252 returned words may have some correct senses judged by the human expert, we got finally 285 correct senses. In order to see how well each algorithm performed, we implemented different thresholds to see how results are varied as we increased our selectivity based on level of match and/or number of common elements matched.