ABSTRACT

Semantic similarity measure plays an effective role and it is a fundamental and effectiveness concept to obtain the most relevant results. WordNet is a most acceptable domain independent lexical database for semantic similarity measurement. Since, its usefulness has been severely hampered due to its limited coverage, bias, and inconsistency because the quality of semantic relationship in WordNet not updated appropriately for the current usage in the modern Information Retrieval (IR). It is mainly due to practical usage of English words and semantic similarity of these words change over the period. It is very difficult and expensive to maintain WordNet as an updated knowledge source. Building the WordNet from scratch is not an easy task for keeping updated with current terminology and concepts. Automated and semi-automated methods for developing such resources are crucial for further resource development and improved application performance. There are a large number of works available in the existing literature to extend WordNet using web resources either English Wikipedia or web search engine.

The primary contribution of this research is to utilize both resources to infer newly evolving relationship between concepts in the WordNet over time. The newly evolving relationship may diminish or increase the value of semantic similarity. Therefore, this research undergoes a different perspective that automatically updates an existing lexical ontology employs knowledge resources such as the Wikipedia and the Web search engine. This methodology has established the recently evolving relations and also aligns the current relations between concepts. It consists of three main phases such as candidate article generation, lexical relationship extraction and generalization and WordNet alignment. In candidate article generation, disambiguation mapping disambiguates ambiguous links between WordNet
concepts and Wikipedia articles and returns a set of word-article pairings. Lexical relationship extraction phase includes two algorithms, Lexical Relationship Retrieval (LRR) algorithm discovers the set of lexical patterns exists between concepts and sequential pattern grouping algorithm generalizes lexical patterns and computes corresponding weights based on its frequencies. Furthermore, Sequential Minimal Optimization (SMO) selects the suitable pattern using the optimal combination of the weight of lexical patterns and page count based concurrence measures. WordNet alignment phase establishes a new relationship that is not available in WordNet and also aligns the existing patterns based on computed weight. The proposed semantic measure is compared with human ratings in well-known benchmark data sets. The proposed measure is applicable to real-world applications to compute the similarity and then evaluate the measures with respect to their performance in applications such as NLP, WSD, and IR systems. Experimental results illustrate that the proposed approach better than existing mechanisms on benchmark datasets.