Abstract

The existing web is made for human usage in order to observe required information and comprehend it. The semantic web alongside to this important purpose, considers the interaction of machine with the web content and automatic concept perception. In this web, contents must be presented to the machine in an understandable way. Most of the work in this area is divided in two categories: concept extraction and relations extraction.

Most of the applications in the field of information extraction, knowledge acquisition, natural language perception and information retrieval require comprehension in the level of concept, semantic, relation and semantic relation. Each of the mentioned entities has specific definition and characteristic. Semantic relations refer to the relations between concepts existing in the mind, which is called conceptual conceptions, or to the relations between terms which are called lexical relations.

Semantic relations in a text can occur in either lower level, such as between words, or in upper levels, i.e. between phrases, sentences, paragraphs, or even at higher levels, such as between two documents or a collection of documents. The mentioned levels have a range of structural definitions, but another level in the area of linguistics, which is yet more important and however not addressed often in the computational linguistics, is the discourse level; which the range of definition is not structural but semantics. At this level, the scale is not a sentence, paragraph or document, but is a semantic unit; and maybe between a few sentences, a number of paragraphs or the whole text.

An important note is the occurrence of these semantic relations in the discourse level, either explicit or implicit. The significance of this is due to the increasing spread of information in the web level, which is often in the form of text. Knowledge acquisition and information extraction from massive amount of data require understanding concepts and the semantic relations between them, especially in the area of linguistics, psychology, cognitive science, and in this thesis computational linguistics.

This thesis presents a new computational method in order to infer and represent the explicit and implicit semantic relations in discourse level, which is a complementary to the preceding methods. The characteristics of the presented method include utilization of natural language processing methods, besides statistical computations, and singular value decomposition, along with unsupervised pattern recognition and extraction. The aim is to help gain precise understanding of text, and implement in various applications in the areas related to linguistics, psychology, cognitive science and especially computational linguistics.

Existing researches show that the presented artificial intelligence and machine learning methods in this area require a high level of knowledge engineering or very large databases for learning and experimentation. The presented method, while innovating in the proposed area, has resolved the previous problems and limitations. The thesis implies, develops and evaluates the proposed method in order to calculate efficiency, precision and accuracy.

The proposed computational method utilizes from linear algebra method of singular value decomposition alongside performing a number of processing and pre-processing in the field of natural language, such as semantic role labeling,

syntactic labeling and statistical techniques for processing natural language, together with coherence and cohesion computing methods.

Applications in ontology creation, evolution and development; knowledge extraction; questioning and answering system; text summarization and information retrieval such as search engines are a few to mention.

Evaluation of the purposed method is done via two approaches: ۱) The application of the method on standard data sets ۲) The implementation of the proposed method in an application field. Evaluation results revealed a satisfactory performance of the proposed method.

Keywords: Discourse level, Semantic Relation, Singular Value Decomposition, Knowledge acquisition, Natural Language Processing, Computational Linguistic, Statistical Computation, Semantic Domain