Abstract

India has a huge amount of agricultural data in the form of textual documents, tables and spread sheets. However, the data is often underutilized by different Government and/or other organizations because of lack in application of data processing techniques to agriculture data. Converting this data into efficient knowledge representation is found to be helpful in answering queries of the underlying domain. Ontology is one such knowledge representation technique. Automatic ontology development focuses on converting the available domain text to machine processable knowledge representation in the form of ontology. Automatic term extraction, and automatic relation extraction constitute important steps prior to ontology design. We propose a regular expression and natural language processing based scheme for automatic term extraction. Automatic relation extraction in the context of automatic ontology development involves identification of relations relevant to the domain. It further involves identification of related pairs of terms corresponding to each relation. Two types of relations viz intra-subdomain and inter-subdomain relations are worked upon in this research. We propose a knowledge based scheme by using expert knowledge for framing of rules for identifying related pairs of terms for each relation. The scheme is further improved by pre-processing the input text using coreference resolution. Multiple ontologies for the same domain may be developed in a distributed manner. Hence automatic ontology merging also constitutes an important part of incremental automatic ontology development. Development of automatic ontology merging scheme involves identifying equivalence, hierarchical and other semantic relations between concepts and individuals of source ontologies. We use two linguistic measures to identify lexical anchors between source ontologies. We also use a mathematical model, viz., Formal Concept Analysis, for identification of additional anchors.