Abstract

Information Retrieval (IR) issues have attracted increasing attention due to the growing availability of the documents. The retrieval of web pages is more challenging due to the ambiguous nature of the unstructured information found in these pages. Ontologies help to overcome the disambiguate nature of the natural language by the use of standard terms that relate to specific concepts. Ontology is a hierarchy of concepts with attributes and relations that defines an agreed terminology to describe semantic networks of interrelated information units. Ontology provides a vocabulary of classes and properties to describe a domain, emphasizing the sharing of knowledge and the consensus about its representation. This research focuses on IR systems moving from a lexical to semantic interpretation to match object and queries on a semantic basis. In natural language, many words are ambiguous giving different meanings based on the context and situation. Therefore, development of web directories, classification of web pages and analysis of topic-specific search are useful. Classification of contents makes an important part of most of the content management and retrieval activities. The underlying objective of this research work is to develop an effective and efficient feature selection and
classification algorithm that can achieve good accuracy in classifying web pages.

References


Index Terms

Computer Science Information Sciences

Keywords

Information Retrieval, Feature Selection, Genetic Algorithm