

IDENTIFYING KEYWORDS AND KEYPHRASES

ASHIWINI C. MADANE

Abstract

Keywords and keyphrases are widely used in large document collections. They describe the content of single documents and provide a kind of semantic metadata that is useful for a variety of purposes. Text mining is powerful tool to find useful and needed information from huge data set. For context based text mining, keyphrases are used. Keyphrases provide brief summary about the contents of documents. In document clustering, number of total cluster is not known in advance. In K-means, if prespecified number of clusters modified, the precision of each result is also modified. Therefore Kea, is algorithm for automatically extracting keyphrases from text is used. In this kea algorithm, number of clusters is automatically determined by using extracted keyphrases. Keameans clustering algorithm provide easy and efficient way to extract test document from large quantity of resources. Keyphrase play important role in text indexing, summarization and categorization. Keyphrases are selected manually. Assigning keyphrases manually is tedious process that requires knowledge of subject. Therefore automatic extraction techniques are most useful.

Keywords: Text mining, Key phrase extraction, keyphrase.