

Finding Related Scientific Documents Using Formal Concept Analysis

Authors : Nadeem Akhtar, Hira Javed

Abstract : An important aspect of research is literature survey. Availability of a large amount of literature across different domains triggers the need for optimized systems which provide relevant literature to researchers. We propose a search system based on keywords for text documents. This experimental approach provides a hierarchical structure to the document corpus. The documents are labelled with keywords using KEA (Keyword Extraction Algorithm) and are automatically organized in a lattice structure using Formal Concept Analysis (FCA). This groups the semantically related documents together. The hierarchical structure, based on keywords gives out only those documents which precisely contain them. This approach opens doors for multi-domain research. The documents across multiple domains which are indexed by similar keywords are grouped together. A hierarchical relationship between keywords is obtained. To signify the effectiveness of the approach, we have carried out the experiment and evaluation on Semeval-2010 Dataset. Results depict that the presented method is considerably successful in indexing of scientific papers.

Keywords : formal concept analysis, keyword extraction algorithm, scientific documents, lattice

Conference Title : ICDMBDDDS 2017 : International Conference on Data Mining, Big Data, Database and Data System

Conference Location : Toronto, Canada

Conference Dates : June 15-16, 2017