Abstract

A common and standard approach to model text document is bag-of-words. This model is suitable for capturing word frequency, however structural and semantic information is ignored. Graph representation is mathematical constructs and can model relationship and structural information effectively. A text can appropriately represented as Graph using vertex as feature term and edge relation can be significant relation between the feature terms. Text representation using Graph model provides computations related to various operations like term weight, ranking which is helpful in many applications in information retrieval. This paper presents a systematic survey of existing work on Graph based representation of text and also focused on Graph based analysis of text document for different operations in information retrieval. In this process taxonomy of Graph based representation and analysis of text document is derived and result of different methods of Graph based text representation and analysis are discussed. The survey results shows that Graph based representation is appropriate way of representing text document and improved result of analysis over traditional model for different text applications.

- Francois Rousseau, Michalis Vazigiannis, Graph-of-word and TW-IDF: New Approach to Ad Hoc IR. Proceedings of the 22nd ACM international conference on Conference on information and knowledge management 2013, pp. 59??68.
Graph based Representation and Analysis of Text Document: A Survey of Techniques


Index Terms

Computer Science Applied Mathematics

Keywords

Information Retrieval Graph Theory Natural Language Processing.