Abstract

It is always difficult to find relevant information in unstructured text documents. In this paper we study the methods of fuzzy search, instant search and proximity ranking and how they can be used in the process of annotation of documents. These various methods can be integrated to give better search results and to achieve efficient space and time complexities. We propose a novel alternative approach which facilitates the generation of the structured metadata automatically using OpenNLP, methods of Instant-fuzzy search and Proximity ranking. It is done by identifying documents which are likely to contain the information of interest. And this information will be subsequently useful for querying the database.

References

- Cetindil, I., Esmaelnezhad, J., Taewoo Kim, Chen Li, "Efficient instant fuzzy search with proximity ranking", Data Engineering (ICDE), 30th International Conference,
IEEE 2014

Index Terms

Computer Science Fuzzy System

Keywords

Document Retrieval Document Annotation Instant-fuzzy search Proximity Ranking OpenNLP.