Abstract

Annotation is a process of adding the information into the Document which is useful for extracting the information. A large number of organizations now days generate a large amount of data which is always present in the textual format. But such collections of textual document which contains a large amount of structured information which is completely hidden in the unstructured information. Information extraction algorithm is too costly because it always works on the top of the text and it does not provide the necessary structured information. In our paper, we present a method to generate the structured attribute by identifying the documents which contain the information of interest and this information in future useful for querying the database. The major contribution of this paper, we propose the algorithm, where it identifies the structured attribute which is present in the document by combining both the query workload and the content of the text document. Our Experiment result shows that our technique gives the better results compared to the methods which only relay on the content of the document and only on the query workload.

References

Index Terms

Computer Science Information Sciences

Keywords

Annotation CADS form CV and QV