Abstract

A large number of systems today generate and share textual descriptions of their products, services, and actions. Such assemblage of textual data contain significant amount of structured information, which reside in the unstructured text. While related information extraction algorithms facilitate the extraction of structured relations, they are often costly and defective, especially when operating on top of text that does not contain any instances of the final targeted structured information. We present a good alternative approach that facilitates the generation of the structured metadata by identifying documents that are likely to have information of interest and this information is going to be subsequently useful for querying the database.

Our approach depends on the idea that instead of writing the query for each requirement it is easier to fetch them by using the annotated form based technique. In this approach user do not need to learn the correct query along with the previous information of the dataset. As a major contribution towards this paper, we present algorithms that identify structured attributes that are
likely to find within the document, by jointly utilizing the content of the text and the query workload. Our experimental evaluation display that our approach generates superior results compared to approaches that go through textual content or only on the query workload, to identify attributes of our interest.

References

3. J. M. Ponte and W.B. Croft: proposed a paper “A Language Modeling Approach to Information Retrieval”.
6. P. Heymann, D. Ramage, and H. Garcia-Molina: proposed a paper “Social Tag Prediction”.
9. B. Sigurbjornsson and R. van Zwol: proposed a paper “Flickr Tag Recommendation Based on Collective Knowledge”.

Index Terms

Computer Science

Applied Mathematics
Keywords

Data mining, ECADS, Annotation, Query Accessing, Content Value, Query Value