Abstract

Rapid increase of use of internet provides knowledge extraction from the web databases and HTML pages associated with it. Although there are various techniques implemented for the access of the annotations of the search results from the web databases. Here in this paper by identifying the problems with the existing techniques for the annotation search results from web databases such as alignment problem or to split composite text node when there are no explicit separators. Here propose an efficient technique which overcomes the above problems by using some supervised learning algorithm such as support vector machine. The technique implemented provides high rate of information by providing high annotations search results from web databases. The proposed method implemented here for the efficient retrieval of text nodes and data units using supervised learning approach using SVM provides efficient precision and recall as compared to the existing approach. The proposed methodology implemented here using SVM based clustering and labeling of search records is compared with existing methodology implemented for the search records. The Result Analysis shows the performance of the proposed methodology. The proposed method shows higher precision and recall as well as has high Accuracy for the prediction of annotated search records from the web databases.
Learning based Clustering for the Automatic Annotations from Web Databases

References

- A Survey of Web Information Extraction Systems Chia-Hui Chang, Member, IEEE Computer Society, Mohammed Kayed, Moheb Ramzy Girgis, Member, Ieee Transactions On Knowledge And Data Engineering, VOL. 18, NO. 10, OCTOBER 2006

Index Terms

Computer Science

Database Management

System
Learning based Clustering for the Automatic Annotations from Web Databases

**Keywords**
Annotations  Wrapper  Semantic Model  HTML Tags  NLP  Ontology  UIUC.