Abstract

The uses of web search engines are very frequent and common worldwide over the internet by end users for different purposes. A web search engine takes the query request from the end user and executes that query on relational database used to store the information on behalf of that web search engine. Based on input queries the dynamic response is generated by search engine, in the form of HTML based pages. Such pages are supported with the web databases. Every web page generated contains many results to display for particular query, called as Search Result Records (SRRs). Sometimes it becomes troublesome to extract relevant data from diverse sources. The SRRs generated may contain data units that are relevant to one common semantic. These SRRs are further required to be assigned with proper labels. The manual methods for record extraction and labeling have a worse scalability. Thus automatic annotation based method is needed to improve the accuracy as well as scalability of web search engines. This paper presents an automatic annotation technique for web search results. The proposed approach first aligns the data units on a result page into different groups such that the data in the same group have the same semantic. Then, each group is annotated from different aspects and aggregates the different annotations to predict a final annotation label for it. The annotation wrapper generated for the search site is automatically constructed and can be used to annotate new result pages from the same web database. Experiments indicate that
the proposed approach is highly effective.

References

- Wei Liu, Xiaofeng Meng, Member, IEEE, and Weiyi Meng, Member, IEEE, ViDE: A Vision-Based Approach for Deep Web Data Extraction
- D. Raghu, V. Sridhar Reddy, Ch. Raja Jacob, Dynamic Vision-Based Approach in Web Data Extraction
- R. Vijay1, Dr. K. Prasadh, A Vision Based Approach for Web Data Extraction using Enhanced Co-citation Algorithm
- Arvind Arasu, Hector Garcia-Molina, Extracting Structured Data from Web Pages, Stanford University
- Yanhong Zhai, Bing Liu, Department of Computer Science University of Illinois at Chicago, Web Data Extraction Based on Partial Tree Alignment
- P. V. Praveen Sundar1 1Research Scholar, Hindusthan College of Arts &Science, Coimbatore, India, Towards Automatic Data Extraction Using Tag and Value Similarity Based on Structural-Semantic Entropy
- T. Seeniselvi1, N. Thangamani. 2 1Associate Professor, Hindusthan college of Arts and Science, India, Semantic Similarity Based Data Alignment and Best Feature Extraction using PSO for Annotating Search Results from Web Databases
- Hongjun Lu, Integrating Database and World Wide Web Technologies, National University of Singapore
- Jungwha Hong, How to Build a Web Database: A Case Study, The University of Texas at Austin.
- Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng, Member, IEEE, and Clement Yu, Senior Member, IEEE, Annotating Search Results from Web Databases

Index Terms

Computer Science Web Services

Keywords

Web Database Annotation Data alignment Annotation Wrapper