Abstract

Websites in today world consist of a large amount of data as per the requirements of the users. So web data extraction systems helps user in extracting the required data from these types of websites. The basic techniques used for web data extraction are manual and web wrapper. Web wrapper further consists of wrapper induction and automatic approaches. A lot of methods are available which uses wrapper induction and automatic methods. This research work provides performance comparison of manual, web wrapper induction and automatic approaches on the basis of methods chosen as manual (By manual efforts), nX1 (web wrapper induction), DEPTA and MDR (Automatic). The results are compared on the basis of various parameters like precision, recall, F-measure and data extraction time.

References


3. Y. Zhai, B. Liu, "Web data extraction based on partial tree alignment", in WWW, pp. 76-85, 2005.


Index Terms

Computer Science  Information Systems

Keywords

Web data extraction, manual, web wrapper, nX1, Depta, MDR.