Abstract

Web contain huge amount of information on Web sites the user can retrieve this with help of the search input query to Web databases & fetch the relevant information. Perhaps Web databases return the multiple search output records dynamically on Web browser, these search record are containing the Deep Web pages in the form of HTML pages. It is time consuming & human efforts are involved. The traditional search engine does not index the hidden Web pages from Web databases, such as (Google, Yahoo etc.). Many existing proposed techniques have addressed the problem of how to extract efficient structure data from Deep Web. The deep web refers to the hidden database used by web sites. But the information extraction & annotation is key challenge in web mining. The information retrieval should be done automatically & arrange in a systematic way for further processing. Various
methodologies like wrapper induction is been induced. The labeling is done to the extracted information as per the concept. Various types of annotators are used on the basis of the data to be annotated. In this paper survey the automatic annotation approach on the basis of different feature of text node and data units.

References


Index Terms

Computer Science

Information Sciences
Keywords
Data Extraction  Data annotation  Annotators  Text nodes  Data Units and Wrapper