To devise vision of the next generation of the web, deep web technologies have gained larger attention in a last few years. An eminent feature of next generation of web is the automation of tasks. A large part of Deep web comprises of online structured domain specific databases that are accessed using web query interfaces. The information contained in these databases is
related to a particular domain. This highly relevant information is more suitable for satisfying the information needs of the users and large scale deep web integration. In order to make this extraction and integration process easier, it is necessary to classify the deep web databases into standard non-standard category domains. There are mainly two types of classification techniques i.e. manual and automatic. As the size of deep web is increasing at an exponential rate with the passage of time, it has become nearly impossible to classify these deep web search sources manually into their respective domains. For this purpose, several automatic deep web classification techniques have been proposed in the literature. In this paper apart from the literature survey, we propose a framework for analysis of automatic classification techniques of deep web. The framework provides a baseline for the analysis of rudiments of automatic classification techniques based on the parameters such as structured, unstructured, simple/advance query forms, content representative extraction methodology, level of classification, performance evaluation criteria and its results. Furthermore, we studied a number of automatic deep web classification techniques in the light of proposed framework.

Reference

- Tiezheng Nie, Derong Shen, Ge Yu, Yue Kou: Subject-Oriented Classification Based on Scale Probing in the Deep Web. WAIM 2008: 224-229
- A helpful guide to search engines: http://www.searchengineguide.com/

**Index Terms**

Computer Science

Information Retrieval

**Key words**

Deep web
web databases
data integration

domain concepts

Survey