Abstract

A large amount of information on the Web is contained in regularly structured objects, which call data record. Data record are important because often present the essential information of it host pages, e.g, list of products and services. It is useful to mine such data record in order to extract information from web pages to provide value-added services.

In this Final project is implemented method for mining data records in web pages automatic with use algorithm is called MDR (*Mining Data Records in Web Pages*). This technique is more effective because just based on two important observations i.e, observe data records in Web pages and a string matching algorithm. In the process of mining data records are three main steps, i.e, building a HTML tag tree of the page, mining data regions in the Web pages using the tag tree and string comparison, and identifiying data records from each data region.

From implementation and analyse stage shown that MDR algorithm is built proved to can find out *data rcords* in Web pages though it is noise.

Keywords: Web Mining, data mining, tag tree HTML, data region, data record.