Abstract

A website has contents that shown based on blocks in the web. In the shown web page not only included main contents from that web page, but also any parts which not relevant with information contents has shown. For example advertisement, navigation panel, user guide, links etc. The data of composer information has shown in the web pages will be construct into regular object structure and has fixed template (data record).Data record will be used as approach for information extraction on the web page(mining data record). 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 DEPTA (*Data Extraction with Partial Tree Alignment*)[1]. This technique is made to develop MDR (*Mining Data Records in Web*)[2] algorithm i.e add to three step i.e *tree edit distance* that implemented *simple tree matching* together and *partial tree alignment*,on the three step before always use *tree (subtree) matching* principle.

From implementation and analyze stage shown that DEPTA algorithm is built proved to can find out *data records* in Web pages and capable to reduce *noise* in web pages better than MDR.

Keywords: Web Mining, partial tree alignment, data region, data record, simple tree matching, mining data record.