Abstract

Data mining is the procedure of extracting or taking out the information from the huge set of data. Web Mining is an important application of data mining, which is to extract knowledge from Web data including Web documents, hyperlinks, usage logs of web sites, etc. A Web Page contains many blocks such as content blocks, copyrights, privacy notes and advertisements. These blocks like advertisements and copyrights etc. don’t come under main content blocks. These blocks are known as noisy blocks or it can be said that these blocks contain noisy information. This noisy information adversely effects web data mining. Eliminating this noisy information will improve web data mining. In this paper, it will be discussed how to identify these noises and how to eliminate them to improve efficiency of web mining. There are many types of algorithms which are used in web mining i.e. Visitor method, Dom Tree. Visitor and Dom Tree both are complex and time consuming methods. We will also discuss removal of noises by using simple LRU algorithm and variants of LRU, which will result into less time consuming algorithm for web mining.
Enhancing Performance of Web Page by Removing Noises using LRU

- K. Rajkumar (2011). "Dynamic Web Page Segmentation Based on Detecting Reappearance and Layout of Tag Patterns for Small Screen Devices".
- K. S. Kuppusamy (2011), "A Model for Web Page Usage Mining Based on Segmentation"
- Swe Swe Nyein (2011) "Mining Contents in Web Page Using Cosine Similarity"
- Shuang Lin, Jie Chen, Zhendong Niu (2012) "Combining a Segmentation-Like Approach And A Density-Based Approach In Content Extraction" TSINGHUA SCIENCE AND Technology issnll1007-0214ll05/18llpp256-264 Volume 17.

Index Terms

Computer Science Web Services

Keywords

Content Extraction DOM Tree LRU Web Mining