A Survey on Web Content Mining Techniques and Its Tools

Yi Yi Aung
yiyiaung123@gmail.com
University of Computer Studies, Mandalay, UCSM

Ei Ei Thu
eieithuet@gmail.com
University of Computer Studies, Mandalay, UCSM

The World Wide Web (WWW) is a rich source of information and continues to expand in size. The complexity at the WWW makes it difficult for data mining. Before the Web, finding information meant asking a friend or an expert, or buying/borrowing a book. However, it is becoming a challenging task to effectively and efficiently retrieve the required web page /information on dynamic and heterogeneous web documents. If we can efficiently use web content mining techniques and tools, we can easily extract useful information from various forms of web page content by eliminating noisy information. This paper presents different web content mining techniques such as structured, unstructured, semi-structured and multimedia and five tools that can collect the appropriate information from websites that the user requires. The focus of this paper is to bring to light the value of Web Content Mining techniques and its tools in virtual society.