Abstract

Classification of Web pages is one of the challenging and important task as there is an increase in web pages in day to day life provided by internet. There are many ways of classifying web pages based on different approach and features. This paper explains some of the approaches and algorithms used for the classification of webpages. Web pages are allocated to pre-determined categories which is done mainly according to their content in Web page classification. The important technique for web mining is web page classification because classifying the web pages of interesting class is the initial step of data mining. The agenda of this paper is first to introduce the concepts related to web mining and then to provide a comprehensive review of different classification techniques.

References


13. Chen H, Fuller SS, Friedman C, Hersh W. Knowledge management, data mining, and text mining in medical informatics. InMedical Informatics 2005 (pp. 3-33). Springer US.


Index Terms
Keywords

Web page classification, Web Mining, Data Mining, uniform resource locator (URL), SVM (Support Vector Machine), KNN (K-Nearest Neighbor), Naïve Bayes, Artificial Neural Network.