Abstract

The use of internet has been increasing day by day. The users can find their resources with the help of different hyperlinks. These usages of Internet have led to the invention of web crawlers. The search engine which helps the user to explore the web is known as Web Crawler. In web crawlers the crawled data can be used to find missing links, community detection in complex networks. The concept of providing accuracy for this is forever in the vein. In this paper, web crawlers: their architecture, process of semantic focused crawling technology, ontology learning, pattern matching, types and various challenges being faced when search engines use the web crawlers, have been reviewed.

The web results more relevant to the user query through keyword expansion have been retrieved by the system. This data is being use further for the efficient association rule mining using Eclat Algorithm which is weaved for the vertical transactions based scheme. This process is being powered with Shannon information gain to identify the important words for the frequent pattern mining, and the whole process is being catalyzed by the fuzzy logic classification for
more pattern identification process.

References


Index Terms

Computer Science  Pattern Recognition

Keywords

Web crawler, Shannon information gain, Association Rules, Eclat Algorithm and Fuzzy.