Abstract

In recent years' growth of World Wide Web become faster to cross all expectations; World Wide Web is becoming most valuable resources to information retrieval and knowledge discovery from Web. It is a fertile area for web mining research; an emerging challenge for web mining is the problem of mining richly qualitative documents, where the objects are linked via multiple types of relations. These links provide additional context that can be helpful for web mining tasks. Traditional link analysis treats all hyperlinks equally and makes the assumption that links are endorsement, so there is need to only extract links which are valuable. Unfortunately, this assumption does not incorporate in present World-Wide Web. Hyperlinks are not identical; they may be created in different contexts and for different purposes. By using novel characteristics of web page and hyperlinks to help a search engine focus on relevant and high quality content. The important hyperlink features—topicality—is proposed.

References


Anchor Information SIGIR’01, New Orleans, Louisiana, USA. ACM 1-58113-331-6/01/0009, 250-257.

Index Terms

Computer Science  Information Sciences

Keywords

Web Mining, Web Structure Mining, Topicality, Information Retrieval, Link Analysis, Anchor Text, WWW.