عنوان مقاله:

New approach for web page classification based on URL and semantic analysis

نویسندگان:

Maide Abedini Bagha - *Young Researchers and Elite club, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

Somayeh Dahmardeh Kemmak - *Islamic Azad University, Zahedan, Iran*

خلاصه مقاله:

Traditional information retrieving methods use keywords occurring in determine the class of web pages, but usually retrieved unrelated web pages. W3 consortium stated that HTML dosnt provide a better description of semantic structure of the web page contents, because of its limited semi structure data, case sensitivity, predefined tags and so on. To overcome these backs, Web developers started to develop web pages on XML, flash kind of new technologies. It makes a way for new research methods. In this article we propose a new approach based on URL and semantic analysis for classifying XML and other types of web page.

کلمات کلیدی:

URL, Web page classification, semantic content, semantic structure