Abstract

In the World Wide Web, or simply the web, the content of information is changing everyday and it is known as dynamic environment. There is more information are uploaded in web and it has grown steadily in recent years. Therefore the several billions of HTML documents, pictures and another multimedia files available on the Internet. Due to the overloaded of information in web, the information extraction is not effectively based on user needs. To overcome the above problem, there is a need of methods to help us extract information effectively from the content of web pages. Nowadays, various web content mining techniques are developed to mine the information and serve people in a simple way: These techniques focuses on the discovery/retrieval of the useful information from the Web contents/data/documents. This paper focus on how to extract the information effectively based on classification and clustering, and detecting phishing websites.

References

No. 3, pp. 1339-1351.

Index Terms

Computer Science

Web Mining

Keywords

Web Content mining  classification  clustering  phishing Websites.