

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

*IJCSMC, Vol. 3, Issue. 2, February 2014, pg.374 – 381*

### **RESEARCH ARTICLE**

# **A Supervised Method for Multi-keyword Web Crawling on Web Forums**

**A.Gowtham<sup>1</sup>, Dr.K.Deepa<sup>2</sup>**

<sup>1</sup>Department of CSE (PG), Sri Ramakrishna Engineering College, India  
E-mail: [gowslm@rediffmail.com](mailto:gowslm@rediffmail.com)

<sup>2</sup>Department of IT, Sri Ramakrishna Engineering College, India  
E-mail: [deepkarun@rediffmail.com](mailto:deepkarun@rediffmail.com)

Abstract – Web forums are used by large number of users to post and share their comments with other users of various websites. The forums consist of many lists of topics on their boards with a large list of threads in each board. The users can create many threads and share their views in posts as well. In this paper a supervised web forum multi-keyword crawler is proposed to crawl relevant contents from the forum pages by reducing the delay. All the forums in the web have navigation paths that lead to the forum threads and these paths are connected by specific types of URLs. Thus the proposed method needs to recognize the various URLs by using the regular expression patterns within the forum. Accurate page classifiers trained by using other forums can be used to classify the regular expression patterns and detect the URLs. The obtained results show that the proposed method is more reliable and accurate comparing to other existing methods.

Keywords - Web crawler, page classification, forum crawler, URL based learning.

Full Text: <http://www.ijcsmc.com/docs/papers/February2014/V3I2201484.pdf>