Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 8, August 2013, pg.243 – 247

RESEARCH ARTICLE

Design of Improved Web Crawler By Analysing Irrelevant Result

Prashant Dahiwale[1], Dr. M.M. Raghuwanshi[2], Dr. Latesh Malik[3]

[1]Research Scholar, Dept. of Computer Science & Engineering, G.H.Raisoni College of Engineering, India

[2]Professor, Dept. of Comp Sc. Engg, Rajiv Gandhi College of Engineering & Research, Nagpur, India

[3]Professor, Dept. of Comp Sc. Engg, G.H.Raisoni College of Engineering, Nagpur, India

[1] prashantdd.india@gmail.com; [2] m.raghuwanshi@rediffmail.com; [3] latesh.malik@raisoni.net

Abstract— A key issue in designing a focused Web crawler is how to determine whether an unvisited URL is relevant to the search topic. Effective relevance prediction can help avoid downloading and visiting many irrelevant pages. In this module, we propose a new learning-based approach to improve relevance prediction in focused Web crawlers. For this study, we chose Naïve Bayesian as the base prediction model, which however can be easily switched to a different prediction model. The performance of a focused crawler depends mostly on the richness of links in the specific topic being searched, and focused crawling usually relies on a general web search engine for providing starting points.

Key Terms: - URL; focused crawler; classifier; relevance prediction; links; search engine; ranking

Full Text: http://www.ijcsmc.com/docs/papers/August2013/V2I8201356.pdf