



RESEARCH ARTICLE

A Novel Architecture of Mercator: A Scalable, Extensible Web Crawler with Focused Web Crawler

Sarnam Singh¹, Nidhi Tyagi²

¹M.Tech, School of Computer Engineering & I.T., Shobhit University (Shobhit Institute of Engineering & Technology), (Deemed – to-be University), Meerut, U.P, India

²Assistant Professor, School of Computer Engineering & I.T., Shobhit University (Shobhit Institute of Engineering & Technology), (Deemed – to-be University), Meerut , U.P, India

¹ sarnammb@gmail.com, sarnamsingh@yahoo.co.in

Abstract— This Paper described A Novel Architecture of Mercator: A Scalable, Extensible Web Crawler with Focused Web Crawler. We enumerate the major components of any Scalable and Focused Web Crawler and describe the particular components used in this Novel Architecture. We also describe this Novel Architecture support for Extensibility and downloaded user’s support information. We also describe how the Focused Web Crawler component integrates with Mercator: A Scalable, Extensible Web Crawler and also describe their functionality of every component and how to work together. We also describe how this Novel Architecture downloaded maximum pages from web in minimum time and sure partially extract web pages which is needed to users.

Full Text: <http://www.ijcsmc.com/docs/papers/June2013/V2I6201372.pdf>