Abstract

Internet forum are important service where user can request and exchange information with other. The Focus(Forum crawler Under Supervision), it is web-scale forum crawler. A Crawler traverses the World Wide Web in a systematic manner with intention of gathering data or knowledge. The goal of Focus is to crawl applicable forum content from the web. Web crawlers following the hyperlinks in Web pages to automatically download a partial snapshot of the Web. Based on this observation, smaller the forum web crawling problem to a URL-type. Although forum have different layouts or styles and different forum software packages. They have always similar implicit navigation path connected by specific URL types to users from entry
pages to thread pages.

References


Index Terms

Computer Science    Artificial Intelligence

Keywords

Eit Path    Forum Crawling    Itf Regex    Url Type    Page Classification    Page Type.