

Abstract

Our goal is to develop a tool for automatic website checking. We created an extensible system that checks the validity of links as well as HTML and CSS syntax. For that, we integrated existing tools and libraries into one application using plugins. Our tool looks for links in HTML and sitemaps. Also, robots.txt including crawl-delay implementation and sitemap discovery is supported. The course of inspection can be affected by configuration rules. Plugin categorization allows for customization of inspection. Not only we verify the validity of standard hypertext links but also links through other HTML tags (images, frames, cascading styles, scripts). In addition, canonical addresses and redirection, duplicates in content or usage of non-semantic tags or attributes are detected. The developed application is extensible for other types of checks.