Google opens the source for its robots.txt parser in Java and testing framework in C++

Last year, Google open sourced the code for the robots.txt parser used in its production systems. After seeing the community build tools with it and add their own contributions to the open source library, including language ports of the original parser written in C++ to golang and rust, Google announced this week it has released additional related source code projects.

Here's what's new for developers and tech SEOs to play with.

C++ and Java. For anyone writing their own or adopting Google's parser written in C++ (a super fast compiled language), Google has released the source code for its robots.txt parser validation testing framework used to ensure parser results adhere to the official robots.txt specification as expected, and it can validate parsers written in a wide variety of other languages.

Additionally, Google released an official port to the more popular Java language. Modern Java is more widely used in enterprise applications than C++, whereas C++ is more typically used in core system applications where performance needs demand it. Some Java-based codebases run applications today for enterprise SEO and or marketing software.

Testing and validation. Requirements for running the test framework include JDK 1.7+ for Apache Maven, and Google's protocol buffer to interface the test framework with your parser platform and development workstation. It should be useful to anyone developing their own parser, validating a port, or utilizing either of Google's official parsers, and especially for validating your development of a port to a new

language.

How difficult would this be to use? We should note these are relatively approachable intern-led projects at Google which ought to be consumable by moderate to higher level programmers in one or more of these languages. You can build a robots.txt parser using practically any programming language. It adds perceived authority, however, when your marketing application runs the exact same parser that governs Googlebot.

Why we care. If you, or your company, has plans to write or has written a crawler which parses robots.txt files for directives looking for important information (not just) for SEO, then this gives you incentive to evaluate whether using Google's parser in C++, Java, or one of the other language ports is worth it. The Java parser in particular should be relatively easy to adopt if your application is already written in Java.

About The Author

Detlef Johnson is the SEO for Developers Expert for Search Engine Land and SMX. He is also a member of the programming team for SMX events and writes the SEO for Developers series on Search Engine Land. Detlef is one of the original group of pioneering webmasters who established the professional SEO field more than 20 years ago. Since then he has worked for major search engine technology providers, managed programming and marketing teams for Chicago Tribune, and consulted for numerous entities including Fortune 500 companies. Detlef has

a strong understanding of Technical SEO and a passion for Web programming.
http://feeds.searchengineland.com/~r/searchengineland/~3/3QWkWiP0f88/google-opens-the-source-for-its-robots-txt-parser-in-java-and-testing-framework-in-c-340920