Solr and Spark: Two Halves Make One Whole Content Indexing Powerhouse

December 12, 2016

The Internet is an undeniably powerful tool greatly because it makes accessing information so easy. That's why search engines are so popular - Google has become a synonym for "search" (and might soon become one for "Internet"). Yet, while Google can lead visitors to a site, it cannot necessarily help those visitors quickly find what they're looking for within it.

For this reason, we here at Six Feet Up have become interested in two separate programs that, when used together, can elevate a website's internal searches or product recommendations, and therefore its number of visitors and conversions. The names of these programs? Solr and Spark.

Solr is an enterprise search appliance that creates software-based indexing of content on a site's internal search engine. Spark is a big data tool for running parallel jobs against a cluster of machines to ask certain kinds of questions and produce certain kinds of answers.

Combining these two technologies allows a site to either provide visitors with most relevant search results or with links to the most popularly trending articles or products. They enable results and suggestions to change over time based on what previous users have searched for, how they've searched, what keywords they've used, and what links they've viewed. Using Solr with Spark makes it possible to combine front end UIQs, so certain keywords or links can be considered trending or popular depending on how many times they've been searched for or acted upon.

This happens because Spark simultaneously processes all of this activity data and sends the information back to Solr as relevancy indices. In essence, a real-time popularity catalog is created between these two programs.

For example, if the software realizes most people searching for soccer shin guards end up clicking on the seventh search result, they will automatically present the seventh link first to future visitors.

These programs work so well together because the team that built Solr also created a plugin for the program that permits Spark to provide the Solr with native data frames. Thanks to this plugin, very little translation is needed between the two for one to bring in data from the other and vice versa.

Spark also has other benefits to the Solr software. For instance, Spark allows Java to run parallel to Solr in order to re-index site content much more quickly using a Spark duster to scale up and down as needed. It can also come in handy for offline computations of pagerank, popularity, or trends.

These powerful uses of Solr and Spark are able to go beyond just web signals as well. Email trackers, UI trackers inside ads, or any other data that can be gathered together as signals will provide site visitors with a better search experience. For instance, clicks made by recipients of email marketing messages can be tracked in order to influence which products or search results are presented in future communications and remarketing efforts.

Google may not be able to help site visitors navigate once they reach a landing page, but the partnership between Solr and Spark is paving the way for a better browsing and eCommerce experience through content indexing.

http://www.sixfeetup.com/blog/Solr-Spark-ab-testing-on-steroids