There’s a method behind StumbleUpon’s madness

August 4, 2008 | MG Siegler | View comments

StumbleUpon is all about site discovery. I used to click on the “Stumble!” button and figured it would return me some random site based on the categories I said I was interested in. But then I noticed that the more I used it, better sites were being sent my way. This is because it’s not actually random, but rather sites are served up based on a series of processes that go on within the StumbleUpon Recommendation Engine.

I had the chance to meet up with co-founder and chief architect Garrett Camp at the StumbleUpon offices last week. He walked me through (in laymen’s terms) what actually goes on in the backend when you click the Stumble button.

As you can see in the chart below, there are three key parts to the Recommendation Engine. There are pages from the topics you marked that interest you, socially endorsed pages and peer endorsed pages. Socially endorsed pages are the ones that users you have befriended on the site like, while peer endorsed pages are ones from users who have similar voting habits (giving a site the thumbs up or thumbs down) as you.

These three factors are why it’s important to not only choose categories you like, but to choose friends with similar interests and to only vote up sites you really enjoy in order to get the best experience out of StumbleUpon.

When a site is first stumbled, it is put through both the Classification Engine and the Clustering Engine.
as shown above. The Classification Engine filters the page by topic and tags. Sometimes a user does this work, but sometimes it’s submitted without any of this information, so the engine has to determine where to put the content. This is a big job when you have over 30,000 pages each day being submitted, as StumbleUpon has.

The Clustering Engine sorts out the votes a site is getting so it can determine which sites are the quality ones that should be served. Again, this sounds simple enough until you realize that StumbleUpon has 5.6 million users. This engine is a key cog in what serves up over 10 million stumbles that take place every day.

Like any good social algorithm maker, Camp wouldn’t divulge all the little details of what goes into the promotion of sites, but he did say that things such as comments on stories and so called “quick stumbles” (when a user quickly hits the stumble button again after landing on a page without voting on it — they dub this a “soft not for me,” or down-vote) are taken into account as well.

This all makes for a system of “quality plus relevance,” as Camp put it.

I was interested to know how this method compared to Digg’s recently launched Recommendation Engine. Camp said he hadn’t to look to closely at it yet, but that it seemed to employ many of the same ideas minus much of the content analysis.

As with any of these recommendation engines, the more data you have, the better it’ll perform. Since it was bought by eBay in May of last year, StumbleUpon has more than doubled its user base, but the company knows that this growth can only last so long given its major restriction: Right now, the vast majority of people who use StumbleUpon use it through a browser plug-in. This limits the service to either Firefox or Internet Explorer users (who happen to have this plug-in). That is why the team is pushing hard to perfect a web-only version of the site.

Creating a way to use the service no matter what browser you are on or what plug-ins you have installed could take the service to an even bigger level in terms of usage, Camp acknowledged.

You can see an example of how the web version of StumbleUpon looks here. (And above.)