Yana Volkovich

When you use a search engine, you want the best possible results. Yana Volkovich is conducting research into how applied mathematics can be employed to improve web searching.

Ranking the pages better

"Websearch is a new area of research in which academics from different areas - such as computer scientists and mathematicians - are active. Their aim is to improve websearch results by using algorithms and study the structure of the Web at the same time. The idea behind my research is to apply mathematics to websearch and, in particular, to the well-known algorithm PageRank."

"Selecting a search at random, most of the hits are often useless. The issue for most users is how to obtain the best search hits as quickly as possible. How can you best sift through all the useless pages to get the best hits? This is where applied mathematics comes in."

"Users generally only look at the first page of search results, occasionally the next one as well. All the rest are ignored. Yet the first hits presented are far from always the best. The first 10 are selected based on the search terms and specified parameters. Just because a page is in the first 10 for a particular search, doesn't mean it will always be in the first 10. It depends on various factors, such as the search terms, and the incoming and outgoing links. Sometimes, a user gets a good first 10, all useful hits. The next time, they might all be useless, some of them spam. PageRank helps to solve this problem. By analyzing how it does this, I'm trying to gain an understanding of websearch and improve PageRank's performance."

"PageRank uses two parameters to show the value of a page. The method first examines the number of links that point to your page hit. It then examines the status of the pages that point to your page hit. The more significant the pointing pages, the higher the score for the page hit. If it's a spam page, the score is much lower than for a valid hit. We're using our stochastic model to try to explain the relationship between PageRank and the various parameters of websearch."

> 'The more significant the pointing pages, the higher the score for the page hit.'

"A major problem we've encountered with PageRank is the ranking method. What is the simplest way to rank pages? A key feature of PageRank is its range, with its huge, dynamic reach. You could even compare the spread with that of a city's population, the size of a data file, or the number of books sold. The average value doesn't tell you too much. The ones you actually encounter can be a large distance away from the average. With PageRank, you also have to consider various parameters that affect the range in a non-trivial way. You don't only have to examine the number of incoming links, but also the PageRank status of the pages that point to your page hit. Hence, PageRank's view of the Web is at the global level. This doesn't make the analysis any easier. We proposed using special techniques from the Theory of Regular Variation and the Extreme Value Theory to analize the ranges."

"The Web has existed for roughly 20 years, and Google has been around for about 10. Research into websearch is new, barely out of its infancy. For users, it's rather uncharted terrain, and most researchers in the field differ in their backgrounds. During the first 10 years of the Internet, it only grew by hundreds of pages annually. Today, that figure is in the billions. The Web is growing and changing extremely rapidly. In my view, this will never stop. Because there's so much data on the Internet, a new field of research has opened up, one that will be around long enough for my career not to encounter any barriers. It's a new and interesting research field, so a lot can still happen in it. It's home to many new ideas and enthusiastic people. For everyone, the field still remains new. With a good idea, you have a strong probability of carrying out worthwhile research. It keeps growing. It's changing the world!"