Search 3.0: Present, Personal, Precise

Nova Spivack

Radar Networks, USA

The next generation of Web search is already beginning to emerge. With it we will see several shifts in the way people search, and the way major search engines provide search functionality to consumers.

To understand movement towards the next generation of Web search, it is helpful to first look at how it has developed to its current state. Beginning with Tim Berners-Lee's foundational proposal in 1989, the Web's evolution may be divided into three distinct phases.

Web 1.0, the first decade of the Web (1989 - 1999), was characterized by a distinctly desktop-like search paradigm. The overriding idea was that the Web is a collection of documents, not unlike the folder tree on the desktop, that must be searched and ranked hierarchically. Relevancy was considered to be how closely a document matched a given query string.

Web 2.0, the second decade of the Web (1999 - 2009), ushered in the beginnings of a shift towards social search. In particular, blogging tools, social bookmarking tools, social networks, social media sites, and microblogging services began to organize the Web around people and their relationships. This added the beginnings of a primitive "web of trust" to the search repertoire, enabling search engines to begin to take the social value of content (as evidenced by discussions, ratings, sharing, linking, referrals, etc.) as an additional measurement in the relevancy equation. Those items that were both most relevant on a keyword level, and most relevant in the social graph (closer and/or more popular in the graph), were considered to be more relevant. Thus results could be ranked according to their social value - how many people in the community liked them and current activity level - as well as by semantic relevancy measures.

In the coming third decade of the Web, Web 3.0 (2009 - 2019), there will be another shift in the search paradigm. This is a shift to from the past to the present, and from the social to the personal, and from the generic to the precise.

Established search engines like Google rank results primarily by keyword (semantic) relevancy. Social search engines rank results primarily by activity and social value (Digg, Twine 1.0, etc.). But the new search engines of the Web 3.0 era will also take into account two additional factors when determining relevancy: timeliness and personalization. The result will be a more precise and customized search capability.

Google returns the same results for everyone, but why should that be the case? In fact, when two different people search for the same information, they may want to elicit very different kinds of results. Someone who is a novice in

a field may want beginner-level information to rank higher in the results than someone who is an expert. There may be a desire to emphasize things that are novel over things that have been seen before, or that have happened in the past; the more timely something is the more relevant it may be as well. These three themes - present, personal, and precise - will define the next great search experience.

To accomplish this, we need to make progress on a number of fronts. First of all, search engines need better ways to understand what content is, without having to do extensive computation. The best solution for this is to utilize metadata and the methods of the emerging semantic web.

Metadata reduces the need for computation in order to determine what content is about - it makes that information explicit and machine-understandable. To the extent that machine-understandable metadata is added or generated for the Web, it will become more precisely searchable and productive for searchers.

This applies especially to the area of the real-time Web, where, for example, short "tweets" of content contain very little context to support good natural-language processing. In such cases, a little metadata can go a long way. In addition, of course, metadata makes a dramatic difference in search of the larger non-real-time Web as well.

Beyond metadata, search engines need to modify their algorithms to be more personalized. Instead of a "one-size fits all" ranking for each query, the ranking may differ for different people depending on their varying interests and search histories.

Finally, to provide better search of the present, search has to become more realtime. To this end, rankings need to be developed that surface not only what just happened now, but what happened recently and is also trending upwards and/or of note. Realtime search has to be more than merely listing search results chronologically. There must be effective ways to filter the noise and surface what's most important effectively. Social graph analysis is a key tool for doing this, but in addition, powerful statistical analysis and new visualizations may also be required to make a compelling experience.

The pace at which Semantic technology is finally developing makes the shift into the third generation of Web search now realizable. Tools operating within the new paradigm of present (Oneriot, Topsy, Twitter), personalized (My6Sense, Siri, Twine), and precise (Bing/Powerset, Hakia, WolframAlpha) are already leading the way in expanding conceptions of how search can function. Search 3.0 will bring a breakthrough in the Semantic indexing of the entire Web, allowing vertical search to be taken to its ultimate conclusion.