Web IR Research: Can we do it without the data?

Amit Singhal
Google, Inc.

Abstract

The field of Information Retrieval (IR) has long tackled the problem of finding useful information. With the prevalence of the Web, this problem has compounded manifold. Even though IR methods and web search should make a perfect combination, IR research has had very little impact on web search. Why? "The devil is in the data." The notion of relevance is quite different on the web, as compared to more traditional IR test collections. In this talk, I will argue that having the right data is critical for developing effective web search algorithms.