

The Next Generation Web Search and the Demise of the Classic IR Model

Andrei Broder

Yahoo! Research, USA

broder@yahoo-inc.com

Abstract. The classic IR model assumes a human engaged in activity that generates an “information need”. This need is verbalized and then expressed as a query to search engine over a defined corpus. In the past decade, Web search engines have evolved from a first generation based on classic IR algorithms scaled to web size and thus supporting only informational queries, to a second generation supporting navigational queries using web specific information (primarily link analysis), to a third generation enabling transactional and other “semantic” queries based on a variety of technologies aimed to directly satisfy the unexpressed “user intent”, thus moving further and further away from the classic model.

What is coming next? In this talk, we identify two trends, both representing “short-circuits” of the model: The first is the trend towards context driven Information Supply (IS), that is, the goal of Web IR will widen to include the supply of relevant information from multiple sources without requiring the user to make an explicit query. The information supply concept greatly precedes information retrieval; what is new in the web framework, is the ability to supply relevant information specific to a given activity and a given user, while the activity is being performed. Thus the entire verbalization and query-formation phase are eliminated. The second trend is “social search” driven by the fact that the Web has evolved to being simultaneously a huge repository of knowledge and a vast social environment. As such, it is often more effective to ask the members of a given web milieu rather than construct elaborate queries. This short-circuits only the query formulation, but allows information finding activities such as opinion elicitation and discovery of social norms, that are not expressible at all as queries against a fixed corpus.