Analyzing Behavioral Data for Improving Search Experience

Pavel Serdyukov
Yandex
Leo Tolstoy 16
Moscow, Russia
pavser@yandex-team.ru

ABSTRACT

Yandex is one of the largest internet companies in Europe, operating Russia's most popular search engine, generating 62% of all search traffic in Russia, what means processing about 220 million queries from about 22 million users daily. Clearly, the amount and the variety of user behavioral data which we can monitor at search engines is rapidly increasing. Still, we do not always recognize its potential to help us solve the most challenging search problems and do not immediately know the ways to deal with it most effectively both for search quality evaluation and for its improvement. My talk will focus on various practical challenges arising from the need to "grok" search engine users and do something useful with the data they most generously, though almost unconsciously share with us. I will also present some answers to that by overviewing our latest research on user model based retrieval quality evaluation, implicit feedback mining and personalization.

I will also summarize the experience we gained from organizing three data mining challenges at the series of workshops on using search click data (WSCD) organized in the scope of WSDM 2012 – 2014 conferences. These challenges provided a unique opportunity to consolidate and scrutinize the work from search engines' industrial labs on analyzing behavioral data. Each year we publicly shared a fully anonymized dataset extracted from Yandex query logs and asked participants to predict editorial relevance labels of documents using search logs (in 2011), detect search engine switchings in search sessions (in 2012) and personalize web search using the long-term (user history based) and short-term (session-based) user context (in 2013).