Mining Web Logs: Applications and Challenges

Ravi Kumar
Yahoo! Research
701 First Ave.
Sunnyvale, CA
ravikumar@yahoo-inc.com

ABSTRACT

Web logs record the primary interaction of users with web pages in general and search engines in particular. There are two sources for such logs: user trails obtained from toolbars and query/click information obtained from search engines. In this talk we will address the task of mining this rich data to improve user experience on the web. We will illustrate a few applications, together with the modeling and algorithmic challenges that stem from these applications. We will also discuss the privacy issues that arise in this context.

Categories and Subject Descriptors

H.2.8 [**Information Systems**]: Database Applications—*Data Mining*

General Terms

Algorithms, Experimentation, Measurement

Keywords

Query logs, Click logs, Toolbar trails

Biography

Ravi Kumar joined Yahoo! Research in July 2005. Prior to this, he was a research staff member at the IBM Almaden Research Center in the Computer Science Principles and Methodologies group. His primary interests are web algorithms, algorithms for large data sets, and theory of computation. He obtained his PhD in Computer Science from Cornell University in December 1997.