



JOURNAL OF THE INSTITUTE FOR OPERATIONS RESEARCH AND THE MANAGEMENT SCIENCES

MARKETING SCIENCE

Volume:

Number:

Year:

Title:

Author:

e-mail:

MktgSci@notes.cba.ufl.edu

Marketing Science Homepage

<http://bear.cba.ufl.edu/centers/MKS>

The Little Engines That Could: Modeling the Performance of World Wide Web Search Engines

Eric T. Bradlow • David C. Schmittlein

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6371
ebradlow@wharton.upenn.edu • schmittlein@wharton.upenn.edu

Abstract

This research examines the ability of six popular Web search engines, individually and collectively, to locate Web pages containing common marketing/management phrases. We propose and validate a model for search engine performance that is able to represent key patterns of coverage and overlap among the engines.

The model enables us to estimate the typical additional benefit of using multiple search engines, depending on the particular set of engines being considered. It also provides an estimate of the number of relevant Web pages *not* found by any of the engines. For a typical marketing/management

phrase we estimate that the “best” search engine locates about 50% of the pages, and all six engines together find about 90% of the total.

The model is also used to examine how properties of a Web page and characteristics of a phrase affect the probability that a given search engine will find a given page. For example, we find that the number of Web page links increases the prospect that each of the six search engines will find it. Finally, we summarize the relationship between major structural characteristics of a search engine and its performance in locating relevant Web pages.

(Capture/Recapture; Hierarchical Bayes; Marketing Information; Probability Models; World Wide Web)