

Connections 2010 Abstract

Author: Anamika Dasgupta

Abstract Title: Google Blog Search Engine v.s. Technorati Blog Directory: A Study on User Relevance Criteria

Affiliation: Long Island University

Abstract:

The increasing popularity of blogs and availability of numerous blog sites have given rise to blog search engines and blog directories such as Google Blog Search engine and Technorati Blog Directory. Although blog search engines and blog directories are information retrieval (IR) systems that are similar to web search engines and web directories, they are quite different in structure, use, and behaviour. Liu, Birnbaum, & Pardo (2008, 1525) pointed out that direct categorization of all the texts written in a blog cannot represent its contents accurately. This is because blog articles are written in an informal erratic style and bloggers do not confine themselves to one topic, creating a multi-topic document. In order to design an effective and improved IR system for blog sites, it is important to know what criteria users use in judging the relevance of information retrieved on blog search engines and directories. So far no single study has focused precisely on relevance criteria used by users in retrieving blog articles using blog search engines and directories.

Thus, this study examined the differences and similarities between Google Blog Search engine and Technorati Blog Directory and investigated the criteria users use in judging the relevance of retrieved results on Google Blog Search engine and Technorati Blog Directory.

Six participants were selected using simple random sampling method. Participants were given 2 task sheets they needed to complete: one task sheet describing a task pertaining to Google Blog Search engine and the other sheet describing a task related to Technorati Blog Directory. Data was collected using a questionnaire that had 3 parts. Part 1 was completed by the participants before they began working on given tasks. Part 2 was completed after they completed Task 1 and Part 3 was completed by them after they completed Task 2. The questionnaire contained both open-ended and closed-ended questions. A focused interview was conducted with 3 of the 6 participants who agreed to participate in the interview.

Results of the study showed that title and excerpts of blog articles are the relevance criteria that were used by participants for selecting blogs. Participants believed that browsing is time consuming and therefore, preferred to have a search box as in Google Blog Search. They believed that subject categories and sub-categories in Technorati Blog Directory were too broad, semantically overlapping, and not similar to the search keywords they would use to search for the topics. This shows that traditional subject categories, similar to Library of Congress subject categories and Sears subject categories are no longer preferred by participants when they are finding information on the Internet. Participants selected relevant links based on various combinations of words along with words indicated in the given activities. This shows that they

evaluate the title of blog articles based on the context, their prior knowledge, and the title of a given topic.

Thus, research on relevance criteria used by bloggers can reveal important information that can be used for the purpose of designing blog search engines and directories.

References:

Liu, Jiahui, Larry Birnbaum, and Bryan Pardo. (2008). Categorizing Blogger's Interests based on Short Snippets of Blog Posts. *ACM CIKM'08*, 1525-1526.