

An Initial Log-based Usage Analysis of Research Networking Systems

Mary Regina Boland^{1,2} BS, Stephen B Johnson³ PhD, Sylvia Trembowelski¹ MS, Daniel C Dine¹ BA, Richard C Steinman¹ BA, Suzanne Bakken^{1,2} DNSc RN, Chunhua Weng^{1,2} PhD

1. The Irving Institute for Clinical and Translational Research, New York, NY
2. Department of Biomedical Informatics, Columbia University, New York, NY
3. Weill Cornell Medical College, Cornell University, New York, NY

Abstract

Despite the proliferation of research networking systems, their adoption status and outcome data are scarce. To bridge this knowledge gap, we present an initial log-based usage analysis of a research networking system. We analyzed five months' log data and correlated usage patterns with user types and tasks. Scientific professionals spend more time performing deep search than generic Google users. Role-specific user support is important because usage patterns vary by user role. Faculty members performed more informational queries than researchers and administrators. We contribute initial knowledge of usage patterns of scientific professionals on a research networking system and a generalizable log analysis method for evaluating research networking systems.

Introduction

Interdisciplinary team science is increasingly needed in modern science. A big barrier to team science is to identify collaborators outside a researcher's own field. To overcome this barrier, research networking has grown rapidly in recent years. A variety of research networking systems have been developed, including Biomedical Experts, Collexis, ResearchGate, and VIVO, and many other home-grown systems. However, there is limited knowledge of how exactly research networking systems are adopted and used and if the current technology and networking features meet the information needs of scientific professionals. Our study here was driven to answer these questions.

Materials and Methods

We conducted a qualitative analysis of user feedback and a time series analysis of the log file for our institutional research networking system. The log data include user login, user type, keyword search, profile lookups for person, organization, or grants, and timestamps of user activities. We collected a 5-month worth of log data and segment log files into a hierarchical information structure so that a user's daily log contains multiple tasks, each beginning with a query, navigational (using person or organization names) or informational (anything except for navigational), followed by sequential profile lookups for persons or organizations or query reformulations. We abstracted usage patterns from this time series analysis and correlated usage patterns with user types.

Results

For 11,846 users invited to use the system, we only received a total of 973 logins by 510 distinctive users, including 43% faculty, 25% scientific administrators, and 24% non-faculty researchers. The adoption rate by busy scientific professionals was around 4%. Anecdotally, this rate may indicate the common low adoption status of most research networking systems. We rank queries by frequency and grouped them by semantic categories. The largest category of keyword searches is diseases, accounting for 30% of the searches, followed by the age group 27%, followed by the term "prevention" account for 18%. Only 3% of the query includes Boolean expressions, this may imply that our researchers may not be familiar with sophisticated search strategies. There were more person lookup (66%) than informational searches (27%). About 61% informational queries involve query reformulation as opposed to 13% of navigational queries, which implies that scientific professional needs support for query formulation to efficiently search research networking systems to meet their information needs.

Conclusion

This study contributes an initial usage analysis of a research networking system as well as an evaluation method and design knowledge for improving the effectiveness and usability of research networking systems.

Acknowledgement

The project described was supported by Grant Number UL1 RR024156 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).