

# How People Search for Governmental Information on the Web

**Bernard J. Jansen**

*The Pennsylvania State University, USA*

**Amanda Spink**

*Queensland University of Technology, Australia*

**Will Berkheiser**

*The Pennsylvania State University, USA*

## INTRODUCTION

People are now confronted with the task of locating electronic information needed to address the issues of their daily lives. The Web is presently the major information source for many people in the U.S. (Cole, Suman, Schramm, Lunn, & Aquino, 2003), used more than newspapers, magazines, and television as a source of information. Americans are expanding their use of the Web for all sorts of information and commercial purposes (Horrigan, 2004; Horrigan & Rainie, 2002; National Telecommunications and Information Administration, 2002). Searching for information is one of the most popular Web activities, second only to the use of e-mail (Nielsen Media, 1997). However, successfully locating needed information remains a difficult and challenging task (Eastman & Jansen, 2003). Locating relevant information not only affects individuals but also commercial, educational, and governmental organizations.

This is especially true in regards to people interacting with their governmental agencies. Executive Order 13011 (Clinton, 1996) directed the U.S. federal government to move aggressively with strategies to utilize the Internet. Birdsell and Muzzio (1999) present the growing presence of governmental Web sites, classifying them into three general categories, (1) provision of information, (2) delivery of forms, and (3) transactions. In 2004, 29% of American said they visited a government Web site to contact some governmental entity, 18% sent an e-mail and 22% use multiple means (Horrigan, 2004). It seems clear that the Web is a major conduit for accessing governmental information and maybe services. Search engines are the primary means for people to locate Web sites (Nielsen Media, 1997).

Given the Web's importance, we need to understand how Web search engines perform (Lawrence & Giles, 1998) and how people use and interact with Web search

engines to locate governmental information. Examining Web searching for governmental information is an important area of research with the potential to increase our understanding of users of Web-based governmental information, advance our knowledge of Web searchers' governmental information needs, and positively impact the design of Web search engines and sites that specialize in governmental information.

## BACKGROUND

There has been limited large-scale research examining Web searching for governmental information. Croft, Cook, and Wilder (1995) present analysis of the use of THOMAS, a governmental system that makes U.S. legislative information available to the public. The researchers report that searchers used very simple queries and appeared to have trouble locating specific bills. The researchers also noted that was some dissatisfaction with the relevance of returned results. Marchionini and Levi (2004) present an on-going study of the Bureau of Labor Statistics, reporting that user interfaces to governmental Web sites require special attention

Hargittai (2003) reports that users look for governmental information in a variety of ways with considerable temporal variance for task completion. The manner in which content is presented often confuses searchers. Hargittai reports that the two major sources of confusion concerned the uniform resource locator and the page design layout. Ceaparu and Shneiderman (2004) investigated alternate ways of organizing governmental statistics, reporting that users were more successful in finding answers when the information was organized into categories rather than an alphabetical listing.

There is a body of research focusing on general Web searching (Jansen, Spink, & Saracevic, 2000; Spink &

Jansen, 2004). Spink and Jansen (2004) report that searching for governmental information on the Excite search from 1997 to 2001 was about 1.5% to 3% of all queries. Jansen and Spink (2005) report that searching for governmental information on AlltheWeb.com, a Norwegian-based search engine, was about 2.0% of all queries in a study of datasets from 2001 and 2002. Jansen, Spink, and Pedersen (2005) report that searching for governmental information on AltaVista in 2002 was about 1.6%. Overall, few Web queries are related to government information; however, as a stand-alone category, governmental queries are a sizeable percentage.

Although these studies provide important insights into Web searching, further research is needed that validates these results for the searching of governmental information on Web search engines. This is especially important as Web searching systems are continually undergoing changes and governmental entities are moving more services to the Web (e.g., <http://usgovinfo.about.com/library/news/aainternet.htm>).

We address this need in the present study by examining a set of queries representing governmental-related information needs to analyze how people are searching for governmental information, including what information they are seeking. We also classify a sub-set of these queries to develop a taxonomy that can assist in the development and organization of governmental Web sites.

## FUTURE TRENDS

### Research Questions

The research questions driving this study are:

1. What are the characteristics of governmental Web searching
2. What types of governmental information are people searching for on the Web
3. How effective are these queries in locating governmental information

### Research Design

#### Data Collection

To address the first research question, we obtained, and qualitatively analyzed, actual governmental-related queries submitted to the AltaVista Web search engine. For this research question, we are interested in examining the characteristics of the governmental-related queries, investigating areas such as the number of terms in queries, the number of queries in a session, and the use of query

operators, among other aspects. Our analyses of searching behavior addressed the following approaches to Web searching behavior.

- a. **Query Length:** The query length is defined as the length, measured in terms, of the entire search query. This may include Boolean operators
- b. **Session Size:** A session is the total amount of queries submitted over a period of time. A session may include only one query, or may extend over a longer period of time and include multiple queries
- c. **Result Pages Viewed:** Result pages viewed is the number of pages returned by the search engine based on the query submitted that were actually visited by the user

The third research question involved classifying a sub-set of these queries with a controlled hierarchical vocabulary. This has implications not only for searchers using the Web to locate governmental information but also search engines and Web sites that serve these users.

To investigate our research questions, we gathered data from the AltaVista search engine. In 2002, Alta Vista was the 9<sup>th</sup> most popular search engine (Sullivan, 2002), had a content collection of 550 million Web pages (Sullivan, 2000), and approximately 5.6 million unique visitors per month. Overall, AltaVista offers a full range of searching options, has an extremely large content collection, and millions of unique visitors per month. After being an independent company for several years, Overture Services purchased AltaVista in 2003 (Morrissey, 2003).

We recorded the queries examined for this study on the AltaVista server on Sunday, September 8, 2002 and span a 24-hour period. We checked news stories from this day to see if any looked as if they may have influenced the investigation, namely the term analysis. There did not appear to be a major news story occurring on this date. However, the date is near the anniversary of the 9-11 attacks.

We recorded the queries in a transaction log that represents a portion of the searches executed on the Web search engine on this particular date. The original general transaction log contains approximately 1,000,000 records. Each record contains three fields:

1. **Time of Day:** Measured in hours, minutes, and seconds from midnight of each day as recorded by the AltaVista server
2. **User Identification:** An anonymous user code assigned by the AltaVista server
3. **Query Terms:** Terms exactly as entered by the given user

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/people-search-governmental-information-web/11614](http://www.igi-global.com/chapter/people-search-governmental-information-web/11614)

## Related Content

---

### Continuous Use Intention of E-Government Portals the Perspective of Older Citizens

Ching Seng Yap, Rizal Ahmad, Farhana Tahmida Newazand Cordelia Mason (2019). *International Journal of Electronic Government Research* (pp. 1-16).

[www.irma-international.org/article/continuous-use-intention-of-e-government-portals-the-perspective-of-older-citizens/231560](http://www.irma-international.org/article/continuous-use-intention-of-e-government-portals-the-perspective-of-older-citizens/231560)

### Towards Digital Governance in UK Local Public Services?

Ian McLoughlin (2009). *Handbook of Research on Strategies for Local E-Government Adoption and Implementation: Comparative Studies* (pp. 122-136).

[www.irma-international.org/chapter/towards-digital-governance-local-public/21458](http://www.irma-international.org/chapter/towards-digital-governance-local-public/21458)

### Natural Resource Information Management at State GovernmentLevel

Lindsay Redlich, Peter Gersekowskiand Michael Middleton Middleton (2008). *Electronic Government: Concepts, Methodologies, Tools, and Applications* (pp. 422-434).

[www.irma-international.org/chapter/natural-resource-information-management-state/9725](http://www.irma-international.org/chapter/natural-resource-information-management-state/9725)

### E-Government Performance Measurement: A Citizen-Centric Approach in Theory and Practice

Forrest V. Morgeson (2012). *Electronic Governance and Cross-Boundary Collaboration: Innovations and Advancing Tools* (pp. 150-165).

[www.irma-international.org/chapter/government-performance-measurement/55178](http://www.irma-international.org/chapter/government-performance-measurement/55178)

### Demystifying the Role of Blockchain Technology in Healthcare and Transaction

Mangesh Manikrao Ghonge, Sulakshana Maneand Devasis Pradhan (2022). *Blockchain Technologies and Applications for Digital Governance* (pp. 60-84).

[www.irma-international.org/chapter/demystifying-the-role-of-blockchain-technology-in-healthcare-and-transaction/293835](http://www.irma-international.org/chapter/demystifying-the-role-of-blockchain-technology-in-healthcare-and-transaction/293835)