

**Amanda Spink, Mohammad Abbas & Bernard. J. Jansen**  
**The Pennsylvania State University**  
**University Park Pennsylvania**

## **Accessing e-commerce Web information: Implications for bridging the digital divide**

**Abstract:** Accessing e-commerce information on the Web is becoming a key aspect of citizenship in the digital age. By 1999 e-commerce queries were the largest group of Web queries, forming some 25% of all queries. Web queries are a primary means for translating people's need for business product and service information for effective e-commerce. This paper reports a study of a sample of e-commerce related sessions submitted to the Excite Web search services in 2001. The study provides insights into the trends in e-commerce related Web searching. Findings include: (1) Business queries often include more search terms, are less modified, lead to fewer Web pages viewed, and include less advanced search features than non-business queries. The great majority of Web business queries posed by the public are short, not much modified, and very simple in structure. Few queries incorporate advance search features and when they do many contain mistakes. Despite getting, as a rule, a large number of Web sites as answers to their queries, Web users for business information view few result pages, tend not to browse beyond the first or second page of results and do not use relevance feedback. A small number of terms are used with very high frequency tend to be more generic, such as "shopping" and "pizza". Alternatively, a great many terms are used only once as the language of Web queries is very rich and varied, with many unique company and product names appearing as queries. (2) Company or product name queries were the most common form of business query. Company (and industry) related queries account for more than half the business-related Web searches. Many people are searching for company web sites. A need exists for new search tools that more easily facilitate company and product searches. Despite the growing business nature of the Web, many users' access to business information is limited due to poor searching tools, and a lack of a standardized approach to search engine functionality and terminology. The business terminology on the Web is also incredibly various and difficult for the average user to predict with total accuracy. Our ongoing research is further investigating the trends in e-commerce related Web searching by looking for the queries relating to commerce from a subset of large number of queries provided by the Fast and Alta Vista search engines.

**Résumé:** En cette ère numérique, l'accès à l'information touchant le commerce électronique sur le Web devient un enjeu majeur pour l'individu. En 1999, les requêtes touchant le commerce électronique constituaient la plus grande partie des demandes sur le Web, soit 25 % de toutes les requêtes. Pour le commerce électronique efficace, l'interrogation du Web est le principal moyen pour traduire les besoins des individus en biens et services informationnels. Cet article rend compte de l'étude d'un échantillon de sessions touchant le commerce électronique soumises au moteur de recherche Excite en 2001. Cette étude donne un aperçu des tendances de recherche en commerce électronique. Les conclusions démontrent que : (1) Les requêtes commerciales comportent souvent plusieurs termes de recherche, ne sont pas souvent modifiées, indiquent moins de sites Web visités et incluent moins l'utilisation des fonctions de recherche avancée que les requêtes ne touchant pas le domaine commercial. La grande majorité des requêtes commerciales des utilisateurs sont courtes, pas souvent modifiées et sont structurées très simplement. Peu de requêtes utilisent les fonctions de recherche avancée et lorsque c'est le cas, les requêtes comportent souvent des erreurs. Malgré le fait que généralement, ils obtiennent un très grand nombre de sites Web comme résultat de recherche, les utilisateurs ne consultent que très peu de sites, n'ont pas tendance à fureter au-delà de la première ou la deuxième page de résultats et n'utilisent pas le contrôle de pertinence.

Un très petit nombre de termes utilisés très fréquemment ont tendance à être très généraux, par exemple « shopping » ou « pizza ». Sinon, un très grand nombre de termes sont utilisés seulement une fois alors que le langage utilisé sur le Web est très riche et varié, avec les nombreux noms de compagnies et de produits uniques pouvant être utilisés lors des requêtes. (2) Les requêtes les plus communément utilisées sont celles comportant le nom de la compagnie ou de leur produit. Les noms des compagnies (ou de fabricants) comptent pour plus de la moitié des recherches commerciales sur le Web. Plusieurs individus recherchent les sites Web des compagnies. Un besoin pressant de créer des nouveaux outils qui faciliteront le repérage des compagnies et de leurs produits sur le Web existe. Malgré le développement croissant de la nature commerciale du Web, l'accès de plusieurs utilisateurs à l'information commerciale est limité par les outils de recherche peu adaptés et par un manque de stratégies normalisées pour utiliser les fonctions des moteurs de recherche et leur terminologie. La terminologie commerciale du Web est aussi incroyablement variée et difficile à utiliser avec précision pour l'utilisateur moyen. Notre recherche en cours examine de façon approfondie les tendances des recherches reliées au commerce électronique sur le Web en étudiant les requêtes reliées au commerce à partir d'un sous-ensemble d'un grand nombre de requêtes fournies par les moteurs de recherche « Fast » et « AltaVista ».

## 1. INTRODUCTION

Web traffic has increased exponentially as people are using Web search engines as a major tool to dig their way through Web-based information. The Web has become a major outlet for all kinds of e-commerce transactions. Since the inception of major businesses on the Web, and easy and relatively less costly access to Web services, customers are using the Web as a transaction medium e-commerce and business information.

Buying and selling products and services over the Web is becoming a part of everyday life for many people who search via many different Web search engines. People are spending increasing amounts of time working with electronic information and engaging in e-commerce. Web searching services are now everyday tools for information seeking and e-commerce. However, many users Web interactions are often frustrating and constrained.

E-commerce is a growing force in the world economy and the Web is becoming a major source of business products, services and information for many people worldwide (Zwass, 2000). Better Web search tools are important for the development of e-commerce, as people use search engines to find business related information on the Web. Many shopping and business Web sites have search accessible on their Web sites. Apart from entering a business related URL, most users have little choice but to interact with a Web search engines, such as Alta Vista or Google.

Large-scale, quantitative or qualitative studies have explored how users' search the Web (Silverstein, Henzinger, Marais & Moricz, 1999; Spink, Wolfram, Jansen & Saracevic, 2001). To support human information behaviors we are seeing the development of a new generation of Web tools, such as Web meta-search engines, to help users persist in electronic information seeking is needed to help people resolve their information problems.

Most Web queries are short, without much modification, and are simple in structure (Spink, Wolfram, Jansen & Saracevic, 2001). Few queries incorporate advance search techniques, and when such techniques are used many mistakes result. However, relevance feedback and some advanced search features are growing in use. Frequently, people retrieve a large number of Web

sites, but view few result pages and tend not to browse beyond the first or second results pages. Overall, a small number of terms are used with high frequency and many terms are used once. Web queries are very rich in subject diversity and some unique. The subject distribution of Web queries does not seem to map to the distribution of Web sites subject content.

The next section of the paper provides an overview of related studies to the analysis of e-commerce queries during Web retrieval.

## 2. RELATED STUDIES

### E-Commerce Web Search

Buying and selling products and services over the Web is becoming a part of everyday life for many people who search via many different Web search engines. Spiteri (2000) compared the effectiveness of six Internet search engines. She found ambiguous and sometimes misleading categories in e-commerce sites, moderate consistency in e-commerce Web site organization, and few opportunities for comparison-shopping. Spiteri (2000) identified two types of consumer Web behavior: (1) *goal-directed* - to find a product or information, and (2) *experiential* - non-directed exploratory browsing or surfing.

However, using Web search engines to find information and conduct e-commerce transactions is challenging for many users (Jansen, Spink, & Saracevic, 2000; Spink, Wolfram, Jansen & Saracevic, 2001). Many Web users find it difficult to conduct e-business via an information system more akin to an automated library catalog than an effective sales transaction system. Studies of business Web queries are important for understanding Web usage and for the development of Web tools and systems to facilitate more effective e-commerce.

Recent studies show that business related queries are increasing as a proportion of all Web queries (Wolfram, Spink, Jansen & Saracevic, 2001). Jansen, Spink and Saracevic (2000) found that in 1997 business-related terms constituted 8.3% of the 63 top terms and 13.5% of queries entered into the Excite search engine. Wolfram, Spink, Jansen and Saracevic found an increase to 24.4% for business-related queries in 1999. Spink, Milchak, Sollenberger and Hurson (2000) found that business-related question queries were the second largest category of question queries behind people and places.

Spink and Guner (2001) sampled 10,000 Excite queries and 10,000 Ask Jeeves question format queries for e-commerce related queries. Findings include: (1) business queries often include more search terms, are less modified, lead to fewer Web pages viewed, and include less advanced search features, than non-business queries, (2) company or product name queries were the most common form of business, and (3) Ask Jeeves business queries in question form were largely limited to the format "Where can I buy....." or the request "I want to buy.....".

Spink and Ozmutlu (2002) also showed that many Ask Jeeves e-commerce queries were in request not question format. The most common format for e-commerce request queries were "I

want to buy....” and “Get me....” Spink, Wolfram, Jansen and Saracevic (2002) found that in 2001 e-commerce queries formed about 25% of Web queries.

The next section of the paper presents the research objectives, research design and the study results followed by a discussion of the study implications. The study of users’ e-commerce related queries to Web search engines is important for understanding information access on the Web. This research can also help design company Web sites and develop better e-commerce search engines. In this paper we explore the extent and nature of business related searching on the Web.

### **3. RESEARCH OBJECTIVES**

Our study objectives were to further investigate the nature of e-commerce related Web queries:

1. Proportion of Web queries that are e-commerce related.
2. Types of e-commerce related information requested.

### **4. RESEARCH DESIGN**

#### **4.1 Data Collection**

We accessed data set from 2001 containing 1 million Excite Web queries. The subsets were created using a computer program that picked out random queries from the original sets. Each data set contained three fields:

- *Identification*: an anonymous code assigned by the Excite server to a user machine.
- *Time of Day*: measured in hours, minutes, and seconds from midnight of 20 December 2001.
- *Query*: the query terms exactly as entered by the user.

#### **4.2 Data Analysis - E-commerce Session Identification**

Our analysis focused on the e-commerce sessions and queries in the data set. The data analyzed included users sessions, queries, and term level of analysis. Basically:

- A *session* is the entire sequence of queries by a particular user.
- A *query* is a set of one or more terms entered into the Web IR system during a single search.
- A *term* is any string of characters bounded by white space.

To determine the extent of e-commerce-related Web queries, we qualitatively analyzed and subject categorized a random sample of 10,000 queries by Excite users from 20 December 2001.

Each query was qualitatively analyzed and classified into one e-commerce-related category previously developed by Spink and Gunar (2001), including.

- Real Estate, Company URLs, Tourism, Travel and Places, Finance, Products, Industry or Association Name, Business Education, Legal Issues, Employment, Trade and Commerce

The categories were developed iteratively from the analysis of the data and evolved as the analysis progressed.

- A *query* was defined as a set of one or more search terms; it may include advanced search features, such as logical operators and modifiers.
- A *e-commerce query* was one that included a company name, product name, brand name, e-commerce-related word or phrase.
- A *user search session* included all queries from the beginning to the end of the users' interaction with the Excite search engine. A session was defined as a consecutive set of e-commerce related queries done by a single user.

The results of our analysis provided insights into the nature and subjects of e-commerce-related queries, sessions and queries.

## 5. RESULTS

### 5.1 E-Commerce Related Query Structure

We identified characteristics of Excite user's e-commerce queries and sessions (Table 1).

**Table 1. Excite 2001 e-commerce sessions.**

Number of e-commerce sessions	388
Maximum queries per e-commerce session	7
Minimum queries per e-commerce session	1
Mean queries per e-commerce session	1.2
Maximum e-commerce query length	11
Minimum e-commerce query length	1
Mean e-commerce query length	2.3
Maximum topics per e-commerce session	5

Minimum topics per e-commerce session	1
Mean topics per e-commerce session	1.1

- We identified 388 e-commerce related sessions in the 10,000 Excite queries.
- The maximum queries users searched in a single e-commerce session was 7, the minimum queries per session was 1, and a mean queries per e-commerce session was 1.2.
- E-commerce queries often included more search terms, are less modified, lead to fewer Web pages viewed, and include less advanced search features than non-e-commerce queries.
- The great majority of Web e-commerce queries posed by the public were short, not much modified, and very simple in structure.
- Few queries incorporate advance search features and when they do many contained spelling mistakes.
- Web users for e-commerce information viewed few result pages, and tended not to browse beyond the first or second page of results and do not use relevance feedback.

## 5.2 Query Subject Categories

Table 2 shows the percentage of queries in each-commerce category.

**Table 2. Subject categories.**

Subject Categories	%
Company URL	27%
Industry or Association Name	23.5%
Trade and Commerce	16.3%
Tourism, Travel & Places	9%
Finance	7.5%
Products	6.5%
Real Estate	5%
Employment	2.5%
Business Education	2%
Legal Issues	0.6%
Total	100%

- Company or product name queries were the most common form of e-commerce query.
- Company (and industry) related queries account for more than half the e-commerce-related Web searches. Many people are searching for company web sites. Despite getting, as a rule, a large number of Web sites as answers to their queries,

- A small number of terms are used with very high frequency tend to be more generic, such as "shopping" and "pizza".
- Alternatively, a great many terms are used only once as the language of Web queries is very rich and varied, with many unique company and product names appearing as queries.

## **6. DISCUSSION**

In general Web search engines are currently trying to cater to the broad rather than specialized nature of human information needs. People are increasingly searching for e-commerce related information using Web search engines. The nature of their e-commerce related information needs are broad and relate to all aspects of e-commerce.

We found that great majority of Web e-commerce queries posed by the public are short, not much modified, and very simple in structure. Few queries incorporate advance search features and when they do many contain mistakes. Despite getting, as a rule, a large number of Web sites as answers to their queries, Web users for e-commerce information view few result pages, tend not to browse beyond the first or second page of results and do not use relevance feedback.

Overall, a small number of terms are used with very high frequency, while there are great many terms that are used only once. The language of Web queries is very rich and varied, with many unique company and product names appearing as queries. A few high frequency terms are more generic, such as "shopping" and "pizza"

Our results show that company (and industry) related queries account for more than half the e-commerce-related Web searches. Many people are searching for company web sites. A need exists for new search tools that more easily facilitate company and product searches. Excite allows industry search, but it would be easier for people to use it if Excite brings this feature to the main page.

Many people appeared to be searching for shopping sites via the search engine, as the number of e-commerce-related queries appeared to be directly shopping-related. They are searching under brands names as well as generic types of products, such as "cars" or "golf clubs".

Many people were seeking advice or product comparisons or they are just not sure what they want or how to find it on the Web. Web services such as "Ask Simon" are attempting to cater to this need. Many users need a conversation with an "expert" to guide them to expressing and finding their real information need.

Many Web search engines and e-commerce sites are beginning to provide some forms of interaction for customers to help them sort through the plethora of Web sites, services, and products that confront them. Some search engines, such as Alta Vista, are providing some form of comparison-shopping.

We also compared the results with two other related studies of large query corpora. Overall, e-commerce-related queries are slightly longer than general queries. People are also increasingly querying for e-commerce information in question format. Ask Jeeves type Web services are growing and encouraging this trend. The development of Web services that focus on helping people with e-commerce-related information needs in question and answer format is the next step in improving the Web as an e-e-commerce tool.

There is a great market for niche Web services and search engines that facilitate more effective and targeted access to e-e-commerce sites. Our study with search log data as the only available data cannot answer other questions about the results of these queries or the performance of different search engines. However, they do provide a snapshot for comparison of public behavior while searching, a behavior that can also serve as a clue for improvement of Web services.

Despite the growing e-commerce nature of the Web, many users' access to e-commerce information is limited due to poor searching tools, and a lack of a standardized approach to search engine functionality and terminology. The e-commerce terminology on the Web is also incredibly various and difficult for the average user to predict with total accuracy.

## 7. CONCLUSION

As the Web continues to become a key centerpiece for e-commerce, new tools and new ways of searching for e-commerce information on the Web are needed. Searching the Web today is somewhat akin to searching a library catalog and equally as frustrating. The Web also lacks a standardized approach to search engine functionality and terminology.

The e-commerce terminology on the Web is also incredibly various and difficult for the average user to predict with total accuracy. As more Web users begin to ask questions rather than producing Boolean queries, methods and approaches are needed to analyze and process question queries in a "question and answer" format to diagnose the users real information requirements to complete the cycle of e-commerce and facilitate the effective sale of goods and services over the Web.

## REFERENCES

- Jansen, B. J. Spink, A. & Saracevic, T. (2000). Real life, real users and real needs: A study and analysis of users' queries on the web. *Information Processing and Management*, 36 (2): 207-227.
- Silverstein, C. Henzinger, M. Marais, H. & Moricz, M. (1999). Analysis of a very large web search engine query log. *Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM SIGIR) Forum*, 33 (3).
- Spink, A. & Guner, O. (2001). E-commerce web queries: Excite and Ask Jeeves study. *First Monday*, 6 (7), July.
- Spink, A. Milchak, S. Sollenberger, M. & Hurson, A. (2000). Elicitation queries to the Excite Web search engine. In *CIKM 2000: Proceedings on the Conference on Information and Knowledge Management, November 2000, Washington, D.C.*, pp. 134-140.



- Spink, A. & Ozmutlu, H. C. (2002). Characteristics of question format web queries: An exploratory study. *Information Processing and Management*, 38 (4): 453-471.
- Spink, A., Wolfram, D., Jansen, B. J. & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science*, 53 (2): 226-234.
- Spiteri, L. F. (2000). Access to electronic commerce sites on the world wide web: An analysis of the effectiveness of six internet search engines. *Journal of Information Science*, 26 (3): 173-183.
- Wolfram, D. Spink, A. Jansen, B. J. & Saracevic, T. (2001). Vox populi: The public searching of the web. *Journal of the American Society for Information Science and Technology*, 52 (12): 1073-1074.
- Zwass, V. (2000). Structure and macro-level impacts of electronic commerce: From technological infrastructure to electronic marketplaces. *Foundations of Information Systems - E-Commerce Paper* [<http://www.mhhe.com/business/mis/zwass/ecpaper.html>]