COVER SHEET

This is the author-version of article published as:

Spink, Amanda and Ozmutlu, H. Cenk and Lorence, Daniel P (2004) Web searching for sexual information: an exploratory study. *Information Processing and Management* 40(1):pp. 113-124.

Accessed from **http://eprints.qut.edu.au**

Forthcoming in the journal Information Processing and Management

WEB SEARCHING FOR SEXUAL INFORMATION: AN EXPLORATORY STUDY

Amanda Spink*
School of Information Sciences and Technology
The Pennsylvania State University
004C Thomas Building, University Park PA 16802
Tel: (814) 865-4454 Fax: (814) 865-5604
Email: spink@ist.psu.edu

H. Cenk Ozmutlu
Department of Industrial Engineering
Uludag University, Bursa, Turkey 16059
Email: hco@uludag.tr.edu

Daniel P. Lorence
Department of Health Policy and Administration
The Pennsylvania State University
Email: dpl10@psu.edu

* To whom all correspondence should be directed.

ABSTRACT

Sexuality on the Internet takes many forms and channels, including chat rooms discussions, accessing Websites or searching Web search engines for sexual materials. The study of Web sexual queries provides insight into sexual-related information-seeking behavior, of value to Web users and providers alike. We qualitatively analyzed 58,027 queries from a log of 1,025,910 Excite Web user queries from 1999. We found that sexual and non-sexual-related queries exhibited differences in session duration, query outcomes, and search term choices. Implications for sexual information seeking and Web systems are discussed.

INTRODUCTION

Sexuality is part of human everyday life behaviors and human information behaviors. The study of everyday life information seeking (ELIS) is a growing area of research that investigates the role of information for people in the information age (Chatman, 1991; Savolainen, 1995; Spink & Cole, 2001). ELIS research also forms a subset of human information behavior (HIB) research that explores all aspect of information related human behaviors, including seeking, searching and use. The permeable relationship between an individual's everyday and occupational information seeking is also embedded with an information environment that includes channels of information and communication (Spink & Cole, 2001). Generally, formal channels involve little or no interaction between the user and information source, while informal channels involve colleagues and friends and involve a certain amount of interaction. The Internet provides a hybrid channel, neither formal nor informal. The seeking of sexual information is an important human element of ELIS. Searching the Web for sexual information represents a hybrid formal and informal channel for information seeking.

There is a need to further understand the nature and characteristics of human information seeking and searching for sexual information. The study reported in this paper explores sexually related human information seeking on the Web. We seek to enhance the social sciences framework for Internet sexuality (Cooper, Scherer, Boies & Gordon, 1999) through the development of an information-seeking framework for Internet sexuality, and specifically Web searching. We examined the characteristics of sexually (including pornographic) related queries, including visual images or textual descriptions of sexual behavior. Sexual and pornographic queries map also overlap. Our

paper first discusses the social sciences framework for human Internet sexuality and then outlines the current studies on sexual-related Web searching.

RELATED STUDIES

Human Internet Sexuality

Pornography or sexually related materials are available widely on the Web. Rimm (1995) found that pornographic images were widely available on the Web, particularly on bulletin boards and USENET. A 1998 study (Willems, 1998) estimated the existence of 22,000 pornographic Web sites. These studies are part of an ongoing debate regarding the availability and accessibility of sexually related material on the Web. Li (2000) provides an overview of this debate, including the issues of obscenity, child pornography and filtering, and the studies conducted particularly in relation to libraries.

Human Internet sexuality is a growing area of research in the social sciences (Cooper, 1998; Cooper, 2002; Cooper, et al., 1999, 2000). Cooper, et al., 1999; Goodson, et al., 1999, 2000) highlights the divergent views of Internet sexuality or "cybersex". The *pathological perspective* focuses on Internet sexuality as deviant, addictive or criminal behavior (Bingham & Piotrowski, 1996; Durkin & Bryant, 1995; King, 1999; Van Gelder, 1985; Young & Rogers, 1998). The *adaptive-perspective – sexual expression* is a more adaptive view that places Internet sexuality within the context of sexual human development and exploration, and love and romance (Cooper & Sportolari, 1997; Leiblum, 1997; Newman, 1997).

Studies also show gender differences in Internet sexuality (Cooper, Morahan-Martin, Mathy & Maheu, 2002; Cooper, Scherer, Boies & Gordon, 1999). Cooper, et al.,

(1999; 2000) found that men are more dominant in online sexual activities, such as sexually related chat rooms and pornographic Website use, Internet abuse and sexual problems. The researchers also found that heavy Internet users, or about 8% of users, who spend the most time online for sex, also reported significant problems associated with compulsive disorders and addiction. The goal of their research is to recommend treatments for outreach prevention programs. Cooper, Scherer, Boies and Gordon (1999) also identify "paraphilics" as dependent on cybersex as a source of stimulation and satisfaction for often unconventional sexual desires.

These studies strongly contribute to our understanding of human sexuality in the new frontier of the Internet.

<div align="center">Sexually-Related Web Searching</div>

Recent studies have also begun to examine the nature of sexually related information seeking on Web search engines. Goodrum and Spink (2001) found that 25 of the most frequently occurring terms in multimedia related queries terms submitted to the Excite commercial Web search engine were clearly sexually related. Spink, Wolfram, Jansen and Saracevic (2001) found that although sexually related searching by users of the Excite Web search engine represented only a small proportion of all queries and terms (>5%), about one in every four terms in the list of 63 highest used terms can be classified as sexual in nature. Spink and Ozmutlu (2002) also found that terms such as "sex", "nude" and "naked" were high frequency terms submitted to the Ask Jeeves commercial Web search engine.

Spink, Jansen, Wolfram and Saracevic (2002) examined large-scale Excite Web query data sets from 1997, 1999 and 2001. They found that sexually related queries decreased as a proportion of all Web queries from the second largest category (16.8%) in 1997 to fourth largest category (7.5%) in 1999 and the fifth largest category (8.5%) in 2001. Web queries related to business, computers and people increased as a proportion of all Web queries. Spink, Ozmutlu, Ozmutlu and Jansen (forthcoming) compared the topics of United States Excite search engine users versus European Fast search engine searches in 2001. They found that that sexually related queries was the fifth largest Excite topic category (8.5%) and the fourth largest category (10.8%).

Previous studies have identified the extent of sexually related Web searches. There is a need for a more detailed analysis of a large-scale set of sexual related queries to further identify the characteristics of sexual Web searching.

RESEARCH GOALS

In this paper we report results from an analysis of the characteristics of sexually related Web queries, using non-sexual Web queries as a comparison. The overall objective of our study was to enhance our understanding of sexually-related information seeking on the Web, including:

1.   Identifying the proportion of sexual related queries.

2.   Comparing the characteristics of sexual and non-sexual queries.

RESEARCH DESIGN

Data Collection

Excite, Inc. is a major Internet media public company that offers free Web searching and a variety of other services [http://www.excite.com]. We extracted a random set of 58,027 queries from a query-log of 1,025,910 queries from 20 December 1999. Each query contained three fields. We could locate each user's initial query and recreate the chronological series of actions during each user session, including: *User Identification*: an anonymous user code assigned by Excite server. *Time of the day*: second, minute, hour, day, month and year are given in adjacent format (on 20 December 1999). *Query terms*: the terms entered by the users.

The researchers had no control over the date and size of the query data set provided by Excite, Inc. At the time of the data analysis, the researchers also had no access to large-scale query data from other commercial Web search engines, e.g., Google, for comparison with the Excite data.

Data Analysis

Sexual Query Classification

Web queries were qualitatively examined for sexual content by two researchers. The primary purpose of the analysis was to classify queries as either a sexual related query or a non-sexual query. To judge the query intent as sexual or pornographic, strong evidence of such intent had to be present in the query log. If queries did not include explicitly sexual (including pornographic) key words, they were not classified for the study.

We judged queries as related to sexual (including pornographic) needs if the queries explicitly requesting sexual information, visual images or textual descriptions of sexual behavior. Sexual and pornographic queries overlap. Sexual queries may consist of one term, e.g., sex, porn, or multiple terms, e.g., "big nude tits", "sex with naked girls" or "fxxxxx pictures". Many sexual terms and queries contain "four letter words", obscene language and expletives that cannot be listed in our paper. Queries that were more health or medical related, such as "sexually transmitted diseases" or "sexual herpes", were not counted as sexual queries.

Inter-coder agreement is defined as the similarity in which each coder in the study decided whether a query was sexual. To check coding consistency, each researcher recoded 10,000 queries previously classified by the other researcher. The researchers met again in order to make final decision about how the classifications. The two researchers discussed each disputed query until a classification consensus for that query was reached.

RESULTS

This paper extends previous results reported by Spink and Ozmutlu (2001).

Terms, Queries and Sessions

The analysis was that sexual sessions were longer than non-sexual sessions in terms of unique queries per session (Table 1).

[Place Table 1 Here]

However, sexual queries were similar in length to non-sexual queries (Table 2).

[Place Table 2 Here]

In addition, sexual queries contained less unique search terms than non-sexual queries that were more varied in search topics.

Except for the queries with 2 terms, non-sexual queries have a higher percentage of occurrences (Figure 1).

[Place Figure 1 Here]

The distributions of terms per query do not differ significantly between sexual and non-sexually related queries. The top non-sexual query terms are aimed to acquire specific Web pages, whereas the top sexual query terms are generally intended for specific topics with no regard to any specific Web address.

We see a lower mean unique terms per query in sexual queries (Table 3).

[Place Table 3 Here]

The most frequently used sexual query terms constitute a higher percentage of all sexual terms, compared to the percentage of top non-sexual terms among all non-sexual terms. Although sexual related searches might include multiple sub-topics, the variety of these topics cannot match that of non-sexual searches. The vocabulary of sexual Web searching is quite constrained

Sexual sessions have a higher probability of having more than 20 queries, compared to non-sexual sessions (Figure 2 & Figure 3).

[Place Figures 2&3 Here]

In addition, sexually related sessions have a lower probability of having 1 or 2 queries. Besides such small differences, both distributions are very similar. Out of 1,957 sexually related sessions, 24.3% included single queries, whereas 19.05% included two queries and 56.5% included three or more queries. For the non-sexual sessions, the

respective ratios were 28.7% for single query sessions, 20.4% for the two-query sessions and 50.7% for the remainder of the sessions.

<u>Session Duration</u>

Although, the distributions of duration in seconds for sexual and non-sexual sessions are still closely related, the duration differences in the short and long lasted sessions increased significantly (Figure 4).

[Place Figure 4 Here]

During sexual related sessions, users tend to view more pages for each query. Sexual related searchers view more Websites than non-sexually related searchers. Some 30% of the non-sexual sessions were less than 20 seconds, whereas almost no sexual related search was less than 20 seconds. Almost 40% of the sexually related sessions were longer than 400 seconds, whereas only 22% of the non-sexual queries were long. Overall, sexually related Web searching seems to take longer than other types of information seeking on the Web. In other words, sexually related session users may more persistent searchers or are willing to spend more time evaluating the results of their queries, before continuing on to the next step of their retrieval process.

DISCUSSION

An interesting feature of sexual searching was the limited vocabulary used in sexual queries compared to non-sexual queries. Sexual queries involved fewer unique terms. Many sexually related terms were repeated frequently in queries, e.g., nude, sex, naked, etc. This finding suggests that despite the complexity of the English language and

the extensive and variant nature of Web terms in general, sexual terms are constrained in range and variance. A quick look at any sex or pornography related Web site reinforces the rather limited nature of the vocabulary used by both sexual related Web site creators and searchers.

We were able to identify some characteristics of sexual related queries and sessions. Sexual related sessions and queries were longer than non-sexual search sessions, and contained more queries. Sexual queries were generally longer than most non-sexual queries and there was a high probability that a sexual session could be longer than 20+ queries. Although, on average users submit more queries for sexual related information, this may not be the only cause for longer durations for sexual related sessions. People seeking sexual information were willing to expend the time and effort to create longer queries and to use more queries.

Another time factor that distinguished sexual searching was the issue of pages viewed. Sexual searchers viewed more pages than non-sexual searchers. Again we see the influence of persistence as a characteristic of sexual information seeking. Most non-sexual searchers do not view much beyond the first or second page of ten Web sites. For example, a typical sexual information seeker was also seeking images of nude women and may view more than 20 pages of Websites.

The sexual sessions may also longer because images often take quite some time to download. It may take longer to make a relevance judgment when searching for sexual Web material. The findings may also relate to the Cooper, et al., (1999) identification of heavy Web users who exhibit "paraphilic" behaviors and are dependent on cybersex as a source of stimulation and satisfaction. We were not able to identify any demographic user

characteristics from the Web query logs, so the relationship between long session and sexually addictive or obsessive behaviors is an area for further research.

Our study does suggest the sexually related information seekers were willing to expend more time and effort in pursuit of their information need than general Web searchers. Considering the limited range of outlets and the restrictions placed on sexual related materials in most societies, it is not surprising that sexually related information seeking is a Web activity that attracts extensive interaction. Web interaction allows consumers to seek sexual information freely, without any judgmental remarks or criticism from their peers. As sexual-related information seeking in its many forms appears to be a major part of human existence, especially on the Web, we need to develop more detailed models of sexual information seeking and searching. Our research suggests some differences between sexual and non-sexual everyday life information seeking models. This is an important area for further research.

How can the existing limitations of Web search engines be overcome for sexual information seeking? In many cases, sexual and non-sexual-related queries have significant differences, in terms of duration of the session, analysis of the results of each query, search term usage, and etc. The findings of this study could be used to improve the performance of Web search. Studies on sexual Web inquiry sessions could also be helpful for researchers in other disciplines, such as human psychology, by providing further insight into human sexuality.

CONCLUSION AND FURTHER RESEARCH

A limitation of our study is our analysis of queries from only one Web search engine. To extend our understanding of sexual Web searching we are currently comparing sexual queries from two other commercial Web search engines, including one European-based and one US-based search engine. We are also working towards research to explore further trends in sexually related Web searching. The authors are also conducting further research into medical and health related Web searching.

ACKNOWLEDGEMENTS

REFERENCES

Bingham, J. E., & Pitrowski, C. (1996). Online sexual addiction: A contemporary enigma. *Psychological Reports, 79*, 257-258.

Chatman, E. (1991). Life in a small world: Applicability of gratification theory to information seeking behavior. *Journal of the American Society for Information Science, 42*, 438-449.

Cooper, A. (1998). Sexuality and the Internet: Surfing into the new millennium. *CyberPsychology & Behavior, 1*(2), 181-187.

Cooper, A. (2002) (Ed.), *Sex and the Internet: A Guidebook for Clinicians*. New York: Brunner-Routledge.

Cooper, A., Boies, S., Maheu, M., & Greenfield, D. (2000). Sexuality and the Internet: The next sexual revolution. In F. Muscarella & L. Szuchman (Eds.), *The Psychological Science of Sexuality: A Research Based Approach* (pp. 519-545). New York: Wiley Press.

Cooper, A., Delmonico, D., & Burg, R. (2000). Cybersex users, abusers, and compulsives: New findings and implications. *Sexual Addiction & Compulsivity: The Journal of Treatment and Prevention, 6*(2), 79-104.

Cooper, A., Morahan-Martin, J., Mathy, R. M., & Maheu, M. (2002). Toward an increased understanding of user demographics in online sexual activities. *Journal of Sex and Marital Therapy, 28*, 105-129.

Cooper, A., Putnam, D., Planchon, L., & Boies, S. (1999). Online sexual compulsivity: getting tangles in the net. *Sexual Addiction and Compulsivity: Journal of Treatment and Prevention, 6*(2), 70-104.

Cooper, A., Scherer, C. R., Boies, S. C., & Gordon, B. L. (1999). Sexuality on the Internet: From sexual exploration to pathological expression. *Professional Psychology: Research and Practice, 30*(2), 154-164.

Cooper, A., & Sportolari, L. (1997). Romance in cyberspace: Understanding online attraction. *Journal of Sex Education and Therapy, 22*(1), 7-14.

Durkin, K. F., & Bryant, C. D. (1995). Log on to sex: Some notes on carnal computer and erotic cyberspace as an emerging research frontier. *Deviant Behavior: An International Journal, 16*, 179-200.

Goodrum, A., & Spink, A. (2001). Image searching on the Excite web search engine. *Information Processing and Management*, *37*(2), 295-312.

Goodson, P., McCormick, D., & Evans, A. (2000). Sex on the Internet: A survey instrument to assess college students' behavior and attitudes. *Cyberpsychology & Behavior, 3*(2), 129-149.

Goodson, P., McCormick, D., & Evans, A. (2001). Sex on the Internet: College students' emotional arousal when viewing sexually explicit materials on-line. *Journal of Sex Education and Therapy, 25*(4), 252-260.

King, S. (1999). Internet gambling and pornography. *Cyberpsychology & Behavior, 2*, 175-193.

Leiblum, S. R. (1997). Sex and the net: Clinical implications. *Journal of Sex Education and Therapy, 22*(1), 21-28.

Li, J. H-S. (2000). Cyberporn: The controversy. *First Monday, 5*(8) (August 2000).

Newman, B. (1997). The use of online services to encourage exploration of egodystonic sexual interests. *Journal of Sex Education and Therapy, 22*(1), 45-48.

Ozmultu, S., Spink, A., & Ozmutlu, H. C. (2002). Trends in multimedia web searching: 1997-2001. *Information Processing and Management*.

Rimm, M. (1995). *Marketing pornography on the information superhighway: A survey of 917,410 images, descriptions, short stories, and animations downloaded 8.5 million times by consumers in over 2000 cities in forty counties, provinces and territories*. (11 July 1995). http://trfn.pgh.pa..us/guest/mrtext.html.

Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of way of life. *Library and Information Science Research, 17*, 259-294.

Spink, A., & Cole, C. (2001). Everyday life information seeking research. *Library and Information Science Research*, 23(1), 301-304.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer*, *35*(3), 133-135.

Spink, A., Yang, Y., Nykanen, P., Lorence, D., Brezinski, K., Ozmutlu, S., & Ozmutlu, H. C. (forthcoming). *Medical or health Web searching: An exploratory study*.

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., & Ozmutlu, H. C. (2001). Sexually-related information seeking on the web. *Proceedings of ASIST 2002: American Society for Information Science and Technology, Washington, D.C., November 2001* (pp. 545-554).

Spink, A., Ozmutlu, S., Ozmutlu, H. C., & Jansen, B. J. (forthcoming). *United States versus European Web searching trends*.

Van Gelder, L. (1985). The strange case of the electronic lover. *Ms. (October)*, pp. 94-124.

Willems, H. (1998). Filtering the net in libraries: The case (mostly) against. *Computers in Libraries, 18*, 55-58.

Young, K. S., & Rogers, R. C. (1998). The relationship between depression and Internet addiction. *Cyberpsychology, 1*(1), 25-28.

Table 1. Data summary.

	Sexual Queries	Non-Sexual Queries
Number of users	1,957	10,326
Total queries	14,379	43,648
Unique queries	6,211	21,049
Repeat queries	6,937	17,527
Zero term queries	1,231	5,072
Mean queries per user session	7.3	4.2
Median queries per user session	4	3
Mean unique queries per user session	3.1	2.03
Median unique queries per user session	2	2

Table 2. Mean terms per sexual and non-sexual query.

	Sexual Queries	Non-Sexual Queries
Mean terms per query	2.2	2.2
Mean unique terms per query	2.9	3.7

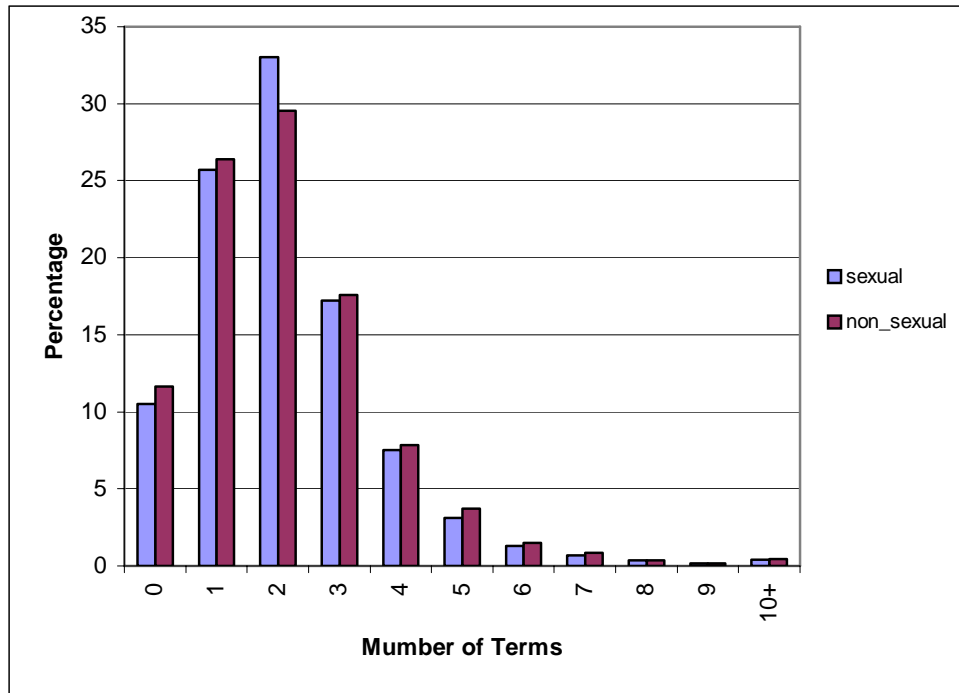Figure 1. Distributions of terms per sexual and non-sexual query.

Table 3. High frequency sexual terms.

Sexual Terms	% of Sexual Queries
Sex	3.57%
Nude	2.78%
Free	2.3%
Pictures	1.59%
And	1.43%
Women	1.14%
Pics	1.09%
Xxx	1.06%
Com	1.04%
Adult	0.99%
Girls	0.92%
www	0.92%
Stories	0.86%
pxxxx	0.83%
Gay	0.68%
Porn	0.67%
Teen	0.64%
Of	0.63%
Photos	0.61%
Naked	0.6%
Playboy	0.52%
Big	0.49%
Black	0.45%
Porno	0.44%
Nudes	0.43%

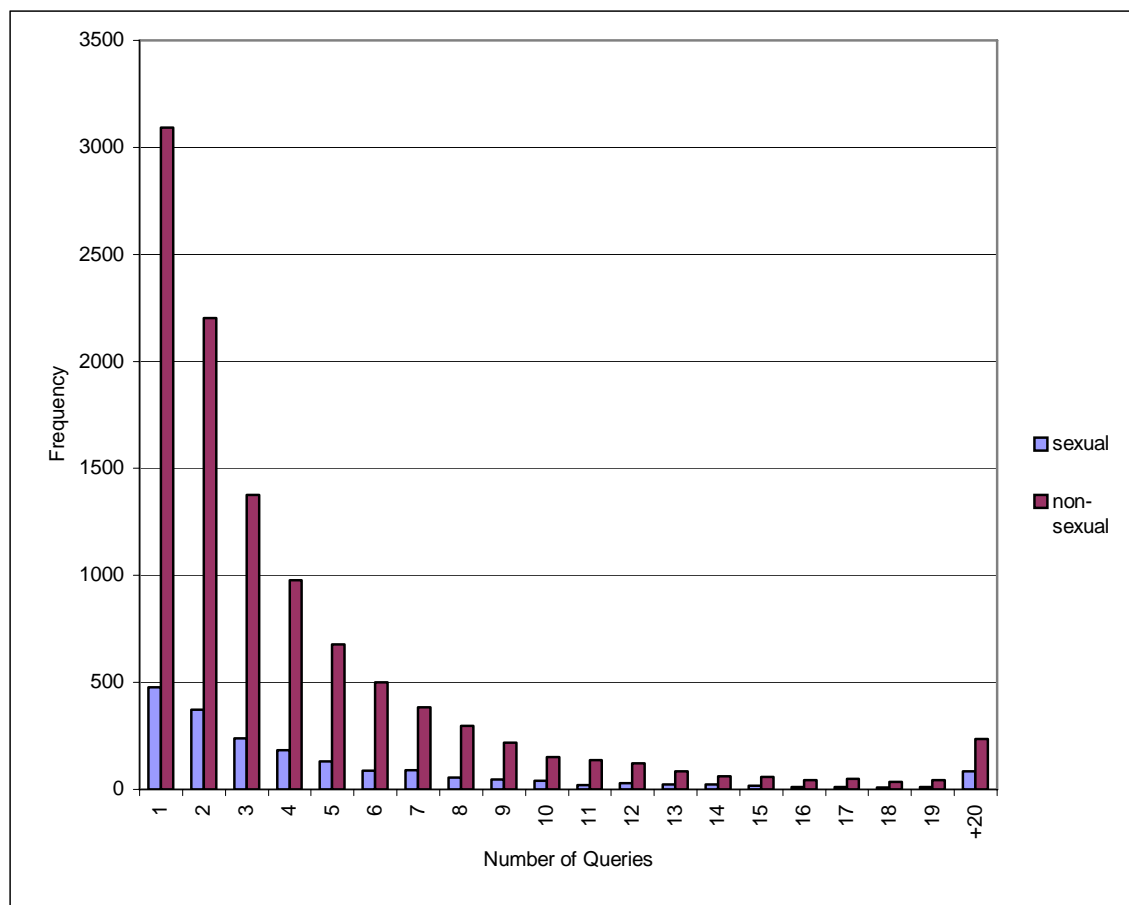Figure 2. Queries per sexual and non-sexual session.

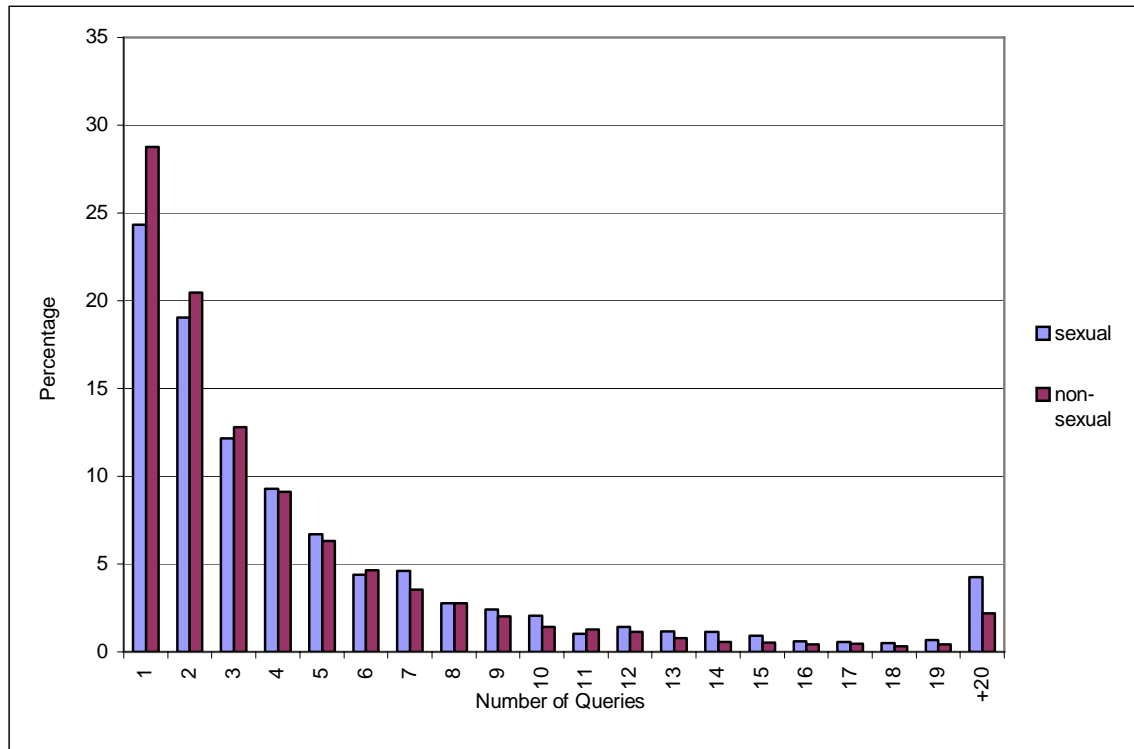Figure 3. Queries per sexual and non-sexual session (as percentage).

Figure 4. Duration of sexual and non-sexual session (seconds).