An Analysis of Internet Use in the Public Library

By

Andrea Görgeç

Shirley Lew,

and Dr. Ann Curry

School of Library, Archival and Information Studies

University of British Columbia

Table of Contents

I.   Introduction

The Internet has had an increasing presence in many Canadian public libraries in recent years. The resources available through the Internet are perceived by many librarians to be equivalent to traditional materials collected by the library and as such, are a necessary, not optional, part of the library collection. (McClure, Bertot, and Zweizig 1993, 3) In addition to this, many library users have come to expect Internet access as part of the public library's regular services. Line-ups of library users waiting to use Internet terminals in libraries across Canada attest to its popularity. Not everyone, however, agrees that the benefits of providing Internet access in the public library outweigh the expenses and difficulties associated with it. Some negative impressions of Internet access in the public library stem from a perception that the Internet is most often being used for inappropriate purposes such as the viewing of email and sexually explicit materials.

But how much do we really know about the way in which the Internet is being used in public libraries? What sites are actually being visited? To date, research on the Internet has focused on two main areas: Internet use patterns of reference librarians and other professional groups and levels of Internet connectivity in public libraries. (Tillotson, Cherry and Clinton 1995, 109) As yet, no studies have been done to explore the way in which public library users are using the Internet in public libraries. We do not have any data to support assumptions or statements made about the Internet and its use. In a 1997 article discussing the American Internet Decency Act,

the author states "although no study has been done on what users are looking at, it is likely that the vast majority of usage is not problematic." ("The Statues of Libertines?" 1997, 296) Statements such as this are indicative of the unsubstantiated conclusions that are being drawn about Internet use. The purpose of the current study is to begin to address this gap in the literature by examining the specific case of Internet use by users of the Burnaby Public Library (BPL) in British Columbia.

As a pilot study, this research project used data gathered from web server logs from the Burnaby Public Library to answer the following research questions.

1. What types of sites are being visited?

2. In what subject areas are people searching the most?

3. Are there differences between branches in terms of the types of sites visited?

4. Are their differences between the days of the week in terms of the types of sites visited?

5. How often are government sites visited?

6. How often are the sites visited magazines or ezines?

7. How often are non-English language sites visited?

8. What non-English languages are used in the above mentioned sites?

Underlying these stated objectives is the question of the feasibility of gathering and analyzing data about the Internet that is meaningful. This study used transaction log analysis (TLA) to analyze Internet use patterns. TLA, as a methodology, has been employed extensively in studies of online library catalogues but not as a tool to analyze

Internet use. Testing the effectiveness of this methodology for the analysis of Internet use was another primary objective of the researchers.

II.   Literature Review

There have been a number of studies done on Internet use in libraries. As mentioned earlier, many of these tend to focus on the impact and presence of Internet access rather than the way it is used as an information source. The others tend to study the use of the Internet by librarians or other groups of professionals. The majority employ data gathered from questionnaires. Due to the restrictions that must be placed upon the number of the responses possible to any one question on a questionnaire, the data provided by these instruments is quite limited. A carefully worded questionnaire can more directly an answer the question of why people are using the Internet than can analysis of the logs representing that use but it is less useful in providing comprehensive answers to the question of what people are actually viewing on Internet terminals.

A study related to the current study that examines Internet use by "actual Internet users" is Perry's 1995 survey. The study is designed to determine Internet use by the international Internet community through questions such as "who uses the Internet" and "what of information is being sought." Perry generalizes global Internet use patterns based upon the results of a survey that was only posted to listservs and that received only 84 responses. His study population is highly specific and precludes such a broad generalization. While his research questions are useful and are relevant

to the current study's research objectives, the response options for each question provide very limited data. For example, respondents are limited to acoustical, graphical, statistical, textual and other categories when asked "what types of information do you look for over the Internet?"

A study conducted by the Seattle Public Library (SPL) (Harvey and Horne 1995) surveys patrons' satisfaction with and use of Internet access and services in the Seattle Public Library system. The purpose of the study and the questions asked are not explicitly expressed but the study does make reference to patrons' responses regarding the information they access through the Internet. The report is highly anecdotal and does not shed light on the specific uses of the Internet by public library users. It selectively highlights only positive comments and uses of the Internet. Other studies of Internet use tend to focus on specific user groups such as that carried out by Lazinger, Bar-Ilan and Peritz (1997) at the Hebrew University of Jerusalem. This study examined the use of Internet resources by various members of the faculty through a survey.

A more comprehensive study of Internet use is one carried out at the University of Toronto Library (Tillotson, Cherry and Clinton 1995). This study employed a useful combination of an online survey, focus groups and analysis of logs of telnet connections to monitor use of the specific Internet resources available through UTLink, the University of Toronto Library's text-based online catalogue and electronic information system. This study is distinguished from others by its analysis of the specific sites visited through UTLink and the number of times each was hit. The log analysis, focus groups and the survey questions were used in combination to answer,

among others, questions that the current study is concerned with, "where do they [library users] go on the Internet? What sites do they access...?" (Tillotson, Cherry and Clinton 1995, 190). The University of Toronto study is largely successful in meeting its goals of a limited study of the use of the Internet resources provided by UTLink but does not attempt to answer the above questions in the broader context of the entire Internet as the current study does. The scope of the University of Toronto study is limited to the 1,325 telnet destinations that library users have access to through UTLink. In addition, the items on the questionnaire that address the purpose and use of the Internet compromise the integrity of the data collected. For example, the response categories for the question asking the users' main purpose in using the Internet resources are not mutually exclusive. There is, for instance, a possibility of confusion over the difference between "to look for information" and almost all of the other possible responses.

III.   Methodology

For the current study, the researchers employed a variation on Transaction Log Analysis (TLA) methodology. TLA has typically been used to study the use of online library catalogues in public and academic libraries. It is a method that traditionally examines transaction logs of interactions between people and an online library catalogue. TLA has been used predominantly to understand the ways in which people search online catalogues by recording and analyzing user search strings. (Peter 1993) The major strength of the methodology is its ability to record data in an unobtrusive

and objective manner. It allows for a system's perspective on online searching that represents the users' actual search strings rather than the search he or she might later recollect having done. (Connaway, Budd, and Kochtanek 1995, 143) Researchers in this area have suggested that TLA be used in the future as a collection development tool to study not only how people are searching the online library catalogue but *what* they are searching. (Peters 1993, 57; Simpson 1989, 61) The researchers of the current study felt it was a logical step to apply this methodology to logs that record the Internet sites visited by library users in order to analyze the ways in which this "database" is being used. This study is as much a test of this assumption as a study of Internet use in the Burnaby Public Library.

For the current study, the researchers used logs extracted from City of Burnaby web server for the Burnaby Public Library (BPL) Internet terminals. These logs record the URLs for the actual pages that library users visited or tried to visit from computer terminals at BPL during the study period. By revisiting the sites that are recorded in the log and classifying them under a predefined list of subject headings, the researchers were able to get a sense of the types of sites and subject areas that Internet terminal users were visiting during this time.

The researchers were provided with individual logs for the seven days between December 2 and December 8, 1998. The logs record data from all Internet terminals in the Burnaby Library System - four terminals in the central Metrotown branch and one each in the Cameron, Kingsway and McGill branches. The Sunday, December 6 log represents use at Metrotown only because the other branches were closed that day.

Hours of operation varied slightly from weekdays to the weekend. All branches except

Cameron were open from 9 am to 9 pm Monday through Friday, 9 am to 6 pm on

Saturday. The Cameron branch was open on Monday from 1 pm to 9 pm, Tuesday

through Thursday 10 am to 9 pm, and Friday and Saturday from 10 am to 6 pm. Only

the Metrotown branch was open on Sunday from 12 pm to 4 pm.

All terminals used in the study were equipped with privacy screens designed to

allow only the person standing directly in front of the screen to view its contents. None

of the terminals used any kind of filtering software to block access to sexually explicit

materials and users were not restricted from accessing email accounts such as those

provided by Hotmail. Internet terminal users at BPL are required to sign on to the

terminals using their library card bar code and telephone number. This information is

used by the computer to time out an Internet session after 30 minutes. However, no

information identifying the users of the terminals appears in the logs.

The Log

The log is a standard transfer log kept by all web servers. The transfer log

provides the following data:

_ the IP address (or unique identifier code) of the computer terminal used at BPL
_ the date and time of the transaction
_ the protocol and URL of the file requested from or transferred to the computer
  terminal
_ the status of the transaction (these are codes such as 200 or 404 that indicate
  whether or not the transaction was successful)
_ the size of the file transferred
      Each line in the log represents a single transaction initiated by a user of a BPL

Internet terminal. The transfer log records the activities that are made through a web server. In most cases, these logs are analyzed by web server owners to monitor the requests for web pages or other Internet resources that are made through their server. They are generally interested in where the requests are coming from and what is being requested. (Stout, 1997) In this situation, the BPL Internet terminals access the Internet through the City of Burnaby web server. Therefore, systems staff at the City of Burnaby were able to extract those lines from their transfer logs that represent the requests made and files transferred through the city's web server by the BPL terminals. This was done by isolating the lines containing the IP addresses of these terminals.

Every time a user makes a request to see a web page by either clicking on a link or typing a URL in the location dialog box a line is recorded. The log records one line for each page that is requested and every file that is transferred to the BPL terminal for the patron to view. This includes all .gif and .jpg files that appear on the page as well as the html file representing the page as a whole. Any additional browser windows that are opened by java script programs are recorded as individual lines as well as requests that are redirected from an incomplete address or old address to a new one.   In short, any transaction involving any kind of file is recorded as one line in the log. Therefore, one page viewed by a BPL patron could be represented by a number of lines in the log. Pages that contain many graphic files will be represented by a greater number of lines in the log than those that are predominantly textual.

Sample

Each day's activities produces thousands of lines. A total of 194,644 lines were logged over the seven days examined in this exploratory study. It was therefore necessary to settle upon a sample of these to be studied. The researchers initially chose a random sample of every tenth line of the logs. After beginning the analysis, however, they decided that it was necessary to revise this to a sample of every twentieth line in order to minimize overrepresentation of a single web page. For example, a page that is saturated with graphic files might record 14 lines in the log making it possible for two lines to appear in the sample that actually represent only one visit to a web page. The researchers will visit the page twice (the first line and the eleventh for example) and record it as two visits. It was felt that a sample of every twentieth line would provide data from more individual pages than the initial sample did.

The Metrotown branch data from December 3 and December 5, however, represent a sample of every tenth line as they were the first to be analyzed. The researchers chose not to reject the data already collected for this branch because of the exploratory nature of the study.

Before the sample was taken the researchers stripped the logs of any lines that were known to contain data that was unclassifiable and therefore unusable (see explanations below of the files removed). This minimized the time wasted on collecting data that would have to be rejected later and increased the chances of useful lines appearing in the sample log.

The files or file types removed were as follows:

1) .gif files

These were removed because they generally appear as buttons or icons on web pages not as pages in themselves therefore their "parent" page would remain in the log and have an equal chance as any other page of being included in the sample.

2) BPL homepage

This page is the default page when a user first logs on to the system.   It does not, therefore, represent a conscious choice on the part of the user to visit the page. Other pages on the BPL web site were left in the log.

3) banner ad files

The researchers removed as many of these files as could be easily identified through URL patterns . These were removed because it was difficult to tell on which site the banner had appeared. These files were unclassifiable in themselves as they always appear on another page. An example of a banner ad file removed is 'http://ad.preferences.com/ '.

4) .js, .class and .cs files

These files were removed because they could not be revisited by the researchers and were therefore unclassifiable.

5) blank lines

Lines were removed that contained no information in the URL field.

These files were removed from the logs using a Unix utility called Egrep. Egrep is able to recognize patterns of character strings and was used to extract all of the files in the logs except those with the specified character strings. For example, Egrep can be commanded to locate all lines with strings of characters ending in .gif and produce a new file from the remaining lines.   A total of 115,264 lines were stripped from the logs leaving 79, 380 lines. The lines were then numbered and every twentieth line was copied into sample logs to be studied (except as noted above). The total number of lines analyzed in this study was 4,712.

Jpg files were not removed from the logs because they often exist as pages in themselves. This is particularly true of sexually explicit materials. The researchers did not want to lose this data on the use of sexually explicit materials and therefore decided to leave .jpg files in the logs. Also, removing the .jpg files did not reduce the size of the logs significantly and it was therefore felt that they would not jeopardize the researchers' ability to extract a good sample representing as many individual web pages as possible.

Limitations

The most significant drawback of TLA data is its inability to tell us anything about the users or their information needs at the time of their search. (Kurth 1993, 98; Simpson 1989, 36) Although the logs provide some idea about the searcher's

experience with the database through the context in which the URLs appear in the log, solid conclusions about the user's intent are not possible without an element of subjective analysis. The problem of context is augmented by the nature of the Internet. The use of hypertext and graphical links encourages browsing of Internet resources in a way that is not seen in online catalogue use for example. The propensity of Internet users towards "surfing" the Internet is not the problem so much as the percentage of hits that will be recorded that are "accidental" and do not represent the searchers intent at all. For example, it is easy to click on a link provided through a search engine or another document and visit (and therefore record) a site that is sexually explicit without intending to do so.

Another factor that complicates the analysis of transaction logs is the unstable nature of the database involved. (Kurth 1993, 99) This is doubly true of the Internet. It can be difficult or impossible to replicate searches conducted by an user at an earlier time because of the dynamic nature of Internet resources. Sites can disappear or be revised radically overnight. A number of times during this study, the researchers either were unable to access a site that the user had successfully accessed or were successful in visiting a site for which the user had received a 404 error message. The latter were classified on the premise that by clicking on a link or entering a URL the user showed at least a nominal intent to visit the site.

Data Collection Process

As each site was visited and classified the data was recorded in a Microsoft Access database. The information collected is as follows:

_ day of the week
_ line number in the log
_ BPL branch
_ classification code assigned
_ subheadings for sites that were magazine, ezine, government or non-English language sites
_ actual language of a non-English language site
_ URL

The researchers were interested in not only the general trends in Internet use but in determining any possible differences in use patterns between days of the week and branches. See sections VIII and IX for a description of the findings.

Considerable thought was given to designing an appropriate classification scheme for organizing the web sites into categories. Following the lead of OCLC in using Dewey Decimal Classification to classify Internet resources in the Net First database, the researchers experimented with Dewey numbers on a test log provided by BPL. It turned out to be a very cumbersome process requiring a lot of effort to force a web site into a very rigid classification scheme. While Dewey classification is appropriate for assigning a shelf address to a physical item in a library, it did not prove to be the best solution for classifying Internet sites in the current study. Much of the material available on the Internet is unique to the Internet and therefore does not have a direct correlation in the books and other materials that are collected for a public library. In addition, the goal in this study was not to enhance access to Internet resources but to bring out patterns of usage. The researchers, therefore, felt that it was

more appropriate to design a subject heading list.

The final subject heading list used for the study represents a combination of subject headings used in popular Internet subject directories such as Yahoo! and Magellan as well as the researchers' own experiences with the Internet and the problems encountered while working with the test log. Subject headings for personal ads, online directories, business products and services and non-profit/philanthropic organizations were added while the study was in process in order to accommodate a need for these. The researchers similarly revised their system for recording information regarding non-English language, magazine, and government sites after beginning the analysis work. In all cases, logs previously classified were re-analyzed to include the new subject headings. See section X for a further discussion of the subject headings and suggestions for improvement for any further studies.

The work of classifying the web sites was, by necessity, a subjective process. It required consistent consultation between the two researchers responsible for the classification both before and during the process itself in order to ensure an acceptable level of consistency in the classification codes assigned.

IV.   The Findings by Broad Category

Because of the exploratory nature of this study and the sample size, statistical

analysis using Chi square is inappropriate. Instead, raw totals and percentages are

given and terms such as significant, high, low, etc. are used descriptively not with

statistical significance in the research context. For clarity, some of the original subject

categories have been grouped together under a broader heading. These categories

are indicated with an asterisk*.

Table I: Broad Subject Categories by Rank

	Subject Category	No. of lines	% of total lines
1.	Email	1445	30.7%
2.	Unclassifiable	600	12.7%
3.	Entertainment	386	8.2%
4.	Business	381	8.1%
5.	Search Engines	353	7.5%
6.	Sex	256	5.4%
7.	Computers	208	4.4%
8.	Sports	174	3.7%
9.	Humanities/Social Sciences	131	2.8%
10.	Education	112	2.4%
11.	Hobbies	115	2.4%
12.	Places & Travel	107	2.3%
13.	Newspapers	86	1.8%
14.	Other News Media	73	1.5%
15.	Personal Ads/Classifieds	66	1.4%
16.	Science	25	1.0%
17.	Directories	20	0.4%
18.	Health	19	0.4%
19.	Philanthropic organizations	15	0.3%
20.	Law	3	0.1%

Table II: Findings by Format or Source

Format or Source	No. of lines	% of total lines
Online Periodicals (magazines & ezines)	116	2.5%
Non-English Language	437	9.3%
Government Web Sites	71	1.5%

1. Email*

This category includes email, chat rooms and electronic greeting cards. These

categories combined comprised the highest number of log lines at 30.7% or a total of

1445 lines of the log sample by a fairly large margin. The second highest category, for

example, is Entertainment at 8.2%. These statistics cannot be taken literally as the

numbers are heavily influenced by the nature and patterns of use of email and chat

room sites.

The process of emailing is particularly unique as it involves a number of

activities and functions that are not commonly found in other kinds of Internet sites.

Every time an email function or activity is performed, it is recorded as a line in the

transaction log. For example, patrons will access the email host homepage, log into

their account, access their messages, read their messages one by one, reply to these

messages, and perhaps perform some file management. Each patron may also have

numerous email profiles. Each step and activity is recorded and logged. In a short

span of time, the process of email can produce a disproportionate number of log lines

compared to other types of activities done through the Internet.

Similarly, the use of chat rooms by one patron can account for a large number of log lines that will naturally inflate its numbers and perceived use.

To illustrate how email sites and chat rooms can become overrepresented compared to other types of sites, a typical chat session from the December 8th log is compared to another session where a variety of sites were visited. Both sessions occurred on the same day, over the same length of time on different terminals in the same branch of the library. There is good evidence to suggest that the chat session is from one patron, as the sites visited come from one particular non-English language chat room and there are no breaks in the session from other sites visited. Neither session had a significant break in time where the Internet was not in use. The following table compares the number of log lines generated for each session:

Chat Room session   from line number 5400 ⅋ 5820 = 420 lines = 21 lines in sample   Non-Chat s

Both of these sessions took place at the same time of day over a period of 1 hour and 45 minutes. From the figures above, chat rooms generated almost double the lines in the log sample than the non-chat session. When multiplied over hundreds of chat room visits over seven days, this amounts to a significant figure.

2. Unclassifiable*

This category accounts for 12.7% or 600 lines of the sample log. It includes the data collected for lines that produced error messages, contained banner ads whose source page could not be determined, contained no data or contained data that on its own was unclassifiable and the source could not be verified.

There were a number of reasons why a site was classified in this category.

Many of these will be discussed in the section discussing problems. On a number of

occasions, the researchers encountered error messages when the log showed that the

patron had been successful in visiting a site. Some sites were protected with

passwords and were therefore inaccessible to the researchers. Many times the file in

the log directed the researchers to a part of a page or to a .jpg file. Often it was

possible to verify the page from which this element had come. The researchers were

able to backtrack through a site by removing portions of the URL moving from right to

left and determine the location of the .jpg or other element on a particular page.

Where it was not possible to do this, the line was put in the unclassifiable category.

Some pages in this category simply contained no data.

A large number of banner ad files still remained in the log even after many of

these were stripped from the raw logs before the samples for each day were taken

because many banner ad files are difficult to identify.

3. Entertainment*

This category accounts for 8.2% or a total of 386 lines of the log sample. These

figures reflect the total numbers of the following subject categories: music,

television/film, radio, games and entertainment miscellaneous. Overall, this category

could be summarized as sites related to mass culture.

For all areas, web pages pertaining to entertainment as an industry and those

providing general entertainment information were included here. For example, music

may contain web sites about a particular artist as well as web sites about the business

side of the music industry. Sites that dealt with the arts as a discipline or area of study were not included. For example, a site that discussed different makes and styles of guitars was classified under the arts rather than musical entertainment. Because the guitar site was not commercially motivated, the researchers felt that it was appropriate to place it in the arts category. The researchers acknowledge that some level of subjectivity was involved when assessing whether a site dealt with mass culture or the arts. In the views of the researchers, a crucial difference lay in the business aspect of this category.

This category also included computer games. Of all of the entertainment categories, game sites recorded the highest number of visits at 28.8% of the total entertainment sites. The popularity of this subject over other entertainment categories surprised the researchers. There appears to be a large and sophisticated community of game users, who use the Internet to share strategies and to offer game reviews and ratings. For example, several game sites detailed how to cheat when playing some computer games. These sites often took some time to classify. The language was unfamiliar and referred to terminology unique to the computer games culture.

4. Business Categories

The business categories, when combined, account for 8.1% or 381 lines in the sample log. This subject area will most likely grow as more businesses become aware of the benefits of Internet exposure and more people become cognizant the variety of business information available on the Internet. Each category is discussed individually

below.

Business – Real Estate

This category accounts for 10.0% (38 lines) of the total number of business

pages. Any page that dealt with real estate from advertisements to services was

classified here. Sites classified in this category include listings for houses and property

and web pages of real estate agencies.

The numbers are relatively low for this category in the overall picture and a third

of the visits recorded were generated by one Internet session. The researchers

anticipated that local real estate information would be accessed but found that some

library patrons also looked for real estate information in other countries, particularly in

Asia. The Asian real estate pages also accounted for a large percentage of the total

Chinese language web pages accessed.

Business – Finance & Investments

This category accounts for 18.1% (69 lines) of the total number of business

pages. Finance includes pages dealing with banks and banking, taxes, credit cards

and stocks and other investments.

The 69 pages classified as finance and investment tended to focus on stock

market and other investment information as well as banking. One page that appeared

a number of times over the study period is a gold investment information page called

Gold-Eagle.

Business – Industry

  This category accounts for 24.1% (92 lines) of the total business pages. Web sites classified in this category dealt with forestry, fisheries, agriculture, heavy transportation, mining, petroleum and other prime industries.   To distinguish this category from science, it was decided that pages whose primary purpose was to sell, promote or profile a particular business industry would be placed in this category. Those that dealt more with specific scientific techniques, procedures or theories would be categorized under science. These criteria were applied to all pages including those hosted by a specific corporation or company.

  The researchers anticipated that this category would show high use of government sites, whose web pages are quite informative in this area. The findings show that only two government web pages were accessed. The researchers also anticipated, but did not find, a lot of web pages about forestry and fisheries which are industries associated with British Columbia. Instead, sites about the heavy transportation and mining industries were two of the more frequently visited subject areas.

  There was an interesting mix of sites visited in this category. One rather unusual web site discussed oil prospecting as a spiritual quest, guided by the hand of God. More indicative of the sites placed in this category are pages about Boeing, Airbus, car manufacturers such as Mercedes and quite a variety of sites about gold mining.

Business – Products & Services

This category accounts for 25.7% (98 lines) of the total business pages. It includes web pages that feature a commercial product or business services such as consulting, accounting, graphic design, telecommunications, delivery services etc. This category should not be confused with shopping, which has been defined by the researchers as web pages whose sole function is to perform online commercial transactions of a variety of products.

After analyzing the wide range of web pages categorized in this area, the researchers feel that future studies should consider a finer breakdown of subject categories for this area. Information on product and service information for computers, for example, is included in the general Computer category. In any future study, all of the categories should be refined to allow for the product and service aspects to be distinguished from other more general types of information.

Employment

This category accounts for 22.0% (84 lines) of the total business pages. Employment includes any web page that provides job postings, career counselling or resume and interview tips. Of particular note is the fact that 32% of the 84 employment web pages accessed were from Canadian federal, provincial or municipal government sites. Of the total number of Canadian government sites visited on any subject, 69% were employment related. The researchers felt that the federal government sponsored job posting service might have benefited from the link provided to this page by the BPL

web site. Web pages of companies who listed their own job postings or who described

the career opportunities of their company were also categorized under employment.

50 Search Engines

This category accounts for 7.5% or 353 lines of the sample log. It includes any

web page that represents an attempt to provide some general organization or access to

Internet resources. This would include anything from search engine/subject directory

sites such as Yahoo! or AltaVista to the BPL subject listing of links to selected Internet

resources.

The numbers for this category are high due probably to the broad definition it

was given. It included the My Yahoo! page that allows a person to customize the

information provided through the Yahoo! subject directory. Some of these sites were in

non-English languages or had been used by people looking for non-English language

sites.

The researchers decided to classify under subject headings any search engines

or subject directories that displayed the search string used and/or the hits returned.

For example, a page from AltaVista containing hits for the Calgary Flames hockey team

and a search string of "Calgary Flames" would be classified under sports.

60 Sex
Softcore Sex

This category accounts for 3.3% or a total of 154 lines of the log sample. The

researchers defined softcore to mean pages that show nude or erotic images with no

artistic or literary intent and whose fundamental purpose is to provoke a sexual response in the viewer.

Again, these figures have to be viewed with consideration of the nature and structure of most sex sites. These kinds of sites are heavy in explicit images (e.g. jpg files) which generally translates to a larger number than usual of log lines every time they are visited. For every image within a site, a single log line is recorded.

The final figures are also exaggerated due to the way in which sex sites are viewed. As mentioned, sex sites are dominated by images with little accompanying text and are therefore designed to be looked at, not read. Analysis of the logs suggest that visits to sex sites are short in length but very active. There is a tendency to browse the pages of a web site quickly and to surf around to other sites at an equally rapid pace. The researchers also observed that this kind of surfing is encouraged by most sex sites. On a typical web page, there are prominently placed links to other pages, always with the promise of "bigger and better" explicit images.

One particular situation illustrates this point well. Over a four hour period, the log for December 4th recorded an almost uninterrupted sequence of sex site use. This amounted to a total of 120 lines in our sample - 2400 lines in the raw logs. The researchers compared this figure to three other sessions from logs recorded on the same day, over the same four hour period and from logs generated from different Internet terminals in the same branch. The other three sessions were typically varied in subject and types of sites. These sessions generated a low of 28 lines in the sample to a high of 46 lines in the sample. Like email and chat rooms, sexually explicit sites can

be overrepresented in the log data. The high number of 46 included a large number of

chat room visits which further demonstrates the points in the previous section.

This category may also be influenced by certain techniques used to attract more

visitors to sex sites. Common search terms and phrases are planted within the sex sites

so that they are captured whenever these terms and phrases are used in a search.

Other sites have manipulated their URL to be extremely similar to well-known sites

such as Amazon.com. Unfortunately, the researchers did not have access to

information describing how a patron came to a site and there cannot quantify the extent

to which this method of planting terms or manipulating URLs has impacted the results.

There is some indication, however, that this effected logs to some degree. For

instance, the majority of sex sites appear in the logs as consecutive chunks. This kind

of pattern occurs in non-sex sites as well. On at least two occasions, sex sites

appeared once and seemingly out of context relative to the pages before and after.

Although there may be other reasons to explain such anomalies, the researchers

considered this problem important to consider in the final analysis and for any future

studies.

Hardcore Sex

This category accounts for 1.9% or a total of 90 lines of the total log sample.

The researchers defined hardcore to mean pages that show explicit images of sexual

intercourse, sex involving violence, children or animals with no artistic or literary intent

and whose fundamental purpose is to provoke a sexual response in the viewer.

The rationale behind having softcore and hardcore as separate categories was to allow a distinction to be made between sites that featured highly controversial or even illegal sexual activity and sites that simply featured nude images where no sexual activity is evident.

In the final analysis, the researchers found no web pages featuring violence, children or animals although some sites suggested through graphically worded hyperlinks that they contained this type of content. The majority of sites classified in this category involved, instead, the portrayal of sexual intercourse or sexual activity between two or more parties.

As discussed above, the statistics for this category are similarly influenced by the way in which sex sites are viewed and the disproportionate number of graphic images associated with them.

The distinction between sites that are hardcore and softcore was occasionally difficult to determine. There were situations where a web page showed a site that was, by the researcher's definition, softcore but which also described in graphic language links to images that would be classified as hardcore. The researchers made a decision to classify only the page which the line from the log displayed and to disregard what the entire site may or may not have contained. This decision was based on the fact that links often did not accurately describe what the viewer would actually see if this option was chosen.

There were also a few cases where the researchers were prevented from viewing pages that required a membership or a subscription. In this situation,

classification was based on the homepage of such sites.

Sex Information & Products

This category accounted for 0.3% or a total of 12 lines of the log sample. The researchers classified sites in this category when they provided instruction, advice or discussion about sex and sex related issues such as STDs, contraceptives etc. This would include the display or description of sexual products and devices.

The researchers chose to include a separate category for this topic because of the controversy that surrounds sexual information. Because BPL's Internet terminals are not filtered, it was also of interest to the researchers to see how many sites of this nature would be accessed by patrons. It also allowed the researchers to treat sexually explicit sites in a way different from sites that are meant to be informative or educational but also deal with sexuality on some level.

Given the criteria, there were web sites that did not neatly fit into any of these sex categories. For example, a site called "Men on the Net" combines information on health and other issues concerning gay men as well as access to erotic stories and sexually explicit photographs.

Of the pages placed in this category, very few discussed health issues. This is consistent with the relatively low number of medical related pages accessed by library patrons in general, as will be discussed later. Most of the pages in this category discussed sex as a lifestyle or were commercial sites for sex products such as condoms.

70 Computers

This category accounts for 4.4% or a total of 208 lines of the log sample. It includes web pages about software, hardware, products, services, web publishing, computer tutorials, web site services etc.

Pages from Microsoft's site had a big presence in this category. Library patrons accessed everything from Microsoft's product information pages to their web publishing tutorial pages. Otherwise, the web pages placed in this category varied widely. The researchers felt that this category should be refined for future studies.

The relatively high numbers for computers may be another reflection on how the Internet can be a valuable source of information that the library may not otherwise be able to support in its collections. Anecdotal information suggests that Burnaby Public Library stocks large numbers of computer books but still has difficulty meeting customer demands. From the experience of the researchers who consulted the Internet for technical questions related to this study, the Internet is a rich source of current computer information that can be accessed quickly and at varying levels of complexity. If the statistics for this category are any indication, the Internet can be a valuable supplement to the computer book collections of public libraries.

80 Sports

This category accounts for 3.7% or a total of 174 lines of the log sample. It includes not only web pages about sports news and athletes but also sports as an industry. For example, sites promoting brand name athletic goods were placed in this

category as well as pages featuring the latest sport scores. It was also decided that web pages from the sports section of online newspapers or news services were placed in this category.

One of the sports which dominated this category was soccer. Of particular interest was the presence of non-English language soccer sites often devoted to specific teams. It was the impression of the researchers that this was another way for library patrons to maintain contact with their home country. For example, web pages from the sports section of non-English newspapers or news services featuring a country's soccer teams were commonly encountered in the logs.

The other sport which had a large presence in this category was World Wrestling Federation (WWF) wrestling. In retrospect, the researchers would consider categorizing these web pages in television entertainment which more accurately describes this activity. Unlike soccer, the wrestling sites dealt with wrestling figures and commercial products associated with these figures rather than current news.

90 Humanities & Social Sciences*

This category accounts for 2.8% or 131 lines of the total sample log. A number of individual categories have been collapsed due to the small numbers represented by each individually. These include History, the Arts, Social Sciences and Language and Literature categories.

This category contains a large number of pages that represent what the researchers, for lack of a better word, refer to as counter culture. Web pages

concerning what would normally be considered alternative lifestyles and interests such as "gothic" culture, vampires etc. were classified under Social Sciences and therefore are included in this category. The Internet is an ideal venue for this type of subculture and allows for the sharing of ideas by people in countries around the world. Much of this material is available only on the Internet as it represents people's personal interests and lifestyle choices. This use of the Social Sciences category was not anticipated but became necessary when this type of material appeared in the logs. In any future studies of this nature, it would have to be accounted for with a subject heading of its own.

The other web pages in this category generally discuss more mainstream subjects such as archaeology, psychology, museums, fine arts and music. The web pages classified under the Social Sciences tended to be more ambiguous than those in the other areas. This indicates a need to re-examine this subject heading and further refine it into more meaningful categories.

100   Education
This category accounts for 2.4% or 112 lines of the sample log. Web pages describing course offerings, scholarships and general information on educational institutions or their departments were classified under this heading.

The Internet provides current and comprehensive information on the subject of education and it was not surprising to see a fairly high level of use. It is a good source of information and web pages from both local and international institutions appeared in

the logs. While information on local institutions is generally easily found in the public

library in the form of course catalogues, information on national and international

institutions can be more difficult to find.

Sites classified in this category were generally unambiguous. The researchers

decided to include information on specific departments under Education rather than

their subject area (e.g. University of British Columbia's Faculty of Science in Education

not Science) because of the emphasis upon the department's position within the

institution rather than the discipline being studied.

110   Hobbies / Personal Interests*
This category accounts for 2.4% or a total of 115 lines of the log sample. It

includes hobbies such as gardening, cooking, carpentry, crafts etc. as well as personal

web pages and shopping sites.

The shopping category was used exclusively for web pages that displayed

commodities for the sole purpose of conducting online commercial transactions. It was

used only 49 times. By definition, this category assumes that the library patron is using

the Internet to purchase goods which is not always the case and the researchers did

not want to misrepresent use in this way. For example, the online bookstore

Amazon.com is in the business of selling books online but the site also offers book

reviews and information about authors. In this situation, if a web page from

Amazon.com was encountered that provided information, it was classified under

language and literature. Similarly, if a web page from this site was related to the

purchasing of a book, it was classified as shopping. The same criteria was applied to similar commercial web sites. Some web sites, for example online auctions or catalogues, were viewed as shopping sites.

There was a surprisingly low occurrence of sites related to hobby activities. It is difficult to determine an explanation for the low numbers but it may be related to the availability of other library resources, such as books, on various hobby activities. For example, very few recipes or gardening web sites were accessed perhaps because the library is well-stocked in these areas.   On the other hand, eclectic hobbies such as virtual reality airplane flight simulation were encountered that are only suited to the multi-media capabilities of the Internet.

The researchers acknowledge that the hobby category is in itself problematic. It involves some subjective interpretation of what constitutes a hobby and what does not. The researchers looked at sites such as Yahoo and Magellan for some guidance when defining this category.

Personal web pages also did not figure prominently in the log sample. There was one particular personal web site that recurred in the logs despite the fact that it did not deal with a specific subject or topic of any kind. This suggests that the the web page was known by the library patron and deliberately accessed.

120   Places & Travel
This category accounts for 2.3% or 107 lines of the sample log. Web pages providing specific information on a particular place including weather or on the travel

industry in general were classified under this heading.

Many of the pages classified under this heading were pages giving comprehensive information about cities or countries. In addition to providing information to would-be travellers or researchers, these pages often provide links to information sources on current news, newspapers, and other information that would also be of interest to people living away from their native country. While many different places are represented in this category, it appeared that there was also a small number of users accessing information on specific countries several times over the week. Many of these were African nations such as Sudan, Algeria and Eritrea. Without information on the users of the Internet terminals this is simply an impression drawn from the sites and the context of the log in which they appeared.

One example of the pages in this category presented an appealing story to the researchers. It is a good example of the kind of "feel" for the person doing the search that was often experienced by the researchers during the analysis of the logs. On the December 7 sample log, a block of consecutive lines began with a weather site for Reno, Nevada.   This site was categorized under Places and Travel. The next line, however, was a page discussing things to do in Las Vegas. The following lines proceeded to explore wedding gowns and Las Vegas chapels. The patron finally seemed to settle upon a Hawaiian themed Las Vegas style wedding!

130   Newspapers*
This category accounts for 1.8% or a total of 86 lines of the log sample. It

includes online newspapers from Canada (in English and non-English languages), the United States and outside of Canada and the U.S. The specific language of any non-English newspapers was also recorded for each web page.

Although the final numbers appear to be low, they are diminished due to the unnatural weight given to email and sex sites for reasons already discussed. In the opinion of the researchers, newspaper sites merit special attention for several reasons. Newspaper sites are heavy in text and lighter in graphics which decreases their presence in the log data. Most importantly, because newspapers are text oriented with multiple subsections, they require more time to read compared to other kinds of sites such as sex pages.

The results of the newspaper category left a strong and positive impression on the researchers as to the value and necessity of Internet access in public libraries. Newspapers in particular suggest that the Internet provides an unique way for patrons to maintain contact with their homeland or with places which are simply of interest to patrons. This was a recurring feeling on the part of the researchers throughout the study. Of the total number of newspapers accessed, 79% (68 lines) of them were from places outside of Canada and the U.S. Of this percentage, 91% of these newspaper sites were also in a language other than English. Languages that dominated this category included Russian, Turkish and Spanish and languages from the former Yugoslavia. Given the large Asian community in Burnaby, this figure would most likely have been higher if the library was equipped with the necessary language software at the time the study was conducted. On the other hand, it could also be that Asian

language print newspapers are widely available through newspaper outlets in the city. The sense that the Internet is used as a vital link to places outside of Canada was reinforced by the fact that many of these non-English newspaper sites were encountered again and again throughout the week.

140   Other News Media*
This category accounts for 1.5% or a total of 73 lines of the log sample. Other news media refers to web pages from television, radio and web-based news services. This category also includes web pages that discuss politics and current affairs but are sites that are not be affiliated with an official news agency.

The results of this category did not show the same consistency as the newspapers category. There was a wide range of news services visited and the majority of them were in English. For example, CBC Newsworld, CNN and BBC were all visited frequently throughout the week. Of particular note was the popular use of headline news services of search engines and directories such as Yahoo and AltaVista. However, many of the news wires offered by search engines and directories were actually provided by television networks such as ABC News.

Non-English language news media still had a significant presence in this category, although to a lesser degree than was seen in the newspaper category. Non-English language web pages accounted for 40% of the total number of pages accessed for other news media. Again, languages of the former Yugoslavia, Russian, Turkish and Spanish were some of the more prevalent languages encountered. The consistent

occurrence of these languages suggest that a regular group of library patrons make heavy use of the Internet terminals at this library.

150   Personals Ads & Classified

This category accounts for 1.4% or 66 lines of the log sample. It includes web pages containing advertisements similar to those that would appear in newspaper classified sections such as ads for personal companionship and other classified listings.   This category was added after the researchers had begun analyzing the logs. This use of the Internet was not anticipated ahead of time and only became apparent after the study began (logs previously classified were re-analyzed to include it). The category is dominated by the personal ads. Very few of the sites contained ad type information typical of a newspaper classified section.

The array of personal ad sites available on the Internet was interesting to observe. These types of sites cater to almost every imaginable group. The researchers came across sites assisting Muslim men and women in finding suitable partners for marriage, Asian dating and marriage sites and a site profiling eastern European women. Many of these sites included photographs of the people placing the ad making them more informative than newspaper personal ads.

160   Science, Health

Science accounts for 1.0% or 45 lines of the sample log while Health accounts for 0.4% or 19 lines of the sample log. Web pages that were related to science subjects such as biology, math, chemistry, physics, geology etc. were included under

the Science category. Any pages pertaining to health topics including pharmaceuticals were categorized under Health.

The researchers anticipated a greater use of both of these categories than actually materialized. In particular, the number of health related web pages accessed is much lower than expected.

A factor that possibly effects the numbers of health and science web pages visited is the period in which the study was conducted. The study period was in the first week of December when students were studying for exams rather than working on assignments. Their need for this type of information to complete assignments would therefore not be as great during this time.

170   Directories, Organizations
Directories accounts for 0.4% or 20 lines of the sample log while Organizations accounts for 0.3% or 15 lines of the sample log.   Directories are any lists of addresses or names of employees, government officials etc. that would normally be found in directory sources in the library. The Organizations category includes philanthropic or charitable organizations or associations that could not be classified in a particular subject area. Directories and Organizations were added as categories after the start of the log analysis because of a perceived need.

Directories were not accessed particularly often but were given a separate heading because of the difficulty of classifying them in any other category and the importance of the Internet for contact information. There was, unfortunately, no way of

knowing how often the Internet was used for contact information when looking at sites which were not specifically directories. For example, someone looking at the homepage of a law firm may have been looking for the firm's address or for information about law. The researchers' own experiences with the Internet suggest that it is a good source for contact type of information but without knowing the motivations of the Internet terminal users it was impossible to tell when a site was accessed for this purpose.

The number of organization web pages that were accessed is greater than is represented by the numbers given above. The researchers preferred to classify organizations within a subject heading whenever possible.   For example, a scientific association's web page would be classified in the science category. Philanthropic and charitable organizations do not generally fit nicely into a subject area and the researchers felt that their presence in the logs could be brought out with a category of their own. In the present study, the category evolved during the course of the classification work and does not present as clear a picture as it could. The sites currently classified in this category include human rights organizations and the OXFAM site.

180   Law
This category accounts for 0.1% or a total of 3 lines in the log sample. The law category was included in the list of subject categories to provide a classification for web pages about law firms, laws and legislation and legal institutions.

The sites included in this category are a HotBot search for the Justice Institute of B.C., a guide to the Employment Standards Act produced by the B.C. government and a web page of a law firm specializing in patent and intellectual property law. The Internet is a good medium for the dissemination of this kind of information. Perhaps a greater use of the law materials available on the Internet will come as users become more knowledgeable about the types of information available on the Internet and more sophisticated in their searching. It would be interesting to observe the use of law information over a longer period of time to see if this low pattern of usage is the norm or a result of the particular study period.

V.   Government Web Pages

Identification of government sites occurred in two ways. First, sites which pertained to no particular subject were categorized as "generic" federal, provincial, municipal or foreign sites. Second, sites which pertained to a subject (e.g. science, employment) were categorized first under that subject and then given a subheading which indicated the level of government involved.

Federal Government

0.7% or 32 lines of the sample log were federal government pages on a variety of topics. Of these, 22% or 7 lines were given the generic government-federal subject category, 66% or 21 lines contained employment related information, and 12% were placed in either business-industry or science. A web page that was accessed often

was the Human Resources Development Canada "Job Bank" page.

Provincial Government

This category accounts for 0.3% or 13 lines of the sample log. Pages in this category are from any one of the provincial government web sites in Canada.

As with the federal government pages, the greatest number of pages pertain to employment: 6 of the 13 pages recorded as provincial government pages were categorized under this category. Four were given the general provincial government heading - two from British Columbia and two from Nova Scotia. One page was categorized as law while two were put in the places and travel category.

The numbers of all of the government categories are low and it is difficult to get a sense of how these sites are being used in the public library. Web site statistics from the government web servers might provide better data on the use of these sites.

Municipal Government

This subheading was only applied to 2 lines or .04% of the sample log. One site that was accessed was the page for the City of Burnaby Parks, Recreation and Cultural Services. The patron visiting this site may have been able to follow a link from the Burnaby Library homepage to reach this site. The BPL web site contains a number of government links.

The other site under this subheading was a City of Vancouver policy report from the Standing Committee of Planning and Environment. This information can at times

be difficult to track down in print if the library patron is unfamiliar with government documents but is relatively easy to find on the Internet using a search engine.

Foreign Government

This category accounts for 0.5% or 24 lines of the sample log. It includes all sites that were from foreign government sources including embassies.

The numbers in this category are slightly inflated by the data from one day. There was one session in which the patron browsed extensively through one Iranian government site. The use of foreign government sites was not spread evenly over the days. The site for the Texas legislature appears in this category as does an embassy site from which a person can download a passport application.

VI.   Non-English Language

The area of non-English language web page use was of particular interest to BPL and the researchers and therefore has been analyzed more closely in this study. Identification of non-English language web pages was accomplished primarily through use of the subheading "non-English language" and the recording of the specific language (e.g. Sports - non-English language - Spanish). Only when the subject of the web page could not be determined was the site given the "generic" category Non-English Language. This occurred in only 39 of 437 instances.

Non-English language web pages accounted for 9.3% or a total of 437 lines of the log sample. Grouped together this way, the significance of the number of non-

English language web pages visited is apparent. To put this figure in context, non-English language would be the third highest category overall following email and entertainment (excluding unclassifiable sites).

A total of sixteen different languages were recorded in the study. The five languages that appeared most frequently in the log sample were (in order of frequency): Russian, languages of the former Yugoslavia, Spanish, German, and Chinese. Russian web pages can be explained in part by the numerous visits to a Russian chat room throughout the seven day period. It is interesting to note that all instances of Russian web pages came from BPL's Metrotown branch location. These findings were corroborated by the staff at this location who are aware of a group of Russian speaking patrons who make extensive use of the library Internet terminals.

Table III: Non-English Languages by Rank

Language	No. of lines	% of non-English language sites
10 Russian	127	29.1%
20 Yugoslavian languages	99	22.7%
30 Spanish	85	19.5%
40 German	37	8.5%
50 Chinese	21	4.8%
60 French	20	4.6%
70 Turkish	18	4.1%
80 Asian languages	8	1.8%
90 Korean	8	1.8%
100 Japanese	3	0.7%
110 Slovakian	3	0.7%
120 Polish	2	0.5%
130 Czech	2	0.5%
140 Hungarian	2	0.5%
150 Swedish	1	0.2%

160 Italian   1   0.2%

Web pages in languages from the former Yugoslavia showed more diverse subject areas than did Russian sites. Online newspapers and other news services characterized the visits to this language group. There was also good use of Yugoslavian language guides to the Internet such as the Serbian Café, which was seen frequently in the log sample. It provided everything from headline news to subject links to Internet sites.

Spanish web pages were unique for several reasons. It was the only language that occurred consistently across all branches. In addition, the kinds of Spanish web pages accessed were so diverse that they cannot be easily generalized into any one subject area. The highest subject areas accessed included newspapers, humanities and social sciences, business and magazines.

German language web pages were dominated by sports-related information including a German soccer chat room. All occurrences of German web pages were from the Metrotown branch. These findings may reflect the use of one individual or at least a small number of people. Similar to the other languages, German newspaper sites were visited on numerous occasions.

Given the large Chinese community in Burnaby, the number of Chinese language pages visited is probably lower than it would have been if BPL was equipped with the language software needed to view Chinese characters at the time the log was collected. Of the sites that were accessed, real estate pages was the highest subject category. This finding is somewhat misleading, however, since all of the real estate

pages viewed and therefore most of the Chinese language pages accessed, come from one particular session. French, the next most frequently accessed language, was spread more evenly through the days indicating a more consistent use. Unlike other languages, there was a definite lack of news related Chinese language sites visited, presumably due to the lack of software. However this may also be related to the fact that Chinese language newspapers are widely available at BPL and at newsstands throughout the city.

The researchers were interested in the very focused use of French language sites. The sites accessed were almost exclusively from Quebec or Algeria. The Quebec web pages ranged from sites about Montreal and Quebec City to French language magazines and newspapers such as Le Devoir and Planète Québec. The Algerian sites were predominantly political in nature including two online newspapers, Liberté and El Watan.

The researchers felt that the presence of non-English language pages in the logs highlights one of the benefits of Internet access in public libraries. The findings discussed here and by subject category reflect how the Internet can be a unique source of information that might not otherwise be available in the library or elsewhere. The pervasive and consistent use of newspapers and other news media in almost all non-English languages suggest that the Internet is a way for library patrons to maintain a link with their native provinces and countries.

VII.   Online Periodicals (Magazines & Ezines)*

Web pages of online magazines and ezines were recorded in two ways for this study. Each was given a separate subject category as well as a subheading. Magazines and ezines were classified under their own category only when the site was too broad to be categorized under a specific subject. All web pages, whether classified under another subject area or under their own category, were given the subheadings for magazines or ezines. For example, whenever possible a magazine site was classified under its subject area and denoted with the magazine subheading in order to bring out both aspects of the site. This richer level of classification allows for librarians to compare patron use of online periodicals versus print magazines.

It was often difficult to conclude, however, with any certainty that some sites were in fact online versions of print magazines, ezines or just an irregularly updated serial. Because of this uncertainty, the researchers decided to collapse magazines and ezines into one category under the heading online periodicals.

The total percentage of online periodical web pages was calculated based on the total number of log lines given the subheadings for magazines and ezines. Online periodicals, therefore, accounted for 2.5% or a total of 116 lines of the log sample. Of this total, the top three subject areas visited were social sciences, computers and sports. These subject areas lend themselves to regular publications and updates. For example, a German-language online sports magazine was encountered a number of times. There was a notable lack of newsstand or well-known magazines that were visited such as Time, Maclean's or Sports Illustrated.

VIII.   Internet Use by Day of the Week

In this study, analyzing patterns of Internet use by day of the week is very limited for several reasons. Because transaction logs were only collected for one week, it is difficult to make comparisons between the days as there is only one set of data recorded for each day. In addition, BPL's hours of operation are different on weekdays and weekends which affects the size and diversity of web sites recorded for each day. For example, the Metrotown branch is open for 12 hours on the weekdays but only 9 hours on Saturday and 4 hours on Sundays. A lower number of patrons was also represented in the logs with shorter opening hours. In order for there to be a meaningful comparison of Internet use by day of the week, a larger sample collected over several weeks would be needed.

Given these limitations, there are a few observations that can be made. The log sample for December 6th (a Sunday) appeared to be unusual compared to the rest of the week. On this day, 51% of Internet use was for email, which is much higher compared to other days. In addition, this was the only day in the week where no sex sites appeared in the log. It is difficult to conclude why patterns of Internet use on December 6th differs in these ways. The high use of email and lack of sex sites on this day does not necessary characterize Internet use on Sundays in general. A much larger sample of Internet use would have to be collected for several Sundays before meaningful patterns can be detected. It is also important to consider the fact that only

the Metrotown branch of BPL is open on Sundays for only four hours.

IX. Internet Use by Library Branch

Comparing Internet use by branch revealed a fair amount of consistency in terms of subject areas visited by patrons. It should be kept in mind, however, that the hours of operation at the Cameron branch are shorter than the other branches. Between Monday and Saturday, it is open a total of ten hours less than the other branches. Metrotown is also the only branch that is open on Sundays. To analyze Internet use per branch, the subject categories were ranked in order of most visited for each branch and then compared. The results are listed in the table below (excluding unclassifiable):

Table IV: Top Broad Subject Categories by Rank per Branch

	Metrotown	Cameron	Kingsway	McGill
1.	Email	Email	Email	Email
2.	Business	Entertainment	Search Engines	Entertainment
3.	Search Engines	Search Engines	Business	Search Engines
4.	Sex	Personals	Entertainment	Computers
5.	Entertainment	Sports	Sports	Business
6.	Computers	Business/Computers	Computers	Hobbies

As shown above, there are a few categories that figure highly in all branches. They differ slightly in the order in which they appear on each list but otherwise the results are very consistent. Email, Search Engines, Business, Entertainment and Computers all appear in each list. Cameron shows a high number of visits to personal ads which is somewhat of an anomaly as this category is ranked 15th overall for all branches combined. Hobbies is high on the list only for the McGill branch.

These rankings would change, however, if the total number of non-English web

pages per branch were configured into the results. Non-English language would be second highest at Metrotown and fifth highest for both Kingsway and McGill. This demonstrates the importance of the access provided to non-English language materials through the Internet.

The diversity of non-English languages per branch may be a reflection of the varying degrees of diversity in the patrons of each branch. Metrotown showed the widest range and highest numbers of non-English language web pages accessed through their Internet terminals. To a lesser degree, Kingsway and McGill showed a relatively wide range of different languages with five languages represented. The Cameron branch had the lowest percentage of non-English language web pages visited and the lowest number of different languages represented. There were only two languages found and the appearance of these can be largely attributed to two isolated Internet sessions. Spanish was the only language that appeared consistently in all branches.

It was also observed that Internet use at the smaller branches was concentrated in a few subject areas compared to the diversity seen at Metrotown. Although purely speculative, the researchers wondered how Internet use was influenced by the number of Internet terminals available in each branch. Because Metrotown has four terminals, time constraints and pressure from others waiting for the terminal may be less of a factor at Metrotown.

Another variation between branches is the higher percentage of sexually explicit materials that were viewed at the Metrotown branch. Metrotown shows a 7.0% use of

sexual material and the next highest is Kingsway with 4.4%, McGill at 2.2% and

Cameron at 1.1%.   These differences may be due in part to the physical arrangement

of Internet terminals at each branch. The terminals at Metrotown are situated in a low

traffic area and away from the view of the reference desk. The Internet terminals at the

other branches are all in highly visible areas. Another factor may be demographic

differences between the library users per branch, such as the age and gender of

Internet users. Further studies would be needed to make any solid conclusions.

Overall, Internet use per branch was fairly consistent. It is the opinion of the

researchers that more conclusive results would require a longer study period where

bigger samples of Internet use would be recorded. It is difficult to measure the extent

to which the results were influenced by unusually long sessions by one individual user

or by other anomalous circumstances.

X.   Research Issues

Many of the problems encountered during this exploratory study have been

discussed in previous sections. The following are notes providing a more general

discussion of the problematic issues faced during the study.

Overrepresentation

_ One of the biggest problem facing the researchers when working with data from the

   transfer logs is the overrepresentation of some types of pages in the logs due to

   either the nature of the pages or the ways in which people use them. The most

   obvious examples of these are email and sexually explicit pages. This problem has

   already been discussed in the explanations of these two categories. It is a concern

for any further study and more work is required to find a way to balance this somewhat uneven representation or to determine the extent to which it affects the other data. Perhaps more attention should be paid to the work being done by web server owners in analyzing their web site statistics to determine if there are ways to extract lines representing one visit to a page in the raw logs as opposed to a line by line analysis. The field in the log that records the time of each transaction might also be used to check the relative time spent on certain areas such as email and sex sites as opposed to simply straight numbers of lines in the log.

Unclassifiable data

_ Another issue to consider is the large number of lines that produced unclassifiable data even after work had been done to strip the logs of these kinds of files. The researchers did not expect to be able to classify everything that was found given the nature of the logs but the percentage of data that remained unclassifiable was slightly higher than was desired. There is still room for improvement in this area. For example, techniques to determine on which page a banner appears to remove even more of these files from the logs before a sample is taken should be explored. There will always be problems with web pages that disappear or that are password protected.

Disappearing and changing web pages

_ The disappearance and revision of web pages raises another concern which is the researchers' inability to determine in many cases if they are viewing the same page

content as the library patron did during the study period. For example, the Burnaby Library web site was revised from the time that the logs were collected to the time of analysis, changing the links available to customers.

Study period

_ It is important to keep in mind the period in which the study is conducted. There is anecdotal evidence to support differences in use from one part of the year to the other. For example, in the spring months more students use the Internet terminals to access old copies of provincial examinations. It is also logical that any events that occur at particular times of the year might effect Internet use. Any further studies should pay careful attention to this consideration particularly if two or more libraries are involved.

Nature of the logs

_ The classification of the sites was slowed down by the fact that many of the lines represent parts of pages that are only classifiable after the researcher has backtracked through the site to determine the location of the file represented in the log line. As mentioned in the methodology section, many files that appear as lines in the log take the researcher to an icon or .jpg file or one section of a frame for a page making it difficult to classify the file without further investigation. The researchers were generally able to classify about 100 sites per hour. However, non-English language or multi-topic pages and files such as the ones mentioned above slowed the process down to approximately one-third this number.

Subjective nature of classification

_ In addition to slowing down the work of classification, multi-topic pages also leave

themselves open to multiple interpretations and therefore classifications. It is very

important for a each subject category to be clearly defined. All people involved in

the classification must consult with each other both before and during the analysis

of the logs in order to ensure a high degree of consistency. While attempting to

minimize inconsistencies in classification, the researchers of the current study

recognize that due to the subjective aspects of the classification process some

inconsistencies exist. These are greatly minimized however by the amount of

communication that occurred while the logs were being classified.

Java script

_ A problem encountered by the researchers while classifying the web pages was the

presence of java script on some pages resulting in the opening of multiple browsers

once a page had been accessed. This occurred with some sexually explicit pages

that were viewed. Often the java script was triggered when the researchers tried to

close the browser or hit the home button to move away from the site. When this

happened, the viewer had no option but to wait until all of the windows had opened

before he/she could shut them down. The java script was not enabled on BPL

terminals during the study period but any further studies would have to take this into

account. Each of these browser windows will be recorded as a separate line in the

log greatly inflating the numbers for sexually explicit materials.

Further refinement of some subject categories

_ A lot of work was spent on developing a useful subject heading list to categorize the

web pages in the logs. While it was largely successful, there are areas in which it

can be improved. One of the primary goals of this exploratory study was to work out

any problems in the methodology. Some of the problem areas were business and

humanities/social sciences. These topics have been discussed in the appropriate

sections above. Further refinement is necessary, however, to bring out more of the

subtleties in the way the Internet is being used. For example, the business -

products, services etc. category was useful but functioned as a catch-all category.

Much product and service information was recorded under other subject categories

such as computers if the products were computer-related or sports for sports related

products such as running shoes. These categories should perhaps be broken down

into, for example, sports (scores, team information etc.) and sports - industry

(athletic equipment) in order to accommodate the business elements of the web

pages as well as the general subject area. A further examination of the categories

used by existing Internet subject directories would also be useful coming from the

more knowledgeable perspective of having examined a large number of pages

during the study.

_ A larger study conducted over a larger period of time would provide more solid data

in all areas. A larger amount of data would hopefully even out some of the

anomalies that no doubt occurred during this study and present a clearer picture of

trends in public use of the Internet in public libraries such as Burnaby Public

Library.

XI. Recommended Future Studies

The researchers consider this pilot study to have successful in meeting its

objectives. In the process of completing the study and analyzing the results, a good

sense of the issues and challenges of conducting such a study were identified. Despite

some problems encountered during the study, the overall impression of the researchers

concerning the use of this methodology in studying Internet use is a positive one. The

problems and mistakes made during the current study are not insurmountable. Once

these have been ironed out, the resulting data will provide the "sense" of Internet use in

the public library that the researchers hoped it would. Any future studies should be

carried out over a longer period of time in order to provide more solid indications of

patterns of use.

Of particular interest would be a Canada wide comparison of Internet use in

public libraries. This would require considerably more organization and would have to

account for differences in the type of Internet access provided by different public

libraries. Even a comparison of the libraries in the lower mainland of Vancouver would

have to consider the differences in the ways in which Internet access is provided. For

example, Burnaby Public Library does not use any filter or blocking software but does

employ privacy screens while the opposite is true of the Vancouver Public Library

system which filters Internet terminals in the children's department, blocks access to

popular email sites and does not use privacy screens. All of these differences and the many others that exist would have to be taken into consideration.

The importance of a Canada wide study is the amount of data it would provide. It would consolidate the findings of any smaller studies to determine if trends do in fact exist in the way in which the Internet is used by patrons of public libraries that are divorced from the influences of specific locations. Equally as interesting, however, would be the local differences that are revealed.

Appendix 1.   Subject Headings

email	computers & computer industry
chat, message boards, bulletin boards	science
	health & medicine
sex - information & products	education
sex - soft core	places & travel
sex - hard core	
	history, religion, and other humanities
search engine/subject directory	painting, dance, music and other arts
	anthropology, psychology, and other social sciences
business - real estate	language & literature
business - personal finance	
business - industry	sports & sport industry
business - products & services	entertainment industry - miscellaneous
employment	entertainment industry - music
	entertainment industry - TV/film
newspaper	entertainment industry - radio
newspaper - Canadian (English/French language)	games
newspaper - Canadian (non-English/French lang.)	
newspaper - American	hobbies
newspaper - outside Canada/US	personal web page
	shopping
other news media	personals/classified
politics & current events	
	directory
magazine	philanthropic, non-profit organizations/NGO's
ezine	
	foreign language
government - federal	404 - patron
government - provincial	404 - researcher
government - municipal	404 - password protected
government - foreign	unclassifiable
law	banner

Subheadings:	
non-English language	
magazines	

ezines	
government - federal	
government - provincial	
government - municipal	
government - foreign	

References Consulted

Connaway, Lynn Silipigni, John M. Budd, and Thomas R. Kochtanek. "An Investigation
   of the Use of an Online Catalog: User Characteristics and Transaction Log   Analysis."

Flaherty, Patricia. "Transaction Logging Systems: A Descriptive Summary." *Library Hi
   Tech* 11, no. 2 (1993): 67-78.

Harvey, Kathy, and Tom Horne. "Surfing in Seattle: What Cyber-Patrons Want."
   American Libraries 26 (1995): 1028-30.

Jul, Erik. "Cataloging Internet Resources: Survey and Prospectus." *Bulletin of the
   American Society for Information Science* 24, no. 1 (1997): 6-9.

Kaske, Neal K. "Research Methodologies and Transaction Log Analysis: Issues,
   Questions, and a Proposed Model." *Library Hi Tech* 11, no. 2 (1993): 79-86.

Kurth, Martin. "The Limits and Limitations of Transaction Log Analysis." *Library Hi Tech*
   11, no. 2 (1993): 98-103*.*

Lazinger, Susan S., Judit Bar-Ilan, and Bluma C. Peritz. "Internet Use by Faculty
   Members in Various Disciplines: A Comparative Case Study." *Journal of the
   American Society for Information Science* 48, no. 6 (1997): 508-18.

McClure, Charles R. "User-Based Data Collection Techniques and Strategies for   Evaluating

Perry, Clifford. "Travelers on the Internet: a Survey of Internet Users." *Online* 19, no. 2
   (1995): 29-34.

Peters, et al. "An Introduction to the Special Section on Transaction Log Analysis."
   Library Hi Tech 11, no. 2 (1993): 38-40.

Peters, Thomas A. "The History and Development of Transaction Log Analysis." *Library
   Hi Tech* 11, no. 2 (1993): 41-66.

Simpson, Charles W. "Opac Transaction Log Analysis: the First Decade." In *Advances in Library Automation and Networking: A Research Annual.* Vol. 2*,* ed. Joe A.   Hewitt, 35

"The Statue of Libertines?" *The Library Association Record* 99 (1997): 296

Stout, Rick. *Web Site Stats: Tracking Hits and Analyzing Traffic.* Berkeley: Osborne/McGraw-Hill, 1997.

Tillotson, Joy, Joan Cherry, and Marshall Clinton. "Internet Use Through the University of Toronto Library: Demographics, Destinations, and Users' Reactions." *Information Technology and Libraries* 14, no. 13 (1995): 190-98.