INFORMATION SEEKING

Searchers Working Paper 10

Vivienne Waller
August 2010

The Searchers

Institute for Social Research
Swinburne University of Technology
Acknowledgements

This Working Paper is part of the Australian Research Council project The Searchers. We would like to thank the many staff at the State Library of Victoria for their assistance and suggestions, in particular Sue Hamilton, Ian Patterson and Janice van de Velde. This research was funded by Australian Research Council grant LP077215, conducted under the guidance of Chief Investigators Professor Julian Thomas and Professor Denise Meredyth and with the assistance of Dr Ian McShane and Dr Denny Meyer. The Institute for Social Research takes all responsibility for the content. The views of the author are not necessarily those of the State Library of Victoria.
Overview of report

This report is the culmination of the quantitative analyses conducted as part of the Searchers project. It brings together previous analyses and adds new empirical data, making a contribution to research on current information-seeking and the role of public libraries in online information provision. It makes broad comparisons between the use of the State Library of Victoria (SLV) catalogue, the SLV website, Google and Wikipedia. Specifically, it compares the subject of searches on the SLV catalogue with the subject of searches on Google, the subjects of searches that lead people to Wikipedia and the subjects that lead people to the SLV website. It provides estimates of the numbers of people who use each information source. Hence, it provides information both on information-seeking (use of Google and the catalogue) and the accessing of specific information resources (SLV website and Wikipedia).

Background to the analysis

The Searchers project examines the strategic challenges for major public libraries presented by the online information environment. The project has three components: research on current information-seeking and the role of public libraries in online information provision, analysis of wider developments in online environments and search models, and evaluation of the policy and strategic implications for public libraries highlighted by the research. The project will provide guidance for the State Library of Victoria (SLV or State Library) in developing appropriate models of information provision and contribute new thinking on the role and position of major libraries in the digital age.

This paper contributes to the series of Working Papers produced as part of the Searchers project. Other titles in the series are:
Searchers Working Paper 3: Virtual Visitors to the website of the State Library of Victoria, August 2008
Searchers Working Paper 4: Accessing the State Library of Victoria through iTunes University (beyond the campus), December 2008
Searchers Working Paper 6: Online visits to large public libraries in Australia - a comparison, December 2008

Searchers Working Paper 8:  Pathways to the SLV Collection. September 2009


This report contributes to the research on current information-seeking, drawing from the analysis in Searchers Working Paper 8: Pathways to the SLV Collection and adding new empirical research on the content of search queries typed into Google and search queries that take people to Wikipedia. As well as comparing the type of use of each of these resources, it estimate the extent of use.

Methodology

Data sources

The data used in the analysis included queries typed directly into Google (Australia) and the State Library of Victoria (SLV) catalogue, and queries that that led Internet users from a search engine to the SLV website or a Wikipedia entry. Internet data was sourced from Hitwise data on Australian Internet use. Catalogue use data came from logs of use of the Main catalogue of the SLV. Table 1 summarises the data sources and timeframe for each element of the analysis.

Table 1: Data sources

<table>
<thead>
<tr>
<th>Search queries that led to SLV website</th>
<th>Queries typed into SLV catalogue</th>
<th>Queries typed into Google (Australia)</th>
<th>Search queries that led to Wikipedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hitwise</td>
<td>Voyager logs</td>
<td>Hitwise</td>
<td>Hitwise</td>
</tr>
<tr>
<td>Timeframe</td>
<td>May 2008</td>
<td>May 2008</td>
<td>April 2009</td>
</tr>
</tbody>
</table>

Hitwise measures Australian Internet use through collecting data directly from Internet Service Providers. It collects the log files of proxy cache servers covering more than one third of Australian Internet subscriptions, including homes, businesses, schools, universities, and libraries.

This study includes the long tail in its analysis. The term ‘long tail’ was brought into popular usage by Chris Anderson (2006). It is used to refer to a zipf distribution whereby a few items account for a sizable proportion of the total, and an enormous number of items (the long tail) each contribute a tiny proportion to the rest. As an example, the top one hundred search queries that brought people to Wikipedia in
April 2009 accounted for just 4% of visits via search engines; hundreds of thousands of search queries accounted for the other 96% of visits with most only accounting for one or two visits each. Analysing the long tail is laborious as it requires manual coding of large numbers of search terms.

**Selection of data for analysis**

**Search queries that led to Wikipedia**

Wikipedia use can be measured in terms of the number of visits to particular pages or in the context of people’s information searches. The number of visits to particular pages is of interest but this measure may be artificially inflated by the repeat visits of zealous contributors. In the context of information searches, the most common way that users arrive at Wikipedia is through clicking on search engine results. In the month of April, 2009, two thirds (66%) of visitors (in Australia) to Wikipedia came directly from a search engine (Source: www.hitwise.com) and almost all (93%) of these came from Google. This paper analyses the search queries that took people to Wikipedia in the four weeks ending 25 April 2009, using data provided by Hitwise (www.hitwise.com).

It is estimated that more than 600,000 search terms in the Hitwise sample took a user to Wikipedia and that at least 400,000 of these search terms were unique. The sample analysed in this study was selected from an extract of 50,000 unique search terms that took a user to Wikipedia in April 2009. The distribution of search terms was mapped and divided into seven groups based on known and estimated search term frequency, as shown in Figure 1.

![Figure 1: Long tail distribution of search queries to Wikipedia](image-url)
Using a proprietary method, Hitwise matches Internet use with Mosaic lifestyle profiles, which divide the Australian population into eleven lifestyle groups. (See http://www.mosaicaustralia.com.au/ for more information on the method and the lifestyle profiles). For each lifestyle group, a random sample of 160, stratified by search term frequency, was selected without replacement. In this way, a total sample of 1760 search terms was drawn. The sample was divided equally across the lifestyle groups to ensure the same precision for each lifestyle group.

The analysis weighted terms by search term frequency. In addition, to avoid the dominance of those search terms at the top of the long tail, search terms were also weighted according to their position in the long tail distribution; for example, terms in the sample that appeared more than 500 times were weighted to contribute a total of 1.7% of visits in the sample (see Figure 2). Reporting on the total frequencies for each subject required an additional weighting taking into account the representation of each lifestyle group in the online Australian population.

**Queries typed into Google (Australia)**

Visits to Google (Australia) accounted for 64.5% of all visits to a search engine in April 2009; (visits to google.com accounted for 17% of all visits to a search engine in that same month). The selection of the sample of Google search queries was done in exactly the same way as the selection of the Wikipedia queries, although the distribution of search queries was quite different and hence the weightings were very different. Three-quarters of queries (74.8%) were unique, that is appeared only once in the month of April 2009. The sample analysed in this element of the analysis was selected from an extract of 60,000 unique search queries that were typed into Google (Australia) in April 2009. This extract accounted for 28.7% of the total search queries entered into Google (Australia) in that month.

**Queries that led to SLV website** (Searchers Working Paper 3)

The data on search terms that resulted in a visit to the SLV website comprised all queries (in the Hitwise data) that led to the SLV website in May 2008. In May 2008, only SLV collection items that included digitized images were visible to search engines. This included manuscripts, maps, pamphlets, artworks and photographs. Catalogue records for books were theoretically ‘visible’ to search engines via the Libraries Australia website, but could not be accessed directly from search engine results.

---

1 The term ‘Google’ is subsequently used to in this report as shorthand for Google (Australia) or www.google.com.au
**Queries typed into SLV catalogue** (Searchers Working Paper 7)

The sample was drawn from transaction logs of all searches undertaken on the Main catalogue in May 2008. The main catalogue of the State Library of Victoria contains records for books, magazines, newspapers, electronic books and journals, video recordings, music, maps and oral history as well as records of several thousand websites. The system used by the State Library at the time of analysis was a Voyager (ExLibris) catalogue. By 2008 this catalogue had more than 1.3 million items. The content analysis in Working Paper 7 is based on a random stratified sample of 1,000 transaction logs from the main catalogue for the month of May 2008. The sample was restricted to members of the public using the catalogue (in other words, it excluded SLV staff searches). The sample was also restricted to the first search undertaken in a session. This reduced double counting of what is essentially part of the one search. This also meant that the sample only included search queries typed in by the user rather than those that were the result of clicking in the resulting record after an initial typed in query. Lastly, the content analysis of the search queries was restricted to those search types that were considered to be about a general topic rather than a specific item and the sample was stratified according to the prevalence of each included search type. (For a fuller description of the methodology, refer to Searchers Working Paper 7.)

While title searches and author browse search types were excluded from the content analysis in Searchers Working Paper 7, they have been added back into the results reported in this paper as queries relating to books or authors.

**Subject coding**

Each of the search queries in the final sample was manually examined and coded to indicate the subject of the query. These codes were created to most closely describe the content of the search queries. The data were combed over numerous times, sorted by code and recoded as necessary. Unless the meaning was immediately obvious, each search query was looked up in Google and/or the SLV catalogue as appropriate. Each search query from the Wikipedia sample was looked up in Google to enable inspection of the relevant Wikipedia entry in the search results.

In Searchers Working Paper 3, there was a substantial group of search terms (9%) that were recorded as leading to the SLV website, but seemed very unlikely to have resulted in search results that linked to the SLV website. These ‘unrelated’ queries were excluded from the following analysis as it is likely that these visitors to the SLV website used tabbed browsing or a bookmark to go directly to the SLV website. In addition, the analysis of unique search queries that brought people to the SLV website included 12% tagged as ‘reference’. In Working Paper 8, search terms with this tag were recoded to the subject of the reference query. The recoded data was used for the analysis in this paper.
The analysis of queries that led to Wikipedia and queries typed into Google were conducted subsequent to the analysis of catalogue queries and terms that led to the SLV website. Care was taken not to force the queries into categories that did not reflect the substance of the query and additional codes were created as required. For example, the code “ecommerce” was not needed for the analysis of library catalogue queries, but was an appropriate code for many of the Google queries.

The resulting codes closely described the content of the search queries and these were amalgamated into the following broad subject groupings. “Other” was used for those subjects that did not have enough queries to warrant their own category. Where the meaning of the term was unknown or ambiguous, the code ‘unknown” was assigned.

- **Popular Culture**
  - popular music, TV show, actor, movie, video game, celebrity, Myspace, radio, Youtube
- **Ecommerce**
  - gambling, airlines, travel, buying/selling, banking/finance, retailer, service, name of product
- **Business-related**
  - the study of business, accounting, management and organizational theory, text book titles
- **Cultural Practice**
  - sport, religion, food, jobs, learning, language, festivals, events, holidays, hobbies, other aspects of cultural practice not elsewhere classified
- **Computing/web**
  - software, email/chat, social networking sites, eg Facebook
- **Health**
  - psychology, mental health, sexual health, general health
- **History**
- **Science (including mathematics)**
- **Place/Building**
- **Contemporary Issues**
  - contemporary affairs, newspapers/news sites, government, organisations
- **Book/Author**
- **High Culture**
  - Fine Art, Classical Music, Architecture
- **Adult**
  - pornography, dating sites
- **Genealogy**
• Unknown
• Other (includes Indigenous and Person- not elsewhere classified, Weather, Time and Public transport)

In most cases, the classification was straightforward. The coding inevitably involves some subjective decisions about the appropriate category in which to place a particular term. However, the contribution of each query to the whole is small.

Google search queries were also classified according to the presumed intention behind the query. Drawing on Broder (2001), search queries were deemed to be either informational, navigational or transactional. Navigational searches are like known item searches, or a shortcut to a particular website. Transactional searches are those searches where the presumed intent of the search engine user is to undertake an Internet transaction, such as a purchase, a download or a communication. It must be noted, however, that there is a degree of judgement involved in distinguishing between these searches. In this study, queries were coded as informational unless it seemed very likely that the user wanted to get to a particular website. So for example, whereas Broder considered that brands or the names of products were navigational, in this study they were coded as informational. This is because it was considered that the user may equally be seeking information on where to buy a product or independent information about that product. Similarly, while Broder considered that the searcher typing the query ‘Don Knuth’ wants to reach this person’s academic home page, in this study, people’s names were coded to the relevant subject (for example, politics, celebrity, popular music).

Limitations

The advantage of using transaction logs for investigating the sort of information people access online is that transaction logs enable the study of a large sample of users, are unobtrusive, and do not affect user behaviour. Because transaction logs record what people actually do, they overcome the limitations associated with relying on what people say they do. The main limitation for a study like this is that the topic in which the user was interested can only be imputed by the researcher on the basis of the search query. A more certain method would be to observe users conducting searches and interview them at this time. In addition, the data used in this study does not provide information on those cases where Wikipedia or the SLV appeared near the top of the search results but were not visited or where Internet users went directly to Wikipedia or to the SLV website for information rather than via a search engine.

There are a number of other caveats to this study. The coding inevitably involves some subjective decisions about the appropriate category in which to place a particular website or term. Although in the long tail the contribution of each site or query to the whole is small, there is a margin of imprecision to the classifications. In interpreting the data, one should focus on the overall pattern, rather than the precise
size of each category. In recognition of the imprecision, the percentage size of the categories is reported to the nearest integer.

Although the Internet Service Providers that provide data to Hitwise include a representative cross-section of sizes, there may be some sample bias, the direction of which is impossible to detect.

Findings


**What are people searching for when they visit Wikipedia?**

![Figure 2: Subject of queries that took Australian search engine users to Wikipedia](image)

**Figure 2:** Subject of queries that took Australian search engine users to Wikipedia
As Figure 2 shows, more than one third (36%) of queries related to popular culture and these most commonly related to popular music. Almost all of the queries coded as popular culture were for names; the names of TV shows, actors, musicians, songs, movies or video games.

Just over one in seven (15%) of queries related to cultural practices. As can be seen from Figure 3, 5% were queries about sport, 1% were about religion, and 9% were about cultural practices – not elsewhere classified. This latter group comprised queries about aspects of everyday life, including product names, foods, holidays and ceremonies as well as words or expressions for which it seemed that the searcher was seeking a definition or some background. Examples of this latter group include the words ‘oxymoron’ and ‘dessert’. At the time of writing, the relevant Wikipedia is the first search result when either of these two queries is typed into the search engine Google.

**What do people search for on Google?**

Only half (52%) of all Google queries were informational. For almost half of the queries (48%), the searcher appeared to have a specific website in mind. Transactional searches are really a subset of navigational searches. In the case of a transactional search, the presumed intent of the search engine user is to go to a particular website to undertake an Internet transaction. These websites included online auction sites, banking sites, queries for downloads and email and social networking sites, where the transaction is communication. One third of the navigational queries (16% of all queries) were transactional.

**Figure 3: Subject of informational searches on Google (Australia)**
It can be seen from Figure 3 that popular culture queries accounted for more than one quarter (29%) of all Google searches coded as informational. These mainly related to popular music (8% of all searches) and TV (8% of all searches).

Cultural practice accounted for one fifth (19%) of all informational searches. As shown in Figure 3, these were mainly queries about sport (6% of all searches). Another 3% were queries about learning (including the names of schools and universities) or jobs, 2% were queries about food, 2% were queries about religion, events and hobbies. The remainder (6% of all searches) were magazine titles and queries about other aspects of cultural practice, such as “April fools jokes”, “girls names”, “aliens”, and “best friend quotes”.

E-commerce also accounted for one fifth of all informational searches. More than half of these (12% of all searches) were queries about products (for example, “vacuum cleaner”, or “Guess jeans”). The remainder were queries about services (for example, “broadband choice”, travel (including airline names), buying and selling (for example, “car auctions”) or banking and finance.

Of course, the distribution of subjects looked up in Google is not an indication of subjects looked up on the Internet. For some subjects, the user may be more likely to go directly to a web page.

Comparing subjects across the different information resources

The results of combining the analyses are shown in Table 2. This compares the subject of the unique terms typed into the catalogue, the terms that brought people to the SLV website, the search queries typed into Google and the search queries that led people to Wikipedia.

There are substantial differences in the spread of subjects being queried or accessed via each information resource. Reflecting the content of the SLV website and the SLV collection, three quarters (76%) of the search queries that took people to the SLV website related to places/buildings (24%), the SLV or other libraries (21%), history (17%) or books/authors//newspapers (14%). In contrast, only 7% of queries typed into Google and 16% of queries that led to Wikipedia related to any of these subjects.

As might be expected, catalogue users were more likely to be looking for books/authors/newspapers than those using other information resources. Two fifths of catalogue users (41%) were looking for books/authors/newspapers. Catalogue users were also the most likely to be looking for information on contemporary issues (13%) or high culture (8%).
Table 2: Comparison of the subject of queries on different information sources

<table>
<thead>
<tr>
<th>Subject</th>
<th>search queries that led to SLV website</th>
<th>search queries typed into catalogue</th>
<th>search queries that led to Wikipedia</th>
<th>information of search queries typed into Google Australia</th>
<th>all search queries typed into Google Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place/building</td>
<td>24%</td>
<td>7%</td>
<td>5%</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>Related to SLV or other libraries</td>
<td>21%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>History</td>
<td>17%</td>
<td>6%</td>
<td>6%</td>
<td>1%</td>
<td>-</td>
</tr>
<tr>
<td>Books/authors/newspapers</td>
<td>14%</td>
<td>41%</td>
<td>5%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Contemporary issues</td>
<td>6%</td>
<td>13%</td>
<td>5%</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>Cultural practice</td>
<td>5%</td>
<td>8%</td>
<td>14%</td>
<td>19%</td>
<td>15%</td>
</tr>
<tr>
<td>Genealogy</td>
<td>4%</td>
<td>2%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>High Culture</td>
<td>3%</td>
<td>8%</td>
<td>2%</td>
<td>1%</td>
<td>-</td>
</tr>
<tr>
<td>Person - not elsewhere classified</td>
<td>2%</td>
<td>1%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Indigenous</td>
<td>2%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Science</td>
<td>2%</td>
<td>2%</td>
<td>6%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Business-related</td>
<td>1%</td>
<td>4%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Popular Culture</td>
<td>-</td>
<td>1%</td>
<td>40%</td>
<td>29%</td>
<td>25%</td>
</tr>
<tr>
<td>eCommerce</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Computing/Web</td>
<td>-</td>
<td>1%</td>
<td>8%</td>
<td>4%</td>
<td>8%</td>
</tr>
<tr>
<td>Health</td>
<td>-</td>
<td>4%</td>
<td>6%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Unknown</td>
<td>-</td>
<td>3%</td>
<td>1%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Weather/Time/Public transport</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Adult</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8%</td>
<td>-</td>
</tr>
<tr>
<td>Other</td>
<td>-</td>
<td>-</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Only 1% of catalogue queries related to popular culture and search engine queries relating to popular culture did not take anyone to the SLV website. However, popular culture was a major subject of investigation, accounting for 29% of Google queries. Wikipedia is notable as a resource for information on popular culture. Two fifths (40%) of queries that took people to Wikipedia related to popular culture.

It can be seen from Table 2 that the main differences between the subject breakdown of informational queries on Google and all queries on Google is the absence of queries coded as ‘Adult’ in the informational queries. This is because all queries coded as ‘Adult’ were navigational or transactional.

Figure 4 summarises the information in Table 2, highlighting the main differences.

---

2 To make the table easier to read, a dash is used in place of 0%.
**Figure 4: Comparing subjects across different information resources**

It indicates that people come to the SLV (website or catalogue) mainly looking for information on books/authors/newspapers, place/buildings, the SLV or other libraries, history, contemporary issues, cultural practice and high culture. In contrast, people looking for information using Google or accessing Wikipedia are likely to be looking for information on popular culture, ecommerce, cultural practice, or computing/web. In other words, the difference is greater than the overlap in the types of information looked for at the library versus Google and Wikipedia.

**Relative use of the various information resources**

These observations about the distribution of the subjects of queries raise the question of whether people are more likely to access information on a particular subject area using a particular information resource (SLV website, Wikipedia, Google, the catalogue). The answer to this, of course, depends on the numbers of people accessing each information resource. For example, even though the proportion of people using Google to look up contemporary issues is low, it would be useful to know how the actual number compares with the number of people looking up contemporary issues on the catalogue. Ideally, one would calculate the numbers of people accessing each subject via each information resource. In theory, one could calculate, for each subject, the probability that someone would turn to the catalogue, Wikipedia or Google to find information. However, this would only be possible if 1) we had access to accurate estimates of the numbers of people using each resource and 2) the subject coding was accurate to several decimal points. As the following shows, neither of these conditions hold:
1) Hitwise covers approximately 40% of all Internet activity in Australia but does not report any absolute numbers. SLV’s own figures on the number of visits to the SLV website vary by up to a factor of 10 depending on whether the source is HBX or WebTrends.

2) The subject coding indicates broad trends not exact sizes. Tiny variations in percentages would translate to large but spurious differences if the percentages were extrapolated to estimates of the numbers of Google and Wikipedia users.

Although it is not possible to compare absolute numbers, it is still possible to get a rough idea of the magnitude of the differences between the number of visitors to the SLV website, users of the catalogue, visitors to Wikipedia and Google users.

There are two ways of doing this. Firstly, it is possible to look at the factor by which visits to Google and Wikipedia are greater than visits to SLV.

![Graph showing relative number of visits to Google (Aust), Wikipedia and SLV]

**Figure 5: Relative number of visits to Google (Aust), Wikipedia and SLV**

Figure 5 plots Hitwise data on the number of visitors to Google divided by the number of visits to the SLV website and similarly for Wikipedia. It can be seen that between May 2008 and April 2009, Google had between two thousand and six thousand times as many users as visitors to the SLV website. In the same period, Wikipedia had approximately 200 times as many visitors. These figures need to be adjusted in order to make a more legitimate comparison. While Google and

---

This is due to differences in the methodologies of WebTrends and HBX analytics.
Wikipedia users are all over Australia, roughly one quarter of visitors to Google and Wikipedia are from Victoria. In April, 2009, 25.5% of visits to Google.com.au and 24.3% of visits to Wikipedia were from Victoria. In the same period, almost two-thirds (64%) of SLV website visitors were from Victoria. Adjusting for this, and the fact that only half (52%) of Google queries are informational, we can very roughly estimate that in April 2009, Wikipedia had in the order of 80 times as many Victorian visits over the month as the SLV. Google had in the order of 800 times as many Victorian visits over the month as SLV. Figure 6 depicts these differences visually. The size of each circle roughly approximates the relative size of the number of queries.

![Diagram](image)

**Figure 6:** Approximate relative sizes of the number of queries using each information resource

In looking at Figure 6, it should be remembered that two thirds (66%) of visitors (in Australia) come to Wikipedia directly from a search engine. Hence the circle representing the total number of visitors to Wikipedia would be half as big again as the circle at D. Similarly, the circle representing the total number of visitors to both google.com and google.com.au would be one quarter as big again as the circle at B.

---

4 The number of catalogue searches is estimated as roughly double the number of web searches that led to the SLV collection. (Source Searchers Working Paper 8).
Another way to get a sense of the relative numbers using different information sources is to look at the Hitwise industry rankings. Hitwise separately identifies reference websites and 4375 of these websites were visited in April 2009. These include Internet reference sites (for example, wisegeek, wikihow, and coolmath) as well the websites of museums, libraries and research institutions. Visits to Wikipedia accounted for one in three (34.3%) of all visits to these reference websites. Visits to the SLV website accounted for 0.087% of all visits to these reference websites. It is not possible to identify which of these reference websites were library websites. Three hundred and sixty five of these websites contained the word ‘library’ or ‘libraries’. However, as there are many library websites which don’t have the word ‘library’ in their title (for example, “Picture Australia” and “Inside a Dog”), it is not possible to identify what proportion of the visits to reference websites were to library sites.

Concluding Comments

The analysis has compared the subjects of catalogue and Google queries and the queries of those accessing information on the SLV website or Wikipedia. This has been one of the first studies to make such comparisons.

The research reported on in this paper shows that the distribution of subjects that people access on Wikipedia, look up on Google, access on the SLV website and look up on the SLV catalogue are quite different from each other. For example, two-fifths of the queries that led to Wikipedia related to popular culture and half of the informational queries on Google related to popular culture or ecommerce. In contrast, two fifths of people using the catalogue were looking for particular books, authors or newspapers. Another one fifth were looking for material related to high culture or specific contemporary issues. The queries that brought web users to the SLV website were most commonly about places or buildings, related to the other SLV or other libraries or history (62% of all query terms). This distribution could change as the SLV catalogue becomes visible to search engines.

As exploratory research, this report has indicated the need for more research on where people go for information on specific subjects, including visits to specific web pages. Further research is also needed to find out whether particular types of people choose among information resources according to the subject of their query or whether particular types of people are more likely to favour particular information resources regardless of the subject of their query. The fact that popular culture queries accounted for such a substantial proportion of Google and Wikipedia queries and almost no SLV catalogue queries indicates that people do turn to different information resources for different subjects.

---

5 This is after removing the website of the Bureau of Meteorology which is classified by Hitwise as a Reference website and which accounted for 19.2% of visits to Reference websites.
Concern has been expressed within the library sector that library reference services and collections are no longer needed because of Google and Wikipedia. This research shows clearly that the number of people using Google and Wikipedia far exceeds the number using the SLV website or catalogue users. The research suggests, however, that the SLV is still an important information resource for particular subjects. The number of searches undertaken in the SLV library catalogue has been increasing in recent years (Searchers Working Paper 8), as has the number of visits to the SLV website. This suggests that the SLV is providing a service to increasing numbers of people. At the same time, the ease of searching the Internet has dramatically increased people’s search activity and hence, the library’s share in search activity has fallen dramatically. This, in itself, need not be a cause for concern for the SLV. What should be of concern to the SLV is that use does not decrease and that the content of catalogue search queries match the content of the collection. It is also important to try to ensure that the SLV website is highly ranked in the results for search engine queries relevant to the SLV or its collection.

The SLV should continue to monitor the subject of catalogue queries and search terms that bring people to the SLV website. If sections of the collection are not being used, it is important to know whether this is because there is little interest in these sections or because people do not think of the SLV as a source of information on these topics. Perhaps promotion of the SLV catalogue and collections could focus on the strengths of the collections; for example history, places, buildings, and high culture. If there are other areas of the collection that are not being accessed, then promotion of the SLV catalogue and collections could emphasise these aspects of the SLV collections.

---

6 Source: Unpublished HBX and Webtrends data.