Waller, V. (2009). What do the public search for on the catalogue of the State Library of Victoria?


Copyright © 2009 Vivienne Waller.

This is the author's version of the work. It is posted here with the permission of the publisher for your personal use. No further distribution is permitted. If your library has a subscription to this journal, you may also be able to access the published version via the library catalogue.

WHAT DO THE PUBLIC SEARCH FOR ON THE CATALOGUE OF THE STATE LIBRARY OF VICTORIA?

Abstract
This study examines what the public search for in the catalogue of the State Library of Victoria (SLV). As well as indicating the type of content being accessed, this gives an indication of what catalogue users expect of the State Library collection. A content analysis was undertaken of a random, stratified sample of 4,000 search queries typed in to the SLV catalogue by the public between 2005 and 2008. This study goes some way to providing empirical evidence that the library catalogue is used to research a different range of topics to those searched via the Internet.

Vivienne Waller
Institute of Social Research
Swinburne University of Technology, Hawthorn Vic 3122.
Email: VWaller@groupwise.swin.edu.au

Introduction
There is much research to show that people start their information searches with a search engine, and in particular with Google1. It is well-established that many people find Internet search engines easier to use than library catalogues and there are predictions that Web 2.0 technologies will make the library catalogue irrelevant2. Even so, use of the catalogue at the State Library of Victoria (SLV) appears to have been steadily increasing over the last few years. As well as being used to look for material on a topic, the library catalogue is also used to locate specific items. With libraries like the SLV where only one tenth of the library’s physical collection is on display, the catalogue is the only way of gaining access to most of the items in the physical collection.

Although there have been numerous quantitative studies of the use of OPACs (Online Public Access Catalogue), these studies tend to focus on search strategies and failure rates3. There is little, if any, quantitative research on the subject of what people are actually looking up on the OPAC. More is known about the type of information people look for using Internet search engines. For example, the ten most frequent Internet search queries are usually for particular web sites that are popular at the time and two thirds of internet users in Britain look for health information online4. Jansen and Spink5 have conducted various studies on the content of search engine queries by web users in Europe and the United States. In a random sample of queries from search engine AlltheWeb.com, two in five queries were classified as relating to ‘people, places or things’, one in eight were related to commerce, travel, employment, or economy, one in six were related to computers or internet and one in ten were
related to sex and pornography. A study of US queries sent in 2001 to Excite classified one in five as relating to ‘people, places or things’, one in four as related to commerce, travel, employment, or economy, one in ten as related to computers or internet and one in eleven as related to sex and pornography. Little is known, however, about the range of topics that people actually look up on a library catalogue. It is possible that people turn to the library catalogue when they cannot find what they are after on the Internet. Alternatively, there may be a group of people who turn to the library catalogue rather than the Internet. The hypothesis of this paper is that it is likely that the library catalogue is used to research particular subjects, either in addition to, or in preference to an Internet search engine. As part of understanding people’s online information seeking, it is necessary to understand what it is that people are searching for on the library catalogue. This paper contributes to filling this gap in the current research on information-seeking through analysing what people search for in the Main catalogue of the State Library of Victoria.

Thus the report comprises an analysis of the content of a sample of search queries undertaken between 2005 and 2008. It is based on an analysis completed in September-November 2008 on logs of searches on the catalogues of the State Library of Victoria. This analysis was conducted as part of a larger project that examines the strategic challenges for public libraries presented by the online information environment.

The article concludes by noting some broad differences between web search queries and catalogue queries. Despite limitations that frustrate a direct comparison, the study reported here goes some way to providing empirical evidence that the library catalogue is used to research a different range of topics to the Internet.

**Methodology**

The State Library of Victoria, Australia is the major reference and research library in Victoria, responsible for collecting and preserving Victoria's documentary heritage and making it available through a range of services and programs. The main catalogue 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 contained more than 1.3 million items. Logs of searches undertaken in August 2005, August 2006, May 2007 and May 2008 were extracted from the main catalogue.

The sample was restricted to members of the public using the catalogue (in other words, it excluded SLV staff searches). The IP address of the computer used to conduct the search enabled identification of whether the user was using a public computer or laptop in the State Library, or using a computer outside the State Library,
for example from home or work or another library.

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.

The sample was also restricted to those search types that were considered to be about a general topic rather than a specific item. Table 1 shows the original categorization that was used to decide whether the user was likely to be looking in the catalogue for a specific item or researching a general topic. Before analysis, this categorization was used to restrict the sample to those search types that were considered to be likely to be about a general topic rather than a specific item. (The limitations of this classification are made clear in the next section which shows the proportion of each type of query which were in fact for a specific item).

Table 1: Original classification of types of search as specific or general

<table>
<thead>
<tr>
<th>General topic</th>
<th>Specific item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boolean Search</td>
<td>Author/Author Browse</td>
</tr>
<tr>
<td>By and/or about a Person</td>
<td>Left Anchored Title (inc Journal title)</td>
</tr>
<tr>
<td>Call Number Browse</td>
<td></td>
</tr>
<tr>
<td>Construct a Search</td>
<td></td>
</tr>
<tr>
<td>Keyword Relevance Search</td>
<td></td>
</tr>
<tr>
<td>Simple Search</td>
<td></td>
</tr>
<tr>
<td>Subject Browse</td>
<td></td>
</tr>
</tbody>
</table>

Hence Author and Title searches were excluded from the sample. Call number browse searches were also excluded because of the small numbers involved and the difficulty of classifying these searches. It should be noted that strictly speaking, a searcher using Author Browse, may not be looking for a specific item. They may just want to see what else a particular author has written, or whether the library has any items written by that author. This is not the same, however, as searching for general information on a topic.

Random stratified samples of 1000 were drawn for each year. These samples were stratified according to the prevalence of each included search type (as the first search undertaken in a session). Table 2 shows the distribution of included search types in the extract of the catalogue used as the sampling pool.

Table 2: Distribution of included search types in catalogue extracts.

<table>
<thead>
<tr>
<th>Year of catalogue extract</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>By and/or about a Person</td>
<td>1.5%</td>
<td>1.4%</td>
<td>1.3%</td>
<td>1.2%</td>
</tr>
</tbody>
</table>
Table 3 shows the corresponding number of records of each search type selected randomly so as to match the distribution of search types in the catalogue extracts.

**Table 3: Size of random samples, stratified by search type.**

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>By and/or about a Person</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Boolean Search</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Construct a search</td>
<td>94</td>
<td>94</td>
<td>74</td>
<td>87</td>
</tr>
<tr>
<td>Keyword Relevance Search</td>
<td>798</td>
<td>799</td>
<td>762</td>
<td>726</td>
</tr>
<tr>
<td>Simple Search</td>
<td>9</td>
<td>7</td>
<td>77</td>
<td>102</td>
</tr>
<tr>
<td>Subject Browse</td>
<td>70</td>
<td>76</td>
<td>63</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

The advantage of transaction logs for investigating what catalogue users search for is that they enable the study of a large sample of users, are unobtrusive, and do not affect user behaviour. The main limitation for a study like this is that the item which the user was looking for or 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.

Each of the 4,000 search queries in the final sample was manually examined and coded. No existing subject classification scheme was applied to the data as this carried the risk of forcing the queries into categories that did not reflect the substance of the query. It would also be quite possible that unexpected types of query would be hidden. A more suitable approach for this type of exploratory research was to draw from the grounded theory technique of open coding. This entailed generating a classification scheme from close examination of the data and the creation of codes that most closely described the content of the search queries. The data were combed over numerous times, sorted by code and recoded as necessary. It should be noted that this approach of generating the coding scheme iteratively from the data gives the final scheme high validity but does not mean that the final scheme is the only valid way of describing the data. A different researcher could have chosen to group the queries in different but equally valid ways; this is the nature of any attempt at descriptive classification.
Unless the meaning was immediately obvious, each search query was looked up in Google and/or the SLV catalogue. In this way, each search query received one or more tags which were progressively refined. The first thousand search queries were coded with a large range of codes, many of which were then discarded because there were too few to keep them as separate codes. In the end fifty-two different subject tags were retained and each search query was given one of these subject tags to indicate the subject of the search.

The subject of SLV catalogue queries will be compared with the subject of Wikipedia queries in a subsequent analysis. This means that while some of the fifty-two subject tags are used rarely in this analysis, it was important that they be retained as separate subject tags to facilitate comparison with search queries leading to Wikipedia. The subject tag “celebrity” is a good example of a tag rarely used in the SLV catalogue, but commonly used to code searches leading to Wikipedia.

The fifty-two different subject tags were amalgamated into twelve separate subject groupings, with those that could not be classified tagged as “unknown”. Several secondary tags enabled identification of whether the search query was about a general topic or a specific item and whether this topic or item related to Australia or Victoria, or related specifically to Aboriginal Australia.

The primary subject groupings are more of the nature of tags than categories as they do not fit into an overall conceptual scheme. Of course such coding inevitably involves some subjective decisions about the appropriate subject grouping in which to place a particular term. The discussion of the results includes descriptions of each category to enable application of the coding scheme to other data. Although the contribution of each 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.

**Results**

**Reseaching a general topic or searching for a specific item?**

As outlined in the methodology section, each search query can be broadly classified as being one of two kinds. A catalogue user may be looking for a specific item and want to find out whether it is part of the SLV collection or, alternatively, may be looking for material relating to a topic of interest. In practice, a catalogue user may well be looking for both specific items and material relating to a particular topic or several topics. Each search query, however, will be formulated to retrieve either a specific item or material related to a topic of interest.
Although, before the analysis began, the search types included for analysis had all been classed as general topic searches, it was found during the course of analysing the content of the search query that one fifth (21%) of the searches examined were for a specific item. Most of these searches were for book titles. In addition, there were searches for journal titles, newspaper titles, film titles and references to specific items in the collection, such as ISBN numbers, map numbers or pictures. For example, a keyword search for ‘ma000154’ yields an historical plan of Deepdene estate and ‘Favourite seaside resorts’ is the title of a picture in the SLV collection, the image of which has been digitised and is accessible via the catalogue. All of these types of search query were tagged as specific items.

Before conducting the content analysis, applying the classification in Table 1 to the transaction logs for the period under study (113,556 records), yielded the result that 57.6% of searches in the Main catalogue were categorised as general searches and 42.4% of searches were for specific items. Using the information gathered from the content analysis to revise these figures, at least half (more than 55%) of all catalogue searches were for specific items.

Figure 1 shows how the proportion of general to specific searches varied according to different types of search. Almost one third of those using command search or construct a search were looking for a specific item, compared to one fifth of those using keyword search. As might be expected, those undertaking a subject browse were least likely to be looking for a specific item (5%).

![Figure 1: Specific or general searches by search type](image)

It should be noted that the proportion of searches identified as being for specific items is likely to be underestimated. This is because only those titles that are fairly clearly recognisable as titles are included. For example, even though there exists a book...
called ‘Michelangelo’, the search term ‘Michelangelo’ would be coded as Art but not as a book title, whereas the search term ”Michelangelo and The Pope's Ceiling” would be coded as Art and a book title. All phrases were looked up on Google and/or the SLV catalogue to see if they were the title of a book or journal. For example, 'annihilating difference' is the title of a book on genocide and hence is tagged as a book title, as well as contemporary affairs and also as political. The search term “Cereal food, baking mix” is the title of a report into manufacturing in Australia and so was tagged as a book title as well as contemporary affairs.

Figure 2 suggests that the library catalogue is more likely to be used as a resource for researching a general topic by those inside the library. (84% of those inside the library were researching a general topic compared to 75% of those outside the library. As Figure 2 shows, those outside the library were more likely to be looking for a book title or another specific item (such as a map) than those using the catalogue inside the library.

![Figure 2: Type of search query by location of user](image)

**Subject of search query**

Figure 3 shows the incidence of the final broad subject groupings of the search queries analysed. It should be noted that no exact search query appeared more than once in the sample.
Almost one in five queries were about contemporary issues. The most common search queries were for material relating to contemporary issues, books/authors/newspapers, art/architecture/music, cultural practice, history and place/building. Each of these subjects accounted for more than one tenth of queries.

Figure 4 repeats the data in Figure 3 but indicates what proportion of search queries in each subject category were for specific items, and what proportion were general searches.
Figure 4: General topic/specific item according to broad subject of search query

Queries about books/authors/newspapers and queries about contemporary issues were most likely to be searches targeting specific items. Queries about places, buildings and persons (not elsewhere classified) were least likely to be searches for specific items.

The following tables go into more detail about each broad subject category. Examples are given to give a better indication of the contents of each broad subject category.
Queries tagged as ‘contemporary affairs’ were predominantly queries about social and political issues, including international issues and issues relating to social welfare. Queries about social and political issues framed in everyday language were tagged as ‘contemporary affairs’, while those framed in more academic language were tagged as ‘policy, social theory, philosophy’. For example “issues in world politics” was tagged with “contemporary affairs” whereas “international relations” was tagged with “policy, social theory, philosophy”. A similar approach was taken with book titles addressing social and political issues; those written for a more general audience were tagged as ‘contemporary affairs’ whereas more academic titles were tagged as “policy, social theory, philosophy”. It should be noted that a high degree of subjective judgment was involved in deciding which of these tags to use. The tag ‘policy, social theory, philosophy’ was also used for theoretical concepts typed in as search queries; for example, “post-colonialism”. This tag was also used for authors who are, perhaps, better known for their theories or philosophy than their books, and who tend to have books written about them; for example “Wittgenstein”.

There was a sufficient number of queries about environmental issues to tag these separately. Examples of this type of query are “Kyoto development” and “water problem”. Most of the queries tagged with “law” contained the word “law” or “legal” (for example “Australian Taxation Law”). Some referred to a legal concept; for example “equitable remedies”. Queries about organizations included queries about government bodies (for example, “vicroads”, non-government and community organisations (for example “Community Radio Federation” and “Confectionery manufacturers association” and a few businesses (for example “fantastic holdings” and “kmart”).

Figure 5: Search queries relating to contemporary issues
Books, authors, newspapers (about 12% of all queries)

Figure 6: Search queries relating to books, authors and newspapers

Figure 6 shows that just over half of the queries in this category were searches for particular authors. It is not possible to gauge the catalogue user’s intent with this sort of search query. Users may have been looking for a specific book by an author, information about an author or doing a search to see what books that author had written or which of these were held by the SLV. Some users may have been looking for biographies of the author. Only two of these search queries were clearly for a book title about a particular author.

Queries tagged with “general literature” include queries such as ‘juvenile literature’, ‘book reviews’, and book titles about writing or literature. Searches for fiction included a mixture of adult and children’s fiction. The “reference tag” was used for dictionaries, directories, and encyclopedia. Less than 2% of search queries were for newspapers.
Art, architecture and music (about 12% of all queries)

Figure 7: Search queries relating to art, architecture and music

Figure 7 shows that just under 10% of all search queries were related to art or architecture. The tag ‘Art and architecture’ was used for search queries related to artists, photographers and architects, as well as search queries related to art theory or specific art works.

The tag ‘Music – classical’ was used to identify queries about music that were not about popular music. Hence, as well as queries about classical music and composers, it includes queries about jazz (for example “Gershwin songbook” and the music of particular ethnic groups, for example “Armenian music”). Only 14% of searches with this tag were for specific items and just 1% of searches with this tag related to indigenous works.
Cultural practice (about 12% of all queries)

Figure 8: Search queries relating to cultural practice

Just over 12% of queries related to cultural practices. Most of these tags are fairly self-explanatory. The tag ‘cultural practice – not elsewhere classified’ was used for non-technical queries about a variety of aspects of culture. This included queries about aspects of everyday Australian life, including product names, hobbies, and travel as well as cultural practices associated with specific ethnicities, for example ‘chinese home altars’. The tag ‘religion’ was used for queries about organized religion as well as more general queries about spirituality (for example, ‘shamanism’). Lifestyle journals such as ‘Vogue’ and special interest journals such as ‘Beer and Brewer’ were given the tag of ‘cultural practice – not elsewhere classified’. The tag ‘Learning English’ was used for queries relating to English exams and ESL materials. The tag ‘Learning - other’ was used for general queries about formal learning, such as “study tips” and “essay”. Queries for text books were tagged with the subject of the text book and so do not appear in this category. Less than 1% of queries were about movies, TV shows, actors or celebrities.
**History (about 10% of all queries)**

![Diagram showing proportion of search queries relating to history]

**Figure 9: Search queries relating to history**

Queries about historical figures (people who could not be considered contemporary) are included in this category. The ‘war-related’ tag was used for queries about military history as well as terms such as ‘nurses and history and war’. Queries about Australian history (including Victorian history) only slightly outnumbered queries about International history (countries other than Australia). The tag ‘History- general’ was used for queries that were not related to a particular country, for example ‘Great Depression’ and ‘Archaeology’.

**Places/buildings (about 10% of all queries)**

![Diagram showing proportion of search queries relating to places/buildings]

**Figure 10: Search queries relating to place or buildings**
Queries about buildings included specific buildings, such as “St John’s church Launceston’ and buildings in general, such as ‘shot tower’. More than half of the queries tagged as maps were queries for specific MMBW maps in the SLV collection. Queries were twice as likely to be for an Australian place or building than one overseas.

Business-related (about 5% of all queries)

The tag ‘Business-related’ was used for queries relating to the study of business, including queries about accounting, management and organizational theory. Many of the queries relating to particular books were for text books.

Science/computing/web (about 5% of all queries)

![Figure 11: Search queries relating to science, computing and the web](image)

Most of these queries related to science, with less than 1% of queries relating to ICT. Queries tagged as ‘ICT’ included queries about particular software programs or programming languages as well as general queries such as ‘computer’. The tag ‘Web’ was used for only a handful of queries, with examples being ‘google’, and ‘wireless security’
**Health (less than 5% of all queries)**

![Chart showing proportions of health-related queries](chart)

Figure 12: Search queries relating to health

Queries relating to Psychology, mental health and sexual health were separately identified. The tag ‘Psychology’ was used for queries about psychological concepts such as ‘multiple intelligences’ as well as queries about communication skills. There were very few queries relating to mental health and only two queries relating to sexual health. The tag ‘Other health’ was used for all other health queries. Some of these related to specific health conditions, such as ‘unstable angina’; others were more general, such as ‘nursing care plans’. The tag ‘Other health’ was also used for queries about alternative health practices, such as ‘ayurvedic’ and ‘kinesiology’.

**Person – not elsewhere classified (less than 3% of queries)**

Some of the search queries tagged with “Person – nec” were fairly common surnames, (for example, “Holmes”) so that one could not be confident whether the searcher was looking for an author with that surname, an historical figure or some other type of person. Some of the names were complete names that did not show up in either the catalogue or Google. These were possibly genealogy-related searches.

**Genealogy (less than 2% of queries)**

Searches for which this tag was used included searches that specifically mentioned ‘genealogy’ and search queries for records of burials, deaths or births. It is likely that many of the search queries tagged with ‘person-not elsewhere classified’, ‘place’ and ‘history’ would also have been genealogical searches; however, it is not possible to identify them as such.

**Unknown (less than 5% of queries)**

It was possible to tag more than 95% of queries with a primary subject tag. Those tagged as ‘unknown’ include those search queries that have multiple meanings, so that it is impossible to know what the person was looking for. For example, the search query ‘Hammerton’ was tagged as ‘unknown’ as Hammerton is the surname of a
historical figure, the surname of a writer, and a parish in Yorkshire.

**Aboriginal Australia**

Just 1% of searches were identified as relating to Aboriginal Australia.

**Queries relating to Australia or Victoria**

![Diagram showing the proportion of search queries in sample](image)

Figure 13: Search queries relating to Australia or Victoria

The tags identifying queries about newspapers, history, buildings and places were refined by tags identifying whether the query related to Victoria or Australia or another country. Figure 13 indicates the number of queries relating to Victoria, the number relating to Australia and the number of queries relating to other countries.

Most queries about buildings were about buildings in Victoria, with no queries about buildings in other parts of Australia. Almost all queries about maps were for locations in Victoria and queries about places in Victoria were more common than queries about places in Australia or overseas. Only a tiny proportion of inquiries were about newspapers. The newspapers are all located in the Newspaper reading room. A user can find their location by browsing or asking a member of staff in the reading room. The Australian newspapers are particularly easy to find. Catalogue queries about newspapers were mainly about International newspapers or Australian (not Victorian) newspapers. As Figure 2 showed, there were equal proportions of catalogue queries about newspapers from users inside the library and users outside the library.

Catalogue queries about history were less likely to be about Victorian history and more likely to be about the history of another country. This is not surprising as the
catalogue is the main point of entry to items in the collection that relate to international history. Users can access materials about Australian and Victorian history via the Australiana index and through visiting the La Trobe Reading room. The Australiana Index has been compiled by Library staff over a number of years. This online index contains references to published material, including newspaper articles about particular topics. It also contains references to unpublished files, such as: Local History; Biography, Subject; Family history letters; and Shipping letters files. These files contain items such as answers to previous enquiries and ephemeral material such as leaflets and pamphlets. The La Trobe reading room contains books, journals and reference materials about Australia and specialist staff at an information desk.

Changes over time

As Figure 14 shows, there was little difference in the subject of search queries across the years 2005-2008. The largest difference was in 2006, where there was a small spike in queries relating to place or buildings and a decrease in queries relating to Art, architecture or music.
As Figure 15 indicates, there were some differences according to the location of the user. Those using the catalogue within the library were slightly more likely to type in business-related search queries and search queries relating to cultural practice. Those using the catalogue outside the library were more likely to conduct search queries relating to books/authors/newspapers.

Comparing catalogue queries and web search queries

It is not possible to make a direct comparison between web search queries and catalogue queries for several reasons. Insufficient detail is available about the construction of existing classifications of web search queries\(^9\) to enable a mapping to the classifications in this paper. In addition, the interpretation of any analysis of web search queries is complicated by the fact that web queries are not only for known items or for information on a topic. Web queries can be conducted for the purpose of conducting a transaction, such as making an online purchase or downloading a particular program. For example, a search for a particular song title may be to find information about the song or may be to enable purchase of the song online. Estimates of the proportion of web queries which are transactional vary from one in ten\(^10\) to one in three\(^11\). Another factor confounding attempts at direct comparison with existing studies of Internet search queries is that these studies have shown marked differences across countries and across time\(^12\). There appears to be no recent analysis of the content of web queries by Australian users.

Despite these limitations, the following comparisons lend support to the hypothesis that the types of subjects researched using the library catalogue are likely to be different from those researched using the Internet.
While it is known that the internet is a major source of information for health information (two thirds of internet users in Britain look for health information online\textsuperscript{13}), less than 5\% of catalogue queries related to health. Less than 2\% of catalogue queries were related to computers or the Internet compared to 16\% of queries on AlltheWeb.com and 9.7\% of queries on Excite\textsuperscript{14}. In the same study of search engine use, Jansen and Spink\textsuperscript{15} found that sex or pornography was the subject of almost one in ten search engine queries. In stark contrast, there was only one catalogue query relating to sex and none relating to pornography.

As would be expected, it seems that catalogue users are much more likely to be looking for a specific item than internet search engine users. More than half of all catalogue searches were for specific items, compared to 10\% of web search queries estimated to be navigational\textsuperscript{16}.

Unfortunately, the classification scheme used in Jansen and Spink\textsuperscript{17} does not allow for any further comparisons of the subject of catalogue queries with the subject of web search queries. A planned analysis of the type of queries most commonly typed into search engines by Australian users will allow a more detailed comparison with this study and perhaps enable some conclusions to be drawn about whether there are subjects which are commonly researched on the catalogue in preference to the Internet.

**Conclusions**

The analysis in this paper gives an indication of the type of content being accessed as well as an indication of what catalogue users expect of the State Library collection. The data indicated that the library catalogue is still used to look for material on topics with up to 45\% of all searches for information on a general topic rather than a specific item. As would be expected, it seems that catalogue users are much more likely to be looking for a specific item than internet search engine users, where only 10\% of queries have been estimated to be navigational.

The data showed that the library catalogue is more likely to be used as a resource for researching a general topic by those inside the library than offsite users. However, it seems that the distribution of topics explored on the library catalogue is quite different from those explored on the Internet. No search query in the catalogue appeared more than once in the sample. This suggests a different distribution to Internet search queries. Search queries typed into the Internet follow a zipf or ‘long tail’ distribution where a small number of queries are typed in numerous times and a large number or queries appear only once or twice.

Although, as this paper has explained, no direct comparison is currently possible, the results of this study lend support to the hypothesis that the types of subjects
researched using the library catalogue are likely to be different from those researched using the Internet. Material relating to contemporary issues was most commonly searched for on the catalogue, accounting for almost one in five queries. Queries about books/authors/newspapers, art/architecture/music, cultural practice, history and place/building were also common, each of these subjects accounting for more than one tenth of queries. However, people rarely used the catalogue to search for material relating to health or computing and didn’t use the catalogue to search for material relating to sex or pornography. Each of these topics have been found to be regularly researched on the Internet.

It is possible that users’ search queries are limited to what they think the collection of the library is likely to hold. The study reported on here has been of queries typed into the Main catalogue of the SLV and has excluded queries typed into the Pictures Catalogue, the Manuscripts catalogue and subscription databases. In the second half of 2009, the SLV is implementing a next generation catalogue (Primo by Ex Libris). This next generation catalogue has federated search functionality and so will enable the user to elect to search across multiple electronic resources. This will enable comparison of the subject of queries made across all electronic resources and the subject of queries limited to a particular subset of the library collection.

Although exploratory, the analysis contained in this report contributes to building a more comprehensive picture of where people search for particular types of information. Future analyses will include an analysis of the type of queries most commonly typed into search engines by Australian users, and an analysis of the type of queries that bring people to Wikipedia. On completion, the findings of these separate analyses will be integrated to present a more complete picture of online information-seeking.

An interesting further topic for investigation would be the relationship between the types of content being accessed and their representation in SLV collections. Such a comparison would provide a useful additional indication of how intensively different parts of the collection are used, and whether some could benefit from further development or expansion.

**Acknowledgements**
The author wishes to thank the many staff at the State Library of Victoria for their assistance and suggestions, in particular Sue Hamilton, Ian Patterson, Janice van de Velde, Anne Beaumont, Kirstie McRobert, Leneve Jamieson and Patrick Gregory. This research was funded by Australian Research Council grant LP077215 and the author is also indebted to Julian Thomas, Denise Meredyth and Ian McShane of the Institute for Social Research, Swinburne University of Technology.
Notes:

1 K Markey 'The Online Library Catalog: Paradise Lost and Paradise Regained' D-Lib Magazine 2007 vol 13 no 1/2.
12 Jansen & Spink op cit p 260
13 Dutton & Helsper op cit p 67
14 Jansen & Spink op cit pp 258-259
15 ibid p258
16 Jansen Booth & Spink op cit p 1251
17 Jansen & Spink op cit p 258