

Searching where for what: A comparison of use of the library catalogue, Google and Wikipedia

Vivienne Waller

Abstract

There is a widespread concern amongst librarians that Google Search and Wikipedia are making library reference services and even library collections redundant. There is however, little research on the types of subjects that people look up on Google, Wikipedia and library catalogues. This exploratory analysis draws together the results of transaction log analyses of the subjects of Google and library catalogue search queries and the subject of search engine queries that took users to Wikipedia. As well as comparing the subject of search queries, it estimates the extent of use of each information resource. Hence this paper contributes to research on current information-seeking and the role of public libraries in online information provision.

1 Introduction: the public library versus the Internet?

There is a widespread concern amongst librarians that Google Search and Wikipedia are making library reference services and even library collections redundant (Abram, 2005). A question frequently asked by those responsible for funding library services has been whether libraries are still needed now that we have the Internet (D'Elia, Jorgensen, Woelfel and Rodger, 2002; Webster, 2006). A widely circulated set of 15 "provocative statements" about university libraries claimed that before 2011 "all information discovery will begin at Google, including discovery of library resources" and that "there will be no more librarians as we know them" (Taiga Forum Steering Committee, 2006). According to Lorcan Dempsey, Chief Strategist at OCLC, Google offers users 'a sufficiently encompassing experience that they need go no further, or do not believe that they need to go further' (Dempsey, 2006). Search engines have been described as "the new interface to information for almost anyone under the age of 35" (Arnold, 2007) and, indeed, empirical data demonstrates the high prevalence of search engine use. In the early 2000s the number of searches on Google already far exceeded the number of searches on a library's catalogue (Markey, 2007). A US

Author

Vivienne Waller is a Lecturer in Sociology at Swinburne University of Technology, Australia.

Email: ywaller@swin.edu.au

Received 30 April 2011

Accepted 29 July 2011

study undertaken in 2007 found that just under half (49%) of Internet users used a search engine every day (Fallows, 2008). Since then, the number of searches undertaken in the US using the major search engines has increased by more than 50%. In the month of April 2010, Americans conducted 15.5 billion searches on major search engines compared to 9.7 billion searches in August 2007 (comScore, 2007, 2010). In Australia in 2009, the Internet was considered to be an 'important' or 'very important' source of information by almost three quarters (73%) of Internet users (Ewing and Thomas, 2010). In the UK in 2009, two thirds of British people (65%) use the Internet first when looking for information relating to a professional, school or personal project and two thirds (64%) of Internet users mainly use a search engine to look for information on the Internet (Dutton, Helsper and Gerber, 2009).

As Tenopir (2003) points out, many librarians consider Google to be their competition. A cursory look at overall comparisons of search engine use and library use suggest that libraries may be no match for the search engine giant, Google. For example, a 2005 OCLC study (2005) found that only 2% of college and university students begin searching for information at a library web site. Griffiths and Brophy (2005) found that more than two-thirds (68%) of students turned to a search engine first compared to 10% using the library catalogue first.

More recently, Wikipedia, the online collaborative encyclopedia has grown to become another force to rival library reference and collection services. Launched in 2001 with one article (<http://en.wikipedia.org/wiki/Wikipedia:About>), as of February 2010, Wikipedia contained more than 3.2 million articles in English and more than 100,000 articles in each of 31 languages. (Source: www.wikipedia.org accessed 1 March 2010). In America, just over 42% of adults use Wikipedia, with one in six (17%) Internet users using Wikipedia on any day (Zickuhr and Rainie, 2011). In Australia, in April 2009, Wikipedia was the twelfth most visited site on the Internet (Source: Hitwise Australia, www.hitwise.com). In a US study by Lim (2009), one third of university students reported that they use Wikipedia for academic purposes, generally for finding background information.

1.1 Lack of empirical data

Despite the concern amongst librarians that Wikipedia and Google will make library's reference services and collection services redundant, there is a lack of empirical data on what people look up on these different information resources.

There exists a body of research that looks at search strategies and compares users' formulation of queries when using search engines or an OPAC (Griffiths and Brophy, 2005; Jansen and Pooch, 2001; Navarro-Prieto, Scaife and Rogers, 1999). There are also a range of studies looking at the type of queries people type into search engines (Jansen, Booth and Spink, 2008; Jansen and Spink, 2005; Park, 2009; Rose and Levinson, 2004; Ross and Wolfram, 2000; Segev and Ahituv, 2010; Spink, Wolfram, Jansen and Saracevic, 2001; Yu and Young, 2004). However, studies examining the use of library OPACs or Wikipedia tend to be restricted to students (for example Campbell and Fast, 2004; Choo, Detlor and Turnbull, 2000; Griffiths and Brophy, 2005; Kim, 2009; Tann and Sanderson, 2009) and some are in experimental settings rather than looking at actual use. For

example, Campbell and Fast (2004) use a hypothetical research scenario to compare student's use of Google and a library catalogue.

Knowledge of for what purposes the broader population are actually using search engines, libraries and Wikipedia will assist libraries to know if indeed they are in competition with Google and Wikipedia. It will also help them to shape their collection and reference services in a way that responds more closely to the needs of different sectors of their publics.

Ideally, it would be possible to track use of public library reference services and use of the library collection. In the absence of data on the subject of reference queries, this article looks at the subject of catalogue searches in one major public library, the State Library of Victoria. One of the stated aims of the State Library of Victoria (SLV) is to 'put information into the hands of all Victorians when and where they want it' (Library Board of Victoria, 2009, p 10) This would seem to make it a particularly pertinent site of comparison. As most of the collection of the SLV is closed stack, it can only be accessed through the library catalogue. In this article, the subject of queries entered into the SLV catalogue is compared with the subject of queries entered into Google and Wikipedia. Each of these analyses has been published separately in Waller (2009, 2011a, 2011b). This paper brings together data from these separate analyses to form a comparison, combining this with previously unpublished estimates of the numbers of people who use each information source. The research increases our understanding of the types of information that people are looking up or accessing on various information sources and the extent to which they are using each resource. Although the analysis is of the search queries of Australian users, the study has implications for web search research and libraries in any country.

2 Methodology

2.1 Data sources

The data used in these analyses was drawn from 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 a Wikipedia entry. Data on visits and visitors to Google and Wikipedia were sourced from one month of Hitwise data on Australian Internet use. Hitwise is a web analytics company which measures Australian Internet use through collecting data directly from Internet Service Providers. It collects the log files of proxy cache servers and covers more than one third of Australian Internet subscriptions, including homes, businesses, schools, universities, and libraries. Hence, it does not include search engine queries that are generated by software agents. 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.

	Queries typed into SLV catalogue	Queries typed into Google (Australia)	Search queries that led to Wikipedia	No of visits to SLV site	Relative % of visits to SLV, Google and Wikipedia
	Voyager logs	Hitwise	Hitwise	HBX analytics	Hitwise
Period of analysis	May 2008	April 2009	April 2009	May 2008	April 2009

Table 1: Data sources

The analyses of Google search queries (Waller 2011a) and queries which led to Wikipedia (Waller 2011b) each included the long tail in its analysis. The term ‘long tail’ was brought into popular usage by Chris Anderson (2006), referring 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 terms that brought people to Wikipedia in April 2009 accounted for just 4% of visits via search engines; hundreds of thousands of search terms accounted for the other 96% of visits with most only accounting for one or two visits each. Similarly, the top one hundred search terms that people typed into Google (Australia) in April 2009 accounted for just 4% of visits to Google (Australia). Most transaction log analyses use random samples or the most popular terms. Either of these methods can be misleading given that search queries follow a zipf distribution. In (Waller 2011a) and (Waller 2011b), both the sampling and the analysis have been weighted to take into account the distribution of the search queries along the long tail.

2.2 Search queries that led to Wikipedia

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 to Wikipedia (from Australia) came directly from a search engine (Source: unpublished data, Hitwise) and almost all (93%) of these came from Google.

It is important here to make a distinction between search terms and search queries. A ‘search query’ is an instance of a user typing a query into a search engine. A ‘search term’ is what is typed into the search engine. It is estimated that more than 600,000 search queries in the Hitwise sample took a user to Wikipedia and that at least 400,000 of these queries used search terms that appeared only once. The sample analysed in this study was selected from an extract of 50,000 different search terms that took a user to Wikipedia in April 2009.

Using a proprietary method, Hitwise matches Internet use with Mosaic Australia lifestyle profiles (More information is available in Waller (2011a, 2011b) and <http://www.mosaicaustralia.com.au/>). For each Mosaic lifestyle group, a random

sample of 160 search terms, stratified by the number of queries associated with each search term, 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.

As described in (Waller 2011a, 2011b) these studies used weighted sampling methods to reflect the distribution of search queries along the long tail. 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. 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.

2.3 Queries typed into Google (Australia)

The selection of the sample of Google search terms was done in exactly the same way as the selection of the Wikipedia search terms, although the distribution of search queries was quite different and hence the weightings were very different. Three-quarters of queries (74.8%) used search terms that 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 different search terms 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.

The extent of any overlap between the Google and Wikipedia samples is negligible; just 2.2% all the search queries typed into Google (Australia) in April 2009 led to Wikipedia.

2.4 Queries typed into SLV catalogue

The State Library of Victoria is a large public research library that provides a legal depository for works published in the State of Victoria (Source: <http://www.slv.vic.gov.au/about-us/history-library> accessed 3 August, 2010). Because almost all of its collection is closed stack, the library catalogue needs to be searched in order to gain access to a particular resource.

The sample of catalogue searches 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 analysis of the subject of queries was 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. Lastly, the sample was restricted to those search types that were considered to be about a general topic rather than a specific item. Title searches for books and journals accounted for 24.5% of all catalogue searches and Author Browse searches accounted for 13.6%

of all catalogue searches. These types of search are clearly searches to obtain a specific item. This can be considered to be different from using the library catalogue to find what resources exist on a particular topic. Hence Author and Title searches were excluded from the sample and the sample was stratified according to the prevalence of each remaining search type (predominantly Keyword Relevance Search). In the period under study, 40% of all searches were conducted from within in the library by members of the public and 60% were conducted by people accessing the catalogue remotely, for example, from home or work.

2.5 Subject coding

In each of the selected samples, all search terms were manually examined and coded with one code 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. Each search query from the Wikipedia sample was looked up in Google to enable inspection of the relevant Wikipedia entry in the search results. Unless the meaning was immediately obvious, each search query from the Google sample and the library catalogue was looked up in Google and/or the SLV catalogue as appropriate.

The analysis of queries that led to Wikipedia and queries typed into Google were conducted subsequent to the analysis of catalogue queries. 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:

- 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 books 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)

The code “Other” was used for those subjects that did not have enough queries to warrant their own category. The code ‘unknown’ was assigned where the meaning of the term was unknown or ambiguous.

For the purposes of the comparison undertaken in this analysis, some of the library catalogue queries that had been coded as cultural practice in Waller (2009) needed to be recoded to the new categories created for the analysis of Google and Wikipedia queries. In particular, the very small number of library catalogue queries relating to celebrities, TV shows and movies were recoded from cultural practice in Waller (2009) to popular culture in this analysis.

Google search queries were also classified according to the presumed intention behind the query. Drawing on Broder (2002), search queries were deemed to be one of 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 Waller (2011a), 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 Waller (2011a) 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 query ‘Don Knuth’ is

navigational, that is that the searcher wants to reach this person's academic home page, in Waller (2011a), people's names were regarded as informational queries and coded to the relevant subject (for example, politics, celebrity, popular music).

2.6 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. Of course, the distribution of subjects looked up in Google and Wikipedia 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. The main limitation for a study like this, however, 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.

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 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 large, medium and small ISPs, there may be some sample bias, the direction of which is impossible to detect.

3 Findings: Comparing subjects across the different information resources

The results of combining the analyses are shown in Table 2. This compares the subject of the general search queries typed into the catalogue with the search queries typed into Google and the search queries that led people to Wikipedia. It can be seen that there are substantial differences in the spread of subjects being queried or accessed via each information resource.

	<i>general search queries typed into catalogue¹</i>	<i>search queries that led to Wikipedia</i>	<i>informational search queries typed into Google Australia</i>	<i>all search queries typed into Google Australia</i>
Place/building	10%	5%	4%	2%
Related to SLV or other libraries	-	-	-	-
History	10%	6%	1%	-
Books/authors/newspapers	12%	5%	2%	1%
Contemporary issues	20%	5%	3%	5%
Cultural practice	12%	14%	19%	15%
Genealogy		-	-	-
High Culture	12%	2%	1%	-
Person - not elsewhere classified	3%	-	-	-
Indigenous	-	-	-	-
Science	4%	6%	2%	1%
Business-related	5%	-	-	-
Popular Culture	-	40%	29%	25%
eCommerce	-	-	20%	24%
Computing/Web	1%	8%	4%	8%
Health	5%	6%	6%	3%
Unknown	4%	1%	6%	3%
Weather/Time/Public transport	-	-	3%	2%
Adult	-	-	-	8%
Other	1%	2%	1%	1%
Total	100%	100%	100%	100%

Table 2: Comparison of the subject of queries on different information sources²

¹ As explained in the Methodology section, this excludes Title searches for books and journals and Author searches as these can be considered to be searches for a specific item.

² To make the table easier to read, a dash is used in place of 0%.

One in five library catalogue users were looking for information on contemporary issues (20%). As described in (Waller 2009), these were predominantly queries about social and political issues, including international issues and issues relating to social welfare, social policy and theory. In contrast, only 5% of Google queries or queries leading to Wikipedia related to contemporary issues. While only 1% of catalogue queries related to popular culture, popular culture was a major subject of investigation, accounting for 29% of Google informational queries and two fifths (40%) of queries that took people to Wikipedia.

A major area of investigation across all information resources was cultural practice, which included queries about sport, religion, hobbies and cultural practices associated with specific ethnicities. Queries about cultural practice accounted for 12% of library catalogue queries, 14% of Wikipedia queries and 19% of Google informational queries. eCommerce was also a major topic of Google informational queries accounting for one fifth of all queries.

As shown in Table 2, the main differences with regard to 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.

Although the catalogue search types of Title search and Author browse were excluded from the analysis, users did use other catalogue search options to enter queries that related to books, authors and newspapers. Just over half of these were searches for an author, while most of the rest were queries relating to types of literature. (See Waller (2009) for more detail on this).

Figure 1 summarises the information in Table 2. In order to highlight the main differences, any subject category responsible for 6% of queries or less has been included in 'Other'.

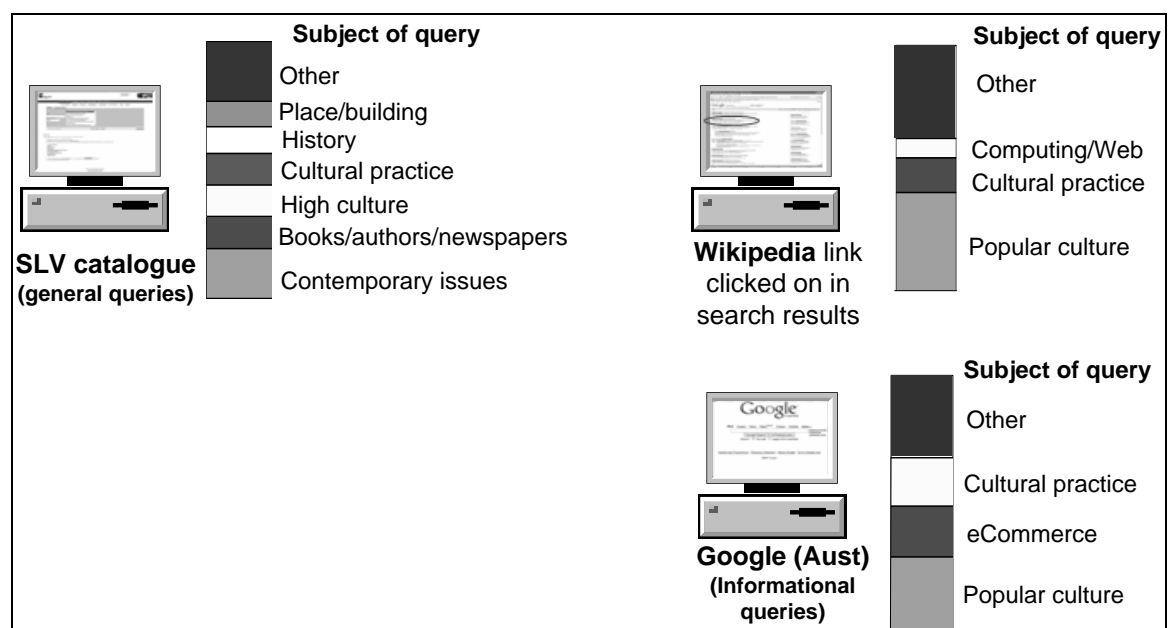


Figure 1: Comparing subjects across different information resources

It is clear from Figure 1 that people who access the SLV catalogue with a general query are mainly looking for information on contemporary issues, books/authors/newspapers, high culture, cultural practice, history or place/buildings. In contrast, people looking for information using Google are likely to be looking for information on popular culture, ecommerce, or cultural practice and those accessing Wikipedia are likely to be looking for information on popular culture, cultural practice, or computing/web. In other words, the distribution of subjects looked up in the library catalogue are quite different from the distribution of subjects looked up on Wikipedia and Google.

2.7 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 particular subject areas via Wikipedia, Google or 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, we would like to know how the actual **number** compares with the number 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 library catalogue, Wikipedia or Google to find information. However, this would only be possible if, firstly, we had access to accurate estimates of the numbers of people using each resource and, secondly, the subject coding was accurate to several decimal points. Neither of these conditions hold. This is because, firstly, Hitwise covers approximately 40% of all Internet activity in Australia but does not report any absolute numbers. Secondly, 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 users of the catalogue, visitors to Wikipedia and Google users. This can be done by using Hitwise data to look at the factor by which visits to Google are greater than visits to Wikipedia and visits to the SLV catalogue. Hitwise data shows the relative difference between visits to Google, Wikipedia and the SLV website rather than the SLV catalogue. It turns out, however, that the number of visitors to the SLV website is of the same order as the number of visitors to the SLV catalogue. In May 2008, there were approximately 120,000 visits to the SLV website (Source: Unpublished HBX Analytics data). In the same month, members of the public typed 117,507 search queries into the catalogue. Because these numbers are of the same order, the number of catalogue visits can be directly substituted for the number of visits to the SLV website, with no transformation needed. In other words, available Hitwise data on the relative difference between visits to Google, Wikipedia and the SLV website can be understood as data on the relative difference between visits to Google, Wikipedia and the SLV catalogue .

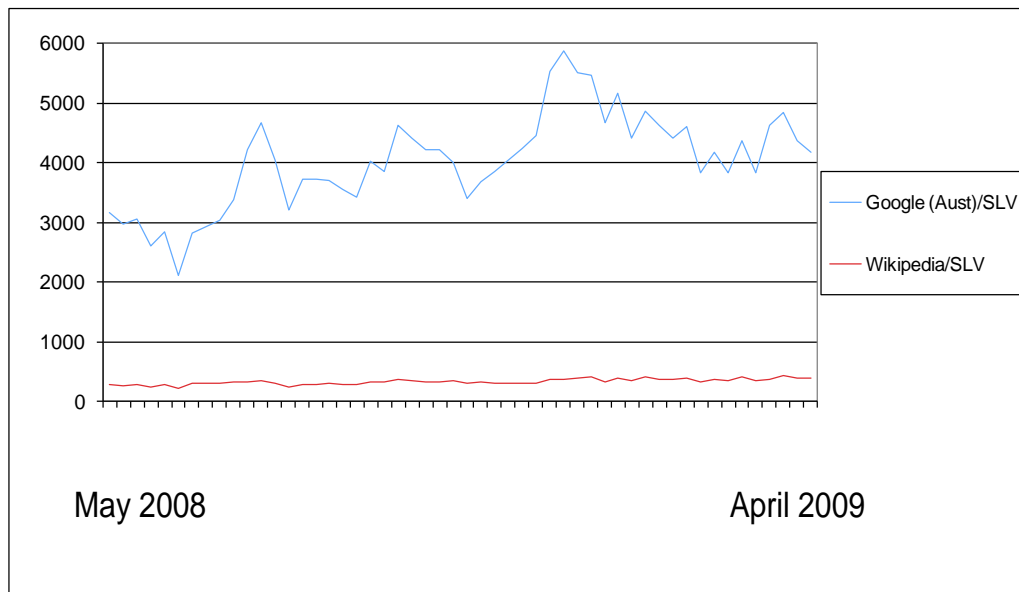


Figure 2: Relative number of visits to Google (Australia), Wikipedia and SLV

Figure 2 plots Hitwise data on the number of visits to Google (Australia) divided by the number of SLV catalogue queries and similarly for Wikipedia. It can be seen that between May 2008 and April 2009, between two thousand and six thousand times more queries were typed into Google (Australia) than into the SLV catalogue. In the same period, the number of Wikipedia queries was approximately 200 times as many as entered into the SLV catalogue. These figures need to be adjusted in order to make a more legitimate comparison. It is assumed that all of those searching from the SLV catalogue are from the State of Victoria, whereas Google and Wikipedia users are from all over Australia. In April, 2009, 25.5% of visits to Google (Australia) and 24.3% of visits to Wikipedia were from Victoria. In addition, for every four visits to Google (Australia) there was a visit to Google.com. 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 100 times as many Victorian visits over the month as the SLV catalogue. Google had in the order of 700 times as many Victorian visits over the month as the SLV catalogue. Figure 3 depicts these differences visually. The size of each circle roughly approximates the relative size of the number of queries.

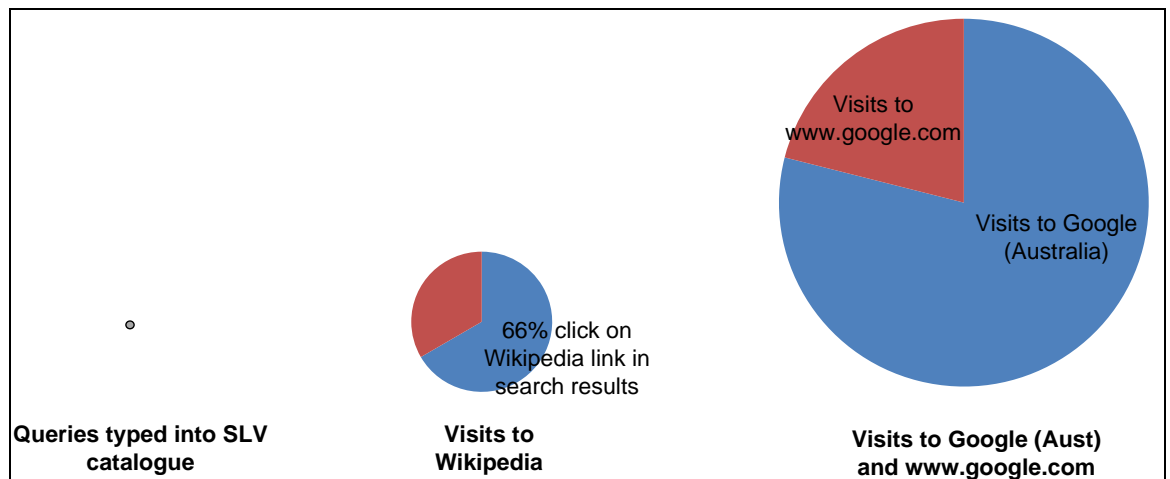


Figure 3: Approximate relative sizes of the number of queries using each information resource

4 Discussion

The distribution of subjects that people access on Wikipedia and look up on Google are completely different from those looked up on the SLV catalogue. This indicates that people turn to different information resources for different subjects. 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; less than five percent of queries on either related to contemporary issues. These findings regarding Australian users are broadly consistent with research from the United States that suggests that in the United States very few internet users seek political information on the web or visit political sites (Hindman, 2009) and one fifth of adults research product information online (Jansen, 2010). In contrast to the spread of topics looked up on Wikipedia and Google, one in five people conducting a general search on the catalogue were looking for material related to contemporary issues and less than 1% were looking for material relating to popular culture or ecommerce.

Estimates of the relative use of each resource indicate that Google is used several hundred times more often than the library catalogue. One way of understanding the implications of this is to look at the incidence of queries about contemporary issues. This is the most frequent subject of the library catalogue search, accounting for one in every five library catalogue searches. For every five library catalogue searches, there are in the order of 3,500 Google informational searches conducted in Victoria, approximately 3% or 100 of which relate to contemporary issues. In other words, Google is used approximately one hundred times more often than the State Library catalogue to look up information on contemporary issues.

Similarly for every five library catalogue searches, there are in the order of 500 searches conducted in Victoria using Wikipedia, approximately 5% or twenty five of which relate to contemporary issues. In other words, Wikipedia is used approximately twenty five times more often than the State Library catalogue to look up information on contemporary issues.

The nature and quality of information that people obtain on contemporary issues is of fundamental importance to democracy (Webster, 2006). This means that it is critical to conduct further research that looks more closely into how people are searching for, interpreting and using information on contemporary issues obtained via Google or Wikipedia.

4.1 Implications for libraries

It is suggested that libraries should monitor the subject of catalogue queries. 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 library as a source of information on these topics. Perhaps, promotion of the catalogue and collections could emphasise these aspects of the collections.

Perhaps the most important aspect to be borne in mind when interpreting this data is the qualitative differences between the library and the Internet as sources of information. Various commentators have pointed out the difficulty in finding authentic meaning when using a search engine, as it is extremely difficult to link the information found to existing bodies of knowledge (Brabazon, 2007; Dreyfus, 2001; Haigh, 2006; Jeanneney, 2007). Moreover, almost two-thirds (62%) of search engine users do not go past the first page of results and less than 10% go past the first three pages of results (iProspect, 2008).

The library has two advantages when it comes to providing access to bodies of knowledge. Firstly, full bibliographic details are provided which enable users to check the source. Secondly, material is organised by subject and this structure of the library catalogue, whether it is the Dewey Decimal System or Library of Congress Subject Headings or some other form, assists users to link to related subjects.

With regard to Wikipedia, commentators have also decried the accuracy (Gorman, 2007; Keen, 2007), depth (Jeanneney, 2007) or coherence (Duguid, 2006) of Wikipedia articles, critical of the lack of overall editorial responsibility and the fact that anybody can contribute to an article. This is in contrast to the library collection which generally has two levels of quality control. Firstly, it generally only contains items that have been through a publishing process and, secondly, the collection itself is selected by librarians.

5 Conclusion

This exploratory analysis has compared the subjects of queries entered into the State Library of Victoria catalogue with Google queries and the queries leading to Wikipedia. This has been one of the first studies to make comparisons between what people look for on the library catalogue, Google and Wikipedia.

The research reported on in this paper has showed that the distribution of subjects that people access on Wikipedia, look up on Google and look up on the SLV catalogue are quite different from each other. There are also enormous differences in the frequency of use. The use of the Google search engine for informational queries is in the order of seven hundred times greater than the State Library catalogue. Wikipedia is used in the order of one hundred times more often than the State Library Catalogue.

Concern has been expressed within the library sector that library reference services and collections are no longer needed because of Google and Wikipedia. While this research shows clearly that the number of people using Google and Wikipedia far exceeds the number of catalogue users, the fact that the distribution of subjects that people access on Wikipedia and look up on Google are completely different from those looked up on the SLV catalogue suggests that the public library is still an important information resource for particular subjects and that public libraries are not in simple competition with Google and Wikipedia. Rather public libraries still have an important role in providing particular types of information or, more importantly, access to particular bodies of knowledge, and should focus their efforts on their collection strengths.

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. More research is needed to gauge whether people tend to choose different information sources for different subjects or whether different people are drawn to different information resources regardless of the subject. 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, indeed, people do turn to different information resources for different subjects. More research is also needed to gauge whether those people using Google and Wikipedia also use public libraries.

References

- Abram, S. (2005) Google's new initiatives are rocking our world. Here's how to rock back. *Library Journal* (02/01/2005).
- Arnold, E. (2007) How to conduct search analytics. *Searcher*, **15**(10), 40-43.
- Brabazon, T. (2007) *The University of Google: Education in the (Post) Information Age*. Hampshire: Ashgate.
- Broder, A. (2002) A taxonomy of web search. *SIGIR Forum* **36**(2), 3-10.
- Campbell, D. G. and Fast, K. V. (2004) Panizzi, Lubetzky and Google: how the modern web environment is reinventing the theory of cataloguing. *The Canadian Journal of Information and Library Science*, **28**(3), 25-38.
- Choo, C. W., Detlor, B. and Turnbull, D. (2000) Information seeking on the web: an integrated model of browsing and searching. *First Monday*, **5**(2).
- comScore. (2007) *Press release: comScore releases August U.S. search engine rankings*. URL: [http://www.comscore.com/Press Events/Press Releases/2007/09/Top US Search Engines](http://www.comscore.com/Press%20Events/Press%20Releases/2007/09/Top_US_Search_Engines) [accessed 18.5.10].
- comScore. (2010) *Press release: comScore Releases April 2010 U.S. search engine rankings, 11 May 2010*. URL: [http://www.comscore.com/Press Events/Press Releases/2010/5/comScore Releases April 2010 U.S. Search Engine Rankings](http://www.comscore.com/Press%20Events/Press%20Releases/2010/5/comScore%20Releases%20April%202010%20U.S.%20Search%20Engine%20Rankings) [accessed 18.5.10].

- D'Elia, G., Jorgensen, C., Woelfel, J. and Rodger, E. J. (2002) The impact of the internet on public library use: an analysis of the current consumer market for library and internet services. *Journal of the American Society for Information Science and Technology*, **53**(10), 802-820.
- Dempsey, L. (2006) The (digital) library environment: ten years after. *Ariadne*, **46**. URL: <http://www.ariadne.ac.uk/issue46/dempsey/> [accessed 08.8.11].
- Dreyfus, H. L. (2001) *On the Internet*. London: Routledge.
- Duguid, P. (2006). *First Monday*, **11**(10), October 2006. URL: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1405/1323> [accessed 08.8.11].
- Dutton, W. H., Helsper, E. J. and Gerber, M. M. (2009) *The internet in Britain 2009*. Oxford: Oxford Internet Institute, University of Oxford. URL: http://www.oii.ox.ac.uk/downloads/index.cfm?File=research/oxis/OxIS2009_Report.pdf [accessed 08.8.11].
- Ewing, S. and Thomas, J. (2010) *CCi digital futures 2010: the internet in Australia*. Melbourne: ARC Centre of Excellence for Creative Industries and Innovation, Institute for Social Research, Swinburne University of Technology.
- Fallows, D. (2008) *Data memo: search engine use*. Washington: Pew Internet and American Life Project.
- Gorman, G. E. (2007) A tale of information ethics and encyclopaedias; or, is Wikipedia just another internet scam? *Online Information Review*, **31**(3), 273-276.
- Griffiths, J. R. and Brophy, P. (2005) Student searching behavior and the web: use of academic resources and Google. *Library Trends*, **53**(4), 539-554.
- Haigh, G. (2006) Information idol: how Google is making us stupid. *The Monthly* (February). URL <http://www.themonthly.com.au/how-google-making-us-stupid-information-idol-gideon-haigh-170> [accessed 08.8.11].
- Hindman, M. (2009) *The Myth of Digital Democracy*. New Jersey: Princeton University Press.
- iProspect. (2008) *iProspect blended search results study*. URL: http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm [accessed 09.3.10].
- Jansen, B. J. (2010) *Online Product Research*. Pew Research Center. URL: <http://www.pewinternet.org/Reports/2010/Online-Product-Research.aspx> [accessed 11.8.11].
- Jansen, B. J., Booth, D. L. and Spink, A. (2008) Determining the informational, navigational and transactional intent of web queries. *Information Processing and Management*, **44**, 1251-1266.
- Jansen, B. J. and Pooch, U. (2001) Web user studies: a review and framework for future work. *Journal of the American Society of Information Science and Technology*, **52**(3), 235 - 246.

-
- Jansen, B. J. and Spink, A. (2005) How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management* **42**(1), 248-263.
- Jeanneney, J.-N. I. (2007) *Google and the myth of universal knowledge : a view from Europe* (T. Lavender Fagan, Trans.). Chicago: University of Chicago Press.
- Keen, A. (2007) *The cult of the amateur : how today's internet is killing our culture*. London: Nicholas Brealey.
- Kim, J. (2009) Describing and Predicting Information-Seeking Behavior on the Web. *Journal of the American Society for Information Science and Technology*, **60**(4), 679–693.
- Library Board of Victoria. (2009) Annual Report 08/09. Melbourne, Australia: State Library of Victoria. URL: http://www2.slv.vic.gov.au/about/information/annual_reports/index.html [accessed 11.8.11].
- Lim, S. (2009) How and why do college students use Wikipedia? *Journal of the American Society for Information Science and Technology*, **60**(11), 2189-2202.
- Markey, K. (2007) The online library catalog: paradise lost and paradise regained? *D-Lib Magazine* **13**(1/2). URL: <http://www.dlib.org/dlib/january07/markey/01markey.html> [accessed 11.8.11].
- Navarro-Prieto, R., Scaife, M. and Rogers, Y. (1999) *Cognitive strategies in web searching*. Paper presented at the 5th conference of Human Factors and the Web, Maryland, USA.
- OCLC. (2005) *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*. URL: http://www.oclc.org/reports/pdfs/Percept_all.pdf [accessed 11.8.11].
- Park, S. (2009) Analysis of characteristics and trends of web queries submitted to NAVER, a major Korean search engine. *Library and Information Science Research*, **31**, 126-133.
- Rose, D. E. and Levinson, D. (2004) *Understanding user goals in web search*. Proceedings of the 13th international conference on World Wide Web (WWW 2004), May 17–22. New York, USA. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.404&rep=rep1&type=pdf> [accessed 10.8.11].
- Ross, N. C. M. and Wolfram, D. (2000) End user searching on the internet: an analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science*, **51**(10), 949-958.
- Segev, E. and Ahituv, N. (2010) Popular searches in Google and Yahoo!: a "digital divide" in information uses? *The Information Society*, **26**(1), 17-37.
- Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. (2001) Searching the web: the public and their queries. *Journal of the American Society for Information Science*, **52**(3), 226-234.

-
- Taiga Forum Steering Committee. (2006) *Taiga forum provocative statements: March 10, 2006*. Accessible from URL: <http://www.taigaforum.org/docs/ProvocativeStatements.pdf> [accessed 15.8.11].
- Tann, C. and Sanderson, M. (2009) Are web-based informational queries changing? *Journal of the American Society for Information Science and Technology*, **60**(6), 1290-1293.
- Tenopir, C. (2003) Use and users of electronic library resources: an overview and analysis of recent research studies. Washington, DC: Council on Library and Information Resources.
- Waller, V (2009) What do the public search for on the catalogue of the State Library of Victoria? *Australian academic and research libraries*, **40**(4), 266-285.
- Waller, V (2011a) "Not just information: who searches for what on the search engine Google?" *Journal of the American Society for Information Science and Technology*, **62**(4), 761-775.
- Waller, V (2011b) "The search queries that took Australian Internet users to Wikipedia" *Information Research*, **16**(2) paper 476. URL: <http://InformationR.net/ir/16-2/paper476.html> [accessed 11.9.11].
- Webster, F. (2006) *Theories of the information society*. London: Routledge.
- Yu, H. and Young, M. (2004) The Impact of web search engines on subject searching in OPAC. *Information Technology and Libraries*, **23**(4 (December)), 168-180.
- Zickuhr, K. and Rainie, L. (2011) *Wikipedia, past and present*. Pew Research Center. URL: <http://www.pewinternet.org/Reports/2011/Wikipedia.aspx> [accessed 11.8.11].
-

Acknowledgements

This research was funded by Australian Research Council grant LP077215. The author would like to thank Hitwise for their generous support and the many staff at the State Library of Victoria for their assistance and suggestions.

Open access and copyright

Library and Information Research is an open access journal. A freely available copy of this paper may be downloaded from the journal's website: <http://www.cilipjournals.org.uk/lir>

Copyright and associated moral rights in works published in *Library and Information Research* are retained by the author(s) but this paper may be used freely, with proper attribution, in educational and other non-commercial settings.