This paper reports on a transaction log analysis of the type and topic of search queries entered into the search engine Google (Australia). Two aspects, in particular, set this apart from previous studies: the sampling and analysis take account of the distribution of search queries, and lifestyle information of the searcher was matched with each search query. A surprising finding was that there was no observed statistically significant difference in search type or topics for different segments of the online population. It was found that queries about popular culture and Ecommerce accounted for almost half of all search engine queries and that half of the queries were entered with a particular Website in mind. The findings of this study also suggest that the Internet search engine is not only an interface to information or a shortcut to Websites, it is equally a site of leisure. This study has implications for the design and evaluation of search engines as well as our understanding of search engine use.

Introduction

As the use of Internet search engines continues to grow to become one of the most frequent activities undertaken on the Internet, it becomes increasingly important to understand what people are using search engines for. In the past, with no single search engine dominant, it has made sense to compare the use of different search engines, as it was possible that different types of users were attracted to particular search engines or that particular search engines were favored for certain subject areas. The situation has changed, with Google currently dominating Internet search in most Western countries. Google’s market share is almost 90% in Australia (Experian Hitwise, 2010), 80% in Europe (comScore, 2008), and 64% in the United States (comScore, 2010). According to the Web analytics company, Comscore (comScore, 2009), “Nearly one out of every ten minutes a person spends online around the world is spent on a Google site,” such as a Google search engine, Google books, YouTube, or Gmail.

This study uses transaction logs to provide an analysis of the type and topic of search queries entered into the search engine Google (Australia) in April 2009. This study addresses many of the limitations of previous studies on search engine use. For example, it uses 1 month’s worth of logs and only includes search engine queries that have actually been typed in. Most transaction log analyses use random samples or the most popular queries. Either of these methods can be misleading, given that search queries follow a zipf distribution. In this study, both the sampling and the analysis take into account the distribution of the search queries along the long tail. One of the most unique aspects of this study is that the analysis matches lifestyle information of the searcher with the search query, with the potential to give insight into who uses search engines for what.

The paper begins with evidence that the Internet and search engines are increasingly used to find information. It then provides a brief overview of the information-seeking literature as a precursor to studies of search engine use which demonstrate that a search engine user may not be looking for information. The discussion compares the findings of the current study with the findings of previous studies of search engine use and suggests a possible interpretation of the results.

While queries relating to popular culture have not been separately identified in previous studies, in this study it was found that such queries accounted for one-quarter of all search queries. This illustrates the importance of an inductive approach to the classification of search queries, using categories that best describe the search query rather than relying on preselected categories. The presentation of the findings includes descriptions of broad categories to enable future application of the coding scheme by other researchers. The findings of this study suggest that the Internet search engine is not only an interface to information or a shortcut to Websites, it is equally a site of leisure, with one in six of all searches estimated to be ‘leisure searches.’

Although the analysis was of the search queries of Australian users, the study has implications for Web search
The Internet as a Source of Information

In the past 10 years the Internet has become widely used as a source of information. It also appears that the search engine is increasingly where Internet users turn to first. A study undertaken by the Pew Internet and American Life Project in 2007 (Wells, 2008) found that almost 60% of respondents would consult the Internet when they needed to address problems. While it is not clear from the question asked whether these respondents were looking up information on the Internet or using the Internet to establish communications with somebody who could assist them, it is known that using a search engine is one of the most common activities on the Internet. Of those Americans who would turn to the Internet first for specific science information, approximately 90% considered that they would use a search engine first (Horrigan, 2006). A U.S. study undertaken in 2007 found that just under half (49%) of Internet users used a search engine every day (Fallows, 2005). Since then, the number of searches undertaken in the U.S. 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, almost three-quarters of Internet users (73%) considered the Internet to be an ‘important’ or ‘very important’ source of information (Ewing & Thomas, 2010).

In Britain, the majority of Internet users have used the Internet to look for health information, sports information, and news (Dutton & Helsper, 2007; Dutton, Helsper, & Gerber, 2009) and those who turn to the Internet for information are increasingly likely to use search engines. According to the Oxford Internet Survey, in 2009, almost two-thirds of Internet users (64%) mainly used a search engine to find information on the Internet compared to one-fifth (19%) 4 years earlier (Dutton et al., 2009).

Because it has been established that many people turn to a Web search engine when they are looking for information, a common assumption is that search engine use is the same as information-seeking.

From Information Search to Search Engine Use

Information search was well established as a field of study many years before the widespread use of Internet search engines. Different models of information search have emphasized different aspects of the searcher and the search environment. For example, some models focus simply on the behavior of the searcher (Wilson, 1981), Kuhlthau (1991) takes into account the feelings, thoughts, and actions of the searcher, while Johnson and Meische (1993) include the information environment. Traditional models of information search have all assumed the existence of an information need, that is, a gap between what a user knows and what they need to know (Belkin, 1990; Marchionini, 1995; Taylor, 1968).

Research on information search has been the theoretical starting point for research on the use of Web search engines. For example, Knight and Spink (2008) draw on a range of models of information search (Bates, 2003; Ellis, 1989; Ingwerson, 1996; Johnson & Meischke, 1993; Kuhlthau, 1991; Marchionini, 1995; Saracevic, 1996; Wilson, 1981) in proposing a model of Web search that includes the information need, the searcher, the setting, and the interaction between these elements. More recently, distinctions have been made between different types of information need. Marchionini (2006) proposes that there is a hierarchy of information needs ranging from Lookup searches which are like “fact retrieval” or “question answering” to “exploratory search” which White, Kules, Drucker, and Schraefel (2006) describe as “those searches where the searcher has a general interest but not specific knowledge.” Exploratory search also includes what Marchionini (2006) calls “social searching,” that is, where searchers uses search to try to expand their social network.

The concepts of task and information need are central to these conceptualizations of information seeking on the Web (Kim, 2009; Rieger, 2009). As Rose and Levinson (2004) point out, underlying this literature on search engine use is the assumption that “Searching is merely a means to an end.” While information-seeking models are useful for understanding information search that is in response to an information need, Web searches are not always conducted to fulfill an information need. The user of a Web search engine may want to go to a particular Website or they may want to do something. To reflect this, Broder (2002) developed a taxonomy of Web searches consisting of informational searches plus navigational and transactional searches. Navigational queries are those queries entered by the search engine user in order to reach a particular site which the user may have visited before or assumes to exist (for example, the name of a government department). Transactional queries are those queries which the search engine user enters in order to perform some Web-mediated activity. Broder’s examples of Web-mediated activity include shopping, downloading files, and accessing databases.

Rose and Levinson (2004) consider that Broder’s category of transactional searches should be replaced with the broader category of ‘resource search.’ The goal of a ‘resource search’ is not necessarily to perform a Web-mediated activity, but to access an online resource other than information. Rose and Levinson include in this category queries for song lyrics, recipes, pornography, and sewing patterns, arguing
that the searcher wants to obtain the resource itself as distinct from information about the resource. It should be noted that not all researchers agree with Broder (2002) and Rose and Levinson (2004) that navigational and transactional (or resource) searches are different from searches conducted to acquire information. For example, Marchionini (2006) regards navigational search and transactional search as types of lookup search. It is considered here, however, that both Broder (2002) and Rose and Levinson (2004) have amply demonstrated that a search engine user may not necessarily be searching for information.

Using Transaction Log Analyses to Study Search Engine Use

Transaction log analyses are commonly used to investigate people’s use of search engines (Jansen, 2006). The main limitation of using transaction logs 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. However, there are several advantages to using transaction logs. They enable the study of a large sample of users, are unobtrusive, and do not affect user behavior. Because transaction logs record what people actually do, they overcome the limitations associated with relying on what people say they do.

Some empirical analyses of Web search transaction logs have drawn on Broder’s taxonomy to classify Web search engine queries according to presumed intent (Jansen et al., 2008; Rose & Levinson, 2004). Although Kammenhuber, Luxenburger, Feldmann, and Weikum (2006) consider that clickstreams are important when trying to gauge searcher intent, Rose and Levinson (2004) found little difference comparing classifications of presumed intent using just the search term and classifications that also made use of the clickstream. Their work suggests that the search term alone is sufficient to classify the presumed intent of the query.

Others have classified the queries according to the topic of the query (Jansen & Spink, 2005; Park, 2009; Ross & Wolfram, 2000; Segev & Ahituv, 2010; Spink, Wolfram, Jansen, & Saracevic, 2001). There is no one standard classification scheme used in these studies. Although library classification schemes such as the Library of Congress Classification could potentially be used, as Park (2009) explains, the classification systems traditionally used in libraries are not well suited for classifying Web queries. Segev and Ahituv (2010) used the Open Directory Project (ODP) classification to classify the 150–200 most popular search queries in different countries. This has the advantage that each query can be typed into the ODP and it automatically suggests a classification. Other researchers have developed their own categories inductively from the data (Park, 2009; Ross & Wolfram, 2000).

Many existing transaction log analyses suffer from a variety of limitations. For example, some are of only 1 day’s worth of logs (Hochstotter & Koch, 2009). More significantly, logs provided by search engines include agent queries which may be difficult to identify and remove (Jansen et al., 2008). Additionally, as pointed out by Jansen (2006), one of the main limitations of search log analysis is the fact that there is no way to collect demographic data.

Another limitation is that most transaction log analyses use random samples (for example, Broder, 2002; Jansen et al., 2008; Park, 2009; Rose & Levinson, 2004) or the most popular queries (Ross & Wolfram, 2000). Either of these methods can be misleading, given the particular distribution of search queries. Search queries entered into a search engine follow a zipf distribution whereby a few items account for a sizable proportion of the total, and an enormous number of items, the long tail (Anderson, 2006), each contribute a tiny proportion to the rest. In a study of 2001 data, Spink and Jansen (2004a) found that 20% of Web search queries were used often, while 10% of the queries appeared only once. The situation has changed dramatically since then with the increased use of search engines. For example, the top 100 search queries that people typed into Google (Australia) in April 2009 accounted for just 4% of visits to Google (Australia); hundreds of thousands of search queries accounted for the other 96% of visits with most only accounting for one or two visits each. In the conclusion to her study, Park (2009) suggests that future research use weighted sampling methods to reflect the distribution of search queries.

There are few, if any, studies of search engine use in Australia apart from Segev and Ahituv’s (2010) cross-national comparison of the most popular Web queries. As the next section shows, the study presented here of Australian search engine use addresses some of the limitations of previous studies. For example, the analysis is of a sample taken from a month of search queries. The sample is selected and weighted to take account of the distribution of search queries and it does not contain software agent queries, only including those queries submitted by humans. The main distinction of this study, however, is that it combines data on the lifestyle of the searcher with the query with the potential to provide an insight into who is using search engines for what.

The research questions this study addresses are:

- What do people in Australia search for on Internet search engines?
- Does the use of search engines vary across different segments of the online population?

Procedures

This study is an analysis of a sample of queries typed into Google (Australia) over a 4-week period in 2009. The data used in the study were obtained from the Web analytics company, Hitwise (www.hitwise.com.au). Hitwise measures Internet use using a network centric methodology and a proprietary method to match Internet use with lifestyle profiles constructed by Mosaic Australia (www.mosaicaustralia.com.au). Hitwise data cover more than one-third of Australian Internet subscriptions, including homes, businesses, schools, universities, and libraries.

This study is different from previous transactions log analyses in that this study uses weighted sampling methods to reflect the distribution of search queries along the long tail. The sample analyzed in this study was selected from an extract of 60,000 different search terms that were typed into Google (Australia) in April 2009. It is important here to make a distinction between search terms and search queries. The term ‘search query’ refers to an instance of a user typing a query into Google (Australia). A ‘search term’ is what is typed into Google (Australia). For example, the search term ‘Facebook’ is the most common search term typed into Google (Australia) in April 2009. This search term accounts for 0.8457% of all search queries typed into Google (Australia) in that period. The extract of 60,000 search terms accounted for 28.7% of the total search queries entered into Google (Australia) in that month. Hence, it included every search term that appeared more than once in April 2009, as well as what amounted to a random selection of those that appeared only once. The distribution of the number of queries associated with each search term was mapped and, as expected, followed a long tail distribution. This was divided into seven groups on the basis of the number of queries associated with each search term, as shown in Figure 1. Three-quarters of queries (74.8%) were unique, that is, appeared only once in the month of April 2009.

As well as including the long tail in its analysis, another unique feature of this study is that it matches search query data with data on the lifestyle of the searcher. It does this through taking advantage of the Hitwise proprietary method to match Internet use with Mosaic lifestyle profiles. The Mosaic lifestyle profiles used were developed in 2008 specifically to describe the Australian population.

Mosaic Australia is a geodemographic classification, which means that it is based on the premise of homophily, that is, that people live near people who are like them. In order to develop the lifestyle profiles, Mosaic Australia used data from more than 1,000 variables from a range of sources, including the 2006 Australian census and consumer surveys. A bottom-up process involved the testing of each variable and selecting on the basis of discrimination, robustness, and correlation with other variables. The final classification system is based on 238 variables which successfully withstood this testing. The data used in the final classification include data on income, education, employment, family structure, housing, consumer habits and preferences, entertainment and media use, internet and computer use, attitudes to government, and the environment. The Mosaic classification identifies 47 Lifestyle Types that can be aggregated into 11 Lifestyle Groups. Hitwise uses the Mosaic classification to classify each Australian Internet user into one of the 47 Mosaic Lifestyle Types. The analysis presented in this paper has used the data at the level of aggregation of the 11 Lifestyle Groups. These Lifestyle Groups are listed in Table 1 alongside the relevant ‘tagline’ which summarizes the salient features. (For a more complete description of each Lifestyle Group see http://www.mosaicaustralia.com.au.)

These Lifestyle Groups have been derived for the purpose of market research rather than academic research. However, this does not detract from the usefulness of the categories for a study like this. For each Lifestyle Group, a random sample of 160 search terms, stratified by search query frequency, was selected without replacement. In this way, a total sample of 1,760 search terms was drawn. The sample was divided equally across the Lifestyle Groups to ensure the same precision for each Lifestyle Group.

To avoid artificial dominance in the sample of those search terms that were used most frequently, it was necessary in the analysis to also weight search queries according to their position along the long tail distribution. This meant that the total percentage share of each frequency group in the sample matched the total percentage share of each frequency group in the population of search terms typed into Google in April.
Although the coding schemes developed by other researchers were not needed for the analysis of library catalog queries (Waller, 2009), it was also developed inductively (Jansen, 2006; Park, 2009; Ross & Wolfram, 2000). Each search term was looked up in Google to enable inspection of the search results (Ross & Wolfram, 2000). Care was taken not to force the search terms into categories that did not reflect the substance of the search term and additional codes were created as required. For example, the major codes “Ecommerce” and Adult “and” “Weather/Time/Public Transport” were not needed for the analysis of library catalog queries in Waller (2009), but were created as appropriate codes to describe a substantial proportion of the search engine queries. Several codes from Waller (2009) were appropriate for only a very small proportion of queries but they were retained as separate codes to enable future comparison with the subject of library catalog queries and Wikipedia queries.

The coding resulted in 78 codes which closely described the content of the search queries. These were amalgamated into 15 broad subject groupings. On the whole, the categories used in this study were mutually exclusive but there were some search terms that could feasibly be placed in more than one category. A coding priority was implicitly established in the effort to best match the code with the presumed topic of interest. The broad subject groupings are listed below in the order of the coding priority. For example, health is listed above Ecommerce. This meant that the search term ‘reflexology’ was coded as ‘general health’ rather than as a service under Ecommerce. Similarly, the search term ‘southbank cinemas’ was coded as ‘movie’ rather than as a building.

- Popular Culture
  - popular music, TV show, actor, movie, video game, celebrity, Myspace, radio, Youtube
- Book/Author
- High Culture
  - Fine Art, Classical Music, Architecture
- Contemporary Issues
  - contemporary affairs, newspapers/news sites, government, organizations
- Health
  - psychology, mental health, sexual health, general health
- Computing/Web
  - software, email/chat, social networking sites, e.g., Facebook
- Adult
  - pornography, dating sites
- Genealogy
- History
- Science (including mathematics)
- Ecommerce
  - Gambling, airlines, travel, buying/selling, banking/finance, retailer, service, name of product
- Weather/Time/Public transport
- Cultural Practice
  - sport, religion, food, jobs, learning, language, festivals, events, holidays, hobbies, other aspects of cultural practice not elsewhere classified
- Place/Building
- Unknown

TABLE 1. Lifestyle groups and their tagline.

<table>
<thead>
<tr>
<th>Lifestyle group</th>
<th>Tagline</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Privileged prosperity</td>
<td>The most affluent families in the most desirable locations</td>
</tr>
<tr>
<td>B. Academic achievers</td>
<td>Wealthy areas of educated professional households</td>
</tr>
<tr>
<td>C. Young ambition</td>
<td>Educated and high-earning young singles and sharers in the inner suburbs</td>
</tr>
<tr>
<td>D. Pushing the boundaries</td>
<td>Young families living in recent developments on the fringes of major cities</td>
</tr>
<tr>
<td>E. Family challenge</td>
<td>Mixed family forms with stretched budgets in outer suburbs</td>
</tr>
<tr>
<td>F. Metro multicultural</td>
<td>Students and professionals living in high density, lower cost suburbs</td>
</tr>
<tr>
<td>G. Learners and earners</td>
<td>Anglo-Australian blue-collar families in provincial settlements</td>
</tr>
<tr>
<td>H. Provincial optimism</td>
<td>Rural landowners and workers in agricultural heartlands</td>
</tr>
<tr>
<td>I. Farming stock</td>
<td>Low income, low-spending households in major regional and outer metro</td>
</tr>
<tr>
<td>J. Suburban subsistence</td>
<td>Older blue-collar workers and retirees in country and coastal locations</td>
</tr>
<tr>
<td>K. Community disconnect</td>
<td></td>
</tr>
</tbody>
</table>
FIG. 2. Lifestyle profile of visitors to Google, compared to their representation in the Australian population (Data source: Hitwise).

It was possible to code 94% of queries with a primary code. Where the meaning of the query was unknown or ambiguous, the code ‘unknown’ was assigned. The presentation of the findings includes descriptions of broad categories to enable future application of the coding scheme by other researchers.

Search terms were also classified according to the presumed intention behind the query. Drawing on Broder (2002), 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 judgment 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 search term ‘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).

There are a number of caveats to this study. First, the coding inevitably involves some subjective decisions about the appropriate category in which to place a particular search term. Although in the long tail the contribution of each site or search term 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.

Second, 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. In particular, Hitwise does not collect data on Internet access that bypasses an Internet Service Provider; for example, when mobile devices are used to access the Internet via the 3G network.

Third, the segmentation of the Australian population is according to household rather than individual user. Some household types only include particular age groups within households, for example, retirees, or people in their 20s and 30s. Within households with families, it is not possible to distinguish the search engine use of the children from that of the adults in the household.

Lastly, it should also be borne in mind that Hitwise measures visits, not visitors, and so the dataset may include a number of queries from the one individual. This is, however, unlikely, given the large size of the set of queries from which the sample was drawn.

Findings

Together, visitors to www.google.com.au (Google Australia) and www.google.com accounted for 82% of all visits to search engines in the 4 weeks ending 25 April 2009. It can be seen from Figure 2 that Lifestyle Groups E (Mixed family forms with stretched budgets in outer suburbs),
TABLE 2. Top 10 search queries entered into Google (Australia), April 2009.

<table>
<thead>
<tr>
<th>Search query</th>
<th>Percent of searches</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook</td>
<td>0.85%</td>
</tr>
<tr>
<td>youtube</td>
<td>0.51%</td>
</tr>
<tr>
<td>myspace</td>
<td>0.36%</td>
</tr>
<tr>
<td>ebay</td>
<td>0.36%</td>
</tr>
<tr>
<td>hotmail</td>
<td>0.35%</td>
</tr>
<tr>
<td>facebook login</td>
<td>0.17%</td>
</tr>
<tr>
<td>youtube</td>
<td>0.14%</td>
</tr>
<tr>
<td>bebo</td>
<td>0.13%</td>
</tr>
<tr>
<td>ebay australia</td>
<td>0.11%</td>
</tr>
<tr>
<td>yahoo</td>
<td>0.10%</td>
</tr>
</tbody>
</table>

H (Anglo-Australian blue-collar families in provincial settlements) and K (Older blue-collar workers and retirees in country and coastal locations) are more likely to visit Google Australia. Young, well-educated households located in metropolitan areas (Lifestyle Groups C and G) are the least likely to visit Google Australia and more likely to visit Google.com. These differences in the likelihood of visiting Google (Australia) do not affect the results of the study because the sample was stratified according to Lifestyle Group. In calculating totals for the entire sample, weightings were used which take into account the relative contribution of each Lifestyle Group in the total online population.

Table 2 shows the top 10 search terms entered into Google in April 2009. These queries are all navigational queries for megasites, including several social networking sites, two email sites, and ebay, a site for buying and selling. These top 10 search terms, however, account for just 3.1% of all queries entered into Google in that month. It can be seen just how misleading it is to look at the top 10 search terms by comparing these with the distribution of topics of all search queries shown in Figure 3.

**Topics of Search Queries**

It can be seen in Figure 3 that together queries about popular culture and Ecommerce accounted for almost half of all search engine queries. Queries about cultural practice accounted for 15% of all queries, and queries relating to computing or the Web and queries about adult sites were the next most common type of query. The subject of query was compared across the Lifestyle Groups. Although differences across the Lifestyle Groups were observed in the distribution of subjects looked up using Google, testing for the statistical significance of differences across Lifestyle Group could only be undertaken for the subjects of Popular Culture, Ecommerce, Cultural practice, and Adult. When comparing across Lifestyle Groups for other subjects, some of the cell sizes were too small for the results of statistical tests to be reliable. Somewhat surprisingly, the distribution of topics of search query did not vary significantly across the different Lifestyle Groups for the broad subjects of popular culture, Ecommerce, cultural practice and adult. Insufficiently large cell sizes precluded a finer grained analysis ($p > 0.01$).

The following graphs provide more detail about five of the most common broad subject categories. Examples are given to give a better indication of the contents of each broad subject category. The graphs also indicate the distribution within the subject category of informational, navigational and transactional queries.

![FIG. 3. Subject of information search.](image-url)
Queries Relating to Popular Culture

One-quarter (24%) of queries related to popular culture. As shown in Figure 4, most of these queries related to video games, TV, popular music, movies, and actors, with queries about video games the most common. Search terms about video games were coded as informational if they were for 'cheats' or general queries for types of games (for example 'ps3 games'). It was decided to code search terms consisting of the names of video games as transactional, as all of these queries led to sites where games could be played online or be downloaded. Search queries with the code 'radio' were all for specific radio stations or radio Websites. Search queries that consisted of the name of a cinema were coded as navigational queries, although it is possible that the searcher wanted to buy tickets for movies, that is, conduct a transaction. Most of the popular music queries were for the artist. Where they were for mp3 downloads, they were coded as transactional.

Queries Relating to Ecommerce

One-quarter (24%) of queries related to Ecommerce. As shown in Figure 5, most of the subcategories are fairly self-explanatory. The queries coded 'name of product' included queries where brands were specified. These queries have been included as 'Ecommerce' as it has been assumed that the
searcher was looking for information about these products in the context of making a purchase.

Just under half of the queries coded ‘service’ were general queries for information about a service, for example, ‘limousine hire melbourne.’ The rest were for a specific service provider and were coded as navigational, as it was assumed that the searcher wished to go to the Website of that service provider. Similarly, it was assumed that searchers who typed in the name of a retailer were after the Website of that retailer. Most of the queries coded as banking/finance, buying/selling, and airlines appeared to be for specific sites. Sites at which transactions can occur, such as bank Websites, online trading sites, online booking sites, online gambling sites, and airline Websites were coded as transactional, although, of course, it is possible that the searcher was just visiting these sites for information, rather than to conduct a transaction.

Queries Relating to Cultural Practice

Queries about sport comprised more than one-quarter of the queries about cultural practice (Figure 6). Most of the ‘learning’ queries were for educational institutions, schools, or universities and were assumed to be navigational. Two Australian public holidays, Anzac Day and Easter, occurred during the time period from which the sample was drawn. Queries about these accounted for most of the queries about festivals, events, or holidays. All of the language queries were transactional, as they were all queries for Web translations (for example, ‘German to English’).

Queries Relating to Computing/Web

Figure 7 shows the subcategories of the search queries relating to computing or the Web. Search terms that were the names of software were coded as informational, although, of course, the searcher may have been seeking to download the software. Queries were coded as transactional if the query included the word ‘download,’ for example, ‘free antivirus download,’ or was for an email, chat, or social networking site.

‘Adult’ Queries

One in 12 (8%) of all queries were coded as ‘Adult.’ Almost all of these were for pornography sites. Following Rose and Levinson (2004), these were coded as transactional. The remaining ‘Adult’ queries were for dating sites and these were also coded as transactional.

Queries Relating to Contemporary Issues

As shown in Figure 8, queries from people wanting information on particular contemporary issues accounted for less than 1% of all Web search queries. Queries about government, including departments, programs, and policies, accounted for less than 2% of all Web search queries.

Type of Search Query

Only half (52%) of all queries were informational (Figure 9). 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 were transactional.
FIG. 7. Search queries relating to computing/Web.

FIG. 8. Search queries relating to contemporary issues.


The type of search query was compared across Lifestyle Groups. Again, surprisingly, the distribution of types of search query did not vary significantly across the different Lifestyle Groups (p > 0.01). In other words, when using a search engine each Lifestyle Group was as likely as the others to be looking for information, navigating to a site, or conducting a transaction.

**Topic of Informational Search Queries**

In order to know how people are using search engines to look up information, it is necessary to report separately on the topic of informational searches. Figure 10 shows the topic of those queries where it appeared that the searcher had no particular destination in mind.

It can be seen from Figure 10 that popular culture queries accounted for more than one-quarter (29%) of all 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. These were mainly queries about sport (6% of all searches). 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.”

Ecommerce 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.

Contemporary issues accounted for only 4% of all informational queries. Queries relating to history, genealogy, high culture, and science were separately identified but together accounted for only 4% of informational queries.

**Discussion**

This study reports on the presumed intent of search queries and the topic of informational queries. Given that half of the searches were navigational and one-third of these are transactional, half of search engines users know where they want to go. With regard to informational queries, more than one-quarter of queries (29%) related to popular culture, one-fifth related to Ecommerce, and another one-fifth related to cultural practice.

Figure 11 displays data from previous studies, as well as the current study, on the distribution of informational, navigational, and transactional queries. For comparison purposes, because Broder’s data excludes adult sites, the data from this current study are repeated excluding adult sites. Despite the difficulties in making direct comparisons, it appears that between 1997 and 2009 the proportion of transactional queries first increased and then decreased. It is not possible to know whether the differences are due to changes in search engine use over the years or differences in coding. As already mentioned, the coding into informational, transactional, and navigational involves some fairly subjective decisions about the likelihood that the searcher had a particular Website in mind or wished to undertake a transaction. However, the initial increase then decrease in the proportion of transactional queries does seem plausible. It is likely that the initial increase reflects the increasing acceptance and take-up of Ecommerce as banks offered Internet banking and more secure methods of funds transfer were established (Pearce, 2003). The frequency of downloading is also likely to have increased as Internet speeds got faster. The decrease after 2003 could be a reflection of increasing concern about Internet security (Informa Research Services Inc, 2005). In response to the increased number of phishing or bogus sites set up to steal people’s financial or personal details, Ecommerce sites recommend Internet users to go directly to their sites via a bookmark or typing in a URL (see, for example, http://www.commbank.com.au/security-privacy/netbank-security/security-checklist.aspx and http://www.nab.com.au/wps/wcm/connect/nab/nab/home/Personal_Finance/12/13/9/1 accessed 24 September 2010). Tann and Sanderson (2009) consider that search engine users are increasingly using search engines for navigational purposes, that is, as shortcuts instead of typing in the URL of the particular Websites they want to go to. While the data in Figure 11 do not refute this, it does not provide clear evidence to support it.
While it is difficult to be conclusive about the extent of changes over time in the distribution of informational, navigational, and transactional queries, it is even more difficult to be conclusive about changes over time in the distribution of topics. This is because there is no standard coding scheme for Web search terms and it is sometimes difficult to know what particular researchers have included in their categories. It does appear, however, that over the last 10 years searching for sexual content has remained relatively stable at between 8–12% of all search engine queries. In an overview of 11 search engine query analyses conducted between 1997 and 2002, Hochstotter and Koch (2009) found that the most popular search queries related to sex. They also found this in their 2005 study of queries from four German search engines. Broder (2002) found that one in eight of queries from the Alta Vista log were sexual. In a random sample of queries from search engine AlltheWeb.com (Jansen & Spink, 2005) were related to sex and pornography. In the same study of U.S. queries sent in 2001 to Excite, it was found that 1 in 11 were related to sex and pornography. In the current study 8%, or 1 in 12 queries, were related to sexual content.

In the current study just under one-quarter of all queries related to Ecommerce. This is a similar proportion to that found by Spink, Jansen, Wolfram, and Saracevic (2002) in an analysis of 2001 logs.

In an analysis of a random sample of queries from search engine AlltheWeb.com, Jansen and Spink (2005) classified two in five queries as relating to ‘people, places or things.’ Using another sample, Spink and Jansen (2004b) found that the proportion of search queries for information on ‘people, places and things’ has increased. Spink and Jansen do not describe the content of this category. However, this current research suggests that the queries about people were probably predominantly queries about popular musicians and actors, while the queries about things were probably predominantly queries about products.

Broder (2002) demonstrated that the need which led to a search query may not be informational, but may be navigational or transactional. This current study suggests another significant category of search which is related to leisure. More than one-quarter (29%) of all informational queries related to popular culture and more than 1 in 10 queries (12%) were for the name of a product. Engaging in popular culture can be regarded as a leisure activity for most people. This is why some major Internet surveys code viewing of popular culture sites as entertainment rather than information acquisition (for example, the World Internet Project; Ewing & Thomas, 2010). With regard to product queries, some of the search engine users would have been searching for information about the product or wanting to buy the product. However, it seems reasonable to suppose that a substantial proportion of those queries relating to products can be regarded not as information-seeking but the virtual equivalent of window-shopping, that is, a leisure activity. This highlights the need for a term to distinguish between information search where finding the information is the goal and information search that is an end in itself. I suggest the term “leisure search” to indicate when the searching for information is itself a leisure activity. To a searcher undertaking leisure search, the question “Did you find what you were looking for?” is irrelevant. The phenomenon of leisure search has existed before the Internet; however, the ease of looking up anything on the Internet has facilitated the expansion of leisure search to become a very common activity. It is possibly similar in nature to browsing...
or surfing the Web as a leisure activity. What makes it different from browsing or surfing the Web is that it involves use of a search engine.

‘Undirected viewing’ is a phenomenon identified by Choo, Detlor, and Turnbull (2000) in their integration of a model of browsing with a model of searching. ‘Undirected viewing’ involves the ‘moves’ of viewing Web pages and following links but not in response to any information need. However, Choo et al. (2000) exclude the use of search engines from this category. In Australia, just under 15% of Internet users engage in ‘undirected viewing’ every day. Another quarter of users do so weekly (Ewing, Thomas, & Schiessl, 2008). In an early qualitative study of how the Internet was being incorporated into people’s home lives, Haddon (1999) flagged, without naming it, this phenomenon of ‘leisure search.’ He commented that “‘Searching’ itself could easily slip into a certain amount of ‘browsing’ if something interesting caught the user’s attention.”

The analysis in this current study bears out the suggestion in Waller and McShane (2008) that the small amount of effort expended in using a search engine means that a multitude of additional searches are being undertaken that are not in response to any particular information need but are somewhat akin to idly flicking through a newspaper to pass the time compared to purposeful reading.

If half of the search queries that are product names (half of 12%) and all of the search queries relating to popular culture (29%) are considered to be leisure searches, then leisure searches could have accounted for one-third of the so-called ‘informational’ searches. In other words, only two-thirds of the ‘informational’ searches were searches for information, or one-third of all search engine queries were for information.

The fact that such a high percentage of people are looking for popular culture has not come to light before because large-scale studies of Internet use (such as Pew Research Center’s Internet and American Life Project, Oxford Internet survey, World Internet Project) ask about the use of the Internet to find information on specific subjects (for example, health, science, the news, sports information, humorous content, jobs, local events, travel plans) but do not ask about popular culture. Other studies on topics of search queries have not singled out ‘popular culture’ as a category but have either bundled it together with high culture (for example, Segev & Ahituv, 2010) or have inadvertently hidden it the category ‘people places and things’ (Jansen & Spink, 2005; Spink & Jansen, 2004a). This current study shows the importance of actually looking at what people do and describing it using categories that best describe it rather than using preselected categories.

The very low proportion of queries relating to particular contemporary issues accords with Hindman’s (2009) finding from a U.S. study that very few Internet users seek political information on the Web. Di Maggio, Hargittai, Celeste, and Shafer (2004) list evidence “that users from more privileged backgrounds are more likely to use the Internet to get ahead and equip themselves to participate in community affairs or politics.” However, in the current study the proportion of queries relating to particular contemporary issues was too low to be able to conduct reliable tests of statistical significance across Lifestyle Groups. Of course, Internet users can go directly to Web pages for information and can also use Website-specific search engines, for example, search engines for government information (Chau, Fang, & Liu Sheng, 2007). As Segev and Ahituv (2010) themselves point out, studies of search engine queries do not take into account the visits to Web pages to find particular types of information.

While one of the unique features of this study was that it matched lifestyle information with each search query, it was a surprising finding of this study that there was no statistically significant difference across the search type or most common topics for the Lifestyle Groups. It could be that the distribution of search types is a reflection of the distribution of search topics and that this does not so much reflect the interests of people but the suitability of search engines for looking up particular types of topic. This would suggest then that general search engines are more suitable for looking up information on popular culture and Ecommerce than, say, information on health, computing, or contemporary issues.

As mentioned earlier, while it is an individual who conducts a search, the Lifestyle Group classification characterizes the household. The searches of children are likely to be different from the searches of adults and the likelihood of children belonging to a household was not even across the Lifestyle Groups. Four of the Lifestyle Groups (C: Young Ambition; G: Learners and Earners; J: Suburban Subsistence; and K: Community Disconnect) are less likely than the others to have children as part of the household structure. This confounding factor may also be involved in the lack of statistically significant differences across the subjects most commonly searched.

Conclusion

Although the analysis reported on was of search queries of Australian users, the implications are relevant to Web search research in any country.

This exploratory study contributes to the literature on search engine use and addresses some of the limitations of previous studies on search engine use. In particular, both the sampling and the analysis takes into account the distribution of search queries along the long tail. More significantly, the analysis matches lifestyle information of the searcher with the search query. The finding that lifestyle differences across searcher intent and topic were not observed to be statistically significant was completely unexpected. It is an open question as to whether the distribution of topics reflects the interests of search engine users or the suitability of search engines for looking up particular types of topic, as this study does not take into account people’s visits to particular sites for information. For example, do Internet users tend to go to particular trusted Websites for information on health, computing, and contemporary affairs, rather than use a search engine? Further qualitative research is needed to understand the factors.
affecting whether people use a search engine or go directly to a Website when pursuing a particular topic.

As Broder has shown, search engines use does not necessarily equate with a search for information. Half of the searches in the current study were navigational and one-third of these were transactional. In other words, half of the search engines users know where they want to go. Hence, in order to better understand how people are using search engines to look up information, it is necessary to report separately on the topic of informational searches, as this study has done. In addition, this study suggested that the Internet search engine is not only an interface to information or a shortcut to Websites, it is equally a site of leisure. It is suggested that future qualitative studies be conducted to study this phenomenon more explicitly. Broder’s taxonomy of Web queries could be expanded to comprise informational, navigational, transactional, and leisure searches.

Acknowledgments

This research was funded by Australian Research Council grant LP077215. The author thanks the anonymous reviewers for helpful comments, Laura De Amicis for assistance in sampling and coding, Hitwise for generous support, in particular, Sandra Hanchard, Ben Carroll, and Bill Tancer, as well as the State Library of Victoria and Dr. Denny Meyer.

References


