Running head: RELIGIOUS INFORMATION SEARCHING

Investigating Religious Information Searching through the Analysis of a Search Engine Log

Rita Wan Chik, Mark Sanderson and Paul Clough

University of Sheffield, Sheffield, UK.

RMIT, Melbourne, Australia

Abstract

In this paper, we present results from an investigation of religious information searching based on analyzing log files from a large general purpose search engine. From approximately 15 million queries we identified 146,217 queries from 74,850 user sessions. We present a method for categorizing queries based on related terms and show differences in search patterns between religious searches and web searching more generally. We also investigate in more depth the search patterns found in queries related to five religions: Christianity, Hinduism, Islam, Buddhism, and Judaism. Different search patterns are found to emerge. Results from this study complement existing studies of religious information searching and provide a level of detailed analysis not reported to date. We show, for example, that sessions involving religious-related queries tend to last longer, that the lengths of religious-related queries are greater and that the number of unique URLs clicked is higher, when compared to all queries. The results of the study can serve to provide information on what this large population of users are actually searching for.

1. Introduction

The Internet is being used as a common means of transmitting information of a religious nature (Kinney, 1995; O'Leary, 1996; Casey, 2001; Helland, 2004). It provides material to reinforce the beliefs of those already engaged with their faith as well as information to those exploring different faith communities. Members of religious groups frequently use the Internet to undertake activities, such as online worship and making requests for prayers, seeking advice and information, sharing experiences and opinions, bestowing blessings and reprimands, listening to sermons, social networking and even shopping for religious merchandise (McKenna and West, 2007; Cheong et al., 2009; Campbell, 2005a; Hoover et al., 2004; Helland, 2002; Foltz & Foltz, 2003; Casey, 2001; Larsen & Rainie, 2001). Højsgaard and Warburg (2005) reported that by the year 2004 there were approximately 51 million religious websites on the Internet, disseminating information and communicating with followers. In a US survey conducted by the Pew Research Centre in 2010, it was found that 41% of users from religious or spiritual groups (e.g., members of a church) were also actively using the Internet to access their groups' websites, to involve themselves in spiritual study and practice groups (Rainie et al., 2011).

Studies showed that search engines have become the primary means for many people to fulfil their information needs (Gan et al., 2008) and their use has become a part of peoples' daily routines (Kelly & Ruthven, 2010). Past research has also shown that an important component of online religious activity is searching for religious information (Casey, 2001; Larsen & Rainie, 2001; Ho et al., 2008; Jansen et al., 2009). For example, the study by Jansen et al. (2009) examined religious searching based upon the frequency with which different search query terms – such as 'bible', 'islam', 'catholic', 'jewish', and 'hindu' – appeared in query logs collected between 1997 and 2005 from three major US search engines. They carried out a longitudinal analysis of users'

searching patterns and found that religious searching remained constant over time with activities involving some kind of religious intent making up around 1-1.5% of the sessions analysed.

As suggested by Dawson (2000), in addition to analysing the religious content of information online, it is also important to investigate people's search purposes and processes for religious information. This is also supported by Jansen et al. (2009) who claim that only a few studies have investigated how people search specifically for religious-related information. The study described in this paper continues to investigate searching patterns for religious information and seeks to complement previous work. We examine queries and clicked items from the query log of a major US web search engine originating from users located in the United States during May 2006. An approach is used to filter out religious-related queries from the query log based upon identifying terms related to each religion that may occur when searching. This results in a dataset comprising of 146,217 queries (85,744 unique) within 60,759 user sessions – around 1% of the complete log. We compare the searching patterns for religious-related queries with patterns measured on all queries to gain insights into this type of information searching behaviour. Our study is similar to that of Jansen et al. (2009), but with a broader investigation of information searching patterns for five major world religions: Christianity, Islam, Hinduism, Buddhism, and Judaism. These were selected as being the most practiced religions in the US (The Religious Composition of the United States, 2007): Christianity, including all related denominations having the largest population (78.4%), followed by Judaism (1.7%), Buddhism (0.7%), Islam (0.6%) and Hinduism (0.4%)[1].

The main contribution of this study is an analysis of users' searching behaviour when locating religious-related information. In describing search episodes – specific interactions with an

[1] People unaffiliated to any faith were 16.1% and all others faiths or people who were unwilling to state their faith were cumulatively 2.1%.

Information Retrieval (IR) system – Belkin (1993) suggests four criteria: the goal of the search interaction; the method of interaction; the mode of retrieval; and the type of resources interacted with during the search. In this study we focus on the methods of interaction and type of resources interacted with during the search. Two research questions are addressed in this paper:

- **RQ1**: How do patterns of search behaviour vary for religious-related queries from patterns for all searches?
- **RQ2**: Do patterns of search behaviour vary for searches related to different religions?

This paper is structured as follows: Section 2 reviews previous related studies followed by a description of the data and the methodology used in this research in Section 3. Section 4 presents results from the data analysis, Chapter 5 provides a discussion of results, and the paper concludes in Section 6 with a summary and avenues for future work in this area.

2. Related Work

In this section we discuss two areas of previous relevant literature: previous studies of religious information searching (Section 2.1) and query log analysis (Section 2.2).

2.1 Religious Information Searching

Religion and the Internet

There have been a number of online religion studies (Campbell, 2005; Helland, 2000; Dawson & Cowan, 2004; Karaflogka, 2006) looking at how religions have embraced the Internet and Web technologies (Kluver & Cheong, 2007) and how religions use them in disseminating its beliefs (Wyche et al., 2006; Wyche & Grinter, 2009). Many religious leaders have used Information and Communications Technology to support religious or church activities and they agree that the

Internet is a valuable tool for disseminating religious information (Wyche et al., 2006; Kluver & Cheong, 2007). Studies of online churches and their websites found differences between smaller and larger churches on the website use, design, and assembly of information (Cheong et al., 2009). Some advantages of using these religious websites are that users could watch their favorite sermons online by their preferred pastors and from other churches and locations (Wyche & Grinter, 2009).

Studies of online communities (i.e., members of religious websites or religious forums) found that among the activities undertaken in these online communities are worship, making requests for prayers, seeking advice/information, sharing experiences/opinions, and bestowing blessings or reprimands (McKenna and West, 2007; Foltz and Foltz, 2003; Casey, 2001). One prime benefit of these online religious communities was that members or religious followers were able to express their opinion or ask sensitive questions allowing for open discussions on faith matters and sensitive or controversial issues (Casey, 2001; Foltz & Foltz, 2003).

Religious information seeking

Several Pew studies on Americans and their Internet use have shown an increase of number of users using the Web for religious purposes: 28 million American in 2001[2] to 82 million American in 2004 with 21% having sought information about how to celebrate religious holidays; 17% having looked for information about where they could attend religious services (Hoover, Clark & Rainie, 2004). Results of another Pew survey, of 2,252 adults (18 years and older, taken in 2010), showed that the three most common types of information sought were health information, news, and religious information (Zickuhr, 2010). The seeking of religious information was between

[2] This is more than those who gambled online, used Web auction sites, traded stocks online, placed phone calls on the Internet, conducted online banking, or used Internet-based dating service (Larsen & Rainie, 2001).

26%-35% of all online activities. Other key online activities included emailing, using search engines, buying products, making travel reservations or purchases, doing online banking, rating products/services/people, making online charitable donations, and downloading podcasts.

In a 2002 study carried out by Fox (2002), it was reported that 25% of Americans used search engines to locate religious information. A survey carried out in November and December 2010 found that church organisations and spiritual groups were the most popular groups among Americans where 40% of American adults admitted being active in such groups and 41% of the adults in this group were Internet users (Rainie et al., 2011). This shows an increasing relationship between Internet users, search engines and religion. Neelameghan and Raghavan (2005) recommended improvement for new methods and tools for the exchange of inter-faith and inter-cultural ideas in helping people with different cultural and linguistic background to publish and seek religious and cultural information.

As discussed above, a number of studies were carried out to investigate users' online activities relating to religion. However, little has been done to explore the searching purposes and processes of these users, the experience in locating their religious information needs (Dawson, 2000; Jansen et al., 2009; Pu et al., 2002) and the challenges these users face (Wyche et al, 2006). As suggested by Wang et al. (2003), "*to solve the fundamental problems of Information Retrieval (such as presenting needs effectively and retrieving useful information efficiently), it is important for researchers and designers to understand what the users are searching for, how they search and what problems they encounter*" (p.743).

2. 2 Query Log Analysis

It is common for many online systems to record the interactions between people using the system

and responses from the system itself. These logs offer potentially valuable information for a wide range of applications, such as the design, personalization, and evaluation of systems. The value of identifying and extracting various patterns and trends from transactions logs has long been discussed (Peters, 1993; Wyly, 1996; Blecic et al., 1999), but as more online services exist and more people interact with them, the analysis of log files has become an important research field in its own right (Jansen, 2008; Silvestri, 2009). In the case of Web search, Jansen (2006:409) defines transaction log analysis as: *"the use of data collected in a transaction log to investigate particular research questions concerning interactions among Web users, the Web search engine, or the Web content during searching episodes."* The results of query log analysis enabled a better understanding of how people interact with search engines and have been used to improve retrieval performance and enhance user interaction (Ozmutlu, Ozmutlu & Spink 2009; Taksa, Spink & Jansen, 2009; Jansen, 2008; Amitay and Broder, 2008; Spink, Wolfram, Jansen & Saracevic, 2001).

Query (or search/transaction) log analysis has been used to examine characteristics of searching episodes in order to isolate trends and identify typical interactions between individual searchers and the system. Interactions often include the queries submitted by users, modifications made to queries, patterns of result list viewing and how information objects are used (Jansen, Taksa & Spink, 2008). Query log analysis has been used for many user behaviour studies, such as analysing general Web searching trends (Silverstein et al., 1991; Jansen et al., 1998; Jansen et al., 2000; Jansen & Spink, 2005). Results from such studies have informed our understanding about typical lengths of search query, as well as users' apparent lack of interest in using sophisticated search operators. Others have examined aspects of user behavior, such as repetition of search (Sanderson & Dumais, 2007) or the topics of queries being issued to search engines (Spink et al., 2001; Jansen et al., 2005a; Jansen & Spink, 2006; Nicholas et al., 2007).

One aspect of log analysis that has received less attention is the study of users issuing a particular type of search (Beitzel et al., 2004, Beitzel et al., 2007). Results from limited past work have indicated that differences exist in the queries issued for a particular topical category when compared against queries from another category or the whole query stream under study, which may suggest differences in searching behaviour (Beitzel et al. (2004). An early comparison was a study by Spink et al. (2002), who compared the searching behaviour of European FAST web search engine users, who were mostly Germans, with Excite web search engine users, who were largely American. They found that *"there are some differences in the topics searched and searching behaviors"* between the two groups of users. Table 1 provides a summary of past studies carried out on specific topics.

More recently, Jansen et al. (2009) looked at religious searching. They studied queries containing terms such as 'bible', 'islam', 'catholic', 'jewish' and 'hindu', as well as terms such as 'peace', 'faith', 'hope', and 'love'. They examined logs from different US search engines, including Excite, Alta Vista, and Dogpile, using data collected between 1997 and 2005. They found that religious and religious-related belief searching remained constant despite the US being described as secularised and factionalised. The study mainly looked at term count, modified queries, and session length. Religious searches were found to make up about 1%-1.5% of all the sessions analysed. This paper provides further evidence of differences between queries for religious-related materials and all queries within a search log, as well as exploring differences in search patterns between five major world religions.

3. Methodology

To study patterns of user searching, a large search engine query log (Section 3.1) was analysed that

allowed the study of individual user patterns as well as aggregated patterns of usage. Query log analysis falls under the broader field of transaction log analysis (Jansen et al., 2009: 2). The particular approach used in this study consists of three stages to identify and analyse religious-related queries[3] and their corresponding sessions. The first stage seeks to generate a list of related concepts/terms for each religion (Section 3.2). This list provides a set of concepts which are used to query the dataset to extract religious queries, e.g. 'Christian' for Christianity. Next, the related terms are used to search the log based on partial matching of the queries, e.g. 'Christian' could match queries including 'Christian books' and 'Christian music' together with further filtering (Section 3.3). The resulting list of religious-related queries is then analysed (see Section 3.4). We chose an approach that emphasizes accuracy over coverage for identifying suitable queries for analysis, focusing on queries we were confident would provide differentiation between the five chosen faiths and would be successful at locating religious-related queries.

3.1 Microsoft Search (MSN) Dataset

We used the Microsoft Search log[4] (MSN Dataset) released by Microsoft in 2006 which contains 14,921,285 queries originating from users located in the United States during May 2006 (Craswell et al., 2009). The logs consist of two files: one recording queries with the fields Time, Query String, Query ID, Session ID, and Result Count; the second file containing any clicks associated with a query. This file has the fields Query ID, Query, Time, URL, and Position.

Although it is not explicitly stated how the sessions were identified, it is believed that queries issued within a limited time frame (approximately 20 minutes) from a computer with the IP address, cookie or search engine toolbar ID were grouped into a session. As an effort to avoid

[3] We consider religious-related queries as those where the user must locate religious information in order to fulfil some specific information need.

identification of personal information from the log. If a user returned to the search engine beyond the time frame, a new session ID was issued to the new group of queries.

The MSN Dataset was shown to contain queries from the MSN Search frontend and external sources, such as third-party APIs. This has resulted in some of the longest sessions being machine-driven rather than initiated by humans (Zhang & Moffat, 2006). We followed a similar approach to Zhang and Moffat by filtering out sessions with zero clicks or with more than 100 queries. Although Stamou and Efthimiadis (2009) demonstrate that queries with no click can be intentional (e.g., the goals of the searcher are satisfied by the results list; or the user only wants to determine the existence of a web page), due to the unreliability and difficulty of correctly determining cases where the user has intended not to click on results from sessions that are machine-generated, we removed all sessions with no clicks. Applying this filtering approach reduced the MSN Dataset from 14,921,285 queries (6,623,637 unique) and 7,470,915 sessions to 12,212,723 queries and 5,684,515 sessions (a 24% reduction in the number of sessions).

3.2 Identifying Related Terms

Queries from the MSN Dataset were filtered based on whether or not they contained related terms for each religion. The compilation process of gathering related terms used a modified snowball sampling technique (Patton, 1990) following approaches adopted in previous query log studies to study specific topics of interest (Jansen et al., 2005; Beitzel et al., 2007; Jansen et al., 2009). Related terms were generated from three online lexical resources: (1) Wikipedia Glossary[5]; (2) WordNet 3.1[6]; and (3) the OneLook reverse dictionary[7]. From all resources a list of related terms

[4] Although the license for using the MSN Dataset has now expired the work in this paper began in 2008 when the license was still valid.
[5] http://en.wikipedia.org/wiki/Wikipedia:Glossary
[6] http://wordnetweb.princeton.edu/perl/webwn
[7] http://www.onelook.com/reverse-dictionary.shtml

(words and phrases) for a given word was derived. For example, in the case of OneLook, the dictionary contains a list of related concepts sorted in descending order of 'relatedness', e.g. for 'Christianity' related words were 'cross', 'Christian' and 'church'. In the case of WordNet, for each religion the list of direct and full hyponyms were used as related terms. For the Wikipedia Glossary, all listed terms for a religion were used as related concepts. The related terms were then aggregated across resources into a single list to produce a unique set of terms per religion. To reduce the likelihood of false hits, we manually removed any lines with single ambiguous words that were more likely to be used in a non-religious sense, such as 'king' and 'month'.

Following the creation of lists of related terms, we asked two assessors familiar with each religion to judge the terms for their relatedness to the religion. The judges also added any variants of terms if missing from the list, e.g. 'divali' was added for the entry 'diwali'. Table 2 summarizes the collected related terms. The first column (Num. terms rated) shows the number of terms which both judges rated. The next column shows the inter-rater agreement (percent agreement) between the two judges. The agreement ranges from 91.3% to 96.3% and only terms that both judges agreed on were kept. Any terms that both assessors rated as 'unrelated' were discarded giving a final set of related terms (Final Count in Table 2). The related terms were then used to search the log for matching queries. Not all terms were found in the log, which is shown in the fifth column (Terms with zero hits) in Table 2. This ranged from 25.2% for Christianity to 71% for Hinduism.

Table 3 shows the top 20 related terms based on the number of hits from the MSN Dataset (individually or co-occurring with other query terms). These show the kinds of seed terms generated from the previous approach for a given topic, which in this case is the name of a religion. The related terms capture key aspects of the religions which users of a general-purpose search engine would be likely to search the web for. In some cases the same related term appears in both lists, e.g. 'Yoga' appears in the lists for Buddhism and Hinduism. The reason for this is that the

religions share similar spiritual meditation practices and therefore we would expect to see the related term in each list. We decided to handle related terms ambiguous with respect to religion in this manner, as without further analysis of the context of the query containing the concept, we would not be able to determine which religion the query would most likely be categorized under.

3.3 Extracting 'Religious Queries'

Using the related terms for each religion we extracted all queries from the MSN Dataset containing the terms. Case-folding was performed on the queries to reduce all letters to lower case and related terms could appear anywhere within a query. For example, for the related term 'Christian' we would match the query 'books for the Christian'. However, due to word ambiguity, false matches may appear in the filtered list of queries. For example, for the term 'Christian' we may also extract 'Christian Dior accessories'. To counteract this, we manually inspected all queries in each list with frequency ≥2 occurrences to develop a stopword list. For example, for the list of queries related to Christianity we added 'Christian Dior'. We then used this to filter out any queries which contain the stopword. With filtering, the total number of queries extracted reduced from 184,568 to 146,217 religious-related queries (a 17% reduction for Christianity; 49% Islam; 18% Buddhism; 16% Judaism; 54% Hinduism) demonstrating the need to post-process the dataset. Table 4 shows various statistics for the dataset. In total, around 1% of queries (and sessions) are religious-related, which agrees with the findings of Jansen et al., (2009).

3.4 Data analysis

Typically log data are analysed at varying levels of abstraction ranging from individual terms, to queries and sessions (Jansen, 2006). In this study we focus on analysing searching patterns at the level of the session (Section 4.1), for individual queries (Section 4.2) and the URLs that are clicked on by users which indicate the resources they have accessed (Section 4.3).

Throughout the analysis, the geometric mean is used to compute the average due to a non-parametric distribution of data, such as session and query lengths. For the session-level analysis, sessions are counted for each religion (and all religions) if *at least one* of the queries in the session is religious-related (i.e., includes an appropriate related term). However, for longer sessions it is common to observe task switching, a common pattern of user behaviour (Spink et al., 2008). In this work we have not analysed the degree of religious-related queries within a session and leave this for future investigation.

In the analysis of queries, the most frequent queries are analysed and classified. A bespoke query classification was developed to distinguish types of query rather than topics/subjects or user intent (Jansen & Booth, 2010). This is because we do not categorise particular occurrences for a given query, but rather consider the query type across all instances. To identify user intent for a sample of queries would require further analysis of the logs which we leave for future work. Table 5 shows the set of categories developed for this study. A deductive approach was followed whereby queries were coded into categories and then refined. During this process, queries were compared with Wikipedia categories to help guide the scheme. After settling on an initial set of categories, the scheme was used by another assessor to categories the most frequent queries again, with any differences discussed and query classification revised. Using the initial scheme the two assessors achieved 83% agreement.

4. Results and Discussion

4.1 Analysis of sessions

The queries and clicked URLs form sessions, a unit of activity (which is demarcated in the logs) that can be analysed (Jansen, 2006). Table 6 shows a breakdown of sessions for all data, all religious-related interactions, and for each religion. Overall, after filtering the logs (as described in Section 3.2), there were 5,684,515 sessions in the MSN Dataset with an average length of 2.64 interactions (a query or click) per session. We found that around 1% of sessions were

religious-related, which corresponds almost exactly to the findings of Jansen et al. (2009). For these sessions the average session length was 4.07 interactions per session compared to 2.64 for all sessions in the MSN Dataset. There was also a higher proportion of sessions with multiple queries and clicks: 53.4% for sessions involving religious queries compared to 32.5% for all queries.

A number of past studies have focused on a session-based analysis of user search behaviour through clustering sessions based on features derived from the sessions, such as number of queries and clicks, click entropy and average query length (Chen & Cooper, 2001; Wolfram et al., 2008; Weber & Jaimes, 2011). In our study we grouped sessions into four disjoint classes based on the number of queries and clicks within the session: single query and single click, multiple queries and single click, single query and multiple clicks, and multiple queries and multiple clicks. The proportion of sessions that fell within each type is shown in Table 6. The categories typically define certain patterns of search behaviour, although further analysis is required to produce a more detailed analysis of session-based patterns of user search behaviour. Table 7 shows example sessions for each of the four types of session.

The first type of session, single query and single click, accounts for 47.6% of queries in the MSN Dataset (24.7% for all religious queries) and typically this is seen to relate to searches for specific items, or as defined in the taxonomy of Broder (2002) a navigational query. For a session, this is also similar to the 'navigational user' category used by Weber & Jaimes (2011) to describe behaviour where the user is using a search engine to navigate to URLs, often known to already exist. For example, many of the queries for 'bible gateway' result in a single clicked URL for the website www.biblegateway.com. This could be viewed as a 'successful' search as the user issues no further queries or clicks in the session (Jones et al., 2008). Wolfram et al. (2008) clustered sessions using cluster analysis from the logs of three large web search engines and also found the largest cluster across logs was with brief user interactions.

The second type of session, multiple queries and single click, accounts for 10.8% of all sessions in the MSN Dataset. There are many possibilities for this particular search behaviour which may depend upon where the user finally issues a click within the session: if at the end of a

session then it may indicate a user attempting to formulate a suitable query before finally clicking on the required result (e.g., looking for a specific website but not issuing the 'right' query). Stamou & Efthimiadis (2010) demonstrated that for some queries, users are able to extract information from the snippets of search results and therefore do not click on search results.

The third category of session is single query and multiple clicks, which accounts for 9.1% of all queries. The user typically clicks on various links from the results list, but does not reformulate their query. This could mean that the user is seeking an answer to a specific question and therefore opens several links to locate the answer (perhaps because they are unable to locate an answer from the snippets alone). However, without further reformulations within the session one might assume the user has completed their search task. For example, one example session begins with the query 'passion of the Christ' and results in a click on the link to Internet Movie Database (IMDB) returned by the search engine at rank position 1, followed by a click on the link at rank position 6 to the official homepage of the film (http://www.thepassionofthechrist.com). One might argue that this, again, represents a more navigational form of search where the user clicks on specific web pages for the query. Other forms of search behaviour include more informational queries (e.g., 'discipleship in the bible') where the user performs a subject search resulting in multiple clicks to sources of information that answer the question or need and result in no further queries (presumably because the session is successful). Clicks for this example include bible study resources (http://biblestudycd.com/) and bible references (http://www.bible.org/page.asp?page_id=1245). This is similar to the 'informational user' category used by Weber & Jaimes (2011) to describe behaviour where the user is using a web search engine to find information on a range of topics.

The final category, multiple queries and clicks, is more indicative of a subject search that involves multiple query formulations and browsing of search results. This type of session can be seen as the opposite of a successful known-item search whereby the user perhaps has no clear or defined goal, the search engine is not returning the results required by the user, the search task could reflect a more difficult type of task, or the user may be less experienced. In some cases the

same query is seen in the logs repeatedly which could be the result of browser caching rather than human-generated search activity.

For all religious queries, there is a correlation (*rho=0.596, p<0.01,* 2-tailed) between the number of queries and clicks indicating that typically the more queries a user issues the more URLs they are likely to click. A similar correlation score is obtained between queries and clicks for each of the religions. From Table 6, one can see there is a difference between the types of search patterns for all queries versus religious queries: we see that the overall session length for religious queries is longer, but also that the number of sessions with multiple queries and clicks is also higher (43.4% compared to 24.7% for the MSN Dataset). The proportion of single query and click sessions is also much lower, which could be attributed to the large number of searches for popular websites that occur with high frequency in the MSN Dataset , e.g. 'Google', 'Facebook' and 'Yahoo'. With users searching on a more specific topic (religion) then this behaviour occurs less. Between religions the behaviour is similar, except for Hinduism where we see a trend of higher single query and click sessions and fewer multiple queries and clicks. The analysis of sessions requires further exploration to better establish the various patterns of search behaviour.

4.2 Query analysis

Table 8 shows a breakdown of overall query statistics for various groups of queries: all queries (MSN Dataset), all religious queries (Religious queries) and then each religion. When computing query length we break words by whitespace (e.g., "Jacob's" would be treated as one word). The first observation is that 85% of the religious queries are comprised of those related to Christianity. This is not surprising as the numbers of followers for Christianity is reported to be higher than any other religion in the US. Next, when comparing between religious-related queries and the MSN Dataset we observe that overall religious queries are longer (geometric mean of 3.45 terms compared with 2.02 terms across all queries). This is further demonstrated by the large numbers of

one-word queries in the MSN Dataset when compared to religious queries (36% vs. 7% respectively).

Out of all religions, we observe that queries for Christianity are the longest, followed by Judaism, Islam, Buddhism, and Hinduism. Only 5% of Christianity-related queries are one word in length; much lower than all other religions. Table 8 also shows the number of queries with zero results returned from the search engine. This is based on results before filtering out sessions with >100 queries and zero clicks (the percentage figures are based on the original numbers of queries). For the MSN Dataset this is 7.6%; for religious queries this is lower at 4.8% indicating fewer, possibly, unsuccessful searches. The lowest figures are for Buddhism. Inspecting the queries with zero result pages shows that causes include misspellings (e.g., "freedom chapel assembly of god *amtiyville* ny") and very long queries (e.g., "*but the answer is not as complicated as some people would have us believe. first of all, ownership of historic christianity is not by any particular group or denomination. no religious group can lay exclusive rights upon christ (as some attempted to)*").

Table 9 shows the top 10 most frequent queries for each religion. Statistics about the number of sessions which the query appears in, the average session length (geometric mean) and type of session are also shown. As expected, the most frequent queries are commonly about locating resources (e.g., religious manuscripts and music), central religious figures (e.g., Jesus, Dalai Lama) and particular websites (e.g., 'bible gateway'). Some queries are clearly navigational, e.g. 'bible gateway', seen through the short overall session length (three interactions) and the high number of single query-click sessions (58%). However, some queries are more a mixture of navigational and informational (indicated by multiple queries and clicks), e.g. 'bible'. This suggests the same query can be categorised in multiple ways depending on the context provided by the click-stream patterns.

Table 10 provides some descriptive statistics on the numbers of queries that involve natural-language questions. Pang and Kumar (2011) demonstrate the amount of natural-language

questions which are posed to general purpose web search engines. They demonstrate that questions in web search queries are both prevalent and increasing. Rose and Levison (2004) also included a category 'question-goal' for a query that could be interpreted as a question. Questions were extracted by filtering queries that matched 'what', 'when', 'why', 'how', 'where' and 'who' at the start of the query (a simplified but more reliable question pattern, Pomerantz, 2005). Example questions for Christianity include "what is grace?", "when will Christ return", "Why is Easter important to the church", "How many times prayer is found in the Bible?", "Where is the nearest Catholic Church to Holmes Beach, Anna Maria Island, FL" and "Who is the antichrist". Example questions for Islam include "what is hijab", "when was the prophet Muhammad born", "why are sunnis Shiites and kurds fighting", "how did Islam expand", "where is Mecca located" and "who is Allah". In the MSN Dataset around 0.3% of queries are questions; for religious queries this rose to 0.6%. For all queries the most common type of question type is 'how'; whereas for religious-related queries the most common type is 'what'.

Table 11 shows the results of classifying the 50 most frequent queries for each religion. Overall, the most common type of query in this sample is spiritual practices and observances (19.2%), followed by queries for resources (17.6%) and then about lifestyle and culture (10%). The results highlight differences between religions for types of queries issued in the top 50 most frequent. For example, many of the Christian-related queries are related to denominations and for online resources, such as bibles, bible study notes and sermons. In contrast, many of the Judaism-related queries concern spiritual practices and observances, such as Pentecost, Bar Mitzvah, and information about the Hebrew/Jewish calendar. Many of the Islam-related queries concern issues and questions, such as 'Islam and Christianity' and 'What is Islam?', as well as resources, such as islam.com and 'about Islam'. The Hinduism-related queries focus more on deities and religious figures, which would be expected due to the belief in multiple Hindu gods. Many of the queries relating to spiritual practices and observances for Hinduism are reflected by queries for yoga and meditation (similar to Islamic queries in the same category).

4.3 Clicked URL analysis

The URLs in the MSN Dataset indicate the range of resources requested by users and which are more popular. Table 12 provides a summary of the clicked URLs for all queries, all religious queries, and by each religion. In total, around 1% of all URLs in the MSN Dataset are clicked on following the input of religious-related queries. In total, 40.7% of clicked URLs are unique. However, for religious queries this rises to around 70% indicating a wider range of links are selected. Table 12 also shows the average rank position of clicked URLs (geometric mean) where we observe that clicks for religious-related queries are typically further down the ranked list than for all queries. This is likely due to the high proportion of navigational-type queries in the MSN Dataset as a whole where items at rank position 1 or 2 are selected. Table 12 also shows the proportion of URLs for different domain name types: business (.com), non-commercial (.org), educational (.edu) and network provider (.net). We focus only on US-centric high-level domain name types. A clear difference is seen between the clicked URLs relating to religious-related queries compared to all queries: there are more commercial sites (41.9% compared to 30.0%) and a far greater proportion of non-commercial organisational sites (21.3% compared to 4.3%). The latter result is related to the nature of religious searches: commonly for resources or services provided by religious organisations.

Wikipedia is a popular online resource that is ranked highly for many queries. In Table 12 we find that 0.5% of all URLs clicked on are for Wikipedia pages (42.7% of Wikipedia URLs being unique) compared to 1.4% of clicks for religious-related queries being Wikipedia pages. The most commonly requested page is the landing page for Wikipedia, with the page for May Day being the most frequently accessed page about a specific topic (reflecting the time at which the log snapshot was recorded). Other pages represent wars (e.g., World War II, Vietnam war, Darfur conflict), people (e.g., Jimmy Hoffa, Hitler), events (e.g., Mother's Day, Hurricane Katrina) and places (e.g., Italy, Mexico). Similarly, for all religious queries, 1.4% of clicked results are

Wikipedia pages with the 10 most popular pages being: Hinduism, Jesus, Islam, Hebrew Calendar, Karma, Christianity, Vatican City, Auschwitz concentration camp, Pentecost and Gospel of Judas. Table shows the top 5 most frequently clicked URLs for each religion.

We also assigned the top 10 most frequently clicked URLs for each religion to the same query categories used for classifying queries (Table 2) to identify common types of resources being clicked. The results are shown in Table 14, where for each religion the three most popular categories of URL type are shown. Across all religions, the categories 'deities and religious figures', and 'beliefs and philosophies' are two most popular categories for frequently clicked URLs. This differs from the overall most frequent query types, which were 'spiritual practices and observances' and 'resources'.

5. Discussion

This paper has explored religious-related searching patterns as observed in a large search engine log. Two research questions were posed in Section 1, which we discuss now. The first research question asked whether differences in search behaviour for religious-related queries compared to all searches could be observed from the search log.

Overall, the results showed that the searching trends of religious queries differ from the trends observed for all queries in the MSN Dataset. The average session length (number of queries and clicks) for religious-related queries is 4.07, almost double the average session length for all queries (2.64). The session length for all queries in the MSN Dataset is very similar to figures obtained in previous studies: 2.02 (Silverstein et al., 1999) and 2.84 (Jansen et al., 2000a). This is also confirmed by the greater number of religious-related sessions with multiple queries and clicks (53.4% compared to 32.5% for all queries). One explanation for longer session lengths for religious-related queries could be the use of more query modification actions in those sessions (Jansen et al., 2000a). Another explanation is that in the query log as a whole, a large proportion of sessions consist of a single query and click (47.6% in the entire dataset vs. 24.7% for religious-related queries). This reflects the large number of queries for specific websites/services, such as Google, Yahoo!, EBay, etc. that dominate the most frequent queries (and therefore the

sessions).

With respect to queries, we observe that the average query length for religious-related queries (3.45) is higher than the average query length for all queries (2.02). This figure is also higher than the 2.35 terms reported in (Silverstein et al., 1999). Overall 35% of all queries in the entire dataset consist of one term, compared to only 6% of religious-related queries. Longer queries on average are likely due to the specialized nature of the topic being searched and involving more subject-searching and query refinements, compared to the large number of navigational queries that typically dominate web searching as a whole. For clicked URLs, we see that the proportion of unique URLs clicked for religious-related queries is much higher (63.7%) compared to all URLs clicked (39.7%). This is not reflected in the proportion of unique queries which is similar for all queries in the MSN Dataset (54.2%) compared to religious-related queries (58.9%). This highlights that the URLs being clicked on for religious-related queries tend to be different compared to all queries where the same URLs are being selected in multiple sessions. The proportion of Wikipedia URLs selected is similar and so is the average rank position of items clicked on.

The second research question asked whether differences between search patterns for different religions could be observed in the log. We observe that the religious-related queries are dominated by Christianity-related searches. The frequency of queries and sessions follows the order of Christianity, Judaism, Hinduism, Buddhism and Islam (Table 3). The high frequency of Christianity and Judaism queries/sessions is most likely linked to the higher numbers of followers (and therefore search engine users) for those religions in the US. There are differences between the religions. For example, we observe that Islam has the highest average session length (4.42); whilst Hinduism has the lowest (3.80). Christianity has the longest length of queries on average (3.59); the lowest query length is, again, Hinduism-related queries (2.49). Between religions the behaviour is similar, except for Hinduism where more single query and single click sessions are observed compared to other religions. The length of queries on average for religions is similar, although Christianity-related queries are generally longer and consist of far fewer queries of 1

word in length. Where we do observe more pronounced differences between religions is the subject of queries being searched most frequently. For example, we find that frequent Christian queries are dominated by queries for resources, Buddhism for queries related to lifestyle and culture; Hinduism for queries about spiritual practices and observances, Islam for issues and questions and Judaism for spiritual practices and observances. This is an interesting area for further investigation with different samples of queries. For clicked URLs we observe that Islamic-related queries result in more unique URLs being selected compared to other religions, particularly Christianity.

6. Conclusions and Future Work

In this paper we present results from an investigation of religious information searching based on analysing logs files from a large general purpose Web search engine. From a total of around 15 million queries we identified 146,217 queries from 74,850 user sessions. We presented a method for categorizing queries based on related terms and show differences in search patterns between religious searches and Web searching more generally. We also investigated in more depth the search patterns found in queries related to five religions: Buddhism, Christianity, Hinduism, Islam, and Judaism. Different search patterns were found to emerge. Results from this study complement existing studies of religious information searching and provide a better understanding of user search patterns within this more specialized search domain.

In the future we want to perform a more fine-grained analysis of religious information searching that takes into account more subtle differences between the religious. For example, in the case of Christianity we would analyse data by denomination (e.g., Roman Catholic, Methodist, Baptist Evangelical, Protestant, etc.). As followers of denominations typically have particular views and ideologies we might expect to observe different patterns of searching behaviour. We would like to conduct a more comprehensive classification of queries and clicked items. This would involve sampling the log to attempt to infer user intent along with a more fine-grained

topical analysis of religious queries. Also, developing an automated approach to domain-specific query classification would be highly beneficial and reduce manual effort for large sample sizes. Our current approach for identifying sets of queries related to specialized domains and topics could be automated by developing techniques to generate seed terms for a given concept that could be used to filter out sets of queries. We plan to investigate this. In future work we also aim to investigate in more detail the different types of sessions observed in the query log to provide a better understanding of the different types of sessions that commonly occur in Web searching. The results from this study should be compared with results obtained on query logs obtained from different search engines and at different times throughout the year, as seasonal changes are likely to affect the subject of queries issued by users (e.g., during religious holidays).

Finally, as a method used to study user behavior, query log analysis poses a number of limitations. Although useful in determining what users search; it is not possible to understand *why* particular searches are conducted the way they are, the information needs that users are attempting to fulfill or user's satisfaction with the search results (Kurth, 1993). Commonly, results from analysing query logs are used in conjunction with alternative methods of data collection, such as interviews, focus groups and questionnaires, to fully understand user searching behaviors (Griffiths et al., 2002; Grimes et al., 2007). This augmentation of our log analysis is another plan for future work.

References

ABDUL-KARIM, N. S. & HAZMI, N. R. (2005) Assessing Islamic Information Quality on the Internet: A Case of Information About Hadith. *Malaysian Journal of Library & Information Science, 10*(2), 15.

AMITAY, E. & BRODER, A. (2008) Introduction to Special Issue on Query Log Analysis: Technology and Ethics. *ACM Transactions on the Web, 2*(4), 1-2.

BAHFEN, N. (2008). Processes and Methods Used in Islamic Internet Identity. In the *17th Bienniel Conference of the Asian Studies Association of Australia*, Melbourne. Retrieved 10 September 2009, from

http://arts.monash.edu.au/mai/asaa/nasyabahfen.pdf.

BEITZEL, S. M., JENSEN, E. C., CHOWDHURY, A., FRIEDER, O., & GROSSMAN, D. (2007). Temporal analysis of a very large topically categorized Web query log. *Journal of the American Society for Information Science and Technology*, *58*(2), 166-178.

BEITZEL, S. M., JENSEN, E. C., CHOWDHURY, A., GROSSMAN, D. & FRIEDER, O. (2004). Hourly Analysis of a Very Large Topically Categorized Web Query Log. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp 321-328).

BELKIN, N. J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. *Information retrieval*, *93*, 55-66.

BLECIC, D., DORSCH, J., KOENIG, M., & BANGALORE, N. (1999). A Longitudinal Study of the Effects of OPAC Screen Changes on Searching Behavior and Searcher Success. *College & Research Libraries*, *60*(6), 515–530. Retrieved from http://crl.acrl.org/content/60/6/515.abstract

BRODER, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, *36*(2), 3.

BROUWER, L. (2004) Dutch-Muslims on the Internet: A New Discussion Platform. *Journal of Muslim Affairs, 24*(1), 9.

CAMPBELL, H. (2005a) *Exploring Religious Community Online: We Are One in the Network (Digital Formations)*, New York, Peter Lang Publishing

CASEY, C. A. (2001) Online Religion and Finding Faith on the Web: An Examination on Beliefnet.Org. *Media Ecology Association, 2*, 32-40.

CHEN, H. M., & COOPER, M. D. (2001). Using clustering techniques to detect usage patterns in a Web-based information system. *Journal of the American Society for Information Science and Technology*, *52*(11), 888-904.

CHEONG, P. H., POON, J. P. H., HUANG, S. & CASAS, I. (2009) The Internet Highway and Religious Communities: Mapping and Contesting Spaces in Religion-Online. *The Information Society, 25*(5), 291-302.

CRASWELL, N., JONES, R., DUPRET, G., & VIEGAS, E. (2009). WSCD09: Workshop on Web Search Click Data 2009, February 9, 2009, Barcelona, Spain. ACM. Retrieved from http://research.microsoft.com/en-us/um/people/nickcr/wscd09/

DAWSON, L. L. (2000) Researching Religion in Cyberspace: Issues and Strategies. In Hadden, J. K. & Cowan, D. E. (Eds.) *Religion and the Social Order.* London, Elsevier Science Inc.

DAWSON, L. L. & COWAN, D. E. (2004) *Religion Online: Finding Faith on the Internet,* London, Routledge

ESS, C., KAWABALTA, A. & KUROSAKI, H. (2007). Cross-Cultural Perspectives on Religion and Computer-Mediated Communication. *Journal of Computer-Mediated Communication, 12*(3), 939-955.

FOLTZ, F. & FOLTZ, F. (2003) Religion on the Internet: Community and Virtual Existence. *Bulletin of Science, Technology & Society, 23*(4), 321-330.

FOX, S. (2002) Search Engines. Pew Internet Project Data Memo. *Pew Internet and American Life Project*. Retrieved September 10, 2009 from: http://www.pewinternet.org/reports/toc.asp.

GAN, Q., ATTENBERG, J., MARKOWETZ, A., & SUEL, T. (2008). Analysis of geographic queries in a search engine log (pp. 49–56). ACM New York, NY, USA.

GRIFFITHS, J. R., HARTLEY, R. J., & WILLSON, J. P. (2002). An improved method of studying user-system

interaction by combining transaction log analysis and protocol analysis. *Information Research*, *7*(4), 7-4.

GRIMES, C., TANG, D. & RUSSELL, D. M. (2007) Query logs alone are not enough. *In*: *Proceedings of the 16th World Wide Web Conference, WWW 2007, May 8-12, 2007, Banff, Alberta, Canada*, ACM.

HELLAND, C. (2000) Online-Religion/Religion-Online and Virtual Communitas. In Hadden, J. & Cowan, D. E. (Eds.) *Religion on the Internet: Research Prospects and Promises*. London, JAI Press/Elsevier Science.

HELLAND, C. (2002) Surfing for Salvation. *Elsevier Science*, *32*, 293-302.

HO, S. S., LEE, W. P. & HAMEED, S. S. (2008) Muslim Surfers on the Internet: Using the Theory of Planned Behaviour to Examine the Factors Influencing Engagement in Online Religious Activities. *New Media & Society*, *10*(1), 93-113.

HØJSGAARD, M. T. & WARBURG, M. (2005) *Religion and Cyberspace*, london, Routledge

HOOVER, S., CLARK, L. S. & RAINIE, L. (2004) Faith Online: 64% of Wired Americans Have Used the Internet for Spiritual or Religious Information. *Pew Internet and American Life Project*. Retrieved February 17, 2009 from

HUURNINK, B., HOLLINK, L., van den HEUVEL, W. & de RIJKE, M. (2010). Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the American Society for Information Science and Technology*, *61*(6), 1180–1197.

ISMAIL, Y. & MHD SARIF, S. (2007) The Coverage of Islamic Management Materials in the Internet Search Engines. *Conference on Management from Islamic Perspectives (ICMIP).* Kuala Lumpur, Malaysia. Retrieved 10 March 2009, from http://www.islamicmanagement.org/files/resources/1199594858.doc.

JANSEN, B. J. (2008). The Methodology of Search Log Analysis. (2008). Handbook of research on Web log analysis. Hershey, PA: IGI Publishing.

JANSEN, B. J. (2006) Search log analysis: What it is, what has been done, how to do it. *Library & Information Science Research*, 28, 407-432.

JANSEN, B. J. & BOOTH, D. (2010). Classifying web queries by topic and user intent. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 4285-4290.

JANSEN, B. J., TAPIA, A. & SPINK, A. (2009) Searching for Salvation: An Analysis of Us Religious Searching on the World Wide Web. *Religion*, *40*(1), 39-52.

JANSEN, B. J., CIAMACCA, C. C. & SPINK, A. (2008a) An Analysis of Travel Information Searching on the Web. *Information Technology &Tourism*, *10*(2), 101-118.

JANSEN, B. J., TAKSA, I. & SPINK, A., (2008b). Research and Methodological Foundations of Transaction Log Analysis. Handbook of research on Web log analysis. Hershey, PA: IGI Publishing.

JANSEN, B. J. & SPINK, A. (2006) How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. *Information Processing and Management*, *42*, 248-263.

JANSEN, B. J. & SPINK, A. (2005) An Analysis of Web Searching by European AlltheWeb. Com Users. *Information Processing and Management*, *41*(2), 361-381.

JANSEN, B. J., JANSEN, K. J., & SPINK, A. (2005). Using the web to look for work: Implications for online job seeking and recruiting. *Internet Research*, *15*(1), 49–66.

JANSEN, B. J., SPINK, A. & PEDERSEN, J. (2005) A Temporal Comparison of Altavista Web Searching. *Journal of the American Society for Information Science and Technology*, *56*(6), 559-570.

JANSEN, B. J., SPINK, A. & PEDERSEN, J. (2003). An Analysis of Multimedia Searching on Altavista. In

Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval (pp 192).

JANSEN, B. J., SPINK, A. & SARACEVIC, T. (2000a) Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web. *Information Processing and Management, 36*, 207-227.

JANSEN, B. J., GOODRUM, A. & SPINK, A. (2000b) Searching for Multimedia: Video, Audio, and Image Web Queries. *World Wide Web, 3*(4), 249–254.

JANSEN, M. B. J., SPINK, A., BATEMAN, J. & SARACEVIC, T. (1998) Real Life Information Retrieval: A Study of User Queries on the Web. *ACM SIGIR, 32*(1), 5-17.

KARAFLOGKA, A. (2006) *E-Religion a Critical Appraisal of Religious Discourse on the World Wide Web,* London, Equinox.

KASMANI, M. F., BUYONG, M. & MAHYUDDIN, M. K. (2009) Dakwah Content and Its Method: An Analysis on Islamic Websites. *YADIM, 11*. Retrieved 6 March 2009, from http://www.surrey.ac.uk/politics/research/documents/CP-FaizalKasmani.pdf.

KELLY, D., & RUTHVEN, I. (2010). Search procedures revisited. In D. Kelly, I. Ruthven, F. Astrom, B. Larsen, & J. W. Schneider (Eds.), *The Janus Faced Scholar - a Festschrift in Honour of Peter Ingwersen* (pp. 59–67). Retrieved from http://strathprints.strath.ac.uk/32874/1/pif_online.pdf

KINNEY, J. (1995) Net Worth? Religion, Cyberspace and the Future. *Futures, 27*(7), 763-776.

KLUVER, R. & CHEONG, P. H. (2007) Technological Modernization, the Internet, and Religion in Singapore. *Journal of Computer Mediated Communication-Electronic Edition-, 12*(3), 1122.

KURTH, M. (1993). The limits and limitations of transaction log analysis. *Library Hi Tech, 11*(2), 98-104.

LARSEN, E. & RAINIE, L. (2001) Cyberfaith: How Americans Pursue Religion Online. *December, 23*, 21.

LARSSON, G. (2005) The Death of a Virtual Muslim Discussion Group. *Online – Heidelberg Journal of Religions on the Internet, 1*(1), 18.

MCKENNA, K. Y. A. & WEST, K. J. (2007) Give Me That Online-Time Religion: The Role of the Internet in Spiritual Life. *Computers in Human Behavior, 23*, 942-954.

NEELAMEGHAN, A. & RAGHAVAN, K. S. (2005) An Online Multi-Lingual, Multi-Faith Thesaurus: A Progress Report on F-Thes. *Webology, 2*(4).

NICHOLAS, D., HUNTINGTON, P., JAMALI, H. R. & DOBROWOLSKI, T. (2007) Characterising and Evaluating Information Seeking Behaviour in a Digital Environment: Spotlight on the 'Bouncer'. *Information Processing and Management, 43*(4), 1085-1102.

O'LEARY, S. D. (1996) Cyberspace as Sacred Space: Communicating Religion on Computer Networks. *Journal of the American Academy of Religion, 64*(4), 781-808.

OZMUTLU, S., OZMUTLU, H. C. & SPINK, A. (2009) From Analysis to Estimation of User Behavior." *Handbook of Research on Web Log Analysis.* IGI Global,. 206-226.

PANG, B., & KUMAR, R. (2011, June). Search in the Lost Sense of "Query": Question Formulation in Web Search Queries and its Temporal Changes. In*Proceedings of the Association for Computational Linguistics (ACL).*Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods* (Second.). California: Sage.

PETERS, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech, 11*(2), 41–66.

POMERANTZ, J. (2005). A linguistic analysis of question taxonomies. *Journal of the American Society for Information Science and Technology*, *56*(7), 715–728.

PU, H.-T., CHUANG, S.-L., & YANG, C. (2002). Subject categorization of query terms for exploring Web users' search interests. *Journal of the American Society for Information Science and Technology*, *53*(8), 617-630.

RAINE, L., PURCELL, K., & SMITH, A. (2011).The social side of the internet. Pew Research Center's Internet & American Life Project, January 2011. Retrieved November 20, 2011, from http://pewinternet.org/reports/2011/The-Social-Side-of-the-Internet.aspx

ROSE, D. E., & LEVINSON, D. (2004). Understanding user goals in web search. *Proceedings of the 13th conference on World Wide Web - WWW '04* (p. 13). New York, New York, USA: ACM Press.

SANDERSON, M. & DUMAIS, S. (2007) Examining Repetition in User Search Behavior. *Lecture Notes in Computer Science*, *4425*, 597-604.

SANDERSON, M. & KOHLER, J. (2004). Analysing Geographic Queries. In *Proceedings of the Workshop on Geographic Information Retrieval, SIGIR.*

SILVERSTEIN, C., MARAIS, H., HENZINGER, M., & MORICZ, M. (1999). Analysis of a Very Large Web Search Engine Query Log. In *ACM SIGIR Forum, 33* (1).

SILVESTRI, F. (2010). Mining query logs: Turning search usage data into knowledge. *Foundations and Trends in Information Retrieval*, *4*(1—2), 1-174.

SPINK, A., JANSEN, B. J., WOLFRAM, D. & SARACEVIC, T. (2002) From E-Sex to E-Commerce: Web Search Changes. *Computer*, *35*(3), 107-109.

SPINK, A., WOLFRAM, D., JANSEN, M. B. J. & SARACEVIC, T. (2001) Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science and Technology*, *52*(3), 226-234.

STAMOU, S., and EFTHIMIADIS, E.N. (2010). Interpreting user inactivity on search results. In Proceedings of the 32nd European conference on Advances in Information Retrieval (ECIR'2010), Cathal Gurrin, Yulan He, Gabriella Kazai, Udo Kruschwitz, and Suzanne Little (Eds.). Springer-Verlag, Berlin, Heidelberg, 100-113.

TAKSA, I., SPINK, A., & JANSEN, B. J. (2009). Web Log Analysis: Diversity of Research Methodologies. In B. J. Jansen, A. Spink, & I. Taksa (Eds.), *Handbook of Research on Web Log Analysis*. IGI Global.

WANG, P., BERRY, M. W., & YANG, Y. (2003). Mining longitudinal Web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology*, *54*(8), 743–758.

WEBER, I., & JAIMES, A. (2011). Who uses web search for what. *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11* (p. 15). New York, New York, USA: ACM Press.

WENK, E. (1999) The Hands of God: Spiritual Values for Coping with Technology in the 21st Century. *Technology in Society*, *21*(4), 427-437.

WYCHE, S. P. & GRINER, R. E. (2009). Extraordinary Computing: Religion as a Lens for Reconsidering the Home. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp 749-758).

WYCHE, S. P., HAYES, G. R., HARVEL, L. D. & GRINTER, R. E. (2006). Technology in Spiritual Formation: An Exploratory Study of Computer Mediated Religious Communications. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp 208).

ZHANG, Y., & MOFFAT, A. (2006). Some observations on user search behavior. In *Proceedings of the 11th Australasian Document Computing Symposium*, 11, 1-8.

ZICKUHR, K. (2010). Generations 2010. *Pew Internet & American Life Project.* Retrieved November 18, 2012 from http://pewinternet.org/~/media//Files/Reports/2010/PIP_Generations_and_Tech10.pdf

The Religious Composition of the United States. Pew Forum on Religion & Public Life / U.S. Religious Landscape Survey. (2007). Retrieved 09 September 2009, from http://religions.pewforum.org/pdf/report-religious-landscape-study-chapter-1.pdf

Table 1. Topic-based Query Logs Analysis studies

Topics	Researcher(s)
Multimedia information searching	Jansen, Goodrum & Spink (2000); Jansen, Spink & Pedersen (2003)
Geographical information searching	Sanderson & Kohler (2004); Gan, Attenberg, Markowetz, & Suel (2008)
Job-related information searching	Jansen, Jansen, & Spink (2005)
Travel information searching	Jansen, Ciamacca, & Spink (2008b)
Audiovisual information searching	Huurnink, Hollink, van den Heuvel, & de Rijke (2010)

Table 2 Agreement between judges for related terms for each religion.

	Num. terms rated	Percent agreement	Final count	Terms with zero hits
Buddhism	333	304 (91.3%)	197	110 (55.8%)
Christianity	481	463 (96.3%)	292	66 (25.2%)
Hinduism	616	568 (92.2%)	566	402 (71.0%)
Islam	729	676 (92.7%)	466	304 (65.2%)
Judaism	352	329 (93.5%)	194	79 (40.7%)

Table 3. Top 20 most frequently matched related terms for each religion (ranked in descending order of frequency in the MSN Dataset).

Buddhism	Christianity	Hinduism	Islam	Judaism
yoga	church	yoga	Islam	jewish
zen	Christian	soma	muslim	israel
karma	Bible	hindu	Medina	holocaust
buddhism	catholic	maya	prophet	hebrew
buddha	baptist	avatar	Pillar	zion
nirvana	God	meditation	Mosque	jerusalem
sutra	saint	karma	Quran	tabernacle
buddhist	ymca	Chakra	Allah	kosher
lama	jesus	hinduism	Iman	auschwitz
tao	cross	tantra	Koran	judaism
guru	Gospel	rama	Mecca	pentecost
tantra	Angel	guru	Jihad	jew
dharma	christ	kama	halal	rabbi
mandala	methodist	rajah	Sunni	Mitzvah
aryan	christmas	Jap	Madina	Torah
dalai	lutheran	yogi	Masjid	shalom
taoism	temple	krishna	Imam	sabbath
lhasa	Trinity	muni	Ramadan	passover
hum	chapel	dharma	Hijab	bar mitzvah
mantra	Grace	mandala	Hajj	yiddish

Table 4. Number of queries and sessions extracted for all queries (MSN Dataset), all religions and each religion.

	Num. queries	Num. unique queries	% unique queries	Num. sessions
MSN Dataset	12,212,723	6,623,637	54.2%	5,684,515
Religious queries	146,217	85,744	58.9%	60,759
Buddhism	4,643	2,797	60.2%	4,973
Christianity	124,786	75,064	60.1%	53,304
Hinduism	6,298	3,997	63.5%	2,205
Islam	2,660	1,806	67.9%	3,665
Judaism	7,830	4,994	63.8%	2,886

Table 5. Categories derived for analysing the most frequent queries.

Category	Description	Example queries
Resources	Queries being used to locate resources of any media type (e.g. text, image or music)	bible,bible.com, Christian music, gospel lyrics, niv bible, sermon central, holocaust pictures, angel pictures, books on hindu mythology
Lifestyle and culture	Queries being used to find out about lifestyle and cultural issues with a religious nature	yoga and burning calories, yoga tee shirt, Christian café (dating website), Hebrew names, cross tattoos
Issues and questions	Queries that deal with issues and seeking some kind of answer	Buddhism vs. Hinduism, suffering in Buddhism,
Organisations - denomination	Queries relating to religious organisations, specifically denominations	salvation army, catholic church, ymca
Organisations - general	Queries to locate a list of organisations, perhaps within a specific geographical region	catholic charities, yoga in Eugene Oregon, karma sutra club,
Organisations - specific	Queries to locate a single organisation located at specific location (e.g. physical entity)	yeshiva university, barnes jewish hospital, vatican
Location	Geographical queries that relate to specific places	Jerusalem,
Theological concept	Queries that relate to theological concepts	devil, hell, angel, creation, post-modernism, trinity, faith
Religious symbol	Queries that relate to religious symbols	cross, alter, statues of Buddha,
Beliefs and Philosophies	Queries that relate to religious beliefs and philosophies	jewish religion, tao, Christianity, catholic saints
Deities and religious figures	Queries that relate to deities and religious figures or leaders	God, Jesus, hindu gods, vishna, muni, dalai lama, buddha
Spiritual practices and observances	Queries that relate to spiritual practices and observances (i.e. things you might do when following a religion)	prayer, yoga, yoga positions, bible study, bar mitzvah, Pentecost, Passover, jewish calendar, bar mitzvah dance, meditation, yoga poses, tantra
Historical event	Queries that relate to historical events	holocaust, holocaust survivors, holocaust timeline, wandering jew, Israel history

Table 6. Breakdown of sessions and types of session based on the number of queries and clicks.

	Num. sessions	Av. session length (geo mean)	Single Click		Multiple Clicks	
			Single query	Multiple queries	Single query	Multiple queries
MSN Dataset	5,684,518	2.64	2,706,210 – 47.6%	612,349 – 10.8%	519,797 – 9.1%	1,846,162 – 32.5%
Religious queries	60,759	4.07	14,989 – 24.7%	7,659 – 12.6%	5,661 – 9.3%	32,450 – 53.4%
Buddhism	2,205	4.18	509 – 23.1%	248 – 11.2%	255 – 11.6%	1,193 – 54.1%
Christianity	53,304	4.10	13,288 – 24.9%	6,726 – 12.6%	4,873 – 9.1%	28,417 – 53.3%
Hinduism	2,886	3.80	637 – 22.1%	380 – 13.2%	294 – 10.2%	1,575 – 54.6%
Islam	1,228	4.42	238 – 19.4%	142 – 11.6%	125 – 10.2%	723 – 58.9%
Judaism	3,665	4.28	822 – 22.4%	439 – 12.0%	386 – 10.5%	2,018 – 55.1%

Table 7. Example sessions types based on the number of queries and clicks.

Single click	**Single query**	(Q_1) bible gateway →(C_1) http://www.biblegateway.com (Q_1) living waters tabernacle →(C_1) http://www.livingwaterstabernacle.com/html/sermons.html
	Multiple queries	(Q_1) Jacob's absolute confidence that God will take his descendants back to Canaa (Q_2) Jacob's confidence God take his descendants back Canaan (Q_3) Genesis 47 Jacob's confidence God take his descendants back Canaan → (C_1) http://ccel.org/w/wesley/notes/notes/Genesis.html (Q_1) jewish community center (Q_2) jewish community center of metropolitan detroit →(C_1) http://www.jccdet.org/
Multiple clicks	**Single query**	(Q_1) passion of the Christ → (C_1) http://www.imdb.com/title/tt0335345/ → (C_2) http://www.thepassionofthechrist.com (Q1) discipleship in the bible → (C_1) http://biblestudycd.com/ → (C_2) http://godsquad.com/squadroom/discipleship/topics/studies.htm → (C_3) http://www.elca.org/Evangelism/dailydiscipleship/index.html → (C_4) http://www.bible.org/page.asp?page_id=1245
	Multiple queries	(Q_1) women's ministry bible studies free (Q_2) bible studies for women → (C_1) http://www.womeninchrist.org/ → (C_2) http://www.nebible.org/womensbiblestudies.htm → (C_3) http://www.wmcc.net/studies.htm → (C_4) http://www.awpministries.org/BibleStudies.htm (Q_3) bible studies for women on Esther → (C_5) http://www.balaams-ass.com/journal/homemake/thought.htm

Table 8. Query statistics for all queries (MSN Dataset), all religious queries and each religion.

	Total number of queries	Average query length (words)			Percent queries of length 1, 2 and 3 words			Num. queries with zero results (non-filtered)
		Geometric mean	Median	Mode	1	2	3	
MSN Dataset	12,212,723	2.02	2	1	35%	27%	18%	1,133,178 – 7.6%
Religious queries	124,422	3.45	4	3	6%	18%	25%	6,827 – 4.8%
Buddhism	3,900	2.58	3	2	15%	31%	26%	194 – 4.2%
Christianity	106,767	3.59	4	3	4%	17%	25%	5,607 – 4.4%
Hinduism	4,973	2.49	3	2	18%	30%	24%	394 – 6.3%
Islam	2,141	2.71	3	3	17%	24%	24%	185 – 7.0%
Judaism	6,641	2.89	3	2	12%	25%	23%	447 – 5.7%

Table 9. Most 10 frequent queries for each religion (ranked by query frequency multiplied by number of sessions).

Query	Query freq. (num. sessions)	Avg. session length	Single query		Multiple queries	
			Single click	Multiple clicks	Single click	Multiple clicks
Buddhism						
yoga	122 (106)	5.1	22.6%	7.5%	15.1%	54.7%
nirvana	75 (62)	6.6	16.1%	4.8%	12.9%	66.1%
karma sutra	102 (32)	5.2	21.9%	28.1%	0.0%	50.0%
buddhism	55 (45)	5.6	26.7%	11.1%	8.9%	53.3%
buddha	29 (22)	6.6	9.1%	9.1%	27.3%	54.5%
karma	24 (19)	4.7	31.6%	5.3%	10.5%	52.6%
taoism	22 (16)	9.4	6.2%	12.5%	0.0%	81.2%
mandala	25 (12)	11	0.0%	25.0%	8.3%	66.7%
dharma	18 (16)	6.2	6.2%	18.8%	12.5%	62.5%
tao	17 (15)	6.2	13.3%	0.0%	20.0%	66.7%
Christianity						
ymca	625	4.7	31.5%	2.1%	12.2%	54.2%
bible	535	3.9	38.0%	9.4%	11.8%	40.8%
salvation army	309	4.7	24.3%	4.8%	20.2%	50.7%
bible.com	227	3.2	50.9%	5.1%	7.5%	36.4%
bible gateway	188	3.1	59.7%	5.7%	3.4%	31.2%
jesus	140	8.2	7.3%	8.2%	17.3%	67.3%
bible verses	136	4.8	15.2%	26.8%	6.2%	51.8%
online bible	110	4	35.4%	10.4%	6.2%	47.9%
christian music	118	7.1	13.8%	6.2%	12.5%	67.5%
angel	97	8.5	8.2%	2.4%	27.1%	62.4%
Hinduism						
yoga	122	5.1	22.6%	7.5%	15.1%	54.7%
rajah	64	5.1	24.5%	2.0%	4.1%	69.4%
avatar	58	6.1	8.2%	4.1%	32.7%	55.1%
soma	55	6.1	11.4%	2.3%	13.6%	72.7%
hinduism	47	5.9	18.6%	16.3%	7.0%	58.1%
karma sutra	51	5.2	21.9%	28.1%	0.0%	50.0%
aquila	37	3.9	35.5%	3.2%	9.7%	51.6%
meditation	33	5.1	21.4%	7.1%	14.3%	57.1%
maya	30	9.1	3.7%	7.4%	14.8%	74.1%
aum	29	3.7	45.8%	0.0%	12.5%	41.7%
Islam						
islam	60	5.8	25.9%	5.6%	7.4%	61.1%
koran	26	5.2	17.4%	21.7%	0.0%	60.9%
pillar	20	7.2	17.6%	5.9%	17.6%	58.8%
iman	18	5.5	26.7%	13.3%	13.3%	46.7%
quran	16	5.4	20.0%	13.3%	6.7%	60.0%
mufti muneer	14	5.9	23.1%	15.4%	7.7%	53.8%
muslim	14	11.1	7.7%	0.0%	0.0%	92.3%
muslim names	12	5	20.0%	10.0%	0.0%	70.0%
nation of islam	12	4	55.6%	0.0%	0.0%	44.4%
mosque	10	11.3	11.1%	0.0%	11.1%	77.8%
Judaism						
holocaust	237	4.6	31.6%	5.3%	15.8%	47.4%
israel	78	6.7	16.7%	3.3%	11.7%	68.3%
pentecost	46	4	33.3%	12.5%	16.7%	37.5%
auschwitz	44	5.8	23.8%	9.5%	4.8%	61.9%
the holocaust	43	22.8	0.0%	0.0%	0.0%	100.0%
holocaust pictures	34	6.2	4.3%	26.1%	4.3%	65.2%
judaism	24	5.8	26.1%	8.7%	8.7%	56.5%
hebrew	27	5.7	17.6%	11.8%	0.0%	70.6%
jerusalem	22	6.5	15.8%	21.1%	5.3%	57.9%
zion	18	5.2	12.5%	6.2%	18.8%	62.5%

Table 10. Query statistics for natural-language question queries.

	Total num. queries	What	When	Why	How	Where	Who
All queries	12,212,723	30,522	7,512	3,779	61,087	7,793	9,461
Religious queries	124,422	704	130	127	445	108	218
Buddhism	3,900	30	4	4	23	5	1
Christianity	106,767	627	127	112	415	104	226
Hinduism	4,973	42	0	4	34	6	4
Islam	2,141	44	5	8	13	3	8
Judaism	6,641	68	15	13	24	9	4

Table 11. Percentage of query categories for the 50 most frequent queries (highest values by column highlighted in bold).

Category	Percentage of queries by category per religion					Total queries per category
	Buddhism	Christianity	Hinduism	Islam	Judaism	
Resources	12%	**40%**	6%	20%	10%	17.6%
Lifestyle and culture	**20%**	4%	16%	2%	8%	10.0%
Issues and questions	6%	0%	6%	**24%**	2%	7.6%
Organisations - denomination	2%	18%	2%	6%	0%	5.6%
Organisations - general	6%	4%	2%	0%	4%	3.2%
Organisations - specific	0%	4%	0%	8%	10%	4.4%
Location	0%	0%	0%	2%	8%	2.0%
Theological concept	6%	8%	12%	10%	0%	7.2%
Religious symbol	6%	2%	2%	0%	2%	2.4%
Beliefs and Philosophies	14%	0%	10%	4%	10%	7.6%
Deities and religious figures	12%	10%	18%	8%	4%	10.4%
Spiritual practices and observances	16%	10%	**26%**	16%	**28%**	**19.2%**
Historical event	0%	0%	0%	0%	14%	2.8%
	100%	100%	100%	100%	100%	100.0%

Table 12. Summary of clicked URLs.

	Num. URLs clicked - % unique	Num. Wikipedia URLs - %unique	% of specific domain name types				Avg. clicked rank position
			.com	.net	.org	.edu	
ll queries	12,224,181 – 40.7%	61,746 – 42.7%	30.0%	2.9%	4.3%	1.1%	1.98
eligious queries	134,383 – 70.0%	1,862 – 42.1%	41.9%	4.7%	21.3%	1.8%	2.49
Buddhism	4,656 – 71.1%	122 – 50.0%	52.5%	5.4%	10.2%	1.8%	2.76
Christianity	113,829 – 62.2%	1,168 – 46.5%	35.6%	5.5%	19.0%	1.8%	2.45
Hinduism	5,676 – 74.0%	118 – 51.7%	54.3%	5.5%	10.9%	2.0%	2.66
Islam	2,349 – 75.7%	83 – 65.0%	43.9%	7.2%	21.2%	2.9%	2.59
Judaism	7,693 – 70.0%	282 – 43.6%	38.0%	4.7%	22.5%	2.8%	2.72

Table 13. Top 10 most frequently clicked URLs for each religion.

Buddhism	**Christianity**
http://www.spaceandmotion.com/karma-sutra-positions.htm	http://www.bible.com/
http://www.yoga.com/	http://www.ymca.net/
http://www.spaceandmotion.com/karma-sutra.htm	http://biblegateway.com/
http://www.yogajournal.com/	http://www.salvationarmyusa.org/
http://yoga.about.com/	http://www.lds.org/
http://www.yogasite.com/	http://bible.gospelcom.net/bible/
http://users.forthnet.gr/ath/nektar/kma/main.htm	http://bible.org/
http://www.tantra.com/	http://www.blueletterbible.org/
http://www.karmasutramusic.com/	http://www.christian-lyrics.net/
http://www.music.msn.com/artist/?artist=16074476	http://www.sermoncentral.com/
Hinduism	**Islam**
http://rajah.com/	http://islam.about.com/
http://en.wikipedia.org/wiki/Hinduism	http://en.wikipedia.org/wiki/Islam
http://www.yoga.com/	http://www.hti.umich.edu/k/koran/
http://www.spaceandmotion.com/karma-sutra-positions.htm	http://www.islam.com/
http://www.aquila.com/	http://www.kitabummuneer.com/info.asp
http://www.yogajournal.com/	http://www.friesian.com/islam.htm
http://yoga.about.com/	http://www.muslim-names.co.uk/
http://www.yogasite.com/	http://www.pillarmusic.com/
http://www.tantra.com/	http://www.carm.org/islam.htm
http://www.yogasite.com/	http://www.i-iman.com/
Judaism	**MSN Dataset**
http://www.ushmm.org/	http://www.yahoo.com/
http://history1900s.about.com/library/holocaust/blholocaust.htm	http://www.myspace.com/
http://www.ushmm.org/wlc/en/	http://www.google.com/
http://www.jewfaq.org/alephbet.htm	http://www.ebay.com/
http://en.wikipedia.org/wiki/Hebrew_calendar	http://www.aol.com/
http://www.holocaustsurvivors.org/	http://www.mapquest.com/
http://www.holocaust-history.org/	http://mail.yahoo.com/
http://en.wikipedia.org/wiki/Holocaust	http://hotmail.com/
http://history1900s.about.com/library/holocaust/blpictures.htm	http://www.bankofamerica.com/
http://holocaust.about.com/	http://www.walmart.com/

Table 14. Three most common types of URL for 10 most clicked URLs by religion.

Religion	**Categorised URLs**
Buddhism	spiritual practices and observances, beliefs and philosophies, deities and religious figures
Christianity	resources, organisations-denomination, deities & religious figures
Hinduism	spiritual practices and observances, deities and religious figures, beliefs & philosophies
Islam	resources, lifestyle and culture, organisation-denomination, beliefs & philosophies,
Judaism	historical event, organisations-specific, lifestyle and culture