

QUT Digital Repository:
<http://eprints.qut.edu.au/>



Tjondronegoro, Dian W. and Spink, Amanda H. and Jansen, Bernard J. (2009) *A study and comparison of multimedia Web searching : 1997-2006*. Journal of the American Society for Information Science and Technology, 60(9). pp. 1756-1768

© Copyright 2010 ASIS&T

A STUDY AND COMPARISON OF MULTIMEDIA WEB SEARCHING: 1997-2006

Dian Tjondronegoro & Amanda Spink
Faculty of Information Technology
Queensland University of Technology
GPO Box 2434, Brisbane, QLD 4001.
dian@qut.edu.au, ah.spink@qut.edu.au

Bernard J. Jansen
College of Information Science and Technology
The Pennsylvania State University
State College PA 16802 USA
jjansen@psu.edu

ABSTRACT

Searching for multimedia is an important activity for users of Web search engines. Studying user's interactions with Web search engine multimedia buttons, including image, audio, and video, is important for the development of multimedia Web search systems. This paper provides results from a Weblog analysis study of multimedia Web searching by Dogpile users in 2006. The study analyzes the: (1) duration, size, and structure of Web search queries and sessions, (2) user demographics, (3) most popular multimedia Web searching terms, and (4) use of advanced Web search techniques including Boolean and natural language. The current study findings are compared with results from previous multimedia Web searching studies. The key findings are: (1) since 1997, image search consistently is the dominant media type searched followed by audio and video; (2) multimedia search duration is still short (>50% less than 1 minute) with few search terms; (3) many multimedia searches are for information about people, especially in audio search; and (4) multimedia search has begun to shift from entertainment to other categories such as medical, sports, and technology (based on the most repeated terms). Implications for design of Web multimedia search engines are discussed.

INTRODUCTION

Searching for multimedia is an important activity for users of Web search engines. Studying user's interactions with Web search engine multimedia buttons, including image, audio, and video, has potential to impact the development of more effective multimedia Web search systems. Image search is most widely studied area and has been given more attention by researchers, mainly due to its application of computer vision and machine learning. However, despite the promising research results, Web image search remains problematic. The precision of content based image search is still an important research problem. Youtube is the most popular video Web platform and tagging is an important method. Query by humming or singing is also an important research direction for audio searching (particularly for music category).

Text-based searching is the dominant method for multimedia Web searching. However, keyword-based Web search engines provide limited query formulation capabilities as text descriptors are yet to be structured using a standardized dictionary. Text based search effectiveness is limited due to gaps in the dimension and richness of contents between text, image, audio and video (Ortiz, 2007). Users often find it difficult to begin their Web search as most multimedia retrieval needs are often not clear as exact query keywords or a set of sample images/sounds. For example, compromise may appear in user's choice of result images due to difficulties in forming Web search queries (McDonald & Tait, 2003). Further focus is needed on users' searching behavior and content-based multimedia information retrieval (Lew et al, 2006). Future multimedia search studies need to understand users' queries, narrowing semantic gap, standardization in image description, and integration of different media (Kherfi, Ziou & Bernardi, 2004).

Previous studies have analyzed different aspects of Web query data logs (Silverman, et al., 1999; Spink, et al., 2002) and multimedia Web search (Tjondronegoro & Spink, 2007). The study reported in this paper is part of a larger ongoing research project to investigating multimedia Web searching behavior (Ozmutlu, et al., 2003; Spink & Jansen, 2004, 2006;

Tjondronegoro & Spink, 2007). Trend analysis is needed and important to understand changes in multimedia Web search over time.

The study reported in this paper provides an analysis of multimedia Web search using a recent Weblog, and compares the findings with previous studies. In particular, the paper examines the characteristics, duration and size of the sessions and queries in multimedia Web search. In addition, most frequently-used repeated terms, most followed-up up terms, and the use of advanced query terms are examined. The paper also discusses implications for the future development of multimedia Web searching technologies.

The next section of the paper outlines the previous studies exploring multimedia Web searching.

RELATED STUDIES

The study of Web search began with the development and popular use of Web search engines in the 1990's. Most Web search research has focused on general Web search and less so on multimedia Web search. A major study of multimedia Web search behavior conducted using Excite Web logs from 1997-2001 (Ozmutlu, et al., 2003) found that multimedia queries are generally longer and largely performed with multimedia interface buttons, i.e., users select particular type of media to search. The use of radio buttons was found to have decreased in multimedia Web searches in the general collection (Ozmutlu, et al., 2003). However, the researchers did not examine queries to the federated collections.

Jansen, et al., (2005) compared Web searching characteristics among Web, image, audio, and video content collections on the AltaVista search engine. They report that of the four types of multimedia Web searching with image searching the most multifaceted task and audio the least complex. The mean terms per query for image Web searching was larger (four terms) than the other categories of multimedia searching, which were less than three terms. The session lengths for image searches were longer than any other type of multimedia searching. Session length for

Boolean usage by image searches was 28 per cent. Sexual or adult content terms appear more frequently in image queries. Overall, multimedia searching is shifting as the content of the Web changes (Spink & Jansen, 2004).

Tjondronegoro and Spink (2007) surveyed the multimedia search functionality for a number of major Web search engines such as Google, Yahoo, MSN and AOL. Less than 5% of investigated retrieval systems provided content-based search methods. Tjondronegoro and Spink (2007) found that: (1) few major Web search engines offer multimedia searching and (2) multimedia Web search functionality is generally limited. They found that despite the increasing level of interest in multimedia Web search, few Web search engines offering multimedia Web search, provide limited multimedia search functionality. Keywords are still the only means of multimedia retrieval, while other methods such as “query by example” are offered by less than 1% of Web search engines examined.

Previous studies examining the characteristics of specialized multimedia searching (Jansen, et al., 2003, Spink & Jansen, 2006) used AltaVista 2002 and Dogpile 2005 Web logs respectively. The study reported in this paper used a Dogpile 2006 dataset. The study reported in this paper is based on Web transaction logs that are commonly used for capturing characteristics of user interactions with Web search engines (Jansen, Spink & Taksa, 2008). Web log analysis can collect data from a great number of users in a relatively inexpensive cost and it is also a least obstructive data collection method which represents the natural and unaltered behavior of searchers.

RESEARCH AIMS

The major aim of the study reported in this paper is to examine multimedia searching trends. The study compares comparison with other major studies which used data from different Web search engines and time. To assess Web multimedia searching trends, the findings from the current analysis is compared with findings from previous studies conducted on Excite 1997,

1999, and 2001 (Ozmutlu, et al., 2003), Altavista 2002 (Jansen, et al., 2004), and Dogpile 2005 datasets (Spink & Jansen, 2006).

The specific aims of the study are to examine the search characteristics and differences in web multimedia button use, including:

- *Session length* – in terms of the actual time spent, and the number of queries per session
- *Number of result pages viewed* and the lowest ranking
- *Searchers' demographics* – in terms of the location and Web browsers used
- *Query length* – the number of terms per query
- *Frequent multimedia terms* – in terms of most occurring and clicked through
- *Categories of frequent multimedia terms*
- *Use of advanced query formulation strategies* – such as Boolean and natural language

RESEARCH DESIGN

Data Collection

Dogpile.com

Infospace, a market leader in the meta-search engine business, owns Dogpile (www.Dogpile.com). Dogpile is the only meta-search engine during the study period to incorporate the indices of the four leading Web search engines into its search results (i.e. Google, Yahoo!, MSN Live Search, and Ask). When a searcher submits a query, Dogpile simultaneously submits the query to multiple other Web search engines, collecting the results from each Web search engine, removing duplicate results, and aggregating the remaining results into a combined ranked listing, using a proprietary algorithm. Hitwise (2005) reports that Dogpile was the 9th most popular Web search engine in 2005 as measured by number of site visits. ComScore Networks (2005) states that in 2005 Dogpile had the industry's highest visitor-to-searcher conversion rate of 83 per cent (i.e. 83 per cent of the visitors to the Dogpile site

executed a search).

Dogpile Web Transaction Logs

This study analyzed the records of searcher–system interactions in a Web transaction log that represents a portion of the searches executed on the Dogpile Web search engine on 15 May 2006. The original transaction log contained 1,228,330 records combined from the specialized multimedia search buttons, namely image, audio, and video. Each record contains nine fields:

- *IP*: user machine’s Internet address which can be used to uniquely identify users
- *Cookie*: automatically assigned by the Dogpile server to identify unique users on a particular computer. A cookie is allocated the first time a user is connected to the search engine until the user left a search session.
- *Time of Day*: the time (in hours, minutes, and seconds) that: a particular query is submitted, or a result is being followed up.
- *Query terms*: the user terms entered into the query box (e.g. “Web search”)
- *Vertical*: the query is being submitted to either the specialized (tabbed) images, video, or audio search, which clearly indicates users’ search intention for a particular type of media.
- *Organic/Sponsored*: the number of clicks on the organic and sponsored search results (indicating that the search result is followed up the user and may be deemed as relevant or interesting). Organic search results are naturally returned by the search engine (based on the search algorithm), while sponsored links are paid by advertisers. When both of these fields are “null”, the record indicates that it is an initial query. If there is no subsequent record showing at least one organic or sponsored link being followed-up in a session, then the user has not found a relevant result (i.e. they left the session after the initial query).
- *Rank*: the search ranking of the result being followed up (e.g. 2, 4, 6, 14, etc) which can provide an insightful information how many search results users are willing to go through.
- *Location*: indicate user’s location.

- *Browser*: the Web browser being used by the user (e.g. Microsoft Internet Explorer 6.0).

Data Analysis

We imported into a relational database, the original flat ASCII transaction log file of 1,228,330 records. Analyses were conducted at multiple levels, using the following metrics similar to that used in other Web transaction log studies (Ozmutlu, et al, 2003; Jansen, 2004; and Spink and Jansen, 2005)

- *Terms level* - any string of characters bounded by some delimiter such as white space.
- *Query level* – a query is a string list of one or more entered terms.
 - *Query length* is measured by counting the number of terms in the query.
 - *Query complexity* examines the query syntax, including the advanced searching techniques such as Boolean and other query operators.
- *Session level* is the entire sequence of queries entered by a searcher with a given data sampling method. This is similar to the definition of a unique visitor used by commercial search engines to measure Website traffic.
 - *Session length* is measured by the number of queries per searcher as each searcher is given a unique identifier within the log, namely the IP address.
 - *Session duration* is the total time the user spent interacting with the Web search engine, measured by subtracting the time user first submits the first query with the time user departs the search engine for the last time (i.e. does not return).
- *Results pages viewed* or *Click-Through* level – a results page is the list of results returned by a search engine in response to a query, which can contain organic or sponsored search results. From a results page, a searcher may click on the URL links to visit one or more results from the results listing, a method which is often referred as page view analysis. The result page viewing patterns of Web searchers are analyzed through the number of results pages clicked. In our Weblog, it is denoted by the sponsored and organic field being greater than 1 (and not

NULL).

Session Identification and Initial Filtering

We generated a unique identifier for each record. The traditional de facto standard for a session is generally 30 minutes without any activity on a given system. However, a recent study (Jansen, et al., 2007) defined the best measure of identifying sessions by a change in terms along with Internet Protocol (IP) address and cookie. Hence, we used three fields (IP address, cookie, and query) to locate the initial query and then recreate the chronological series of actions in a session. A session change is identified within the same IP address and cookie combination when the current query does not contain at least one term that has been used in the previous query.

To obtain human (not agent) sessions, all sessions with more than 101 queries submitted (in a session) were deemed to be conducted by agents. This threshold is used, as it is almost 50 times greater than the reported mean search session for human Web searchers (Ozmutlu, et al., 2003). After being filtered, the study focused on 361,319 human sessions. Our study reports on: (1) analyzing the duration, size, and structure of search sessions and queries and the demographics of the users, (2) click-through analysis to study the lowest ranked results being followed up and to find the most popular multimedia Web searching terms, and (3) the use of advanced Web search techniques including Boolean and natural language.

RESULTS

Multimedia Session Level Analysis

Multimedia Search Duration

Table 1 and Figure 1 show the characteristics of multimedia search sessions. Overall, image search dominates (i.e. 55% of multimedia sessions), 31% were audio sessions and 14% were video queries.

[Place Table 1 Here]

[Place Figure 1 Here]

Table 1 shows that based on the average Web session duration, video search is the longest at 6.9 minutes, while audio is 6.8 minutes, and image is 4.94 minutes. The top 3 most common session duration for all multimedia searches indicate that 80% of the users spend less than 5 minutes (with 56 to 62% of them are less than 1 minute), while only 6% and 3% users spend 5 to 10 minutes or 1 to 2 hours respectively. Table 1 shows that multimedia search is generally very short, and there are only a very small portion of users spending around 1 or 2 hours. By observing the average number of queries submitted per session, we can learn that users have entered more queries on image search (2.8 queries) followed by video (2.4) and audio (2.3) respectively. The most common session size is 2 queries (greater than 40%), followed by 1 query (greater than 23%). It shows that users generally do not spend a lot of time refining their queries.

Search Result Clicks

Table 2 shows the statistics on search results being followed up per session. From the total number of sessions that have at least one follow up (per session), only 56% of them are image sessions, compared to 80% in both audio and video sessions.

[Place Table 2 Here]

Users do not generally have to click on the results of image search due to the visual thumbnails being sufficient for users to judge the relevance, while users need to actually view the temporal information in audio and video to understand the contents. This can also demonstrate that image search users are not interested in the full size of the image or reading the document that contains the image to get more complete information, and as such, they are only looking for visual stimuli.

Nevertheless, the Web logs cannot show whether the 44% (i.e. the rest) of the image sessions with no follow up actually means browsing intention or unavailability of relevant results. Figure 2 shows that the majority of multimedia sessions (i.e. 89% of image search, 81% of audio

search, and 73% of video search) followed up one to five results, and that the majority of the links followed up are the organic search results.

[Place Figure 2 Here]

However, given the small number of sponsored results (which are returned for a single query), we should not conclude that users do not use sponsored results. There are generally more interests in sponsored results for image search than in audio and video.

Table 3 shows the statistics of the lowest ranking results being followed up per session to indicate users' reliance on the ranking algorithm of the Dogpile search engine and their patience in scrolling for long list of search results.

[Place Table 3 Here]

Figure 2 shows that that 38% of users are only willing to go as far as the results ranked at fifth. However, there are also 31 to 36 % users that go beyond rank sixteenth results and are willing to browse a long list of results, or the results are spanned across multiple pages and users only look at the first 2 pages. Combined with the majority of users spending less than 5 minutes on multimedia Web searches, there are two challenges for designing a search results interfaces: 1) support for more effective and semantic browsing (whereby users are presented with thumbnails or a summary of the multimedia contents), and 2) improve the ranking algorithm to speed up the process of finding the right information or support pleasant scrolling. For example, Web browser plug-in software such as "cool iris" (cooliris.com) provides a visual browsing environment for image search.

Dogpile User Demographics

Table 4 and Table 5 show the demographics of Dogpile multimedia search users in 2006.

[Place Table 4 Here]

[Place Table 5 Here]

The top 3 browsers used are Microsoft Internet Explorer (86%), Mozilla (9%) and Apple

(4%). These statistics show that most multimedia searchers rely on Microsoft Internet Explorer due to its integration with Windows (as the most used Operating Systems) and the strong supports for multimedia file formats and Direct X. The top four countries of multimedia Web search session are English speaking countries: United States (82%), United Kingdom (9%), Canada (5%), and Australia (4%); followed by Saudi Arabia, Germany, Singapore, India, New Zealand, and Malaysia.

Query Level Analysis

Table 6 shows the typical length and occurrences of multimedia search queries.

[Place Table 6 Here]

Figure 3 displays the distribution in a graphical diagram.

[Place Figure 3 Here]

Query length for multimedia searching generally ranges between 1 to 4 terms with two terms per query most commonly used. There are a total of 178,399 unique search terms in image, 94,682 in audio, and 49,978 in video, were used in multimedia queries. The average terms per query for audio search are slightly longer at 3.1, followed by audio and video both at 2.3 terms per query.

The 2006 trends (Figure 3a) clearly show that the distribution statistics of query length in video and image search is almost equivalent.

[Place Figure 3a Here]

Query length of greater than or equal to 6 is rarely used for image and video searches (with percentage of less than 2%). However, in audio search, query length of 5 and 6 are often used with percentages of 10.6% and 6.2% respectively. This could be due to the fact that for audio search, users generally use song title or artist name or a combination of both like “ghost riders in the sky johnny cash” and “lionel richie just for you”. Compared to 2002 (Figure 3b), the query length distribution is almost the same for image, audio and video search. There are two

notable differences between 2002 and 2006, including: 1) 2002 audio query length was greater than or equal to 5 and much less than 2006; 2) the percentage of 2002 image query length equal to 9 is 27% (which was deemed as an anomaly of the data collection).

Top 25 Terms - Click-Through Analysis

Table 7 shows the top 25 terms that produced results from which users clicked on results as indicated by how many times they are entered by users (not necessarily unique) and the number of results viewed (i.e. clicked).

[Place Table 7 here]

All top 25 terms in image queries are sexual. This is also the case for video queries, except for “anime”, which is Japanese movie animation. In audio queries, most queries are either song title or artist (or performer), where “ridin dirty” is the most popular being entered 289 times (including the shortened version), which users clicked 762 organic links. The most popular performers searched for are “Shakira” and “Eminem”. Sponsored links are generally not followed (i.e. clicked by users), the average is 1-2 clicks per session with a maximum of 5 clicks. Compared to Spink et al., (2004) it is quite consistent that the most frequent sexual terms in Web search are limited and mostly based on celebrities’ name or “girls”. However, they are not major proportion of all search queries.

Table 7 also suggests that some queries co-occurred in multiple Web multimedia searches. An analysis of overlapping queries (i.e. unique terms which are used in multiple multimedia types searches) shows 12,730 overlapping queries used both in image and video searches. There are 5,742 queries co-occurring in image and audio searches, and 5,754 co-occurring in video and audio.

Concepts (or Topics) of the Most Followed-up Queries

We further examine the search intent for the most followed-up queries by semi-

automatically categorizing the submitted search terms (used in each query). Out of the most clicked-through search queries, 1485 (out of 45,578) audio, 2818 (out of 63,815) image, and 911 (out of 25,216) video queries were able to be automatically classified into semantic categories using Open Calais tools (refer to <http://www.opencalais.com/calaisAPI> for the complete list of the extracted entities and their definitions). We manually corrected some of the categorization results by adding some new categories (namely 'sexual', "music" and "movie or game" for image and video, "song" and "sound" for audio). The corrections are particularly required as many of the *IndustryTerm* search terms are incorrect. The detection of other topic categories such as person, sports game, and country are sufficiently accurate, as such, we did not need to correct them.

Some examples of the categorized multimedia search terms are:

- *Holiday*: "memorial day", "labor day", and "independence day"
- *Medical condition*: "mad cow disease", "eating disorder", and "liver cancer"
- *Natural disaster*: "avalanche", "hurricane", and "tsunami"
- *Natural feature*: "Caribbean islands" "Australian island", and "French beach"
- *Region*: "East Africa", "Mediterranean", and "Antarctic".

Users often opted to use incomplete words from song title and performing artist (which can be a person's name or a group name). Hence we actually performed the search on Dogpile Web search engine and examine the first page of the results to predict the intended search purposes. For example, "tatiana" refers to "Brian Tatiana (artist)"; "Tater White Salad" refers to "Ron White (artist) – Tater Salad (song title)"; and "breakaway" refers to "Breakaway (song title) – by Kelly Clarkson (Artist)". The more difficult task is to interpret generic terms as band group's name such as "silver chair" and "Beatles".

Moreover, an assumption that the search was on the context of audio must be made in order to interpret some ambiguous terms. For instance, "car squealing", "car stopping", "car racer", "loud car exhaust", "bass", "bass solo" are identified as "sound (or sound effect)" search

terms. Another assumption that users are searching for the soundtrack (theme song) is made when the search terms refers to the titles of movie, game, animation, and TV program. For example, “metal alchemist” refers to the soundtrack of the animation of that title, “snake eater” refers to the theme sounds and songs from the game of “metal gear solid – snake eater”. Some search terms are best described as song genre, such as “metal rap” and “love metal”. However, singular terms which are too ambiguous are left it in the original category. For example “im” could refer to “I’m okay” or other song title which has the words “I’m”, while “pleasure” could refer to “Struggle for pleasure”; but without the existence of other clarifying terms, an assumption cannot be made. Hence, industry term category in audio (as well as image and video) search still includes ambiguous and unclassified terms.

During the manual correction of image search term categories, we also focused on re-categorizing the search terms which are classified as IndustryTerm. However, image search terms are more diversified and thus we decided to avoid using too specific category as it will not reveal a converging result. Instead we have defined new categories, including “sexual”, “movie or game”, and “music”, which can be easily identified. As for the rest of the incorrectly classified terms (as IndustryTerm), we determine that many of the search terms are for landscapes (e.g. “yosemite”, “Melbourne South bank”), and visual objects (“food recipes”, “teeth”, “gas cooker”, “energizer bunny”). However, these categories cannot be determined only with assumptions and without manually examining the existence of other clarifying terms, For example as users may actually search for desktop wallpaper showing “Yosemite”, while they may wanted to get individual pictures of “teeth”, but an animation of “energizer bunny”.

Predicting the search intent of video terms was found to be the most challenging task as they can refer to wide range of topics, including music video whereby audio terms such as singer or band group’s name and song title can be used. Moreover, movie video search can use actor’s name or movie title. The previous assumptions for image and audio search terms cannot be used as video contains both media. For example, a search on “car” can mean the sounds of car

or video scenes which contain car, while a search on “cable car” can mean the music or video which contains cable car. Hence, we only corrected terms which clearly belong to the existing categories, including “sexual,” “person”, “movie/game title”, “music (i.e. song title)” and “sports”. Table 8 lists the topics (or concepts) of the most clicked-through queries.

[Place Table 8 Here]

Queries for image, audio and video are about *people*, including singer’s or band group’s name, politicians, sportsperson and movie actors. The second and third most important categories are: *medical condition* and *sexual* for image search; *song* and *medical condition* for audio search; *medical condition* and *sports game* for video search.

Use of Advanced Web Search Features

Advanced query terms are usually combined using Boolean expressions such as AND (often symbolized by +, or &), NOT (symbolized by -, or ~), OR (symbolized by |), and Quotation (i.e. “”). Table 9 presents the Top 10 advanced image, audio, and video query terms which returned the most relevant results for users, as indicated by the number of click-through (either organic or sponsored).

[Place Table 9 Here]

At first, “round and brown” could be seen as a method by which users are trying to search images based on the shape and color (i.e. features based query). However, after we examine the search results, it is more likely to be a sexual search. Most of the advanced audio query terms are song title. It can be noted that some of the top 10 advanced video terms co-occurred in the image search, including “romeo and Juliet”, “dead or alive” which are movie or game, and “round and brown” which is sexual. Natural language is often started by “when”, “how”, “what”, “where”, “does”, “do”, “is”, and “are”. Table 10 presents the Top 10 natural language query terms in multimedia search which returned the most relevant results for users, as indicated by the number of click-through (either organic or sponsored).

[Place Table 10 here]

The low number of click-throughs that use natural language may be interpreted as users not finding many relevant results due to not many images that are described using natural language. Alternatively, users may have solved their information problem in the first few clicks.

Most image search queries using natural language started with “how” (6 out of the top 10) and users were often looking for instructional images. The majority of image searches were for a descriptive diagram (e.g. how does a liver cancer look like). Thus, multimedia search engines should aggregate results on the information about the disease (text), with images that describe the different stages or types (images), and documentary videos. The natural language terms used in video search were looking for tutorials, such as “how your body work”, “how to get back in shape”, “how to tell the sex between frogs” and “how cheese is made”. Most song title queries are in natural language form, such as ‘what a wonderful world”, and “when I’m gone”. Therefore, the number of natural language queries for audio search is higher than video and image queries.

DISCUSSION

What are the trends in multimedia searching? Image search dominates the multimedia Web queries with more audio searches than video searches. People Web search is the most frequent search for image, audio and video. Medical conditions (i.e. diseases) are the second most frequent topic in all multimedia Web search types as users are looking for descriptive media to explain diseases beyond textual information. However, based on the top 25 most clicked through multimedia Web queries, sexually related Web searches are the most frequent genre in image and video Web searches. The most frequent audio Web searching terms are music related, such as song/album title or artist. There are also many co-occurring terms in popular image and video web searches, showing that users use similar searching strategy for defining their needs for visual information.

This study shows 12,730 overlapping queries that are used both in image and video Web

searches, 5,742 in image and audio searches, and 5,754 video and audio. Web search engines should explore providing integrated search using a single interface to all multimedia types. Moreover, cross-media search results can be aggregated based on the semantic meaning. For example, movie title search can comprehensively return video clips (e.g. teaser trailers, or the music video of its theme song), audio (e.g. the song or sound effect clips from the movie), and images (e.g. on the actors and the actions).

There are not many users of natural language in image and video search, and they are commonly used for answering “how to” and “what is “questions. Boolean (“AND”) search is used for image and video Web search to combine their problems, such as “dead sea and jordan river” and “piano and rose”. For audio web search, Boolean keywords and natural language are often used as song titles. Thus, we are currently able to draw a precise analysis on users’ intention in using advanced searching.

Compared to the earlier Dogpile 2005 study (Spink & Jansen, 2006), the current study shows that similar to overall Web searching, most content collection Web searches are short, contain few terms, and that results pages are viewed except for image searches. Similarly to our findings, Spink and Jansen (2006) found that image search is the most popular (67%), followed by audio (22%), and video (11%) respectively. However, currently video search is more dominant at the expense of (less) audio search. This slight increase could be due to the increasing proportion of users with high band-width Internet connections and the increasing popularity of video as rich content types for social networking and advertising. Video and audio search in information and entertainment genre may increase as the uptake of portable media players, such as iPod, increases. Compared to the earlier study, the current found that more users across image, audio, and video content collections entered two queries per session (i.e. 42% in 2006 compared to 16% in 2005), instead of only one query.

Spink and Jansen (2005) found that Dogpile searchers generally use three terms per query (mean=2.85) and spend less than one minute interacting with the Web search engine. Our

study confirms this trend. Users' terms-per-query are now slightly shorter (mean=2.6) and slightly more (57 to 62 per cent) users spend less than one minute. Consistent with Spink and Jansen (2006) currently multimedia queries are generally one to four terms (two terms being the most commonly used).

Unlike the Spink and Jansen (2006) study, the current study did not see audio sessions being longer but with fewer queries per session. Video is currently the longest with fewer queries per session than image search, but longer than audio searches. The lowest ranked result is now being followed up in a session and people are now willing to go beyond 16th ranked results. The same proportion of users examined only the first results page (i.e. rank 1 to 5) as Spink and Jansen (2006) where users' generally examined only the first results page and only people seeking images examined further results pages. Considering that multimedia searchers still spend a very short period of time (i.e. less than one minute), this mounts more serious challenge for the future multimedia search engine to provide a better browsing interface.

The analysis of categorized topics in the search terms shows a consistent finding with the previous result which found that the most frequent queries was for popular people and celebrities, places, or things. We also found that sexual and pornographic queries represented a very low proportion of all queries. However, we found that the top 25 most followed up queries in image and video search are sexual.

Compared to the multimedia search trends study of 1997-2001 (10 years ago) that used Excite Web log (Ozmutlu, et al., 2003), the proportion of multimedia searching is still consistent; averaged over 3 years, 50% of MM search is for image, 28% for video, and 22% for audio. The mean queries per session is also quite consistent, being 3.2 for images, 2.4 for audio, and 2.6 for video. For each query, the mean number of terms used is: 4.1 to 4.4 for audio, 3.4 to 4.9 for images, and 3.5 to 4.1 for video.). The most frequently used terms for music search are "music", "mp3", "sound", and "songs"; while the popular terms in video search are "video", "movie", "free", and "MPEG"; and finally for image search "free", "art", "pics" are most frequently used. However,

it is important to note that the Weblog used for this study was prior to the use of separate buttons for each multimedia type. The analysis for image, audio, and video searching depends on the detection of some typical terms such as “mp3” for music. Hence, the results of their analysis cannot be directly compared with this study.

Jansen, et al., (2004) examined the effect of specialized multimedia searching buttons using Altavista 2002 data. They found that image searching appears to be the more multifaceted task, and audio appears to be the least complex. The most dominant session size is one query which is still consistent with our study. However, they found that the mean terms per query for image searching was notably larger (3.21 terms) than the other categories of searching, which were less than 3 terms (audio is 1.62, and video is 1.09). In contrast, our study found that two terms per query is mostly used for all image, audio and video searches, demonstrating that users have incorporated longer search terms per query in audio and video searching.

Furthermore, based on the most occurring terms, their study noted that the shift in Web searching trends (from entertainment to commercial content and an increased variety of topics) has not happened yet in multimedia searching which are mostly for entertainment and many of the terms are sexual related. We noted a difference in our study on the most repeated terms (shown in Table 5), which used a more recent Weblog (i.e. Dogpile 2006). We found that some multimedia searching users have started to shift (from entertainment) into other categories such as medical, sports and technology.

Moreover, more Web search providers are now moving towards a “federated” search interface to enable users searching for information without worrying about what type of media it is represented in (i.e. media-agnostic searching). For example, when general users search for a movie (based on the title or its main actors’ names), they would be interested in images (e.g. photos of actors and snapshots from key scenes), audio (e.g. the sound tracks from the movie), and video (e.g. the teaser or key video segments). As a future work, when portable devices become more popular, we should benefit from a specialized study on multimedia searching from

mobile devices which generally have small screen size, thereby users should be presented with more effective interface, and more requirements for contextualized and personalized search results.

Users' terms-per-query is now slightly even shorter (mean=2.60) and there are slightly more (57 to 62 per cent) users spend less than one minute. Combined with the fact the most users are only willing to follow up the first five results of the first and second page, the future multimedia search interfaces need to be more effective for: 1) query formulation, 2) browsing the search results (to determine the relevant ones), and 3) to view the multimedia document which caters the spatial and temporal dimension in text, audio, and video. Given that there are still some proportion of users (although small) that would spend more than 10 minutes in a search session, there are two types of users in multimedia search: (1) those users who know precisely what they need (i.e. only willing to spend less than one minute to decide whether the search results provide what they need and leave), and (2) those users who know their search topic and are willing to browse for more information to decide what they actually want.

Implications

To support the first type of users (i.e. specific and on-the-go), the search interface should enable them to narrow down the results and the ranking is improved based on the context, interests (e.g. user profile or search history) and the particular searching purpose at that time. For example, the system should allow users start their image search for "Paris" using the keyword, and let users select the pictures/keywords that represent "Eiffel Tower" to filter out other pictures such as "Paris Hilton" (using image similarity algorithms). To support the second type of users (i.e. generalists and browsers), the search interface should provide users with more relevant (and still interesting) results with their initial query while providing a fun and intuitive interface to browse results. For example, query expansion based on concept mapping can be used to bring more results to a specific query and allow users to see more related results which are still conceptually related.

Most image, audio and video queries are seeking information on a *person*; including singer's or band group's name, politicians, sportsperson and movie actors. Hence, the improvement of face detection (from computer vision field) may provide faster and more accurate results. More visually relevant search results are important for image and video Web search. However, manual annotation for people's appearance in images and videos is an expensive and tedious task. Moreover, music annotation should at least include the name of the singers and song composers or producers to allow users in getting more relevant music from their favourite (i.e. of their current interest) artist. The other most important categories are: *medical condition* and *sexual* for image search; *song* and *medical condition* for audio search; *medical condition* and *sports game* for video search.

To support better medical-related searches in image and video, Web search engines can benefit from the advancements in medical imaging (from computer vision field) to semi-automatically annotate visual information with the types of disease or anomaly in medical images which include external organs (e.g. skin and hair images), and internal organs (e.g. X-Ray images). Moreover, amid all the ethical issues, there is potential in allowing expert users (preferably doctors and experienced practitioners) to annotate medical-related images, so that users can search and learn from more examples of a particular disease or condition, particularly for educational purposes.

Lastly, sports game related video search will benefit from richly annotated and content-based segmented video segments. For example, many users in the (arguably) most popular video sharing portal Youtube.com have contributed in the creation of video compilations on "the top ten goals of a particular player", "most memorable match of a particular team", and "highlights of a particular sport game". It is worth to note that Youtube.com video collections are indexed and searched by Google.com, thus Dogpile.com search (which combine the results from Google) should return more results from such richly annotated sports-related video to improve the relevant search results.

Limitations

While examining the characteristics of specialized multimedia searching, we have compared the current results (based on Dogpile 2006 dataset) with previous studies that used Excite 1997-2001, AltaVista 2002 and Dogpile 2005 Web logs (Ozmutlu, et al., 2003, Jansen, et al., 2003, Spink & Jansen, 2006). As such, our comparative analysis to study the trends from 1997 to 2006 assumes that multimedia Web searching users consistently share the same interests regardless of the search engines used. In addition to this variant, we have used a new method for session identification (by adding a change in query term from the initial term as an indication for a session change within the same IP and cookie) and an additional calculation method for session duration (by using the timestamp field incorporated within the latest web log data).

There are some limitations in using web log analysis to accurately calculate session duration and the time spent for browsing/viewing search results. The timestamp for each query (within a session) is recorded when a user submits the query, thus there is a possibility that users may leave the computer for conducting other activities (such as resting) between submitting two subsequent queries. Moreover, web log data cannot be used to determine when a user leaves the current search session, as there is only a record of the last query being submitted. Hence, the calculation of session duration in this study does not take into account the viewing time (of the last search results) prior to leaving the session. The problem with the current method in identifying a user's session in a particular computer (i.e. via cookie and IP address) is the inability to learn whether a user opens multiple browser windows to submit each query. As a result, a study of multi-tasking search cannot be conducted.

Without other user data, such as from questionnaire or observation, we cannot determine successful searching, including: 1) whether a user is satisfied with the results prior to leaving a search session, and 2) whether a query with zero (number of) click through means that users

are satisfied with their image/video search after browsing the visual thumbnails or they did not find any relevant result. Furthermore, a web log cannot indicate the reasoning process behind the initial and subsequent query terms (to reach the original or updated search goal(s)). As such, a study of users' search intent evolution while searching for multimedia is limited. Finally, classification of query terms cannot accurately determine the intended concepts or topics (being searched), while a fully manual process of classification (that would avoid false detections) is tedious, lengthy, and inaccurate, especially when the query is too short and without any clarifying terms.

CONCLUSION AND FURTHER RESEARCH

Our current study has provided insights into the current state and trends in multimedia web searching. We are currently conducting further research into multimedia Web searching by investigating users' genre interests and searching for personal names. For example, the personal names can be for music, sports, politics, and movie or games. Using larger scale data and other data analysis techniques, we also need to study multitasking during multimedia Web search sessions. Effective user models are needed to help provide users with better search support.

REFERENCES

- ComScore Networks (2005), available at:
www.comscore.com/press/release.asp?press=325.
- Hitwise (2005), available at:
www.clickz.com/stats/sectors/search_tools/article.php/3528456.
- Jansen, B.J., Spink, A., & Kathuria, V. (2007). How to Define Searching Sessions on Web Search Engines. *Lecture Notes in Computer Science*. Vol 4811/2007. 92-109. Springer Berlin.

Jansen, B. J., Spink, A., & Pedersen, J. (2003). An analysis of multimedia searching on AltaVista. *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, (pp. 186-192). Berkeley, California: ACM.

Jansen, B. J., Spink, A., & Pedersen, J. (2004). *Web Search: Comparison of Searching for Web, Image, Audio, and Video Content*. <http://jimjansen.blogspot.com/2008/08/comparison-of-searching-for-web-image.html>.

Jansen, B. J., Spink, A., & Pedersen, J. (2005). The effect of specialized multimedia collections on Web searching. *Journal of Web Engineering*, 3(3/4), 182-199.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). 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., Spink, A., & Taksa, I. (Eds.), (2008). *Handbook of Research on Web Log Analysis*. Idea Group.

Kherfi, M. L., Ziou, D., & Bernardi, A. (2004). Image retrieval from the world wide web: Issues, techniques and systems. *ACM Computing Surveys (CSUR)*, 36, 35-67.

Lew, M. S. (2000). Next-generation Web searches for visual content. *Computer*, 33(11), 46-53.

McDonald, S., & Tait, J. (2003). Search strategies in content-based image retrieval. *SIGIR2003: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, Netherlands*.

Ortiz, S. (2007). Searching the visual web. *Computer* 40(6), 12-14.

Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2003). Multimedia web searching trends: 1997-2001. *Information Processing & Management*, 39(4), 611-621.

Silverman, C., Henzinger, M., Marais, H., & Morris, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33.

Spink, A., & Jansen, B. J. (2004). *Web Search: Public Searching of the Web*. Dordrecht: Springer.

Spink, A., & Jansen, B. J. (2006). Searching multimedia federated content web collections, *Online Journal Review*, 30(5), 485-495.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: web search changes. *IEEE Computer*, 35, 107-109.

Spink, A., Ozmutlu, H. C., & Lorence, D. P. (2004). Web searching for sexual information: An exploratory study. *Information Processing and Management*, 40(1), 113-124.

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science*, 53(2), 226-234.

Tjondronegoro, D., & Spink, A. (2008). Search engine multimedia functionality. *Information Processing and Management*, 44(1), 340-357.

Wolfram, D., Spink, A., Jansen, B. J., & Saracevic, T. (2001). Vox populi: The public searching of the web. *Journal of the American Society for Information Sciences and Technology*, 52(12), 1073-1074.

Table 1: Comparison of audio, video and images Web search sessions.

Variables	Image	Audio	Video
Total sessions	196,755 (54.4%)	113,769 (31.4%)	50,795 (14.06%)
Average Duration per Session (in seconds)	296.1 (4.9 mins)	410.4 (6.8 mins)	414.5 (6.9 mins)
Less than 5 minutes	88.9 %	86.9 %	82.5 %
5 to 10 minutes	5.1 %	5.9 %	8.12 %
10 to 15 minutes	1.5 %	1.7 %	2.7 %
15 to 30 minutes	1.1 %	1.2 %	1.9 %
30 to 60 minutes	0.2 %	0.3 %	0.3 %
1 to 2 hour	2.2 %	2.6 %	3.3 %
2 to 3 hour	0.2 %	0.3 %	0.3 %
3 to 4 hour	0.08 %	0.1 %	0.1 %
4 to 5 hour	0.04 %	0.09 %	0.08 %
More than 5 hours	0.3 %	0.4%	0.4 %
Average Queries Per Session	2.8	2.2	2.4
1 query	31.05 %	28.7 %	23.7 %
2 query	41.1 %	40.5 %	42.2 %
3 query	14.2 %	14.6 %	15.5 %
4+ query	13.5 %	16.01 %	18.4 %
Maximum Queries Per Session	78	48	48
Standard Deviation for Queries Per Session	2.22	1.6	1.8

Figure 1: Distribution of session duration and size in multimedia search.

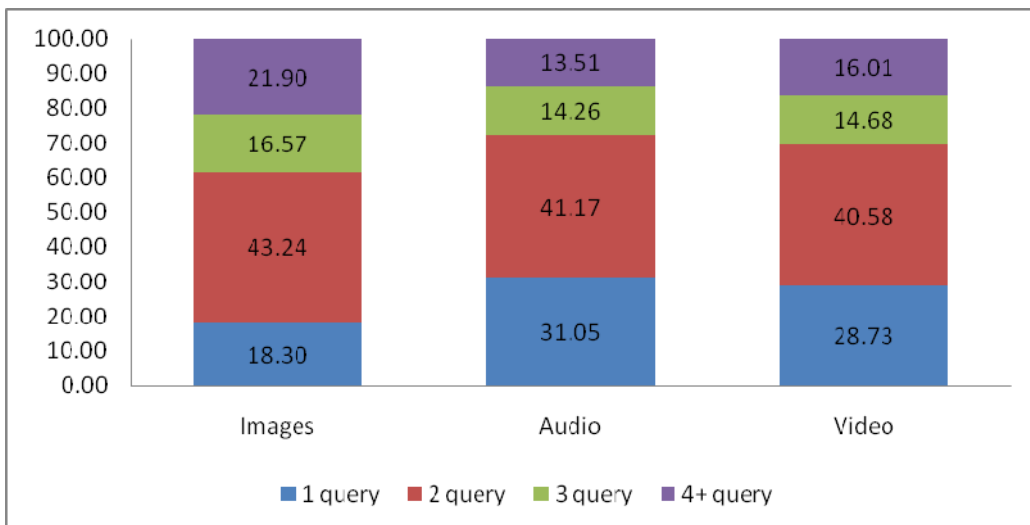
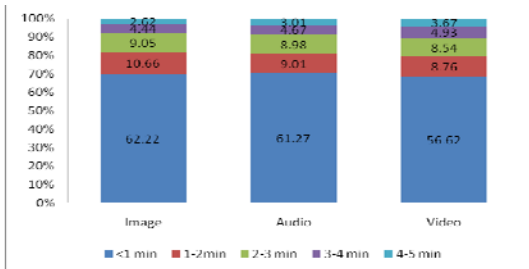
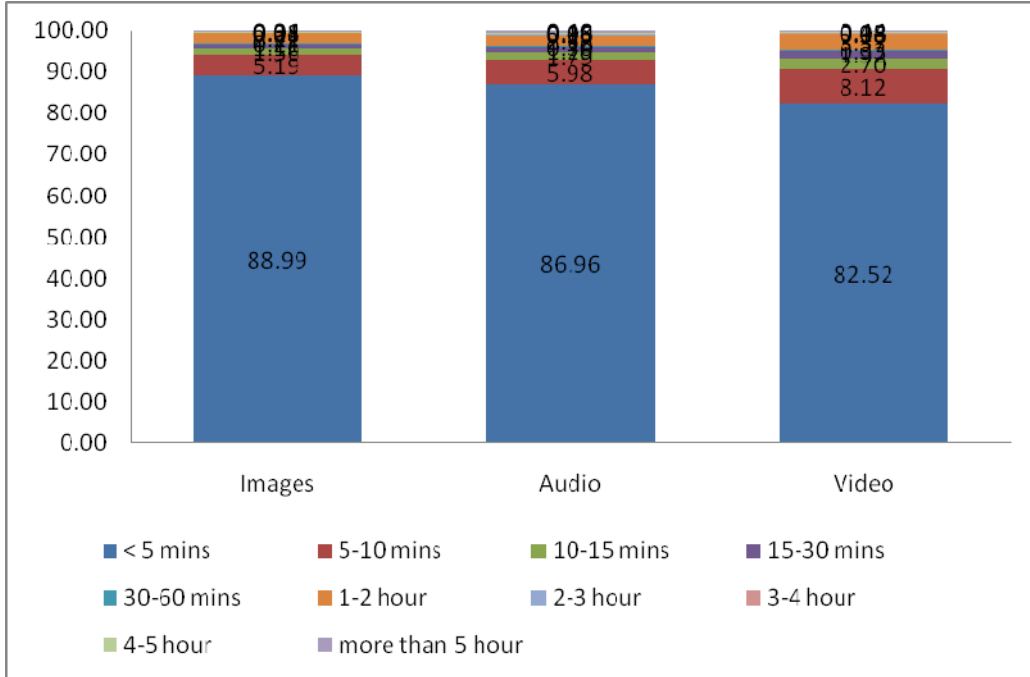


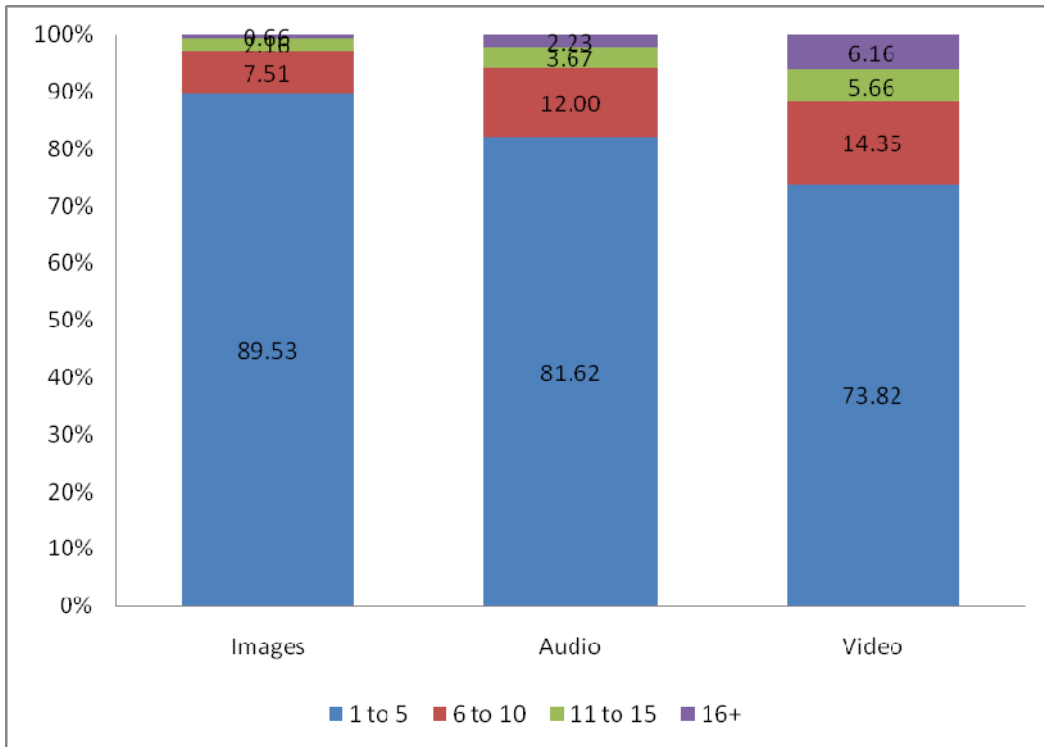
Table 2: Statistics of search results followed-up in a session.

Organic Links Visits In a Session:	Images	Audio	Video
1 to 5	87,492	66,218	27,298
6 to 10	7,553	9,772	5,335
11 to 15	1,828	2,986	2,104
16+	1,154	2,205	2,290
Total	98,027	81,181	37,027
Sponsored Links Visits In a Session:			
1 to 5	2,725	224	141
6 to 10	10	0	0
11 to 15	2	0	0
16+	0	0	0
Total	2,737	224	141
Number of Sessions that have at least 1 result followed up	100,764.00	81,405.00	37,168.00
Percentage of the sessions with at least 1 follow up	55.7%	80.8%	79.7%

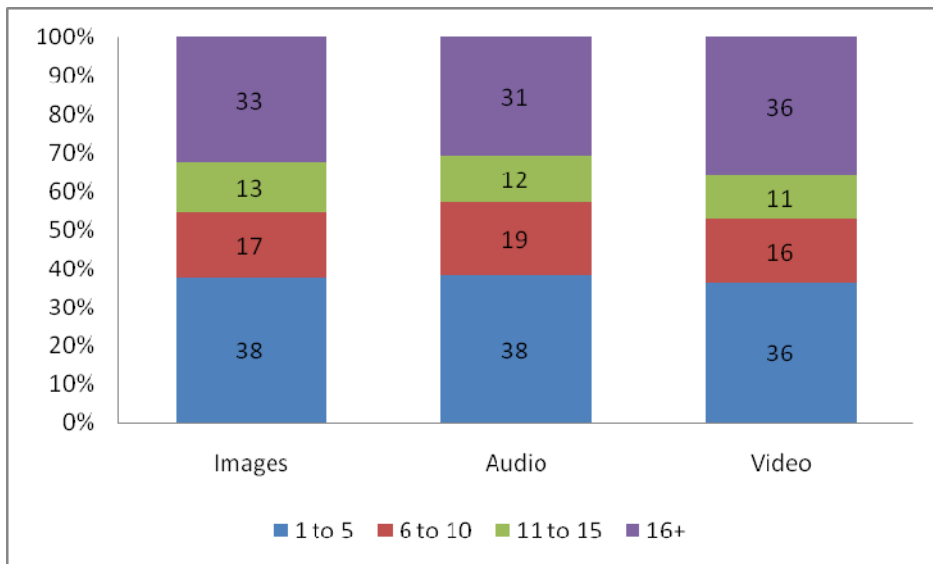
Table 3: Statistics of lowest ranking followed-up in a session.

Lowest Rank followed up in a session	Images	Audio	Video
0 (not followed up)	84,612	21,753	10,279
1 to 5	41,248	35,032	14,695
6 to 10	18,633	17,571	6,615
11 to 15	14,144	11,028	4,657
16+	35,647	28,385	14,549
Total	109,672.00	92,016.00	40,516.00
Percent Followed up	55.7 %	80.8 %	79.7%

Figure 2: Distribution of links followed up and the lowest ranks.



Results Followed-Up per Session



Lowest Ranking Followed-Up per Session

Table 4: Browsers used in multimedia sessions.

Browser	Microsoft	Mozilla	Apple	Netscape	Opera	Media Players	Others
Multimedia Sessions	309,631	31,416	14,258	2,675	1,120	127	116

Microsoft: Microsoft Internet Explorer

Mozilla: (firebird, firefox, mozilla)

Apple: Safari, iCab

Media players: iTunes Media, Real Player, Web TV browser

Others: Curl, Galeon, IBM Planet, K-Meleon, LibWWW, Konqueror, WGet.

Table 5: Session locations.

Location	CountOfid
usa	281,207
gbr	30,807
can	15,965
aus	12,969
sau	1,615
deu	1,315
sgp	1,299
ind	1,184
nzl	1,149
mys	753
irl	751
phl	680
jpn	582
zaf	565
fra	507
mex	506
are	447
chn	389
swe	378
ita	353
nld	322
isr	316
pri	290
nor	288
esp	279
	354,916
Sampling from Total Sessions	98.227882

Country names based on ISO 3166-3 (Alpha 3)

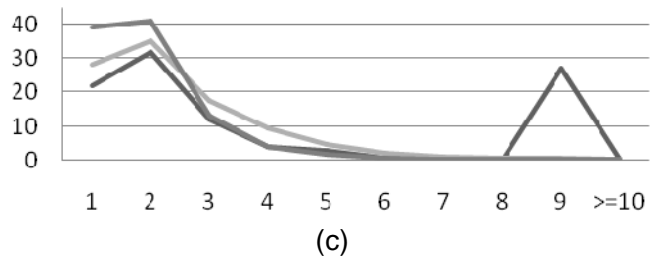
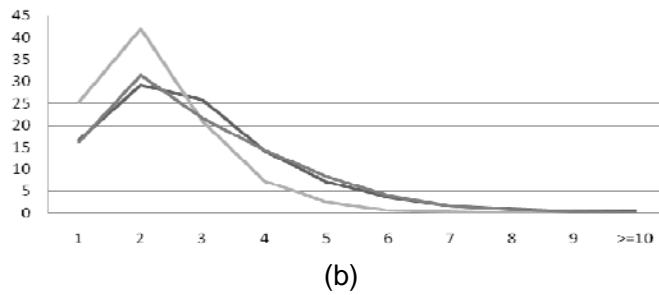
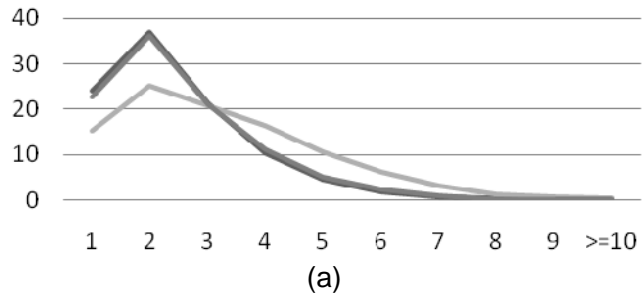
Highlighted indicates top 10: 1) United States, 2) United Kingdom, 3) Canada, 4) Australia, 5) Saudi Arabia, 6) Germany, 7) Singapore, 8) India, 9) New Zealand, 10) Malaysia

Table 6: Multimedia search query length (number of terms).

Query Length	Image Queries	Image Queries (%)	Audio Queries	Audio Queries (%)	Video Queries	Video Queries (%)
1	31,748	23.9%	13,663	15.2%	9,575	22.8%
2	49,094	36.9%	22,601	25.2%	15,200	36.1%
3	28,424	21.45	18,621	20.7%	8,869	21.1%
4	13,744	10.3%	14,685	16.3%	4,713	11.2%
5	6,014	4.5%	9,531	10.6%	2,164	5.1%
6	2,390	1.8%	5,552	6.2%	956	2.3%
7	918	0.7%	2,857	3.2%	397	0.95
8	386	0.3%	1,250	1.4%	65	0.2%
9	180	0.1%	613	0.7%	58	0.15
>=10	135	0.1%	489	0.5%	55	0.1%
Total	133,033	100%	89,862	100%	42,052	100%

Figure 3: Multimedia query length distribution.

(a) Dogpile 2006 (this paper), (b) Dogpile 2005 (Spink, Jansen 2004), (c) Altavista 2002 (Wolfram et al, 2001). Y-axis: Percentage, X-axis: Number of terms



- Occurrences in Image Search
- Occurrences in Audio Search
- Occurrences in video Search

Table 7: Top 25 most frequently occurring and clicked-through multimedia query terms.

Image			Audio			Video		
Total Queries	562,380		Total Queries	370,890		Total Queries	208,115	
Query	Frequency (%)	Links Followed	Query	Frequency (%)	Links Followed	Query	Frequency (%)	Links Followed
pussy	0.08%	498	ridin dirty	0.07%	427	pussy	0.6%	1371
boobs	0.04%	251	ridin	0.06%	335	boobs	0.3%	652
sex	0.04%	237	shakira	0.06%	318	hentai	0.2%	514
hentai	0.03%	196	eminem	0.06%	343	sex	0.2%	468
porn	0.03%	181	50 cent	0.06%	362	porn	0.1%	370
tits	0.02%	176	dani california	0.06%	266	preteen	0.1%	320
paris hilton	0.02%	134	temperature	0.04%	231	tits	0.1%	344
milf	0.02%	129	panic at the disco	0.04%	191	paris hilton	0.1%	310
ass	0.02%	122	ms new booty	0.04%	197	milf	0.1%	315
penis	0.02%	113	sean paul	0.04%	199	ass	0.1%	260
thongs	0.02%	107	system of a down	0.04%	203	carmen electra	0.1%	249
carmen electra	0.02%	106	hips dont lie	0.04%	253	penis	0.1%	226
anime	0.01%	103	chamillionaire	0.04%	194	hot girls	0.1%	225
jessica alba	0.01%	103	chris brown	0.04%	218	incest	0.1%	231
lesbians	0.02%	99	bad day	0.04%	191	big tits	0.1%	225
hot girls	0.02%	99	taking back sunday	0.04%	173	blowjob	0.1%	210
big tits	0.01%	99	slipknot	0.03%	187	thongs	0.09%	199
nude	0.01%	92	rihanna	0.03%	168	lesbians	0.09%	188
preteen	0.02%	92	grillz	0.03%	199	upskirt	0.09%	182
incest	0.01%	90	panic! at the disco	0.03%	152	anime	0.08%	201

Table 8: Categories of most repeated queries in image, audio and video search.

Image		Audio		Video	
Total Classified Queries	2,818	Total queries	1,485	Total queries	911
Query	Frequency (%)	Category	Frequency (%)	Query	Frequency (%)
Person	56.2%	Person	75.7%	Person	64.1%
IndustryTerm	21.7%	Song	8.9%	IndustryTerm	13.3%
MedicalCondition	7.4%	MedicalCondition	2.7%	MedicalCondition	5.3%
Sexual	2.6%	IndustryTerm	2.7%	SportsGame	4.06%
Country	2.5%	City	1.6%	Sexual	2.8%
SportsGame	1.9%	SportsGame	1.5%	Technology	2.5%
Technology	1.8%	Sound	1.5%	City	1.4%
City	1.6%	Technology	1.4%	Song	1.2%
Holiday	0.9%	Holiday	1.2%	MovieOrGame	0.9%
NaturalFeature	0.8%	Country	0.8%	Holiday	0.9%
Region	0.8%	Region	0.4%	Region	0.6%
MovieOrGame	0.4%	Currency	0.2%	Country	0.6%
NaturalDisaster	0.2%	Organization	0.2%	Organization	0.3%
Currency	0.1%	TV	0.2%	Tutorial	0.3%
Organization	0.1%	Product	0.2%	Currency	0.3%
PhoneNumber	0.1%	NaturalFeature	0.1%	Facility	0.2%
Product	0.1%	Event	0.07%	NaturalFeature	0.2%
Music	0.1%			NaturalDisaster	0.2%
				Product	0.1%
Total	100%		100%		100%

Table 9: Top 10 advanced query used in image, audio and video search.

Image			Audio			Video		
Query	Num of queries	Organic Links Followed	Query	Num of queries	Organic Links Followed	Query	Num of queries	Organic Links Followed
romeo and juliet *	22	24	nice and slow	4	73	romeo and juliet *	31	34
ivy and harley	4	12	angels and airwaves	53	72	round and brown	26	27
dead sea and jordan river	3	12	simple and clean	37	54	dead or alive kabuki *	22	24
the reflecting pool and national mall	2	11	wait and bleed	24	41	aladdin and sex	23	23
dead or alive*	7	9	coheed and cambria	30	39	dead or alive *	17	20
mother and child	2	9	guns and roses	22	33	baucis and philemon	14	19
round and brown	7	8	beauty and the beast	20	29	piano and rose	11	18
chronicles of narnia the lion the witch and the wa*	2	8	bob and tom	17	24	mac and bumble	17	17
skull and crossbones	8	7	pomp and circumstance	16	24	monty python and the holy grail	16	17
hutus and tutsis	3	7	cassie me and you	15	22	tarzan and jane porn	16	16

Table 10: Top 10 natural language queries in image, audio and video search.

Image			Audio			Video		
Query	# Queries	Links Followed	Query	# Queries	Links Followed	Query	# Queries	Links Followed
how your body works	2	4	what you know	77	89	how to self suck	19	19
do not enter	2	3	what hurts the most	52	62	how your body works	4	6
how to have sex	2	3	do it to it	22	35	do not enter	4	5
what is adequate water	2	3	when im gone	22	34	what is adequate water	4	5
when hell freezes over	2	3	what you know about that	23	30	how to get in shape	3	3
how to draw manga	1	3	where is the love	13	28	how to tell the sex between frogs	3	3
how to draw a dog	2	2	what you know t i	20	26	how to draw flames	3	3
how to get in shape	2	2	how bad do you want it	4	26	how is cheese made	3	3
how to build a volcano	1	2	is this love bob	2	25	how deep is a cunt	3	3
where is the plant mouth	1	2	what a wonderful world	15	21	how are you feeling today	3	3