World Wide Web **3:** 249–254, 2000.

Searching for multimedia: analysis of audio, video and image Web queries

Bernard J. Jansen [a], Abby Goodrum [b] and Amanda Spink [c]

[a] *Computer Science Program, University of Maryland (Asian Division), Seoul, 140-022 Korea*
E-mail: jjansen@acm.org
[b] *College of Information Science and Technology, Drexel University, 3141 Chestnut St., Philadelphia, PA 19104, USA*
E-mail: Abby.Goodrum@drexel.edu
[c] *School of Information Sciences and Technology, The Pennsylvania State University, 511 Rider I Building, 120 S. Burrowes St., University Park, PA 16801, USA*
E-mail: spink@ist.psu.edu

The development of digital libraries has enhanced the integration of textual and multimedia information in many document collections. The World Wide Web provides the connectivity for many digital library users. Studies exploring the searching characteristics of Web users are an important and a growing area of research. Most Web user studies have focused on general Web searching, regardless of subject matter or format. Little research has examined how Web users search for multimedia information. Our study examines users' multimedia searching on a major Web search service. The data set examined consisted of 1,025,908 queries from 211,058 users of Excite ®, a major Web search service. From this data set, we identified and analyzed queries for audio, image, and video queries. Our findings were compared to results from general Web searching studies. Implications for the design of Web searching services and interfaces are discussed.

1. Introduction

The World Wide Web (Web) is an immense repository of multimedia information [Angelides and Dustar 1997; Lesk 1997a, b] that includes combinations of text, image, video, film or audio artifacts. Many museums and repositories of multimedia information are going online [Takahashi *et al.* 1998]. One can now access world famous art galleries via the Web, such as Monet's work at `http://sunsite.unc.edu/wm/paint/auth/monet/first/`. As of 22 August 1999, Alta Vista `http://www.altavista.digital.com` had indexed approximately 9,983,032 images on the Web[1]. Lawrence and Giles [1999] estimated there are 180 million images on the publicly indexed Web and 3Tb of image data, not including other types of multimedia files, such as audio and video. The hypertext transfer protocol (HTTP) lends itself to the easy transfer of audio, video, and image formats integrated with textual information.

In general, Web users search for multimedia information as they search for textual information [Schauble 1997]. The simplest image search algorithm used by information retrieval (IR) systems locates multimedia files by searching for file extensions and matching the filename to terms in a query [Witten *et al.* 1994]. Some Web IR systems may retrieve on-line documents with embedded multimedia files. The multimedia filename may not match the query terms, but the Web document may contain text that does.

[1] With Alta Vista, one can select the image radio box and enter a '*' (e.g., the wildcard character) into the search box. This will return the number of images in the Alta Vista inverted file index.

Some Web search services provide special mechanisms for multimedia searching. Excite `http://www.excite.com` and Yahoo `http://www.yahoo.com` are two such Web IR systems. The advantage of their approach is that multimedia searching is performed in an identical way to text searching. No additional burden is placed on the searcher. If the searcher desires a multimedia document, the searcher enters a query and specifies a multimedia attribute. For example, a user searching for recordings of *Jimmy Buffet* songs could enter *"Jimmy Buffet songs"* or *"audio of Jimmy Buffet songs."* This may retrieve lyric sheets of Jimmy Buffet's songs, rather than the actual audio files. The searcher could also use audio file extensions, such as avi or wav. The same procedures are utilized for video or image retrieval, using appropriate terms and file extensions for each media. The disadvantage of this approach is the placement of contextual knowledge burden on the searchers, who may not be familiar with multimedia formats.

Web multimedia searching mechanisms include the following: Alta Vista `http://www.altavista.com` searchers can narrow a query to specifically search for an image; Lycos `http://www.lycos.com/` searchers can search for pictures and audio files in MP3 format only; HotBot `http://wwww.hotbot.com` provides searching for image, video, and the MP3 audio format. Some Web search services specialize in multimedia collections. Webseek `http://www.ctr.columbia.edu/webseek/` allows users to search by term or select from general categories of images and video. Webseek returns thumbnail images as the document result list. The system also provides tools for content-based searching for images and videos using color histograms generated from visual scenes.

In this paper we present an overview of research in Web multimedia searching and present results of audio, video, and images data analysis followed by a more in-depth comparison between multimedia and general Web searching characteristics.

2. Related studies

A growing body of research is analyzing users' general Web searching characteristics, with fewer studies examining queries by users seeking multimedia information. Jansen and Pooch [1999] provide an in-depth review of Web user searching studies in general (i.e., without regard to textual or multimedia). Spink *et al.* [1999] conducted research into Web user searching intents. Multimedia searching research has typically focused on image retrieval utilizing indexed image collections [Enser 1995; Goodrum and Kim 1998; Hastings 1995; O'Connor *et al.* 1999; Turner 1990]. Some image research has focused on the design of multimedia IR systems [Aslandogan *et al.* 1997]. Other researchers have investigated audio and video retrieval [Brown *et al.* 1996]. Smith *et al.* [1998] provide analysis on the demand for seeking video when designing a multimedia classroom.

Goodrum and Spink [1999] specifically analyzed users' image queries, terms and sessions using the same data used in our study. Twenty-eight (28) terms were used to identify queries for both still and moving images, resulting in a subset of 33,149 image queries by 9,855 users. They provide data on: (1) *image queries* – the number of search terms, and the use of visual modifiers, (2) *image search sessions* – the number of queries per user, modifications made to subsequent queries in a session, and (3) *image terms* – their rank/frequency distribution and the most highly used search terms. Goodrum and Spink [1999] found a mean of 2.64 image queries per user containing a mean of 3.74 terms per query. Image queries contained a large number of unique terms. The most frequently occurring image related terms appeared less than 10% of the time, with most terms occurring only once. This can be contrasted to earlier work by Enser [1995] who examined written queries for pictorial information in a non-digital environment.

In our study, we analyze a large set of Web multimedia queries from Excite, including image, audio and video queries. We investigate the searching characteristics of Web users seeking multimedia information with implications for Web search services. Our study design generally adheres to the format and definitions of Web studies outlined by Jansen and Pooch [1999]. This analysis is part of a larger ongoing study of Web searching behavior by Jansen *et al.* [1998, 1999] utilizing transaction logs of searches conducted by Excite users.

2.1. Research questions

Our study addresses the following research questions.

1. What are the characteristics of Web users' queries for multimedia, audio and image information?
2. What are the similarities and differences between Web users' multimedia and general search queries?

3. Research design

3.1. Excite data set

Excite `http://www.excite.com` searches are based on the exact terms a user enters in a query; however, capitalization is disregarded, with the exception of logical commands AND, OR, and AND NOT. There is no stemming. The Excite transaction log data set consisted of 1,025,908 records (table 1).

Each record in the transaction log contained three fields:

- *Identification*: an anonymous code assigned by the Excite server to a user machine.
- *Time of Day*: measured in hours, minutes, and seconds from midnight of 16 September 1997.
- *Query:* the query terms exactly as entered by the user.

Our analysis focused on the user's sessions, queries, and term level of analysis. Basically, a *session* is the entire sequence of queries by a particular user. A *query* is the one or more terms entered into the Web IR system. A *term* is any string of characters bounded by white space.

3.2. Data analysis

The data set was loaded into a database management application. Queries in this application were developed that contained multimedia terms. Specifically, there were:

- *audio query* – containing 27 audio related terms,
- *video query* – containing 13 video-related terms,
- *image query* – containing 30 image-related terms.

Figure 1 shows the specific terms used in each query. The queries were case insensitive.

These queries were executed against the database. If a user session contained a query not including any of these terms, that query would not appear in the analysis. Since it is difficult to determine an information need based on a single term, the result lists were reviewed, and the queries that were obviously not multimedia related were removed. When in doubt, the query was not removed from the result lists. We feel confident that majority of the queries in this analysis relate to multimedia searching.

Table 1
Basic data.

1,025,910 queries	211,063 users	2,216,986 terms	2.16 = mean terms per query	4.86 = mean queries per session

Generally, the queries were not altered in anyway. Research by Jansen *et al*. [2000] shows that the cleaning of the query terms (i.e., removing non-alphanumeric characters such as +, −, :, etc.) results in minor changes to the overall results. We did remove leading and trailing + and " characters in the term analysis. Also, as discussed by Jansen and Pooch [1999], concerning Web transaction logs, we are making an assumption in this analysis that the user identification field denotes a searcher, while technically it denotes a computer on which a cookie has been placed. This impacts the analysis, especially on lengthy sessions. These sessions may indicate that the machine is a common use computer.

Audio terms	Image terms	Video terms
au	art'	.avi
.au	bitmap	.mjpeg
audio	Bmp	.mov
av	.bitmap	.mov8
.av	.bmp	.mpeg
band	Camera	.mpg
cd	Cartoon	animated
concerts	Gallery	clip
lyrics	gif	clips
mpz	.gif	drivers
multimedia'	image	mjpeg
music	images	mov
noise	jpeg	movie
song	jpg	movies
songs	pcx	mpeg
sonic	.jpeg	mpg
sonics	.jpg	plugins
sound	.pcx	quicktime
sound card	photo	video
sound cards	photographs	viewers
soundblaster	photograph	avi
sounds	photos	
soundwave	pic	
speakers	pics	
track	.pic	
vocals	.pics	
wav	picture	
.wav	pictures	
	png	
	.png	
	tif	
	tiff	
	.tif.tiff	

Figure 1. Specific terms used in each query.

4. Results

Table 2 provides an overview of the results of the data set analysis. Table 2 presents the median, mean, standard deviation, maximum, and minimum for session length and queries length in each of the three-multimedia categories.

4.1. Audio queries

Findings related to audio queries were:

- 3,810 audio queries representing: 0.37% of all queries were submitted by 1,525 users representing 0.73% of all users from the data set containing 15,661 total terms and 2,101 unique terms.
- Total number of terms represented 0.73% of all total terms in the data set.
- Mean session length for audio searching was 2.44 queries
- Mean query length was 4.11 terms.
- Top occurring audio term was *music*, with 1365 occurrences in the set of audio queries.
- Number of audio queries was extremely small (0.37%) compared to the total set of all queries. This is surprising given the large number of music compact diskettes (CD) and tapes sold each year. The lack of audio search terms may have been due to economic and technical issues concerning delivery of commercial recordings via the Web [Kirsch 1998]. With the acceptance of the MP3 audio standard for Web delivery of commercial audio, the number of audio queries on a particular search engine will probably increase. Already, *MP3* is a top query term on Web IR systems [Nielsen/NetRating 1999].

4.2. Video queries

Findings related to video queries were:

- 7,630 video queries represented 0.74% of all queries submitted by 2,613 users containing 24,514 total terms and 2,725 unique terms representing 1.14% of all total terms in the data set.
- Mean session length for video searching was 2.91 queries.
- Mean video query length was 3.21 terms.

Table 2
Comparison of statistics from the three multimedia categories.

	Audio queries		Video queries		Image queries	
	Number	%	Number	%	Number	%
Total	3810	0.37%	7630	0.74%	27144	2.65%
	Queries/user	Terms/query	Queries/user	Terms/query	Queries/user	Terms/query
Median	2	4	2	3	2	3
Mean	2.44	4.11	2.91	3.32	3.27	3.46
Std dev	2.95	2.67	3.85	1.96	5.49	2.2
Max	51	37	70	44	267	33
Min	1	1	1	1	1	1

Table 3
Top 10 multimedia terms in each category.

Rank	Audio			Video			Image		
	Term	Number	%	Term	Number	%	Term	Number	%
1	music	1365	8.72	movies	1707	6.96	pictures	10571	11.26
2	sound	485	3.10	video	1696	6.92	photos	3507	3.74
3	audio	457	2.92	movie	1289	5.26	pictures	1508	1.61
4	lyrics	340	2.17	videos	860	3.51	pics	1500	1.60
5	cd	333	2.13	clips	428	1.75	photo	1241	1.32
6	song	227	1.45	clipart	219	0.89	gallery	950	1.01
7	songs	225	1.44	pictures	204	0.83	images	875	0.91
8	wav	211	1.35	mpeg	133	0.54	art	809	0.86
9	band	204	1.30	animated	117	0.48	camera	679	0.72
10	sounds	117	.90	avi	117	0.48	photography	579	0.62

- Top occurring video term was *movies*, with a frequency of 1,707 occurrences in the set of video queries.
- Almost twice as many *video* queries as *audio* queries and the other statistics such as number of users and number of terms were along the same line of about twice what the audio queries were. However, the number of video queries was still quite small compared to over data set. As a category, the 0.74 percentage is similar as reported by Jansen *et al.* [1999] where *pictures* was the fifth top ranking category of terms (ranking behind the four categories of *sexual*, *modifiers*, *locations*, and *economic*). So, although the percentage is small, video may represent one of the larger classifications of queries relative to all other categories.

4.3. Image queries

Findings related to image queries were:

- 27,144 image queries representing 2.65% of all queries submitted by 8,310 users representing 4.37% of all users from the data set containing 93,847 total terms and 8,009 unique terms representing 4.37% of all total terms in the data set.
- Mean session length was 3.27 queries.
- Mean query length was 3.46 terms.
- Top occurring image term was *pictures*, with 10,571 occurrences in the set of image queries.
- The number of image queries was by far the largest of the three-multimedia categories. There were seven times more image queries than audio queries and image queries were 3.5 times more frequent than video queries. Also, 4.37% of the users searching for some type of image are not an insignificant number of users. With a user population of this size, it seems it would be worthwhile for an IR system to provide some mechanism to facilitate image searching.

Overall, multimedia queries formed a small proportion (less than 4%) of users' queries. However, when Excite users were searching for multimedia, they were more likely to search for images than audio or video. Audio queries were the smallest proportion of multimedia queries, but they were slightly longer than video or image queries.

4.4. Term analysis

Table 3 lists the top ten terms used for multimedia searching. The terms are listed from the top ranked term to the tenth ranked term by frequency of occurrence in queries from that category. Number is the frequency of occurrence, e.g., the number one ranked audio term (e.g., *music*) occurred 1,365 times in the audio queries. The % is the percentage that this number represents of all terms from all queries in that multimedia query category.

In table 3 there is surprising little overlap between categories, with 'pictures' being the only top term to appear in more than one category. This is surprising because these multimedia formats are typically used in combination, especially audio and image files, and one would expect some overlap among the categories. Across all three categories, there appear to be three or four terms that dominate – *music*, *movies*, *video*, *movie*, and *pictures*. For Web site designs, these terms should be included in the meta-data of appropriate Web sites. For Web IR systems that desire to cater to multimedia searcher, these terms probably should be available via some interface mechanism.

5. Discussion

In analyzing trends in multimedia queries, one can compare and contrast-searching characteristics, such as those listed in table 2, in each of the three categories – video, audio and image. From table 2, we note that the median session in all cases was 2 queries and the average was varied from 2.44 for audio queries to 3.27 for image queries. These figures are generally higher than those reported from general Web searching, where the mean session was 1 query and the mean was 2.84 queries.

With respect to the query level of analysis, the median query length varied with 3 terms for video and image queries and 4 terms for audio queries. The mean query length ranged from 3.32 terms for video queries to 4.11 terms for audio queries. We compared these statistics to general Web searching characteristics, using data from [Jansen *et al.* 1999].

As might be expected, these figures are higher than general Web searching, where the reported median was 1 term and the average was 2.21 terms. The higher figure is expected due to the need by the searcher to add a multimedia term to the query. However, these findings suggest that multimedia searching may place additional cognitive load on the searcher that the IR system should address.

Our findings also highlight several key aspects of multimedia searching. First, the number of users searching for multimedia documents, especially images, suggests a need to provide Web mechanism to facilitate this searching. Second, multimedia sessions and queries are still short compared to traditional IR system searching, but longer than general Web sessions and queries. There is little query reformulation for the majority of users. This may also suggest either a problem with the Web IR system or that the precision of the Web IR system has satisfied the searcher's information need. Third, there appear to be a small number of multimedia terms that occur frequently and a large number of terms that occur very infrequently.

Our analysis of these Web multimedia queries indicates that multimedia searching may be challenged by a lack of representational congruity. There are four areas that affect the outcome of IR system interaction with respect to representational congruity:

1. The extent to which document representations share congruence with the documents for which they stand.
2. The extent to which queries share congruence with the information needs for which they stand.
3. The extent to which queries and document representations share congruence with each other.
4. The extent to which representations of retrieved items support user's relevance judgments.

Problems arise when either documents or information needs cannot be expressed in a manner that will provide congruence between the representation and its referent. In this case, the problem seems to be in representing audio and image information needs with textual queries, or with representing retrieved multimedia documents as short textual abstracts. In order to express a non-textual information need in only textual terms, the user takes on an additional cognitive load. In order to make relevance judgments, the user must visually inspect the full record in order to know if the retrieved document contains the requested multimedia information. The use of textually bounded systems for the retrieval of multimedia results in an increase in the contextual load placed on the user, as is evidenced by the number of terms and the number of queries needed to retrieve multimedia objects on the Web.

6. Conclusion and further research

Although it may not be possible at this time to provide users with non-textual mechanisms for querying a search engine's database, it is not difficult to provide tools to assist users in specifying a multimedia information need and retrieving information with media file extensions. What is more challenging at this time is the provision of multimedia surrogates in the retrieved item list. Extracting thumbnails and sound bites from web pages and ranking them for relevancy are areas of much needed future research.

Acknowledgements

The authors gratefully acknowledge the assistance of Excite, Inc., in providing the data for this research. Without the generous data sharing by Excite, Inc., this research would not be possible. We also acknowledge the generous support of our institutions for this research.

References

Angelides, M. and S. Dustdar (1997), *Multimedia Information Systems*, Kluwer, Boston.

Aslandogan, Y., C. Thier, C. Yu, J. Zou, and N. Rishe (1997), "Using Semantic Contents and WordNet in Image Retrieval," In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Publications, pp. 286–295.

Brown, M., J. Foote, G. Jones, K. Spärck Jones, and S. Young (1996), "Open-Vocabulary Speech Indexing for Voice and Video Mail Retrieval", In *Proceedings of the 4th ACM International Multimedia Conference, ACM Multimedia 96*, pp. 307–316.

Enser, P.G.B. (1995), "Progress in Documentation: Pictorial Information Retrieval," *Journal of Documentation 51*, 2, 126–170.

Goodrum, A. and C. Kim (1998), "Visualizing the History of Chemistry: Queries to the CHF Pictorial Collection," Report to the Chemical Heritage Foundation Pictorial Collection, `http://www.chemheritage.org/Publications/ChemHeritage/Goodrum/goodrum.htm`

Goodrum, A. and A. Spink (1999), "Visual Information Seeking: A Study of Image Queries on the World Wide Web," In *Proceedings the 1999 Annual Meeting of the American Society for Information Science*, Washington, DC, November 1999, pp. 665–674.

Gordon, M. and P. Pathak (1999), "Finding Information on the World Wide Web: The Retrieval Effectiveness of Search Engines," *Information Processing and Management 35*, 2, 141–180.

Hastings, S.K. (1995), "Query Categories in a Study of Intellectual Access to Digitized Art Images," In *Proceedings of the 1995 Annual Meeting of the American Society for Information Science 32*, pp. 3–8.

Jansen, B.J. and U. Pooch (1999), "Web Use Studies: A Review of Current and Frame for Future Research," *Journal of the American Society for Information Science*.

Jansen, B.J., A. Spink, and T. Saracevic (1998), "Failure Analysis in Query Construction: Data and Analysis From a Large Sample of Web Queries," In *Proceedings of the Third ACM Conference on Digital Libraries*, Pittsburgh, PA, pp. 289–290.

Jansen, B.J., A. Spink, and T. Saracevic (1999), "Relevance Feedback," In *Proceedings of WebNet 99: The World Conference of the World Wide Web, Internet, and Intranet*, October 1999, Hawaii.

Jansen, B.J., A. Spink, and T. Saracevic (2000), "Real Life, Real Users and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing and Management 36*, 2, 207–227.

Kirsch, S. (1998), "Everything You Need to Know about the Internet," Retrieved from the World Wide Web on 23 August 1999 at `http://topgun.inforseek.com/stk/presentationa/sigir.ppt`.

Lawrence, S. and C.L. Giles (1999), "Accessibility of Information on the Web," *Nature 400*, 107–109.

Lesk, M. (1997a), "Going Digital," *Scientific American 276*, 3, 58–60.

Lesk, M. (1997b), *Practical Digital Libraries: Books, Bytes, and Bucks*, Morgan Kaufman, San Francisco, CA.

Nielsen/NetRating (1999), Retrieved from the World Wide Web on 24 August 1999 from `http://www.nielsen-netrating.com/`.

O'Connor, B., M. O'Connor, and J. Abbas (1999), "Functional Descriptors of Image Documents: User-Generated Captions and Response Statements," *Journal of the American Society for Information Science 50*, 8, 681–697.

Schauble, P. (1997), *Multimedia Information Retrieval*, Kluwer, Boston.

Smith, T., A. Ruocco, and B.J. Jansen (1998), "Digital Video in Education," In *Proceedings of the 13th SIGCSE Technical Symposium on Computer Science Education*, pp. 122–126.

Spink, A., J. Bateman, and B.J. Jansen (1999), "Searching the Web: A Survey of Excite Users," *Internet Research: Electronic Networking Applications and Policy 9*, 2, 117–128.

Takahashi, J., T. Kushida, J. Hong, S. Sugita, Y. Kurita, R. Rieger, W. Martin, G. Gay, J. Reeve, and R. Loverance (1998), "Global Digital Museum Multimedia Information Access and Creation on the Internet," In *Proceedings of the 3rd ACM Conference on Digital Libraries*, pp. 244–253.

Turner, J. (1990), "Representing and Accessing Information in the Stockshot Database of the National Film Board of Canada," *The Canadian Journal of Information Science 15*, 1–22.

Witten, I.H., A. Moffat, and T.C. Bell (1994), *Managing Gigabytes: Compressing and Indexing Documents and Images*, Van Nosstrand-Reinhold, New York.