SEARCHING FOR MULTIMEDIA: AN ANALYSIS OF AUDIO, VIDEO, AND IMAGE WEB QUERIES

Bernard J. Jansen
Computer Science Program
University of Maryland (Asian Division)
Seoul, 140-022 Korea
E-mail: *jjansen@acm.org*

Abby Goodrum
College of Information Science and Technology
Drexel University
3141 Chestnut St.
Philadelphia PA 19104
E-mail: goodruaa@drexel.edu

Amanda Spink
School of Information Sciences and Technology
The Pennsylvania State University
University Park PA 16801
E-mail: spink@ist.psu.edu

Go to Publication List

ABSTRACT

The development of digital libraries has led to the integration of textual and multimedia information in many document collections. The World Wide Web provides the necessary connectivity for many users of these digital libraries. Studies exploring the searching characteristics of Web users are an important and growing area of research. Most Web user studies have focused on Web searching in general, regardless of subject matter or format. Little research has examined how Web users search specifically for multimedia information. This study examines users' multimedia searching on a major Web information retrieval system. The data set examined consisted of 1,025,908 queries from 211,058 users of Excite ®, a major Web search engine. From this data set, terms were used to identify queries for audio, image, and video queries. The queries were isolated and examined at various levels of analysis. Our findings were compared to data from previous, more general, Web searching studies. Implications for the design of Web information retrieval systems and interfaces are discussed.

INTRODUCTION

The World Wide Web (Web) is an immense repository of multimedia information (Angelides & Dustdar, 1997; Lesk, 1997). Multimedia information may include combinations of text, image, video, film or audio artifacts. Many museums as repositories of multimedia information are going online (*Takahashi, Kushida, Hong, Sugita, Kurita,*

*Rieger, Martin, Gay, Reeve & Loverance. 1998). One can now visit 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 (Jansen, 1999). 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 must search for multimedia information as they would 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 the query (Witten, Moffat & Bell, 1994). Some Web IR systems may retrieve on-line documents that are primary textual but with embedded multimedia files. The multimedia filename may not match the query terms, but the Wed document may contain text that does.

Many Web IR systems provide no special mechanism for multimedia searching. Excite (http://www.excite.com) and Yahoo (http://www.yahoo.com) are two such Web IR systems. The advantage of this approach is that multimedia searching is performed in an identical manner to text searching. No additional burden is placed on the searcher. If the searcher desires a multimedia document, the searcher enters the query and specifies some 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 query might very well 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 would be utilized for video or image retrieval, using appropriate terms and file extensions for each media. The disadvantage of this approach is that it places more contextual knowledge burden on the searchers, who may not be familiar with multimedia formats. Cognitive load is further challenged by necessitating that users translate a non-semantic information need into a textual query. This creates what some authors refer to as a lack of representational congruity (Goodrum in review), or as a *semantic gap* (Gudivada & Raghavan, 1995). This problem is usually exacerbated by the presentation of retrieved items as text-only entries in a list rather than as thumbnail images, sound bites, or video keyframes.

Some Web IR systems provide mechanisms for users searching for multimedia, e.g., by radio boxes or media specific search syntax. 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://www.hotbot.com) provides searching for image, video, and the MP3 audio format. Some Web IR systems 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. Both Webseek and Alta Vista returns thumbnail images and file names in the document result list. Webseek also provides tools for

content-based searching for images and videos using color histograms generated from the visual scenes.

The next section of the paper discusses related research to our study.

RELATED STUDIES

There is a growing body of research analyzing users' general Web searching characteristics, with fewer studies specifically 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, Bateman, Jansen (1999) present research concerning the intent of Web searchers on a Web IR system.

Multimedia searching research has typically focused on the retrieval of images utilizing indexed image collections (Enser, 1995; Goodrum & Kim, 1998; Hastings, 1995; O'Connor, O'Connor & Abbas, 1999; Turner, 1990). Some image research has focused on the design of multimedia IR systems (*Aslandogan, Thier, Yu, Zou & Rishe,* 1997). Other researchers have investigated audio and video retrieval (*Brown, Foote, Jones, Spärck Jones &Young,* 1996). Smith, Ruocco and Jansen (1998) provided 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 set used for our study. In Goodrum and Spink (1999), 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 provided data on: (1) *image que*ries -- 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. They 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 percent of the time, with most terms occurring only once. This analysis contrasted with earlier work by Enser (1995), who examined written queries for pictorial information in a non-digital environment.

In this research, we focus on a large set of Web multimedia queries from Excite, including image, audio and video queries. We sought to investigate the searching characteristics of Web users as they search for multimedia information with implications for Web IR system design. The design of this study generally adheres to the format and definitions for Web studies outlined by Jansen and Pooch (1999). This analysis is part of a larger ongoing study of Web searching behavior by Jansen, Spink and Saracevic (1998,1999a, in press) utilizing transaction logs of searches conducted by Excite users.

The next section of the paper discusses the research questions addressed by this study.

RESEARCH QUESTIONS

This 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?

The next section of the paper describes the research design used in our study.

RESEARCH DESIGN

Excite Data Set

Founded in 1994, Excite, Inc. is a major Internet media public company that offers free Web searching and a variety of other services. The company and its services are described in more detail at its Web site (http://www.excite.com). Excite searches are based on the exact terms that a user enters in the query; however, capitalization is disregarded, with the exception of logical commands AND, OR, and AND NOT. There is no stemming. An online thesaurus and concept linking method called Intelligent Concept Extraction (ICE) is used to find related terms in addition to terms entered. Some of the advanced search features are:

- Boolean operators AND, OR, AND NOT, and parentheses can be used in ALL CAPS and with a space on each side. When using the Boolean operator the ICE (concept-based search mechanism) is turned off.

- A set of terms enclosed in quotation marks (no space between quotation marks and terms) returns Web sites with the terms as a phrase in the exact order they were entered.

- A + (plus) sign before a term (no space) requires that the term must be in an answer. A – (minus) sign before a term (no space) requires that the term must NOT be in an answer. We denote plus and minus signs, and quotation marks, as modifiers.

- There is a clickable option *More Like This*, that is a relevance feedback mechanism to find similar documents.

For a complete explanation of Excite's searching capabilities see (http://www.excite.com).

The transaction log data set consisted of 1,025,908 records. Each action record contained three fields, which were:

- *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 terms. 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.

The next section of the paper discusses the data analysis techniques used in our study.

Data Analysis

The data set was loaded into a database management application. Queries that contained multimedia terms were developed in this application. Specifically, the queries and the number terms utilized in the queries 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.

Figure 1: List of terms used to identify queries.

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	

These queries were executed against the database of 1,025,908 Web queries. If a user session contained a query that did not use any of these terms, that query would not appear in the analysis. Since it is difficult to determine an user's 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.

Generally, the queries were not altered in anyway. Research by Jansen, Spink and Saracevic (in press) 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. This impacts the analysis, especially on lengthy sessions. These sessions may indicate that the machine is a common use computer.

In the next section of the paper, we present results in separate sections of audio, video, and images data analysis followed by a more in-depth comparison between multimedia and general Web searching characteristics.

RESULTS

Table 1 provides an overview of the results of the data set analysis.

Table 1: 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 1 presents the median, mean, standard deviation, maximum, and minimum for session length and queries length in each of the three-multimedia categories. The findings are discuss below.

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).

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.
- 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, Spink & Saracevic, 1999a) 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.

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 is 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.

The next section of the paper examines the terms most frequently used to find multimedia information on the Web.

Term Analysis

Table 2 lists the top ten terms used for multimedia searching.

Table 2: Top 10 multimedia terms in each category.

	Audio			Video			Image		
Rank	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

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.

As shown in Table 2, 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 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, discounting stemming. These terms are *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 searchers, these terms probably should be available via some interface mechanism.

The next section of the paper discusses the major findings of our study.

DISCUSSION

In analyzing trends in multimedia queries, one can compare and contrast-searching characteristics, such as those listed in Table 1, in each of the three categories - video, audio and image. From Table 1, 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, Spink and Saracevic (1999a). 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 also suggest that multimedia searching may place additional cognitive load on the searcher by requiring that non - semantic information needs be represented textually. This representational congruity at the query level is an issue that the IR system should address to better assist users during the search process.

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 and possibly for viewing of results. Second, multimedia sessions and queries are still short compared to traditional IR system searching, but longer relative to general Web sessions and queries. There is little query reformulation for the majority of users. This may 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. Web IR systems should capitizalize on the frequently occurring terms and offer thesaurus-type assistance for infrequently occurring query terms.

CONCLUSION & FURTHER RESEARCH

Our analysis of these Web multimedia queries indicates that users engaged in 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 (e.g., how well file names and surrounding text on a web page represent embedded sound and image files.)
2. The extent to which queries share congruence with the information needs for which they stand (e.g., how well the usually textual queries represent the multimedia needs of the users).
3. The extent to which queries and document representations share congruence with each (e.g., the degree of match between the filenames and other text used to index multimedia and the terms used in queries.)
4. The extent to which representations of retrieved items support user's relevance judgments (e.g., how well the entries, usually textual, in the results list represent the underlying image documents and how this affects the user's interaction with the system.).

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 the case of multimedia searching, there are problems in representing audio, video, and image information needs with textual queries, and with representing retrieved multimedia

documents as short textual abstracts. 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. 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 many times visually inspect the full record in order to know if the retrieved document contains the requested multimedia information.

Although it may not be possible at this time to provide users with non-textual mechanisms for querying a Web IR system's database, tools can be provided 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. The provision of extracted thumbnails and sound bites from web pages for relevance judgments and query reformulation are areas of potential benefit for future research.

ACKNOWLEDGMENTS

REFERENCES

Angelides, M., & Dustdar, S. (1997). *Multimedia information systems*. Kluwer: Boston.

Aslandogan, Y., Thier, C., Yu, C., Zou, J., & Rishe, N. (1997). Using semantic contents and WordNet in image retrieval. *Proceedings of the 20th annual international ACM SIGIR conference on research and development in information retrieval (pp. 286 - 295).*

Brown, M., Foote, J., Jones, G., Spärck Jones, K., &Young, S. (1996). Open-vocabulary speech indexing for voice and video mail retrieval. *Proceedings of the fourth ACM international multimedia conference on Proceedings 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. & Spink, A. (1999) Visual Information Seeking: A study of image queries on the World Wide Web. *Proceedings the 1999 Annual Meeting of the American Society for Information Science*, Washington, DC. November, 1999.

Goodrum, A. (in review) "Multidimensional Scaling of Video Surrogates," Journal of the American Society of Information Science.

Goodrum, A., & 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

Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: The retrieval effectiveness of search engines. *Information Processing and Management*, 35(2), 141 – 180.

Gudivada, V.V., & Raghavan, V.V. (1995) "Content-based image retrieval systems," *IEEE Computer*, 28(9), 18-22.

Hastings, S. K. (1995). Query categories in a study of intellectual access to digitized art images. *Proceedings of the 1995 Annual Meeting of the American Society for Information Science, 32,* 3-8.

Jansen, B. J. & Pooch, U. (under review). Web use studies: A review of current and frame for future research. Submitted to *Journal of the American Society of Information Science*.

Jansen, B. J. (1999). *Note on retrieval of number of images indexed at Alta Vista.* 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.

Jansen, B. J., Spink, A., & Saracevic, T. (1998). Failure analysis in query construction: Data and analysis from a large sample of Web queries. *Proceedings of the Third ACM Conference on Digital Libraries, Pittsburgh, PA*. (pp. 289-290).

Jansen, B. J., Spink, A., & Saracevic, T. (in press). Real life, real users and real needs: A study and analysis of user queries on the Web. *Information Processing and Management.*

Jansen, B., Spink, A., & Saracevic, T. (1999). The Use of Relevance Feedback on the Web: Implications for Web IR System Design. *Proceedings of WebNet 99: The World Conference of the World Wide Web, Internet, and Intranet, October, 1999, Hawaii*.

Kirsch, S. (1998). Everything you need to know about the Internet. Retrieved from the World Wide Web on 23 August 1999 at http://topgun.infoseek.com/stk/presentations/sigir.ppt.

Lawrence, S., & Giles, C. L. (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.

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

O'Connor, B., O'Connor, M., & Abbas, J. (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., Ruocco, A., & Jansen, B. (1998). Digital Video in Education. *Proceedings of the Thirtieth SIGCSE Technical Symposium on Computer Science Education,* 122 – 126.

Spink, A., Bateman, J., & Jansen, B. (1999). Searching the Web: A Survey of Excite Users. Internet Research: Electronic Networking Applications and Policy.

Takahashi, J., Kushida, T. Hong, J., Sugita, S., Kurita, Y., Rieger, R., Martin, W., Gay, G. Reeve, J., & Loverance, R. (1998). Global digital museum multimedia information access and creation on the Internet. *Proceedings of the third ACM Conference on Digital Libraries, 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., Moffat, A., & Bell, T. C. (1994*). Managing gigabytes: Compressing and indexing documents and images*. Van Nosstrand Reinhold: New York.