

# An Analysis of Multimedia Searching on AltaVista

**Bernard J. Jansen**

School of Information Sciences and  
Technology  
The Pennsylvania State University  
2P Thomas Building, University Park  
PA USA 16802  
[jjansen@ist.psu.edu](mailto:jjansen@ist.psu.edu)

**Amanda Spink**

School of Information Sciences  
University of Pittsburgh  
135 N. Bellefield Avenue  
Pittsburgh PA 15260 USA  
[aspink@mail.sis.pitt.edu](mailto:aspink@mail.sis.pitt.edu)

**Jan Pedersen**

Chief Scientist, Overture  
Web Services Division  
1070 Arastradero Road  
Palo Alto, CA USA 94304  
[jan.pedersen@av.com](mailto:jan.pedersen@av.com)

## ABSTRACT

Web searching is a significant activity for many people seeking multimedia information. Major Web search engines, such as Alta Vista, are essential tools in the quest to locate relevant online information. As such, it is important that we understand how searchers utilize these Web information systems. This paper presents research that examines characteristics of multimedia Web searching on Alta Vista. More specifically, the research questions driving this study are: (1) What are the characteristics of multimedia searching on Alta Vista? and (2) How does this multimedia searching compare to Web searching in general? The results of our research show that multimedia searching is complex relative to general Web searching and that searching specific multimedia collections does not necessarily reduce the searching complexity. We discuss the implications of the findings for the development of online multimedia retrieval systems.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - *Search process*; H.3.5 [Information Storage and Retrieval]: Online Information Services— *Web-based services*.

## General Terms

Human Factors, Experimentation, Measurement

## Keywords

Multimedia searching, Web searching, search engines

## 1. INTRODUCTION

Web searching has become a daily behavior for many people, with the Web now the first choice for many people seeking information [1]. Most major search engines support some type of multimedia searching, and there are several image and multimedia specific search engines, including ImageScape [2] and Webseek [3]

PERMISSION TO MAKE DIGITAL OR HARD COPIES OF ALL OR PART OF THIS WORK FOR PERSONAL OR CLASSROOM USE IS GRANTED WITHOUT FEE PROVIDED THAT COPIES ARE NOT MADE OR DISTRIBUTED FOR PROFIT OR COMMERCIAL ADVANTAGE AND THAT COPIES BEAR THIS NOTICE AND THE FULL CITATION ON THE FIRST PAGE. TO COPY OTHERWISE, OR REPUBLISH, TO POST ON SERVERS OR TO REDISTRIBUTE TO LISTS, REQUIRES PRIOR SPECIFIC PERMISSION AND/OR A FEE.

MIR '03, NOVEMBER 7, 2003, BERKELEY, CALIFORNIA, USA.  
COPYRIGHT 2003 ACM 1-58113-778-8/03/00011...\$5.00.

Given the Web's importance, we need to understand how people use and interact with Web search engines to locate multimedia information. Examining Web multimedia searching is an important area of research with potential to increase our understanding of multimedia searching, advance our knowledge of user information needs, and positively impact the design of online multimedia systems. This understanding will assist in addressing of the many challenges of multimedia retrieval [4].

To our knowledge, there has been limited large-scale research examining Web multimedia searching [5, 6]. Although the limited studies conducted provide important insights into multimedia Web searching, further research is needed to validate these results across search engines and time. This is especially important as Web information systems are continually undergoing incremental changes. Research is needed to evaluate the effect of these changes on system performance and user searching behaviors over time.

We address these needs in the present research by examining Web searchers using Alta Vista<sup>1</sup>, a major U.S. Web search engine. We examine the effect of multimedia radio buttons, an innovation in multimedia searching, and one that has not yet been studied. We analyze multimedia searching characteristics, including session duration, query length, results pages viewed and term usage.

We describe our research design and our analysis of the Alta Vista Web search engine data, along with a discussion of results. We then discuss the key findings and the implications of our research results for online multimedia system users and system designers. We conclude with directions for future research.

## 2. RELATED STUDIES

There is a growing body of research examining the use of Web search engines [7-9]. Jansen and Pooch [7] present an extensive review of the Web searching literature, comparing Web searchers with searchers of traditional information retrieval systems and online public access catalogues. The researchers report that Web searchers exhibit different search characteristics than do searchers of other information systems.

Spink, Jansen, Wolfram and Saracevic [8] provide a four year analysis of searching on the Excite search engine. They report that Web searching has remained relatively stable over time, although they noted a shift from entertainment to commercial searching. The researchers show that on the Excite search engine Web

---

<sup>1</sup> <http://www.altavista.com>

searching sessions are very short as measured by number of queries. Users view a very limited number of results pages<sup>2</sup>, with the majority of Web searchers, approximately 80%, viewing no more than 10 to 20 Web documents. These characteristics have remained fairly constant across multiple studies.

In the area of multimedia searching, there has been dramatically less published research [5, 6]. Goodrum and Spink [5] analyzed image queries, and Jansen, Goodrum, and Spink [6] analyzed audio, image, and video sessions and queries from the Excite search engine from multiple years. The researchers noted that multimedia sessions and queries are generally longer than general Web queries indicating an increased cognitive load for multimedia searching. Both of these studies were on the general database and did not examine the effect of separate multimedia collections on Web searching.

In this research, we seek to address the issue of no investigation of Web searching within specific multimedia databases by examining the searching patterns of Alta Vista users. We present results from analysis of both general searching and multimedia content collections.

### 3. RESEARCH QUESTIONS

More specifically, the research questions driving this study are:

- 1) What are the characteristics of Alta Vista multimedia searching?
- 2) How does multimedia searching compare to Web searching in general?

The broader goal of our study is to gauge the effect of separate multimedia collections on Web searching characteristics and thereby effect the design of multimedia retrieval systems.

### 4. RESEARCH DESIGN

In 2002, Alta Vista was the 7<sup>th</sup> most popular search engine [10] in terms of unique visitors and had a content collection of nearly approximately a billion Web pages [11]. Alta Vista supported several query operators including AND, OR, NOT, NEAR, MUST APPEAR, MUST NOT APPEAR, and PHRASE operators [12].

From this information, we see that AltaVista offers a full range of searching options, has an extremely large content collection, and millions of unique visitors per month. After being an independent company for several years, Overture Services purchased AltaVista in 2003 [13], and Yahoo! [14] then purchased Overture Services.

#### 4.1 Data Collection

Many Web searchers are interested in multimedia content, such as audio, images, and video files, although the interest is predominantly in images. The Alta Vista search engine uses a straightforward ontology designed to assist users in searching for multimedia, namely radio buttons that target searches for specific audio, image or video content collections. Alta Vista resolves these multimedia searches against separate content collections.

---

<sup>2</sup> When a Web search engine user submits a query, the search engine returns the results in blocks, of usually about 10 results. These “chucks” are referred to as *Results Pages* and are presented to the user sequentially from the top most ranked results page to the maximum number of results retrieved by the search engine.

Each collection (i.e., audio, image, and video) has different indexing features and relevance ranking functions based on textual attributes of the particular multimedia file. Many have proposed content based indexing [15, 16]; however, these contextual based indexing methods such as the methods used by Alta Vista, have proven quite effective within the Web searching environment.

To address our research questions, we obtained, and quantitatively analyzed, actual queries submitted to Alta Vista main search box in 2002. We also analyzed queries in from each of the three multimedia content collections. The queries were submitted to Alta Vista on 8 September 2002 and span a 24-hour period. We recorded the queries in four transaction logs (*general*, *audio*, *image*, and *video*) and represent a portion of the searches executed on the Web search engine on this particular date.

The original general transaction log contains approximately 3,000,000 records. Each record contains three fields:

- (1) *Time of Day*: measured in hours, minutes, and seconds from midnight of each day as recorded by the Alta Vista server,
- (2) *User Identification*: an anonymous user code assigned by the Alta Vista server, and
- (3) *Query Terms*: terms exactly as entered by the given user. Using these three fields, we located the initial query and then recreated the chronological series of actions in a session.

We adopt the terminology outlined by Jansen and Pooch [7] in this research. Specifically,

- (1) A *term* is any series of characters separated by white space or other separator.
- (2) A *query* is the entire string of terms submitted by a searcher in a given instance, and
- (3) A *session* is the entire series of queries submitted by a user during one interaction with the Web search engine.

When a searcher submits a query, then views a document, and returns to the search engine, the Alta Vista server logs this second visit with the identical user identification and query, but with a new time (i.e., the time of the second visit). This is beneficial information in determining how many of the retrieved results pages the searcher visited from the search engine, but unfortunately it also introduces duplicate queries in the transaction log.

To address this issue, we collapsed the four data sets by combining all identical queries submitted by the same user to give us the unique queries for analyzing sessions, queries and terms, and pages of results viewed [7]. We utilized the complete un-collapsed sessions in order to obtain an accurate measure of the temporal length of sessions and the number of results pages visited.

For analysis of the multimedia data sets, we followed the same procedure and compared the results from the analysis of the general Alta Vista transaction log to the results from the analyses of the audio, image and video transaction logs.

## 5. RESULTS

We present the aggregate results for the analysis in Table 1.

**Table 1. Comparison of general, audio, image, and video searching in 2002**

	General	Audio	Image	Video
<b>Sessions</b>	369,350	3,181	26,720	5,789
<b>Queries</b>	1,073,388	7,513	127,614	24,265
<b>Terms</b>				
<i>Unique</i>	297,528 (9.5%)	6,199 (33.4%)	71,873 (14.1%)	8,914 (19.1%)
<i>Total</i>	3,132,106 (100%)	18,544 (100%)	510,807 (100%)	46,708 (100%)
<b>Mean terms per query</b>	2.91 (sd=4.77)	2.47 (sd=1.62)	4.00 (sd=3.21)	1.92 (sd=1.09)
<b>Terms per query</b>				
<i>1 term</i>	218,628 (20%)	2,128 (28%)	27,808 (22%)	9,465 (39%)
<i>2 terms</i>	330,875 (31%)	2,652 (35%)	40,472 (32%)	9,979 (41%)
<i>3+ terms</i>	523,885 (49%)	2,733 (36%)	59,334 (46%)	4,814 (20%)
<b>Mean queries per user</b>	2.91 (sd=4.77)	2.36 (sd=3.85)	4.78 (sd=10.44)	4.19 (sd=6.14)
<b>Users modifying queries</b>	193,468 (52%)	1,496 (47%)	14,838 (56%)	3,350 (58%)
<b>Session size</b>				
<i>1 query</i>	175,882 (48%)	1,685 (53%)	11,882 (44%)	2,439 (42%)
<i>2 queries</i>	75,343 (20%)	657 (21%)	4,759 (18%)	1,016 (18%)
<i>3+ queries</i>	118,125 (32%)	839 (26%)	10,079 (38%)	2,334 (40%)
<b>Results Pages Viewed</b>				
<i>1 page</i>	781,483 (72.8%)	5,551 (73.9%)	80,455 (63.0%)	13,357 (55.0%)
<i>2 pages</i>	139,088 (13.0%)	1,070 (14.2%)	14,498 (11.1%)	3,905 (16.1%)
<i>3+ pages</i>	150,904 (14.1%)	892 (11.9%)	32,661 (25.65)	1,949 (28.9%)
<b>Boolean Queries</b>	61,065 (6%)	210 (3%)	35,955 (28%)	299 (1%)
<b>Terms not repeated in data set</b>	176,196 (6%)	3,720 (20%)	35,955 (7%)	5,292 (11%)
<b>Use of 100 most frequently occurring terms</b>	592,699 (19%)	4,889 (26%)	26,621 (5%)	17,745 (38%)

In comparing the four types of searching (general, audio, image, and video), we see that the use of unique terms in audio searching (33%) is substantially higher than the other types of Web searching (10% to 19%), indicating that searching for audio medium utilizes a broader jargon.

The mean terms per query for image searching was notably larger (4 terms per query) than the other categories of searching, which were all less than 3 terms.

Video searchers also viewed more pages of results than other searchers, with 45% of video searchers viewing more than one results pages.

The session lengths for image searchers were longer than any other type of searching, although video sessions were also

relatively lengthy. The session lengths of image searches when combined with the longer queries may indicate the image searching is a more difficult cognitive task than other types of searching.

Another indicator of the complexity of image searching is Boolean usage, which was 28%. This is more than four times the next highest category of general Web searching.

Audio searching had, by far, the highest percentage of terms not repeated in the data set (20%) and the highest percentage usage of most utilized terms (26%).

To our knowledge, there has been no previously published analysis of Web searching using a multimedia ontology such as AltaVista's (i.e., radio buttons). However, there has been previous

research on multimedia searching on the Web [5, 6, 17]. These studies report on multimedia searching using the standard interface methodology for general Web searching (i.e., text box).

Table 2 presents a comparison of the means and standard deviation using data from Jansen and colleagues [6]. If we view

increased query length and increased session length as indicators of searching complexity, we see from Table 2 that the use of a relative simple interface (i.e., radio buttons for specific multimedia type) has reduced the complexity of audio and video searching. It has not had the same effect on image searching.

**Table 2. Comparison of audio, image, and video searching with and without multimedia interface**

	Audio		Image		Video	
	With	Without *	With	Without *	With	Without *
<b>Mean Terms Per Query</b>	2.47 (sd=1.62)	4.11 (sd=2.67)	4.00 (sd=3.21)	3.46 (sd=2.20)	1.92 (sd=1.09)	3.32 (sd=1.96)
<b>Mean Queries Per User</b>	2.36 (sd=3.85)	2.44 (sd=2.95)	4.78 (sd=10.44)	3.27 (sd=5.49)	4.19 (sd=6.14)	2.91 (sd=3.85)

Note: \* Data from [6]

We next examined query lengths. Research shows that increased query length measured in number of terms is directly related to retrieval effectiveness [18]. Table 3 presents the query length percentages for each data set.

**Table 3. Query lengths in each data set.**

Query Length	Percentage of All Queries			
	General	Audio	Image	Video
0	0.03%	0.1%	0.04%	0.03%
1	20.4%	28.3%	21.8%	39.0%
2	30.8%	35.3%	31.7%	41.1%
3	22.8%	17.5%	11.9%	13.2%
4	12.0%	9.5%	3.9%	4.1%
5	5.9%	4.7%	2.7%	1.6%
6	2.5%	2.2%	0.5%	0.5%
7	1.2%	1.0%	0.2%	0.2%
8	0.5%	0.5%	0.1%	0.1%
9	3.5%	0.5%	27.0%	0.1%
>= 10	0.4%	0.4%	0.2%	0.1%

From Table 3, we see that generally the percentages are similar at each length although with some variation. A notable exception is image queries with length of 9 terms at 27%. Naturally this skewed the average. We cannot account for this clustering at this query length. We examined this in query length in more detail. There were several individual images sessions, not a single user. We hypothesize that this is an

anomaly of the data collection. Over a longer collection period, the percentage would be more evenly spread at the longer query lengths for image searchers.

We next examined the use of terms. Table 4 presents the term – frequency for the top twenty-five most frequently occurring terms within each data set, after removal of stop words.

**Table 4. Top 25 Most Frequently Occurring Terms in Each Data Set.**

General		Audio		Image		Video	
Total Terms	3,132,106	Total Terms	18,544	Total Terms	510,807	Total Terms	46,708
Term	Frequency	Term	Frequency	Term	Frequency	Term	Frequency
free	0.6%	mp3	1.0%	nude	0.6%	sex	3.7%
sex	0.2%	music	0.8%	sex	0.4%	free	1.2%
pictures	0.2%	you	0.6%	girls	0.3%	teen	1.0%
new	0.2%	sounds	0.5%	pictures	0.2%	nude	1.0%
nude	0.2%	free	0.5%	p***y	0.2%	f**k	0.9%
music	0.2%	sex	0.5%	naked	0.2%	porn	0.9%
school	0.2%	john	0.5%	teen	0.2%	girls	0.9%
how	0.2%	me	0.4%	women	0.2%	p***y	0.9%
lyrics	0.2%	song	0.4%	pics	0.2%	f***ing	0.8%
home	0.2%	love	0.4%	free	0.1%	c*m	0.8%
pics	0.2%	sound	0.4%	black	0.1%	gay	0.7%
download	0.2%	by	0.3%	girl	0.1%	lesbian	0.7%
online	0.1%	my	0.3%	porn	0.1%	video	0.7%
american	0.1%	on	0.3%	big	0.1%	black	0.7%
state	0.1%	songs	0.3%	hot	0.1%	anal	0.6%
county	0.1%	download	0.3%	t*ts	0.1%	hardcore	0.6%
university	0.1%	wav	0.3%	young	0.1%	big	0.6%
car	0.1%	world	0.3%	f***ing	0.1%	hentai	0.5%
texas	0.1%	theme	0.2%	flag	0.1%	t*ts	0.5%
real	0.1%	orgasm	0.2%	sexy	0.1%	young	0.5%
games	0.1%	midi	0.2%	gay	0.1%	girl	0.5%
software	0.1%	star	0.2%	c*m	0.1%	videos	0.5%
art	0.1%	your	0.2%	a**	0.1%	movies	0.4%
map	0.1%	down	0.2%	world	0.1%	asian	0.4%
florida	0.1%	pink	0.2%	map	0.1%	women	0.4%

In previous studies of general Web searching trends, there is a documented shift away from entertainment to commercial and increased information searching on a variety of topics [8]. These changes have paralleled the increase availability of commercial content on the Web [19].

This does not appear to be the case with multimedia searching on the Web, as evidenced by the most frequently occurring terms, most of which are entertainment related. Many terms are sexual in nature. Although this may cause offense for some, it is important for researchers and practitioners in the field of multimedia retrieval

to clearly understand the motivation and information need of many currently searching for multimedia on the Web. It remains to be seen whether or not Web multimedia searching follows the same trend as general Web searching, which is shift to other topical domains.

The other notable area is the frequency of term usage. The term frequencies between general and image searching are nearly identical. Audio and video frequencies are substantially higher. However, one might expect this given the clustering in the entertainment domain.

## 6. DISCUSSION

This research had the goals of identifying the characteristics of Alta Vista Web multimedia searching and measuring the effect of radio buttons on multimedia searching. From our analysis, multimedia appears to require greater interactivity between the user and search engine, relative to general Web searching. The increase in query and session lengths and the increase in the number of results pages being viewed indicate this greater interactivity. Overall, the interactions between Web searchers and systems are still relatively simple as evidenced by the low use of query operators. There are indications that the use of query operators has little effect on search engine results [20, 21], which may account for their low usage.

However, the range of information needs appears to be broadening based on the high percentage of unique terms and large number of terms not repeated in the data set. Other Web studies also report a trend toward a broadening of information needs [8, 22]. The increased interactivity is actual welcome news for search engine developers, as it indicates a move by Web searchers to more carefully refine their information needs.

There is a sharp decrease in the number of pages viewed, especially between the first and second and the second and third results pages, with very few users viewing more than four or five pages of results. Alta Vista users have a low tolerance for reviewing large numbers of results, although again the trend is to view more. Given that over 70% of Web users utilize search engines to locate other Web sites [23], the implications are rather clear for content providers. Certainly for those publishing multimedia content on the Web or engaged in Web e-commerce in the multimedia area, the need to be ranked within the top 10 or 20 results remains critical in order to direct visitors to one's Web site.

At the term level of analysis, the most frequently occurring terms represent a small percentage of overall term usage. The most frequently used term (*free*) accounted only for approximately 0.6% of all term usage. The use of sexual terms was extremely low, in the general set, and the diversity of terms was quite large. Even in the multimedia searching, with more target topic, the frequency of top term usage was quite low. Again, this diversity reinforces our initial term analysis findings -- that these Web users are searching for an increasing variety of information topics.

For second research question, we examined how multimedia searching on Alta Vista compares to general Web searching. Generally, it appears that the use of separate searching interfaces aids users in multimedia searching. Multimedia searching using Alta Vista's "radio button" ontology is less complex in terms of query and session length than searching for multimedia content without such an ontology. However, even with the use of the multimedia radio buttons and specific multimedia content collections, searching for multimedia is more complex than general Web searching. This indicates the need for and the possibility for further system improvements in this area.

Of the four types of searching (general, audio, image and video), image searching appears to be the more multifaceted task. The mean terms per query for image searching was notably larger (4 terms) than the other categories of searching, which were less than 3 terms. The session lengths for image searchers were longer than any other type of searching, although video sessions were also relatively lengthy. Boolean usage by image searchers was 28%, over four times the next highest category of general Web searching. These

results certainly indicate the need for more efforts to increase the ease of multimedia searching.

### 6.1 Strengths

This study contributes to the multimedia searching literature in several important ways. First, the data comes from users submitting real queries. Accordingly, it provides a realistic glimpse into how Web users search, without the self-selection issues or altered behavior that can occur with lab studies or survey data. Second, the sample is quite large, with over 350,000 sessions from general transaction log. Also, there were 34,000 multimedia additional sessions. Third, we obtained data from a very popular search engine as measured by both document collection and number of unique visitors to ensure that our results were generalizable. Fourth, it appears to be the first published analysis of Web searching on specific multimedia databases.

### 6.2 Limitations

As with any research, there are limitations that one should recognize. The sample data comes from one major Web search engine, introducing the possibility that the queries do not represent the queries submitted by the broader Web searching population. However, Jansen and Pooch [7] have shown that characteristics of Web sessions, queries, and terms are very consistent across search engines.

Other potential limitations are that we do not have information about the demographic characteristics of the users who submitted queries, so we must infer their characteristics from the demographics of Web searchers as a whole. Also, the data was collected on a specific date, introducing the possibility of bias due to these particular dates not being representative. However, a comparison of the collected body of Web transaction log research [8, 22, 24-28] shows a great deal of similarity among Web searchers, indicating particular dates have little effect on session lengths, query lengths, Boolean usage, etc. However, particular dates do have an effect on the usage of popular terms [29].

## 7. FUTURE RESEARCH

There are several avenues for future research. Certainly, we need more analysis in this field on a wider variety of Web search engines, ideally on the most popular search engines such as Microsoft Search, Google, America Online, or Yahoo!. However, access to the user data and the willingness of search engines to provide the access hampers this type of research. Additionally as Web search engines introduce additional searching interfaces changes, researchers should continue the evaluation of these changes to gauge their effect on Web searchers, using either transaction logs analysis or lab studies. Finally, we must continue the trend analysis of Web searching in order to predict future behavior and identify future user needs.

## 8. CONCLUSION

The results of our research provide important insights into the current state of multimedia searching and Web information system usage for users, search engines developers, and Web sites designers. Certainly, system design work needs to continue, especially in the area of multimedia searching. The short session lengths and short queries are challenging issues for designers of Web information systems. This approach does not seem to be a successful strategy to maximize either recall or precision, the standard metrics for information retrieval system performance.

Given the large number of users that Web search engines continue to attract, it appears that these searchers are finding relevant information with this searching strategy. This may indicate the need for new metrics to evaluate Web information systems.

## 9. ACKNOWLEDGEMENT

We thank Alta Vista for providing the Web search engine data set without which we could not have conducted this research.

## 10. REFERENCES

- [1] J. I. Cole, M. Suman, P. Schramm, R. Lunn, and J. S. Aquino, "The UCLA Internet Report Surveying the Digital Future Year Three," vol. 2003: UCLA Center for Communication Policy, 2003.
- [2] M. S. Lew, "Next Generation Web Searches for Visual Content," *IEEE Computer*, vol. 33, pp. 46-53, 2000.
- [3] J. R. Smith, "Webseek at Columbia University," vol. 2003, 2003.
- [4] M. S. Lew, N. Sebe, and J. P. Eakins, "Challenges of Image and Video Retrieval," in *Lecture Notes in Computer Science*, vol. 2383/2002, M. S. Lew, N. Sebe, and J. P. Eakins, Eds. Heidelberg: Springer-Verlag, 2002, pp. 1-6.
- [5] A. Goodrum and A. Spink, "Image Searching on the Excite Search Engine," *Information Processing & Management*, vol. 37, pp. 295-311, 2001.
- [6] B. J. Jansen, A. Goodrum, and A. Spink, "Searching for Multimedia: Video, Audio, and Image Web Queries," *World Wide Web Journal*, vol. 3, pp. 249-254, 2000.
- [7] B. J. Jansen and U. Pooch, "Web User Studies: A Review and Framework for Future Work," *Journal of the American Society of Information Science and Technology*, vol. 52, pp. 235-246, 2001.
- [8] A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic, "From E-sex to E-commerce: Web Search Changes," *IEEE Computer*, vol. 35, pp. 107-111, 2002.
- [9] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz, "Analysis of a Very Large Web Search Engine Query Log," *SIGIR Forum*, vol. 33, pp. 6-12, 1999.
- [10] D. Sullivan, "Nielsen / NetRatings Search Engine Ratings," vol. 2002: SearchEngineWatch.com, 2002.
- [11] D. Sullivan, "Search Engine Sizes," vol. 2000, 2000.
- [12] AltaVista, "Special search terms," vol. 2003: AltaVista, 2003.
- [13] B. Morrissey, "Overture to Buy AltaVista," vol. 2003: Internet Advertising Report, 2003.
- [14] B. Berkowitz, "Yahoo! to Buy Overture in \$1.63 Bln Deal," in *Reuter*. Los Angeles, CA, USA, 2003.
- [15] C. Djeraba, "When Image Indexing Meets Knowledge Discovery," in *Proceedings of Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, USA, 2000.
- [16] J. Z. Wang, "SIMPLIcity: A Region-based Image Retrieval System for Picture Libraries and Biomedical Image Databases," in *Proceedings of ACM Multimedia*, Los Angeles, CA, 2000.
- [17] C. Ozmultu, A. Spink, and Z. Ozmultu, "Multimedia Web Searching Trends: 1997 - 2001," *Information Processing & Management*, vol. 39, pp. 611 - 621, 2003.
- [18] N. Belkin, C. Cool, W. B. Croft, and J. Callan, "The Effect of Multiple Query Representations On Information Retrieval Systems," in *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1993.
- [19] S. Lawrence and C. L. Giles, "Searching the World Wide Web," *Science*, vol. 280, pp. 98-100, 1998.
- [20] B. J. Jansen and C. M. Eastman, "The Effects of Search Engines and Query Operators on Top Ranked Results," in *Proceedings of The IEEE 4th International Conference on Information Technology*, Las Vegas, NV, 2003.
- [21] C. M. Eastman, "30,000 Hits May be Better than 300: Precision Anomalies in Internet Searches," *Journal of the American Society for Information Science and Technology*, vol. 53, pp. 879-882, 2002.
- [22] B. J. Jansen and A. Spink, "An Analysis of Web Searching By European Alltheweb.com Users," *Information Processing and Management*, Forthcoming.
- [23] Alexa Insider, "Alexa insider's page," vol. 2000: Alexa Insider, 2000.
- [24] G. Abdulla, B. Liu, and E. Fox, "Searching the World-Wide Web: Implications from Studying Different User Behavior," in *Proceedings of the World Conference of the World Wide Web, Internet, and Intranet*, Orlando, FL, 1998.
- [25] F. CACHEDA and Á. Viña, "Experiences retrieving information in the World Wide Web," in *Proceedings of the 6th IEEE Symposium on Computers and Communications*, Hammamet, Tunisia, 2001.
- [26] C. Hölscher and G. Strube, "Web Search Behavior of Internet Experts and Newbies," *International Journal of Computer and Telecommunications Networking*, vol. 33, pp. 337-346, 2000.
- [27] A. Montgomery and C. Faloutsos, "Identifying web browsing trends and patterns," *IEEE Computer*, vol. 34, pp. 94-95, 2001.
- [28] E. Selberg and O. Etzioni, "The MetaCrawler architecture for resource aggregation on the Web," *IEEE Expert*, vol. 12, pp. 11 - 14, 1997.
- [29] D. Wolfram, "Term Co-occurrence in Internet Search Engine Queries: An Analysis of the Excite Data Set," *Canadian Journal of Information and Library Science*, vol. 24, pp. 12-33, 1999.