

Understanding Searches of an On-line Contemporary Art Museum Catalogue

A Preliminary Study: Fall 2006

J. Trant, Archives & Museum Informatics

comments to http://conference.archimuse.com/blog/jtrant/searching_museum_collections_on_line_what_do_peo

A. Summary

Unassisted searching of the collections catalogues of libraries, some archives, and some museums has developed in tandem with the information networks that support remote access to databases. Many art museums now offer a searchable collections catalogue as a part of their institutional Web sites. These on-line catalogues represent a significant investment in the dissemination of collections information and offer the potential for improved access to collections. But there have been few assessments these catalogues.

Are searches of on-line museum catalogues successful? The queries made of on-line museum catalogues have not been studied, either on their own or in the context of other types of unassisted searching. The kinds of terms used in unassisted museum Web site searches have not been compared to the standards for museum documentation that form the basis for museums' Web-accessible catalogues. Tools for assessing the effectiveness of on-line catalogues of museum collections are lacking.

This paper begins to address these lacunae through a study of the searches of a contemporary art museum's on-line catalogue. By asking: a. "what do users search for?" b. "do they find it?", and c. "if not, why not?" the study assesses whether this museum's on-line catalog is successful in providing information resources in response to their queries.

The study makes several contributions:

1. The method used is replicable. Searches of other on-line museum catalogues may be analysed in a similar way, enabling results to be compared across like institutions, and possibly some general conclusions drawn, and
2. The analysis of the success or failure of users' searches provides insight into the effectiveness of current art museum documentation practices, in the context of delivering collections information to the public.
3. Opinions on the effectiveness current documentation methods are offered, alternative strategies to improve access to on-line collections (such as social tagging and folksonomy) are introduced, and issues for further study identified.

This paper will 1) examine the literature of access to museum collections, and 2) review approaches to search log analysis and query term analysis in the context of access to on-line information resources. Methodological approaches will be assessed for their applicability to 3) an analysis of the searches of the on-line collections catalogue of the Solomon R. Guggenheim Museum, New York.

B. Review of the Literature

1. Studies of Access to Museum Collections

Published studies of access to museum collections tend to focus either on vocabulary and terminology control as an aspect of museum data standardization, or explore log analysis as a part of a broader study of museum Web site use. Several formative studies explored the questions users asked of museums, in a grounded attempt to understand the access requirements on-line museum collections databases were to support. There are no published studies that *systematically* examine the queries made an actual museum's on-line collections catalogue in detail.

A. Questions Asked of Museums

One of the earliest, studies of museum-goers' information requirements, The Catechism Project, conducted by Helen McCorry and Ian O. Morrison (McCorry and Morrison 1993) was designed to inform the creation of on-line museum databases. The authors reviewed the questions users ask of museums, solicited from their network of colleagues in many different types of museums. Over 100 responses to about 300 requests for "examples of the questions [museum professionals] ... are asked and the questions which they ask themselves". A total of 1,013 questions were submitted (well over the 5 per respondent requested). Questions were categorized by type, and then further analysed to see if they could be answered by information that might be found in a database. Two thirds of the questions (695 or 69%) were thought likely to be answerable by a database query. The other third required "looking up dictionaries of biographical or local histories, checking files of similar information held in the museum, sending out pre-prepared information sheets, following policy, identifying objects, supplying front-of-house information, and simply providing the sort of information that one has in one's head after working with the same collection for a number of years". A content analysis was then performed on the database-related questions to determine which data elements were required to answer them. The common element was object name: 30%. This was followed by a cluster of other data elements: "place names associated with the object such

as place of origin or manufacture are almost as important as the names of makers, owners or institutions". All together these accounted for about 30% of queries. Descriptive information (materials, techniques, manufacture, dimensions, iconography and inscriptions) accounted for another 16%, followed by data about procedures (acquisition, exhibition, loan), at 13%. Significantly, "a majority of queries required more than one field to answer in full". The authors, contributing to a developing dialogue on museum collections documentation standards conclude with a recommendation to focus standards efforts on 'what where who and when', rather than on descriptive elements. The extent to which these findings may have reflected practices history rather than art museums has not been explored.

In the context of the development of the Consortium for the Computer Interchange of Museum Information (CIMI) Z39.50 application profile to support distributed searching of museum collections, Janney and Sledge reviewed ~1500 user questions (Janney and Sledge 1995), drawn from Jane Sledge's Points of View work at the Getty Art History Information Program (Sledge 1995, Sledge and Case 1995), Cody Janney's work at Corbis, and the Catechism Project reports. These questions were analysed to determine the categories of information present in questions that users might ask of museums. Sample questions are grouped under categories drawn from the CIDOC Data Model (Documentation Standards Working Group - Committee on Documentation (CIDOC) of the International Council of Museums (ICOM) 1995). This report presents a useful summary of the museum information universe from users' perspectives, questions were both predictable ("Do you have any drawings signed "Charles Edouard Jeneret"?") and surprising ("Do you have any interior scenes by Mr. Rococo that are suitable for framing?"). Data groups most commonly asked for included "Classification, Date, Event, Material / Method, People, Place, Object, Style and Subject". Questions that required outside resources were also often asked. Categories of information were often paired in searches: "Creator/Title, Subject/Date, Classification/Date, Material/Object/Style, Subject/Place, Object/Place, and Classification/ Style (Janney and Sledge 1995, Section 2. Common Data Groups).

2. User Studies of Museum Web Sites

The Museums and the Web conference series has been the venue for a number of reports on studies of museum Web site use that use log data as input. Studies that use Web site transaction logs to derive a profile of numbers of visits, unique hosts, countries of origin and other transaction-based data begin with a study of ArtsEdNet, a project of The J. Paul Getty Trust Education Institute (Borland and Wongse-Sanit 1997). A study from the Science

Learning Network, reported in 2000, shows how much data can be mined from site logs, and puts that data in context with results using other evaluative methods (Semper et al. 2000). This study begins to acknowledge the role of searching, noting the significance of Web text search engines (such as Google, or Yahoo!) in driving traffic to the site. The use of search engines as a point of access to museum Web resources was confirmed by another paper from 2000, in which Nordbotten noted that “85% [of the Web exhibition sessions studied] began from a keyword search using a search engine” (Nordbotten 2000, Section: Web site users).

Peacock addresses the use of log data to assess user experience in a paper presented in 2002 (Peacock 2002), and posits a set of “log based diagnostics for evaluating the performance of museum sties” that include both “traffic indicators” and “satisfaction indicators”. The latter includes “index of search terms used”. These ideas are explored further in a paper that surveys collections search functionalities available, and contrasts “databased” and “curated” modes of information presentation (Doolan, Peacock, and Ellis 2004). Log-based statistics also formed a major part of the evaluations of the MarsQuest online project (Harold et al. 2005) and the Canadian Heritage Information Network’s Virtual Museum of Canada (Carey and Jeffrey 2006).

In a report of a broad survey of museum Web implementations in 2005, Bearman and Trant note that most museums use log data to assess their impact, reporting “more than a dozen different analysis packages were used to gather statistics” (Bearman and Trant 2006, section II. A). However, 20% of museums report collecting log data but not analyzing it, pointing to the need for methods and tools that are straightforward to use and produce replicable results.

While most acknowledge the importance of searching as a means of access to museum Web sites – a fact that is confirmed by a user survey that notes 63% of Web site visitors go “to search museum collections” (Kravchyna and Hastings 2002, Conclusion) – none of these papers studies the terms that are used for searching.

B. Data Standards, Vocabulary and Terminology Control in Museums

Searching of museum databases is often referenced in theory in discussions of museum documentation standards (Art Information Task Force (AITF) 1995, Bearman 1990, 2006, Heller 1968, Mack and Llewellyn 1998, McKenna and Patsatzi 2005, Roberts 1985, Shatford Lane 1998, Sherwood 1998, Stam 1989a) but needs for particular kinds of access have not been substantiated by use-based studies.

Similarly, the role of vocabulary and authority control to improve access to museum collections has been discussed widely, in international conferences (Roberts 1990), and by authors from the Getty Vocabulary Program (in studies such as Baca 2002, Baca et al. 2006, Harpring 1999, 2002). These studies build on classic (Cranfield) approaches to information retrieval (Sorgel 1994, Voorhees 2001) and assert that consistent vocabulary will improve retrieval results by increasing precision (Hourihane 2002).

Others have looked at the possibility of improving search results through thesaurally-aided term expansion (Bates 2002, Bearman 1994, Bearman and Peterson 1991, Dalmau et al. 2005, Tudhope et al. 2006), acknowledging that controlled vocabulary may impede accessibility by users unfamiliar with institutional policies and procedures for collections documentation.

C. User Information Needs

Studies of users have examined what art historians, in particular, might “need” as access to museum information (Bakewell, Beeman, and Reese 1988, Stam 1989b); access to information about works of art through end-user searching is – as expected – one of those key needs. But how museum on-line collection users actually search has not been studied.

3. Studies of Search Logs and Searching

The information science literature has acknowledged the role that search log analysis can play in developing an understanding of user search needs and behaviours.

The Getty Online Searching Project studied the search behavior of a group of humanities scholars at the Getty center for the History of Art and the Humanities. Results reported by Bates and her collaborators note a stronger reliance on names, places, works as subjects and chronological terms than a comparative study in the sciences (Bates 1996a, 1996b, Bates, Wilde, and Siegfried 1993, 1995, Siegfried, Bates, and Wilde 1993).

Jansen introduces the methods and goals of search log analysis (Jansen 2006), defining terms, and offering a method for the collection, preparation and analysis of search data that can be extracted from the transaction logs of web servers. From this type of study, one can “isolate trends and identify typical interactions between searchers and the system”, by examining the “physical expressions of communication exchanges between the searcher and the system”

(Jansen 2006, 410). He reviews the data available in a standard web transaction log, methods of preparation of data, to clean and parse it into a form suitable for analysis. Challenges include removing corrupt data and identifying multiple interactions by the same user as search sessions. Several levels of analysis are outlined, focused on terms entered (and their co-occurrence), queries formulated (and their refinement), and sessions of interaction (and their duration and characteristics beyond searching). In analysis, data is often queried for analysis, and a number of standard queries, to determine basic metrics describing searches, are given in Appendix A. Applications that gather both user- and server-side data are discussed. In summarizing the benefits of search log analysis, Jansen notes that this mode of data collection is direct, unobtrusive, and relatively low cost. However, data available is limited to that captured by logging software, and prone to some difficulties in interpretation (more than one user can be associated with an IP address for example). As search log analysis studies only computer/user interaction, demographic data is lacking and user motivation is obscured (Jansen 2006, 424).

Jansen and Spink report a comparison of nine search engine transaction logs, studied to develop a profile of web searching (Jansen and Spink 2006). They distribute queries into broad categories identified by Spink earlier (Spink et al. 2002), and show a trend towards more searches for “people places or things” away from “entertainment or recreation” (Jansen and Spink 2006, 259), a change also reported in an earlier study (Spink et al. 2002) and echoed elsewhere (Wolfram et al. 2001). These studies build on earlier work (Jansen, Spink, and Saracevic 2000, Ozmutlu, Spink, and Ozmutlu 2004, Spink et al. 2001).

Queries that indicated a multimedia resource was desired were studied by Jansen, Goodrum and Spink (2000), using a log segmenting technique based on format-related query terms. That they contained more terms was not surprising, as some of those terms were used to limit results to the desired formats. This result is echoed in a follow-on report (Goodrum and Spink 2001), that also notes a low co-occurrence of terms (outside those that identify format); “over half the terms were used only once” (Goodrum and Spink 2001, 304). Multimedia searches have also been studied using similar techniques (Jansen, Spink, and Pedersen 2003), and analysed over time (Ozmutlu, Spink, and Ozmutlu 2000).

Jansen and Pooch (2001) suggest a framework for web-searching studies, based on a review of the literature to date. They point out the difficulty of comparing data cross studies, given different metrics involved, and note challenges in defining sessions, and gathering reliable

session data. They compare ‘traditional IR, OPAC and Web-User studies, and present the results in useful tables-by-type. But the metrics available across all three types of study (such as terms per query) offer limited insight into user behaviour.

Other studies of large query logs include Silverstein et al. (1999) who focuses on search term correlation, Jones et al, who study the use of a digital library of computer-related literature (Jones et al. 2000), Wen, Nie and Zhang’s study of query clustering based on logs of user behaviours (Wen, Nie, and Zhang 2002), Wang, Berry and Yang (2003) who study five years of longitudinal data and Poydoatou (2006), who studies the use of a metadata registry rather than a data access system,

A recent study for the Arts and Humanities Research Council focused on “Log Analysis of Digital Resources in the Arts and Humanities” (Warwick et al. 2006). The data set examined was drawn from logs of the Arts and Humanities Data Service (one year’s use), the HUMBLE Humanities Hub (one year’s use) and Artifact (3 month’s use, as this was all that was available). The report spends much time discussing results of qualitative methods introduced into the study to compensate for a lack of log data, and notes that obtaining log data was “time-consuming” (Warwick et al. 2006, 13). Very little is actually reported about the log data analysis, though there is an observation that names and subjects often recurred. Looking at the parsed data appended to the report as Appendix A1.2, it appears that the team did its analysis by page [file] name (without directory structure) and so had problems disambiguating actual resources used. Unacknowledged is the fact that they were comparing usage of very different resources (a ‘hub’ or web index for the Humanities, a ‘hub’ for the arts and creative industries, and an archive of digital humanities data).

Several recent studies of OhioLINK logs show how log data can be used to build a profile of the user interactions with large networked resources. Raw logs for OhioLINK from June 2004 to December 2004 were analysed (for both on- and off-campus users) and usage statistics like items viewed, number of users, number of pages viewed and number and length of sessions were calculated(Nicholas et al. 2006). In addition, the group looked at ‘site penetration – ‘how deep into the service users penetrate or burrow; how much they take away with them’ (Nicholas et al. 2006, 488). Resource-use was also characterized by subject area. Huntington, Nicholas, Jamali and Watkinson (2006) show that it is possible to use sub-domain names as recorded in the DNS addresses shown in a web log to provide finer-grained analysis of

electronic resource usage in OhioLINK, imputing department through abbreviations in machine naming.

A short paper, profiling 404 queries asked of Google Answers™ looked at the information content of ‘real’ questions posed of this ask-the-expert service and identified categories of ‘bibliographic metadata’ [a slight misnomer] are identified that are found in 80% of the queries: artist, date, nationality, title, size, provenance, publisher and other (Cunningham, Bainbridge, and Masoodian 2004). Other studies of Web queries that segment interest on the basis of search terms include Pu et al. (2002), who propose an algorithmic approach to categorizing large volumes of search log data, and report the an experiment in auto-categorization built on manually created seed data, and Rose and Levinson (2004) who propose a taxonomy of information seeking goals, and present the results of its application to a large query set using manual classification.

Studies of domain-specific search behaviour include studies of retrieval in American History (Choi and Rasmussen 2002, 2003), Psychology / American History (Yi et al. 2006), and government documents (Chau, Fang, and Sheng 2005).

The utility of transaction log analysis for those making digital library design and management decisions is reviewed by Covey (2002). Sandore (1993) discusses how log analysis results can inform system design.

C. Searches of the Guggenheim Collection On-line.

1. Searching the Guggenheim Collection On-line

Logs of the searches of the Solomon R. Guggenheim Museum, New York’s on-line catalogue (at <http://www.guggenheimcollection.org/index.html>) have been studied to build knowledge of how anonymous Web users search art museum collections.

The Guggenheim is unusual, in that it has a number of different sites, each with a separate presence on the Web. The search data studied here is presented on the museum’s flagship site, the Frank Lloyd Wright-designed museum in New York City, where the majority of the museum’s permanent collections are held. Data in the on-line catalogue represents holdings from the Solomon R. Guggenheim Museum, the Guggenheim Museum Bilbao and the Peggy Guggenheim Collection, Venice.

2. The Search Dialog

Any Web user can search selected data about the Guggenheim Collection from the museum's web site (Figure 1).



Figure 1. The main page of the Guggenheim Museum, New York's web site. Select "Collection Online" to search.

Searches of the collection are made by entering a term or phrase in the search box on the left hand margin of the Collections Search page (Figure 2), and pressing 'enter' or clicking 'go'. It is also possible to search the entire Guggenheim Web site, using the search box on the far right of the navigation bar that appears at the top of every second-level page on the Guggenheim site (Figures 2 – 7).

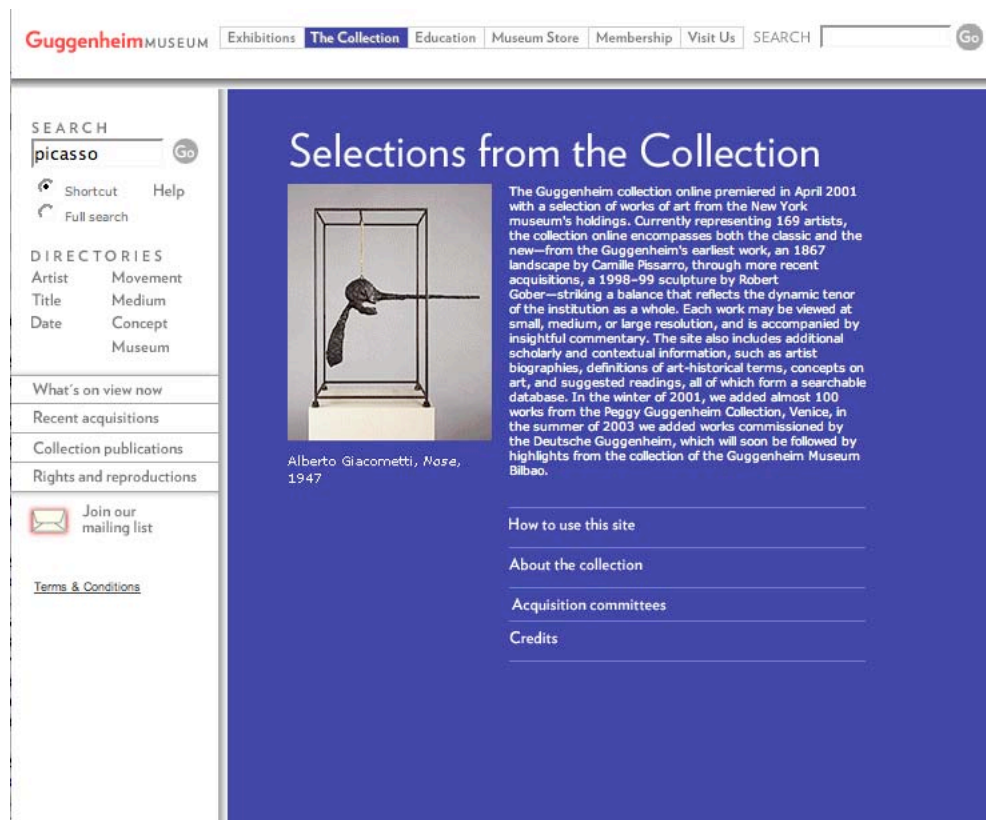


Figure 2. The main collections search page of the Guggenheim Museum, New York's web site. Search terms are entered in the box in the left margin.

The two choices of search offered – “shortcut” and “full search” – affect the way results are presented, not how the search is constructed. A ‘Shortcut’ search goes right to “the page most relevant to the word you entered” (Guggenheim Museum New York 2006), usually an artists page, listing works of art with a thumbnail image, title, and date (Figure 3).

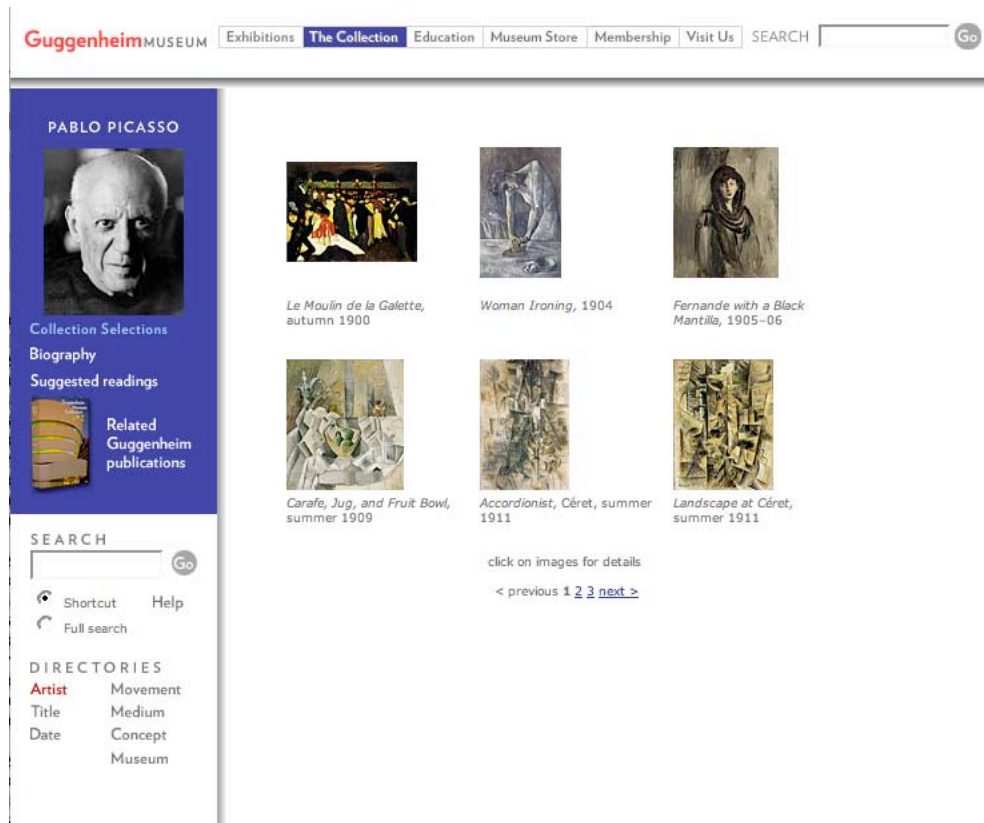


Figure 3. The collections search result page of the Guggenheim Museum, New York's web site. A 'shortcut' search for "picasso" goes directly to a list of works, shown as thumbnail images with title and date below.

Clicking on a thumbnail in the Shortcut search result page takes you to the page for that particular work (Figure 4).

The screenshot shows the Guggenheim Museum website interface. At the top, there is a navigation bar with links for Exhibitions, The Collection, Education, Museum Store, Membership, Visit Us, and a search bar. Below this, a blue sidebar on the left contains a portrait of Pablo Picasso and links for Collection Selections, Biography, Suggested readings, and Related Guggenheim Products. The main content area features a large image of the painting 'Le Moulin de la Galette' with navigation links for '<Previous Picasso work' and 'Next Picasso work'. Below the image is a detailed caption: 'Le Moulin de la Galette, autumn 1900. Oil on canvas, 34 3/4 x 45 1/2 inches. Solomon R. Guggenheim Museum, Thannhauser Collection, Gift, Justin K. Thannhauser. 78.2514.34. © 2005 Estate of Pablo Picasso/Artists Rights Society (ARS), New York.' An 'Enlarge' button with a magnifying glass icon is positioned to the right of the image. Below the caption is a paragraph of text providing context about Picasso's time in Paris and the painting's significance. At the bottom of the page, the name 'Jan Avgikos' is listed.

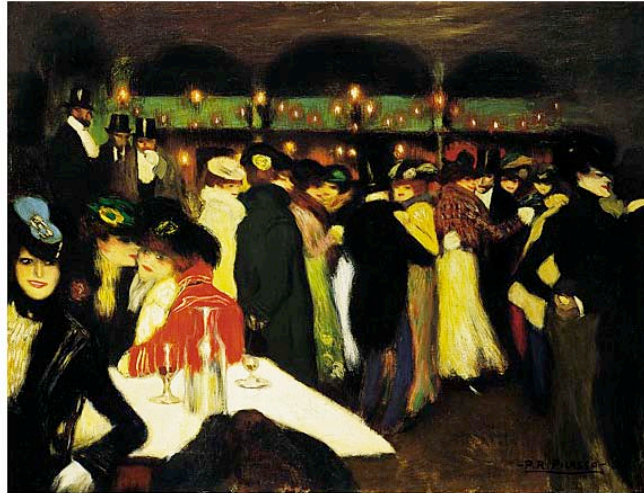
*Figure 4. The page for Picasso's *Le Moulin de la Galette, autumn 1900* from the Guggenheim Museum, New York's web site, showing the full detail available for each work.*

Each work page includes a larger image, a caption and a note. All these data are searched together in both types of search. It is possible to enlarge the image, by clicking on 'enlarge' or on the magnifying glass (Figure 5).

[<Previous Picasso work](#)

[Back to entry](#)

[Next Picasso work>](#)



Le Moulin de la Galette, autumn 1900. Oil on canvas, 34 3/4 x 45 1/2 inches. Solomon R. Guggenheim Museum, Thannhauser Collection, Gift, Justin K. Thannhauser, 78.2514.34, © 2005 Estate of Pablo Picasso/Artists Rights Society (ARS), New York.

*Figure 5. The enlarged image page for Picasso's *Le Moulin de la Galette*, autumn 1900 from the Guggenheim Museum, New York's web site, showing the largest image available for each work.*

Choosing the option “Full Search” returns a listing of all pages where the search term appears (Figure 6).

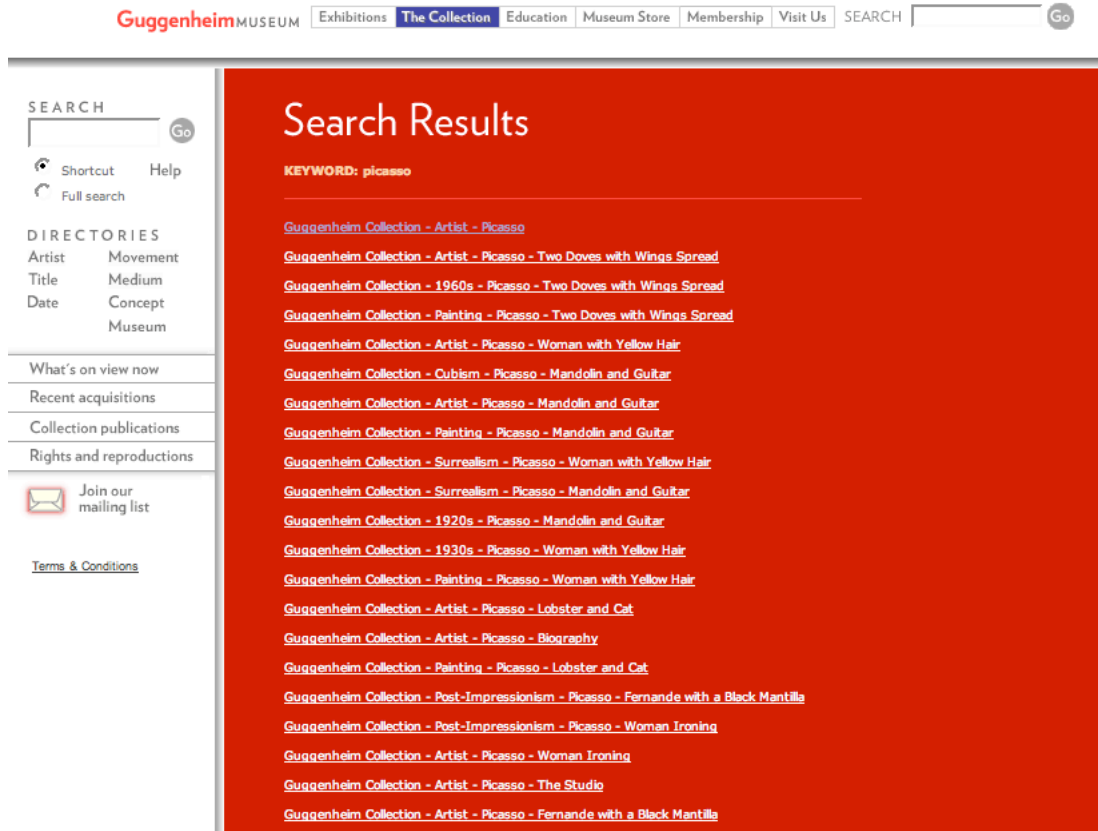


Figure 6. The collections search result page of the Guggenheim Museum, New York's web site. A 'full search' for "picasso" produces a list of pages where that keyword appears. Clicking on a title goes to that page.

Clicking a page title in the list takes you to that page (Figure 7).

The screenshot shows the Guggenheim Museum website interface. At the top, there is a navigation bar with links for Exhibitions, The Collection, Education, Museum Store, Membership, and Visit Us. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a blue sidebar on the left containing text about an exhibition in 1910-11 and a search section with a search bar and a 'Go' button. Below the search section, there are 'DIRECTORIES' for Artist, Title, Date, and Museum, with 'Movement' highlighted in red. The main content area features a portrait of a woman in a black mantilla, with a '<Previous work>' link to the left and a 'Next work>' link to the right. Below the image is an 'Enlarge' button with a magnifying glass icon. The text below the image provides details about the painting: 'Pablo Picasso, *Fernande with a Black Mantilla (Fernande à la mantille noir)*, 1905–06. Oil on canvas, 39 3/8 x 31 7/8 inches. Solomon R. Guggenheim Museum, Thannhauser Collection, Bequest of Hilde Thannhauser. 91.3914. © 2005 Estate of Pablo Picasso/Artists Rights Society (ARS), New York.' Below this text are two paragraphs of descriptive text about the painting and Picasso's work.

Figure 7. The page for Picasso's, *Fernande with a Black Mantilla (Fernande à la mantille noir)*, 1905–06, a "Post-Impressionist" Movement page from the Guggenheim Museum, New York's web site. The data presented for the work is similar, but the navigational surround differs, depending on the type of page.

Unfortunately, the only way to navigate among pages listed this way is to use the back button to return to this list of works and choose another one. Contextual navigation on the work pages is based on the position of that page within the themed structure of the Web site, i.e. whether its is an Artists page, a Work of Art page, or a page for a Date, Movement, Medium, or Concept.

3. *Browsing the Collection*

It is also possible to browse the Guggenheim Collection on-line by Artist, Title, Date, Movement, Medium, Concept or Museum by choosing one of the “Directories” that appear on every Collections page, below the search box. Data about this type of access to on-line information about the Guggenheim Collection was not available, and this type of navigation into the collection is not included in the results presented here.

The presence of these categories – common as they are to contemporary art history – may influence the search terms entered.

D. The Data Set: Guggenheim Collection Search Log Reports

The Guggenheim Museum, New York, provided access to on-line reports summarizing the searches of their on-line collections catalogue. Each search term is logged as it is entered into the search box on any Guggenheim Collections Web page. Data is recorded as entered, and reports are created by the Content Management System serving the catalogue – SPS, developed by Sam Nelson and distributed under the GNU General Public License (clevernamehere.com 2006). Reports offer abstracted data presented month by month, as summarized in Table 1.

Name	Definition
Sequence	The numerical sequence of the search during the month
Date	The date and time of the search, presented as day, month, date, time, year in the form NNN MMM DD HH:MM:SS YYYY
Search Phrase	The search string as entered. Accented characters were replaced by a space. Quotation marks were included, preceded by an escape character, e.g. \”
Hits	A summary of the results found, presented as N:N:N, where these equaled “represent total hits:exact keyword matches:partial keyword match”, e.g. a search reported as “0:0:0” found no results. (Cherry 2006)

Table 1: Guggenheim Collections Search Log Report: Data Available

This study looks at searches made between September 1, 2005 and August 31, 2006. Sample data from the first 20 searches for the month of December are shown Table 2.

	Date	Search Phrase	Hits
1	Thu Dec 1 00:13:53 2005	auguste rodin	7:0:0
2	Thu Dec 1 00:22:13 2005	Autumn Rhythm	5:0:0
3	Thu Dec 1 00:27:24 2005	chagall	34:1:0
4	Thu Dec 1 00:37:34 2005	RUSSIA The Majesty of the Tsars Treasures from the Kremlin Museum	0:0:0
5	Thu Dec 1 00:38:48 2005	russia	50:0:0
6	Thu Dec 1 00:39:26 2005	wine fountain	0:0:0
7	Thu Dec 1 00:52:49 2005	paintings	1823:0:0
8	Thu Dec 1 01:01:42 2005	fear	50:0:0
9	Thu Dec 1 01:16:19 2005	edgar degas	24:0:0

10	Thu Dec 1 01:53:52 2005	Theo van Doesburg	42:1:1
11	Thu Dec 1 02:07:10 2005	van gogh	55:1:0
12	Thu Dec 1 02:07:59 2005	morisot	8:0:0
13	Thu Dec 1 02:09:53 2005	chinese	18:0:0
14	Thu Dec 1 02:36:17 2005	georg beselitz	0:0:0
15	Thu Dec 1 02:42:51 2005	one quart	1:0:0
16	Thu Dec 1 02:44:11 2005	the artist	2703:0:0
17	Thu Dec 1 03:30:27 2005	drawings	586:0:0
18	Thu Dec 1 03:42:26 2005	commere	0:0:0
19	Thu Dec 1 03:45:56 2005	andre	47:1:0
20	Thu Dec 1 03:47:17 2005	Gabo	11:0:0

Table 2: Sample Data from the Guggenheim Collections Search Log Report: September 2005: First twenty searches in the chronological log for December 2005

E. Analysis of the Data Set

The log data provided by the Guggenheim was downloaded from their Intranet, and prepared for analysis. First the monthly reports were loaded into Excel, to parse the data into individual elements. Summaries of data in Excel were compared to reports available in the Guggenheim’s Intranet to ensure accurate data transfer (record numbers and totals were identical). Data was then exported in comma delimited text format for import into statistical analysis software. Comma delimited files were reviewed in a text editor to insure integrity and completeness. CSV formatted data was imported into SPSS (on PC) and Stata (on a Macintosh); Stata and Excel were used for the majority of the data analysis.

1. How many searches are there?

The Guggenheim’s own reporting tool aggregates searches by month, as shown in Table 3.

	Sep 05	Oct 05	Nov 05	Dec 05	Jan 06	Feb 06	Mar 06	Apr 06	May 06	Jun 06	Jul 06	Aug 06
Total Search Queries	9,227	13,120	13,376	9,952	11,406	11,209	13,431	12,039	11,165	8,053	6,761	7,260
Distinct Words	4,309	5,787	5,831	4,777	5,193	4,902	5,375	5,083	4,984	3,949	3,192	3,364
Distinct Phrases	4,674	6,411	6,540	5,094	5,650	5,470	5,933	5,504	5,364	4,092	3,264	3,452

Table 3: Guggenheim Collections Searches: September 2005 - August 2006

For the time period under study – September 1, 2005 to August 31, 2006 – a total of 126,999 queries were made of the Guggenheim Museum’s Collection on-line (Table 4).

2. When are searches made?

An average month saw 10,583 queries, but queries were not evenly distributed over time. The highest month (March 2006: 13,431 searches) was 199% of the number of searches in the lowest month (July 2006: 6,761 searches).

	Total	Monthly			
		High	Low	Average	Mean
Total Search Queries	126,999	13,431	6,761	10,583	11,187
Distinct Words	56,746	5,831	3,192	4,729	4,943
Distinct Phrases	61,448	6,540	3,264	5,120	5,417

Table 4: Guggenheim Collections Searches: September 2005 - August 2006: Aggregate Totals

The curve of queries over time shown in Chart 1 reflects the cycle of the academic year, with peaks in October/November and March and troughs in December and June/August.

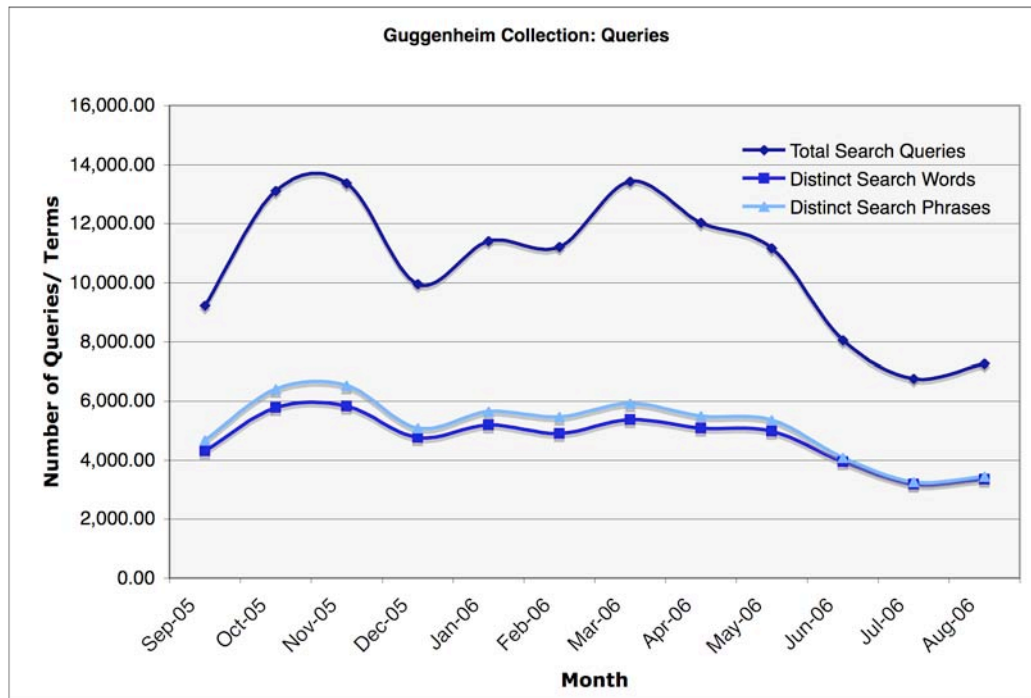


Chart 1: Guggenheim Collections Searches: September 2005 - August 2006: Total Searches, Distinct Words, and Distinct Phrases

Searches are not distributed equally over the month, though. *Chart 2* shows distinct peaks and valleys that reflect weekly cycles. (September data may be artificially low before September 7; only 5 searches were recorded on September 2, and 5 on September 5. After 189 searches on September 7, the pattern returns to that of previous/subsequent months.).

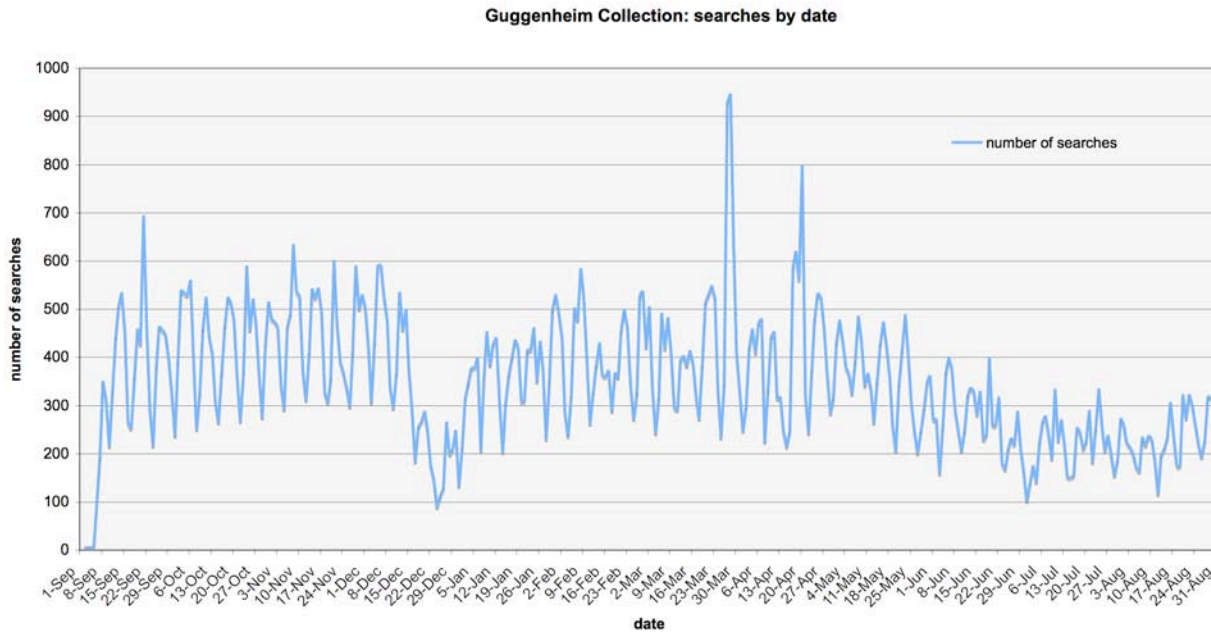


Chart 2: Guggenheim Collections Searches: September 2005 - August 2006: Searches by Date

A review of total searches by day of the week (Chart 3) confirms that there are fewer searches at the end of the week than the beginning. Monday, Tuesday and Wednesday account for 17% of the searches each. This falls very slightly on Thursday (16%) and then significantly on Friday (12%), Saturday (9%) and Sunday (12%).

Guggenheim Collection: Searches by Day

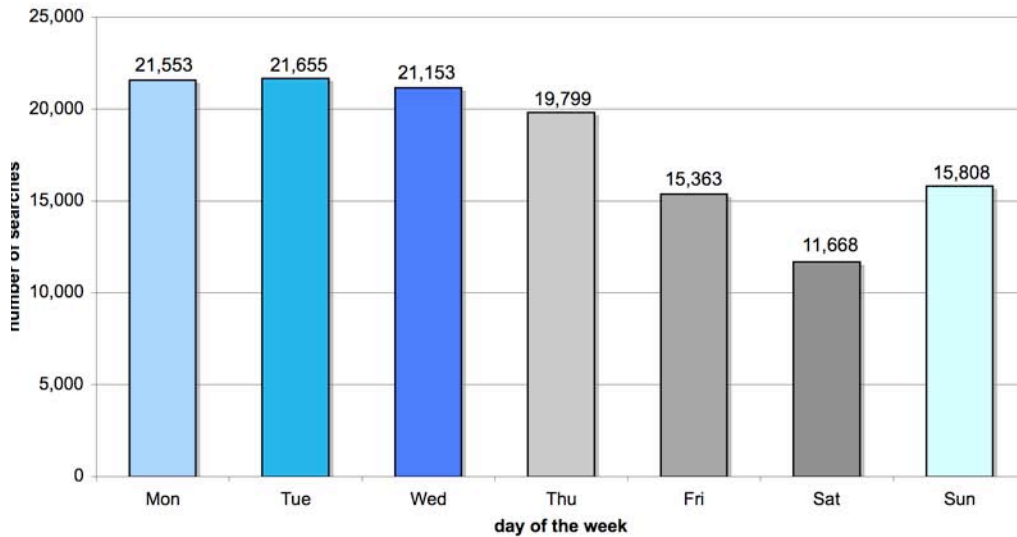


Chart 3: Guggenheim Collections Searches: September 2005 - August 2006: Searches by Day of the Week

Other significant dips in search numbers cluster around statutory holidays. Once the anomalous data are removed for early September the lowest 10 days of searching (Table 5) cluster around Christmas (Dec 23-26), New Years (Dec 31) and July 4 (July 1, 2, 4). The museum is closed December 25th and July 4th each year.

Sequence	Date	Number of Searches
1	24-Dec-05	89
2	1-Jul-06	101
3	25-Dec-05	112
4	12-Aug-06	116
5	26-Dec-05	128
6	31-Dec-05	133
7	2-Jul-06	138
8	4-Jul-06	141
9	23-Dec-05	146
10	14-Jul-06	150

*Table 5: Guggenheim Collections Searches: September 2005 - August 2006:
Lowest Numbers of Searches Per Day (once anomalous early September data removed)*

3. What is searched for?

Search terms were predominantly single terms or very short phrases. Chart 1 plots a count of distinct words and a count of distinct phrases against the total number of searches, data also shown in Table 4. The small variation between these two numbers reflects the fact that most searches are single words.

A large proportion of searches – 56% – were comprised of a single word. When searches of one, two or three words are combined, these make up 95% of all searches (Chart 4). The shortest search had no value (it is possible to search with a ‘blank’ search box); the longest search was made up of 205 words (an entire ‘curatorial note’ entered possibly to find its source; this was one of a series of such queries). Distribution of search words per query is shown on Chart 5. It is likely that the size of the query box (Figure 1) encourages short searches, as it only displays approximately 12 characters.

Guggenheim Collection: number of words per search by percentage

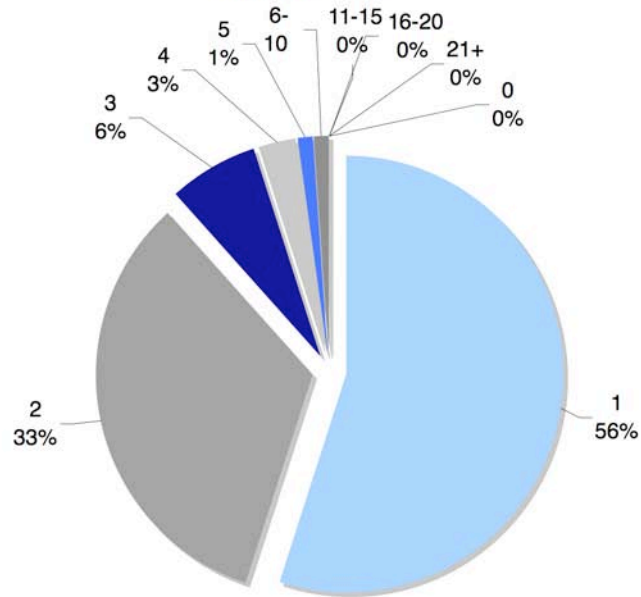


Chart 4: Guggenheim Collections Searches: Number of Words per Search by Percentage

Guggenheim Collection: search words per query

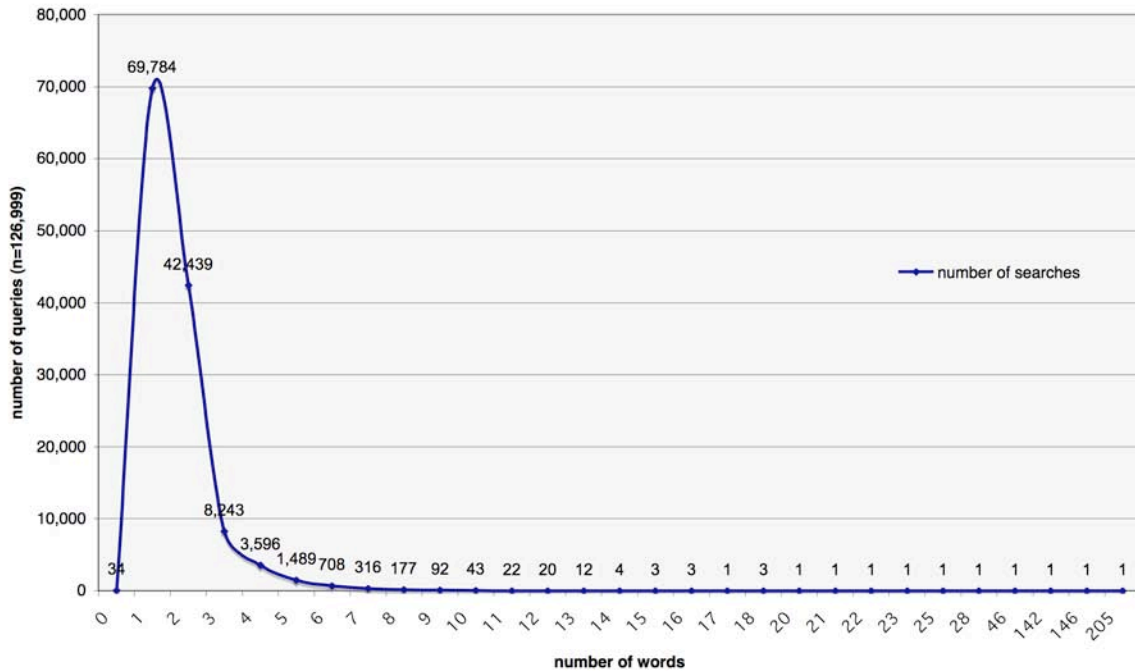


Chart 5: Guggenheim Collections Searches: September 2005 - August 2006: Number of Words per Search

4. Search Term Distribution

The terms searched were reviewed to see what kinds of things users were searching for in the Guggenheim Collections on-line. Data was converted to lower case for analysis (the Guggenheim's Search function is indifferent to case, but the logging software recorded different values for Van Gogh, Van gogh and van Gogh and van gogh).

A. What kinds of terms are searched for?

The terms used to search the Guggenheim Collections On-line were very diverse. A plot of their frequency conforms to a very steep Zipf distribution; even the most commonly searched term – picasso – accounts for only 2.86% of all searches.

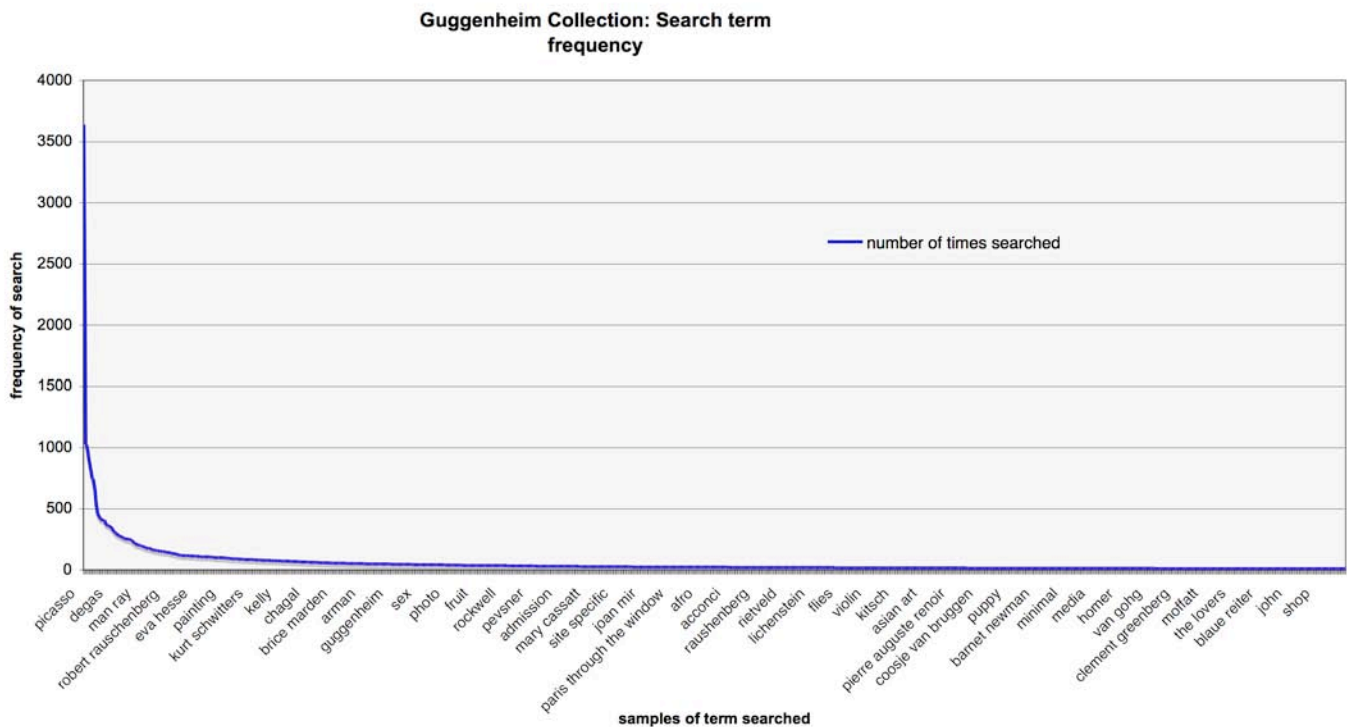


Chart 6: Guggenheim: Search Term Frequency: All Searches.
All searches plotted on curve, only sample data values shown.

A plot of the top fifty searches by percentage of all searches (Chart 7) shows that this curve is exceptionally steep, with a very strong drop-off from 1.74% to 0.81% between the second and third most-common search terms (kandinsky and pollock).

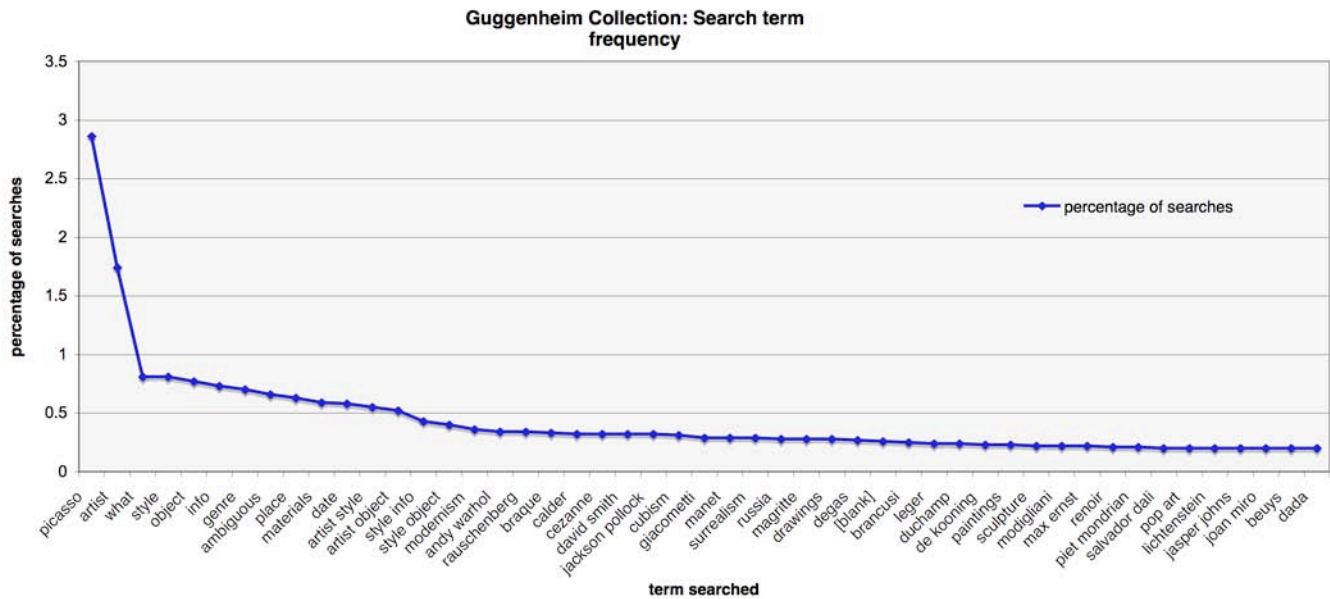


Chart 7: Search term frequency: top 50 search terms, by percentage of total searches

To determine the nature of searching of the Guggenheim Collection on-line, all searches made more than ten (10) times were categorized according to the following rubric:

Category	Definition
artist	An artist's name; e.g. "kandinsky"
what	The title or subject of a work; e.g. "The Ninth Wave" or "several circles"
style	The name of an artistic style or movement; e.g. "conceptual art"
object	The type of object or work of art; e.g. "drawing"
info	Requests for additional information beyond that about works of art; e.g. "when did he die"
genre	Specific modes of visual expression; e.g. "still life" or "self portrait"
ambiguous	Problematic terms that are not possible to categorize or that could belong to multiple categories; e.g. "all" or blank searches. Also includes 'junk' characters
place	The proper name of a geographic place; e.g. "egypt" or "new york"
materials	The materials a work of art is made of and/or the techniques used: e.g. "oil on canvas" or "cardboard"
date	A specific date; e.g. 1950 or 1970s

Table 6: Categorizing Types of Searches

When a search combined multiple concepts, these were categorized using multiple terms, in the order the concepts appeared in the query.

Chart 8 and Chart 9 show the results of a content analysis of the unique (lower case) search strings for all searches made ten (10) times or more.

Searches of the Guggenheim Collection on-line were predominantly for a specific *artist* by name; 63.23% of unique search strings were artists' names. This was followed by *what* (17.79%), *style* (7.26%), and *object* (4.67%). A relatively small percentage of distinct

searches (3.19%) asked for information (*info*) rather than works of art, and therefore were outside the scope of the Guggenheim collections. There were twice as many times as many *genre* terms (1.19%) than either *materials* searches (.67%) or *place* (0.59%) but these are a very small proportion of total searches. *Date* searches were also very infrequent (.07%).

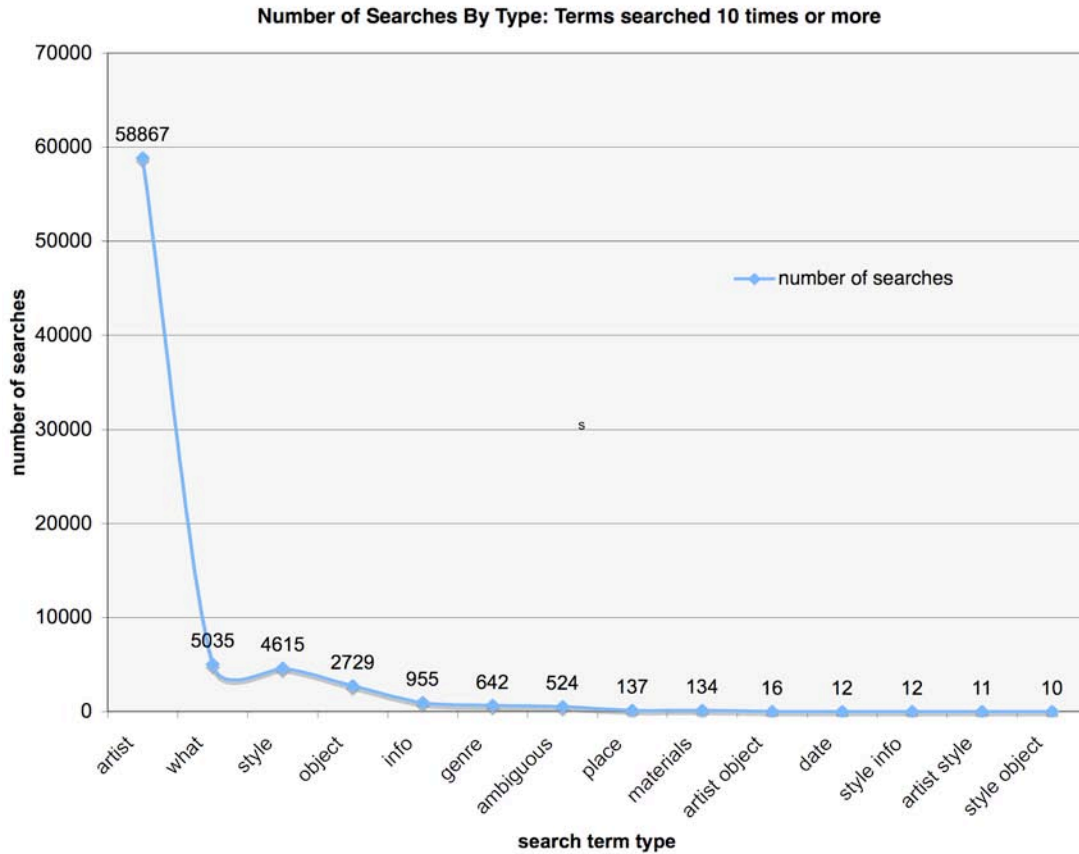


Chart 8: Total Searches by Type: terms searched 10 times or more

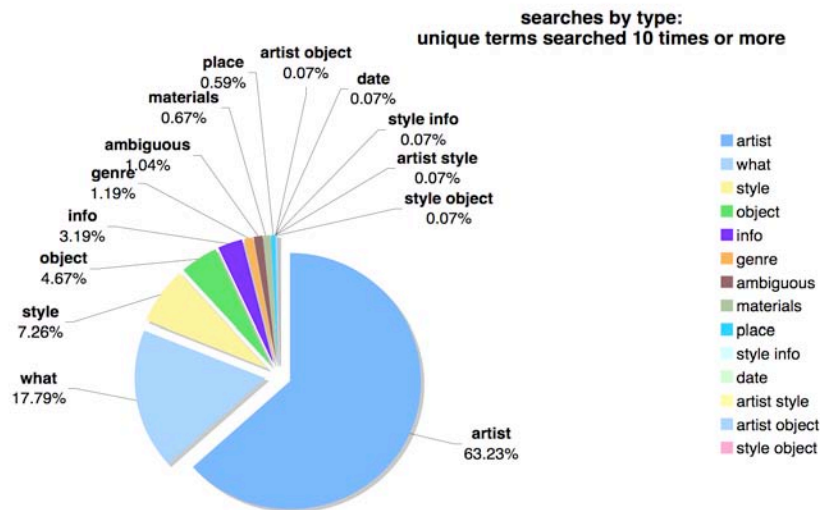


Chart 9: Percentages of searches by type for unique terms searched 10 times or more

Searches in this group rarely combined categories (.07% of terms): concepts combined included *artist/object*, *style/info*, *artist/style* and *style/object*, all pairings mentioned by Janney and Sledge (1995).

B. Do searches relate to the current exhibition program?

To establish if there was a relationship between the content of searches of the Guggenheim Collections on-line and the on-site exhibition program, a selection of searches was examined that related to temporary exhibitions held at the Solomon R. Guggenheim Museum, New York. The two cases chosen were David Smith and Zaha Hadid, because the “solo” nature of the exhibitions pointed to forms of the artists’ names as indicative search terms. There was a very strong relationship between the dates of the exhibition and searches of the on-line catalogue.

The exhibition *David Smith: A Centennial* held from February 6 – May 14, 2006. The distribution of the searches (for “david smith”, “smith, david” or “smith”) over the sample period shows a strong upward curve in the months of the exhibition (Chart 10). There is a jump of 184% between January and February 2006 – around the exhibition’s opening – followed by a 141% increase (72 to 102 searches) in the first month of the show. Stronger interest is maintained in the mid-point of the show (March: 102, April: 105), and then interest declines as the show closes (May: 72).

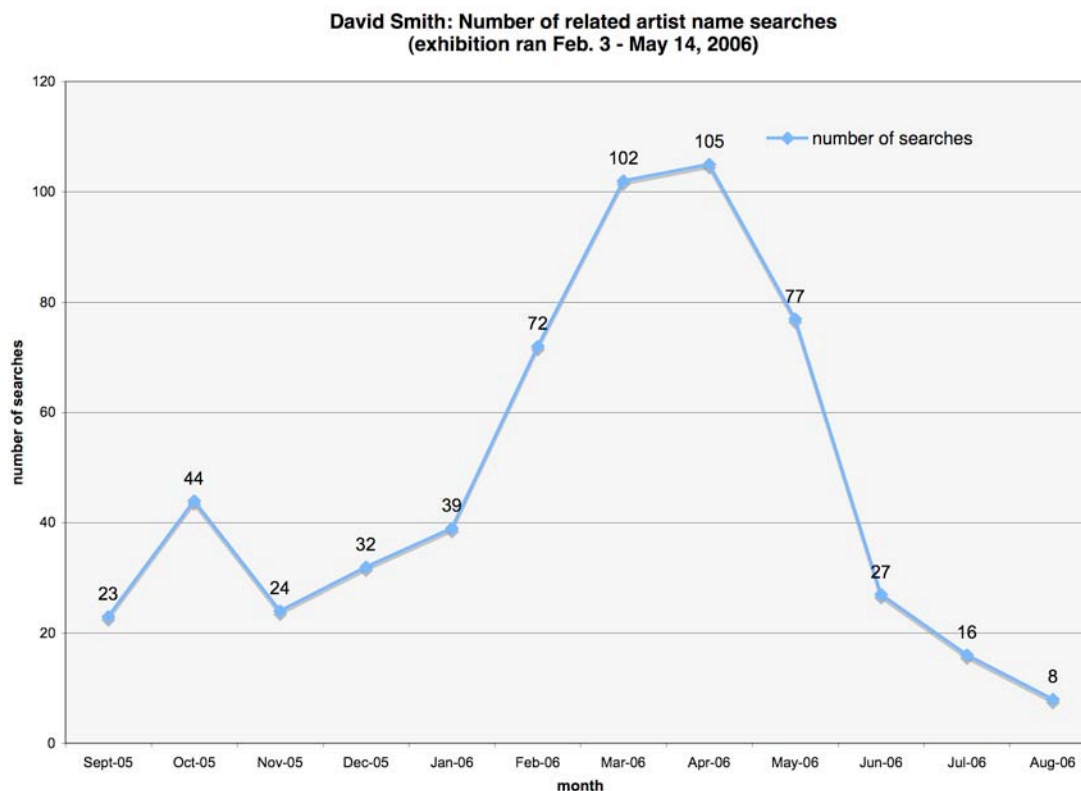


Chart 10: Searches Related to David Smith

This relationship is even more apparent when searches related to Zaha Hadid are examined (Chart 11). The exhibition *Zaha Hadid* ran at the Guggenheim Museum, New York June 3 – October 25, 2006. Searches (for “zaha”, “hadid”, “zaha hadid” or “zaha haddid”) go from five or under per month before April 2006, to 9 in May 2006. Then there is a significant increase in June, July and August. There is a jump of 500% between May and June, and then a steady increase of 120% to a plateau in July and August. Ironically, *all* of these searches were unsuccessful (producing zero results) because Hadid is not represented in the Guggenheim’s Permanent Collection.

That the exhibition program influences searching of the collections catalogue is supported by the presence of a number of searches related to “motorcycles”, including specific manufacturers and models such as “ducati”. This content – atypical of the collection of a museum of modern art – was featured in *The Art of the Motorcycle* held June 26 - September 20, 1998 at the Solomon R. Guggenheim Museum, New York.

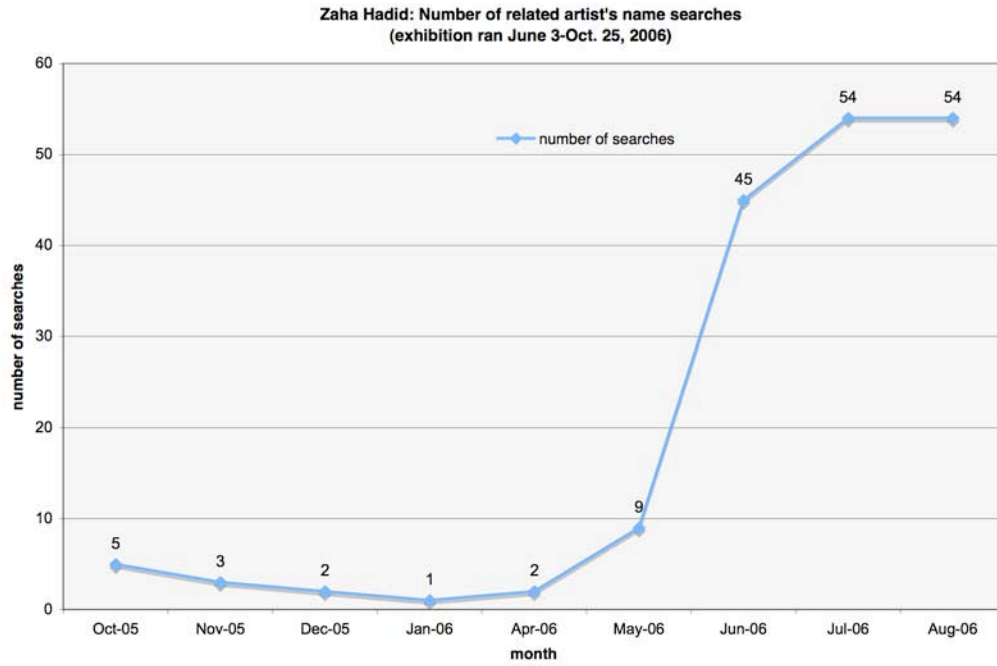


Chart 11: Searches Related to Zaha Hadid

5. *Are searches successful?*

Slightly less than three-quarters of the searches (73.15%) made of the Guggenheim Collection on-line found at least one result that matched the search string entered. More than one quarter of all searches (26.85%) were unsuccessful. This proportion stays consistent over time (Chart 12).

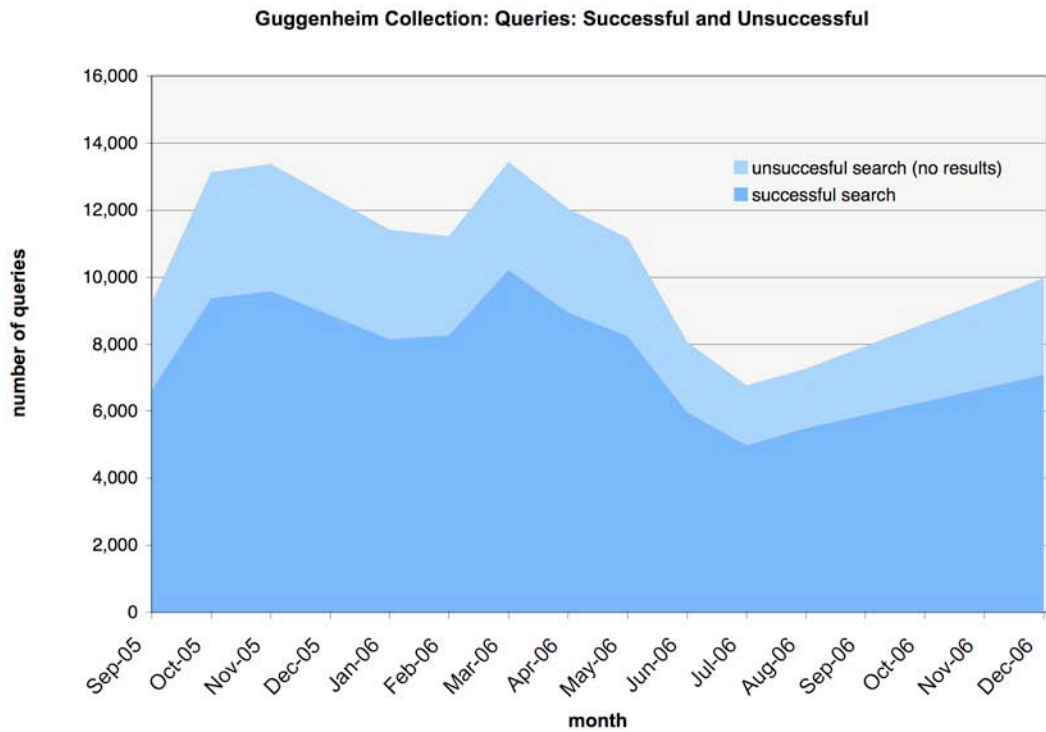


Chart 12: Guggenheim: Queries, successful and unsuccessful

C. Unsuccessful Searches

Unsuccessful searches made five times or more were analysed using the same categories as the overall searches (see D.3 What kinds of terms are searched for?). As shown in Chart 13, *artist* searches also predominated in this group, followed by *what* and *style*. Chart 14 shows the relative distribution of unique unsuccessful searches made five times or more.

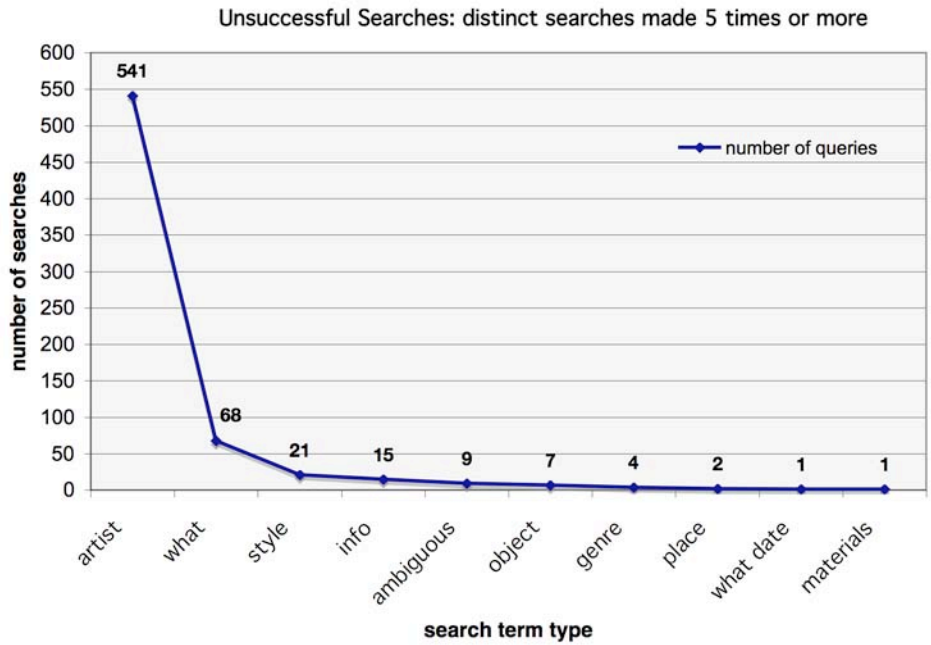


Chart 13: Unsuccessful Searches by Type: Searches made 5 times or more

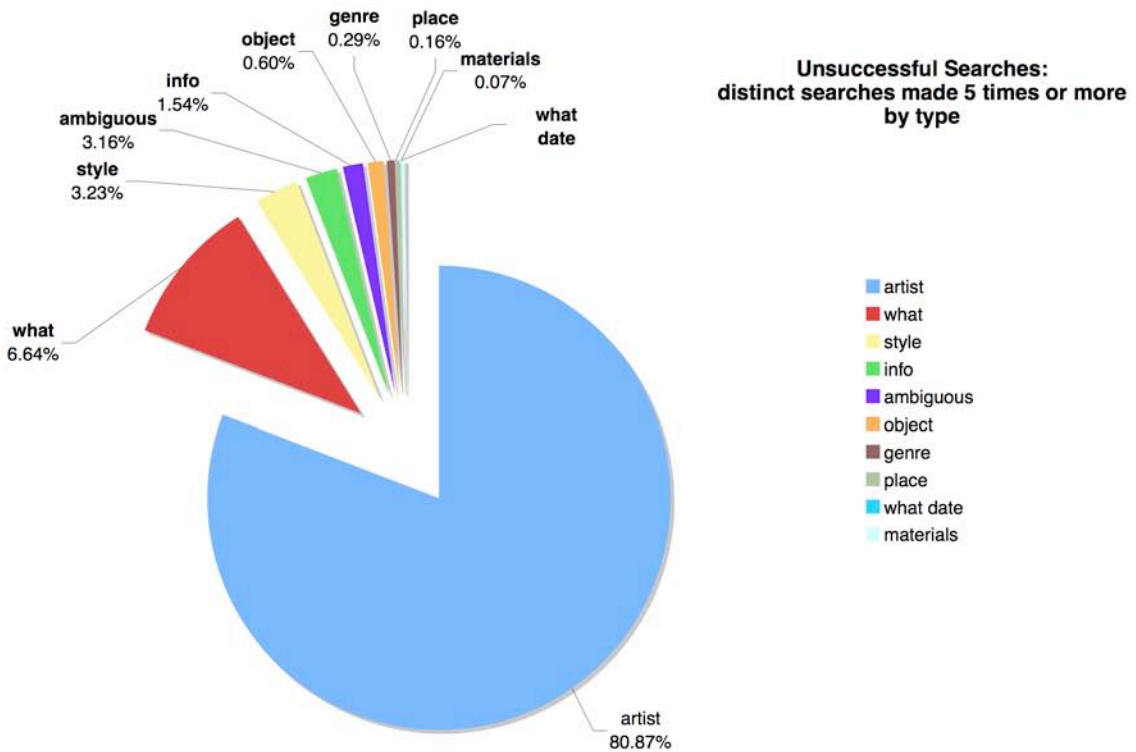


Chart 14: Unsuccessful searches by type: distinct searches made 5 times or more

D. Are some kinds of searches more successful?

The distribution of unsuccessful searches over categories was slightly different than successful searches (Table 7).

search type	successful searches searched 10 times or more		unsuccessful searches searched 5 times or more	
	unique terms	percentage	unique terms	percentage
artist	853	79.76%	541	80.87%
what	240	7.23%	68	10.16%
style	98	5.97%	21	3.14%
object	63	3.76%	7	1.05%
info	43	1.29%	15	2.24%
genre	16	0.81%	4	0.60%
ambiguous	14	0.72%	9	1.35%
place	9	0.19%	2	0.30%
materials	8	0.19%	1	0.15%
date	1	0.02%	0	
artist/style	1	0.02%	0	
artist/object	1	0.02%	0	
style/info	1	0.02%	0	
style/object	1	0.02%	0	
what/date	0		1	0.15%
	1349		669	

Table 7: Unique Search Terms: Successful and unsuccessful by category

Chart 15 shows that about same number of *artist* searches were successful as unsuccessful (79.76% vs. 80.87%). A greater percentage of *what* searches were unsuccessful (10.16% vs. 7.23 %); Both *style* and *object* searches were more likely to be successful, as were *date*, and combined categories. A greater number of *ambiguous* searches was unsuccessful (to be expected as ‘junk’ data was included here).

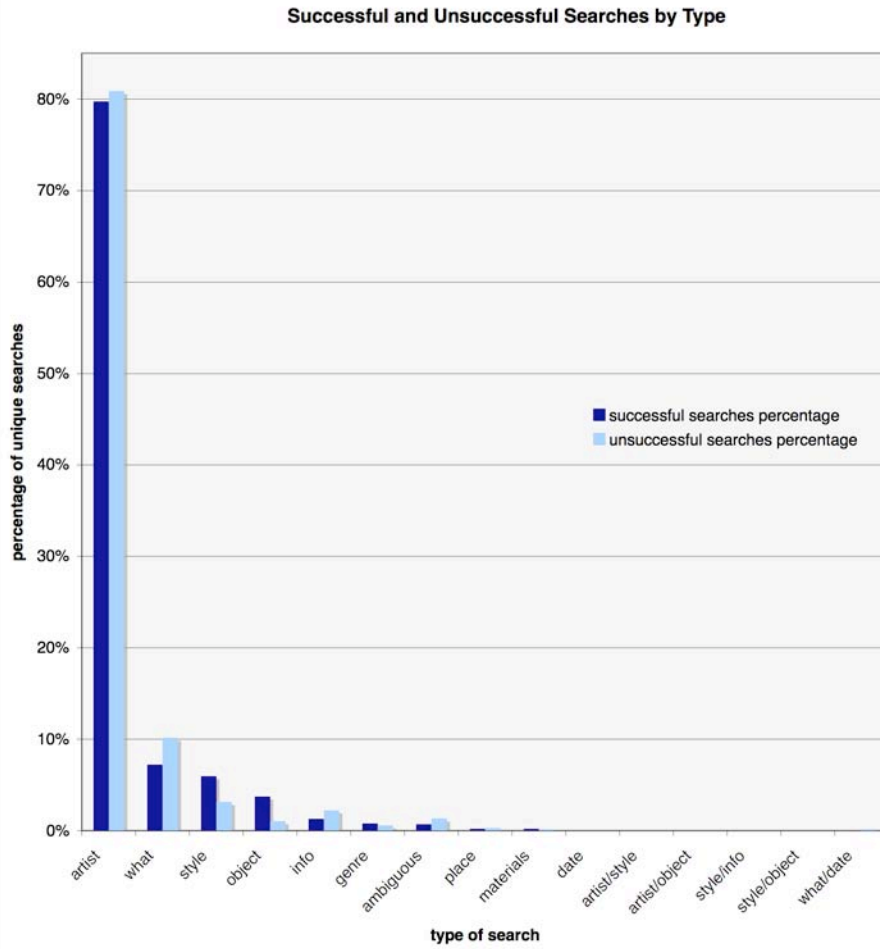


Chart 15: Successful and Unsuccessful Searches by Type

F. Why were searches unsuccessful?

When the unsuccessful searches are examined some striking patterns appear. Unsuccessful searches made 10 times or more were categorized according to the following rubric:

Type	Definition
art: wrong period	An art-related search for something outside the scope of the Guggenheim's collection; e.g. "michaelangelo"
not found	A search for something that was within the scope of the Guggenheim's collection but was not in the catalogue; e.g. "zaha hadid"
not found (now ok)	A search for something that was within the scope of the Guggenheim's collection but was not in the catalogue when the search was made. Can be found December 2006; e.g. "Dadaism"
obscene	A profanity
out of scope	Searches for topics outside the scope of the Guggenheim collections catalogue; e.g. "admissions"
spelling	A search term with a spelling mistake in it: e.g. "rauschenberg" (for Rauschenberg)
wrong collection	A search for a specific work of art, not held by the Guggenheim Museum; e.g. "weeping woman"

Table 9: Types of Unsuccessful Searches

Chart 17 shows reasons for search failure by type. Almost half (46%) of the failed searches were for content likely to be found in the Guggenheim Collections on-line catalogue but not. A large proportion of searches (36%) were unsuccessful searches because of a spelling error. Confusion about the nature of the Guggenheim Collection accounted for 9% of failed searches. Requests for information outside the scope of the collections catalogue (such as admissions details) made up 3% of failed searches. Requests for works from a collection other than the Guggenheim were about 1% of this group of failed searches. Obscenities accounted for less than 1% of failed searches.

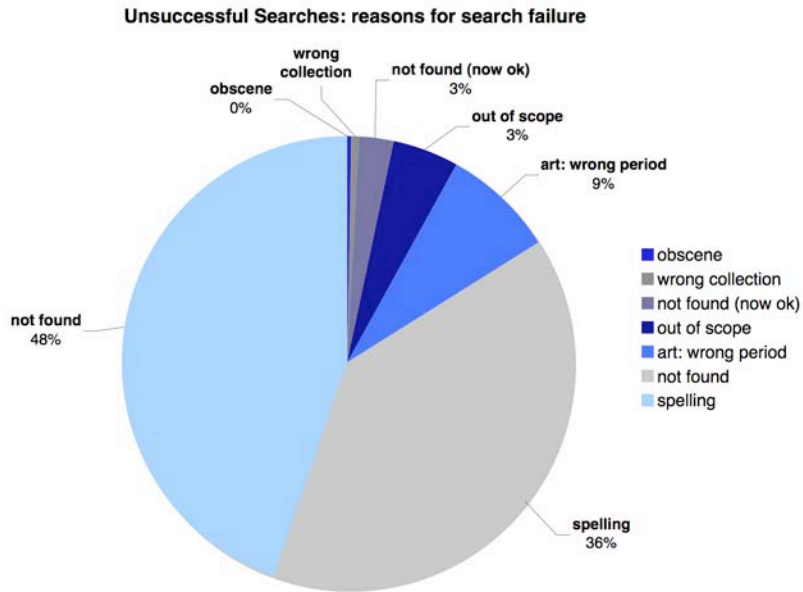


Chart 17: Unsuccessful Searches: Reasons for search failure

Categories of searches are not distributed evenly across these types of error, however. *Chart 18* shows striking clusters of errors. Almost all (98%) of the spelling errors were mis-spelled artists' names. Half (50%) of the unsuccessful artist's name searches failed because of a spelling error.

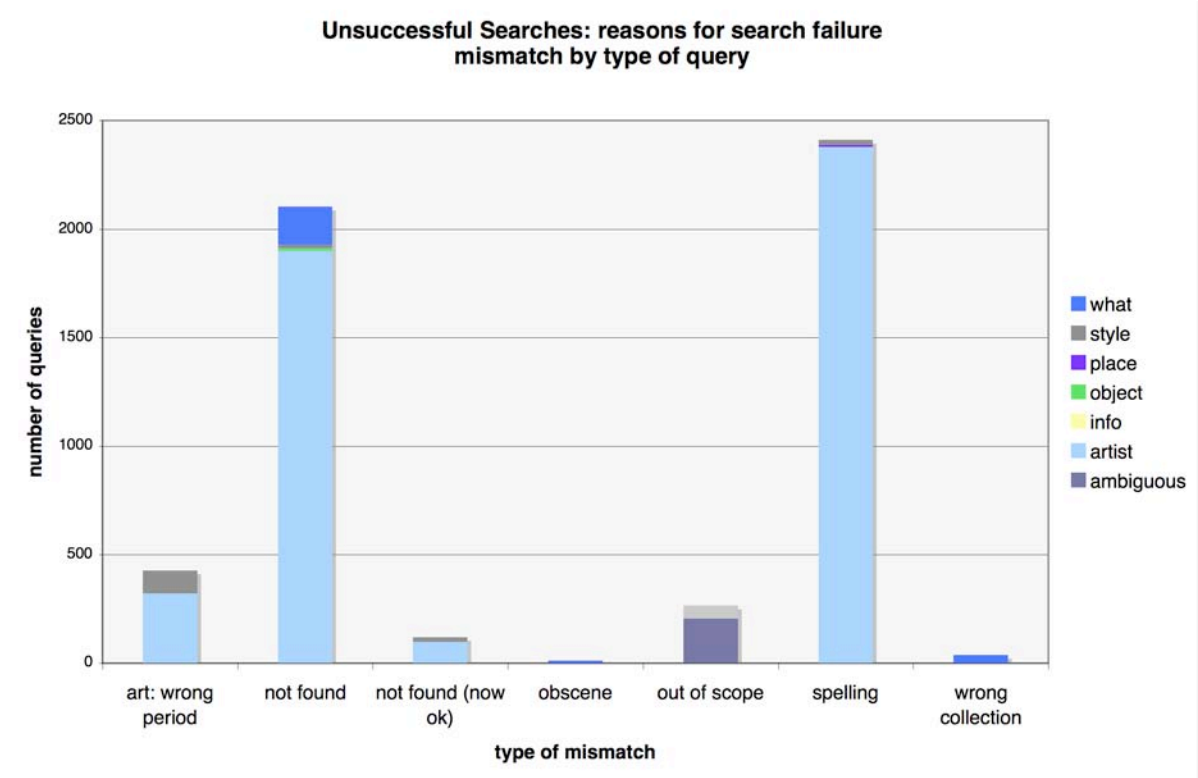


Chart 18: Unsuccessful Searches: mismatch by type of query

G. Anomalous Search Results

There were a number of anomalous entries in the search logs, that appear to have resulted from a failure of the search system. For example, the term “nelson” generated 3132 results, many of which appeared not to be related at all. An examination of the source code for the page revealed the following <meta> tag, containing the name of the developer “Sam Nelson”:

```
<META NAME="generator" CONTENT="SPS: Sam's Publishing System v4.1 by Sam Nelson sps@clevernamehere.com">
```

Similarly, the term “window” retrieved 3780 results, which include matches in the following code string that generates part of the interface, despite its being included inside

<SPS_NOSEARCH> </SPS_NOSEARCH> tags.

```
function popUp(URL, x, y) {
  var myWin = window.open(URL,myWin, "resizable=yes,scrollbars=yes,height="
  + y + ",width=" + x);
}
```

F. Analysis of Results

A review of the analysis of the search logs points to areas where changes in museum practice or approach might improve levels of service for users of the on-line collection catalog of the Guggenheim Museum.

1. Terminology and users

In summarizing their analysis of museum goers' questions, Janney and Sledge note that "the public may not know the technical vocabulary used by museum professionals." The difference between public and professional vocabularies reinforces perceived distance between the public and institutions. "It is important that users be able to ask questions in their own voice, otherwise they may feel too intimidated to ask at all. Questions should be answered in a vocabulary that matches that of the user" (Janney and Sledge 1995, Section 5. Additional Observations and Conclusion).

There is no adaptation to the user in the Guggenheim Collection on-line search facility. The language of the data is that of the contemporary art scholar and curator – though enabling browsing by specialized period or style terms does help in their definition. As the relationship of query length to successful result shows, the more discursive a query is, the less likely a result will be found. Some types of query, such as those in Table 10, are looking for a different kind of information than a museum object catalogue record. Appropriate resources could be suggested.

what was the title of the painting that van gogh first sold
what was this work of art about
what was vincent van gogh typical subject
what was wrong with matisse when he went into a wheelchair

Table 10: Sample "What" questions entered into the query box (once each)

2. Categories of Information

The majority of the queries identified by type fell easily into one of the categories identified by Janney and Sledge (Janney and Sledge 1995). The preponderance of artist-name queries is counter to the distribution of questions in the Catechism project (McCorry and Morrison 1993), but not surprising in a contemporary art museum, where the discourse focuses on individual creative output. The pairing of categories shown in the content analysis of the unsuccessful queries shows a sophisticated user-base with a strong knowledge of the content of this collection. It is likely that the distribution of types of queries would be very different in a history museum or a science museum. Available museum documentation is able to support this kind of query readily.

The concentration on queries by artist’s reflects the findings of Cunningham, Bainbridge and Masoodian, who found an emphasis on artist (82%), date (50.5 %), nationality (38.5%) and Title (31.4%) in queries identified as ‘bibliographic’ or likely to be satisfied by some kind of metadata about an information object (Cunningham, Bainbridge, and Masoodian 2004).

However, the failure of the “what” queries points to a lacunae in museum documentation practices, that may be addressed through other methods of collection description than professional cataloguing. It may be that the folksonomy that results from user tagging of art collections addresses this facet of description more effectively (Trant 2006b, Trant 2006a).

3. Artists’ Names

The failure of such a large number of artists searches because of mis-spelling points to the need for a ‘spell checking’ or suggestion function that draws on alternate forms of artists names and collocates common mis-spellings with the correct spelling.

Users try to correct themselves when they make errors, as is shown in the sequence of searches for Gauguin in Table 11

Date and Time	Search Term	Results
Mon Oct 3 10:23:16 2005	gaugain	0
Mon Oct 3 10:22:44 2005	gaugin	0
Mon Oct 3 10:22:39 2005	gauagin	0

Table 11: Series of mis-spelled queries for the artist "Gauguin"

Unfortunately, this person never guessed the spelling correct, and left without seeing any of the Guggenheim’s Collection. It would not be difficult to collocate common errors, such as those shown in Table 12.

Vasily Kandinsky	Number of queries
kadinsky	109
wassily kandinsky	101
kandisky	65
kandinskij	17
kandinsky	16
kadinski	10
vassily kandinsky	10
	328

Table 12: Misspellings of Vasily Kandinsky's name

4. *Query Routing*

Searches could be sent to the full Guggenheim Web site when a collection search produces no results. This would enable users to find information on artists (like Zaha Hadid) featured on the Web site but not found in the on-line collections catalogue.

G. Future Research

This study points to a number of areas for further research.

1. *Other modes of navigation*

Searching is only one way of gaining access to on-line museum collections. Other modes of navigation, including browsing, need to be studied to build a full picture of the user experience. (Bates 1989, Mat-Hassan and Levene 2005)

2. *Searching in other places*

The data set studied was constructed from searches made when users were already on the Guggenheim Web site, and after they had selected a search of the collection. Searchers are assumed to know that they are on a museum's Web site, and to be familiar with the kinds of data that are available. But what of searches that cross multiple museum web sites? (Rose and Levinson 2004) Referrer data from Web logs could be used to identify the terms that are used to find the Guggenheim site itself, and works within it. Searches of the full Guggenheim web site could be compared to those of the collection to see if the concept of a collections catalogue is appreciated by users.

3. *Search dialogs*

Without full Web transaction log data that enables clustering of searches by session it is not possible to review how various queries relate to each other, or how a users search strategy develops over time. Query logs should be reviewed in the context of session-related (time- and machine-based) log data (Bates 1979).

4. *User groups*

A study of Web transaction logs may enable the identification of particular user groups (e.g. students) whose use of the Guggenheim Collection on-line may be distinct (Huntington et al. 2006). These groups could be studied in depth by segmenting their queries from the rest of the log data. Other methods, such as observed query tasks (Toms et al. 2003) might enable the development of a further understanding of the needs of sub-groups and generate further recommendations for improving access to the Guggenheim Collections on-line

5. *User behaviour*

Search logs provide no insight into why users chose to search the Guggenheim Collection online. Interviews, observation, focus groups or surveys of users of the collection would all provide more qualitative information about the use of this resource, and might generate insight into how to direct users towards the most appropriate resources on the Guggenheim web site.

6. *Users Tasks*

Searching can be seen as one facet of a larger research process, and needs to be studied in light of users' objectives (Bearman and Trant 1998, Toms et al. 2003). Understanding particular information-seeking behaviour within the context of overall goals was might inform systems-design (Borgman et al. 2005).

H. Conclusions

A study of searching in a single institution cannot provide the basis for many broad conclusions. It is apparent that the online collection of the Solomon R. Guggenheim collection is used on an academic year cycle by individuals familiar with the scope and content of the museum's collection. That they search predominantly by artist's name reflects the focus of twentieth century art-making and critical theory on the creative output of single individuals. That search behaviour is tied so tightly to subject may call into question the need for museum-wide metadata repositories (Cole and Shreeves 2004). Or it may point to an adaptive user base that has figured out what kind of searching is possible and successful. Further in-depth studies of users of the Guggenheim's online collections catalogue and comparative studies are needed.

Comparative studies of end-user searching both of other art museum's on-line catalogues and of the catalogues made available history, cultural, and natural history museums, and science centres are also essential to put this data into context, and to draw broader conclusions about needs for access to museum collection's online. If museums wish to broaden their audiences, from a narrow group of scholars to the general public, supporting searches in a more approachable manner, without demanding the name of an artist (Picasso) might be more fruitful. Comparisons of search terminology to user supplied tags in the context of the *steve.museum* project should be informative.

I. References

- Art Information Task Force (AITF). 1995. Categories for the Description of Works of Art (CDWA): J. Paul Getty Trust.
- Baca, Murtha, ed. 2002. *Introduction to Art Image Access: Issues, tools, standards, strategies*. Los Angeles, CA: The J. Paul Getty Trust.
- Baca, Murtha, Patricia Harpring, Elisa Lanzi, Linda McRae, Ann Baird Whiteside, and on behalf of the Visual Resources Association. 2006. *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images*.
- Bakewell, Elizabeth, William O. Beeman, and Carol McMichael Reese. 1988. *Object, image, inquiry : the art historian at work : report on a collaborative study by The Getty Art History Information Program (AHIP) and the Institute for Research in Information and Scholarship (IRIS), Brown University*. Edited by M. Schmitt. Santa Monica, CA: The J. Paul Getty Trust (AHIP).
- Bates, M.J. 1979. Information Search Tactics. *Journal of the American Society for Information Science* 30 (4):205-214.
- . 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 407-424.
- . 1996a. The design of databases and other information resources for humanities scholars: The Getty online search project report no. 4. *Online and CD-ROM Review* 1994:331-340.
- . 1996b. The Getty-end-user online searching project in the humanities: Report no. 6: Overview and conclusions. *College & Research Libraries* 57:514–523.
- . 2002. The cascade of interactions in the digital library interface *Information Processing and Management* 38 (2):381-400.
- Bates, M.J., D.N Wilde, and S. Siegfried. 1993. An analysis of search terminology used by humanities scholars: The Getty online searching project report no. 1. *Library Quarterly* 63:1–39.
- . 1995. Research practices of humanities scholars in an online environment: The Getty Online Searching Project report no 3. *Library and Information Science Research* 17 (1):5-40.
- Bearman, D. 1994. Thesaurally Mediated Retrieval. *Visual Resources* 10:295-307.
- Bearman, D., and Toni Peterson. 1991. Retrieval Requirements of Faceted Thesauri in Interactive Information Systems. In *Advances in Classification Research: Proceedings of the 1st ASIS SIG/CR Classification Research Workshop*, edited by S. Humphrey and B. Kwasnik. Medford NJ: Learned Information.
- Bearman, David. 1990. *Archives & Museum Data Models & Dictionaries, Archives and Museum Informatics technical report, no. 10*. Pittsburgh, Pa.: Archives & Museum Informatics.
- . 2006. Representing Museum Knowledge. In *Museum Informatics*, edited by P. Marty and K. Jones: Dekker.
- Bearman, David, and Jennifer Trant. 1998. Unifying our cultural memory: Could electronic environments bridge the historical accidents that fragment cultural collections? In *Information Landscapes for a Learning Society, Networking and the Future of Libraries*, edited by L. Dempsey, S. Criddle and R. Heseltine.
- . 2006. *Survey of Museum Web Implementations, 2005*. Archives & Museum Informatics 2006 [cited December 12, 1006 2006]. Available from <http://www.archimuse.com/research/mwbenchmarks/report/mwbenchmarks2005.html>.
- Borgman, Christine, Laura Smart, Kelli A. Millwood, Jason R. Finley, Leslie Champeny, Anne J. Gilliland, and Gregory H. Leazer. 2005. Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries.

- Journal of the American Society for Information Science and Technology* 56 (6):636 - 657.
- Borland, Candace M., and Naree Wongse-Sanit. 1997. ArtsEdNet: Assessing an Arts Education Web Site. In *Museums and the Web 1997: Proceedings*, edited by D. Bearman and J. Trant. Los Angeles, California, USA: Archives & Museum Informatics.
- Carey, Sheila, and Robyn Jeffrey. 2006. Audience Analysis in the Age of Engagement. In *Museums and the Web 2006: Proceedings*, edited by D. Bearman and J. Trant. Albuquerque, New Mexico, USA: Archives & Museum Informatics.
- Chau, Michaela, Xiao Fang, and Olivia R. Liu Sheng. 2005. Analysis of the query logs of a Web site search engine. *Journal of the American Society for Information Science and Technology* 56 (13):1363-1376.
- Cherry, Rich. 2006. Email Message about Guggenheim Search Log Data. New York, August 31, 2006 2:49 PM -0400.
- Choi, Youngok, and Edie M. Rasmussen. 2002. User's relevance criteria in image retrieval in American history. *Information Processing and Management* 38 (5):695 - 726.
- . 2003. Searching for Images: the Analysis of Users' Queries for Image Retrieval in American History. *Journal of the American Society for Information Science and Technology* 54 (6):498-511.
- clevernamehere.com. 2006. *SPS (Sam's Publishing System)*, 2006 2006 [cited December 12 2006]. Available from <http://sps.clevernamehere.com/>.
- Cole, Timothy W., and Sarah L. Shreeves. 2004. Search and discovery across collections: the IMLS digital collections and content project. *Library Hi Tech* 22 (3):307-322.
- Covey, Denise Troll. 2002. Usage and Usability Assessment: Library Practices and Concerns: Council on Library and Information Resources (CLIR).
- Cunningham, Sally Jo, David Bainbridge, and Masood Masoodian. 2004. How people describe their image information needs: a grounded theory analysis of visual arts queries. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*. Tuscon, AZ, USA.
- Dalmau, Michelle, Randall Floyd, Dazhi Jiao, and Jenn Riley. 2005. Integrating thesaurus relationships into search and browse in an online photograph collection. *Library Hi Tech* 23 (3):425-452.
- Documentation Standards Working Group - Committee on Documentation (CIDOC) of the International Council of Museums (ICOM). 2006. *CIDOC Relational Data Model*, 1995 [2002-urls] 1995 [cited January 4 2006]. Available from <http://www.willpowerinfo.myby.co.uk/cidoc/model/relational.model/>.
- Doolan, John, Darren Peacock, and Derek Ellis. 2004. Searching For Meaning, Not Just Records. In *Museums and the Web 2004: Proceedings*, edited by D. Bearman and J. Trant. Washington DC / Arlington VA, USA: Archives & Museum Informatics.
- Goodrum, Abby, and Amanda Spink. 2001. Image searching on the Excite web search engine. *Information Processing and Management* 37 (2):295-311.
- Guggenheim Museum New York. 2006. *Collections Search: Help* [web site] 2006 [cited December 12 2006]. Available from http://www.guggenheimcollection.org/site/search_help.html.
- Harold, James, Paul Dusenbery, Randi Korn, and Christopher Randall. 2005. Hands-On versus On-Line: Evaluating MarsQuest On-line. In *Museums and the Web 2005: Proceedings*, edited by D. Bearman and J. Trant. Vancouver, British Columbia, Canada: Archives & Museum Informatics.
- Harpring, Patricia. 1999. How Forcible are Right Words! *: Overview of Applications and Interfaces Incorporating the Getty Vocabularies. In *Museums and the Web 1999: Proceedings*, edited by D. Bearman and J. Trant. New Orleans, Louisiana, USA: Archives & Museum Informatics.

- . 2002. The Language of Images: Enhancing Access to Images by Applying Metadata Schemas and Structured Vocabularies. In *Introduction to Art Image Access: Issues, Tools, Standards and Strategies*, edited by M. Baca. Los Angeles: J. Paul Getty Trust.
- Heller, Jack. 1968. The Value of a Computerized Data Bank as an Adjunct to a Card Catalogue. In *Computers and their Potential Applications in Museums*. The Metropolitan Museum of Art, New York: published for The Metropolitan Museum of Art by ARNO press.
- Hourihane, Colum. 2002. It Begins with the Cataloguer: Subject Access to Images and the Cataloguer's Perspective. In *Introduction to Art Image Access: Issues, Tools, Standards, Strategies*, edited by M. Baca. Los Angeles: J. Paul Getty Trust.
- Huntington, Paul, David Nicholas, Hamid R. Jamali, and Anthony Watkinson. 2006. Obtaining subject data from log files using deep log analysis: case study OhioLINK. *Journal of Information Science* 32 (4):299-308.
- Janney, Kody, and Jane Sledge. *A User Model for CIMI Z39.50 Application Profile*, September 1995 [cited May 3, 2006. Available from http://www.cimi.org/old_site/documents/Z3950_app_profile_0995.html].
- Jansen, Bernard J. 2006. Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research* 28 (3):407-432.
- Jansen, Bernard J., Abby Goodrum, and Amanda Spink. 2000. Searching for multimedia: analysis of audio, video and image Web queries. *World Wide Web Journal* 3 (4):249-254.
- Jansen, Bernard J., and Udo Pooch. 2001. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology* 52 (3):235-246.
- Jansen, Bernard J., and Amanda Spink. 2006. How are we searching the world wide web?: a comparison of nine search engine transaction logs. *Information Processing and Management* 42 (1):248-263.
- Jansen, Bernard J., Amanda Spink, and Jan Pedersen. 2003. An analysis of multimedia searching on AltaVista. *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*:186 - 192
- Jansen, Bernard J., Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management* 36 (2000):207-227.
- Jones, Steve, Sally Jo Cunningham, Rodger McNab, and Stefan Boddie. 2000. A transaction log analysis of a digital library. *International Journal on Digital Libraries* 3 (2):152-169.
- Kravchyna, V., and S.K. Hastings. 2002. Informational Value of Museum Web Sites. *First Monday* 7 (2).
- Mack, Vanessa, and Richard Llewellyn. 1998. Australian University Museums On-Line (AUMOL). *Archives and Museum Informatics* 12 (1):81-88.
- Mat-Hassan, Mazlita, and Mark Levene. 2005. Associating search and navigation behavior through log analysis. *Journal of the American Society for Information Science and Technology* 56 (9):July.
- McCorry, Helen, and Ian O. Morrison. 1993. Report on the Catechism Project.
- McKenna, Gordon, and Efthymia Patsatzi. 2005. *SPECTRUM: The UK Museum Documentation Standard*. Version 3.0 ed. Cambridge: MDA.
- Nicholas, David, Paul Huntington, Hamid R. Jamali, and Carol Tenopir. 2006. What deep log analysis tells us about the impact of big deals: case study OhioLINK. *Journal of Documentation* 62 (4):482-508.
- Nordbotten, Joan, ed. 2000. *Entering Through the Side Door - a Usage Analysis of Web Presentations*. Edited by D. Bearman and J. Trant, *Museums and the Web 2000: Proceedings*. Minneapolis, Minnesota, USA: Archives & Museum Informatics.

- Ozmutlu, Seda, Amanda Spink, and Huseyin C. Ozmutlu. 2000. Multimedia web searching trends: 1997-2001. *Information Processing and Management* 39 (11):611-621.
- . 2004. A day in the life of Web searching: an exploratory study. *Information Processing and Management* 40:319-345.
- Peacock, Darren. 2002. Statistics, Structures & Satisfied Customers: Using web log data to improve site performance. In *Museums and the Web 2002: Proceedings*, edited by J. Trant and D. Bearman. Boston, Massachusetts, USA: Archives & Museum Informatics.
- Polydoratou, Panayiota. 2006. Using web logs transactions to assess a metadata registry system's use: the case of MetaForm. *OCLC Systems & Services: International digital library perspectives* 22 (1):67-79.
- Pu, Hsiao-Tieh, Shui-Lung Chuang, and Chyana Yang. 2002. Subject categorization of query terms for exploring Web users' search interests. *Journal of the American Society for Information Science and Technology* 53 (8):617 - 630.
- Roberts, D. Andrew. 1985. *Planning the documentation of museum collections*. Cambridge: Museum Documentation Association.
- , ed. 1990. *Terminology for Museums: Proceedings of an International Conference held in Cambridge, En glad, 21-24 September 1988*. Cambridge: Museum Documentation Association.
- Rose, Daniel E., and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM.
- Sandore, B. 1993. Applying the results of transaction log analysis. *Library Hi Tech*:87-97.
- Semper, Rob, Noel Wanner, Roland Jackson, and Martin Bazley. 2000. Who's Out There? A Pilot User Study of Educational Web Resources by the Science Learning Network (SLN). Paper read at Museums and the Web 2000.
- Shatford Lane, Sara. 1998. Modelling Relevance in Art History: Identifying Attributes that Determine the Relevance of Art Works, Images, and Primary Text to Art History Research, Library and Information Science, University of California at Los Angeles, Los Angeles, CA.
- Sherwood, Lyn Elliot. 1998. Standards for access to museum content: practical solutions or technical chimeras? *Computer Standards & Interfaces* 20 (2-3):111-115.
- Siegfried, Susan, Marcia J. Bates, and Deborah N. Wilde. 1993. A profile of end-user searching behavior by humanities scholars: The Getty Online Searching Project Report No. 2. *Journal of the American Society for Information Science* 44 (5):273-291.
- Silverstein, Craig, Hannes Marais, Monika Henzinger, and Michael Moricz. 1999. Analysis of a very large web search engine query log. *ACM SIGIR Forum* 33 (1):6-12.
- Sledge, Jane. 1995. Points of View. In *Multimedia Computing and Museums, Selected papers from the Third International Conference on Hypermedia and Interactivity in Museums (ICHIM95/MCN95)*, edited by D. Bearman. Pittsburgh: Archives & Museum Informatics.
- Sledge, Jane, and Mary Case. 1995. Looking for Mr. Rococo: Getty Art History Information Program Point of View Workshop. *Archives and Museum Informatics* 9 (1):124-129.
- Sorgel, Dagobert. 1994. Indexing and Retrieval Performance: The Logical Evidence. *Journal of the American Society for Information Science* 45 (8):589-599.
- Spink, Amanda, Bernard J. Jansen, Dietmar Wolfram, and Tefko Saracevic. 2002. From E-Sex to E-Commerce: Web Search Changes. *Computer* 35 (3):107-109.
- Spink, Amanda, Dietmar Wolfram, Bernard J. Jansen, and Tefko Saracevic. 2001. Searching the Web: The public and their queries. *Journal of the American Society for Information Science* 52 (3):226-234.
- Stam, Deirdre C. 1989a. Public Access to Museum Information Pressures and Policies. *Curator* 32 (3):198-198.

- . 1989b. Tracking Art Historians: On their Information Needs and Information seeking Behavior. *Art Libraries Journal* 14 (3):13-16.
- Toms, Elaine G., Luanne Freund, Richard Kopak, and Joan C. Bartlett. 2003. The effect of task domain on search. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*. Toronto, Ontario, Canada: IBM Press.
- Trant, J. 2006a. Social Classification and Folksonomy in Art Museums: early data from the steve.museum tagger prototype. In *ASIST SIG-CR workshop on Social Classification*. Austin, TX.
- Trant, Jennifer. 2006b. Exploring the potential for social tagging and folksonomy in art museums: proof of concept. *New Review of Hypermedia and Multimedia* 12 (1):83-105.
- Tudhope, Doug, Ceri Binding, Dorothee Blocks, and Daniel Cunliffe. 2006. Query expansion via conceptual distance in thesaurus indexed collections. *Journal of Documentation* 62 (4):509-533.
- Voorhees, Ellen. 2001. The Philosophy of Information Retrieval Evaluation. *Lecture Notes in Computer Science: Evaluation of Cross-Language Information Retrieval Systems : Second Workshop of the Cross-Language Evaluation Forum, CLEF 2001, Darmstadt, Germany, September 3-4, 2001. Revised Papers* 2406:355.
- Wang, Peiling, Michael W. Berry, and Yiheng Yang. 2003. Mining longitudinal web queries: trends and patterns. *Journal of the American Society for Information Science and Technology* 54 (8):743 - 758.
- Warwick, Claire, Melissa Terras, Paul Huntington, Nikoleta Pappa, and Isabel Galina. 2006. The LAIRAH Project: Log Analysis of Digital Resources in the Arts and Humanities. Final Report to the Arts and Humanities Research Council: School of Library, Archive and Information Studies, University College London
- Wen, Ji-Rong, Jian-un Nie, and Hon-Jiang Zhang. 2002. Query Clustering Using User Logs. *ACM Transactions on Information Systems (TOIS)* 20 (1):59-81.
- Wolfram, Dietmar, Amanda Spink, Bernard J. Jansen, and Tefko Saracevic. 2001. Vox populi: the public searching of the Web. *Journal of the American Society for Information Science and Technology* 52 (12):1073-1074.
- Yi, Kwan, Jamshid Beheshti, Charles Cole, John E. Leide, and Andrew Large. 2006. User search behavior of domain-specific information retrieval systems: An analysis of the query logs from PsycINFO and ABC-Clio's Historical Abstracts/America: History and Life *Journal of the American Society for Information Science and Technology* 57 (9):1208-1220.