

Hansi Hoskins. A Server-Side User Study of UNC-Chapel Hill's 19th-Century American Sheet Music Collection Database. A Master's paper for the M.S. in L.S. degree. April, 2010. 32 pages. Advisor: Diane Kelly

This study describes a server-side user study of UNC-Chapel Hill's 19th-Century American Sheet Music Collection Database. The study was based on the transaction logs of the database and was conducted to learn about the search pathways, and habits of the users of the database.

The information collected from the transaction logs included IP addresses, time-stamps, URLs, and referrers. This data provided information about the amount of time users spend actively searching the database, the how users are querying the database, the results they are receiving from their query, and the documents they chose from the results page.

Headings:

Music Information Retrieval

Sheet Music Databases

Server-Side User Study – Sheet Music Database

19th-Century American Sheet Music

A Server-Side User Study of UNC-Chapel Hills's 19th-Century American Sheet
Music Collection Database

by
Hansi Hoskins

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April, 2010

Diane Kelly, Faculty Advisor

INTRODUCTION

Many libraries across the country maintain large collections of popular sheet music dating from the 19th century and early 20th century. Sheet music publishing during the 19th and 20th century was quite prolific, but when the popularity of sheet music waned, libraries became the recipients of the outdated media. Recently, there has been a movement to digitize these collections due to the fragility of the paper on which the music was printed (Dougan, 2006). Much of the music is now public domain, so libraries are free to make the music available to the public via internet databases. Institutions such as the University of Michigan, Duke University, the Library of Congress, Harvard University, Brown University, University of Chicago, Yale University, the New York Public Library, the Detroit Public Library, the San Francisco Public Library, Boston College, and the University of Michigan all house historical sheet music collections and are in the process of digitizing these bodies of work (Wheeler, 2005).

The vast number of collections prompted the creation of the Sheet Music Consortium (SMC) which went public in 2003. The SMC provides libraries that have digitized historical sheet music collections the ability to combine their resources in order to provide the public with a way to search multiple collections at once. Current participating institutions include Duke University, Indiana University, John Hopkins University, Library of Congress, Maine Music Box, National Library of Australia, and the University of California Los Angeles (Sampsel, 2007).

The University of North Carolina at Chapel Hill is also one of the many institutions to decide to digitize their 19th Century American Sheet Music Collection. While still in the process of uploading information, the majority of UNC's 19th Century Sheet Collection has been available online since 2009. This collection contains around 3,500 pieces of music. Each item is represented in the database by 20 metadata fields and a scanned image of each of the pages of music pertaining to the piece. Patrons can browse the collection by title, composer, publisher, topic, date, or publisher location. A free text search option which searches within the metadata fields is also available, as well as links to other online sheet music collections, including the sheet music consortium. Figure 1 shows the homepage of the database.

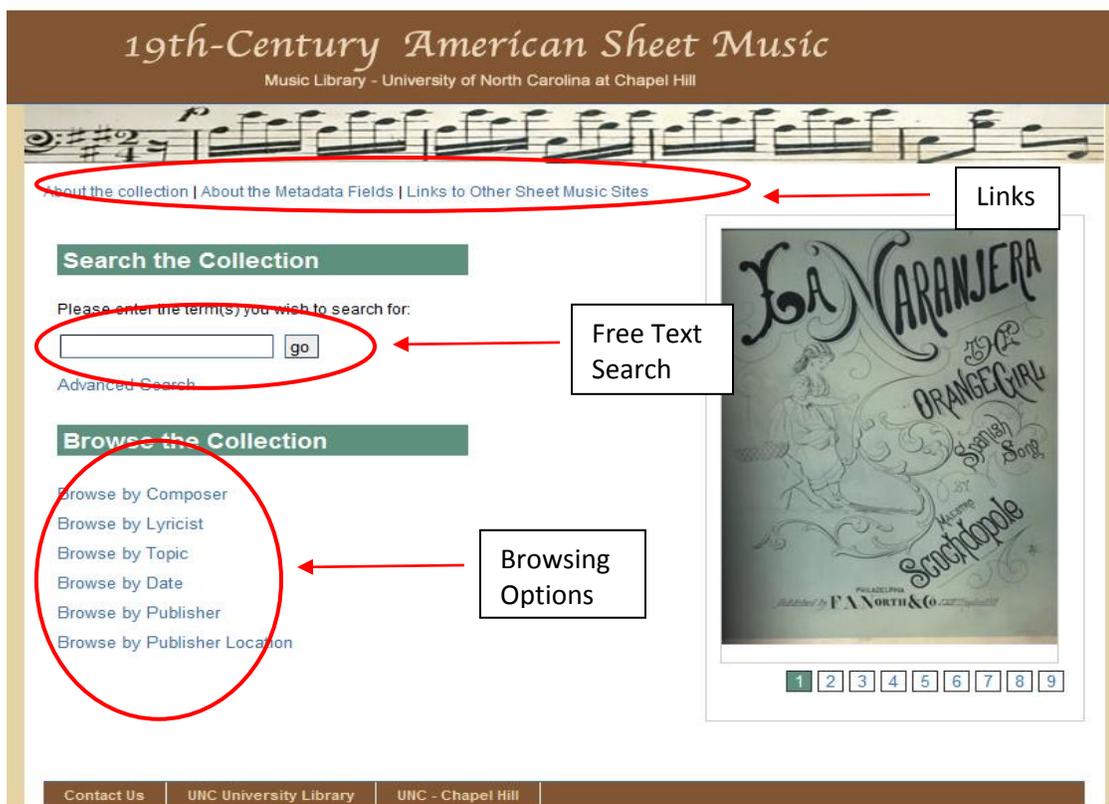


Figure 1 Screen Shot of 19th-Century American Sheet Music Collection Homepage

Although this database has been available to patrons for two years, a user study of this particular resource has never been conducted. Understanding the ways in which people search for music is a necessary part of creating and maintaining a music information retrieval system that is able to fulfill the needs of its users. Knowing how patrons are using this database is essential if UNC Chapel Hill wants to keep their database relevant and up to date, or if they want to incorporate new search methods.

LITERATURE REVIEW

The goal of this literature review is to provide the reader with a working knowledge of the field of music information retrieval in order to provide a basis for understanding the necessity of conducting a user study on the 19th Century Sheet Music Collection at the University of North Carolina at Chapel Hill.

Music Information Retrieval

In order to create and analyze an historical sheet music database one must have a working knowledge of the field of music information retrieval. Music information retrieval (MIR) is an interesting and complex field of study. According to J. Stephen Downie, “MIR is a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world’s vast store of music accessible to all” (2004, p.12). MIR concerns catalogers, music librarians, computer scientists, musicologists, and the general public. It is an important field of study which has implications for all realms of alternative media searching.

The origins of music information retrieval lie in the thematic catalogs of musical works. Thematic catalogs, such as *A Dictionary of Musical Themes* (Barlow, Harold & Morgenstern, Sam, 1948) and *A Dictionary of Opera and Song Themes* (Barlow, Harold & Morgenstern, Sam, 1950), contained short passages of music which represented larger works. With the advent of digital searching came many new opportunities for MIR research (Downie, 2000). Unfortunately, there was no standard for evaluating and talking about the many creative solutions researchers were inventing to solve the major issues of

digital music information retrieval. Due to this need for a scholarly community concerned with music information retrieval, the International Society for Music Information Retrieval was born.

MIR researchers are supported by several professional organizations, including the International Society for Music Information Retrieval (ISMIR). Since its first official conference in 2000, ISMIR has been providing music information scholars with a supportive research community and bibliographic materials (<http://php.indiana.edu/~donbyrd/DonMusicIRBibliography.HTML>). One of the most important things that ISMIR has achieved is creating an evaluation standard for music information retrieval research (Downie, 2004). In his article “The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future,” Downie details the process of creating a standard evaluation system. The need for a standardized way to evaluate music information retrieval systems arose from the numerous systems which were being created to accommodate content-based searching. Other systems for evaluation existed already, such as the paradigm that is used by the text retrieval community, but due to various music specific issues, a system based on the text retrieval paradigm would not work. The four main problems fingered by the ISMIR community include the complexity of music, the complexity of music queries, the evaluation metrics of relevance, precision, recall in relation to music, and intellectual property law. These four problems are the basic issues each new system must overcome, and as such they became the basis for evaluating the systems.

Creating a Music Information Retrieval System

In order create a way to be able to retrieve music in any of its formats (scores, recordings, etc.) from an information retrieval system, the physical item must be digitally represented in the system. “Music information is a multifaceted amalgam of pitch, tempo, rhythmic, harmonic, timbral, textual (e.g., lyrics and libretti) editorial, praxis, and bibliographic elements” (Downie, 2004). The complexity of music is the most difficult problem for database creators to overcome. There are two ways to describe musical materials: defining the item by its bibliographic properties, and describing the music itself. Defining the item by its physical properties (title, composer, performer, album art, etc.) is done by providing metadata tags. However, music is made up of more than just bibliographic properties—there is the actual music to consider as well. In an attempt to discover the properties of music that people use to identify individual pieces, Kim and Belkin devised a study in which participants listened to seven classical works of music and listed terms they associated with the pieces along with terms they would use to search for the pieces. The results of this study showed that there were seven main categories of search terms people were applying to music, including movements, neutral concepts, emotions, nature, objects, occasions or filmed events, and musical features. The importance of this study is that it illustrates the need for a search tool that allows users to search for music using search terms that are not necessarily music specific (Kim and Belkin, 2002).

Most current research in music information retrieval focuses on providing patrons with a way to conduct content-based queries. Content-based queries are queries which focus on the music itself. This type of query could be formatted in several ways.

One way is the ‘real-world’ query. This type of query allows the user to format their search in full sentences that describe the music. Lesaffre, DeVoogt, Leman, Baets, Meyer, and Martens discuss one research project which explored how people describe musical content (Lesaffre, DeVoogt, Leman, Baets, Meyer, Martens, 2008). The research team conducted a preliminary survey which questioned people about their favorite music, whether they were musically trained, and if they played an instrument. From the list of favorite songs, a database was created in which the final participants added semantic annotations to the pieces of music. The study yielded interesting results, which suggested that when semantic descriptors were added to music items in a music information retrieval system, users would be able to conduct content-based searches that would yield accurate results. Pickens discusses an alternative system in which music is reformatted as a text document in order to be able to apply text based queries to musical entities. The study was successful in creating a searchable database of monophonic, reformatted music, but stated that in order to index polyphonic music another system would be needed (Pickens, 2000).

Another option that would provide users with a method for conducting content-based searches is a query-by-humming or query-by-singing search tool that could be incorporated into a music information retrieval system. A query-by-humming/singing system would allow users to search for music by singing or humming. The problem associated with creating this type of search tool is navigating the range of users’ musicality. A system such as this would have to be able to match a user’s interpretation of a piece of music to an actual piece of music in the database. A user’s query could include wrong notes, wrong rhythms, and transposed musical keys, making the retrieval

process difficult (Byrd and Crawford, 2002). Pardo and Birmingham looked at how well a query-by-humming system could be expected to perform by comparing three QBH systems to three human counterparts. This study showed that the two types of QBH systems which were tested (Hidden Markov Model and probabilistic string matching) did not achieve human level performance (Pardo and Birmingham, 2002).

Content-based searching is very much so the future of MIR. However, in order to determine whether or not there is a need for this type of technology used in correspondence with UNC Chapel Hill's 19th-Century American Sheet Music Collection a user study of the current data base must be conducted.

User Studies

One way to determine whether a musical information retrieval system such as a historical sheet music database is meeting the needs of its users is by conducting a user study. User studies have been conducted in various library settings for many years. According to Inskip, Macfarlane, and Rafferty, "The purpose of evaluation is to measure the performance of an information retrieval (IR) system and to help determine how effective it is at meeting the information needs of the users" (Inskip, Macfarlane, and Rafferty, 2010:p. 518). User studies provide valuable information to librarians, such as whether patrons are finding what they need, the type of searches they are conducting, how much time they are spending on each search, and much more.

Before the invention of the computer, user studies focused on analyzing user surveys, questionnaires, and interviews. This type of data provided researchers with plenty of information about the attitudes and opinions of users, but it did not provide objective

information about users' search habits (Ford, 1977). Once the computer was introduced to library settings and card catalogs went digital, new research opportunities presented themselves. Digital databases were capable of retaining transaction logs that provided researchers with the chance to analyze the large amounts of objective data needed to view trends in search behavior. The next large change in user studies came with the invention of the Internet and Google. When library catalogs went online, patrons were able to search from the comfort of their homes at any time—day or night. The introduction of Google to the internet changed the way in which library patrons formatted their searches and how they used library search engines (Martin, 2006; King, 2008; Nicholas 2008).

The ways in which researchers analyze search data varies depending on the type of data and what they want to discover. Transaction logs are one of the most common forms of collected data used in user studies of sheet music databases. Transaction log analysis provides researchers with objective data on a large scale (Sheble and Wildemuth, p. 167, 2009). There are two types of transaction logs: data- server-side logs and client-side logs. Both types of transaction logs have their benefits and their disadvantages. Server-side logs are easy to gather, data capture is unobtrusive, and the data is objective in nature. Client-side logs are resource intensive, but this type of data can provide more information about users' actions. Once the data is mined and analyzed, trends can be seen that will show weaknesses in the database. A good overview of search log analysis is detailed by Jansen (Jansen, 2006). Jansen discusses the recent history of search log analysis and the benefits of using search log analysis to discover information about users query habits. He also presents a methodology for implementing a search log analysis study and the limitations of such a study.

The digital era drastically changed the ways in which user studies are conducted. In addition to being able to analyze transaction logs, researchers are also able to access a larger group of users through online user surveys and questionnaires. The combination of subjective and objective data has provided researchers with a large amount of information about how users search for information. This information is then analyzed and used to improve information retrieval systems and user services. Other benefits of conducting user studies include learning about how to increase collection usage and user satisfaction.

METHODOLOGY

For this study, data was collected from the search logs of UNC-Chapel Hill's 19th-Century American Sheet Music Database (<http://www.lib.unc.edu/dc/sheetmusic/?CISOROOT=/sheetmusic>). Ninety-four days worth of data was collected from the period of time between November 29, 2010 and March 2, 2011. Four fields were extracted from the search logs by David Romani, Library Systems Department staff at UNC-Chapel Hill, including IP addresses, date and time stamps, the requested URL, and the referrer. The URL and the referrer were retrieved using the following command lines:

```
ack 'CISOROOT=/sheetmusic.+results\.php' /logs/Content5/access.log | awk -F" " '$11
!~/^"-\/ { print $1"\t"$4"\t"$7"\t"$11}' | sed 's/\t[/^/ >> file.txt
```

```
ack 'results\.php.+CISOROOT=/sheetmusic' /logs/Content5/access.log | awk -F" " '$11
!~/^"-\/ { print $1"\t"$4"\t"$7"\t"$11}' | sed 's/\t[/^/ >> file.txt
```

```
sort -u file.txt | sort -n > sorted.txt
```

The data was then reformatted from a .txt file to an .xlsx file in order to be able to analyze it in Microsoft Excel.

In order to mask the IP addresses, individual numeric codes were added to each entry representing the IP addresses. Once the IP addresses were replaced, individual sessions were identified based on the identifying number assigned to each IP address and the time stamp. When each session had been identified, I was able to follow the URL that

corresponded with the beginning of each session and follow the pathway of each user.

The URL took me to the results page of the user's query. On this page the query is present. I recorded this query along with the number of results the query returned. Figure 2 shows a screen shot of a results page from the 19th-Century American Sheet Music Collection.

The screenshot shows the search results page for the 19th-Century American Sheet Music Collection. The search term "War" is circled in red, and a red arrow points from a box labeled "Search Term" to it. The number of results, "1-20 of 27 item(s)", is also circled in red, with a red arrow pointing from a box labeled "Number of Results" to it. The page displays a list of search results with columns for Image, Composer, and Title. The first result is "Over the river 'neath the shade of the trees" by Armand, and the second is "A Soldier's Life from the Bohemian Girl" by Balfe, Michael William.

Figure 2 Screen Shot of Search Results Page

The referrer statement indicated when the user opened a link to one of the pieces of music on the results page. I followed these statements and recorded the volume information as well as the metadata field in which the original query was found. The result of this process can be seen in Figure 3.

Page Description	
Title	Page 1
Repository	University of North Carolina at Chapel Hill. Music Library
Digital Collection	19th-Century American Sheet Music
Document Description	
Composer	Armand
Title	Over the river 'neath the shade of the trees
Publisher	A.E. Blackmar
Publisher Location	New Orleans, LA
Publication Date	1866
Publisher's Number	280_3
Pages	4
Instrumentation	voice; piano
Topic	American Civil War
Language	English
Inscription	L young L young
Price	3
Volume	Old Series 44,29
Repository	University of North Carolina at Chapel Hill. Music Library
Digital Collection	19th-Century American Sheet Music

Figure 3 Screen Shot of Chosen Search Result

Once all the data was collected, I separated the fields into two sections: time data and search data. As stated earlier, the time data consists of the date and the time stamp of each transaction completed by a user. In order to determine the amount of time each user was spending per session, this data was analyzed by subtracting the first time stamp of each user session from the last time stamp of the session. These subtotals were then added together to determine the total amount of time spent on the database. The subtotals were also averaged to determine the average amount of time users are spending on the site.

The second section of data that was analyzed was the search data. This data consists of the volume and item number of each requested document, the metadata field in which the query was located within the requested document, the query entered to begin the search, and the number of results each query returned. The volume information was subtotaled to determine how many times each individual piece of music was being looked

at, and which volumes were being utilized the most. The queries were separated into four categories: dates, people, text strings, and volumes. These categories were then subtotaled to determine the preferred query category. The actual queries were also subtotaled to determine which queries were used most frequently.

Other analyses that were applied to the data included comparing the results data and the time data in order to determine whether there is a correlation between the amounts of time users were spending on the database and the number of documents they considered relevant to their search. A comparison was also done of the number of relevant documents users were finding on their first trip to the database versus repeat trips.

RESULTS

Users

Over the course of the data collection period, 884 IP addresses were recorded as having queried the 19th-Century American Sheet Music Collection database. This number does not designate individual people as people are not inseparable from their computers. It does designate individual search sessions during which queries were entered and which documents may have been retrieved. Of the 884 IP addresses recorded, 80 were recorded as having returned to the website at least once.

Time

One important goal of this study was to determine how much time users were spending browsing the UNC-Chapel Hill's 19th-Century American Sheet Music database and how that time affected their searching. The data collected provided answers to following questions regarding time usage:

- What is the average amount of time users are spending on the database?
- What is the total amount of time users spent on the database?
- Are there any apparent trends in usage over the course of the 94 day period?

The total amount of time logged on the database was 93 hours 47 minutes and 4 seconds. The average amount of time users spent actively searching the site was 5 minutes and 27 seconds with a standard deviation of .0114. The median amount of time spent searching was 18 seconds.

Of the five months during which data was collected, December had the highest number of search sessions. During this time 870 queries were used to locate documents within the database. One reason for this may be that December coincides with the end of the fall semester at UNC-Chapel Hill, meaning that many final projects are due and students might be using the database more in order to complete their research. Figure 4 illustrates the breakdown of queries entries per month.

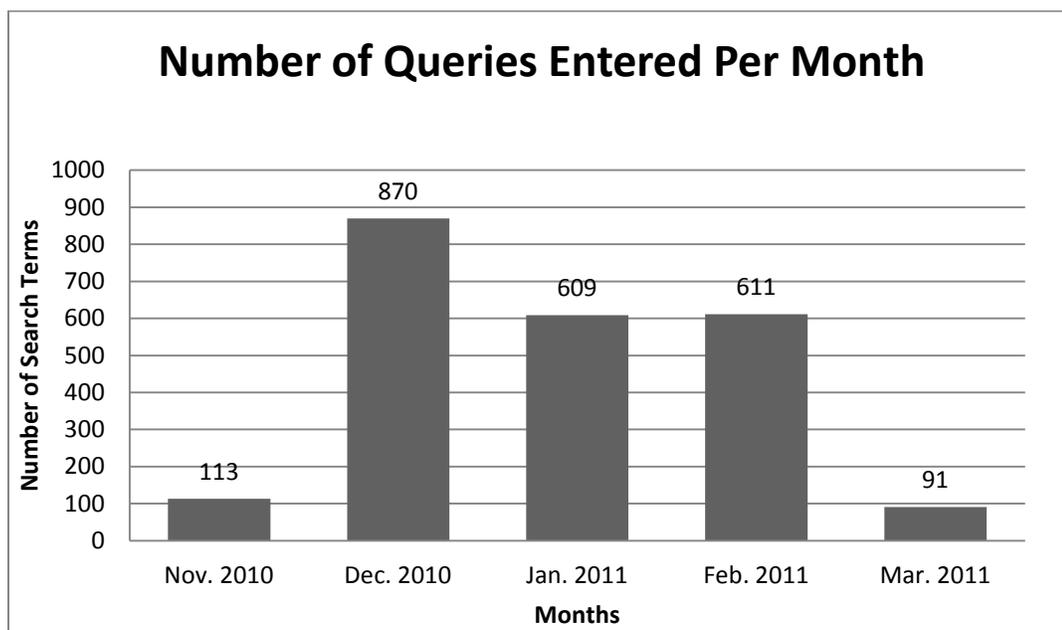


Figure 4 Number of Queries Entered Per Month

Figure 4 also shows that usage during November and March appear much lower than December through February. This is because the data only covered two days from both November (the 29th and the 30th) and March (the 1st and the 2nd).

Searching the Database

The second goal of this study was to discover what users are doing once they enter the database. The data that was collected to help answer this question included the

queries people used, the documents they chose from the results page, and where the query was found within the chosen document. During the period of time that data was collected, 2,292 search queries were used to retrieve results from the database. Of those queries, 1,327 were used more than once. The top twenty queries can be found in Table 1.

Table 1 Top Twenty Queries Used

Search Term (number of results returned)	Number of Times Search Term Was Used
1815 (6 items)	44
A Gentleman of Maryland (1 item)	27
Bach, J.S. (1 item)	22
Oesten, Theodore (24 items)	21
1900 (2 items)	19
A Lady of Baltimore (4 items)	15
Burgmuller (43 items)	15
Mozart, Wolfgang Amadeus (19 items)	15
Oesten (25 items)	15
1816 (10 items)	13
American Civil War (14 items)	13
Beyer (78 items)	12
Wyman, Addison P. (9 items)	12
A. & J.P. Ordway (8 items)	11
A. Fiot (163 items)	11
Buck, Francis (21 items)	11
Burgmuller, Friedrich (27 items)	10
Herz (58 items)	10
A Lady of Charleston (2 items)	9
Carnival of Venice (3 items)	9

Of the top 20 searches, three are publication dates, thirteen are people, two are publishers, and two are other text strings. Publication dates are a metadata field that is based on the publication date of the piece of music. The publication dates can be identified as such by comparing the number of search results returned by the search in the publication date field to the number of results provided by the 'Browse by Publication

Date' option. It is important to note that dates other than publication dates might be found in other metadata fields, but based on the number of results returned, none of the dates used to query the database were typed into the search box.

Searches for people can include lyricists, performers, composers, arrangers, dedicatees, and inscribers. These searches are formatted as names, possibly including titles (e.g. Mr., Mrs., Hon., etc). Users who are searching for people might utilize the 'Browse by Composer' or 'Browse by Lyricist' options, or they may enter a name into the free text search box.

The publisher category consists of the publishers of the sheet music. Publishers are considered corporate entities and therefore differ from people. They can be distinguished from people by looking through the list of names of publishers found in the 'Browse Publisher' option on the website.

The fourth category, other text strings, is split up into three sub-categories: topics, volumes, and text strings of two or more words, which can include titles, subtitles, or lyrics. Topics can be found using the browsing option, but volumes and longer text strings cannot. These four categories represent the majority of searches attempted by users. Figure 5 illustrates the percentage of each of the preferred query categories out of the total number of queries. As the chart shows, searches which fall under the 'People' category are the most popular and account for 59% of the total number of searches followed by Text Strings and Dates.

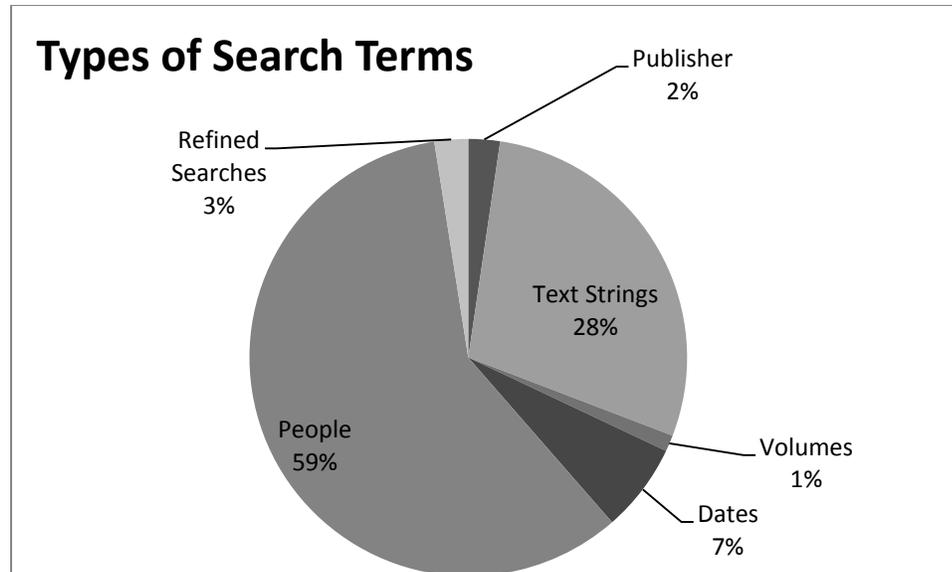


Figure 5 Percentage of Types of Queries Utilized by Users

These percentages can be explained by the placement of the browsing options of the website. On the homepage of the website, the browsing options are listed in the following order:

- Browse by Composer
- Browse by Lyricist
- Browse by Topic
- Browse by Date
- Browse by Publisher
- Browse by Publisher Location

When comparing the percentages of the types of entered queries to the order of the browsing options it seems like the percentages correspond with the ordering of the options. The most common type of query was for 'People' and the top two browsing options are 'Browse by Composer' and 'Browse by Lyricist'. This trend continues for the

other preferred query types as well. This connection suggests that the website interface may affect the search process of users.

Search Options

There are two ways to search the 19th-Century Sheet Music Collection database. A user can choose to enter a free text search into a text box, or they can browse the collection via six different metadata fields including composer, lyricist, topic, publication date, publisher, and publisher location. When analyzing the data, it was difficult to distinguish between free text searches and browsing searches. However, there were some cases when it was apparent the user was utilizing the free text search, such as when they would search for titles (as there is no browsing option for title), or when there would be two different numbers of results returned for the same query. For example, a commonly used term found under the 'Browse by Topic' option is 'Christmas'. When someone searches for items pertaining to 'Christmas' by using the browse option they will receive one result, but if they type Christmas into the free text search box they receive 118 results. This is because the 'Browse by Topic' option only searches the 'Topic' metadata field whereas the free text search looks at all of the metadata fields. This is a big issue for the database because it affects the efficacy and utility of the search. Free text searches were also apparent when peoples' names were misspelled or incomplete.

Sometimes, people would narrow their search if the query yielded too many results by choosing one of the query options in the 'Refine Your Search' box on the left side of the results page. This option was utilized by 36 times out of the 2,292 total searches. This option is illustrated in Figure 6.

19th-Century American Sheet Music
Music Library - University of North Carolina at Chapel Hill

sheetmusic home | browse | 19th Century American Sheet Music Home | advanced search | help | my favorites | home

Search results for **Christmas** search again

results 1-20 of 118 item(s)
select all : delect all : add to favorites

Refine your search

Composer
Beethoven, Ludwig Van (5)
Bellak, James (5)
Jullien, Louis Antoine (5)
Labitzky, Joseph (5)
D'albert, Charles (4)
Show more...

Topic
God (1)
Night (1)
Christmas (1)

Image: ▲ **Composer:**

1. [View in 3D](#) Sontag polka piano

2. [View in 3D](#) Brigand polka piano

3. [View in 3D](#) Operatic waltz quadrille piano

In this case, the user can refine their search by Composer or Topic.

Figure 6 Screen Shot Illustrating Refined Search Options

An advanced search option (located as a link underneath the free text search box) is also available to users. The advanced search option will link a user to UNC-Chapel Hill's Digital Collections homepage. At this page the user can choose the digital collection(s) and metadata field(s) they want to query. This option was utilized 85 times by users.

Information was also gathered on the volumes in which the relevant documents were found. The 125 physical volumes that contain the musical pieces which have been scanned and uploaded are located in the Music Library of UNC Chapel Hill. The volumes were all donated to the Music Library by individuals. Each volume consists of musical pieces that were collected and bound together by the previous owner of the sheet music. The volumes are distinguishable by their volume number. The volume numbers that have been assigned to the volumes consist of either a roman numeral, a series designation (Old Series or New Series) and an Arabic numeral, or a proper name (Person or White) and an Arabic numeral. Knowing which volumes are being most utilized could be helpful to the

music librarians at UNC-Chapel Hill as it provides them with information about collection usage. The top ten most used volumes can be seen in Table 2.

Table 2 Top Ten Volumes Utilized by Users

Volume	Number of Times A Document Was Found Within The Volume
I	17
Old Series 32	13
Old Series 88	10
LXX	9
Old Series 69	9
XX	9
Old Series 40	7
Old Series 66	7
XXIX	6
LI	5

Other Findings

As stated above, of the 884 recorded search sessions, 80 IP addresses were recorded as having returned to the database at least once. When comparing the first recorded search sessions of an IP address to the return sessions of the same address, an interesting trend was discovered. Once a user enters a search query they are presented with a page of results. From this results page, they may choose one of the documents by clicking on the thumbnail image of the cover of the sheet music. The documents that they click into can then be considered relevant to the user because the user chose the document. This does not necessarily mean that the document was relevant to the search; it simply means the user found something they wanted to look at and selected the item. IP addresses that were recorded as having returned to the database more than once were 3 percent more likely to find documents that they wanted to select than those that had visited the site only once. This could indicate that users with more experience using the

site are more likely to find documents they want to select, or it could simply mean that users are repeating searches to find documents they found during previous sessions. It could also mean that repeat searcher rely more heavily upon the database than those who only visit the site once. This information is illustrated in Figure 7.

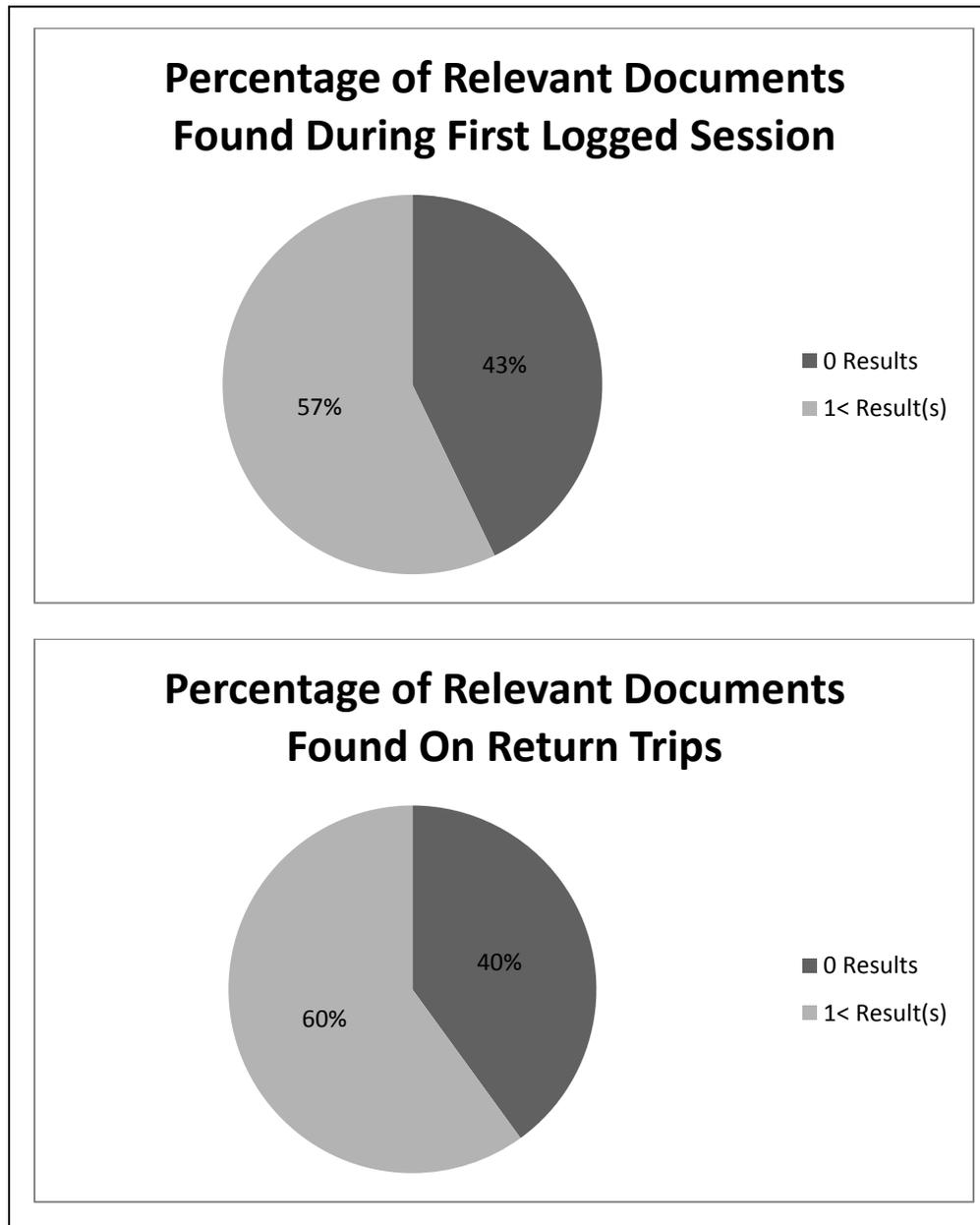


Figure 7 Comparison of the Number of Documents Selected During First Logged Search Session VS. Return Sessions

completing numerous searches during each session. This can be presumed because the search logs provide information about when a user interacts with the database by clicking on something or entering text into a free text search. The more queries that are entered into the database and the more documents that are chosen by the user, the longer the active search session will become.

Another interesting discovery was that there were several instances of a user having translated a results page to a language other than English by using Google Translate. Two sessions were translated into Portuguese and one session was translated to Italian. Knowing that the site is being translated into languages other than English has implications for user accessibility. When considering possible improvements to the website, providing a translation tool on the website could be a valuable addition to users whose first language is not English.

Issues with the data

Although the data collected provides answers to many questions about the usage of UNC-Chapel Hill's 19th-Century American Sheet Music Collection which have previously gone unanswered, as with any research project, there are limitations to this study.

First of all, IP addresses do not accurately designate individual users. This means that we still do not know how many people are regularly using this database, let alone anything about the individual users. We do not know who they are (students, professors, or researchers), how they learned about this site, why they are using the database (other than to simply find sheet music), or what they do with the information. In order to get

this information, a second study should be implemented to secure user-side information. This type of study, which could utilize surveys, questionnaires, and interviews, would also inform collection managers about user satisfaction with the services provided by the database.

Other limitations of this study are that the period of time during which data was retrieved covered three full months and two partial months including the months during which the fall semester ends and the spring semester begins. This study does not show how the university calendar can affect the usage rates of the database. In order to have a complete picture of the usage of the database, data would need to be collected over the course of an entire year. If data was collected throughout the entire year, trends in usage based on the scholastic year would be accounted for.

CONCLUSION

Maintaining the 19th Century Sheet Music Collection at the University of North Carolina at Chapel Hill allows the university to preserve this informative form of media for future generations to enjoy and access. The collection is currently being cataloged and uploaded using the database software, CONTENTdm. This software allows patrons to view scanned images of the actual sheet music along with corresponding metadata tags online.

This study used server-side data retrieved from the web search logs of the 19th-Century American Sheet Music Collection in order to shed light on the search pathways and habits of users of the database. Because a study of this kind has never been conducted in regards to this collection, the data collected contained information that had never before been reviewed. The findings provided new information regarding length of search sessions, peak usage during the period which data was collected, the effects of time on the number of documents selected per search session, the types of key words used to find documents, the documents chosen by users, and the most common volumes utilized via the database.

In addition to providing information about users' search habits, this study provides a foundation upon which additional studies can be conducted. Studies conducted during different times throughout the school year, user-side studies which focus on the user's perspective of the database, and other studies concerned with the content of the database would all benefit from the information found in this report.

BIBLIOGRAPHY

- Barlow, H. & Morgenstern, S. (1948). *A dictionary of musical themes*. New York: Crown Publishers.
- Barlow, H. & Morgenstern, S. (1950). *A dictionary of opera and song themes*. New York: Crown Publishers.
- Birmingham, W., Dannenberg, R., Wakefield, G., Bartsch, M., Bykowski, D., Mazzoni, D., Meek, C., Mellody, M., & Rand, W. MUSART: Music retrieval via aural queries. *Proceedings of the Second International Conference on Music Information Retrieval*, 73, 8.
- Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. *Information Processing and Management*, 38(2), 249-272. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57538813?accountid=14244>
- Dougan, K. (2006). Online sheet music projects and metadata from a public service perspective. *Music Reference Services Quarterly*, 9(1), 1-11. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57684973?accountid=14244>
- Downie, J.S. (2004). Music information retrieval research bibliography [web site]. *Canadian Journal of Information and Library Science*, 28(2), 45-46. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57549702?accountid=14244>
- Downie, J.S., & Nelson, M. (2000). Evaluation of a simple and effective music information retrieval system. *Proceedings of ACM SIGIR 2000 Conference on Research and Development in Information Retrieval*. New York: Association for Computing Machinery.
- Ford, G. (1977). *User studies : an introductory guide and select bibliography*. University of Sheffield: The Centre.

- Haus, G., Longari, M., & Pollastri, E. (2004). A score-driven approach to music information retrieval. *Journal of the American Society for Information Science and Technology*, 55(12), 1045-1052. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57622533?accountid=14244>
- Inskip, C., Macfarlane, A., & Rafferty, P. (2010). Creative professional users' musical relevance criteria. *Journal of Information Science*, 36(4), 517-529.
- Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research*, 28, 407-432.
- Lesaffre, M., De Voogdt, L., Leman, M., De Baets, B., De Meyer, H., & Martens, J. (2008). How potential users of music search and retrieval systems describe the semantic quality of music. *Journal of the American Society for Information Science and Technology*, 59(5), 695-707. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57657096?accountid=14244>
- Kim, J.Y., & Belkin, N.J. (2002). Categories of music description and search terms and phrases used by non-music experts. *Proceedings of the Third International Conference of Music Information Retrieval*, 209-214.
- Lippincott, A. (2002). Issues in content-based music information retrieval. *Journal of Information Science*, 28(2), 137-142. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57549458?accountid=14244>
- Martin, P. (2006). Google as teacher: Everything your students know they learned from searching google. *College & Research Libraries News*, 67(2), 100-101. Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57635311?accountid=14244>
- Pardo, B., & Birmingham, W. (2003). Query by humming: How good can it get? In *The MIR/MDL Evaluation Project White Paper Collection*, 107—109.
- Pickens, J. (2000). A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. *Proceedings of the First International Conference on Music Information Retrieval*.
- Sampsel, L.J. (2007). Sheet music consortium. *Digital Media Reviews, Notes*, 64(1), 663-667.
- Sheble, L. & Wildemuth, B.M. (2009). Transaction Logs. In B.M. Wildemuth (Au.), *Applications of social research methods to questions in information and library science* (pp. 166-177). Westport, CT.: Libraries Unlimited.

- Taheri-Panah, S. (2004). Music information retrieval systems: Why do individuals use them and what are their needs? *Proceedings of the Fifth International Conference on Music Information Retrieval*, 455-460.
- Wheeler, M. B., & Venetis, M. J. (2005). Evaluation of web access to historical sheet music collections and music-related iconography. *First Monday*, 10(10) Retrieved from <http://search.proquest.com.libproxy.lib.unc.edu/docview/57642546?accountid=14244>