Analysis of User Needs and Information Features in Natural Language Queries seeking Music Information

Jin Ha Lee

Information School, University of Washington, Box 352840, Mary Gates Hall, Seattle, WA 98195. Phone: (206) 685-9937. Fax: (206) 616-3152. Email: jinhalee@uw.edu

Abstract

Our limited understanding of real-life queries is an obstacle in developing music information retrieval (MIR) systems that meet the needs of real users. This study aims, by an empirical investigation of real-life queries, to contribute to developing a theorized understanding of how users seek music information. This is crucial for informing the design of future MIR systems, especially the selection of potential access points, as well as establishing a set of test queries that reflect real-life music information seeking behavior.

Natural language music queries were collected from an online reference website and coded using content analysis. A taxonomy of user needs expressed and information features used in queries were established by an iterative coding process. This study found that most of the queries analyzed were known-item searches, and most contained a wide variety of kinds of information, although a few features were used much more heavily than the others. In addition to advancing our understanding of real-life user queries by establishing an improved taxonomy of needs and features, three recommendations were made for improving the evaluation of MIR systems: (i) incorporating user context in test queries, (ii) employing terms familiar to users in evaluation tasks, and (iii) combining multiple task results.

Introduction

One of the major issues in current Music Information Retrieval (MIR) research is the lack of empirical studies of real-life users and their music information seeking and retrieval processes. The absence of rigorous and comprehensive studies of the MIR system users in the field (Byrd & Crawford, 2002; Downie & Cunningham, 2002; Futrelle & Downie, 2003; Cunningham, 2003) has led researchers to rely on anecdotal evidence and a priori assumptions of typical usage scenarios for designing MIR systems (Downie & Cunningham, 2002; Cunningham et al., 2003). Without rich understanding of user needs and behaviors, the MIR community is running the risk of developing ill-suited systems for the users (Cunningham et al.,

2003). Futrelle and Downie (2003) argued that the almost complete non-existence of studies on the needs of potential MIR system users is the very reason why current MIR research is weak on evaluation and application to real users.

In the general field of information seeking and retrieval, we have well-developed literatures, and various information seeking theories and models exist. To name a few, there are Bates' (1989) Berrypicking model of information seeking, Kuhlthau's (1991) Information Search Process model, Dervin's (1992) sense-making, Savolainen's (1995) Everyday Life Information Seeking (ELIS) model, Ingwersen's (1992) cognitive model, Wilson's (1999) model of information seeking and more. These theories and models, however, mostly concentrate on describing the general processes of information seeking/searching from a broad perspective. Therefore, applying these theories and models only provides limited information for information seeking/searching behaviors in specific contexts such as music information seeking and retrieval. MIR presents some distinctive issues, thus warranting further investigation.

One clue that MIR differs in important respects from the general text IR can be found in the nature of how people encounter and interact with music. People often first encounter new music from various media (e.g., a movie, TV commercial, radio) without ever knowing the artist name and song title, or they often forget the information over the course of time. This brings challenges to known-item searches for music as those users must attempt to describe the sought music other than using the bibliographic information. In addition, the universal appeal of music to an extremely broad and diverse group of people can make seeking music information more difficult and frustrating. For instance, people who had no formal music education or people who seek music from different culture or music in non-native languages can experience difficulties describing the music they seek (Lee et al., 2005a; Orio, 2006). Furthermore, the concept of subject in music has always been uncertain and more difficult to grasp than in text (McLane, 1996; Byrd & Crawford, 2002; Kim & Belkin, 2002). One may conduct a search using lyric words, but not all musical works are accompanied by lyrics. Even for vocal music, it is often difficult to comprehend the lyric terms, thus they are often misheard, especially for foreign songs.

These unique aspects of music are the very reasons why common bibliographic access points (e.g., author, title, and subject) alone may not be sufficient for users seeking music objects or information about those objects, and it is particularly interesting to study the users' behaviors in this context. It is generally acknowledged that the multidimensional nature of music requires multiple features in order to represent a music object, but there is little research as to which features are in fact most useful for searching (McLane, 1996; Downie, 2003; Goodrum, 2003). Among the several conditions required for a feature to be deemed useful for searching, the first and foremost condition is that users have to be able to make use of that

feature, that is, they need to be able to provide some accurate and helpful information about that particular feature. Studying real-life examples of music information searches is essential for improving our understanding of this issue.

The lack of empirical data regarding the various aspects of real-life queries is also an obstacle in designing and evaluating usable MIR systems. This is especially so because the common assumptions of MIR researchers regarding the nature of music queries are found to be remarkably different from the real-world situation (Futrelle & Downie, 2003). For the evaluation tasks of various MIR systems and techniques (e.g., Music Information Retrieval Evaluation eXchange (MIREX) (Downie, 2008), a set of test queries must be selected for each task. Although some degree of arbitrariness is inevitable in selecting these test queries, the method of selection can certainly be improved. One way is to make them more realistic by using knowledge of what kinds of information we can reasonably assume users will have in hand when they are conducting particular searches in MIR. The empirical data that is currently available to us, however, is very limited.

The objective of this study is to improve our understanding of the common patterns and parameters in real-life music information seeking. More specifically, this study aims to further our understanding of the kinds of information features provided by users in real-life music information queries. Music information queries here mean the natural language query statements of the users searching for music objects or information about those objects (i.e., metadata), not bounded by the limited set of features provided in currently available catalogs or MIR systems. In their queries, people generally provide information they believe will be useful for the search, not necessarily all they can possibly say about the sought object. Thus, analyzing queries will also help understand how users conceptualize the search by learning which information users think is relevant for each search task. This is part of a broader agenda which is best characterized by Downie's assertion (2003) that discovering which facets of music information are essential, potentially useful, and superfluous to the construction of robust MIR systems is one of the central questions confronting future MIR research. This study provides much needed information to the MIR community which, in addition to improving our general understanding of real-life queries and helping us identify potential new access points, can also provide an empirical basis for MIR system design and evaluation.

Review of Related Works

User Studies in Music Information Retrieval

Despite the increasing interests in user studies in the MIR domain, there exist only a limited number of prior studies of MIR system users' needs and behaviors. Nonetheless, these studies provided invaluable preliminary information on the needs, uses, and information behaviors of the MIR system users, and also helped raise general awareness of the importance of

understanding users among MIR researchers. Studies that are based on transaction log analysis (Itoh, 2000; McPherson & Bainbridge, 2001) and semi-structured interviews (Cunningham et al., 2003; Taheri-Panah & MacFarlane, 2004; Laplante & Downie, 2006) are more commonly found, although some studies employ methods such as ethnographic observations (Cunningham et al., 2003; Cunningham et al., 2004), diary study (Cunningham et al., 2007) or user experiments (Kim & Belkin, 2002). The studies based on surveys or interviews (e.g., Downie, 1994; Lee & Downie, 2004; Taheri-Panah & MacFarlane, 2004; Inskip et al., 2008) generally focus on collecting quantitative information about what users *say* they did or might do in certain MIR scenarios. This study shifts the focus to what users *actually* do in real-life music information seeking and retrieval tasks. This study also intends to be distinct from MIR system usability studies focusing on user behavior exhibited in specific MIR systems and analyzing transaction logs. While these kinds of studies can provide information on what actions people take in an existing system, they do not inform us about their motivations, degree of success or failure of the search, or their general search strategies that are not bound to a given set of usable features (Cunningham, 2002).

Analyzing natural language music queries was suggested as an alternative method and a few studies employing this method published some preliminary findings (Downie & Cunningham, 2002; Bainbridge et al., 2003; Lee et al., 2005a). Instead of the search statements entered in a specific MIR system, these studies analyze users' natural language queries in which users can provide any kind of information which may or may not be usable in existing MIR systems. These studies provided preliminary information regarding the basic types of information features identified in queries, some descriptions of those features, and the categories of music information needs along with some quantitative data.

Downie and Cunningham (2002) analyzed 161 music-related information requests posted to the rec.music.country.old-time newsgroup, and categorized the types of information used to characterize the user's information need, types of music information requested, intended uses for the information, and additional social and contextual elements present in the requests. Although the queries analyzed in this study were limited to a single genre, it clearly demonstrated how this kind of study can inform the development of effective and usable MIR system interfaces, and indicate the types of document representations required to support specific user needs.

Following this study, Bainbridge et al. (2003) analyzed 502 music queries posted to the Google Answers website to learn how users of MIR systems express their needs in real-life situations. Bainbridge et al. analyzed a considerably larger number of queries, and the scope of research questions was broadened as well. Authors presented 10 main categories of the need description types, and the bibliographic metadata category was further divided into 10 subcategories. They found that users

experienced difficulty in coming up with clear descriptions for several of the categories such as date or genre, indicating a need to support 'fuzzy' metadata values and query-by-example (more things like this) features.

Building upon this study, Lee et al. (2005a) did a comparative analysis of 107 music queries from the Korean knowledge search portal Naver 지식 iN and 150 music queries from Google Answers to explore the challenges in cross-cultural/multilingual music information seeking and retrieval. They found that users experienced difficulty in precisely describing bibliographic metadata, genre, and lyrics, and relied on contextual metadata such as the information about the use of the sought music in other cultural objects (e.g., movie, TV commercials) or association-based concepts (e.g., "Give me some music similar to this particular song(s) or artist(s)"). A main interest of this study is to explore what alternative access points are used by the inquirer in situations like this when standard MIR access points such as title and creator are not available or deemed unreliable.

Downie and Cunningham (2002), and Lee et al. (2005a) both report the needs identified from the analyzed music information queries. Downie and Cunningham (2002) described the needs in two dimensions – desired information and intended uses, whereas Lee et al. (2005a) listed the types of needs by the users' underlying objectives (e.g., identify work/artist, get recommendations). A few features were used in all of the studies and labeled the same, but for most features, the terms or phrases used to describe them vary across the studies. Even when the feature with the same label is used in multiple studies, we cannot be certain if they are actually referring to the same kind of information due to the lack of definition and/or description of the feature in each study.

In particular, these studies contributed in identifying the basic types of real-life music information needs and the types of information features that people use when seeking music information. Nevertheless, our current understanding of the real-life music information queries is still astonishingly poor. This study attempts to address some limitations of the prior studies, namely the deficiency of formal definitions of the categories of needs and the features for representing MIR queries and limited information regarding the use patterns of the features.

The deficiency of a formal taxonomy of user tasks and queries in MIR is perceived as a major barrier in appropriate evaluation design (Goodrum, 2003; Downie, 2004). In the prior studies of natural language music information queries (Downie & Cunningham, 2002; Bainbridge et al., 2003; Lee et al., 2005a), authors already started to identify the general types of needs expressed and the types of information elements provided by users. However, proper definitions and detailed explications of each feature (referred to as "category" in some studies) were not sufficiently provided in these studies. We still do not have a taxonomy of standardized terms and definitions for each feature, meaning we will not be able to reliably

reproduce the results or easily compare results from different studies. This taxonomy is necessary as part of the formal model of users' music information seeking behaviors and without this kind of model, experiments and evaluation of information retrieval (IR) techniques will not yield valid results, nor contribute to the development of IR theory (Ingwersen, 1992).

Building on the results from prior studies, especially on the category of needs description types in Bainbridge et al. (2003), this research aims to contribute to building a taxonomy to represent the information needs expressed and features used in users' music information queries. Existing taxonomies of reference questions and searches are reviewed in order to gain insights into how to establish this taxonomy. Some definitions of musical terms are based on authoritative sources such as Grove Music Online, *The Harvard Dictionary of Music*, *The Penguin Dictionary of Music*, and so on. The definitions of the features that are not found in dictionaries of music are sought from other reference sources, or developed based on the uses in empirical query data. Categories of features are refined so that they are sufficiently specific, mutually exclusive to each other, and exhaustively cover all information features used in the analyzed queries as a whole. These refined categories are the building blocks of the taxonomy.

The categories in the taxonomy resulting from this research are not meant to be complete, but comprehensive. The taxonomy should not be final, but revised and updated by the MIR research community members. The main purpose of building this taxonomy is to establish the starting point for this collaboration by providing a basis for critical examination and discussion of the features in queries. It is also aimed to contribute in generating results from future query analyses that are more easily comparable to each other.

Studies of Real-Life Investigation of Human-mediated Search

Many of the earlier major investigative studies of real-life human-mediated search examined aspects of the reference interview such as communication and interactions between the user and the intermediaries (Ingwersen & Kaae, 1980; Cochrane, 1981; Ingwersen, 1982). More recently, one of the most notable large-scale projects was conducted by Spink et al. (2002a) who investigated the processes of mediated information retrieval searching during human information-seeking processes and published a series of articles characterizing certain aspects of these processes including uncertainty (Wilson et al., 2002); successive searching (Spink et al., 2002b); cognitive styles (Ford et al., 2002); and user-intermediary interaction (Ellis et al., 2002). The research project involved observational, longitudinal data collection based on questionnaires, interviews, recorded search transaction logs and search processes of 198 information seekers participating in a mediated

online search with a professional intermediary using the Dialog Information Service in the United States and United Kingdom (Spink et al., 2002a).

The main differences between prior research of human-mediated search such as the project led by Spink et al. and this study are as follows: First, this study involves an Internet-based reference service rather than reference interviews for using a commercial database such as Dialog, thus dealing with a broader user group seeking music information for a wider range of purposes than for work or research. Second, the search type that researchers mainly dealt with in prior studies was "subject search" due to the nature of the information service based on a commercial online database. Previous studies of music reference questions (Christensen et al., 2001), multimedia retrieval (Hertzum, 2003), and query analysis (Downie & Cunningham, 2002; Bainbridge et al., 2003; Lee et al., 2005a) suggest the strong presence of "known-item search" as well as "subject search," thus a broader range of searches will be included in this research. Also collecting query documents from the Web has certain advantages over observing the off-line or mixed on/off-line music information seeking and retrieval situations. In an off-line situation, non-verbal cues may be involved in the information seeking and retrieval process (e.g., inquirer's look, hesitation, confidence) that can be hard for the researcher to pick up and thoroughly record. However in an online situation, the query documents collected from the website contain all the information (including the answer and comments when available) that is communicated between the user and the mediated searcher, providing rich context information of the information seeking and retrieval processes.

Additionally, Spink and Saracevic (1997) discussed the effectiveness of search terms used during mediated real-life online searching and found that search terms from users' written question statements (equivalent to "query" in this study) and term relevance feedback were the most productive sources of terms contributing to the retrieval of items judged relevant by users. One of their findings was that less than one-twentieth of all search terms retrieved relevant answers only, and more than one-third of all search terms produced nothing but non-relevant answers or had no retrieval at all. The finding suggests that certain information features may be relatively more important in retrieving the sought information than others. This study is an attempt to learn what those features would be for various music information seeking tasks.

Method

Queries are an essential component of the IR process and also an excellent source for collecting information to identify and evaluate the kinds of information features that are relevant for various user tasks (Downie & Cunningham, 2002; Cunningham, 2003), especially for searches in which conventional access points are unusable (e.g., searching by lyric or

melody). In order to arrive at a definitive understanding of IR processes, many IR researchers believe that empirical studies of *real users* performing *real tasks* in *real environments* are necessary (Downie & Cunningham, 2002; Petrelli et al., 2002; Goodrum, 2003). Following this logic, the queries asked by real users based on their real needs in an operational system from a natural setting are collected for this study.

Content analysis was employed to systematically collect and organize queries into a standardized format that allows one to make inferences about the characteristics and meaning of recorded material (Krippendorff, 2004). The main advantages of content analysis are that the results can be expressed in quantitative terms as well as qualitative judgments and interpretation (Case, 2002) and it provides a measure for checking the intercoder reliability.

Data Collection

Since the early 1990s, expert question-answers services such as Google Answers, Yahoo! Answers, ChaCha, and Mahalo began to appear on the Internet and have been extremely popular. The particular website selected for the source of query documents was the Google Answers website, an online reference service provided by Google. The rationale for selecting Google Answers is that the amount of information the users provide in their queries and the quality of the replies are impressive because it is a fee-based service and queries are not completely limited to easy-to-answer questions (Katz, 2002/2003). On Google Answers, the user pays a non-refundable listing fee of $0.50 per question plus an additional price he sets for the question reflecting how much he is willing to pay for an answer. The price range is set from $2 up to $200 (Google, 2003). The questions are then answered by the Google Answers Researchers who are search experts hired by Google. Queries from Google Answers tend to be rather difficult which makes it more appropriate for this particular study, as the author is interested in discovering features that can be useful for searching beyond the common bibliographic access points such as title or artist name.

Upon receiving the approval from the Institutional Review Board (IRB) at the University of Illinois, 2208 queries were collected from the Google Answers' music category in 2005. These were all the queries posted under the music category on the site as of April 27, 2005. A total of 3318 queries were posted under the same category before Google discontinued this service in December 2006. A query document consists of the following three main components: 1) Question (referred to as "query" in this research) where the user describes their information needs and provides information features as search clues to the intermediaries, 2) Answer where the Google Researcher provides an answer to the query, and 3) Comments that can be

posted by any user who viewed the web document on the Google Answers website. The question (query) part is what is encoded, although the whole document including the answer and comments were reviewed during the coding process.

Since the main interest of this study is to identify and study the information features found in the queries comprehensively, the diversity of features is more important than their representativeness of the data in hand. Rather than adopting a particular sampling measure to analyze part of our data set and make inferences about the whole query data collection, all the queries available were analyzed. The queries that are of interest to this study are those in which the user is searching for music objects or information about those objects.

Data Coding

A total of 1,705 out of 2,208 queries were coded by the author from February to November 2007. 503 queries (22.8%) were discarded as they were deemed to be off-topic or unfitting (e.g., queries asking about legal issues, business aspects, tools/equipments related to music). This number is comparable to what was originally estimated (approximately 20% discard rate) based on the previous study of Google Answers queries (Bainbridge et al., 2003). The data coding process involved identifying the needs expressed in queries, and marking up all the instances of information features in the query text with proper tags. An example of a coded query is presented below.

Original query: I heard this song by a female singer in an ARBY's. I believe it is from the 70s or early 80s. The main chorus of the song says, "over and over again." Kind of a sad, slow, easy listening love song.

Coded query: I heard this song by **<number>**a **<gender>**female**</gender>** **<role>**singer**</role></number>** **<placeref association="contact">**in an ARBY's**</placeref>**. I believe it is from **<date association="music">**the 70s or early 80s**</date>**. **<lyricdesc>**The main chorus of the song says**</lyricdesc>**, **<lyric>**"over and over again."**</lyric>** Kind of a **<affect>** sad**</affect>**, **<tempo>**slow**</tempo>**, **<genre>**easy listening**</genre>** **<about association="music">**love song**</about>**.

The categories of needs and features from the previous MIR user studies (Downie & Cunningham, 2002; Bainbridge et al., 2003; Lee et al., 2005a) were used to establish the initial set of features (categories) as the pre-coding scheme. These features were regarded as tentative and subject to revision based on the further analysis of queries. The revision included adding new categories, deleting categories, re-labeling, and refining (sub-dividing) categories. The main reason for starting with the categories from previous studies rather than developing them from scratch was to maintain some comparability of the features with the ones used in the previous studies. By an iterative coding process, the features in the taxonomies were modified several times and refined to a sufficient level so that they are exclusive, unambiguous, and comprehensive when

taken together for expressing the information provided in the music information queries. The whole process of category development can be found in Lee (2008).

Results

Information Needs

Table 1 and Table 2 present the final sets of 26 needs by their FORM and TOPIC accordingly, including 10 different types by their FORM, and 16 by their TOPIC. Information needs are categorized in two different dimensions, FORM and TOPIC allowing more than one way to compare the usage of features across different types of needs. The FORMs of needs refer to different acts of providing information to the user, regardless of the TOPICs. The TOPICs are the things that users are seeking, in other words, the different kinds of information that is required to answer the user's question. Downie and Cunningham (2002) also had two dimensions of needs in their study, desired information and intended uses, similar to TOPIC and FORM respectively. The categories in previous taxonomies are somewhat similar to the FORMs of needs. In studies of taxonomies such as Pomerantz (2005) and in the FRBR (Functional Requirements for Bibliographic Records) report by the IFLA Study Group (1998), the forms of expected answers or the main four "user tasks" are employed to categorize the different needs. The TOPICs are added as another dimension in this study to observe the variations in the frequency of occurrences among different topics, and the co-occurrence patterns of the FORMs and TOPICs of needs.

Information Features

Table 3 and Table 4 present the final set of features resulting from the iterative coding process. The final set consisted of 102 features (including feature-attribute pairs). There were 50 features without attributes, and 7 features with a total of 52 attribute values. There were two major changes in the structure of the whole scheme. First, the features are grouped into ten main classes: OBJECT:Music, OBJECT:Recording, OBJECT:Score, OBJECT:Related work, ARTIST, SUBJECT, CIRCUMSTANCE, RESPONSE, USER, and OTHER[1] [Table 4]. Second, attribute values were added to seven features, namely PERSONNAME, CORPORATENAME, TITLE, DATE, PLACEREF, LINK, and ABOUT. The features with attributes were separated from the rest of the features and were organized in a different table [Table 3]. The main reason for this division was to simplify the structure as the number of categories of features grew significantly larger than the initial set.

[1] The class OBJECT:Printed material was originally included as one of the 11 classes, but it was omitted from the table and the instruction prepared for coders because the features that would fall under this class all had attributes (e.g., PERSONNAME, CORPORATENAME, TITLE) and ended up being listed in a separate table.

If the features were kept in a simple list as they were before, there would have been more than a hundred different features and coders would have to go over several pages to find the exact feature they need to use. Some of these features were essentially the same in their forms (how they appear in the queries), but were different in what they were referring to. For example, PERSONNAME "John Smith" can sometimes be referring to an artist, or an author of a book related to music, or an actor who played a role in the movie featuring a particular musical work. If they were listed separately as three different features ARTIST-NAME, AUTHOR-NAME, and OTHER-PERSONNAME, coders then would have to look up three different locations under different classes in the whole scheme in order to be sure of which feature to use to code this information. Using the revised scheme, however, coders can find all the attribute values under a single feature PERSONNAME, making it easier to select the appropriate attribute value to pair with the feature. All the categories of features are also accompanied with detailed definitions and examples for an easier understanding of the features. Many of the features, particularly the ones related to the class OBJECT:Music and the class OBJECT:Recording, were adopted from the definitions of attributes in the FRBR (IFLA Study Group on the FRBR, 1998) report.

Distribution of Needs and Features across All Queries

Each of the 1705 queries was assigned with at least one type of need by TOPIC and one type of need by FORM. More than one need could be assigned when appropriate. Table 5 shows the distribution of needs across all analyzed queries. The overall distribution of needs shows that most of these queries were identifying or locating work(s)/recording(s)/artist(s). The most dominant needs by FORM were IDENTIFICATION (43.8%) and LOCATION (33.3%), and by TOPIC, WORK (49.1%), ARTIST (36.4%), and RECORDING (16.7%). REPRODUCTION follows as the next most commonly expressed need by FORM (10.9%) which often appeared together with the TOPIC-LYRICS (10.4%). The connection between REPRODUCTION and LYRICS is discussed in later sections. Other types of needs appeared much less frequently compared to the needs just identified. Table 6 and Table 7 present the frequencies of occurrences of all features across all analyzed queries showing the counts of all queries that contained each feature. Table 6 lists all the features and Table 7 shows the breakdown of the distribution of feature-attribute value pairs for the features with attributes. The distribution of features shows that users rely heavily on a relatively small number of features; only 7 of all features were used in more than 25% of all query data, and 15 were used in more than 10%. This is most likely related to the highly skewed distribution of the needs. Since most queries were seeking help in identifying/locating musical works/artists/recordings and obtaining lyrics, it seems natural that the distribution of the features used is also highly skewed. The traditional bibliographic access points such as

PERSON NAME (53.0% of all queries) and TITLE (50.9%) had strong presence overall. Various aspects of artist(s) such as ROLE, GENDER, and OTHER ARTIST DESCRIPTION also were frequently used. The majority of names were artist names (95.6% of all PERSONNAME instances) and most of the TITLE instances were those of a musical work (68.9%), a related work (24.9%), and an album recording (22.0%). The most prominent relationship among people was based on their similarity (7.2% of all PERSON NAME instances) and among works was based on which works "used" musical works (21.5% of all works). Dates related to music were most commonly provided (36.2% of all DATE instances), but dates associated with related works (26.2%) or recordings (18.9%) were also often used. As for the PLACE REFERENCE, most were describing the place where the user encountered the music (46.7% of all PLACEREF instances) rather than the places related to artists (20.6%) or music (15.6%). LINK instances were mostly directing to audio/video examples of music (42.8% of all LINK instances) and aboutness was mostly of music (77.8% of all ABOUT instances), but sometimes of related works as well (24.8%). Further discussion about needs and features is provided in the next section.

Discussion on Needs and Features

Types of Queries Based on Expressed Needs

Most of the Google Answers queries turned out to be queries seeking help in identifying (43.8%) musical works (49.1%) and artists (36.4%), and locating (33.3%) music recordings (16.7%). Various taxonomies of questions and types of searches exist in the LIS literature; nonetheless it is still difficult to find a specific type of question or type of search among the categories of those taxonomies that would adequately describe these kinds of music queries.

The majority of Google Answers queries could be broadly described as "known-item searches." In the sense that users are seeking identification and/or location information about specific information objects, there is little doubt that these queries are, in nature, closer to "known-item searches (KIS)" rather than "subject searches." However, one may still question if those queries are truly KIS or not due to the fact that they do not conform to some of the key assumptions that LIS researchers have about the concept of known-item search. Lee et al. (2006) surveyed various conceptual and operational characterizations of known-item search in LIS literature and investigated what these definitions imply or infer about knowledge and knowing, and how they pertain to the nature of the item being sought. The authors evaluate each definition/characterization of KIS by asking the following questions to understand the necessary part of the definition:

1. Is a special epistemic relationship required (between the seeker and the sought item)?
2. Is knowledge beyond a simple knowledge of existence required?

3. Is knowledge required at all?
    a. Is the existence required? / b. Is belief required?
4. What exactly is being sought?

The authors start with a common characterization of a KIS as "a search for a particular item that the user knows from previous contact" and revise this characterization after investigating the assumption entailed in each of the questions raised above. Through various counter examples, they show that all of these assumptions can be challenged. As a result, among the features that are ordinarily associated with a KIS, the only aspect that could be confidently said to be essential to the concept was that the search is concerned with a particular thing.

Many of the music queries on Google Answers (especially the ones containing incorrect information about the sought object) can in fact work as a counter example to the second requirement of knowledge beyond a simple knowledge of existence. If someone were looking for music he/she previously heard, but *all* of the information he/she thinks is relevant to finding the item and is attempting to use in the search is incorrect, the search does seem to be a known-item search yet it is difficult to say the user really knows the object beyond its existence. Even when the user is doubtful about the existence of the song (e.g., "Does such a song exist?? Can anybody remember it or is it a figment of my imagination perhaps?"), the search itself is not at all different from when the user believes that the song exists.

One aspect that is unique to these music queries is that they are often searches for information about the object (i.e., metadata) rather than the object itself and for that reason, they may also serve as a counterexample to the assumption that the 'item' is the object being sought. Without a doubt, many users will use the metadata to find music recordings themselves from various sources, but at least in some cases, the users simply wanted to know the title and/or artist of the music they have heard and remembered for many years. In such cases, the actual "seeking" (i.e., obtaining) of the object might never occur and the search will end when the correct metadata is acquired. In such cases, what the users are seeking is not necessarily an item, but information about that item (or manifestation, expression, work) yet the nature of the search and the search process closely resembles other known-item searches.

Interesting Types of Searches

Close examination of the Google Answers queries revealed some interesting types of searches that may be more common in the particular context of MIR. Examples include "dormant" searches, long-term searches, and searching based on information from other people. Dormant searches refer to relatively inactive searches which may become active again,

typically triggered by a certain event. For example, a user may first hear some music and want to know what it is (they may even go on to attempt to search for it unsuccessfully); they give up due to lack of information about the music, or in time simply forget about the search. However, when the user happens to hear the music again, perhaps from some other source such as radio, television, or a movie, it reminds them about their previous music information need and reactivates their search for the music. Dormant searches are different from latent searches (i.e., searches based on latent user needs). A latent information need refers to a need that users are unaware of, but will benefit from once it is fulfilled (i.e., need that is recognized after the fact) (Bodoff, 2006; Kim & Jarvenpa, 2008). For dormant searches, however, the user already recognizes the need; the problem is that the need cannot be fulfilled until more clues are obtained and the user is reminded of the search. An example of dormant searches is provided below:

[Q.] I've heard the spooky tune, The Death March, several times tonight for Halloween. There are not words, just music. I've also heard the tune used in B-rated movies or cartoons to signify that someone or something has died. What is the origin of this tune? Who wrote it, when, and for what reason?

Another interesting aspect about music searches on Google Answers is that some of them have been going on for a long period of time, up to several years (e.g., "it has been on my mind for ten or so years and I'd really like to put it to rest") or sometimes even for most of the user's lifetime if the sought music is from the user's childhood ("The vague memory of one particular song on the tape has remained with me for the past 30 years"). Then it is not surprising that the information users have about the sought music is often vague, incomplete, and incorrect. What is surprising is that the users seem to remember enough about the music, even after many years, to be able to determine if the suggested answer is correct or not, and in most cases, they seem fairly certain about their decision to accept or reject the answer. However, it does occasionally happen that several Google Answers researchers are certain that the answer suggested is correct yet the user keeps rejecting the answer. These problems are often due to the fact that there are many different versions of the same song and when the user is searching for a song based on information about one particular version, the information may not be true for other versions of the same song. Even when the correct song is found, if the version found is different from what the user expected, the user may listen to the song and deny that it is the right song (especially if the version is in a different genre, tempo, or vocal).

Many users also seem to search for music based on information from other people, especially when they are searching for lullabies or folk songs. For example, they heard a song repeatedly from a family member when they were younger and later they try to search for the right song often based on the fuzzy memory they have from their childhood (e.g., "My grandfather, who was born in 1899, used to sing me to sleep with this song and I can't remember the words"). This type of search can turn

out to be quite difficult because often the information that was relayed to them in the first place might not be precise or accurate. The difficulty is amplified as there are frequently many different versions of lullabies and folk songs available. Other interesting cases include queries where the users have somehow obtained the music without any metadata and are trying to identify what music it is that they have (e.g., "I've somehow acquired a song which I have no information about").

Comparison of Features with Previous Studies

The kinds of features that Google Answers users provided in their queries turned out to be numerous, although several of them seemed to emerge as the key features. The features PERSONNAME (53.0%), TITLE (50.9%), and GENRE (37.2%) from the bibliographic facet (as referred to by Downie (2003)) were among the ones that were most commonly used. DATE (45.6%), ROLE (33.8%), LYRIC (27.6%), and PLACEREF (25.6%) were also used in more than one-fourth of the queries. One of the main interests for conducting this research was to discover new features in addition to the conventional attributes provided in typical bibliographic records, thus we will compare the features found in music queries to the conventional attributes in the FRBR model. The findings of this study may help extend the FRBR model to include the kinds of attributes and relationships commonly found in music information queries in addition to the conventional attributes found in typical bibliographic records already represented in FRBR. Table 8 shows which features were new and unique to this particular study compared to previous studies on music information queries, as well as the attributes of the Group 1 entities in the FRBR report. When we compare the scheme derived from this study to the attributes of FRBR Group 1 entities, we can see that the classes of features that are unique to this scheme are OBJECT:Related work, RESPONSE, and USER. Categories under CIRCUMSTANCE are also unique to the studies of music queries, and do not appear in the FRBR attribute set. Overall, it seems that features related to how a user encountered the music, especially through which related work they heard it, how they responded to and felt towards the music and information about the users themselves and their previous search attempts, are what is lacking in the traditional descriptions of bibliographic entities as represented by FRBR.

One notable aspect of these features is that they are not musical features (i.e., features derived from music itself), but rather they are extra-musical features (Downie, 2000) describing relationships among multimedia works as well as the user context. The analysis showed that information about related works had slightly higher positive effects on queries being answered than information surrounding the circumstance of the user's contact of music. This may be due to the limitation of the currently available search systems. There are better sources of information that can be exploited for information about related works, such as IMDB, official websites of movies and TV shows listing the music tracks used, music rights clearance

databases, and various websites relying on user-generated content such as Wikipedia, and YouTube. Conversely, information about the circumstances with regards to where, when, and how someone heard the music may be helpful in narrowing the potential song candidates or verifying what is suspected to be the right answer, but tools for searching music databases using such information seem to be close to non-existent. This is likely because it is extremely difficult to develop and maintain such databases or systems given the diversity of sources which would have to be monitored and mined, and the fact that almost all the information will be relative to a particular user's experience.

Note that in the FRBR model, the user is situated outside of the bibliographic universe as an entity who is navigating the bibliographic universe, not as a type of *Person* in Group II entities; however in our taxonomy, the user and the object are situated much more closely. The information derived from the user's encounter with the music such as the information about the circumstance surrounding when the music was heard (e.g., particular event, date, media, place) and the reactions (e.g., mood, similarity, meaning) are important features. Whereas the FRBR model is largely based on bibliographic "facts," in queries, the users' "beliefs" and assertions based on those beliefs play important roles. The bibliographic information also tends to be more stable and static compared to the users' information that can be vague, incomplete, and dynamic.

The major difference between FRBR research and this study was that the attributes and relationships were derived from empirical music information seeking and retrieval data rather than bibliographic records. One limitation of the FRBR model is that although FRBR provides a comprehensive set of attributes, there are still certain types of attributes missing in the model. For instance, the kinds of attributes that are derived from the context of using music in other cultural objects and/or events (e.g., used in a TV show, commercials, place or event where music was heard) are not included in FRBR. Although the context of the entity is already listed as one of the external attributes in FRBR, the context information with which FRBR is concerned is mostly from the context of creation, not the context of use, as we can infer from the descriptions of the context attribute from FRBR at the work level and expression level.

The attributes that relate to the user's reaction to an entity (e.g., mood) are also absent and the list of relationships defined in FRBR may also need to be extended. For instance, the similarity-based relationship (similar artists/works) which seems important in the context of MIR is not represented in the FRBR model. The kinds of information features that people use as clues to find music objects may be derived from sources other than the music object itself, which would imply that even very comprehensive bibliographic information may not be enough.

Recommendations for Music Information Retrieval Evaluation eXchange (MIREX)

The main motivation for this study was to improve our understanding of real-life music queries in order to develop a better set of tasks and queries for evaluating various MIR systems and techniques. To understand the progress made so far in MIREX, it is necessary to review the tasks that were carried out in the last three years. Table 9 summarizes all the MIREX tasks run from 2005 to 2007.

Downie (2008) differentiates the tasks representing micro-level MIR research from macro-level research. Tasks such as "Audio Onset Detection" represents micro-level MIR research as they tend to be very narrow and technical tasks (i.e., onset detection seeks to accurately locate music events in audio files), and the results are often used as part of other more complex tasks. Tasks such as "Symbolic Melodic Similarity," however, represent macro-level research as they are concerned with higher level retrieval tasks based on many different aspects of music.

Another way to understand this notion is to consider how close the tasks resemble the real-life user tasks. For common users, tasks such as "Audio Onset Detection" or "Multiple Fundamental Frequency Estimation & Tracking" would not be very relevant as they are the kinds of tasks that are more useful for MIR system developers. Common music users would probably relate more to tasks such as "Query-by-Singing or Humming" or "Audio Cover Song Identification" which are more similar to the kinds of needs that would arise in their everyday lives and would directly provide results that they would be interested in. Looking at the past MIREX tasks and how they evolved over the years, it seems like the later tasks tend to be more concerned with the needs of real-life users, although there are some exceptions (such as "Multi F0 Estimation"). The tasks also seem to reflect a growing interest in music similarity and mood in recent years in the MIR community. Based on the findings of this research, a few recommendations can be made regarding the directions for future MIREX tasks.

Incorporate more user context in test queries

The first recommendation is to consider the user context more for tasks such as music similarity evaluation. There has been a great deal of interest in music similarity among MIR researchers recently. Evaluation tasks based on audio and symbolic music similarity have been run in the past MIREX, thus the discussion among MIR researchers surrounding what is really meant by "music similarity" has also been a topic of debate for several years now. Music similarity has been typically associated with playlist generation or music recommendation, but the analysis of Google Answers data shows that it is also used for seeking a particular music object. The SIMILAR attribute was used in conjunction with ARTIST and TITLE features to code references to known artists or works used to describe attributes of the sought music. In the analyzed queries,

65 queries (7.2%) contained references to similar artist(s) whereas only 16 queries (1.8%) referenced similar musical work(s). Although they are not labeled as "similar," the attribute value "example" associated with features TITLE (2.9%) and PERSONNAME (2.5%) were also based on the similarity among musical works and artists as they exemplified certain traits in a similar group of music/artists.

MIREX started incorporating music similarity evaluation contingent upon subjective human evaluation in 2006 (Downie, 2008). The instructions given to human evaluators explained the tasks as follows (Downie et al., 2007; Jones et al., 2007):

> *Evaluate how well various algorithms retrieve results that are…*
>
> *Symbolic Music Similarity (SMS): MELODICALLY similar to a given query. You will find in the candidate files a variety of different instrumentations as set by the creators of the MIDI files. We need you to look beyond the differences in timbre and instrumentation in assigning your grading scores.*
>
> *Audio Music Similarity (AMS): MUSICALLY similar to a given query. You will be presented with files from a number of different music genres. Please assign the scores according to what you find "sounds" similar and do not take into account whether you like the music or not.*

The practical objective of determining the similarity between two musical works has not been clear to the evaluators or even to researchers. The past evaluations have focused on determining the similarity between two music files after listening to them, and users were not provided with any metadata such as artist names or genre labels when they were evaluating these music files. However in Google Answers data, similarity among artists was used approximately four times more than similarity among musical works. This may be due to the fact that most of these queries were for searching music objects and information rather than generating playlists or browsing a music collection. Nevertheless, it certainly demonstrates that the notion of music similarity in users' minds is a lot more complex than just two musical works sounding similar. The similarity between two artists can be based on many different reasons, independent of the features that can be directly extracted from the music itself (e.g., looks, influences, perspectives, similar/same appearances and origins in time and space).

One of the most difficult issues here is how to put this notion of similarity into evaluation. MIREX organizers found that narrowing the concept of similarity by specifying which aspect human evaluators are supposed to focus on produced better results (Downie et al., 2007; Jones et al., 2007). By analyzing the 7,602 similarity judgments collected for SMS and AMS evaluation tasks for MIREX 2006, Downie et al. (2007) found that compared to the AMS evaluators, the SMS evaluators seemed to have "a more concrete understanding of melodic similarity than musical similarity" as they more often gave the

perfect similarity scores, auditioned the music files fewer times and also modified their score less often than AMS evaluators. For obtaining more reliable evaluation data, the notion of similarity needs to be narrowed down, but the similarity notion that is more relevant to users in their casual searches for music objects seems to be based on a complex mix of the various aspects of the artists. One suggestion for resolving this issue would be to specify the objective of the task more clearly, in other words, what the similarity evaluation result will be used for.

A great deal of effort has been made regarding how to evaluate the music similarity of two music files in the past "Audio Music Similarity and Retrieval" task. The concept of music similarity has been extremely difficult to deal with, because there are so many different aspects one can consider when we say a musical work or an artist is similar to another work/artist. As an example, take a look at the following four cases of real-life query examples:

[Q1] It has a female voice similar to Lora Logic or Linder Sterling from Ludus(I always thought this was a song by Ludus but now I have their complete discography and it isn't there).

[Q2] Sounds like the Temptations or some band from the 1960s.

[Q3] The whole thing sounds a little John Mayersque. In that genre, with Jack Johnson, Jason Mraz, etc.

[Q4] They were the sort of hyper skinny 26 year old white guys you think of when you think of Red Hot chili peppers… sort of a repetitive pop song in the spirit of, say, FatBoy Slims "wonderful night"…

In all four cases, the users provided information about artists/works similar to the ones they are seeking. However, what exactly was similar between the mentioned artists/works and the sought artists/works are different for each case. In the first case, we can see that the similarity between the artists originates from the voice of the artist, therefore the similarity in vocals would be more important than similarity in instrumentation, genre, etc. Compare it to the second case where the similar artist was mentioned as a representative band from the 1960's. In this case, the overall characteristics of the sounds of 60's bands would be the most important thing to consider. In the third case, the similar artists were mentioned to refer to "that genre" which can be described by several exemplary artists. In the fourth case, two different bands were mentioned: the first was mentioned to explain the physical resemblance of the sought artist with the referenced one, and the second was mentioned to explain the similarity in musical style between the two bands. These cases show that our natural perception of music similarity in everyday life is much more complex to be evaluated on a single specific aspect.

Another possible way to improve the evaluation results would be to incorporate more description of the user's context in the search. Consider the four examples again. Although we can attempt to answer them by using similarity information, the way the user will evaluate the results will be very different for each case. For instance, for the first case, the candidate music

files containing similar vocals would receive a higher score whereas for the second case, the overall timbre or instrumentation might be more important in determining the relevance score. As another example, take a look at the following two queries:

[Q5] I need a list of the 200 or so best known celebration songs, anthems of a sort from 1975 to the present. The type of song is like celebration by Kool and the gang.

[Q6] What is the name of this song and who sang it? …She was more country artist than star, along the lines of Allison [*sic*] Krauss or Emmylou Harris (I thought it might have been Allison Krause in the duet, but if so, can't find it.)[...]

If we were to use some similarity algorithm to generate the results for the first query, they should not include the particular song and the artist mentioned by the user as it would not provide any new additional information. However, for the second query, if the system were to find a song by one of the artists mentioned by the user (i.e., "Allison Krauss"), it would make sense to return that result to the user since it could be the very song that the user is looking for, unlike the first case. These query examples suggest that the user context information not only helps us understand which aspect of music similarity is of interests to the user, but can also help us improve the way the system provides the returned results. This is exactly why it is important to consider user context in better evaluating the results of these tasks.

Employ terms that are familiar to and are more likely to be used by real-life users

The second recommendation is to incorporate more terms and phrases that are more likely to be used by real-life users as descriptors, especially for things such as affect/mood of music or genres. Here, we will discuss Mood Classification Task as an example. Examining the 92 queries which contained AFFECT/MOOD feature, Affect/mood information seems to work as important "hooks" in the minds of the searchers when they are trying to remember and describe the sought music. There were a total of 128 unique mood terms used in these queries. Among these, 25 terms were used more than one time.

In the past "Mood Classification" task, human evaluators were asked to listen to the music clips provided and select the mood cluster that best fits the music in order to generate the ground truth. There were multiple clusters of moods that the evaluators could choose from when categorizing the music clips. For each cluster, several terms were used to refer to a particular mood of the songs as a whole [Table 10]. The task organizers explained that these mood clusters were derived from the AMG (All Music Guide) mood repository (Hu & Downie, 2007). Some of the terms provided for describing these clusters do not seem like terms that common users would use to describe the mood of music (e.g., rollicking, campy). In fact, when the terms in these clusters are compared against the 127 unique instances of affect/mood description given by users in Google Answers queries, only seven of the thirty-one terms overlap with the user's mood labels. The exact numbers of times

they appear in the Google Answers data are as follows: rousing (1), fun (3), sweet (2), humorous (1), silly (1), quirky (2), and aggressive (1). Presumably, the quirkiness of the terms is a result of running a cluster analysis on a large number of mood terms (thus naturally including some uncommon ones) used by AMG. However it is questionable if using these terms is a better option over using the terms that lay people would actually use to describe the sought music. One way to improve the mood classification would be to consider incorporating the terms that are actually used by real-life music users. For instance, would it be possible to substitute some of the terms used for cluster 3 to "sad" or "melancholy"? For cluster 2, could we use terms like "happy" or "playful" instead of "rollicking"? It is possible that these terms may not be the best choice for deriving mood clusters that are most distinct from each other, but they may be terms that users can easily understand and relate to, which should, at least in theory, help the evaluation process. Another way to improve the mood classification would be to consider what kinds of mood clusters would be most meaningful and relevant for the users by examining the mood labels provided in real-life queries and observing their browsing behaviors of music by moods.

Explore ways to combine and evaluate the results from multiple tasks for a test query

The last recommendation is to explore ways to combine and evaluate the results from multiple tasks for users' queries. Suppose a user wants to identify a particular musical work of which he/she knows some information. Although users tend to heavily rely on a small number of features when posing their queries, the analysis also showed that there is a diversity of kinds of features that were used. For instance, the user might be able to hum the melody of the sought music, as well as specifying another song that sounds similar to the one that he/she is seeking. The user might also be able to describe the mood of the song and guess which genre it will fit into. The way that MIREX tasks are organized now, this one search scenario will be partitioned into four different tasks – "Query by Singing/Humming," "Audio Music Similarity and Retrieval," "Music Mood Classification," and "Audio Genre Identification." However, in a real-life situation, the results from all of these tasks must be consolidated to one set of results to be given to the user. Therefore, to make the tasks more realistic and useful for real-life users, it will be necessary to move beyond focusing on evaluating the results for each task and find a way to combine the efforts from multiple tasks.

Conclusion

A pervasive theme emerges from these analyses that user queries are very complex and it is difficult to derive simple user patterns based on needs expressed and features provided by users. To summarize the findings, there were various FORMs and TOPICs of needs that were present in Google Answers data, but the most important needs in the Google Answers were to

identify a musical work and/or an artist, locate a recording, and obtain the lyrics of songs, which can be referred to as known-item searches in a broad sense.

In addition, the study suggests that users seeking music objects and information about these objects on Google Answers make use of a large number of different types of information as they seek music objects and music information, although a few key features are much more heavily used than the rest. Only seven features were used in more than 25% of the queries, and fifteen features were used in more than 10% of the queries. Person name, Title, Date, Genre, Role, Lyric, and Place reference were some of the most heavily used features. Date information was used surprisingly often, even more so than lyric information. The study also reveals that the kinds of information features provided in Google Answers queries tend to be highly personal and user-centered, compared to the traditional descriptions of bibliographic entities.

Some recommendations are made for improving MIR systems and system evaluation, including: (i) incorporating user context in test queries, (ii) employing terms familiar to users in evaluation tasks, and (iii) combining multiple task results. Information about related multimedia works may also help improve the current MIR systems.

Acknowledgements

This research is conducted as part of the HUMIRS (Human Use of Music Information Retrieval Systems) project funded by the Andrew W. Mellon Foundation and the National Science Foundation [Grant No. NSF IIS-0327371]. The author would also like to thank Allen H. Renear, J. Stephen Downie, Linda C. Smith, David Dubin, and Sally Jo Cunningham for their support as doctoral committee members and M. Cameron Jones for his assistance in collecting and processing the data.

References

Bainbridge, D., Cunningham, S. J., & Downie, J. S. (2003). How people describe their music information needs: a grounded theory analysis of music queries. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR),* (pp. 221-222). Baltimore, Maryland, USA: Johns Hopkins University.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*(5), 407-424.

Bodoff, D. (2006). Relevance for browsing, relevance for searching. *Journal of the American Society for Information Science and Technology, 57*(1), 69-86.

Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. *Information Processing and Management, 38*(2), 249-272.

Case, D. O. (2002). *Looking for information: A survey of research on information seeking, needs, and behavior*. Amsterdam ; Boston: Academic Press.

Christensen, B., Du Mont, M., & Green, A. (2001). Taking note: assessing the performance of reference service in academic music libraries. A progress report. *Notes, 58*(1), 39-54.

Cochrane, P. (1981). Study of events and tasks in pre-search interviews before searching. In *National Online Meeting proceedings - 1981* (2nd ed., pp. 133-147). Medford, N.J.: Learned Information.

Cunningham, S. J. (2002). What people do when they look for music: implications for design of a music digital library. In *Proceedings of the 5th International Conference on Asian Digital Libraries: ICADL 2002* (pp. 177-178). London, UK: Springer-Verlag.

Cunningham, S. J. (2003). User studies: a first step in designing an MIR testbed. In J. S. Downie (Ed.), *The MIR/MDL Evaluation Project White Paper Collection* (3rd ed., pp. 19-21). Champaign, Illinois: GSLIS.

Cunningham, S. J., Reeves, N., & Britland, M. (2003). An ethnographic study of music information seeking: implications for the design of a music digital library. In *Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries: JCDL 2003* (pp. 5-16). Houston, Texas: IEEE Computer Society.

Cunningham, S. J., Jones, M., & Jones, S. (2004). Organizing digital music for use: an examination of personal music collections. In *Proceedings of the 5th ISMIR,*(pp. 447-454). Barcelona, Spain: Universitat Pompeu Fabra.

Cunningham, S. J., Bainbridge, D., & Falconer, A. (2006). More of an art than a science: supporting the creation of playlists and mixes. In *Proceedings of the 7th ISMIR,* (pp. 240-245). Victoria, Canada.

Cunningham, S. J., Bainbridge, D. & McKay, D. (2007). Finding new music: a diary study of everyday encounters with novel songs. In *Proceedings of the 8th ISMIR*. (pp. 83-88). Vienna, Austria.

Dervin, B. (1992). From the mind's eye of the user: the sense-making qualitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.

Downie, J. S. (1994). The Musifind musical information retrieval project, phase ii: user assessment survey. In *Proceedings of the 22nd Annual Conference of the Canadian Association for Information Science*, 149-166.

Downie, J. S. (2003). Music information retrieval. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 37, pp. 295-340). Medford, NJ: Information Today.

Downie, J. S. (2004). The creation of music query documents: framework and implications of the HUMIRS Project. In *Proceedings of the Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing: ACH/ALLC 2004*. Goteborg, Sweden.

Downie, J. S., & Cunningham, S. J. (2002). Toward a theory of music information retrieval queries: system design implications. In *Proceedings of the 3rd ISMIR,* (pp. 299-300). Paris, France: IRCAM - Centre Pompidou.

Downie, J. S., Lee, J. H., Gruzd, A. A., & Jones, M. C. (2007). Toward an understanding of similarity judgments for music digital library evaluation. In *Proceedings of the ACM IEEE Joint Conference on Digital Libraries 2007* (pp. 307-308). Vancouver, British Columbia, Canada.

Downie, J. S. (2008). The Music Information Retrieval Evaluation eXchange (2005-2007): a window into music information retrieval research. *Acoustical Science and Technology, 29*(4), 247-255.

Ellis, D., Wilson, T. D., Ford, N., Lam, H. M., Burton, R., & Spink, A. (2002). Information seeking and mediated searching. Part 5. User-intermediary interaction. *Journal of the American Society for Information Science and Technology, 53*(11), 883-893.

Ford, N., Wilson, T. D., Foster, A., Ellis, D., & Spink, A. (2002). Information seeking and mediated searching. Part 4: Cognitive styles in information seeking. *Journal of the American Society for Information Science and Technology, 53*(9), 728-735.

Futrelle, J., & Downie, J. S. (2003). Interdisciplinary research issues in music information retrieval: ISMIR 2000-2002. *Journal of New Music Research, 32*(2), 121-131.

Goodrum, A. A. (2003). If it sounds as good as it looks: lessons learned from video retrieval evaluation. In J. S. Downie (Ed.), *The MIR/MDL Evaluation Project White Paper Collection* (pp. 97-102). Champaign: GSLIS.

Google. (2003). Google Answers: Frequently Asked Questions. Retrieved March 28, 2006, from http://answers.google.com/answers/faq.html

Hertzum, M. (2003). Requests for information from a film archive: a case study of multimedia retrieval. *Journal of Documentation, 59*(2), 168-186.

Hu, X.,& Downie, J. S. (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata. In *Proceedings of the 8th ISMIR,* (pp. 67-72). Vienna, Austria.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records: final report*. München: K.G. Saur.

Ingwersen, P. (1982). Search procedures in the library: analysed from a cognitive point of view. *Journal of Documentation, 38*(3), 165-191.

Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.

Ingwersen, P., & Kaae, S. (1980). *User-librarian negotiations and information search procedures in public libraries: analysis of verbal protocols, final research report*. Copenhagen: Royal School of Librarianship.

Inskip, C., Macfarlane, A., & Rafferty, P. (2008). Music, movies and meaning: communication in film-makers' search for pre-existing music, and the implications for music information retrieval. (pp. 477-482). In *Proceedings of the 9th ISMIR*. Philadelphia, USA.

Itoh, M. (2000). Subject search for music: Quantitative analysis of access point selection. In *Proceedings of the First ISMIR*. Amherst, MA: University of Massachusetts at Amherst.

Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: a study of user queries on the Web. *ACM SIGIR Forum, 32*(1), 3-17.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the Web. *Information Processing and Management, 36*(2), 207-227.

Jones, M. C., Downie, J. S., Ehmann, A. F. (2007). Understanding human judgments of similarity in music information retrieval. In *Proceedings of the 8th ISMIR,* (pp. 539-542). Vienna, Austria.

Katz, B. (2002/2003). Digital reference: an overview. *The Reference Librarian,* (79/80), 1-17.

Kim, J.-Y., & Belkin, N. J. (2002). Categories of music description and search terms and phrases used by non-music experts. In *Proceedings of the 3rd ISMIR,* (pp. 209-214). Paris, France: IRCAM - Centre Pompidou.

Kim, Y., & Jarvenpaa, S. L. (2008). Formal boundary spanning and informal boundary spanning in cross-border knowledge sharing: a case study. In *Proceedings of the 41st Annual Hawaii international Conference on System Sciences* (January 07 - 10, 2008). HICSS. IEEE Computer Society, Washington, DC, 337-346.

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Kuhlthau, C. (1991). Inside the information search process: information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361-371.

Lankes, R. D., Nicholson, S., & Goodrum, A. (2003). *The digital reference research agenda: compiled from the Digital Reference Research Symposium, August 2002, Harvard University*. Chicago: ACRL.

Laplante, A., & Downie, J. S. (2006). Everyday life music information-seeking behaviour of young adults. In *Proceedings of the 7th ISMIR,* (pp. 381-382). Victoria, Canada.

Lee, J. H., & Downie, J. S. (2004). Survey of music information needs, uses, and seeking behaviours: preliminary findings. In *Proceedings of the 5th ISMIR,* (pp. 441-446). Barcelona, Spain: Universitat Pompeu Fabra.

Lee, J. H., Downie, J. S., & Cunningham, S. J. (2005a). Challenges in cross-cultural/multilingual music information seeking. In *Proceedings of the 6th ISMIR,* (pp. 1-7). London, UK.

Lee, J. H., Renear, A., & Smith, L. C. (2006). Known-item searching: variations on a concept. In *Proceedings of the 69th ASIS&T Annual Meeting*. Austin, Texas.

McLane, A. (1996). Music as information. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 37, pp. 225-262). Medford, NJ: Information Today, Inc.

McPherson, J. R., & Bainbridge, D. (2001). Usage of the MELDEX digital music library. In *Proceedings of the 2nd ISMIR,* (pp. 19-20). Bloomington, IN: Indiana University.

Orio, N. (2006). Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval, 1*(1), 1-90.

Petrelli, D., Beaulieu, M., Sanderson, M., & Hansen, P. (2002). User requirement elicitation for crosslanguage information retrieval. *The New Review of Information Behaviour Research: Studies of Information Seeking in Context, 3*.

Pomerantz, J. (2005). A linguistic analysis of question taxonomies. *Journal of the American Society for Information Science and Technology, 56*(7), 715-728.

Savolainen, R. (1995). Everyday life information seeking: approaching information seeking in the context of “Way of Life”. *Library and Information Science Research, 17*, 259-294.

Spink, A., & Saracevic, T. (1997). Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science, 48*(8), 741-761.

Spink, A., Wilson, T. D., Ford, N., Foster, A., & Ellis, D. (2002a). Information seeking and mediated searching. Part 1: Theoretical framework and research design. *Journal of the American Society for Information Science and Technology, 53*(9), 695-703.

Spink, A., Wilson, T. D., Ford, N., Foster, A., & Ellis, D. (2002b). Information seeking and mediated searching. Part 3: successive searching. *Journal of the American Society for Information Science and Technology, 53*(9), 716-727.

Taheri-Panah, S., & MacFarlane, A. (2004). Music Information Retrieval systems: why do individuals use them and what are their needs? In *Proceedings of the 5th ISMIR,* (pp. 455-460). Barcelona, Spain: UPF.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249-270.

Wilson, T. D., Ford, N., Ellis, D., Foster, A., & Spink, A. (2002). Information seeking and mediated searching. Part 2: Uncertainty and its correlates. *Journal of the American Society for Information Science and Technology, 53*(9), 704-715.