

# CHALLENGES IN CROSS-CULTURAL/MULTILINGUAL MUSIC INFORMATION SEEKING

**Jin Ha Lee**

Graduate School of Library  
and Information Science  
University of Illinois at  
Urbana-Champaign  
jinlee1@uiuc.edu

**J. Stephen Downie**

Graduate School of Library  
and Information Science  
University of Illinois at  
Urbana-Champaign  
jdownie@uiuc.edu

**Sally Jo Cunningham**

Dept. of Computer Science  
University of Waikato  
New Zealand  
sallyjo@cs.waikato.ac.nz

## ABSTRACT

Understanding and meeting the needs of a broad range of music users across different cultures and languages are central in designing a global music digital library. This exploratory study examines cross-cultural/multilingual music information seeking behaviors and reveals some important characteristics of these behaviors by analyzing 107 authentic music information queries from a Korean knowledge search portal Naver 지식 (knowledge) iN and 150 queries from Google Answers website. We conclude that new sets of access points must be developed to accommodate music queries that cross cultural or language boundaries.

**Keywords:** User behaviors, cross-cultural music information seeking, multilingual music information seeking, Korean users.

## 1 INTRODUCTION

Futelle and Downie [1,2] criticize the overemphasis on Western music in MIR research and assert the necessity for developing MIR systems that are also suitable for non-Western music. Another aspect of this “multi-cultural challenge” in MIR research concerns the difficulties that non-Western users experience when they seek Western music information. In a broad sense, this problem is applicable to all people who seek music information from cultures or language of which they are not native.

In everyday life, it would be rare for people to seek textual materials in a language that they do not know (unless there is help available for translation); however, it is quite common for people to seek music from different cultures or music in non-native languages. Witness, for example, the emergence of ‘World Music’ as a category in the recording industry and in music stores. While music may be “a universal language” [3], there

still exist important differences in the way people describe and search for music according to their individual cultural and linguistic backgrounds. In short, this multi-cultural aspect of music can make music information seeking a difficult and frustrating task.

Currently, there is little known about cross-cultural information seeking behaviors in the general field of information retrieval (IR), and much less in the MIR domain. Research on Cross Language Information Retrieval (CLIR), the retrieval of information written in one language based on a query expressed in another [4], has primarily been restricted to studies involving precision/recall measurements over artificial queries, rather than examining use of CLIR systems in real-life settings. Strictly speaking, the cases that we discuss in this paper are best considered as examples of *multi-lingual information seeking* rather than cross language information seeking, as the searchers in this study describe their music information needs in multiple languages.

## 2 PROBLEMS AND OUR APPROACH

Historically, music libraries have provided access points for music retrieval via bibliographic information (e.g., composer, title) and/or some type of genre scheme [1]. One limitation of these traditional access points is the assumption of users’ prior musical knowledge [3]. To overcome this limitation, much of current MIR research focuses on supporting music searches by exploiting content-based characteristics of the music itself rather than its metadata (e.g., query-by-humming systems). Another assumption exists that audio MIR systems are more flexible than symbolic ones because the audio representation features are culturally more neutral, but this assumption has not been specifically tested [2].

Notwithstanding the limitations of bibliographic information, bibliographic information still remains the primary communication mechanism between music collections and users. Previous MIR user studies indicate that music queries often included bibliographic information. Of the analyzed queries, 81.3% in [5] and 75.2% in [6] included some sort of bibliographic description. For both studies, the second and the third most commonly provided data in music queries were genre and lyric information. Vignoli [7], and Lee and Downie [8] also found that genre was frequently viewed by users as an important organization and retrieval method.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2005 Queen Mary, University of London

Would these three elements – bibliographic, genre, and lyric information – also be used predominantly in cross-cultural/multilingual music information seeking? What if people cannot remember, pronounce, or spell the name of the artist, song, or album? What if they do not know the ‘proper’ name for a musical genre, or perhaps know that genre label in a language other than English? What if they cannot understand, memorize or replicate the lyrics? The following **translated** example from a Korean music Q&A site helps us illustrate these questions:

*“First of all, it’s very long. Just the first movement seems to go over 10 minutes. The mood is dark and sad overall. The main melody is played by violin and the piano briefly appears. I only know that the composer’s name is four characters when it is spelled out in Korean. Does anybody know the title of this work?”<sup>1</sup>*

In the above example, known difficulties in searching for classical music – those characteristics such as instrumentation and ‘mood’ are rarely searchable metadata – are exacerbated because the searcher is working in a different language, culture, and script than the culture from which the music originated.

This exploratory study examines music information seeking behaviors of those seeking music from different cultures and in non-native languages. It should be noted that the purpose of this study is not to provide generalizable statements for all cross-cultural/ multilingual music information seeking; rather the main contribution of this work is to provide seed data and observations about these kinds of music information seeking behaviors in order to frame future research agendas.

In the following, we present a brief analysis of the types of music information queries and the kinds of information that searchers provided in those queries. Additionally, we present a number of real queries as examples to illustrate the unique difficulties inherent in this type of information seeking. By doing so, we provide empirical evidence on why it appears to be important to consider certain features in developing MIR systems, and some insights on how to better design MIR systems for international users.

All example queries presented in this paper were originally in Korean, were translated by the first author, and are presented in *italics*. Terms in **bold italics** denote that they originally appeared in English. Terms left in the original Korean script are onomatopoeic words or Korean transliterations of English words.

### 3 DATA COLLECTION

Petrelli et al [4] asserted the necessity for “in the field evaluations” in CLIR, and stated that “only with real users performing real tasks and possibly in real environments, we can arrive at a definitive understanding of

---

<sup>1</sup> The answer to this query is: *The Devil’s Trill* by the Italian composer, Tartini, whose name is, in fact, four characters in Korean (타르티니).

CLIR.” The same logic applies for this study: We sought music information requests made by real users based on their real needs in an operational system. Downie and Cunningham [6] suggest Internet newsgroups, websites, and archives of mailing lists as sources of authentic music queries expressed in natural language, unencumbered by query syntax or prescribed metadata. Following this suggestion, we selected Naver 지식 (knowledge) iN<sup>2</sup>, a popular Korean ‘knowledge search’ portal as our data source for collecting cross-cultural/multilingual music queries. ‘Knowledge search’ system in Korea refers to a web-based system where people build up a sort of online knowledge base by asking and answering questions, and sharing their personal knowledge [9]. Naver started its 지식 iN service in 2002 and currently maintains a well-developed and widely used user-contributed Q&A system for collaborative information seeking [10]. Under the music category alone, there were already 306,719 questions posted and answered as of April 17, 2005. The predominately Korean user-base pose queries asking about Western (English lyrics), Japanese, Chinese, Korean music, etc. In order to compare the music information search behaviors of Koreans and Americans, we also selected Google Answers as our second data source to collect Western music queries. Google Answers is a web-based online reference service in a Q&A format which was also used in the previous MIR user study by Bainbridge et al [5].

As an exploratory study, we looked at a limited number of queries to suggest interesting aspects for future research. We analyzed a total of 107 queries posted under the category “Pop” on Naver. Two days were arbitrarily selected – April 1 & 2, 2005 – and all “Pop” queries posted on those dates were collected. Two queries were discarded as they were off-topic. We also analyzed a total of 150 queries from Google Answers – March & April, 2005 – under the music category. Nine were discarded as they were off-topic or redundant queries with the same content posted by the same searcher.

## 4 OBSERVATIONS AND DISCUSSIONS

### 4.1 Comparison: Naver and Google Answers

Searchers frequently provided vague, incorrect, and incomplete information when they described their music information needs in both Naver and Google Answers queries. Surprisingly, many of these queries were answered and the answers were verified to be correct or relevant by the searchers. Naver affords the searcher the ability to signify correct/useful answers, and Google Answers allow searchers to comment on the answers as they rate them.

Some interesting types of queries that caught our attention that were unique to Naver data set include requests for writing the pronunciations of English lyrics in Korean such as:

---

<sup>2</sup> kin.naver.com

“Can someone write the lyrics for ‘Lady Marmalade’ sung by ‘Moulin Rouge’ in Korean? *Gitchi gitchi ya ya da da (hey hey hey)* - 키치 키치 야 야 다 다 (헤이 헤이 헤이) – Like this.”

Not knowing the language or meaning of songs certainly does not stop people from wanting to sing them. The influence of karaoke culture was evident in a number of Naver queries analyzed. Naver searchers often sought suggestions of songs that are easy to sing in their own vocal ranges and singing styles: this behavior was not observed in the Google Answers data set. Asking for appropriate songs for special occasions (e.g., birthday, parties) and for specific person(s) (e.g., significant others, friends, co-workers) were also quite common:

“I have three friends whose birthday is in April and I want to sing a song for them. Can you recommend any good songs related to friendship?”

The types of queries posted on two websites are summarized in Table 1 (Naver) and Table 2 (Google Answers). The queries are categorized by the music information needs expressed in them. The most common music information need in both Naver (48.6%) and Google Answers (32.6%) data sets was “Identify artist/work.” The rest of the information needs were ordered fairly differently for two data sets: “Get recommendations” was the second most common information needs in the Naver data set (36.2%), whereas only 7 queries (5.0%) from Google Answers sought music recommendations. In the Google Answers data set, queries seeking general information related to music (e.g., about an artist, work, instrument) (25.5%) and queries for locating a musical work (17.0%) were the second and third most common music information needs, contrary to the low proportions in the Naver data set.

Several assumptions can be made about the reasons for this difference between two data sets. First, the cultural difference between Koreans and Americans may affect the information seeking and sharing behaviors of the two groups. Korea’s collectivist culture and the United States’ individualist culture are often contrasted in the studies of organizational behaviors. Hofstede and Bond found that individualists rely more on their own experience when making decisions than do collectivists in the organizational setting [11]. If we consider online communities as a type of organization, we may speculate that this cultural difference affects the degree of how much an individual is willing to request and accept music recommendations from other members of the community. Second, the cost of using the service may also affect the kinds of information requests submitted. Naver 지식 iN service is semi-free to users: Naver users earn points by answering other people’s questions, which in turn they can use to ‘buy’ opportunities to ask their own questions. On the other hand, the Google Answers service asks for a minimum charge (\$2.50) per question to its users. Searchers using Google Answers

service may choose other methods without charge for those queries (e.g., getting recommendations) that do not necessarily need an expert’s searching skills. These assumptions should be further tested in future studies employing qualitative methods (e.g., interviewing the searchers using both systems).

Table 1. Naver queries (105): music information needs categories

Music Information Need	Percentage
Identify artist/work	48.6
Get recommendations	36.2
Acquire lyrics	5.7
Request translation	2.9
Locate specific version of work	1.9
Seek information	1.9
Request transliteration	1.9
Locate work	1.0
Total	100.0

Table 2. Google Answers queries (141): music information needs categories

Music Information Need	Percentage
Identify artist/work	32.6
Seek information	25.5
Locate work	17.0
Acquire lyrics	6.4
Get recommendations	5.0
Ready reference	4.3
Locate specific version of work	2.8
Others	2.8
Seeking score/tab	2.1
Request translation	0.7
Request research	0.7
Total	100.0

In the following tables, we compared the features (types of information) that searchers provided in the music queries from Naver (Table 3) and Google Answers (Table 4) websites. The number of features provided in a query is highly variable depending on the information needs expressed in the query. Therefore, we delimited our analysis to the single most common category “Identify artist/work” in order to control this variability. 24 feature categories were found in the queries from both data sets.

By examining the relative frequency of the categories of *user-provided* information found in the queries between the more Western-centric Google, and the Korean-centric Naver, we hope to make the first steps toward classifying those features that we might consider “universal” and those “culturally determined”. This information can be very useful in the creation of *system-provided* access points and interfaces that afford access to both the “universal” and the “culturally determined” approaches to MIR query construction.

Table 3. Features provided by searchers in 51 Naver “Identify artist/work” queries

Feature	Percentage
Associated use (e.g., movie, ad)	58.8
Phonetic sound of lyrics	31.4
Audio/Video example	23.5
Lyric word	21.6
Name of artist	19.6
Description of musical style	19.6
Gender of artist	17.6
Name of musical genre	15.7
Mood/Affect of music	15.7
Description of work	13.7
Media (e.g., radio, TV)	11.8
Title of work	9.8
Tempo of music	9.8
Description of artist	7.8
Date of music released/heard	7.8
Name of musical instrument	5.9
Voice range of artist	5.9
Place where music heard	5.9
Subject of music	2.0
Nationality of music/artist	2.0
Similar work/artist	2.0
Storyline of work	2.0
Name of related event	0.0
Name of region	0.0

Table 4. Features provided by searchers in 46 Google Answers “Identify artist/work” queries

Feature	Percentage
Date of music released/heard	50.0
Name of musical genre	50.0
Lyric word	45.7
Associated use (e.g., movie, ad)	39.1
Name of artist	30.4
Description of musical style	30.4
Title of work	28.3
Gender of artist	23.9
Media (e.g., radio, TV)	21.7
Name of musical instrument	19.6
Description of artist	15.2
Audio/Video example	13.0
Nationality of music/artist	13.0
Similar work/artist	13.0
Description of work	10.9
Storyline of work	10.9
Mood/Affect of music	10.9
Name of region	10.9
Tempo of music	8.7
Place where music heard	6.5
Name of related event	4.3
Subject of music	2.2
Voice range of artist	2.2
Phonetic sound of lyrics	0.0

The statistical analysis was performed using Fisher’s exact test. The proportion of each feature used in each data set was pair-wise compared between two data sets and tested for significant differences. When the p-value was smaller than 0.05, we considered the difference to be statistically significant. Table 5 lists the features that had p-values (two-tailed) less than 0.05.

Table 5. Result of the statistical analysis

Variable	P-value (2-tailed)
Date of music released/heard	.000
Phonetic sound of lyrics	.000
Name of musical genre	.000
Lyric word	.017
Name of region	.021

In the queries where searchers were trying to identify an artist or a musical work, Google Answers searchers provided *date*, *genre*, *lyrics*, and *region* information significantly more often than Naver searchers. Naver searchers exploited the *phonetic sound of lyrics* significantly more than Google Answers searchers. It is through this preliminary examination that we begin to see the distinctions between the “universal” and the “culturally determined” query approaches users construct. For example, the *phonetic sound of lyrics* approach is found in a non-trivial percentage of the Naver queries. This implies that we must design MIR systems that can afford access using this “culturally determined” query approach.

In Section 4.2, we discuss difficulties that Naver searchers encountered with commonly provided access points in MIR, and describe what they did to successfully find music information despite these limitations.

## 4.2 Use of Traditional Access Points in Naver Queries

### 4.2.1 Bibliographic information

Naver searchers often failed to provide any bibliographic metadata such as composer, performer, or title in their queries, which are traditionally the primary access points of MIR systems. Searchers sometimes managed to remember the characteristics of the performer such as “*a sexy blonde singer who was wearing a skirt made with beads*” – details that might be helpful in jogging the memory of a human music expert, but not useful in a MIR system. Even when bibliographic metadata is known, searchers did not always spell the artist name or work title in English: Instead, a mix of English names and Korean pronunciation of those names was often used. Inconsistency in transliterating names from English to Korean frequently resulted in fatal search errors, as the following case illustrates:

“*I heard this song in my friend’s car. I asked what it was and he said he was sure (about the information), but I tried searching and got no results! Think he made*

a mistake? He told me the singer was *메리호키스* and the title was *도즈월드데이즈*, but I can't find it! I tried *메리허키즈*, *메리호키즈*, *더즈월드데이즈*, *더즈월드데이즈* as search terms, but didn't even find a similar one! Please tell me if somebody knows this<sup>1</sup>.”

Naver searchers often exploited “context metadata” to compensate for the difficulties in providing reliable bibliographic information. Context metadata indicates the “extrinsic aspects, uses, and relationships of music” and one category of context metadata is “associative metadata” which regards the use of music in other multimedia objects such as movies, commercials, etc. [8]. Associative metadata was found in 30 (58.8%) of the “Identify work/artist” queries from Naver:

“This pop song has been around for a while. I really want to know what it is! It was used in a TV commercial where Cha In-Pyo (Korean actor) was running somewhere, and it was also in the **OST** for Powerpuff girls. It goes like *파이틀 오버~ 먼먼머 파이틀 오버~ 먼먼먼먼먼먼먼 파이이야이야이야이야 와우?*.”

“What's the song played when the TV show Happy Sunday at 6pm on **KBS2** starts? I heard *오 미키?*.”

“Can you tell me the title of this song? The first part goes like *라라라~ 라라라라라~* and it's a pop song by a female singer. I heard it is played a lot in night-clubs...I only remember the words *everynight~ every-day~*.”<sup>4</sup>

TV shows, commercials, movies, and music videos seemed to be major sources for Naver searchers to encounter new pop music. It is not surprising that OST (Original Soundtrack) is often one of the main categories of music on Korean music-related websites [see Section 4.2.2]. Associative metadata was also found in 39.1% of the “Identify work/artist” queries from Google Answers, which is a smaller proportion than Naver queries, but still a significant proportion.

#### 4.2.2 Genre information

Studies of music genre and classification schemes [7], [13] conclude that genre definitions appear to be neither consistent nor objective, and there is no consensus on genre descriptions among users [4]. People generally find it difficult to clearly distinguish one genre from another, especially when they are faced with an unfamiliar genre label [12]. Pachet [14] added that the terms among different genre classifications are not consensual and the taxonomy structures do not match. A comparison of Korean and American genre classifications well illustrates this problem.

In Korea, the term *pop music* denotes very different songs than in America. “POP” is loosely used to indicate any Western music from outside of Korea, thus

people simply do not have much information about specific genres and how to distinguish among them. The Recording Industry of Association of Korea (miak.or.kr) collects record sales statistics under two major genres – *가요* (Korean popular music, also sometimes referred as K-POP) and POP (Overseas) – with no subgenres listed. Major Korean music Websites like ClickBox<sup>5</sup> has *가요* (K-POP), POP, OST (movies), 클래식 (Classical), 뮤직비디오 (music video), and 연예방송/뉴스 (Entertainment news) as the main categories. Another website, iLikepop,<sup>6</sup> has *가요* (K-POP), POP, J-POP (Japanese popular music), OST (movie-drama), 기타 (etc.), and 뮤직비디오 (music video) as the main categories. Genre can sometimes be useful in suggesting additional similar songs that a user may enjoy [7], however this may not be the case for Korean websites with very broad genres that include vastly different items in each category. For instance, on iLikepop.com, the artists found under the POP category include Britney Spears, Marilyn Manson, Wu-Tang Clan, Moscow Boys Choir, and the Chicago Symphony Orchestra. Consequently, it can be confusing and difficult for Korean people to apply genre labels to pop songs. In fact, many queries failed to provide ‘proper’ genre information and when searchers did suggest a genre, they often expressed uncertainty.

“I'm curious about **Linkinpark**'s music style. Is that alternative metal?”

Some Naver searchers explicitly stated that they wanted to explore a new genre and they were just starting to learn about these different kinds of music.

“I'd like to listen to pop, but I don't know much about it. Among Sweetbox's songs, 'life is cool' is my favorite. Please suggest pop songs like this that are easy to listen to.”

For Naver searchers, the association-based concept “Give me some music similar to this particular song(s) or artist(s)” [7] was a common key to finding new music. Among all Naver queries analyzed, 25 (23.8%) asked for suggestions similar to given titles/artists. Korean people may feel more comfortable asking for similar music by providing one or more titles/artists as examples, rather than providing detailed genre information (e.g., “...songs like **Sk8ter boi** by Avril Levine, just **Sk8ter boi** and not any of the rest of her songs”). Descriptions of personal music preference were often added:

“*비운세* – **crazy in love**, *제니퍼로페즈* – **play**, *스테이시오리코* – **stuck**, *에미넴* – **without me**. I really like these kinds of music, but I don't know any other songs like them. If you know these kinds of songs, please let me know as many as you can! I especially like the ones by *스테이시* and *비운세*.”

<sup>1</sup> Mary Hopkin (메리홉킨) - Those Were The Days (도즈워드데이즈)

<sup>2</sup> Knack – My Sharona

<sup>3</sup> Toni Basil – Mickey

<sup>4</sup> Kylie Minogue – Can't get you out of my head

<sup>5</sup> Clickbox.co.kr

<sup>6</sup> iLikepop.com

“Can someone suggest some depressing songs? For example, do you know **‘I need some sleep’** by **Eels**? That kind of style or I also like **Portishead’s ‘Wandering stars’** and **‘Roads.’** I want pop songs and not just ‘kind of’ depressing, but seriously depressing songs!”

“I’m starting to like pop songs lately and want some suggestions. 1) I like female singers. 2) I don’t like rock – I like ballad and lively songs. 3) I want the names of the singers. 4) Also I heard a very cheerful carol song by some female singer the other day – if you know what it is, let me know. 5) Give me more descriptions than just titles. If you know songs similar to **‘as long as you love me,’** please recommend as well.”

Bainbridge et al [5] have suggested compensating for users’ widely varying music descriptions by providing query-by-example facilities in MIR systems to allow users to ask for ‘more things like this.’ In [5], they found that few music queries posed to Google Answers included an audio representation of the desired music, likely due to limitations in the interface. In our data set, 12 of analyzed Naver queries (11.4%) used Naver’s embedding and linking features to provide an audio/video example while only 6 of analyzed Google Answers queries (4.3%) included links to audio/video representation of sought music. Interestingly, when there was no pre-existing example on the Web to link to, a few Naver searchers even created their own examples by capturing part of the audio, or recording their own performance like the query below.

“I heard this song in the late 70s or early 80s, so when I was really young. I can’t remember the singer or the title. The unique voice and saxophone were really impressive. I tried to play the part that I remember. I think the basic melody is at least 85% correct. Please help me find this song in my memory. <Audio file included>”

In Korea, 12 years of formal music education is a state-mandated requirement, and we speculate that this may have some effects on music information seeking behaviors like such given above. The majority of Korean people know how to read sheet music and play some sort of musical instrument, thus rather than providing inaccurate or uncertain bibliographic or textual information, some may feel more confident in providing self-made music examples. This assumption should be further tested in the future user studies of Naver searchers.

#### 4.2.3 Lyric information

23 of analyzed Naver queries (21.9%) and 25 of analyzed Google Answers queries (17.7%) included lyric information. In the Naver data set, 10 queries included lyric information in English and 15 queries included transliteration of English lyric words in Korean. The average number of lyric words given in a single query was 6.1 for the Naver data, and 33.96 for the Google Answers data. One of the Google Answers queries had a

transcription of lyrics based on an audio example of the sought music which highly influenced the average number of lyric words – without this outlier, the average number of words drops to 19.38, which is still much higher than the Naver queries. The Naver query with the longest lyric information in English had only 14 words, shorter than the average number of lyric words (19.38) in Google Answers queries. The kinds of lyric words that Korean searchers were able to catch and remember were common words (e.g., baby, love, eyes, funky) that may lack discriminating power in IR. In the Naver queries, it seemed unlikely that English lyric information alone would effectively help others to successfully answer the queries. In most successful cases, other kinds of information were additionally provided to help the answerer as the following example illustrates:

*Q: This was a popular dance song in late 80s. They often play this song in the nightclubs. I used to listen to it without much thought, but since when I heard it again on **mbc 일요일밤에 대단한도전** (TV show) where Lee Yun-seuk was riding rollerblades, I’ve been wanting to know the title of this song. I don’t know the lyrics very well because the rhythm’s fast, but I think **funky funky** was repeated and also pronunciation like **송송** [shong shong] was repeated. It was sung by a male singer. I’m sorry that my hints are poor, but I would appreciate it if someone can tell me what this song is.*

*A: I’m just guessing since you said late 80s. If it is French, it may be correct...*

**Debut de Soiree – Nuit De Folie**

**Year: 1988**

**Country: France**

**Et tu chantes, chantes, chantes, ce refrain qui te plait** (에뚜 송뜨 송뜨 송 쎄 레퀘테켓) – It is possible that this is the **송송** [shong shong] part...**La cadence du funk au plus haut t’emmene** - I think this may be the funky part...If the song that you are looking for is in English or the singer is black, this is probably not the one. Check it on **벅스 뮤직** (a Korean music website).

## 5 CONCLUSION AND FUTURE WORK

Our study has revealed some interesting characteristics of cross-cultural/multilingual music information seeking behaviors. Meeting the needs of a broad range of music users across different cultures and languages is a challenging task, but is nonetheless necessary for the realization of truly global music digital libraries. Our analysis supports the idea that research should commence on the establishment of new kinds of access points beyond the bibliographic, genre, and lyric information in order to accommodate queries that cross cultural and/or language boundaries.

One option may be to consider developing metadata that are culturally neutral and thus can easily be picked up by multicultural/multilingual users. For example, the

gender of performer is often used as a query feature but rarely formally included in any metadata system. Research into capturing and better understanding the universal aspects of music mood and affect would also enhance the cross-cultural/multilingual retrieval of music.

In the future works, we will address the following limitations of this study. First, we will explore new sources other than Naver 지식 iN and Google Answers for data collection. These Q&A websites can develop certain “culture” and group norms over time, which may reinforce users to behave in certain ways. In order to eliminate this variation in users’ information searching behaviors due to the individual website’s own culture and study the true variation between Western and Non-western music information searchers, queries must be collected from more than one representative website for each group. Second, the queries themselves do not provide in-depth information on the reasons *why* searchers asked certain kinds of questions, and why they did or did not provide certain information in their music information need descriptions. Studies with more qualitative approaches employing interviews or a focus group like [12] should be done to obtain such information.

## 6 ACKNOWLEDGEMENTS

We thank the Andrew W. Mellon Foundation for their moral and financial support for the HUMIRS (Human Use of Music Information Retrieval Systems) project. This project is also supported by the National Science Foundation (NSF) under Grant Nos. NSF IIS-0340597 and NSF IIS-0327371. We also thank Dr. Swicegood in the Sociology department at the University of Illinois at Urbana-Champaign (UIUC) for statistical consultation, and M. Cameron Jones at the Graduate School of Library and Information Science at UIUC for assisting in data collection.

## REFERENCES

- [1] Downie, J. S. Music information retrieval. In Annual Review of Information Science and Technology 37, ed. Blaise Cronin, Medford, NJ: Information Today, 2003, 295-340.
- [2] Futrelle, J., and Downie, J. S. “Interdisciplinary communities and research issues in music information retrieval”, Proceedings of the Third International Conference on Music Information Retrieval, Paris, France, 2002, 215-221.
- [3] Lippincott, A. Issues in content-based music information retrieval. Journal of Information Science, 28, 2002, 137-142.
- [4] Petrelli, D., Beaulieu, M., Sanderson, M., and Hansen, P. User requirement elicitation for cross-language information retrieval. The New Review of Information Behaviour Research: Studies of Information seeking in context, 3, 2002.
- [5] Bainbridge, D., Cunningham, S. J., and Downie, J. S. “How people describe their music information needs: A grounding theory analysis of music queries”, Proceedings of the Fourth International Conference on Music Information Retrieval, Baltimore, USA, 2003, 221-222.
- [6] Downie, J. S., and Cunningham, S. J. “Toward a theory of music information retrieval queries: System design implications”, Proceedings of the Third International Conference on Music Information Retrieval, Paris, France, 2002, 299-300.
- [7] Vignoli, F. “Digital music interaction concepts: a user study”, Proceedings of the Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 2004.
- [8] Lee, J. H., and Downie, J. S. “Survey of music information needs, uses, and seeking behaviours: preliminary findings”, Proceedings of the Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 2004, 441-446.
- [9] Park, J., and Jeong, D. An empirical study on web-based Question-Answer services. Journal of the Korean Society for Information Management, 21(3), 2004, 83-98.
- [10] Lee, T., and Kang, S. “A study on knowledge management of web search engine based on knowledge search community”, The Tenth Proceedings of the Korean Society for Information Management Conference, 2003, 143-152.
- [11] Earley, P. C., and Gibson, C. B. Taking stock in our progress on individualism-Collectivism: 100 years of solidarity and community. Journal of Management, 24(3), 1997, 265-304.
- [12] Cunningham, S. J., Reeves, N., and Britland, M. “An ethnographic study of music information seeking: implication for the design of a music digital library”, Proceedings of the Third ACM/IEEE-CS joint conference on Digital libraries, 2003, 27-31.
- [13] Fabbri, F. Browsing music spaces: categories and the musical mind. Presented at the Third Triennial British Musicological Societies Conference (Guildford, UK, 1991). Retrieved from <http://www.theblackbook.net/acad/others/fabbri990717.pdf>.
- [14] Pachet, F. Content management for electronic music distribution. Communications of the ACM, 46(4), 2003, 71-75.