Evidence Based Library and Information Practice

Evidence Summary

Music Information Seeking Behaviour Poses Unique Challenges for the Design of Information Retrieval Systems

A Review of:
Lee, J. H. (2010). Analysis of user needs and information features in natural language queries seeking music information. *Journal of the American Society for information Science and Technology*, *61*, 1025-1045.

Reviewed by:
Cari Merkley
Librarian
Mount Royal University
Calgary, Alberta, Canada
Email: cmerkley@mtroyal.ca

Abstract

Objective – To better understand music information seeking behaviour in a real life situation and to create a taxonomy relating to this behaviour to facilitate better comparison of music information retrieval studies in the future.

Design – Content analysis of natural language queries.

Setting – Google Answers, a fee based online service.

Subjects – 1,705 queries and their related answers and comments posted in the music category of the Google Answers website before April 27, 2005.

Methods – A total of 2,208 queries were retrieved from the music category on the Google Answers service. Google Answers was a fee based service in which users posted questions and indicated what they were willing to pay to have them answered. The queries selected for this study were posted prior to April 27, 2005, over a year before the service was discontinued completely. Of the 2208 queries taken from the site, only 1,705 were classified as relevant to the question of music information seeking by the researcher. The off-topic queries were not included in the study.

Each of the 1,705 queries was coded according to the needs expressed by the user and the information provided to assist researchers in answering the question. The initial coding framework used by the researcher was informed by previous studies of music information retrieval to facilitate comparison, but was expanded and revised to reflect the evidence itself. Only the questions themselves were subjected to this iterative coding process. The answers provided by the Google Answer researchers and online comments posted by other users were examined by the author, but not coded for inclusion in the study.

User needs in the questions were coded for their form and topic. Each question was assigned at least one form and one topic. Form refers to the type of question being asked and consisted of the following 10 categories: identification, location, verification, recommendation, evaluation, ready reference, reproduction, description, research, and other. Reproduction in this context is defined as "questions asking for text" and referred most often to questions looking for song lyrics, while evaluation typically meant the user was seeking reviews of works (p. 1029). Sixteen question topics were outlined in the coding framework. They included lyrics, translation, meaning (i.e., of lyrics), score, work, version, recording (e.g., where is an album available for purchase), related work, genre, artist, publisher, instrument, statistics, background (e.g. definitions), resource (i.e. sources of music information) and other.

The questions were also coded for their features or the information provided by the user. The final coding framework outlined 57 features, some of which were further subdivided by additional attributes. For example, a feature with attributes was title. The researcher further clarified the attribute of title by indicating whether the user mentioned the title of a musical work, recording, printed material or related work in their question. More than one feature could appear in a user query.

Main Results – Overall, the most common questions posted on the Google Answers service relating to music involved identifying works or artists, finding recordings, or retrieving lyrics. The most popular query forms were identification (43.8%), location (33.3%), and reproduction (10.9%). The most common topics were work (49.1%), artist (36.4%), recording (16.7%), and lyrics (10.4%). The most common features provided by users in their posted questions were person name (53%), title (50.9%), date (45.6%), genre (37.2%), role (33.8%), and lyric (27.6%). The person name usually referred to an artist's name (in 95.6% of cases) and title most often referred to the title of a musical work. Another feature that appeared in 25.6% of queries was place reference, almost half of which referred to the place where the user encountered the music they were enquiring about. While the coding framework eventually encompassed 57 different features, a small number of features dominated, with seven features used in over 25% of the queries posted and 33 features appearing in less than 10%. The seven most common features were person name, title, date, genre, role, lyric, and place reference.

Lee categorized most of the queries as "known-item searches," even though at times users provided incorrect information and many were looking for information about the musical item but not the item itself (p. 1035). Other interesting features identified by the author were the presence of "dormant searches," long standing questions a user had about a musical item, sometimes for years, which were reawakened by hearing the song again or other events (p. 1037). Multiple versions of musical works and the provision of information gleaned third hand by users were also identified as complicating factors in correctly meeting musical information needs.

Conclusion – While certain types of questions dominated among music queries posted on the Google Answers service, there were a wide variety of music information needs expressed by users. In some cases, the features provided by the user as clues to answering the query were very personal, and related to the context

in which they encountered the work or the mood a particular work or artist evoked. Such circumstances are not currently or adequately covered by existing bibliographic record standards, which focus on qualities inherent in the music itself. The author suggests that user context should play a greater role in the testing and development of music information retrieval systems, although the instability and variability of this type of information is acknowledged. In some cases this context could apply to other works (film, television, etc.) in which a musical work is featured. Another potential implication for music information retrieval system development is a need to re-evaluate the terminology employed in testing to ensure that it is the language most often employed by users. For example, the 128 different terms used in this study to describe how a musical item made the user feel did not significantly overlap with terms employed in a previous music information retrieval task involving mood classification conducted through MIREX, the Music Information Retrieval Evaluation Exchange, in 2007. The author also argues that while most current music information retrieval testing is task-specific – e.g., how can a user search for a particular work by humming a few bars or searching for a work based on its genre, in real life, users come to their search with information that is not neatly parsed into separate tasks. The study affirms a need for systems that can combine tasks and/or consolidate the results of separate tasks for users.

Commentary

This study reaffirms the value of evaluating information retrieval systems with data gleaned from empirical studies of users in their natural habitat. As the author of the study rightly points out, what is particularly valuable in this instance is that the queries used in this study were not shaped by their interaction with a particular database or existing bibliographic records, but rather contained the information that users thought

would be most helpful in tracking down the answer to their question.

The types of questions logged may suggest that in many cases the information need users were attempting to satisfy was personal rather than academic, which may play a role in the potential applicability of the results to certain contexts. However, the high level at which the data is presented in the study and the potential overlap between these spheres make it difficult to achieve a clear determination on this issue. The Google Answers service may have attracted a particular type of music seeker, but the fact that users are expressing their questions in free form makes it a particularly rich source of data on how users articulate their information needs.

The field of music information retrieval research is complex, and involves experts from a variety of fields, of which information science is one (Downie, 2008). Throughout the study, the author draws on existing work on information retrieval while clearly making the case for the unique challenges faced by individuals working to facilitate user access to the rich body of music information objects in existence. Another source of research on this particular issue is the proceedings of the annual conference of the International Society of Music Information Retrieval (2010). Non-music specialists may also find value in the methodology employed to answer other types of research questions. The author provides considerable detail on the coding framework created to support content analysis of the Google Answers questions and addresses some of the advantages and challenges posed by use of web resources as artifacts of information seeking. The author highlights the advantages of content analysis as a methodology, such as the ability to express results both numerically and qualitatively. The author also clearly addresses the issue of the representativeness of the data sample, and refrains from making sweeping generalizations based on the data. Finally, the author's call for more empirical studies on user behaviour and less reliance on anecdotal evidence when creating information systems

will strike a chord with information professionals generally, not just those working with music.

References

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.

International Society of Music Information Retrieval. (2010). ISMIR - The International Society for Music Information Retrieval. Retrieved November 22, 2010 from http://www.ismir.net/