Subject Search for Music:
Quantitative Analysis of Access Point Selection

Background

Non-book materials have unique characteristics in subject analysis that influence their information retrieval process. It is obvious that online search studies intended for books are not appropriate for music materials. There are not enough empirical studies based on what objects focus on music materials. This study aims to explore useful tools for subject search of music scores in online environment. To comprehend problems regarding the issue, the author surveyed 21,177 online catalog search logs in an academic library of music.

Smiraglia (1989, p. 64-73) describes the characteristics of music and divided them into 4 groups according to the subject analysis of textual materials: 1) intellectual form; 2) topicality; 3) the intended audience; 4) physical form. In the study such attributes to music are considered as subject aspects for searching. The elements include topic, genre, form of composition, technique or style, medium of performance, physical format or version, language of text, historical style period or time. The author examined these elements for analysis in the survey.

Purpose

The purpose of this survey is to comprehend how the searchers selected a variety of access points for examining what search tools are most effective and necessary for searching music, and in what way they should be provided, especially for an unknown item search. The survey was carefully treated not to include the data restricted by the system requirements from which they were collected, because the purpose of the study is not to explore problems on system designing, or user interface. Manual analysis of data carried out here excludes the problems resulting in the system limitations that may differ in each online system, such as misspelling correction, stemming standardization, truncation, stop word lists, and so on. Instead, the researcher focused on which access point the user chose, not on how s/ he input terms.

Methods

Samples of transaction logs were taken from the OPAC system in the Library at Kunitachi College of Music in Tokyo, Japan. The data were collected for 82 days from September 1st to December 20th, 1999. This is a part of a second semester that does not include an examination period and is not as disorganized as at the beginning of a school year (April in Japan). The number of queries conducted by library users in the sample data was 39,811 of which search sessions for music search were 21,177. The number of initial queries was 11,435. A log for a query consisted of "ID","Date","Time","Terminal no.","Database no.","Language","Search field","Input terms","Number of results","Time taken for searching (second)".

Initial queries of each search session were used to analyze types of access points. The reason for it is because an initial query can be considered to most express the searchers' intention, and access points of the sequential queries in the same session are the ones chosen from the results of interaction with the initial search results. Only when a user made simple mistakes, for example, inputting personal name as a title search, the initial access points were ignored. These were corrected to the proper ones by the researcher. If access points of the successive queries seemed appropriate for the searcher's intention, they were taken into account. After adequate access points were determined, subject terms were focused for further analysis and examined according to the categories of music characteristics. Finally, searches by subject terms after zero-hit results or information overload of initial queries were analyzed.

Results and Discussion

There were two major findings in the quantitative analysis of access point selection. First, more than half of the initial searches were combinations of access points. It became clear that searchers

tend to choose more than two access points at the beginning of their search. Second, the combinations of search types varied according to the searchers' intentions. Less usage of only a single access point means that its selection is not efficient enough for music search.

As to searches by subject terms, terms for medium of performance and genre were most frequently used in initial queries as well as in secondary ones to refine initial queries. When initial searches with title/ uniform title or opus number were modified in order to raise their recall rate, subject terms for medium of performance were selected in secondary queries. It was found particularly for unique musical instruments like marimba or electronic organ. Searchers selected related terms of initial input words to refine their initial search results instead of changing types of access points. The narrower term of the subject word used in an initial query was selected, e.g., "trombones (3)" for initial search term "brass trios". In both cases to raise recall rate and precision rate, subject terms were used with various access points to construct search queries. In order to specify initial queries by adding other access points to them, publisher's name, editor, edition and physical format of music were important elements.

Selection of genre terms were seen when initial searches failed to find a portion of a large musical work or a piece in larger collections of a certain genre. In this case, personal name and title search was not effective enough. Therefore, searchers selected subject terms expressing genre to expand the search results. When genre terms were used in their initial queries, they were added terms of medium of performance in the secondary queries to limit search results. Also, genre terms were selected to modify initial queries with words expressing general topic, for example, "death" to which adding "ceremonial music."

Quantitative analyses show that personal name or title search is not as useful in music information retrieval as it is in the textual one. In most cases, subject terms provide more comprehensive searches compared with those of title words as access points. However, such subject terms which are related to the attributes to music can be accessible from more than one input fields in online search system. These different choices for a particular element of music hinder effective searches. It is closely related to the organization of bibliographic records which describe intellectual contents and physical characteristics of the bibliographic entity. For instance, the medium of performance is accessible in title, subject headings or classification numbers that are not clearly distinguished between them. In addition, it is necessary to supply various types of access points for all attributes of music that were found important in online information retrieval for music. Proper control of these elements may improve subject searches of music. A qualitative survey must be conducted to study that which cannot be covered by the transaction log analyses. This further study will clarify the issues of the exact search purpose of each library user and his/ her satisfaction level in search results through it in order to devise a way of effective searching.

Mari Itoh
Graduate School of Library and Information Science
Aichi Shukutoku University
989043x@asu.aasa.ac.jp