This is the published version of a paper presented at *XXV Bobcatss Symposium, Tampere, Finland, January 2017*.

Citation for the original published paper:


N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-68106
THE IMPORTANCE OF QUALITY CLASSIFICATION AND SUBJECT INDEXING OF HEALTH INFORMATION IN PUBLIC LIBRARIES: A COMPARATIVE ANALYSIS

KRISTINA BERKETA
Department of Information Sciences
University of Zadar
Zadar, Croatia
kberketa@student.unizd.hr

MADELEINE LUNDMAN
Department of Library
and Information Science
School of Cultural Sciences
Linnaeus University
Växjö, Sweden
ml222vc@student.lnu.se

MARTINA DRAGIJA IVANOVIĆ
Department of Information Sciences
University of Zadar
Zadar, Croatia
mdragija@unizd.hr

KORALJKA GOLUB
Department of Library
and Information Science
School of Cultural Sciences
Faculty of Arts and Humanities
Linnaeus University
Växjö, Sweden
koraljka.golub@lnu.se

Keywords: classification, subject indexing, health information, public library

Abstract

The general aim of this paper is to explore the practice of subject indexing and classification of health information in public libraries. The pilot study encompasses two public libraries, each from a different European country, Croatia and Sweden: Zadar Public Library, and Växjö Public Library. The research questions are: 1) how important is the terminology for subject headings; and, 2) how many subject terms are enough to describe an information resource in a catalog record? The research was done in two phases. First, an exploratory evaluation of subject searching in the catalogs was conducted. Second, after the analysis of the catalogs, questions for interviews were formed. Four semi-structured interviews were conducted, two per library, of which one interview with a subject librarian and one with a reference librarian. Finally, a comparative analysis between the two libraries from the two countries was conducted. Results imply that the subject headings and classification used were very generic and did not cater for specific, topically narrower queries. Also, subject indexing was considered important, especially to the librarians for searching purposes through ensuring consistency in the catalog, thereby making it easier to find resources on the same topics.
Introduction

We, librarians, and information scientists, like to live in a belief that library catalogs are more useful than Google search. In reality, a library catalog represents a selection of information resources but the access to them by end users seems to be difficult. Library catalogs still struggle with simple keyword search which users expect to implement with ease as they do in web search engines. This paper attempts to enlighten the need for better subject access and classification with the focus on health information.

Literature review

Health information

Health information includes different categories, for example, medical information, dietary information, pharmaceutical information, fitness information, information on how to fill out medical forms etc. It is important for the public to be aware that their library provides access to health information (Rubenstein, 2015). But to understand what is needed to fully comprehend it librarians and users must have sufficient health information literacy. “Health literacy is a foundational and essential notion that connects a variety of key issues in health policy, health services research, health communication, and health care delivery (Luo & Park, 2013).” It has been reported that public libraries are often the first access point to extend, confirm, or refute the information learned from the health professionals. The user may approach the reference desk with health-related question, for themselves, a family member, or a close friend. Some come to the library prior to their doctor’s visit to research their symptoms, or after the visit to the doctor to gain further information (Alcock, 2000).

According to Luo & Park (2013), public libraries play an important role in patrons’ quest for health information by providing relevant print and electronic resources, free computer and internet services which allow patrons easy access to the online health information, and reference services which help patrons address their health information seeking needs. What is most important for libraries is to keep up with current health information in their physical and online collections. Also, when presenting the information to the users, librarians should keep in mind the delicate nature of that kind of information and should approach the user without any judgement or personal views. Study conducted by Ingham (2014) concluded that most users visited the library for health information because they were already aware that the library has that information.

While reviewing related literature, some issues arose. Could it be that the library catalog is inefficient in helping the librarians find health information they already have in library collections? This is the question which we tried to answer in this pilot study.

Subject indexing and classification of health information

One important part of access to health information is health literacy; in the words of Flaherty and Kaplan (2016): “A primary component of health literacy is the ability of an individual to obtain basic health information.” Libraries have been identified by users as a place for trustworthy health information resource (ibid.). Users have high expectations of libraries, especially public libraries where we can encounter all sorts of different user questions.

In this exploratory research, we wanted to examine what kind of subject access to health information is offered to the users, and whether the information available in the catalog is sufficient to assure that users are finding the answer they are seeking. According to Golub (2016) subject searching is one of the most common and yet the most challenging type of searching a library catalog. Kumar (2012) says that today’s searching strategies are those supported by Google; research has shown that users execute the same search strategies for finding information in OPACs as they do in search engines. Further, we see a rising trend that only search results retrieved on the first page are viewed (ibid.).

We would like to emphasize the importance of quality subject indexing and classification. “‘Good’ indexing can be defined in a very pragmatic way as indexing that allows items to be retrieved from a database in searches in which they are useful response and prevents them from being retrieved when they are not (Lancaster, 1998).” When users encounter a failed search in OPAC the usual response is to leave the search because they presume that the information does not exist in the library collections (Kumar, 2012). Expected challenges with subject searching are: concepts take various names (synonyms) and a name could take various forms; most users have difficulty with conceptualizing information need and transforming it into search query (Golub, 2016). One way in which libraries could approach this problem is by improving the subject access because users use broad keywords for which search engines always retrieve answers regardless of the quality. Yu and Young (2004) studied the impact of web searches on subject searching in OPACs and reported that users value quality of the results more than the process but they expect the search to be quick and easy. Good subject indexing and classification must be achieved so that users can retrieve the right answer with minimal effort.

Research

The general aim of this pilot study is to explore the practice of subject indexing and classification of health information in public libraries. Specifically, we want to determine to what degree librarians value good subject indexing and classification in their catalogs as an every
The research was conducted in two phases. Firstly, each of the public libraries' catalogs was searched for the same ten health information topics randomly selected by the authors. Secondly, after the analysis of the search results and observed issues in catalogs, questions for the interview were formed. Four semi-structured interviews were conducted, two per library, one interview with a subject librarian and one with a reference librarian. We wanted to see different points of view on the same subject and problems. Finally, a comparative analysis between the two libraries from the two countries was conducted.

**Methodology**

The empirical data for the study were collected in two phases, from December 5 to December 8, 2016. The first phase was the search of the two library catalogs using ten selected health information topics, which were represented by one major term and one closely related term, the latter where it was estimated that users would use a different form of the concept. The selected terms were as follows: diabetes/sugar disease, cancer/oncology, depression/angst, mammography/breast cancer, gluten intolerance/celiac, stroke, stress, high blood pressure, pregnancy, exercise/yoga. The terms were typed in library catalogs search box and the number of the retrieved results was recorded. Also, additional topics automatically suggested by the catalog for search modification were observed.

The second phase of the study comprised the interviews of four librarians, two in each library. The interview questions were formed after the initial search of the catalog and observing initial problems. The subject librarian was interviewed to further explore the observations resulting from the first phase, and to learn about the subject indexing and related library policies. The reference librarian was selected because the authors wanted to find out how every day interaction with users and their questions on health information were influenced by the subject indexing.

**Library catalogs**

**Zadar Public Library**

Vero, the library catalog (http://161.53.142.3/cgi-bin/wero.cgi?q=&x=44&y=21) is a third generation FRBRred catalog. The records are created by the library and some are downloaded from the union catalog. This seems to be a problem for the catalog because the union records are not thoroughly examined for inconsistencies among the records. The Vero catalog offers its users a one-box search interface; there is not any help offered to the user or explanation how to do a proper search. The only instruction for the user on how to use the catalog is to type in author, title, subject, words, ISBN, ISSN, or publisher under the search box. There is no option for advanced search immediately available; only after making the first search, one of the options that appear in the results interface is an advanced search.

The results interface offers many possibilities of a faceted catalog. Search results are first ranked by relevance, which is not clear how is it measured; an option to rank by title and date is also available. Users have the possibility to limit the search by subject (it is unclear what is meant by this), title or format, and the branch location of the title. The catalog offers users to refine their search by author/contributor, topic (it is unclear what is meant by this), language and year. We can only assume that the topics are derived from the subject fields of the records, no explanation is given. The catalog also offers users an option to "try a new search", by clicking on automatically suggested authors/contributors and terms (also unclear what this refers to). Again, it is not clear how the catalog connects the searched query with the given suggestions.

**Växjö Public Library**

Växjö has the catalog integrated on the library website (https://bibliotek.vaxjo.se/web/arena) and the library system used is BOOK-IT provided by Axiell Sweden. The records in the catalog are included when purchasing books and they are provided by BURK (a service from BTJ Sweden, http://www.btj.se) and LIBRIS (The Swedish National Catalog). The records are consistent throughout the catalog and follow the same pattern. The catalog offers its users a one-box search interface, like the Vero catalog, but also with no help or instructions of how to search. The only help provided is an instruction “search for books, movies, music etc.” On the left-hand side of the screen, there is a cogwheel that leads to an advanced search option, where one can search by free text (same as start page), but also by author, subject, keywords and title. It is not specified what is meant by the latter two fields. In the advanced search option, there is an option to also refine the search by choosing library, category, language etc.

The results interface offers many possibilities of a faceted catalog, but not as detailed as the Vero Catalog. The search results are first ranked by relevance, then there is an option to rank by title, year of publication and author. The search can also be refined by different facets: author, media, subject, language, target group, year of publication, keywords, and tags. Users have the option to choose from available facets.
Results

Catalog search
The study of the catalogs was conducted using ten terms. The terms were searched in Croatian/Swedish: dijabetes/diabetes (diabetes), šeőerna bolest/sockersjuka (sugar disease); rak/cancer (cancer), onkologija/onkologi (oncology); depreœija/depression (depression), anksioznost/ängst (angst); mamografija/mammografi (mammography); rak dojke/bröstcancer (breast cancer); netolerancija gluten/glutenintolerans (gluten intolerance), celi-jaklija/celiaki (celiac); moždani udar/stroke (stroke); stres/stress (stress); visoki krvni tlak/högt blodtryck (high blood pressure); tvrdnoœa/graviditet (pregnancy); tjelovježba/träning (exercise), yoga/yoga (yoga). Further in text, terms in English will be used for better understanding.

Table 1 below shows the results of the search. The results show considerably higher number of results retrieved by the search in Växjö than in Zadar for most of the cases. The exact reason for this difference could not be explained, but one may assume that this could be because Växjö may have a more complete collection, a better library system or better subject indexing practice.

The table below also lists the terms which were searched. The authors additionally used a second term, a layman form more used according to the knowledge of the authors. The problem of different terminology for health information and its relatedness during retrieval and for deriving automatic suggestions was unclear. For example, in Zadar 13 results were retrieved by term “diabetes”; “sugar disease” retrieved additional 27 results. The terms which were then automatically suggested after search by “diabetes” were all beginning with the phrase “sugar disease”; while when searching for “sugar disease”, “diabetes” was not offered as a related subject term, and results retrieved differed from the previous. The problem with terminology was also viewed on an example of a homonym. In the Croatian language “rak” can refer to cancer, the disease, but also represents the astrological sign. Thus, the search results included records classified by UDC number 615 Pharmacology. Therapeutics. Toxicology, 133 The paranormal. The occult. Psi phenomena or 82 Literature. No disambiguation help is offered.

There was also a problem in Växjö while using different synonyms. For example, in Sweden, the term “diabetes” has become more used than the Swedish term “sockersjuka” that was used more before. The results also show that “diabetes” retrieved more records (65 results versus 43 for “sockersjuka”). One would think that both searches should give the same results, but unfortunately, it is important to find the “right” words and terminology to get the best result. The results were all classified by the Swedish Classification System (SAB) with Veo for “diabetes” and Vep for “cancer” when classifying textbooks; in addition, some books had Hcf for children’s books and Lz for biographies.

When examining retrieved records, it has been noticed that approximately two subject headings are assigned to most records. The Zadar library downloads subject headings from the union catalog. There are no official national subject headings or indexing policies in place. Subject headings create a very general index, e.g. the most relevant record retrieved on “stroke” is indexed by subject heading Stress--Manual and by UDC number 616 Pathology. Clinical medicine.

In Växjö catalog records are downloaded from BURK or LIBRIS, which already have the subject headings included, while allowing for the possibility to add or remove subject headings by the local librarian. Subject headings in the catalog are also few and general, e.g. the most relevant record on “stress” was indexed with four subject headings, personal training, personal development, psychology, and applied psychology and classified under Dok for applied psychology.

Interviews

Health information queries
In Zadar, the librarians encounter health information queries daily, many users come to the library especially because of their medical collection. They have a great variety of questions and interests related to health. But also, they look for very specific health related information. In Växjö the librarians do not encounter many health information queries, which they believe is because of the Internet. The reference librarian in Växjö (RLV) encounters questions about health information very seldom. When they cannot find relevant books, the librarians refer users to articles (e.g. Läkartidningen, a Swedish medical journal) and to 1177.se (a Swedish online database of health information for the general public).

Terminology

The subject librarian in Zadar (SLZ) said that they create two kinds of subject headings, controlled, and freely formed. The controlled ones are downloaded from the
union catalog and can be modified. As discussed above, they are very general in their focus. When indexing with both types of subject headings, the librarian looks up the terminology first in general dictionaries and then in different medical dictionaries. Also, the librarian looks at users’ queries (e.g., from the reference interview) when forming freely formed headings, making subject access more user friendly.

Both librarians in Växjö find the terminology very important for quality retrieval by both librarians and users. To find relevant information, the subject librarian in Växjö (SLV) uses both classification and controlled subject headings. The website 1177.se is also an important help for finding the right term for indexing. RLV also uses classes as a tool for the use by librarian, but stated that it is not useful for the patrons.

Subject headings

SLZ said that they are not a special library of a medical institution and by their estimate one, two or possibly three subject terms are enough. But she emphasized that if they observe that users frequently seek out questions about a particular medical problem then they assign additional index terms to existing records. SLV said that the subject terms used (often two or three) are not enough to describe the resource. The subject headings themselves are usually so generic that they are hardly of any help. SLV also stated that their search engine goes through the table of contents as well as the title, subject and so on. Therefore, even if the subject headings are not enough, these other fields may help retrieve the right resources. The subject headings do not usually matter to the patrons, when they want a specific book, they find that anyway.

RLV also said that the subject terms used are not enough to describe information resources at hand. The subject terms in the records are general while the questions the patrons have are specific, which makes it hard to find the material through subject headings. The patrons will probably need help from the librarians to find information regarding specific questions. RLV also has the same opinion about general subject terms in their catalog but sees their freely formatted headings as extra useful to the users.

Improving subject access to health information

SLV said that there is a lot one can do to improve the subject access to health information, but all the things are time-consuming and not prioritized in their work. It is also a question about the need: since the Växjö Library gets so few questions about health information, this is not really an issue. SLV still gave some examples on how to improve subject access to health information: more (specific) subject headings, good classification (so that it is easy to find resources on the shelves as well as in the catalog), complete records with links to further information, developing the library catalog so it is more like Google, to get help with suggestions and misspelling. On the other hand, SLZ said that the lack of a controlled vocabulary, whether medical or any kind, is a big problem in the Croatian subject indexing practice. RLV said that a person with medical background or knowledge would be most useful when indexing health information.

Both librarians in Växjö stated the importance of the information to be up to date, especially when it comes to health information, since it can be dangerous otherwise (RLV). Växjö library weeds all sections of the collection often to get rid of outdated information, so most of the health information books were only a few years old (text books). The librarians also thought that it is important to inform the patrons about 1177.se if they did not know about it. RLV said that the library can provide relevant information about health when it comes to the more general questions. For more specific questions, if they cannot do an interlibrary loan or find articles, the librarians refer users to the local hospital library. RLZ is aware that they hold outdated books on health, but their practice is to warn the users about those resources.

Concluding remarks

Public libraries are recognized as places that offer access to health information. This pilot study aimed at exploring indexing practices of two different public libraries, in two different countries. By analyzing the catalog of the two libraries first and seeing that both libraries index their health information using the minimum number of subject headings and classes, we asked ourselves whether that suffices for an end user and her finding the right information.

According to the librarians, the subject terms used in both catalogs are enough when it comes to answering more general queries, but lacks specific subject terms when the patrons have more specific questions. The subject headings used in the catalogs are not enough, but with the help from the librarians (Växjö) and freely formatted subject headings (Zadar), the user can still often find what they want. We could conclude that a practice of expanding the subject headings with more than two subject headings assigned could be more informative to the user.
References


