Creating Intellectual Access through the Organization of Knowledge: An Introduction to a Project-Specific Approach

Amanda Wakaruk

Introduction

Creating access to information about knowledge for the purpose of creating new knowledge epitomizes the efforts of librarianship. For centuries, librarians have been charged with collecting, organizing, and disseminating recorded knowledge. In this role, we have acted as intermediaries between users and creators, employing structures that are dependent on a consistent and shared language, the components of which are rarely obvious or intuitive for either party. This pursuit has contributed greatly to the formation and evolution of library and information science as a discipline. Academic activity in this area has been most active over the past century. Today, we are fortunate to have tools built on a strong theoretical foundation, tools that allow library users to access knowledge created by a wide range of disciplines and for interests as varied as the human species itself.

The Gender and Work Database (GWD) provided an opportunity to build upon a body of knowledge within library and information science and to develop new tools specific to the needs of the project.
The first section of this paper provides a focused, brief introduction to the area of library and information science concerned with the organization of knowledge. The second section discusses our application of this work in the context of the GWD.

**Existing Structures for Intellectual Access**

Birger Hjorland, research professor at the Royal School of Library and Information Science in Copenhagen, describes knowledge organization as an interdisciplinary field of study that includes intersections with the research of library and information science. The goal of library and information science as it pertains to knowledge organization is the development of knowledge on the optimization of subject access points for purposes of information retrieval (87-88). This intellectual access is created by providing accurate and consistent representation of works that embody knowledge.

The representation of individual works for ease of access has long been the responsibility of a subsection of librarians known to their colleagues as cataloguers or, more recently, bibliographic services librarians. These information professionals create surrogates for individual works, such as books or websites or periodicals. From the mid-twentieth century forward, these surrogates have typically been stored in the form of records in a database. The architecture of the databases allows for the retrieval of records based on specific combinations of search criteria, commonly referred to as a “search string.” Two common types of databases used for this purpose are library catalogues and periodical indexes. The retrieval elements within the database records represent access points, which normally include fields for author and title. Subject representation is the part of the surrogate that provides access by topic. The assignment of subject
representation is usually completed through the use of subject headings and classification structures. The topic of the work is labelled so that others seeking information on that topic may find it. Obviously, the process of creating subject representation or naming a piece of intellectual output can be, and usually is, influenced by societal and individual bias. Furthermore, the success of this process is affected by the level of common understanding and vocabulary used by both creators and users. Hope Olson, professor at the School of Information Studies, University of Wisconsin, describes this professional responsibility in her book The Power to Name:

Naming is the act of bestowing a name, of labeling, of creating an identity. It is a means of structuring reality. It imposes a pattern on the world that is meaningful to the namer. Each of us names reality according to our own vision of the world built on past meanings in our own experience. Each of us creates our own structure through naming. Naming is, therefore, not a random process even when it is varied (4).

Olson goes on to describe the naming process as a construction of information rather than the seemingly more modest representation of information (6). I suggest that this construction is most evident in the primary tool used by librarians to assign subject headings: the controlled vocabulary.

On a practical level, prescriptive vocabularies allow for the control of synonyms, homonyms, alternate spellings, and obsolete terminology. This helps create a common language and understanding for both user and creator, improving search precision and recall. Precision and recall in this context are generally accepted measurement ratios used by researchers in library
and information science to evaluate information retrieval systems. In short, they provide quantitative relevancy measurements. It is generally accepted that a controlled vocabulary produces a higher level of precision and recall than free-text searching alone.

Controlled vocabularies also represent the communicative bridge between librarians and their users, and it is with this application that we find an opportunity to educate the user about the organization of knowledge. The work of librarians in this area remains heavily influenced by two nineteenth-century contributors to the discipline: Charles Cutter and Melvil Dewey. Dewey was concerned with classifying the entire universe of knowledge and his work produced, in part, the Dewey Decimal Classification, which remains the most widely used classification system in the world. And while the art of cataloguing dates back to ancient times, Charles Cutter is generally regarded as the first person to codify cataloguing rules. This codification included rules for subject representation: rules generally accepted as being the progenitor of the most influential controlled vocabulary in the world, the Library of Congress Subject Headings (LCSH). Both Dewey and Cutter were concerned with providing access to the universe of knowledge with a universal language. While their intent—reducing confusion and collocating materials—was for the benefit of the user and can be viewed only as altruistic, the biases of the world they lived in continue to persist in their influential tools.

The LCSH is the most widely used controlled vocabulary in North America and, increasingly, throughout the world. It was created more than one hundred years ago and continues to be maintained by librarians at the Library of Congress in Washington, DC. Its purpose is to provide universal access to works received by that institution. Its influence and
power come, in part, from the distribution of “cataloguing copy” or database records, records that represent manifestations of a work. This “copy” is distributed to thousands of academic, public, and special libraries around the world and includes subject assignments. As a result, a high percentage of library catalogues in North America use subject headings created by Library of Congress librarians as intellectual access points. It is also important to understand that the driving principle behind the creation of LCSH is that of “literary warrant.” Literary warrant is, quite simply, the idea that terms used in the materials themselves be used as access points in the surrogates of those same materials. Thus the derivation of terms to be included in LCSH is based on an empirical process. Of course, this process is not without its practical challenges. For example, it may take years before a phrase or term is represented in the literature to an extent that warrants its inclusion as a subject heading. Also, while retroactive changes are made to catalogue records held at the Library of Congress, changing the copy that has been distributed to other libraries is the responsibility of the receiving institution and such changes are all too often considered to be too expensive for implementation. When one considers the type and content of literature produced in North America prior to the late twentieth century, it is not surprising that, for many years, “mankind” was the subject heading that represented the concept of “humankind” in LCSH. It is a bit more disconcerting to learn, however, that “man” was not replaced with “human beings” in LCSH until 1996 (Gerhard 136). This and many other improvements were realized because groups of librarians worked with their colleagues at the Library of Congress to affect changes in the terminology. These efforts continue today and in fact have accelerated over the past few decades.
Like most nineteenth- and early twentieth-century controlled vocabularies, LCSH offers a vehicle for a precoordinated system of retrieval. This means that indexing terms are brought together prior to the search and that users must follow the arrangement of those terms when constructing their own searches. Most considered this a necessity in the days of card catalogues, but with advancing technologies came a new system of information retrieval: the postcoordinate system. A postcoordinate system allows the user to combine the indexing terms at the time of the search. Thankfully, such systems have forced the evolution of other tools for lexical control.

One of the most important tools in a postcoordinate information retrieval system is the thesaurus. Thesauri, in this context, represent a system of knowledge organization (Hjorland 104) that provide both user and indexer with a conceptual map of knowledge representing a controlled indexing language. This type of controlled vocabulary is organized so that relationships between terms are explicit, and it is these semantic relationships that make up the units of knowledge organization as they are understood in library and information science. At a practical level, the construction of these relationships is guided by professional standards, most notably ANSI/NISO standard Z39.19, titled Guidelines for the Construction, Format and Management of Monolingual Thesauri. Common mappings outlined in the standard include hierarchical, equivalency, and associative relationships, better known as broader and narrower terms, “used for” and “use” conditions, and “see also” references. Given this utility, the terms used by the indexer are matched with terms common to, and employed by, the user. This creates a higher level of relevant retrieval and, if employed appropriately, improves precision and recall. This area of thesaurus development is closely linked to advances in information science. In addition, it builds on thesaurus design features that date back to eighteenth-century attempts to
standardize the English language. Other influences include mid-twentieth century work around faceted thesauri and, later still, the public services-oriented “person-in-situation” or “holistic cognitive” approach to thesaurus construction.

The pursuit of representing knowledge in an accurate yet universal manner has driven the creation of many different tools. Library of Congress Subject Headings have been used to help provide thousands if not millions of library users with consistent and reliable intellectual access to a wide range of works. The evolution of controlled vocabularies includes recent improvements in the design and construction of thesauri, making these tools relevant and useful in an online environment. Both provide a reflection of a specific and subjective conceptual reality, underscoring the fact that the power of lexical control is not something to be undervalued. The importance and power of defining the world around us was considered at all stages of developing the library component of the GWD.

**Library Tools of the Gender and Work Database**

As a practicing reference librarian, my professional development activities have, until very recently, revolved around the application and utility of tools based on, or informed by, the type of structures noted in the first section of this paper. My goal with the GWD was, and continues to be, to create a tool that builds on these foundations yet is flexible enough to guide users through the discovery and creation of knowledge within the study of gender and work. Achieving this goal required careful analysis of the problems associated with existing tools for lexical control. It also necessitated striking a balance between the needs of the individual module and its experts and the broader, cross-cutting needs of our intended users and the GWD itself. In
the end, we strove to create a conceptual map informed by the quickly evolving study of gender and work that would allow for ease of access in practice.

We began our work by determining the level of access that would be appropriate for the library component of the GWD. My experience working with library users and a review of research in this area convinced us that we would need to apply a controlled vocabulary as part of a postcoordinate system. This decision behind us, we began to look for a suitable controlled vocabulary. My hopes of using an existing vocabulary were short lived.

As noted earlier, the LCSH is one of the most accessible controlled vocabularies, but it is not without its problems. In addition to compounding delays brought on by the steadfast adherence to the principle of literary warrant, there were other practical problems including the level of specificity and coverage offered by the subject headings in Library of Congress records. For example, it is generally accepted that a topic must make up 20 percent of the work before subject assignment will occur. This normally results in the omission of book chapters from the subject headings of edited books, books that are treated at the level of the whole and not the level of contribution. In addition, the “rule of specific entry” states that when assigning subject headings, the term that would be considered the most specific descriptor would be used alone. For example, the subject heading “temporary employment” would be used to describe a work on that topic; the broader subject headings of “employment” or “precarious employment” would not be used. Therefore if a researcher searched only for “precarious employment,” he or she would not find work that was primarily about temporary employment even though, conceptually, “temporary employment” is a component of “precarious employment.”
Other perhaps more compelling problems involve a series of documented issues related to the bias of LCSH. Research published by Olson in 1991 demonstrates that the LCSH is a reflection of a patriarchal society. Specifically, she refers to sex-based assumptions and a lack of inclusive language reflected in the choice of terminology. Gerhard, Su and Rubens, all practicing academic librarians, drew on the work of Olson and others to publish their 1998 article entitled “An Empirical Examination of Subject Headings for Women’s Studies Core Materials.” Both works raise issues related to sexist language, complexities arising from the interdisciplinary nature of women’s studies materials, and inconsistent levels of assigned subject heading specificity. One quick but classic example, also used by Olson, is that of the LC subject heading “Woman – Theology” being placed as a narrower phrase to the heading “Man – Theology,” implying that the former is a subset of the latter and, some would argue, not of equal standing. In a later work, Olson and her research assistant Rose Schlegl analyze the level of systemic bias found in subject access standards as discussed in library and information science literature. Building from earlier critiques of subject access for marginalized materials, Olson and Schlegl identify five types of negative bias: first, treatment of the topic as an exception (e.g. “women as physicians” as an exception to the broader “physicians,” implying that such an occurrence is an anomaly); second, the ghettoization of the topic (i.e. gathering and isolating the topic rather than integrating it); third, the omission of the topic (i.e. a lack of currency within subject access standards); fourth, an inappropriate structure of the standard (e.g. poor allocation of space in a classification system or subject hierarchy to cultures that are not of European or white North American origin), and finally, biased terminology. Their work also identified that when subject
headings were the standard of focus, terminological bias was the most prevalent problem (Olson and Schlegl 66-70).

Thankfully, Gerhard, Su, and Rubens identified methods that could help remedy some of these problems. Olson’s later work, published in 2002, also builds on solutions to these problems. These solutions include efforts to continue working with librarians at the Library of Congress to alter existing controlled vocabularies. And while changes have been made to some of the most egregious problems, it quickly became apparent that the current state of this vocabulary was not going to be adequate for our needs with the GWD. Furthermore, the discovery that other controlled vocabularies, specifically those dedicated to women’s or labour studies, were equally problematic was also a disappointment. It needs to be stressed, however, that obsolescence was the primary concern with the discipline-specific thesauri. In the end, it was clear that we would need to construct our own subject access headings if we were to remain true to the goals of the GWD. Naturally, we strove to learn from existing problems before we began our work. To this end, Olson and Schlegl offered three principles for practicing librarians: first, users must not be regarded as homogeneous; second, striving for objectivity can result in separate but equal treatment when equitable treatment tailored to the specific context is preferred, and third, standards lose their effectiveness if they are not carefully and equitably applied, including adaptation to a specific context (78). These principles, and a desire to remain user-focused and consistent, drove our work around the creation of the GWD thesaurus, our controlled vocabulary and system for knowledge organization.
Constructing a thesaurus demands the consideration of many factors, including an understanding of the end-users’ needs, their habitual vocabularies, and their likely interpretation of word orders. This includes an awareness of both the individual user’s construction of knowledge and the broader context within which that construction takes place e.g. sociocultural and historical systems. While these factors were considered, our thesaurus terms and concepts were collected using a deductive method whereby I relied heavily on input from our module experts, individuals whose work often involves studying the same population we are trying to represent within the GWD. Traditionally, librarians have attached great importance to conducting their work in a neutral and objective manner. In contrast, my work on this project often involved deferring to the module experts’ decisions about the ways in which knowledge should be labelled, a practice recognized, although not always supported, in the library literature.

Our users are our current and future researchers and I have experienced first hand the blinding frustration caused by confusing or ambiguous language within controlled vocabularies. As a reference librarian, I recognize that frustration as an often impenetrable barrier to learning and research, a situation I am keen to avoid. As you can imagine, striving to balance literary and scholarly warrant with structural and use warrant within one database analyzing many different slices of the intersection of gender and work provided opportunities for numerous fruitful discussions about the selection of appropriate terminology. The language of labour unions is markedly different than that of the health care industry or migration yet we strove to provide access to information about all these areas, and their related statistical tables, with a single vocabulary. Our goal was to achieve standardization through consensus, and the results are
available to you through the thesaurus. I am sure it will look quite different five years from now, given that we are attempting to map an area of knowledge that is rapidly changing.

**Summary**

Our goal has been to provide access to a body of knowledge that has, until this point, been out of reach to all but the most ardent and committed of library users. Our work has been built upon previous research in library and information science. It has also been guided by both the successes and challenges of the English-speaking world’s most widely used controlled vocabulary, the LCSH. Manipulating vocabulary to show relationships and essentially construct maps of knowledge is exciting and influential work. The goal of the thesaurus is two-fold: to facilitate understanding at a conceptual level by communicating on a linguistic level, and to build a vocabulary that can be used to create intellectual access points for the topic of a manifestation of intellectual output. Put another way, the thesaurus supports information retrieval and the production of knowledge.

In this paper, I have purposely excluded many elements that were required to make the library component of this project a reality. Compiling and managing the thesaurus itself would not have been possible without freeware created and kindly distributed by Timothy Craven, a professor in the Faculty of Information and Media Studies at the University of Western Ontario. In addition, we are fortunate to have a system where we enjoy a high level of interoperability, that is, you can search the thesaurus for terms and then, with the quick click of a mouse, conduct
a search for relevant materials in our library database. This did not happen without the dedication and expertise of our technical, and sometimes unpaid, staff. Other omissions from this paper include reference to research in the discipline of library and information science related to the information-seeking process, collections management, service provision, linguistics, the communication process, and the actual methods of knowledge organization themselves e.g. indexing and subject analysis. Methodology specific to knowledge organization is rooted in fundamental theories of epistemology, a discussion of which goes beyond the introduction that this paper was created to provide. However, a bibliography that will assist those interested in learning more about the contributions of library and information science to this area of research follows the list of works cited in this paper. For now, it is important to note that these areas will be drawn on as work on the GWD library progresses, including the evolution of the thesaurus itself. Finally, I am hopeful that our efforts will go beyond the immediate users of the GWD and will be ultimately considered by other librarians, perhaps even the librarians at the Library of Congress, as a resource for improving existing controlled vocabularies.
Works Cited


Selected Bibliography


