Interdisciplinary Searching as a Use Case for Vocabulary Mapping

1.0 Introduction

The ISKO 2020 conference theme of “Knowledge Organization at the Interface” relates to the challenge of facilitating searching at the interface among disciplines. As Dextre Clarke (2019) and Zeng (2019) make clear, the publication of ISO 25964-2: Information and Documentation - Thesauri and Interoperability with Other Vocabularies - Part 2: Interoperability with Other Vocabularies (ISO 2013) brought into focus the principles and practice of vocabulary mapping. But given the effort required to accomplish mapping among two or more vocabularies, it is important to ask, echoing Ford (2019) in his advocacy for use cases, “Who will use this and why?” This paper explores interdisciplinary searching as an important use case for vocabulary mapping.

There is increasing recognition of the importance of interdisciplinary research, as noted by Gibson (2012, 213): “The structures of the academy are straining toward integration across disciplines in order to solve transdisciplinary problems and ‘grand challenges’. ” However, as Palmer and Fenlon (2017, 429) observe, “we are far from realizing the potential of information systems and services for fueling interdisciplinary research.” In order to better support interdisciplinary research, there is a need for continuing evolution of discovery tools.

But what type of research does the label “interdisciplinary” encompass? Wasserstrom (2006) notes a range of possibilities, including: team-based interdisciplinarity, to which several scholars bring different skills; cross-over interdisciplinarity, referring to fields like bioethics which have roots in two disciplines; and exploratory interdisciplinarity, for scholars who apply material from other fields on occasion. The term may also encompass explicitly interdisciplinary fields such as area studies, ethnic studies, gender and women’s studies, and environmental and resources studies. As explained by Linköping University’s Institutionen för Tema (Department of Thematic Studies), which encompasses Child Studies, Gender Studies, Environmental Change, and Technology and Social Change:

In an increasingly complicated world, the need for interdisciplinarity is greater than ever, researchers that can see both breadth and depth in major societal issues. At the Department of Thematic Studies, natural science, social science, technology and the humanities meet in the common aim to increase
understanding for and find solutions to important future questions. (https://liu.se/en/organisation/liu/tema)

In this context interdisciplinary research attempts to integrate insights from multiple disciplines to address a problem that defies explanation by a single discipline. This has implications for education as well. A recently published report (National Academies of Sciences, Engineering, and Medicine 2018, x) observes that there is a concern that “an education focused on a single discipline will not best prepare graduates for the challenges and opportunities presented by work, life, and citizenship in the 21st century” and recommends “an approach to higher education that intentionally integrates knowledge in the arts, humanities, physical and life sciences, social sciences, engineering, technology, mathematics, and the biomedical disciplines”.

Because the most widely used knowledge organization systems were developed “when a discipline-based view of the universe of knowledge was common within both information science and the wider academy” (Szostak, Gnoli, and López-Huertas 2016, 1), interdisciplinary researchers must navigate an array of information resources that have tended to become fragmented and specialized, each often with its own distinct controlled vocabulary. In this environment interdisciplinary searching becomes “an arduous undertaking for the end-user” (McCulloch 2004, 298).

Given this context, in this paper consideration of interdisciplinary searching as a use case for vocabulary mapping includes: (1) discussion of the challenges posed by interdisciplinary searching and the need for continuing evolution of discovery tools; (2) analysis of the guidance provided by ISO 25964-2 in creating vocabulary mappings that could aid in interdisciplinary searching; (3) approaches to assessing the value added by vocabulary mapping from a user perspective.

2.0 Challenges of interdisciplinary searching

The most recent edition of the text Interdisciplinary Research: Process and Theory (Repko and Szostak 2020) includes a chapter on “Conducting the Literature Search” and observes that subject-oriented indexes and databases are of particular interest in this process. The reader is cautioned that each of these has its own thesaurus and is advised: “As researchers move from discipline to discipline in search of different insights on the same topic, they need to check each discipline’s thesaurus to find the term(s) to search for” (136).

The recognition of the need to support interdisciplinary searching is not new (see, for example, Smith, 1974; Weisgerber, 1993), but there have not been a large number of studies of interdisciplinary information seeking reported in the literature. Some key findings of completed studies include:

Bates (1996): “scholars in interdisciplinary fields may have to engage in both substantially more information seeking—and of a different kind—than scholars in a conventional discipline” (159). Information scatter contributes to the problems faced by interdisciplinary researchers.

Palmer (1996): “information probing” is an important type of information work for interdisciplinary researchers. “Probing is investigative in nature and takes place outside of the scientist’s core knowledge domain” (169). With each new domain, there are new terms and concepts to learn and analytical approaches to understand.
Spanner (2001): “One of the most frequently cited problems in crossing over into other disciplines is the problem of acculturation to non-affiliate disciplines, particularly in adjusting to conflicting vocabularies” (355).

In summary, the challenges confronting those engaged in interdisciplinary searching include the scatter of potentially relevant literature, some of which is outside the searcher’s core knowledge domain. More databases have to be searched, which involves having to navigate conflicting vocabularies.

3.0 Vocabulary mapping

Given the challenges posed by interdisciplinary searching, there is a strong interest in mechanisms to achieve subject or semantic interoperability (Zeng 2019). Dextre Clarke and Zeng (2012, 22) provide a timeline of landmark thesaurus standards in the English language, culminating in a new two-part international standard (ISO 2011; ISO 2013) developed by a working group with members from 15 countries, a chairman from the United Kingdom, and a Secretariat run by the National Information Standards Organization (NISO) in the US. As the title of this international standard, Thesauri and Interoperability with Other Vocabularies, makes clear, there is a strong focus on interoperability, “the goal of taking vocabularies which in most cases were intended to stand alone and relating them to each other in sufficient detail to permit searches drawn from one vocabulary to be effective in another” (NISO 2005, 82).

The Basel Register of Thesauri, Ontologies & Classifications (http://bartoc.org/) makes visible the wide range of extant thesauri, listing 766 ranging from The Art and Architecture Thesaurus to the Zoological Record Thesaurus. Where such thesauri are used as indexing vocabularies for specific discipline-oriented databases, a mapping can allow translation of a search statement from its initial formulation in one vocabulary to an equivalent statement in other vocabularies, where such correspondences exist. If vocabularies have only partial overlap in subject scope, mappings can be established only for those concepts covered in common, simplifying the task of mapping. Creating such mappings enhances the possibilities that formerly disconnected databases will be used in combination, especially to explore interdisciplinary topics. The goal is to enhance the ability of interdisciplinary researchers both to find what they want and to discover related information that they would not have known to look for.

As Zeng (2019) explains, the principles and practice of mapping are the prime focus of ISO part 2 (2013). The scope includes interoperability of thesauri with each other as well as with classification schemes, taxonomies, subject heading schemes, ontologies, terminologies, name authority lists, and synonym rings. Mapping establishes relationships between the concepts of one vocabulary and those of another, but challenges arise because vocabularies can differ with regard to structure, domain, language, or granularity.

Turning to the standard itself, Clause 5 Objectives and identification explains that the purpose of mapping is “to enable an expression formulated using one vocabulary to be converted to (or supplemented by) a corresponding expression in one or more other vocabularies” (16). When overlap between two vocabularies is small, selective mapping is carried out (Clause 6.5, 19). Three categories of mappings are distinguished: Equivalence mappings (Clause 8), Hierarchical mappings (Clause 9), and Associative mappings (Clause 10).
Equivalence mappings (EQ) encompass simple equivalence, 1 to 1 matching of concepts found in two or more different vocabularies (e.g., mobile phones EQ cell phones), and compound equivalence (1 to many). In compound equivalence a preferred term in one vocabulary may be represented in another vocabulary by a combination of two or more concepts/terms. The standard distinguishes between intersecting compound equivalence (e.g., genetically modified wheat EQ genetic modification + wheat) or cumulative compound equivalence (e.g., inland waterways EQ rivers | canals). In the case of exact equivalence, terms may be used interchangeably. In Clause 11 the standard also makes provision for identification of inexact equivalence (~EQ) when the most closely matching concepts in two or more vocabularies are not exactly the same. Such a mapping may lead to additional relevant items without too many irrelevant ones.

Hierarchical mappings (BM Broader Mapping/NM Narrower Mapping) between concepts are established when one is clearly broader than the other (e.g., rats BM rodents; rodents NM rats).

Associative mappings (RM Related Mapping) between concepts indicate situations where concepts do not qualify for equivalence or hierarchical mappings, but are semantically associated to such an extent that documents indexed with the one are likely to be relevant in a search for the other (e.g., e-learning RM distance education).

Clause 14 identifies how such mappings are accomplished: “Traditionally the identification of mappings is an intellectual process. It needs one or more experts familiar with the relevant subject field(s), fluent in the language(s) of the vocabularies to be mapped, and having a good understanding of the structure and conventions of the vocabularies” (38). Computer assistance may be used to automate the process in part by employing a matching algorithm and presenting candidate mappings for review by an expert. Because thesauri may continue to be revised and updated, Clause 15.3 highlights the maintenance that is needed if there are changes that affect the validity of the mapping.

Once created, Clause 12 discusses the uses of mappings in information retrieval as part of the indexing process or at the time of search. In interdisciplinary searching, mappings could be applied automatically in extending a search from one database to another, or human mediation could be used to select among options for search expansion. At a minimum equivalence relations enable translation of search terms from one thesaurus into those of another. Inexact equivalence, as well as hierarchical and associative mappings, offer possible avenues for search expansion.

Clause 16 discusses displays of mapped vocabularies, noting that the standard “does not seek to constrain the presentation of mappings data to end users” (45). The user need not be aware of the mapping process to make use of displays. For example, mapped terms “may be presented in the style of a tag cloud...without explicit designation of the type of mapping in each case.” (45).

4.0 Evaluation of mapping

Mappings connecting concepts in two or more thesauri will vary in the proportion of concepts involved, depending on the subject scope of each thesaurus. But even selective mappings require intellectual effort (and cost) to develop and maintain. As Kemp (2018, 82) notes, if metadata creators do not understand how metadata is used, it can be difficult
to make the case for enriching it. Research by Hider, Mitchell, and Parkes (2019) investigated the “retrieval power” added by subject indexing when compared to searching free text in a database and concluded that professional indexing using controlled vocabularies enhanced the yield of relevant resources. This approach can be extended to consider the retrieval power added by extending a search from the subject indexing employed by one database to the subject indexing via mapping in one or more additional databases when conducting an interdisciplinary search. The effectiveness of the mapping in retrieval can best be measured by submitting it to tests in the form of subject searches since its intent is to facilitate improved response to such queries. Such tests typically use measures of recall and precision based on relevance assessments of retrieved documents to gauge performance.

One exemplary study in this vein was carried out by Mayr and Petras (2009), who sought to assess the performance of a German Federal Ministry for Education and Research terminology mapping project, creating selective mappings (“cross-concordances” in their terminology) among multiple controlled vocabularies. Noting that while many mapping projects are undertaken, “the actual effectiveness and usefulness of the project outcomes is rarely evaluated stringently” (Mayr and Petras, 2009, 47), so they sought to determine how effective and helpful the mappings are in actual searches. Their experiments show the positive effects of mappings for search in heterogeneous databases, with interdisciplinary mappings having a higher positive impact on search results (Mayr and Petras 2009, 51). An early study by Smith (1974) demonstrated that using a mapping to extend searches to other databases beyond the major medical database for topics falling within nuclear medicine, bioengineering and computer applications, and physiology and biophysics yielded an increase in the variety of potential relevant resources by including technical reports, conference proceedings, and additional journals. In his expansive review of the role of thesauri in new information environments, Shiri (2012) found that study results demonstrate the usefulness of thesauri in both providing users with alternative search terms for query expansion and improving retrieval performance.

Sunny and Angadi (2018) sought to systematically review the literature evaluating the effectiveness of thesauri in digital information retrieval systems. In addition to studies that reported positive effects of an online thesaurus on retrieval performance, they found several studies focused on the usability of online thesauri. These studies reported positive reactions of participants in terms of identification and use of thesaurus terms; use in searching, browsing and navigation; and ease of use and helpfulness (65).

While evaluations of mappings between thesauri have been limited in number to date, the studies of single thesauri suggest needed directions for evaluation of mappings as they are developed and implemented, addressing both retrieval performance and usability. For interdisciplinary searching, semantic interoperability via mappings not only increases the success chances for distributed searches over databases with different thesauri, but it also can provide a view of a different disciplinary framework and domain-specific language, if the mapped vocabularies are made visible. Precision and recall of retrieval could be enhanced if mappings are differentiated using all the types of mapping described in Clauses 8 to 10 and if degrees of equivalence are marked as recommended in Clause 11. Especially where a given concept has no exact equivalent, provision of
Inexact, broader, narrower, and associative mappings could be helpful in enabling selection of the best option for cross-browsing and cross-searching of various databases. In such query expansion, the user’s initial query statement is enhanced by additional search terms in order to improve retrieval performance.

In their discussion of new approaches to interdisciplinary knowledge organization, Szostak et al. (2016, 216) outline an ambitious research agenda to answer such questions as: What success do users have with the sorts of complex queries that interdisciplinary users often have? What about queries that members of one group might make about the practices or beliefs of other groups or disciplines? Evaluations of mappings can assess their efficacy in such situations.

5.0 Conclusion

The need for support for interdisciplinary searching is clear: “Search and communication absorb a great deal of the interdisciplinary researcher’s time. And failure to identify relevant information limits scholarly discovery” (Szostak et al. 2016, 220). Now that standards have formalized approaches to achieving semantic interoperability through mapping thesauri, next steps are implementation of more mappings and evaluation of their efficacy and usability. Echoing Szostak et al. (2016, 220), the goals of these efforts are to:

- Facilitate interdisciplinary searches by interdisciplinary scholars and students, both when they know what they are looking for and when they are seeking novel connections.
- Clarify terminology across disciplines.
- Facilitate the communication of research results to all relevant audiences.

As more thesaurus-enhanced user interfaces are implemented as search and browsing tools in a broad range of systems (Shiri 2012), from bibliographic and full-text databases to digital libraries, portals, open archives, subject gateways, and linked data repositories, the potential benefits of embedding mappings will increase. To take an example related to the conference theme of Knowledge Organization at the Interface, consider the researcher who is exploring approaches to enhance access to the Web for users with disabilities. Possible databases within scope include Inspec and Library and Information Science Source. In Inspec one finds the terms “handicapped aids” and “user interfaces”. In Library and Information Science Source one finds the terms “assistive computer technology” and “user interfaces (computer systems)”. A mapping of concepts common to Inspec and Library and Information Science Source would facilitate exploration of the computer science and engineering as well as library and information science perspectives on this topic. Extending the mapping to search PubMed via Medical Subject Headings could provide links to “communication aids for disabled” and “user-computer interface”. Given such a mapping, an evaluation study could involve the researcher in assessing items retrieved in response to a search request mapped across the three thesauri/databases for relevance. A mapping that included hierarchical and associative relationships that could be viewed by the researcher for possible search refinement or expansion (e.g., in Library and Information Science Source, “assistive computer technology” has “computers & people with disabilities” as a BT and “accessible websites for people with disabilities” as an RT) could be assessed for usability as well as retrieval performance.
In “As We May Think” Bush (1945) envisioned the memex, an information system with “a mesh of associative trails”. In the digital Web-based environment of 2020, vocabulary mapping offers one approach to creating a mesh of associative trails among a broad range of systems employing the thesauri included in the mapping. While this holds promise of facilitating interdisciplinary research, more evaluation is needed in order to assess its full potential.

References


