Bridging the Gap Between Search and Engineering Domain Content: A User Focused Assessment of Controlled Vocabularies

by

Ashleigh Pirone

Bachelors of History, Edinboro University, 2009

Submitted to the Graduate Faculty of the

School of Computing and Information in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF COMPUTING AND INFORMATION

This dissertation was presented

by

Ashleigh Pirone

It was defended on

November 12, 2021

and approved by

Judith Brink, Head of Bevier Engineering Library

Daqing He, Professor, Associate Chair

Marcia Rapchak, Teaching Assistant Professor

Dissertation Director: Alison Langmead, Associate Professor

Bridging the Gap Between Search and Engineering-Domain Content: A User-Focused Assessment of Controlled Vocabularies

Ashleigh Pirone, PhD

University of Pittsburgh, 2021

This Search, revise, and repeat. This is often the cycle users find themselves in when searching through repositories that aggregate dispersed information. Unfortunately, frustration is more often the result of these searches than success. To be faced with a repository of information and not find the desired information is not a new dilemma. In the past ten years, literature has shown that knowledge organization has suffered from outdated methods and practices, some ramifications of which being an emphasis on browsability of information rather than searchability, outdated or offensive terminology, underestimating the user's natural language, and emphasis on using the "right" subject in search, rather than the words that are natural to the user, which is a more user-centered design. Most scholars advocating for change within the library classification space, do not suggest doing away with subject indexing, but updating the indexing process to better accommodate users' expectations for a "Google-like search" without loss of the specificity of precision offered by subject indexing. This study explores to what degree controlled subjects differ from users' natural language, finding through the goodness-of-fit chi-square assessment that overall the terminology does not match users' language in the subset of engineering-domain control terms analyzed. The findings from the user-focused card sorting survey are then used to suggest updates to the controlled terms to better align them with users' natural language and potentially increase engineering researchers' search effectiveness and satisfaction.

Table of Contents

1.0 Introduction 1

1.1 Inspiration 2

1.2 Exploration of the Problem 3

1.3 Research questions 12

1.4 Overview of study design 14

2.0 Review of Literature 19

2.1 Review method and structure 19

2.2 Tacit versus explicit knowledge 22

2.3 Information science in knowledge management 24

2.4 Empiricism version pragmatism 28

2.5 KO and the engineering community 35

2.6 Knowledge management ontology structures 47

2.7 Current state of controlled and uncontrolled terms IR 55

2.8 Summary of Reviewed Literature 74

3.0 Outline of Methods 78

3.1 Warrant as a framework 78

3.2 Overview of study design 82

3.3 Population, sample size, and distribution 84

3.4 Hub or knowledge graph mapping model 87

3.5 Chi-squared goodness-of-fit Assessment 92

3.6 Dataset creation method 96

3.7 Card sorting method 100

3.8 Position of the researcher and ethical considerations 104

4.0 Data Collection and Analysis 106

4.1 Cleaning and preparing the dataset 106

4.1.1 Establishing a mapping technique and target vocabulary 107

4.1.2 Preparing Control Dataset 114

4.2 RQ1: Control data matches to controlled vocabularies 121

4.3 RQ2: Controlled vocabulary matches (scientific warrant) 124

4.4 Creating User Natural Language Dataset 129

4.5 RQ3: Users' Natural Language matches to Control Terms 136

4.6 RQ3 Initial Findings: 153

5.0 Discussion of Findings 155

5.1 Discussion on Study Design Methods 155

5.1.1 Lack of machine learning terminology resources 156

5.1.2 Lack of participation in survey 157

5.1.3 Lack of methods for controlled vocabulary analysis 158

5.2 Answering the main research question 160

5.2.1 Summary and Codification of Findings 160

5.2.2 Search compared to Browse Pattern 162

5.2.3 Exploring the main causes for misalignment 163

5.2.4 Additional Suggestions for More Precise Subject Search 185

5.2.5 Additional Observations from Findings 187

5.3 Options for User' Natural Language in Controlled Vocabularies 188

6.0 Conclusion 190

6.1 Summary of Findings 191

6.1.1 Additional Suggestions of Improvement: 193

6.2 Positioning the Findings 195

6.3 Further Research 197

Appendix A 200

Bibliography 201

List of Tables

Table 1: Overview of the Study Design..... 15

Table 2: Wittgenstein's knowledge organization. 33

Table 3: Knowledge Organization Literature Analysis. 52

Table 4: Natural Language Processing Levels. 63

Table 5: Studies on natural language user tags. 65

Table 6: Overview of the Study Design. 81

Table 7: Overview of the Study Design- Detail 83

Table 8: Sample assessment of terms to determine RQ1-4 findings. 95

Table 9: Examples of Match Categories for Mapping. Researcher created. 110

Table 10: Mapping for Baseline. Researcher created..... 113

Table 11: Control dataset coverage across databases. Researcher created. 117

Table 12: Mappings Control terms and controlled vocabulary terms..... 120

Table 13: Control to Source match summary. Researcher created. 121

Table 14: Fuzzy search Chi-Square Assessment for Literary Warrant. 123

Table 15: Browse pattern Chi-Square Assessment for Literary Warrant 123

Table 16: Control to Source Controlled Vocabulary Matches. 125

Table 17: Fuzzy search Chi-Square Assessment for Scientific Warrant. 126

Table 18: Browse pattern Chi-Square Assessment for Scientific Warrant. 127

Table 19: Summary for Chi-Square Expected Distribution. 129

Table 20: Fuzzy Search Assessment for Classification Rule. 138

Table 21: Browse pattern Assessment for Classification Rule. 138

Table 22: Fuzzy Search Assessment for Classifier Calibration. 139

Table 23: Browse pattern Assessment for Classifier Calibration. 139

Table 24: Fuzzy Search Assessment for Data Streams..... 140

Table 25: Browse pattern Assessment for Data Streams. 140

Table 26: Fuzzy search Ass. for Complexity in Adaptive Systems..... 141

Table 27: Browse pattern Ass. for Complexity in Adaptive Systems..... 142

Table 28: Fuzzy search Ass. for Computational Complexity of Learning. 143

Table 29: Browse pattern Ass. for Computational Complexity of Learning. 143

Table 30: Fuzzy Search Assessment for Deep learning. 144

Table 31: Browse pattern Assessment for Deep learning..... 145

Table 32: Fuzzy search Ass. for Evaluation of Learning Algorithms. 146

Table 33: Browse pattern Ass. for Evaluation of Learning Algorithms. 146

Table 34: Fuzzy Search Assessment for Feature Selection 147

Table 35: Browse pattern Assessment for Feature Selection. 148

Table 36: Fuzzy Search Assessment for Precision and Recall. 148

Table 37: Browse pattern Assessment for Precision and Recall..... 149

Table 38: Fuzzy Search Assessment for Regression. 149

Table 39: Browse pattern Assessment for Regression. 150

Table 40: Fuzzy Search Ass. for Topology of Neural Networks. 151

Table 41: Browse pattern Ass. for Topology of Neural Networks. 151

Table 42: Fuzzy Search Assessment for Weight. 152

Table 43: Browse pattern Assessment for Weight. 152

Table 44: Summary of Chi-Square calculations for RQ3. 153

Table 45: Final Chi-Square calculations RQ3 (fuzzy search). 154

Table 46: Final Chi-Square calculations RQ3 (browse pattern). 154

Table 47: Summary of RQ1-4 Findings. Researcher created. 161

Table 48: Summary of RQ4 Findings. Researcher created. 162

Table 49: Summary of Suggested Updates. 164

Table 50: Summary Ch-Square Assessment for RQ3- Findings are Ho is rejected. 165

List of Figures

Figure 1: Twelve general method types as stated by Szostak (2003). 41

Figure 2: Relationships between controlled vocabulary types. Researcher created. 50

Figure 3: Distinction between taxonomy and ontology. 53

Figure 4: Medical repository using natural language. 70

Figure 5: Taxonomy of Social IR. Recreated with permission. 72

Figure 6: Triple mapping between vocabulary terms. Researcher created. 89

Figure 7: Protege codification of terminology mapping..... 90

Figure 8: Goodness-of-fit Chi-Test Calculation. Researcher created. 93

Figure 9: Goodness-of-fit Chi-Test Calculation. 94

Figure 10: Google Trends volume of search on methodology types..... 98

Figure 11: Closed Card Sorting. Reseacher created..... 101

Figure 12: Open Card Sorting. Researcher created image..... 102

Figure 13: Example Prompt 1 Classification Rule. Researcher created..... 103

Figure 14: Comparison between mapping models. Recreated with permission. 108

Figure 15: All vocabualry matching totals. Researcher created. 111

Figure 16: Protege codification of terminology mapping. Researcher created. 112

Figure 17: Example Search in Google Scholar..... 116

Figure 18: Codification in Protege. Researcher created. 118

Figure 19: RQ3 Response Calculations..... 132

Figure 20: Matching color coding. 134

Figure 21: Partial Match 1 and 2..... 135

Figure 22: Equivalents for Classification Rule. 167

Figure 23: Equivalents for Classifier Calibration..... 168

Figure 24: Equivalents for Clustering from Data Streams. 170

Figure 25: Equivalents for Complexity of Adaptive Systems. 172

Figure 26: Equivalents for Comp. Complexity of Learning. 174

Figure 27: Equivalents for Deep Learning. 175

Figure 28: Equivalents for Eval. of Learning Algorithms. 177

Figure 29: Equivalents for Feature Selection. 178

Figure 30: Equivalents for Precision and Recall..... 180

Figure 31: Equivalents for Regression. 181

Figure 32: Equivalents for Topology of Neural Networks. 183

Figure 33: Equivalents for Weights..... 184

1.0 Introduction

Search, revise, and repeat. This is often the cycle users find themselves in when searching through repositories that aggregate dispersed information. Unfortunately, frustration is more often the result of these searches than success, with estimates of 36% to 45% frustration in knowledge-intensive search environments ("SAE Digital Library Survey," 2014; Field & Allan, 2009). To be faced with a repository of information and not find the desired information is not a new dilemma. In the past ten years, literature has shown that knowledge organization has suffered from outdated methods and practices, some ramifications of which being an emphasis on browsability of information rather than searchability, outdated or offensive terminology, underestimating the user's natural language, and emphasis on using the "right" subject in search, rather than the words that are natural to the user, which is a more user-centered design. Lumley (2015) characterizes this as a shift from a model of knowledge organization

> in which everything is defined as it is, to a contemporary notion of classification as epistemology, in which everything is interpreted as it could be [Mai, 2011]. The challenge for libraries now is how they can contribute to the findability (IR) of documents, given the availability of competing services such as Google, which allow users the flexibility of natural language searching (p. 2).

Lumley (2015) highlights a few key shifts in knowledge organization—namely the search engine and the freedom to use natural language within a search. Lumley, and most scholars who advocate for change in the library classification space, does not suggest doing away with subject indexing, but rather updating the indexing process to better accommodate users' expectations for a "Google-like search" without the loss of the precision offered by subject indexing. To address this challenge, most research focuses on changing the headings themselves or adding alternate

headings. This may be a useful stopgap, but not a solution that can scale to the needs of modern library researchers. This study seeks to explore the extent to which controlled subjects differ from users' natural language, as well as the ways libraries might supplement controlled subjects with users' natural language for more user-centered search.

1.1 Inspiration

The researcher became aware of the engineering community's frustration with knowledge organization in 2015. During a presentation on engineering-domain taxonomy (which turned into an impromptu interview) that the researcher was honored to give to an audience of engineering-domain librarians and more than 400 scientists and engineers, the audience was asked what their biggest frustrations were in finding content. They indicated that the many vocabularies in the various engineering disciplines were difficult to use because it felt like they were built primarily for librarians instead of end-users. The engineers also indicated that the vocabularies often did not make use of the natural language they used in search, that searching journal repositories often did not produce answers or resources to address their queries, and that there were relatively few research aids to help them navigate search in these repositories.

The researcher followed up on this conversation by asking the Special Libraries Association engineering division listserv if these frustrations were felt beyond the sizable community originally asked. The division board and the education branch (about 45 members in all) emphatically replied yes. A few respondents elaborated further: With so much engineering research now available through various journal repositories, such as IEEE's open journals, the methods of retrieving relevant knowledge become more imperative for the entire community. It

made the researcher wonder how much the library community connected their indexing to how modern search engines used that indexing, and how well the indexing serves their users.

The researcher set out to learn more and found many studies that suggest making changes to the traditional library indexing practice. These changes focused primarily on correcting biased or cultural inequity in subject indexing (Hope Olson and Sandy Berman are the most notable), but few suggested enhancing vocabularies with natural language. This study will explore the null hypothesis that controlled vocabulary terms match the users' natural language to the expected match rate, while the alternative hypothesis is that controlled vocabulary terms match significantly less than the expected match rate.

1.2 Exploration of the Problem

Knowledge organization—or vocabulary creation and subject tagging, in the case of this study—is defined by Hjorland (2008) as the act of "document description, indexing, and classification (p.86)." Document description, indexing, and classification are all forms of abstraction, which is the process of "ignoring the inessential details of things" and focusing on the core substance of content (Booch, 2016). The act of distilling a piece of content down to its essential focus, identifying descriptive terms to capture those topics, and then tagging the content with those terms is called "abstracting." Abstracting summarizes what the content is about so it can be retrieved using that information. Abstraction is synonymous with knowledge organization (Hjorland, 2013), which encompasses indexing and controlled vocabulary creation.

When a vocabulary term is created there is a process known as "warrant," which is when a term is created with information retrieval as its primary goal. This information retrieval can be

based on either terminology found in content (literary warrant), the industry-standard terminology often found in controlled vocabularies (scientific warrant), and most important to this study, the terminology focused on how the user would search for information (user warrant). In this study, the term "user-centered" implies that the user is the main focal point for the creation of the tags and the indexing (user warrant). This is a change from the traditional focus on literary warrant, which is more concerned with what is "correct" from the literature or from other controlled vocabularies rather than "what is helpful" to the user's information retrieval practices. The use of warrant to determine if controlled vocabulary terms are user-centered will be more thoroughly explained later in this chapter, but it is important to note that controlled vocabulary terms can be focused on different aspects of information retrieval.

Controlled vocabulary term creation is only one aspect of indexing. The other is abstracting, or tagging the content with the controlled terms. Empirical abstraction has been the norm, focusing on content with similar characteristics, like placing books about Ireland next to each other on a shelf. This can be considered a literary warrant method of abstraction. Pragmatic abstraction, on the other hand, focuses on organizing based on the different approaches a person might take to find the content, and is, therefore, a user-centered approach similar to user warrant. Empirical abstraction usually places content under a primary topic where "the ideal situation is one in which one the heading will suffice to express the subject of the work" (Chan, 2007 p.210). Pragmatic abstraction can assign multiple topics—essentially creating multiple paths for a researcher to find information. Take for instance a search for the article "Irish Tourism: Image, Culture, and Identity." An empirical method would place this piece of content with others focused on Ireland, whereas pragmatic methods could place this article with content focused on Ireland, tourism, culture, identity, or a combination of these.

Empiricism most likely maintains a stronghold in abstraction because it is useful in physical organization, but digital repositories can group content in infinite ways. This may seem to be a nonessential distinction, but empirical organization limits access to knowledge and encumbers tacit knowledge retrieval. But there is a balance to be struck. Explicit knowledge capture and empirical indexing methods may limit a user's information retrieval options, but pure tacit knowledge capture and pragmatic indexing methods—usually instantiated as user-created tags called a folksonomy—may not give enough structure for information retrieval. Both have their use in practice but rarely, if ever, supplement one another. If there were a way to map the controlled vocabularies of the empirical method along with the pragmatic user-centric terminology, this study argues that such a dynamic would help the user in their information retrieval process. More specifically, this study positions knowledge graph mapping via ontology as one method that may be able to bridge the gap.

It is no easy task to gather tacit knowledge from people, often because they do not realize what they know, something Mather and Leeds (2014) call "unconscious competence." Unconscious competence refers to knowledge that a user does not consider sharing because they think it is common knowledge. For example, someone attending a book club meeting might not share which book they read that week with the person sitting next to them because there is a tacit assumption that, since they are in the same book club there is no need to share such information. Addressing unconscious competence requires self-reflection, and communication of that self-reflection, to facilitate knowledge transfer. Vocabulary construction can also suffer from unconscious competence when more common labels for a given subject, such as those found in users' natural language, may not be deemed important enough to share or capture. To address this, ethnographic survey methods in which users are given the opportunity to share their opinions and

interpretations on terminology, particularly the terminology they would use to search for specific knowledge topics, may help to surface unconscious competence and uncover the tacit information that may not be codified in a vocabulary. Mezghani, Exposito, and Drira (2015) created a generic two-step process for capturing tacit knowledge for information organization: The domain-specific community is first surveyed for user-centered terminology, which is then aggregated into a mapping model or knowledge graph/ontology structure that a machine can use to create the mapping between controlled vocabulary terms and users' natural language. This process allows for user-centered terminology—based on the tacit knowledge of the users—to be captured and used in systems that implement semantic search, recommendation engines, and query expansion.

A knowledge graph connects controlled vocabulary terms with relationships. In mapping, these relations can be Match or Partial Match, and in semantics, they can be partOf, hasPart, or isA relations. These relationships are based on tacit information and are usually noted in the framework of the vocabulary and combine to make a web-like structure. There are scenarios where someone may be searching for information only—for example, they want to know how many types of cats are in the genus *Felis*. That information can be represented by a hierarchy of broader and narrower terms. However, if someone wanted to know all the different words used across the globe for "cat," a simple hierarchy is unlikely to help. This is an example of where a knowledge graph is handy.

With a knowledge graph, one subject can have a multitude of relationships to other subjects or synonyms. This type of search is called query expansion, or semantic search, and focuses on expanding the user's query into similar or related terminology, either in the full text or metadata of an article. This is the best method for searching when a user does not know exactly what they are looking for or if they have little knowledge of a particular subject. Knowledge graphs are often

displayed in research interfaces, but they can also be built into the search itself, similar to Google's knowledge graph (Zou, 2020). In scholarly research especially, interfaces that use knowledge graphs can assist researchers in perusing and locating knowledge more efficiently because they do not need to learn a new controlled vocabulary for every database they search, nor do they need to worry if their query is an exact match to the subject tags. This alleviates stress and a frustrating search experience.

Take, for example, a scenario in which someone is searching a repository for information on drones. The term *drone* is an explicit controlled vocabulary term, according to the Transportation Research Board Thesaurus and the Library of Congress Subject headings (LCSH). *Unmanned aircraft* is listed as a synonym in the vocabularies but is not used to abstract and tag content in line with empirical indexing methods (i.e., the content can only be about *drones* OR *unmanned aerial vehicles*, not both). Empirically, *unmanned aircraft* would not be used in information retrieval. This is an issue because, according to a survey of the top engineering digital libraries conducted by the Society of Automotive Engineers (SAE) International, 77% of knowledge seekers in the engineering-domain would expect to find this content tagged as *unmanned aircraft* and not *drones* (SAE Digital Library Survey, 2014). In this situation, the controlled vocabulary term *drone* does not capture the natural language of the researcher. What this means for the engineering researcher is that using their preferred natural language term, *unmanned aircraft*, will likely miss a large portion of content they would have deemed relevant in their search results.

Now, consider another scenario in which a researcher's query is expanded with mapped natural language that allows either *drones* OR *unmanned aerial vehicles* to retrieve content. If both the controlled term *drones* and the natural language term *unmanned aircraft* are mapped together

and have been identified as synonyms of one another, the researcher can search using just one of the terms to find the content they seek—effectively doubling their chances of retrieving the content they are looking for. Also, if a knowledge graph also included semantic relations between subject tags, it would enable the researcher to not only encounter content on drones and unmanned aircraft but also find associated topics like *electric propulsion* or *automatic pilot systems*, or make complex search queries such as content focused on *UAV architecture structures for fly-by-wire applications.*

So far, this study has covered the acts of abstraction and using a graph structure in knowledge organization. Vocabulary terms can be controlled, in which case professionals create and manage the terminology with strict guidelines or with uncontrolled vocabulary terms, where anyone can create the terminology and there are few rules involved with how they are created or maintained—think Twitter tags. Controlled vocabularies such as those of the LCSH and Medical Subject Headings (MeSH) are preferred in knowledge organization because they are consistent, are easily managed, and have a long tradition in information science. Uncontrolled vocabularies, on the other hand, consist of terms that are found in natural language and do not necessarily follow formal spellings, meanings, or grammar—and can include slang. Uncontrolled vocabularies are preferred in applications of self-expression, such as on social media. To facilitate the strongest information retrieval, the users' query (using their natural language), the unstructured text of the content, and the controlled vocabulary tags all need to align to some degree. Natural language is the most important of these aspects because it is the terminology used the most for information retrieval—it is the terminology used in everyday speech and general search engines like Google (Hjorland, 2012). By including natural language in the search for knowledge, organization moves from empirical and explicit abstraction toward a user-centered method of abstraction that deems

the users' natural language just as important as the terms found in the content (literary warrant) and the controlled vocabulary tags (scientific warrant).

Unfortunately, many knowledge organization methods (primarily the empirical methods) do not attempt to capture natural language. If an alignment of terms exists, it is usually not a purposeful alignment with the users' natural language, which creates a situation where the user may not retrieve the information they seek because the words they used are not included in the content (string matches) or the controlled vocabulary tags. This indicates that empirical methods put an emphasis on explicit terminology to facilitate information retrieval over user-centered terms. This suggests that knowledge organization is not user-centered, which is a significant issue considering they are the main users of information repositories. Knowledge organization that fails to accommodate the user's natural language may cause undue frustration and limit scholars' access to knowledge. As Vonda N. McIntyre warned in the novel *Starfarers* (1989), knowledge organization without a focus on human functionality is likely to drown users in irretrievable knowledge. And as Ernest Cline's novel *Ready Player One* (2011) highlights, even in a future with infinite digital information access, if something is not tagged for user discoverability it could be lost forever. To determine if controlled terms are user-centric, this study proposes to examine the extent to which users' natural language matches the controlled terms through an examination of the warrant the terminology aligns most with, and what that might mean for users' search satisfaction.

Knowledge management, indexing, and organizing information specifically have been—and still are—largely manual processes. Starting with S. Ranganathan, G. Vico, and M. Dewey, and progressing into more reformist interpretations from R. Hagler, P. Otlet, and J. Anderson, knowledge organization only started to have a more digital component after pioneers such as Tim

Berners-Lee and Barry Smith explored how to organize knowledge in a digital ecosystem that had the potential to make more information available than ever before (Boyne, 2006; Fricke, 2012; Hjorland, 2011; Mai, 2011; Smiraglia and Heuvel, 2013). In 2008, Smith first wrote about ontologies as knowledge organization structures (defined here simply as how one subject is related to another subject explicitly; though Smith would not agree on such a simple definition). That same year, Berners-Lee published a linked data framework for information sharing (hello world!). Also among these major knowledge organization accomplishments, Gnoli (2008) outlined ten long-term research questions for modernizing knowledge organization. The questions covered how knowledge organization principles can be applied outside the traditional library context, how domain-specific terminology can be applied to epistemological ontologies, how domain-specific terminology can be applied in a multidisciplinary way for cross-domain search, how to scale knowledge organization with knowledge evolution, and who should develop the terms for knowledge organization. These disciplines Smith (medical), Berners-Lee (computer science), and Gnoli (library science) had a lot to say about knowledge organization in 2008. The first two breakthroughs from Smith and Berners-Lee could help answer some of what Gnoli asked, but graph theory was and remains a rarely applied method in library abstraction (excluding linked open vocabularies, of course). Tackling all of the questions posited by Gnoli (2008) is unrealistic. However, this research will explore how domain-specific terminology can be applied to epistemological ontologies in order to map users' natural language—in this case, library users researching engineering-domain topics.

Domain knowledge is a key factor in knowledge organization because it connects the user to information based on their domains' conceptualization and labeling of the topic. Domain knowledge is unique to a particular discipline, although some domains will share terms that may

have different meanings. For example, the word *domain* itself can refer to a territory, a discipline, or an internet identification. The meaning of this term can change depending on the context in which it is being used. Domain-specific controlled vocabularies often share terms with the same string of characters, but most do not map contextual meanings between terminologies, which limits cross-domain search. Some families of terms have limited value to cross-disciplinary search, such as automotive terms to a medical researcher. However, other types of terms, for example those used for study designs and methods of research, have been shown to have great benefit to multiple disciplines (Hjorland, 2016; Smiraglia, 2012). As Bekhuis, Demner-Fushman, and Crowley (2013) state, such terms are crucial for researching and synthesizing the knowledge contained in scholarship and projects across domains. With this in mind, the present study will focus primarily on the engineering-domain and investigate terms that focus on study design and methods. The literature review for this study demonstrates a pre-existing understanding that the user prefers to use their own language in a search and trusts that they will retrieve the research they seek, even if they do not know the "correct" subject headings tagged to that content—especially for terms that are shared across disciplines. This effectively helps them research across multiple domains, find studies that directly pertain to the methods they will use, and identify work that contains the knowledge most pertinent to their research without missing content due to query-subject misalignment.

1.3 Research questions

To assess how well controlled vocabularies in the engineering-domain align with the natural language of its users, and to investigate the ramifications of these findings in light of research discoverability, this study will address the following four research questions:

- **RQ1:** To what extent do expert terminologies match controlled vocabulary terms (literary warrant test)?
- **RQ2:** To what extent do controlled vocabularies match one another (scientific warrant test)?
- **RQ3:** To what extent do expert and controlled vocabularies match users' natural language (user warrant test)?
- **RQ4:** To what extent do controlled vocabularies match users' natural language without fuzzy search also being applied?

A fundamental principle from the library and information science that will be used to organize this study is the use of warrant. There are three main types of warrant: literary, scientific, and user. All warrant types are equally valid, but it comes down to which warrant type can indicate who the indexing is serving during the search process. Literary warrant is most commonly used in subject tagging (indexing) and derives headings and indexing directly from reading the literature and extracting terms that correspond to the terminology of expert researchers (Hulme, 1911; ANSI/NISO Z39.19, 2005; Bullard, 2017). By comparing controlled vocabularies to the literature written by scientific experts in the field (RQ1), this study will explore the extent to which controlled vocabularies are derived from literary warrant. Scientific warrant is focused on controlled vocabularies—usually derived from the consensus of terminology from a domain—and is often associated with standardized terminology from a given domain or discipline (Bliss, 1929). By comparing controlled vocabularies to each other (RQ2), this study will explore the extent to which the vocabulary terms are derived from scientific warrant. User warrant is derived from the most prevalent terms from users and is centered capturing the users' terminology for search

(Ranganathan, 1964). By comparing controlled vocabularies to the user's natural language (RQ3), this study will explore the extent to which controlled vocabularies are derived from user warrant. This study is focused on user terminology and how well the engineering-domain controlled vocabulary serves the users in this space. Identifying the type of warrant will be the mechanism used to determine if the terminology is based on the literature, the consensus of the controlled vocabularies, or the user.

Tagging content serves as a bridge between the search engine and the content for the user. Controlled tags are one of two pieces of metadata that Google does not have (the other being abstracts) on its scholarly platform, Google Scholar. This sets libraries apart from Google. As leading cataloger and classification expert Lois Mai Chan (2007) states,

> At the end of a subject search on the Internet using full-text searching-- the most common Internet search mode-- the user is presented with a list of the leading documents in an enormously long list of links to information sources that contain his or her query term. There is no mechanism for telling that user of other documents on the same topic whose authors use a synonym for the submitted term. Furthermore, the number of hits in the list of retrieved documents is often at least in the thousands...databases designed on similar patterns of library catalogs result in a smaller number of hits and a higher on-target rate [called precision] (p. 11-12).

Chan goes on to explain the differences in bibliographic control, or the standardization of the document metadata that can be used to filter search results more precisely. The top two elements she mentions are title and subjects. Titles are a known item search—the book title is exactly what the user searches for and is what the user expects to retrieve—but subjects are an unknown item search. The user does not know exactly what to expect from their search, and they will get little indication if they used the "wrong" word to retrieve their content. If subject tags do not successfully connect the users' search to the content via the tags on said content, the research is not as successful and users get frustrated.

Worse yet, they may miss important information because they did not search with the "right" word. This research will shed light on how strong the bridge between users' queries and content is, and if it is found to be lacking, how controlled vocabularies can be supplemented to make the bridge stronger.

1.4 Overview of study design

Using the different types of warrant as a framework for identifying if a subject tag is focused on the literature, other controlled vocabularies, or the user, I have developed a three-step research plan: First, a baseline is established to understand (1) what specific vocabulary terms are candidates for analysis (creating the Control list for the remainder of the study), (2) which vocabularies serve as the target to be mapped to, and (3) assessing if the observed distribution matches what is expected from the baseline with a chi-squared goodness-of-fit assessment for each term and an aggregate of all terms. Then, the Research Questions are explored to investigate the extent to which controlled terms match those from the literature (RQ1 determining literary warrant), how controlled vocabularies match one another (RQ2 determining scientific warrant), and how well engineering-domain controlled terms match users' natural language (RQ3 determining user warrant). RQ4 is assessed throughout to determine if a browse or fuzzy search pattern is better for bridging the user's natural language and the controlled vocabulary terms in a search. An overview of the study design can be found in Table 1.

Table 1: Overview of the Study Design

<table>
<tr><td colspan="3">H_0= Controlled vocabulary terms match the users' natural language to the expected match rate.
H_a= Controlled vocabulary terms match significantly less than the expected match rate.</td></tr>
<tr><td>Step</td><td>Focus</td><td></td></tr>
<tr><td>Pre-stage</td><td colspan="2">Establish which terms will be assessed throughout the study, selecting engineering vocabularies for Control terms, and establishing a baseline expected distribution of matches to be used throughout study</td></tr>
<tr><td>Step 1 (RQ1)</td><td>RQ1: To what extent do expert terminologies match controlled vocabulary terms?

Assessing literary warrant via literature-to-control term matching</td><td rowspan="3">RQ4: To what extent do controlled vocabularies match the user's natural language without fuzzy search also being applied?</td></tr>
<tr><td>Step 2 (RQ 2)</td><td>RQ2: To what extent do controlled vocabularies match one another?

Assessing scientific warrant via controlled vocabulary matching</td></tr>
<tr><td>Step 3 (RQ 3)</td><td>RQ3: To what extent do expert and controlled vocabularies match users' natural language?

Assessing user warrant via user to control term matching</td></tr>
</table>

The findings from each RQ build on each other, first to help establish the chi-square expected distribution, and second to understand the current state of the types of warrant used in controlled vocabulary creation and indexing. Identifying the type of warrant being used will be the primary way this study measures what the focus of the subject term is—literature, other controlled vocabularies, or users for RQs 1-3. The match rate from a baseline vocabulary mapping, as well as the match rates from RQ1-2, will be used to determine the chi-squared goodness-of-fit expected match rate for answering the main null hypothesis of this study (RQ3). If the RQ match rates in RQ1-3 fall below the expected match rate, the null hypothesis is rejected. If the match rates are significantly lower than the expected match rate, the alternate hypothesis is accepted. In addition to the chi-squared goodness-of-fit assessment, additional underlying methods will include social discourse theory, user-centered design, and card sorting (Hider, 2015; Hobbs, 1996; Lambropoulos & Zaphiris, 2007).

Each stage follows a similar approach, in which controlled vocabulary terms are compared to a source dataset—Encyclopedia of Machine Learning and Data Science (Phung, Webb, and Sammut, 2020) (RQ1), engineering-controlled vocabulary (RQ2), and surveyed user entries (RQ3)—and are then deemed to be an exact match (EM), a partial match (PM), or a no match (NM) by the researcher. These are recorded in an Excel table format for use throughout the study. Another supplemental dataset that is more machine- and data-focused was produced from WebProtege, a knowledge graph modeling tool that allows knowledge graph files (OWL/RDF) to be exported; vocabulary formats such as SKOS; and CSV for other common data analysis tools. This is housed on Bioportal as a creative commons dataset for scholarly reuse, located here: https://bioportal.bioontology.org/ontologies/MLTX

Once the matches are identified and the logic for determining a match is noted, a goodness-of-fit test is used to understand how likely the control is to match the source for each RQ assessment. For each RQ assessment, two chi-square tests are performed to determine if browse or search patterns are more successful for matching the source to the control terminology (RQ4). In this case, the first test where fuzzy search counts EM/PM as a match, and the second test where browse counts only EM as a match. Because vocabularies are only the container for the individual subjects, and individual subjects are what literature and users are focused on, a chi-square test is done for each term for each RQ, as well as an aggregate test for each RQ.

In the process of organizing knowledge, explicit and tacit information is contained in the text of the content. But seekers of knowledge do not have the time to read the entirety of content for any given research topic. This is where subject access comes in. Subject access allows for knowledge resources to be searched based on the main topics of a Work, and tagging that content based on its “aboutness” is called essentialism (Fricke, 2013; Horjland, 2008). Subject access also

helps researchers quickly deduce if the content is worth looking into. Subject access is facilitated by the vocabulary of subject terms. The knowledge contained in the full text of resources (like journal articles) is captured by assigning terms to content, and vocabulary terms are often derived from a repository of content. For instance, a database collection of journal articles focused on medical procedures may include terms such as *cardiovascular*, *surgical*, or *biopsy*. Machines are gaining popularity for identifying and tagging subjects in content based on explicit and recognized as a string of characters, but many still perform this task manually. A major underpinning of this study is that human understanding, consisting of both explicit and tacit knowledge, is more complex than string matching and requires gathering information from the users themselves.

While many of the articles reviewed for this study discuss the appropriateness of certain controlled vocabulary terminology from cultural, gender, sexuality, and belief system standpoints, many others allude to controlled vocabularies and indexing being outdated and point to specific terms that should be changed. Very few, on the other hand, took a user-centered approach to understanding how users actually define and search for specific topics. Even if the intended audience was described as the user, none surveyed the words the users actually used to search for the topic. Because of this, there was little baseline data that could be extracted from previous studies. In that same vein, most user vocabulary studies used either open or closed card sorting methods, neither of which fits the needs for the current study because traditional card sorting methods depend heavily on a user interacting with a vocabulary to suggest alternate labels—essentially, looking solely at scientific warrant and not user warrant. User warrant is addressed in this study because it is more focused on how the users' search for content is supported, not how the user searches for other subject terms via a browsable controlled vocabulary—hence RQ 4, which looks at the browse and search patterns for terminology. To be sure users are the main

resource for user-centered vocabulary, a hybrid card sort methodology is used to gather users' natural language terminology based on a literature prompt and use their entries to determine matches to the controlled vocabulary set. More details on this hybridized approach are in the Methods section.

The next chapter will explore literature related to knowledge organization, first looking at how previous studies have defined knowledge, and how knowledge is codified via warrant. This will be followed by a survey of the ways knowledge organization fits within the larger context of knowledge management to suggest the potential value of the current study. The chapter concludes with an analysis of overall knowledge organization literature to illustrate why switching from an empirical information model to a pragmatic knowledge model may help reduce user frustration and add more precise access points for information retrieval.

2.0 Review of Literature

2.1 Review method and structure

This section will review the literature supporting this study. This body of literature was selected using multiple methods, including reference analysis, topic coverage, and citation and impact (bibliometrics) factors. That said, many of the studies reviewed did not solely focus on the current state of user-centered vocabulary design -which indicates a lack of user-centered vocabulary research- or on controlled vocabularies focused on the engineering-domain -perhaps because many open source vocabularies are more general and not dedicated to one specific domain, but rather the literature reviewed here will build the case as to why user-centered design is important to user satisfaction in information retrieval. From the literature, the following review took on a topical structure rather than chronological or by author. Another result of reading this body of material was a reinforcement of the idea that a graph-like structure could facilitate better mapping across dispersed datasets, in this case, vocabularies, increasing search effectiveness through query expansion. More on these methods of information organization is described in the proceeding Methods section. The literature review topics and their scope are outlined here:

Tacit versus explicit knowledge:

In this subsection, knowledge is defined and the differences between tacit and explicit knowledge are reviewed. It can be seen that previous knowledge organization studies have primarily focused on explicit knowledge, most likely because explicit knowledge can be controlled and therefore easier to define, harness, and study. However, tacit knowledge is largely uncontrolled but, as the literature indicates, may have a positive impact on knowledge organizational approaches because it keeps the user directly in mind, that is, it is more user-

centered. It is for this reason that the present study will harness users' natural (tacit) language for knowledge organization.

Role of information science in knowledge management:

After positioning the importance of tacit knowledge capture, knowledge management literature is reviewed which indicates the need for more knowledge management and organization studies to better understand why it is necessary, and best practices. In addition to this, the literature indicates that information science may have a greater impact on knowledge organization that has yet to be realized. Based on this, the potential value of the current study is to bring more applications of information science principles to knowledge management studies–particularly those principles focused on benefits that span different domains such as improved knowledge organization and an emphasis on the user's knowledge-seeking preferences, for example, knowledge graph structures.

Knowledge organization methods, empiricism version pragmatism:

Building on the need for more user-centered knowledge organization, the literature reviewed in this subsection indicates a historical paradigm shift from an empirical information model, to a pragmatic knowledge model for knowledge organization. This shift points to a growing need to organize not only the raw information of research but also the knowledge that the information represents. Here, the literature suggests that pragmatic methods show promise in adapting to the user's natural language because it is centered on gathering tacit as well as explicit knowledge from specific user groups. This, coupled with the suggestion within the literature that tacit knowledge is more user-centered, indicates that pragmatic or user-centered methods seem the most beneficial to use in the current study.

Knowledge organization and the engineering user community:

Drawing on the preceding subsections' stress on user-centered design methods, the engineering user group is now examined–particularly these users' preferences for knowledge search. The reviewed literature suggests that the most useful function to engineering users is multidisciplinary knowledge search – particularly with regards to study and design methods. This is, therefore, the primary vocabulary domain proposed in the current study. An underlying theme of this literature, which had a significant effect on the current study, is the implication that ontologically or graph-like structured vocabularies may more effectively capture natural language terms for more robust knowledge search in the engineering community than previous information models. The studies reviewed also indicate that, because graph structures more closely align with human cognition, structuring vocabularies as graphs may likely increase user satisfaction with information retrieval exercises. This understanding is carried into the next section where graph structures for knowledge capture and information retrieval are reviewed.

Knowledge management structures:

In this subsection, the topics are viewed through the lens of information and knowledge organization structures. Because many of the studies reviewed thus far have highlighted the benefits of ontology graph-like structures in the context of knowledge organization, only a summary of benefits graph may offer is represented here. Of particular importance for this section is the ability to connect terms through more traditional vocabulary relationships such as *see also* and *use for*, which may assist in connecting controlled and uncontrolled terminology within vocabularies even if a full graph cannot be created Supporting this, the literature reviewed also indicates that the graph structures used so often in mapping and knowledge organization work may be of great assistance in the vocabulary space. Indeed, the literature supports the position that graph structures may be the best framework for codifying a hybrid controlled/uncontrolled vocabulary.

Current state of controlled and uncontrolled terms in information retrieval

The culmination of the preceding sections is an analysis of the use of controlled and uncontrolled vocabulary terms for better knowledge organization. Setting aside the arguments for eliminating controlled vocabularies altogether or supporting only abstraction terms that are approved and controlled (contrary to their *preferred term* label), much of this section reviews literature focused on both the benefits and disadvantages of controlled and uncontrolled terms. While many acknowledge that user natural language is essential, few studies merge the two cohesively. There are two exceptions to this –van Damme et al. (2007) and Sharif (2009), both of which are key studies that influenced the current research and that argue that graph-like structures are effective at mapping dispersed vocabularies and supplementing vocabularies with users' natural language.

The subtopics outlined above will be the framework upon which the following literature review is organized. These topics build to establish a baseline of literature for analyzing the current state of user-centered controlled vocabulary, how these vocabularies align with one another and with the natural language of users, for a more accessible and satisfying search experience.

2.2 Tacit versus explicit knowledge

A study on knowledge organization must begin by defining just what *knowledge* is. Knowledge is a unique phenomenon. Unlike many constructs, knowledge can be used without depleting it; transferring knowledge enriches both the giver and receiver without loss from either person; and knowledge is everywhere but knowing how to find it or use it takes time and can be limited to one's access to resources and sources respectively (Adolf & Stehr, 2014). Gascoigne and Thornton (2013), define knowledge as "information that facilitates action (p. 20)." Adolf and Stehr (2014) indicate that knowledge is either "technical (scientific) or formal (academic)," and that individual experiences build to form something distinctly separate from knowledge. Lai (2004), who has a detailed list of knowledge definitions, disagrees with Adolf and Stehr (2014) by stating that knowledge is the combination of information, context, experience by individuals and communities to form personal and communal knowledge bases, although the author points out that philosophers continue their attempts to define knowledge to the present day. Ajiferuke (2003) agrees with this definition by stating that explicit knowledge is that which has been expressed in documents, and which are captured in databases, while tacit knowledge is based on the know-how, skills, or expertise of users. The majority of literature reviewed for the current study focuses on explicit knowledge, but if knowledge organization does not also focus on tacit knowledge, explicit knowledge only addresses half of a user's needs. One term that has become synonymous with ontologies is a knowledge graph, as Bergman (2016) notes, "because the word 'ontology' is a bit intimidating, a better variant has proven to be the *knowledge graph* (because all semantic ontologies take the structural form of a graph)." While knowledge graphs and ontologies are not identical, they are indeed similar enough that in the current study these graph-like models used to capture knowledge as data access points will be used synonymously.

Lai (2004) and Sharma (2015) also make the distinction between tacit and explicit knowledge. Here, Lai (2004) specifies tacit knowledge as residing in the human mind and which is created through individual perceptions and experience with the world -with personal interaction being the main component of giving and receiving tacit knowledge. Sharma (2015) simplifies the distinction where tacit knowledge is that of humans and explicit knowledge is primarily used by computers. Gascoigne and Thornton (2013), writing about tacit knowledge theory, define it as "understanding in practice" (p.23). They further state that tacit knowledge is not based on scientific but rather on natural rules and which has an "ontological aspect (p. 22)" where understanding is represented by the relationship established between two entities. The idea that tacit knowledge is conducive to ontological structures is a likely reason that more knowledge organizations use these graph-like structures more than other more static information retrieval methods such as synonym rings and taxonomies. That said, the reviewed knowledge organization studies did not use tacit knowledge for selecting the vocabulary terms, only for knowledge transfer methods. Becerra-Fernandez and Sabherwal (2014), go on to define explicit knowledge as characterized by rules and codification which agrees with Gascoigne and Thorton's (2013) definition but seems to conflict with Ajiferuke's (2003) idea of explicit knowledge. However, Ajiferuke's (2003), Lai (2014), Gascoigne and Thornton (2013), and Becerra-Fernandez and Sabherwal (2014) do agree that explicit knowledge is structured, externalized, conscious, controlled, and formal.

Wittgenstein, a commonly cited author in knowledge organization, also explores the difference and relation between explicit and tacit knowledge. He asks: "What really comes before our mind when we understand a word?" (Wittgenstein, 1953 p. 139-41, as cited in Gascoigne & Thornton, 2013). In this early example from Wittgenstein, he is referring to verbal words but when words are set down in text they then have both explicit and tacit characteristics. The actual text,

the characters, and syntax of written language are explicit knowledge, whereas the meaning behind the words, which is formed by logic and contextual understanding, is tacit knowledge (Hobbes, 1963). In this way, natural language can be seen as tacit knowledge because it is dependent on an individual's experiences, the context in which it is used, and is informal and uncontrolled, whereas controlled vocabulary can be seen as explicit knowledge because it is structured, formal, and controlled. Wittgenstein's discourse is reflected in the 80/20 rule of knowledge management, which states that roughly 80% of individual, group, and organizational knowledge is tacit uncontrolled knowledge, and 20% of knowledge is codified and controlled (Liebowitz, 2012). Tacit information is most effectively gathered through pragmatic user-centered methods, although most often it is not captured in any formal sense. Explicit information, on the other hand, is more commonly recorded because it is easily identifiable in content. Both explicit information, such as controlled vocabularies, and tacit knowledge, such as users' natural language, are integral to developing and sharing knowledge –which is the primary focus of knowledge management and organization.

2.3 Information science in knowledge management

On the surface, stating that knowledge can be managed sounds a bit presumptuous. Defined practically, knowledge management is "the generation, representation, storage, transfer, transformation, application, embedding, and protecting" of knowledge (Schultze & Leidner 2002). Knowledge organization is a subset of knowledge management and is reviewed here to show how more studies focused on information science applications of knowledge management and organization may have a greater impact on the domain – an impact that the literature suggests has

yet to be fulfilled. Wilson (2002) examined knowledge management literature and found that it primarily was used in computer and information system contexts (33%), but did not particularly draw from the information science domain. Updating Wilson's (2002) statistics in 2016, the primary contexts of study have remained the same although the order has slightly changed - business (28%), engineering (22%), information science (21%), computer and information systems (16%), and artificial intelligence (13%). One article in particular highlights why a knowledge organization study from the information science field may have a positive impact on the domain. Kebede (2010), in his article *Knowledge management: An information science perspective*, states that even though knowledge is at the forefront of many information science studies, "IS [information science] is not playing as influential a role as it should be (p.416)." Souza, Tudhope, and Almeida (2012) have a similar statement,

> IS takes upon itself the task of organizing and facilitating the retrieval of the wealth of information that arises from the knowledge produced in all other fields, and this involves the creation of epistemological and ontological surrogates…[which are] dependent on representation products, modeled through successive abstractions over the relevant characteristics of a chosen world or domain, or the information gathered and processed about these, registered in information systems and documents (Souza, et al., 2012, p. 179).

Additional support of Kebede (2010) and Souza, Tudhope, and Almeida's (2012) statements can be seen in Ajiferuke (2003), who found 86% of information professionals from the Special Libraries Association (SLA) work with knowledge management systems, which indicates IS is highly involved in knowledge management and therefore may have contributions for the domain.

In addition to Ajiferuke (2003) and Corrall (1999), Durst and Edvardsson (2012) both indicate the scarcity of knowledge management studies in information science. Corrall (1999)

specifically calls out knowledge management as an emerging information professional role due to the community's focus on recording knowledge and Durst and Edvardsson's (2012) systematic review goes further and states that while knowledge management, perception, and transfer are strong themes in the literature, knowledge organization and utility are lacking. Jashapara (2005) indicates by amplifying the role of information science, a more robust knowledge management theoretical framework may emerge and have a positive impact on the domain and Martin (2008) documents the progress of knowledge management while highlighting knowledge organization as an aspect that requires further research. Orzano, McInerney, Scharf, Tallia, and Crabtree (2008) indicate knowledge management is a "subspecialty" of information science and that effective knowledge organization is critical in the medical field especially to improve the quality of patient care. Sarrafzadeh, Martin, and Hazeri (2006) surveyed information professionals which concluded that 82% of respondents felt that knowledge management could benefit from more information science involvement and that their repositories would benefit from better knowledge organization. Summers, Oppenheim, Meadows, McKnight, and Kinnell (1999) found that knowledge management can benefit from organization and abstraction theory, but the methods used for the two have thus far not been effectively aligned.

Taking the aforementioned studies into consideration, a search for knowledge organization and knowledge management IS dissertation research in the last fifteen years was conducted to identify the saturation of dissertation studies in both areas. The result was less than 15 dissertations focused on IS knowledge management (analysis from ProQuest Dissertation Database conducted by the researcher, April 2016). Attempting to identify the level of overlap between knowledge management and organization studies in IS, a similar search was conducted on journal articles. The analysis found that 26% (n=933 out of n=3586) of the literature focused on knowledge

management in IS only. Comparatively, 69% (n=196 out of n=284) of content focused on knowledge organization in IS, but with 92% less content overall (analysis from Web of Science conducted by the researcher, April 2016). This indicates that although the literature points to knowledge organization as a significant component of knowledge management, it does not receive nearly as much IS coverage within the literature on its own. Additionally, while knowledge management has more coverage across disciplines, IS dominates the literature in knowledge organization, most likely because of the historical prominence of IS knowledge organization of vocabularies such as the Library of Congress Subject Headings (LCSH) or the Medical Subject Headings (MeSH).

Kebede (2010) puts forth a few reasons explaining the disproportion of studies focused on knowledge management in IS, most of which focus on IS researchers, who Kebede (2010) indicate have admitted difficulty in grasping the key concepts in knowledge management which they feel excludes them from effective research, and a lack of techniques, frameworks, and tools focused on IS specifically. He stresses that this is harming information science research and is limiting the benefits the field could otherwise have on knowledge management studies. Kebede states that "knowledge has always been the ultimate concern of the profession (p. 418)" and quotes from ASIST (1999), which states that:

> Our ability to transform data into information, and to transform information into knowledge that can be shared, can change the face of work, education and life. We have increasing capacity to generate or gather, model, represent and retrieve more complex, cross disciplinary and multi format data and ideas from new sources and at varying scales. The transformational power of information can only be capitalized upon through knowledge acquisition, classification, utilization and dissemination research, tools and techniques (ASIST, 1999; as cited in Kebede, 2010, p. 418).

Kebede further states that information aligned with personal experience and understanding (tacit knowledge) coupled with logical reasoning creates knowledge and the goal of information science is "to help facilitate human access to information and knowledge for effective decision making and problem-solving (Kebede, 2010, p. 421)." Kebede's (2010) study can be seen as a call for more knowledge management studies from the IS community, and which the aforementioned literature analysis supports. Taking this and viewing it through the lens that IS already uses methods which may benefit knowledge organization studies and that IS has a historical emphasis on "knowledge acquisition, classification, utilization and dissemination [of] research, tools and techniques," Kebede's call for more IS focused knowledge management and organization studies indicate the current study will be a positive contribution to the field.

2.4 Empiricism version pragmatism

Knowledge organization methods have been shifting from information (empirical) models to more user-centered (pragmatic) models over the past few decades, which has seen general knowledge organizational studies increase over 900%. One reason for the increased focus on knowledge organization, which appears in much of the literature, is the exponential growth of domain-specific knowledge within content, such as scholarly articles, and knowledge transfer, which is focused on the perpetuation of knowledge between researchers as well as between different repositories and domains. However, as previously stated, reviewing dissertations from the last fifteen years indicates that knowledge organization dissertation topics are few (temporal analysis conducted on *knowledge organization* journal articles through Scopus and Web of Knowledge, 2016). Out of these, five focused on pragmatic methods of knowledge organization.

Huang (2009) examines the patterns of thinking when organizing knowledge to match terms with the user's task and situation for greater relevancy in information retrieval. Wyatt (2009) focuses on how knowledge organization aligns with cognitive processes of different demographics and asks if knowledge organization methods differ among domain specializations, primarily between biology and history. Ziemba (2011) focuses on how knowledge discovery may be increased in agricultural and natural resource digital collections through ontological concepts rather than hierarchical subject abstraction. Loehrlein (2012) focuses on how people rank user-centered taxonomy terms before and after they are used in a knowledge organization system. White (2012) analyzes the organization of scientific data sets, comparing metadata and subject term creation to identify which taxonomy works best for specific content types. Identifying only five dissertations focused on pragmatic knowledge organization indicates the present research will be a positive addition to this body of knowledge. That said, these dissertations cover many of the same topics as their journal counterparts. Themes running throughout include modernity of organization models, terms, and theory, the prevalence of search engines over repository search engines, interoperability and mapping of dispersed terminologies, cognition during a search, and human information interaction, to name a few. These topics point to a growing need to organize not only the raw and empirical information of research but also the knowledge that the information represents in a pragmatic, user-centered design.

The issues surrounding knowledge organization, coupled with the extensive collection of knowledge resources available, have spurred initiatives to switch from an empirical information model to a pragmatic user-centered knowledge model of organization. Initiatives such as the establishment of the Association of Research Libraries (ARL) Joint Task Force and the Elsevier Hunter Forum are two examples that highlight the shift from repositories as information

(empirical) aggregators to user-oriented (pragmatic) knowledge repositories. As the ARL Joint Task Force states:

> There is a perception that science librarians, more than ever before, need to be actively engaged with their user communities. They need to understand not only the concepts of the domain, but also the methodologies and norms of scholarly exchange… This new paradigm suggests a shift in focus from managing specialized collections (the 'branch library' model) to one that emphasizes outreach and engagement (2007, p. 5).

Additionally, the American Library Association (ALA) Midwinter Meeting gathered 100 information professionals at the Elsevier Hunter Forum in 2016 to investigate to what extent knowledge organization was centered on the users. During this forum, Kelechi Okere stated that "libraries have been at the consumption end (of knowledge generation) [and] they do a great job of gathering and organizing information, and making it available to researchers and students (Elsevier Hunter Forum, 2016)." Okere then went on to state that this characteristic is broadening into more knowledge management areas as well. MJ Tooey supported this sentiment in the same forum where they stated that information professionals:

> have the skill set to organize things…understand controlled vocabulary...[and] understand ontologies. We understand organization of information in a way that seems as if it's bred into us. The evolution from the print word and subject heading to data made a lot of sense – we understand it. So there started to be conversations about what we could do to help people organize, access, and store data (Elsevier Hunter Forum, 2016).

These are just two examples where the shift from empirical information management to user-centered knowledge management has been discussed.

The studies reviewed thus far are only highlights of the paradigm shift within knowledge organization methods. For more discussions exploring aspects of the switch from an empirical to

a more pragmatic model see Carter and Whittaker (2015), who express the different methods for knowledge organization in distinctive or special repositories; Parker and McKay (2016) speak to the "next" shift where information and knowledge management, intranets, search engine optimization, business analysis, taxonomy, information governance, and learning management methods collide; Vargas, Vanderkast, Garcia, and Gonzalez (2015) dive into the idea of disruptive technology and its effect on knowledge and how blended information professionals address knowledge needs through new methodological frameworks; Beck (2015) indicates medical knowledge management based on methodologies and systematic review has changed the role of medical information methods used; and Ma, Lyu, and Wang (2014) indicate that information professionals are starting to use more user-centered tools for analysis of user keywords and vocabulary data in addition to empirical methods. Hjorland (2015), Hjorland (2012), Hariri (2013), Georgas (2014) Miller and Pellen (2014), among others, also support more discussion on switching from an empirical information model to a more user-centered knowledge organization model.

Knowledge organization, whether based on empirical or pragmatic methods, has a rich history focused on connecting users to the content they seek. In information science, one of the primary methods for organizing knowledge are subject headings, or controlled lists of "aboutness" tags used to summarize what a Creative Work is about. One researcher, in particular, appears in much of the literature. With over 7,000 citations, Hjorland is one of the most influential researchers in knowledge organization. Hjorland also focuses on engineering-domain topics and therefore his work is of particular interest to the current study. In Hjorland (1998), he noted that semantics, which is the meanings of words and phrases in particular contexts, was not only associated with terminology but also with the relationships between the terms –drawing on previous work by Harter, Nisonger, and Wenig (1993). Hjorland's early work can be seen as pivotal in emphasizing

the combination of semantics and knowledge organization because it has influenced over 6,000 articles. As such, Hjoralnd and his counterparts -Mai (1999a; 1999b; 2000; 2003; 2004a; 2004b; 2005; 2008; 2010; 2011a; 2011b), and Smiraglia (2001a; 2001b; 2002a; 2002b; 2012; 2013a; 2013b; 2014), among others- continue to stress the importance of melding semantic contextual search, domain centered terminology, and bettering knowledge organization and indexing methods. Of note, many of these authors have been researching these themes in knowledge management for years and yet many of their findings have not found wide adoption in knowledge organization practice. The reason for this is unclear.

In his 1998 work, Hjorland outlines the different components of a scientific research article and how these components, both elements contained in the article and added information such as controlled vocabulary tags, hold semantic meaning that may be used to enhance knowledge retrieval. Hjorland (2015) goes on in a later study to show how user-centered knowledge organization is semantic and is more in line with the words users use in their keyword queries, called natural language. This contrasts with Wittgenstein's earlier methods focused on empirical-based methods of organization. Interestingly, while Wittgenstein first adhered to empirical methods, the majority of his career was later focused on more user-focused methods such as user mental models and decision making during information retrieval exercises. Hjorland (1998) outlines the differences between Wittgenstein's knowledge organization methods (see Table 2):

Table 2: Wittgenstein's knowledge organization.

Empirical knowledge organization	***Pragmatic user-centered knowledge organization***
Terminology always has the same meaning no matter its context.	Terminology is flexible depending on context and is domain-specific.
Indexing is descriptive and not theoretical.	Indexing is theoretical where each term is "meant to mark family resemblance between objects' concepts (Hjorland, 1998, p. 27)."
The volume the term is represented within the document signals its importance.	There is no absolute universality between categories and document vocabulary.
The more controlled vocabulary assigned to a document the better.	Essentialism, that of assigning controlled vocabulary to capture the essence of the context, should be used when assigning terms.

A notable member of the library classification community also has commented on Wittgenstein's early empirical methods. Dewey, Ratner, and Post (1939) stated that "empiricism has also misread the significance of conceptions…[and] such ideas are dead, incapable of performing regulative office in new how situations (p.883)," implying that empirical terminology assignment is not flexible enough to deal with new phenomena which severely limits its usefulness in information retrieval. Hjorland (1998) uses Dewey et al.'s (1939) criticism to highlight how empirical knowledge organization may lead to obsolescence because of its static nature, while pragmatic user-centered design may be better equipped to adapt to the user's natural language which is always changing. Pragmatic knowledge organization is centered on gathering tacit as well as explicit knowledge from specific user groups to establish a knowledge graph. This conflicts with the traditional a priori methods of empirical knowledge organization where the knowledge of users is inferred (Hjorland, 2012). Both methods are valid but pragmatic knowledge organization directly involves the users while empirical methods only speculate, therefore pragmatic methods tend to align better with users' natural language and expectations because they derive their logic from the users themselves.

How users review search queries also supports a shift for a more user-centered knowledge organization. As Hjorland (2015) states, each controlled vocabulary is based on the domain, search philosophy, and knowledge needs of its users. If this is the case, and the user's input is not directly sought, then the controlled vocabulary is based on assumptions. Assumptions, no matter how well they are researched, are not true assessments until users are consulted. Mai (2011a) explores this in their research into the modernity of knowledge organization. In their study, Mai (2011a) identifies epistemological pluralism as a sign of modernity in organizing knowledge. Mai indicates that pluralism is how users identify what they know (their tacit information), how user groups establish a common understanding between themselves through discourse, and how the users then connect that to the knowledge in the world around them. Mai (2011a) does caution that the literature and researchers rarely have terminology consensus and that true modernity in knowledge organization takes a pragmatic approach where multiple perspectives terminology are considered. Smiraglia (2013), who also explores pragmatic methods in knowledge organization, also cautions that although pragmatic methods may meet current needs they also may lose value over time. Hecker (2012) and Pando and Almeida (2016) suggest a postmodern approach to pragmatic organization that incorporates natural language and controlled vocabularies in a way that resolves some of these issues. However, the postmodern approach is more difficult to capture in controlled vocabularies and resonates with fewer users as a whole.

Although Mai and Hjorland agree that pragmatic user-focused design is the most appropriate method to bridge the gap between user queries and the formalism of language present in Works, they disagree on how to incorporate pragmatic methods into controlled vocabularies and indexing. In Hjorland's (2011) study, they ask if controlled vocabularies are necessary after Google. They indicate that formally controlled vocabularies, and the assignment of them to Works,

should remain with the IS professionals because users do not have the necessary objectivity to organize knowledge themselves. Mai (2011b) takes a different approach and indicates that formal controlled vocabularies can be enhanced with the user's natural language through folksonomic methods. Smiraglia (2013; 2010) also mentions the usefulness of user tags in knowledge organization, which they describe as *noetic tagging* where all tags stem from the way an individual perceives the world. However, Smiraglia (2013), unlike Mai (2011b), does not explore adding folksonomic terms to controlled vocabularies. The difference between Mai (2011b) and Hjorland (2011) is that Mai believes in a higher dependency on direct user input while Hjorland believes pragmatic design is the ultimate goal but with the IS professionals being the intermediary. Hjorland, Mai, and Smiraglia all suggest pragmatic methods to incorporate explicit as well as tacit knowledge. Pragmatic user-centered design, therefore, is a viable method to investigate a more user-centered knowledge organization process within the present study. These topics point to a growing need to organize not only the raw information of research but also the knowledge that the information represents. This is especially important in the engineering-domain which regularly sees exponential growth in literature published each year.

2.5 KO and the engineering community

As the number of information increases, so too does the difficulty of finding knowledge within it. Reiterating much of what Lancaster (1999) described in his *paperless society*, Rayward (2014) and Sharma (2015) state that society is currently in a third information revolution fostered by the increase in digital information. Adolf and Stehr (2014) state that the current society is a *knowledge society* where “the modern economy is knowledge-based and that worlds of work,

politics, and everyday life are transformed by and based on knowledge and information (p. 2)." Castells (1996) describes this as the rise of the *networked society* based on computer technology, while Gaines (2013) shows through a historical analysis of information exchange that the issue of having too much knowledge, identifying reputable information, and finding knowledge amongst the multitude of information artifacts has been an issue dating back more than two millennia. Regardless, "information overload" will only increase as more research is conducted. With the exponential growth of information, effectively searching and synthesizing the knowledge contained within said information assumes greater importance. Traditionally, libraries have been the repositories of knowledge but there has been an emergence of more diverse types of knowledge repositories in the past decade. According to a 2016 study on *How readers discover content in scholarly publications,* the majority of researchers from all domains, countries, and socioeconomic brackets tended to start their search for knowledge via Google (Gardner & Inger, 2016). Hjorland (2015), Hjorland (2012), Hariri (2013), Georgas (2014) Miller and Pellen (2014), support this assertion in their exploration of the evolving knowledge ecosystem, particularly in the engineering-domain. As a result, more Google-like search is expected in all forms of information retrieval but what Google does not do is tailor search to specific domains of research. Instead, Google serves as a starting point for research, being a broad search engine focused on everything from the Wikipedia article that can get a user started on a topic of research, to what plethora of diseases your symptoms point to. While making information retrieval on scholarly Works more accessible is important, losing the specificity of scholarly metadata and indexing should not be underestimated.

Studies into engineering-domain terminologies are especially needed because, as Adolf and Stehr's (2014) description of a knowledge society states, when information is expected to grow

exponentially, it becomes increasingly important for researchers to have effective ways to search for knowledge. Adding to this is the lack of engineering-domain terminological studies. Maurer and Shakeri (2016) affirm that "librarians assert within the library literature that the science, technology, engineering, and mathematics (engineering) disciplines, in particular, are not well served" by controlled vocabularies and that "little research has been conducted to document the amount of topical access within bibliographic records from different academic disciplines (p.214)." They go on to state that studies into underserved domains such as engineering could be used to improve users satisfaction during knowledge search (Maurer & Shakeri, 2016). In agreement with Hislop (2013) and Oborn and Dawson (2010), Hjorland (1998) states that the most useful information to an engineering user is being able to identify the different meanings of terms, especially terms that are shared across disciplines.

Searching knowledge repositories may seem simplistic. There are 12 common pieces of information that facilitate search. Out of these, *domain* and *subject tag* are abstracted through assigning controlled vocabulary terms, and *author keyword* and *free-text search* are uncontrolled abstraction by authors or users –both of which are typically domain-specific and have a high level of domain bias (Hjorland, 2013; Maurer & Shakeri, 2016). This causes complexity in information retrieval because controlled vocabulary and textual terminology usually differ across domains and repositories and researchers use different natural language across disciplines. If the controlled vocabularies were mapped and the user's natural language was also included, there would likely be a much higher probability of successful information retrieval across the massive amount of content available to researchers. That said, uncontrolled terms are based on tacit rather than explicit description information which makes them more difficult to quantify, capture, record, and share. Hjorland (2011; 2012) supports this statement and Cleverley and Burnett (2015) Iyer and Bungo

(2011), Rodriguez-Gonzalez, et al. (2012), Meij and Rijke (2007), Lu, et al. (2010), Pirmann (2012), Sharif (2009), van Damme et al. (2007), Vandic, van Dam, and Frasincar (2012), Wetzker, Bauckhage, Zimmermann, and Albayrak (2010), Willoughby, Bird, and Frey (2015), and Zhu and Dreher (2012) further argue that capturing tacit knowledge through natural and controlled vocabulary is imperative for facilitating a better bridge between content and satisfying users' natural language queries.

Hjorland (2012; 2013; 2016) has expressed that user-centered methods can have a positive effect on information retrieval because the words users enter into their query more closely align with their preferences. Hjorland finds that "users need 'maps' of information structures…[which] uncover the more or less hidden meanings, interests, and goals in the document (p. 28)" and that by having a resource that maps and expands the distinctions between terms, more cross-disciplinary research may occur. In Hjorland's (2013) study, he outlines how knowledge organization based on user group cognition is more effective than empirically-based knowledge organization methods based on an individual's subjectivity. Hjorland's study draws heavily on the Ingwersen and Jarvelin (2005) model IRiX, which indicates better retrieval experience and satisfaction may be achieved if the organization is based on the user's cognitive viewpoint and need rather than the intended use of the content, and Robson and Robinson (2012), which found that the more knowledge organization structures mimic the way users think, the higher the likelihood of achieving research goals and retrieval satisfaction. Hjorland (2013) concludes that pragmatic knowledge organization may be used to greater effect in domain-specific organization research. That said, Baeza-Yates and Saez-Trumper (2016) indicate that individual user terms typically only represent 50% of the community's tacit knowledge and that further research is needed to determine how to achieve the wisdom of crowds rather than the wisdom of a few.

Similar to Baeza-Yates and Saez-Trumper (2016), in their 2013 study, Hjorland distinguishes individual and community knowledge capture. They state that empirical studies only focus on individual users but fail to take into account the needs (versus the wants) of knowledge seekers for particular domains. Horjland uses Rosa, et al.'s (2005) study, which indicates students prefer Google over repository search and that users try to avoid using LCSHs but they do not go into detail as to why this may be. Hjorland points out that card sorting is an effective means to gather user-centered terminology; which supports the planned methods for the current study to record user natural language through card sorting methods. The user in the current study will be considered anyone working or studying in the engineering-domain. The users are members of the knowledge domain with which they associate and can have various levels of expertise. Even experts in a field need to start research on a new topic they are not an expert in every now and again. A user domain in this sense is,

> best understood as a unit of analysis for the construction of a KOS [knowledge organization system]. That is, a domain is a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics (Smirgalia, 2012, p. 114).

From Smirgalia's statement, it can be seen that not only are ontologies useful for capturing tacit information and mimicking how knowledge works within the human mind, these graphs also mimic the way domain communities create a common language of understanding. Wang, Bales, Rieger, and Zhang (2016) go so far as to state that domain knowledge like engineering is naturally structured within an ontological framework. Focusing on how controlled vocabularies are assigned to scientific articles for knowledge retrieval, Szostak (2003) argues that the aboutness of a Work, they define this as the 6 aspects (Who, What, When, Where, Why, and How) of scientific

scholarship, are important to capture through the indexing process because they contain both the tacit and explicit knowledge the researcher is likely to use while searching. Szostak (2003) stresses that out of all the 6 aspects, the theory and method terminology "are arguably the two aspects of scholarship that are most in need of classification (p. 20)" because these contain the most tacit information that aligns with how the researcher naturally researches. The literature reviewed implies that ontologically structured vocabularies may effectively capture terms related to these 6 research aspects, particularly those related to theory and methods and that this is important to facilitate more robust information retrieval in the engineering community. The benefits of these graph frameworks are further reviewed later in this study.

Cleverly and Burnett (2015) and Szostak (2003) also explore how the engineering community searches for knowledge based on their knowledge needs. After surveying 52 engineers from 32 organizations, Cleverly and Burnett (2015) found that a research need model, called BRIDGES, may help engineers to more effectively meet their research needs and "stimulate new needs, improving a system's ability to facilitate serendipity (p. 97)." The model includes terminology types most used by engineers. They include Broad terms, which are large containers of topics; Rich terms, which are akin to sentiment and synonyms; Intriguing terms, which are unanticipated terms that lead to serendipitous results; Descriptive terms, which are terms based on essentialism and themes; General terms, which are akin to natural language terms used in a general store; Expert terms, which are specialized or jargon terms; and Situational terms, which are akin to named entities or methodological in nature (Cleverly & Burnett, 2015). Szostak takes a more topical approach and provides what he states is a list of the majority of methods used in engineering scholarship. They are (see Figure 1):

Experiments (including natural and quasi-experiments
Surveys
Interviews
Mathematical models (and simulations)
Statistical analysis (often, but not far from always, associated with models)
Including secondary [that is, collected by others] data analysis
Ethnographic/observational analysis [some would distinguish "interactional" analysis in which the investigator interacts with those under observation[
Experience/intuition [some would treat this as an important subset of observational analysis, since we are in effect "observing" ourselves here]
Textual)content, discourse) analysis
Classification (including evolutionary analysis)
Mapmaking
Hermeneutics/semiotics (the study of symbols and their meaning)
Physical traces (as in archaeology)
Some would treat "evaluation" of programs as distinct, though it can be seen as a combination of some of the above methods. Similar arguments can be made with respect to "demography," case study, feminism, and perhaps also hermeneutics. Certainly, "case studies" involve the use of one or more of the above methods.

Figure 1: Twelve general method types as stated by Szostak (2003).

Szostak (2003) indicates that by using these methods, or a combination of them, scholars may be able to more effectively search for studies across multiple domains, have a more satisfactory search experience, find studies that will directly pertain to the methods they use, or which are the most suitable to their research. Additionally, they suggest that engineering studies lend themselves to term-relationship (graph-like) connections because of the emphasis on the scientific process and cause and effect. They go on to state that theory and method terms may also help strengthen the scientific process by way of highlighting alternative or complementary methods from other disciplines. Szostak (2003) concludes that by including study and design method types in controlled vocabularies, scholars would stop "re-inventing the wheel" and would more effectively identify fresh lines of inquiry. In agreement with Szostak (2003), the medical community has seen several studies, mostly in the systematic review and clinical studies space, which indicate the need for more effective search on method types to ensure reproducibility and

to decrease duplication of research. Many of these studies point to study design and methodology as the knowledge that is lacking effective means for subject retrieval which supports the vocabulary topic selected for the present research. See Hackett (2016), Bekhuis, et al, (2013), Vampati and Schurer (2012), Vallas, Gibert, Sanchez, and Batet (2010), and Stevens, Goble, Bechhofer (2000) for examples.

Although ontological frameworks for knowledge organization are discussed in a later section, there are quite several studies that indicate engineering researchers, in particular, prefer a search that supports cause and effect, symptom and treatment, and generally relationship-type connections between topics. For example, Gilchrist (2003), Angele, Erdmann, and Wenke (2008), and Giess, Wild, and McMahon (2007) found that using graph structures for subject browse, as well as query expansion in search, aligns with the way those in the engineering field tend to search and think about information. Some studies have shown that to normalize the search experience, terminology across different engineering sub-disciplines and repositories of information that varies can be mapped based on semantic similarity for a better search experience. Some studies include: Du, Lau, Ma, and Xu (2015) investigate using ontologies to map engineering-domain terminologies across different domains and repositories for better knowledge transfer and search; Giess et al., (2007), who show that engineers prefer the context-rich knowledge search afforded by query expansion based on terminology mapping; Hjorland (2002), who indicates engineering-domain terminological studies may use ontologies to link domain terms with natural language in a domain “according to semantic and pragmatic criteria (p.451)” furthering the researchers access to knowledge; Angele et al. (2008) who use an ontology structure for an automotive knowledge management system; and El-Diraby (2013), who use an ontology structure to map the knowledge from the construction domain for enhances search within a specialized repository. These

terminology mappings can lead to hypergraphs, or higher levels of semantic relatedness (Bendersky and Croft, 2012; Devezas and Nunes, 2019), which is not explored in the current study but is a prime area to expand upon for later research. Some studies that cover the higher-order semantic knowledge between topics in the engineering-domain are Hjorland (2013), who indicates that the labeling of the natural world is conducted the same as content tagging and therefore can equally benefit from knowledge being mapped through ontological relationships; Gilchrist (2015), who describes ontology structures as a means for replicating knowledge found in content; Parinov and Kogalovsky (2014) as well as Yan and Zhu (2015), which both explore how ontological term-relationship structures from scientific articles can be used to represent the scientometric knowledge within studies.

Adding to these, Balatsoukas and Demian (2010) show that ontological knowledge organization is particularly helpful to engineers who have expressed their preference for understanding the context of knowledge retrieval to "gain orientation and inspiration (p. 453)." Balatsoukas and Demian (2010) go on to say that engineers are more familiar with granularly presented and organized knowledge resources because of the highly structured content they are accustomed to interacting with. Balatsoukas and Demian (2010) define granularity as "the size, decomposability, and extent to which a resource is intended to be used as a part of a larger resource…more granular digital resources are larger and are composed of smaller pieces (p.454)" such as the abstract, introduction, sections, and subheadings are granular pieces of the overall article. This can be seen as a mirror of whole-part or broader-narrower relationships in knowledge organization structures. Balatsoukas and Demian (2010) indicate the benefits of facilitating granular search are improving engineers' performance, satisfaction, judgment of relevance, better design solutions, and more productive design-development work. In these examples, ontology

knowledge graph structures were used as a way of connecting controlled terms with relationships, by way of forming a knowledge structure more closely related to how people think. However, these studies also indicate that natural language terminology has not yet been utilized to a significant degree, and capturing tacit knowledge remains elusive. Attempts to facilitate a more satisfactory search, through linked knowledge graph vocabularies such as the Medicine Medical Subject Headings (MeSH) and the National Cancer Institute Thesaurus, have made progress towards expanding tacit knowledge in a controlled vocabulary, however, the user's natural language may provide a more satisfactory search for knowledge that is included.

The literature focused on knowledge organization that incorporates the natural language of the engineering community is limited, although the studies that do focus on this indicate that including natural language helps to make information retrieval more accessible and a better search experience. Zhang, Ogletree, Greenberg, and Rowell (2015) analyzed the use of controlled vocabularies in engineering and found that 73% of users preferred using natural language free text to search, and Cleverley and Burnett (2015), examining how a research team interacted with the knowledge organization of a 13,000 volume repository, found that geoscientists' satisfaction with knowledge search grew after natural language was added to the controlled vocabulary. In addition, Cleverley and Burnett (2015) stated that the ontology structure and natural language additions increased "the propensity of a search UI [user interface] to facilitate unexpected, insightful, and serendipitous discoveries (p.36)." Adding to this, Gardner and Inger (2016) surveyed the users of scholarly publishing platforms, accumulating some 40,000 user responses, and found that researchers in the engineering community, particularly chemists, reported they suffer from too much information and not enough resources to extract knowledge from them. This is not surprising with projections of engineering research growing 50% every nine years (Noorden, 2014). Gardner

and Inger (2016) also found that engineering scientists struggled with current awareness of technology and scholarship in their domains due to poor knowledge discovery in repositories. Taking into account that users prefer natural language search and that those researching in the engineering field are in danger of being buried in the information they cannot search through, graph structures seem to be a viable means to align the user's preference to search with natural language and the consistency offered by controlled vocabularies.

In addition, some studies included natural language in knowledge organization and also conducted user studies to supply supporting evidence. Zhang, et al. (2015), indicate that engineering users slightly preferred using a single controlled vocabulary (76%) over free-text (75%) and that 70% of engineering users would also like the option to search from multiple controlled vocabularies or have a normalized topical search experience, in the future. Interestingly, Zhang, et al. (2015) found that while authors and indexers assigning controlled vocabulary tags had preferences for knowledge search similar to user's, those who developed the controlled vocabularies preferred using free text (77%), one controlled vocabulary (76.8%), and software-generated controlled vocabulary tags (70%) instead. It is interesting to note that the developers who build controlled vocabularies have different preferences for how these vocabularies should work than their users. Kotis and Vouros (2006) describe a "living ontology" based on the flexibility of a research team's natural language, while Walk, et al., (2014) analyze the changelogs of a collaborative ontology engineering project where natural language is used to accommodate changes in terminology used by the research teams. Additional engineering studies which focus on knowledge organization that includes natural language in some form include Ciccarese, et al., (2011), which propose using an extensible ontology framework to map the many biomedical ontologies to the natural language of the research community; Hjorland (2016), which state that

different domains and repositories require different knowledge organization structures and reiterates there is strong evidence that natural language terminology assists in better knowledge access and retrieval (especially when it is aligned with a controlled vocabulary). To create a more connected knowledge organization graph, Cleverley and Burnt (2015) suggest enhancing engineering-domain thesauri information with an ontology structure, and Broughton (2006), Hjorland (2012), and Giess et al., (2008) maintain that even though faceted knowledge organization is typical, these controlled vocabularies lack the specificity many engineering researchers are familiar with. Pirolli and Kairam (2013) and Nielson (2011) found that engineering users can learn from the natural (author keywords) and controlled vocabulary terms assigned to content. Cleverley (2016) indicates researchers' personalities (tacit knowledge) also play a role in how they search for knowledge.

The studies reviewed here focused on knowledge organization in the form of controlled vocabularies in the engineering community, although they were limited in scope (most surveyed engineering groups with less than 100 participants). More than any other engineering sub-domain, medical/biomedical had the most literature focused on knowledge organization, primarily ontological graph-like frameworks. The other engineering sub-domains, such as chemical, civil, electrical, mechanical, agricultural, and materials engineering, have very little research, at least publicly available research, involving knowledge organization let alone anything detailing whether these engineering communities prefer natural language or not. This signals that more research into knowledge organization methods for the greater engineering-domain may shed light in this area. This review into knowledge organization studies focused on the engineering-domain also indicates that research exploring the engineering community's preference between natural language and controlled vocabularies may be a valuable contribution to the field, especially if the research uses

an ontological framework and user-centered design methods. All of which will be discussed below through a review of the literature.

2.6 Knowledge management ontology structures

The idea that tacit knowledge is conducive to ontological structures is a likely reason why more knowledge organizations use these graph-like structures than other common organizing formats like hierarchies. As Bergman (2016) notes,

> Every knowledge structure used for knowledge representation (KR) or knowledge-based artificial intelligence (KBAI) needs to be governed by some form of conceptual schema. In the semantic Web space, such schema are known as "ontologies," since they attempt to capture the nature or *being* (Greek ὄντως, or ontós) of the knowledge domain at hand. Because the word 'ontology' is a bit intimidating, a better variant has proven to be the *knowledge graph* (because all semantic ontologies take the structural form of a graph).

Bergman (2016) goes on to state that knowledge graphs follow Peirce's triadic logic, or semiosis, which states that the most basic way to categorize things, concepts, terminology, and ideas is in threes. Smiraglia (2013) also makes this statement and takes this a step further by indicating that threes, or triples, also mimic the way people think and structure knowledge in their own minds. Bergman (2016) alludes to this as well by stating that "traditional classification schemes have a dyadic or dichotomous nature, which does not support the richer views of context and interpretation inherent in the Peircean view." This may also suggest why Berners-Lee used an ontology structure for his semantic web research (Marco, 2016b). Indeed, the graph or web structures have a many-to-many relationship that is otherwise not possible with traditional data models like hierarchies or static tables, furthering the notion that a graph structure may be able to

connect many controlled terms to many user terms for the same contextual meaning called a class, or atom axiom (W3C, 2012).

The medical community has also embraced ontological knowledge organization methods, as can be seen in Conway, et al (2016), Unger, et al. (2016), Dos Reis, et al. (2015), Huang, et al. (2015), Jeong, Kim, Park, and Kim (2014), and Alfonso-Goldfarb, Waisse, and Ferraz (2013), among others. Conway, et al (2016), describes an ontological schema called SKOS (Simple Knowledge Organization System) which is used to connect terms from multiple terminologies to assist in knowledge sharing; Unger, et al. (2016) find that ontologies provide an organizational structure which "converts information known about individual entities into an interconnected network in which concepts can be linked by many types of relations (e.g., taxonomic, thematic) (p. 202)" and examines methods for mapping controlled vocabularies using ontologies; Dos Reis, et al. (2015), identify ontology mapping issues that may hinder knowledge transfer between different repositories and that mapping knowledge terminologies with ontology structures "remains cornerstone to annotate *electronic health record* (EHR) content to facilitate its sharing and retrieval (p. 167);" Huang, et al. (2015) investigate a tool for domain experts to construct knowledge graphs rather than ontology engineers and state that "ontologies for knowledge organization and access has been widely recognized in various domains (p. 1);" Jeong, et al. (2014), use an ontology structure to connect knowledge organization terms; and Alfonso-Goldfarb, et al. (2013), document how knowledge organization in engineering has evolved and expresses that with the abundance of content published today, ontology structures are often used to connect dispersed and heterogenous topics for better information retrieval. The knowledge organization literature from the medical community often is concerned with knowledge transfer and access, which these studies use terminological mapping techniques to mitigate. In these examples, the

authors used an ontology structure because it enables more term-to-term relationships for a more satisfactory search experience as well as using a structure that more closely resembles the way knowledge is structured in the natural world.

These are just a few examples of those who favor ontologies for information discovery. Others already mentioned in previous sections represented case studies where an ontological structure was used for knowledge organization, scholarly collaboration, and knowledge sharing between different domains (term mapping) and/or different repositories (interoperability). In these examples, ontology knowledge graph structures were used as a way of connecting controlled terms with relationships, by way of forming a knowledge structure more closely related to how people think. Engineering is a multidisciplinary field, borrowing methods and terminology from across domains and often changing the context of a term, for instance, the term “bank” can mean a waterway, a financial institution, or turning an aircraft left or right, coupled with technology terminology having a high degree of change, from users and content alike, due to terminology changing to reflect the emerging understanding on the technology. Having a framework to map synonyms or context and semantic matches across terminology is of huge importance to the engineering community, as Hjorland (2013) and Gilchrist (2015), as well as Angele et al. (2008), Rodriguez-Gonzalez, et al. (2012), Martinez-Gonzalez and Alvite-Diez (2014), Walk, et al. (2015), and Holsapple and Joshi (2001), van Damme, Hepp, and Siorpaes (2007) agree, that ontologies may be able to retain the consistency of controlled vocabularies while also being flexible enough to include user’s natural language. Van Damm and Sharif (2009) support using ontologies for aligning controlled vocabularies and the user’s natural language as well. In addition to this, Hjorland (2015), Cabitza, Colombo, and Simone (2013), Mangisengi and Essmayr (2003), and McKerlich, Ives, and McGreal (2013) state that ontology structures may be used to capture

tacit knowledge through their term (node)-relationship organization. Therefore, it can be seen that the literature identifies ontologies as a promising framework to test the merging of users' natural language and engineering-domain vocabularies to create a stronger bridge between users' queries and content.

Because natural language is so unstructured, knowledge organization outside of graph are not used as often in the literature. Dong, Wang, and Liang (2015) describe knowledge organization as a spectrum from unstructured with less semantics to a highly structured organization with rich semantics. The literature reviewed also follows this spectrum where folksonomies are considered the least structured, thesauri are more structured, and taxonomies and ontologies are the most structured. Madalli, Balaji, and Sarangi (2015) represent the relationship between the different knowledge organization structures as spheres that overlap, as can be seen in Figure 2 (inspired by Madalli et al. (2015).

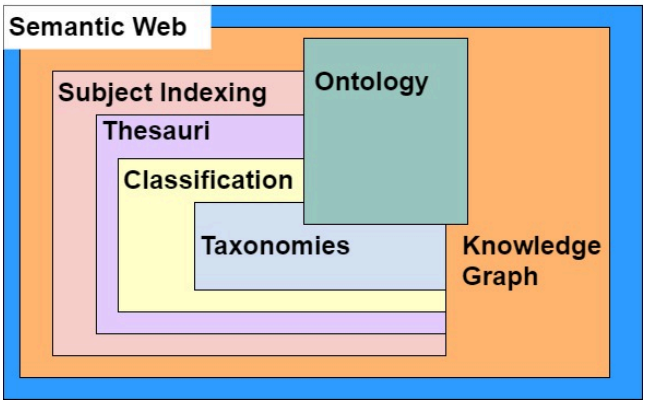


Figure 2: Relationships between controlled vocabulary types. Researcher created.

Madalli et al. (2015) show five types of controlled vocabularies used in knowledge organization whereas Raghavan and Krishnamurthy (2013) group controlled vocabularies into three larger groupings: *term lists*, which include authority files, glossaries, dictionaries, and

gazetteers; *classification/categories*, which include subject headings, classification schemes, and taxonomies; and *relationship lists*, which include thesauri, ontologies, and semantic networks. Both listings of knowledge organization structures contain similar controlled vocabulary types. The semantic web is structured with a controlled schema like Dublin Core and SKOS and can be used to link controlled vocabularies together with semantic relationships. Subject indexing languages would include subject headings like the LCSH whereas classification schemes would include the Library of Congress Classification. Taxonomies are usually organized with facets and have a hierarchical structure similar to subject indexing languages. Differing from subject indexing languages, taxonomies are not restricted to subject or topical indexing. Ontologies overlap many of these because an ontology structure often has components found in other types of controlled vocabularies (Garshol, 2004; Madalli et al., 2015).

Unlike Raghavan and Krishnamurthy (2013), Madalli et al. (2015) do not mention thesauri in their analysis. Thesauri are defined as:

> Controlled and structured vocabulary in which concepts are represented by terms organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms…the purpose of a thesaurus is to guide both the indexer and researcher to select the same preferred term or combination of preferred terms to represent a given subject. For this reason a thesaurus is optimized for human navigation and terminological coverage of a domain (ISO-25964-2, 2013, p.14).

While thesauri can be considered controlled vocabularies, they are not a specific knowledge organization structure. Instead, they borrow knowledge from the organizational structures of other vocabulary types. In this way, they too overlap different controlled vocabulary types similar to ontologies (Hedden, 2016). Essentially, ontologies and thesauri express relationships between terms in a network structure whereas taxonomies, classification schemes,

and subject indexing languages are most often hierarchical and express relationships only as linear broader/narrower or parent/child relationships. While all are acceptable structures for knowledge organization, a literature analysis shows that ontologies and thesauri saturate most studies in knowledge organization (see Table 3, Web of Science, search conducted by researcher, 2016).

Table 3: Knowledge Organization Literature Analysis.

2016 Knowledge Organization Literature	n=	%
Ontology/ies	87	13%
Thesauri/thesaurus	70	11%
Taxonomy/taxonomies	30	5%
Subject heading/s	12	2%
Controlled vocabulary/ies	26	4%

One reason for this may be that ontologies and thesauri allow for relationships and for flexible knowledge representation, which as Hjorland (2015), Cabitza et al. (2013), Smiraglia (2013), Bergman (2016), Mangisengi and Essmayr (2003), and McKerlich et al. (2013) have shown, is more akin to how users think and search. These Works suggest, while not directly indicating, that ontologies are best suited for mapping multiple vocabularies and other data such as user's natural language, while others have taken this a step further by showing how terminology mapping can be used for better information retrieval through use cases such as query expansion (Azad and Deppak, 2019; Khalid and Wu, 2020), recommendation systems (Wang et. al., 2018; Sun et. al., 2020), and other information retrieval exercises.

Ontologies are initially constructed from hierarchy structures like taxonomies but are then enhanced with more complex semantic relationships like *methodFor* and *theoryOf* instead of only

broader/narrower or parent/child relationships. This is one reason why visualizing ontologies is so difficult (Brachman et al., 1991). The distinction between the typical taxonomy parent/child relationships and the more complex relationships that can be created with an ontology structure can be seen in Figure 3.

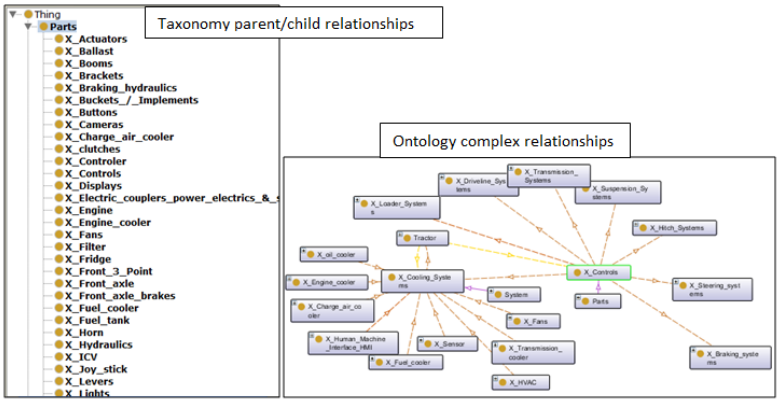


Figure 3: Distinction between taxonomy and ontology.

As Anderson and Hofmann (2006) state, Ranganathan, Linnaean, and Bliss established the roots of a faceted taxonomy where terms are structured hierarchically and may be aligned to form more complex subject tags but these tags are not naturally formed (dog -- food --recipe is an example) so they are more difficult to string match in search. Anderson and Hofmann (2006) also state that hierarchies add meaning to terms based on where they are placed within the hierarchy and Pirolli and Kairam (2013) state that this helps users learn from the architecture of vocabulary but this is an artificial organization or assumed logic (as Hjorland (2015) points out) of the organizer of the terms, i.e. the indexer in many cases, which may limit the exploration and innovation of learners. Cleverley and Burnett (2015) found that faceted search significantly improved knowledge research, and which Giess et al., (2007) agree, and Niu and Hemminger

(2015) found that facets were used 90% of the time to refine repository search queries. They concluded that:

> Facets improve the interactions between searchers and catalogs under some specific situations. Faceted searches demonstrate statistically significant improvements in search accuracy for complex and open-ended tasks. In addition, facets are used statistically more frequently in complex and open-ended tasks than in simple and close-ended tasks (Niu & Hemminger, 2015, p. 1045).

Taxonomy structures are used the most in popular journal repository search tools such as the Web of Science, Scopus, and Engineering Village. Currently, only a few journal repositories allow knowledge seekers to use knowledge graphs, such as NASA, IHS Workbench, and some medical repositories like Elsevier's Embase (Antelman, Lynema, & Pace, 2006; Cleverley & Burnett, 2015; Niu & Hemminger, 2015). With the amount of evidence that suggests ontology structures improve user knowledge search and satisfaction, the reason for this shortage is unclear.

Ontologies are initially based on hierarchical structures but then are enhanced with customizable relationships derived from unstructured or structured text, adding more semantic context to how things are related. Ontologies define a domain of knowledge and are "a more complex type of thesaurus, in which instead of simply "related term" relationships, there are various customized relationship[s] (Hedden, 2016)." Ontologies and thesauri are similar and in the literature, they are sometimes used interchangeably. Kless, Milton, Kazmierczak, and Lindenthal (2015) found, and which Hjorland (2016) agrees, the similarities and differences between the two are that both use real-world universal concepts, have similar relationship types (such as *whole-part*), and also exhibit specific domain relationships. Kless, et al., (2015) indicate the differences between the two include:

- Thesauri concepts are bundles of terms used primarily for navigation whereas ontologies have separate instances that are more explicitly defined

- Precision of terms in ontologies is preferred over user interpretation, which is what thesauri are focused on
- Ontologies allow term membership to define meaning, i.e computer logic models called reasoners
- Thesauri *is-a* relationships cannot be used for logical inference and are less specific than the more complex relationship designs in ontologies
- Ontologies can more easily be structured as polyhierarchical
- Ontologies are more redundancy-free
- Thesauri and ontologies use *whole-part* and associative relationships differently (such as thesauri whole-part relationships are always bi-directional and ontological relationships are not).

Kless, et al. (2015) summarize that thesauri have strong knowledge organization characteristics and are strongly focused on user-knowledge interaction, but to be used as an ontology, significant structural changes are needed. Nielson (2011) highlights that because thesauri are so user-centered they are useful in engineering research collaboration and team projects because they can accommodate the team's natural language. Nielson (2011) also found that including more natural language in thesauri increases researchers' access to content. Based on the overwhelming evidence that ontology structures support terminology mapping as well as align with how users in the engineering-domain tend to interact and think about information, the current study will use an ontology knowledge structure to record the results of this study and discover if graph-like structures can be used to connect both controlled and user terminology effectively.

2.7 Current state of controlled and uncontrolled terms IR

The core of the present research is focused on assessing whether controlled vocabularies in the engineering-domain align with the natural language of its users. As such, this section will focus on literature investigating the current state of controlled vocabularies as well as the literature addressing how well controlled vocabularies meet the needs and expectations of users.

Traditionally, knowledge organization within traditional knowledge repositories like libraries has primarily focused on indexing content with controlled vocabularies like the Library of Congress Subject Headings (LCSH) for information retrieval, search intelligence, and browsability of content topics. Specifically, controlled vocabularies are used in part to:

> improve the effectiveness of information storage and retrieval systems, Web navigation systems, and other environments that seek to both identify and locate desired content via some sort of description using language. The primary purpose…is to achieve consistency in the description of content objects and to facilitate retrieval (ANSI/NISO Z39.19-2005, 2010, p.1).

Controlled vocabularies have been used for this purpose for quite some time, but with the increase in published content and greater access to digital resources, controlled vocabularies have become even more crucial for indexing content (Joudrey, Taylor, & Miller, 2015). According to the publisher, Emerald Group published materials are estimated to grow to 2 million articles each year, roughly 5,000 new Works a day, from more than 10 million researchers around the world, and these estimates will grow exponentially each year (Leonard, 2016). Part of this growth is due to the growth of open access materials, which often are in pre-publication status and therefore not tagged with controlled terms (Hook, Calvert, and Hahnet, 2019). The growth of more knowledge resources, coupled with the diversity of those now able to access the content due to open access, points to make sure the bridge between content and users queries is flexible enough to handle the influx of new terminology and greater user diversity. Tang (2016), a taxonomist at LinkedIn, states that incorporating user's natural language into discoverability "has a unique and democratic advantage [because] terms that are used most often as categories are probably being chosen by a large audience" and using a crowdsourced natural language vocabulary may give "valuable insight on what people want and what they are looking for." With the accumulation of new research,

knowledge organization that gives every terminological advantage to the researcher is that much more important because it helps the researcher wade through the sea of scholarship at their disposal using the language they are most comfortable with.

Traditional controlled vocabularies, which are empirically based, have struggled to hold relevancy in the Google-culture of the 21st century. In Rolla (2009), the research did a comparative analysis between LibraryThing user assigned terms and LCSH controlled terms and found the two vocabulary methods, pragmatic vs empirical, differed greatly. Similar to Rolla (2019), Leblanc (2020), and Samanta and Rath (2020) both explore supplementing controlled vocabulary indexing with user-generated tags, finding that the users' natural language helps users' understand how librarians index content and from user-generated tags there were 93% more unique terms than controlled vocabulary indicating that while users' natural language complements subject vocabularies, the users' natural language seems to cover different needs than what subject vocabularies address. It is important to note in neither of these studies are the user's natural language is suggested as being part of the vocabulary, only as a supplementary uncontrolled field similar to author keyword tags. Rolla (2009) concluded that user tags can enhance knowledge discovery but they cannot completely replace typical controlled vocabularies. Lu et al. (2010), Adler (2009), and Heymann and Garcia-Molina (2009) come to similar conclusions, but they also compared how users assigned LCSH terms as well as expert indexers. Both studies found that 50% of controlled terms had equivalent natural language terms in a general terminology and that terms with annotations such as synonyms, scope notes, and definitions, had a higher equivalency rating than terms that did not. This would suggest that categorical or broader terminology may have a greater alignment between controlled vocabulary and natural language terms, especially when there are definitions associated with the term. Lu et al. (2010) also found that 2.2% of users'

preferred natural language exactly matched their equivalent LCSH terms and that expert indexers often used different terms to tag content than users. Iyer and Bungo (2011), analyzing the difference between natural language and controlled tagging in the medical domain, also found that terminology between expert indexers and users differed, which Hjorland (2013) also established, indicating less than 1% of tags were exact matches to controlled terms and 46% were partial matches. These studies indicate that controlled vocabularies rarely match the natural language of users without interpretation (which is where a term is a partial match), that natural language does have benefits for information retrieval, but in all of these studies the users' natural language was always treated like an uncontrolled field separate from the controlled subject tags.

Additional research into how well current controlled vocabularies meet the needs of users has been conducted. While total abandonment of controlled vocabularies is not the suggestion from most literature (see Rolla, 2009), change and modernization are. For example, efforts to address user dissatisfaction and antiquated controlled vocabulary terms have resulted in marginal changes, including the sexual orientation terminology (see Berman, 2000; Dabrinski, 2013; Olson, 2007), Roberto (2008) express there are issues with the LCSH handling of gendered headings, which Johnson (2008) and Tierra (2008) support, Exner (2008) and Webster and Doyle (2008) find that North American Indian names are non-existent in the LCSH, Weinberg (2008) found that non-Roman scripts were also poorly represented, and Exner (2008) and Lu, Zhang, and He (2016) found that inconsistent international subject headings were becoming more problematic as repositories grew across domains and cultures. Additionally, Hall-Ellis (2008) found that there were few Spanish subject headings, Strottman (2007) explored the lack of cultural and regional subject headings, and Jiang (2007), Park (2007), and Powell (2007) discussed the need for accurate multilingual subject headings in libraries. Powell (2007) in particular raised the question of LCSH

appropriateness in classifying the content of other cultures. Also, as Del Fiol, et al. (2015) and Mishra, White, Ieong, and Horvitz (2014) found, the effort involved with deciphering controlled vocabulary terms may be harmful in time-sensitive and critical decision situations which are often found in medical, safety, religious, and financial research. Recently, libraries have started to discontinue using these terms even if they are perpetuated in the LCSH, such as Harvard University libraries discontinuing the use of the term “illegal immigrants” (Burgess, 2021) or the University of Oklahoma petitioning the LCSH to update the term “Tulsa Race Riot” to “Tulsa Race Massacre” (OU, 2021). Adding user tags to content, studied by Clements and Liew (2016), Williams (2013), and Yang, (2012) to name a few, have also been heavily explored because users now expect, thanks in part to Google-like search capabilities, to search with their own natural language vocabulary. These additional studies show that empirical controlled vocabularies have been struggling to modernize their language and retain their effectiveness for bridging the gap between users' natural language queries and content, especially for users in underrepresented demographics and terminology they may use in search, further placing users at a disadvantage.

The disconnect between controlled vocabulary terms and user’s natural language has resulted in universities and some public libraries abandoning more traditional vocabularies, and Mai (2011b) and Olson (2007) indicate the ridged parent-child relationships in traditional frameworks and empirical indexing may also be factored in the limited diversity and user-centered terminology. Clements and Liew (2016), interviewing around 700 indexers about their perceptions on user tags, found that an overwhelming amount preferred indexing with LCSH terms, and the majority were not aware of user tag benefits. Adding to this, Yang (2012), found that only 47% of common vocabulary management tools support user tags but found that out of public, academic, school, and special repositories, 49% (n=149 out of 307) of special repositories enabled user tags

in content discoverability. Williams (2013), studying the quality of user tags in public repositories, found that BISAC controlled vocabulary terms only matched 2.8% of user tags and that adding user tags to content metadata resulted in a "more comprehensive [and] representative list of terms (p.7)" for users' content discoverability. The studies conducted by Clements and Liew (2016), Williams (2013), and Yang, (2012) indicate that while user tags seem to be preferred by users and may enable more robust terminologies, especially users of special repositories, professional indexing tends to weight controlled vocabularies higher than user tags if they allow for user tags at all. The inverse to this is tagging mechanisms solely generated by the user, which Agarwal and Sureka (2016), Baldoni, Baroglio, Patti and Rena (2016), and Dame (2016) indicate also exhibit some undesirable characteristics, such as hate-speech and misspellings which were especially present in automated natural language tag generation from social media text, which controlled terminology stewardship would help to mitigate. Even so, authors in favor of user tags indicate that the more vocabularies that can include user-centered natural language, the greater the user self-expression and improved search satisfaction.

Traditionally empirical controlled vocabularies index or tag what the content is about so that users can effectively search for the explicit information within the content. Fagbola (2016), drawing upon the investigations of Fidel (1986), Bajpai (1999), Cleveland and Cleveland (2000), Chowdhury and Chowdhury (2003), Taylor and Jourdney (2009), and Rubin (2010), define knowledge organization as two parts, indexing and abstracting, which are used for

> distilling information into an abbreviated, but comprehensive representation of an information resource(s). They are knowledge organisation tools which usually provide detailed and accurate maps and road signs in the information superhighway…They are information retrieval systems (a device interposed between a potential user of information and the information itself) which provide opportunities to access and retrieve information (p.156).

Fagbola (2016), goes on to state that controlled vocabulary abstraction is a retrieval mechanism “to the millions of subjects housed by such a library which helps or assists a user with a collection of published materials as it relates to a given discipline (p.174).” For instance, if a user was searching Elsevier’s Engineering Village, one of the most widely used engineering scholarship aggregators, for content on the potential dangers to cyclists due to electric vehicle engines being quieter than internal combustion engines, users can search for serendipitous articles with free-text or controlled vocabulary terms to find what they need. Using current empirical methods, controlled vocabulary terms can be strung together to form concepts that connect the researcher’s needs to the content they seek. If a researcher were using a repository that only used empirical methods, terms such as *electric vehicles*, *pedestrian safety*, and *noise* could be used to search for the appropriate content. In Engineering Village, this results in 49 pieces of content to review instead of the 180,000 articles returned with free-text search only. This indicates aligning keyword search, coupled with controlled vocabularies, may positively influence the volume of content found during a specific repository search, something a Google search struggles to do (using the same query, and using Google’s more research focused search Google Scholar returned over 30,000 results, most of which were not the correct context for this search).

That said, the issue with the above approach is that controlled vocabularies focus on the explicit information contained in a piece of content and not necessarily the broader tacit knowledge it contains, explicit subject search does not account for alternative forms from other vocabularies, and most controlled vocabularies do not include subject matches to users natural language -all of which poses a risk to the researcher who might miss retrieving content if they do not use the preferred subject label in their search. The researcher does not only need to be able to effectively articulate what they are searching for in the empirical method, which is rare, but they also need to

"crack the code" to decipher what controlled vocabulary terms the repository uses in an empirical search. Furthermore, searching with only keyword and controlled vocabulary, tacit knowledge may be missed because the peripheral content addressing the user's need may not use the keywords or controlled vocabulary the researcher used. These are two of the biggest issues with empirical knowledge organization today and one of the main areas of inquiry for the current study. As Hjorland (2011; 2012), Cleverley and Burnett (2015), Iyer and Bungo (2011), Jimenez, et al. (2012), Meij and Rijke (2007), Lu, Park, and Hu (2010), Pirmann (2012), Sharif (2009), van Damme et al. (2007), Vandic, van Dam, and Frasincar (2012), Wetzker, Bauckhage, Zimmermann, and Albayrak (2010), Willoughby, Bird, and Frey (2015), and Zhu and Dreher (2012) state, capturing the natural tacit terminology of the user, and aligning that with the controlled vocabulary mechanism of search, is imperative for facilitating a more satisfactory knowledge search for engineering researchers but in none of these studies do the researchers posit a way to align natural language and controlled terms, nor do they outline a reproducible way to measure the alignment between or a way to gather users natural language for use in controlled vocabulary.

Gilchrist (2015) suggests Liddy's (1998) natural language processing levels may be used to extract tacit knowledge from content automatically, but as Agarwal and Sureka (2016), Baldoni, Baroglio, Patti and Rena (2016), and Dame (2016) cautioned, automated mining of users natural language can have some undesirable consequences. Additionally, Wiki data is another popular source for automatically mining users' natural language and while it is a good resource to consult, it too has had issues with hate speech, outdated and offensive terminology, and biases (Shaik, Illievski, and Mostatter, 2021; Gerritse, Hasibi, and Vries, 2021). Even if automated extraction has some drawbacks, the methodology behind the identification of natural language is useful even for traditional linguistic terminology identification. These levels are organized by the least to most

difficult for interpreting the meaning of natural language (see Table 4, Libby's (1998) Natural Language Processing Levels, expressed in Gilchrist (2015)):

Table 4: Natural Language Processing Levels.

Phonological	Interpretation of speech sounds within and across words.
Morphological	Componential analysis of words, including prefixes, suffixes, and roots.
Lexical	Word level analysis including meaning and parts of speech.
Syntactic	Analysis of words in a sentence to assess the grammatical structure of the sentence.
Semantic	Determining possible meaning, including disambiguation of words in context.
Discourse	Interpreting structure and meaning from text larger than a sentence.
Pragmatic	Understanding the purposeful use of language in situations, particularly those aspects of language which require contextual knowledge.

Gilchrist suggests using Libby's (1998) methods for extracting tacit information from content and interpreting meaning from the terminology gathered Gilchrist (2015) concludes that pragmatic interpretation is the most difficult to capture through controlled vocabularies because it is unstructured and domain-dependent. That said, if the user's natural language was included in vocabulary, there is a greater chance that more diverse knowledge would be retrieved and would meet more variations of user queries.

Studies have also found other benefits to using controlled vocabularies and natural language, including serendipitous search, knowledge sharing, and time-sensitive search. Giess et al., (2008) found that in the engineering-domain the terminology between teams of researchers is diverse, and often authors and researchers use different terminology across disciplines. Added to this is the difference between the natural language used among different languages and geographic regions. Giess et al., (2008) also found that engineers in particular often were not familiar enough

with a new topic to know the keywords to use in searching a repository and sometimes used the controlled vocabulary to learn the vocabulary of a new topic –what Cleverly and Burnett (2015) call serendipitous search. Howard, Culley, and Dekonicck (2010) indicate that faceted organization of knowledge, in their study they used an ontology framework, helps engineers stimulate creativity which aids in the discovery of hidden connections between topics, or more tacit knowledge. Howard et al. (2010) also found that those in the engineering field shared their search strategies, which include controlled vocabulary and natural language, with one another to learn more quickly and mentor those new to the field. In addition to serendipitous search and knowledge sharing, Ellsworth, et al. (2015) found that 91% of respondents from the Mayo Clinic indicated they searched for knowledge resources within three hours of their patient interaction, which indicates that knowledge organization is crucial to time-sensitive research. Supporting Ellsworth, et al. (2015), Goh, Giess, and McMahon (2009) indicate that timely and satisfactory retrieval of knowledge resources is also important in manufacturing when recurrent, systematic, or safety-critical error reports are needed to address an issue. The added benefit of serendipitous search, knowledge sharing, and time-sensitive search lends more importance to studies focused on aligning natural language and controlled vocabularies for specific domains.

In knowledge organization, uncontrolled natural language terms are called folksonomies. Folksonomies typically have a crowdsource component to their aggregation. When folksonomies are recorded in some metadata fields, they are very rarely respected as true subject tags because of their uncontrolled nature. With pragmatic user-focused design being primarily concerned with abstracting based on user needs, it is very much in line with users tagging content in their own natural language (Hjorland, 2008; 2013). There has been much research conducted on folksonomic

social tagging and how it affects the satisfaction in the retrieval of knowledge resources, as can be seen in Table 5.

Table 5: Studies on natural language user tags.

Tagging Site	Contributing research
del.icio.ous	Cui et al., 2011; van Damme et al., 2007
Flickr	Cagliero et al., 2013; Madden et al., 2012; Saab, 2011; Schmitz, 2006; van Damme et al., 2007;Vandic et al., 2012; Xia et. al., 2014
Twitter	Correa et al., 2011.
Wikipedia	Kamran et al., 2011; Suchanek et al., 2007; Tapscott et al., 2006; van Damme et al., 2007; Xia et al., 2014; Zaidan et al., 2011
YouTube	Belem et al., 2011; Kaplan et al., 2010; van Damme et al., 2007; Madden et al., 2012; Mahapatra et al., 2013
Other	Dong, Wang, & Liang, 2015; Huang, Lin, & Chan, 2012; Wang, Jiang, Huang, & Tian, 2012

Zaiden et al. (2011) in particular look at the knowledge of the crown method which is an underlying component to folksonomy tags. In their study, Zaiden et al. (2011) use crowdsourced enhancement from the Semantic MediaWiki, reported to be the most used semantic wiki by academics, to explore how knowledge within groups is shared collaboratively through user tags. This study is heavily influenced by Piaget's (1995) contribution, which suggests that group interaction motivates the sharing of knowledge, and which Hjorland (1998), Hislop (2013), and Oborn and Dawson (2010) also agree with. Zaidan et al. (2011) and Kitzie et. al (2020) also indicates that knowledge sharing can occur through linguistic exchanges, like writing or speaking to one another, and that this tacit knowledge is important to the dissemination of knowledge in communities, social media being one of the most used applications of folksonomies.

In part, folksonomic tags have become mainstream due to their use on Twitter, Facebook, LibraryThing, and LinkedIn, among others. Interestingly, Panke and Gaiser (2009) found that 56%

of self-taggers use 1-3 terms, and 41% use 4-7 terms to tag content which aligns with the standard practice for indexing with controlled vocabularies (D. Myers, personal communication, January 12, 2016). Panke and Gaiser (2009) and Semanta et. al. (2020) both found that users tend not to tag with terms associated with methods and procedures, according to Hjorland (2015) and Szostak (2008), is needed for engineering information retrieval. Additionally, Panke and Gaiser (2009) found that users tag with terms associated with domain topics the most followed by media and genre type tags, again very much like indexing with controlled vocabularies (Panke & Gaiser, 2009). Domain-specific tags may be more common because, as Zaidan et al. (2011) state, peer collaboration in disciplinary teaching is essential for the learning process and helps record and transfer knowledge through interactions and communications. From this research, it can be seen that harnessing natural language through user tags to record knowledge from researcher interaction, particularly in domain-specific contexts, may be one way to extract users' natural language for enhancing controlled vocabularies.

Thus far, the literature reviewed shows that broader level terms tend to align between user-generated terms and controlled vocabularies but Balatsoukas and Demian (2010) indicate for specific engineering terminology, the more granular terms the better because of the highly structured research those in the engineering field are accustomed to interacting with. Balatsoukas and Demian (2010) define granularity as “the size, decomposability, and extent to which a resource is intended to be used as a part of a larger resource…more granular digital resources are larger and are composed of smaller pieces (p.454)” such as the abstract, introduction, sections, and subheadings are granular pieces of the overall article. Similarly, Cleverley and Burnt (2015) show that 24% of the researcher’s time is spent on seeking information which is made difficult because 80% of the terminology used for research by engineers in the oil and gas community does not

agree with the controlled vocabularies. Szostak (2008) explores this issue and finds that cross-disciplinary vocabulary usage is possible via ontological frameworks, term-to-term mappings, and crosswalks. He also states that the researchers' methods and testing designs should be terms within the vocabulary and using these in indexing is important for triangulation, reproducibility, and de-duplication of efforts within engineering research. Szostak (2008) further states term mapping controlled vocabulary terms from multiple domains is "critical for interdisciplinary scholarship" and "as research and knowledge become more interdisciplinary, the academic subjects represented in our research libraries become increasingly ill-suited to the conduct of research (Szostak 2008, p.319; 320)." Mapping terms have been used as a means for organizing vocabulary terms to help users access knowledge resources across domains and may be used as a vehicle to connect controlled vocabularies to the natural language of its users. See Panajotu (2012), Azuara, Gonzalex, and Ruggia (2013), Bekhuis et al. (2013), Bedford, Greenberg, Hodge, White, and Hlava (2010), White (2013), Szekely, et al. (2013), Dumontier and Wild (2012), and Samwald et al. (2011) for more examples where controlled vocabularies have been mapped for increasing access, furthering discovery, and incorporating natural language in query expansion.

Mapping terms is equivalent to establishing a relationship between terms from different sources. These relationships, usually exact or partial equivalencies, can be recorded through taxonomy hierarchies, thesaurus notations, mapping tables such as crosswalks, or term-relationship ontology frameworks (Alemu et al., 2012). For example, while searching in the repository's discovery portal, the researcher could use the Transportation Research Board (TRB) controlled term *Magnetic levitation* or the SAE International Digital Library term *Electromagnetic suspension* to search for content on Maglev trains if the two terms had been mapped to one another the users' query would retrieve content which has both tags, even if the user did not know they are

synonymous. In equivalency analysis, some have reported as low as 9-8% agreement between terms (in an information science vocabulary) and as high as 62% (in an English to Chinese medical controlled vocabulary) but these term to term mapping assessments usually look at only controlled vocabulary terms, not how well the control terms match the users' natural language (Wang, Bales, Reiger, and Zhang, 2016; Xu et al., 2015). Hjorland (2015), Szostak (2008), Wang, Bales, Reiger, and Zhang (2016), and Cleverley (2016), agree that by assessing alignment between interdisciplinary vocabularies there is a high likelihood that researchers will be able to tap into new and previously inaccessible knowledge resources. Hjorland (2015) and Szostak (2008) state that this will in all likelihood make for richer and more robust studies as well as decrease the likelihood of duplicating research and findings. As Szostak (2008), de Andrade and de Lara (2016), White (2013), and Shiffrin and Borner (2004) indicate, mapping terms may connect the silos that have been generated by dissimilar vocabulary for interdisciplinary knowledge access. These studies indicate that including term equivalency mappings into vocabularies may assist the researcher in their knowledge discovery and facilitate greater cross-disciplinary knowledge transfer.

The issues surrounding controlled vocabularies, coupled with the extensive volume of knowledge resources available, have spurred initiatives to switch from an empirical information model to more user-centered models. A major part of this switch is incorporating users' natural language into knowledge organization models. Folksonomies have become a popular tagging technique but because users' natural language is not consistent or reliable enough on their own, as Gruber (2008), Kroski (2005), Guy and Tonkin (2006) Madden, Ruthven and McMenemy (2012), Mika (2007), Suchanek, Kasneci, and Weikum (2007), Xia, Peng, Feng, and Fan (2013), have found, user tag inconsistency, as explored by Thomas, Caudle, and Schmitz (2010), are not

considered robust enough to be weighted the same as controlled subject tags. Many of the other studies reviewed would agree that folksonomies on their own are not strong enough to provide effective information retrieval but instead need to supplement the more controlled terminology of subject vocabularies. As Schill, Truyen, and Coppens (2007) state,

> The issue is not whether an individual tagger has correctly identified ("tagged") a reference. What tagging essentially does is link a concept to its social practice… [natural language tags] connect the objects involved and the correlated concepts to activity clusters in a community (p.107).

Sharif (2010) extends this and proposes that folksonomic tags will have a significant role in the Semantic Web and that ontology structures may support natural language folksonomy tags as well as more controlled terminology but the application of natural language mapped to controlled terms was not undertaken. Lawson (2009), Kakali and Papatheodorou (2010), and Pirmann (2012) also support this sentiment.

This inconsistency and lack of natural language to controlled term assessment methods may be why some studies have dismissed using folksonomies at all. Maggio et al., (2009) found that user tags were not widely used in the medical community because they fostered more classification errors, although reviewing Kipp (2011), Batch and Yusof (2015), and Gasson (2015), folksonomy tags seem to be gaining popularity for information literacy benefits. See Figure 4 for an example.

Figure 4: Medical repository using natural language.

Yang (2012) found that only 47% of repositories turn on user tags in their knowledge organization tools and Porter (2011) summarizes that the most commonly-cited disadvantage of user tags "are a result of their lack of semantic and linguistic control, which, ironically, are also their greatest strengths (p.251)." Almost all studies reviewed indicated there are a great many advantages to incorporating user's natural language in information retrieval but their unstructured, uncontrolled, and inconsistent nature is difficult to overcome.

That said, more studies have suggested using natural language alongside vocabularies is promising, albeit still difficult to implement effectively. Griffis and Ford (2009) found that "subject liaisons," or subject matter experts, offered high-quality tagging keywords that facilitated better search in digital repositories. Strader (2009) builds upon the work of Maggia (2009) by analyzing the similarities in author keywords and controlled vocabulary terms. Tuominen, Talja, and Savolainen (2002), Martin-Moncunill, Garcia-Barriocanal, Sicilia, and Sanchez-Alonso (2015), and Engerer (2016) agree that recording natural language is equally important to knowledge discovery as controlled vocabularies. Porter (2011) and Annfinsen, Ghinea, and Cesare (2011) support Cleverly and Burnett (2015)'s findings that folksonomies and controlled

vocabularies “stimulate new needs, improving a system’s ability to facilitate serendipity (p. 97)” which they also report as the discovery model used by engineers. Annfinsen et al. (2011), one of the most cited articles on folksonomic applications in library contexts, focuses on implementing folsonomic tags in an academic library setting to assist users with limited knowledge of a subject area and who were trying to develop research questions. Annfinsen et al. (2011) found that “through the use of tagging and the development of folksonomies users felt more able to browse resources, as well as search for specific material (p.65).” More studies, conducted by Pera, Lund, and Ng (2009), Noorhidawati, Hanum, and Zohoorian-Fooladi (2013), Oyieke (2015), Ibba and Pani (2016), among others, have all explored case studies where folksonomic tagging was used alongside controlled vocabularies to the effect of more user satisfaction. Acknowledging that user tags are beneficial to user satisfaction during information retrieval and that folksonomies on their own are not consistent enough for satisfactory knowledge organization, some have argued a hybrid approach is needed. These studies show users' natural language and vocabularies can benefit from one another, but most do not indicate to what extent the two align or what impact that potential alignment has on user satisfaction and discoverability.

Authors who support a hybrid controlled and uncontrolled vocabulary that includes natural language include Strader (2009), which conclude that a hybrid approach improves information retrieval; Pirmann (2012) assessed systems’ capabilities to accommodate user tags and found that even though there is potential in folksonomies many repository systems cannot accommodate them; and Kakali and Papatheodorou (2010), which finds through an academic case study that folksonomies are so beneficial to user satisfaction when search for knowledge that current knowledge organization methods need to be reassessed to include them. Of particular note, Bouadjenek, Hacid, and Bouzeghoub (2016) explore Social Information Retrieval (SIR), which

enhanced the information retrieval process through social media content and review the major contributions to SIR to create a taxonomy of SIR functions. Bouadjenek et al.'s(2016) study is of interest because it outlines the major motivations behind aligning controlled and natural vocabularies, many of which have been highlighted in the current review. Figure 5 shows the full SIR taxonomy derived from the literature. While all of these features can be influenced by the current study, this study is primarily concerned with the search aspect of the taxonomy tree, which is where "social information is used to improve the classic IR process, e.g., documents re-ranking, query reformulation, and user profiling (Image is reproduced with permission from Bouadjenek, Hacid, and Bouzeghoub (2016); Bouadjenek et al., 2016, p.5)."

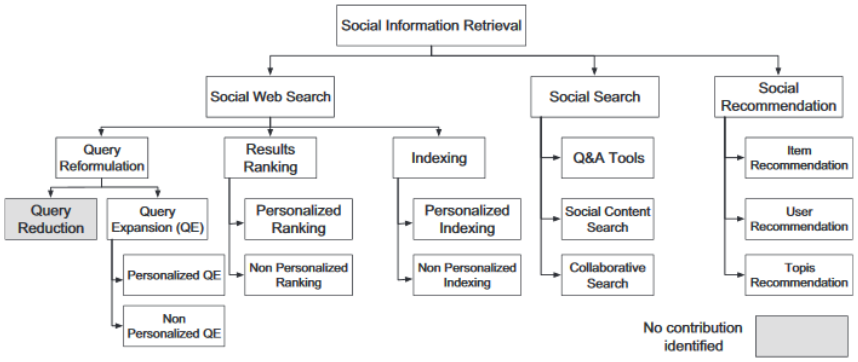


Figure 5: Taxonomy of Social IR. Recreated with permission.

Further examples of the hybrid approach can be seen in van Damme et al. (2007) and Sharif (2009), both of which will be discussed in detail because they are the closest in purpose, methodology, and scope to the present study.

Van Damme et al. (2007) developed a hybrid classification scheme called a folksontology, in which folksonomic tags are converted into an ontology. The study examines tags from Flickr, del.icio.us, and Technorati. After extracting the folksonomic tags, the authors supplement tags

with controlled vocabulary terms derived from -mainly Wikipedia, Leo dictionaries, Google suggestions, and Wordnet. Mapping methods are used to connect the folksonomy to the controlled vocabulary through an ontological framework of entity relationships. The findings are not focused on the accuracy of folksontologies so much as the process in which they built the ontology. The processes they use differ from Braun (2007), Maier and Schmidt (2007), and Specia and Motta's (2007) ontology generation model by using a "mash-up" of lexical enhancements (van Damme et al., 2007). Even though van Damme does not specifically assess the user-centered design, nor do they study terms in the engineering-domain, much of their procedure can be used for the current study.

Similar to van Damme et al. (2007), Sharif (2009) explores how to map folksonomic tags to controlled vocabularies by way of an ontological framework. They explain that folksonomies and controlled vocabularies are not mutually exclusive but they are at opposite ends of the knowledge organization spectrum; they can be used together. Sharif (2009) uses the ontology creation tool Protégé, as will be proposed here, and derives the folksonomy from natural language sources such as social media, surveys, and general human interaction. Their rationale is that this will increase the amount of tacit knowledge increased for further knowledge retrieval improvement. They do not, however, use user-centered design to map controlled vocabulary terms to their equivalent natural language counterparts. They instead use automated word co-occurrence to find equivalency. They propose visualizing the ontology so that both the controlled and uncontrolled terms are accessible. The mapping model is then crowdsourced to assess which term is preferred based on the user's individual needs. This is another example that supports using a pragmatic user-centered rather than an empirical method in the hybrid approach. Sharif (2009)

concludes that aligning folksonomies with controlled vocabularies in an ontological framework increases search precision and recall, which may lead to more user knowledge search satisfaction.

2.8 Summary of Reviewed Literature

A review of the literature suggests that tacit information is most effectively gathered through pragmatic methods and captured in ontological frameworks. The literature also suggests that although knowledge organization is a component of knowledge management, it does not receive nearly as much coverage within the body of work currently available and thus further research may be beneficial. The literature lists the possible benefits of more information science knowledge management studies as better knowledge acquisition, classification, utilization, and dissemination of research. It is anticipated that the current research will positively influence the assessment of user-centered engineering-domain controlled vocabularies. The studies reviewed also suggested that capturing tacit knowledge through natural and controlled vocabulary is imperative for facilitating a more user-centric information discovery experience.

As the amount of information increases, so too does the difficulty of finding knowledge within it. The reviewed literature suggests that when a user performs their search, they prefer to use their own language and trust that they will retrieve the research they seek, even if they do not know the “correct” subject headings tagged to that content, especially terms which are shared across disciplines. This effectively helps them to research across multiple domains, to find studies that directly pertain to the methods they will use, and identify work that contains knowledge most pertinent to their research without missing content because of query-to-subject misalignment. The studies directly concerned with engineering knowledge organization and search suggest theory and

method terms may help strengthen the scientific process and that by including theory and method types in controlled vocabularies, scholars would stop “re-inventing the wheel” and would more effectively identify fresh lines of inquiry (Szostak, 2003). The studies reviewed also indicate that because ontological structures more closely align with human cognition, it is likely that structuring vocabularies as ontologies may increase user-centered vocabulary furthering researchers’ satisfaction with knowledge discovery. Not only do ontologies better represent tacit knowledge than other structures for knowledge organization, but studies also suggested that ontology structures may be a promising framework to test the merging of user’s natural language and controlled vocabularies because their structure allows mapping between controlled and user-centered terminology. This is valuable because studies also indicate that controlled vocabularies struggle to stay current and updated, and with supplemental user-centered terminology, they may prove to be more effective in search, especially for those in underrepresented languages and demographics whose terminology is not typically well-represented in controlled vocabularies.

This suggests that aligning natural language with controlled vocabularies may positively influence the volume of appropriate content found during a specific repository search. It also suggests that if the user’s natural language is aligned with controlled vocabularies, there is a greater chance that more diverse knowledge would be retrieved and would meet more variations of user queries. Aside from this, studies have found other benefits to using controlled vocabularies and natural language together, including during serendipitous search, knowledge sharing, and time-sensitive search. The added benefit to serendipitous search, knowledge sharing, and time-sensitive search lends more importance to studies focused on aligning natural language and controlled vocabularies for specific domains, because it increases the discoverability of information for a wider variety users who may use different terminology than the subject terms to search.

From this research, it can be seen that harnessing natural language through user tags to record knowledge from researcher interaction, particularly in domain-specific contexts, may be one successful way to include tacit information in knowledge organization and retrieval. Domain natural language also allows for greater term flexibility. Acknowledging that user tags are beneficial to user satisfaction during a knowledge search and that folksonomies on their own are not consistent enough for effective knowledge organization, some have argued that a hybrid approach is needed where using pragmatic methods, aligning folksonomic user natural language with controlled vocabularies in an ontological framework, will likely increase search precision and recall (i.e. the effectiveness of the knowledge search) as well as the effectiveness and satisfaction of user search.

The studies reviewed here have revealed a disconnect between current knowledge organization methods and users' needs but most suggest approaching this problem with two separate approaches: either update controlled vocabularies on a case-by-case basis, an effort that does not seem to have had much success or have users tag the content themselves, which also has its disadvantages. Very few of these studies have suggested *aligning* users' natural language and controlled terms, which in turn leaves a gap in the body of research. How often do users' natural language match controlled terms, and can they be more closely aligned in a hybrid way that falls outside what most have previously tried? This is the main question this research seeks to answer.

The current research will undertake to understand to what extent are vocabularies user-focused, and what findings can be derived from that assessment. The literature reviewed here seems to indicate that it is likely that controlled vocabularies will align to some degree but no repeatable method has yet to be suggested that measure to what extent, and if misalignment is identified, what impact this has, and what steps may be taken to mitigate the misalignment. The

proposed methods for approaching these questions will be explained in the following Methods section.

3.0 Outline of Methods

In this chapter, the three stages proposed for this study will be outlined, as well as the methods used to conduct the research. Overall, this study uses vocabulary matching, user-centered design card sorting methods, and chi-squared goodness-of-fit methods. As Hjorland (2015) states, pragmatic knowledge organization is semantic and is more in line with the user's natural language –particularly those in specific domains. Because pragmatic methods emphasize the user and domain group as the focal point, the methods used here are also influenced by utility theory and Hobbes' theory of the good. Utility theory states that a user's conceptualization of "good" is centered on the satisfaction of their preferences (or, as Hobbes states, their desires) (Chung, 2012; Hobbes, 1963). Applied to users' impressions on knowledge organization terminology, these theories suggest that user satisfaction is directly tied to their preferences -or user warrant, as Hider (2015) labels it. Therefore, user warrant will serve as one of the main indicators to explore the hypothesis of this study into how well controlled vocabulary terms from the engineering-domain match users' natural language.

3.1 Warrant as a framework

Warrant is a methodology that is essential to knowledge organization. Warrant is defined genetically as the authority an indexer invokes during vocabulary creation and content tagging to justify and verify decisions about terms, their relationships, and their assignment (Beghtol,1986). Barite (2014), outlines nine aspects to warrant:

1. control of synonyms and variants,
2. control of equivalences,
3. control of homonyms and polysemy,
4. control of abbreviations,
5. writing scope notes,
6. writing definition notes,
7. writing historical notes,
8. control of hierarchical relationships, and
9. control of associative relationships

While many might assume vocabulary creation and tagging content is not a critical or weighty exercise, the simple act comes with great responsibility. Defining what words have merit enough to be added to a controlled vocabulary, and dictating how these words are defined and how they should be used to tag content, is the act of defining what something is and is not. This responsibility has been reviewed by Olson (2007) where they assesses how sub-divisions and hierarchical organization stems from a male logic perspective (and also important to the current study, she expresses a more explicit relationship than "related" to help break down the male organization of knowledge); Drabinski (2013) where they walk through Queer theory and how it shifts the primacy of indexers from authoritative definers, stating terminology is too volatile to be officially fixed in controlled vocabularies, to more holistic interpreters and educators for imparting information literacy; Adler (2016) who advocates for taxonomic reparations for underrepresented peoples information journey, from black, indigenous, and LGBTQ+, for more inclusivity and decrease bias in terminology and indexing activities; and Martin (2021) outlines the need for a controlled vocabulary and knowledge management code of ethics to help align literary and user warrant, which may help curb the biases in the indexing as well as create better access points for users to access information. These studies are a small sample of research showing the importance and gravity of subject creation and how literary warrant, while useful for summarizing what content is about, needs to be supplemented to meet the needs of the end-user.

There are several warrant types (defined by Barite in ISKO, 2017); however, the three types used in this study are literature and scientific, both using authoritative sources from literature and controlled vocabularies respectively, to derive terms and indexing logic, and the last is user warrant which is the main focus of this study. The different types of warrant focus on the primary form of evidence, or where the term and tagging logic derive when making their terminology decisions. Warrant has been used to determine the intent or rationale behind controlled vocabulary in several studies, a few to note are Bullard (2017), Wilson (1993), Choi (2018), and Fox (2015), all of which cover various warrant types but all include literary warrant. The main difference between user warrant and literary warrant is whether the indexer believes that users are qualified to determine the best language for the system (Clare Beghtol, 1986; Hjørland, 2013b). The current study will use warrant in a similar way by using warrant as the framework to interpret the matching and chi-square analysis for each Research Question (RQ). Each RQ will build evidence to understand if user warrant, evidenced with user terminology matching controlled terms, was used for the vocabulary creation. The expectation is they will match because vocabulary indexing exists to help users find the information they seek but is this truly the case? This study seeks to explore the hypothesis in the following way.

To assess whether controlled vocabularies in the engineering-domain align with the natural language of its users, and investigate the ramifications of these findings in light of research discoverability, this study will address the following four research questions, each of which will build upon each other to help address the null hypothesis (Table 6):

Table 6: Overview of the Study Design.

H_0= Controlled vocabulary terms match the users' natural language to the expected match rate. H_a= Controlled vocabulary terms match significantly less than the expected match rate.		
Step	**Focus**	
Pre-stage	Establish which terms will be assessed throughout the study, selecting engineering vocabularies for Control terms, and establishing a baseline expected distribution of matches to be used throughout study	
Step 1 (RQ1)	**RQ1:** To what extent do expert terminologies match controlled vocabulary terms? Assessing literary warrant via literature-to-control term matching	**RQ4:** To what extent do controlled vocabularies match the user's natural language without fuzzy search also being applied?
Step 2 (RQ 2)	**RQ2:** To what extent do controlled vocabularies match one another? Assessing scientific warrant via controlled vocabulary matching	
Step 3 (RQ 3)	**RQ3:** To what extent do expert and controlled vocabularies match users' natural language? Assessing user warrant via user to control term matching	

The research questions are designed to build upon one another to understand how the more common approaches, first literary and then scientific warrant (ANSI/NISO Z39.19, 2005; Bullard, 2017; Barité, 2018), match one another to form a baseline match-to-no match threshold for terminology, followed by assessing how well users' natural language matched the baseline. All warrant types are equally valid; however, which warrant type specifically can indicate who the indexing is serving during the search process. This study is focused on the user terminology and how well the engineering-domains' controlled vocabulary serves the users in this space. First, a pre-mapping step is required to identify which terms will be used through the study for assessment, selecting which engineering vocabularies will be used, and establishing baseline expected distribution of match to no match to be used throughout the study. Following this, RQ1 explores whether the control terms match expert vocabulary to test for literary warrant and findings on how to better align the two; RQ2 explores how vocabularies match one another to test for scientific warrant and findings on how to better align these sources; and RQ3 explores how the control set or baseline created from RQ1-2 matches the users' natural language to test for user warrant and

suggested changes to align the two. The mapping done in each research question will help determine the expected match percentage to be used in the chi-square assessments throughout. Each question will have its own section with analysis at the aggregate and individual term levels, findings, and discussion, and the findings from each research question will build to accept or reject the null hypothesis focused on how well engineering-controlled vocabularies align with users' natural language.

3.2 Overview of study design

This study is organized into one pre-stage baseline exercise, followed by three main stages with one search question being covered for each section. The table below (Table 7), shows the research question, research focus, planned approach to answering the research question; and lastly, the methods which will be used for each stage of research.

Table 7: Overview of the Study Design- Detail

H_0= Controlled vocabulary terms match the users' natural language to the expected match rate. H_a= Controlled vocabulary terms match significantly less than the expected match rate.			
Step	**Focus**	**Approach**	**Method**
Pre-stage	Establish baseline expected distribution of matches to be used throughout the study	Based on the mapping, the distribution of match to no match is recorded for chi-square assessment in RQ1-3	Hub/knowledge graph matching
	Select engineering vocabularies to use throughout the study	Using a general vocabulary (SAE vocabulary) to map common engineering vocabularies (LCSH, TRT, NASA, and INSPEC) to determine which vocabulary will be used as the mapping source for the results of this study	
	Establish which control terms will be assessed throughout the study	Using one of the most recent encyclopedias for machine learning, identify topics with the most literature (showing value potential for assessment to researchers) and gather their definitions	
Concurrent step (through RQ1-3)	**RQ4:** To what extent do controlled vocabularies match the user's natural language without fuzzy search also being applied?	Based on RQ findings, compare match to no match likelihood in both the fuzzy search and browse patterns for each term in RQ	Matching; Chi-squared goodness-of-fit
Step 1 (RQ1)	**RQ1:** To what extent do expert terminologies match controlled vocabulary terms? Assessing literary warrant via literature-to-control term matching	Map and compare encyclopedia (literary) terms to Control terms; Perform chi-square test to assess match likelihood	Matching; Chi-squared goodness-of-fit
Step 2 (RQ 2)	**RQ2:** To what extent do controlled vocabularies match one another? Assessing scientific warrant via controlled vocabulary matching	Map and compare vocabulary to vocabulary (scientific) terms to one another; Perform chi-square test to assess match likelihood	Matching; Chi-squared goodness-of-fit
Step 3 (RQ 3)	**RQ3:** To what extent do expert and controlled vocabularies match users' natural language? Assessing user warrant via user to control term matching	Create a card sort survey to gather users' natural language Match and compare users' entries (user warrant) terms to Control terms; Perform chi-square test to assess match likelihood.	Bi-directional card sort; Matching; Chi-squared goodness-of-fit

The following subsections will begin with the population sample size and distribution method, followed by the rationale and methods behind the steps listed in Table 3.2, mainly why a hub or knowledge graph model is used, how matching is conducted, why the goodness-of-fit test is used and how it is interpreted, and lastly how the dataset for the study is created with matching and card sorting methods. The last sections will detail the position of the researcher as well as the ethical considerations of this study.

3.3 Population, sample size, and distribution

This study will focus on the population within the engineering community, which is defined as anyone studying, working, or interacting with engineering knowledge or knowledge repositories. The engineering community is the population target of this research; however, multidisciplinary research is quite common especially when searching for content on research methodologies which may indicate that it is likely researchers outside the engineering community will also benefit from this study. The results of this study may serve to help those in the general information and knowledge organization research space, those researching classification and knowledge organization, as well as librarians and information professionals in the engineering field.

Unlike traditional power studies that require upwards of 300 or more responses, user experience card sorts have sample sizes that are much smaller, between 15 to 20 responses according to Tullis and Wood (2004), data is extrapolated into 90% of the entire sample. Some card sorts, such as Baxter, et al., (2015) who used a sample of around 400, are closer to power study sizing, but the typical range seems to be similar to Spencer (2009) who used 18 and 21

participants and Olson and Wolfram (2007) who used a sample size of 64. Because of the range of card sorting sample size, it is assumed that by keeping within the sample size range indicated by the literature, that the current study's sample size (targeting 50 responses based on typical response rates from ASEE membership communications) is acceptable. Engineering communities such as the University of Pittsburgh and Carnegie Mellon Engineering Schools, as well as engineering societies such as the Institute for Electrical and Electronics Engineers (IEEE), Society of Automotive Engineers (SAE), American Society for Engineering Education (ASEE), and the Special Libraries Association (SLA), were consulted on the research questions, the perceived value of this study, as well as the best method of distribution for the study. From these universities and societies, the ASEE was selected as the main distribution of the survey because their membership is part of the engineering community, they have three main membership distribution methods through Facebook and Twitter, and their membership is quite active compared to other societies considered. Another main contributing factor is that ASEE was one of the only distribution methods contacted by the researcher that was willing to send the survey to their membership on behalf of the researcher.

Distribution through the membership communication is ideal due to the ASEE membership requirements that would assist in targeting the sample population, ASEE membership includes educators from Preschool-12th grade, engineering students, or "persons occupying or having occupied responsible positions in engineering instruction, research, or practice, and other persons having a demonstrated professional interest in engineering education. They may come from within or without the academic environment" (ASEE, 2021). The contacts at the societies were consulted to select divisions from their membership that represent the target sample group. The following

divisions were selected based on education, machine learning, information retrieval, and/or a general engineering focus (full descriptions for each can be found on the ASEE website (2021)):

- Computers in Education
- Computing & Information Technology
- Design in Engineering Education
- Educational Research and Methods
- Electrical and Computer Engineering
- Engineering Libraries
- Engineering Technology
- Software Engineering Division
- Systems Engineering Division
- Technological and Engineering Literacy/Philosophy of Engineering
- Women in Engineering

The researcher provided the text for ASEE to send to their membership. The text was:

> Calling all those interested in improving tags and engineers' search for machine learning/data mining research materials!
>
> Please consider participating in a survey geared toward gathering the natural vocabulary used for finding machine learning (ML) concepts. The data will be used in an upcoming dissertation study from the University of Pittsburgh focused on understanding how well users' natural language aligns with controlled vocabulary tags, in the hopes of improving search and to educate researchers on different search terms for ML techniques.
>
> To be eligible for the survey, you must have worked in the engineering field for at least 2 years or have a masters' degree in an engineering field. Please consider taking and sharing the survey, all entries are anonymous.
>
> As a thank you for participating, your email (if you choose to provide one) will be eligible for a drawing for a $50 Amazon gift card.
>
> [link to survey]
>
> #ML #machinelearning #seo #vocabularybuilding

The survey link was provided in the text and stayed open for one month. Entries are anonymous and the data is stored in the secure University of Pittsburgh Qualtrics system. Participants had the opportunity to volunteer their emails and had the opportunity for a $50 Amazon gift card at the end of the survey. These emails were contained separately from the Qualtrics survey and were only used to send the participants the gift card if they were selected.

While exploring users' natural language alignment with controlled vocabulary, there does need to be a basic understanding of information literacy in engineering to effectively understand the search prompts of this study so a master's level degree in engineering, or at least 2 years' experience in the engineering field, were required to participate. Because of the distribution method, participants were also required to be members of ASEE from the selected divisions. The membership of ASEE are considered literate in engineering terminology to some extent due to their selection of membership in an engineering-specific society, but even experts need help searching for topics new to them so while the sample likely has a degree of familiarity with engineering and machine learning terminology, there is always a degree of difference in the terminology in the search process and therefore results from the survey are likely to be different, but not overly so because the sample will have the same education or experience level. Roles from the sample may include authors, librarians, faculty, post-doctoral student, researchers, research assistants, and library or information professionals, among others who are involved in education and research in the engineering field.

3.4 Hub or knowledge graph mapping model

A hub or knowledge graph dataset will be created to capture the mapped data, as well as the suggested changes based on warrant analysis. The data will be modeled in the most common controlled vocabulary model called Simple Knowledge Organization System (SKOS) and the format of the file will be in RDF/OWL for linked data or downloaded as a CSV for easier use by consumers of this research. This helps with reproducibility and sharing of the results of this study. Both van Damme et al. (2007) and Sharif (2009) use preexisting terminologies as their target

dataset for natural language term analysis. Source terminology, the terms that will be matched to the target vocabulary that will serve as the pref labels in the model, will be mapped to the selected target vocabulary using a hub structure, as specified in ISO 25964, and which Binding and Tudhope (2015) use for effective mapping and Orgel, Hoffernig, Bailer, and Russegger (2015) use to map cultural heritage terminology. Hub structure is used to describe the matching process, and the knowledge graph is a way to model the mappings. Mapping and matching are two very similar aspects of this work, mapping being the conceptual model and matching being the codification of the data into that model, and therefore will be used interchangeably throughout this study.

The mapping is based on ISO 25964, as well as mapping techniques outlined by Binding and Tudhope (2015) and Faith, Tseytlin, and Bekhuis (2014). How one term is related to another is a loose form of Category Theory, defined as a category, or an abstracted form of a real-world concept, having a relation, or arrow, to an object or set of objects (Simmons, 2011). In the case of this study, the category will be the target vocabulary term, the matching type will be the relation or arrow to the equivalent terms from source vocabularies being mapped to the target vocabulary, as Figure 3.2 shows. Category Theory models how a source vocabulary term can be mapped to a target vocabulary term. Or put into W3C wording, a subject, predicate, object triple structure of graph data (W3C, 2004). This mapping structure will be used to identify and map exact, partial, and no match controlled vocabulary terms and users' natural language labels to one another. An example of this is in Figure 6:

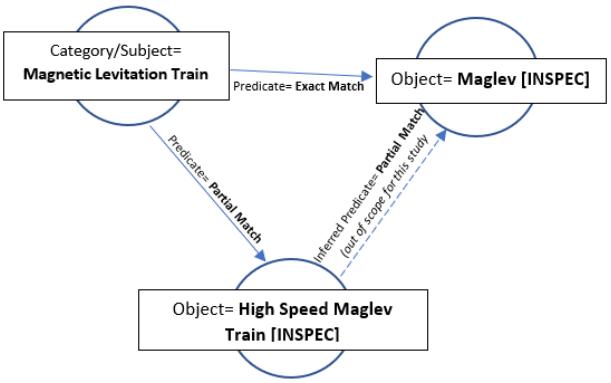


Figure 6: Triple mapping between vocabulary terms. Researcher created.

Where the target vocabulary term is the Category or subject, and the source vocabulary and users' natural language terminology are the mapped objects, with the match type as the arrows. Unfortunately, the subject-predicate-object wording is so similar to subject vocabulary terminology that they often get confused so while this is the mapping technique used in this study, the researcher will refrain from describing the mapping with the W3C terminology. The codification of these triples, or matches, will be done using the SKOS controlled vocabulary namespace. The matching itself, deciding if the strings match between target and source vocabularies (including user entries), will be conducted by the researcher. There are automated means to assess string matches such as multi-dimensional scaling (MDS) and hierarchical cluster analysis, as used by Olson and Wolfram (2007), which would be useful in larger mapping projects, however with the limitations on how many terms a user can review in the RQ3 survey, the data volume in this study does not require automation to be effective. Also, it is the hope of the researcher that the methodology used in this study can be used by librarians in the field, who do not always have access to or knowledge of tools to perform cluster analysis, and therefore a manual approach might be welcome to that audience and for those that can perform cluster analysis, may do so.

A hub structure uses a source vocabulary as a hub with which to map other target vocabularies (which are the spokes). Because the target vocabulary is the only vocabulary to match the source vocabulary strings to, the amount of mapping involved with four or more vocabularies significantly decreases (n^2-n compared to 2n) (Binding & Tudhope, 2015). The mapping will be recorded using Protégé, a free open source editor often used in vocabulary mapping projects (see Bioportal online for many examples) created and maintained by Stanford University, and which can output linked data formats such as RDF/OWL or plain data formats like CVS. The classes in the vocabulary mapping will be derived from the literature dataset, most likely a dictionary or encyclopedia, to organize the mapping in Protege (see below Figure 7 for example).

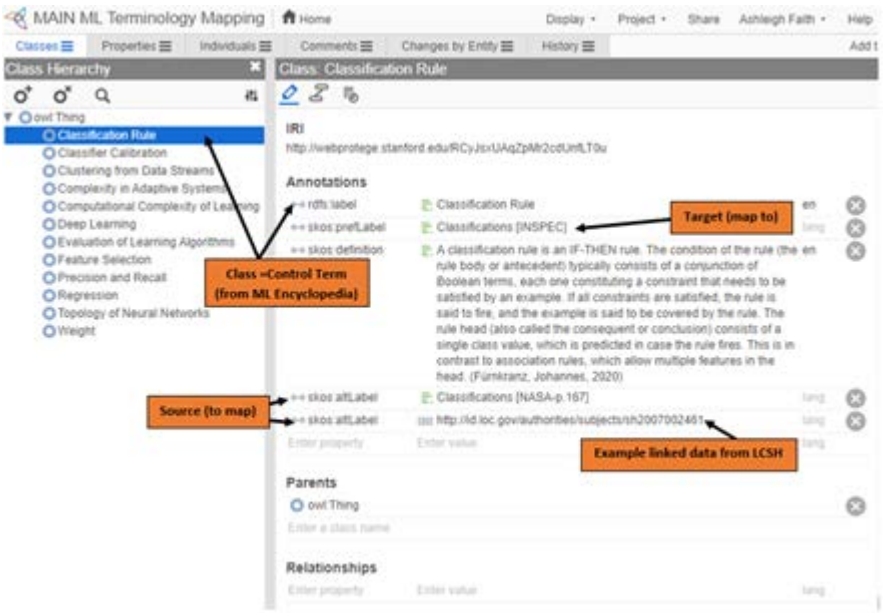


Figure 7: Protege codification of terminology mapping.

The mapping technique, modeled after those used by Binding and Tudhope (2015) and Faith, Tseytlin, Bekhuis (2014), will use equivalent (exact and partial mappings -EM and PM respectively) and no match (NM) as indicators for each term mapping. Both exact (EM) and partial

matches (PM) are assessed via string match by the researcher. Exact is exactly the same string, with stemming exceptions, and partial matches are almost the exact same string. The Target Vocabulary term is the prefLabel and is the vocabulary Source Vocabularies map to (in this case INSPEC is the Target vocabulary, rationale for this is described in the next chapter), and PMs and EMs from Source vocabularies (TRT, NASA, and LCSH) are mapped to their corresponding Class in Protege. The LCSH is the only open linked data vocabulary assessed in this study and therefore will be the only vocabulary with a linked data URI (Figure 3.3 shows what this looks like in Protege). Protege is where the resulting mappings and suggested updates for vocabularies will be recorded. Specifics on mapping types, primarily used for the chi-square assessments, will be documented in Excel tables. In the mapping documentation, the first portion of the PM will be noted as PM1 and the second part of the term, if it has a second part, will be counted as PM2. These were determined by using the definitions for each category documented in ISO 25964 Part 1, where no match is nothing in the target string matches the source, partial is at least part of the target string is found in the source, and an exact match is the exact target sting (stemming is considered a match) is found in the source.

There is also the question of search engine assistance in the user search process. Before information discovery was digital, materials would have a physical location within the library. This required precision in tagging the card in the card catalog which served as the main access point to finding content. This legacy can still be felt in some discovery systems that support controlled vocabulary browsing to find materials. But if the terminology in these browse vocabularies did not match the user's query, or if the user did not understand the organization of the vocabulary, it may prove very difficult to browse for information. To understand the role of the search engine in users' query-to-literature and controlled vocabulary tagged content retrieval

two assessments, fuzzy search and browse assessments, will be conducted for RQ4. For a fuzzy search pattern, EMs and Pms will be counted as a match, the search engine approximates a string match between query and content (Lai, Lien-Fu, et al, 2011) giving the user more assistance in searching for a term they may not be familiar with. The browse pattern will treat only EMs as a match and is defined as a more traditional card catalog or hierarchical browse behavior where the user would need to find the term in a vocabulary to search with it. With this in mind, each research question will explore to what extent fuzzy search or browsed search assist in the matching likelihood of the users' natural language to controlled terms.

This mapping technique will be used throughout RQ1-4 and be assessed with a chi-squared goodness-of-fit assessment to determine the likelihood of the user's natural language matching the controlled vocabulary.

3.5 Chi-squared goodness-of-fit Assessment

The chi-squared goodness-of-fit test is used to identify how confident the observed data matches the expected data, or is there a significant difference between what is expected and what is observed -in the case of this study the expected match rate compared to the observed matches in RQ1-4 (Greene and Manuela, 2005). In knowledge management, the goodness-of-fit tests are common to understand if there is a significant difference between categorical data such as seen in Mohan and Rajgoli (2017) where the goodness-of-fit was used to compare the differences between several cited authors per paper and citations per author between three Astronomical journals; Wang and Wolfram (2015) where the goodness-of-fit was used to compare journal discipline tagging

between the tags the journals documented and those assigned by Web of Science; Olson and Wolfram (2007) where the goodness-of-fit was used to compare inter-indexer consistency; Green (2005) used goodness-of-fit to compare hierarchy level of categories between four different controlled vocabularies; and Schneider and Borland (2004) indicate the goodness-of-fit is one of the most popular methods comparing similarity distance for author co-citation, scientific literature, and citation impact assessments. These studies will be consulted to assist the use of chi-squared goodness-of-fit assessments in the current study, the specifics of which are as follows.

For each term, the total EM, PM1-2, and NM will be calculated in preparation for the chi-squared goodness-of-fit assessment. The baseline expected match will be used to calculate the expected column, and the total match/no match will be in the observed column. The corresponding chi-square score is then used to estimate the p-value. While SPSS can be used to calculate this assessment, the researcher calculated using an excel formula created with the assistance of the University of Pittsburgh Statistics Department. It is as follows in Figure 8.

$H_0: p_1 = p_1^*, p_2 = p_2^*, \ldots, p_r = p_r^*$

Category	Observed	Probability	Expected
1	O_1	p_1^*	$E_1 = np_1^*$
2	O_2	p_2^*	$E_2 = np_2^*$
⋮	⋮	⋮	⋮
r	O_r	p_r^*	$E_r = np_r^*$
Sum	n	1	n

$$\chi_0^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \Rightarrow \begin{cases} \text{if } \chi_0^2 \geq \chi_\alpha^2(r-1), \text{ then reject } H_0 \\ \text{if } \chi_0^2 < \chi_\alpha^2(r-1), \text{ then fail to reject } H_0 \end{cases}$$

Figure 8: Goodness-of-fit Chi-Test Calculation. Researcher created.

For this study, the category is Match and No Match; the observed is the sum of the matches between target control terms and the source data (literature for RQ1, other controlled vocabularies for RQ2, and finally user entries for RQ3); the probability is established from baseline vocabulary matching exercise in the Pre-step before the RQ assessments, and validated through the RQ1-2 assessment matching exercises which were calculated to be 50% match to no match (further details on the calculation of this in the following chapter), and expected is calculated from the total observed divided by the expected distribution probability. The chi-square is determined by the following calculations discussed in Greene and Manuela (2005), and calculated in Excel as shown in Figure 9:

	B	C	D	E	F	G	H
6	Category	Observed N	Probability	Expected N		chi-squre	=(C7-E7)^2/E7+(C8-E8)^2/E8
7	Match	n	0.5	=$C$9*D7		df	=COUNTA(B7:B8)-1
8	No Match	n	0.5	=$C$9*D8		p-value	= CHISQ.DIST.RT(H6,H7)
9	Sum	=SUM(C7:C8)	=SUM(D7:D8)	=SUM(E7:E8)		result	=IF(H8>0.05,"Accept Ho","Reject Ho")

Figure 9: Goodness-of-fit Chi-Test Calculation.

The degree of freedom for a chi-square test is calculated as df =number of categories minus 1, in this case, df is 2-1 = 1. For p-value, the significance level is .05 to limit the chance of Type 1 and Type 2 errors and assessed with the chi-square p-value lookup in Excel.

Interpreting the chi-squared goodness-of-fit, if the p-value is higher than the significance level of .05, the null hypothesis is accepted, and if lower it is rejected. The extent of the source (literature for RQ1, other vocabularies for RQ2, and users' natural language for RQ3) matches the

target control terms is calculated by dividing the match by the total observed for each term, as well as the aggregate for all terms at the end of each RQ section. The aggregate alignment testing for RQ showing for each type of warrant is their alignment between the source warrant dataset and the target control terms based on the expected match distribution from the Pre-step assessment, as well as showing if the expected distribution holds across the Pre-step assessment and RQ1,2, and 4 (yellow highlight in Table 3.4). The null hypothesis of this study is assessed in RQ3 (green highlight in Table 8).

Table 8: Sample assessment of terms to determine RQ1-4 findings.

Control Terms	Pre-step Match %	RQ1 Literary Warrant	Match %	RQ2 Scientific Warrant	Match %	RQ3 Ho User Warrant	Match %	RQ4 Ho Fuzzy Search	Match %
Term 1	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%
...	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%
Term 12	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%
Aggregate	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%	Accept/ Reject	%

The acceptance of the hypothesis will indicate the control terms match the type of warrant for each RQ assessment, culminating in the main hypothesis testing for determining if users' natural language matches engineering-domain controlled terms, and to what extent.

3.6 Dataset creation method

This study will heavily use controlled vocabularies as a dataset. While there are many vocabularies used in the engineering space, there are a few that are more commonly used to tag content than most. All but one (INSPEC) are open vocabularies created by government entities which also contributes to their popularity, as well as increasing the potential impact of this research because of the amount of content tagged with these subjects. Because of studies such as Hjorland (2013) and Szostak (2008), this study will mainly focus on open controlled vocabularies that contain engineering study and design methodological terms. INSPEC is the only exception to this and is included because it is so highly used by the engineering community and because it covers the database search within libraries from sources such as Elsevier, EBSCO, ProQuest, and Ex Libris. The three main engineering specific vocabularies are INSPEC, created by the Institution of Electrical Engineers (IET), and the main vocabulary assigned to the world's largest collection of engineering research materials Engineering Village; NASA Thesaurus, created by NASA for tagging the NASA Technical Reports Server (NTRS), used by all aerospace engineers to some extent; and the Transportation Research Thesaurus (TRT), created by the National Academies of Sciences, Engineering, and Medicine which is tagged to the top government Transportation Research Institute Database (TRID). The fourth vocabulary that is not only focused on engineering is the Library of Congress Subject Headings (LCSH), which is used to tag most engineering library repositories in English-speaking countries. The user terms gathered from RQ3 will also be treated as a source vocabulary during the mapping process.

Two others that were considered were the Association for Computing Machinery (ACM) and the Medical Subject Headings (MeSH) vocabularies because of their focus not only on engineering but also on computing methods for research. ACM was not used because it was not

available to search as a vocabulary, rather it is only available as a filter from the ACM database search which made vocabulary assessment too cumbersome. MeSH is also a strong candidate because it has an entire branch of the vocabulary dedicated to methods however the vocabulary is not typically used on engineering content so it was excluded. With MeSH at least, it is mapped to many LCSH terminology via linked data so there is potential to traverse the mappings documented in this study to find the MeSH equivalent if that is of interest to readers. Each vocabulary contains hundreds or thousands of terms so it is unreasonable to assess each one, however, vocabularies are collections of terms organized as hierarchies of broader and narrower terms so the individual branches for machine learning in each vocabulary can be assessed without loss of context. From these vocabularies, a set of terms was selected for assessment in this study based on one of the fastest-growing areas of methodology in engineering, machine learning.

To determine candidate sub-disciplines in engineering for targeting branches in vocabularies, an emphasis on usefulness in information retrieval, common methods taught in engineering methodology and analytics courses, as well as popularity of search were considered. First, methods were gathered from available syllabi online for “Introduction to Engineering Analysis” courses, and University of Pittsburgh Engineering faculty were asked what methods would be most beneficial for those learning or teaching in engineering. Three main types were identified in this way: Machine learning (specifically linear algebra and other statistical methods); Prototyping (both digitally as in digital twins or physical prototyping like rapid prototyping); and Experimental observations (such as those in ethnography and user studies). From these, a Google Trends search was conducted to see how popular each was in a general Google search. As seen in Figure 10 (Retrieved by Researcher on July 6, 2021), Machine learning was the most commonly searched sub-discipline.

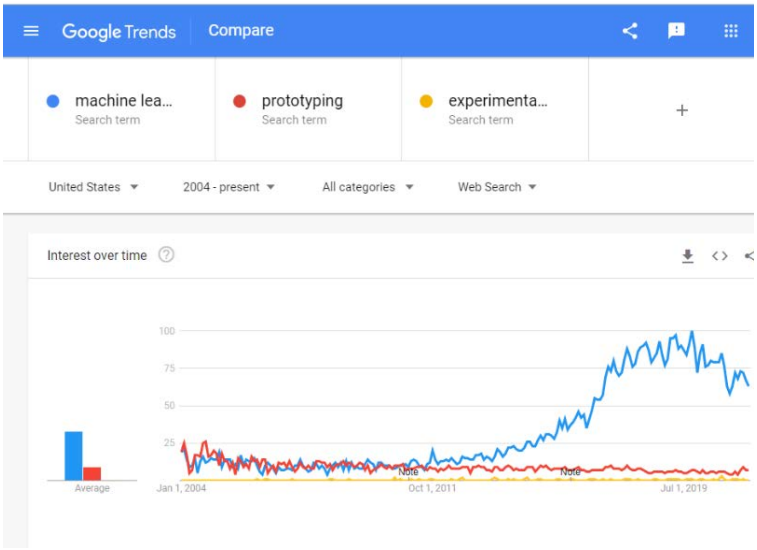


Figure 10: Google Trends volume of search on methodology types.

Due to the speed of machine learning terminology changes, it is a logical place to start investigations into natural language and controlled vocabulary alignment and has high potential to help researchers struggling to search with terminology that is rapidly changing because of its popularity in search.

From a natural language, perspective terminology is constantly in flux, not only for technology that rapidly changes such as machine learning. Because of this, it is expected that this study, which will only use a subset of terms from the machine learning branches of the selected vocabularies, will not determine if the user's natural language will match for the entirety of all controlled subjects -this study is seeking to understand if the null hypothesis is accepted for this selected subset, and a finding of this work may be that this type of assessment can be reproduced with other controlled vocabularies. It is expected that if this study method should prove effective for determining users natural language alignment with controlled terminology, and the resulting natural language supplementary alternate terms, that the assessment here will be conducted on additional sets of terms and regularly to continue to realign with users natural language. The

number of terms to assess should not total more than 60 based on Baxter, Courage, and Cain (2015), who suggest using a maximum of 90 terms, although there have been other studies that used more than 500 successfully. Blanchard and Banerji (2016) reviewed multiple studies which used card sorting methods and found using more than 60 terms had negative effects on the studies. In all of these cases, the researcher or a librarian/indexer conducted the mappings, not an end-user who will not be as familiar, or willing to give up the amount of time, to assess that many terms. In this study, where each term will have at least four mappings, assessment for even 60 terms is too time-consuming from a mapping perspective, and even from a user survey perspective, the average prompt is estimated to take at least 3-5 minutes to complete based on trial survey responses, indicating an iterative approach is needed for a full vocabulary assessment, which is not the scope of this study. Therefore, the sample of terms assessed in this study will be no more than 12 control terms (48 total mappings for each warrant assessed) in order to keep the survey at least under 60 minutes to complete.

In addition to controlled vocabulary and term selection, a baseline vocabulary and a literary dataset focused on machine learning methods terminology (for RQ1) are also needed. The baseline vocabulary will help determine which out of the four selected vocabularies will serve as the vocabulary the other three will be mapped to as an outcome of the findings from this study (in line with ISO 25964 suggested mapping technique). The Society of Automotive Engineers (SAE), a high-level categorical vocabulary rich in engineering category data available online, was selected. This approach was selected, instead of starting with an existing machine learning methodology vocabulary, because 1.) a dedicated machine learning methods vocabulary does not currently exist as a controlled vocabulary, and in fact may be a resulting artifact of this study, and 2.) a target vocabulary is needed to house the mappings from other controlled vocabularies, as well as the

users' natural language equivalents gathered in the later part of this study. The second aspect of mapping SAE as a baseline is to determine the expected observations for the chi-squared goodness-of-fit test. Using the method described in this study, the distribution between match and no match will help determine what the expected distribution would be for the chi-square in RQ1-3, with RQ1-2 being used to verify the distribution holds for all three warrant types.

3.7 Card sorting method

The most widely used method for assessing user terminology preferences and satisfaction is card sorting (Baxter, Courage, & Cain, 2015; Blanchard & Banerji, 2016). For instance, Khoo and Hall (2013) employ user card sorting to tag search/browse themes from the Internet Public Library to identify user mental models for search and to assess if users prefer natural language search in digital libraries (they found that users did prefer natural language search in libraries). Another example is Kupersmith (2012), who includes card sorting and category membership tests as test methods for repository term satisfaction. The two most applicable works to this study are van Damme et al. (2007) and Sharif (2009), which explore how to map folksonomic tags to controlled vocabularies by way of an ontological framework. Both use open sorting exercises to assess users' preferences in natural and controlled vocabulary selection in knowledge organization systems. This study will borrow from their research designs, as well as Olson and Wolfram (2007), who use the chi-squared goodness-of-fit to assess the likelihood of matching alignment between indexer and the other studies already noted in the chi-squared goodness-of-fit assessment section noted above.

Both closed and open sorting have one main issue in regards to this study, and that neither allows users to describe what words they would use to search for content. One very popular card sorting tool is from OptimalSort (Figure 11 and 12 below), which is where the study design was first tested. In both open and closed card sorts (there is also hybrid but that is just a combination of the two) for this study, the cards would be controlled vocabulary terms pre-populated for the user to cluster and the categories would be the labels for the controlled vocabulary "synonym" clusters. Closed card sorts would not be suitable for this study because they are totally pre-populated and do not allow the user to input their own natural language.

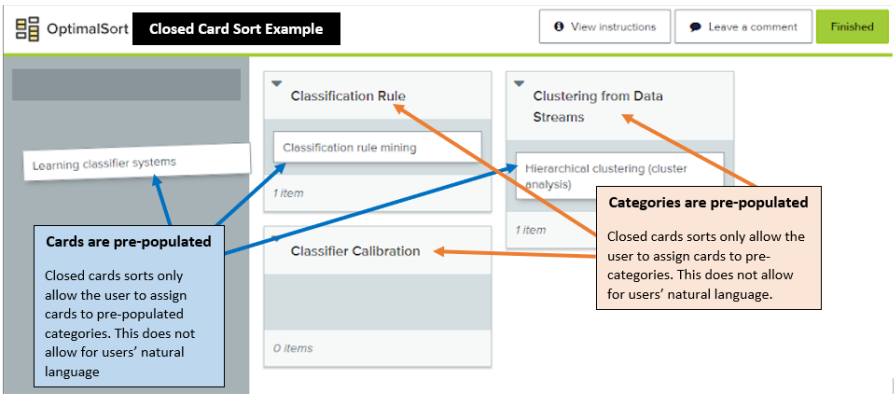


Figure 11: Closed Card Sorting. Reseacher created.

In open card sorts, users organize the controlled terms into clusters of similar terms and then label the clusters. This only tests how the user would organize the controlled terms and how they would classify how those terms are related, i.e. how they cluster. This does not test what words the user would use to search for content on specific topics. Both types of card sorts also do not allow for individual search prompts, instead of focusing on testing how users would interact

with a vocabulary in a browse pattern which does not go to the granularity needed for testing users' terminology for specific search topics.

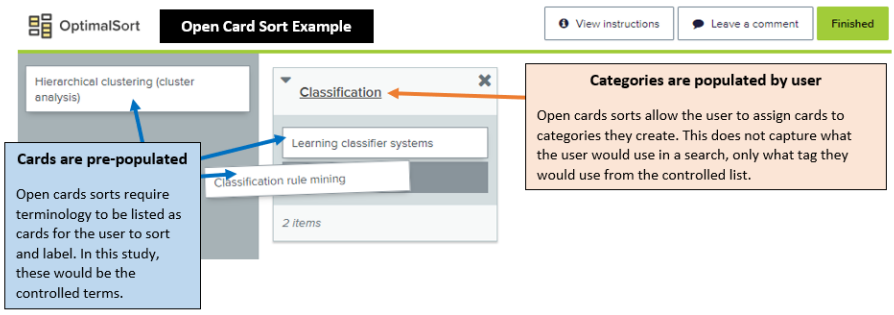
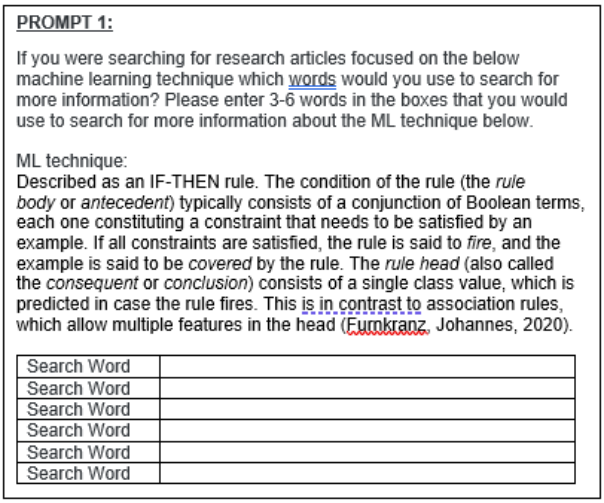


Figure 12: Open Card Sorting. Researcher created image

If traditional card sorts could be updated to have empty cards for users to input their natural language entries, the open card sort categories could be pre-populated with the search prompts to facilitate gathering users' natural language for specific topical content. Because this was not possible in popular card sorting tools, the researcher created this new card sort method in Qualtrics.

For simplicity of documenting the method throughout this study, the researcher has labeled this updated card sorting method as a bi-directional card sort due to its facilitation of testing how users' natural language (user warrant) match both controlled vocabulary (scientific warrant) and expert terminology from literature (literary warrant). Updating the open card sort method as described above, in Qualtrics the search prompt was used as the category and open cards were given to the user to note what words they would use to search for content similar to the prompt. Open card sorts assess how users interpret the relationship between cards, i.e. the label they assign the cluster, which is used to understand what label a group of synonyms can have or a broader/narrower organization in the vocabulary hierarchy. In the bi-directional card sort, the words a user would use to search for a topical prompt are gathered. As can be seen in Figure 13.



PROMPT 1:

If you were searching for research articles focused on the below machine learning technique which words would you use to search for more information? Please enter 3-6 words in the boxes that you would use to search for more information about the ML technique below.

ML technique:
Described as an IF-THEN rule. The condition of the rule (the *rule body* or *antecedent*) typically consists of a conjunction of Boolean terms, each one constituting a constraint that needs to be satisfied by an example. If all constraints are satisfied, the rule is said to *fire*, and the example is said to be *covered* by the rule. The *rule head* (also called the *consequent* or *conclusion*) consists of a single class value, which is predicted in case the rule fires. This is in contrast to association rules, which allow multiple features in the head (Furnkranz, Johannes, 2020).

Search Word	
Search Word	
Search Word	
Search Word	
Search Word	
Search Word	

Figure 13: Example Prompt 1 Classification Rule. Researcher created.

Assessment of the users' entries is focused on comparing what the user entered in the card sort to what the literature term is (literary warrant) and what the controlled term is (scientific warrant) for each prompt, the sum of which will be calculated using a chi-squared goodness-of-fit test to determine to what extent the users' natural language matches the controlled terms. If the users' and controlled terminology match to the level determined by mapping exercises in RQ1-2, then the null hypothesis is accepted. If they do not match as expected, the higher percentage between match/no match can be assumed to be the alternate hypothesis that the user's natural language matches to a much lower percentage than expected, and discuss what this might mean for user search. In all other aspects, from the sample size, ethnographic methods, and distribution, this study will use the same as a traditional open card sort.

3.8 Position of the researcher and ethical considerations

The position of the researcher is the facilitator of the card sorting exercises and as a taxonomist for mapping terms. There is little bias involved while facilitating the card sorting exercises because it is largely dependent on the users' own selections. However, creating the card sorting tasks and mapping terms is based on the researchers' experience with engineering terminologies and therefore may pose a risk of bias. That said, Rose (1985) has stated that "there is no neutrality. There is only greater or lesser awareness of one's biases (p. 77)" which indicates that complete neutrality is near impossible to achieve. To minimize the risk of bias, the researcher will use strict string matching as much as possible, and when a decision needs to be made during matching, the decision is documented for each decision and term this occurs in. This helps minimize bias as well as helping to make this research reproducible by other researchers.

For ethical considerations, ensuring the ethical treatment of participants and bearing in mind the effect of research findings on the body of knowledge if it were skewed, methods and survey of this study have been reviewed by the University of Pittsburgh Internal Review Board (IRB) and all statistical analysis has been reviewed by the University of Pittsburgh Statistics Consulting Center. No ethical or legal issues arose during the IRB application process or during the data gathering and survey portions of this study. Due to the emphasis on an individual's natural language, all references to individuals have been omitted to ensure complete data anonymity. Informed consent was obtained as part of the first portion of the survey to participants. The consent stated the contact information of the researcher and information regarding the nature and use of the research, the privacy statement and measures of the study, and where interested participants can access the information gathered. Participants were eligible for the opportunity to receive a $50

gift card if interested, and the emails of those interested were gathered in a separate survey that could not be connected to their survey responses.

The next chapter will focus on how the study used these methods to gather the dataset; test for a baseline for matching expectation; and finally how for each Step from RQ1-3, how the Step was conducted, the findings for each, implications and suggested updates derived from the findings, and how each finding builds to help prove out whether users' natural language matches controlled vocabulary terms or not.

4.0 Data Collection and Analysis

The following section outlines how the dataset was created, how data was collected, and lastly, data analysis and initial findings are presented. RQ1 will be presented first, showing how controlled vocabularies have been mapped and compared to Encyclopedia entries, followed by RQ2 where controlled vocabulary terms are matched to one another, and finally a term-by-term analysis comparing Control terms to users' natural language. Initial findings based on this step-by-step analysis are then presented for further discussion.

4.1 Cleaning and preparing the dataset

The results presented here are based on systematic, manual data analysis of the controlled vocabulary headings, as well as categorization of survey results, determining which category the responses fall into via Category Theory. Research aims were addressed by 1.) mapping controlled terminology together to determine if they match and to what extent, as well as 2.) identifying patterns in the users' natural language data and comparing them to the controlled vocabulary terms.

Many branches of the vocabularies were considered for this research, however, the theme of "machine learning" was the sub-discipline for vocabulary term selection due to its prevalence in general search, all sub-disciplines of engineering, its tendency to have many shorthand or natural language variations on its terminology, and its focus on methodology and testing measures within research. Based on this criterion, it is likely that the findings of this study will prove fruitful and be of great benefit to the engineering research community due to the exponential growth of

machine learning methods in this discipline. Due to the size of an entire vocabulary, this research will focus on a sample of terminology only to keep the study manageable. However, the methods and assessment for this study have been structured so that any information professional working with controlled vocabularies and search can replicate the process with different disciplines, terminology, and populations, as well as repeating this process as part of subject heading schedules, data governance, or vocabulary health assessments.

4.1.1 Establishing a mapping technique and target vocabulary

The first step is to select a target vocabulary to map the controlled vocabulary as well as user natural language to. Using the same method Binding and Tudhope (2015) and Faith, Tseytlin, and Bekhuis (2014) used as part of the Evidence in Documents, Discovery, and Analytics (EDDA) project, the researcher would identify a target vocabulary or hub, to map source vocabularies to, similar to semantic or graph-like mapping structures of target sameAs source mappings. A hub mapping structure will be used because more than 3 vocabularies will be mapped in this research– the users' natural language is being considered a vocabulary or data source in mapping, and as suggested by Binding and Tudhope (2015), a hub mapping is more effective than a crosswalk, which would require much more effort (12 mappings per subject rather than 8 using a hub structure) with little to no additional benefits, as can be seen in Figure 14 (Recreated with permission from Springer Nature and Copyright Clearance Center. International Journal on Digital Libraries. Improving interoperability using vocabulary linked data. Ceri Binding et al. (2015).

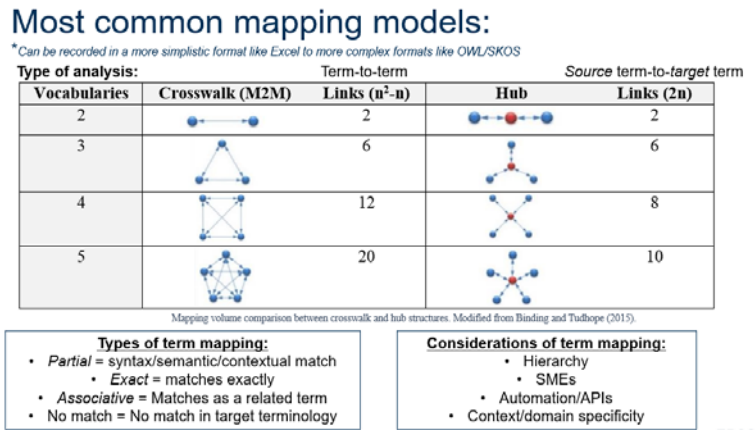


Figure 14: Comparison between mapping models. Recreated with permission.

The EDDA vocabulary is one of the few vocabularies dedicated to methodologies used in research studies so it was a prime target to use in this research as a target vocabulary. It is comprised of testing, design, and method terminology mapped from Journal of the Medical Library Association (JMLA), Medical Subject Headings (MeSH), National Cancer Institute (NCI) Thesaurus (NCI), Elsievers' Excerpta Medica dataBASE (Embase) vocabulary called Emtree, the Health Technology Assessment (HTA) Database Canadian Repository [international repository for health technology assessment], and Robert Sandieson's synonym ring for research synthesis. Mostly composed of prefTerm labels, the terminology was mapped to their synonymic representations in each vocabulary, as well as definitions, if they exist, to help with disambiguation. However, a review of the EDDA vocabulary (see Figure 15) by the researcher found that topical coverage did not cover machine learning methods to a great degree, instead of focusing on the types of studies used for a general research article. A different target vocabulary would need to be identified.

For source vocabularies, the most commonly used open-source vocabularies for engineering content were selected, which are LCSH, Transportation Research Board Thesaurus

(TRT), and the NASA Thesaurus as well as the top licensed vocabulary from Engineering Village called INSPEC -all of which have a user interface with a search bar to assist in the string searches for this study.

To close out the mapping technique, it was decided that the EDDA methodology of recording source mappings as annotations to the target vocabulary would be used to record these mappings. The target would serve as classes in a Web Ontology Language (OWL) file, similar to the OWL file for EDDA on Bioportal, and the source mappings would be recorded as altLabels. The OWL file would be edited and modeled using Protege, a free open-source ontology editor, and a knowledge management system created and maintained by Stanford University, and which can output RDF/OWL or CVS formats.

Having already established EDDA would not serve as a target to map to, each source vocabulary was assessed as a potential target vocabulary due to their prevalent use in the engineering library and indexing space. These are not as easily reviewed for engineering terminology coverage as EDDA, being orders of magnitude larger than EDDA. So to test if one of the source vocabularies could serve as the target, the researcher started with a high-level categorical vocabulary rich in engineering, that is the Society of Automotive Engineers (SAE) category data available on their website, to assess baseline coverage of engineering terminology. At this stage, the control dataset has not yet been selected. The SAE vocabulary is being used to assess the robustness of general engineering terminology first before the machine learning terminology is selected. This approach was selected, instead of starting with an existing machine learning methodology vocabulary, because 1.) a dedicated machine learning methods vocabulary does not currently exist as a controlled vocabulary, and in fact may be a resulting artifact of this study, and 2.) a target vocabulary is needed to house the mappings from other controlled

vocabularies, as well as the users' natural language equivalents gathered in the later part of this study. The second aspect of mapping SAE as a baseline is to determine the expected observations for the chi-squared goodness-of-fit test. Using the method described in this study, the distribution between match and no match will help determine what the expected distribution would be for the chi-square in RQ1-3, with RQ1-2 being used to verify the distribution holds for all three warrant types.

The SAE vocabulary was downloaded and the methodology terminology was selected for analysis and added to an excel sheet for further assessment. Out of 730 SAE subjects, only 127 were deemed testing or design methodology in nature. The researcher then searched for these strings, or their synonyms, in the source vocabularies and noted where a partial match, exact match, or no match was found. These were determined by using the definitions for each category documented in ISO 25964 Part 1, where no match is nothing in the target string matches the source, partial is at least part of the target string is found in the source vocabulary, and exact match is the exact target sting (stemming is considered a match) is found in the source vocabulary.

Table 9: Examples of Match Categories for Mapping. Researcher created.

Match Category	Target Vocabulary Term	Source Vocabulary Term
Partial Match (PM)	Magnetic Levitation Train	Maglev Train
Exact Match (EM)	Magnetic Levitation Train	Magnetic Levitation Train
No Match (NM)	Magnetic Levitation Train	High Speed Rail Train

These matching categories, seen in Table 9, also will be used in the mapping phase of this research once the sample machine learning methodology vocabulary is selected.

One thing to note: when mapping existing controlled vocabulary, there is a degree of warrant where the person mapping ideally has deep knowledge on the subjects, vocabularies, and content those vocabularies are used on while they are mapping so that they can assess matches even if they are not string matches (Hjorland (2008), National Information Standards Organization (ISKO), (2010). This is a skill that is often associated with literary warrant when indexing, and which is also used when cataloging and creating taxonomies. Part of literary warrant is scientific warrant, the consulting of the consensus of authoritative works such as controlled vocabularies (Bliss, 1929). This takes advantage of contextual metadata that may be associated with vocabulary terms to make a mapping determination, such as broader/narrower relations, definitions, synonyms, and scope notes. The researcher will use this warrant during the target vocabulary identification process, as well as the vocabulary and user natural language mapping process later in this study. The researcher will make sure to note when warrant is used to make a decision, how that decision was made, and the outcome in the discussion below. The results of this mapping are as follows in Figure 17:

LCSH Terms	*n*	*%*
Match (M)	44	35%
Partial Match (PM)	40	31%
No Match (NM)	43	34%
M/PM	84	66%
Total	127	100%

TRBT Terms	*n*	*%*
Match (M)	45	35%
Partial Match (PM)	29	23%
No Match (NM)	53	42%
M/PM	74	58%
Total	127	100%
Overall	127	100%

NASA Terms	*n*	*%*
Match (M)	50	39%
Partial Match (PM)	21	17%
No Match (NM)	106	83%
M/PM	71	56%
Total	127	100%

INSPEC Terms	*n*	*%*
Match (M)	48	38%
Partial Match (PM)	46	36%
No Match (NM)	33	26%
M/PM	94	74%
Total	127	100%

Figure 15: All vocabualry matching totals. Researcher created.

NASA had the highest number of exact matches (50), followed by INSPEC (48), TRT (45), and then LCSH (44), but taking into account partial matches the highest is INSPEC (94), LCSH, (84), TRT (74), and NASA (71).

When assessing a target vocabulary candidate, vocabularies with more opportunity for a match, seen by the total partial and exact match criteria, are ideal to use as the target -i.e. the vocabulary to map all other vocabularies to, because a vocabulary with a high match rate has the highest probability of mapping terminology across vocabularies. For this reason, INSPEC will be used as the target vocabulary (i.e. INSPEC terms will be used as the preferred label for the mappings) to record the outcomes from this study. INSPEC does not always have a match though, so in the few cases where INSPEC does not have a match, the ML Encyclopedia or Class will be used as the preferred label. The findings of this study will be recorded in Protege, and the specific mapping categories will be documented in Excel for use in the chi-square assessment. An example can be seen in Figure 18:

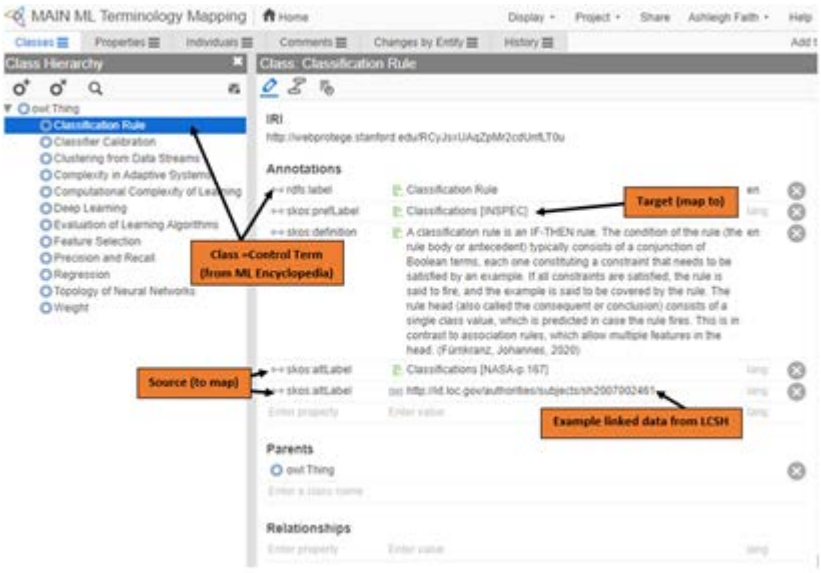


Figure 16: Protege codification of terminology mapping. Researcher created.

Interestingly, NASA is the top contributor of exact match terminology but has very few partial matches. This may be due to the SAE vocabulary having more aerospace terminology than the more general engineering vocabulary. Also of note, the LCSH has a surprisingly low coverage of engineering terminology. This mapping exercise was very enlightening about the perceived versus actual coverage of engineering vocabulary in the most used engineering vocabularies and could be continued in later research outside of this study.

Now having a baseline mapping, the distribution from SAE to the three selected vocabularies is as follows (Table 10):

Table 10: Mapping for Baseline. Researcher created.

	LCSH	TRT	NASA	INSPEC	TOTAL	%
NM	43	53	106	33	235	42%
M/PM	84	74	71	94	323	58%
TOTAL	127	127	177	127	558	

The average distribution between match and no match is 58% to 42% across all vocabularies. Because the SAE vocabulary is general, this distribution will be compared to the distributions in RQ1 and RQ2 assessments to determine if this finding holds across more specific terminology in the other warrant types. The final expected distribution will be calculated based on the SAE baseline, RQ1, and RQ2 distribution findings. The final expected distribution will be used to assess if the match rate between users' natural language and controlled vocabularies is as expected, proving user warrant is used and therefore the vocabularies are user-centered. The controlled vocabularies will not be user-centered if the resulting match rate for RQ3 does not meet the expected distribution.

4.1.2 Preparing Control Dataset

This aspect is focused on determining which machine learning methods terminology will be used as the control dataset throughout this study. The control dataset will be used as the terminology to compare both controlled vocabulary and users' natural language during the mapping process, as well as the source for the bi-directional card sort prompts. Most controlled vocabularies have hundreds if not hundreds of thousands of terms or headings to choose from so a sample of 10-15 machine learning terms will be used in this study, determined based on how many prompts an end-user can review in under 1 hour. This is only a small subset of any vocabulary. With this in mind, this study is structured in a way that additional portions of a vocabulary can be assessed iteratively with the methods laid out herein.

The control dataset needs to have definitions or descriptions of the method to create the bi-directional card sort prompts. Unfortunately, reliable definitions in controlled vocabularies are lacking, with many vocabularies having scope notes instead of definitions as in the case of the TRT, or not having any scope note/definition at all, as in INSPEC. The difference between a scope note and a definition is a scope note is used to direct librarians and subject indexers on how to use the heading to tag information, whereas a definition is how the subject is defined. Ideally the target, or even source, vocabularies would have definitions to be used in the survey to gather users' natural language preferences. Due to this lack of descriptive text in the source vocabularies, this study will instead use the Encyclopedia of Machine Learning and Data Science (Phung, Webb, and Sammut, 2020) because it is one of the most recent authoritative encyclopedias on machine learning methodology that also has brief descriptions of the methodology for use as prompts in the survey for this study. The title of this encyclopedia will be shortened within this research to Encyclopedia for ease. Please note, the online version of this Encyclopedia is a "Living Edition" with live updates

to the content so for the sake of reproducibility, this resource was accessed between May and August of 2020.

The Encyclopedia has over 200 methods so the researcher needed to determine which methods had the most impact on machine learning research. These 200 methods will be considered candidates until the control dataset is established. Impact on research was determined by searching for the topic in the most used engineering research database called Engineering Village. Using the University of Pittsburgh subscription to this database, the topics from the Encyclopedia were entered into the Engineering Village keyword search. Engineering Village allows for the bibliometrics of the search results to be downloaded and analyzed and this feature was used to gather the amount of articles within the database that mentioned the subject from the Encyclopedia within the last five years. Five years was used as the threshold because a longer time period was deemed too long. After all, machine learning techniques and terminology would be too different for the technology available to researchers, and a two-year threshold, also considered, did not have a drastic difference in the volume of records. Some of the Encyclopedia entries had qualifiers in addition to the subject, in which case these were not entered into the query since they were superfluous. This did not change the context of the method for most of the candidate control dataset entries, however, the entry *Machine Learning and Game Playing* was too broad to provide meaningful results as a complex query, and gave too many false results when shortened so it was excluded from the final control dataset.

Example query on Engineering Village where only the highlighted subject was changed for each subject search:

((((abduction) WN ALL)) AND (({ja} WN DT) AND ((2020 OR 2019 OR 2018 OR 2017 OR 2016) WN YR)))

Understanding that not all research is published traditionally, a search with the same parameters was conducted on Google Scholar to offset any limitations introduced from using only one source. Only three additional search parameters were introduced in the Google Scholar query, that being “in machine learning,” due to the breadth of scope of that “database,” compared to Engineering Village which is a dedicated engineering research database, as well as excluding citations and patents, to reduce noise and redundancy of a broad Google index. These research volume counts are not to indicate one method is more important than the other, or even to indicate an impact factor of some sort. These counts are merely to indicate how often these methods show up in various academic searches, showing their potential to be included in this study.

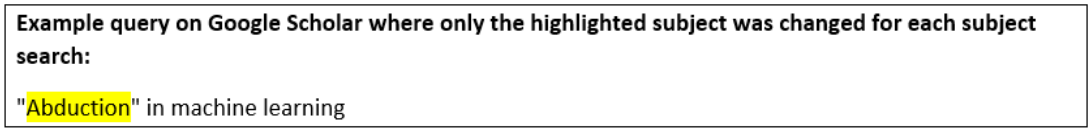


Figure 17: Example Search in Google Scholar.

Based on this analysis, and barring any special search engine optimization from these sources, the methods that have the highest potential in the machine learning domain are as follows, and will be considered the control dataset used for the duration of this study used to analyze and map controlled vocabulary and users’ natural language to, and will be called the “Control” or “Control dataset.”

Table 11: Control dataset coverage across databases. Researcher created.

ML Encyclopedia Chapter	Search Topic	Volume of Articles in EV	Volume of Articles in Google Scholar	Total Potential
~~Machine Learning and Game Playing~~	~~Machine Learning~~	~~43,088~~	~~704,000~~	~~747,088~~
Classification Rule	Classification Rule	68,463	612,000	680,463
Weight	Weight	138,376	388,000	526,376
Topology of a Neural Network	Neural Network	65,877	232,000	297,877
Regression	Regression	59,944	229,000	288,944
Clustering from Data Streams	Clustering	87,651	150,000	237,651
Deep Learning	Deep Learning	25,075	146,000	171,075
Precision and Recall	Precision and Recall	125,088	140,000	265,088
Feature Selection	Feature Selection	70,000	96,000	166,000
Classifier Calibration	Classifier	57,512	123,000	180,512
Evaluation of Learning Algorithm	Learning Algorithm	96,000	102,000	198,000
Computational Complexity of Learning	Computational Complexity	68,000	87,000	155,000
Complexity in Adaptive Systems	Adaptive Systems	74,000	88,000	162,000

Once the sample was selected, the encyclopedia chapters related to that method were downloaded and the definitions, identified as the first paragraph for the method, were also extracted for use later in the survey setup.

Mapping controlled vocabulary to create a baseline

Each method from Control was added to a Web Protege project as a class. The label was noted as rdf:label and the definition were noted as skos:definition, following the W3C standards for Resource Description Framework (RDF) and Simple Knowledge Organization System (SKOS). The researcher then used the same mapping method described above in the *Establishing a mapping technique* section, first mapping the target vocabulary (INSPEC) as the prefLabel, and the source vocabularies (TRT, NASA, and LCSH) that have a match or partial match are also mapped in via AltLabel with match category, source vocabulary, and source vocabulary ID, if it exists, is listed in brackets as can be seen in Figure 20.

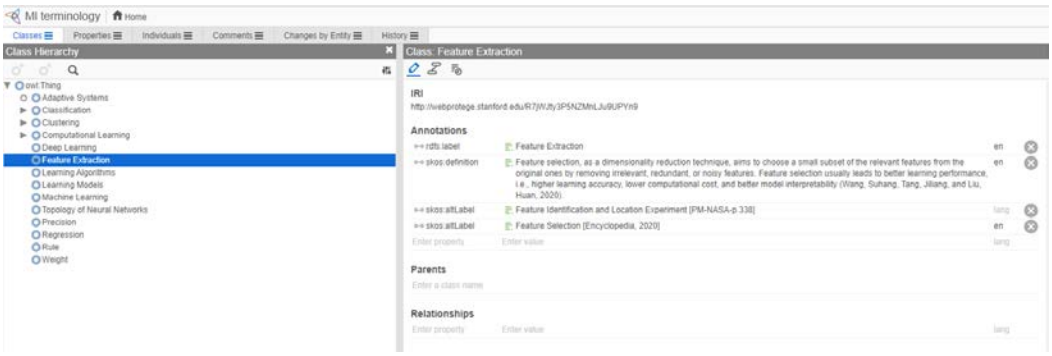


Figure 18: Codification in Protege. Researcher created.

If this vocabulary mapping were to be used as a W3C linked data vocabulary, the linked data URIs would be used instead of the vocabulary ID but that is not the purpose of this study. This process was repeated for all entries from the Control dataset.

During the mapping process, three entries could not be mapped like the others from LCSH, *Weight*, *Deep Learning*, *Neural Network*. *Weight* was too broad and generic on its own, and the LCSH search engine does not allow for fuzzy search so even adding machine learning as context would not give any results, therefore this was categorized as a NM. In the example above, *Deep Learning* was not found in LCSH but the Encyclopedia prompt mentions this in conjunction with neural networks so that was searched instead and deemed a partial match, however later when the researcher searched for the Encyclopedia method *Topology of Neural Networks* in LCSH, the same PM from *Deep Learning* was found. The researcher assessed definitions from both and determined that *Deep Learning* on its own was more akin to Neural Networks in Computer Science, whereas *Topology of Neural Networks*, was more about the topology method and neural networks were just the use case for that method -thus deeming *Neural Networks (Computer Science)* as a PM for *Deep Learning* and *Topology* as a PM for *Topology of Neural Networks*.

When mapping NASA, many of the machine learning methods were present but two potential matches, *Classification* and Classifier Calibration were similar enough in the NASA thesaurus that the NASA GS (notation in NASA thesaurus for a narrow term) and related terms were used by the researcher to determine the NASA term *Classifier* was not the same context as the Encyclopedia method and therefore resulted in a NM from NASA for *Classifier Calibration*. A similar dilemma occurred with the method *Classification Rule*, where in the NASA thesaurus two terms were similar enough that the NASA Scope Note was used to determine the NASA term *Classifications* was more a partial match to *Classification Rule* than the more generic *Classifying*, although the researcher can see how using either of these NASA terms would be valid.

When mapping INSPEC, the entry *Precision and Recall* had no exact or partial match, however, INSPEC allows for fuzzy search so the researcher was able to find terms that perhaps represent a use case for *Precision and Recall,* such as the INSPEC terms *Recommender Systems* and *Search Engines*, but for this study, those are represented as a Related term and not a match. For *Neural Networks*, INSPEC had many more specific terms for this which indicates this is a topic of research the vocabulary is often tagging and provides more access points and opportunities for researchers to find research on this topic. The method Clustering also had more specific types in INSPEC, as well as one broader which was Clustering Tools, but staying within the scope of the original Encyclopedic entries intended context, Pattern Clustering was identified as a Partial Match from INSPEC. Below (Table 12) is the result of the category mappings to Control from source vocabularies:

Table 12: Mappings Control terms and controlled vocabulary terms.

ML Encyclopedia Chapter		LCSH			TRT		NASA		INSPEC
	LCSH	Term	TRT	Date	Term	NASA	Term	INSPEC	Term
Classification Rule	PM	Classification rule mining	NM	2007	NA	PM	Classifications	PM	Classifications
Classifier Calibration	PM	Learning classifier systems	NM	2009	NA	NM	NA	NM	NA
Clustering from Data Streams	PM	Hierarchical clustering (Cluster analysis)	NM	2014	NA	PM	Cluster Analysis	PM	Pattern Clustering
Complexity in Adaptive Systems	PM	Adaptive computing systems	PM	2003	Adaptive Control	PM	Acive Control	PM	Adaptive Systems
Computational Complexity of Learning	PM	Computational complexity	NM	1997	NA	NM	NA	PM	Computational Complexity
Deep Learning	PM	Neural Networks (Computer Science)	NM	2016	NA	PM	Neural nets	PM	Neural nets
Evaluation of Learning Algorithm	NM	NA	NM		NA	PM	Backpropagation (artificial i	PM	Learning (artifical intelegenc
Feature Selection	NM	NA	NM		NA	PM	Feature Identification and Lo	PM	Feature Extraction
Precision and Recall	NM	NA	NM		NA	NM	NA	NM	NA
Regression	NM	Regression Analysis	PM		Regression Analysis	NM	Regression Analysis	NM	Regression Analysis
Topology of a Neural Network	PM	Topology	PM	1992	Topology	PM	Topology	PM	Topology
Weight	NM	NA	PM		Weighting	EM	Weight	NM	NA
									TOTALS

While mapping, the researcher also gathered any definitions that were present. NASA surprisingly had many definitions, which were added to the Protege project, but many also were very specific to the aerospace industry, such as *Active Control* (a PM to *Complexity in Adaptive Systems*), which NASA defines as "The automatic activation of various control surface functions in aircraft (NASA Thesaurus, 2020). These definitions were deemed too short and not containing enough context to be used for the survey prompts. The high volume of methods coverage in NASA also reflects how the aerospace engineering sub-domain has a high degree of machine learning terminology in research. This is not surprising because most vocabularies are created with a specific discipline, user base, or content corpus in mind, for the TRT it is mostly ground vehicles that may not need as much machine learning granularity as the NASA thesaurus, which focuses on aerospace vehicles that have more, while INSPEC is more of a general engineering vocabulary so more generalized methods were present in that vocabulary. LCSH is not focused on engineering alone, although it is used in many libraries to index engineering content, this does mean it is more generalized in its coverage of engineering topics. The mapping results are as follows, and have also been recorded in the Protege project.

4.2 RQ1: Control data matches to controlled vocabularies

This section will build on the baseline dataset of Encyclopedia terms (Control) by mapping controlled vocabulary terms (called Source) to their EM or PM equivalencies. There were very few exact matches to the Control dataset which may indicate the controlled vocabulary does not usually match the natural language of the experts in the field. This is not surprising because subject vocabulary terms are structured differently than what an encyclopedia would offer. Both serve as browse functionality to researchers, used as a means to increase the precision of their search, and many encyclopedia entries or chapters are tagged with subjects and indexed as separate from the main work. It begs the question, do the machine learning experts who wrote the Encyclopedia entries, and the controlled subjects match, and to what extent?

To test Research Question 1, the null hypothesis being expert vocabulary and controlled vocabularies do not differ significantly, largely due to vocabulary terms often being derived from expert reference materials such as encyclopedias so they likely are not significantly different. In this case, a chi-squared goodness-of-fit test is performed for each Control-to-Source vocabulary match set, as seen in Table 13.

Table 13: Control to Source match summary. Researcher created.

ML Encyclopedia Chapter	LCSH	LCSH Term	TRT	Date	TRT Term	NASA	NASA Term	INSPEC	INSPEC Term	Out of Existing Terms Across all Vocabularies PM1	PM2	EM	NM
Classification Rule	PM	Classification rule mining	NM	2007	NA	PM	Classifications	PM	Classifications	3	0	0	1
Classifier Calibration	PM	Learning classifier systems	NM	2009	NA	NM	NA	NM	NA	1	0	0	3
Clustering from Data Streams	PM	Hierarchical clustering (Cluster analysis)	NM	2014	NA	PM	Cluster Analysis	PM	Pattern Clustering	3	0	0	1
Complexity in Adaptive Systems	PM	Adaptive computing systems	PM	2003	Adaptive Control	PM	Acive Control	PM	Adaptive Systems	0	3	0	1
Computational Complexity of Learning	PM	Computational complexity	NM	1997	NA	NM	NA	PM	Computational Complexity	2	0	0	2
Deep Learning	PM	Neural Networks (Computer Science)	NM	2016	NA	PM	Neural nets	PM	Neural nets	0	0	0	4
Evaluation of Learning Algorithm	NM	NA	NM		NA	PM	Backpropagation (artificial i	PM	Learning (artifical intelegence	0	1	0	3
Feature Selection	NM	NA	NM		NA	PM	Feature Identification and Lo	PM	Feature Extraction	2	0	0	2
Precision and Recall	NM	NA	NM		NA	NM	NA	NM	NA	0	0	0	4
Regression	NM	Regression Analysis	PM		Regression Analysis	NM	Regression Analysis	NM	Regression Analysis	4	0	0	0
Topology of a Neural Network	PM	Topology	PM	1992	Topology	PM	Topology	PM	Topology	4	0	0	0
Weight	NM	NA	PM		Weighting	EM	Weight	NM	NA	1	0	1	2
									TOTALS	20	4	1	23

For instance, for controlled terms mapped to *Classification Rule (Row 3 in Figure 4.7)*, two of the vocabularies (NASA and LCSH) are only similar to the Control term *Classification rule* because of the first part of Control term *Classification*, which is a Partial Match in the first position (PM1 in blue). An example of a Partial Match second position (PM2 in orange) would be the *Adaptive* mappings from LCSH, TRT, and INSPEC where the first part of the terms are not the same as the Control term, but the latter portion of the term is -both mention *Adaptive*. A count for no match (NM in white) would count any vocabularies that did not contain the term at all. In this case, 5 terms across all source vocabularies (example *Neural Nets* from NASA and *INSPEC* and *Active Control* from NASA) could be mapped to Control because of additional metadata in the vocabulary to help with the mapping, but their labels were different though not to merit a match in this assessment between Control and Source vocabulary matching and therefore count as a NM her. A total for each set of mapped terms is calculated to assess for significance using a chi-square test.

For the fuzzy search pattern for how well the terminology from the literature (literary warrant) matches controlled vocabulary, PMs and EMs are counted as a match for a total of 25 matches compared to 23 No Matches, giving a probability of p=0.77 and a chi-square score of 0.08, signaling the null hypothesis for RQ1 is accepted for the fuzzy search pattern, meaning the expert and controlled vocabulary terms statistically matched with 52% matched, compared to the 50% evenly distributed expectation found through the engineering-domain controlled vocabulary general mapping assessment done between SAE, INSPEC, NASA, TRT, and LCSH (summarized in Figure 14).

Fuzzy Search Pattern Findings: Null Hypothesis Accepted $\chi 2$ (1) = 0.08, p= 0.77

Table 14: Fuzzy search Chi-Square Assessment for Literary Warrant.

Category	Observed N	Probability	Expected N	chi-squre	0.083333333
Match	25	0.5	24	df	1
No Match	23	0.5	24	p-value	0.772829993
TOTAL	48	1	48	result	Accept Ho

For the browse pattern for how well the terminology from the literature (literary warrant) matches controlled vocabulary, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 14 matches compared to 34 No Matches, giving a probability of p= < .001 and a chi-square score of 8, signaling the null hypothesis for RQ1 is rejected for the browse search pattern, meaning the expert and controlled vocabulary terms did not statistically match with 29% match, compared to the 50% match expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 44, p= < .001

Table 15: Browse pattern Chi-Square Assessment for Literary Warrant

Category	Observed N	Probability	Expected N	chi-squre	8.333333333
Match	14	0.5	24	df	1
No Match	34	0.5	24	p-value	0.003892417
TOTAL	48	1	48	result	Reject Ho

This indicates that using a controlled vocabulary to tag content in a browse pattern would not match the expert researcher's vocabulary on its own, and would benefit from a search engine for fuzzy matches and query expansion. The null hypothesis assertion that vocabulary terms are often derived from expert reference materials such as encyclopedias may be true logically, but statistically, the nature of publishing a new vocabulary with updated terms, or the time it takes to publish a new Encyclopedia, may slow down the process of updating terminology that changes quickly such as machine learning terminology leading to the null hypothesis being rejected for the browse pattern.

From the match to no match between literature terminology and controlled vocabularies, the average distribution is 52% to 48% across all vocabularies. Compared to the distributions in the baseline SAE mapping (58% to 42%), RQ1 findings therefore support the baseline finding that there is a general 50/50 distribution for terminology matches. These findings will be further compared to the findings from the RQ2 assessment to determine if this finding holds across the main types of warrant, literary and scientific, to use as the baseline distribution for user warrant, estimating to what extent users' natural language matches controlled vocabulary terms.

There is a risk the expert vocabulary used in this assessment might be too small; however, one can argue that there are so few reference books on modern Machine Learning Methodology that the terminology expressed within its chapters are approved and shared to some degree with their peers who selected the authors for each chapter, as well as the peer reviewers and book reviewers of the Encyclopedia. That said, this methodology should be continued with additional reference materials as a source, or expand to a bi-directional card sort for experts only. Both are out of the scope of this research; however, these are fruitful lines of inquiry for future research.

Now that the baseline mapping is complete, and the alignment between the Control and Source vocabulary terms are complete, the researcher then moved on to the assessment for Research Question 2: how often and to what extent do controlled vocabulary terms match each other?

4.3 RQ2: Controlled vocabulary matches (scientific warrant)

Using the baseline mapping, the researcher first assessed if the vocabularies had any significant difference between them and to what extent, without matching to the Control set

(Research Question 2). The null hypothesis for Research Question 2 is that controlled vocabularies will match more often than they do not (even distribution), largely based on assertions that controlled vocabularies have standards that guide how they are created as well as being highly controlled lists with indexing professionals researching and curating them regularly.

The researcher categorized each set of controlled vocabulary terms for each Control dataset and calculated the matches across each row to indicate how well vocabulary terms matched for the same Control term across source vocabularies.

Table 16: Control to Source Controlled Vocabulary Matches.

	LCSH		TRT		NASA		INSPEC	Out of Existing Terms Across all Vocabularies			
LCSH	Term	TRT	Term	NASA	Term	INSPEC	Term	PM1	PM2	EM	NM
PM	Classification rule mining	NM	NA	PM	Classifications	PM	Classifications	1	0	2	1
PM	Learning classifier systems	NM	NA	NM	NA	NM	NA	0	0	0	4
PM	Hierarchical clustering (Cluster analysis)	NM	NA	PM	Cluster Analysis	PM	Pattern Clustering	3	0	0	1
PM	Adaptive computing systems	PM	Adaptive Control	PM	Acive Control	PM	Adaptive Systems	3	1	0	0
PM	Computational complexity	NM	NA	NM	NA	PM	Computational Complexity	0	0	2	2
PM	Neural Networks (Computer Science)	NM	NA	PM	Neural nets	PM	Neural nets	1	0	2	1
NM	NA	NM	NA	PM	Backpropagation (artificial i	PM	Learning (artifical intelegenc	0	2	0	2
NM	NA	NM	NA	PM	Feature Identification and Lo	PM	Feature Extraction	2	0	0	2
NM	NA	NM	NA	NM	NA	NM	NA	0	0	0	4
NM	Regression Analysis	PM	Regression Analysis	NM	Regression Analysis	NM	Regression Analysis	0	0	4	0
PM	Topology	PM	Topology	PM	Topology	PM	Topology	0	0	4	0
NM	NA	PM	Weighting	EM	Weight	NM	NA	2	0	0	2
							TOTALS	12	3	14	19

For instance, for controlled terms mapped to *Classification Rule (Row 3 in Figure 4.8),* two of the vocabularies (NASA and INSPEC) have the same term *Classifications* (in green) which would be counted as Exact Matches (EM), *Classification rule mining* from LCSH is only similar to the NASA and INSPEC terms because of the first part of LCSH term *Classification,* which is a Partial Match in the first position (PM1 in blue) to the INSPEC and NASA matches. An example of a Partial Match second position (PM2 in orange) would be the *Feature Selection* mappings from NASA and INSPEC where the first part of the terms are not the same, but the latter portion of the term is -both mention *artificial intelligence*. A count for no match (NM in white) would count any vocabularies that did not contain the term at all. In this case, only 1 term from LCSH (*Learning classifier system*) could be mapped to Control, but no other vocabulary had a term to compare it

to. While this would count in the PMs for the Control dataset assessment, this assessment is only looking at the matches between vocabularies and therefore this is counted as a NM. A total for each set of mapped terms is calculated to assess for significance using a chi-square test.

For the fuzzy search pattern for controlled vocabularies matching one another, PMs and EMs are counted as a match for a total of 29 matches compared to 19 No Matches, giving a probability of p=0.14 and a chi-square score of 2, signaling the null hypothesis for RQ2 is accepted for the fuzzy search pattern, meaning the controlled vocabulary terms matched across other controlled vocabulary sources with 54% matched, compared to the 50% evenly distributed expectation. The RQ2 findings support the baseline and RQ1 findings that an even distribution between match and no match can be expected in assessment of RQ3.

Fuzzy Search Pattern Findings: Null Hypothesis Accepted $\chi 2$ (1) = 2, p= 0.14

Table 17: Fuzzy search Chi-Square Assessment for Scientific Warrant.

Category	Observed N	Probability	Expected N		chi-squre	2.083333333
Match	29	0.5	24		df	1
No Match	19	0.5	24		p-value	0.148914673
TOTAL	48	1	48		result	Accept Ho

For the browse pattern for controlled vocabularies matching one another, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 14 matches compared to 34 No Matches, giving a probability of p= < .001 and a chi-square score of 8, signaling the null hypothesis for RQ2 is rejected for the browse pattern, meaning the controlled vocabulary terms did not match across other controlled vocabulary sources with 27% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 8, p < .001

Table 18: Browse pattern Chi-Square Assessment for Scientific Warrant.

Category	Observed N	Probability	Expected N		chi-squre	8.333333333
Match	14	0.5	24		df	1
No Match	34	0.5	24		p-value	0.003892417
TOTAL	48	1	48		result	Reject Ho

Finding that the controlled vocabularies did not often agree with one another in the browse pattern is surprising. The assertion that controlled vocabularies have standards that guide how they are created as well as being highly controlled lists with indexing professionals researching and curating them regularly may hold for standardized terms within one vocabulary, but this finding seems to imply that vocabulary management does not stress interoperability with other vocabularies, even from the same discipline like NASA and TRT. This may be due to vocabularies, while from the same discipline, are still used on different content corpora and for different use cases and therefore adds a layer of uniqueness to all vocabularies that may explain why even if the vocabularies are expertly curated, from the same discipline, they are likely to have significant terminology differences, even for a term that is contextually the same.

This last is an important finding because the difference between browse and fuzzy search patterns has a significant impact on the discoverability of research. When a user is browsing a controlled vocabulary they can read and "translate" the subject headings into what they understand and thereby continue their search unhindered. For example, if searching a corpus of research using the NASA Thesaurus heading *Cluster analysis*, which can be used to retrieve content about clustering methods in machine learning but only if the content was tagged with that exact heading. This dichotomy between Boolean and Semantic search has been reviewed by many (a primary source used in this study was Azad and Deppak's 2019 review of query expansion methods and results), with the general finding that query expansion on sting matches increases the effectiveness of information retrieval "significantly for unstructured queries, while only marginal improvement

was observed for structured queries (p.5).” Depending on exact string matches alone increases the risk the researcher will miss important content on Clustering that was tagged with a different heading, like *Pattern clustering* in INSPEC, if fuzzy matches are not also queried by the search engine. This researcher experienced this issue in NASA and LCSH when mapping to Control because they did not have a fuzzy search functionality (NASA is a flat document, and the LCSH has a search engine but it does not expand a keyword query). The fuzzy search pattern, which is more common in aggregated search like Google Scholar or INSPEC, can expand the users’ query to subject equivalents like the NASA heading *Cluster analysis* and LCSH heading *Hierarchical clustering (Cluster analysis)* if the vocabulary mappings are used in the search index. This is especially useful for contextual, rather than string matches, such as is the case of *Deep Learning* (Control) mapping to *Neural Nets*, found in LCSH, NASA, and INSPEC. This point will also be raised in the Control and Source matching assessment, which is the next step in this study.

Before moving into the RQ3 assessment, a final measure for expected distribution for the remaining chi-square tests will be established from the previous match to no match findings. The summary is as follows (Table 19):

Table 19: Summary for Chi-Square Expected Distribution.

Summary of Distribution	Baseline	RQ1	RQ2	Avg.
NM	42%	48%	60%	50%
M	58%	52%	40%	50%

Comparing distributions in the baseline SAE mapping, RQ1, and RQ2, all fall within the expected 50/50 match distribution range, showing that matching across different warrants remains constant and can be used as the distribution for the chi-squared goodness-of-fit test for user warrant in RQ3.

4.4 Creating User Natural Language Dataset

With the controlled vocabularies mapped the researcher needed to gather users' natural language (called Natural Language or Entries). This was by far the most difficult task to complete, largely due to distribution methods falling through. At first, the researcher set out to use LinkedIn Groups, specifically those focused on engineering, but one year into research LinkedIn shut down their Groups feature. It has since returned, but the disruption in access to this feature blocked distribution. Next, the researcher reached out to the Special Librarians Association (SLA) Engineering Division, which is composed of information professionals working with the engineering discipline. Unfortunately, after the researcher reached an agreement of distribution with SLA, the organization went through a change in management, and the rules for distributing to members, as well as activity between members and non-members of the association, blocked the researcher from working with that group. After this, the researcher reached out to the University of Pittsburgh Engineering School to ask faculty if they would assist in distributing the

survey to their students. Outside of one faculty member, no other faculty responded. Working with University of Pittsburgh Engineering School Librarians, Carnegie Mellon Engineering Librarians, and Duquesne University Engineering Librarians, the researcher attempted to send out the survey through a network of engineering librarians, but many, while wanting to help, were limited in what they could send to their students in the form of a research study survey. Finally, thanks to University of Pittsburgh Engineering School Librarian Judy Brink, the researcher was put into contact with the American Society for Engineering Education (ASEE), who agreed to distribute the survey via their social media and membership accounts. With distribution finally secured, the research sent the survey out to ASEE for their distribution

The survey to gather users' Natural Language was constructed in the University of Pittsburgh Qualtrics tool. Each Control term was used to designate a survey block, or "card" similar to an open card sort. Each block contained the Control prompt, gathered from the Encyclopedia, with the main Control term removed at least from the first sentence of each prompt to remove unconscious influences; directions for the user; and 6 keyword boxes for users' Entries.

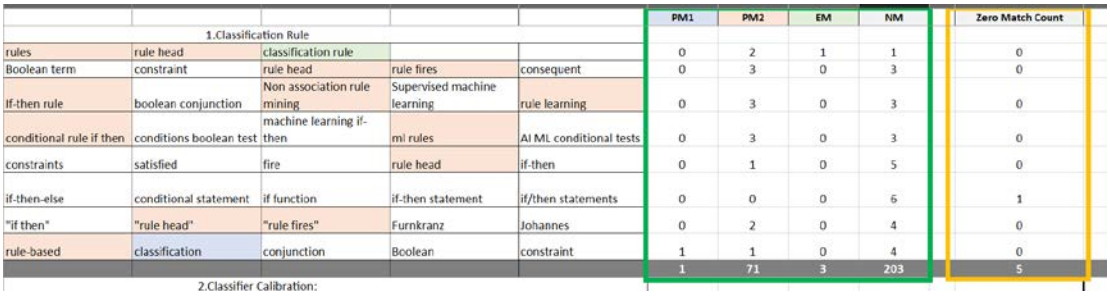
Prompts, how they are written, and what the author chose to emphasize, may affect what entries users select. However, users' are likely in real-world scenarios to have had some prompt, whether from a class assignment, something they heard on the news, or a new project at work, to start their search whether they are a new researcher to engineering or a seasoned professional in the field. Outside prompts to research, coupled with the variation of words found in the literature itself, mitigates the risk that the survey prompts should not overly influence the users' entries. In the discussion section, prompts may be referenced to explore why users' may have selected some terms, but this is no different than what the user would encounter outside of this study so it will not be a focus for the assessment of this study.

The structure of the survey followed a bi-directional card sort method where at least 6 empty “cards” are given to the participant, along with a prompt, to enter the Natural Language the user would use to search for research on the topic described in the prompt. At the end of the survey, each participant is given a random chance to be chosen to receive a $50 gift certificate as a token of appreciation for their participation in the study. This will not affect whether the participant is considered for the survey and a separate survey was used to gather the emails of those who participated, to ensure anonymity. After 1 month of the survey being distributed by ASEE, the responses were gathered.

The assessment for RQ3 started by downloading the survey results from Qualtrics as a CSV file. Next, the researcher made headers in row 1 for the Control label for each prompt. The researcher then eliminated responses that did not contain any data, most often when a user opened the survey but did not respond to any prompts. These are not relevant to this study due to the complete lack of any information in these responses. There were a lot of empty responses in the first few days of the survey being distributed. After consulting with the Qualtrics survey support team, it was found that participant responses were not cached or saved in any way unless the participant advanced to the next page of the survey. The original design of the survey had all prompts on one page, or survey block, which is why there were over 30 responses that the survey did not record. This is extremely unfortunate, but the survey was updated so that each prompt was on a separate page/block, and therefore even partial responses have been retained after the error was corrected. The original CSV download had result blocks organized horizontally.

To make the chi-square assessment more manageable, each response block was organized vertically, with the categories (PM1-Blue, PM2-orange, EM-green, and NM white) sums to the left of each block. Each Control term was calculated for a total count for each type for each

participant response by hand by the researcher. Each row represents a participant, and each cell (Entry) represents a term the participant would use to describe or search for the method represented by the survey prompt given. This study is concerned with how well a users' natural language term (as in one tag) matches the Control term (again one tag), which aligns with one-to-one mapping for controlled vocabularies assessment at the instance or node level. See Figure 4.9for a comparison between an entire users' response being calculated as Match (M) or No Match (NM) versus each entry from the user being counted as a M/NM for each prompt. Therefore, each Entry (up to 6) for each participant is assessed and calculated for matching, and the matches are used to assess if the users' term and the Control term have a significant probability of matching via chi-square assessment, where the yellow box represents total matches per response per participant vs the green box that represents total matches per Entry per participant as seen in Figure 22.



					PM1	PM2	EM	NM	Zero Match Count
1.Classification Rule									
rules	rule head	classification rule			0	2	1	1	0
Boolean term	constraint	rule head	rule fires	consequent	0	3	0	3	0
If-then rule	boolean conjunction	Non association rule mining	Supervised machine learning	rule learning	0	3	0	3	0
conditional rule if then	conditions boolean test	machine learning if-then	ml rules	AI ML conditional tests	0	3	0	3	0
constraints	satisfied	fire	rule head	if-then	0	1	0	5	0
if-then-else	conditional statement	if function	if-then statement	if/then statements	0	0	0	6	1
"if then"	"rule head"	"rule fires"	Furnkranz	Johannes	0	2	0	4	0
rule-based	classification	conjunction	Boolean	constraint	1	1	0	4	0
					1	71	3	203	5
2.Classifier Calibration:									

Figure 19: RQ3 Response Calculations.

For matching assessment, the researcher first identified EM for Control terms. Identifying and counting these matches manually would be quite extensive so the researcher decided to use Excel Find and Replace functionality to reformat matches, and then use the formula COUNTIF to calculate the color-coded matches (shown in Figure 4.9). Color coding was easier to do with find and replace functions in Excel where the exact (EM), partial 1 (PM1), and partial 2 (PM2) were

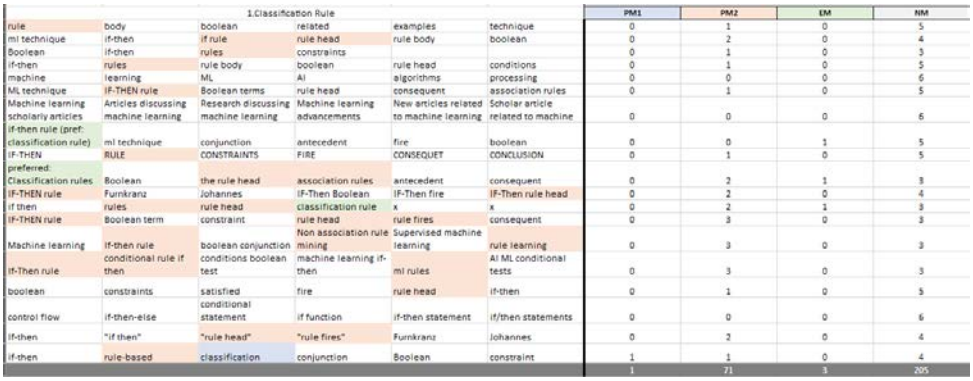
searched and replaced with the color-coded format. This necessitated PM1 and PM2 being assessed first, followed by EM due to the risk of EM being picked up in the partial find and replace actions.

The researcher quickly realized the original text entries were being overridden with this strategy so switched to Conditional Formatting to make sure data was not overwritten. Not wanting to depend on keyword matching alone, the researcher then went through each section to make sure nothing was missed, as well as identify matches that were similar enough to merit a match, such as a user entry *calibrated* as opposed to the Control *Calibration*, that would have otherwise been missed with a keyword identification alone. Because each prompt would have different Entries to trigger the conditional formatting, this approach was done for each prompt separately. The formatting is not necessarily needed for the chi-square to be performed; however, if this method is used on a small dataset (estimated below 50 responses) the color-coding helps manual counting of results, and even if tallies are not manual, having the color-coded visuals is helpful when discussing results to readers not familiar with data.

The researcher then used COUNTIF to calculate the matches for each participant for each Control term and realized Excel does not have a function to count Conditional Formatting. Having found no other way, the researcher resorted to counting the matches manually. Counting the total EM, PM, and NM for more than a handful of responses would be arduous if done manually, as well as increasing the possibility of human error, so in future iterations of this study, a python script will be created to string search, count, and print the sum of the matches to make future assessments more efficient and to extend the maximum response volume, as can be seen in Figure 23. Full mapping assessment can be found at:

https://docs.google.com/spreadsheets/d/1we9PNWifNozObrw-

VNxEWONNp6P45SIK/edit?usp=sharing&ouid=105068306933040109159&rtpof=true&sd=true



1.Classification Rule						PM1	PM2	EM	NM
rule	body	boolean	related	examples	technique	0	1	0	5
ml technique	if-then	if rule	rule head	rule body	boolean	0	2	0	4
Boolean	if-then	rules	constraints			0	1	0	3
if-then	rules	rule body	boolean	rule head	conditions	0	1	0	5
machine	learning	ML	AI	algorithms	processing	0	0	0	6
ML technique	IF-THEN rule	Boolean terms	rule head	consequent	association rules	0	1	0	5
Machine learning scholarly articles	Articles discussing machine learning	Research discussing machine learning	Machine learning advancements	New articles related to machine learning	Scholar article related to machine	0	0	0	6
if-then rule (pref: classification rule)	ml technique	conjunction	antecedent	fire	boolean	0	0	1	5
IF-THEN	RULE	CONSTRAINTS	FIRE	CONSEQUET	CONCLUSION	0	1	0	5
preferred: Classification rules	Boolean	the rule head	association rules	antecedent	consequent	0	2	1	3
IF-THEN rule	Furnkranz	Johannes	IF-Then Boolean	IF-Then fire	IF-Then rule head	0	2	0	4
if then	rules	rule head	classification rule	x	x	0	2	1	3
IF-THEN rule	Boolean term	constraint	rule head	rule fires	consequent	0	3	0	3
Machine learning	if-then rule	boolean conjunction	Non association rule mining	Supervised machine learning	rule learning	0	3	0	3
If-Then rule	conditional rule if then	conditions boolean test	machine learning if-then	ml rules	AI ML conditional tests	0	3	0	3
boolean	constraints	satisfied	fire	rule head	if-then	0	1	0	5
control flow	if-then-else	conditional statement	if function	if-then statement	if/then statements	0	0	0	6
If-then	"if then"	"rule head"	"rule fires"	Furnkranz	Johannes	0	2	0	4
if-then	rule-based	classification	conjunction	Boolean	constraint	1	1	0	4
						1	71	3	205

Figure 20: Matching color coding.

The structure of the results is each prompt is a Control term and has a result block from participants of the survey. Each row is a participant, and each result block is their responses, or Entries, for the prompt. Each prompt is a textual example of content where the current Control term is used. For this study, these are the prompts derived from the *Data Science Dictionary*, but any content could be used as a prompt so long as it is a meaningful representation of the Control term's context, such as a WordNet or Wikidata definition, full-text extractions, or scope notes from a thesaurus or the LCSH. Each row contains up to 6 Entries from the participant that are then analyzed for match categories by the researcher, and the total sum of each matched category is calculated. Each Natural Language term for each participant is assessed for a match category PM1, PM2, EM, or NM and each Control block is disjoint from any other Control block so no chi-square assessments will impact any other assessment in this study. For multi-word terms, the first position is treated as PM1 and the second position is PM2. For clarification, there is a key for how the

researcher assessed PMs for each Control block in the analysis below. The matches are counted as character string matches with the flexibility only for stemming, for example:

Complexity in Adaptive Systems=

Complexity (and its string variations such as complex) =(PM1)

and Adaptive (and its string variations such as Adapt) =(PM2)

Figure 21: Partial Match 1 and 2.

For some Control terms, especially simple one-word terms, there may not be both a PM1 or a PM2. Additionally, broad one word, or Entries posed as a question, will not be counted as a PM1 or PM2 because broad one word PMs would be too ambiguous in search, for example, “complexity” in the Control term Complexity in Adaptive Systems, and questions are outside the scope of this research. That said, one-word Entries for one-word Control terms or PMs that are specific, such as “classifier” from Control term Classifier Calibration, can be counted as a match because the singletons retain specific meaning in search.

For the survey, a total of six Natural Language tags was selected because the average journal article has on average between 3 to 10 subject tags assigned by indexers (NISO Z39.4-2021), indicating how many entry point opportunities would be typical in a scholarly research search. Three Entries was deemed too little, and ten took too much time for participants to enter and lead to more surveys being abandoned, so an average of 6 opportunities for each Control term was used per participant. This matching and calculation was done for each Control term block. The sum of all match types per Control term are then used in the chi-square assessments for each

Control block to determine if there is a significant difference between Control and users' Natural Language.

4.5 RQ3: Users' Natural Language matches to Control Terms

The survey sample was a total of 53 responses, meeting the target sample size for this study. Using the dataset derived from the survey (generating the Natural Language dataset), the researcher assessed if the Control dataset had any significant difference from the user's natural language (Research Question 3). This is the main question posed in this study. Many controlled vocabularies draw their terminology from the terminology used within research, as can be seen by the findings of RQ1, and before this study, the researcher assumed controlled vocabularies drew from existing controlled vocabularies, such as LCSH and MeSH, but RQ2 dispelled this theory, at least for machine learning terminology in the engineering discipline. More research would be needed to assess if this proves true for different terminology and disciplines. The null hypothesis is expecting at least 50% of terms will match but the literature review alludes to, but does not provide evidence for, the tendency for controlled vocabulary to not align with users' natural language, leading to the alternate hypothesis that 50% of the terms will not match. Countering the literature suppositions is the prevalent teachings from Chan (2007) and International Society for Knowledge Organization's definition of literary warrant (2021) which is that vocabularies are derived from authoritative literature resources. If this holds true, would controlled vocabularies not also match the language used by researchers in the field? If the controlled vocabulary terms do not match the users' natural language, how might this affect the search satisfaction of the users? RQ3 seeks to demonstrate how well controlled vocabularies align with users' natural language.

And based on these findings, what impact might controlled vocabularies have on the discoverability of research that may or may not make the search experience frustrating to the engineering community. To find out, a chi-square assessment was completed on each Control term, matching the users' natural language and the controlled vocabularies terms via Category Theory matching.

The following subsections assess each Control block using the matching technique used for the previous sections, PM1, PM2, EM, NM, as well as assessing the total matches for the fuzzy and browse search patterns. Chi-squared goodness-of-fit tests were conducted for each term and the hypothesis is accepted or rejected based on the 50% expected match rate. Furthermore, a key for how each Control block was assessed is noted, as are any decisions the researcher had to make per block. At the end of this section, all Control blocks are calculated to conduct a chi-square assessment on the full Control dataset in answer to RQ3.

Control: Classification Rule

KEY: PM1= Classifi; PM2= Rule; EM= contains classification rule

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 75 matches compared to 203 No Matches, giving a probability of p= < .001 and a chi-square score of 58.9, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 27% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 58.9, p= < .001

Table 20: Fuzzy Search Assessment for Classification Rule.

Category	Observed N	Probability	Expected N
Match	75	0.5	139
No Match	203	0.5	139
TOTAL	278	1	278

chi-squre	58.9352518
df	1
p-value	1.62947E-14
result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 3 Matches compared to 275 No Matches, giving a probability of p=< .001 and a chi-square score of 266, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 1% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 266, p=< .001

Table 21: Browse pattern Assessment for Classification Rule.

Category	Observed N	Probability	Expected N
Match	3	0.5	139
No Match	275	0.5	139
TOTAL	278	1	278

chi-squre	266.1294964
df	1
p-value	7.91573E-60
result	Reject Ho

The top user entries: If-THEN (62); Boolean (34)

Control: Classifier Calibration

KEY: PM1= Classifi; PM2= Calibrat; EM= contains classifier calibration

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 109 matches compared to 159 No Matches, giving a probability of p=.002 and a chi-square score of 9.3, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not match the Control term with 40% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 9.3, p= 0.002

Table 22: Fuzzy Search Assessment for Classifier Calibration.

Category	Observed N	Probability	Expected N		chi-squre	9.328358209
Match	109	0.5	134		df	1
No Match	159	0.5	134		p-value	0.002256344
TOTAL	268	1	268		result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 5 Matches compared to 263 No Matches, giving a probability of p=5.8 and a chi-square score of 248, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not match the Control term with 1% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 248, p= < .001

Table 23: Browse pattern Assessment for Classifier Calibration.

Category	Observed N	Probability	Expected N		chi-squre	248.3731343
Match	5	0.5	134		df	1
No Match	263	0.5	134		p-value	5.87646E-56
TOTAL	268	1	268		result	Reject Ho

The top user entries: Classifier (70); scores (25); Linear classifier (19)

Control: Clustering from Data Streams

KEY: PM1= Cluster; PM2= Data Streams; EM= contains clustering from data streams

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 114 matches compared to 147 No Matches, giving a probability of p= 0.04 and a chi-square score of 4.1, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 43% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 4.1, p= 0.04

Table 24: Fuzzy Search Assessment for Data Streams.

Category	Observed N	Probability	Expected N
Match	114	0.5	130.5
No Match	147	0.5	130.5
TOTAL	261	1	261

chi-squre	4.172413793
df	1
p-value	0.041087225
result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 1 Match compared to 260 No Matches, giving a probability of p=< .001 and a chi-square score of 257, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 0.4% matched, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 257, p=< .001

Table 25: Browse pattern Assessment for Data Streams.

Category	Observed N	Probability	Expected N
Match	1	0.5	130.5
No Match	260	0.5	130.5
TOTAL	261	1	261

chi-squre	257.0153257
df	1
p-value	7.6757E-58
result	Reject Ho

The top user entries: Clustering (103); Data (62); Grouping (40); Stream (39)

Control: Complexity in Adaptive Systems

KEY: PM1= Complex; PM2= Adaptive; EM= contains complexity in adaptive

This term is complex, meaning it has multiple concepts present, including a qualifying term, in this case, complexity seems to be the qualifier, but after reviewing the users' entries and the term prompt, it is clear complexity is the main topic in this term. Qualifiers make EMs difficult to identify because the core concept is at first ambiguous and therefore the Control term qualifier might not be obvious to a researcher. There was one participant that was very close to having an

EM with the entry A*daptive system and complex system* but because EMs need to be almost (i.e. stemming is allowed) exactly the same string to be counted as such, this entry was instead counted as a PM2.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 121 matches compared to 252 No Matches, giving a probability of p= 0.5 and a chi-square score of 0.39, signaling the null hypothesis is accepted for the fuzzy search pattern, meaning the users' natural language statistically matches the Control term with 48% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Accepted $\chi 2$ (1) = 0.39, p= 0.5

Table 26: Fuzzy search Ass. for Complexity in Adaptive Systems.

Category	Observed N	Probability	Expected N		
Match	121	0.5	126	chi-squre	0.396825397
No Match	131	0.5	126	df	1
TOTAL	252	1	252	p-value	0.528733325
				result	Accept Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 8 Matches compared to 244 No Matches, giving a probability of p=< .001 and a chi-square score of 221, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 3.2% matched, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 221, p=< .001

Table 27: Browse pattern Ass. for Complexity in Adaptive Systems.

Category	Observed N	Probability	Expected N		chi-squre	221.015873
Match	8	0.5	126		df	1
No Match	244	0.5	126		p-value	5.42998E-50
TOTAL	252	1	252		result	Reject Ho

The top user entries: Adaptive Systems (40); Internal Complexity (36); External Complexity (30)

Control: Computational Complexity of Learning

KEY: PM1= Comput complex; PM2= complex learning; EM= contains computational complexity in learning

This term is complex, meaning it has multiple concepts present, including a qualifying term, in this case, "complexity." These qualifiers make EMs difficult to identify because the core concepts are *Computational Learning* and *Computational Learning*, but the Control terms qualifier might not be obvious to a researcher. There was one participant that was very close to having an EM with the entry *machine learning complexity theory* but because EMs need to be almost (i.e. stemming is allowed) exactly the same string to be counted as such, this entry was instead counted as a PM2. Note, *complexity* was entered a few times on its own in participants' entries, as was *learning* and *computational* but these were counted as a NM because the entry had very little context to add to an otherwise very common word. Also, ML is short for machine learning but the researcher is only counting fuzzy matches for PMs which means acronyms that would contain a PM, but which are not written out, will not count toward a PM match.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 30 matches compared to 223 No Matches, giving a probability of p= < .001 and a chi-square score of 147, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural

language does not statistically match the Control term with 11% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 147, p= < .001

Table 28: Fuzzy search Ass. for Computational Complexity of Learning.

Category	Observed N	Probability	Expected N
Match	30	0.5	126.5
No Match	223	0.5	126.5
TOTAL	253	1	253

chi-squre	147.229249
df	1
p-value	6.99224E-34
result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 3 Matches compared to 250 No Matches, giving a probability of p=< .001 and a chi-square score of 241, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 1% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 241, p=< .001

Table 29: Browse pattern Ass. for Computational Complexity of Learning.

Category	Observed N	Probability	Expected N
Match	3	0.5	126.5
No Match	250	0.5	126.5
TOTAL	253	1	253

chi-squre	241.1422925
df	1
p-value	2.21634E-54
result	Reject Ho

The top user entries: Learning (94); Complexity (53); PAC (24); Oracle (24); Query-based (24)

Control: Deep learning

KEY: PM1= Neural Net; PM2= NA; EM= Deep learn

This is a simple term, meaning it is one word or short phrase term. This makes finding PMs difficult because there is so little room for fuzzy matching to identify a shorter string match for both PM1 and PM2 so, building on the findings from the Baseline mapping procedure previously described in this study, Neural Networks will be treated as a PM for the Control term Deep Learning, and Topology will be treated as a PM for the Control term Topology of Neural Networks. This is not ideal because Neural Network should instead be treated as an alternate form for Deep Learning in controlled vocabulary used for browse, but for the sake of this study, they will be treated as PMs, which are only counted as matches in the fuzzy search pattern, to understand if Control terms significantly differ from users' natural language in fuzzy search.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 32 matches compared to 223 No Matches, giving a probability of p= < .001 and a chi-square score of 143, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 12% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 143, p= < .001

Table 30: Fuzzy Search Assessment for Deep learning.

Category	Observed N	Probability	Expected N	chi-squre	143.0627451
Match	32	0.5	127.5	df	1
No Match	223	0.5	127.5	p-value	5.69524E-33
TOTAL	255	1	255	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 23 Matches compared to 232 No Matches, giving a probability of p=< .001 and a chi-square score of 171, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 9% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 171, p=< .001

Table 31: Browse pattern Assessment for Deep learning.

Category	Observed N	Probability	Expected N
Match	23	0.5	127.5
No Match	232	0.5	127.5
TOTAL	255	1	255

chi-squre	171.2980392
df	1
p-value	3.85184E-39
result	Reject Ho

The top user entries: Pattern Recognition (37); NN (35)

Control: Evaluation of Learning Algorithm

KEY: PM1= evaluation of learning; PM2= learning algorithm; EM= evaluat learning algor

Notte, there were quite a few instances where the participants switched the concept order in the complex term, instead of having evaluation of learning algorithm, many had learning algorithm evaluation. While this seems like a mundane switch in word order, in both browse and fuzzy search the order of words is important and can change the meaning of a query. To test this theory, the researcher did a search in INSPEC and Google Scholar (the same databases used to assess the impact of the machine learning methodology terms from the Encyclopedia) for both forms of the term and found that the search engines did indeed treat these as different queries with slightly different contexts. For this reason, only Evaluation of Learning was used as PM1, Learning Algorithms PM2, and Evaluat of Learning Algor as EM.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 66 matches compared to 203 No Matches, giving a probability of p= < .001 and a chi-square score of 24.8, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 32% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 24, p= < .001

Table 32: Fuzzy search Ass. for Evaluation of Learning Algorithms.

Category	Observed N	Probability	Expected N	chi-squre	24.83251232
Match	66	0.5	101.5	df	1
No Match	137	0.5	101.5	p-value	6.25338E-07
TOTAL	203	1	203	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 4 Matches compared to 203 No Matches, giving a probability of p=< .001 and a chi-square score of 187, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 2% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 187, p=< .001

Table 33: Browse pattern Ass. for Evaluation of Learning Algorithms.

Category	Observed N	Probability	Expected N	chi-squre	187.3152709
Match	4	0.5	101.5	df	1
No Match	199	0.5	101.5	p-value	1.22566E-42
TOTAL	203	1	203	result	Reject Ho

The top user entries: Learning Algorithm (73); Properties (27); Suitability (21)

Control: Feature Selection

KEY: PM1= features with context; PM2= selection with context; EM= contains feature selection

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 32 matches compared to 173 No Matches, giving a probability of p= < .001 and a chi-square score of 96.9, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 15% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 96.9, p= < .001

Table 34: Fuzzy Search Assessment for Feature Selection

Category	Observed N	Probability	Expected N	chi-squre	96.9804878
Match	32	0.5	102.5	df	1
No Match	173	0.5	102.5	p-value	7.00139E-23
TOTAL	205	1	205	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 17 Matches compared to 188 No Matches, giving a probability of p=< .001 and a chi-square score of 142, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 8% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 142, p=< .001

Table 35: Browse pattern Assessment for Feature Selection.

Category	Observed N	Probability	Expected N	chi-squre	142.6390244
Match	17	0.5	102.5	df	1
No Match	188	0.5	102.5	p-value	7.04948E-33
TOTAL	205	1	205	result	Reject Ho

The top user entries: Dimensionality Reduction Technique (44); Feature (33)

Control: Precision and Recall

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 11 matches compared to 191 No Matches, giving a probability of p= < .001 and a chi-square score of 160, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 5% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 160, p= < .001

Table 36: Fuzzy Search Assessment for Precision and Recall.

Category	Observed N	Probability	Expected N	chi-squre	160.3960396
Match	11	0.5	101	df	1
No Match	191	0.5	101	p-value	9.27084E-37
TOTAL	202	1	202	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 2 Matches compared to 200 No Matches, giving a probability of p=< .001 and a chi-square score of 194, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 1% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 194, p=< .001

Table 37: Browse pattern Assessment for Precision and Recall.

Category	Observed N	Probability	Expected N	chi-squre	194.0792079
Match	2	0.5	101	df	1
No Match	200	0.5	101	p-value	4.09238E-44
TOTAL	202	1	202	result	Reject Ho

The top user entries: Information Retrieval (68); Confusion Matrix (24)

Control: Regression

KEY: PM1= regression with additional concepts; PM2= NA; EM= contains regression or regression with simple qualifier

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 70 matches compared to 207 No Matches, giving a probability of p= < .001 and a chi-square score of 21, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 33% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 21, p= < .001

Table 38: Fuzzy Search Assessment for Regression.

Category	Observed N	Probability	Expected N	chi-squre	21.68599034
Match	70	0.5	103.5	df	1
No Match	137	0.5	103.5	p-value	3.21128E-06
TOTAL	207	1	207	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 58 Matches compared to 149 No Matches, giving a probability of p=< .001 and a chi-square score of 40, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 28% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 40, p=< .001

Table 39: Browse pattern Assessment for Regression.

Category	Observed N	Probability	Expected N	chi-squre	40.00483092
Match	58	0.5	103.5	df	1
No Match	149	0.5	103.5	p-value	2.53336E-10
TOTAL	207	1	207	result	Reject Ho

The top user entries: Parametric (57); Regression Function (51)

Control: Topology of Neural Networks

KEY: PM1= topology; PM2= deep learning; EM= neural network

This term is complex, meaning it has multiple concepts present, including a qualifying term, in this case "of Neural Networks." These qualifiers make EMs difficult to identify because the core concepts are *Topology* and *Neural Networks*, but the Control terms qualifier might not be obvious to a researcher. There is the question of the Control term Deep Learning which also uses Neural Network as a PM. Each Control term is treated as disjoint of any other Control term so these two analyses will not impact the chi-square assessment, but Deep Learning will be treated as a PM2 because of the findings from the Control mapping. Topology will be PM1 and Neural Network treated as EM.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 44 matches compared to 174 No Matches, giving a probability of p= < .001 and a chi-square score of 77, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 20% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 77, p= < .001

Table 40: Fuzzy Search Ass. for Topology of Neural Networks.

Category	Observed N	Probability	Expected N	chi-squre	77.52293578
Match	44	0.5	109	df	1
No Match	174	0.5	109	p-value	1.3119E-18
TOTAL	218	1	218	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 1 Match compared to 218 No Matches, giving a probability of p=< .001 and a chi-square score of 214, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 0.5% matched, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 214, p=< .001

Table 41: Browse pattern Ass. for Topology of Neural Networks.

Category	Observed N	Probability	Expected N	chi-squre	214.0183486
Match	1	0.5	109	df	1
No Match	217	0.5	109	p-value	1.8248E-48
TOTAL	218	1	218	result	Reject Ho

The top user entries: Topology (50); Neurons (36); Networks (36)

Control: Weight

KEY: PM1= weights with qualifiers; PM2= weighted with qualifiers; EM= weight

Note, because weight is so ambiguous and the stemmed term implies different contexts, weight with qualifiers will be counted as PM1, and weighted plus a qualifier will be PM2, and weight will be treated as EM.

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 45 matches compared to 209 No Matches, giving a probability of p= < .001 and a chi-square score of 67.7, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural

language does not statistically match the Control term with 21% match, compared to the 50% evenly distributed expectation.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 67.7, p= < .001

Table 42: Fuzzy Search Assessment for Weight.

Category	Observed N	Probability	Expected N	chi-squre	67.75598086
Match	45	0.5	104.5	df	1
No Match	164	0.5	104.5	p-value	1.85035E-16
TOTAL	209	1	209	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 7 Matches compared to 209 No Matches, giving a probability of p=< .001 and a chi-square score of 181.9, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 3% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 181.9, p=< .001

Table 43: Browse pattern Assessment for Weight.

Category	Observed N	Probability	Expected N	chi-squre	181.937799
Match	7	0.5	104.5	df	1
No Match	202	0.5	104.5	p-value	1.8295E-41
TOTAL	209	1	209	result	Reject Ho

The top user entries: Connection (46); Neurons (34); Networks (25)

4.6 RQ3 Initial Findings:

The researcher then calculated the total match and no match for both the browse and fuzzy search patterns. Resulting in the following chi-square calculations for the entire Control dataset (Table 44):

Table 44: Summary of Chi-Square calculations for RQ3.

Control Term	Fuzzy Matchign Ho	Browse Ho	Search M	Search NM	Browse M	Browse NM
Classification Rule	Reject	Reject	75	203	3	275
Classifier Calibration	Reject	Reject	109	159	5	263
Clustering from Data Streams	Reject	Reject	114	147	1	260
Complexity in Adaptive Systems	Accept	Reject	121	131	8	244
Computational Complexity of Learning	Reject	Reject	30	223	3	250
Deep Learning	Reject	Reject	32	223	23	232
Evaluation of Learning Algorithm	Reject	Reject	66	137	4	199
Feature Selection	Reject	Reject	32	173	17	188
Precision and Recall	Reject	Reject	11	191	2	200
Regression	Reject	Reject	70	137	58	149
Topology of a Neural Network	Reject	Reject	44	174	1	217
Weight	Reject	Reject	45	164	7	202
TOTAL			**749**	**2062**	**132**	**2679**

For the fuzzy search pattern, PMs and EMs are counted as a match for a total of 749 matches compared to 2,062 No Matches, giving a probability of $p = < .001$ and a chi-square score of 613, signaling the null hypothesis is rejected for the fuzzy search pattern, meaning the users' natural language does not statistically match the Control term with 26% match, compared to the 50% evenly distributed expectation. This finding therefore accepts the alternate hypothesis that the terms do not match to a significant degree.

Fuzzy Search Pattern Findings: Null Hypothesis Rejected $\chi 2\ (1) = 613$, $p = < .001$

Table 45: Final Chi-Square calculations RQ3 (fuzzy search).

Category	Observed N	Probability	Expected N	chi-squre	613.2938456
Match	749	0.5	1405.5	df	1
No Match	2062	0.5	1405.5	p-value	2.1495E-135
TOTAL	2811	1	2811	result	Reject Ho

For the browse pattern, EMs are counted as a match and PMs and NMs are counted as No Match for a total of 132 Matches compared to 2,679 No Matches, giving a probability of p=< .001 and a chi-square score of 2307, signaling the null hypothesis is rejected for the browse search pattern, meaning the users' natural language does not statistically match the Control term with 4% match, compared to the 50% evenly distributed expectation.

Browse Search Pattern Findings: Null Hypothesis Rejected $\chi 2$ (1) = 2307, p=< .001

Table 46: Final Chi-Square calculations RQ3 (browse pattern).

Category	Observed N	Probability	Expected N	chi-squre	1697.766636
Match	132	0.5	1097	df	1
No Match	2062	0.5	1097	p-value	< .001
TOTAL	2194	1	2194	result	Reject Ho

In all but one case (fuzzy search pattern for Complexity in Adaptive Systems), the Control dataset did not match the users' natural language to a significant degree. The percentage matched in the fuzzy search pattern was always greater than subject headings alone in the browse pattern (an average 26% match versus 5% matching in the browse pattern), indicating there is at least a greater chance to discover research on a given machine learning methodology topic when using a search engine that uses fuzzy search.

The next chapter will focus on exploring the results from this analysis, understanding how they shed light on the null hypothesis of this study, and what ramifications may arise from these findings.

5.0 Discussion of Findings

This chapter provides discussion on the methods used in this study, as well as further interpretation of the major results and patterns from the analysis chapter, compiled to answer the main research question (RQ3). The findings of RQ3 answers to what extent controlled vocabularies in the engineering-domain match the engineering user's natural language. The findings of RQ3 are then used to suggest updates to controlled vocabularies for including more user language, which strengthens the bridge between users' queries and content. Furthermore, the findings allude to three additional suggestions to strengthen vocabularies for more precise subject search.

5.1 Discussion on Study Design Methods

Starting the research itself was a challenge. There were three main areas focused on in the research methods for this study that will be mentioned here:

1. Lack of machine learning terminology resources and the problems searching for materials
2. Lack of participation in survey and scope of vocabulary review
3. Lack of methods for controlled vocabulary analysis

Each will be discussed below to explore the difficulties found in these areas during this study, as well as how these issues may be overcome or addressed in future studies.

When searching for the original Control dataset focusing on literary warrant, the researcher searched in the University of Pittsburgh library system as well as external Google Scholar for

materials dealing with taxonomies, controlled vocabularies, ontologies, dictionaries, encyclopedias, or lists of any sort that described machine learning methods. The issue here is that machine learning is itself a method for creating and assigning taxonomies so queries tended to produce results that showed how machine learning was used to *make and create* taxonomies, not that revealed the vocabulary and terminology used *in the machine learning community*. When the search was broadened to include materials dealing with machine learning types, broad level types such as supervised, unsupervised, etc. were retrieved. Again, this was too broad to be used in this study because these are more like classifications of machine learning types, not the methods themselves. Searching on machine learning algorithms retrieved lists of different algorithms and their use but these can be used for many different methods so they are not suitable for this study.

5.1.1 Lack of machine learning terminology resources

The difficulty this researcher had finding controlled lists containing descriptions of methods used in machine learning research is a finding of this study in and of itself. Having to search so exhaustively to find even one list of machine learning methodologies with descriptions of those techniques implies that an engineering user trying to do the same would encounter frustrations as well. The literature review alluded to reference materials for methodologies lacking, and the finding of the present study confirms that indeed, reference materials on methodologies used in machine learning were scarce. This makes the research process, especially for new students, that much more difficult. The lack of standardization also may contribute greater difficulty in finding research using the same method (which may be labeled differently in each study) in a general search. Poor multidisciplinary search indicates the value of standardizing the methodologies into a more formal, controlled vocabulary to facilitate better recall of studies using

similar methods. While controlled vocabularies may need improvement, which is a theme of the present study, they are the basis of any mapping project.

Machine learning terminology, and a good bit of other technical terminology for that matter, is constantly changing because this technology updates at a fast pace, which may account for the lack of standardized vocabulary because traditional vocabulary creation needs to wait until a term is "settled" into its generally accepted form and context. But this means newer research methods will not get the benefit of subject search in information retrieval. The impact of this can be seen in this study in the case of the labels Deep Learning and Neural Network which seemed to be in flux, based on their use in literature and controlled vocabularies. According to standard indexing and controlled vocabulary rules, terms that can mean the same thing should not be used (Chan, 2007) but it seems with areas that have a quickly evolving vocabulary, controlled subjects may have difficulty keeping up. These issues all made the identification of the Control set difficult to create.

5.1.2 Lack of participation in survey

There was also considerable difficulty obtaining enough participation in the survey portion of this work which may allude to difficulties with full vocabulary assessments. Most participants found the survey too long, even with only 12 prompts. The current study is small in comparison (12 control terms) to the effort needed to assess an entire controlled vocabulary (the typical vocabulary can contain over 1,000 subjects). With these two elements at play, assessing an entire vocabulary would not be feasible on a term-by-term basis. Two approaches could alleviate these issues. The first would be to assess vocabulary terms with the users who would use them the most, specifically the library or research team researching the specific database or library system. Users

that have the most interest for a specific term might be more willing to spend time on a list of subjects that are in that subject interest. Users of a particular discovery system would also have more incentive to participate because it would directly affect their work, unlike a survey to a more general population. Focusing on the users of a particular discovery system is also more user-centered because the culture of the researchers for an institution or library would be unique, potentially with different natural language, so assessing the vocabulary with the specific groups using that specific discovery system would be beneficial from both a user-centered and participation perspective. Focusing on a specific discovery system or ecosystem would also assist in better targets for natural language assessment. If a library is more focused on robotics, perhaps the terminology used in robotics would be targeted for natural language assessment to assist in the core research of that institution. Another strategy to scope the analysis more effectively may be to assess newer terminology, those that might be in flux in users vernacular like Precision and Recall, or in the case of literature or other vocabulary terminology being in flux, like Deep Learning and Neural Network. Terms that are more established, such as Adaptive Systems or Complexity Theory, may not need to be assessed or assessed too often because they are more set or understood. Lastly, an analysis schedule may also help scope the work in a way to spread out the assessment, while also maintaining assessment yearly to ensure user-entered terminology is accounted for. These approaches may help to break the assessment apart into more manageable pieces.

5.1.3 Lack of methods for controlled vocabulary analysis

Finally, the methods for vocabulary assessment itself were problematic. The main methods for vocabulary user-centered design are open and closed card sorts. Interestingly, most vocabulary assessment methods are focused only on how users' will use or interact with the vocabulary, not

as much about how vocabularies match one another, or if vocabularies align with users' natural language. The issue with traditional card sorts are they function on the browse pattern alone and require terminology to be pre-populated (closed card sorts are used to establish hierarchy whereas open card sorts are used to label clusters of terms) Neither is suited for gathering new terminology without the aid of existing terms. In the fuzzy search pattern, users typically do not use any controlled vocabulary, preferring a more Google-like search, therefore making traditional card sorts unsuitable for this study. The researcher went to considerable lengths to identify an existing methodology to accommodate gathering users' natural language without the aid of pre-populating terms. As such, this study used a new design called a bidirectional card sort, of the researcher's own design, that combines the traditional card sorting methods with aspects of a diary study where in place of pre-populated terms, definitions, or abstract level text defining the terms are used as prompts to the participant, and the "cards" are open for the participants to enter in what they would use to search for content that is similar to the prompt. The prompt in this is the bidirectional pivot where both controlled vocabularies and users preferred terms can be assessed. Improvements to this method can certainly be made, such as:

1. Automation where the user's natural language could be coded as EM, PM1, and PM2 and automatically counted,
2. Creating a full-text corpus of definitions for prompts or,
3. Having a pre-mapped corpus of different vocabularies to consult instead of mapping the vocabularies during the assessment process.

These, as well as the other improvements mentioned in this section, will be considered for future studies based on this work. These updates may also help others interested in user-centered vocabulary assessment methods start to perform and include more user-centered terminology.

Moving on, the remaining sections of the discussion will focus on the research questions presented as the basis of this study.

5.2 Answering the main research question

The structure of the preceding discussion is as follows: First, a summary of the findings from RQ1-4 and a description of how the findings of this study have been codified for reuse. Next, the findings from the comparison of search vs. browse patterns are discussed. Then, the causes of misalignment will be discussed and updates to address it are suggested. In this section, each Control term will be discussed separately, focusing on findings from RQ1-3. In each section, the ramifications of information discovery will be addressed, as well as how the Control can be updated to be more user-centered. The last section will summarize and discuss how the individual Control sections culminate to address the hypothesis of engineering-domain controlled vocabularies being in misalignment with users' natural language. Furthermore, findings may suggest that adding users' natural language to controlled vocabularies will add more user-centered design to information discovery, primarily for the scope of this study which is machine learning methodologies for the engineering community, leading to a more satisfactory search experience.

5.2.1 Summary and Codification of Findings

Throughout the analysis, there were suggested changes to each term to add more user-centered terminology. These were modeled in Web Protege and the resulting data artifact was uploaded to Bioportal, an open data resource for vocabulary mappings and ontologies where

datasets can be explored and downloaded, to foster greater reproducibility of this research as well as serving as a public dataset to continue expanding on this research. This work can be found on Bioportals open terminology dataset portal at the following link, https://bioportal.bioontology.org/ontologies/MLTX.

From the term by term discussions, this study has identified the following summary of findings from the RQ1-4. The null hypothesis of the current study was rejected. With only 26% of the terms matching between controlled terms and users' natural language, it is clear that at least in the subset of terms analyzed in this research, these vocabulary terms are not user-focused. The findings of RQ1-2 point to the main focus of controlled vocabularies as being from literature and scientific warrant. Exploring these findings, and suggested changes to address the identified issues, are described in depth below (Table 47):

Table 47: Summary of RQ1-4 Findings. Researcher created.

Control Term	Fuzzy Matching Ho	Browse Ho	Search M	Search NM	Search % Match	RQ1	RQ2	RQ3	Top control	Top user entries	Browse M	Browse NM	Browse % Match
Classification Rule	Reject	Reject	75	203	27.0%	Literary	Scientific	Fail	Rule	IF-THEN; Boolean	3	275	1.1%
Classifier Calibration	Reject	Reject	109	159	40.7%	Fail	Fail	Fail	Classifier	Classifier; Rule; Linear Classifier	5	263	1.9%
Clustering from Data Streams	Reject	Reject	114	147	43.7%	Literary	Scientific	Fail	Clustering	Clustering; Data; Grouping; Stream	1	260	0.4%
Complexity in Adaptive Systems	Accept	Reject	121	131	48.0%	Literary	Scientific	User	Adaptive System	Adaptive Systems; Internal Complexity; External Complexity	8	244	3.2%
Computational Complexity of Learning	Reject	Reject	30	223	11.9%	Literary	Scientific	Fail	Computational Complexity	Learning; Complexity; Oracle; PAC; query-based	3	250	1.2%
Deep Learning	Reject	Reject	32	223	12.5%	Fail	Scientific	Fail	Neural Net	Pattern Recognition; NN	23	232	9.0%
Evaluation of Learning Algorithm	Reject	Reject	66	137	32.5%	Literary	Scientific	Fail	Computational Complexity	Learning Algorithm; Properties; Suitability	4	199	2.0%
Feature Selection	Reject	Reject	32	173	15.6%	Literary	Scientific	Fail	Feature	Dimensionality Reduction Technique; Feature	17	188	8.3%
Precision and Recall	Reject	Reject	11	191	5.4%	Fail	Fail	Fail	Information Retrieval	Information Retrieval; Confusion Matrix	2	200	1.0%
Regression	Reject	Reject	70	137	33.8%	Literary	Scientific	Fail	Regression	Parametric; Regression Function	58	149	28.0%
Topology of a Neural Network	Reject	Reject	44	174	20.2%	Literary	Scientific	Fail	Topology	Topology; Neurons; Networks	1	217	0.5%
Weight	Reject	Reject	45	164	21.5%	Literary	Scientific	Fail	Weight		7	202	3.3%
TOTAL			749	2062	26.6%	3	2	11			132	2679	4.7%

Before discussing the findings of RQ1-3, RQ4 will be addressed.

5.2.2 Search compared to Browse Pattern

As mentioned previously, RQ4 was found to favor more matches if a fuzzy search was used to supplement controlled vocabulary tagging so it will not be mentioned again after this section. The difference between search and browse patterns was assessed throughout this study to determine the usability of subject headings alone to guide users to the information on any given Control term. This evidence helps to answer RQ4, whether controlled vocabularies alone would match the users' natural language without fuzzy search also being applied. Many discovery interfaces have started to move away from a hierarchical display of controlled vocabularies in favor of a fuzzy search pattern that utilizes subject tags. Evidence was found within this study to support that subject search seems to benefit from fuzzy search patterns so this trend seems to be in the positive direction; however, the terminology assessed in both the Fuzzy Search and Browse patterns did not match to a significant degree, outside of the term Complexity in Adaptive Systems, so no correlation can be made between better controlled vocabulary search in either pattern, only that the Fuzzy Search pattern provided a higher match probability than a search without.

Table 48: Summary of RQ4 Findings. Researcher created.

Control Term	Fuzzy Matchign Ho	Browse Ho	Search % Match	Browse % Match
Classification Rule	Reject	Reject	27.0%	1.1%
Classifier Calibration	Reject	Reject	40.7%	1.9%
Clustering from Data Streams	Reject	Reject	43.7%	0.4%
Complexity in Adaptive Systems	Accept	Reject	48.0%	3.2%
Computational Complexity of Learning	Reject	Reject	11.9%	1.2%
Deep Learning	Reject	Reject	12.5%	9.0%
Evaluation of Learning Algorithm	Reject	Reject	32.5%	2.0%
Feature Selection	Reject	Reject	15.6%	8.3%
Precision and Recall	Reject	Reject	5.4%	1.0%
Regression	Reject	Reject	33.8%	28.0%
Topology of a Neural Network	Reject	Reject	20.2%	0.5%
Weight	Reject	Reject	21.5%	3.3%
TOTAL			**26.6%**	**4.7%**

In every chi-square test for RQ1-3 (see Table 5.2), the browse pattern never matched the Control dataset to a significant degree and had a much lower match percentage than the Fuzzy Search pattern. Where once the vocabulary hierarchy was the only means to search within the library, the card catalog being a major contributor to this use case, the modern library now can supplement the controlled vocabularies with fuzzy search, helping to bridge the gap between users and the content tags. Due to the physical nature of the catalog, vocabularies had to be more rigid and controlled because the content could only belong to one physical location in the stacks, and therefore have one primary tag or location to belong. With the addition of a search engine and more information at hand than ever before, the modern library user can find content with more specific terminology, as well as being able to access content from any given tag. This stretching of the catalog, as well as the expanding information literacy of the modern user, necessitates the need for fuzzy search to help supplement controlled vocabulary search, as can be seen through the findings from RQ4 where if a fuzzy search was not available, there would be a much lower match rate to users' natural language and that of the controlled vocabularies. Likely because the variability of user terminology is so broad that exact matches are highly unlikely (4% chance) and fuzzy matches are more likely (26% chance) than browse patterns alone. With this in mind, the remaining discussion will only focus on the fuzzy search pattern findings.

5.2.3 Exploring the main causes for misalignment

The misalignment between users' natural language and controlled vocabulary was the most common issue found. In addition there were three additional issues that decrease precision and make the bridge between users' natural language and content tags more difficult. Causes 2-4 were

found in all three warrant types which indicates these are broader issues in controlled vocabulary, even outside of making vocabularies more user-centered.

The following section will use the findings of this study to suggest updates to controlled vocabularies for more user-centered terms and more precise information search. The following is a summary of the suggested updates for each term (see Table 5.3):

Table 49: Summary of Suggested Updates.

Suggested Alignment Updates			Suggestions to Improve Alignment			
Control Term	Top control	Top user entries	Map Natural Language Equivalents	Qualifiers	Similarity of Terms	Ambiguous
Classification Rule	Rule	IF-THEN; Boolean	X		X	
Classifier Calibration	Classifier	Classifier; Rule; Linear Classifier	X	X	X	X
Clustering from Data Streams	Clustering	Clustering; Data; Grouping; Stream	X	X		
Complexity in Adaptive Systems	Adaptive System	Adaptive Systems; Internal Complexity; External Complexity		X	X	
Computational Complexity of Learning	Computational Complexity	Learning; Complexity; Oracle; PAC; query-based	X	X	X	X
Deep Learning	Neural Net	Pattern Recognition; NN	X		X	X
Evaluation of Learning Algorithm	Computational Complexity	Learning Algorithm; Properties; Suitability	X	X		X
Feature Selection	Feature	Dimensionality Reduction Technique; Feature	X			X
Precision and Recall	Information Retrieval	Information Retrieval; Confusion Matrix	X			
Regression	Regression	Parametric; Regression Function	X			X
Topology of a Neural Network	Topology	Topology; Neurons; Networks	X	X	X	
Weight	Weight		X			X

Cases Found to Lead to Misalignment of Terms:

1. Vocabularies not aligned to users' natural language
2. Qualifiers making the Control too specific to match users queries
3. Being too similar to other Control terms
4. Ambiguous terminology

Case 1: Vocabularies not aligned to users' natural language

For RQ3, the match between users' natural language and controlled vocabulary was a 26% match rate compared to the 50% expected match rate. As stated in the analysis section, the following is the aggregation of the Fuzzy Search Pattern Findings (Null Hypothesis Rejected $\chi 2$ (1) = 613, p= < .001) and supports rejecting the null hypothesis (Table 5.4).

Table 50: Summary Ch-Square Assessment for RQ3- Findings are Ho is rejected.

Category	Observed N	Probability	Expected N	chi-squre	613.2938456
Match	749	0.5	1405.5	df	1
No Match	2062	0.5	1405.5	p-value	2.1495E-135
TOTAL	2811	1	2811	result	Reject Ho

This finding is surprising because even though some variation was expected, the researcher was not expecting there to be such a large discrepancy between the term matches for content, the tags, and what language users would use themselves. The following section will suggest specific updates for each vocabulary term to be more user-centered, followed by three additional vocabulary suggestions based on the findings of this study. A summary of these suggestions is listed in Figure 5.13.

In the following subsections, each Control term will summarize the findings for each research question and discuss what changes could be made to make the term more user-centered. These suggestions will be recorded in the Protege dataset derivative of this work. The final subsection will assess the findings as a whole to determine if in the scope of this study, do controlled vocabulary terms align with users' natural language and based on the discussed suggested updates, what learnings can be taken away from this study.

Classification Rule

RQ1: 3 out of 4 control vocabulary terms matched Control, two of which were similar to one another. The TRT did not cover this topic. Classifications were more common among vocabularies and are a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 3 out of 4 control vocabulary terms were matched when compared to each other. The TRT did not cover this topic. Classifications were more common among vocabularies and are a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 27% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing “rule” were most prevalent (71), indicating a preferred label emphasizing the rules aspect of this term might be more user-centered. In addition, the highest volume user entries (regardless of a match to Control) were IF-THEN and Boolean. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, alternative labels containing IF-THEN and Boolean could be considered.

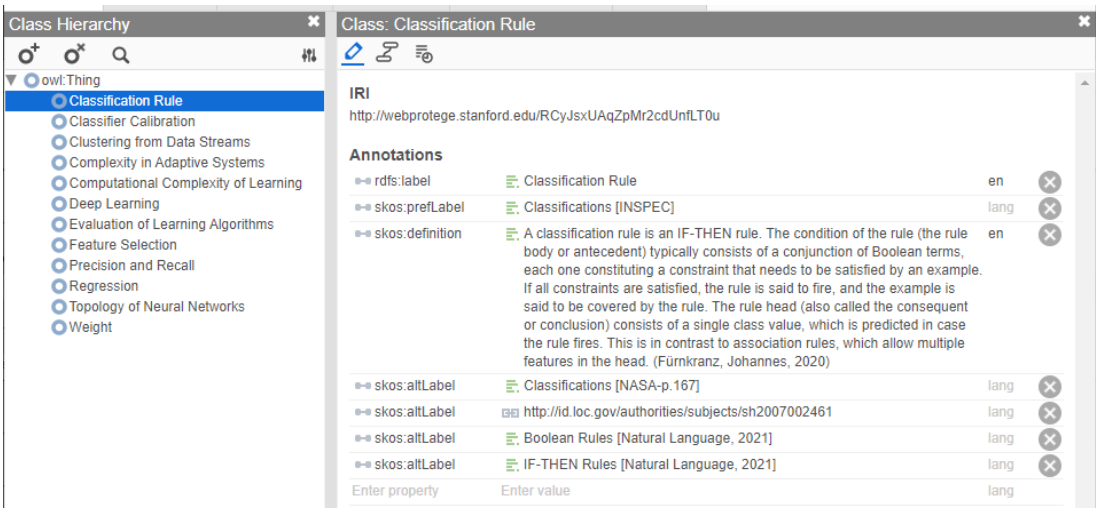


Figure 22: Equivalents for Classification Rule.

Classifier Calibration

RQ1 and RQ2: 1 out of 4 control vocabulary terms matched Control. The TRT, NASA, and INSPEC did not cover this topic. The clarifying portion of this term (Calibration) implies that the main topic is Classifier, which is a closer match to the previous Control term and may explain why many of the controlled vocabularies did not cover this topic. Comparing matches between the expert Encyclopedia terminology and the controlled vocabularies as well as comparing vocabulary to vocabulary, the majority did not match. This indicates the terminology was not likely derived from literary warrant or scientific warrant. Additional findings here may be that complex or combination terms may be better suited as alternative or narrower terms of the parent they are describing, in this case, a variant of Classifier.

RQ3: 40% of users' natural language matched the Control, a higher percentage but not statistically significant, leading to a rejection of the null hypothesis. Out of the matches, entries containing “classifier” were most prevalent (80), indicating a preferred label emphasizing the classifier aspect of this term, or as RQ1-2 suggested a broader/narrower relationship, might be

more user-centered. In addition, the highest volume user entries (regardless of a match to Control) were Classifier, Rules, and Linear Classifier. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. However, this one was very close and if the PM2 portion (Calibration, the clarifying aspect of this term) were not counted in the matching assessment, this term may have significantly matched. To supplement with more user-centered terminology, using the broader or root form of this term without the clarifying aspect, or including Rule or Linear Classifier as an alternative or related labels, could be considered.

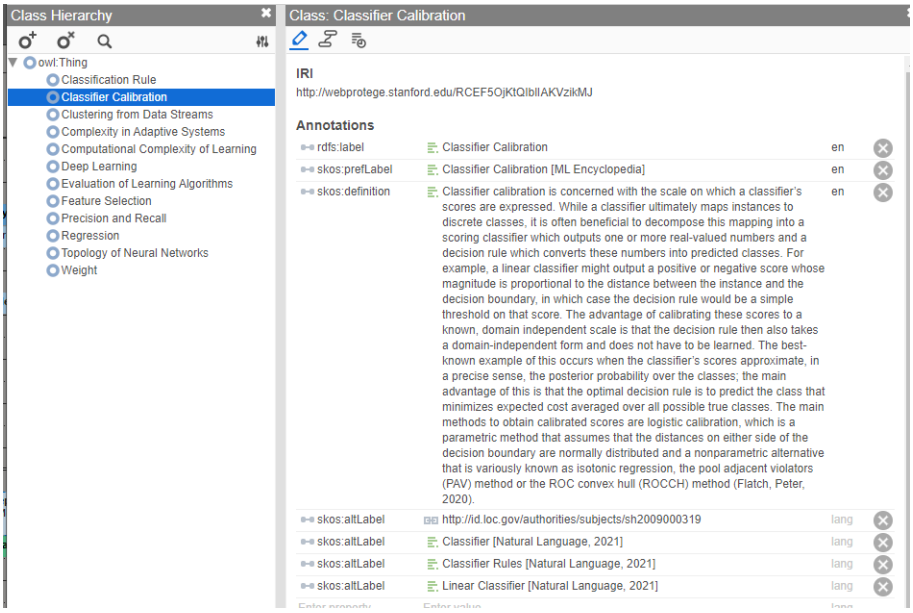


Figure 23: Equivalents for Classifier Calibration.

Clustering from Data Streams

RQ1: 3 out of 4 control vocabulary terms matched Control, two of which were similar to one another. The TRT did not cover this topic. Clustering was more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between the expert

Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 3 out of 4 control vocabulary terms were matched when compared to each other. The TRT did not cover this topic. Variations on clustering were more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 43% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing “cluster” were most prevalent (95), indicating a preferred label emphasizing the clustering aspect of this term might be more user-centered. In addition, the highest volume user entries (regardless of a match to Control) were Clustering, Data, Grouping, and Stream. These are so close to the Control which can also be seen by the high percentage that matched; however, it was not enough to be significant and therefore the rejected null hypothesis indicates the terminology was not likely derived from user warrant.

This Control block seemed to have quite a few user entries where the query was broken into individual words. If this were to be used in practice, a brute-force keyword search may still result in the user finding the information they seek, but at a much higher cost in time and frustration due to multiple queries to get the desired results. This also raises the question, do users typically search with one or multiple words for any given query? The Control dataset leans to multi-word queries, and even sometimes question queries, but this is not the focus of the current study, although it is certainly an additional line of inquiry to pursue at a later date. To supplement with more user-centered terminology, alternative labels containing IF-THEN and Boolean should be considered.

As with the previous control term Classifier Calibration, this Control term seems to suffer from having a qualifying statement that may confuse users. Unlike Classifier Calibration, which has a distinct verb as its qualifier, Clustering from Data Streams could be focused more heavily on either Clustering or from Data Streams in regards to literary or scientific warrant. Consulting the majority of the user entries, the emphasis seems to be on Clustering rather than Data Streams, but the Control term is ambiguous in that respect and therefore not as strong a tag as it could be. To supplement with more user-centered terminology, using the broader or root form of this term (Clustering) without the clarifying aspect (Data Streams), or including Grouping as an alternative label, could be considered.

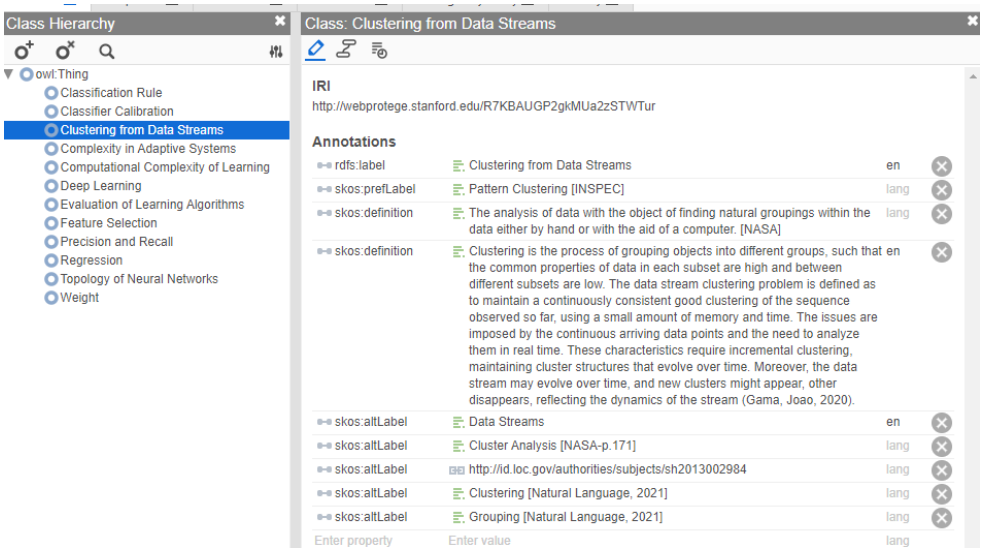


Figure 24: Equivalents for Clustering from Data Streams.

Complexity in Adaptive Systems

RQ1: 3 out of 4 control vocabulary terms matched Control, two of which were similar to one another. Adaptive System was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia

terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 4 out of 4 control vocabulary terms were matched when compared to each other. Adaptive and Control were the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 48% of users' natural language matched the Control, leading to the only accepted null hypothesis of the study. Out of the matches, entries containing “Complexity” were most prevalent (92), indicating a preferred label emphasizing the complexity aspect of this term is more user-centered than the adaptive systems aspect (21 matches). Complexity is a potentially ambiguous term, however, complexity in computer science is a well-researched and documented term which may be why it matched so highly. In addition, the highest volume user entries (regardless of a match to Control) were Adaptive Systems, Internal Complexity, and External Complexity. Complexity is heavily used in the prompt so it is no surprise users’ would select that as a term they would use in search but it does highlight that more specific types of computational complexity might be helpful to the end-user because complexity on its own is too ambiguous in multidisciplinary search, as can be seen by the users adding Internal and External to their selections. The accepted null hypothesis indicates the terminology was likely derived from user warrant. To supplement with additional user-centered terminology and potentially avoid ambiguous information retrieval in a multidisciplinary search, alternative labels containing more specific complexity types such as internal and external could be considered.

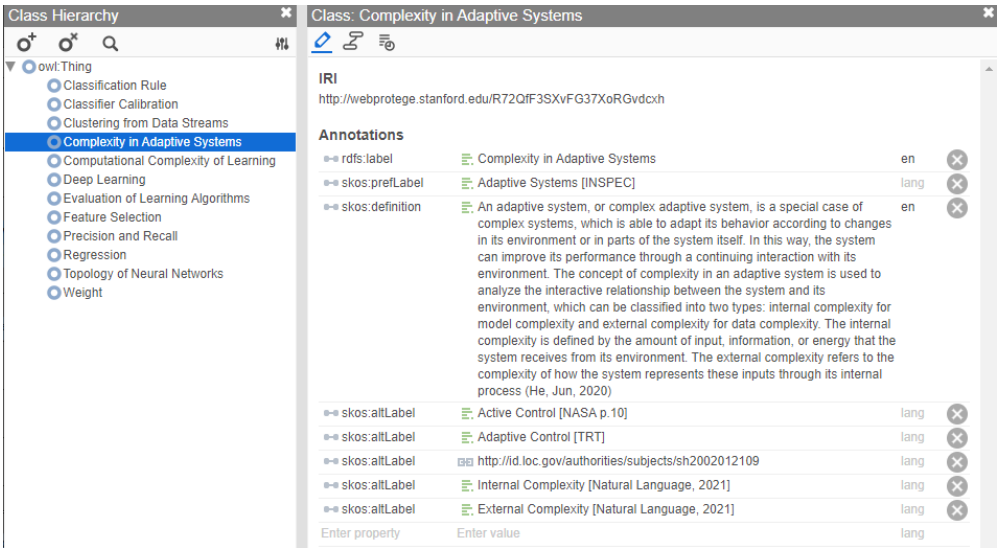


Figure 25: Equivalents for Complexity of Adaptive Systems.

Computational Complexity of Learning

RQ1: 2 out of 4 control vocabulary terms matched Control, two of which were similar to one another. TRT and NASA did not cover this topic. Computational Complexity was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 2 out of 4 control vocabulary terms were matched when compared to each other. Even though only 2 out of the 4 vocabularies covered this term, the vocabularies that did cover matched exactly when compared to each other. This is only one of three terms that matched exactly across all vocabularies. Where the other two examples had all four vocabulary terms match (Regression Analysis and Topology), indicating a strong retrieval of indexed content my result, the Computational Complexity term is not as strong, merely because it is not as prevalent in the consensus materials (i.e. engineering controlled vocabularies) and therefore would not connect

indexed content to the degree the other two vocabulary terms would. This matching indicates the terminology was likely derived from scientific warrant.

RQ3: 11% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing “complexity” were most prevalent (19), indicating a preferred label emphasizing the complexity aspect of this term might be more user-centered. This is very similar to the interpretations from the Control Complexity in Adaptive Systems where both Control terms are primarily about complexity with a clarifying aspect to the Control term, Adaptive Systems and Learning respectively. This is also the second to lowest user entry match, likely because “of Learning” is very ambiguous and does not convey machine learning with many users selecting complexity theory as their preferred term match. In addition, the highest volume user entries (regardless of a match to Control) were Learning and Complexity, followed by Oracle, PAC, and query-based (these last three were all at 24 count). This last is interesting that the distinct keywords were focused on by the users, and not the term Class which is very prevalent in the prompt. To a user not as familiar with machine learning terminology, the word Class may seem to be an ambiguous term and therefore ignored by the users in this example. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, using the broader or root form of this term without the clarifying aspect (Complexity or Complexity Theory), or refining the Learning aspect of this term which seemed to be the most confusing part for users, could be considered. Because Complexity in Adaptive Systems had similar user terminology, a seeAlso note could also be considered.

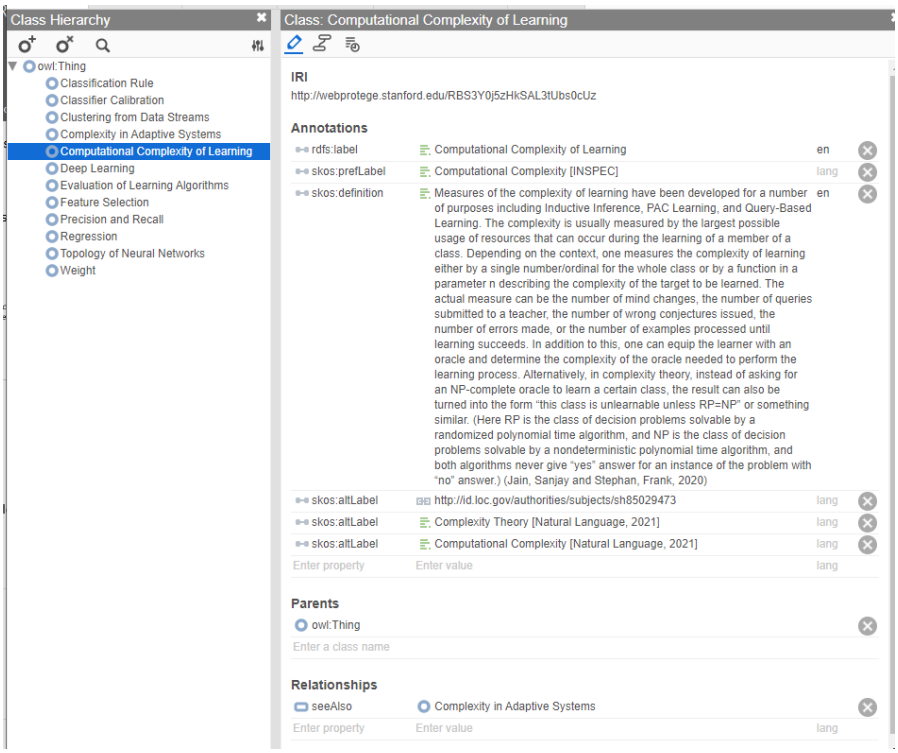


Figure 26: Equivalents for Comp. Complexity of Learning.

Deep Learning

RQ1: 0 out of 4 control vocabulary terms matched Control. The TRT did not cover this topic. Neural network was the more common among vocabularies, but it is not a close enough match to the expert Encyclopedia terminology for even the fuzzy search pattern to pick up without additional query expansion resources. This indicates the terminology was not likely derived from literary warrant.

RQ2: 3 out of 4 control vocabulary terms were matched when compared to each other. The TRT did not cover this topic. Neural net was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 12% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries with exact matches to Deep Learning were most prevalent (23). In addition, the highest volume user entries (regardless of a match to Control) were Pattern Recognition and NN. Strangely, NN is not defined in the prompt and most did not identify the acronym (which stands for neural network). This detail, and the association of a synonym within the prompt, may have caused users to miss this aspect of the term, and focus more on pattern recognition. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, changing the preferred label to be Neural Networks, adding this and NN as alternative labels, and pattern recognition as a seeAlso term could be considered.

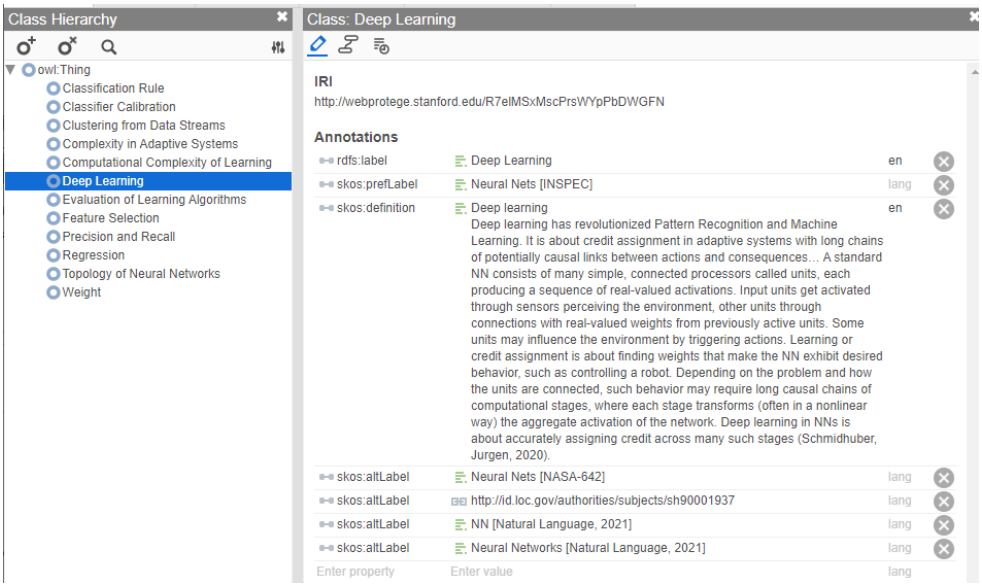


Figure 27: Equivalents for Deep Learning.

Evaluation of Learning Algorithm

RQ1: 2 out of 4 control vocabulary terms matched Control, two of which were similar to one another. The TRT and NASA did not cover this topic. Computational Complexity was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 2 out of 4 control vocabulary terms were matched when compared to each other. The TRT and NASA did not cover this topic. Computational Complexity was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 32% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing “learning algorithm” were most prevalent (62), indicating a preferred label emphasizing the learning algorithm aspect of this term might be more user-centered. In addition, the highest volume user entries (regardless of a match to Control) were Learning Algorithm, Properties, and Suitability. Here again, a qualifier (evaluation) seems to have been missing from users’ entries but they did pick up on Learning Algorithm which would explain the higher percentage matched, albeit not significant enough to accept the null hypothesis. Interestingly, “learning” was often mentioned in user entries without the inclusion of the “machine” aspect of the common entry users gave throughout the study. This is most likely because the phrase learning algorithm, without noting machine, is mentioned at least three times in the prompt. The rejected null hypothesis indicates the terminology was not likely derived from user

warrant. To supplement with more user-centered terminology, removing the qualifier, or adding alternative labels containing Learning Algorithm, Properties, and Suitability could be considered.

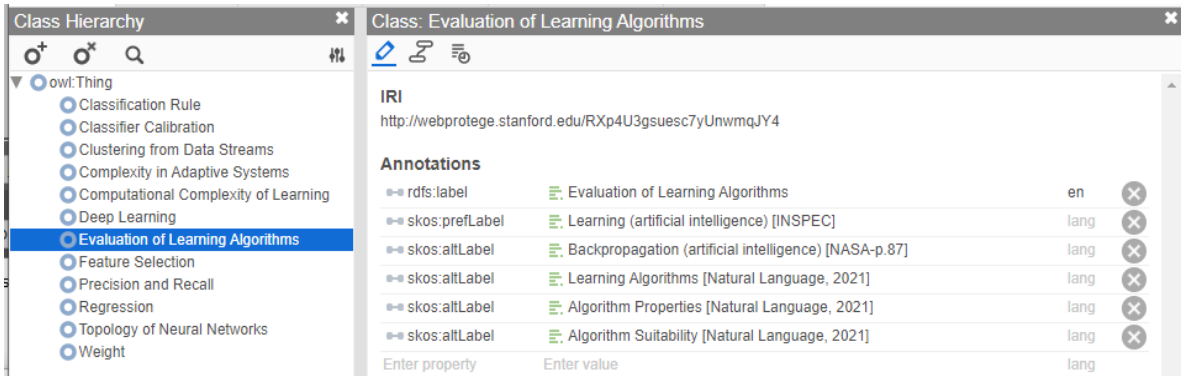


Figure 28: Equivalents for Eval. of Learning Algorithms.

Feature Selection

RQ1: 2 out of 4 control vocabulary terms matched Control, two of which were similar to one another. The LCSH and TRT did not cover this topic. Feature was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 2 out of 4 control vocabulary terms were matched when compared to each other. The LCSH and TRT did not cover this topic. Feature was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled vocabularies, the majority matched. This indicates the terminology was likely derived from scientific warrant.

RQ3: 15% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing the exact match Feature Selection or “feature”

on its own were most prevalent (12 and 17 respectively), indicating a preferred label emphasizing the feature aspect of this term might be more user-centered. In addition, the highest volume user entries (regardless of a match to Control) were Dimensionality Reduction Technique and forms of Feature. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, alternative labels containing Dimensionality Reduction Technique and forms of Feature could be considered.

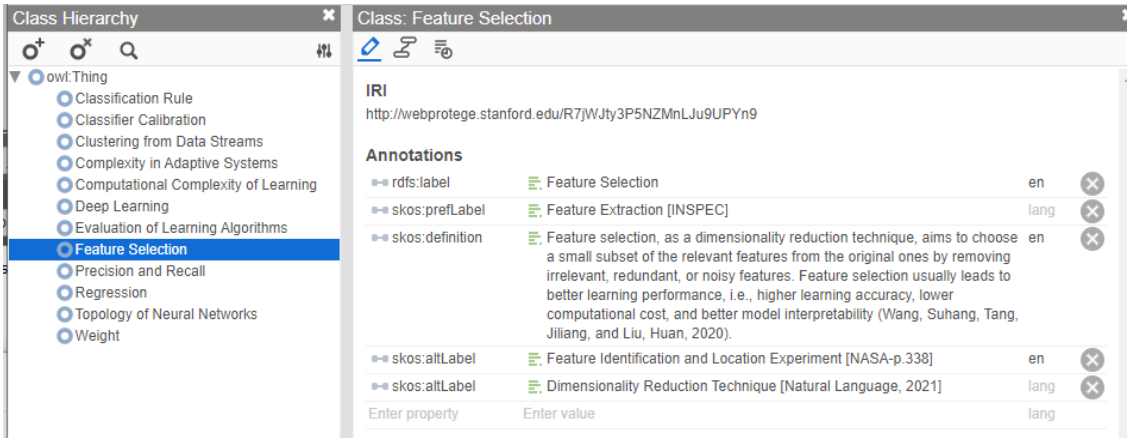


Figure 29: Equivalents for Feature Selection.

Precision and Recall

RQ1 and RQ2: 0 out of 4 control vocabulary terms matched Control, and none of the controlled vocabularies had this terminology. This is a very well-known methodology in machine learning so it was surprising to find this missing in all the controlled vocabularies. Searching both INSPEC and Google Scholar, there are more than 170,000 results in both using “precision and recall” so the terminology is used so the researcher used the controlled vocabularies to find an alternative form (starting with common relations like f-score and confusion matrix and broader terms like search engine) for this method but nothing was found. The only somewhat related term

was found in INSPEC (information retrieval system evaluation) but this is too broad to be a match for precision and recall. Comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, and comparing matches between vocabularies, was therefore not possible which indicates the terminology was not likely derived from literary or scientific warrant.

RQ3: 5% of users' natural language matched the Control (the lowest of all Control terms), leading to a rejection of the null hypothesis. Out of the matches, entries containing “precision” were most prevalent (5), but this is still very low. In addition, the highest volume user entries (regardless of a match to Control) were Information Retrieval and Confusion Matrix. Wondering if these existed in controlled vocabularies, only Information Retrieval was found but it did occur exactly the same in all 4 controlled vocabularies. However, information retrieval is not synonymous with precision and recall, nor is a confusion matrix but they are highly related. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, adding precision and recall or confusion matrix as narrower terms to information retrieval, or as alternative labels, could be considered.

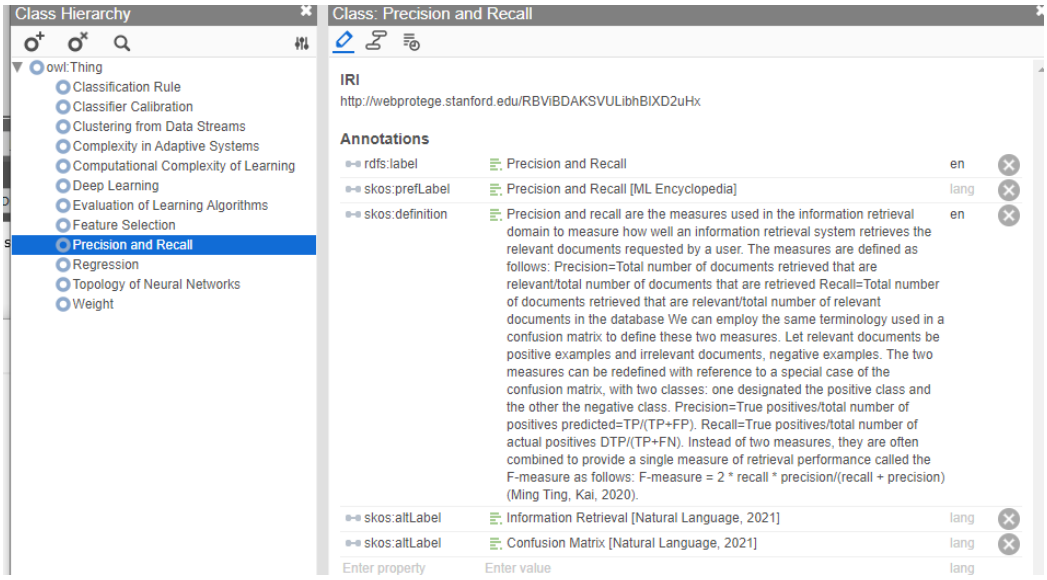


Figure 30: Equivalents for Precision and Recall.

Regression

RQ1: 4 out of 4 control vocabulary terms matched Control, all of which were the same, Regression Analysis. This is a close enough match to Control that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. This indicates the terminology was likely derived from literary warrant.

RQ2: 4 out of 4 control vocabulary terms were matched exactly when compared to each other. This is only one of three terms that matched exactly across all vocabularies. This would indicate a strong emphasis on this terminology in the scientific domain, specifically in controlled vocabularies. Another implication is that if a user is familiar with this term, they will potentially have a less frustrating search experience and have a lower chance of missing content in a multi-database search because all of the vocabularies match. This indicates the terminology was likely derived from scientific warrant.

RQ3: In RQ3, 33% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, exact matches for Regression were most prevalent (58). Because this is a single term, matches with anything containing regression would be considered a match which may account for the higher match percentage. In addition, the highest volume user entries (regardless of a match to Control) were Parametric and Regression Function. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, alternative labels containing Parametric or Regression Function could be considered.

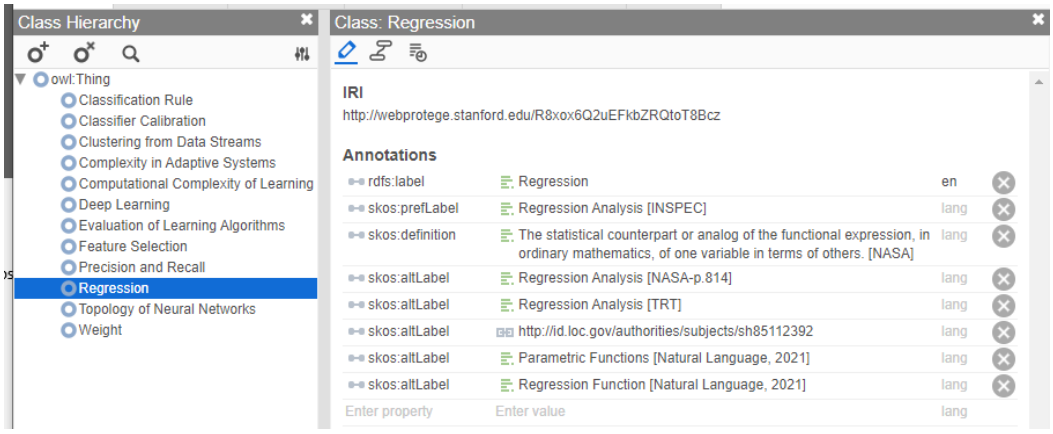


Figure 31: Equivalents for Regression.

Topology of Neural Networks

RQ1: 4 out of 4 control vocabulary terms matched Control, all of which were the same, Topology. This is a close enough match to Control that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. Because Topology is a broad term, when searching the literature, it is likely

best to couple this term with additional contextual terminology. This indicates the terminology was likely derived from literary warrant.

RQ2: 4 out of 4 control vocabulary terms were matched exactly when compared to each other. This is only one of three terms that matched exactly across all vocabularies. This would indicate a strong emphasis on this terminology in the scientific domain, specifically in controlled vocabularies. Another implication is that if a user is familiar with this term, they will potentially have a less frustrating search experience and have a lower chance of missing content in a multi-database search because all of the vocabularies match. That said, Topology is a broad term, and knowing Neural Networks is also available in these vocabularies, in the scientific warrant application of this controlled vocabulary term, it is suggested that it be coupled with Neural Networks when tagging content. This indicates the terminology was likely derived from scientific warrant.

RQ3: 20% of users' natural language matched the Control, leading to a rejection of the null hypothesis. This is likely because this is a combination term. Out of the matches, entries containing “topology” were most prevalent (38). In addition, the highest volume user entries (regardless of a match to Control) were Topology, Neurons, and Networks. This adds further evidence that the user may benefit from Topology and Neural Network terminologies being tagged to content because they did select both topology and neural network terminologies, just not as a combination term. The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, splitting the combination term into two distinct terms (that have been found in this study to exist in the controlled vocabularies already), or adding alternative labels containing neural network or topology network could be considered. In addition, Deep Learning, previously examined in this study, was found to have NN

(neural networks) as a potential alternative form and therefore can also be considered as an alternate or related form to Topology of Neural Networks, to align closer to users' natural language.

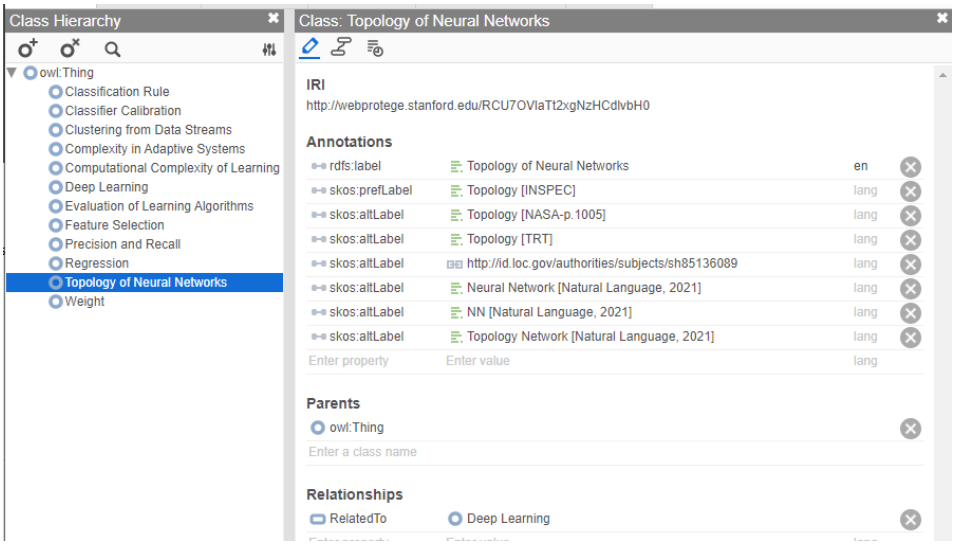


Figure 32: Equivalents for Topology of Neural Networks.

Weight

RQ1: 2 out of 4 control vocabulary terms matched Control, two of which were similar to one another. The LCSH and INSPEC did not cover this topic. Variations on the word “weight” are more common among vocabularies and are a close enough match that in the fuzzy search pattern comparing matches between the expert Encyclopedia terminology and the controlled vocabularies, the majority matched. However, this is a highly ambiguous term so more context would be needed for robust indexing of this term. This indicates the terminology was likely derived from literary warrant.

RQ2: 2 out of 4 control vocabulary terms were matched when compared to each other. The LCSH and INSPEC did not cover this topic. Weight was the more common among vocabularies and is a close enough match that in the fuzzy search pattern comparing matches between controlled

vocabularies, the majority matched. In a multidisciplinary search, this term would likely be too ambiguous and would need more context, such as scope notes or broader/narrower terminology. This indicates the terminology was likely derived from scientific warrant.

RQ3: 21% of users' natural language matched the Control, leading to a rejection of the null hypothesis. Out of the matches, entries containing “weight” were most prevalent (28), but variations on the term weight, such as weighting and weighted were also common and would add more context than the term weight on its own. In addition, the highest volume user entries (regardless of a match to Control) were Connection, Neurons, and Networks (very similar to topology or neural network user terms). The rejected null hypothesis indicates the terminology was not likely derived from user warrant. To supplement with more user-centered terminology, alternative labels or adding to the pref label terminology containing Weighted Connection, Neurons, and Networks could be considered. In addition, Deep Learning and Topology of Neural Networks had a lot of similarities between the controlled vocabulary and user vocabulary entries. This indicates a strong relationship between them, which could be represented as a see also relationship.

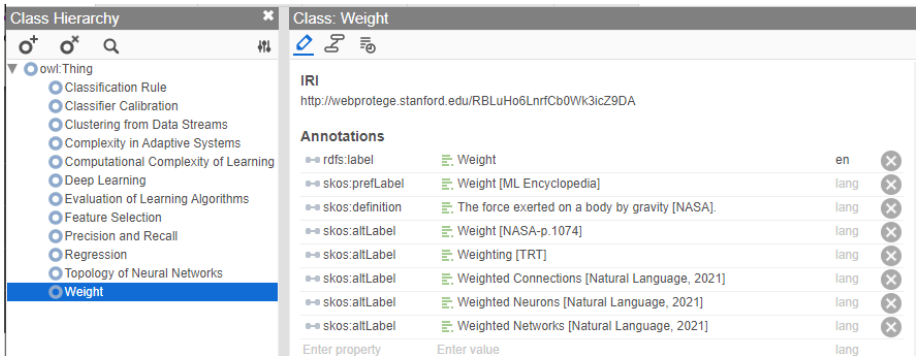


Figure 33: Equivalents for Weights.

5.2.4 Additional Suggestions for More Precise Subject Search

Case 2: Term Qualifiers

The findings from this study show that qualifying terms decreases the precision of mapping the users' terminology to controlled vocabularies. For the user, qualifiers may not show up in the full text that can lead to low recall in search. The suggestion here would be to treat terms as post-coordination and have the term split into its distinct parts (for instance Clustering and Data Streams as separated tags) to have a better chance of matching the users' query, as well as giving the most opportunity to the indexer to "mix and match" terms to tag a variety of concepts without making a multitude of stand-alone tags. Generally speaking, term qualifiers are valuable, and being specific in research is a good thing, eliminating noise during the research process. However, coupling the method with a contextual topic in the same term makes the likelihood of a user using those exact words highly unlikely (this study found $p = < .001$), even with fuzzy search being used. For example in the case of the term Classifier Calibration (40% match rate in RQ3 and the term had no match in RQ1-2), users did not seem to note the qualifier for the term, implying the term would not be as effective for those users in search. Indeed, what does calibration mean in this context, how is it done, what kind of calibration is happening? In this case, users preferred the root of the term Classifier over the qualified term. This last is important from a recall perspective where more research can be retrieved by using word/tag co-occurrence on separated tags than would otherwise be available with a combo term that is more rigid in context.

Case 3: Similarity of Terms

Another common issue in term alignment was terminology being too similar to other Control terms. In this situation, users are likely to retrieve the incorrect context for their search. The suggestion here would be to add these terms as Use For terms under one preferred term if they

are very similar like Neural Network and Deep Learning, or making these narrower terms to a broader term that encompasses the overarching topic like Computational Complexity. This serves the user twofold by helping the user determine if their query was interpreted by the search engine correctly by associating the retrieved search tags to their expected terminology, as well as having a higher chance of precision by eliminating the number of tags the search engine needs to interpret for the same concept. If there are too many tags that mean almost the same thing, this will decrease the effectiveness of precision because different tags will be used for the same concept. This was tested for the specific terms above by searching for both Control terms in Google Scholar and INSPEC, both of which returned different results but for nearly the same context. This was seen the most in Deep Learning and Neural Networks, Computational Complexity terms Computational Complexity of Learning and Complexity in Adaptive Systems, and Classification terms Classification Rule and Classifier Calibration.

Case 4: Ambiguous Terms

Ambiguous terminology such as Weight or Regression can have a multitude of meanings if taken out of the context of the user's specific research. The suggestion here would be to add a more specific preferred label to add more context for better precision. An example of this using the term Weighted would be updating to Weighted Connections or Weighted Networks (coupling the controlled term with the common user entries), or to be combined with more specific tags to be used for word/tag co-occurrence in search. The users' entries for these were quite varied because ambiguous terms do not have a clear context for the user and therefore make the users' search experience labored and difficult as they struggle to discover the "right" word to use. These can decrease the effectiveness of precision because there are too many contexts the term can be associated with. For instance, the term "Weight" without any context retrieved over a million

search results in INSPEC and the results had many different contexts, decreasing the effectiveness of this tag. Searching on “Weighted Networks” retrieved under 200,000 results, making this a much more precise and effectives subject search. In this scenario, the user searching may be missing content from their search results if the term is not updated. Examples of ambiguous terms that could benefit from additional context are Deep Learning (12% match rate), Computational Complexity of Learning (11% match rate), and Precision and Recall (5% match rate) being the lowest.

5.2.5 Additional Observations from Findings

In addition to the issues discussed thus far, in many cases the vocabularies did not cover a topic at all. This was unexpected because the terms selected are some of the most researched topics in machine learning, and, without any subject coverage, a search on one of these topics gets no subject precision assistance. The suggestion here would be to measure the rate of change for subjects, through mechanisms like Google Trends or search logs, to update terms that are influx while the terms that have “settled” into their prefered label do not need as much attention. A few example terms that had this issue were Deep Learning, Precision and Recall, and Classification. A survey of historical notes from the vocabularies that contained this metadata (LCSH and INSPEC) indicated the most recent update to any of the Control terms was 2016 (*neural networks* in LCSH), 4 years before the terminology in the Encyclopedia was published. This indicates the controlled vocabulary labels, definitions, and other contextual metadata can be out of date. This shows that some terms may not be aligned with any of the warrant types if they do not match literature, other vocabularies, or users. In this case, there is little search benefit for the user. This is an unfortunate consequence of the vast amount of terms in many vocabularies (hundreds of thousands of terms in

some cases like the LCSH). The sheer amount of terms makes it difficult to assess, update, maintain, and correct all existing controlled vocabulary effectively, a sentiment found in much of the literature reviewed for this work.

5.3 Options for User' Natural Language in Controlled Vocabularies

The findings indicate that adding users' natural language as a supplement to controlled library terms may increase the effectiveness of users' search, and with these additional access points, create a stronger bridge between the users' query and content. While the user's natural language can certainly be added as reference notations in controlled vocabularies via Use For or See Also (653 or 69X MARC fields) that may deter those like INSPEC, which is not a library, from adopting this mapping method. As mentioned in previous chapters, using a hub model for mapping, codified as an ontology or knowledge graph, helps make term mapping more effective as well (limiting the amount of mappings one has to do). While this study is not focused on testing how users' natural language is currently used in digital library search engines, other studies, most notably Wise, et. al. (2020), and Reinanda, Meij, and Rijke (2020) indicate ontologies can be used as embedded knowledge graphs to facilitate query expansion for this purpose. More datasets are needed for these types of studies. As such, the suggestions outlined throughout this study have been added to Bioportal, a well-known area for open-source research ontological and terminological datasets, for scholarly reuse in query expansion, terminology mapping, and semantic search studies.

This study has found that controlled vocabulary terms were derived primarily from literature (scientific being a subset of literary derivatives), but this study also shows the

discrepancy between these and the users' natural language. This misalignment strongly suggests including users' natural language in controlled vocabularies via Use For or graph mapping, to add another layer of much-needed access points into content. One of the most encouraging aspects of the mapping explored in this research is it does not sacrifice the specificity and authoritative control offered by controlled vocabulary preferred labels. The benefits, impact, and further research plans are concluded in the next section.

6.0 Conclusion

Search, revise, repeat. These are the words that started this study, indicating that there is a problem with information retrieval focusing specifically on engineering content, often leading to user frustration because their query retrieved unintended or incorrect results. It is the goal of the library, particularly in the content tagging stage, to break down barriers to information by adding more access points (subjects) to content to enhance subject retrieval in a search. But in the grand scheme of things, some libraries may be missing the mark because of the long-standing professional emphasis on literary warrant as opposed to user warrant, thereby not being as user-centered as they thought they were. Through a review of the literature, common issues with subject vocabularies showed that outdated, ineffective, and often exclusionary terminology makes information retrieval difficult for the end-user. Experts also pointed out that specific domains that have unique terminology, such as engineering, often suffer from subject-user misalignment, and that capturing the user's terminology may lead to a better understanding of the users' warrant—the mental model used when making information retrieval decisions. The literature also points to adding natural language to subjects, especially if a mapping framework such as a knowledge graph may help supplement subjects and lead to more user-focused information retrieval. But how aligned are the subjects to the users' terminology? That was the main question (RQ3) this study sought to answer, as well as provide evidence about whether controlled vocabularies in the engineering-domain align with the natural language of its users for a more user-centered vocabulary and indexing experience.

6.1 Summary of Findings

This study proceeded through four stages of term analysis: 1.) pre-mapping stage to determine baseline matching percentage and select the target vocabulary for mapping; 2.) an analysis of RQ1 focused on determining if controlled vocabularies matched the literature (literary warrant); 3.) an analysis of RQ2 focused on determining if controlled vocabularies matched each other; 4.) and using the matching baseline established in RQ1-2 for RQ3, an analysis to determine how often controlled vocabulary matched users' terminology. The expectation was that all three term analyses would show a match between the target and source at least 50% of the time. This was to be expected because of the baseline engineering-domain terminology mapping (SAE compared to LCSH, INSPEC, TRT, and NASA), as well as the pervasive understanding that subject tags are meant to be a bridge between the user's query and the text of the desired content to be retrieved. The findings of all three assessments (RQ1-3) were that terminology was focused more on the terms used in literature (literary warrant) and other vocabularies (scientific warrant) than the users' terminology (user warrant). In addition, RQ4 was assessed throughout all three stages to find that subjects matched more often for fuzzy search than browse behaviors (aggregate match of 26% versus 4% match in the browse pattern), indicating that subject search does benefit from the fuzzy search assistance of modern search engines.

Overall, while controlled terms did match the literature (RQ1 found 52% match between control and literature terminology), and controlled vocabularies matched over half of the time (54%) with one another in the fuzzy search pattern, the final assessment of these experiments was surprising. From the subset of terms used in this study, the evidence led to a rejection of the null hypothesis. The findings indicate that users' natural language did not match controlled vocabularies, the aggregated findings showing only a 26% match rate between controlled

vocabularies and users' natural language compared to the expected 50% match rate. For almost all term examples (Complexity in Adaptive Systems being the only exception), the users' natural language did not match the subject terminology (aggregate match of 26% across all control terms versus the 50% match expected from the baseline matching). This is disappointing because, why do subject tags exist if not to assist the end-user in retrieving the information they seek? This points to a gap in how abstraction is applied in digital information retrieval systems. It also points to a disconnect from what the tagging is thought to do versus what it is actually doing. This research suggests a way forward, which is to start the user-centered terminology assessment to determine the scope of the potential updates, then to update the vocabulary with the findings. Updates could include those to the vocabulary itself through label updates, Use For updates, or if possible, use a knowledge graph mapping similar to the one from this research to add query expansion logic to a scholarly search engine.

One important thing to note is that this study does not seek to undermine the value of controlled vocabulary and expert indexing. On their own, search engines can only index the strings from content—if the full text and metadata can even be indexed by the search engine. In many cases, subjects help tag what the content is about if there is no uncontrolled text for the search engine to use, or if the text is overly verbose, interpretive (such as commonly found in magazine titles that are "punchy," like puns or other creative titles), or have an ambiguity to their meaning. In this case, the subject tag creates a more reliable access point for the search engine to use. But if subject tags are derived from only the literature, and that terminology is not known to the user or the user is more familiar with subject tags that do not match the database subjects, the tags fail to connect the users' query to the content. This study does not claim that the inclusion of natural language in subject vocabularies will stop the revisionary behavior in search, but it does show that

by including more user terminology, there may be more content access points that have a higher chance to match the users' query, thus making their information retrieval that much easier and inclusive.

Some key factors that lead to misalignment between controlled terms and users' natural language are the use of qualifiers and ambiguous terms (same word, different meaning) terminology. This was revealed when the Control vocabulary was too specific to match users' queries, leading to overly specific tags that reduced search recall. This was also the case when Control terms were too ambiguous to single out one context for the term—potentially retrieving the correct term from the wrong context and reducing precision. The qualifier scenario could be resolved through the post-coordination of terms (instead of having qualifiers, the subjects are tagged to form the search/tagging context), and the ambiguity scenario could be resolved with using a more specific tag to narrow the context for better precision. Another cause for misalignment was terminology being to similar to another, such as Classification terms Classification Rule and Classifier Calibration which leads to lower precision because the context is spread over too many similar terms. This could be resolved with label changes and synonym mapping, either through knowledge graph mappings or useFor notations in the controlled vocabulary. These changes could be made to control vocabulary terms for improved search, regardless of adding users' natural language.

6.1.1 Additional Suggestions of Improvement:

Updates to controlled vocabularies were not the only findings from this study. Additional areas of insight included the lack of methods for analyzing controlled vocabularies' alignment with

users' natural language as well as uncovering the misalignment between controlled vocabularies, at least in the engineering-domain space.

It was found that there are few studies documenting methods to analyze the effectiveness of subjects with users natural language queries. This study used a bidirectional card sort, a new form of card sort, to facilitate this study but there are a few areas for improving this method in future research. Using a bidirectional card sort with a specific user group, either at a specific library or institution or a research team, would make for even stronger connections between the users' terms and the controlled terms because it would be connected to specific users', not just a representative group like ASEE in this study. For instance, if the natural language of the users of the University of Pittsburgh's Bevier Engineering Library could be compared to the LCSH terms assigned to that library's catalog, the resulting user-centered findings would be more beneficial to the Bevier patrons because it would benefit that specific user group in their daily search. Further testing will likely focus on specific users and specific databases (and therefore the vocabularies associated with those databases). Another improvement would be to expand the number of terms assessed to increase the power of the findings, either using branches of a specific vocabulary or the most popular searches for a specific institution.

Another unfortunate finding is that controlled vocabularies often do not match one another in the browse pattern. While the browse pattern was not completely in alignment for any of the assessments (RQ1-3), it is especially disappointing that controlled vocabularies did not match one another. It implies that a search on aggregated engineering-domain databases without the vocabularies being mapped together is likely not meeting the precision that could otherwise be expected. In this scenario, if the synonymous terminologies from the engineering-domain were mapped, the result would likely be better search precision. In comparison, literary and user warrant

both depend on the unstructured text, and therefore the exact match rate is expectedly lower than structured text such as controlled vocabularies (scientific warrant). In a multi-database search, controlled terms across different vocabularies compound the users' information retrieval work required to discover content. Without the synonymous terms across vocabularies being mapped, the user must not only investigate (or often guess) what subjects make sense for their research needs, but they also need to identify the subjects to search from each database. This scenario highlights why different controlled vocabularies across different databases need a level of agreement to assist both the user and the search engine in retrieving the most relevant results for the user. This points to a need for more alignment assessment between vocabularies. Not all databases will have the same content, so 100% alignment is not expected, but for the terms that do have overlap, it would be best for the user if they did not have to guess which synonymous term to use in which database search.

6.2 Positioning the Findings

There is tension between Google and librarians from a user-centered perspective. Google receives billions of queries a day about anything from the best cars of the year to self-determining health concerns—and the majority of its users have access to a library. And yet, libraries are not used as often as Google and do not receive the volume or breadth of questions that Google does. The fact of the matter is that users prefer Google (Hjorland, 2012) for many reasons, including the fact that it feels more accessible to them, even if Google cannot answer scholarly research questions nearly as well as librarians. (Vaccarrii, 2011, found that librarians answered user questions 30% more accurately than Google.) But more and more, Google-like behavior is

expected of the digital library, specifically answering questions and bridging literature terminology with the users' natural language. Mayr, Frommholz, Cabanac, et. al (2018) and Hu, Duan, and Dang (2020) propose an approach similar to the one used in this study, that of an ontological knowledge graph, to help expand users' queries into synonyms and questions within the context of digital information retrieval. This study adds to that proposal by also including an expansion on users' natural language as a supplement to the library-controlled vocabulary.

While libraries should not strive to be "Google-like," since Google is a general search platform and the library (at least in higher education) is a dedicated scholarly research space, they have always focused on breaking down the barriers to entry. The library strives to be a bedrock for learning and innovative scholarship, but this mission seems somewhat misaligned with the expectation for users to "know the right word" to find the most comprehensive and relevant list of results. This is a much larger question, and one that many others, such as Olson, Berman, and Hjorland, have explored; and still, there is no definitive answer. This study is positioned as another line of inquiry to pursue, opening a new method to assess the user warrant in libraries, and even suggesting ways to better align vocabularies with the users' natural search expectations ingrained from Google, while not losing the specificity and dependency of controlled subject tags.

This study has uncovered a few areas of misalignment that are impacting the effectiveness and satisfaction of search in the engineering-domain. However, the areas of improvement are not insurmountable obstacles, and the tools and experience exist to remedy the issues identified. If those that maintain and contribute to controlled vocabulary would begin to include these updates--especially the vocabularies that are open, such as LCSH--this would serve as a fulcrum for change. Open vocabularies are used by millions of institutions around the world, so if the changes are propagated through open vocabularies, these would spread more user-centered vocabulary terms

to all institutions using those vocabularies. Even on a micro-level, by understanding how to analyze and interpret the match rate between users' natural language and controlled vocabularies, the researcher hopes that more institutional controlled vocabularies will include the users' natural language to improve the accessibility and information retrieval of content, thereby creating a better bridge between the user and the content.

Lexicographers have their own term when creating a new term, called "settling." But language is messy and is constantly changing, so a word rarely "settles" and rather continues to grow into the context in which it is used. New words are created, old words become outdated and potentially offensive, and users' behaviors and expectations evolve. This study shows that we can honor the traditional library organization and the controlled terminology that has been a foundation of library studies since Dewey's first published catalog in 1876. At the same time, we can look to the needs and expectations of the 21st-century user, harnessing the abilities that a digital ecosystem offers and making research more accessible to the end-user.

6.3 Further Research

While the findings of this study point to trends in subject indexing terminology, as well as suggested updates to the process, the nature of terminology is such that, for each controlled term, a vocabulary would need its own user-centered analysis to determine if the findings of this study hold true for the entirety of a vocabulary. This study assessed 12 terms from a specific domain, and the null hypothesis was tested on natural language gathered from 53 participants. The participant volume meets the stated 15-20 participant threshold outlined by card sorting methods; however, the entirety of a vocabulary should ideally be tested using the total term count as the

population to derive the sample size. The survey increases in time commitment the more terms that are added for testing, so a study on a full vocabulary would likely benefit from an iterative approach. One such method would be using the highest level branches of the vocabulary to group like-context terms for testing in each iteration. Even so, the findings of this study focus on some of the most popular machine learning methods of research, with hundreds of thousands of research articles tagged with each term. So while the entirety of a vocabulary or the state of subject indexing in general cannot be concluded based on this one study, the impact of the findings can potentially help make the discovery of machine learning research of these 12 methodologies more effective for the engineering-domain user. Furthermore, this study shows that the method used here is a viable way to assess the alignment of controlled vocabularies and supplement controlled terms with the users' natural language.

One somewhat alarming discovery, though not directly tied to the scope of this study, was the lack of alignment between the most commonly used engineering-domain controlled vocabularies (27% exact match between LCSH, NASA, INSPEC, TRT). Some misalignment was not surprising, but the extent discovered in this study indicates there is a serious precision problem on multi-database or multi-vocabulary search in the engineering-domain. This finding shows there is a need to proactively align the vocabularies in this domain in a way that does not place the burden on the shrinking resources the owners of these vocabularies are likely facing. In many cases, undergoing a vocabulary alignment project such as this will take time, effort, and capital that many of these government entities (TRT, LCSH, NASA) may not have. Or, the vocabulary is proprietary (INSPEC) and the owners may not feel the need to align with open vocabularies. Unfortunately, indexing teams are shrinking, not growing. Even the Library of Congress now contracts with third parties to maintain the LCSH because its team has shrunk so much. This study

analyzed only a small subset of terms, so perhaps the controlled vocabulary-to-controlled vocabulary misalignment issue found in this study is not pervasive. If it is, however, this finding may point to an underlying controlled vocabulary crisis no one seems to be talking about. Further analysis on the underlying issues and how the library community can address them is needed and is likely an additional path of study.

This study is therefore just the first of many. Future research opportunities include expanding the terminology analyzed as well as delving into more terminology domains. The additional research will also serve to refine the methods used in this study for gathering and comparing user terminology to controlled vocabularies.

Appendix A

All mapping dataset for Baseline, RQ1, RQ2, and RQ3 Datasets located at: https://docs.google.com/spreadsheets/d/1we9PNWifNozObrw-VNxEWONNp6P45SIK/edit?usp=sharing&ouid=105068306933040109159&rtpof=true&sd=true

Bibliography

"Criteria for Indexes." NISO Z39.4-2021. NISO, 2021.

"IEEE Membership FAQ." (2018). Accessed November 2, 2018 at https://www.ieee.org/about/today/at-a-glance.html#ieee-quick-facts

"Literary Warrant (IEKO)." (2021). International Society for Knowledge Organization. https://www.isko.org/cyclo/literary_warrant

Adler, M. (2009). Transcending library catalogs: A comparative study of controlled terms in Library of Congress Subject Headings and user-generated tags in LibraryThing for transgender books. *Journal of Web Librarianship*, *3*(4), 309-331.

Adler, Melissa. "The case for taxonomic reparations." *KO KNOWLEDGE ORGANIZATION* 43.8 (2016): 630-640.

Adolf, M. & Stehr, N. (2014). *Knowledge*. Abingdon, Oxon: Routledge.

Agapova, O., Buchel, O., Freeston, M., Frew, J., Smith, T., Zeng, M, ... & Ushakov, A. (2002, July). Structured models of scientific concepts for organizing, accessing, and using learning materials. In *International Conference on Digital Libraries: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries* (Vol. 14, No. 18, pp. 407-407).

Agarwal, S. & Sureka, A. (2016). Spider and the flies: focused crawling on Tumblr to detect hate promoting communities.

Ajiferuke, I. (2003). Role of information professionals in knowledge management programs: empirical evidence from Canada. *Informing Science, 6*, 247-257.

Alfonso-Goldfarb, A., Ferraz, M. & Waisse, S., & Ferraz, M. From shelves to cyberspace: organization of knowledge and the complex identity of history of science. *Isis, 104*(3), 551-560.

Allen, R., Fink, E., Goodlander, G., Knoblock, C., Szekely, P., Yang, F., Zhu, X. (2013). Connecting the Smithsonian American Art Museum to the linked data cloud. *The Semantic Web: Semantics and Big Data: 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013.* Proceedings, 8(2014), 593–607.

Almeida, C. & Pando, D. (2016). Knowledge organization in the context of postmodernity from the theory of classification perspective. *Knowledge Organization*, *43*(2).

Ananiadou, S., Chen, L., Fan, Y., Qian, Y., Wei, J., Xu, Y ... & Tsujii, J. (2015). Bilingual term alignment from comparable corpora in English discharge summary and Chinese discharge summary. *BMC Bioinformatics*, *16*(1), 149.

Anderson, J., & Hofmann, M. (2006). A fully faceted syntax for Library of Congress subject headings. *Cataloging & Classification Quarterly*, 43(1), 7–38.

Anfinnsen, S., Cesare, S. & Ghinea, G. (2011). Web 2.0 and folksonomies in a library context. *International Journal of Information Management*, *31*(1), 63-70.

Angele, J., Erdmann, M., & Wenke, D. (2008). Ontology-based knowledge management in automotive engineering scenarios. In *Ontology Management* (pp. 245-264). Springer United States.

ANSI/NISO. (2005). ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Content and Collection Management Topic Committee.

Antelman, K., Lynema, E., & Pace, A. (2006). Toward a twenty-first century library catalog. *Information Technology and Libraries*, *25*(3), 128.

ARL Joint Task Force on Library Support for E-Science. (November, 2007). *Agenda for Developing E-Science in Research Libraries." Final Report and Recommendations to the Scholarly Communications Steering Committee, the Public Polices Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee*. Report.

Asher, A., Duke, L., & Wilson, S. (2013). Paths of discovery: comparing the search effectiveness of EBSCO discovery service, summon, Google scholar, and conventional library resources. *College & Research Libraries 74(5)*, 464–88.

Azad, Hiteshwar Kumar, and Akshay Deepak. "Query expansion techniques for information retrieval: a survey." *Information Processing & Management* 56.5 (2019): 1698-1735.

Azuara, F., González, J., & Ruggia, R. (2013). Proposal of a controlled vocabulary to solve semantic interoperability problems in social security information exchanges. *Library Hi Tech*, *31*(4), 602–619.

Baeza-Yates, R. & Saez-Trumper, D. (2016). Wisdom of the crowd or wisdom of a few?: An analysis of users' content generation. 69-74.

Baeza-Yates, R., & Saez-Trumper, D. (2015). Wisdom of the crowd or wisdom of a few? An analysis of users' content generation categories and subject descriptors. *Proceedings of the 26th ACM Conference on Hypertext & Social Media. ACM, 2015*, 69–74.

Balaji, B., Madalli, D., & Sarangi, A. (2015). Faceted ontological representation for a music domain. *Knowledge Organization*, *42*(1), 8-24.

Balatsoukas, P., & Demian, P. (2010). Effects of granularity of search results on the relevance judgment behavior of engineers: building systems for retrieval and understanding of context. *Journal of The American Society for Information Science and Technology*, *61*(2), 453–467.

Baldoni, M., Baroglio, C., Patti, V., & Rena, P. (2012). From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, *6*(1), 41-54.

Bales, S., Rieger, J., Wang, P., & Zhang, Y. (2004). Survey of learners' knowledge structures: Rationales, methods and instruments. *Proceedings of the ASIST Annual Meeting*, *41*, 218–228.

Banerji, I. & Blanchard, J. (2016). Evidence-based recommendations for designing free-sorting experiments. *Behavior research methods*, 1-19.

Barifah, Maram, Monica Landoni, and Ayman Eddakrouri. "Evaluating the user experience in a digital library." *Proceedings of the Association for Information Science and Technology* 57.1 (2020): e280.

Barité, Mario. 2014. "Control de Vocabulario: Orígenes, Evolución y Proyección". *Ciência da Informação* 43, no. 1: 95-119. ttp://revista.ibict.br/ciinf/issue/view/111/showToc

Batch, Y. & Yusof, M. (2015). Organizing information in medical blogs using a hybrid taxonomy-folksonomy approach. *Journal of Web Engineering*, *14*(3-4), 181-195.

Batet, M., Gibert, K., Sanchez, D., & Vallas, A. (2010). Using ontologies for structuring organizational knowledge in Home Care assistance. *International Journal of Medical Information*, 79(5), 370-387.

Baxter, K., Caine, K. & Courage, C. (2015). *Understanding your users: a practical guide to user research methods*. (Second 2nd ed.). Waltham, MA: Morgan Kaufmann.

Becerra-Fernandez, I. & Sabherwal, R. (2015). *Knowledge management: Systems and processes* (Second ed.). Armonk, New York: M.E. Sharpe, Inc.

Bechhofer, S., Goble, C., & Stevens, R. (2000). Ontology-based knowledge representation for bioinformatics. *Briefings in Bioinformatics, 1(4), 398-414.*

Beck, C. (2016). *Research partner and search methodology expert: the role of the librarian in systematic reviews* (Doctoral dissertation, University of British Columbia).

Bedford, D., Greenberg, J., Hlaya, M., Hodge, G., & White, H. (2010, October). *Knowledge Organization: Evaluating Foundation and Function in the Information Ecosystem*. Panel presented at ASIS&T '10 Annual Meeting, Pittsburgh, Pennsylvania

Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification

Bekhuis, T., Crowley, R., & Demner-Fushman, D. (2013). Comparative effectiveness research designs: an analysis of terms and coverage in Medical Subject Headings (MeSH) and Emtree. *Journal of Medical Library Association*, *101*, 92-100.

Bendersky, Michael, and W. Bruce Croft. "Modeling higher-order term dependencies in information retrieval using query hypergraphs." *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 2012.

Bergman, M. (2016). A knowledge graph for knowledge. *LinkedIn*. Accessed October 7, 2016, from https://www.linkedin.com/pulse/knowledge-graph-mike-bergman?articleId=6189464181712965632

Bona, J. & Ceusters, W. (2016). Ontological foundations for tracking data quality through the Internet of Things. *Studies in Health Technology and Informatics,* 221, 74-78.

Borner, K. & Shiffrin, R. (2004). Mapping knowledge domains. *PNAS, 101*(1), 5183–5185.

Bouadjenek, M., Bouzeghoub, M., & Hacid, H. (2016). Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. *Information Systems, 56*, 1-18.

Bouton, C., Hajagos, J., Jentzsch, A., Kallesoe, C. S., Samwald, M., Willighagen, E., ... & Stephens, S. (2011). Linked open drug data for pharmaceutical research and development. *Journal of Cheminformatics*, 3(1), 19.

Brachman, R., McGuinness, D., Patal-Schneider, P., & Resnick, L. (1991). Living with CLASSIC: When and how to use a KL-ONE-like language. In J. Sowa (Ed.), *Principles of Semantic Networks: Explorations in the Representation of Knowledge* (pp. 401–456). San Mateo: Morgan Kaufmann Publishers.

Brophy, P. & Griffiths, J. (2005). Student searching behavior and the web: Use of academic resources and Google. *Library Trends, 53(4),* 539–54.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Journal of Documentation*, *58*(1), 49–72.

Burnett, S. & Cleverley, P. (2015). Retrieving haystacks: a data driven information needs model for faceted search. *Journal of Information Science*, *41*(1), 97-113.

Cabitza, F., Colombo, G., & Simone, C. (2013). Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities. *International Journal of Human-Computer Studies*, *71*(1), 24-45.

Carter, L. & Whittaker, B. (2015). Area studies and special collections: shared challenges, shared strength. *Portal: Libraries and the Academy*, *15*(2), 353-373.

Castells, M. (1996). *The network society* (Vol. 469). Oxford: Blackwell.

Castine, M., Conway, M., Fana, F., Khojoyan, A., Mowery, D., Scuba, W., ... & Jupp, S. (2016). Developing a web-based SKOS editor. *Journal of Biomedical Semantics*, *7*(1), 1.

Chan, L. & Zeng, M. (2002). Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis. In *Proceedings of the 68th IFLA Council and General Conference* (August 18-24, 2002).

Chan, L. & Zeng, M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology, 55*(5), 377-395.

Chan, L. (2007). *Cataloging and Classification*. The Scarecrow Press.

Chan, L., & Zeng, M. (2006). Metadata interoperability and standardization-a study of methodology part I: Achieving interoperability at the schema level. *D-LIB Magazine*, *12*(6).

Chan, L., Lin, X., & Zeng, M. (1999). Structural and multilingual approaches to subject access on the web. In *Proceedings of the 65th IFLA Council and General Conference* (August 20-28, 1999).

Chan, Y., Huang, S., & Lin, S. (2012). Investigating effectiveness and user acceptance of semantic social tagging for knowledge sharing. *Information Processing & Management*, *48*(4), 599-617.

Chung, H. (2012). Unleashing the leviathan: Against common interpretations and a contemporary decision-theoretic reconstruction of Hobbes's Moral and political philosophy. (Doctoral dissertation, Cornell University).

Ciccarese, P., Ocana, M., Castro, L., Das, S., & Clark, T. (2011). An open annotation ontology for science on web 3.0. *Journal of Biomedical Semantics*, *2*(2), 1.

Cimino, J., Ellsworth, M., Herasevich, V., Homan, J., Peters, S., & Pickering, B. (2015). Point-of-care knowledge-based resource needs of clinicians. *Applied Clinical Informatics*, *6*(2), 305-317.

Cleverley, P. & Burnett, S. (2015). Retrieving haystacks: A data driven information needs model for faceted search. *Journal of Information Science*, *41*(1), 97-113.

Cleverley, P. (2016). Personality influences the way you search, but not necessarily your search performance. *LinkedIn*. Accessed April 4, 2016, from https://www.linkedin.com/pulse/personality-influences-way-you-search-necessarily-your-paul-cleverley.

Cooksey, D. & Soranzo, A. (2015). Testing taxonomies: Beyond card sorting. *Bulletin of the American Society for Information Science and Technology, 41*(5), 34-39.

Coppens, H., Schiltz, M., & Truyen, F. (2007). Cutting the trees of knowledge: Social software, information architecture and their epistemic consequences. *Thesis Eleven*, *89*(1), 94-114.

Corrall, S. (1999). Knowledge management: Are we in the knowledge management business. *Ariadne, 18*, 16-18.

Crabtree, B., McInerney, C., Orzano, A., Scharf, D., Tallia, A. (2008). A knowledge management model: Implications for enhancing quality in health care. *Journal of the American Society for Information Science and Technology*, *59*(3), 489-505.

Cui, H., Cole, H., Endara, L., Huang, F., & Macklin, J. (2015). OTO: Ontology Term Organizer. *BMC Bioinformatics*, *16*(1), 1.

Culley, S., & Dekoninck, E. & Howard, T. (2011). Reuse of ideas and concepts for creative stimuli in engineering design. *Journal of Engineering Design*, *22*(8), 565-581.

Da Silveira, M., Dinh, D., Dos Reis, Pruski, C., & Reynaud-Delaitre, C. (2015). Recognizing lexical and semantic change patterns in evolving life science ontologies to inform mapping adaptation. *Artificial Intelligence in Medicine*, *63*(3), 153-170.

Dame, A. (2016). Making a name for yourself: Tagging as transgender ontological practice on Tumblr. *Critical Studies in Media Communication*, *33*(1), 23-37.

Dawson, S. & Oborn, E. (2010). Knowledge and practice in multidisciplinary teams: Struggle, accommodation and privilege. *Human Relations*, *63*(12), 1835-1857.

de Andrade, J. & de Lara. (2016). Interoperability and mapping between knowledge organization systems: Metathesaurus — Unified Medical Language System of the National Library of Medicine. *Knowledge Organization*, *43*(2), 107–113.

Del Fiol, G., Islam, R., Jones, M., Samore, M. H., & Weir, C. R. (2015). Understanding complex clinical reasoning in infectious diseases for improving clinical decision support design. *BMC Medical Informatics and Decision Making*, *15*(1), 1.

Devezas, José, and Sérgio Nunes. "Hypergraph-of-entity." *Open Computer Science* 9.1 (2019): 103-127.

Dewey, J., Ratner, J., & Post, E. (1939). *Intelligence in the modern world: John Dewey's philosophy*. New York: Modern Library.

Dextre-Clarke, R. I., Karlova, N., Lee, J. H., Perti, A., & Thornton, K. (2014). Facet analysis of video game genres. *iConference 2014 Proceedings*.

Dextre-Clarke, S. & Zeng, M. (2012). From ISO 2788 to ISO 25964: The evolution of thesaurus standards towards interoperability and data modelling. *Information Standards Quarterly (ISQ)*, *24*(1).

Dimitrov, D. & Rumrill Jr., P. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159-165.

Dong, H., Liang, H., & Wang, W. (2015). Learning structured knowledge from social tagging data: A critical review of methods and techniques. *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, China, 2015*, 307-314.

Du, J. & Evans, N. (2011). Academic users' information searching on research topics: Characteristics of research tasks and search strategies. *The Journal of Academic Librarianship*, *37*(4), 299-306.

Du, W., Lau, R., Ma, J., & Xu, W. (2015). A multi-faceted method for science classification schemes (SCSs) mapping in networking scientific resources. *Scientometrics*, *105*(3), 2035-2056.

Dumontier, M., & Wild, D. (2012). Linked data in drug discovery. *IEEE Internet Computing*, (6), 68-71.

Durst, S., & Edvardsson, R. (2012). Knowledge management in SMEs: a literature review. *Journal of Knowledge Management*, *16*(6), 879-903.

El-Diraby, T. E. (2012). Domain ontology for construction knowledge. *Journal of Construction Engineering and Management*, *139*(7), 768-784.

Elsevier Hunter Forum. (January, 2016). ALA Mid-Winter Meeting, Boston, MA.

Engerer, V. (2016). Exploring interdisciplinary relationships between linguistics and information retrieval from the 1960s to today. *Journal of the Association for Information Science and Technology*.

Faith, Ashleigh, Eugene Tseytlin, and Tanja Bekhuis. "Development of the EDDA study design terminology to enhance retrieval of clinical and bibliographic records in dispersed repositories." *Proceedings of the 2014 International Conference on Dublin Core and Metadata Applications*. 2014.

Fisher, A., MacLellan, C., Nugent, R., Unger, L., & Ventura, S. (2016). Developmental changes in semantic knowledge organization. *Journal of Experimental Child Psychology*, *146*, 202-222.

Frické, Martin. "Reflections on classification: Thomas Reid and bibliographic description." *Journal of documentation* (2013).

Friedman, A., & Smiraglia, R. P. (2013a). Nodes and arcs: concept map, semiotics, and knowledge organization. *Journal of Documentation*, *69*(1), 27-48.

Gaines, B. (2013). Knowledge acquisition: Past, present and future. *International Journal of Human-Computer Studies*, *71*(2), 135-156.

Gaiser, B. & Panke, S. (2009). ``With my head up in the clouds" Using social tagging to organize knowledge. *Journal of Business and Technical Communication*, *23*(3), 318-349.

Garcia, A., Gonzalez, J., Vabderjastm E., & Vargas, G. (2015). The blended librarian and the disruptive technological innovation in the digital world.

Garcia-Barriocanal, E., Martín-Moncunill, D., Sánchez-Alonso, S. & Sicilia, M. (2015). Evaluating the practical applicability of thesaurus-based keyphrase extraction in the agricultural domain: Insights from the VOA3R project. *Knowledge Organization*, *42*(2).

Gardner, T. & Inger, S. (2016). How readers discover content in scholarly publications.

Garshol, L. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, *30*(4), 378-391.

Gascoigne, N., & Thornton, T. (2013). *Tacit knowledge*. Durham, UK: Acumen.

Gasson, S. (2015). Practice-based information and data management: A network approach. *iConference 2015 Proceedings*.

Gentner, D., Goldwater, M., & Rottman, B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science*, *36*(5), 919-932.

Georgas, H. (2014). Google vs. the library (part II): Student search patterns and behaviors when using Google and a federated search tool. P*ortal: Libraries and the Academy*, *14*(4), 503-532.

Gerritse, Emma J., Faegheh Hasibi, and Arjen P. de Vries. "Bias in conversational search: The double-edged sword of the personalized knowledge graph." *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 2020.

Giess, M., Goh, Y., & McMahon, C. (2009). Facilitating design learning through faceted classification of in-service information. *Advanced Engineering Informatics*, *23*(4), 497-511.

Giess, M., McMahon, C., & Wild, P. (2008). The generation of faceted classification schemes for use in the organization of engineering design documents. *International Journal of Information Management*, *28*(5), 379–390.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies-an etymological note. *Journal of Documentation*, *59*(1), 7-18.

Giunchiglia, F., & Zaihrayeu, I. (2007). *Lightweight ontologies,‖ DIT, University of Trento*. Tech. Rep. DIT-07-071. Accessed August 28, 2016, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.1717&rep=rep1&type=pdf

Giunchiglia, F., Marchese, M., & Zaihrayeu, I. (2007). Encoding classifications into lightweight ontologies. In *Journal on data semantics VIII* (pp. 57-81). Springer Berlin Heidelberg.

Green R. (2005). Vocabulary alignment via basic level concepts, http://www.oclc.org/research/grants/reports/green/rg2005.pdf.

Greenberg, J., Ogletree, A., Rowell, C. & Zhang, Y. (2015). Controlled vocabularies for scientific data: Users and desired functionalities. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, 54.

Greene, Judith, and Manuela D'Oliveira. *Learning to Use Statistical Skills in Psychology*, McGraw-Hill Education, 2005. *ProQuest Ebook Central*, https://ebookcentral.proquest.com/lib/pitt-ebooks/detail.action?docID=290384.

Gruber, T. (1992). Ontolingua: A mechanism to support portable ontologies. Technical Report KSL-91-66, Stanford University, Knowledge Systems Laboratory.

Hackett, P. (2016). Facet theory and the mapping sentence as hermeneutically consistent structured meta-ontology and structured meta-mereology. *Frontiers in Psychology*, 31(7), 471.

Hall, C. & Khoo, M. (2012, September). What would 'Google' do? Users' mental models of a digital library search engine. In *International Conference on Theory and Practice of Digital Libraries* (pp. 1-12). Springer Berlin Heidelberg.

Hariri, N. (2013). Do natural language search engines really understand what users want? A comparative study on three natural language search engines and Google. *Online Information Review*, *37*(2), 287-303.

Harter, S., Nisonger, T., & Weng, A. (1993). Semantic relationships between cited and citing articles in library and information science journals. *Journal of the American Society for information Science*, 44, 543-552.

Harvard University. " ." *Harvard Library ends use of subject heading 'illegal alien.'* (2021). https://news.harvard.edu/gazette/story/newsplus/harvard-library-ends-use-of-subject-heading-illegal-alien/

Hazeri, A., Martin, B., & Sarrafzadeh, M. (2006). LIS professionals and knowledge management: Some recent perspectives. *Library Management*, 27(9), 621–635.

He, S., Liu, X., Pang, M., Shi, J. & Zhu, L. (2015). Keywords co-occurrence mapping knowledge domain research base on the theory of Big Data in oil and gas industry. *Scientometrics*, *105(1)*, 250-260.

Hecker, A. (2012). Knowledge beyond the individual? Making sense of a notion of collective knowledge in organization theory. *Organization Studies*, *33*(3), 423-445.

Hedden, H. (2016). *The Accidental Taxonomist*, (2nd edition). Information Today.

Hemminger, B. & Niu, X. (2015). Analyzing the interaction patterns in a faceted search interface. *Journal of the Association for Information Science and Technology*, *66*(5), 1030-1047.

Hepp, M., van Damme, C., & Siorpaes, K. (2007). Folksontology: An integrated approach for turning folksonomies into ontologies. *Bridging the Gap between Semantic Web and Web*, *2*(2), 57-70.

Heuvel, C. & Smiraglia, R (2013b). Classifications and concepts: towards an elementary theory of knowledge interaction. *Journal of Documentation*, *69*(3), 360–383.

Heymann, P., & Garcia-Molina, H. (2009). Contrasting controlled vocabulary and tagging: Do experts choose the right names to label the wrong things?. Retrieved from http://ilpubs.stanford.edu:8090/955/1/cvuv-lbrp.pdf

Hider, P. (2015). A survey of the coverage and methodologies of schemas and vocabularies used to describe information resources, *Knowledge Organization, 42*(3), 154–164.

Hislop, D. (2013). *Knowledge management in organizations: A critical introduction*. Oxford University Press.

Hjorland, B. & Albrechtsen, H. (1999). An analysis of some trends in classification research. *Knowledge Organization,* 26 (3), 131-137.

Hjorland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25 (1/2), pp. 16-31.

Hjorland, B. (2008). Core classification theory: A reply to Szostak. *Journal of Documentation, 64*(3), 333-342.

Hjorland, B. (2012). Facet analysis: The logical approach to knowledge organization. *Information Processing & Management*, *49*(2), 545–557.

Hjorland, B. (2013). User-based and cognitive approaches to knowledge organization: A theoretical analysis of the research literature. *Knowledge Organization, 40*(1), 11-27.

Hjorland, B. (2015). Theories are knowledge organizing systems (KOS). *Knowledge Organization*, *42*(2), 113–128.

Hjorland, B. (2016). Does the traditional thesaurus have a place in modern information retrieval?. *Knowledge Organization, 43(3),* 145-159.

Hobbes, T. (1963). *Leviathan*. Cleveland: World Pub. Co.

Holman, L. (2011). Millennial students' mental models of search: Implications for academic librarians and database developers. *Journal of Academic Librarianship, 37(1),* 19.

Holsapple, C. & Joshi, K. (2001). Organizational knowledge resources. *Decision Support Systems*, *31*(1), 39-54.

Hook, Daniel, Mark Hahnel, and Ian Calvert. "The ascent of open access." *Digital Science* (2019).

Horvitz, E., Ieong, S., Mishra, N., & White, R. (2014). Time-critical search. *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, 747-756.

Hu, Xin, Jiangli Duan, and Depeng Dang. "Scalable aggregate keyword query over knowledge graph." *Future Generation Computer Systems* 107 (2020): 588-600.

Huang, Q., Jiang, S., Tian, Q., & Wang, S. (2012, June). Multi-feature metric learning with knowledge transfer among semantics and social tagging. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2240-2247). IEEE.

Huang, X. (2009). *Relevance, rhetoric, and argumentation: A cross-disciplinary inquiry into patterns of thinking and information structuring.* (Doctoral dissertation, University of Maryland, College Park).

ISO. (2013). ISO 25964-2 – Information and documentation- Thesauri and interoperability with other vocabularies- Part 2: Interoperability with other vocabularies.

Iyer, H. & Bungo, L. (2011). An examination of semantic relationships between professionally assigned metadata and user-generated tags for popular literature in complementary and alternative medicine. *Information Research, 16*(3).

Jashapara, A. (2005). The emerging discourse of knowledge management: a new dawn for information science research? *Journal of Information Science*, *31*(2), 136-148.

Jeong, S., Kim, H., Kim, J. & Park, Y. (2014). Clinical data element ontology for unified indexing and retrieval of data elements across multiple metadata registries. *Healthcare Informatics Research, 20*(4), 295.

Joudrey, D., Taylor, A., & Miller, D. (2015). *Introduction to Cataloging and Classification*. ABC-CLIO.

Kairam, S. & Pirolli, P. (2013). A knowledge-tracing model of learning from a social tagging system. *User Modeling and User-Adapted Interaction*, *23*(2-3), 139-168.

Kakali, C., & Papatheodorou, C. (2010). Exploitation of folksonomies in subject analysis. *Library & Information Science Research*, *32*(3), 192-202.

Kashyap, M. (2003). Likeness between Ranganathan's postulations based approach to knowledge classification and entity relationship data modelling approach. *Knowledge Organization*, *30* (1), 1-19.

Kebede, G. (2010). Knowledge management: An information science perspective. *International Journal of Information Management, 30*(5), 416-424.

Khalid, Shah, and Shengli Wu. "Supporting scholarly search by query expansion and citation analysis." *Engineering, Technology & Applied Science Research* 10.4 (2020): 6102-6108.

Kinnell, M., Meadows, J., McKnight, C., Oppenheim, C., & Summers, R. (1999). Information science in 2010: A Loughborough University view. *Journal of the Association for Information Science and Technology*, *50*(12).

Kipp, M. (2011). Tagging of biomedical articles on CiteULike: A comparison of user, author and professional indexing. *Knowledge Organization*, *38*.

Kitzie, Vanessa L., et al. "Using the World Café Methodology to support community-centric research and practice in library and information science." *Library & Information Science Research* 42.4 (2020): 101050.

Kless, D., Milton, S., Kazmierczak, E., & Lindenthal, J. (2015). Thesaurus and ontology structure: Formal and pragmatic differences and similarities. *Journal of the Association for Information Science and Technology, 66*(7), 1348-1366.

Kotis, K., & Vouros, G. A. (2006). Human-centered ontology engineering: The HCOME methodology. *Knowledge and Information Systems, 10*(1), 109-131.

Krishnamurthy, M. & Raghavan, K. (2013). An overview of classification technologies. *Journal of Library & Information Technology, 33(4),* 306-313.

Kupersmith, J. (2012). Library terms that users understand. LAUC-B and library staff research. Accessed October 21, 2016, from http://escholarship.org/uc/item/3qq499w7

Lai, L. (2004). *Knowledge organization by information technology consultants: Exploring and discovering the organizational aspect of knowledge management.* (Doctoral dissertation, University of Pittsburgh).

Lai, Lien-Fu, et al. "Developing a fuzzy search engine based on fuzzy ontology and semantic search." *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE, 2011.

Lancaster, F. (1999). Second thoughts on the paperless society. Library Journal- New York, *124*, 48-51.

Leblanc, Elina. "Participatory indexing in the eyes of its potential users: An example of a co-design of participatory services in an academic digital library." *International Conference on Theory and Practice of Digital Libraries*. Springer, Cham, 2020.

Leeds, O. & Mather, R. (2016). Realization of the knowledge-based organization. In *Impact: The Journal of Innovation Impact, 7*(1), 222.

Leidner, D., & Schultze, U. (2002). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly, 26*(3), 213–242.

Leonard, C. (2016). How discoveryability is shaping academic research. *LinkedIn*. Accessed April 21, 2016, from https://www.linkedin.com/pulse/how-discoverability-shaping-academic-research-chris-leonard

Liebowitz, J. (2012). *Knowledge management handbook: Collaboration and social networking,* (2nd edition). Hoboken: CRC Press.

Loehrlein, A. (2012). *Priming effects associated with the hierarchical levels of classification systems.* (Doctoral dissertation, Indiana University).

Lu, C., Park, J., & Hu, X. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science, 36*(6), 763-779.

Luchins, D. (2012). Clinical expertise and the limits of explicit knowledge. *Perspectives in Biology and Medicine*, *55*(2), 283-290.

Lyu, P., Ma, F., & Wang, X. (2014) Knowledge discovery of complex networks research literatures. In Chen & Larsen. *Library and information sciences: Trends and research*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-54812-3

Mai, J. (1999a). A postmodern theory of knowledge organization. In *Proceedings of the ASIS Annual Meeting, 36.* 547-56.

Mai, J. (1999b). Deconstructing the indexing process. *Advances in Librarianship, 23*, 269-298.

Mai, J. (2000). *The subject indexing process: an investigation of problems in knowledge representation* (Doctoral dissertation, University of Texas at Austin).

Mai, J. (2003). The future of general classification. *Cataloging & Classification Quarterly*, *37*(1-2), 3-12.

Mai, J. (2004a). Classification in context: relativity, reality, and representation. *Knowledge Organization*, *31*(1), 39-48.

Mai, J. (2004b). Classification of the web: challenges and inquiries. *Knowledge Organization, 31*(2), 92.

Mai, J. (2005). Analysis in indexing: document and domain centered approaches. *Information Processing & Management, 41*(3), 599-611.

Mai, J. (2008). Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowledge Organization*, *35*(1), 16-29.

Mai, J. (2010). Classification in a social world: Bias and trust. *Journal of Documentation, 66*(5), 627-642.

Mai, J. (2011a). The modernity of classification. *Journal of Documentation*, *67*(4), 710-730.

Mai, J. (2011b). Folksonomies and the new order: Authority in the digital disorder. *Knowledge Organization*, *38*(2), 114-122.

Malcolm, N., Wittgenstein, L., & Wright, G. (2001). *Ludwig Wittgenstein: A memoir*. Oxford University Press.

Mangisengi, O., & Essmayr, W. (2003, September). P2P knowledge management: an investigation of the technical architecture and main process. In *Database and Expert*

Systems Applications, 2003. Proceedings. 14th International Workshop on (787-791). IEEE.

Marco, F. (2016a). Enhancing the visibility and relevance of thesauri in the web: Searching for a hub in the linked data environment. *Knowledge Organization*, *43*(3).

Marco, F. (2016b). The evolution of thesauri and the history of knowledge organization: Between the sword of mapping knowledge and the wall of keeping it simple. *Brazilian Journal of Information Science*, *10*(1).

Martin, B. (2008). Knowledge management. In C. Blaise (Ed.), *Annual review of information science and technology (ARIST), 42,* 371–424. Medford, NJ: Information Today, Inc.

Martin, Jennifer M. "Records, responsibility, and power: An overview of cataloging ethics." *Cataloging & Classification Quarterly* 59.2-3 (2021): 281-304.

Martínez-González, M. & Alvite-Díez, M. (2014). On the evaluation of thesaurus tools compatible with the semantic web. *Journal of Information Science, 40*(6), 711-722.

Materska, K. (2014, October). Information heuristics of information literate people. In *European Conference on Information Literacy* (59-69). Springer International Publishing.

Maurer, M. & Shakeri, S. (2016). Disciplinary differences: LCSH and keyword assignment for ETDs from different disciplines. *Cataloging & Classification Quarterly, 54*(4), 213-243.

Mayr, Philipp, et al. "Introduction to the special issue on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL)." *International Journal on Digital Libraries* 19.2 (2018): 107-111.

McKay, D. & Parker, R. (2016). It's the end of the world as we know it... or is it?. *Marketing and Outreach for the Academic Library: New Approaches and Initiatives, 7*, 81.

McKerlich, R., Ives, C., & McGreal, R. (2013). Measuring use and creation of open educational resources in higher education. *The International Review of Research in Open and Distributed Learning*, *14*(4).

Meij, E. & De Rijke, M. (2007, September). Thesaurus-based feedback to support mixed search and browsing environments. In *International Conference on Theory and Practice of Digital Libraries* (pp. 247-258). Springer Berlin Heidelberg.

Miller, W., & Pellen, R. (Eds.). (2014). *Googlization of libraries*. New York: Routledge.

Mohan, B. S, and Iqbalahmad U Rajgoli. "Mapping of Scholarly Communication in Publications of the Astronomical Society of Australia, Publications of the Astronomical Society of Japan, and Publications of the Astronomical Society of the Pacific: A Bibliometric Approach." *Science & technology libraries (New York, N.Y.)* 36.4 (2017): 351–375. Web

Mosa, A., & Yoo, I. (2013). A study on PubMed search tag usage pattern: association rule mining of a full-day PubMed query log. *BMC Medical Informatics and Decision Making*, *13*(1), 1.

Musen, M., Noy, N., Singer, P., Strohmaier, M., Tudorache, T., & Walk, S. (2014). Discovering beaten paths in collaborative ontology-engineering projects using markov chains. *Journal of Biomedical Informatics, 51*, 254-271.

Nickles, T. (2016). Fast and frugal heuristics at research frontiers. In *Models and Inferences in Science* (pp. 31-54). Springer International Publishing.

Nielson, K. (2011). Self-study with language learning software in the workplace: What happens. *Language Learning & Technology*, *15*(3), 110-129.

Noorden, R. (2014). Global scientific output doubles every nine years. *Newsblog*. Nature.com. Accessed April 3, 2016, from http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html

Olson, H. (2007). How we construct subjects: A feminist analysis. *Library Trends*, *56*(2), 509-541

OU. "Library of Congress Accepts OU Libraries' Proposal to Change Subject Heading to 'Tulsa Race Massacre.'" (2021). https://www.ou.edu/web/news_events/articles/news_2021/library-of-congress-accepts-ou-libraries-proposal-to-change-subject-heading-to-tulsa-race-massacre

Panajotu, K. (2012). The use of the NATO glossary of terms and definitions: Allied administrative publications– AAP-6 (2010) in language training. *Education*, *11*(2), 249-255.

Panzer, M., & Zeng, M. (2009, September). Modeling classification systems in SKOS: some challenges and best-practice recommendations. In *International Conference on Dublin Core and Metadata Applications*.

Parinov, S. & Kogalovsky, M. (2014). Semantic linkages in research information systems as a new data source for scientometric studies. *Scientometrics*, *98*(2), 927-943.

Phung, D., Webb, G. I., & Sammut, C. (Online service). (2020). Encyclopedia of machine learning and data science. New York, NY: Springer US.

Pirmann, C. (2012). Tags in the catalogue: Insights from a usability study of LibraryThing for libraries. *Library Trends*, *61*(1), 234-247.

Pirolli, P. & Kairam, S. (2013). A knowledge-tracing model of learning from a social tagging system. *User Modeling and User-Adapted Interaction*, *23*(2-3), 139-168.

Porter, J. (2011). Folksonomies in the library: Their impact on user experience, and their implications for the work of librarians. *Australian Library Journal*, *60*(3), 248-255.

Qualtrics. (2019). “Calculating Sample Size.” Accessed on May 21, 2019 at https://www.qualtrics.com/blog/calculating-sample-size/

Ramakrishnan, C., Sheth, A. & Thomas, C. (2005). Semantics for the semantic web: The implicit, the formal and the powerful. *International Journal on Semantic Web and Information Systems (IJSWIS), 1*(1), 1-18.

Rayward, W. (2014). Information Revolutions, the information society, and the future of the history of information science. *Library Trends, 62*(3), 681-713.

Reinanda, Ridho, Edgar Meij, and Maarten de Rijke. *Knowledge graphs: An information retrieval perspective*. Now Publishers, 2020.

Robson, A., & Robinson, L. (2012). Building on models of information behavior: Linking information seeking and communication. *Journal of Documentation*, *69*(2), 169–193.

Rodriguez-Gonzalez, A., García-Crespo, A., Colomo-Palacios, R., Gomez-Berbís, J., & Jimenez-Domingo, E. (2012). Using ontologies in drug prescription: the semMed approach. *Intelligence Methods and Systems Advancements for Knowledge-Based Business*, 247.

Rolla, P. (2009). User tags versus subject headings: Can user-supplied data improve subject access to library collections? *Library Resources & Technical Services, 53*(3), 174.

Rosa, C., Cantrell, J., Cellentani, D., Hawk, J., Jenkins, L., & Wilson, A. (2005). Perceptions of Libraries and Information Resources: A Report to the OCLC Membership. Dublin, Ohio: OCLC.

Samanta, Kalyan Sundar, and Durga Sankar Rath. "User-Generated Social Tags Versus Librarian-Generated Subject Headings: A Comparative Study in the Domain of History." *DESIDOC Journal of Library & Information Technology* 40.3 (2020).

Savolainen, R., Talja, S., & Tuominen, K. (2002, July). Discourse, cognition, and reality: Toward a social constructionist metatheory for library and information science. In *Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (CoLIS): Emerging Frameworks and Methods. Libraries Unlimited, Greenwood Village* (pp. 271-283).

Schurer, S. & Vampati, U. (2012). Development and applications of the Bioassay Ontology (BAO) to describe and categorize high-throughput assays. In *Assay Guidance Manual,* ed. Sittampalam. Methesda, MD: Eli Lilly & Company and the National Center for Advancing Translational Science.

Shaik, Zaina, Filip Ilievski, and Fred Morstatter. "Analyzing Race and Country of Citizenship Bias in Wikidata." *arXiv preprint arXiv:2108.05412* (2021).

Sharif, A. (2009). Combining ontology and folksonomy: An integrated approach to knowledge representation, 1–13. Accessed August 28, 2016, from http://eprints.rclis.org/15058/

Sharma, R. (2015). Knowledge management in the digital era: A challenge for librarians. *International Journal of Academic Research in Commerce & Management,*1(1).

Simmons, Harold. An Introduction to Category Theory. Cambridge: Cambridge University Press, 2011.

Smiraglia, R. (2001a). Further progress toward theory in knowledge organization. *Canadian Journal of Information and Library Science, 26*(2-3), 31-50.

Smiraglia, R. (2001b). *The nature of 'a work:' Implications for the organization of knowledge*. New York: Scarecrow Press.

Smiraglia, R. (2002a). The progress of theory in knowledge organization. *Library Trends, 50*(3), 330-349.

Smiraglia, R. (2002b). *Works as entities for information retrieval*. New York: Routledge.

Smiraglia, R. (2012). Epistemology of domain analysis. *Cultural Frames of Knowledge. Wurzburg: Ergon-Verlag*, 111-24.

Smiraglia, R. (2013). The epistemological dimension of knowledge organization. *IRIS-Revista de Informação, Memória e Tecnologia, 2*(1), 2-11.

Smiraglia, R. (2014). *The Elements of Knowledge Organization*. New York: Springer Cham Heidelberg.

Smith, T. & Zeng, M. (2006). Building semantic tools for concept-based learning spaces: Knowledge bases of strongly-structured models for scientific concepts in advanced digital libraries. *Journal of Digital Information, 4*(4).

Solvberg, I., Nordbo, I., & Aamodt, A. (1992). Knowledge-based information retrieval. *Future Generation Computer Systems, 7*(4), 379-390.

Souza, R., Tudhope, D., & Almeida, M. (2012). Towards a taxonomy of KOS: Dimensions for classifying knowledge organization systems. *Knowledge Organization, 39*(3), 179-192.

Spencer, D. (2009). *Card sorting: Designing usable categories*. Rosenfeld Media.

Sun, Feipeng, et al. "A vocabulary recommendation system based on knowledge graph for Chinese language learning." *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2020.

Szostak, R. (2003). Classifying scholarly theories and methods. *Knowledge Organization, 30*(1), 20-35.

Szostak, R. (2008). Classification, interdisciplinary, and the study of science. *Journal of Documentation, 64*(3), 319–332.

Tang, J. (2016). Folksonomies: The opportunistic vocabulary. *LinkedIn*. Accessed October 8, 2016, from https://www.linkedin.com/pulse/folksonomies-opportunistic-vocabulary-jeanette-a-tang-m-s-l-s-?trk=hb_ntf_MEGAPHONE_ARTICLE_POST.

Tinker, A., Pollitt, A., O'Brien, A., & Brakevelt, P. (1999). The Dewy Decimal Classification and transition from physical to electronic knowledge organization. *Knowledge Organization*, 26 (2), 0-96.

Tudhope, D. (2012). 2012 Update on ISO 25964 Thesauri and interoperability with other vocabularies. NKOS workshop.

Tullis T, & Wood L. (2004). How many users are enough for a card-sorting study?, In *Proceedings of the Usability Professionals' Association 2004 conference,* Minneapolis, MN, June 7-11, 2004.

US News & World Report. (2019). "*Best Global Engineering Schools*." Accessed on May 21, 2019 at https://www.usnews.com/education/best-global-universities/engineering (2019)

Vakkari, Pertti. "Comparing Google to a digital reference service for answering factual and topical requests by keyword and question queries." *Online Information Review* (2011).

Vandic, D., van Dam, J., & Frasincar, F. (2012). Faceted product search powered by the Semantic Web. *Decision Support Systems*, *53*(3), 425-437.

Wang, Hongwei, et al. "Ripplenet: Propagating user preferences on the knowledge graph for recommender systems." *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 2018.

Wetzker, R., Zimmermann, C., Bauckhage, C., & Albayrak, S. (2010, February). I tag, you tag: translating tags for advanced user models. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 71-80). ACM.

White, H. (2013). Examining scientific vocabulary: Mapping controlled vocabularies with free text keywords. *Cataloging & Classification Quarterly*, *51*(6), 655–674.

White, H. C. (2012). *Organizing scientific data sets: Studying similarities and differences in metadata and subject term creation.* (Doctoral dissertation, The University of North Carolina at Chapel Hill).

Willoughby, C., Bird, C., & Frey, J. (2015). User-defined metadata: Using cues and changing perspectives. *International Journal of Digital Curation, 10*(1), 18-47.

Wilson, T. (2002). The nonsense of knowledge management. *Information Research*, *8*(1), 8-1

Wise, Colby, et al. "COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature." *arXiv preprint arXiv:2007.12731* (2020).

Wyatt, J. (2009). *Knowledge organization and domain performance.* (Doctoral dissertation, Fordham University).

Yan, E. & Zhu, Y. (2015). Identifying entities from scientific publications: A comparison of vocabulary-and model-based methods. *Journal of Informetrics*, *9*(3), 455-465.

Yang, S. (2012). *Tagging for subject access: A glimpse into current practice by vendors, libraries, and users.* Information Today, Inc.

Zeng, M. & Zumer, M. (1995). Comparison and evaluation of OPAC end-user interfaces. *Cataloging & Classification Quarterly*, *19*(2), 67-98.

Zeng, M. (1999). Metadata elements for object description and representation: A case report from a digitized historical fashion collection project. *Journal of the Association for Information Science and Technology*, *50*(13), 1193.

Zeng, M. (2008). Knowledge organization systems (KOS). *Knowledge Organization*, *35*(2-3), 160-182.

Zhu, D. & Dreher, H. (2012). Exploring semantic characteristics of socially constructed knowledge repository to optimize Web search. *Advancing Information Management through Semantic Web Concepts and Ontologies*, 250.

Ziemba, L. (2011). *Information retrieval with concept discovery in digital collections for agriculture and natural resources.* (Doctoral dissertation, University of Florida).

Zou, Xiaohan. "A survey on application of knowledge graph." *Journal of Physics: Conference Series*. Vol. 1487. No. 1. IOP Publishing, 2020.