INFO 202-16 Information Retrieval System Design
Project 2
Group 8- Holly Goodrich, Shelly Herbert, Kristin Holzhauer, Jenelle Klavenga
Fall 2018
San Jose State University

Part A. Design Database

Part B. Create Content

Part C. Query & Evaluate

Part A: Design Database

A1. Descriptor Vocabulary

abstract	digital collections	information retrieval
abstracting	digital libraries	information science
access	effectiveness	information seeking
archives	electronic records	information systems
art	electronic resources	keywords
Boolean logic	evaluation	librarians
browsing	George Boole	librarianship
cataloging	history	libraries
catalogs	index	metadata
classification	indexers	monitoring
Claude Shannon	indexing	Mortimer Taube
collections	information behavior	museums
cultural heritage institutions	information design	online catalogs
data	information needs	paradigms
design		

A2. Statement of Purpose

The target users of this database are students in an MLIS program with an interest in information science. The goal of this database is to create a usable system that allows for detailed searching of articles related to the field of information science. The title field will allow students to search titles for certain words or phrases. The author field allows students to search for a specific author's name. Publication information will assist students in searching for a specific publication. Year published allows students to search for articles by publication date, if more current information is needed. The abstract field allows students to search for key terms within the abstract of the article, which would help them find articles on their specific topics.

Descriptors allow students to search the database for articles using key terms that are more specific and common in natural language.

A3. Data Structure

group8_articles
---

group8_article_id	int(11)	textbox	group8_article_id
title	text	comment	Title
author	text	comment	Author
publication_information	text	comment	Publication_Information
year_published	varchar(20)	textbox	Year_published
abstract	text	comment	Abstract
descriptors	text	comment	Descriptors

A4. Rules

Field Name: Title
Required: Yes
Field type: Comment
Definition: Enter the title as written on the document. Format the title using APA guidelines.
Example: Information search tactics

Field Name: Author
Required: Yes
Field type: Comment
Definition: Enter the author's name putting last name first followed by a comma and then first name. Enter the author's middle initial, if available, followed by a period after the first name. For multiple authors, list names in the order they are listed in article.
Example: Bates, Marcia J.

Field Name: Publication Information
Required: Yes
Field type: Comment
Definition: Enter the publication information as found on the document using APA guidelines.
Example: Journal of the American Society for Information Science, 30(4), 205-214.

Field Name: Year Published
Required: Yes
Field type: Textbox

Definition: Enter the four digit number for the year the article was published.
Example: 1999

Field Name: Abstract
Required: No
Field type: Comment
Definition: Enter all text as written in the article. If no abstract is present, leave the field blank.

Field Name: Descriptors
Required: Yes
Field type: Comment
Definition: Enter 3-6 key terms or phrases from the controlled vocabulary list that are applicable to the article's key themes.

Part B. Create Content

B1. Search Page URL: **https://libr202.sjsu.edu/webdata_pro/student/4149/cgi-bi**

B2. Database Records

group8 article id	Title	Author	Publication Information	Year published	Abstract	Descriptors
1	The invisible substrate of information science	Bates, Marcia J.	Journal of the American Society for Information Science, 50(12), 1043-1	1999	The explicit, above-the-water-line paradigm of information science is well known and widely discussed. Every disciplinary paradigm, however, contains elements that are less conscious and explicit in the thinking of its practitioners. Elucidates the key elements of the below-the-water-line portion of the information science paradigm. Highlights the role of information science as a meta-science: conducting research and developing theory around the documentary products of other disciplines and activities. Views the mental activities of the professional practice of the field as centering around representation and organization of information, rather than knowing information. Argues that such representation engages fundamentally different talents and skills from those required in other professions and intellectual disciplines. Also considers methodological approaches and values of information science.	information design. information science. paradigms. theories.
2	Vocabulary as a central concept in library and information science	Buckland, Michael	Proceedings of the Third International Conference on Conceptions of	1999	The nature and role of vocabulary in information systems is examined. "Vocabulary" commonly refers to the stylized adaptation of natural language to form indexes and thesauri. Much of bibliographic access, filtering, and	indexing. information retrieval. metadata. searching. vocabularies

			Library and Information Science. Dubrovnik, Croatia, 23-26 May 1999		information retrieval can be viewed as matching or translating across vocabularies. Multiple vocabularies are simultaneously present. A simple query in an online catalog normally involves at least five distinct vocabularies: those of the authors; the cataloger; the syndetic structure; the searcher; and the formulated query.	
3	A brief history of information architecture	Resmini, Andrea & Rosati, Luca	Journal of Information Architecture, 3(2), 33-46	2012	Information architecture (IA) is a professional practice and field of studies focused on solving the basic problems of accessing, and using, the vast amounts of information available today. This article covers the history of information design from the 1960s, when the concept was first introduced in the computing world, to today, in a world where relationships with people, places, objects, and companies are shaped by semantics and not only by physical proximity. (This article is a reprint of parts of Chapter 2, “Towards a Pervasive Information Architecture”, from Andrea Resmini and Luca Rosati's “Pervasive Information Architecture”, a book published by MorganKauffman. The text was partially edited for clarity by the authors.)	history. information design. access.
4	On the shoulders of giants : From Boole to Shannon to Taube: The origins and development of computerized information from the Mid-19th Century to the present	Smith, Elizabeth S.	Information Technology and Libraries, 12(2), 217-226.	1993	This article describes the evolvement of computerized information storage and retrieval, from its beginnings in the theoretical works on logic by George Boole in the mid-nineteenth century, to the application of Boole’s logic to switching circuits by Claude Shannon in the late 1930s, and the development of coordinate indexing by Mortimer Taube in the late 1940s and early 1950s. Thus electronic storage and retrieval of information, as we know it today, was the result of two major achievements: the advancement of computer technology initiated to a large extent by the work of Shannon, and the development of coordinate indexing and retrieval by the work of Taube. Both these achievements are based on and are the application of the theoretical work of George Boole.	Boolean logic. Claude Shannon. George Boole. history. information science. Mortimer Taube.
5	Information retrieval as a trial-and-error process	Swanson, Don R.	Library Quarterly, 47(2), 128-148.	1977	Recognition of the essential role of trial and error in access to scientific literature may point the way toward improved information services and may illuminate inconsistencies that have beset many retrieval experiments. This paper examines three important and well-known information retrieval experiments, with a focus on certain internal inconsistencies and on the high variability of search results. In these	information systems. information. retrieval. libraries. indexing.

					experiments, retrieval systems are evaluated in terms of their ability to select relevant documents and reject those that are irrelevant. It is suggested that this criterion is inadequate because of ambiguities inherent in the concept of relevance and that closer attention to trial-and-error processes may be helpful in developing better criteria. Specific examples of how one might improve document retrieval, library use, and citation indexing are offered.	
6	Historical note: Information retrieval and the future of an Illusion	Swanson, Don R.	Journal of the American Society for Information Science, 39(2), 92 -98	1988	The article presents the historical aspects of information retrieval (IR). Some experimental tests of information systems have yielded good retrieval results and some very poor results. Automatic keyword-in-context indexing was not yet born, though its conceptual origin dates back at least to the year 1247. The decade of the 1950s was marked by endless disputes among proponents of various breeds of indexing schemes, classification systems, and bird-dogs. The Blair-Maron experiment, based on a collection of some 400,000 documents, offers valuable insight into the nature and the subtlety of the conceptual problems of IR-the problems of meaning. An important unintended consequence of specialization is the failure of the various branches of science to fertilize one another.	access. history. information design. information retrieval.
7	A survey of stemming algorithms in information retrieval	Moral, Cristiano, de Antonio, Angelica, Imbert, Ricardo, & Ramirez, Jaime	Information Research, 19(1).	2014	Background. During the last fifty years, improved information retrieval techniques have become necessary because of the huge amount of information people have available, which continues to increase rapidly due to the use of new technologies and the Internet. Stemming is one of the processes that can improve information retrieval in terms of accuracy and performance. Aim. This paper provides a detailed assessment of the current status of the stemming process framed in an information retrieval application field by tracing its historical evolution. Method. Papers presenting the first approaches for stemming were reviewed to extract their main features, benefits and drawbacks. Additionally, papers dealing with stemmers for non-English languages or with some more recent proposals were also consulted and compiled. Finally, experimental papers defining the most well-known methods and metrics aimed at evaluating and classifying stemmers were also taken into account to expose their contributions and results. Results. Even if not all researchers agree on the benefits and drawbacks of using stemming in an	history. information retrieval. stemming. stemmers. effectiveness.

					information retrieval process in general terms, many of them agree on its benefits in specific contexts, such as when the language is highly inflective, when documents are short or when there is limited space for storing data. Some researchers also state that the nature of the documents can influence the performance and the accuracy of the stemmer. Conclusions. Despite many researchers having investigated this field over many years, there are still some open questions, such as how to evaluate a stemmer independently of the information retrieval process, or how much a stemmer improves an information retrieval application in terms of speed. As a summary, some guidelines are also provided to help readers to determine which is the best stemmer for their needs and the tasks they have to carry out.	
8	Indexing and access for digital libraries and the internet: Human, database, and domain factor	Bates, Marcia J.	Journal of the American Society for Information Science, 49(13), 1185-1205.	1998	Presents information on a study which looked at indexing and access to digital libraries and the Internet. Factors important in the design of access mechanisms; Skills of an indexer; Reference to pervious literature; Information on folk classification.	access. digital libraries. indexer. indexing. classification.
9	The design of browsing and berrypicking techniques for the online search interface	Bates, Marcia J.	Online Review, 13(5), 407-424.	1989	First, a new model of searching in online and other information systems, called "berrypicking", is discussed. This model, it is argued, is much closer to the real behavior of information searchers than the traditional model of information retrieval is, and , consequently, will guide our thinking better in the design of effective interfaces. Second, the research literature of manual information seeking behavior is drawn on for suggestions of capabilities that users might like to have in online systems. Third, based on the new model and research on information seeking, suggestions are made for how new search capabilities could be incorporated into the design of search interfaces. Particular attention is given to the nature and types of browsing that can be facilitated.	browsing. design. information behavior. information retrieval. information seeking. searching
10	Library data in a modern context	Coyle, Karen	Library Technology Reports, 46(1), 5-13.	2010	The article examines the state of library data in a modern context. It recounts the history of modern library cataloging practice, which included Anthony Panizzi's 91 rules and the creation of the Online Public Access Catalog (OPAC) in the 1980s. A description of metadata, which is processed by computers to be understandable to humans, is discussed as well as library bibliographical metadata. Also	cataloging. data. history. metadata. online catalogs.

					explained is the process of cataloging and the use of the library catalog. The role of the World Wide Web as an information platform, including social networking sites, is noted.	
11	Metadata for all: Descriptive standards and metadata sharing across libraries, archives, and museums	Elings, Mary W. & Waibel, Gunter	First Monday, 12(3).	2007	Integrating digital content from libraries, archives and museums represents a persistent challenge. While the history of standards development is rife with examples of cross-community experimentation, in the end, libraries, archives and museums have developed parallel descriptive strategies for cataloguing the materials in their custody. Applying in particular data content standards by material type, and not by community affiliation, could lead to greater data interoperability within the cultural heritage community. In making this argument, the article demystifies metadata by defining and categorizing types of standards, provides a brief historical overview of the rise of descriptive standards in museums, libraries and archives, and considers the current tensions and ambitions in making descriptive practice more economic [1].	archives. history. libraries. metadata. museums. standards.
12	A cognitive process model of document indexing	Farrow, John F.	Journal of Documentation, 47(2), 149-166.	1991	Classification, indexing and abstracting can all be regarded as summarisations of the content of a document. A model of text comprehension by indexers (including classifiers and abstractors) is presented, based on task descriptions which indicate that the comprehension of text for indexing differs from normal fluent reading in respect of: operational time constraints, which lead to text being scanned rapidly for perceptual cues to aid gist comprehension; comprehension being task oriented rather than learning oriented, and being followed immediately by the production of an abstract, index, or classification; and the automaticity of processing of text by experienced indexers working within a restricted range of text types. The evidence for the interplay of perceptual and conceptual processing of text under conditions of rapid scanning is reviewed. The allocation of mental resources to text processing is discussed, and a cognitive process model of abstracting, indexing and classification is described.	abstract. abstracting. classification. indexers. Indexing. paradigms.
13	Issues in the development of a thesaurus for patients' chief complaints in the emergency department	Haas, Stephanie W. & Travers, Debbie A	67th Proceedings of the ASIS&T annual meeting, 41, 411-417.	2004	When a patient visits the Emergency Department (ED), the reason the patient is seeking care is recorded as the Chief Complaint (CC). Beyond its role in the patient's care, there is interest in the CC for secondary uses. Clinicians and epidemiologists can use CC for research. ED clinicians and administrators	design. electronic records. thesauri. vocabularies

					incorporate CC data into quality monitoring and improvement efforts. Public health officials can use it as data for health surveillance. But there is no controlled vocabulary for recording CC, or standard for a CC component in the patient record. Travers (2003) completed a crucial first step toward the creation of a thesaurus for CC by analyzing a corpus of CCs to determine the nature of the language used by triage nurses, and the concepts that were expressed. Her analysis also illuminated many issues concerning the content and structure of a CC thesaurus that must be discussed before the thesaurus can be developed. Using Cimino's 1998 article, “Desiderata for Controlled Medical Vocabularies in the Twenty-First Century”, as a framework, we discuss these issues and the resulting decisions that the thesaurus development team, along with other stakeholders, will encounter.	
14	Subject matter categorization of tags applied to digital images from art museums	Klavans, Judith L., LaPlante, Rebecca, & Golbeck, Jennifer	Journal of the Association for Information Science & Technology, 65(1), 3-12.	2014	The use of social tagging by cultural heritage institutions has grown in recent years. This paper forms a better understanding of the nature of those tags by examining tags assigned to a collection of 100 images of works of art from the steve.museum project using subject matter categorization. We present the results of experiments showing that the majority of these tags are generic in nature and describe the people and things in the image. These new results are not consistent with subject matter analyses of queries, tags, and index terms for other image collections, which suggests that the nature of social tags may be largely dependent on the type of collection and user needs. This information can be used by cultural heritage institutions to better manage and use the data provided through tags	art. digital collections. indexing. museums. social tagging.
15	Learning from librarians and teens about YA library spaces	Agosto, Denise E., Kuhlmann, L. Meghann, Pacheco Bell, J., & Bernier, Anthony	Public Libraries, 53(3), 24-28	2014	The article discusses the results of the empirical study of the physical and spatial aspects of young adult (YA) library services in the U.S. as of May 2014. Topics highlighted include ways for public libraries to improve their library services for YA users, the need for design revisions to adapt to how users are using and interacting with their libraries and analysis of video data gathered during the study. Also mentioned is the importance of access to technology for YA users.	librarians. libraries. public libraries. spaces. young adults.

16	People, places, and questions: An investigation of the everyday life information-seeking behaviors of urban young adults	Agosto, Denise E. & Hughes-Hassell, Sandra	Library & Information Science Research, 27(2), 141-163.	2005	This article presents preliminary findings from a research grant on the everyday life information-seeking (ELIS) behaviors of urban young adults. Twenty-seven teens aged 14 through 17 participated in the study. Qualitative data were gathered using written activity logs and semi-structured group interviews. A typology of urban teens' preferred ELIS sources, media types, and query topics is presented. The typology shows friends and family as preferred ELIS sources, cell phones as the preferred method of mediated communication, and schoolwork, time-related queries, and social life as the most common and most significant areas of ELIS. The results indicate a heavy preference for people as information sources and that urban teens hold generally unfavorable views of libraries and librarians. The conclusion lists questions that information practitioners should consider when designing programs and services for urban teens and calls for researchers to consider this often-ignored segment of the population as potential study participants.	design. librarians. public libraries. services. spaces. young adults
17	How does search behavior change as search becomes more difficult?	Aula, A., Khan, R.M., & Guan, Z.	CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems in Atlanta, GA. New York, NY: AMC.	2010	Search engines make it easy to check facts online, but finding some specific kinds of information sometimes proves to be difficult. We studied the behavioral signals that suggest that a user is having trouble in a search task. First, we ran a lab study with 23 users to gain a preliminary understanding on how users' behavior changes when they struggle finding the information they're looking for. The observations were then tested with 179 participants who all completed an average of 22.3 tasks from a pool of 100 tasks. The large-scale study provided quantitative support for our qualitative observations from the lab study. When having difficulty in finding information, users start to formulate more diverse queries, they use advanced operators more, and they spend a longer time on the search result page as compared to the successful tasks. The results complement the existing body of research focusing on successful search strategies.	information behavior. information seeking. search engines. searching. strategies. users.
18	Information search tactics	Bates, Marcia J.	Journal of the American Society for Information Science, 30(4), 205-214.	1979	As part of the study of human information search strategy, the concept of the search tactic, or move made to further a search, is introduced. Twenty-nine tactics are named, defined, and discussed in four categories: monitoring, file structure, search formulation, and term. Implications of the search tactics for research in search strategy are considered. The search tactics are intended to be practically useful in	information retrieval. information seeking. monitoring queries. searching. strategies.

					information searching. This approach to searching is designed to be general, yet nontrivial; it is applicable to both bibliographic and reference searches and in both manual and on-line systems.	
19	Manipulating search engine algorithms: the case of Google	Bar-Ilan, Judit	Journal of Information, Communication and Ethics in Society 5(2/3), 155-166.	2007	Purpose - To investigate how search engine users manipulate the rankings of search results. Search engines employ different ranking methods in order to display the "best" results first. One of the ranking methods is PageRank, where the number of links pointing to the page influences its rank. The "anchor text," the clickable text of the hypertext link is another "ingredient" in the ranking method. There are a number of cases where the public challenged the Google's ranking, by creating a so-called "Google bomb" - creating links to pages they wanted to be highly ranked for given query. Google is chosen as the search engine, because it is currently by far the most popular search engine. Design/methodology/approach - PageRank, one of the major parameters of Google's ranking algorithm is described, and the author explains how this algorithm is exploited by communities of users to promote a certain web page for a specific query. This process is called "Google bombing." Recent reaction of Google to this phenomenon is also described. Findings - Specific examples of "accomplished Google bombs" show that the public is able to manipulate search results. Originality/value - Google, instead of being an unobtrusive information retrieval tool has become highly influential in the web scenery. Some users pay for search engine optimization, while others utilize the power of the crowd to influence Google's rankings. This paper supports the claims of Introna and Nissenbaum regarding the power of search engines.	information retrieval. information seeking. queries. results. search engines. searching.
20	TREC: Improving information access through evaluation.	Voorhees, Ellen M.	Bulletin of the American society for Information Science and Technology, 32(1), 16-21.	2005	Evaluation is a fundamental component of the scientific method: researchers form a hypothesis, construct an experiment that tests the hypothesis and then assess the extent to which the experimental results support the hypothesis. National Institute of Standards & Technology proposed that instead of simply building a single large test collection, it organize a workshop that would both build a collection and investigate the larger issues surrounding test collection use. This was the genesis of the Text Retrieval Conference (TREC). TREC assumes the Cranfield paradigm of retrieval system evaluation, which is based on the abstraction of a test collection: a set of documents, a set of	evaluation. information needs. information retrieval. information systems. paradigms. relevance.

					information needs that TREC calls topics and a set of relevance judgments that say which documents should be retrieved for which topics. TREC has been able to build on the text-retrieval field's tradition of experimentation to significantly improve retrieval effectiveness and extend the experimentation to new sub-problems.	

Part C. Query & Evaluate

C1. Topics & Queries

Topic 1: How do users act when searching for information?

Queries in descriptor field:

users AND searching AND behavior	This retrieved 1 record.	The record ID was: 17
seeking AND behavior	This retrieved 2 records.	The record IDs were: 9, 17

Queries in abstract field:

users AND searching AND behavior	This retrieved 1 record.	The record ID was: 9
seeking AND behavior	This retrieved 2 records.	The record IDs were: 9, 16

Topic 2: How can information retrieval systems be evaluated?

Queries in descriptor field:

information retrieval AND systems AND evaluation	This retrieved 1 record.	The record ID was: 20
retrieval AND success	This retrieved zero records.	
retrieval AND failure	This retrieved zero records.	
information retrieval	This retrieved 7 records.	The record IDs were: 2, 6, 7, 9, 8, 19, 20

Queries in abstract field:

information retrieval AND systems AND evaluation	This retrieved zero records.	
retrieval AND success	This retrieved zero records.	
retrieval AND failure	This retrieved 1 record.	The record ID was: 6
information retrieval	This retrieved 6 records.	The record IDs were: 2, 5, 6, 7, 9, 19

Topic 3: How has the technological age changed information system design and retrieval?

Queries in descriptor field:

information systems AND history	This retrieved zero records.	
information retrieval AND history	This retrieved 2 records.	The record IDs were: 6, 7
information systems AND information retrieval AND history	This retrieved zero records.	
information retrieval AND technology	This retrieved zero records.	
Information design	This retrieved 3 records.	The record IDs were: 1, 3, 6

Queries in abstract field:

information systems AND history	This retrieved zero records.	
information retrieval AND history	This retrieved zero records.	
information systems AND information retrieval AND history	This retrieved zero records.	
information retrieval AND technology	This retrieved zero records.	
Information design	This retrieved 1 record.	The record ID was: 3

Topic 4: How does searching online affect a person's information seeking behavior?

Queries in descriptor field:

information seeking AND information behavior	This retrieved 2 records.	The record IDs were: 9, 17
information behavior AND search engines	This retrieved 1 record.	The record ID was: 17
information behavior AND searching	This retrieved 2 records.	The record IDs were: 9, 17
information seeking AND search engines	This retrieved 2 records.	The record IDs were: 17, 19
information seeking	This retrieved 4 records.	The record IDs were: 9, 17, 18, 19

Queries in abstract field:

information seeking AND information behavior	This retrieved zero records.	
information behavior AND search engines	This retrieved zero records.	

information behavior AND searching	This retrieved zero records.	
information seeking AND search engines	This retrieved zero records.	
information seeking	This retrieved 1 record.	The record ID was: 9

Topic 5: I would like to know more about information science history.

Queries in descriptor field:

history AND libraries	This retrieved 1 record.	The record ID was: 11
history AND information science	This retrieved 1 record.	The record ID was: 4
history AND librarians	This retrieved zero records.	
history	This retrieved 6 records.	The record IDs were: 3, 4, 6, 7, 10, 11

Queries in abstract field:

history AND libraries	This retrieved 1 record.	The record ID was: 11
history AND information science	This retrieved zero records.	
history AND librarians	This retrieved zero records.	
history	This retrieved 3 records.	The record IDs were: 3, 10, 11

Topic 6: I would like to know more about how young adults use libraries

Queries in descriptor field:

young adults AND libraries	This retrieved 2 records.	The record IDs were: 15, 16
young adults AND librarians	This retrieved 2 records.	The record IDs were: 15, 16
young adults AND users	This retrieved zero records.	
young adults	This retrieved 2 records.	The record IDs were: 15, 16

Queries in abstract field:

young adults AND libraries	This retrieved 1 record.	The record ID was: 16
young adults AND librarians	This retrieved 1 record.	The record ID was: 16
young adults AND users	This retrieved zero records.	
young adults	This retrieved 1 record.	The record ID was: 16

Topic 7: I'd like to know more about information retrieval system design.

Queries in Descriptor field:

systems AND design	This retrieved zero records.	
information AND retrieval AND design	This retrieved 2 records.	The record IDs were: 2, 9.

Queries in Abstract field:

systems AND design	This retrieved 2 records.	The record IDs were: 9, 18.
information AND retrieval AND design	This retrieved zero records.	

Topic 8: I'd like to know more about digital content being used in libraries.

Queries in Descriptor field:

digital AND libraries	This retrieved 1 record.	The record ID was: 8.
digital AND collections	This retrieved 1 record.	The record ID was 14.
digital AND content AND libraries	This retrieved zero records.	

Queries in Abstract field:

digital AND libraries	This retrieved 2 records.	The record IDs were: 8, 11
digital AND collections	This retrieved zero records.	
digital AND content AND libraries	This retrieved zero records.	

C2. Evaluation

Topic	**Relevant articles to the Topic (article #'s)**	**Queries**	**Relevant articles retrieved with Abstract field search (article #'s)**	**Relevant articles retrieved with Descriptor field search (article #'s)**	**Recall**	**Precision**	**Eff**
Topic 1: How do users act when searching for information?	9, 16, 17, 18	*users AND searching AND behavior*	9	17	Abstract: .25 Descriptor: .25	Abstract: 1 1.0 Descriptor: 1.0	Abstract: .4697 Descriptor: .4697

		Seeking AND behavior	9, 16	9,17	Abstract: .50 Descriptor: .50	Abstract: 1.0 Descriptor: 1.0	Abstract: .6464 Descriptor: .6464

Analysis & Takeaways: The first descriptor field search returned one article focused on user responses to difficulties encountered during harder searches using search engines. This is a relevant result, though the search question may have been intended to return a more general result. If that is the case, it would indicate that different search terms should be tried or the question reconsidered. A more specific question might be helpful. The second descriptor field search returned the same article as the first and found another focused on a model of information seeking that better reflected users actual search processes. This is a relevant article and shows some challenges with having both information seeking and searching as descriptor terms. These are different concepts and in certain articles the distinction would be important. For this search, however, there is little difference between these terms and the fact multiple terms must be used may frustrate the user or lead to missed information. It is worth considering whether both terms are necessary for the vocabulary or setting a clear rule about how best to assign the terms. The first abstract field search retrieved the model of information seeking article only. The second search returned that same article and an article about the information behavior seeking of young adults which had not turned up in previous searches. This was a relevant article that was not retrieved by the field search may indicate an error on the part of the indexer. If the abstract information accurately represents this article, either information seeking or searching should have been assigned as terms. The descriptors include services and design which do not appear to be as essential to the article as information seeking or a related term. Overall precision for all searches was very high but recall was lower. User would find relevant articles using these search terms but would likely miss several relevant articles regardless of terms used or abstract vs. field descriptor searches.

Topic	Relevant articles to the Topic (article #'s)	Queries	Relevant articles retrieved with Abstract field search (article #'s)	Relevant articles retrieved with Descriptor field search (article #'s)	Recall	Precision	Eff
Topic 2: How can information retrieval systems be evaluated?	5, 6, 7, 20	*information retrieval AND systems AND evaluation*	none	20	Abstract: 0 Descriptor: .25	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor: .4697

		retrieval AND success	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0
		retrieval AND failure	6	none	Abstract: .25 Descriptor: 0	Abstract: 1.0 Descriptor: 0	Abstract: .4697 Descriptor: 0
		Information retrieval	5, 6, 7	6, 7, 20	Abstract: .75 Descriptor: .75	Abstract: .50 Descriptor: .43	Abstract: .6047 Descriptor: .5599

Analysis & Takeaways: Overall most results for these searches were poor. The terms used in the first 3 sets of search terms did not match well with either the terms used in the abstracts or the descriptor terms for these articles. Use of a single broad term greatly improved recall but significantly reduced precision. The earlier searches while less successful in terms of finding articles but did return only relevant results. However, they often returned no results. This set of searches illustrates the challenges of designing searches that return both relevant and comprehensive results and shows the trade-offs between the two. The search would have been more successful if vocabulary terms like access or effectiveness were used. This may point to issues with the vocabulary terms available or the ways they are being used. It may also point to the many topics covered by the articles in this sample and the difficulty of judging relevance. Not all the articles I identified as relevant would be considered that way by all searchers (or indexers) this demonstrates another challenge in database design. Also, this was a broad search and one that was difficult to break down into precise search terms. The searcher may need to think of related questions or terms that might open up more possibilities.

Topic	**Relevant articles to the Topic (article #'s)**	**Queries**	**Relevant articles retrieved with Abstract field search (article #'s)**	**Relevant articles retrieved with Descriptor field search (article #'s)**	**Recall**	**Precision**	**Eff**

Topic 3: How has the technological age changed information system design and retrieval?	3, 4, 6, 7, 8	information systems AND history	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0
		information retrieval AND history	none	6, 7	Abstract: 0 Descriptor: 0.40	Abstract: 0 Descriptor: 1.0	Abstract 0: Descriptor: 0.5757
		information systems AND information retrieval AND history	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract:0 Descriptor: 0
		information retrieval AND technology	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0
		information design	3	3, 6	Abstract: 0.20 Descriptor: 0.60	Abstract: 1.0 Descriptor: .67	Abstract: 0.4343 Descriptor: 0.6333

Analysis & Takeaways: Overall the search results for this topic were low. The search keywords did not perform well in the abstract searches and only increased for both the abstract and descriptor fields when a single phrase search was done. Even with a single phrase search in the abstract field only 1 relevant document was found. This search highlights the difficulty in locating results that cover multiple topics. My topic search contained 3 different ideas- history, information system design, and information retrieval. However, the best results were given when only 1 of those ideas was used as a search keyword. This may show signs of weakness in the vocabulary. It may also show some gaps in the abstract information from each article with this content topic. Anyone searching this database for a topic similar to mine (3 ideas) may do better focusing on just one of the ideas for their search.

Topic	**Relevant articles to the Topic (article #'s)**	**Queries**	**Relevant articles retrieved with Abstract field search**	**Relevant articles retrieved with Descriptor field search (article #'s)**	**Recall**	**Precision**	**Eff**

			(article #'s)				
Topic 4: How does searching online affect a person's information seeking behavior?	9, 12, 17, 19, 29, 35, 37, 42	information seeking AND information behavior	none	9, 17	Abstract: 0 Descriptor: 0.25	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor : 0.4697
		information behavior AND search engines	none	17	Abstract: 0 Descriptor: 0.13	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor : 0.3813
		information behavior AND searching	none	9, 17	Abstract: 0 Descriptor: 0.25	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor : 0.4697
		information seeking AND search engines	none	17, 19	Abstract: 0 Descriptor: 0.25	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor : 0.4697
		information seeking	9	9, 17, 19	Abstract: 0.125 Descriptor: 0.38	Abstract: 1.0 Descriptor: .75	Abstract: 0.3813 Descriptor : 0.5240

Analysis & Takeaways: The results for this search were better than my previous search in the descriptor field. The results for the abstract field were still very low and again only became better with a single phrase search. I found it interesting that there were a number of articles that could be deemed relevant to this topic and yet the search produced results for 4 of the 6 with #9, 17, and 19 being the highest in frequency. It is interesting that when the descriptor field yields better results, the abstract field is low and vice versa. This shows the difficulty in creating a database that can produce significant search results for both fields. It is also interesting that more results weren't available as these were common LIS terms that were relevant to many of the articles we chose. We may have seen better potential for this with a larger sample size or more descriptors having been assigned to each article. Again with this search someone using this database would do better searching with a broad idea or with one key term or phrase.

Topic	Relevant articles to the Topic (article #'s)	Queries	Relevant articles retrieved with Abstract field search (article #'s)	Relevant articles retrieved with Descriptor field search (article #'s)	Recall	Precision	Eff
Topic 5: I would like to know more about information science history.	3, 4, 6,7, 10, 11	history AND libraries	11	11	Abstract: .17 Descriptor: .17	Abstract: 1.0 Descriptor: 1.0	Abstract: .4131 Descriptor : .4131
		history AND information science	none	4	Abstract: .17 Descriptor: .17	Abstract: 1.0 Descriptor: 1.0	Abstract: .4131 Descriptor : .4131
		history AND librarians	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor : 0
		history	3, 10, 11	3, 4, 6, 7, 10, 11	Abstract: .50 Descriptor: 1.00	Abstract: 1.0 Descriptor: 1.00	Abstract: .6464 Descriptor :1.00

Analysis & Takeaways: Most of the search results for this topic did not yield much in the way of results. Several searches gave no results, and the large majority gave less than 20% of possible results. Searching was most successful when a broad connecting term, "history" in this case, was used to search for a large number of articles that fell into various historical aspects of information sciences. For the most part, abstract searches were unsuccessful. Having the term "history" as a common search term throughout each search seemed to limit results because it was not frequently used in abstracts. Including an additional, mandatory term limited results further. While searching solely for "history" yielded all possible results because it was a shared descriptor. The experience has illustrated the need for more diverse search terms. Rather than connecting all searches with the "history" term, the inclusion of related terms while excluding "history" may have yielded better results. Perhaps the use of "archives" or "museums" or the inclusion of

historical figures such as George Boole might have helped, however using “archives” or “museums” may have pulled more results that were not actually relevant. For example, one article may have archives included in the descriptor or abstract but not actually address the history of information science.

Topic	Relevant articles to the Topic (article #’s)	Queries	Relevant articles retrieved with Abstract field search (article #’s)	Relevant articles retrieved with Descriptor field search (article #’s)	Recall	Precision	Eff
Topic 6: I would like to know more about how young adults use libraries	15, 16	young adults AND libraries	16	15, 16	Abstract: .50 Descriptor: 1.00	Abstract: .50 Descriptor: 1.00	Abstract: .5000 Descriptor: 1.00
		young adults AND librarians	16	15, 16	Abstract: .50 Descriptor: 1.00	Abstract: .50 Descriptor: 1.00	Abstract: .5000 Descriptor: 1.00
		young adults AND users	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor : 0
		young adults	16	15, 16	Abstract: .50 Descriptor: 1.0	Abstract: .50 Descriptor: 1.0	Abstract: .5000 Descriptor: 1.0
Analysis & Takeaways: This topic was reasonably successful. In most cases, half or all applicable records were found. The success of this is not related to altered search terms, though. It is because there are only two articles discussing young adults. Having fewer records to find creates the appearance of a more successful search. In reality, each search provides the same or fewer actual records as the information science history topic searches. What this search really drove home is how small the current database is. Young adult fiction is a very popular genre, and meeting the needs of teenagers is a pretty							

pressing goal for many libraries. The topic provides plenty of room for pulling articles that are not relevant to how young adults use libraries, however both articles that pulled with the young adult descriptor were relevant. It would be a more comprehensive database if we were able to include articles on young adult literature as a genre as well as young adults as a user group. The inclusion of young adult librarianship might also provide an interesting variety among articles with the young adult descriptor or with young adult in the article abstract. Without this diversity within subjects, it becomes difficult to rigorously test the database for precision.

Topic	Relevant articles to the Topic (article #'s)	Queries	Relevant articles retrieved with Abstract field search (article #'s)	Relevant articles retrieved with Descriptor field search (article #'s)	Recall	Precision	Eff
Topic 7: I'd like to know more about information retrieval system design	2, 3, 4, 5, 6, 9, 17	Systems AND design	9	none	Abstract: .14 Descriptor: 0	Abstract: .50 Descriptor: 0	Abstract: .2966 Descriptor: 0
		information AND retrieval AND design	none	2, 9	Abstract: 0 Descriptor: .29	Abstract: 0 Descriptor: 1.0	Abstract: 0 Descriptor: .4980

Analysis & Takeaways: My searches did not find very many results. There were several articles that had information about information science retrieval system design, however my choice of search terms was apparently not very good because I had fewer results. I'm not sure if the vocabulary list should be expanded or if my choice of search terms were lacking. I thought I had chosen main words that would be important to the search, however with there being few results, it could be indexer error or authors of the vocabulary error. The topic itself was probably too vague for the search. It could be more in depth about what specifically in information science retrieval design is being researched. Information science retrieval design is quite a long term, so maybe cutting out science would have helped. I'm unsure how to make this topic and search more productive. Does the searcher need more of a history of the topic or the components of design, or both? History could have been added as a search term and more results may have become available.

Topic	Relevant articles to the Topic (article #'s)	Queries	Relevant articles retrieved with Abstract field search (article #'s)	Relevant articles retrieved with Descriptor field search (article #'s)	Recall	Precision	Eff
Topic 8: I'd like to know more about digital content being used in libraries	8, 11, 14	*digital AND libraries*	8, 11	8	Abstract: .67 Descriptor: .33	Abstract: 1.0 Descriptor: .33	Abstract: .7667 Descriptor: .3300
		digital AND collections	none	14	Abstract: 0 Descriptor: .33	Abstract: 0 Descriptor: .33	Abstract: 0 Descriptor: .3300
		digital AND content AND libraries	none	none	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0	Abstract: 0 Descriptor: 0
Analysis & Takeaways: While this is a handy topic, I believe the number of articles in the database were not enough to search for this topic. There were only three relevant articles, however between my three search parameters, we did get all three article results. While technically this search did work and it did find the relevant articles, I'm unhappy with only having those three articles. There definitely needs to be a larger sample of articles in the database to make this more user friendly. Also, I don't think that content was a great word to use for this search. At this second, I cannot think of what other term I could have used to make it a better topic to search. It could be how libraries use digital content instead of digital content being used in libraries. Or it needs to be made more specific like what type of digital content, eBooks, streaming content, online resources such as Ebscohost.							

C3. Reflections

Our group worked well together throughout both projects. We did our best to divide responsibilities evenly and everyone was willing to help out where needed. It was nice to have a group to work with for many of the components of the project. Having a collective reflection of our ideas made a more user friendly and comprehensive end product. For this project, we all read the selected articles. We collectively created the original and final controlled vocabulary list. We each assigned vocabulary terms from the final controlled vocabulary list to the articles we read. Holly created the database. We all did individual searches and wrote the analysis for each search. We individually completed our personal reflections. Kristin edited and combined the project components into a Word document for submission. Much of the other work and details for the project were completed and added collaboratively in weekly Zoom sessions.

What worked and what didn't in the process of creating your subject term vocabularies?

It was nice to read and assign initial terms to the articles individually so we weren't influenced by others' opinions on the articles. Once these were initially assigned we had the benefit of working collaboratively to brainstorm on the relevancy of each term to ensure that each final assignment was the most relevant for the controlled vocabulary list. This collaboration allowed us to create a more holistic list of terms since it wasn't based on just one person's way of thinking or one person's ideas. The collaboration also led to a more user friendly vocabulary list since each person's search habits were taken into consideration in the making of the list.

We found it difficult to avoid multiple levels of specificity. Determining the best-fit and most relevant terms was difficult. We wanted to continually add terms that may not have been as relevant to the overall content of the document. When we had similar terms it was difficult to decide which term to eliminate and which term to use.

What challenges did you encounter when working to determine the main subjects of the articles?

Many times articles are not focused on just one idea or topic. It was difficult to determine the most relevant key terms to assign each article when the content was detailed and complex. Not having a detailed background in library science made it difficult to understand which concept was the priority concept of the article to focus on when assigning terms from the vocabulary list. The abstract of each article didn't always fully relate to the content of the actual article. You cannot rely solely on the information in the article's abstract to determine the final and most relevant key terms.

What improvements would you make in a future iteration of your database? For example: What about additional fields? More ways to provide subject access? More ways to provide other forms of access for users?

We feel it would have been nice to be able to assign more terms to each article. Even assigning 8 terms instead of a maximum of 6 terms would have provided more recall in the results of our searches. A database that is limited to 20 records is difficult to make user friendly as well. We feel having a much larger database lends itself to better search results. It is hard to evaluate the effectiveness of the database having such a small sample size.

What did you learn from doing the queries and evaluating the search results?

We have all learned how difficult it is to create a user friendly, effective, and accurate database. We learned how hard it is to develop a vocabulary list that yields high search results. Through this project we have a better understanding of the balance between recall and precision and that often improving the result of one means lowering the score in the other.

Holly's Individual Reflection

This project demonstrated how difficult making objects findable truly is. According to Morville (2005) findability includes "the quality of being locatable or navigable" and "the degree to which a particular object is easy to discover or locate" (p. 4). Attempting to describe articles in a way that makes them findable is difficult for many reasons. One is that language is inherently imprecise, as Morville says, "words are imprecise, ambiguous, indeterminate, vague, opaque…" (p. 51). The more precise one needs to be the harder it becomes.

In an everyday conversation, two people can think of slightly different things when a word is used. While miscommunication is not uncommon in these scenarios, usually they will understand each other well enough. When an indexer is trying to tell a group of users how to find a document based on a set of terms this level of ambiguity will often lead to problems. Since the process is inherently subjective, no concrete rules can definitively create a system in which these perfectly match ((Rowley & Hartley, 2016, p. 128). Information retrieval systems must address this but because language is always somewhat ambiguous and complex, there will always be failures. As Morville (2005) points out, "most categories we employ in everyday life are defined by fuzzy cognitive models rather than objective rules" (p. 133). Again, this may work well if we are attempting to communicate on a general level, but it can cause great difficult when we need to use the same terms in mostly the same way for the system to work. This was something I struggled with during the process of creating and testing our database.

Another area plagued by this kind of subjectivity is determining what an article is about. This cannot be determined simply by whether a word occurs in a document (Morville, 2005, p. 53; Rowley & Hartley, 2016, p. 126). In fact, as we learned in the exercise, the best words may

not be used in the article at all. Reducing thousands of words to three to six descriptive and relatively unambiguous terms while still expressing the "aboutness" of the article is challenging. Attempting to create a series of terms that still describes this article, collects similar articles, and does not include unrelated articles is daunting. After completing this project, I have greater appreciation for the subjective nature of the process of creating and assigning controlled vocabulary terms.

Kristin's Individual Reflection

This project has humbled me and challenged me in its complexity in many ways. I have always considered myself to be somewhat technologically savvy and able to understand the abstract components of projects. My over-analytical, type-A, perfectionist self didn't like the obscure and abstract nature of this way of thinking, however. This project has given me a newfound respect for anyone who professionally creates information retrieval systems.

Establishing findability is incredibly difficult. Databases and search engines rely on reducing large articles to just a small sampling of key terms. As Morville (2005) states if, "words are imprecise, ambiguous, indeterminate, vague, opaque…" (p. 51), then it is surprising databases and search engines can wind up being effective in recall and precision at all. This project highlighted this ambiguity in language very well. Many of the articles' "aboutness" was not found in the words used to describe the article in the abstract. In fact, some of the best terms weren't even used in the article at all. As creators of our database we had to go beyond the words used in front of us to determine terms that would increase findability. Each of our ideas were different making the final choice for our work very difficult. Ultimately relying on a group

consensus made it possible to narrow down terms and discover the best approach, but indexers tasked with making those same decisions aren't always given the benefit of group consensus.

I appreciate the nature of the course work and projects we have had to work on thus far in the class. I now know the value of effective search engines and databases and no longer take for granted the amount of work that has been involved in making them successful. In my professional work, I will be able to better understand the inner-workings of the systems I am asked to use for retrieval and dissemination of information for my library users. I am thankful to have worked in a group and appreciate having team members to collaborate with.

Shelby's Individual Reflection

This project really instilled a respect for people who create databases for a living. The creation of our controlled vocabulary required me to consider the language I use more carefully, and it impressed upon me the value of having a group to work with when creating databases. I often found myself lapsing into my own style of speaking when creating the vocabulary. For instance, I typically use the word patrons instead of users, however others tended to prefer users. Having the experiences of others allowed us to make our vocabulary more accessible to a wider variety of people.

Narrowing down the key terms for each article was also challenging, and I experienced particular difficulty balancing the six term limit and findability. Articles frequently had multiple important themes, and choosing only six descriptors limited precision.

The descriptor cap does have its positive aspects, though. As Weedman (2009) states, "Standards are constraints that both enable and limit design work" (p. 1504). By setting a rule that prohibits more than six descriptors, the chances that a given term may be overused are

reduced. For example, most of the articles included in our database could be tagged with the information science descriptor, however most articles had ideas that were more pressing than information science. This caused the information science descriptor to be eliminated for most articles. While this may seem like a small thing, allowing a descriptor, in this case information science, to be added to most or all records in a database renders the descriptor meaningless for the purposes of searching. Overall, the project helped me better appreciate the database creators at my library, and it afforded me a better understanding of how to use language effectively for information retrieval.

Jenelle's Individual Reflection

Working with my group has been fantastic. Everyone is very supportive and helps out in any way they can, which is good because I have learned that I never want to be an indexer or create a database on my own. I know that it is a great tool, however, I am going to stay far, far away from them.

I think that, with the articles we used, we each took away different things from the same articles. I think this is why we were able to come up with many different search terms for the same articles. The "aboutness" of the articles was different for each of us that read the same articles (Weedman, 2018). It's difficult to pinpoint directly what the aboutness of the articles is, because it differs depending on a person and their likes and interests (Weedman, 2018).

Creating a database means thinking outside the box and being open-minded about what different people may look for when searching. Where one person may find the three terms, another may find different terms. The findability of a database depends also on a person's interests and can be very different from another person's, however both could find what they

need out of the same article, depending on the search terms used (Morville, 2005).

I know when I have been doing my own searches on a topic for a research paper for another class, I have not yet hit the nail on the head with search terms. I have tried just about everything I can think of, and I know that there is information out there, I just haven't found that perfect combination of search terms to help me find what I need. It's possible to be too broad as well as too narrow when doing a search, but no matter what, I keep going and do my best searches possible.

References

Morville, P. (2005). *Ambient findability.* Sebastopol, CA: O'Reilly Media.

Rowley, J. & Hartley, R. (2016). *Organizing knowledge: An introduction to managing access to information* (4th ed.). New York, NY: Routledge.

Weedman, J. (2009). Design in the information sciences. In M. J. Bates, & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed., p. 1493). Boca Raton, FL: CRC Press. Retrieved from https://www-taylorfrancis-com.libaccess.sjlibrary.org/books/e/9780849397110

Weedman, J. (2018). Lecture 3 supplement: Subject metadata. In V. M. Tucker (Ed.), *Information retrieval system design: Principles & practice* (5.1 ed., pp. 141-142). San Jose State University.