**Assignment 2: Subject Analysis**

<table>
<thead>
<tr>
<th>Part A. Design Database</th>
<th>pagenumber</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1/A2. Postco Vocabulary (final version)</td>
<td>2</td>
</tr>
<tr>
<td>A3. Statement of Purpose</td>
<td>3</td>
</tr>
<tr>
<td>A4. Data Structure</td>
<td>4</td>
</tr>
<tr>
<td>A5. Rules</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part B. Create Content</th>
<th>pagenumber</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Records</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part C. Write Reflections</th>
<th>pagenumber</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. Reflections</td>
<td>15</td>
</tr>
</tbody>
</table>
Part A. Design Database

A1/A2. Postco Vocabulary (final version)

Abstracting
Accessibility
Archives
Automated indexing
Cataloging
Classification
Communication skills
Data longevity
Data organization
Database design
Digital records
Document contents
Document preservation
Document selection
Documentation
Folksonomy
Human technology interaction
Indexers
Indexing
Information access
Information professionals
Information representation
Information retrieval
Information retrieval system
Information science
Information seeking
Information sharing

Information technologies
Information visualization
Internet
Keywords
Librarians
Libraries
Lifelong learning
Mental models
Metadata
Museums
Online Search
Readability
Reference interview
Relevance
Search engines
Search methods
Social interactions
System performance
Teenagers
Text analysis
User assistance
User behavior
User interface
User satisfaction
User studies
User training
Web usability
A3. **Statement of Purpose**

With specialization in information retrieval subject, the database is designed to facilitate search of scholarly articles through core descriptive information and well-organized subject access.

**Target Users**

Primary target users are students who are currently attending San Jose State University for the Master Program of Library and Information Science, or people who graduated from similar programs. They are expected to be familiar with subject related technical terms, as well as Boolean, proximity and truncation searches. Target users are also expected to have basic understanding and related experiences of searching their queries in academic databases, such as Dr. Martin Luther King Jr. Library online catalog, and through web search engines, such as Google. Users of this database already have a solid foundation in the subject of information retrieval, and want to learn more about specific area(s) for personal use or academic research. For example, if users plan to conduct research on user behavior, they can search “user studies” to explore existing scholarly articles. Users may also search “information system design” to improve personal knowledge of IR system or increase search efficiency.

**Objectives**

The database includes seven fields: Record ID, Author, Title, Source, Year, Abstract and Subject. The first five fields offer an easy way to search a specific article. Additionally, natural language used in Title and Abstract fields, together with controlled vocabularies in Subject field, provides users diverse ways to explore articles within a specific subject.
The database aims to contain an extensive collection of bibliographical records for articles related to the subject of information retrieval. It will serve as a supplemental readings list for students in information retrieval related courses in San Jose State University as well as a well-developed information resource for users in general.

### A4. Data Structure

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Type</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>bibliographic_database_id</td>
<td>textbox</td>
<td>Check the most recently added bibliographic_database_id by clicking “Search/modify” on the data entry page. Enter the next whole number.</td>
</tr>
<tr>
<td>bibliographic_database_id</td>
<td>textbox</td>
<td>Check the most recently added bibliographic_database_id by clicking “Search/modify” on the data entry page. Enter the next whole number.</td>
</tr>
<tr>
<td>Record ID</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Author</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Title</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Source</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Year</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Abstract</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
<tr>
<td>Subject</td>
<td>textbox</td>
<td>Enter name of author with exact spelling as shown in articles. Name should be entered using the following format: Last name followed by a comma, first name followed by space, and middle name initial followed by a period (e.g.:</td>
</tr>
</tbody>
</table>
Author, Author A.). If author’s first name is only available in initial, enter as: Last name followed by a comma, first name initial followed by a period, middle name initial followed by a period (e.g.: Author, A.A.). Enter only applicable information. Enclose author’s full name in double-quotes. If an article has multiple authors, enter all authors’ names and separate each name entry by a comma. Do not include additional space in between words unless instructed to. If no author information available, enter: Not applicable.

Field Name: Title
Type: Comment
Rule: Enter the full title exactly as it is shown in articles. Include subtitles. Capitalize only the first letter of titles, subtitles and proper nouns. If no title is available, enter: Not applicable

Field Name: Source
Type: Comment
Rule: Source may include journal name/ book title, edition, volume number, issue number, publishing location, publisher, page number and URL. Enter only applicable information following APA format guide. Do not italicize text.

Field Name: Year
Type: Textbox-Numeric
Rule: Enter the year of the publication in four digits. Do not include month or date.

Field Name: Abstract
Type: Comment
Rule: Enter the abstract exactly as provided in articles. If no abstract is available, compose a paragraph that summarizes the article. If an abstract is more than one paragraph, separate paragraphs by an additional line of spacing.
Field Name: Subject
Type: Textbox

Rule: Determine and enter three to six descriptors to represent the major subject components of articles in alphabetical order. Each term should reflect a single concept. Keep in mind that descriptors should serve to aggregate and discriminate relevant records. Capitalize the first letter of each term and separate each by a comma. Do not include space in between values. If the term is countable noun, plural form should be used.

Part B. Create Content

B. Records

Link to database:
https://libr202.sjsu.edu/webdata_pro/student/242/cgi-bin/webdata_pro.pl

<table>
<thead>
<tr>
<th>Record ID</th>
<th>Author</th>
<th>Title</th>
<th>Source</th>
<th>Year</th>
<th>Abstract</th>
</tr>
</thead>
</table>
| 1         | “Bates, Marcia J.” | The invisible substrate of information science | Journal of the American Society for Information Science, 50, 1043-1050 | 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. The purpose of this article is to elucidate key elements of the below-the-water-line portion of the information science paradigm.

Particular emphasis is given to information science’s role as a meta-science—conducting research and developing theory around the documentary products of other disciplines and activities. The mental activities of the professional practice of the field are seen to center around representation and organization of information rather than knowing information. It is argued that such representation engages fundamentally different talents and skills from those required in other professions and |

Subject: Data organization, Information professionals, Information representation, Information science
<table>
<thead>
<tr>
<th>Page</th>
<th>Citation</th>
<th>Title</th>
<th>Journal</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>&quot;Rothenberg, Jeff&quot;</td>
<td>Ensuring the longevity of digital documents</td>
<td>Scientific American, 272(1), 42-47</td>
<td>1999</td>
</tr>
</tbody>
</table>

Discussion in the research community and among the general public regarding content indexing (especially subject indexing) and access to digital resources, especially on the Internet, has underutilized research on a variety of factors that are important in the design of such access mechanisms. Some of these factors and issues are reviewed and implications drawn for information system design in the era of electronic access. Specifically the following are discussed: Human factors: Subject searching vs. indexing, multiple terms of access, folk classification, basic-level terms, and folk access; Database factors: Bradford’s Law, vocabulary scalability, the Resnikoff-Dolby 30:1 Rule; Domain factors: Role of domain in indexing.

To prevent loss of information with historical and cultural value, information preservation has always been an important topic. Maintenance of physical media and digital record remains as two core aspects in digital content preservation. However, longevity of digital information is determined by a more diverse combination of elements. Context of the document is essential in understanding its meaning. Additionally, digital record’s software dependency requires preservation of appropriate technology to read and interpret the preserving data. To provide the bootstrap for future users, an easily readable annotation should also be included to instruct how to read the records and what is its significance.

Collaborative tagging describes the process by which many users add metadata in the form of keywords to shared content. Recently, collaborative tagging has grown in popularity on the web, on sites that allow users to tag bookmarks, photographs and other content. In this paper we analyze the structure of collaborative tagging systems as well as their dynamical aspects. Specifically, we discovered regularities in user activity, tag frequencies, kinds of tags used, bursts of popularity in bookmarking and a remarkable stability in the relative proportions of tags within a given url. We also present a dynamical model of collaborative tagging that predicts these stable patterns and relates them to imitation and shared knowledge.
<table>
<thead>
<tr>
<th></th>
<th>Authors</th>
<th>Title</th>
<th>Journal/Publication Details</th>
<th>Year</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>“Huston, Mary M.”</td>
<td>Windows into the search process: An inquiry into dimensions of online information retrieval</td>
<td>Online Review, 227-243</td>
<td>1991</td>
<td>From diverse users’ points of view, contextual frameworks are elaborated for the nature of the information technology, the information universe, and the information search. Within these conceptual parameters, established theories on search strategy are reviewed and cognitive models of information-seeking are highlighted. Future directions for research on users’ search processes are discussed in terms of the role for online retrieval in the future information environment.</td>
</tr>
<tr>
<td>6</td>
<td>“Agosto, Denise E.”, “HUGHES-HASSELL, Sandra”</td>
<td>People, places, and questions: An investigation of the everyday life information-seeking behaviors of urban young adults</td>
<td>Library &amp; Information Science Research, 27, 141-163</td>
<td>2005</td>
<td>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.</td>
</tr>
<tr>
<td>7</td>
<td>“Jansen, Bernard J.”, “Spin, Amanda”, “Koshman, Sherry”</td>
<td>Web searcher interaction with the Dogpile.com metasearch engine</td>
<td>Journal of the American Society for Information Science and Technology, 589, 744-755</td>
<td>2007</td>
<td>Metasearch engines are an intuitive method for improving the performance of Web search by increasing coverage, returning large numbers of results with a focus on relevance, and presenting alternative views of information needs. However, the use of metasearch engines in an operational environment is not well understood. In this study, we investigate the usage of Dogpile.com, a major Web metasearch engine, with the aim of discovering how Web searchers interact with metasearch engines. We report results examining 2,465,145 interactions from 534,507 users of Dogpile.com on May 6, 2005 and compare these results with findings from other Web searching studies. We collect data on geographical location of searchers, use of system feedback, content selection,</td>
</tr>
</tbody>
</table>
sessions, queries, and term usage. Findings show that Dogpile.com searchers are mainly from the USA (84% of searchers), use about 3 terms per query (mean 2.85), implement system feedback moderately (8.4% of users), and generally (56% of users) spend less than one minute interacting with the Web search engine. Overall, metasearchers seem to have higher degrees of interaction than searchers on non-metasearch engines, but their sessions are for a shorter period of time. These aspects of metasearching may be what define the differences from other forms of Web searching. We discuss the implications of our findings in relation to metasearch for Web searchers, search engines, and content providers.

This paper reports on part of a study of real users’ behavior in selecting documents from a list of citations resulting from a search of an information retrieval (IR) system. Document selection involves value judgments and decision making. Understanding how users evaluate documents and make decisions provides a basis for designing intelligent IR systems that can do a better job of predicting usefulness.

Twenty-five faculty and graduate students in an academic department participated in the study with requests presented to the department’s library. After a reference interview to establish the user’s profile (information needs, tasks at hand, expectations, and experience), a search was done on Dialog. From the resulting printouts of full citations participants selected the documents for which they wished to see the full text. Participants were asked to read aloud (information attended) and think aloud (whatever comes to their minds, decisions and their reasons).

Analysis of the verbal protocol data identified the criteria, the sources of information, and the patterns of document selection behavior. Among the factors identified are topical relevance, field relevance (field is broader than topic), novelty, expected usefulness (from previous experience of the author or journal), orientation, recency, availability, special requisites and time constraints. The design of an intelligent IR system requires an enormous knowledge base including not only the domain knowledge, but also the individual user’s preferences and document selection behavior. Such a system could

8

“Wang, Peiling”, “Soergel, Dagobert”

Beyond topical relevance: Document selection behavior of real users of IR systems


1993

Database design,Document selection,Information retrieval,Information retrieval system,User studies

“Wang, Peiling”, “Soergel, Dagobert”

Beyond topical relevance: Document selection behavior of real users of IR systems


1993

Database design,Document selection,Information retrieval,Information retrieval system,User studies
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
<th>Details</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Metadata for all: Descriptive standards and metadata sharing across libraries, archives, and museums</td>
<td>“Elings, Mary W.”; “Waibel, Gunter”</td>
<td>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.</td>
<td>2007</td>
</tr>
<tr>
<td>10</td>
<td>Probabilistic design principles for conventional and full-text retrieval systems</td>
<td>“Maron, M.E.”</td>
<td>In order for conventionally designed commercial document retrieval systems to perform perfectly, the following two (logical) conditions must be satisfied for every search: (1) There exists a document property (or combination of properties) that belongs to those (and only those) documents that are relevant. (2) That property (or combination of properties) can be correctly guessed by the searcher. In general, the first assumption is false, and the second is impossible to satisfy; hence no conventional IR system can perform at a maximum level of effectiveness. (We are painfully aware of the current poor performance values for Recall and Precision. Furthermore, Recall deteriorates rapidly as document corpora continue to grow in size.) However, different design principles can lead to improved performance. This article presents a view of the document retrieval problem that shows that since the relationship between document properties (whether they be humanly assigned index terms or words that occur in the running text) and relevance is at best probabilistic, one should approach the design problem using probabilistic principles. It turns out that a front end designed to permit searchers to attach probabilistically interpreted weights to their query terms could be adapted for conventional IR systems. Such an enhancement could lead to improved performance.</td>
<td>1988</td>
</tr>
<tr>
<td></td>
<td>Author(s)</td>
<td>Title</td>
<td>Journal/Book/Conference</td>
<td>Year</td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>12</td>
<td>“Hearst, Marti A.”</td>
<td>Interfacing for searching the Web</td>
<td>Scientific American, 276(3), 68-72</td>
<td>1997</td>
</tr>
</tbody>
</table>

To become life-long learners, students must acquire information retrieval skills for future as well as current information needs. This article describes how the Library Use Instruction Program at Northwest Missouri State University incorporates a heuristic model in which students continually evaluate and refine their information seeking practices while progressing through all levels of courses in diverse disciplines. Collegial partnerships with departmental faculty and ongoing instructional assessment are essential to the success of the program.

Data organization, Information visualization, Online search, Search engines, Text analysis, User interfaces

Communicating skills, Librarians, Libraries, Information seeking, Reference interview

To become life-long learners, students must acquire information retrieval skills for future as well as current information needs. This article describes how the Library Use Instruction Program at Northwest Missouri State University incorporates a heuristic model in which students continually evaluate and refine their information seeking practices while progressing through all levels of courses in diverse disciplines. Collegial partnerships with departmental faculty and ongoing instructional assessment are essential to the success of the program.

Data organization, Information visualization, Online search, Search engines, Text analysis, User interfaces

Communicating skills, Librarians, Libraries, Information seeking, Reference interview

This paper questions the ubiquitous practice of supplying minimalistic information to users, of making that information functional only, of assuming that the Shannon-Weaver communication model should govern online systems, and of ignoring the social implications of such a stance. Help systems that provide fast, temporary solutions without providing any background information lead to the danger of users completing tasks that they do not understand at all. (Word will help us write a legal pleading, even if we have no idea what one is.) As a result, we have help

Documenting, Informing, Visualization, User assistance, User interfaces, User training
systems that attempt to be invisible and to provide tool instruction but not conceptual instruction. Such a system presents itself as a neutral tool, but it is actually an incomplete environment, denying both the complexity and alternative modes of thinking about the subject at hand.

---

<table>
<thead>
<tr>
<th>Page</th>
<th>Author(s)</th>
<th>Title</th>
<th>Journal</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>“Marchini, Gary”</td>
<td>Interfaces for end-user information seeking</td>
<td>Journal of the American Society for Information Science</td>
<td>1992</td>
</tr>
<tr>
<td>16</td>
<td>“Lynch, Clifford”</td>
<td>Searching the Internet</td>
<td>Scientific American</td>
<td>1997</td>
</tr>
<tr>
<td>17</td>
<td>“Applegate, Rachel”</td>
<td>Models of user satisfaction: Understanding false positives</td>
<td>RQ</td>
<td>1993</td>
</tr>
</tbody>
</table>

Essential features of interfaces to support end-user information seeking are discussed and illustrated. Examples of interfaces to support the following basic information-seeking functions are presented: problem definition, source selection, problem articulation, examination of results, and information extraction. It is argued that present interfaces focus on problem articulation and examination of results functions, and research and development are needed to support the problem definition and information extraction functions. General recommendations for research on interfaces to support end-user information seeking include: attention to multimedia information sources, development of interfaces that integrate information-seeking functions, support for collaborative information seeking, use of multiple input/output devices in parallel, integration of advanced information retrieval techniques in systems for end users, and development of adaptable interfaces to meet individual difference and multicultural needs.

The Net is not a digital library. But if it is to continue to grow and thrive as a new means of communication, something very much like traditional library services will be needed to organize, access and preserve networked information. Even then, the Net will not resemble a traditional library, because its contents are more widely dispersed than a standard collection. Human indexing, classification and selection skills much be complemented by the computer-based approaches to automate the task of indexing and storing information. Only a synthesis of the two differing perspectives will help organize the information on the Web and allow this new medium to remain viable.

What does it mean to say that users are “satisfied” with online searches or other products of information retrieval systems? A review of research in library and information science, computer systems, marketing, and psychology reveals three models of user satisfaction. In the Material Satisfaction Model, product performance...
(e.g., recall and precision) determine whether a user’s stated question is answered (material satisfaction). In Emotional Satisfaction Model - Simple Path, users are happy or emotionally satisfied when their questions have been answered. In the Emotional Satisfaction Model - Multiple Path, users’ happiness depends not only on questions answered (material satisfaction) but also on factors such as setting and expectations. This last model allows an understanding of the phenomenon of “false positive” emotional satisfaction: users who are happy with bad searches. The choice of model has implications for user services and for research into user-system interaction.

| 18 | “Borgman, Christine L.” | The user’s mental model of an information retrieval system: An experiment on a prototype online catalog | International Journal of Man-Machine Studies, 24, 47-64 | 1986 |

Providing navigational aids to assist users in finding information in hypertext systems has been an ongoing research problem for well over a decade. Despite this, the incorporation of navigation aids into Web search tools has been slow. While search engines have become very efficient in producing high quality rankings, support for the navigational process is still far from satisfactory. To deal with this shortcoming of search tools, we have developed a site specific search and navigation engine that incorporates several recommended navigational aids into its novel user interface, based on the concept of a user trail. Herein, we report on a usability study whose aim was to ascertain whether adding semi-automated navigational aids to a search tool improves users’ experience when “surfing” the Web. The results we obtained from the study revealed that users...
of the navigation engine performed better in solving the question set posed than users of a conventional search engine. Moreover, users of the navigation engine provided more accurate answers in less time and with less clicks. Our results indicate that adding navigational aids to search tools will enhance Web usability and take us a step further towards resolving the problem of “getting lost in hyperspace”.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>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.</td>
<td>Information representation, Indexers, Abstracting, Classification, Indexing, Document contents</td>
<td></td>
</tr>
</tbody>
</table>
C. Reflections

Dividing the work was not so hard. We worked as a team for assignment 1 and already knew each other’s skill inventory. Yu Ting was responsible for SOP and formatting, and H. constructed database with WDP software. Other works, which include postco vocabulary list construction, rule development, and record indexing, were divided equally among us, and final products were reviewed and edited by both of us. As in the last assignment, this has been a pleasant and rewarding group work experience.

During postco term construction, we aimed to merge narrow terms into broader terms to ensure ability of aggregation. However, we went too far on generalizing, neglecting the need of discrimination. We had to go back to articles to re-select terms. We ensured at least one unique concept per article for the purpose of discrimination. Before finalizing, we re-assigned terms on the list to articles. This final step helped us evaluate effectiveness of each term and fine-tune the whole list.

We also found abstracts as a useful tool to determine the main subjects of articles. However, abstracts can also be misleading in some circumstances. Sometimes, abstracts serve more similar to an introductory overview, including concepts that are not necessarily core ideas throughout the entire article. In other cases, abstracts summarize articles with details, mixing major concepts with minor ideas. Thus, instead of grabbing terms directly from abstracts, we utilized it as hints for possible topics, and then compared with article content. By doing so, we were able to achieve a better determination.

Our current database structure enables search through basic document information and subjects. To allow more complex search queries and provide more accurate search
results, we suggest adding ISSN, DOI and document format fields to existing structure. Furthermore, provision of thesaurus will clarify preferred vocabularies and display relationships among terms. Classification may also serve to provide another form of access. These two additions will allow a more diverse search process and help users to browse information especially when information need is not clearly defined.

Due to limitation of WebData Pro functionality, we used textbox type for Subject field with a link to our supplemental controlled vocabulary list. However, we believe that the list type would be a better choice. List field provides a clear view of controlled vocabularies. Users will have no need to guess or go through the hassle of external links.

*Reflection by Yu Ting*/ Controlled vocabulary selection is one of the most difficult parts in this assignment. If a keyword reflects a concept that is only used in one specific article, it will lose the functionality for aggregating documents that share similar concepts. On the other hand, if a keyword covers a concept that is too broad, it fails to discriminate undesired information. This process teaches us the importance of the balance between a controlled vocabulary’s uniqueness and broadness, which directly influences findability of documents.

*Reflection by H.* Controlled vocabulary is an effective method of subject access in terms of aggregating articles with a similar subject and discriminating the ones with a particular subject. Utilization of controlled vocabulary is useful when the aboutness of desired articles is certain but exact keyword is unknown. Whereas, natural language search can yield better search results when users know what words might be presented in the records. When both subject accesses are provided to users, the optimal findability is possible.