12-13-2013

Formative Evaluation of Near-semantic Search Interfaces

Jim Hahn
University of Illinois at Urbana-Champaign

Chris Diaz
University of Iowa

Internet Reference Services Quarterly, Volume: 18, 3-4, (2013) pp.175-188, DOI: 10.1080/10875301.2013.856367

Comments

This study was made possible through the collaboration between colleagues and campus units. The authors wish to acknowledge the Research and Publication Committee of the University of Illinois at Urbana-Champaign Library and the University of Illinois Campus Research Board, which provided support for the completion of this research. We would also like to thank Yinan Zhang, for his development work and information retrieval expertise, as well as Jennie Archer, Amanda Hatland, and Rubayya Hoque, who assisted in the recruitment, observation, and interviewing of the usability test participants. Address correspondence to Jim Hahn, University of Illinois, Undergraduate Library, 1402 W. Gregory Dr, Urbana, IL 61801, USA. E-mail: jimhahn@illinois.edu

This is an Author's Accepted Manuscript of an article whose final and definitive form, the Version of Record, has been published in *Internet Reference Services Quarterly* [13 December 2013] [copyright Taylor & Francis], available online at: http://www.tandfonline.com/10.1080/10875301.2013.856367.

Formative evaluation of near-semantic search interfaces

Jim Hahn

Orientation Services & Environments, Undergraduate Library,

University of Illinois, Urbana, Illinois, USA

Chris Diaz

Scholarly Communications & Collections, University of Iowa, Iowa City, Iowa, USA

Abstract

This paper reports findings from a usability study of an experimental near-semantic search interface. We present user preferences for exploring semantic recommendations and model a new user interface for subject suggestions based on the results. The resulting catalog search engine (Deneb 2.0) is in public beta implementation, available at: http://dunatis.grainger.uiuc.edu/deneb-2

Keywords

Semantic web, data mining, usability

Introduction

Researchers at the University of Illinois at Urbana-Champaign conducted an evaluation of a new experimental search interface. As a formative study, results are used to shape a production implementation of a subject index that was vetted by user preferences and needs for subject results. The suggestions used in this experiment are derived from bibliographic subject data in the library catalog. This

paper presents a user-centered perspective on designing semantic web search interfaces[1].

The Deneb experiment explores the semantic use of bibliographic subject data in catalog searching. Both Deneb versions 1.0 and 2.0 present an alternative interface[2] to the widely implemented VuFind overlay (http://vufind.org/). For example, when a user does a search for "health care" in Deneb, the engine searches the subject field of each item record to arrange the facets according to the frequency of the term in entire library catalog, which bring up facets for semantically related terms, such as "Delivery of health care", "health care services," and "insurance coverage" that would otherwise be hidden. When the user does a search in the University of Illinois' installation of VuFind (http://vufind.carli.illinois.edu/), for instance, the facets are organized, from top to bottom, as such: Format, Location, Author, Topic, Subject Area, Language, Genre, Era, Region, and Title. While VuFind covers Topic and Subject Area as facets, these do not surface the semantic connections among terms that Deneb subject results expose.

Deneb is an experiment to streamline the facet arrangement, encourage the use of the catalog's subject terms, and evaluate the usability of such an arrangement from the user's perspective. Development of the Deneb 1.0 index is the result of

---

[1] The semantic web is defined in the *Encyclopedia of Library and Information Sciences* as "an extension, in progress, to the World Wide Web (WWW), designed to allow software processes, in particular artificial agents, as well as human readers, to acquire, share, and reason about information" (O'Hara & Hall, 2011, p4663).

[2] Deneb stands apart from the VuFind installation at Illinois; as an alternative layer, it is derived from a snapshot of catalog data, and has no direct connection to the ILS (Voyager). Through unique bibliographic identifiers all Deneb links point back to the VuFind catalog. Deneb 1.0 and 2.0 are powered by the open source search engine Apache Lucene (http://lucene.apache.org/).Deneb therefore could work with any ILS technology using unique bibliographic identifiers – e.g. most contemporary integrated library systems.

collaboration among library practitioners and information retrieval researchers. These disciplines have common foundations dating back to the founding of the online library catalog. In data mining, it is asserted that knowledge or semantic meaning can be derived from applying mathematical models to identify structure and patterns in datasets (Ince, 2013).

Yinan Zhang, a PhD candidate in the Department of Computer Science at Illinois, developed the Deneb 1.0 Catalog through original data mining research. Motivations for the Deneb experiment stemmed from a grant stream funded to produce subject suggestions based on bibliographic metadata. Researchers began by investigating VuFind and attending to how related subjects are suggested. The subject suggestions in the VuFind catalog are organized by facets and defined by the results of the search, rather than the total corpus of the library catalog. The Deneb 1.0 search engine was developed to provide subject recommendations based on the total corpus of subject metadata.

In order to provide broader context to a catalog query, implementation of the Deneb 1.0 subject index started by first cleaning up the library subject data[3]. This included the de-duplication and simplification of Library of Congress Subject Headings (LCSH), where compound subjects were given more streamlined entries. After the corpus of the Illinois library catalog was cleaned, the associations in their entirety were mined for their frequency of occurrence and weighted based on the total number of times these subjects occur in the entire catalog. Thus, when a search

[3] The research and development test dataset (pre-Deneb 1.0) was 33,372 titles consisting of 3,556 terms (Wang et al., 2013, p442). The test data set expanded over the time of our grant to encompass the entirety of the catalog's 13 million volumes in both Deneb 1.0 and 2.0.

is performed in the Deneb 1.0 index, the subject recommendation is based on total number of items in the catalog, rather than the limited number of query results; these are near-semantic[4] subject results to the catalog query. Research and development of the algorithm used in Deneb 1.0 are documented in *A Phrase Mining Framework for Recursive Construction of a Topical Hierarchy,* in the proceedings of the 19th International Conference on Knowledge Discovery and Data Mining (Wang et al., 2013).

This paper progresses next with a literature review on other work in bibliographic data analysis, semantic retrieval, and semantic interfaces. Next we detail the formative evaluation methods used in this study. Our results for the test of the Deneb interface follows. We conclude with a discussion on implications and design choices made as a result of the user feedback, and their applications within an academic library environment.

Literature Review

This section reviews literature to provide a context for our development of the Deneb 1.0 search engine and interface. It begins by looking at the current state of bibliographic subject data structures and proceeds with impetus for term-weighting systems for next generation interfaces and their affordances for subject recommendations. We also examine work on semantic retrieval as it applies to semantic search interfaces. This review provides context to the research problem of

[4] We use "near-semantic" in this paper to denote that the research and development process deviates from traditional semantic web experimentation, particularly in the aims and uses of publishing linked open data, which was not a research goal of our study.

designing recommendation interfaces that are useful and used by library patrons for research tasks.

MARC Subject Data

Looking at how subject metadata is assigned is instructive for understanding its importance. Research in this area includes Moen and Bernardino's work on MARC utilization. In order to undertake a study of actual use of MARC in records a test data set was constructed that had 419,657 MARC records. Moen and Bernardino found subject cataloging (650 subfield a) to be the most frequently occurring field with 602,362 occurrences in their dataset (Moen & Bernardino, 2003, p. 5, Table 9). OCLC research from a larger data set (http://experimental.worldcat.org/marcusage/) shows 281,926,809 occurrences of 650 subfield a. The total number of records in WorldCat with a 650 field is 113,967,827 out of a total of 292,447,199 MARC records, or roughly 39%, in April 2013 (http://experimental.worldcat.org/marcusage/650.html). These data points serve to underscore that semantic web development for libraries should rightly attend to subject-based classifications. Svenonius (2000) argues that:

> While it is possible to build semantic structures from scratch, doing so would take a long time—as long as it has taken to create the authority files, multilingual thesauri, and classificatory structures developed in the context of traditional bibliographic systems. It would make better economic sense to adapt traditional structures to this purpose, but, however it is done, it will have either to reference or to recreate the intellectual foundation of information organization. (p. 198)

The Deneb 1.0 experiment adapts traditional bibliographic data structures – subject description in particular – to the area semantic search.

Term weighting

Our decision to apply term weighting to subjects was influenced by previous research on library systems. For example, (Zhang, Smith, Twidale, & Gao, 2011) write that weighting systems over subject indexing and tagging "enable search algorithms to be more discriminating and browsing better oriented, and thus make it possible to provide more granular access to information" (p. 79). As an alternative to the current VuFind catalog, our goal is to give a subject hierarchy prominent role in the search process. This, we assert, will help the user better understand his or her research topic. In our test case, we study the preferences and user needs for the display or presentation layer of this access in subject metadata.

Semantic Search

The Deneb interface initially uses subject-based faceting as a way to explore semantic meaning. In a user study on how faceted search interfaces help understand user queries – Kules, & Capra find that "facets play an important role in the exploratory search process," (2009, p. 313), through an original eye tracking study the researchers found that "participants spent the most time looking at the result items returned in response to their queries," and that, "at some stages of the search process participants' interaction with the facets appeared to be as important (or more important) than the results themselves." (p. 320). This serves to underscore the design choice for targeted recommendations inside of hierarchical subject facets.

The paper *The Usability of Semantic Search Tools: a Review* offers an analysis of various interfaces to semantic search including a review of the different types of interaction observed in semantic search. Form-based systems are one mode the authors suggest:

> Form-based interfaces can side-step mapping issues by getting users to select terms from the lists of valid terms. They support user understanding of the domain by, quite literally, showing the user what is there, and what they can expect valid searches to look like. (Uren et al. 2007, p. 368).

Form-based systems offer a complementary approach to navigating an information universe of linked data. We propose navigating subject recommendations in tiled hierarchical views. Designing interfaces to semantic search is an important area in need of scholarly study since semantic organization of information may increase findability. In her article on this topic, (Hedden, 2008) writes, "One way semantic indexing is distinguished from traditional subject indexing of documents is that it focuses on concepts rather than the documents as a whole," (p. 40). In order to better understand from an empirical perspective the user needs for semantic indexing and interface requirements, we conducted a formative evaluation of the Deneb 1.0 interface that can be replicated as new semantic search engines emerge.

Methodology

This is a usability study comprised of original research data gathered from student interviews and observations. Student interviewing and observation occurred during the spring, 2013, school term. Test participants were recruited using a convenience sample of those users who were in the University of Illinois

Undergraduate Library. Investigators interviewed and observed each participant once.

There were a total of eight participant observations and interviews. Participants were given a $5 gift card for the local café as compensation. According to usability expert Jakob Nielsen, a majority of usability issues within a given tool can be uncovered after small numbers of participants complete the study (Nielsen, 1993). The chosen methodology would not definitively uncover all issues related to semantic search interfaces since eight study participants do not represent statistical significance by quantitative standards. The usability data do provide us with a deep qualitative data set to shape and improve the Deneb interface, as our Results section shows.

Research participants used the prototype subject suggestion interface (http://dunatis.grainger.uiuc.edu/deneb/) for evaluating results of near-semantic search. Investigators collected two sources of data, observations of how students interact with the interface and a debriefing interview after the observation portion of the study. See Appendix for observation log and debriefing interview questions.

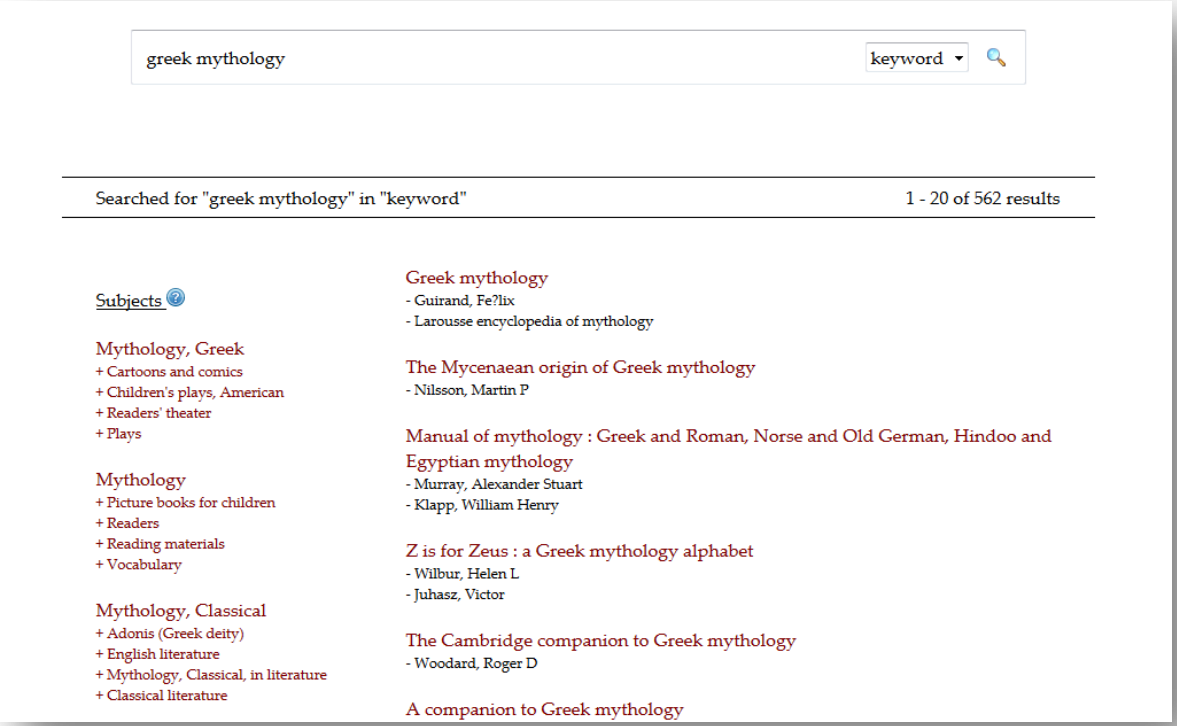

Figure 1. Results page from a keyword search of Deneb 1.0.

Results

The results presented are centered on the research objective of improving and shaping the Deneb 1.0 interface. This section is organized around the themes of improving semantic search catalogs including the subject suggestions in Deneb 1.0 and title results in Deneb 1.0.

Subject Suggestions

Formatting: Student test participants had trouble differentiating between title results and the subject suggestion facets. It was suggested that we develop boxes or a color scheme that makes it easy to distinguish between title and subject columns. Another student asked that we label subjects and results as different

sections. One student remarked, “I don’t know where to look first.” At the time of the study, the Deneb 1.0 interface did not have a heading for the title results – only subject recommendations were labeled.

Usefulness of subject suggestions: Overall, if students liked the subject recommendation, it was because they had performed a broader initial search. Study participants indicated they liked the recommendations and thought the subject suggestions were useful – unless the student was searching in a very specific area – in which case the subject suggestions did not help the student explore a topic. This is due to a familiarity with the topic area and a lack of a need for broader subject exploration. Student participants were interested in knowing the significance of subject recommendations ordering and rank. Students asked that the interface explain why the subject term is relevant to the search.

Confusion about the ordering of the subject terms occurred when the subject suggestions were irrelevant to the query. In one case, a student searched for “Red Blood Cells”, and the subject heading for the color “Red” appeared in the subject recommendation list.

Interactions with subject suggestions: Another theme with the subject suggestion column is concerned with the definitions of the subjects – users wanted to know the definitions of the words to understand whether they were going to be useful or not in exploring their topic. Students need to know what the suggested subjects mean and also suggested that we include tool tips so that they could hover over the subject to get a translation or definition of the subject. Another student asked that the tool tips perform a search in databases and let them know if items

covering this subject are available. Some students remarked that the subject suggestions were helpful in learning the vocabulary of the topic and the library catalog. One student said that subject recommendations helped her "search smarter."

Title Results

Most test users tended to look at the titles to understand more about their research subject. It may make sense to incorporate the subject data into the title results so that students could expand the topics attached to the titles if they are interested in learning more about what the subject encompasses.

Title Recommendations: Rather than solely being interested in title recommendations, the test participants asked if they could see, from the search result page, a list of similar books inside of the book search result or in the book information.

Interactions with title results: Test users asked for a set of features that would allow them to interact with the title results in the list, such as hovering over a title to get an abstract, summary, or table of contents of the item.

Title results design and style: Students preferred the search results of individual books to have similar styling to catalog results – which set each title apart by columns and also expected book cover images in the results. A common criticism of the Deneb interface is the lack of visual cues, such as item format icons, a color scheme, and clear separation between the title results and the subject recommendations, so that the users would not be overwhelmed with text.

Context: Students suggested that additional contextual information would be helpful. Information that gets pulled into the title results on related books – optionally doing additional searches for more books in a new tab.

Discussion and Conclusion

In this section we detail changes made to the Deneb 1.0 interface, preview the Deneb 2.0 interface, and detail unexpected findings of the study. We connect trends in our interface redesign with a discussion on exploratory search and its value in designing such systems. This paper concludes with a treatment on the future development of semantic interfaces for libraries.

Deneb 2.0

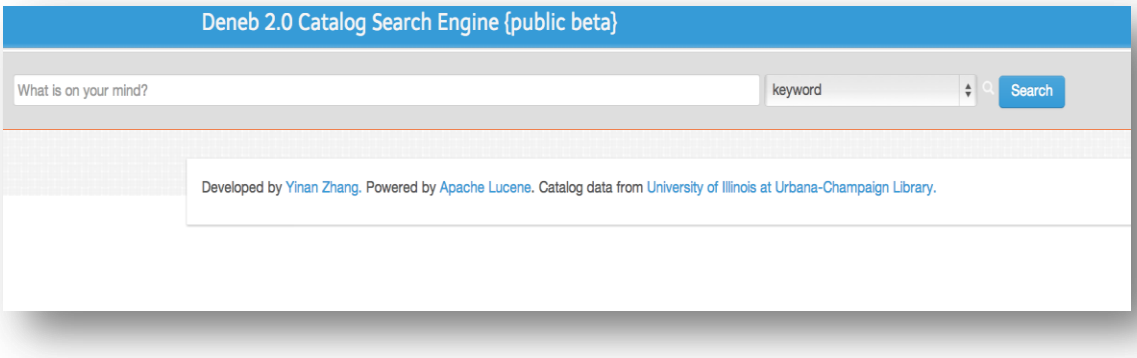

Figure 2. New Search Screen: Deneb 2.0 Catalog Search Engine.
Available at: *http://dunatis.grainger.uiuc.edu/deneb-2*.

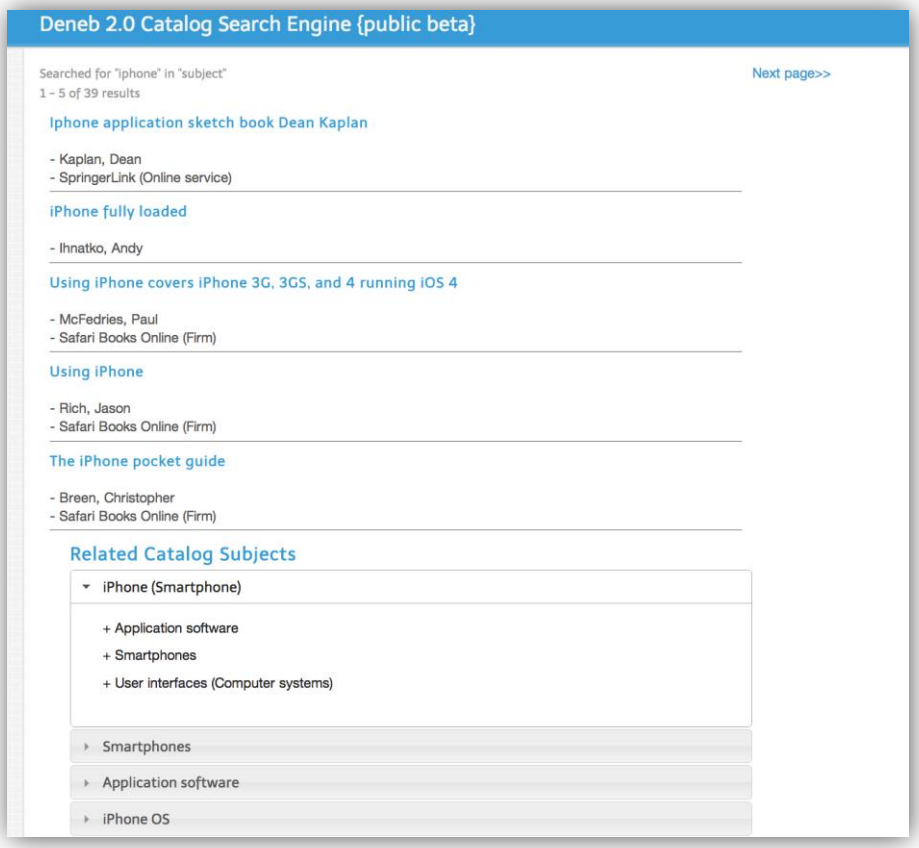

Figure 3. Deneb 2.0 Search Results

We implemented the second interface (Figure 2 & 3), to more closely resemble contemporary Web practices that would aid in search and discovery. The new interface is a “simple search results page” Bootstrap theme available from the BootWrap (https://wrapbootstrap.com) Website. After obtaining this template researchers implemented the title and subject suggestions inside of the simple search results framework. Many features that our students requested could be implemented here, particularly the critical interface problems of formatting, style, and drawing attention to the subject recommendations.

The Deneb 2.0 interface has design implications beyond the University of Illinois. For example, the near-semantic interface borrows from contemporary Web

design, particularly Bootstrap CSS (http://twitter.github.io/bootstrap/base-css.html) and JQuery UI (http://jqueryui.com) to implement a subject recommendation. The findings of this search interface study indicate that designers of semantic web systems have significant front-end design responsibilities that will help users understand and traverse linked data associations.

The redesign of the interface meets student desires for formatting and display by using a color scheme that differentiates the catalog title results from the subject catalog results. The color scheme for related catalog subjects is offset in grey and the subject results expand in an accordion view. This subject expansion performance may help the students first see the broad subject areas; then, if a detailed subject area is of interest, the student is able to expand the related subject results.

Traditional facet views are discarded in this public beta implementation. A number of participants in this study did not recognize the subject suggestion facet – an unexpected finding that led to the design of an integrated results view that shows both the first five title results and the weighted subject recommendation into one result view. This implementation is shown in the image below (Figure 3). This implementation uses contemporary Web design where suggestions, as advertising, are built into the search result lists. While facets are popular for narrowing, our findings indicate that topic facets are counter-productive in helping a user see the broader interconnection of resources through subject data.

Exploratory Search

Semantic search systems use the underlying data structures - in this case, bibliographic metadata - to produce results. Traditionally, this has been done in the form of facets or form-based interfaces. While this study provides feedback on the usability of a near-sematic search system, it is worth exploring the environment in which such systems will be used. This section examines related research on exploratory search and considers its applications in providing better internet services to an average library user with a broad-based topic in mind.

Semantic search interfaces present educational opportunities on the use of effective keywords and the discovery of related information. This has been discussed in current research on exploratory search, which "assumes that the user has some broader information needs that cannot be simply solved by a single 'relevant' Web page, but requires multiple iterations of search/ analysis interleaved with browsing and analyzing the retrieved information" (Ahn et al. p.383). To serve a user's information needs, exploratory search characterizes three search activities: lookup, learn, and investigate (Marchionini, 2006). White et al. (2007) offer descriptions of these activities:

- Lookup: carefully specified queries that retrieve precise results
- Learn: cognitive processing and the interpretation of new knowledge
- Investigate: critical assessment for integration into knowledge bases (p. 2878)

Our methods for studying the usability of the Deneb interface draw from the "goal of inducing an exploratory (rather than directed) style of search, wherein searchers will individually interpret the tasks, their relevance, and the results - while

maintaining some level of experimental control," as Kules & Capra (2012) describe exploratory search test design (p. 116). Reference staff in first and second year undergraduate settings are often assisting patrons in finding research materials on broad, sometimes thinly understood topics. One intended outcome of the effective use of the Deneb system is to augment the learning and investigative aspects of research.

A novice user of a library catalog is likely to employ exploratory search behavior when engaging the library catalog. This presents a teachable moment. By integrating subject suggestions within the search results, a semantically rich interface can seamlessly instruct the user to employ the bibliographic data structures, i.e. subject metadata, to enhance his or her understanding of a research topic and search more effectively. The Deneb interface presents a design alternative to the standard VuFind facets because this study's participants used the title column of results to refine their search strategies. This design perspective favors the semantic richness of the library catalog markup to improve library catalog searching.

To offer a pedagogical perspective, Marchionini (2006) situates exploratory search within the realm of learning objectives: "Using terminology from Bloom's taxonomy of educational objectives, searches that support learning aim to achieve: knowledge acquisition, comprehension of concepts or skills, interpretation of ideas" (p. 43). These learning objectives recall the ACRL's Information Literacy Competency Standards for Higher Education, particularly Standard Three, which states: "The information literate student evaluates information and its sources

critically and incorporates selected information into his or her knowledge base and value system" (Association of College and Research Libraries, 2004). The promotion of information literacy across library systems and services is encouraged to be a shared, central mission among all library staff. By leveraging semantic encoding through subject data mining, and implementing these systems through user-centered design, we can transform the library catalog into an instructional technology that works toward this shared mission.

Future of Semantic Search Interfaces

Facebook's Graph Search uses automatic query completion tools in order to help users see the connections that make up the nodes and arcs within linked data. Designers of Graph Search implemented easy to understand icons that help users understand which types of queries can be performed. The semantic auto-completion of queries is still quite new for most of the Web-going public, but semantic uptake and broader adoption of semantic queries and semantic tools will be greatly aided by interfaces that help users explore topics in novel ways. Implementing plain text-based search result views will not offer users a compelling reason to explore semantic web technologies.

At the same time we need to create links among data elements that both users and librarians agree to be useful – the LCSH are one such area. An example implementation of easy-to-use query completion tool that utilized controlled vocabularies is reported in the Code4Lib Journal article *Better search through query expansion using controlled vocabularies and Apache Solr* – the author shows how

query formation itself is used to assist users in understanding meanings of topics (Williams, 2013).

Exploring the semantic web is compelling when the performance of the presentation layer is done well. One way to go about making sure your library is poised to adequately deliver semantic web technologies and affordances is by forming the interfaces by way of ongoing iterative user studies – we provided one such example of how to iterate through a subject suggestion interface. Research into discovery layers and single search box capabilities underscore the need for ongoing evaluation of interface and search result design (Lown, Sierra, & Boyer, 2013). The future of the semantic web in libraries is one of iterative steps forward where both interface and data elements are reworked over time through sustained study.

References

Ahn, J., Brusilovsky, P., Grady, J., He, D., & Florian, R. (2010). Semantic annotation based exploratory search for information analysts. *Information Processing & Management* , 46(4), 383-402. http://dx.doi.org/10.1016/j.ipm.2010.02.001

Association of College and Research Libraries (2004). Information literacy competency standards for higher education, Retrieved from:

www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm

Hedden, H., (2008). How semantic tagging increases findability. *Econtent,* 31(8), 38–43.

Ince, D. (2013). Data mining. In dictionary of the Internet. : Oxford University Press. Retrieved from: http://www.oxfordreference.com/view/10.1093/acref/9780191744150.001.0001/acref-9780191744150-e-824

Kules, B., & Capra, R. (2012). Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal Of The American Society For Information Science & Technology*, 63(1), 114-138. doi:10.1002/asi.21647

Kules, B., Capra, R., Banta, M., Sierra, T. (2009). What do exploratory searchers look at in a faceted search interface? *Proceedings of the Joint Conference on Digital Libraries, June 15-19, 2009, Austin TX, USA, 313-322.*

Lown, C., Sierra, T., & Boyer, J. (2013). How users search the library from a single cearch box. *College & Research Libraries* , 74(3), 227-241.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM,* 49(4), 21-46.

Moen, W., & Benardino, P. (2003). Assessing metadata utilization: An analysis of MARC content designation use. *International Conference On Dublin Core And Metadata Applications,* 171-180. Retrieved from: http://dcpapers.dublincore.org/pubs/article/view/745

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

O'Hara, K. & Hall, W. (2011). Semantic web. In encyclopedia of library and Information sciences, Third Edition. Taylor and Francis: New York, 4663-4676.

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass: MIT Press.

Uren, V., Lei, Y., Lopez, V., Liu, H., Motta, E. Giordanino, M. (2007). The usability of semantic search tools: A review. *The Knowledge Engineering Review,* 22(4), 361-377.

Wang, C. Danilevsky, M., Desai, N. Zhang, Y. Nguyen, P. Taula T., Han, J (2013). A phrase mining framework for recursive construction of a topical hierarchy. *In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* August 11-14, Chicago, IL. DOI=10.1145/2487575.2487631 http://dl.acm.org/citation.cfm?doid=2487575.2487631

White, R.W., Hearst, M., Drucker, S.M., Schraefel, M.C., & Marchionini, G. (2007). Exploratory search and HCI: Designing and evaluating interfaces to support exploratory search interaction. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '07), ACM, New York, NY, USA, 2877-2880. DOI=10.1145/1240866.1241100 http://doi.acm.org/10.1145/1240866.1241100

Williams, S. (2013). Better search through query expansion using controlled vocabularies and Apache Solr. *Code4Lib Journal,* 20, Retrieved from http://journal.code4lib.org/articles/7787

Zhang, H., Smith, L.C., Twidale, M., Gao, F.H. (2011). Seeing the wood for the trees: Enhancing metadata subject elements with weights. *Information Technology and Libraries,* 30(2), 75-80.

Appendices

Search Interface - Interview Questions

How easy is the Website to use? What would make it easier to use?

What was hard to do with the Website?

What was confusing?

What was surprising?

What do you wish you could have done with the Website while you were using it?

How useful do you find the Website? What would make it more useful?

Would you recommend it to friends?

What do you actually want from a library search Website? Is there something else that should be here that is not here?

Is this Website a worthwhile tool for the library to develop?

Catalog Interface Study - Investigator Log (observations)

How easy to use is the Website?

What unexpected things occur?

How do students react when the Website does not work as they expect?

Do students make use of the subject recommendation features?

Note any additional observations of student use of the Website: