Information Retrieval in Domain-specific Databases: An Analysis to Improve the User Interface of the Alcohol Studies Database

Ronald Jantz

Academic libraries are becoming more directly involved in the design and publishing of electronic information resources, including bibliographic databases, electronic journals, and digital archives. As a result, librarians are dealing with many user interface design issues that computer scientists and information specialists in other fields have encountered. Transaction log analysis can provide a rich source of information on user behavior and insights as to how user interfaces can be improved. This article describes the methodology and results of the log analysis for the Alcohol Studies Database (ASDB), a domain-specific database supported by the Center of Alcohol Studies and Rutgers University Libraries (RUL). The goals of this study were to better understand user search behavior, to analyze failure rates, and to develop approaches for improving the user interface.

The ASDB content was developed by the Center of Alcohol Studies at Rutgers University, and the Web site and user interface were designed by the Scholarly Communication Center of RUL. An overview of the ASDB is provided as a backdrop for the in-depth analysis of the user interface and transaction logs. As part of the design to provide Web access to the ASDB, the author developed a statistical gathering and reporting subsystem. Activated in October 2000, this subsystem has since provided more than two full years of statistics on search behavior. As a result of the log analysis, specific improvements have been made to the ASDB user interface. A unique aspect of this article is the summary of the transaction logs before and after improvements to the user interface, which illustrates how the changes have affected search behavior and search results.

Introduction

Academic libraries are becoming more directly involved in the design and publishing of electronic information resources, including bibliographic databases, elec-

Ronald Jantz is the Government & Social Sciences Data Librarian in the Alexander Library at Rutgers University; e-mail: rjantz@rci.rutgers.edu. The task of providing Web access to the Alcohol Studies Database has been a collaborative effort. Penny Page and Valerie Mead, librarians at the Center of Alcohol Studies, have developed the content and indexing for the database. The author designed the original architecture and Web-based user interface. Mike LeBlanc, computer science student at Rutgers University, developed the improved user interface under the author's direction. This team, as a whole, has participated in numerous discussions on how to improve the ASDB user interface and in the testing of the resulting improvements.

tronic journals, and digital archives. These new roles represent a challenging future for librarians who want to utilize their technology and design skills.[1] The Scholarly Communication Center of Rutgers University Libraries (RUL) and the Center of Alcohol Studies (CAS) at Rutgers have collaborated to provide Web access to the Alcohol Studies Database (ASDB). The ASDB contains more than 60,000 citations of documents indexed by the CAS since 1987.[2] The primary focus of the database is on research and professional materials dealing with beverage alcohol and its use and related consequences. Although a growing amount of literature on other drug use/abuse has been added in recent years, this material represents only a small percentage of the database and is not indexed to the same depth as the alcohol literature. In addition to the research and professional material, the database includes a small collection of educational and prevention materials, including audiovisuals suitable for students and educators K–12, parents, community workers, and the general public.

At the outset, the author, working with librarians at the CAS, wanted to quickly develop a Web-based user interface for the ASDB. Originally, as a nonnetworked database, a controlled vocabulary of indexing terms had been developed and each article was extensively indexed with these terms. This vocabulary became the key component of the online search interface. The initial Web-accessible database and user interface were completed late in 1999 using the approach and technology described in an article by the author in 2001.[3] As a result, users at Rutgers University and throughout the world gained access to this important and freely available collection of medical and scientific research dealing with the use of alcohol and the related consequences. Subsequent to this introduction on the Web, the author designed a statistical gathering and reporting subsystem that was implemented in October 2000. The transaction logs now contain more than two full years of search statistics that have

assisted researchers in making decisions about how to improve the user interface. Based on the transaction log analysis and extensive ASDB team discussions, an improved user interface was launched in February 2002. This article summarizes the data from the transaction logs and compares search results from the initial user interface and the improved user interface.
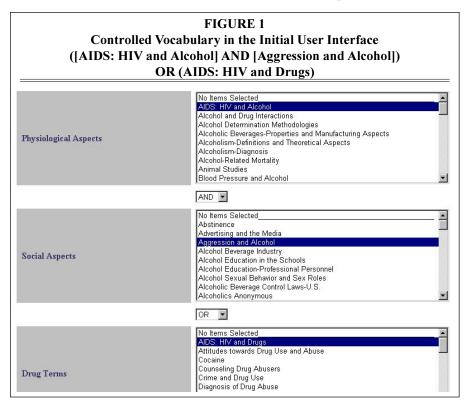
User Interface

Overview

A partial image of the initial user interface showing the controlled vocabulary pick-lists is shown in figure 1. This partial image shows three primary subject-related pick-lists labeled as follows: physiological aspects, social aspects, and drug terms. Each pick-list has some thirty or more controlled vocabulary terms that can be selected by the user to form a query. In addition to these lists, a user can select items from two additional pick-lists, format and special populations (not shown), that will further constrain the query. Finally, author and title word or phrase searching also is available. Online help instructions are available on the top navigation bar and "example" links to screen images are provided for each type of search box to demonstrate clearly how one would specify a query.

Complex Queries

The user can form simple or quite complex Boolean queries with the ASDB interface. For example, one could simply do a search for a specific author or a search on a word or phrase that might be found in the title of an article. However, the user also has the capability to form complex Boolean operations by selecting multiple items from any one of the three primary pick-lists. Multiple items selected within a pick-list default to a Boolean "or" and the user also can use the toggle switch between the major pick-lists to select either an "and" or an "or" between these major categories. The default Boolean operation between search boxes is an "and." The example in figure 1 illustrates a more complex Boolean search with the

---

FIGURE 1
Controlled Vocabulary in the Initial User Interface
([AIDS: HIV and Alcohol] AND [Aggression and Alcohol])
OR (AIDS: HIV and Drugs)

Physiological Aspects	No Items Selected
	AIDS: HIV and Alcohol
	Alcohol and Drug Interactions
	Alcohol Determination Methodologies
	Alcoholic Beverages-Properties and Manufacturing Aspects
	Alcoholism-Definitions and Theoretical Aspects
	Alcoholism-Diagnosis
	Alcohol-Related Mortality
	Animal Studies
	Blood Pressure and Alcohol

AND

Social Aspects	No Items Selected
	Abstinence
	Advertising and the Media
	Aggression and Alcohol
	Alcohol Beverage Industry
	Alcohol Education in the Schools
	Alcohol Education-Professional Personnel
	Alcohol Sexual Behavior and Sex Roles
	Alcoholic Beverage Control Laws-U.S.
	Alcoholics Anonymous

OR

Drug Terms	No Items Selected
	AIDS: HIV and Drugs
	Attitudes towards Drug Use and Abuse
	Cocaine
	Counseling Drug Abusers
	Crime and Drug Use
	Diagnosis of Drug Abuse

---

properly parenthesized result shown at the top of the figure. After forming a query, the user can then select "search" at the bottom of the screen, which will yield a set of summary results, each of which can then be selected to view the full bibliographic citation.

Results Display

After a user selects the "search" button, each resulting bibliographic record is displayed in summary form, ordered by publication date with the most recent first. Within publication year, there is a secondary ordering by author. Note that relevance orderings are not appropriate because there are no abstracts or full-text content that can be used to make relevance decisions.

Approach and Methodology

The ASDB is a domain-specific database that contains bibliographic records of more than 60,000 citations primarily to journal articles and books relating to the beverage alcohol and its use and related consequences. The use of transaction logs is one primary method of improving user interfaces and, thereby, also improving the information retrieval performance for users. Transaction logs have been used successfully to improve user interfaces of traditional OPACs in libraries.[4] This article discusses the use of transaction logs to improve the user interface for the ASDB. The logs analyzed herein represent usage from October 2000 through September 2001 for the initial user interface and usage from February 2002 through April 2002 for the improved user interface. The ASDB is a research-oriented database and, by Web standards, it is not heavily used; however, the transaction log contains a significant statistical representation and usage continues to grow as more people discover the availability of the ASDB. At the writing of this article, the author and colleagues were seeing between 1,300 and 1,500 searches a month during the standard academic fall and spring semesters.

The objectives of this analysis were to understand user behavior, analyze failure rates, and identify improvement areas for the user interface. The analysis methodology used in this article is similar to that described by Jansen and Spink.[5] Although many improvement areas were discovered through the analysis, a specific objective was to reduce the number of searches that resulted in either zero hits or greater than 100 hits. These types of outcomes were considered potential failures of the user interface.

Based on identified improvement areas, an improved user interface was launched in February 2002. Data from the initial and the improved user interfaces are compared to determine how the changes have improved the ability of searchers to use the ASDB. The following levels of analysis as reported by Jansen and Spink will be used.

Session

The session is the entire sequence of queries entered by the user. Heuristics will be used to define a session because the ASDB does not employ any type of "log-in" scenario that would accurately register each user. For the purposes of this article, the session will be defined as those queries submitted consecutively by a single IP address and not separated by more than twenty minutes. The twenty-minute interval was arrived at by visual inspection of the intervals that occur in the transaction log. Although it is conceivable that another user may have started another session with the same IP address and within a twenty-minute time frame, this condition is highly unlikely. It should be noted that a session can have a single query.

Query

Sessions are composed of queries. A query within the context of the ASDB is defined when a user selects the "search" button and an entry is written into the transaction log. For the purposes of this article, the concepts of initial query and modified query will be used. The *initial query* is the first query in a session, and the

modified query is a subsequent query in a session that is different from the initial query. Query length is measured by the number of terms used, and query complexity is determined by the use or absence of Boolean expressions.

Term

Within the ASDB, a term is defined as any controlled vocabulary term that is selected from one of the five pick-lists in the user interface (i.e., physiological aspects, social aspects, drug aspects, special populations, and special format). A term also can be an author's name or words/phrases entered into the "title phrase" search box and which might be separated by the Boolean operators of AND/OR.

The Statistical Gathering and Reporting Subsystem

The author designed the statistical subsystem to capture as many data as possible about the user search behavior. Every aspect of the user query is captured, including search terms and how the user has toggled the AND/OR selection between major subject areas in order to create a Boolean expression. Each search is associated with a unique user identification, although users always remain anonymous. In addition, the results of each search are recorded, including the number of results generated and a time-date stamp. Because users do not register to search the ASDB, some mechanism was needed to identify a user session. The time-date stamp in conjunction with the IP address is used to track the concept of a "session" as discussed above. It should be noted that the statistical system only records data from users who conduct a search of the ASDB. Any data regarding users who are just visiting and who do not conduct a search, sometimes referred to as "tourists," is not recorded.[6]

The reporting subsystem provides several types of summary reports, including total number of searches, searches with zero results, searches with more than 100 results, searches by month, and searches by major domain. In addition, the database administrator can select a more de-

tailed report to see all the fields and options for a particular search.

Transaction Log Summary Statistics
User Demographics
Table 1 shows the demographics by domain for the users of the ASDB. Although usage is predominantly from the United States, users are coming to the ASDB from all over the world.

Zero-hit Outcomes
Many studies have analyzed user difficulties with the syntax and semantics of Web searching. One paper has reported that more than 30 percent of searches of a university Web site resulted in zero-hit

TABLE 1
User Demographics for the ASDB

Domain	Country	Percent
edu	USA	26.9
com	USA	18.2
net	USA	16.0
us	USA	3.1
ca	Canada	2.6
au	Australia	2.3
org	USA	1.6
uk	United Kingdom	1.5
nz	New Zealand	0.4
jp	Japan	0.3
se	Sweden	0.3
nl	Netherlands	0.3
mil	USA	0.3
at	Austria	0.3
ie	Ireland	0.2
il	Israel	0.2
gov	USA	0.2
it	Italy	0.2
no	Norway	0.2
gr	Greece	0.1
be	Belgium	0.1
es	Spain	0.1
dk	Denmark	0.1
br	Brazil	0.1
fr	France	0.1
mx	Mexico	0.1
my	Malaysia	0.1
za	South Africa	0.1
All others		24.0

outcomes.[7] In an earlier paper, T. Peters reported that 40 percent zero-hit outcomes are common in his specific academic library OPAC.[8] Table 2 shows the distribution of hits in four different ranges, including zero-hit outcomes. The table indicates that the initial user interface of the ASDB is incurring 33.6 percent zero-hit outcomes where N = 10,267 is the total number of searches. The improved user interface has a marked reduction in zero-hit outcomes at 27.8 percent.

Sessions and Queries
Table 3 provides summary-level statistics for sessions and queries for both the initial (N = 10,267) and improved user interfaces (N = 3,375). From these summary statistics, it is obvious that the sessions are relatively short (e.g., 2.45 queries in the initial UI). From an examination of session length, it is apparent that 71.1 percent of all sessions in the initial UI have either one or two queries and 80.6 percent have one or two queries in the improved user interface. In other analyses, researchers have found similar results, speculating that users are either unwilling or unable to expend the effort to develop effective search strategies.[9]

Analysis: Zero-hit Outcomes
The zero-hit outcomes are a fruitful area for examination and will generally reveal a wealth of information regarding the effectiveness of a user interface. This analysis will proceed by examining the zero-hit outcomes of the initial user interface in more detail. Table 2 shows that 33.6 percent of the searches (3,454 out of 10,267) using the initial user interface resulted in zero hits. In the improved user interface, zero-hit outcomes have been significantly reduced to 27.8 percent. Of the zero-hit outcomes in the initial user interface (N = 3,454), 595 searches attempted some type of author search.

Author Searching
There were obvious syntactical and semantic errors with author searching. Generally, the semantic errors will be more difficult

TABLE 2
Overall Search Outcomes

Measure	Initial UI (N = 10,267)	Improved UI (N = 3,375)
Zero-Hit Outcomes (%)	33.6	27.8
Outcomes - GE 1, LT 100 Hits (%)	33.9	34.2
Outcomes - GE 100, LT 1,000 Hits (%)	32.5	20.7
Outcomes - GE 1,000 Hits (%)	9.4	17.3

to detect and correct. For example, there were a few users who confused the author search field with a keyword search field and searched on phrases such as "Advocacy 1992" or used a term that was obviously subject related rather than an author. These errors were uncovered by visual inspection of the logs and are reported in table 4 as "incorrect AU semantics." In addition, a number of author search syntax errors were evident in the transaction log that could clearly be eliminated or minimized by improving the user interface. The following illustrate some specific examples that do not follow the conventions that are documented as part of the ASDB user interface:

1. typing in the first name "first" (e.g., "Bill Wilson");

2. typing initials with no blanks (e.g., "Epstein, J. A.");

3. omitting comma delimiters that separate the last name from the first initial (e.g., "borg s").

Errors of this type comprise a significant percentage (32.6%) of the author search zero-hit outcomes as shown in table 4 as "incorrect AU Syntax." In the improved user interface, more restrictive syntax checking was implemented with a request to the user to reformulate the author search according to the required syntax conventions for author searching. Although some types of incorrect syntax still have not been detected, the improved user interface has significantly reduced the zero-hits due to incorrect author syntax to only 12.8 percent (table 4).

Title Searching

In the initial ASDB user interface, phrase searching was implemented; however, quoted phrases and the use of an asterisk as a truncation symbol were not permitted. Given the prevalence of these conventions in existing Web search engines, many ASDB users tried using these symbols. Of the zero-hit outcomes (N = 3,454), 2,343 searches attempted some type of title search and 3.5 percent used special conventions that were not supported by the ASDB. (See "incorrect Title Syntax" in table 4.) In the improved user interface, both quoted phrases and use of the asterisk are flagged with a message to the user to reformulate the query using approved syntax conventions. This checking has virtually eliminated zero-hits due to the use of these conventions.

TABLE 3
Session and Query Summary Statistics

Measure	Initial UI (N = 10,267)	Improved UI (N = 3,375)
Mean Queries per Session	2.45	1.93
Session Length		
Minimum	1	1
Maximum	44	54
% 1 Query	48.3	60.7
% 2 Queries	22.8	19.9
% 3+ Queries	28.9	19.4

TABLE 4
Incorrect Author and Title Searches

Measure	Initial UI	Improved UI
Incorrect AU Syntax (%)	32.6 (N = 595)	12.8 (N = 180)
Incorrect AU Semantics (%)	9.4 (N = 595)	12.2 (N =180)
Incorrect Title Syntax (%)	3.5 (N = 2343)	0%

Keyword Searching

Perhaps one of the most confusing parts of the user interface was allowing keyword and phrase searching *only* in the title field. The rationale behind this decision was the assumption that users would use the controlled vocabulary in lieu of general keyword/phrase searching and some users would want to search the title field only. However, it is clear from a visual inspection of the title phrase searching that users continue to use the title phrase search as a general keyword search box. One simple example illustrating that users are not obtaining the proper results is evident in the search using the phrase "military and alcohol," which returned eighteen results in the initial UI. For the improved user interface, keyword searching has been offered across all fields in the database and the title-specific search option eliminated. If one reruns the search "military and alcohol" in the improved UI, 130 results are returned. It is probably a reasonable assumption that the user in this case did not want those citations in which the terms "military" and "alcohol" appeared only in the titles of the citations. Hence, the change in keyword searching for ASDB has likely improved recall for the great majority of users.

Analysis: Controlled Vocabulary
Frequency of Use

In the ASDB, there are more than 150 controlled vocabulary terms across the three primary areas of physiological aspects, social aspects, and drug terms. Table 5 shows the fifty most frequently selected terms from the 10, 267 searches using the initial user interface. Although the data shown in table 5 are not used quantitatively in this study, the qualitative assessment was that use of this highly technical vocabulary could not be abandoned, thereby leaving users with only a free-text searching capability. Many users of the initial and improved user interfaces took advantage of the controlled vocabulary; however, some unexpected results were encountered, as discussed in the following sections.

Use of Subject Terms

As shown in table 6, data have been collected and summarized for the initial and the improved user interfaces that illustrate the percentage of users who did not use the controlled vocabulary. The "No Subjects" row includes the percentage of users who did not use any of the controlled vocabulary from the three main areas of physiological aspects, social aspects, and drug aspects; however, they may have used some of the other special pick-lists or the free-text search. The "Free Text Only" row shows the percentage of users who did not use the subject vocabulary and other special pick-lists such as those for population and format. These queries (23.6% in the initial UI and 39.3% in the improved UI) used only the free-text searching fields. It should be noted that the two rows in table 6 are not mutually exclusive. For example, a free-text-only search would be counted in both rows; hence, summing to greater than 100 percent in a column is possible.

It is obvious from table 6 that a significant number of users do not use the controlled vocabulary in either the initial or improved user interfaces. However, the dramatic result is the increased number of users who did not use the controlled vocabulary in the improved user interface. The author suggests that this result stems from two major changes in the user interface as we moved from the format of the initial UI to that of the improved UI. First, general keyword searching was introduced in contrast to only allowing keyword searching in the title field. In all like-

TABLE 5
Controlled Vocabulary: Frequency of Term Selection

Term	Freq	Term	Freq
Alcohol and Drug Interactions	408	Crime and Alcohol	106
Alcoholism: Diagnosis	347	Social and Cultural Aspects of	
Alcoholism: Etiology, Definitions,		Alcohol: Miscellaneous	104
and Theoretical	280	Crime and Drug Use	102
Alcohol-Related Mortality	237	Counseling Drug Abusers	101
Aggression and Alcohol	203	Blood Pressure and Alcohol	98
Alcohol Determination		Alcoholism Treatment Programs	
Methodologies	165	and Facilities	96
Attitudes toward Drinking and		Alcoholism Treatment:	
Alcoholism	163	Miscellaneous	92
Alcoholics Anonymous	161	Drug Abuse Treatment Programs	
Family Aspects and Alcohol	155	and Facilities	90
Advertising and the Media	154	Driving and Drinking: Management	
AIDS, HIV, and Alcohol	153	of Offenders	88
Drinking Experiments	149	Intoxication and Alcohol Poisoning	81
Stress and Alcohol: Physiological		Driving Skill and Alcohol	80
Aspects	149	Memory and Alcohol	75
Brain Pathology and Alcohol	136	Withdrawal and Post-alcohol	
Alcohol Education in the Schools	136	Phenomena	74
Alcoholism: Miscellaneous	136	Cocaine	72
Drug Abuse Treatment:		Heredity: Human Studies	69
Miscellaneous	130	Historical Aspects	65
Alcoholic Beverage Control Laws:		Recidivism and Relapse in	
U.S.	119	Alcoholism Treatment	63
Alcoholic Beverages: Properties,		Statistics: Alcoholic Beverages	61
Manufacturing Aspects	118	Treatment Outcome Studies	61
Attitudes toward Drug Use and		Stress and Alcohol: Psycho-social	
Abuse	118	Aspects	60
Abstinence	117	AIDS, HIV, and Drugs	60
Fetus and Alcohol: Human Studies	116	Diagnosis, Drug Abuse	60
Sexual Behavior, Sex Roles, and		Individual Therapies	59
Alcohol	115	Alcohol Education: Professional	
Detoxication and Treatment of		Personnel	57
Withdrawal	113	Cognitive and Perceptual Functions	57
Alcohol Beverage Industry	107		

lihood, users were more inclined to use the more familiar keyword searching rather than the less familiar method of selecting items from the controlled vocabulary list. Second, it was known that there would be trade-offs in the presentation styles of the two user interfaces. In the initial UI (figure 1), pick-lists were chosen in which the user could only see a small subset of the terms without scrolling. In the improved UI, the user is first presented with major subject areas (figure 2) and then must do a mouse click to see the terms of the controlled vocabulary in a checkbox format (figure 3). The advantage of this approach is that the user can see all the subject terms whereas she or he could only see a limited subject list (without scrolling) in the initial UI. Thus, the improved UI approach requires more mouse clicks for the user to select the subject terms. Perhaps more important, in the
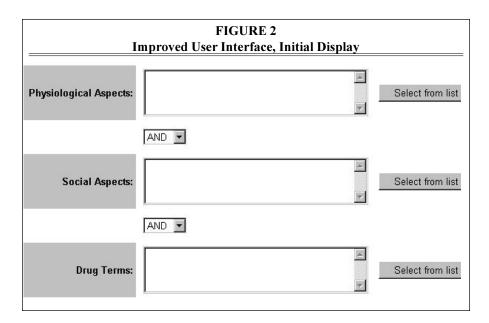
TABLE 6
Percentage Of Users Who Did Not Use the Controlled Vocabulary

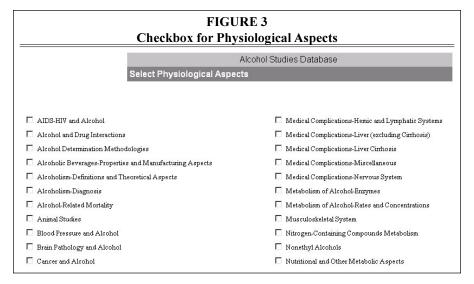| Measure	Initial UI (N = 10,267)	Improved UI (N = 3,375)
No Subjects (% of queries)	33.7	62.6
Free Text Only (% of queries)	23.6	39.3

improved UI, users do not see any terms on the first search page and it is suspected that they more naturally gravitated to the use of the obvious keyword searching capability rather than take the time to explore the subject terms available.

Summary

The improved UI has significantly reduced the number of zero-hits that users incur from 33.6 percent to 27.8 percent. This result is due primarily to the improved error checking for author searching and the checking for special syntax conventions that users might have seen on the Web, but which are not available in the ASDB. However, when one examines the distribution of outcomes with non-zero hits, there are 41.9 percent with greater than 100 hits in the initial UI and 48.0 percent with greater than 100 hits in the improved UI. One phenomenon that

is occurring in the improved UI is that users are selecting many more controlled vocabulary terms to OR together, which is resulting in searches with more hits. In all probability, this search behavior stems from users being able to see the complete selection of controlled vocabulary terms in the checkbox format. It is difficult to put a value judgment on these outcomes, although it is unlikely that users are examining results beyond their first 100 hits.

With the change in subject term display format from a pick-list to checkbox style, users have dramatically reduced the use of the ASDB controlled vocabulary from 33.7 percent not using any of the controlled vocabulary terms to 62.6 percent. This result was clearly unexpected by the author and the CAS librarians and not an altogether desired result. Although there is less usage of the controlled vocabulary, users who use the keyword

FIGURE 2
Improved User Interface, Initial Display

FIGURE 3
Checkbox for Physiological Aspects

Alcohol Studies Database

Select Physiological Aspects

☐ AIDS-HIV and Alcohol	☐ Medical Complications-Hemic and Lymphatic Systems
☐ Alcohol and Drug Interactions	☐ Medical Complications-Liver (excluding Cirrhosis)
☐ Alcohol Determination Methodologies	☐ Medical Complications-Liver Cirrhosis
☐ Alcoholic Beverages-Properties and Manufacturing Aspects	☐ Medical Complications-Miscellaneous
☐ Alcoholism-Definitions and Theoretical Aspects	☐ Medical Complications-Nervous System
☐ Alcoholism-Diagnosis	☐ Metabolism of Alcohol-Enzymes
☐ Alcohol-Related Mortality	☐ Metabolism of Alcohol-Rates and Concentrations
☐ Animal Studies	☐ Musculoskeletal System
☐ Blood Pressure and Alcohol	☐ Nitrogen-Containing Compounds Metabolism
☐ Brain Pathology and Alcohol	☐ Nonethyl Alcohols
☐ Cancer and Alcohol	☐ Nutritional and Other Metabolic Aspects

searching also are searching all controlled vocabulary terms. Although this approach is yielding many relevant results, we are still struggling with the classic information retrieval problem of the difference between the user vocabulary and the indexer's vocabulary.[10] The other observation here is that user overhead in terms of mouse clicks appears to be more of an issue than originally expected. It appears that initial impressions from the first search page had a very strong impact on user behavior leading them to keywords when they did not see any controlled vocabulary on the first page. Although only one mouse click away, the controlled vocabulary has become more inaccessible for a great many users.

A related phenomenon is that of session length between the two user interfaces. In the initial UI, the percent of single query sessions was 48.3 percent whereas this statistic jumped to 60.7 percent in the improved UI. There are possibly two factors that could account for this behavior. First, the reduced number of zero-hit outcomes in the improved UI suggests that more users have received appropriate results in a single-session query. The other factor is that many users may have considered the improved UI more complex than the initial UI and thus did not continue to explore how to use the ASDB effectively.

In addition to the major user interface changes made above, it should be noted that the improved UI of the ASDB also includes the ability to e-mail citations, to select a "print-friendly" interface, and to page results in order to limit the size of the html page returned to the user's browser. Steps also are under way to implement the linking to the full text of journals that RUL has licensed.

Conclusions

For most Web database projects, it is unclear who the user community will be and frequently all the designer can assume is that users are all those "out there" on the Web. With the ASDB, the domain demographics suggest a very diverse user community. It is unlikely that one will have the luxury of understanding the user's search behavior prior to developing the user interface. Therefore, it is very important to capture usage statistics via transaction logs. These logs enable the designer to learn more about the user and to make incremental changes to improve the user interface.

Users will frequently make assumptions about the user interface syntax given their experience with Web search engines or other database products. They frequently assume these conventions are standard and universal. In designing user

interfaces, the librarian must either support a variety of conventions or provide error checking and feedback to assist the user in learning the syntax of the specific search engine. Rarely will one get the user interface "right" on the first iteration. The designer should plan on making improvements after the transactions logs have been reviewed and there is more information on the types of users and their search behavior.

With respect to the specific ASDB analysis, a heuristic of zero-hits and greater than 100 hits has been used as an indicator that the user interface can be improved. Although certain of the searches that fall under this general classification are legitimate, this indicator can serve as a useful, low-cost tool for identifying and making user interface improvements. Certainly the reduction of zero-hit outcomes in the ASDB is an improvement. However, it appears that the improved user interface might be more complex for our user community given the decrease in usage of the controlled vocabulary, the increase in single-query sessions, and significantly more outcomes with greater than 1000 hits. This conclusion suggests that there are some possible future improvements that would help users. The

researchers suspect that many of the users are casual information searchers who are accustomed to a basic keyword search interface and that the professional information specialists would find the controlled vocabulary most useful. Thus, offering a "basic" and "advanced" user interface is likely to help considerably in meeting the needs of two quite different user populations. However, with the basic keyword search, the researchers are still left with the problem of bridging the user vocabulary and the controlled vocabulary. In these small specialized databases, linking the keyword search terms with the controlled vocabulary is a further improvement that is likely to help considerably and is one area of continuing investigation.

Many librarians are entering the information profession with technology skills or are acquiring and using technology skills while on the job. As a result, these librarians will likely be confronted with user interface design issues and the resulting questions of effective information retrieval. Analysis of transaction logs is an excellent method for better understanding user search behavior and also an effective tool for identifying improvement areas in the user interface.

Notes

1. J. Prinsen, "A Challenging Future Awaits Libraries Able to Change," *D-Lib Magazine* 7, no. 11 (2001). Available online from http://www.dlib.org/dlib/november01/prinsen/11prinsen.html.

2. P. Page, R. Jantz, and V. Mead, The Alcohol Studies Database (2000). Available online from http://www.scc.rutgers.edu/alcohol_studies.

3. R. Jantz, "Publishing Databases on the Web: A Major New Role for Librarians and Research Libraries," in *Creating Web-accessible Databases: Case Studies for Libraries, Museums, and Other Nonprofits*, ed. J. Still (Medford, N.J.: Information Today, Inc., 2001), 7–26.

4. D. Blecic, N. Bangalore, J. Dorsch, et al., "Using Transaction Log Analysis to Improve OPAC Retrieval Results," *College and Research Libraries* 59, no. 1 (1998): 39– 50.

5. B. Jansen and A. Spink, "Methodological Approach in Discovering User Search Patterns through Web Log Analysis," *Bulletin of the American Society for Information Science* 27, no. 1 (2000): 15–17.

6. M. Cooper, M., "Usage Patterns of a Web-based Library Catalog," *Journal of the American Society for Information Science and Technology* 52, no. 2 (2001): 137–48.

7. P. Wang, W. Hawk, and C. Tenopir, "Users' Interaction with World Wide Web Resources: An Exploratory Study Using a Holistic Approach," *Information Processing and Management* 36, no. 2 (2000): 229–51.

8. T. Peters, "When Smart People Fail: An Analysis of the Transaction Log of an Online Public Access Catalog," *Journal of Academic Librarianship* 15, no. 5 (1989): 267–73.

9. S. Jones, S. Cunningham, R. McNab, and S. Boddie, "A Transaction Log Analysis of a Digital Library," *International Journal of Digital Libraries* 3 (2000): 152–69.

10. R. Jantz, "An approach to Managing Vocabulary for Databases on the Web," *Cataloging & Classification Quarterly* 28, no. 3 (1999): 55–66.