

Test and Evaluation of an Electronic Database Selection Expert System

Wei Ma and Timothy W. Cole

Introduction

As the number of electronic bibliographic databases available continues to increase, library users are confronted with the increasingly difficult challenge of trying to identify which of the available databases will be of most use in addressing a particular information need. Academic libraries are trying a variety of strategies to help end-users find the best bibliographic database(s) to use to meet specific information needs. One approach under investigation at the University of Illinois has been to analyze and then emulate in software the techniques most commonly used by reference librarians when helping library users identify databases to search. A key component of this approach has been the inclusion in the system's index of database characteristics the controlled vocabulary terms used by most of the databases characterized (Ma 2000).

In the spring of 2000, a prototype version of the database selection tool developed at Illinois was evaluated and tested with a focus group of end users. The prototype tested, named the "Smart Database Selector," included the Web-accessible, three-form interface shown in Figure 1. The multiple form structure of the interface allowed users to search for relevant databases by keyword or phrase, by browsing lists of librarian-assigned subject

terms describing the available databases, or by specifying database characteristics only. The three search forms work independently. Figure 2 indicates the indices and logic behind each form. The prototype system characterized a total of 146 databases. Partial or complete database controlled vocabularies were included for 84 of the 146 databases characterized.

This paper reports on this initial testing and evaluation of the Smart Database Selector. We describe the test design, methodology used, and performance results. In addition to reporting recall and precision measures, we briefly summarize query analyses done and report on user satisfaction measures estimated.

The Objectives

The evaluation described here focused only on search Form 1—Keyword Search. Search Forms 2 and 3 were not evaluated in this Usability Test. The evaluation was intended to satisfy the following objectives:

- Measure the system's performance in suggesting useful electronic resources relevant to users' queries.
- Discover how users, both library professionals and library end-users (frequent and non-frequent), utilize the system.

Wei Ma is an assistant reference librarian, and Timothy W. Cole is an interim mathematics librarian at the University of Illinois at Urbana-Champaign

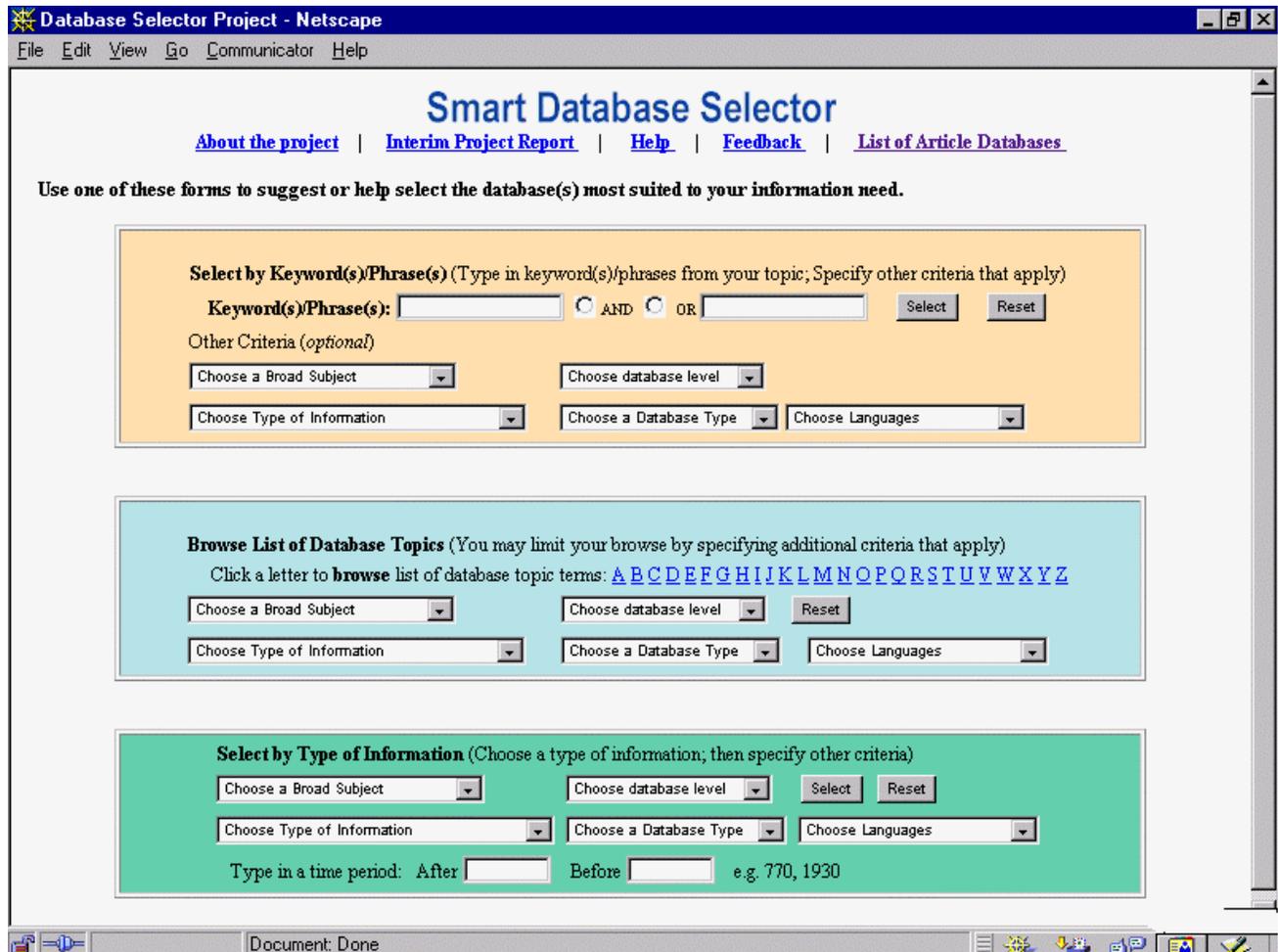


Figure 1. The tested interface of the Smart Database Selector

- Identify ways in which the system can be improved.
- Solicit user reactions to the usefulness of the system.

Evaluation Methodology

3.1 Focus Group Usability Testing

Two groups of participants were recruited—a library user group (undergraduates, graduates and teaching faculty) and a library staff group (librarians and staff who had at least two years of library public service experience). Table 1 shows populations sampled.

We solicited participants through e-mail announcements, postings in student dormitories and academic office buildings, and class announcements. Participants were selected from over 100 UIUC affiliated respondents. The library user group included students majoring in humanities, social sciences, science and engineering, and representing all levels from freshmen to senior, first year and senior gradu-

ates, and teaching faculty. This group also contained American-born and foreign-born individuals. The library staff group included professional librarians, full-time library staff, and students from Graduate School of Library Information Science (GSLIS) who had received at least two years of professional library training and had at least two years of library working experience.

Focus Groups	Population sampled	Numbers of participants:	
		#	%
Library Users Group	Graduates/Faculty	8	36.4
	Undergraduates	8	36.4
Library Staff Group	Librarians and staff	6	27.3
Total		22	100

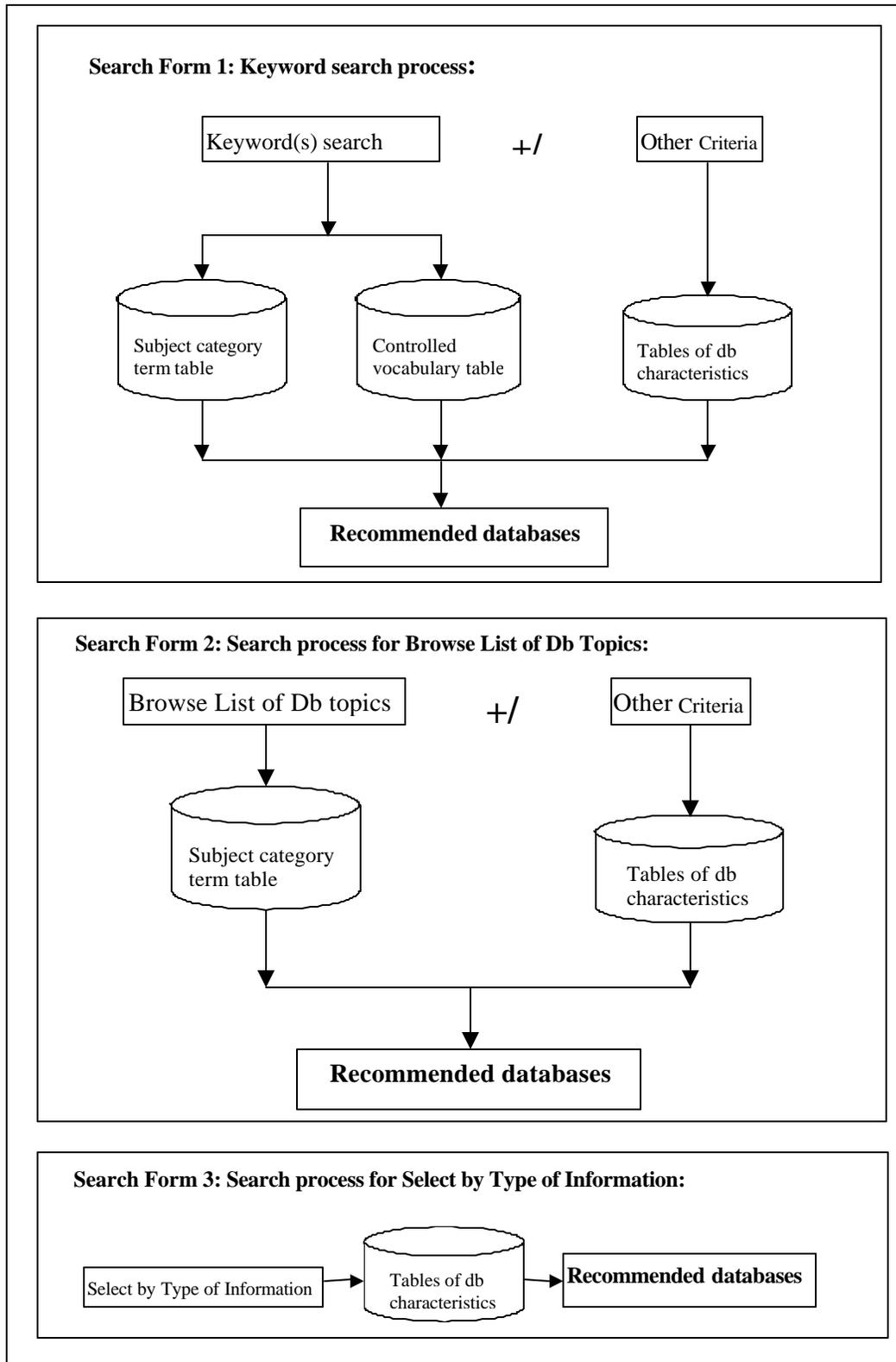


Figure 2: Search Processes & Data Structures for Each Form

Search scenarios for usability testing were selected from real reference questions collected at the Library's central reference desk. The two groups were given different (though overlapping) sets of search questions. Questions given to the library user group are provided below. The library staff group's search scenarios included a few tougher reference questions, and questions with more conditions.

Usability testing took place from mid-April to early June of 2000. Testing was done in groups of 1 to 4 individuals. Each session took approximately 1 to 2 hours. Before beginning, participants were told the purpose of testing and the general outlines of what they were expected to do. No specific instructions or training were given on how to use the Smart Database Selector tool (a brief online help page was available).

After completing initial demographic background questionnaires, participants were asked to pick a topic of their own choosing and use the Selector to suggest databases. Then, they were asked to use the Selector to identify resources for the pre-defined search questions. The entire test process was carefully observed. At the end of each session participants were asked to complete a Post-Test Questionnaire providing qualitative feed back on usability of the interface. Before leaving, a brief interview was typically conducted—e.g. asking if she/he felt the Selector was use-

ful, if he/she like the Selector, if he/she had suggestions for improvements, etc. Transaction logs were kept of all searches done by participants during testing. Full details regarding search arguments submitted, search limits applied, and results retrieved (i.e., lists of database names) were recorded in transaction logs for later analysis.

3.2 Data Analysis

System performance was measured in terms of recall and precision. Recall is defined as the proportion of available relevant material that was retrieved while precision is the proportion of retrieved material that was relevant. Both measures were reported as percentages. Thus:

$$\text{Recall} = \frac{\text{Number of Relevant Items Retrieved}}{\text{Total Number Relevant Items in Collection}} * 100$$

$$\text{Precision} = \frac{\text{Number of Relevant Items Retrieved}}{\text{Total Number of Items Retrieved}} * 100$$

In averaging recall and precision measures over focus group populations, standard deviation was calculated to indicate the variability in these measures user to user and search to search. For these analyses, standard deviation was calculated as:

$$\text{Standard deviation} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

where

N is the number of samples used to calculate the average recall or precision,

x_i is each recall or precision value, and

\bar{x} is the calculated average recall or precision value.

Together, higher recall and precision measures imply a better, more successful search. Smaller standard deviations imply greater consistency in search success across user population. However, an actual successful search depends not only on performance, but also on a user's search strategy, query construction, and proper searching behavior.

Recall and precision measures were not calculated for searches that retrieved 0 hits. All failure searches (0 hit searches, 0 recall searches, and too many hit searches) were analyzed on per search basis to determine most common reasons for search failures.

Results

System Performance

Usability test participants did a total of 945 searches. Of

- | Questions for library user group | |
|---|---|
| 1) | Need articles on U. S. invasion of Normandy |
| 2) | What is being done to deal with aging nuclear power plants? (need to find full-text articles about this topic for an undergraduate presentation). |
| 3) | Need to find articles and statistics on the current trend of women studying computer science. |
| 4) | Need to find articles that compare different literary reviews and criticism on Wilde, Oscar's "The Importance of being Earnest". |
| 5) | Water quality in developing countries (need to do in depth research as for a research project presentation) |
| 6) | "Rosencrantz and Guildenstern are Dead". (wanted articles which compared the play and the film) |
| 7) | A comparative look at the role of the church in the Mexican American community (need to do in depth research as for a graduate dissertation) |
| 8) | Alcohol abuse in European countries. (need to find full-text articles about this topic quickly for an undergraduate term paper assignment) |
| 9) | Wanted articles which outline the current events on stock market investment for a short undergraduate presentation. |
| 10) | Gun control in the United States (need to find full-text articles about this topic for a undergraduate term paper) |

these, 672 keyword searches (i.e., using form 1 of the interface) were done in an effort to answer one of the predefined search questions. The rest of the searches used Form 2 or Form 3, which were not included in this evaluation, or were done to answer a question of a participant's own devising. Of the keyword searches analyzed, 457 retrieved at least one database name. Recall and precision measures were calculated for each of these 457 searches. Recall and precision quality varied significantly question to question, user to user, and search to search. To aggregate our results for presentation, recall and precision averages (arithmetic mean) and standard deviations from those averages were calculated.

Per search recall and precision measures calculated for searches done to answer predefined test questions were averaged on a per question basis over different user groups. Table 2 shows per question average recall and precision measures for searches done by library user group participants (a total of 297 searches that retrieved results). Averages for library staff participants are shown in table 3 (a total of 160 searches that retrieved results). Averages were calculated for other user group breakdowns (e.g., undergraduate vs. faculty/graduate users, frequent vs. infrequent library users, native English language users vs. users whose

native language was not English), but are not shown here. Review of these data did not turn up any meaningful differences by user group.

Examining tables 2 and 3, it is clear that (as anticipated) recall from end-user keyword searches is much better when such searches can be done against database controlled vocabularies than when such searches are done only against summary descriptions generated by librarians. Almost none of those databases for which we did not include controlled vocabulary terms in the Selector's index were discovered via end-user keyword searches, even though a number of these databases were judged relevant to a particular question. Conversely, average per question precision varied from a high of 81.4% to a low of 10%. While this precludes any general conclusion for the full range of keyword searches done, high precision measures for some searches does suggest that the inclusion of controlled vocabulary when characterizing electronic resources does not by itself result in poor precision.

Six of the predefined questions asked of library user group participants were also asked of library staff participants (though not in the same order). On two of these six questions, recall measures for library staff were the same as those obtained by end-users. On the other four questions in

Table 2: Per Question Recall & Precision, Library User Group

Question	Average Recall (R) and Precision (P) +/- Standard Deviation, Calculated Considering:					
	All Databases		Only Databases For Which Controlled Vocabulary Terms WERE SEARCHED		Only Databases For Which Controlled Vocabulary Terms Were NOT AVAILABLE	
	1	R: 26.8% (+/- 11.7%) P: 74.8% (+/- 28.9%)	R: 49.1% (+/- 21.4%) P: 74.8% (+/- 28.9%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)		
2	R: 44.4% (+/- 26.3%) P: 55.0% (+/- 35.1%)	R: 49.4% (+/- 29.3%) P: 55.9% (+/- 35.1%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
3	R: 33.8% (+/- 20.1%) P: 72.5% (+/- 20.7%)	R: 44.0% (+/- 27.9%) P: 69.0% (+/- 27.4%)	R: 3.3% (+/- 6.8%) P: 10.3% (+/- 24.2%)			
4	R: 9.6% (+/- 11.6%) P: 30.4% (+/- 37.8%)	R: 14.5% (+/- 17.4%) P: 30.4% (+/- 37.8%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
5	R: 28.8% (+/- 24.4%) P: 44.4% (+/- 28.1%)	R: 43.1% (+/- 36.6%) P: 44.4% (+/- 28.1%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
6	R: 48.1% (+/- 9.8%) P: 79.0% (+/- 33.8%)	R: 48.1% (+/- 9.8%) P: 79.0% (+/- 33.8%)	N / A			
7	R: 20.0% (+/- 18.9%) P: 10.0% (+/- 11.6%)	R: 33.3% (+/- 31.6%) P: 10.0% (+/- 11.6%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
8	R: 34.7% (+/- 23.5%) P: 35.6% (+/- 31.0%)	R: 52.0% (+/- 35.3%) P: 35.6% (+/- 31.0%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
9	R: 59.2% (+/- 21.9%) P: 50.7% (+/- 24.8%)	R: 70.0% (+/- 25.5%) P: 50.7% (+/- 24.7%)	R: 5.0% (+/- 22.4%) P: 5.0% (+/- 22.4%)			
10	R: 34.8% (+/- 15.8%) P: 57.1% (+/- 36.6%)	R: 60.8% (+/- 27.6%) P: 57.1% (+/- 36.6%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			

Table 3: Per Question Recall & Precision, Library Staff Group

Question	Average Recall (R) and Precision (P) +/- Standard Deviation, Calculated Considering:					
	All Databases		Only Databases For Which Controlled Vocabulary Terms WERE SEARCHED		Only Databases For Which Controlled Vocabulary Terms Were NOT AVAILABLE	
	1	R: 39.3% (+/- 32.1%) P: 28.9% (+/- 28.2%)	R: 49.1% (+/- 40.1%) P: 28.9% (+/- 28.2%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)		
2	R: 44.6% (+/- 20.7%) P: 81.4% (+/- 29.6%)	R: 49.6% (+/- 23.0%) P: 81.4% (+/- 29.6%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
3	R: 58.3% (+/- 16.2%) P: 71.9% (+/- 18.6%)	R: 70.0% (+/- 19.4%) P: 71.9% (+/- 18.6%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
4	R: 53.3% (+/- 17.2%) P: 94.0% (+/- 19.0%)	R: 53.3% (+/- 17.2%) P: 94.0% (+/- 19.0%)	N / A			
5	R: 77.8% (+/- 44.1%) P: 77.8% (+/- 44.1%)	R: 77.8% (+/- 44.1%) P: 77.8% (+/- 44.1%)	N / A			
6	R: 20.9% (+/- 18.8%) P: 46.1% (+/- 37.4%)	R: 28.4% (+/- 24.8%) P: 45.3% (+/- 38.4%)	R: 5.9% (+/- 17.6%) P: 7.1% (+/- 24.4%)			
7	R: 44.4% (+/- 20.3%) P: 67.6% (+/- 22.0%)	R: 66.7% (+/- 30.5%) P: 67.6% (+/- %)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
8	R: 54.3% (+/- 14.4%) P: 38.5% (+/- 25.3%)	R: 54.3% (+/- 14.4%) P: 38.6% (+/- 25.2%)	N / A			
9	R: 32.0% (+/- 14.0%) P: 12.1% (+/- 6.6%)	R: 53.3% (+/- 23.3%) P: 12.1% (+/- 6.6%)	R: 0.0% (+/- 0.0%) P: 0.0% (+/- 0.0%)			
10	R: 53.7% (+/- 26.1%) P: 43.2% (+/- 22.3%)	R: 62.2% (+/- 27.3%) P: 44.3% (+/- 21.5%)	R: 11.1% (+/- 33.3%) P: 1.1% (+/- 3.3%)			
11	R: 62.5% (+/- 23.1%) P: 22.4% (+/- 15.7%)	R: 62.5% (+/- 23.1%) P: 22.5% (+/- 15.6%)	N / A			
12	R: 30.6% (+/- 11.8%) P: 47.2% (+/- 25.4%)	R: 39.5% (+/- 14.8%) P: 46.9% (+/- 25.7%)	R: 3.7% (+/- 11.1%) P: 5.6% (+/- 16.7%)			

common, library staff recall was significantly better. Library staff precision measures were better in five out of the six cases. These results suggest that library staff did tend to formulate “better” search queries for resource discovery using the Smart Database Selector tool.

Table 4 shows impact on performance of combining keyword searches with limit criteria. While the general trends were as expected (use of optional limiting criteria results in better precision but also some loss of recall), the magnitude of the effects on recall and precision were not as expected. Optional limiting criteria were used on 197 out of the 297 keyword searches by library user group participants that retrieved results. Library user group participants had a difficult time using the limits effectively on some questions (in some instances recall was lowered dramatically when optional limits were applied), suggesting a need to rework the limit options provided (which has now been done). Library staff participants used limits a greater percentage of the time and seemed able to use them somewhat more effectively.

Participants were allowed to submit multiple searches in an effort to identify best resources for a particular question. Most users did so. The assumed advantage of this strategy was to net a higher percentage of the universe of relevant databases. To see if this strategy indeed helped users discover more relevant databases, an “effective recall” was calculated across all the searches done by each participant for each question. The results of this analysis are shown in tables 5 and 6. Effective “per user” recall was calculated by combining all sets retrieved by a given user from all searches by that user regarding a particular question and then calculating what percentage of the question-specific relevant databases were represented in the combined superset. The results show that effective recall averages per user was generally higher than per search recall averages presented in tables 2 and 3.

Summary of Failure searches:

Of the 672 searches analyzed, 215 searches did not re-

Table 4: Per Question Recall & Precision Searches With/Without Limits, Library User Group

Question	Average Recall (R) and Precision (P) +/- Standard Deviation, Calculated Considering:					
	All keyword searches		Keyword Searches in conjunction WITH LIMITS		Keyword Searches ONLY	
	1	R: 26.8% (+/- 11.7%) P: 74.8% (+/- 28.9%)	R: 18.9% (+/- 9.4%) P: 75.6% (+/- 30.7%)	R: 31.2% (+/- 10.6%) P: 74.3% (+/- 28.6%)		
2	R: 44.4% (+/- 26.3%) P: 55.9% (+/- 35.1%)	R: 34.3% (+/- 25.6%) P: 69.9% (+/- 36.9%)	R: 62.3% (+/- 16.9%) P: 31.0% (+/- 6.7%)			
3	R: 33.8% (+/- 20.1%) P: 72.5% (+/- 20.7%)	R: 18.4% (+/- 11.7%) P: 75.2% (+/- 28.0%)	R: 48.1% (+/- 14.9%) P: 70.0% (+/- 11.1%)			
4	R: 9.6% (+/- 11.6%) P: 30.4% (+/- 37.8%)	R: 5.9% (+/- 8.0%) P: 26.7% (+/- 38.9%)	R: 29.6% (+/- 5.7%) P: 50.0% (+/- 25.0%)			
5	R: 28.8% (+/- 24.4%) P: 44.4% (+/- 28.1%)	R: 20.1% (+/- 17.8%) P: 44.0% (+/- 30.3%)	R: 50.0% (+/- 26.4%) P: 45.2% (+/- 23.5%)			
6	R: 48.1% (+/- 9.8%) P: 79.0% (+/- 33.8%)	R: 46.9% (+/- 12.5%) P: 79.5% (+/- 37.6%)	R: 50.0% (+/- 0.0%) P: 78.3% (+/- 28.4%)			
7	R: 20.0% (+/- 18.9%) P: 10.0% (+/- 11.6%)	R: 11.8% (+/- 14.7%) P: 10.0% (+/- 13.7%)	R: 42.5% (+/- 7.1%) P: 10.0% (+/- 1.0%)			
8	R: 34.7% (+/- 23.5%) P: 35.6% (+/- 31.0%)	R: 29.6% (+/- 25.9%) P: 40.7% (+/- 35.3%)	R: 47.6% (+/- 6.3%) P: 22.6% (+/- 6.4%)			
9	R: 59.2% (+/- 21.9%) P: 50.7% (+/- 24.8%)	R: 61.5% (+/- 19.0%) P: 57.2% (+/- 22.3%)	R: 50.0% (+/- 33.3%) P: 24.8% (+/- 17.2%)			
10	R: 34.8% (+/- 15.8%) P: 57.1% (+/- 36.6%)	R: 32.3% (+/- 17.3%) P: 67.0% (+/- 36.5%)	R: 42.9% (+/- 0.0%) P: 24.9% (+/- 2.2%)			

trieve any results (0 hit searches), 64 of the 672 retrieved results, but per search recall was 0 (databases recommended did not contain relevant items), 9 of the 672 keyword searches retrieved more than 25 databases (too many hit searches). We defined these searches as failure searches. Table 7 summarizes population of failure searches.

Every failure search was analyzed and classified for the purpose of identifying the causes of the problem. Some failure searches involved more than one category of error types. Table 8 categorizes the failure searches.

The results shown in Table 8 indicate a couple of the difficulties when evaluating a system such as this using focus group usability testing:

- Focus group participants lacked of motivation or desire to find the information.
- Participants may not be familiar with all of the topics or subject areas of the pre-defined search scenarios.

From transaction logs and on-site usage, the authors ob-

served and found differences between real-time users and focus group usability test participants using the system. Real-time users, when using the system to recommend electronic resources, would usually focus on the keyword formulation and then move on to the specific resources suggested. They are more interested in searching the suggested databases for relevant citations. However, focus group participants were more interested in seeing how the system suggested different set of electronic resources with different

Table 5: Per User Recall for Each Question, Library User Group

Question	Average Number of Searches Done	Average Recall (+/- Standard Deviation) Calculated Considering:			
		All Databases		Only Databases For Which Controlled Vocabulary Terms WERE SEARCHED	
		1	3.1	30.7% (+/- 10.3%)	56.2% (+/- 18.8%)
2	2.7	59.4% (+/- 18.8%)	66.0% (+/- 20.9%)		
3	2.3	41.7% (+/- 16.7%)	53.7% (+/- 23.0%)		
4	3.5	15.6% (+/- 13.6%)	23.3% (+/- 20.3%)		
5	2.6	36.2% (+/- 25.9%)	54.3% (+/- 38.9%)		
6	2.5	44.8% (+/- 15.5%)	44.8% (+/- 15.5%)		
7	2.3	25.0% (+/- 19.7%)	41.7% (+/- 32.8%)		
8	1.9	42.7% (+/- 22.4%)	64.1% (+/- 33.6%)		
9	1.9	64.3% (+/- 15.5%)	75.7% (+/- 17.5%)		
10	2.0	40.2% (+/- 7.6%)	70.3% (+/- 13.4%)		

Table 6: Per User Recall for Each Question, Library Staff Group

Question	Average Number of Searches Done	Average Recall (+/- Standard Deviation) Calculated Considering:			
		All Databases		Only Databases For Which Controlled Vocabulary Terms WERE SEARCHED	
1	4.8	56.7%	(+/- 28.1%)	70.8%	(+/- 35.1%)
2	2.3	56.7%	(+/- 19.7%)	63.0%	(+/- 21.9%)
3	2.6	60.0%	(+/- 8.6%)	72.0%	(+/- 10.3%)
4	1.7	61.1%	(+/- 13.0%)	61.1%	(+/- 13.0%)
5	3.3	100.0%	(+/- 0.0%)	100.0%	(+/- 0.0%)
6	3.5	38.9%	(+/- 27.4%)	50.0%	(+/- 26.6%)
7	4.0	45.8%	(+/- 13.4%)	68.8%	(+/- 20.0%)
8	4.4	58.3%	(+/- 19.5%)	58.3%	(+/- 19.5%)
9	1.7	33.3%	(+/- 15.6%)	55.6%	(+/- 25.9%)
10	1.5	58.3%	(+/- 28.0%)	66.7%	(+/- 28.7%)
11	1.9	60.0%	(+/- 21.1%)	60.0%	(+/- 21.1%)
12	1.6	33.3%	(+/- 13.3%)	42.6%	(+/- 17.0%)

Conclusions

The focus group usability testing helped us identify needed interface design changes. In particular it led us to simplify optional limiting criteria features. Based on results of evaluations described above and on other evaluative work performed, we have proceeded with the following system/interface improvements:

- Reduced default, “Basic Search” interface to two, simplified forms (derived from forms 1 and 2 as shown in figure 1 of this paper), and added an “Advanced Search” interface, also comprised of 2, some-

keywords and optional criteria. They paid little attention to the evaluation of the suggested electronic resources.

User Satisfaction Measures

Measures of user satisfaction were estimated based on post-test questionnaire responses of focus group participants and on brief in-person interviews conducted at the end of each test session. Tables 9–11 summarize answers to post-test questionnaires. Overall user reaction to the usefulness of the system was positive. Responses of 17 out of the 22 participants supported the usefulness of the system, with 2 of the participants remaining neutral. Twenty of the 22 participants favored use the Selector alone or in combination with the current Library menu system. Undergraduate group and library staff participants appeared to like the system most. Brief interviews after test sessions revealed that those who did not favor use of the Selector tend to feel they already know which resources to search for research topics of interest to them (and therefore they do not need software to recommend databases to them).

what more complex forms (derived from forms 1 and 3 as shown in figure 1).

- Collapsed optional limiting criteria to three categories on the Basic Search interface and four categories on the Advanced Search interface.
- In order to improve search consistency and recall, revised procedures for assigning subject category terms (based on producers’ database descriptions) by implementing a system-specific controlled vocabulary.
- Provided more on screen search hints and more meaningful instructions in how to use the tool.

In fall of 2000, with the above revisions in place we performed a second usability test. We hope to analyze these tests and publish results soon.

References

Ma, Wei and Timothy W. Cole, “Genesis of an Electronic Database Selection Expert System,” *Reference Services Review* 28 (2000): 207–22.

Table 7: Failure searches logged by user groups

		Library User Group		Library Staff Group		Total
		#	%	#	%	
Total number of keyword searches						672
Failure searches	0 hit searches	161	23.9%	54	8.0%	215 (31.2%)
	0 recall searches	53	7.9%	12	1.8%	65 (9.7%)
	Too many hit searches	8	1.2%	1	0.1%	9 (1.3%)
Total		222	33%	67	9.9%	289 (42.2%)

Table 8: Breakdown of Failure Search Causes

Failure searches	Error Types			
	Misspelling	Improper Keywords Used	Improper Limiting Criteria	Incorrect Use of Search Form
0 Hits	31	129	81	8
0 Recalls	3	46	30	
Too Many Hits		9		
Total (289):	34	184	111	8

Table 9. Answers to the Post-Test Question 1.

Question 1: I found that the Smart Database Selector was very useful in recommending appropriate databases for topics or term papers.

Participant Population	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Total
Undergraduates	1	6	1			8
Graduates/faculty	1	5		1	1	8
Library staff	1	3	1	1		6
Total	3 (13.6%)	14 (63.6%)	2 (9.0%)	2 (9.0%)	1 (4.5%)	22

Table 10. Answers to the Post-Test Question 2.

Question 2: Compare the Smart Database Selector to the current Library's static menu page, I would use

Participant Population	this Selector only	this Selector & the current Library menu	not this Selector	other	Total
Undergraduates	4	4			8
Graduates/faculty	2	4	1	1	8
Library staff		6			6
Total	6 (27.2%)	14 (63.6%)	1 (4.5%)	1 (4.5%)	22

Table 11. Answers to the Post-Test Question 3

Question 3: Overall, I found that the Smart Database Selector's interface easy to use.

Participant Population	Strongly agree	Agree	Neither agree or disagree	Disagree	Strongly disagree	Total
Undergraduates	1	3	1	2	1	8
Graduates/faculty		7	1			8
Library staff		3	1	1	1	6
Total	1 (4.5%)	14 (63.6%)	3 (13.6%)	3 (13.6%)	2 (9.0%)	22