

CALIFORNIA DIGITAL LIBRARY

Core Collection Interviews

FINAL REPORT

1 July 2004

Prepared by

Alex Wright
alex@agwright.com

with: Rosalie Lack, Ellen Meltzer, Steve Toub, Roy Tennant

Contents

Executive Summary 3
Introduction 4
Methodology 4
Findings 5
Appendix A: Interview Notes 9
Appendix B: Moderator’s Script 10

Executive Summary

This report documents the findings from a series of interviews conducted with ten UC-Berkeley undergraduate students. All interviews took place on the Berkeley campus in March, 2004.

This study's research objective was to model student behaviors and attitudes towards Web-based research tools (both free and licenced), and to explore opportunities for new online research tools such as Scholar's box and metasearch tools in the current UC environment.

Key findings:

- Undergraduates conduct "transactional" research, typically in 2-3 week cycles
- Upper year students tend to rely on known resources
- Many students prefer licensed databases to Google for scholarly research purposes
- Expectations of search interfaces derive from Google and other Web resources
- No use of citation management software

Recommendations:

- **Embed metasearch/Scholar's Box features** within existing research tools wherever possible, rather than try to drive students towards adopting new software or Web sites.
- Keep search query interfaces **simple and Google-like**. Support advanced search features by following standard Web search conventions.
- Support **iterative searching** by making search history terms visible, allowing "search within these results" feature.
- **Improve discoverability of controlled vocabularies** by exposing metadata (e.g. subject headings, keyword lists) in search results as well as in abstract-level records.
- **Allow students to stay within a preferred source** for the duration of a search session.
- Allow users to **save search results and citations** online or send them to an email address. Permit registered users to save citations online indefinitely.

Introduction

As part of the California Digital Library's Core Collection initiative, Alex Wright and Rosalie Lack conducted a series of interviews with current UC-Berkeley undergraduates in March 2004.

The interviews focused on understanding student use of online search tools, including UC library resources (e.g. OPACs and online databases) and public Web sites (such as Google and Yahoo!).

This study's primary research objectives included:

- **Prioritizing user goals** for new CDL development initiatives such as the Personal Library Manager
- **Understanding user expectations** of search features and user interface design conventions
- **Generating qualitative insights** about undergraduate research behaviors, and opportunities for improving the user experience

Methodology

From March 17-18, 2004, we conducted a series of interviews with ten current undergraduates at UC-Berkeley. Interviews took place on the UC campus; participants were compensated for their participation.

The interviews employed the "friendly dyad" format¹, with pairs of students teaming up for 60-90 minute interview sessions conducted by a moderator (Alex Wright) with one observer (Rosalie Lack).

Students participated anonymously, but were asked to identify themselves by year, concentration, and first language. Following is the participant breakdown:

Concentration	Year	First language
Psychology	4 th	English
Asian Studies	3 rd	English
Political Science	3 rd	English
Psychology	4 th	English
Asian Studies	2 nd	English
Neurobiology	4 th	Vietnamese
Anthropology	4 th	English
Social Welfare/Psychology	4 th	Spanish
Women's Studies	4 th	English
Legal Studies	3 rd	Hmong

For the complete Moderator's Script, see [Appendix B](#).

¹ Due to schedule changes, one interview had one participant; one interview had three.

Findings

General Research Behaviors

Most of the students interviewed approach secondary research as a transaction, rather than as an ongoing undertaking (unlike, we presume, faculty and graduate students). Students typically conduct research over a 2-4 week period. The notable exceptions were students writing honors theses, who expressed a stronger interest in productivity tools.

Students rely almost exclusively on online search tools to conduct academic research. Although all students have at least a basic familiarity with GLADIS and online databases, many students rely primarily on one or two authoritative resources in their field of concentration (e.g. PsycInfo for Psychology majors).

Although all students are familiar with Google, Yahoo! and other popular Web sites, most shy away from relying on these services for academic research. They perceive the licensed campus resources to be more reliable, selective and authoritative. Many students said professors and graduate instructors have discouraged them from using public Web resources for academic research.

Students exhibited a wide disparity in styles and approaches to note-taking and citation management, typically using a combination of paper notes, Microsoft Word and other ad hoc methods for managing research. None of the undergraduates interviewed used citation management software.

Almost all students had consulted personally with a librarian, faculty or graduate student in formulating research strategies for a current or past project. Students placed a high premium on access to personal, individualized research support from librarians or other subject experts.

The most commonly used current search tools included:

Library resources:

- GLADIS
- Melvyl
- Pathfinder
- Public library OPACs (Oakland and Berkeley)
- PsycInfo
- ERIC

Public Web sites:

- Google
- Yahoo!
- Amazon
- Ask Jeeves
- American Historical Anthropology
- FDA Web site

Search Expectations

Based on their experience using current search tools, users voiced a strong preference for simple, uncluttered search interfaces like Google. Although some users would limit searches by date for journal article searches, in general they exhibited limited interest in advanced search features.

Users frequently perform search transactions iteratively, refining query terms and sometimes even scanning abstracts and articles for keywords or authorized subject terms. They may not know what a "controlled vocabulary" is, but they know one when they see it.

While users voiced interest in the general proposition of federated search tools, they do not perceive a high value in results federation alone, unless the service provided qualitatively better search results or some other combination of new features. Students were mostly comfortable with existing sources, and voiced a strong desire to stay within a selected resource for the duration of a search transaction.

Citation Management

Depending on concentration, users need to create citations in different formats (e.g. APA for Psychology, MLA for humanities). At this point I don't think we have a firm grip on which citation formats are most important; but hoping we can derive this through additional interviews and by looking at Refworks, Endnote etc.

Overall, We believe that current search tools are well understood and the demand for horizontal search, while important, carries less value to undergraduates than the ability to manage notes and citations.

Environment

Users conduct their research from multiple locations, usually using multiple computers: e.g. home, library, other campus computers. PDA use was extremely limited.

Some users mistakenly believe that they must be on campus to search licensed databases (i.e. some users were not aware of the campus proxy servers)

Users frequently email citations to themselves (one interviewee used XDrive for remote storage).

Software

The most commonly used software tools included:

- MS Word
- MS Excel
- MS Powerpoint
- Adobe Acrobat
- SPSS
- Blackboard
- WebCT

Note: CDL made a conscious decision to focus on upper-year students. All participants had already declared majors, and as such were often well acquainted with available research sources in their fields of study. We can hypothesize that lower-year students are more likely to use generalist resources such as GLADIS, Pathfinder and Expanded Academic Index.

Recommendations

Based on the interview findings, I recommend that CDL focus its Core Collection initiative around the following high-level features:

1. Embed PLM features within existing online services such as GLADIS/Melvyl and licensed databases.

Users voiced a strong preference for using known and proven search tools. While they are willing in principle to consider a new “meta-search” tool, they are far more interested in seeing improvements to the tools they already use.

PLM development should focus, at least initially, on adding value to these existing tools, through the development of features such as:

- **“Shopping basket”** style results list for saving and exporting citations
- **Emailing/exporting citations** in popular formats (e.g. MLA, APA)
- **Saving search history terms** for future reference, or repurposing a search query in other sources

2. Meta search interface

To drive user adoption, any new meta-search interface will have to exhibit a clear and compelling benefit over tools already in use. Federating search results alone will not deliver sufficient benefit for users to abandon existing search services.

Based on the interview findings, I believe the biggest opportunities for differentiation include the development of a simple, predictable interface across search databases, citation management, advanced search, and improved capabilities for iterative searching.

The interview findings also suggest the following key design principles:

- Keep search input forms **simple and Google-like**
- Keep all additional non-keyword search parameters on an **Advanced Search** page, following standard search form conventions as established at Google, Yahoo!, Silverplatter et al.
- Support **iterative searching** by:
 - **Making search history terms visible** to help users keep track of queries
 - **Allow “search within these results”** feature, letting users expand/constrain sources using a previous query term.
 - **Improve discoverability of controlled vocabularies** by exposing metadata (e.g. subject headings, keyword lists) in search results as well as in abstract-level records. Make it easy for users to launch new queries using those terms.

- o **Allow students to stay within a preferred source** for the duration of a search session (or even as a saved user preference).

3. Citation Management

Allow users to **save search results and citations** online for future reference.

Standardize the “mailto” user interface across databases, especially within search results, abstracts and articles.

Support **easy export to different citation formats**, especially MLA, APA and other standard citation formats. See current citation management tools like Refworks and EndNote for comparison.

4. Additional Features

I would also recommend that CDL explore the feasibility of the following additional features:

- Improved access to personal research support (e.g. Ask a Librarian feature)
- Ability to share search results with small groups (e.g. for group projects)
- Ability to set up “alerts” for new search term matches (honors thesis students only)

Appendix A: Interview Notes

[For privacy reasons the interview notes have been removed]

Appendix B: Moderator's Script

I Introductions

Thank you for agreeing to participate in this interview today. We're conducting this study on behalf of the California Digital Library in an effort to understand how students conduct their academic research. Your input will help the CDL in developing new research tools for the entire University of California system. Your participation is completely anonymous; your name will not be used in the final report.

First, tell us a little about yourself:

Participant 1

Name: _____

() Freshman () Sophomore () Junior () Senior

Major: _____

Is English your first language? Y ___ N ___

Participant 2

Name: _____

() Freshman () Sophomore () Junior () Senior

Major: _____

Is English your first language? Y ___ N ___

II General Research Habits

In a typical semester, how often do you find yourself conducting research online? For example:

- Research paper
- Senior thesis
- History 101 paper

Have you had library instruction?

When you have an assignment due (e.g. a paper), how far in advance do you typically start the research process?

Can you think of a recent assignment where you had to conduct research online? Tell us about that process:

- What sources did you use?
- Did you rely solely on online sources, or did you also use books, journals or other print material?

- Did you consult a librarian?
- Who else did you ask for help? (e.g., a professor, TA, other students)
- Did you find what you were looking for?
- What kinds of challenges or difficulties did you encounter?

If you had to write a 5-10 page paper about a topic you didn't know much about – e.g. plant regeneration for a biology class – where would you start?

If an item were only available in print, would that eliminate it from the list of items you would consider?

How do you take notes? (e.g. using a computer, paper notebooks)

How do you create bibliographies or footnotes?

Have you worked on research projects where you had to collaborate with other students?

III Online Sources

What sources do you typically use?

- Web search engines (e.g. Google, Yahoo)
- Melvyl
- Pathfinder
- Spiro
- Academic Index
- Reserves desk
- Web of Science
- UC eLinks
- Searchlight

How often do you use each one?

What do you like about these services? What do you dislike?

What other Web sites do you use (e.g. Amazon, Ebay, Friendster)? What do you like/dislike about them?

IV Environment

Where do you typically conduct research:

- Dorm room / Apt
- Library
- Student center
- Computing Center

Do you have a laptop? A PDA?

Do you ever do research using computers other than your own (e.g. a public terminal at the library?)

What kinds of software do you typically use?

- MS Word
- MS Powerpoint
- Blackboard
- Refworks
- Web CT
- CourseWeb
- Blogs
- Wikis
- Etc.

V “Blue Sky”

In a perfect world, what kinds of research services would you like to see?

- Cross-database searching?
- Integration of Web content and library catalogs?
- Topic guides?
- Abstracts
- Bibliography/citation tools
- etc...

If there were a service that offered better search results than the services you currently use, would you be willing to wait for the results? How long?

If you could get a better list of search results (with relevant items ranked first, and duplications removed) delivered to you via email, would you be willing to wait 15 minutes? If not, how long would you be willing to wait?

Think about an ideal search tool. Now, tell us about it.