UCLA European Integration Portal:

Metasearch Assessment

Table of Contents

Introduction

The European Integration Portal is one of the services being developed as part of the California Digital Library's MetaSearch Infrastructure Project. The CDL is working to create metasearch tools and software that campus libraries can use to craft search portals tailored to specific audiences and needs. A primary goal of the metasearch service is to assist users in efficient discovery of information across a range of resources.

This document presents preliminary findings from a round of needs assessment conducted at UCLA on June 1-2, 2005.

The purpose of these interviews was to document the research behaviors and needs of faculty and graduate students in the area of European Integration in order to inform the development of the European Integration Portal.

The key questions that were explored by this round of assessment include the following:

1. What are the research *behaviors* of users who possess domain expertise?
2. What are the research *needs* of users who possess domain expertise?
3. Can the MetaLib product play a role in research for users who possess domain expertise? How do we position this product?

Methodology

This round of needs assessment consisted of five group interviews. Each interview consisted of two to four interviewees, a facilitator, and two observers. A total of 14 multidisciplinary participants with domain knowledge in European studies were recruited by Kati Radics of UCLA for the interviews. Four were professors and ten were graduate students; four of the five interview groups included one faculty member.

The group interview format was chosen in order to increase the comfort level of the interview subjects by distributing the focus of attention among the group instead of putting the spotlight on one individual. Felicia Poe of CDL assumed the role of interview facilitator and used a list of prepared questions as a guide for the discussion. Jane Lee of CDL and Bo-Gay Tong Salvador of UCLA observed and took notes on interviewees' responses to questions using laptop computers. At the close of each interview session, the group was asked to express their familiarity with and interest in eleven features and services relating to the Metasearch project.

Analysis

What are the research behaviors of users?

The search behavior of researchers depends on their goals. Their goals, in turn, depend upon several parameters, including the following:

> - New to field ←→ Domain expert
> - Starting coursework ←→ Finishing dissertation ←→ Years of experience in the field
> - Seeking general information about a topic ←→ Looking for a gem

Often, the level of experience and comfort with computers is also a strong factor, but we did not find this to be true in this instance. Everyone we spoke with had trouble using the various databases. Everyone had trouble knowing which ones to use. These are universal challenges to using information resources that are not solely dependent on a user's ability level.

Types of resources

Interviewees reported using many types of resources due to the interdisciplinary and international nature of their research areas. They rely heavily on catalogs – both foreign and domestic. Those who utilize statistics in their research reported finding them online easily. For more contemporary research questions, there is a greater reliance on journals than books. Researchers "definitely need various kinds of articles from different kinds of genres."

Finding a starting point for research

> - Researchers start close to home, i.e. with campus OPAC or Melvyl, and expand outwards.
> - Researchers use the bibliography and table of contents of a good source to find leads to others.

Given that it is difficult to find good sources, especially in a interdisciplinary area, researchers in European studies employ a variety of tactics at the outset of their search. The most common strategy expressed by interviewees is to start close to home and then expand outward. This means beginning with a broad search within the UCLA catalog and Melvyl. A number of interviewees reported accessing WorldCat as well, and experienced researchers also examine the main journals in their field. Because of the importance of getting a physical copy as quickly as possible, researchers want to know where sources are physically located. Local accessibility, including full-text online, is highly prized; the next best option is interlibrary loan.

At this early stage of research, the ultimate goal is to get at least one good source in order to look at its footnotes, bibliography, and chapter headings for leads on other sources or keywords to use. In order to find this source, scholars rely on searching, browsing, and recommendations from colleagues and fellow students. Often, they employ the process of elimination, going through a personal list of trusted, dependable resources one-by-one.

Search

- Interviewees greatly prefer a basic search interface with fielded searching and optional limits.
- Discovering effective search terms is a process of trial and error.
- Experienced researchers are looking for gems.

Interviewees overwhelmingly prefer a basic search interface to an advanced search interface. However, the basic search screen must offer fielded searching, including keyword, author, and title, and the ability to apply optional limits for date and language. The author field is particularly important to one researcher we spoke with, because he usually remembers names over titles. One graduate student complained that one database he uses recently removed some optional limits from the basic search screen, thereby leaving him in limbo between advanced and basic search. The one interviewee who prefers starting with advanced search likes having control and the "possibility of playing around with things." Others desire a greater level of control only if they are searching for a known item, such as one they found in a bibliography.

When beginning their search, researchers use keywords to cast a wide net. Finding effective search terms is a process of trial and error. Some researchers reported occasionally employing Internet search engines to hunt for good keywords and to find general information about a topic. Several researchers expressed dissatisfaction with using Library of Congress Subject Headings as a guide, calling them "virtually useless" and "too selective."

When searching for sources, researchers enter a term and quickly scan the results to decide whether or not the term worked. If the search term they tried doesn't seem to have worked, then they return to the search screen and try another. They repeat this process until they get a promising result. When this happens, they look through the results hoping to find a good source.

> *… for all of us doing research, we want access to everything. We will make the decisions ourselves.... Everything – that's our dream.*

Unlike users looking for general information on a topic new to them, for whom any reliable information will suffice, our researchers are looking for gems. Because of this, they are willing to sift through *all* of the returned results – even if they number in the hundreds. They want to find the uncommon, the elusive, so they need to feel like they have seen everything related to their research topic/question. They want more rather than less. Speaking about an undergraduate class she taught, one graduate student observed, "I see undergrads and they see 50 results and they want to pull out their hair….The joy of research for me is to see all these things."

Regarding relevance ranking, researchers do value and desire this feature, but they recognize that a system's determination of relevance may not correspond to their own. One researcher noted, "My relevance changes from week to week." All researchers, however, value the merging and deduplication of records.

Metasearch

- Researchers welcome the ability to work within a single interface to search several different resources.
- Researchers feel that working with emailed results would complicate their trial-and-error strategy, which depends on fast interactions and nearly instantaneous feedback.

When presented with the idea of a metasearch, interviewees responded positively. They welcomed the ability to enter a search term into a single interface and retrieve results from different resources, such as catalogs and article databases. Reactions to the hypothetical 2-minute wait time for ranked, merged, and deduplicated results, however, were mixed. Some interviewees admitted being impatient. They don't like to wait and would probably move on if they felt that the system was taking too long. A few interviewees suggested that receiving results in batches may alleviate feelings of impatience and information overload. Most researchers expressed a willingness to wait for high-quality ranked results as long as

1) they knew that the system was still working,
2) they could multitask and work on something else in the meantime, and
3) they were given an idea of how much longer they'd have to wait.

The option of having search results emailed to them instead of waiting did not appeal to most researchers. They felt that working with emailed results would complicate their trial-and-error strategy, which depends on fast interactions and nearly instantaneous feedback. Once source selections have been made, however, researchers do value the ability to email their choices to themselves.

Although all researchers viewed metasearch as a potentially useful service, some observed that it probably works best for topic searching and thus might be more appropriate for undergraduates. As one professor explained, "You drag along things that you keep building on." That is, as one gains experience in a field, he or she begins to accumulate tidbits of information and inklings of questions. Because of her many years of research experience in her profession, this researcher felt as though she did not need to start new threads of research very often. She implied that experienced researchers may not need to perform general topic searches often – a feeling that was reiterated by other professors.

EndNote

- EndNote for search – A way to "connect to different things quickly"
- EndNote for citation management

The graduate students who use EndNote rave about it. Although it can be difficult to configure, once EndNote is set up correctly, researchers appreciate the ability to search different resources within a single interface. One researcher called EndNote "one way to weed out and pull in resources" and "create your own library." The biggest problem she observed is that EndNote does not indicate where items are located, which makes it difficult to get the item. Researchers also noted that EndNote makes it easy to take citation information from a library and put it into a Word document. They appreciate not having to type in references separately.

Browsing for discovery

- JSTOR is "useful for trying to root something up" and provides a positive browsing experience.
- Researchers browse the table of contents of journals hoping for serendipitous discovery of sources.

Researchers have mixed opinions about browse. A spectrum of opinions was voiced – from a professor who considers browse more important than search to a graduate student who equates browsing with procrastination. Most, however, recognize the potential value of browse. A number of interviewees report browsing the stacks in the library for inspiration or in hopes of discovering good sources. Browsing online received mixed reviews. Those hoping for a re-

creation of physical browsing were not satisfied. They feel that the online space cannot hope to simulate collocation adequately or to provide an environment for the kind of serendipitous discovery that comes from wandering among shelves filled with books.

Several researchers did, however, point to JSTOR as an instance where virtual browsing is better than physical browsing. JSTOR not only gives users the option of browsing journal titles alphabetically or by discipline, but it also provides table of contents information. With these features, our researchers found that they could browse online much more quickly and effectively than if they had to locate and physically page through a stack of journals. The main drawback of JSTOR is that it does not provide access to the most recent journal articles.

Researchers also appreciate the ability to browse by author in some systems. Browsing by topic can be helpful if one knows which topics to look under. Some interviewees report using the browse by topic feature of some interfaces *after* finding a good source.

The role of serendipity

- "Research is usually serendipitous."

Speaking of his research habits, one professor reported,

> *Mine [my research] is always scattered among different things. Usually I'm squeezing something in between other things. Usually it's because someone has told me about something…. Research is usually serendipitous – often more kind of random work. [That's] why I tend to like wandering around the library and wandering around the papers or journals.*

Sometimes one finds the best information in the least expected places. A different professor recounted how he once found a source on fascism in a bibliography. At first, he thought that it wouldn't be relevant to his research because it was published in the 1930's. He decided to look at it anyway and discovered that it had greater relevance than more recent publications. Informal conversations can also provide tips to valuable sources. "A lot of it is word-of-mouth and then you go look for something."

Databases

- Because it is easy to feel overwhelmed by the vast array of databases available, researchers rely on a handful of familiar resources.
- Researchers realize that they are probably missing out on some gems, but without good recommendations on which databases to explore, they are unlikely to venture out into the unknown on their own.

Knowing which databases to use is a difficult task. Because it is easy to feel overwhelmed by all of the choices, most researchers rely on a handful of resources that they have found useful in the past. One researcher admitted that there is always a risk of missing things, but he does not have enough time or energy to search everything all the time. Navigating to and within databases is not a trivial task, so researchers have even more reason to stay with the databases they already know how to use. Researchers realize that they are probably missing out on some gems, but without good recommendations on which databases to explore, they are unlikely to venture out into the unknown on their own.

Interviewees offered suggestions – aside from getting recommendations from librarians and colleagues – on how to help them discover new resources and use them effectively. They referred to Amazon's recommender system as a model for how new databases could be introduced to them. In addition, a keyword guide/recommender for each database would be extremely helpful, especially since search terms may not work the same way in all databases.

Presentation of databases

- Choice and control are important to experienced researchers.
- Since is often impossible to know what a database covers from its title, each database should have a short description of its contents.

One way to provide database suggestions is to cluster them into logical groupings. During each session, interviewees were presented with two search interfaces with the same set of databases. (See Appendix A.) In column 1, databases are clustered by category, and the user selects one or more categories in which to search. The system performs a metasearch across all databases in the chosen categories. In column 2, databases are presented in an alphabetical list. The user selects the individual databases to search, and the system performs a metasearch across those databases.

Interviewees identified advantages and disadvantages to each option. Dissatisfaction with column 1 stemmed from suspicion of terms and categories that are not one's own. Some researchers felt that all categorization schemes are highly subjective and may actually hinder discovery of new resources instead of help. Others felt that by presenting topic clusters, column 1 provides database recommendations that would be especially helpful to those new to the field.

The following quotes are representative of the opinions expressed:

> *We are in age of interdisciplinary studies. If you use this scheme [column 1] you are in a sense taking a step back. If doing research on an advanced level, [you] need to think on interdisciplinary categories or non-categories. What about the box that isn't here?*

Versus

> *You may not be familiar with a database, but the librarian knows about it so it's included. So it forces you to expand how you think.*

Some researchers proposed a compromise solution that involved presenting databases by category but allowing users to choose individual databases regardless of category. This would retain the recommender aspect while giving users complete control over which databases are searched.

Managing Information

> - "My problem with online research is that at some point there is just too much information."
> - Email is an important information management tool, and "Don't pollute my inbox" is a strong sentiment that many hold.

Researchers have come up with several strategies to prevent information overload. Email is an important storage vehicle, and "Don't pollute my inbox" is a strong sentiment that many hold. Researchers value the ability to email results to themselves, but they do not want the system to deliver search results directly to their inboxes. They want absolute control over what ends up in their email. It is important that every record in their inbox has survived an initial cut.

Other methods of managing information include printing records, using browser history as memory, and gathering and saving records in a system's "basket" feature. One researcher keeps track of bibliographic records using simple text files. Some use commercial products, such as EndNote to manage bibliographic information.

Survey of features and services

> - Researchers want recommendations for databases and search terms and help managing information.

At the end of each interview session, participants were asked to indicate their level of familiarity with and interest in eleven potential features and services of the metasearch portal. Felicia Poe guided the interviewees through each line of the survey, explaining what each service or feature was. Interviewees were asked to mark "Unfamiliar with item" if they had not heard of the item prior to the explanation.

The results of the survey show that researchers want two main things: recommendations for databases and search terms and help managing information. They also want full-text results. Not surprisingly, researchers were either unfamiliar with or not interested in RSS feeds and content from weblogs. (See Appendix B for complete results.)

Appendices

Appendix A: Database Presentation

<u>Column 1</u>	<u>Column 2</u>

Column 1

Search [   ]

☐ **Economics**
- Resource A
- Resource B
- Resource C
- Resource D
- Resource E

☐ **Government and Politics**
- Resource F
- Resource G
- Resource H
- Resource I
- Resource J

☐ **Environment**
- Resource K
- Resource L
- Resource M
- Resource N
- Resource O

Column 2

Search [   ]

☐ Resource A
☐ Resource B
☐ Resource C
☐ Resource D
☐ Resource E
☐ Resource F
☐ Resource G
☐ Resource H
☐ Resource I
☐ Resource J
☐ Resource K
☐ Resource L
☐ Resource M
☐ Resource N
☐ Resource O

▇ Very Familiar   ▇ Somewhat Familiar   ▇ Not at All Familiar

Appendix B: Survey and Results

Survey

Please indicate your level of interest in the following features and services by marking the appropriate box.

	Interested	Neutral	Not Interested	Unfamiliar with item
1. Ability to browse resources by topic	9	3	0	1
2. Suggested keywords for expanded searching	9	3	1	0
3. Alert services	4	3	2	4
4. Database recommender	13	0	0	0
5. RSS feeds	0	0	0	13
6. Content from weblogs	0	0	6	7
7. Links to email listservs	4	5	2	2
8. Ability to export to citation management systems, e.g. EndNote	10	1	0	2
9. Limitation of results to full-text articles	9	4	0	0
10. Ability to save items to a basket within a session	12	1	0	0
11. Ability to log in and view saved items at any time	11	0	1	1

Appendix C: Questions and Objectives

1. Please introduce yourself and describe your area of expertise.
   Objective: Give participants an opportunity to speak in order to break the ice.

2. Is there anything unique about your area of expertise that makes research particularly challenging?
   Objective: Determine sources of pain, which might help identify a role for MetaLib.

3. After identifying a research question, how do you get started? What are your strategies for finding information?
   Objective: Determine research behaviors of users.

4. How do you stay current? What resources do you use?
   a. Can you envision the qualities of the best possible service that would help keep you current?
   Objective: Determine whether or not users use non-traditional resources.

5. There are many different kinds of sources – such as databases and journals – that one may use for research. How do you decide which ones to use?
   Objective: Determine research behavior of users.

6. What kinds of difficulties do you run into in the course of doing research?
   Objective: Determine research behavior of users. Determine "points of pain".

7. Are you satisfied with the number and variety of sources that you currently use? Are you confident that you are looking in all the right places?
   Objective: Determine potential research needs. Determine if there is a reason to change current practices and adopt new methodologies.

8. [Present clustering mockup.] How would you prefer to view your sources? Why?
   Objective: Determine whether or not users are interested in clustering.

9. Do you prefer to browse by topic or do a keyword search?
   Objective: Determine users' preferences for information gathering.

10. What is frustrating about viewing or using search results?
    a. Explore features, e.g. ranked results, basket, check boxes, etc.
    Objective: Determine "points of pain". Determine what types of features are useful.

11. Would you rather get your search results in small batches so they come back quickly or all at once with everything ranked, which would take longer? (How long would you be willing to wait?)
    Objective: Determine whether users value ranking more than speed.

Appendix D: Interview Schedule

Wednesday, June 1, 2005

    9:00-10:15AM
    • Professor
    • Graduate student
    • Graduate student

    1:30-2:45PM
    • Graduate student
    • Professor

    3:00-4:15PM
    • Professor
    • Graduate student
    • Graduate student

Thursday, June 2, 2005

    9:00-10:15AM
    • Professor
    • Graduate student
    • Graduate student
    • Graduate student

    11:00AM-12:15PM
    • Graduate student
    • Graduate student