21. Search Models and UIs for IR

INFO 202 - 10 November 2008

Bob Glushko

Plan for Today's Lecture

The "Classical" Model of Search and the "Classical" UI for IR

Web-based Search

Best practices for UIs in query specification, query results, and query reformulation
A Range of Information Needs

- Citations to documents
- Answers to specific questions
- Comparison of similar items
- Familiarization (browsing and "building upon")
- Knowledge discovery / data mining

Some General Usability Goals for IR Systems


- Strive for Consistency
- Provide Shortcuts
- Offer Informative Feedback
- Design for Closure
- Provide Simple Error Handling
- Permit Easy Reversal of Actions
- Support User Control
- Reduce Short-term Memory Load
The "Classical" IR System: Searching for Citations or Documents

Early in the digital era (1970s) large specialized bibliographic collections were created for academic articles, legal cases and opinions, news articles, etc.

The users of these systems were reference librarians, paralegals, journalists, and other professionals willing to be trained in using them.

Since these systems predated the PC and graphical terminals, their user interfaces used "command lines".

The command languages were complex and powerful, typically supporting Boolean, adjacency, and term stemming operators.

\[(\text{PCR OR POLYMERASE(}w\text{)CHAIN(}w\text{)REACTION? OR DNA(}w\text{)SEQUENC?}) \text{ AND} \]
\[(\text{CANCER? OR PRECANCER? OR NEOPLASM? OR CARCINO?})\]

The Classical Model of Search

Assumes a "go to the library and use the IR system" approach, where the user is either a trained searcher or has a trained searcher acting as an intermediary actually doing the search.

Assumes the information need remains the same throughout the search process.

Goal is to maximize recall and precision simultaneously.

The value is in the retrieved document set.
Processing Model in Classical Search

Query Specification in "Classic" Search UI

?s (biometric? and security)/TI,LP
  1621 BIOMETRIC?/TI,LP
  40268 SECURITY/TI,LP
S1  678 S (biometric? and security)/TI,LP
?s S1 and CS=(HARVARD AND MEDIC?)
S2  52 S1 and CS=(HARVARD AND MEDIC?)
?T S2/3/1-10
Challenges Posed by the "Classic" IR UI

Untrained users (e.g. with public access terminals in libraries) generally couldn't use these systems effectively on their own.

But effective use is essential in narrow and specialized domains where high recall is mandated.

Classic Query Specification "Webified" -- 1998
Classic Source and Field Selection "Webified" -- 1998

Classic Query Specification "Webified" -- 2007
Searching the Web

The size and scope of the Web is vastly greater than any "classic" bibliographic or document collection.

But the scope of what people search for is also vastly greater than in classical IR systems.

People expect to get information and documents, rather than just citations.

Challenges Posed by Searching the Web

UIs must accommodate differences among people in:

- Knowledge / life experience
- Cultural background and expectations
- Reading / scanning ability and style
- Methods of organizing and looking for things

UI functionality is constrained by heterogeneity of content (e.g., can't assume complete and consistent metadata or structure).

So what this means in practice is that the default web search UI is the simplest possible one: just a search box...

Additional challenges arise because people have multiple devices on which they want to search the Web and many have limited capabilities for UIs.
Google's Default Search UI

Ask's Default Search UI
Microsoft Live Search Default Search UI

Yahoo! Default Search UI
"Ordinary" People Just Don't Get IR

1 in 7 never type URLs in the address bar and others use it wrong
Some use "URL style" (no spaces between words) when entering words into query forms
Only 1 in 6 uses quotes in query forms, and many of these do so incorrectly
Almost no use of any advanced search syntax or functions
They don't appreciate the "vocabulary problem," so if their first query term doesn't work, they just give up rather than trying other terms

"Ordinary" People Don't Understand Boolean Operators

For most people, Boolean semantics are counterintuitive or backwards
- Boolean AND narrows a search, but natural language "and" implies a request for more information
- Likewise, Boolean OR is a union that widens a search, while "or" implies a mutually exclusive choice in everyday language
"Ordinary" People Don't Understand Text Processing in IR Systems

If very frequent words ("stop words") are removed by the search engine, a query like "To be or not to be" won't find anything.

A query like "boat fire" is different from "fire boat" but many users don't realize that term order matters.

The Search Process and Interface Components

Hearst says "the heart of the search process is an iterative cycle of query specification, inspection and interpretation of query results, and query reformulation"

- Query Specification: Selecting and structuring search terms
- Query Results: Ordered list of documents or other objects matching the query
- Query Reformulation: If nothing in the results satisfies the query, users modify their initial queries and submit new ones
Small Details Matter

UIs for IR require great care in small details because of the text-intensive nature of search

- Tension between more information and clutter

How and where to arrange components of the interface and results matters a lot (cf. Tidwell's principles)

- People don't read instructions or help text
- People scan / skim rather than read

Best Practices in Query Specification

Provide advanced capabilities for defining queries and constraining results, but progressively disclose them to hide complexity

Query suggestions

Query expansion and contraction (also used in query refinement stage) that is TRANSPARENT to the user

DWIM / spelling correction
Google Advanced Search

Windows Live Advanced Search
Query Refinement That No One Knows About

Search Window with Query History
Search Window with Subject Drop Down Menu

Search Window with Terse and Verbose Instruction
Query Suggestions in Yahoo!

Yahoo! Search

Search Results

Also try: weather forecast, berkeley, california, weather in berkeley, wind

Extended weather forecast for Berkeley, CA

- Currently: 55°F
- Tuesday: Sunny, 72°F
- Wednesday: Partly Cloudy, 67°F

Yahoo! Shortcut - About - Get local weather on your cell

1. Berkeley Weather Forecast and Conditions California (94704)
Berkeley weather forecast and weather conditions for 94704. ... Parks in My Area. Nearby Intersections. Event Venues Near Me. ... 
www.weather.com/weather/local/94704 - 132k - Cached - More from this site

2. Berkeley Weather Forecast and Conditions California (94707)
Berkeley weather forecast and weather conditions for 94707. ... Parks in My Area. Nearby Intersections. Event Venues Near Me. ... 
www.weather.com/weather/local/94707 - 132k - Cached - More from this site

Query Suggestions in Ask

Ask.com

Search Results

Weather in Berkeley, CA

- Forecast: 55°F
- Morning clouds. Cool.

Detailed Weather Report | 7-Day Forecast | Seasonal Weather

Current weather in Berkeley, CA

Current weather for a different location

The Weather Channel - Berkeley, CA (94720)
Berkeley weather forecast and weather conditions for 94720. Today's Berkeley California weather plus a 36 hour forecast and Doppler radar from

www.weather.com/weather/local/94720

Local Weather Forecast for Berkeley, CA (94702) - weather.com

Local Weather Forecast for Berkeley, CA (94720). Today's weather plus a 36 hour forecast and Doppler radar

www.weather.com/weather/local/94702Cached

More Results from www.weather.com

Maps

Berkeley, CA 94701

Current Time

Berkeley, CA

09:44:01 AM

Mon, 16 Jul 2007

News

County road proposal the wrong solution

Inland Packet

11 hours ago

Images
Query Suggestions in AltaVista

Google Presents a Spelling Correction
"Do What I Mean" mechanisms try to be "smart" and determine the searcher's unstated intentions or goals

Examples:

- Automatically suggest spelling corrections
- Automatically augment my query with related terms, synonyms, abbreviations, etc.
- Pop up "the paperclip" that tells me what kind of help I need

CRITICAL POINT: Users love DWIM when it works, but DESPISE it when it doesn't

Best Practices in Results Presentation

Present ranked results (people won't look past the first page) but don't show ranks

Sort of search results according to important criteria (date, author)

Group results according to well-organized category labels (see Flamenco)

Highlight query terms

Present query terms in context

Counter-intuitive failure to help: visualization
Search Query Term Highlighting

TIP
Firefox's Find bar can highlight all occurrences of the find text in the document if you click the Highlight All button.

NOTE
The Firefox Find bar—or, as it's sometimes called, the Find toolbar—has a number of controls. From the left they include an area in which to type characters to look for, a Find Next button, a Find Previous button, a Highlight All button, a Match Case button, and a message area.

Firefox always highlights the first occurrence of the term typed. Clicking the Highlight button tells Firefox to highlight all occurrences of the search term found on the page. Pressing F3 or clicking Find Next takes you to the next occurrence of the find text.

Using the Find bar's Find area, you can type a word(s). As you type, Firefox searches for what you type, finding and highlighting the first occurrence. As you type more letters, the search moves to the next match. For example, I typed a and Firefox found the first a. It

Search Query Term in Context

how to prevent cheese from molding

how to prevent cheese from molding

Preventing Moldy Cheese
It's the air that makes the mold on cheese so keeping air from getting at ... In the bowl I add a little corn starch which will prevent it from sticking. ... www.stretcher.com/stories/5601123b.cfm - 42k - Cached - Similar pages

How to Prevent Cheese Mold | eHow.com
How to Prevent Cheese Mold. It is frustrating to go into the refrigerator to get cheese for a recipe or sandwich and find that it has gone green with mold. ... www.ehow.com/how_1049885_prevent-cheese-mold.html - 49k - Cached - Similar pages
Grouping of Search Results

Interviews with lay users often reveal a desire for better organization of retrieval results

Useful for suggesting where to look next

Variety of techniques - categories vs clusters, single vs complex (faceted) category structure

Comparing the Techniques -- Clustering

Clustering is an automated technique for assigning results to groups (exclusively)

More flexible than pre-defined categories

Disambiguates ambiguous terms

Automatically generated labels can be unintuitive and occur at different levels of description

Potentially useful if the user wants a summary of the main themes in the subcollection

Potentially harmful if the user is interested in less dominant themes
Results Clustering by Clusty.com

Results Clustering by Vivisimo.com
Comparing the Techniques -- Categories

Human-created categories, but items can often be automatically assigned (to multiple categories)

Usually restricted to a fixed set

Intended to be readily understandable to those who know the underlying domain

Provide a novice with a conceptual structure
Results Categorization by Google

Vitamin A - Google Search - Mozilla Firefox

Refine results for Vitamin A:
- Drug uses
- Interactions
- For patients
- From medical authorities
- Side effects
- Warnings/recalls
- For health professionals

Results Categorization by Google [2]
Results Organized Using Faceted Categories (Flamenco)

Best Practices in Query Reformulation

(Do what works in query specification within the context of current results)

Make it easy to conduct iterative search by modifying queries to search within the current results

Get relevance feedback from searcher ("more like this")
Not all Suggested Refinements Work

<table>
<thead>
<tr>
<th>Refinement category</th>
<th>Examples (query / refinement)</th>
<th>Percent frequency of use as refinement</th>
<th>Percent frequency of use for new search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>triassic / triassic period</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Modifier</td>
<td>buckets wholesale / plastic buckets</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Elaboration</td>
<td>jackson pollack / museum of modern art</td>
<td>24</td>
<td>27</td>
</tr>
<tr>
<td>Location</td>
<td>vietnam / ho chi minh city</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Alternative</td>
<td>job listings / job openings</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Hyponym</td>
<td>birds of prey / falcons</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Morphological variant</td>
<td>norse myth / norse myths</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Syntactic variant</td>
<td>map of sudan / sudan map</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Acronym</td>
<td>usa maps / united states of america</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Spelling</td>
<td>stationary catalog / stationery</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Change</td>
<td>skateboards / mountainboards</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>

Query Refinement Based on Relevance Feedback

Many studies show that if users engage in relevance feedback, the results are much better.

But the explicit effort required to rate relevance is usually a roadblock.

This is one motivation for using "social" or indirect methods of assessing relevance.

- Chapter 1: Information retrieval using the Boolean model
- Chapter 2: The term vocabulary & postings lists