IR User Interfaces and Visualization

Norbert Fuhr

June 5, 2008
Human-Computer interaction

[Shneiderman]: ‘Well designed, effective computer systems generate positive feelings of success, competence, mastery and clarity in the user community. When an interactive interface is well-designed, the interface almost disappears, enabling users to concentrate on their work, exploration, or pleasure.’
Design Principles

- offer information feedback
e.g. query formulation vs. documents retrieved
- permit easy reversal of actions
e.g. keep track of choices during search,
provide browsable info relevant for current stage of search
- reduce working memory load
- provide alternative interfaces for novice and expert users
  (easy to learn v.s. powerful interfaces)
Visualization techniques

- icons
- color highlighting
- brushing and linking (connect two or more views of the same data)
- panning and zooming
- focus-plus-context
- magic lenses (transformations on the focussed data)
- tree views
- use of animation
The information access process

goals + tasks

1. start with an information need
2. select a system and a collection to search on
3. formulate a query
4. send the query to the system
5. receive the results in the form of information items
6. scan, evaluate and interpret the results
7. either stop, or
8. reformulate the query and go to step 4.
Standard model for the search process

1. Information Need
2. Query
3. Send to System
4. Receive Results
5. Evaluate Results
6. Reformulate
   - No
   - Done?
      - Yes
         - Stop
      - No
         - Reformulate
Alternative models

**empirical studies:**
- information seeking consists of a set of interconnected but diverse searches
- search results for a goal trigger new searches
- context of the problem is carried over
- main value of searching is accumulated learning and acquisition of information during the search

Bates **Berripicking model:**
1. continuous shift of information need and queries during search process
2. information need is not satisfied by a single set of documents instead: series of selections and bits of information acquired throughout the search
Erweitertes Modell für interaktives Retrieval


Dokumente

Repräsentation

Selektion

Projektion

Mengendarst.

Dokumente mit Rep. selektierte Dok. projizierte Dok. Ergebnismenge

Formale Anfrage Interaktion

Benutzer

Resultatkorb
Non-search parts of the information access process

80% fall into
- finding trends
- making comparisons
- aggregating information
- identifying a critical subset
- assessing
- interpreting

20 %:
- cross-referencing
- summarizing
- finding evocative visualizations for reports
- miscellaneous
Starting points
lists of collections

LEXIS-NEXIS source selection screen
Source and document type selection in Daffodil
Category or directory overviews based on classification system (e.g. Yahoo!)

MeSHBrowse interface
Cone tree
Figure 5.
Horizontal Cone Tree visualization of a directory hierarchy
HiBrowse: view category labels hierarchically and according to facets (disease, therapy, groups)
Automatically derived collection overviews

Scatter-Gather clustering retrieval results
Three-dimensional overview based on document clustering
Kohonen feature map (2D overview)
Kohonen feature map: SOMlib
Examples, dialogs and wizards

retrieval by reformulation start with example, iterative query reformulation

Interactive dialog mimic human search intermediary

Wizard step-by-step shortcut through sequence of menu choices

guided tour sequence of navigational choices through hypertext links
Query specification

**HCI styles:**
- command language
- form filling
- menu selection
- direct manipulation
- natural language
Form filling
Form filling: syntax and spell check in Daffodil
Form filling: proposal of related terms in Daffodil
Graphical approaches to query specification

Simple filters

Magic lens interface for query specification (lenses act as filters)
Graphical approaches to query specification
Boolean queries

Block-oriented diagram visualization of Boolean queries starting from natural language query, direct manipulation of blocks
Filter-flow visualization of Boolean queries
VQuery Venn diagram visualization
WOB model: matrix for Boolean query formulation
Presenting retrieval results

a) Single result elements: Surrogates
b) Organization of result set
Document surrogates

- ordering criterion: RSV / attribute value (e.g. date)
- document entry:
  - title
  - metadata (date, source, length)
- document summary
  - query-independent
  - query-dependent
- clicking on document icon/title retrieves complete document
Information Retrieval

... Information Retrieval. Publikationen im Netz. Information Retrieval von C.J. van Rijsbergen (1979)
Intelligente Information Retrieval Systeme von A ...
http://www.inf-wiss.uni-konsta ...
Treffer aus der Katalog-Rubrik > Veranstaltungen

Wikipedia, der freien Enzyklopädie. Information Retrieval (IR) (... Information Retrieval (IR) (Informationswiedergewinnung, gelegentlich Informationsbeschaffung) ist ein Fach ... Methoden des
Information Retrieval werden in Internetsuchmaschinen (z ...
http://de.wikipedia.org/wiki/I ...

Information Retrieval
... Information Retrieval. 1995 this conference was merged into HIM ... Jürgen Krause, Christa
Wormser-Hacker (Eds.): Information Retrieval '93: Von der Modellierung zur Anwendung ...
http://www.informatik.uni-trie ...

Information Retrieval
University of Duisburg, Information Systems. ... Information Retrieval. Fachgebiet:
Informationssysteme ... Fuhr:93c. Titel: Information Retrieval. Autor(en): Fuhr, N ...
http://www.is.informatik.uni-d ...
Example of ranked list of titles + surrogate info


AN IMPROVED TREATMENT OF EXTERNAL BOUNDARY FOR THREE-DIMENSIONAL FLOW COMPUTATIONS? Semyon V. Tsykinov Veer N. Vatsar NASA Langley Research Center, Hampton, VA Abstract We present an innovative numerical approach for setting highly accurate nonlocal boundary conditions at the external computational

National Aeronautics and Space Administration Langley Research Center Hampton, Virginia 23681-2199NASA Technical Paper 3631Multistage Schemes With Multigrid for Euler and Navier-Stokes Equations Components and Analysis R. C. Swanson Langley Research Center Hampton, Virginia Eli Turkel Tel-Aviv University

A Distributed Garbage Collection Algorithm Terence Critchlow UUCS-92-11 Department of Computer Science University of Utah Salt Lake City, UT 84112 USA July 30, 1992 Abstract Concurrent Scheme extends the Scheme programming language, providing parallel program execution on a distributed network. The
Query term hits within document content

- highlighting
- KWIC (keyword in context)
- TileBars
- SeeSoft
27th Annual International ACM SIGIR Conference Workshop on Peer-to-Peer Information Retrieval

SIGIR is the major international forum for the presentation of new research results and the demonstration of new systems and techniques in the broad field of information retrieval.

This SIGIR workshop on Peer-to-Peer Information Retrieval focus on new methods of resource representation, resource selection, and data fusion in peer-to-peer networks. The workshop particularly encourages papers that address heterogeneous peer-to-peer networks, as well as papers about methods that cope with partial and uncertain information. However, more broadly, papers are solicited on any topic related to information retrieval in peer-to-peer networks.
KWIC: keyword in context

KWIC is an acronym for Keyword In Context, the most common format for concordance lines. It is the most common format for concordance lines. In Context, the most common format for concordance lines. Encyclopedia, The Free Encyclopedia, The Encyclopedia, The Free Encyclopedia, The Encyclopedia, The Free Encyclopedia.
Organization of result set

place document set in the context of other information types, in order to make it more understandable
relate document set to

▶ query terms
▶ Boolean query structure
▶ collection overviews
▶ descriptive metadata
▶ hyperlink structure
▶ document structure
▶ other documents
TileBars: Weight and location of query terms

User Query:
(Enter words for different topics on different lines.)
osteoporosis
prevention
research

Run Search  New Query  Quit
Search Limit: 50 100 250 500 1000
Number of Clusters: 3 4 5 8 10

Mode: TileBars
Cluster  Titles

FR88513–0157
AP: Groups Seek $1 Billion a Year for Aging Research

SJMN: WOMEN’S HEALTH LEGISLATION PROPOSED
AP: Older Athletes Run For Science

FR: Committee Meetings
FR: October Advisory Committees; Meetings
FR88120–0046
FR: Chronic Disease Burden and Prevention Models; Program
AP: Survey Says Experts Split on Diversion of Funds for AIDS

FR: Consolidated Delegations of Authority for Policy Developm
SJMN: RESEARCH FOR BREAST CANCER IS STUCK IN P
VIBE: result documents in term space
InfoCrystal: Query term hits between documents
Eingabefeld

Definition eines Suchkriteriums
Definition eines zweiten Suchkriteriums

Definition eines dritten Suchkriteriums
Probabilistisches Retrieval
Probabilistisches Retrieval mit manueller Umgewichtung
Eibl: Evaluierungsergebnisse

**Precision**
- Visualisierung: 66
- Messenger: 61,6
- freeWAIS: 51,8

**Recall**
- Visualisierung: 15,6
- Messenger: 12,7
- freeWAIS: 10,5
SuperBook: context via table of contents
Problems with SuperBook:

- Users wander off by following links
  - SB uses automatic linking to any other occurrence of the same word in the hypertext
  - more discriminating links required
- Users form better mental models when a hierarchical structure is given
Evaluation of Superbook:

- comparison with paper and standard online information access
- comparison based on set of tasks: browsing topics of interest, citation searching, searching for question answering, browsing for writing summary
- results:
  - SB users faster in searching tasks when search terms occur in text
  - SB not faster when terms occur neither in the document text nor in the tables of contents
Categories for results set context

DynaCat interface for viewing category labels that correspond to query types
Using hyperlinks to organize retrieval results
Cha-Cha: SuperBook on the Web

based on link structure of web site (shortest paths from root)
Mapuccio: graphical depiction of link structure (between retrieved pages)
### Envision Query Window

**Query History:**
- 1. User interface design
- 2. A:Card, Stuart K.

**Query #2.1**

**Authors:**
- Card, Stuart K.

**Words in Title:**
- human-computer interaction

**Content Words:**

**Best 75 Items Found**

<table>
<thead>
<tr>
<th>Icon Label</th>
<th>Relevance Rank</th>
<th>Icon Color</th>
<th>Est. Relevance</th>
<th>Icon Shape</th>
<th>Item Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Item</td>
</tr>
</tbody>
</table>

| V-Axes: | Author | Card, S. K. | 37 | 68 |
|         |        | Card, Stuart |   |    |
|         |        | Card, Stuart K. | 2 | 22 |
|         |        | Carroll, John W. |   |    |
|         |        | Entin, Eilson E. |   |    |
|         |        | Fuchs, Henry | 43 |    |
|         |        | Hartson, H. R. |   |    |
|         |        | Henderson, D. Austin | 15 | 22 |
|         |        | Hill, Ralph D. |   |    |

**Shape Legend**
- Book
- Article
- Proceedings Article

**Color Legend**
- Least Relevant
- More Relevant
- Most Relevant
- Not Useful

**X-Axes:**
- Pub. Year

### Envision Item Summary: Query #2.1

<table>
<thead>
<tr>
<th>Icon #</th>
<th>Useful</th>
<th>Est. Rel</th>
<th>Author/Editor</th>
<th>Year</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>369</td>
<td>Card, Stuart K.</td>
<td>1989</td>
<td>Understanding key constraints governing human-computer interfaces</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>348</td>
<td>Card, Stuart K.</td>
<td>1983</td>
<td>The Psychology of Human-Computer Interaction</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>302</td>
<td>Card, Stuart K.</td>
<td>1994</td>
<td>Human Limits and the VDT Computer Interface</td>
</tr>
</tbody>
</table>

**Envision search results display.**
TableLens visualization (for structured data)
Perspective Wall
Using relevance judgments
Example UI for relevance feedback

how to handle docs shown before?
users often reluctant to give RF—'more like this' in Web search engines
User control over relevance feedback:

Control: No relevance feedback

Opaque: Users give relevance judgments, see revised ranking

Transparent: Users see how system reformulates the query, + revised rankings

Penetrable: Users may modify the query reformulated by the system, then see resulting ranking

evaluation:
penetrable > opaque, transparent > control
Fetching relevant information in the background

- assistants / recommender agents
- learn (persistent) user profile based on relevance judgments when browsing Web pages, annotate links in new Web pages according to estimated relevance
- Letizia: recommendations for further actions (follow links)
- Butterfly: prefetching of potentially relevant information
Group relevance judgments

social recommendation approach:

1. find users with similar interests (relevance judgments)
2. recommend documents not seen yet, but judged positively by similar users
Interface support for the search process

.Interfaces for string matching

Mozilla’s string matching interface
Example: string matching over a list (Netscape)
combination of search formulation and search results
  ▶ re-use same window: result may belong to former query
  ▶ new window: user has to reorganize windows
  ▶ stacked windows
Window management

divide functionality into different views (displayed simultaneously): monolithic (tiled windows) vs. flexible (Overlapping windows)

▶ overlapping windows: much user activity for moving between functionally related windows

▶ workspace: grouping of functionally related windows

▶ multiple virtual workspaces: 3D spatial metaphor, with virtual doors in between

▶ persistent sessions
Example diagrams of monolithic layouts
The DLITE interface
queries, sources, documents and groups of retrieved docs as objects
query formulation = creation of query object
submit query = move query object to collections
tool-tips
animation (e.g. for indicating illegal operations)
The Butterfly interface (browsing citations)
The SketchTrieve interface
Retaining search history

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Datenbank</th>
<th>Suchfrage</th>
<th>Treffer</th>
<th>Titelliste (Suchfragen)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>INSPEC</td>
<td>MULTIMEDIA AND RETRIEVAL</td>
<td>6935</td>
<td>Anzeige</td>
</tr>
<tr>
<td>2.</td>
<td>INSPEC</td>
<td>MULTIMEDIA SAME RETRIEVAL</td>
<td>5057</td>
<td>Anzeige</td>
</tr>
<tr>
<td>3.</td>
<td>INSPEC</td>
<td>USER SAME INTERFACE</td>
<td>23244</td>
<td>Anzeige</td>
</tr>
<tr>
<td>4.</td>
<td>INSPEC</td>
<td>2 AND 3</td>
<td>346</td>
<td>Anzeige</td>
</tr>
<tr>
<td>5.</td>
<td>INSPEC</td>
<td>&quot;USER INTERFACE&quot;</td>
<td>15363</td>
<td>Anzeige</td>
</tr>
<tr>
<td>6.</td>
<td>INSPEC</td>
<td>&quot;MULTIMEDIA RETRIEVAL&quot;</td>
<td>156</td>
<td>Anzeige</td>
</tr>
<tr>
<td>7.</td>
<td>INSPEC</td>
<td>5 AND 6</td>
<td>8</td>
<td>Anzeige</td>
</tr>
</tbody>
</table>

Suchbegriff eingeben
oder Suchschrittknopfung eingeben mit AND (einschränkend), OR (erweiternd), NOT (ausschließend)

Search history in INSPEC
VISAGE interaction history visualization
Integrating Scanning, selection, and querying

Query history revision in the Melvyl bibliographic catalog
The Cat-a-Cone interface for integrating category and text scanning
Starting point: searching over category labels