# Interactive Information Retrieval

Norbert Fuhr

University of Duisburg-Essen

German IR Autumn School, Schloss Dagstuhl, 2012

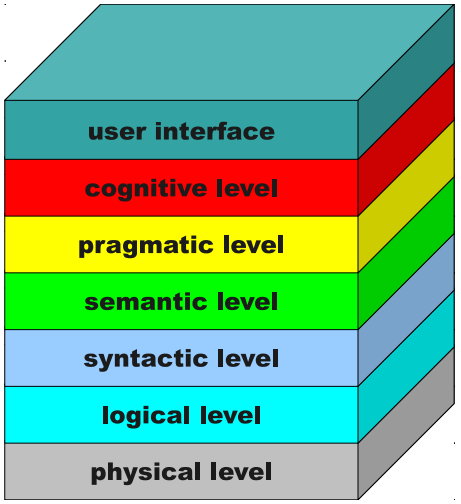
UNIVERSITÄT DUISBURG ESSEN

## Contents



## Architecture Levels of IR Systems

Architecture Levels of IR Systems



## Levels

- physical level: data structures and algorithms
- **logical level**: IR models
- syntactic level: focus on syntactic properties of objects (text as character sequence, image as pixel matrix, ...)
- semantic level: objects and their relationships → *ontologies, ...*
- pragmatic level: meaning of a document/application issue → *definition of relevance*
- **cognitive level**: user support during search *process*
- **user interface**

## Remarks on architecture levels

most systems comprise only some of these levels

- users have to compensate the missing levels (especially cognitive and pragmatic level)
- missing separation of levels (e.g. no logical or physical data independence as in DBMS)

## Models for Interactive Retrieval



## Classical Probability Ranking Principle

defines optimum retrieval for probabilistic models:

ranking documents according to decreasing values of the

probability of relevance

yields

optimum retrieval quality

**Restrictions**

- Relevance judgments of documents are independent
- Focus on user's assessment of result list

## Interactive Retrieval

- User has a rich set of interaction possibilities
  - (re)formulate query
  - selection based on summaries of various granularity
  - select related terms from list
  - follow document link
  - relevance judgment
- Information need changes during a search

No theoretic foundation for constructing IIR systems!

## Requirements for an IIR-PRP

[Fuhr 08]

- Consider the complete interaction process
- Allow for different costs for different activities
- Allow for changes of the information need

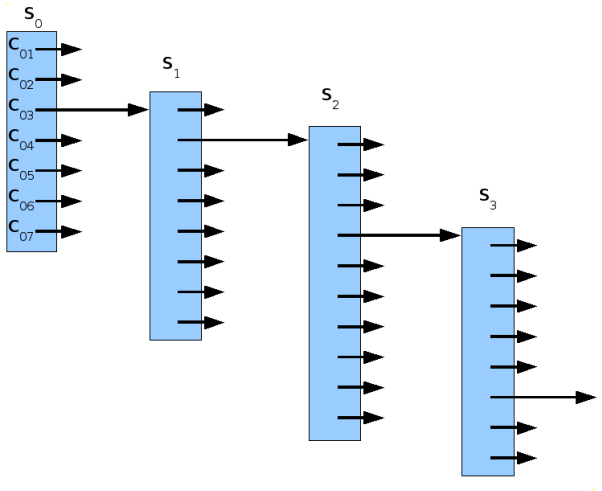
## Basic Assumptions

- Focus on a functional level of interaction
  (usability issues disregarded here)
- System presents list of choices to the user
- Users evaluate choices in linear order
- Only positive decisions/choices are of benefit for a user

## Examples of decision lists

- ranked list of documents
- list of summaries
- list of document cluster
- KWIC list
- list of expansion terms
- links to related documents
- ...

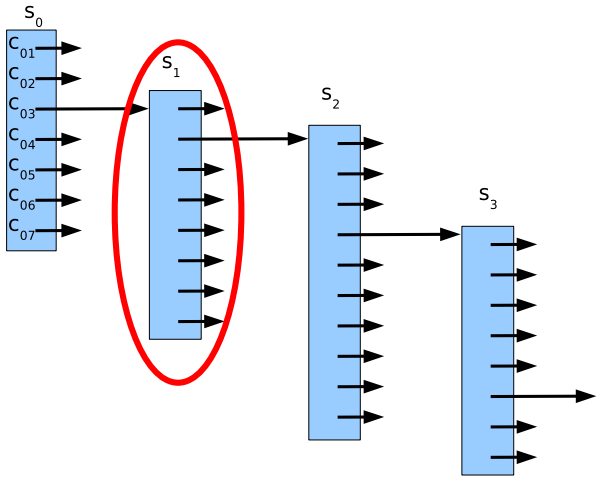
## Abstraction: Situations with Lists of Choices




## Basic ideas

- A user moves from situation to situation
- In each situation $s_i$, the user is presented a list of (binary) choices $< c_{i1}, c_{i2}, \ldots, c_{i,n_i} >$
- The user decides about each of these choices sequentially
- The first positive decision moves the user to a new situation $s_j$
- A decision may be wrong, requiring backtracking

## Probabilistic model focusing on single situation




## Expected Benefit of a choice

$p_{ij}$ probability that the user will accept choice $c_{ij}$

$q_{ij}$ probability that this decision was right

$e_{ij}$ $< 0$: effort for evaluating the choice $c_{ij}$

$b_{ij}$ $> 0$: resulting benefit from positive, correct decision

$g_{ij}$ $\leq 0$: cost for correcting a wrong decision

**Expected benefit of choice** $c_{ij}$

$$E(c_{ij}) = e_{ij} + p_{ij}\,(q_{ij}\,b_{ij} + (1 - q_{ij})g_{ij})$$

## Example

Web search: 'Java' $\rightarrow$ $n_0$=290 mio. hits

System proposes extension terms:

| term | $n_i$ | $p_{ij}$ | $b_{ij}$ | $p_{ij}b_{ij}$ |
|---|---|---|---|---|
| program | 195 mio | 0.67 | 0.4 | 0.268 |
| blend | 5 mio | 0.02 | 4.0 | 0.08 |
| island | 2 mio | 0.01 | 4.9 | 0.049 |

benefit $b_{ij} = \log \frac{n_0}{n_i}$

## Strategies for maximizing expected benefit

$$E(c_{ij}) = e_{ij} + p_{ij}\left(q_{ij}b_{ij} + (1 - q_{ij})g_{ij}\right)$$

(assume that benefit $b_{ij}$ and corr. effort $g_{ij}$ are given)

1. minimize effort $|e_{ij}|$ —
   but keep $p_{ij}$ (selection prob.) and $q_{ij}$ (success prob.) high
2. maximize $p_{ij}$: user should choose $c_{ij}$ whenever it is appropriate —
   but keep success probability $q_{ij}$ high
   $\rightsquigarrow$ increased effort $e_{ij}$
3. maximize $q_{ij}$ by avoiding erroneous positive decisions
   $\rightsquigarrow$ increased effort $e_{ij}$

## Further remarks

$$E(c_{ij}) = e_{ij} + p_{ij}\left(q_{ij}b_{ij} + (1 - q_{ij})g_{ij}\right)$$

- Expected benefit should be positive
  choices with negative values should not be presented to a user.
- Methods for estimating parameters $p_{ij}$, $q_{ij}$, $b_{ij}$, $e_{ij}$, $g_{ij}$:
  Issue of further research
- In the following, let $a_{ij} = q_{ij}b_{ij} + (1 - q_{ij})g_{ij}$
  ("average benefit")

$$E(c_{ij}) = e_{ij} + p_{ij}a_{ij}$$

## Example for Expected Benefit

After formulating a query, a user may choose to perform the following actions with the corresponding parameter triple $(e_{ij}, p_{ij}, a_{ij})$

1. $(-1.0, 0.3, 8)$ add expansion term to the query
2. $(-2.0, 0.4, 10)$ look at the first result list entry
3. $(-10.0, 0.4, 25)$ immediately go to the first document
4. $(-5.0, 0.3, 20)$ look at an aggregated summary of the top ranking documents

In which order should these choices be presented to the user?

1. $(-1.0 + 0.3 \cdot 8) = 1.4$
2. $(-2.0 + 0.4 \cdot 10) = 2$
3. $(-10.0 + 0.4 \cdot 25) = 0$
4. $(-5.0 + 0.3 \cdot 20) = 1$

## Expected benefit of a choice list

situation $s_i$ with list of choices $r_i =< c_{i1}, c_{i2}, \ldots, c_{i,n_i} >$

expected benefit of choice list:

$$\begin{aligned} E(r_i) &= e_{i1} + p_{i1}a_{i1} + \\ &\quad (1-p_{i1})\,(e_{i2} + p_{i2}a_{i2} + \\ &\quad (1-p_{i2})\,(e_{i3} + p_{i3}a_{i3} + \\ &\quad \ldots \\ &\quad (1-p_{i,n-1})\,(e_{in} + p_{in}a_{in})\ )) \\ &= \sum_{j=1}^{n} \left( \prod_{k=1}^{j-1} (1-p_{ik}) \right) (e_{ij} + p_{ij}a_{ij}) \end{aligned}$$

## Ranking of choices

Consider two subsequent choices $c_{il}$ and $c_{i,l+1}$

$$E(r_i) = \sum_{\substack{j=1 \\ l \neq j \neq l+1}}^{n} \left( \prod_{k=1}^{j-1} (1-p_{ik}) \right) (e_{ij} + p_{ij}a_{ij}) + t_i^{l,l+1}$$

where

$$\begin{aligned} t_i^{l,l+1} &= (e_{il} + p_{il}a_{il}) \prod_{k=1}^{l-1} (1-p_{ik}) + \\ &\quad (e_{i,l+1} + p_{i,l+1}a_{i,l+1}) \prod_{k=1}^{l} (1-p_{ik}) \end{aligned}$$

analogously $t_i^{l+1,l}$ for $< \ldots, c_{i,l+1}, c_{il}, \ldots >$

## Difference between alternative rankings

$$\begin{aligned} d_i^{l,l+1} &= \frac{t_i^{l,l+1} - t_i^{l+1,l}}{\prod_{k=1}^{l-1}(1-p_{ik})} \\ &= e_{il} + p_{il}a_{il} + (1-p_{il})(e_{i,l+1} + p_{i,l+1}a_{i,l+1}) - \\ &\quad \left(e_{i,l+1} + p_{i,l+1}a_{i,l+1} + (1-p_{i,l+1})(e_{il} + p_{il}a_{il})\right) \\ &= p_{i,l+1}(e_{il} + p_{il}a_{il}) - p_{il}(e_{i,l+1} + p_{i,l+1}a_{i,l+1}) \end{aligned}$$

Necessary condition for the maximum expected benefit of the list: $d_i^{l,l+1} \stackrel{!}{\geq} 0$, which leads to

$$a_{il} + \frac{e_{il}}{p_{il}} \geq a_{i,l+1} + \frac{e_{i,l+1}}{p_{i,l+1}}$$

## PRP for Interactive IR

$$a_{il} + \frac{e_{il}}{p_{il}} \geq a_{i,l+1} + \frac{e_{i,l+1}}{p_{i,l+1}}$$

$\rightsquigarrow$ **Rank choices by decreasing values of**

$$\varrho(c_{ij}) = a_{il} + \frac{e_{il}}{p_{il}}$$

## Expected Benefit vs. Ranking Criterion

$$\begin{aligned} \text{expected benefit:} \quad E(c_{ij}) &= p_{ij} a_{ij} + e_{ij} \\ \text{ranking criterion:} \quad \varrho(c_{ij}) &= a_{il} + \frac{e_{il}}{p_{il}} \end{aligned}$$

| choice | $p_{ij}$ | $a_{ij}$ | $e_{ij}$ | $E(c_{ij})$ | $\varrho(c_{ij})$ |
|---|---|---|---|---|---|
| $c_1$ | 0.5 | 10 | -1 | 4 | 8 |
| $c_2$ | 0.25 | 16 | -1 | 3 | 12 |

Expected benefits of the 2 possible lists:

$$\begin{aligned} E(< c_1, c_2 >) &= 4 + 0.5 \cdot 3 = 5.5 \\ E(< c_2, c_1 >) &= 3 + 0.75 \cdot 4 = 6 \end{aligned}$$

## IIR-PRP vs. PRP

$$a_{il} + \frac{e_{il}}{p_{il}} \geq a_{i,l+1} + \frac{e_{i,l+1}}{p_{i,l+1}}$$

Assumptions for classical PRP:

1. constant effort for each document $e_{ij} = -E$, $E > 0$
2. constant benefit from each relevant document $a_{il} = B$

$$\begin{aligned} B - \frac{E}{p_{il}} &\geq B - \frac{E}{p_{i,l+1}} \\ \Rightarrow p_{il} &\geq p_{i,l+1} \end{aligned}$$

$\rightsquigarrow$ Classic PRP still holds!

## IIR-PRP: Observations

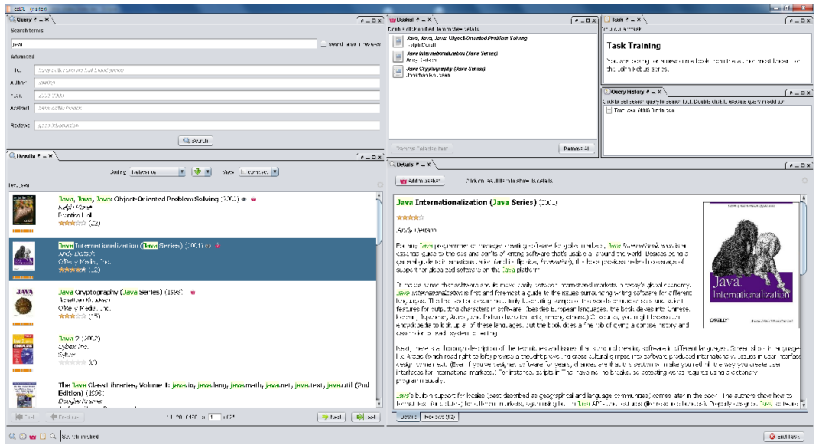
$$\text{Rank choices by} \quad a_{ij} + \frac{e_{ij}}{p_{ij}}$$

- $p_{ij}$ 'probability of relevance' still involved
- tradeoff between effort $e_{ij}$ and benefit $a_{ij}$
- difference between PRP and IIR-PRP due to variable values for $e_{ij}$ and $a_{ij}$
- IIR-PRP looks only for the first positive decision

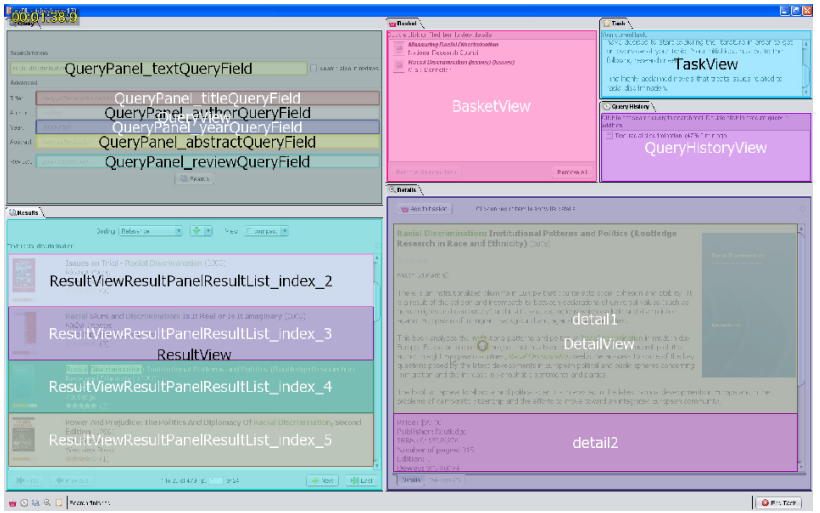
## Parameter estimation

1. Selection probability $p_{ij}$:
   focus of many IR models,
   but models for dynamic info needs required
2. Effort parameters $e_{ij}$, $g_{ij}$ +success probability $q_{ij}$:
   most research needed
3. Benefit $b_{ij}$:
   - information value ?
   - saved effort

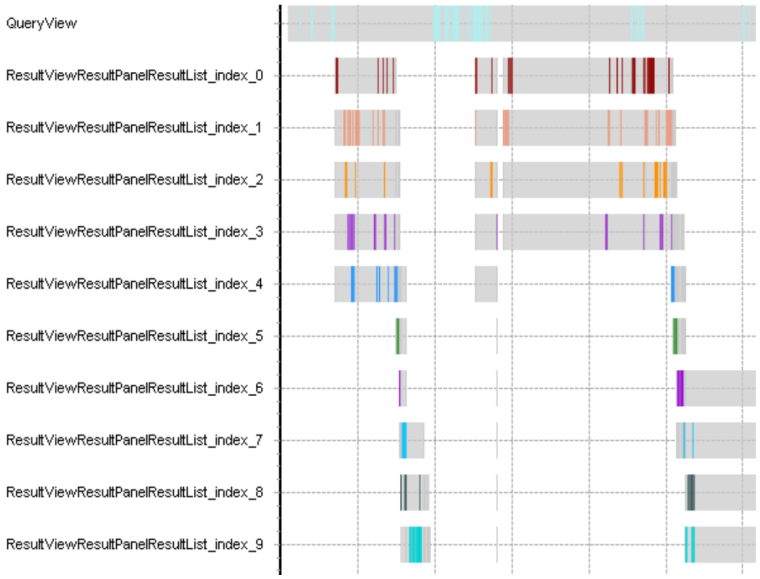
## User Interface



## Areas of Interest for Gaze Tracking




## AOI Sequence

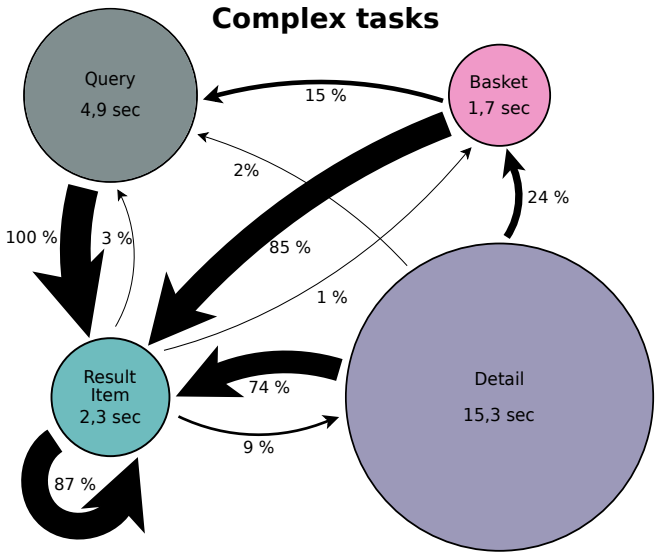



## Task setting

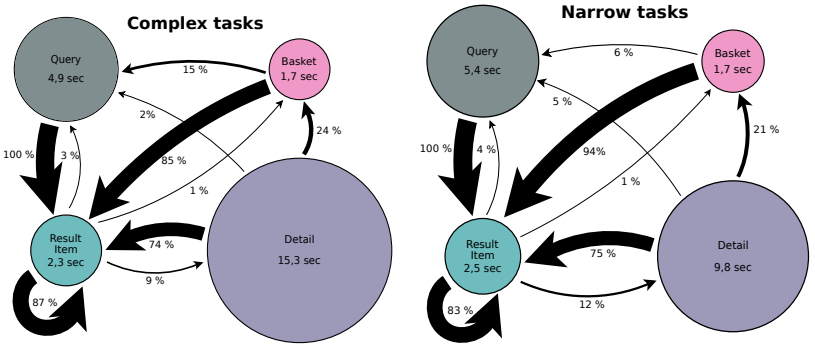
[Tran & Fuhr 12]

- Retrieval experiments with 12 subjects
- Users had to work on two tasks for 15 minutes each
  - complex tasks: consideration of user reviews necessary for judging relevance
  - narrow tasks: reading of book abstracts sufficient
- Users judge individually about relevance (by placing items in the basket

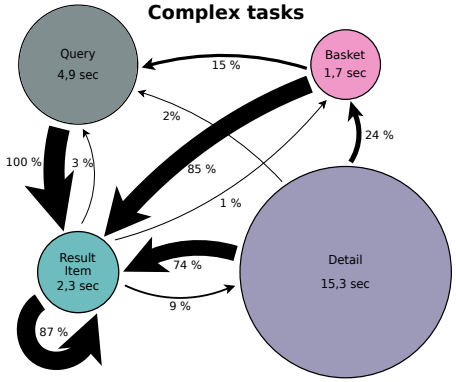
## Markov Model for Complex Tasks




## Markov Models Complex vs. Narrow Tasks




## Estimating the iPRP parameters




- Effort: time spent in a situation
- Acceptance probability = transition probabilities
- Benefit?

## Expected time for reaching the basket

- effort in states $t_q$, $t_r$, $t_d$ and $t_b$
- $p_{XY}$: transition probability from state $X$ to state $Y$
- expected times $T_q$, $T_r$ and $T_d$ for reaching the basket state

$$
\begin{aligned}
T_q &= t_q + p_{qr} T_r \\
T_r &= t_r + p_{rq} T_q + p_{rr} T_r + p_{rd} T_d \\
T_d &= t_d + p_{dq} T_q + p_{dr} T_r
\end{aligned}
$$

## Expected times and benefits

| | complex | narrow |
|---|---|---|
| $T_q$ | 127.9 | 120.8 |
| $T_r$ | 123.0 | 115.4 |
| $T_d$ | 109.5 | 102.4 |
| $b_q$ | 4.9 | 5.4 |
| $b_r$ | 17.7 | 14.7 |
| $b_d$ | 15.9 | 10.7 |

## IPRP for Nonbinary Choices

- $c_{ij}$: choice
- $m_{ij}$ alternatives $c'_{ijk}$, $k = 1, \ldots, m_{ij}$
- $a'_{ijk}$ corresponding benefits
- $q_{ijk}$ selection probabilities such that

$$\sum_{k=1}^{m_{ij}} q_{ijk} = 1 \quad \text{and} \quad \sum_{k=1}^{m_{ij}} q_{ijk}\, a'_{ijk} = a_{ij}.$$

Example: examining a result item in a complex task:

$$b_r = \frac{0.03(T_q - T_r) + 0.01(T_r - 0) + 0.09(T_d - T_r)}{0.03 + 0.01 + 0.09} = 17.7s$$

# Information Seeking Behavior

Information Seeking Behavior and Information Searching
Ellis' Behavioral Model of Information Seeking Strategies

## Models of information searching

- **classic IR**
  - content-oriented search in unstructured documents
  - vague information needs, uncertain representations
  - system-oriented view, assume static information need
- **Interactive information retrieval**
  - focus on user interaction with information system
  - dynamic information need
- **2 views on interactive IR:**
  1. Information Seeking Behavior
  2. Information Searching

## Information Seeking Behavior

- broader view than content-oriented search
- model user's actions, motivations and strategies for satisfying an information need
- questions of interest:
  - what triggers an information need?
  - what are users doing for solving this problem?

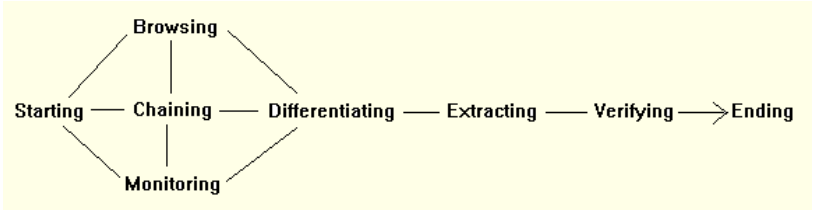
## Information Searching

- focus on user's interaction with information sources
- regard classic IR systems as well as other sources (e.g. personal communication)

## Ellis' Behavioral Model of Information Seeking Strategies

[Ellis 89]

- general model of search behavior
- based on empirical studies in social sciences and engineering companies
- general categories or properties of search behavior: Starting, Chaining, Browsing, Differentiating, Monitoring, Extracting, Verifying, Ending




## Categories of search behavior according to Ellis (1)

**Starting**

- get overview of literature/locate key authors in a field, e.g. by
  - selection of information source (e.g. personal collection, digital library, Web search engine)
  - review articles
  - personal contacts

## Categories of search behavior according to Ellis (2)

**Chaining**

- follow different forms of referential connections between sources (in both directions)
  - citations
  - Web links
  - same author/ research team
  - same conference / journal issue
  - same category
- factors considered:
  - topical relevance
  - popularity of author
  - timeliness
  - citation frequency
  - cost and time for document acquisition
- leads to finding new sources or even to reformulation of information need

## Categories of search behavior according to Ellis (3)

**Browsing**

- starts from information sources and retrieved documents
- semi-goal-oriented search by browsing in promising areas
- scanning of tables of contents, references, lists of people and organizations
- browsing is used when relevant information is available in a comprehensive way

## Categories of search behavior according to Ellis (3)

**Differentiating**

- judging of sources according to type, quality, importance, usefulness
- leads to information filtering
- e.g. comment vs. report, specification vs. manual

## Categories of search behavior according to Ellis (4)

**Monitoring**

- maintain awareness of developments and technologies in a field
- by following particular sources
  - formal channels: scientific journals, conferences, alert profiles
  - informal channels: personal contacts, actual practice (field research, experimental work)

## Categories of search behavior according to Ellis (5)

**Extracting**

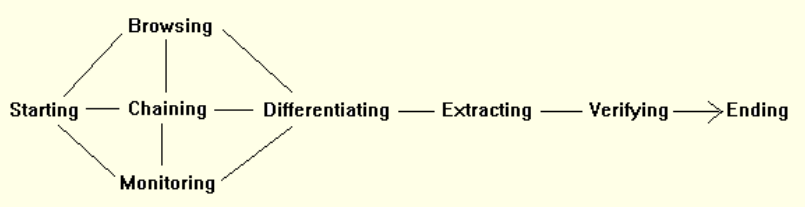
- working through sources to locate material of interest
- material: documents, new sources, passages
- cognitive capture of information by the user
- user's background knowledge important

## Categories of search behavior according to Ellis (6)

**Verifying**
check information wrt. correctness and reliability

**Ending**
end of search, linking of new information with previous knowledge

## Process model




- no strict sequential process
- starting, browsing, chaining and monitoring are search procedures
- differentiating is a filtering step

## Extension by Meho/Tibbo

- Repeated Ellis' study, new analysis (especially wrt. new technologies)
- confirmation of Ellis' model
- but: extension by new categories

## Additional categories

**Accessing**

- Access to full texts (instead of surrogates)
- acquisition of contents via different channels and with different costs

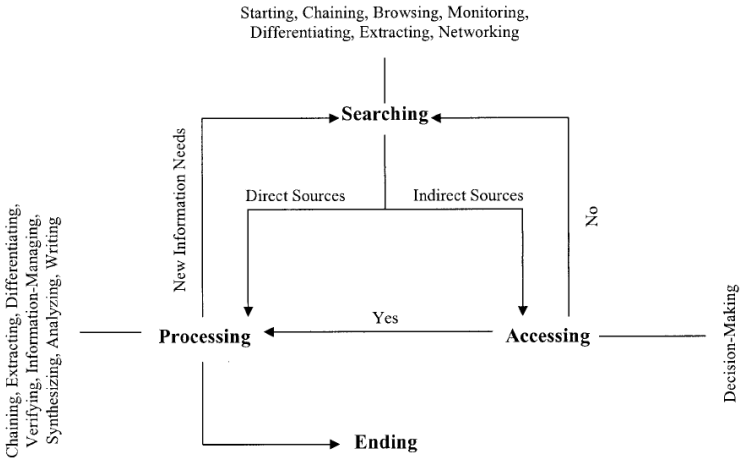
**Networking**

- personal communication with various persons
- discussion and evaluation of retrieved information via internet/intranet fora

**Information Managing**

- filing, storing and organizing retrieved and used information

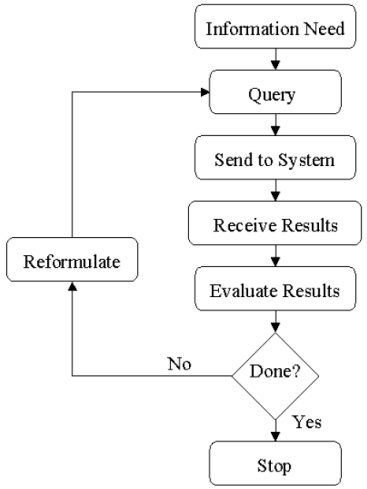
## Phases in search behavior




# Information Searching

Simple models of the search process
Patterns of search behavior
Search activities
Anomalous State of Knowledge
Ingwersen's Cognitive Model
Classification of search activities

## Classical search process model



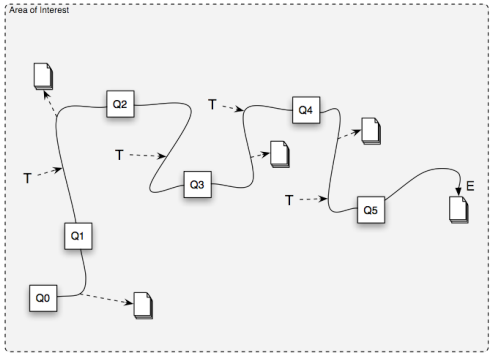
## Empirical studies

- information search consists of a sequence of connected, but different searches
- search result may trigger new searches
- only task context remains the same
- main goal of a search is accumulated learning and collection of new information while searching

## Berrypicking Model

[Bates 90]

- continuous change of information need and queries during search
- information need cannot be satisfied by a single result set
- instead: sequence of selections and collection of pieces of information during search




## Support for Berrypicking

- Filing of single results
- Adding terms/items to the query

## A taxonomy of Web search

[Broder 2002, Rose & Levinson 2004]

Navigational: to reach a particular site

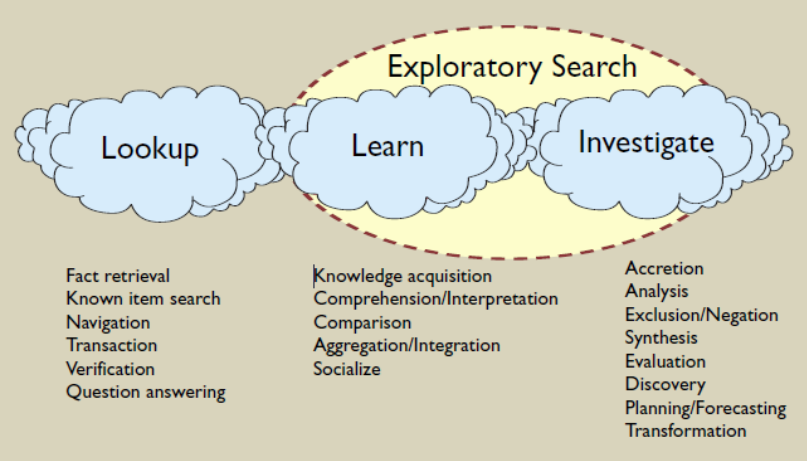
Informational: to acquire information assumed to be present on one or more web pages

Transactional: to perform some web-mediated activity

Resource: to get access to an online resource

## Search activities

[Marchionini 1995]




## Search modes

- **Lookup**
  - Locating
  - Verifying
  - Monitoring
- **Learn**
  - Comparing
  - Comprehending
  - Exploring
- **Investigate**
  - Analyzing
  - Evaluating
  - Synthesizing

## Search Modes: Lookup

**Locate** Find a specific (possibly known) item

**Verify** Confirm that an item meets some specific, objective criterion

**Monitor** Maintain awareness of the status of an item for purposes of management or control

## Search Modes: Learn

**Compare** To identify similarities & differences between a set of items

**Comprehend** To generate independent insight by understanding the patterns within a data set

**Explore** To investigate an item or data set for the purpose of knowledge discovery

## Search Modes: Investigate

**Analyze** To examine an item or data set to identify patterns & relationships

**Evaluate** To use judgement to determine the value of an item with respect to a specific goal

**Synthesize** To create a novel or composite artefact from diverse inputs

## Anomalous State of Knowledge (ASK)(1)

[Belkin 80]

### Classic IR systems: "best match" principle

- system returns those documents that fit best to the representation of the information need (e.g. query statement)
- only feasible, if user can give precise specification of her information need (like e.g. in DBMS)

## Anomalous State of Knowledge (ASK)(2)

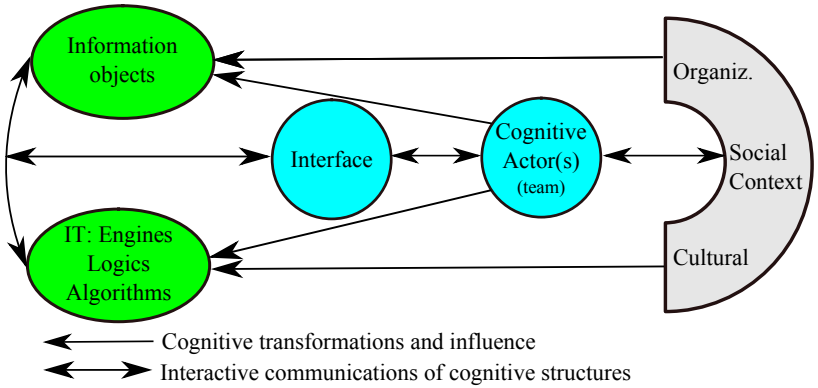
### ASK-Hypothesis

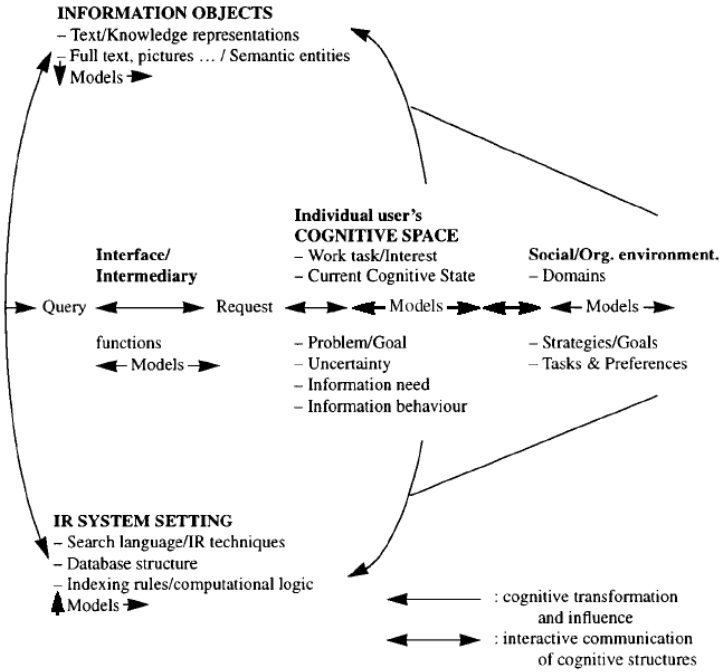
- information need results from user's *anomalous state of knowledge (ASK)*
- user is unable to precisely specify information need for removing the ASK
- instead: describe ASK
- requires capture of cognitive and situation-specific aspects for resolving this anomaly

## Ingwersen's Cognitive Model

- Global perspective
- comprises all factors influencing a search
  - social context
  - IR system
  - information objects
  - user interface
  - user
- focuses on *cognitive structures* – manifestations of human cognition, reflexions and ideas

## Ingwersen's Cognitive Model




## Classification of search activities

[Cool & Belkin 2002]

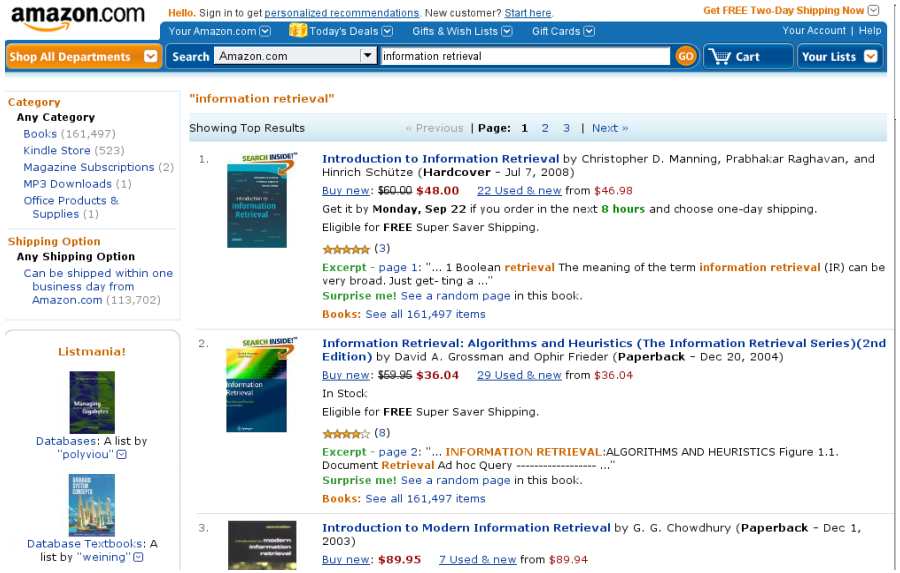
- Access:
  - method: scanning ... searching
  - mode: recognition ... specification
- object interacted with
  - level: information ... meta-information
  - media: text, images, speech, video, ...
  - quantity: 1 object, set of objects, database
- common dimensions of interaction
  - information objects: parts ... complete objects
  - systematics: random ... systematic
  - degree: selective ... exhaustive
- interaction criteria
  (e.g. precision, authority, date, person)

## Polyrepresentation

[Ingwersen 94]

- representation of information objects in different forms
- representations should correlate with cognitive structures
- example: document can be represented by
  - title (specified by the author)
  - keywords (by indexer)
  - other documents citing the current doc (extern)
  - annotations (extern)
- retrieval system should supports several representations (thus, also several cognitive structures)
  $\rightarrow$ *intentional redundancy*
- good search result, when several representations point to the same document (Overlap)

## Polyrepresentation: the Amazon case

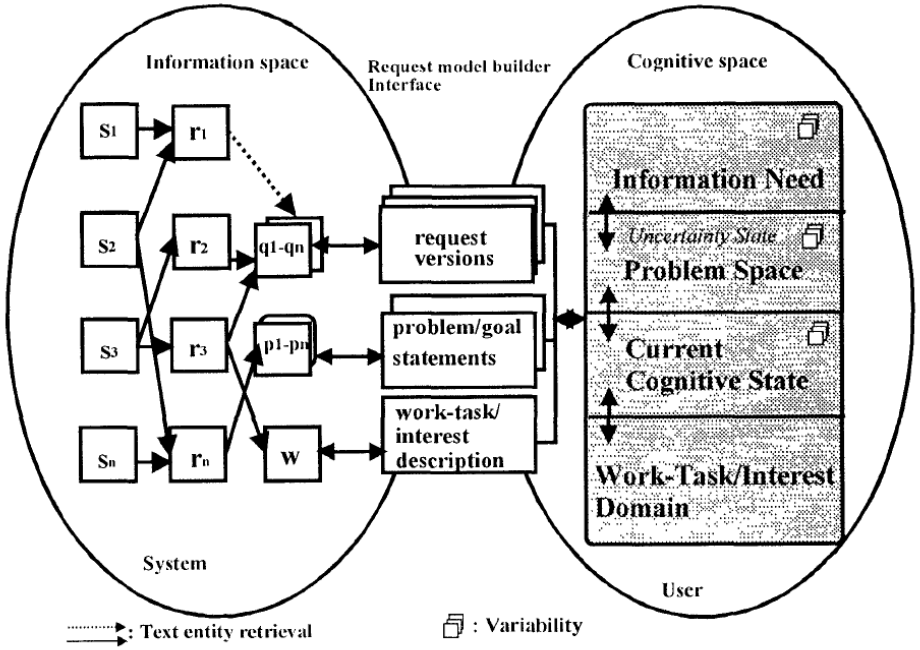



## Polyrepresentation of the cognitive user space

cognitive space can be represented via polyrepresentation

- requests
- problems/goals
- work task

## Global Polyrepresentation Model




## Strategic Support

Information Seeking Behavior & Information Searching
Levels of search activities
Degrees of system involvement
Proactivity in IR Systems

## Information Seeking Behavior & Information Searching

- searching consists of sequence of different phases
- experienced searchers employ a variety of actions in different phases
- these actions should be supported by the system as much as possible
- each phase should be supported appropriately by the system

### Bates' model for strategic system support

- Levels of search activities
- Degrees of system involvement

## Levels of search activity

1. Move: An identifiable thought or *action* that is a part of information searching.
2. Tactic: One or a handful of moves made to *further* a search.
3. Stratagem: A larger, more complex set of thoughts and/or actions than the tactic, all designed to exploit the file structure of a *particular search domain* thought to contain desired information.
4. Strategy: A *plan*, which may contain moves, tactics, and/or stratagems, for an entire information search.

## Monitoring Techniques

| | |
|---|---|
| CHECK | To review the original request and compare it to the current search topic to see that it is the same. |
| WEIGH | To make a cost-benefit assessment, at one or more points of the search, of current or anticipated actions. |
| PATTERN | To make oneself aware of a search pattern, examine it, and redesign it if not maximally efficient or if out of date |
| CORRECT | To watch for and correct spelling and factual errors in one's search topic. |
| RECORD | To keep track of followed and of desirable trails not followed or not completed. |

## File Structure Tactics

| | |
|---|---|
| BIBBLE | To look for a bibliography already prepared, before launching oneself into the effect of preparing one; more generally, to check to see if the search work one plans has already been done in a usable form by someone else. |
| SELECT | To break down complex search queries into subproblems and work on one problem at a time. |
| SURVEY | To review, at each decision point of the search, the available options before selection. |
| CUT | When selecting among several ways to search a given query, to choose the option that cuts out the largest part of the search domain at once . |
| STRETCH | To use a source for other than is intended purposes. |
| SCAFFOLD | To design an auxiliary, indirect route through the information files and resources to reach the desired information. |
| CLEAVE | To employ binary searching in locating an item in an ordered file. |

## Search Formulation Tactics

| | |
|---|---|
| SPECIFY | To search on terms that are as specific as the information desired . |
| EXHAUST | To include most or all elements of the query in the initial search formulation; to add one or more of the query elements to an already-prepared search formulation. |
| REDUCE | To minimize the number of the elements of the query in the initial search formulation; to subtract one or more of the query elements from an already-prepared search formulation. |
| PARALLEL | To make the search formulation broad (or broader) by including synonyms or otherwise conceptually parallel terms. |
| PINPOINT | To make the search formulation precise by minimizing (or reducing) the number of parallel terms, retaining the more perfectly descriptive terms. |
| BLOCK | To reject, in the search formulation, items containing or indexed by certain term(s), even if it means losing some document sections of relevance |

## Term Tactics

| | |
|---|---|
| SUPER | To move upward hierarchically to a broader (superordinate) term. |
| SUB | To move downward hierarchically to a more specific (subordinate) term. |
| RELATE | To move sideways hierarchically to a coordinate term. |
| TRACE | To examine information already found in the search in order to find additional terms to be used in furthering the search. |
| VARY | To alter or substitute one's search terms in any of several ways. See remaining term tactics for some specific variations. |
| REARRANGE | To reverse or rearrange the words in search terms in any or reasonable orders. |
| CONTRARY | To search for the term logically opposite that describing the desired information. |
| RESPELL | To search under a different spelling. |
| RESPACE | To try spacing variants |
| NEIGHBOR | To seek additional search terms by looking at neighboring terms, whether proximate alphabetically, by subject similarity, or otherwise. |
| FIX | To try alternative affixes, whether prefixes, suffixes, or infixes. |

## Idea Tactics

| | |
|---|---|
| RESCUE | In an otherwise unproductive approach, to check for possible productive paths still untried. |
| BREACH | To breach the boundaries of one's region of search, to revise one's concept of the limits of the intellectual or physical territory in which one searches to respond to a query. |
| FOCUS | To look at the query more narrowly, in one or both of two senses: (1) to move from the whole query to a part of it or (2) to move from a broader to a narrower conceptualization of the query. |

## Example stratagems

### Subject Search

| | |
|---|---|
| Journal Run | Having identified a journal central to one's topic of interest, one reads or browses through issues or volumes of the journal. |
| Citation Search | Using a citation index or database, one starts with a citation and determines what other works have cited it. |
| Area Scan | After locating a subject area of interest in a classification scheme, one browses materials in the same general area. |
| Footnote Chase | One follows up footnotes or references, moving backward in time to other related materials. |

## Tactics suggested in response to searcher request

| SEARCHER COMMAND | SYSTEM RESPONSE LIST |
|---|---|
| Too many hits | SPECIFY<br>EXHAUST<br>PINPOINT<br>BLOCK<br>SUB |
| Too few hits | NEIGHBOR<br>TRACE<br>PARALLEL<br>FIX<br>SUPER<br>RELATE<br>VARY |
| No hits | RESPACE<br>RESPELL<br>REARRANGE<br>CONTRARY<br>SUPER<br>RELATE<br>NEIGHBOR<br>TRACE |
| Need other terms<br>or<br>wrong terms | NEIGHBOR<br>TRACE<br>SUPER<br>SUB<br>RELATE |
| Revise terms | SPACE<br>RESPELL<br>FIX<br>REVERSE<br>CONTRARY<br>SUPER<br>SUB<br>RELATE |
| Revise search formulation | SPECIFY<br>EXHAUST<br>REDUCE<br>PARALLEL<br>PINPOINT<br>BLOCK |

## Degrees of system involvement

0 – No system involvement All search activities human generated and executed.

1 — Displays possible activities. System lists search activities when asked. Said activities may or may not also be executable by system (higher levels).

2 — Executes activities on command System executes specific actions at human command.

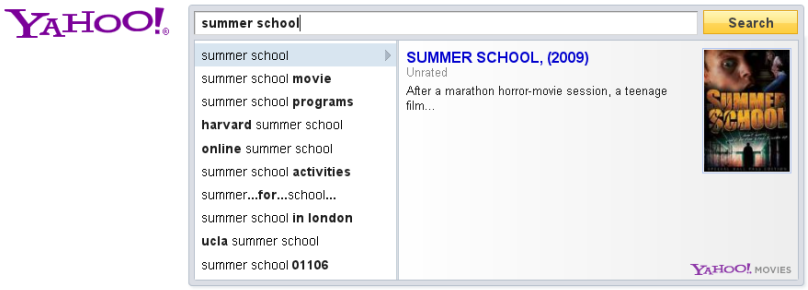
3 — Monitors search and recommends System monitors search process and recommends search activities:

a) Only when searcher asks for suggestions.
b) Always when it identifies a need.

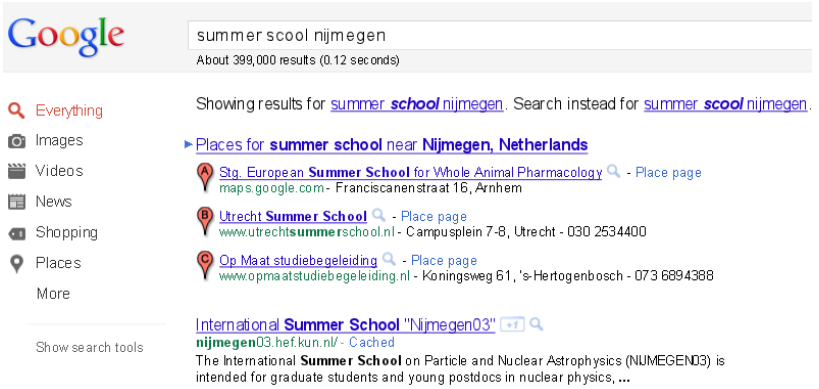
4 — Executes automatically. System executes actions automatically and then:

a) Informs the searcher.
b) Does not inform the searcher.

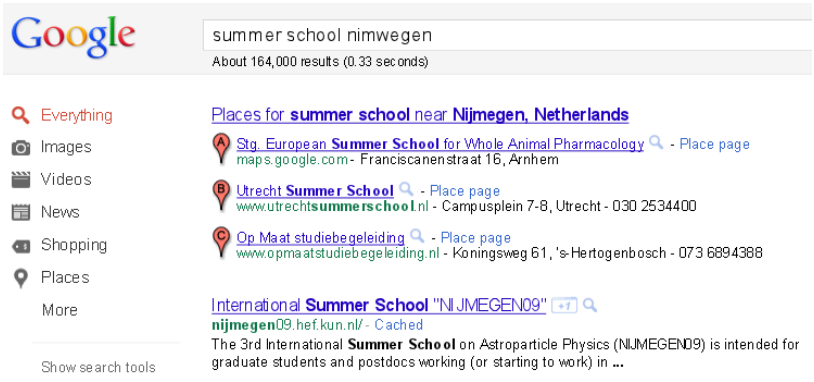
## Monitors search and recommends




## Executes automatically + informs



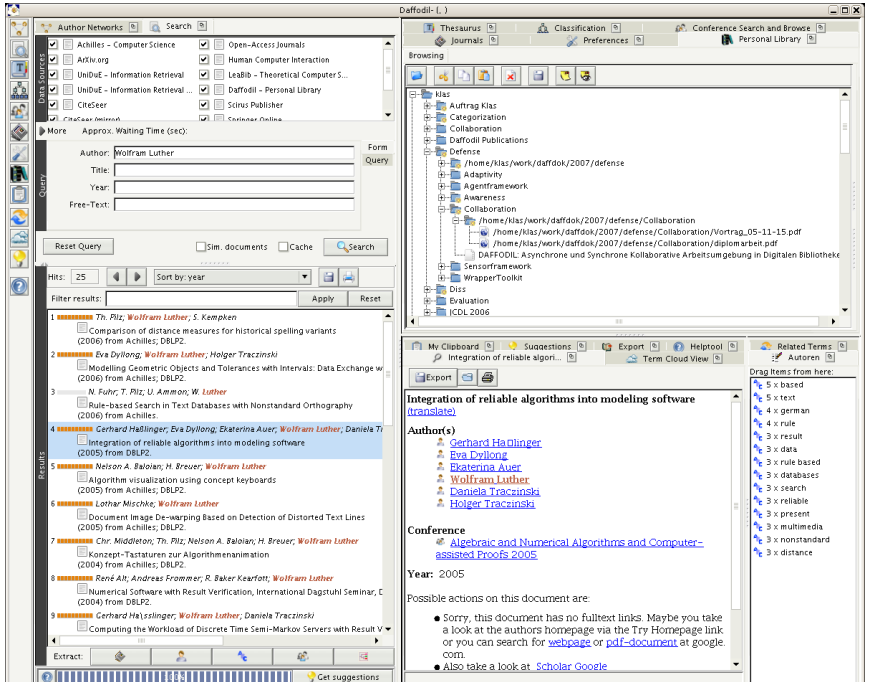

## Executes automatically + does not inform




## Combination of search activities and system support

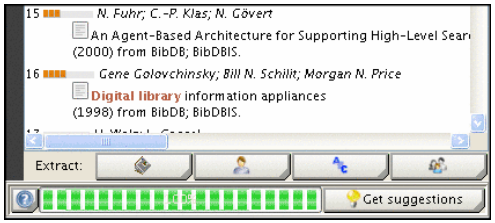
| Search Activity/ System involvement | Move | Tactics | Stratagem | Strategy |
|---|---|---|---|---|
| No system involvement | | | | |
| Displays possible activities | | | | |
| Executes activities on command | | | | |
| Monitors search and recommends | | | | |
| Executes automatically | | | | |

## Daffodil desktop

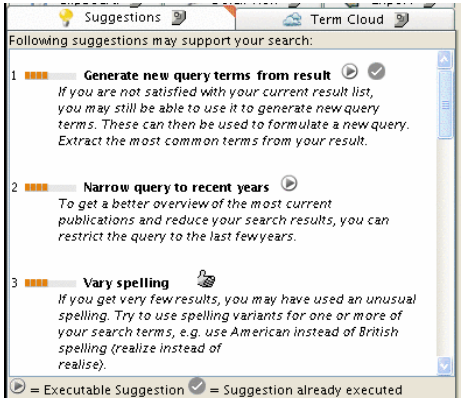


## Daffodil: Search Continuation

- proposal based on automatic analysis of the curent search result
- case-based reasoning
- availability of suggestions indicated as button at the bottom of result list window




## Daffodil: Search Continuation 2




- suggestions displayed as ranked list
- descriptive title, explanation, success rate
- execute on or more suggestions, with following feedback
- icons indicate the state of suggestions: (executable, used, useful)

## Evaluation of search suggestions

- 24 test subjects, half of them w/ suggestion component
- each subject worked on 3 tasks
- case base contained 30 different suggestions

**Results:**
supported users

- are more content with the search process ($p = 0,067$)
- are significantly more satisfied with the result
- find more relevant documents
- use significantly more often Daffodil's advanced seearch tools

(unsupported users mainly restrict on reformulating queries)

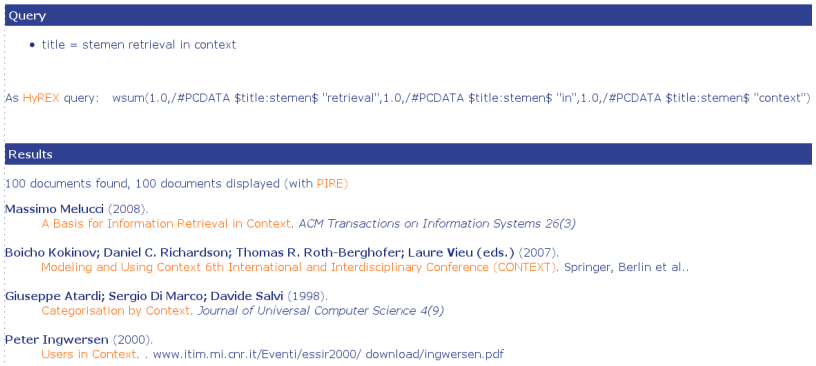
# From Cognitive Models to IR Interfaces

Session support in the user interface
Design Patterns for Search Modes
Support for seeking behavior according to Ellis/Meho/Tibo

## Session Support

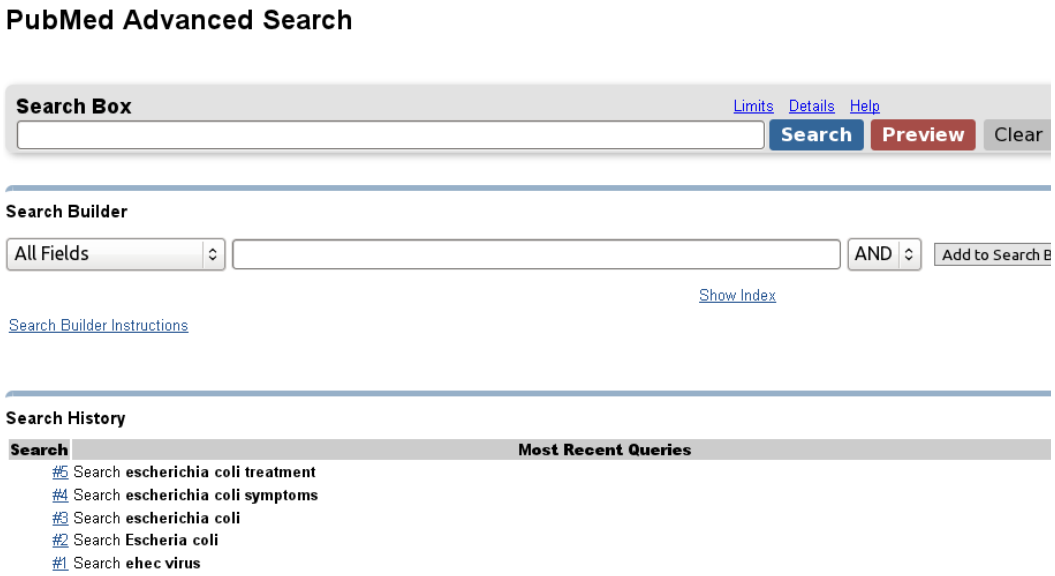
- Show results together with the query
- Allow editing of the previous query
- Show search history
- Allow for combination of queries
- Filing of single results
- Storing of sessions

## Show results together with the query

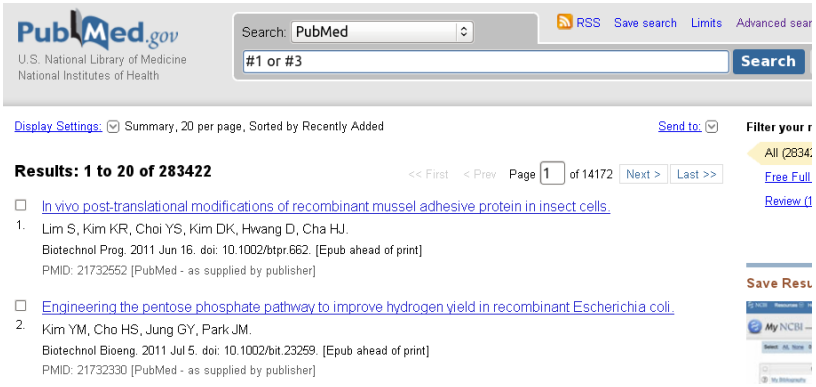



## Allow editing of the previous query

## Show search history

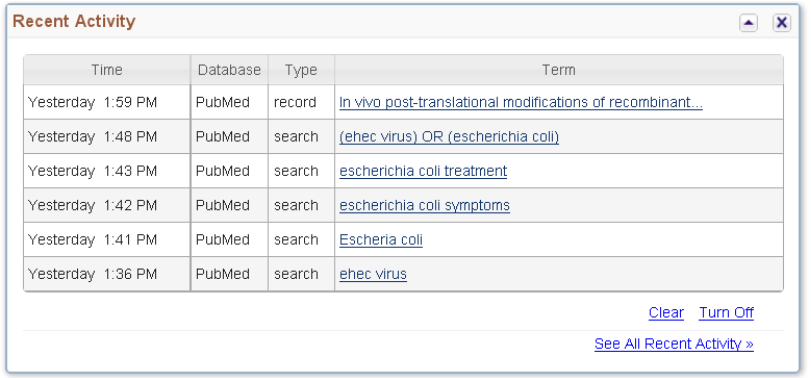



## Allow for combination of queries




## Filing of single results

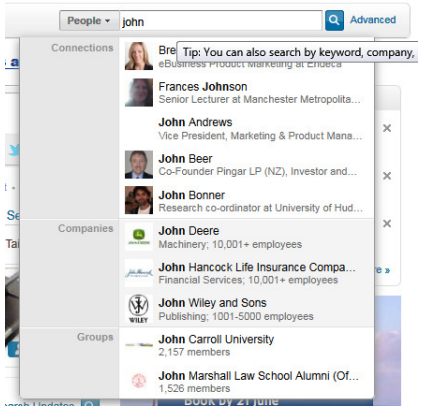

## Save sessions for continuing later



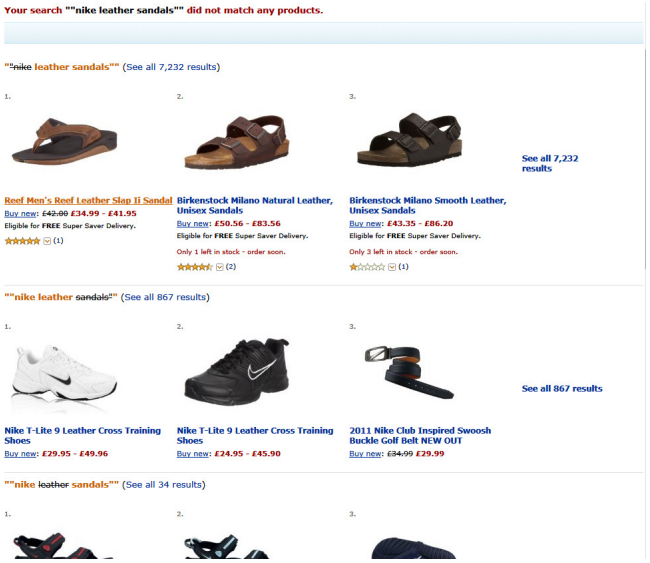
# Design Patterns for Search Modes

Support for Marchionini's search modes

# Locating: Autocomplete

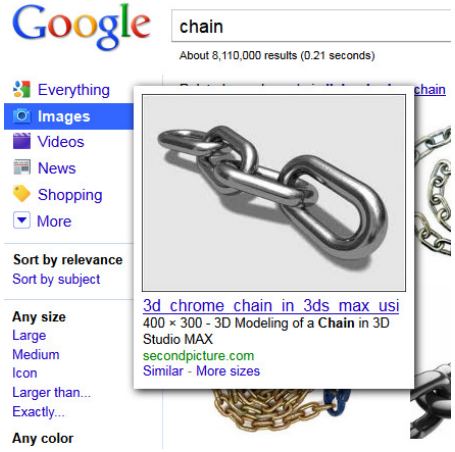



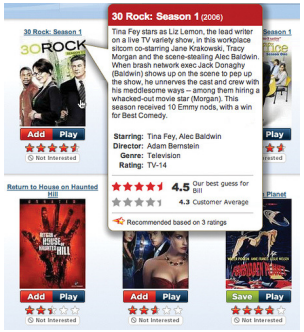

# Locating: Partial Matches



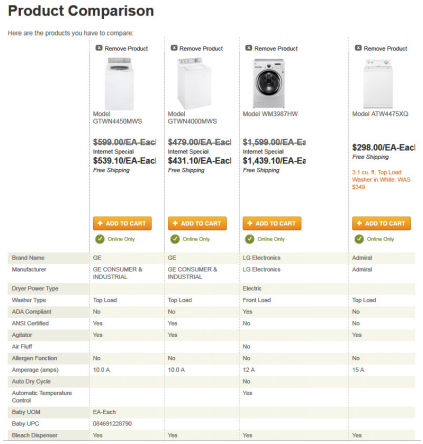

# Verifying: Instant Results

## Verifying: Detail Overlay



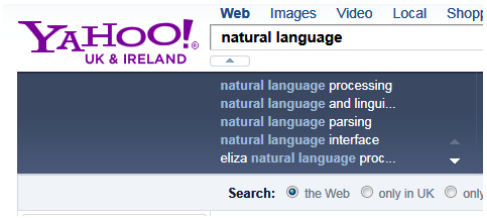




## Comparing: Parallel Views



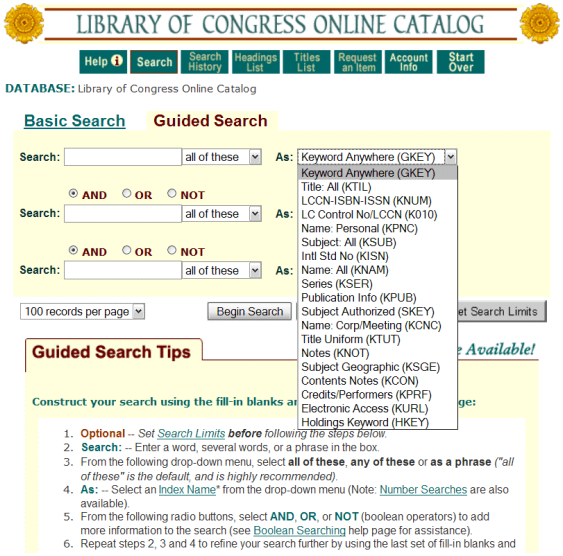


## Exploring: Autosuggest

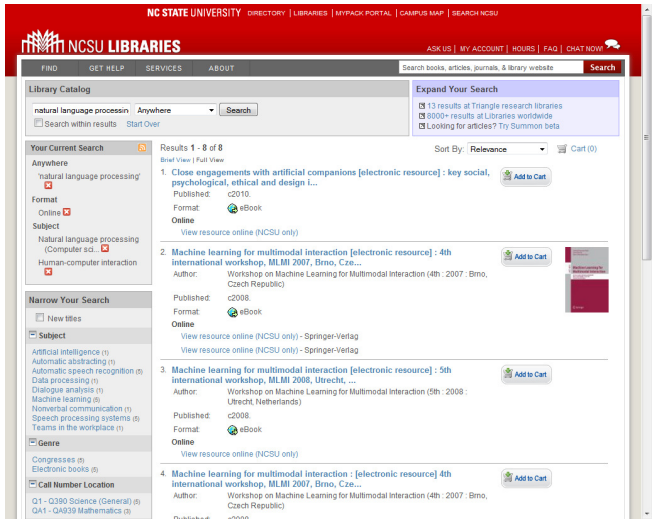




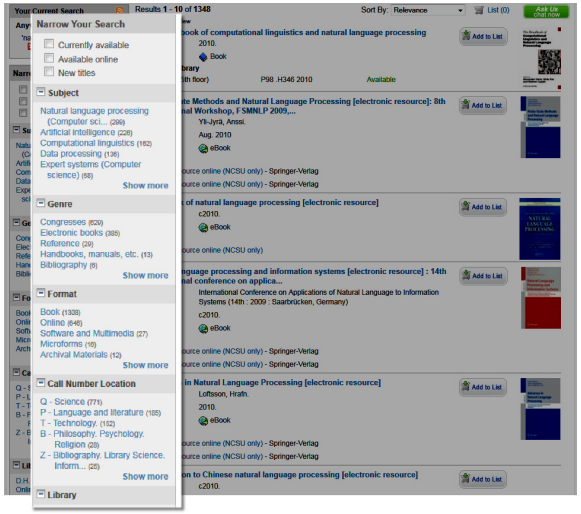
## Exploring: faceted search

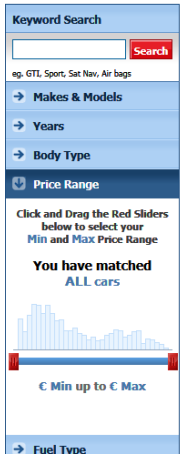


## Exploring: faceted search

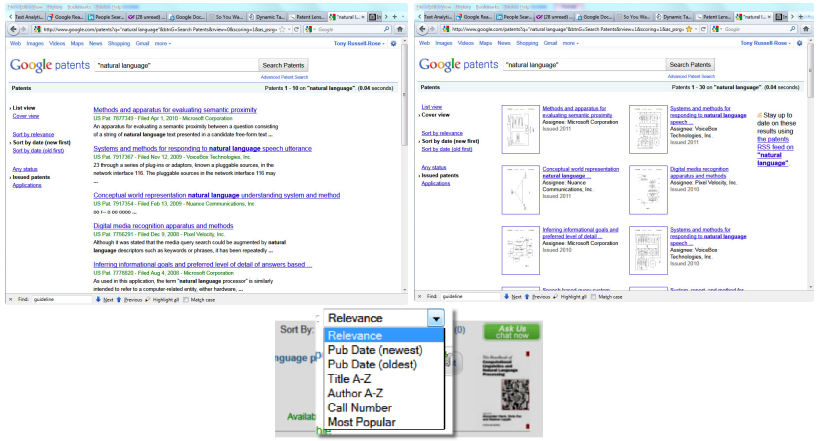



## Comprehending: facet menus

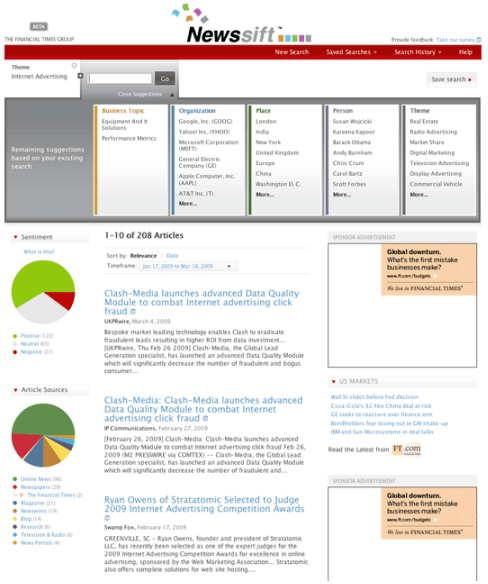



## Analyzing: alternate views



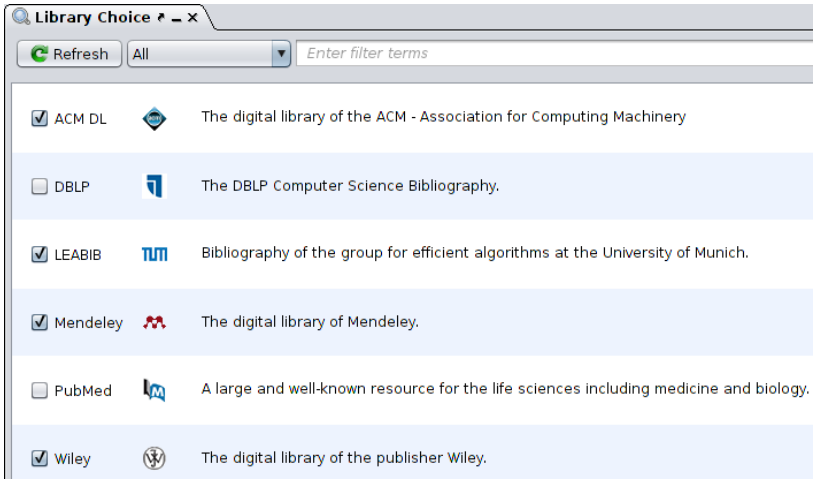

## Analyzing: data visualization



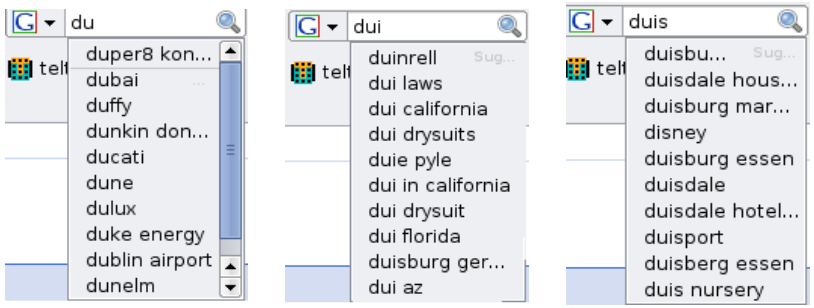
## Support for seeking behavior according to Ellis/Meho/Tibo

| | |
|---|---|
| Starting | Resource selection<br>identifying popular authors<br>entering search terms |
| Browsing | Sort result list by different criteria<br>highlighting, also user-defined |
| Chaining | links in results pages<br>comparison of result pages |
| Monitoring | storing and periodical execution of queries |
| Extracting | searching in the result page |
| Inform. Mgmt. | Collate/organize result items<br>Annotate items (Interpret) |

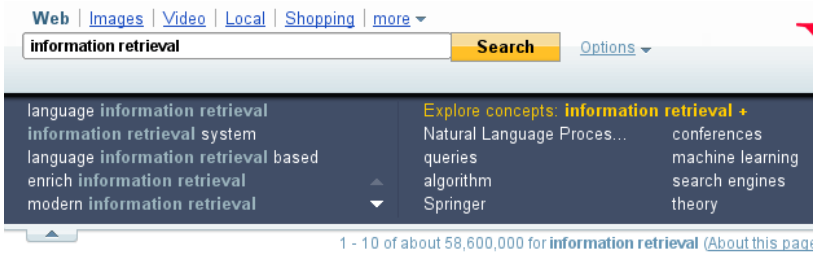
## Starting: Resource Selection



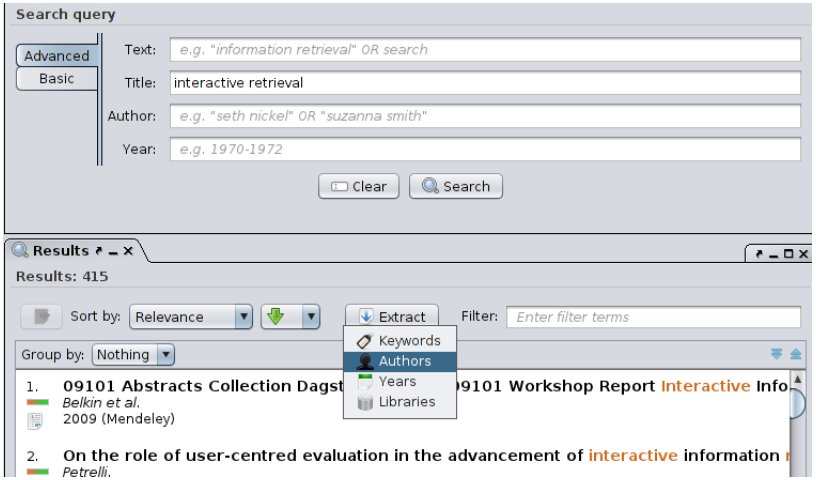

## Starting: Search term completion



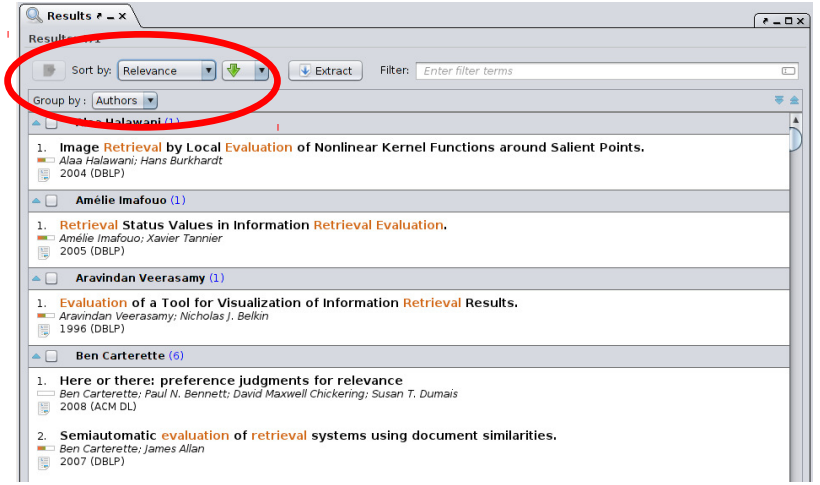

## Starting: Related Terms



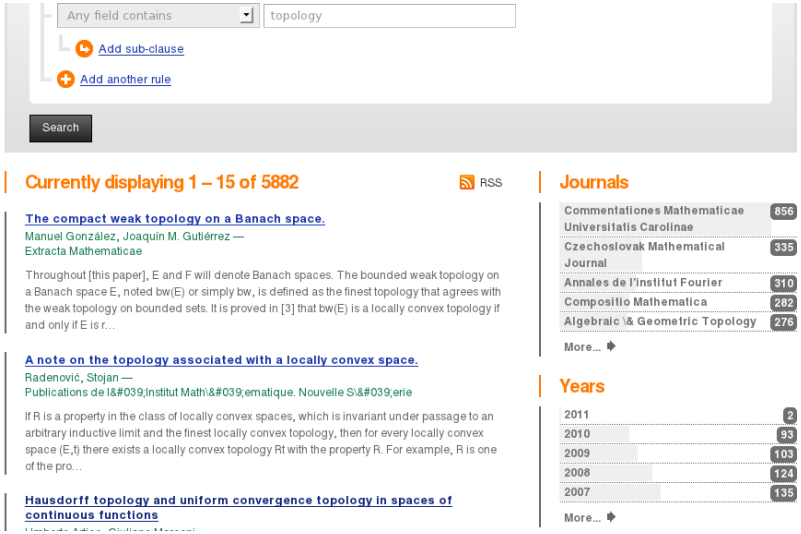
## Starting: Identify important authors
### 'extract authors' in ezdl



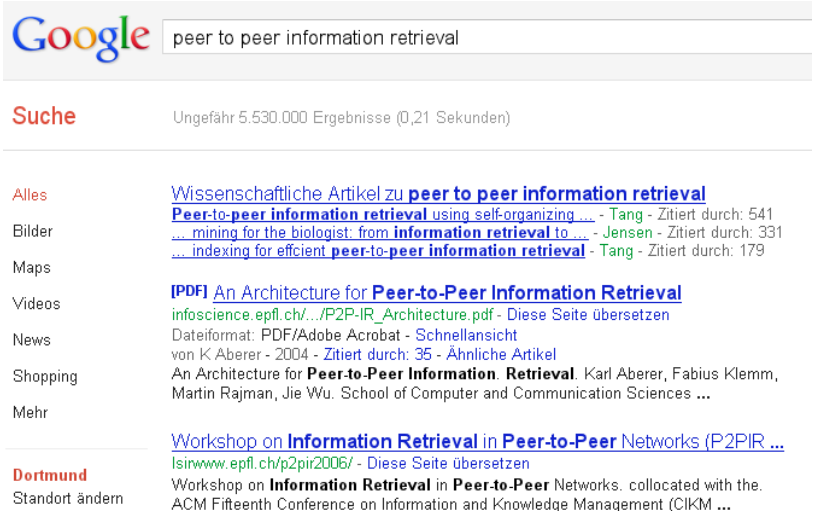

## Browsing: Sort/group results by different criteria




## Browsing: Meaningful Surrogates




## Browsing: Highlighting in the Result List



## Chaining: Clickable Entries in Result Pages

**User term feedback in interactive text-based image retrieval**

Full Text: Pdf Buy this Article

Authors: Chen Zhang Michigan State University, East Lansing, MI
Joyce Y. Chai Michigan State University, East Lansing, MI
Rong Jin Michigan State University, East Lansing, MI

2005 Article

Bibliometrics
· Downloads (6 Weeks): 6
· Downloads (12 Months): 48
· Citation Count: 7

Published in:

· Proceeding
SIGIR '05 Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval
©2005 table of contents ISBN:1-59593-034-5
doi>10.1145/1076034.1076046

## Chaining: backward/forward chaining of references

Abstract | Authors | References | Cited By | Index Terms | Publication | Reviews | Comments | Table of Contents

7 Citations

Zhenyu Zhang , Tong Lu , Feng Su , Ruoyu Yang, A new text detection algorithm for content-oriented line drawing image retrieval, Proceedings of the 11th Pacific Rim conference on Advances in multimedia information processing: Part I, September 21-24, 2010, Shanghai, China

Mouna Torjmen , Karen Pinel-Sauvagnat , Mohand Boughanem, Using textual and structural context for searching Multimedia Elements, International Journal of Business Intelligence and Data Mining, v.5 n.4, p.323-352, October 2010

Woo-Cheol Kim , Ji-Young Song , Seung-Woo Kim , Sanghyun Park, Image retrieval model based on weighted visual features determined by relevance feedback, Information Sciences: an International Journal, v.178 n.22, p.4301-4313, November, 2008

Dilip Kumar Limbu , Andy Connor , Russel Pears , Stephen MacDonell, Contextual relevance feedback in web information retrieval, Proceedings of the 1st international conference on Information interaction in context, October 18-20, 2006, Copenhagen, Denmark

Mouna Torjmen , Karen Pinel-Sauvagnat , Mohand Boughanem, Towards a structure-based multimedia retrieval model, Proceeding of the 1st ACM international conference on Multimedia information retrieval, October 30-31, 2008, Vancouver, British Columbia, Canada

Joyce Y. Chai , Chen Zhang , Rong Jin, An empirical investigation of user term feedback in text-based targeted image search, ACM Transactions on Information Systems (TOIS), v.25 n.1, p.3-es, February 2007

Daniela Petrelli, On the role of user-centred evaluation in the advancement of interactive information retrieval, Information Processing and Management: an International Journal, v.44 n.1, p.22-38, January, 2008

## Forward Chaining in Web Searches

Google link:www.is.inf.uni-due.de/courses/index.html

Suche 7 Ergebnisse (0,23 Sekunden)

Alles
Bilder
Maps
Videos
News
Shopping
Mehr

Das Web
Seiten auf Deutsch
Übersetzte Seiten

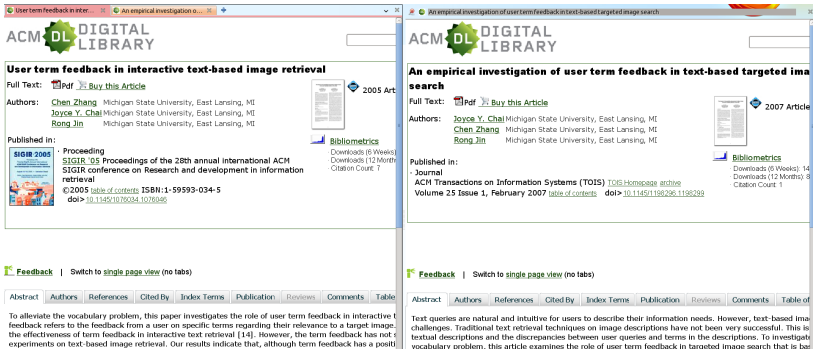
DAFFODIL: Strategic Support Evaluated
www.is.informatik.uni-duisburg.de/bib/docs/Klas_04ta.html
DAFFODIL: Strategic Support Evaluated. UDEFakultäten IngenieurwissenschaftenAbteilung InKoInformationssystemeDAFFODIL: Strategic Support Evaluated ...

DiLAS: a Digital Library Annotation Service
www.is.informatik.uni-duisburg.de/bib/docs/Agosti_etal_05.html
UDEFakultätenIngenieurwissenschaftenAbteilung InKoInformationssystemeDiLAS ...

Teaching
www.is.informatik.uni-duisburg.de/.../index.htm... - Diese Seite übersetzen
UDEFakultätenIngenieurwissenschaftenAbteilung InKoInformationssysteme ...

Comparing different architectures for query routing in peer-to ...
www.is.informatik.uni-duisburg.de/bib/.../Nottelmann_Fuhr_06.html
Zitationsschlüssel: Nottelmann/Fuhr:06; Titel: Comparing different ...

## Differentiating: comparison of result pages



## Extracting: Highlighting in the result page

# Sheffield July 29

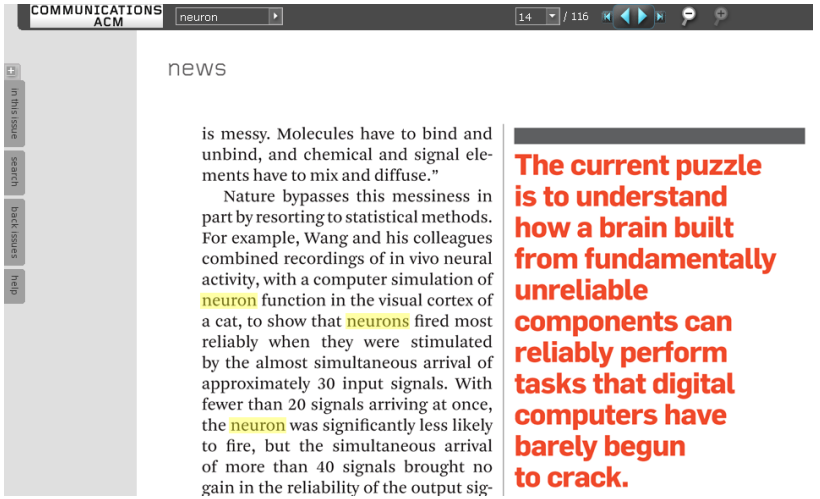
## 27th Annual International ACM SIGIR Conference

### Workshop on Peer-to-Peer Information Retrieval

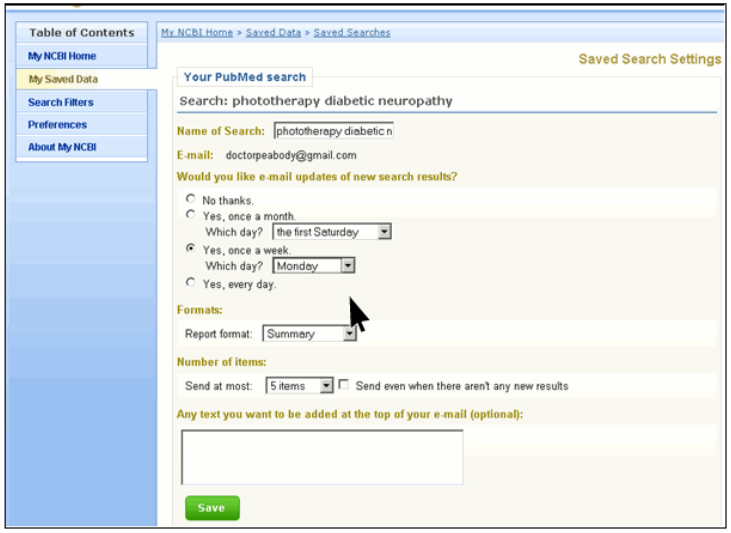
SIGIR is the major international forum for the presentation of new research results and the demonstration of new systems and techniques in the broad field of information retrieval.

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.

## Extracting: Searching in the result page



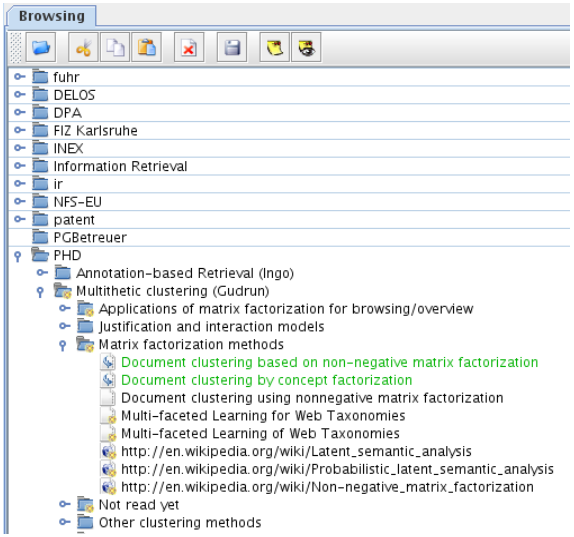

## Monitoring
### 'Saved Searches' at PubMed




## Monitoring
### Watchthatpage.com

## Information management: Organizing and Annotating results


## Information management: personal library in Daffodil




## Summary

## Summary

- Interactive PRP for analysis and design of IIRS
- Information seeking behavior vs. searching
- cognitive models:
  - search as iterative process
  - large variation in search tasks
  - search influenced by many factors
- systems:
  - strategic support through high-level search functions (especially for typical cognitive actions)
  - proactive support
- user interface design on cognitive models

# References I


**Fuhr, N.**
(2010).
Skriptum Information Retrieval, Kapitel 6 'Interaktives Retrieval'.
*http://www.is.inf.uni-due.de/courses/ir_ss10/folien/kapiir.pdf*

**Bates, M. J.**
(1989).
The design of browsing and berrypicking techniques for the online search interface.
*Online Review 13(5)*, pp. 407–424.

**Bates, M. J.**
(1990).
Where Should the Person Stop and the Information Search Interface Start?
*Information Processing and Management 26(5)*, pp. 575–591.

**Belkin, N. J.**
(1980).
Anomalous states of knowledge as a basis for information retrieval.
*Canadian Journal of Information Science 5*, pp. 133–143.


# References II


**Belkin, N.**
(1996).
Intelligent information retrieval: Whose intelligence?
In: *ISI '96: Proceedings of the Fifth International Symposium for Information Science*, pp. 25–31. Universitätsverlag Konstanz.

**Cool, C.; Belkin, N. J.**
(2002).
A Classification of Interactions with Information.
In: Bruce, H.; Fidel, R.; Ingwersen, P.; Vakkari, P. (Hrsg.): *Emerging frameworks and methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (COLIS4)*, pp. 1–15. Libraries Unlimited, Greenwood Village.

**Ellis, D.**
(1989).
A behavioural approach to information retrieval system design.
*Journal of Documentation 45(3)*, pp. 171–212.

**Fuhr, N.**
(2008).
A Probability Ranking Principle for Interactive Information Retrieval.
*Information Retrieval 11(3)*, pp. 251–265.
http://dx.doi.org/10.1007/s10791-008-9045-0.


# References III


**Hearst, M. A.**
(2009).
*Search User Interfaces*.
Cambridge University Press.

**Ingwersen, P.**
(1992).
*Information Retrieval Interaction*.
Taylor Graham, London.

**Ingwersen, P.**
(1994).
Polyrepresentation of Information Needs and Semantic Entities, Elements of a Cognitive Theory for Information Retrieval Interaction.
In: Croft, B. W.; van Rijsbergen, C. J. (Hrsg.): *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 101–111. Springer-Verlag, London, et al.

**Kriewel, S.; Fuhr, N.**
(2007).
Adaptive Search Suggestions for Digital Libraries.
In: *Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers (ICADL 2007)*, pp. 220–229.


# References IV


**Kriewel, S.; Fuhr, N.**
(2010).
An evaluation of an adaptive search suggestion system.
In: *32nd European Conference on Information Retrieval Research (ECIR 2010)*.

**Meho, L. I.; Tibbo, H. R.**
(2003).
Modeling the information-seeking behavior of social scientists: Ellis's study revisited.
*Journal of the American Society for Information Science and Technology 54(6)*, pp. 570–587.

**Robertson, S. E.**
(1977).
The Probability Ranking Principle in IR.
*Journal of Documentation 33*, pp. 294–304.

**Schaefer, A.; Jordan, M.; Klas, C.-P.; Fuhr, N.**
(2005).
Active Support For Query Formulation in Virtual Digital Libraries: A case study with DAFFODIL.
In: Rauber, A.; Christodoulakis, C.; Tjoa, A. M. (Hrsg.): *Research and Advanced Technology for Digital Libraries. Proc. European Conference on Digital Libraries (ECDL 2005)*, Lecture Notes in Computer Science. Springer, Heidelberg et al.

# References V


**Tran, V. T.; Fuhr, N.** (2012).
Using Eye-Tracking with Dynamic Areas of Interest for Analyzing Interactive Information Retrieval.
In: *Proceedings of the 35th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 1165-1166.

**Turpin, A. H.; Hersh, W.**
(2001).
Why batch and user evaluations do not give the same results.
In: Croft, W. B.; Harper, D.; Kraft, D. H.; Zobel, J. (Hrsg.): *Proceedings of the 24th Annual International Conference on Research and development in Information Retrieval*, pp. 225–231. ACM Press, New York.

**Zuccon, G.; Azzopardi, L.; van Rijsbergen, C. J.** (2011).
The interactive PRP for diversifying document rankings.
In: Ma, W.-Y.; Nie, J.-Y.; Baeza-Yates, R. A.; Chua, T.-S.; Croft, W. B. (eds.): *SIGIR*, pages 1227–1228. ACM.