CS 572: Information Retrieval

Challenges and Open Research Questions

Acknowledgements
Some slides in this lecture adapted from various IR researchers, credits noted on each slide.
Search Challenges (2002)

UMass CIIR report, 2002

- **Global information access:** Satisfy human information needs through natural, efficient interaction with a system ... [over the world’s data]... in any language.

- **Contextual retrieval:** Combine search technologies and knowledge about query and user context ... to provide the most “appropriate” answer for a user’s information needs.
Search Challenges (2012)

- **Beyond ranked list**: enrich querying & results
- **IR for all**: empower the user to search & learn
- **Capture context**: current task, time, etc.
- **Beyond document retrieval**: complex data & result integration
- **Domain search**: Verticals, apps, restricted data
- **Evaluation**: for new search types, tasks
Search Challenges (2016)

• **Semantics**: query vs. document semantic mismatch
• **Domains/real-time content**: ranking, relevance?
• **Monetization**: spam vs. “promoted” content
• **Usefulness/utility**: indirectly through behavior
• **Result presentation**: 10 blues links → “universal search”
• **Ethics**: influence on health, politics, society
SEMANTICS
Where we want to be

Thomas Cruise Mapother IV (born July 3, 1962), widely known as Tom Cruise, is an American film actor and producer.
Where we are today

- There are many forms of annotation
  - Microdata (Schema.org, Good Relations), microformats, RDFa, folksonomy tags ...
  - Mostly manual, with some automation when generating templated pages
    - Automation saves typing, but rarely performs disambiguation

- Annotation schemes identify entity boundaries just fine, and offer *some* semantics
  - But no consistent vocabulary, and reconciliation is a challenge
  - How do we know two web pages refer to the *same* entity?
Existing Knowledge Graphs

- Freebase
- Wikidata
- Yago
- DBpedia
- Facebook’s Entity Graph
- OpenIE (Reverb, OLLIE)
- Microsoft’s Satori
- Google’s Knowledge Graph
• **Freebase** is an open, [Creative Commons](https://creativecommons.org) licensed repository of structured data.

![Freebase Person Type](image)

Relations are typed too!
The never-ending quest for knowledge

- **Freebase**
  - (as of Oct’12)
  - 25M objects
  - 437M facts

- **Knowledge Graph**
  - (as of May’12)
  - 500M objects
  - 3,500M facts

---

http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html
Challenge: Text annotation / disambiguation

• Task: Identify entities on a page and map them to corresponding entities

  — “Albert Einstein was a German-born theoretical physicist ...”
  — “Clinton briefly considered dedicating his life to music”
Challenge: Surface Data in web search

Augmenting the presentation with relevant facts
Surfacing facts proactively

1. Search for "Maui population" on Google.
2. Click on the first result: "144,444 (2010)".
3. Click on the "Details" link for more information.
4. Verify the population data on the official Maui County website.

Additional search terms:
- "Maui County QuickFacts" for quick facts and demographic data.
- "Maui County, HI - Official Website" for the official county website.

Other relevant searches:
- "Maui vacation" for travel information.
- "Maui restaurants" for restaurant recommendations.
- "Maui attractions" for popular tourist spots.

Maps and images:
- "Maui Map" and "Maui Islands" for geographical context.

Wikipedia links:
- "Maui - Wikipedia" for comprehensive information.
- "Hawaiian Islands - Wikipedia" for information on the islands of Hawaii.

Other resources:
- "Maui Travel Guide" for travel tips and guidance.
- "Fodor's Travel Guides" for detailed travel advice.

Feedback:
- "See results about Maui County, Hawaii" for more information about the county and its location.

Additional notes:
- "144,444 (2010)" indicates the population as of 2010.
- "Maui has been voted "Best Island" by readers of Condé Nast Traveler for more than a dozen years." for accolades on the island.
Exploratory search

**Kilauea Lighthouse**

Built in 1913 as a navigational aid for commercial shipping between Hawai`i and the Orient, Kilauea Point Lighthouse stands as a monument to Hawai`i's ... 6 Google reviews - Write a review

3500 Kilauea Road, Kilauea, HI 96754
(808) 828-0168

**Lighthouses of the U.S. - Hawai`i**
[www.unc.edu/~rowlett/lighthouse/hi.html](http://www.unc.edu/~rowlett/lighthouse/hi.html)

There is no state lighthouse preservation society in Hawai`i, and local preservation efforts have just begun to appear in the last decade. We follow Hawaiian ... 6 Google reviews

**Hawaii Lighthouses**

Kilauea Lighthouse, Barbers Point, Cape Kumukahi, Diamond Head, Kauhola Point, Kilauea Point, Lahaina, Makapuu, Minor light of Hawai`i, Coconut ...

6 Google reviews

**Hawaii Map - Lighthouse Friends**
[www.lighthousefriends.com/hi.html](http://www.lighthousefriends.com/hi.html)

Hawaii Lighthouses. Click on a lighthouse name or icon for more information on that lighthouse. The lighthouses, About Us, Our friends, The Maps, Links to other ...
Triggering actions in Web search

Yahoo! search results for "pulp fiction quotes, pulp fiction soundtrack, more..."

Pulp Fiction (1994) movies.yahoo.com
Yahoo B+, Critics A An inside look at a memorable community of criminals. Prizefighter Butch Coolidge has decided to stop payment on a... more
Add to Netflix Queue
Running Time: 2 hr 29 min
Directed by: Quentin Tarantino
Starring: John Travolta, Samuel L. Jackson, Uma Thurman, Harvey Keitel, ... more

Add to Netflix Queue
Try this search on my phone

Wikipedia - Pulp Fiction
en.wikipedia.org/wiki/Pulp_Fiction_(film) - Cached
Entity-based actions

[Image of a search result for Anna Karenina]

10,800,000 RESULTS

Anna Karenina - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Anna_Karenina
Main characters - Plot introduction - Plot summary - Style - Major themes
Anna Karenina is a novel by the Russian writer Leo Tolstoy, published in serial installments from 1873 to 1877 in the periodical The Russian ...

Anna Karenina (2012) - IMDb
www.imdb.com/title/tt1781769
Drama · R · 130 min
Director: Joe Wright. Actors: Keira Knightley; Anna Karenina · Aaron Taylor-Johnson: Count Vronsky · Kelly Macdonald: Dolly · Jude Law: Alexei Karenin · Michelle ...

Anna Karenina: Leo Tolstoy, Constance Garnett; Amazon.com: ...
www.amazon.com/Anna-Karenina-ebook/dp/B008476UXW
User rating: 4.1/5 · By Leo Tolstoy, Constance Garnett · Kindle Edition
Some people say Anna Karenina is the single greatest novel ever written, which makes about as much sense to me as trying to determine the world’s ...

Watch trailer

2hr 10min · R · Drama
Synopsis: A bold, theatrical new vision of the epic story of love, adapted from
Summary

• Annotation is just an interim step towards **text understanding**
  • And most annotation schemes offer limited semantics

• **Large-scale knowledge repositories** provide a lexicon of entities and relations, which are critical for text understanding

• Research challenges
  • Information extraction at Web scale
  • Information fusion from imperfect extractors
  • Large-scale inference

• Deeper understanding of text will allow
  • Proactively fetching relevant information
  • Entity-based services
SEARCH INTERFACES
Web Search at 15

What’s available

- Number of pages indexed
  - 7/94 Lycos –
  - 95 – 10^6 millions
  - 97 – 10^7
  - 98 – 10^8
  - 01 – 10^9 billions
  - 05 – 10^10 …

- Types of content
  - Web pages, newsgroups
  - Images, videos, maps
  - News, blogs, spaces
  - Shopping, local, desktop
  - Books, papers, many formats
  - Health, finance, travel …

How it’s accessed
Support for Searchers

- The search box
- Spelling suggestions
- Query suggestions
- Auto complete
- Inline answers
- Richer snippets
- But, we can do better

Search in the future will look nothing like today's simple search engine interfaces, [Susan Dumais] said, adding, "If in 10 years we are still using a rectangular box and a list of results, I should be fired." [Mar. 7, 2007, NYTimes, John Markoff]
Future Challenges

• Dynamic information environments [Adar et al., Elsas et al.]
  – Content changes (e.g., news, blogs, lifelogs ... much more general)
  – People re-visit, re-query, re-find
  – IR opportunities ... crawling, doc and user representation, ranking, etc.
  – Interesting historically and socially

• Data/Evaluation
  – Data as valuable resource
  – Large-scale log data
  – Operational systems and a “Living Laboratory”
  – IR opportunities ... representations, ranking, etc.

• Thinking outside the traditional IR boxes
  – Better understanding of users and application domains
  – Collaborations across disciplinary boundaries
Information Dynamics

My Homepage

Susan Dumais

Research Activities:
I am interested in algorithms and interfaces for improved information retrieval, as well as general issues in and human-computer interaction. I led Microsoft Research in the area of web search, and a variety of information access and management issues, including personal information management and web search.

What's New:

Workshops, Collaborations and Papers:

We're Hiring at MSR and LiveLabs...
We're looking for great folks to advance the state of the art and influence new products in the search area. We have internships and permanent positions in several areas including: internet search, desktop search, personalization, and novel interfaces for search.

Research Activities:
- I am interested in algorithms and interfaces for improved information retrieval, as well as general issues in and human-computer interaction. I led Microsoft Research in the area of web search, and a variety of information access and management issues, including personal information management and web search.

Workshops, Collaborations and Papers:
Information Dynamics

Content Changes

User Visitation/ReVisitation

Today’s Browse and Search Experiences

But, ignores ...
Not All Searches Need to Be Fast

• Long-term tasks
  – Long search sessions
  – Multi-session searches

• Social search
  – Question asking

• Technologically limited
  – Mobile devices
  – Limited connectivity
  – Search from space
Using human computation to improve search

CROWDSOURCING
Replace Components with People

• Search process
  – Understand query
  – Retrieve
  – Understand results

• Machines are good at operating at scale

• People are good at understanding

with Kim, Collins-Thompson
Understand Query: Query Expansion

• Original query: *hubble telescope achievements*

• Automatically identify expansion terms:
  – *space, star, astronomy, galaxy, solar, astro, earth, astronomer*

<table>
<thead>
<tr>
<th></th>
<th>space</th>
<th>star</th>
<th>astronomy</th>
<th>galaxy</th>
<th>solar</th>
<th>astro</th>
<th>earth</th>
<th>astronomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>hubble</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>telescope</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>achievements</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
p(term_j | \text{query}) = \prod_{i \in \text{query}} \frac{\text{vote}_j,i}{\sum_j \text{vote}_j,i}
\]
Understand Results: Filtering

• Remove irrelevant results from list
• Ask crowd workers to vote on relevance
• Example:
  – hubble telescope achievements
People Are Not Good Components

• Test corpora
  – Difficult Web queries
  – TREC Web Track queries

• Query expansion generally ineffective

• Query filtering
  – Improves quality slightly
  – Improves robustness

• Not worth the time and cost

• Need to use people in new ways
People Can Provide Rich Input

- Test corpus: Complex restaurant queries to Yelp
- Query understanding improves results
  - Particularly for ambiguous or unconventional attributes
- Strong preference for the tabulated results
  - People asked for additional columns (e.g., star rating)
  - Those who liked the traditional results valued familiarity
Create Answers from Search Results

- Understand query
  - Use log analysis to expand query to related queries
  - Ask crowd if the query has an answer
- Retrieve: Identify a page with the answer via log analysis
- Understand results: Extract, format, and edit an answer

with Bernstein, Dumais, Liebling, Horvitz
Why is the moon sometimes orange?

The moon appears orange when it is low in the sky due to the scattering of light by the atmosphere. Moonlight must pass through much more atmosphere when it is near the horizon. The air molecules scatter away the blue, green and purple light, leaving only yellow, orange, or red. Dust, smoke, or pollen can also cause light to scatter and change the color of the moon when directly overhead.

Reference: home.hiwaay.net/.../moonorange
Create Answers to Social Queries

- Understand query: Use crowd to identify questions
- Retrieve: Crowd generates a response
- Understand results: Vote on answers from crowd, friends

with Jeong, Morris, Liebling
Searching versus Asking

- Friends respond quickly
  - 58% of questions answered by the end of search
  - Almost all answered by the end of the day
- Some answers confirmed search findings
- But many provided new information
  - Information not available online
  - Information not actively sought
  - Social content

with Morris, Panovich
Shaping the Replies from Friends

- Larger networks provide better replies
- Faster replies in the morning, more in the evening
- Question phrasing important
  - Include question mark
  - Target the question at a group (even at anyone)
  - Be brief (although context changes nature of replies)
- Early replies shape future replies
- Opportunity for friends and algorithms to collaborate to find the best content

with Morris, Panovich
Supporting Search through Structure

- Provide search recipes
  - Understand query
  - Retrieve
  - Process results
- For specific task types
- For general search tasks
- Structure enables people to
  - Complete harder tasks
  - Search for complex things from their mobile devices
  - Delegate parts of the task

with Liebling, Lasecki
Algorithms + Experience

Taking Our Sweet Time to Search - Marian Dörk
marian.doerk.de/slowsrch/chi2013.pdf · PDF file
Taking Our Sweet Time to Search Marian Dörk CultureLab NewcastleUniversity
marian.doerk@ncl.ac.uk Peter Bennett BristolInteraction&Graphics UniversityofBristol

Slow Search | August 2014 | Communications of the ACM
cacm.acm.org/magazines/2014/8/177023-slow-search
We live in a world where the pace of everything from communication to transportation is getting faster. In recent years a number of “slow movements” have emerged that ...

Evaluation Campaigns and TRECvid - ResearchGate - Share ... www.researchgate.net/profile/Wessel_Kraaij/publication/29661745... · PDF file
Evaluation Campaigns and TRECvid ... for work-related use have exploded in size, IR ... paigins and that they can overly-influence the research agenda. This is no ...

Nitrogen and hydrogen related infrared absorption in CVD ...
www.researchgate.net/...hydrogen_related_infrared_absorption_in_CVD... · PDF file
Using these measurements we point out that the oscillator strength of the different IR modes ... additional peaks which were confirmed as nitrogen related using ab ...

DEVELOPING HYPOTHESIS AND RESEARCH QUESTIONS
www.public.asu.edu/~kroz/www500/hypothesis.pdf · PDF file
DEVELOPING HYPOTHESES & RESEARCH QUESTIONS Definitions of hypothesis "it is a tentative prediction about the nature of the relationship between two or
"Atari Jaguar FAQ"
Atari Jaguar FAQ, Atari Archives, Atari Jaguar, Atari Jaguar64? ... (6)
http://www.digiserve.com/eescape/showpage.html?page=atari64

"Jaguar Interactive II -- The Premier 24-Hour Atari Jaguar"
... 09:45 26/Jul/04, Jaguar Collector, ... 18:07 26/Jul/04, Jaguar Collector, ... (6)
http://www.atarihq.com/interactive/

"Atari Jaguar VLM"
Atari Jaguar VLM. Much thanks to Joe Britt for the pix and modification details. Atari's Virtual Light Machine (VLM), was developed ... (8)
http://www.audiosubstrates.com/technosynth/vidsynth/vag_vlm/vag_vlm.htm

"AtariAge - Atari Jaguar History"
... However, after the Summer CES that year, Atari announced that the Panther was cancelled so that they could concentrate on a new machine, the 64-bit Jaguar. ... (6)
http://www.atariage.com/jaguar?SystemID=JAGUAR

"Slashdot | New Atari Jaguar Game Running $1,225 on eBay"
... New Atari Jaguar Game Running $1,225 on eBay. Games. ... Bill Kendrick writes, "This long-awaited Atari Jaguar game Battle Sphere has finally been released. ... (8)
http://slashdot.org/articles/04/03/20/190232.shtml

"Slashdot | New Atari Jaguar Game Running $1,225 on eBay"
... New Atari Jaguar Game Running $1,225 on eBay. Games. ... Bill Kendrick writes, "This long-awaited Atari Jaguar game Battle Sphere has finally been released. ... (8)
http://slashdot.org/articles/04/03/20/190232.shtml
Kaki & Aula Key Findings
(all significant)

• Category use takes almost 2 times longer than linear
  – First doc selected in 24.4 sec vs 13.7 sec

• No difference in average number of docs opened per search (1.05 vs. 1.04)

• However, when categories used, users select >1 doc in 28.6% of the queries (vs 13.6%)

• Num of searches without 0 result selections is lower when the categories are used

• Median position of selected doc when:
  – Using categories: 22 (sd=38)
  – Just ranking: 2 (sd=8.6)
Conclusions from Kaki Study

• Simplicity of category assignment made groupings understandable
  – (my view, not stated by them)

• Keyword-based Categories:
  – Are beneficial when result ranking fails
  – Find results lower in the ranking
  – Reduce empty results
  – May make it easier to access multiple results
  – Availability changed user querying behavior
DynaCat, Pratt, Hearst, and Fagan.

• Medical Domain
• Decide on important question types in an advance
  – What are the adverse effects of drug D?
  – What is the prognosis for treatment T?
• Make use of MeSH categories
• Retain only those types of categories known to be useful for this type of query.
Query: What are the ways to prevent breast cancer?
(83 different references retrieved)

Behavior and Behavior Mechanisms (14 refs)
- Attitude (9 refs)
- Behavior (8 refs)
- Psychology, Social (3 refs)

Biochemical Phenomena, Metabolism, and Nutrition (5 refs)
- Diet (5 refs)

Chemicals and Drugs (52 refs)
- Amino Acids, Peptides, and Proteins (2 refs)
- Antineoplastic and Immunosuppressive Agents (18 refs)

Behavior and Behavior Mechanisms
- Attitude
  - Attitude to Health
    - Por La Vida intervention model for cancer prevention in Latinas.
    - Breast cancer prevention education at a shopping center in Israel: a student nurse community health project.
    - Future challenges in secondary prevention of breast cancer for women at high risk.
    - A study of diet and breast cancer prevention in Canada: why healthy women participate in controlled trials.

- Knowledge, Attitudes, Practice
  - Por La Vida intervention model for cancer prevention in Latinas.
DynaCat Study, Pratt, Hearst & Fagan

• Design
  – Three queries
  – 24 cancer patients
  – Compared three interfaces
    • ranked list, clusters, categories

• Results
  – Participants strongly preferred categories
  – Participants found more answers using categories
  – Participants took same amount of time with all three interfaces
Figure 3. Results from the validated, user-satisfaction questionnaire. The mean values across all 15 subjects are shown on the y axis. The x axis shows a brief summary of the questions asked. Subjects answered the questions using a scale from 1 to 5, where 1 meant almost never and 5 meant almost always (the ideal answer). The difference between DynaCat and the cluster tool was statistically significant ($p < 0.05$) for all questions, as was that between DynaCat and the ranking tool, with the exception of the question about sufficient information where $p = 0.11$. 

DynaCat study, Pratt et al.
The Idea of Facets

• Facets are a way of labeling data
  – A kind of Metadata (data about data)
  – Can be thought of as properties of items

• Facets vs. Categories
  – Items are placed INTO a category system
  – Multiple facet labels are ASSIGNED TO items
The Idea of Facets

• Create INDEPENDENT categories (facets)
  – Each facet has labels (sometimes arranged in a hierarchy)

• Assign labels from the facets to every item
  – Example: recipe collection

Cooking Method
  - Stir-fry

Course
  - Main Course

Cuisine
  - Thai

Ingredient
  - Chicken
  - Bell Pepper
  - Curry

Example: Recipe collection

[Diagram showing cooking method, course, cuisine, and ingredients connected to a central image of a dish]
Using Facets

• Now there are multiple ways to get to each item

- Preparation Method
  - Fry
  - Saute
  - Boil
  - Bake
  - Broil
  - Freeze

- Desserts
  - Cakes
  - Cookies
  - Dairy
    - Ice Cream
    - Sherbet
    - Flan

- Fruits
  - Cherries
  - Berries
    - Blueberries
    - Strawberries
    - Bananas
    - Pineapple

Fruit > Pineapple
Dessert > Cake
Preparation > Bake

Dessert > Dairy > Sherbet
Fruit > Berries > Strawberries
Preparation > Freeze
### Aduna Spectacle

#### Search

<table>
<thead>
<tr>
<th>Location</th>
<th>My Aduna Files</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Past year</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Older than 1 year</td>
<td>20</td>
</tr>
<tr>
<td>Type</td>
<td>Microsoft PowerPoint Presentation</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>OpenOffice Presentation</td>
<td>2</td>
</tr>
<tr>
<td>Author</td>
<td>Aduna sales</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Hans Nederbragt</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Jeroen Wester</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>sales</td>
<td>1</td>
</tr>
<tr>
<td>Language</td>
<td>Dutch</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>English</td>
<td>12</td>
</tr>
<tr>
<td>Size</td>
<td>10 - 50 KB</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>100 - 500 KB</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>500 KB - 1 MB</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1 - 5 MB</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>5 - 10 MB</td>
<td>2</td>
</tr>
</tbody>
</table>

### Results 1 - 10 of 21

1. **OpenRNA**
   - **Source:** My Aduna Files
   - **Microsoft PowerPoint Presentation**
   - **Significant Terms:** content, applications, sorts, open, standards, reference, users, software, retrieved, connect

2. **Presentatie GX 16 december 2004**
   - **Source:** My Aduna Files
   - **OpenOffice Presentation**
   - **Significant Terms:** search, information, metadata, aduna, exploration, based, navigation, spectacle, support, rdf

3. **Presentatie RIVM 9 december 2004**
   - **Source:** My Aduna Files
   - **OpenOffice Presentation**
   - **Significant Terms:** information, search, sdc, data, metadata, rdf, aduna, based, sdc, semantic

4. **Introduction Aduna for Gartner**
   - **Source:** My Aduna Files
   - **Microsoft PowerPoint Presentation**
   - **Significant Terms:** search, based, rdf, navigation, sesame, information, web, exploration, spectacle, semantic

### Selections

- Search: "semantic web"
- Search: metadata
- Type: Presentations
- Remove all

### Suggestions

- access
- aduna
- applicatie
- applications
- architectuur
- based
- browser
- browsing
- classification
- content
- control
- data
- demonstratie
- engine
- exploration
- extractie
- februari
- gebruiker
- guided
- informatie
- informatiemonitor
- information
- machine
- main
- navigation
- niveau
- open
- presentatie
Flamenco Usability Studies

• Usability studies done on 3 collections:
  – Recipes: 13,000 items
  – Architecture Images: 40,000 items
  – Fine Arts Images: 35,000 items

• Conclusions:
  – Users like and are successful with the dynamic faceted hierarchical metadata, especially for browsing tasks
  – Very positive results, in contrast with studies on earlier iterations.

Flamenco Study Post-Interface Assessments

Adjectives to Describe Interfaces

All significant at p<.05 except “simple” and “overwhelming”

Creative Facet Visualization

- Aduna Autofocus
Creative Facet Visualization

• Fathumb mobile search interface
  • http://research.microsoft.com/vibe/projects/FaThumb.aspx
Creative Facet Visualization

• Hutchinson et al.

[Image of a simple search interface with various categories and book images]
Summary: Grouping Search Results

Grouping search results seems beneficial in two circumstances:

1. General web search, using transparent labeling (monothetic terms) or category labels rather than cluster centroids.

Effects:
- Works primarily on ambiguous queries, (so used a fraction of the time)
- Promotes relevant results up from below the first page of hits
  - So important to group the related items together visually
- Users tend to select more documents than with linear search
- May work even better with meta-search
- Positive subjective responses (small studies)
The Document Lens
Fisheye Menu

Fisheye Views of Documents

2. Workspace Awareness

When people work together, they maintain an awareness of others that helps them coordinate activity and find opportunities to collaborate. This awareness, which we call group awareness (Gutwin and Greenberg, 1995a,b; Gutwin, Stark and Greenberg, 1995), is part of the "glue" that allows groups to be more effective than
Multiple Focii

2. Workspace Awareness

When people work together, they main awareness of others that helps them be active and find opportunities to collaborate. This awareness, which we call group awareness (Owens and Greenberg, 1993a,b; and Stark and Greenberg, 1996), is part of the "glue" that allows groups to be more effective than.
Does it always work?
TileBars

• All this zooming is making me dizzy
  – What about a more abstract representation of the document?
• Show a graphical representation of term distribution and overlap in search results
• Simultaneously Indicate:
  – Relative document length
  – Query term frequencies
  – Query term distributions
  – Query term overlap
Technique

Blocks indicate “chunks” of text, such as paragraphs

Blocks are darkened according to the frequency of the term in the document
Example

Topic: reliability of DBMS (database systems)
Query terms: DBMS, reliability

- Mainly about both DBMS and reliability
- Mainly about DBMS, discusses reliability
- Mainly about, say, banking, with a subtopic discussion on DBMS/Reliability
- Mainly about high-tech layoffs
"Color transformations. (guide to color-translation)

"Oracle makes inroads into imaging market.

"Mac-Sun-PC network takes off at Harvard.

"Forecast calls for faster weather data via satellite.

"Picture this: DEC adds scanner, imaging to hardware mix.

"Database stands ready as probe nears Venus.

"Challenging for the multimedia lead. (IBM)

"Ventura Publisher. (Software Review) (Xerox)

"Congress shuns FTS 2000, picks MCI. (Federal)

"Broadband ISDN waits in the wings. (integrator)

"DEC beefs up visualization to catch the eye.

"Image-processing system targets OS/2 workstations. (RSC)
TileBars Summary

• Compact, graphical representation of term distribution for full text retrieval results
  – Simultaneously display term frequency, distribution, overlap, and doc length
  – However, does not provide the context in which query terms are used

• Do they help?
  – Users intuitively understand them
  – Lack of context sometimes causes problems in disambiguation
health and safety of staff and inmates, the bureau or prisons will restrict areas and circumstances in which smoking is permitted within its institutions and offices.

(a) All areas of Bureau of Prisons facilities and vehicles are no smoking areas unless specifically designated as smoking areas by the Chief Executive Officer consistent with the guidelines set forth in this rule.

(b) Chief Executive Officers shall limit smoking areas to a minimum number of locations, consistent with effective operations. Under no circumstances shall smoking be permitted in the following areas, except as noted in §Section; 551.162(a):

(1) Elevators,
(2) Storage Rooms and Warehouses,
(3) Libraries,
(4) Corridors and Halls,
(5) Dining Facilities,
(6) Kitchen and Food Preparation Areas,
(7) Medical/Dental Care Delivery Areas,
(8) Institution/Government Vehicles,
(9) Administrative Areas and Offices,
(10) Auditoriums,
(11) Class and Conference Rooms,
(12) Gymnasiums and Exercise Rooms, and
(13) Restrooms.

$\&$Section; 551.161
Definition.

For purpose of this rule, smoking is defined as carrying or inhaling a lighted cigar, cigarette, pipe or other lighted tobacco products.
PadPrints

- Tree-based history of recently visited Web pages
  - History map placed to left of browser window
  - Node = title + thumbnail
  - Visually shows navigation history

- Zoomable: ability to grow and shrink sub-trees
PadPrints Screenshot
PadPrints Thumbnails
Zoomable History
The Importance of Interfaces

• The user interface is the part of the system that the user interacts with:
  – Interaction is an integral part of the information seeking process
  – Search experience is affected by the quality of the interface

• Interfaces should
  – Help users get started
  – Help users keep track of what they have done
  – Help users make sense of what the system did
  – Suggest next choices

• It is very difficult to design good UIs
• It is very difficult to evaluate search UIs
Challenge: Search Everywhere on All Devices

- **Search Behavior for mobile:**
  - Multi-touch, multi-modal behavior models
  - Grounding search in physical and social context

- **Incorporate sensor data**
  - Speech-based search: acoustics, other sensors
Looking Forward

• Behavior, Context Models for Ubiquitous Search

• New tools for analysis of cognitive processing
Future Search Interfaces: Visions

• Intelligent Assistant:
  – Dialogue-based
  – Explicit interaction

• Augmented reality:
  – Information in Context
  – Explicit & Implicit
Challenges/Opportunities

- Mobile devices with eye tracking
- User interfaces, query specification
- Result presentation
- Feedback modalities and interpretation
- Shared Tools (UIs, benchmarks, evaluation)?
- Privacy-preserving analysis
Concentrate on the MentalPlex spiral and project an image of what you wish to find.

Note: This page posted for April Fool's Day - 2000.
Modern Information Retrieval, Chapter 10:

http://people.ischool.berkeley.edu/~hearst/irbook/print/chap10.pdf
Ethics

• How Google Could Rig the 2016 Election - POLITICO ... 
  http://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548

• Cyberchondria is a growing concern among many healthcare practitioners as patients can now research any and all symptoms of a rare disease, illness or condition, and manifest a state of medical anxiety.

• https://en.wikipedia.org/wiki/Cyberchondria/