

Search User Interfaces

Marti Hearst
UC Berkeley

Chapter Structure

- 1: Design of Search Interfaces
- 2: Evaluation of Search Interfaces
- 3: Models of the Search Process
- 4: Query Formulation
- 5: Retrieval Results
- 6: Query Reformulation
- 7: Supporting the Process of Search
- 8: Integrating Navigation and Search
- 9: Personalization
- 10: Information Visualization and Search
- 11: Visualization for Text Analysis
- 12: Future Trends in Search Interfaces

Chapter Structure

- 1: Design of Search Interfaces
- 2: Evaluation of Search Interfaces
- 3: Models of the Search Process
- 4: Query Formulation
- 5: Retrieval Results
- 6: Query Reformulation
- 7: Supporting the Process of Search
- 8: Integrating Navigation and Search
- 9: Personalization
- 10: Information Visualization and Search
- 11: Visualization for Text Analysis
- 12: Future Trends in Search Interfaces



1. The Design of Search Interfaces



The paradox of web search

- Why is search difficult?
 - Must support an enormous range of use cases.
 - Over an enormous collection of information.
 - Used by an wide range of people.
 - Requires reading, which requires one's full attention.
 - Ideas can be expressed many different ways.

The paradox of web search

- Why is search easy?
 - On the web, the collection is so big that the same information is often stated many different ways (in English).
 - Many people often look for the same information as many other people.
 - A very simple interface, along with highly tuned algorithms, has proven highly effective in many cases.



2. The Evaluation of Search Interfaces



Evaluation Techniques

- Informal (discount) usability testing
- Formal laboratory studies
- Field studies
- Longitudinal studies
- Log file analysis
- Large-scale testing (A/B testing, bucket testing, parallel flights)

After 1 Week

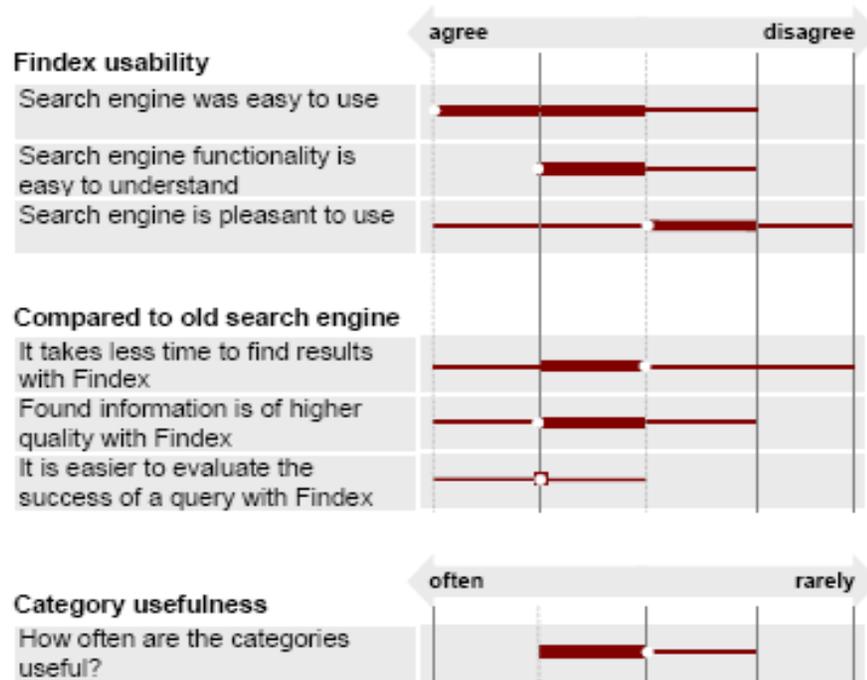


Figure 5. Answer distributions as box plots for the first questionnaire. Thin line represents range of observations, thick 50% of them and white dot the median.

After 2 Months

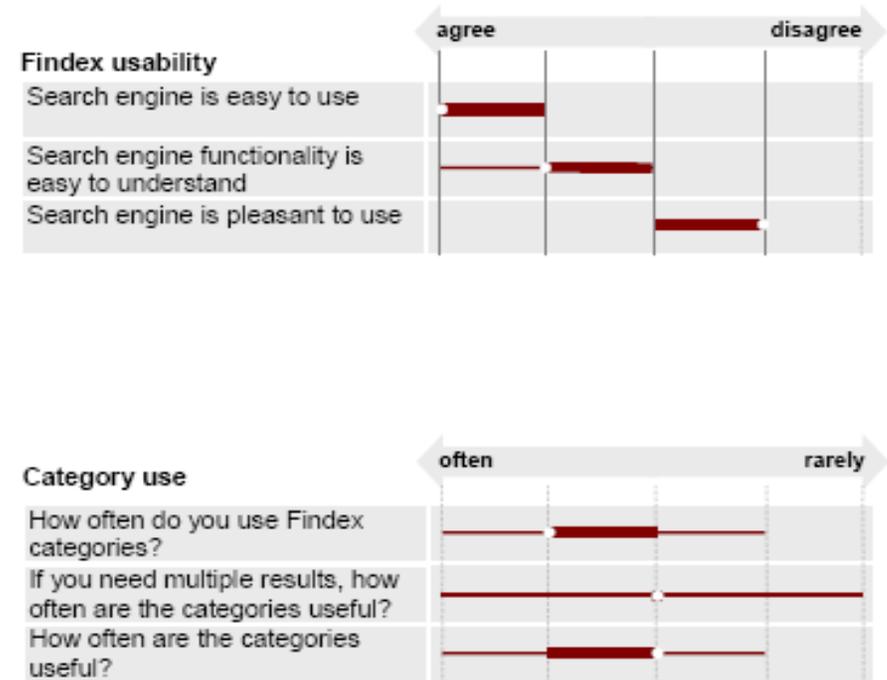


Figure 6. Answer distributions as box plots for the second questionnaire. Thin line represents range of observations, thick 50% of them and white dot the median.

Evaluation Essentials

- Matching participants to tasks
 - I love cooking, but I hate recipes!
 - Students don't care about campus administration
- Participants should be highly motivated
 - Allow for selection among topic choices
 - Spool's technique:
 - Let them dream about spending money.

Evaluation Essentials

- There is more variation in one system across tasks than across systems.
 - Some studies are now focusing on evaluating one interface across different task types (Woodruff et al. 2001, Baudisch et al. 2004)
- Differences in the cognitive abilities of individual participants is a better predictor of performance than differences in systems.
 - So need to have a large participant pool, and vary the ordering of the study conditions.

Evaluation Essentials

- Don't evaluate your own design
- Or if you do:
 - Don't "leak" which one you think is best
 - "We are assessing these designs", not "we designed this and are now assessing it."
 - Use generic names for the different designs
 - Plan to have a strong, state-of-the-art baseline for comparison.
 - Use the same collections for each design.
 - Make all designs aesthetically pleasing.



6. Query Reformulation



Cognitive Principles

- Recognition over Recall
- Multiple Means of Expression (The Vocabulary Problem)
- Anchoring
- Addressing these:
 - Modern query and query reformulation aids.
 - Modern site navigation and search aids.

Recognition over Recall

- It is easier to recognize some information than generate it yourself.
 - Learning a foreign language
 - Recognize a face vs. drawing it from memory

Multiple Means of Expression

- People remember the gist but not the actual words used.
- People can agree on the meaning of a label, even though with no other cues they generate different labels.
 - The probability that two typists would suggest the same word: .11, and the probability that two college students would name an object with the same word: .12. (Furnas et al. 1987)

The Vocabulary Problem

- There are many ways to say the same thing.
 - How much does that camera cost?
 - How much for that camera?
 - That camera. How much?
 - What is the price of that camera?
 - Please price that camera for me.
 - What're you asking for that camera?
 - How much will that camera set me back?
 - What are these cameras going for?
 - What's that camera worth to you?
- The interface needs to help people find alternatives, or generate them in the matching algorithm.

The Problem of Anchoring

- Ariely discusses this in *Predictably Irrational*
 - Tell people to think of the last 2 digits of their SSN
 - Then have them bid on something in auction
 - The numbers they thought of influenced their bids
- Anchoring in search
 - Start with a set of words
 - Difficult to break out and try other forms of expression
 - Example from Dan Russell:
 - Harry Potter and the Half-Blood Prince sales
 - Harry Potter and the Half-Blood Prince amount sales
 - Harry Potter and the Half-Blood Prince quantity sales
 - Harry Potter and the Half-Blood Prince actual quantity sales
 - Harry Potter and the Half-Blood Prince sales actual quantity
 - Harry Potter and the Half-Blood Prince all sales actual quantity
 - all sales Harry Potter and the Half-Blood Prince
 - worldwide sales Harry Potter and the Half-Blood Prince

Query Suggestion Aids

- Early systems showed huge numbers of choices
 - Often machine-generated
 - Often required the user to select many terms
- Newer approaches
 - Can select just one term; launches the new query
 - Queries often generated from other users' queries
 - Have good uptake (~6% usage) Anick and Kantamneni, 2008

Dynamic Query Suggestions

- Shown both dynamically, while entering initial query, and static, after the query has been issued.

The image displays two examples of dynamic query suggestions in a search interface. The first example shows a search bar with the text 'be' and a dropdown menu of suggestions including 'Being There', 'Being John Malkovich', 'Being Julia', 'Beijing Bicycle', 'Beiderbecke Connection', 'The Unbearable Lightness of Being: Special Edition', 'The Importance of Being Earnest', 'The Beiderbecke Affair', 'The Business of Being Born', and 'Devious Beings'. The second example shows a search bar with the text 'emb' and a dropdown menu of suggestions including 'EMB (Emb)', 'EMB (MGC71745, Gp70, AL022799, MGC21425)', 'EMB (Ethambutol)', 'EMB (Methylurethane)', 'Emb1', 'Emb2', 'EMBBA (Embba)', 'Embryo', 'Embarin (Allopurinol)', and 'Embutox (Butoxone)'. The interface also includes a 'Search' button, a 'Winners' section, and a 'NEXTBIO' logo.

Post-Query Suggestions

- Shown after the query has been issued.

YELLOWSTONE

- Yellowstone **Lodging**
- Yellowstone **Map**
- Yellowstone **Hotels**
- Yellowstone **Weather**
- Yellowstone **Vacations**
- Reference**

Web Images Video Local Shopping More ▾

yellowstone

yellowstone national park
yellowstone volcano
yellowstone lodging
yellowstone park
west yellowstone

Explore related concepts:
Yellowstone National Park
Yellowstone Vacation
Yellowstone Park
National Park Service

Old Faithful
geysers
national park
camping

RELATED SEARCHES

Yellowstone **National Park**
Pictures

Old Faithful Inn

Yellowstone **Volcano**

History of Yellowstone
National Park

Yellowstone **RV**

Yellowstone **National Park**
Campgrounds

Suggesting Destinations

- Record search sessions for 100,000's of users
- For a given query, where did the user end up?
 - Users generally browsed far from the search results page (~5 steps)
 - On average, users visited 2 unique domains during the course of a query trail, and just over 4 domains during a session trail
- Show the query trail endpoint information at query reformulation time
 - Query trail suggestions were used more often (35.2% of the time) than query term suggestions (White et al. 2007)



Showing Related Documents

- Can be a “black box” and so unhelpful
- But in some circumstances, works well
 - Related articles in a search tool over biomedical text

Related Links

- ▶ Creating an online dictionary of abbreviations from MEDLINE. [J Am Med Inform Assoc. 2002]
- ▶ Automatic extraction of acronym-meaning pairs from MEDLINE databases [Medinfo. 2001]
- ▶ Resolving abbreviations to their senses in Medline. [Bioinformatics. 2005]
- ▶ ALICE: an algorithm to extract abbreviations from MEDLINE. [J Am Med Inform Assoc. 2005]
- ▶ Ranking the whole MEDLINE database according to a large [BMC Bioinformatics. 2005]
- ▶ [See all Related Articles...](#)

Relevance Feedback

- A darling of the ranking community
 - User selects relevant documents, these are used to re-rank the results list.
- But ... doesn't really work in practice
 - Works best for tasks requiring high recall
 - Results are unreliable, which is poor for usability
 - Relevance judgements are difficult to make
 - The more you know, the easier it is to make relevance judgements, and typically the less you need this tool.



8. Integrating Navigation and Search



Integrating Navigation and Search

- Key points:
 - Show users structure as a starting point, rather than requiring them to generate queries
 - Organize results into a recognizable structure
 - Aids in comprehension
 - Suggests where to go next within the collection
 - Eliminate empty results sets
- Techniques:
 - Flat lists of categories
 - Faceted navigation
 - Document clustering



10. Visualization in Search

Will visual search ever succeed?

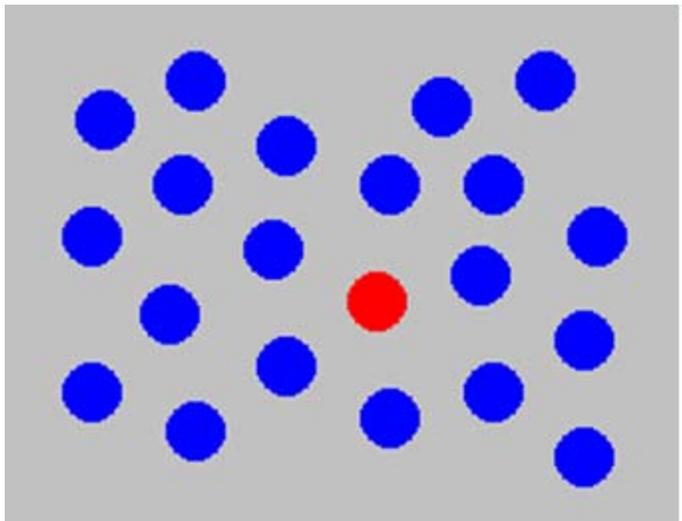


Properties of Text

- Reading requires full attention.
- Reading is not pre-attentive.
- Can't graph nominal data on axes.

Pre-attentive Properties

- Humans can recognize in under 100ms whether or not there is a green circle among blue ones, independent of number of distractors.

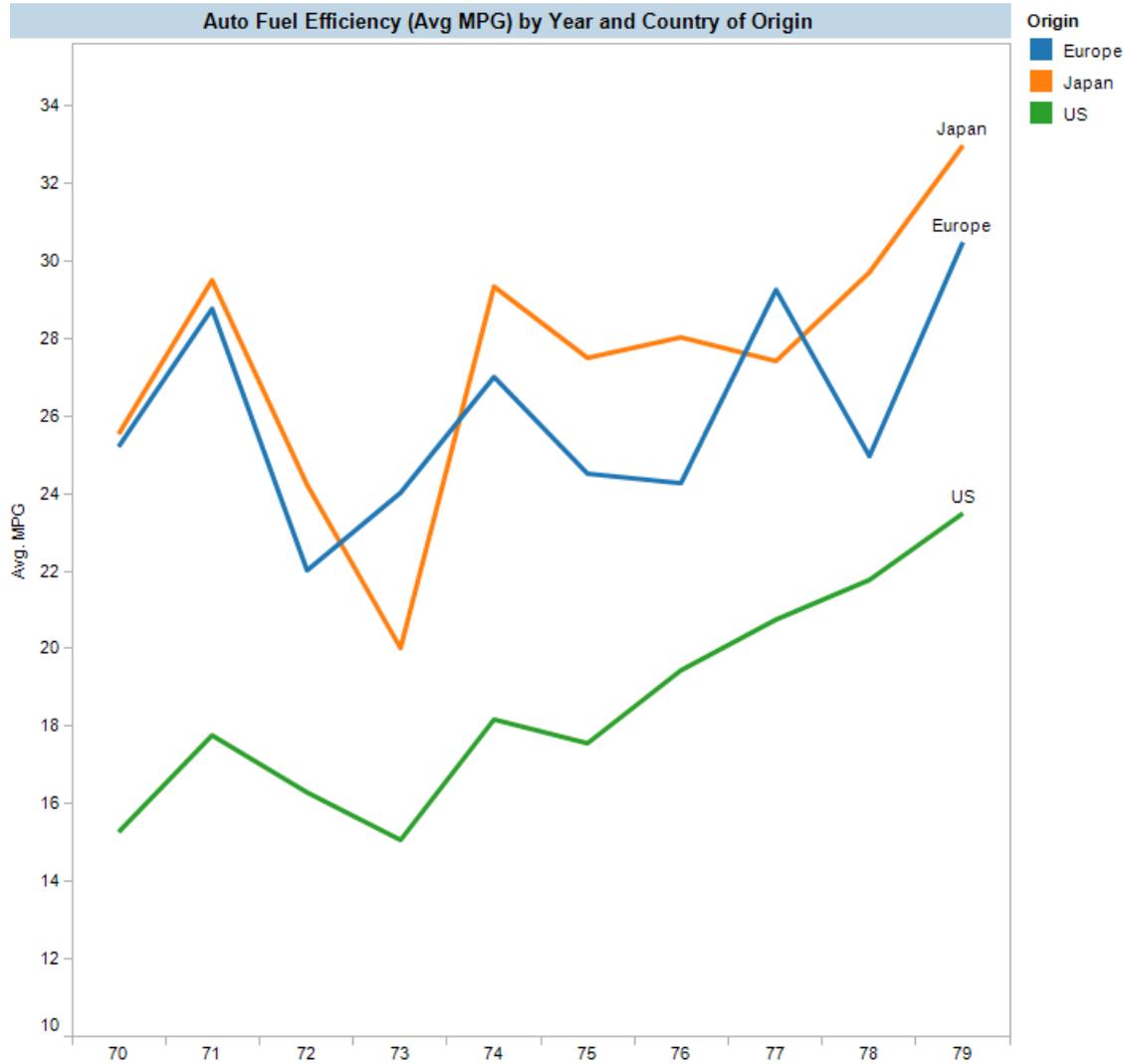


- This doesn't work for text.

Text is NOT Preattentive

SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

Quantitative Data is Easy to Visualize



**Auto data:
Comparing Model year
vs. MPG by Country**

Nominal Data is Difficult to Visualize



**Auto data:
Comparing MPG by
Model name by
Country**

**A non-sensical
visualization.**

Search Results: Thumbnail Images of Pages



[Dziadosz and Chandrasekar, 2002](#) Showing thumbnails alongside the text made the participants much more likely to assume the document was relevant (whether in fact it was or not).

Search Results: Thumbnail Images of Pages

- Results tend to be negative.
 - E.g., Blank squares were just as effective for search results as thumbnails, although the subjective ratings for thumbnails were high. (Czerwinski et al., 1999)
- BUT
 - People love visuals,
 - Technology is getting better (see SearchMe)
 - Making important text larger improves search for some tasks (Woodruff et al. 2001)
 - Earlier studies maybe used thumbnails that were too small. (Kaasten et al. 2002, browser history)
 - Showing figures extracted from documents can be useful. (Hearst et al. 2007)

Search Results: Thumbnail Images of Pages



BioText SEARCH ENGINE [Home](#) | [About BioText](#) | [Contact Us](#)

Search:

Search Over: Full Text & Abstracts Figure Captions (List) Figure Captions (Grid) Tables Sort By: Relevance Results Page: 20

Results 61.80 of 417 searching full text < 1 2 3 4 5 6 7 >

ABSTRACTS FULL-TEXT EXCERPTS FIGURES

The small molecule Mek1/2 inhibitor U0126 disrupts the chordamesoderm to notochord transition in zebrafish
 Hawkins, T., Cavodeassi, F., Erdélyi, F., Szabó, G., Lele, Z. (2008) *BMC Developmental Biology*.

ABSTRACT
 Key molecules involved in notochord differentiation and function have been identified through genetic analysis in **zebrafish** and mice, but MEK1 and 2 have so far not been implicated in this process due to early lethality (Mek1^{-/-}) and functional redundancy (Mek2^{-/-}) in the knockout animals. ... [Show Full Abstract](#)

FULL-TEXT EXCERPTS
 ...develops from the chordamesoderm, a derivative of dorsal mesoderm, and is ultimately incorporated into the forming vertebrae as the nucleus pulposus. Here we report the analysis of **zebrafish embryos** treated with the MEK1/2 inhibitor U0126 which causes an almost 100% penetrant, dose-dependent and reproducible phenotype consisting of short trunk and tail, lack of dark pigmentation, and abnormal...
 ...in **zebrafish** [33]. Consistent with those analyses, treatment of [Show Full Excerpts](#)

VIEW FULL ARTICLE: [HTML](#) | [PDF](#)

Distinct Functions for Different scl Isoforms in Zebrafish Primitive and Definitive Hematopoiesis
 Qian, F., Zhen, F., Xu, J., Huang, M., Li, W., Wen, Z. (2007) *PLoS Biology*.

ABSTRACT
 The stem-cell leukemia (SCL, also known as TAL1) gene encodes a basic helix-loop-helix transcription factor that is essential for the initiation of primitive and definitive hematopoiesis, erythrocyte and megakaryocyte differentiation, angiogenesis, and retroviral development. Here we report that the **zebrafish** produces, through an

FIGURES FROM ARTICLE:

[View all figures \(7\) and tables from this article.](#)

FIGURES FROM ARTICLE:

Search Results: Thumbnail Images of Pages





12. Future Trends in Search Interfaces



Future Trends in Search Interfaces

- Longer, more natural queries
- Better Mobile Interfaces
 - Audio (spoken) queries and results
- Social / Collaborative search
- Longer term:
 - Video and audio dominating text
 - Dialogue / conversational interactions

Future Trends: Longer, more natural queries

- The research suggests people prefer to state their information need rather than use keywords.
 - But after first using a search engine they quickly learned that full questions resulted in failure.
- Average query length continues to increase
 - Major search engines are now handling long queries well.
- Information worded as questions is increasing on the web.
 - From social question-answering sites and forums.

Naturally-worded queries and social media



how do I make a pb&j

Search

[Advanced Search](#)

Web [+ Show options...](#)

Results 1 - 10 of about 285,000 for

[How to Make a Peanut Butter and Jelly Sandwich: 4 steps \(with ...](#)

Get a good amount of peanut butter on a knife or spoon (this amount varies ... Consider removing the dry crust with a well-shaped cookie cutter or special ...

[www.wikihow.com](#) > ... > - [Cached](#) - [Similar](#)

[Discussion:Make a Peanut Butter and Jelly Sandwich - wikiHow](#)

Nov 14, 2009 ... I totally agree with 71.228.33.204. I mean, come on? ... just say how to **make** it ... im 15 and my 5 year old neice knows how ... im glad its up ...

[www.wikihow.com/Discussion:Make-a-Peanut-Butter-and-Jelly-Sandwich](#) - [Cached](#) - [Similar](#)

[How to Make a Peanut Butter and Jelly Sandwich | eHow.com](#)

How to **Make a Peanut Butter and Jelly** Sandwich. With Video! **Peanut butter and jelly** sandwiches are popular amongst children and adults. Yes, anyone can **make** ...

[www.ehow.com](#) > ... > - [Cached](#) - [Similar](#)

[How to Make a Triple-Decker Peanut Butter and Jelly Sandwich ...](#)

How to **Make** a Triple-Decker **Peanut Butter and Jelly** Sandwich. Great for bag lunches, this variation of the classic **peanut butter and jelly** sandwich won't ...

[www.ehow.com](#) > ... > - [Cached](#) - [Similar](#)

[+ Show more results from www.ehow.com](#)

[Video results for how do I make a pb&j](#)

[YouTube - How To Make A PB&J Sandwhich](#)

4 min 2 sec - Nov 7, 2008

[www.youtube.com](#)

[Weekend Project: Make a peanut butter and](#)

[... astrology.yahoo.com](#)

Future Trends: Social Search

- Social ranking (see also Ch.9, Personalization)
 - Explicitly recommended
 - Digg, StumbleUpon
 - Delicious, Furl
 - Google's SearchWiki
 - Implicitly recommended
 - Click-through
 - People who bought...
 - Yahoo's MyWeb (now Google Social Search)

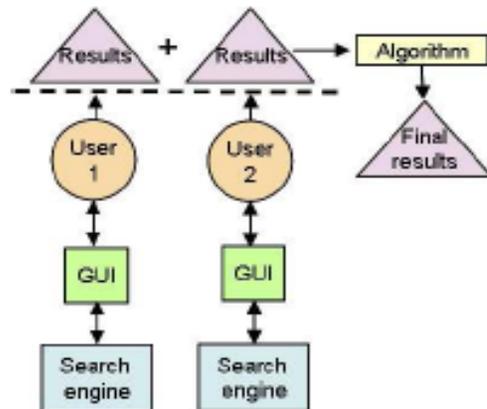


Research on Collaborative X

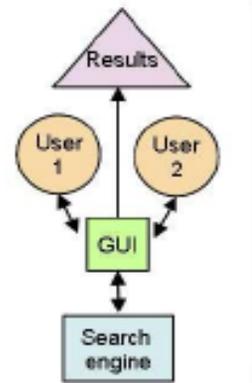
Collaborative Search

Collaborative Visualization

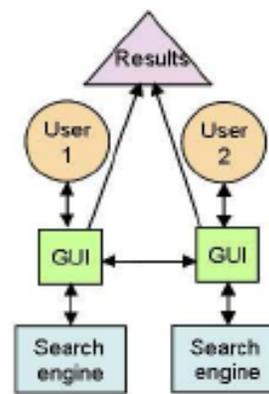
Collaborative Search



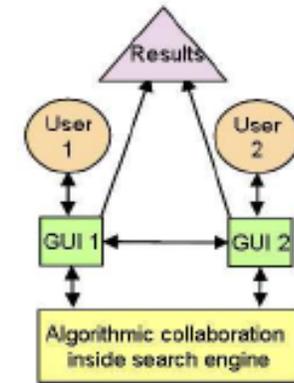
(1) Baeza-Yates, 1997



(2) Smeaton, Foley, 2005



(3) Morris, 2007

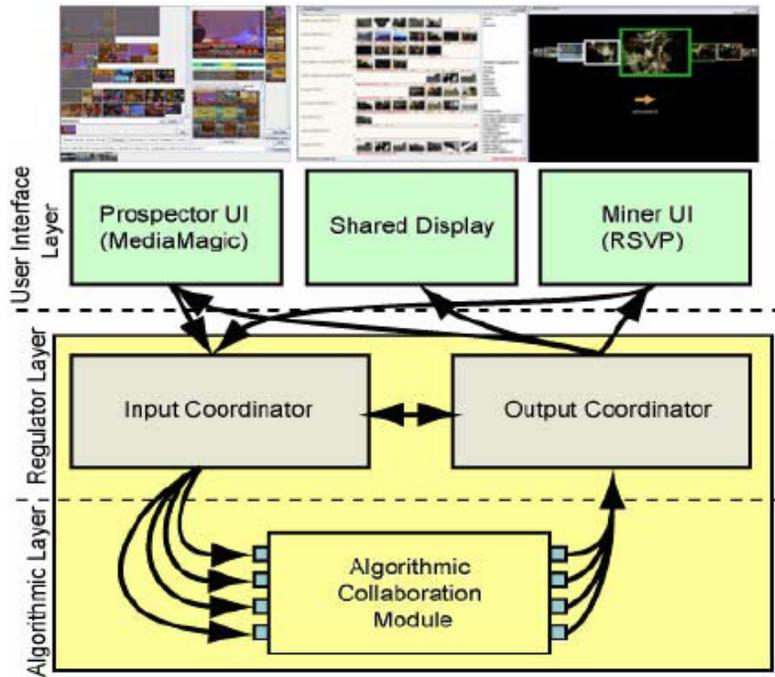


(4) Pickens et al. 2008

Collaborative Search



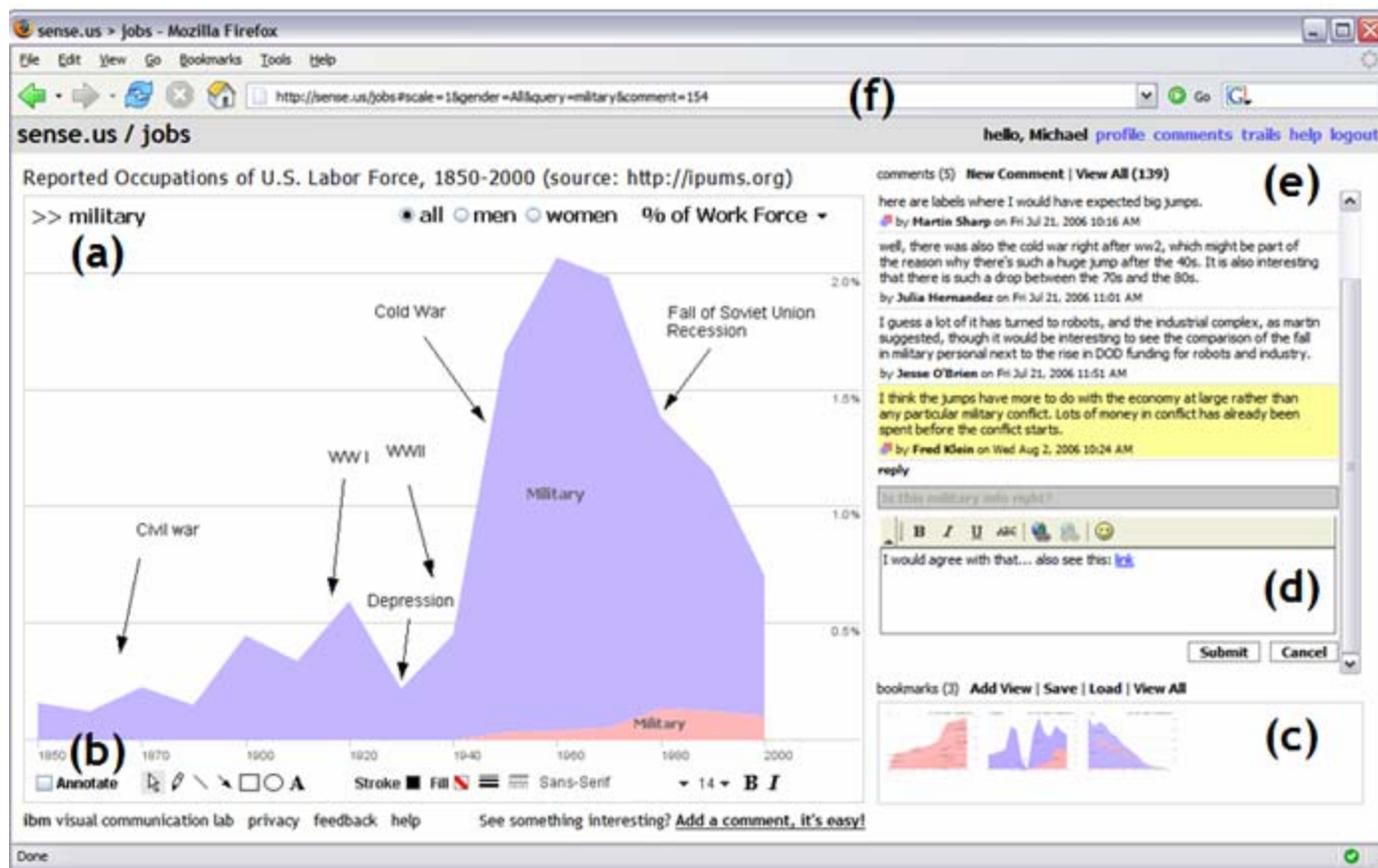
The screenshot shows a 'Shared Display' window with a list of explored items and their corresponding image thumbnails. The items include 'major spoor (MVED) 13', 'railway (MVED) 12', '(Image Only) 11', 'train station subway (MVED) 10', 'wall AT edition looking (MVED) 9', '(Image Only) 8', 'locomotive (MVED) 7', 'rail (MVED) 6', 'railroad (MVED) 5', and '(Image Only) 4'. To the right, there are sections for 'RSVP User Queries' and 'System suggestions'. Below the main window is a 'Shared Display System' window showing a sequence of images with a central image highlighted in a green box and an 'ADVANCE' button below it.



Pickens et al. 2008

Collaborative Visualization

Sense.us (Heer): collaborative analysis around viz



Future Trends: The Decline of Text

- The “cultural heavy lifting” in America is moving from text to audio and video.
 - Video and audio are now easy to produce and share.
 - Pew: Use of video sharing sites doubled from 2006-2009
 - YouTube: Video “responses” arose spontaneously
 - Videos for presidential debates were mundane.
 - Millions of video views; no where near this number for article readings
 - Pew: Marketing emails with podcasts 20% more likely to be opened.
 - Movies with subtitles do poorly in the U.S.
 - NYTimes: news web sites are starting to look like tv.
- The main impediment is the need for better search and scanning of audio and video information.

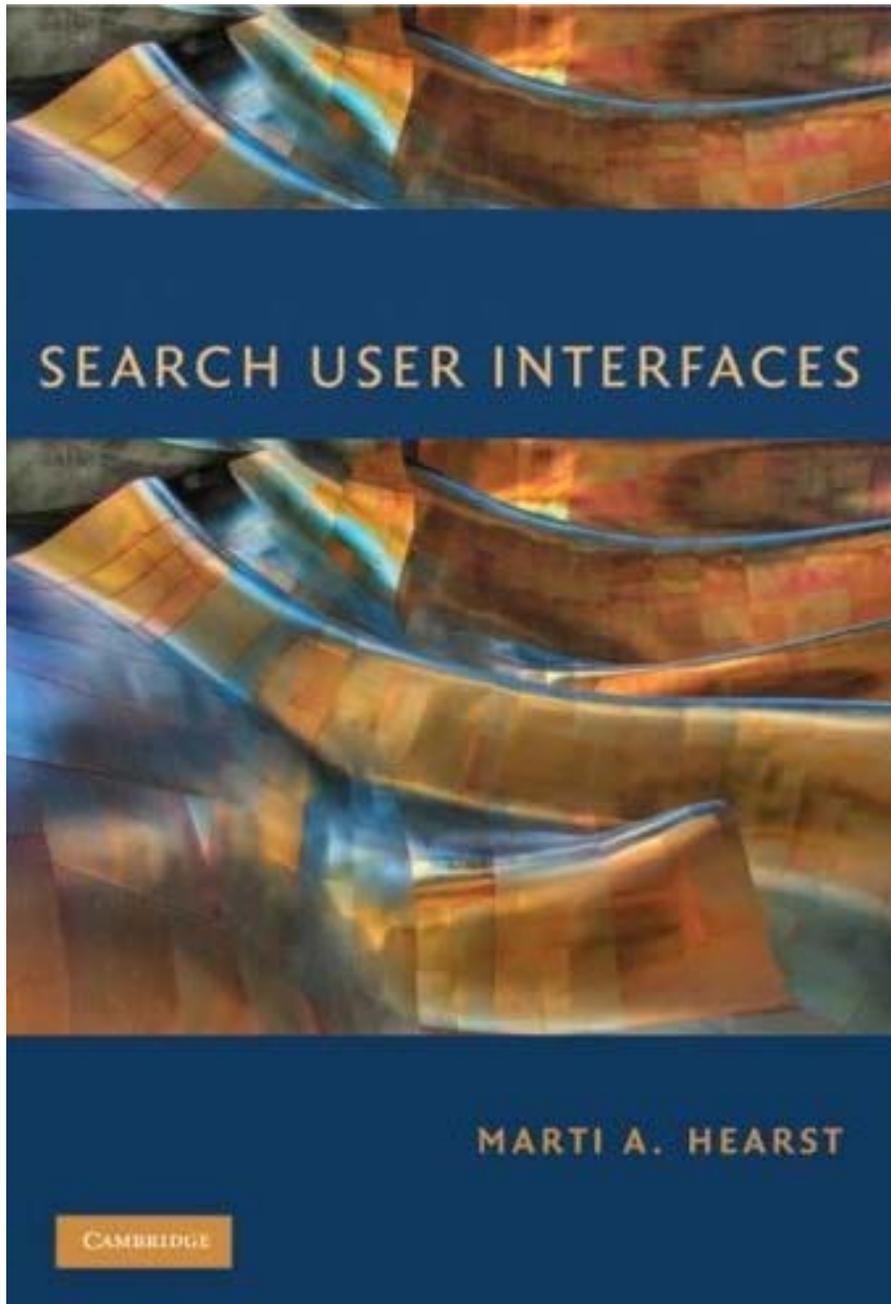
Far Future Trend: Dialogue

- We're still far away.
- SIRI is promising as a move forward; based on state-of-the-art research.



Future Trends ... not so much?

- Personalization
- Visualization ... some breakthroughs are needed.



Thank you!

Full text freely available at:
<http://searchuserinterfaces.com>