



# IS2140 Information Storage and Retrieval



## Unit 8: Interactive IR



**Daqing He**  
School of Information Sciences  
University of Pittsburgh

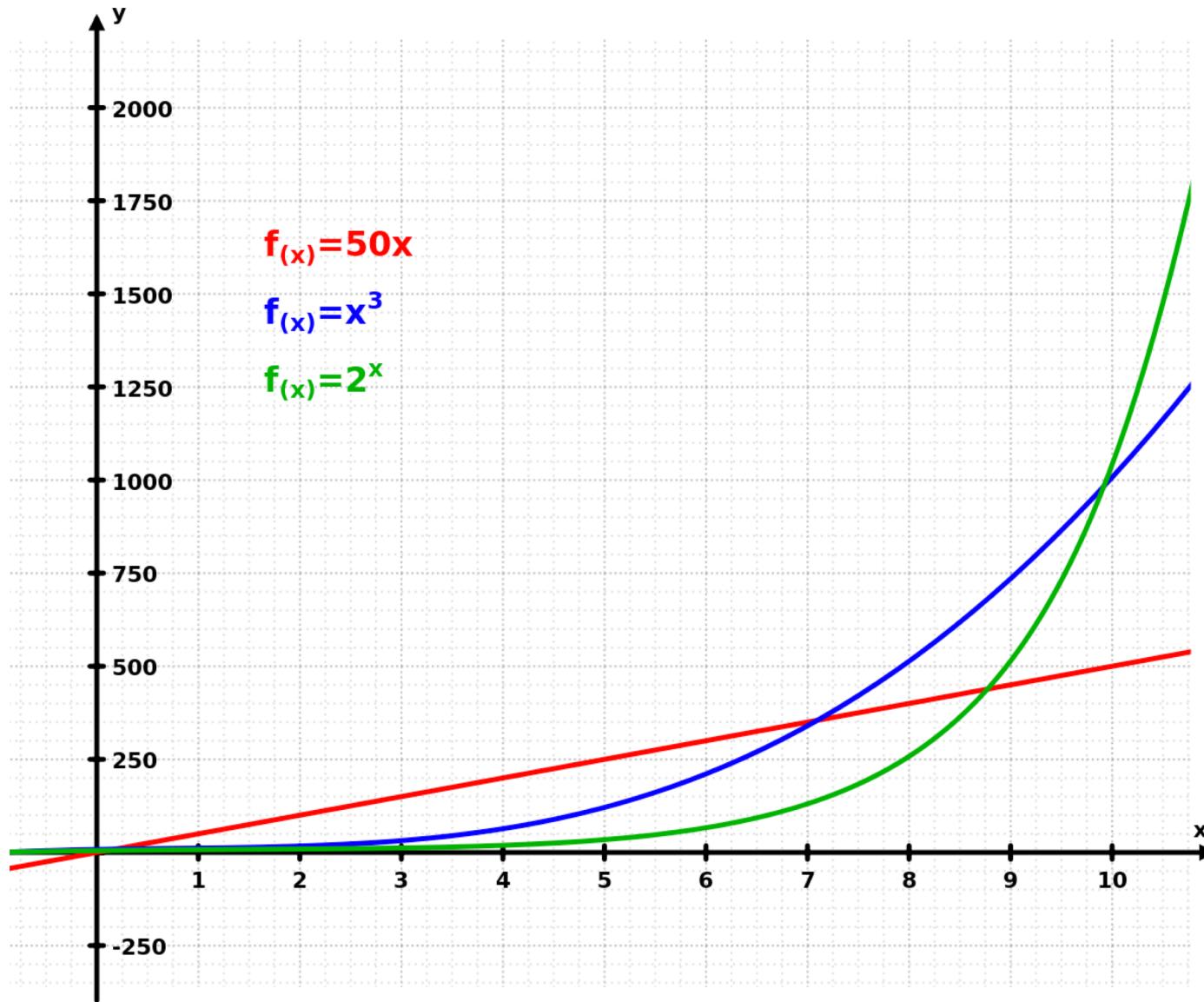
October 22, 2018

# Muddiest Points

- Evaluation
  - How do we judge whether the system is good or not when users perform different qualities of queries, some queries with high relevance results and the others with low relevance results?
  - When we calculate the mean average precision(MAP) for a search engine, we use the average precision for each query result and then calculate the mean value as the MAP. But why is the ranking of each retrieval result be different if the ranking algorithm within the search engine is the same?

# Muddiest Points

- Evaluation Measures
  - I think the measurements for precision and ranked retrieval are very interesting, but I didn't catch the answer for which type of search best suits each measurement.
  - Precision at fixed rank points is unfair for different queries, so how can we determine the the rank point for a query?
  - In discounted cumulative gain (DCG), we discount each result document's graded relevance based on the ranking. Mathematically, why do we use the exponential function as the numerator instead of linear function? Why don't we just say the discounted gain for a document i is  $\text{rel}(i)/\log_2(1+i)$ ?



# Muddiest Points

- Evaluation Measures
  - In the slide of page 22. The Precision decreases 0.5 for the second item when it fails to predict the first one correctly. But this doesn't apply to recall. Can I interpret the changes of recall and precision in this way that the ranking level has more impact on the precision than on the recall?
  - When we use the MAP as evaluation, can I see the area under the curve(on slides P16) as the score? Is the area larger the performance better?
    - <https://datascience.stackexchange.com/questions/806/advantages-of-auc-vs-standard-accuracy>

# Muddiest Points

- Relevance
  - How to determine the threshold of relevance? Or how to determine the threshold of average precision value?
  - About "Aboutness" and "Usefulness", there are some ways measuring "aboutness", but how to measure if we only concerns about "usefulness"?
  - Has anyone found value in incorporating feedback that is richer than a binary 'relevant' or 'not relevant'? For example, with a multi-term query a document could be relevant to one of the terms but not the other. Or one could incorporate feedback along other dimensions than relevance, like whether a document is authoritative or whether it's targeted at entry-level vs expert readers in a topic.

# Muddiest Points

- Implicit Relevance Feedback
  - In implicit feedback, can we return different top k relevant terms based on the users' demonstrated interest.
  - Could we use implicit feedback to evaluate the results when the ground truth is not available?

# Agenda

- What is Interactive Information Retrieval
- Design for Interactive Information Retrieval
- Query Formulation and Reformulation
- Result Presentation
- Mobile Web Search
- Evaluate Interactive Information Retrieval Systems

# Learning Objectives

- Clarifying the characteristics of interactive information retrieval, distinguishing it from static IR;
- Illustrating the design principles in IIR systems;
- Recognizing the interactive features in a search system
- Classifying the methods to evaluate an IIR system;
- Comparing and evaluating IIR systems.

# Relevance Feedback

# Types of Relevance Feedback

- Interactive relevance feedback: feedback information obtained from the user
  - Explicit relevance feedback
    - users explicitly mark relevant and irrelevant documents in the search results
    - We have shown the examples
  - Implicit relevance feedback
    - system attempts to infer user intentions based on observable behavior
- Blind relevance feedback or pseudo relevance feedback
  - feedback in absence of any evidence, explicit or otherwise
  - System assumes that the top ranked documents as relevant docs

# Pseudo Relevance Feedback

- Also called blind relevance feedback
  - Avoid obtain user explicit or implicit feedback information
  - Use top returned N docs in the initial result for RF
  - Assume all of them are relevant

**Any problem of this?**

# Does Pseudo RF Work?

- A study by Jimmy Lin
- Retrieval engine: Indri
- Test collection: TREC, topics 301-450
- Procedure:
  - Used topic description as query to generate initial hit list
  - Selected top 20 terms from top 20 hits using  $tf.idf$
  - Added these terms to the original query

# Pseudo RF Example

Number: 303

Title: Hubble Telescope Achievements

Description:

Identify positive accomplishments of the Hubble telescope since it was launched in 1991.

Narrative:

Documents are relevant that show the Hubble telescope has produced new data, better quality data than previously available, data that has increased human knowledge of the universe, or data that has led to disproving previously existing theories or hypotheses. Documents limited to the shortcomings of the telescope would be irrelevant. Details of repairs or modifications to the telescope without reference to positive achievements would not be relevant.



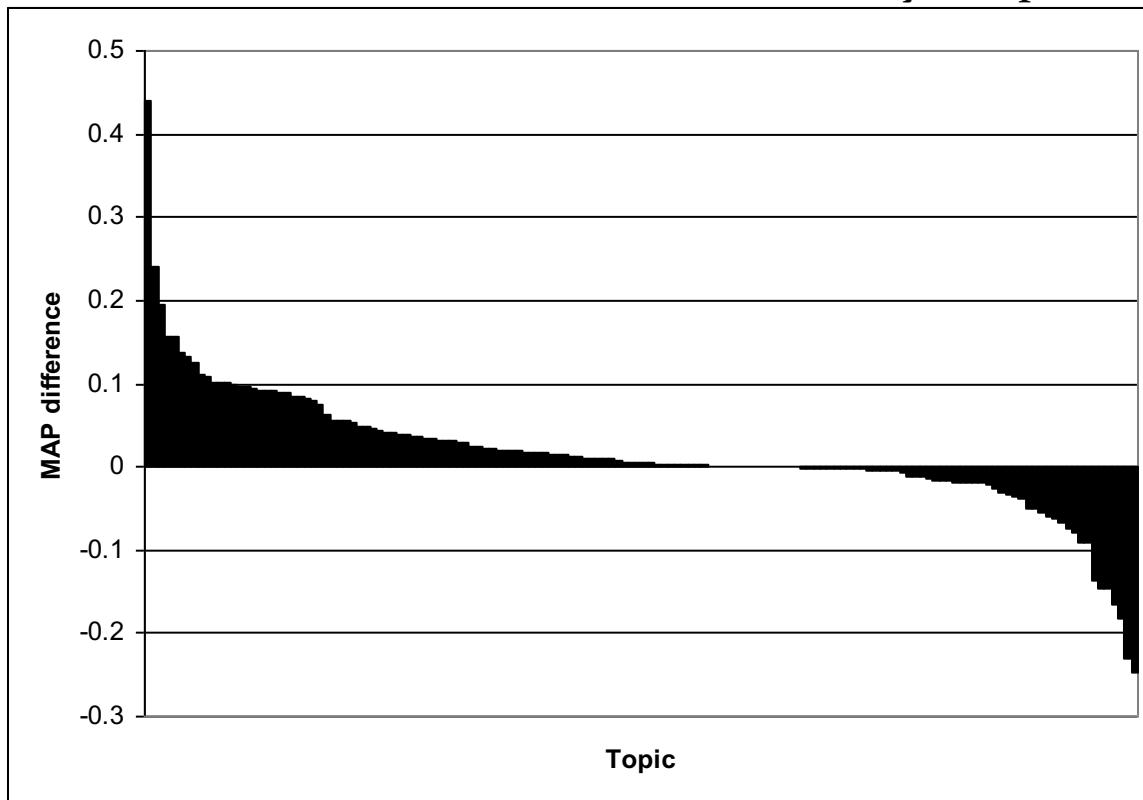
Terms added

telescope	1041.33984032195
hubble	573.896477205696
space	354.090789112131
nasa	346.475671454331
ultraviolet	242.588034029191
shuttle	230.448255669841
mirror	184.794966339329
telescopes	155.290920607708
earth	148.865466409231
discovery	146.718067628756
orbit	142.597040178043
flaw	141.832019493907
scientists	132.384677410089
launch	116.322861618261
stars	116.205713485691
universe	114.705686405825
mirrors	113.677943638299
light	113.59717006967
optical	106.198288687586
species	103.555123536418

# Experiment Results

	MAP	R-Precision
No feedback	0.1591	0.2022
With feedback	0.1806 (+13.5%)	0.2222 (+9.9%)

Blind relevance feedback doesn't always help!



# How to use Relevance Feedback?

- Assume that there is an optimal query
  - relevance feedback helps to bring user's query closer to the optimal one

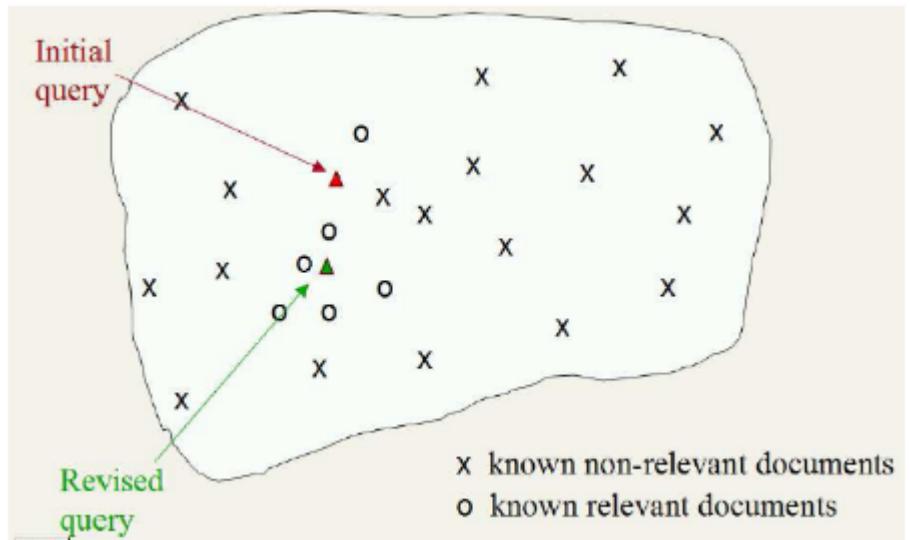
- How?

- Term reweighting:

- boost weights of terms from relevant documents

- Query expansion:

- Add terms from relevant documents to the query



# Relevance Feedback in Vector Space Model

- Rocchio Algorithm  $\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$

- Often alpha = 1, Beta = 0.75, Gamma = 0.15
- Ide modifications
  - Ide: Beta and Gamma are 1, no normalization on Dr and Dnr

$$\vec{q}_m = \alpha \vec{q} + \beta \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \gamma \sum_{\forall \vec{d}_j \in D_{nr}} \vec{d}_j$$

- Ide dec-hi: Beta and Gamma are 1, no normalization on Dr and Dnr, negative result only consider the highest ranked non-relevant doc

$$\vec{q}_m = \alpha \vec{q} + \beta \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \gamma \max_{non-relevant}(\vec{d}_j)$$

- Ide dec-hi is the most effective when there are a few relevant docs

# Rocchio in Example

query vector =  $\alpha \cdot$  original query vector

+  $\beta \cdot$  positive feedback vector

-  $\gamma \cdot$  negative feedback vector

Typically,  $\gamma < \beta$

Original query

0	4	0	8	0	0
---	---	---	---	---	---

$\alpha = 1.0$

0	4	0	8	0	0
---	---	---	---	---	---

Positive Feedback

8	4	8	0	0	2
---	---	---	---	---	---

$\beta = 0.5$

4	2	4	0	0	1
---	---	---	---	---	---

(+)

Negative feedback

0	0	4	4	0	1
---	---	---	---	---	---

$\gamma = 0.25$

0	0	1	1	0	.25
---	---	---	---	---	-----

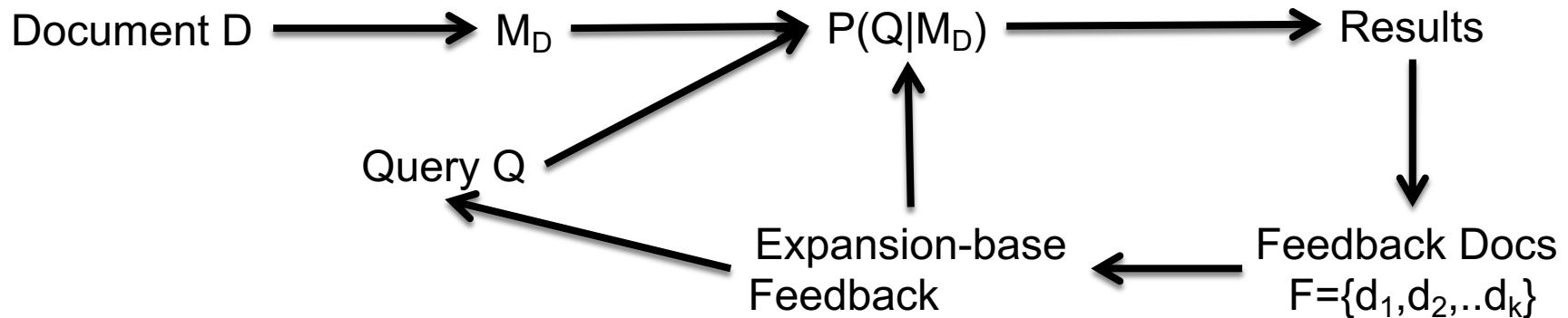
(-)

New query

4	6	3	7	0	.75
---	---	---	---	---	-----

# Relevance Feedback in Language Models

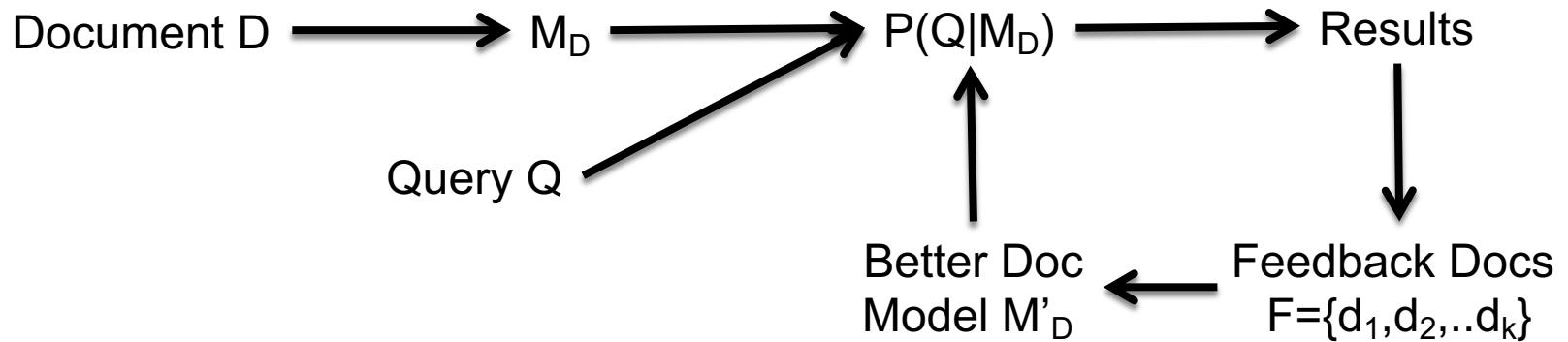
- Expansion-based Feedback in Query Likelihood LMs



- Select new query terms from the feedback documents
  - Such as using TF-IDF or BM25 to select top weight terms
  - So the new terms are  $\{nq_1, nq_2, \dots, nq_n\}$
- Apply new query terms into new query
  - Combine with original query  $Q' = \{q_1, q_2, \dots, q_m, nq_1, nq_2, \dots, nq_n\}$ , or
  - Simply viewed as the new query  $Q' = \{nq_1, nq_2, \dots, nq_n\}$

# Relevance Feedback in Language Models

- Model-Interpolation Feedback in Query likelihood LMs



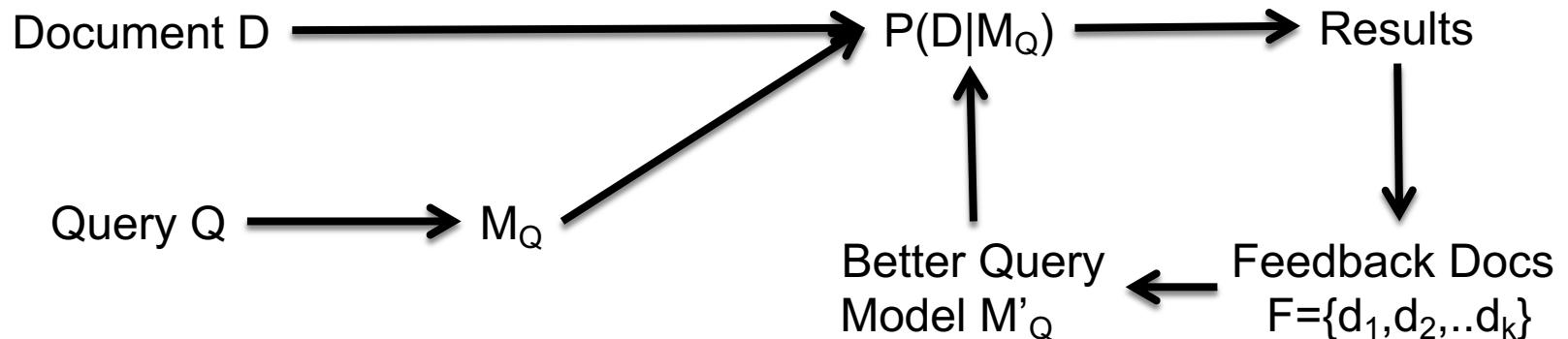
- Maybe hard to generate a query model, but can interpolate a better document model

$$P(Q|M'_D) = \alpha P(Q|M_D) + (1-\alpha) P(Q|F)$$

- We still can use JM smoothing or Dirochlet Prior smoothing for  $P(Q | M_D)$

# Relevance Feedback in Language Models

- Model-Interpolation Feedback in Document likelihood LMs



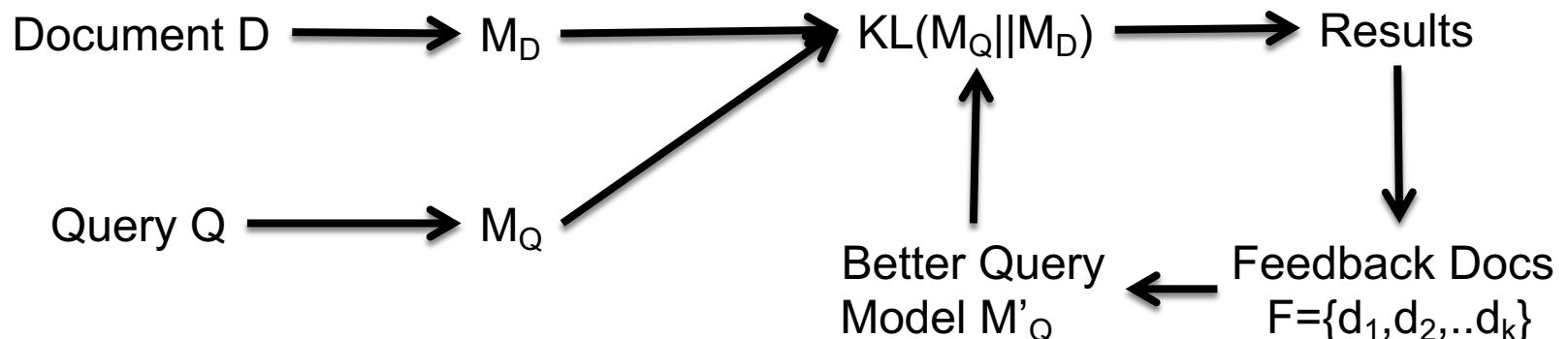
- Based on the feedback documents, we can estimate a better query model through interpolation

$$P(D|M'_Q) = \alpha P(D|M_Q) + (1-\alpha) P(D|F)$$

- We still can use JM smoothing or Dirochlet Prior smoothing for  $P(D | M_Q)$

# Relevance Feedback in Language Models

- Model-Interpolation Feedback in Model Comparison LMs



- Based on the feedback documents, we can estimate a better query model through interpolation

$$M'_Q = \alpha M_Q + (1-\alpha) M_F$$

- We still can use JM smoothing or Dirochlet Prior smoothing for both  $M_Q$ , and sometimes for  $M_F$

# What is Interactive IR

# A System Oriented View of IR



Query

Search Engine

Ranked List

harbin china

About 1,900,000 results (0.28 seconds)

Advanced search

Search

[Haerbin, Heilongjiang China](#) maps.google.com



[Harbin - Wikipedia, the free encyclopedia](#) ☆

Harbin had established its status as the center of northeastern China and as .... As a result, Harbin is China's base for the production of commodity grain ...  
History - Administrative Divisions - Climate - Economy  
[en.wikipedia.org/wiki/Harbin](#) - Cached - Similar

[Harbin Travel Guide: Harbin Hotel, Map, Heilongjiang China](#) ☆

Harbin (Capital of Heilongjiang Province) travel information about attractions, accommodation, transport, dining, shopping, climate and other tips of ...  
[www.travelchinaguide.com/cityguides/heilongjiang/harbin/](#) - Cached - Similar

[Harbin Ice Festival, bing xue jie, Heilongjiang](#) ☆

Harbin Ice Festival, established in 1985, is held annually from January 5 and lasts for over one month.  
[www.travelchinaguide.com/attraction/.../harbin/ice\\_snow.htm](#) - Cached - Similar

# Scenario



# Example

Jaguar



[jaguar.co.uk](http://jaguar.co.uk)



[jaguarlandrover.com](http://jaguarlandrover.com)



[wikipedia.org/Jaguar](http://wikipedia.org/Jaguar)



1 >

[bigcatrescue.org](http://bigcatrescue.org)



[fender.com/jaguar](http://fender.com/jaguar)

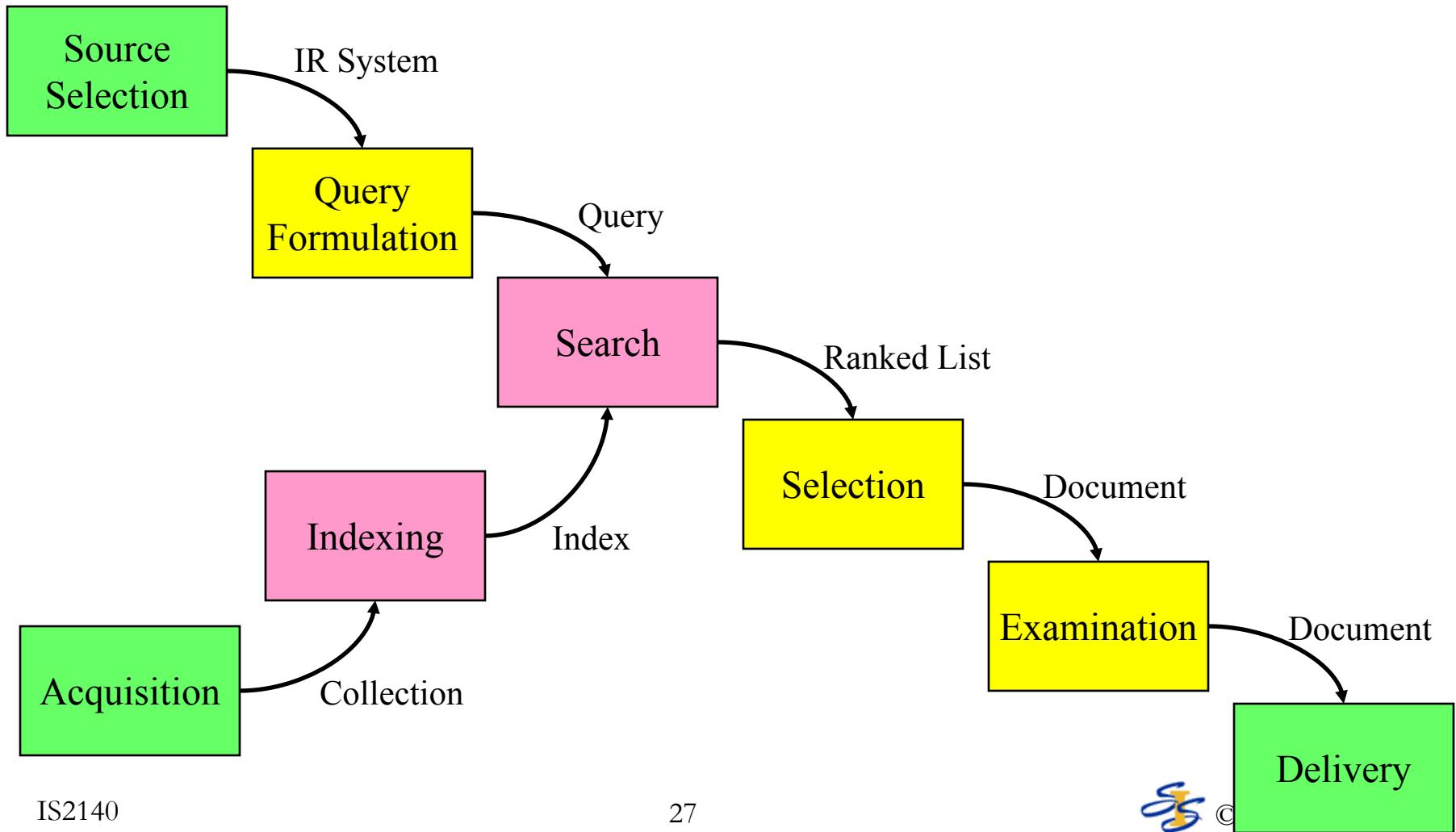


[wikipedia.org/Jaguar\\_car](http://wikipedia.org/Jaguar_car)



< 2 >

# Supporting the Search Process



# Interactive IR Systems

- “In interactive IR, users and information retrieval systems are the two partners. Users interact with information retrieval systems via *interfaces of these systems*.”



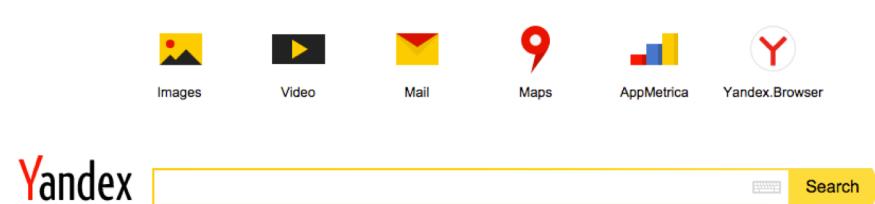
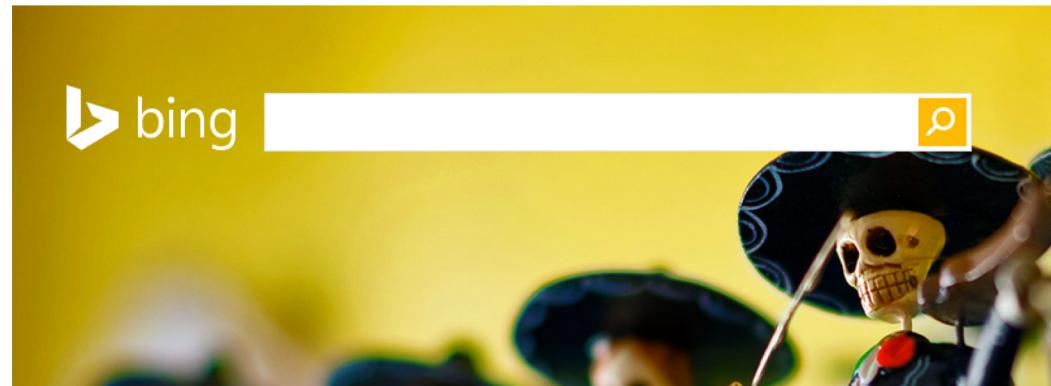
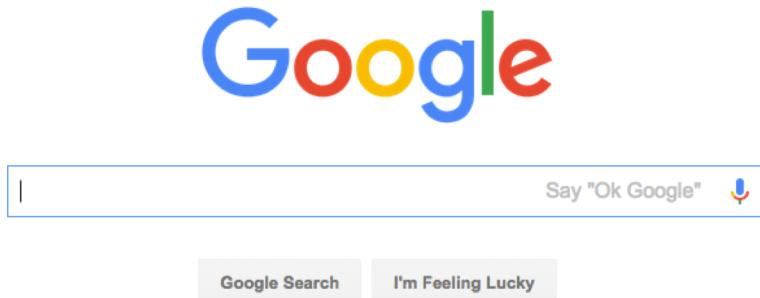
IS2140 Xie, H. (2007). "Shifts in information-seeking strategies in information retrieval in the digital age: Planned-situational model" *Information Research*, 12(4) paper colis22. [Available at <http://InformationR.net/ir/12-4/colis/colis22.html>]

# Design Interactive IR

# General Design Principles

- Pay attention to **usability**: how easy it is to use (5 components)
  - *Learnability*: How easy is it for users to accomplish basic tasks the first time they encounter the interface?
  - *Efficiency*: How quickly can users accomplish their tasks after they learn how to use the interface?
  - *Memorability*: After a period of non-use, how long does it take users to reestablish proficiency?
  - *Errors*: How many errors do users make, how severe are these errors, and how easy is it for users to recover from these errors?
  - *Satisfaction*: How pleasant or satisfying is it to use the interface?

# Modern Search Interface



[My Page](#) [\(edit\)](#)
[add / delete pages](#)
 Web  Images  News  Video

Search the Web:

[Search](#)
[Got Speedbar?](#)
[Yellow Pages](#) | [White Pages](#)
**Best Fares**

Expedia

Flight  Packages  
 Hotel  Cruise  
 Car  Activities

Shop with confidence:  
Best price guaranteed!

**Explore Excite**

- |                                   |                                      |                             |
|-----------------------------------|--------------------------------------|-----------------------------|
| <a href="#">Air/Travel</a>        | <a href="#">Entertainment</a>        | <a href="#">Money</a>       |
| <a href="#">Autos</a>             | <a href="#">Family</a>               | <a href="#">Movies</a>      |
| <a href="#">Beauty</a>            | <a href="#">Fashion</a>              | <a href="#">News</a>        |
| <a href="#">Careers</a>           | <a href="#">Games</a>                | <a href="#">Real Estate</a> |
| <a href="#">Casino</a>            | <a href="#">Gifts</a>                | <a href="#">Recipes</a>     |
| <a href="#">Celebrities</a>       | <a href="#">Health</a>               | <a href="#">Recreation</a>  |
| <a href="#">Computers</a>         | <a href="#">Holidays</a>             | <a href="#">Shopping</a>    |
| <a href="#">Crafts</a>            | <a href="#">Home &amp; Garden</a>    | <a href="#">Software</a>    |
| <a href="#">Credit Score</a>      | <a href="#">How To Video</a>         | <a href="#">Sports</a>      |
| <a href="#">Dating</a>            | <a href="#">Lifestyle</a>            | <a href="#">Technology</a>  |
| <a href="#">Local</a>             | <a href="#">Today</a>                |                             |
| <a href="#">Event Tickets</a> NEW | <a href="#">Online Education</a> NEW | <a href="#">Video</a>       |
| <a href="#">Home Security</a>     |                                      |                             |

**Best Fares**

Expedia

Flight  Packages  
 Hotel  Cruise  
 Car  Activities

Shop with confidence:  
Best price guaranteed!

**Weather**

Please enter your city or zip code:

Submit

**My Ebay**
**Business**

- [Bachelors in Business Management](#)
- [Accounting School](#)
- [Bachelors in Communications](#)
- [Business Administration Jobs](#)
- [Marketing Schools](#)
- [MBA in Entrepreneurship](#)
- [MBA in Finance](#)
- [Online Accounting Degree](#)
- [Online Marketing Degree](#)
- [International Business Jobs](#)
- [Risk Management Degree](#)
- [Supply Chain Management Degree](#)
- [Bachelors in Business Administration](#)
- [Masters in Human Resources](#)
- [Online Economics Degree](#)
- [Organizational Psychology Programs](#)
- [MBA Operations Management](#)
- [Bachelor Degree in Finance](#)

**Excite Today**

Add your favorite Excite Today content to this module!

[Click here to begin!](#)
**Email and Community**
[Email](#) - No new messages.

[Speedbar](#) - Download Now!

**My Games**

Click edit to select your favorite games!

**Daily Fortune**

Your fortune for today...

The light of a hundred stars cannot equal the light of the moon.

[Crystal Ball](#) | [Love Match](#)

[7 Key Stats You Should Not Ignore About Online Education](#)

[Getting A Degree With A 9 To 5 Job](#)

[How Can A New Qualification Help Further Your Career](#)
**Top Concert Tickets**

- [Concerts Tickets](#)
- [Broadway Tickets](#)
- [5 Seconds of Summer Tickets](#)
- [Adele Tickets](#)
- [Ariana Grande Tickets](#)

**My News**
[Court adjourns for week after jury sees video of shooting](#)

CHARLESTON, S.C. A Court has adjourned for the week in the trial of a white former South Carolina police officer charged with murder in the ...

**Top News**
[Court adjourns for week after jury sees video of shooting](#) November 5, 2:20 am (ET)

Screenshot on Nov.6, 2016



Google Search

I'm Feeling Lucky

# General Design Principles

- Which one do you prefer?
- Why?
- Reasons from Hearst's “Search User Interface” book
  - Users in search are usually engaged in some larger task, and do not want their flow of thought interrupted by an intrusive interface.
  - Search is a mentally intensive task. It is not possible to read and to think about something else at the same time. Thus, the fewer distractions while reading, the more usable the interface.
  - The interface design must be understandable and appealing to a wide variety of users of all ages, cultures and backgrounds, applied to an enormous variety of information needs. **Everyone can use it!**

# General Design Principles

- design desiderata for search user interfaces
  - Offer informative feedback.
  - Support user control.
  - Reduce short-term memory load.
  - Provide shortcuts for skilled users.
  - Reduce errors; offer simple error handling.
  - Strive for consistency.
  - Permit easy reversal of actions.
  - Design for closure: user has a sense when the search is complete

<http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>

# General Design Principles

- Offer efficient and informative feedback
  - Show search results immediately
  - Show informative document surrogates; highlight query terms
  - Allow sorting of results by various criteria
  - Show query terms suggestions
  - Provide simple and meaningful history mechanism
  - Integrated search and navigation
- More on this topic can be found at  
[http://searchuserinterfaces.com/book/sui\\_ch1\\_design.html](http://searchuserinterfaces.com/book/sui_ch1_design.html)

# Interactive IR System

- Typically, the Search Engine Result Pages (SERP)

The image shows a search engine interface with a red arrow pointing downwards from the 'Ranked List' label to the search results.

**Query:** information retrieval

**Ranked List:**

- Document surrogate:** Information retrieval - Wikipedia, the free encyclopedia  
[en.wikipedia.org/wiki/Information\\_retrieval](https://en.wikipedia.org/wiki/Information_retrieval) ▾ Wikipedia  
Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be ...  
Information retrieval applications - Category:Information retrieval - Relevance
- Introduction to Information Retrieval**  
[informationretrieval.org/](http://informationretrieval.org/) ▾  
The book aims to provide a modern approach to information retrieval from a computer science perspective. It is based on a course we have been teaching in ...
- Information Retrieval - School of Computing Science - University of ...**  
[www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html) ▾ University of Glasgow  
This chapter has been included because I think this is one of the most interesting and active areas of research in information retrieval. There are still many ...
- Information Retrieval - Springer**  
[link.springer.com/journal/10791](http://link.springer.com/journal/10791) ▾ Springer Science+Busine... ▾  
Subscription e-journal dedicated to theory and experimentation in information retrieval. Sample copy available.
- Information Retrieval – incl. option to publish open access - Springer**  
[www.springer.com/.../information+retrieval/.../.../...](http://www.springer.com/.../information+retrieval/.../...) ▾ Springer Science+Busine... ▾  
The journal provides an international forum for the publication of theory, algorithms, and experiments across the broad area of information retrieval. Topics of ...
- 11-741: Information Retrieval**  
[boston.iti.cs.cmu.edu/classes/11-741/](http://boston.iti.cs.cmu.edu/classes/11-741/) ▾  
30+ items - The Information Retrieval core components of the course ...  
2 1/16. Introduction to ad-hoc search: Boolean retrieval (pdf). Ch 1.  
6 1/30. Information needs and queries (pdf1, pdf2). Ch 19.4.

IS2140

 © Daqing He

# Interactive IR System

- IIR system may be sophisticated in complex search scenarios

Query

information retrieval

Search

Choose: find experts for reviewing

Welcome hanshuguang Log out

Time cosuming: 0.609 seconds.

Research Communities	Apply
<input checked="" type="checkbox"/> SIGIR;CIKM;ECIR	(453)
<input checked="" type="checkbox"/> WWW;CIKM;WSDM	(62)
<input checked="" type="checkbox"/> JCDL;CIKM;DL	(50)
<input checked="" type="checkbox"/> MM;MIR;	(36)
<input checked="" type="checkbox"/> SIGDOC;ICAIL;CIVR	(35)
<input checked="" type="checkbox"/> SIGMOD;VLDB;DBTEST	(31)
<input checked="" type="checkbox"/> KDD;ICML;CIKM	(30)
<input checked="" type="checkbox"/> GIS;MDM;CIKM	(21)

All Communities

Affiliations	Apply
<input checked="" type="checkbox"/> Microsoft Research	(18)
<input checked="" type="checkbox"/> IBM Research Center for Software ..	(16)
<input checked="" type="checkbox"/> University of Massachusetts, Amher..	(12)
<input checked="" type="checkbox"/> University of Glasgow, Glasgow, S..	(11)
<input checked="" type="checkbox"/> Microsoft Research Asia, Beijing...	(10)
<input checked="" type="checkbox"/> Dublin City University, Dublin, I..	(10)
<input checked="" type="checkbox"/> U of Massachusetts Amherst, Amher..	(10)
<input checked="" type="checkbox"/> University of Illinois at Urbana-C..	(10)

All Affiliations

Seniority	Apply
<input checked="" type="checkbox"/> 1 - 2 Years	(347)

Navigation



# Interactive IR System

- IIR system may be sophisticated in complex search scenarios

The screenshot displays the Interactive IR System interface, divided into several sections:

- Chat**: A sidebar on the left containing a "Chat Box" with a history of messages between users DT0101 and DT0102.
- Topic Statement**: A search bar with the query "history of olympics" and a "Search" button.
- Web Search**: Results for the query "history of olympics", including links to "Ancient Olympic Games First Olympics in Olympia | Greek History" and "Olympic Games - Wikipedia, the free encyclopedia".
- Team Workspace**: A large workspace area titled "Query" containing a "Search History" panel with items like "history of olympics" and "olympic games". It also shows a list of search results from the "Collaborative Search System >> Training task" for the query "history of olympics". These results include links to "History of the Olympics - Creating the Modern Olympic Games", "News for history of olympics", "Origin and History of the Olympic Games | Go for the Gold Stud", "The Olympic Games — History.com Articles, Video, Pictures and", "The Olympics - EnchantedLearning.com", "Olympics Timeline: Ancient Greece - 1940s | Infoplease.com", and "databaseOlympics.com - Olympic results for all Winter and Summer...". Each result has a "View" dropdown, a "Comment" button, and a "Delete" button.
- Workspace**: A large area on the right labeled "Workspace" containing the same search results and history as the Team Workspace section.

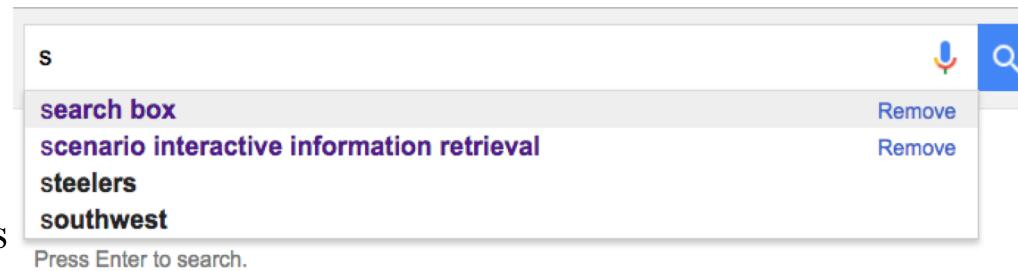
# Query Formulation and Reformulation

# Query Formulation

- User's goals
  - Identify the right query for the current need
    - conscious/formalized need => compromised need
- How can the user achieve this goal?
  - Infer the right query terms
  - Infer the right composition of terms
- More recently ...
  - Spoken queries
  - Queries from Scanning / taking pictures
  - Search by drawing
  - etc.

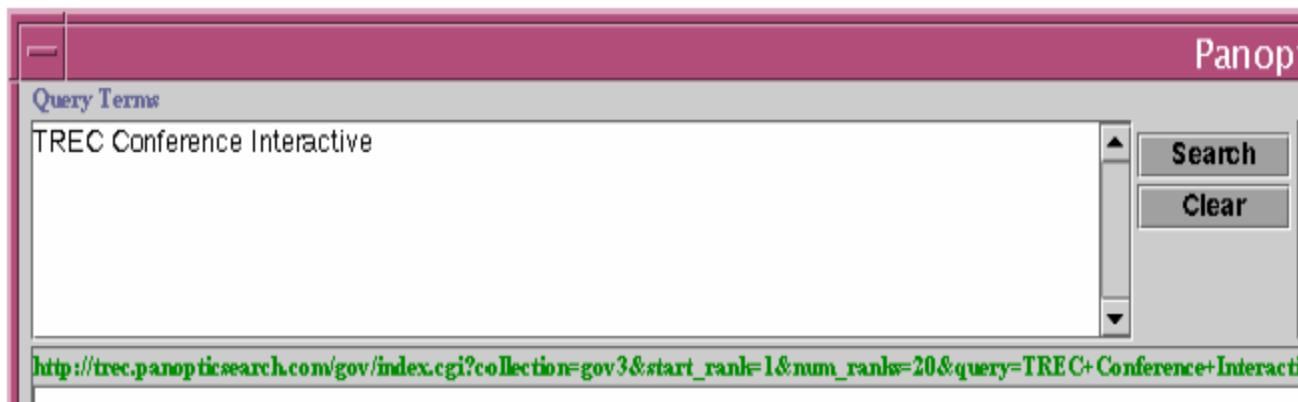
# Query Formulation -II

- What IR system can do?
  - Ask more from the user
    - Encourage long/complex queries
      - Provide a large text entry area
      - Use forms filling or direct manipulation
    - Obtain contextual information from users
      - Location
      - Date and time
      - Devices...
  - Infer from relevant items
    - Infer from previous queries, term auto suggestions
      - <http://www.google.com/>
    - Infer from previous retrieved documents

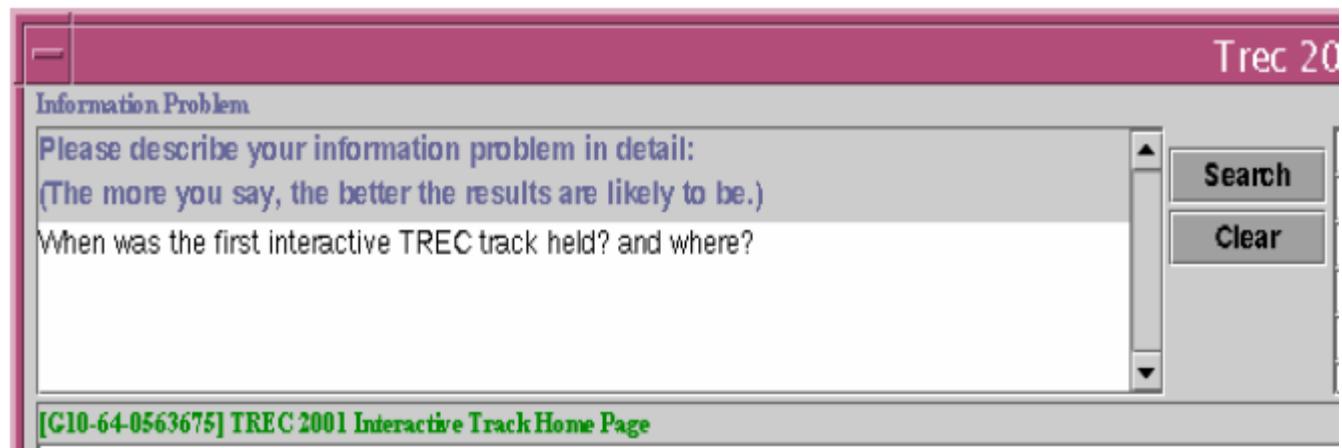


# leading longer queries

- 4.19 words on average in the top versus 6.02 words



A search interface which asks searchers to describe their information problems at length will lead to longer queries than one which asks searchers to simply input a query as a list of words or phrases



# leading longer queries



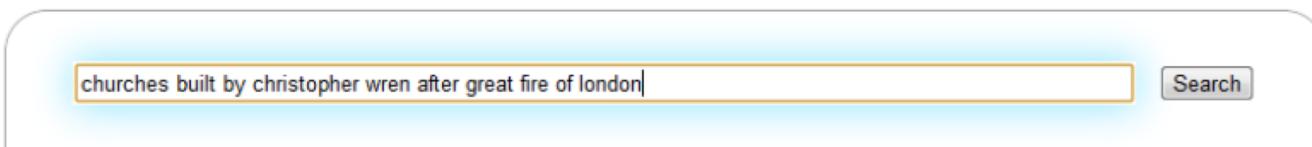
**Figure 1. An empty query box has a pink halo.**



**Figure 2. As the person starts to type, red hue starts to fade.**



**Figure 3. As the query gets longer, the halo becomes progressively bluer.**



**Figure 4. Long queries are displayed with a blue halo.**

Agapie, E., Golovchinsky, G., & Qvarfordt, P. (2013, April). Leading people to longer queries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3019-3022). ACM.

# Form-Based Query Formulation (Google)

**Google** [Advanced Search](#) [Advanced Search Tips](#) 

Use the form below and your advanced search will appear here

**Find web pages that have...**

all these words:

this exact wording or phrase:  [tip](#)

one or more of these words:  OR  OR  [tip](#)

**But don't show pages that have...**

any of these unwanted words:  [tip](#)

**Need more tools?**

Results per page:  

Language:  

File type:  

Search within a site or domain:   
(e.g. youtube.com, .edu)

[⊕ Date, usage rights, numeric range, and more](#)

[Advanced Search](#)

## Topic-specific search engines from Google:

[Google Book Search](#)  
[Google Code Search](#) New!  
[Google Scholar](#)  
[Google News archive search](#)

[Apple Macintosh](#)  
[BSD Unix](#)  
[Linux](#)  
[Microsoft](#)

[U.S. Government](#)  
[Universities](#)

# Highlight suggested query terms

[Barack Obama](#)  
www.barackobama.com/ ▾ Barack Obama ▾  
Official re-election campaign website of President **Barack Obama** provides the latest updates, election news, videos, local events and ways to volunteer and ...

[Barack Obama - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/Barack\_Obama ▾ Wikipedia ▾  
His father, **Barack Obama**, Sr., was a Luo from Nyang'oma Kogelo, Kenya. Obama's parents met in 1960 in a Russian class at the University of Hawai'i at ...

Google barack obama |

barack obama facts  
barack obama twitter  
barack obama wiki  
barack obama quotes

About 672,000,000 results (0.22 seconds)

Search a location

Burnham Park, Baguio City  
Burnham Suites Hotel, Bag  
Burnham Hotel, F. Calderon  
Burnaby, BC, Canada  
Metrobank - Baguio Burnha

powered by Google

# Query Reformulation

- Historical queries
- Related queries

Searches related to **pittsburgh**

[pittsburgh basketball](#)

[pittsburgh pirates](#)

[pittsburgh university](#)

[pittsburgh post gazette](#)

[pittsburgh steelers](#)

[pittsburgh airport](#)

[pittsburgh penguins](#)

[pittsburgh travel](#)

# Query Reformulation

- Providing facets. E.g., Google Scholar, LinkedIn

The image shows two search interfaces side-by-side. On the left is Google Scholar, and on the right is LinkedIn.

**Google Scholar:** The search term "information systems" is entered. The results page shows a sidebar with facets: "Articles", "Case law", "My library", "Any time" (with options for "Since 2014", "Since 2013", "Since 2010", and "Custom range..."), "Sort by relevance", "Sort by date", and checkboxes for "include patents" and "include citations". A red dashed box highlights the sidebar area.

**LinkedIn:** The search term "software engineering" is entered. The results page includes a sidebar with facets: "All" (People, Jobs, Companies, Groups, Universities, Posts, Inbox), "Location" (All, United States, San Francisco Bay..., Washington D.C. Me..., Greater New York Ci..., Greater Seattle Area), "Company" (All, EMC, KP Recruiting Group, Pivotal Software, Inc., Adobe, MedeAnalytics Inc.), and a "Filter" button. A red dashed box highlights the sidebar area.

**Google Search Results:** The search term "information systems" is entered. The results include a snippet from a book by J Castro, M Kolp, and J Mylopoulos, followed by a snippet from a book by Z Pawlak. Below these are sections for "Information systems theoretical foundations" and "Books Information systems developer".

**LinkedIn Results:** The search term "software engineering" is entered. The results show 101,586 results. It lists several job posts for "Software Engineering Intern" (Summer 2015) and "Software Engineer, PhD University Graduate". It also shows sections for "5,042,315 People results for software engineering" (listing profiles like Pavan Reddy, Huai-Yuan Hsu, Constantine Aaron Cois) and "1,260 Groups results for software engineering" (listing groups like "Computer & Software Engineering Professionals" and "Information Technology: Software, Hardware, Computer, De...").

# Ask's Query Refine Suggestions



information retrieval



Advanced

## Narrow Your Search

Information Retrieval System  
Information Retrieval Methods  
Information Retrieval Tutorial  
Modern Information Retrieval  
Computers Information Retrieval  
Evaluation of Retrieval Systems  
Information Retrieval Papers  
Information Retrieval Theory  
Multimedia Information Retrieval  
Information Retrieval Software

More »

## Expand Your Search

Natural Language Processing  
Gigabyte Definition  
Teoma

## information retrieval

Showing 1-10 of 1,662,000

### Dog Weight Information

Sponsored Results

Weight is an Important Part of Your Dog's Health - Don't Ignore it.

[www.StopCanineObesity.com](http://www.StopCanineObesity.com)

### EMC Backup & Recovery

Lower Costs w/EMC Next-Gen Recovery Solutions. Free Taneja Group Paper

[www.emc.com](http://www.emc.com)

### Retrieve History

Top Rated Data Recovery Software. Free Download. 100% Guaranteed.

[Data-Recovery-Professional.com](http://Data-Recovery-Professional.com)

### SIGIR: Information Retrieval

QUOTATION Addresses issues ranging from theory to user demands in the application of computers to the acquisition, organization, storage, ...

[www.acm.org/sigir/](http://www.acm.org/sigir/) · Cached

### Introduction to Information Retrieval

Introduction to Information Retrieval: Table of Contents ... 11

Probabilistic information retrieval pdf slides ...

[www-csli.stanford.edu/~schuetze/information-retrieval-b...](http://www-csli.stanford.edu/~schuetze/information-retrieval-b...) · Cached

### Information Retrieval

Information Retrieval Group, University of Glasgow ... probabilistic retrieval. This chapter has been included because I think this is one ...

[www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html) · Cached

### Information Retrieval Research

An up-to-date overview of research in the field of information retrieval. ...

Modern Information Retrieval Ricardo Baeza-Yates & Berthier ...

[www.searchtools.com/info/info-retrieval.html](http://www.searchtools.com/info/info-retrieval.html) · Cached

### Information Retrieval Links

## Synonyms

### Synonyms of 'retrieval'

3 synonyms - Roget's II: Thesaurus

- The act of getting back or regaining: recoup, recovery, repossession.

## Dictionary

### Definitions of 'retrieval'

(rī-trēvəl) (n.) - 3 definitions

The American Heritage® Dictionary

retrieval (n.) The act or process of retrieving.

retrieval (n.) The process of accessing information from memory or other storage devices.

## W Encyclopedia



### Information retrieval

Information retrieval (IR) is the science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational stand-alone databases or hypertextually-networked databases...

More »

# Visualizing Related Terms

- <http://www.quintura.com/>

The screenshot shows the Quintura search interface. At the top, there's a search bar with 'pittsburgh' typed in, with tabs for 'Web', 'Images', and 'Video'. Below the search bar are navigation icons for home, back, forward, embed, and share. The main content area displays a network of related terms centered around 'greater pittsburgh area'. These terms include 'convention', 'pa', 'federal', 'international chamber', 'in pittsburgh', 'police', 'the pittsburgh arts council', 'region', 'chapter', 'community', 'automobile', 'detector', 'plumbing', 'postal', 'properties', and 'hotel'. To the right of this network, a list of six search results is shown:

1. [Pittsburgh](http://greaterpittsburgh.com/)  
The Greater Pittsburgh Region's Greatest Web Site  
<http://greaterpittsburgh.com/> - 15Kb
2. [Visit Pittsburgh](http://www.visitpittsburgh.com/)  
The Greater Pittsburgh Convention and Visitors  
hotels, activities, attractions, restaurants, and events  
<http://www.visitpittsburgh.com/> - 70Kb
3. [Greater Pittsburgh Chamber of Commerce](http://www.pittsburghchamber.com/)  
Business and professional people working together  
everyone.  
<http://www.pittsburghchamber.com/> - 22Kb
4. [Pittsburgh metropolitan area - Wikipedia, the free encyclopedia](http://en.wikipedia.org/...Pittsburgh_metropolitan_area)  
It has been suggested that this article or section be merged into Metropolitan Statistical Area (also called Greater Pittsburgh).  
[http://en.wikipedia.org/...Pittsburgh\\_metropolitan\\_area](http://en.wikipedia.org/...Pittsburgh_metropolitan_area)
5. [Greater Pittsburgh Federal Credit Union](http://www.greaterpittsburghfcu.org/)  
You are invited to attend the 54th Annual Members  
Union. Please join us at the Pittsburgh Athletic ...  
<http://www.greaterpittsburghfcu.org/> - 17Kb
6. [Greater Pittsburgh Plumbing](http://www.greaterpittsburghplumbing.com/)  
Greater Pittsburgh Plumbing full service plumber  
service, emergency service, free estimate, free service  
<http://www.greaterpittsburghplumbing.com/> - 19Kb

# Search Result Presentation

# Interactive IR System

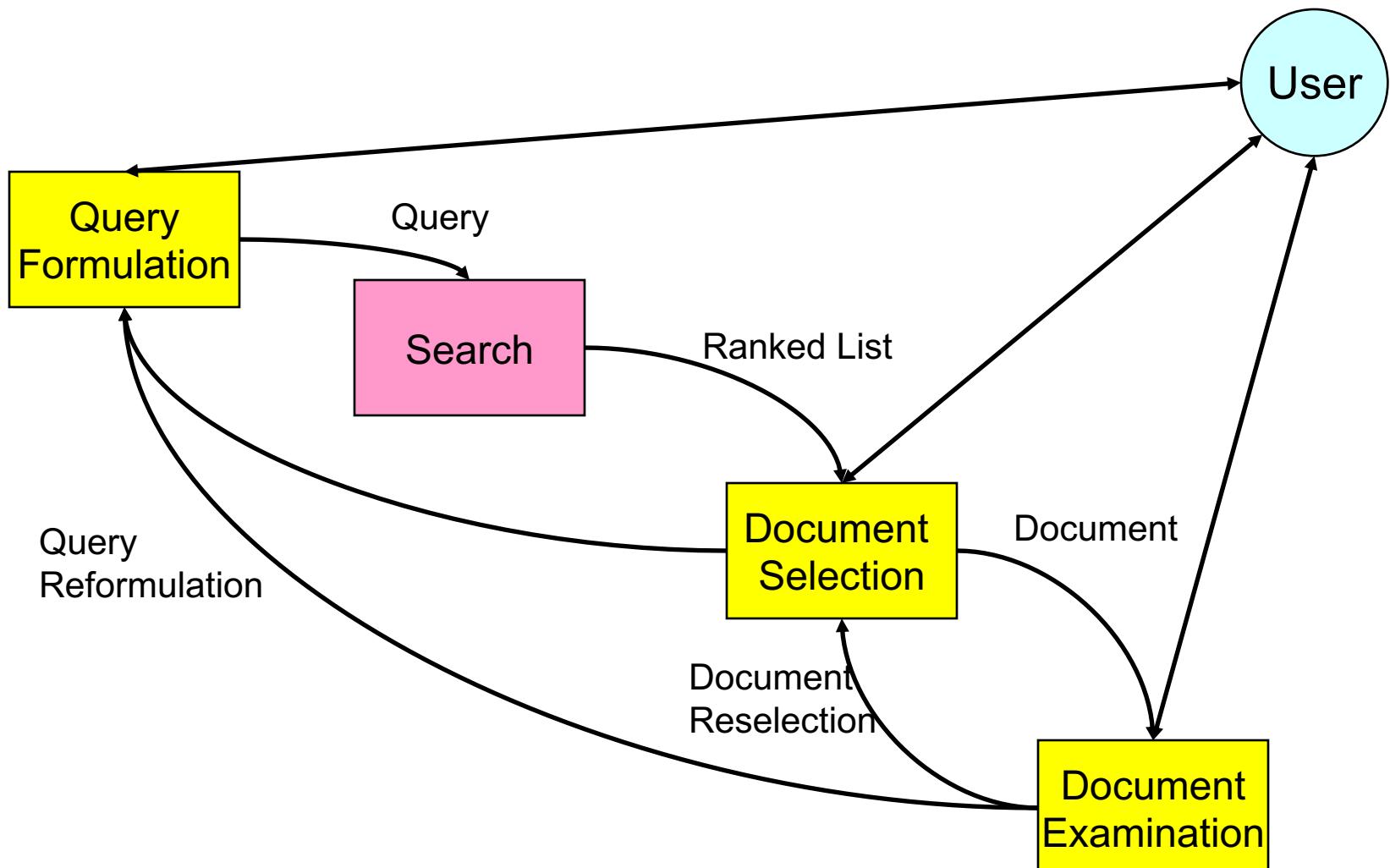
The screenshot shows a search interface with a red dashed box highlighting the search bar containing 'information retrieval'. Below the search bar is a navigation menu with 'Web' selected, followed by Books, Images, Videos, News, More, and Search tools. The main area is titled 'Ranked List' and displays approximately 16,300,000 results found in 0.27 seconds. The first result is a link to 'Information retrieval - Wikipedia, the free encyclopedia' from en.wikipedia.org. The snippet for this result describes information retrieval as the activity of obtaining information resources relevant to an information need from a collection of information resources. Subsequent results include links to 'Introduction to Information Retrieval' (informationretrieval.org), 'Information Retrieval - School of Computing Science - University of ...' (www.dcs.gla.ac.uk/Keith/Preface.html), 'Information Retrieval - Springer' (link.springer.com/journal/10791), and 'Information Retrieval – incl. option to publish open access - Springer' (www.springer.com/...information+retrieval/...). The bottom section shows course material for '11-741: Information Retrieval' from boston.iti.cs.cmu.edu/classes/11-741/.

Query

## Document surrogate



# Search Results Presentation



# Post-Search Analysis

- Depends on the types of information needs
  - Simple cases: identification of the relevant items
  - Complicated cases: exploration, learning and investigation
- Post-search analysis types: looking for information to explore
  - Trends
  - Comparisons
  - Aggregation and Scaling
  - Identifying a Critical Subset
  - Assessing
  - Interpreting
  - others: cross-reference, summarize, find evocative visualizations ...

# Search Results Presentation

- User's goals:
  - Identify docs for use => completing the search tasks
  - Identify new query terms/combinations => continue the search
    - Useful terms for future queries
- System goals:
  - minimize time/effort for deciding which docs to examine
  - Assist user to identify relevant documents
  - Assist user to identify potential useful terms
    - for clarifying the right information need
    - for generating better queries

# Solutions

- Search Interfaces should
  - Show the roles of the query terms in the retrieved documents
  - give hints about the roles certain terms play in the collection
  - give hints about what will happen if various terms are combined
  - show explicitly why documents are retrieved in response to the query
  - summarize compactly the subset of interest
- Technologies
  - Summarization: generate document surrogates
  - Presentation: Textual vs graphical, One dimensional vs two dimensional vs three dimensional

# Document Surrogate

- Design of document surrogate influences user clicks
  - Where possible, all the query terms should appear in the surrogate, reflecting their relationship to the corresponding Web page
  - When the query terms are present in the title, they need not appear in the summary
  - Length and complexity of URLs should be reduced, and URLs should be selected and displayed in a manner that emphasizes their relationship to the query.

Clarke, C. L., Agichtein, E., Dumais, S., & White, R. W. (2007). The influence of caption features on clickthrough patterns in web search. In *SIGIR* 207:135-142

# Google's KWIC Document Surrogates

Google™ information retrieval  Advanced Search Preferences  
 Search the Web  Search Chinese (Simplified) and Chinese (Traditional) and English pages

Web Books Scholar Results 1 - 30 of about 9,420,000 Chinese (Simplified) and Chinese (Traditional) and English pages for [information retrieval](#). (0.24 se)

Try your search on [Yahoo](#), [Ask](#), [AllTheWeb](#), [Live](#), [Lycos](#), [Technorati](#), [Feedster](#), [Wikipedia](#), [Bloglines](#), [Altavista](#), [A9](#)



## [Information Retrieval: Uncertainty and Logics ...](#)

by Cornelis Joost van Rijsbergen - 1998 - 358 pages  
[books.google.com](#) - [About this book](#) - [More book results »](#)

## [Information retrieval - Wikipedia, the free encyclopedia](#)

Information retrieval (IR) is the science of searching for **information** in documents, searching for documents themselves, searching for metadata which ...  
[en.wikipedia.org/wiki/Information\\_retrieval](http://en.wikipedia.org/wiki/Information_retrieval) - 68k - [Cached](#) - [Similar pages](#) - [Filter](#)

## [Category:Information retrieval - Wikipedia, the free encyclopedia](#)

From Wikipedia, the free encyclopedia. Jump to: navigation, search. The main article for this category is **Information retrieval**. ...  
[en.wikipedia.org/wiki/Category:Information\\_retrieval](http://en.wikipedia.org/wiki/Category:Information_retrieval) - 24k - [Cached](#) - [Similar pages](#) - [Filter](#)

## [Information Retrieval](#)

An online book by CJ van Rijsbergen, University of Glasgow.  
[www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html) - 7k - [Cached](#) - [Similar pages](#) - [Filter](#)

## [Information Retrieval](#)

Online text of a book by Dr. CJ van Rijsbergen of the University of Glasgow covering advanced topics in **information retrieval**.  
[www.dcs.gla.ac.uk/~iain/keith/](http://www.dcs.gla.ac.uk/~iain/keith/) - 5k - [Cached](#) - [Similar pages](#) - [Filter](#)

## [Introduction to Information Retrieval](#)

The book aims to provide a modern approach to **information retrieval** from a computer science perspective. It is based on a course we have been teaching in ...  
[www-csli.stanford.edu/~hinrich/information-retrieval-book.html](http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html) - 9k -

### Sponsored Links

#### [Information Retrieval](#)

Search your intranet, file shares & more with solutions from Google  
[www.google.com/enterprise](http://www.google.com/enterprise)

#### [EMC Backup & Recovery](#)

Lower Costs w/EMC Next-Gen Recovery Solutions. Free Taneja Group Paper  
[www.emc.com](http://www.emc.com)

#### [Info. Retrieval Schools](#)

Find Degrees and School Programs in **Information Retrieval** Tech  
[www.Degree-Finder.com](http://www.Degree-Finder.com)

#### [Information Retrieval](#)

Find **Information Retrieval** in Local Directory!  
[www.usdirectory.com](http://www.usdirectory.com)

# Document Surrogate

- Design of document surrogate influences user clicks
  - My own experience at **Baidu** (2009)



[左宗棠鸡 百度百科](#)

湘菜名菜，主料为鸡，为美国人认知中最著名的中国菜之一。以湖南名  
将左宗棠的名字命名，但其实与他无关。

[菜品介绍](#) [做法](#) [营养价值](#) [食用指南](#)

[baike.baidu.com/](http://baike.baidu.com/) 2013-12-26 ▾

[左宗棠鸡 百度百科](#)



湘菜名菜，主料为鸡，为美国人认知中最著名的中国菜之一。以湖南名  
将左宗棠的名字命名，但其实与他无关。

[菜品介绍](#) [做法](#) [营养价值](#) [食用指南](#)

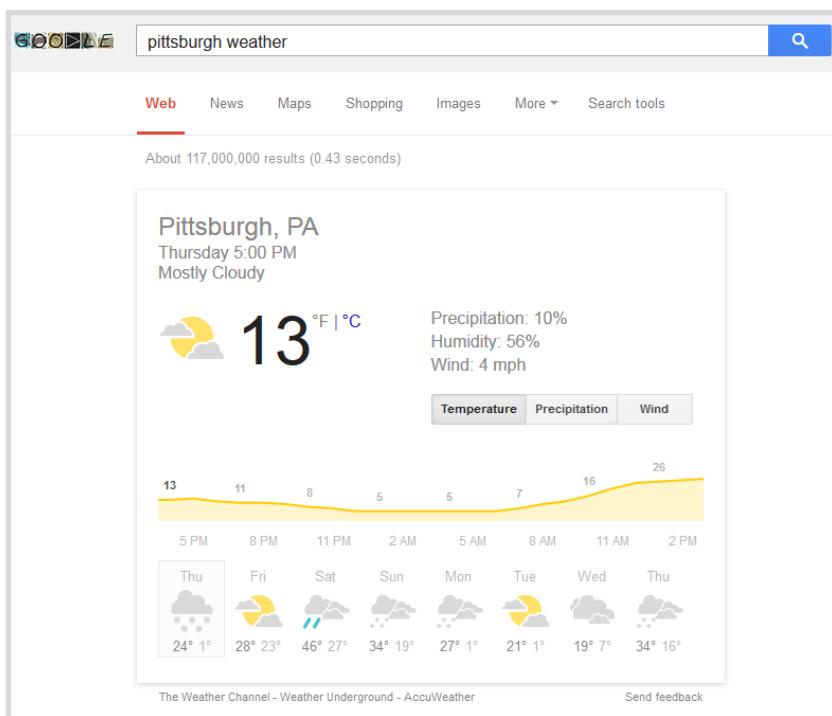
[baike.baidu.com/](http://baike.baidu.com/) 2013-12-26 ▾

↑ +10% clicks

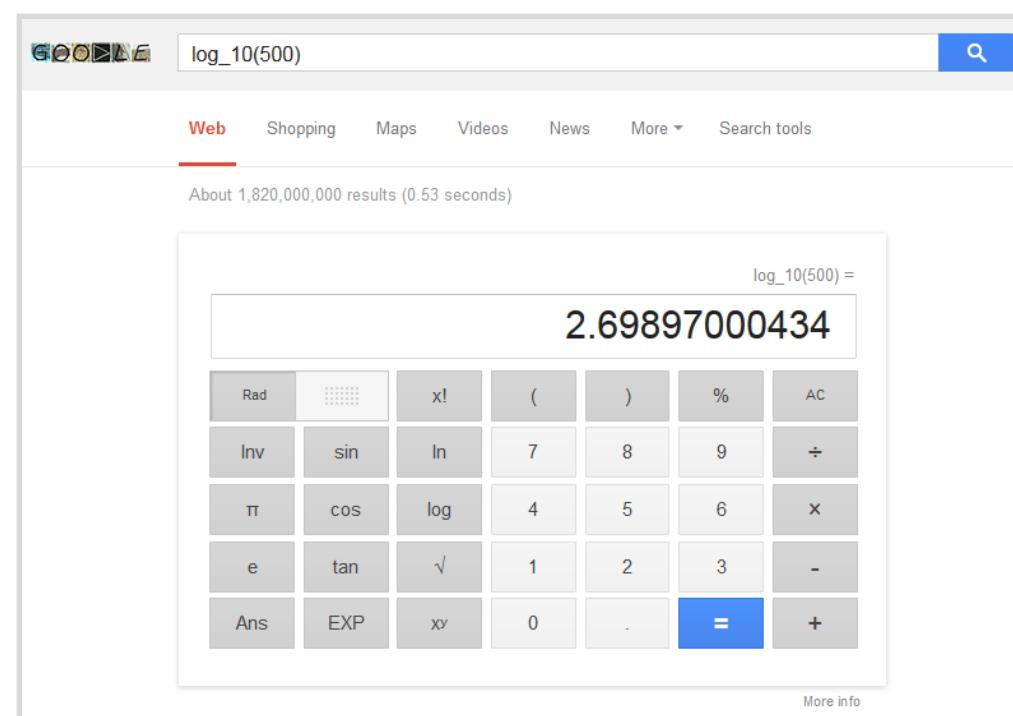
# Document Surrogate

- The end goal isn't necessarily the document set

Pittsburgh Weather



lg 500



# Taxonomy for Result Presentation

- One dimensional lists
  - Content: title, source, date, summary, ratings, ...
  - Order: retrieval status value, date, alphabetic, ...
  - Size: scrolling, specified number, RSV threshold
- Two dimensional displays
  - Construction: clustering, starfields, projection
  - Navigation: jump, pan, zoom
- Three dimensional displays
  - Contour maps, fishtank VR, immersive VR

# Results Presented as a Ranked list

Google™   Advanced Search Preferences  
 Search the Web  Search Chinese (Simplified) and Chinese (Traditional) a

Web Books Scholar Results 1 - 30 of about 9,420,000 Chinese (Simplified) and Chinese (Traditional) a  
Try your search on [Yahoo](#), [Ask](#), [AllTheWeb](#), [Live](#), [Lycos](#), [Technorati](#), [Feedster](#), [Wikipedia](#), [Bloglines](#), [Altavista](#)

 [Information Retrieval: Uncertainty and Logics ...](#)  
by Cornelis Joost van Rijsbergen - 1998 - 358 pages  
[books.google.com](#) - [About this book](#) - [More book results »](#)

[Information retrieval - Wikipedia, the free encyclopedia](#)  
Information retrieval (IR) is the science of searching for information in documents, searching for documents themselves, searching for metadata which ...  
[en.wikipedia.org/wiki/Information\\_retrieval](http://en.wikipedia.org/wiki/Information_retrieval) - 68k - [Cached](#) - [Similar pages](#) - [Filter](#)

[Category:Information retrieval - Wikipedia, the free encyclopedia](#)  
From Wikipedia, the free encyclopedia. Jump to: navigation, search. The main article for this category is **Information retrieval**. ...  
[en.wikipedia.org/wiki/Category:Information\\_retrieval](http://en.wikipedia.org/wiki/Category:Information_retrieval) - 24k - [Cached](#) - [Similar pages](#) - [Filter](#)

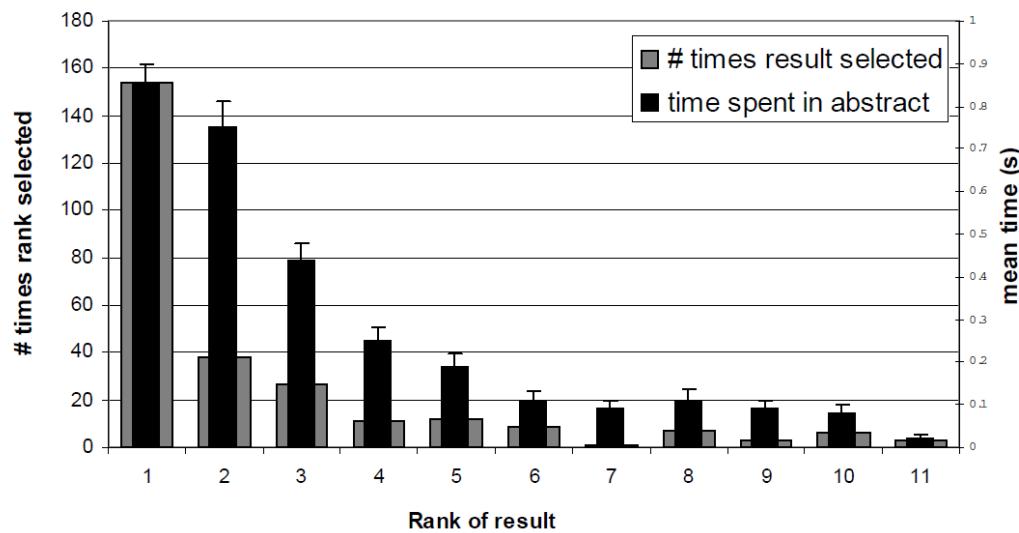
[Information Retrieval](#)  
An online book by CJ van Rijsbergen, University of Glasgow.  
[www.dcs.gla.ac.uk/Keith/Preface.html](http://www.dcs.gla.ac.uk/Keith/Preface.html) - 7k - [Cached](#) - [Similar pages](#) - [Filter](#)

[Information Retrieval](#)  
Online text of a book by Dr. CJ van Rijsbergen of the University of Glasgow covering advanced topics in **information retrieval**.  
[www.dcs.gla.ac.uk/~iain/keith/](http://www.dcs.gla.ac.uk/~iain/keith/) - 5k - [Cached](#) - [Similar pages](#) - [Filter](#)

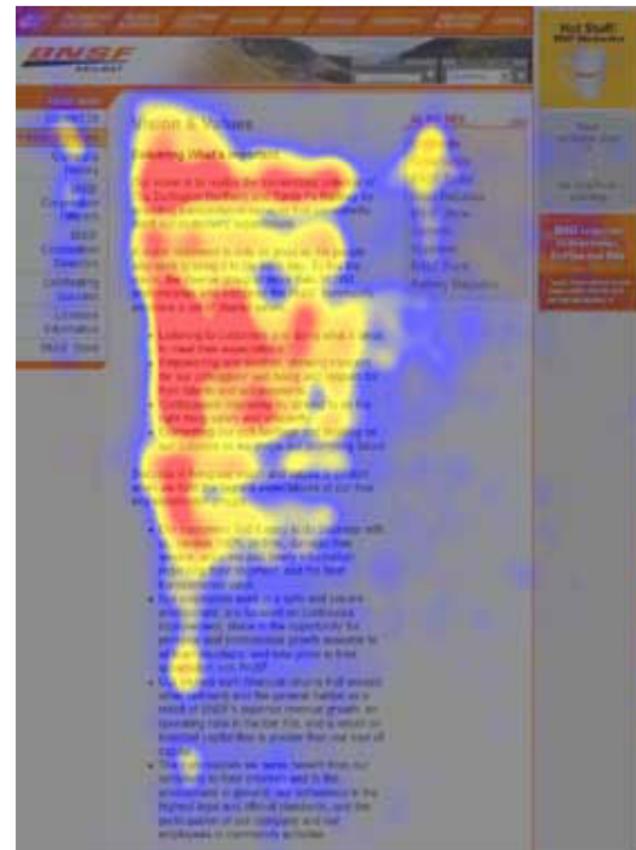
[Introduction to Information Retrieval](#)  
The book aims to provide a modern approach to **information retrieval** from a computer science perspective. It is based on a course we have been teaching in ...  
[www-csli.stanford.edu/~hlinrich/information-retrieval-book.html](http://www-csli.stanford.edu/~hlinrich/information-retrieval-book.html) - 9k - [Cached](#) - [Similar pages](#) - [Filter](#)

# Ranked List

- Ranked Lists of search result representation biased on top results



Laura A. Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking analysis of user behavior in WWW search. In SIGIR '04. ACM, New York, NY, USA, 478-479.



<http://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/>

# Blend Results and Media Types

- Web search is not just search for Web pages anymore
  - Information exist in many different sources and media types
  - Search engines increasingly automatically collect them and display together
    - Example: search “pittsburgh” in google, bing, yahoo,
    - Especially in [www.hakai.com](http://www.hakai.com)

wuhan

search

## ▲ Web

Wuh  
The c  
Wuh  
http://

Wuh  
Havin  
city is  
http://

Intro  
Wuh  
Gorge  
http://

Wuh  
() is the  
the eas  
out of t  
"thorou  
expres:  
approxi  
was the  
shek, n

## ▲ Gallery

## ▲ Credible

Wuhar

The me  
Towns  
two). Tl  
three p  
bridges  
and hill

Wuh  
bad r  
seriou  
http://

Wuhar  
() is the  
the eas  
out of t  
"thorou  
expres:  
approxi  
was the  
shek, n

## ▲ Pubmed

[Stud  
2007  
He Q  
Depa

The r  
3.1%  
Wuh  
hike  
CRU

Foxc  
Foxc  
Wuh  
abou  
incre  
DigiT

[Stud  
Wuh  
2004  
Li Q,  
Institu  
OBJE  
admin  
METI

Braz  
The I  
Wuh  
Com  
news  
Stee

YALL

## ▲ News

Sort

## ▲ Twitte



## ▲ Video

## Wuhan China Cymbals



## wuhan s series splashes 8,10,12..



## 湖北武汉旅游 Travel China Tours.



## Wuhan 武汉



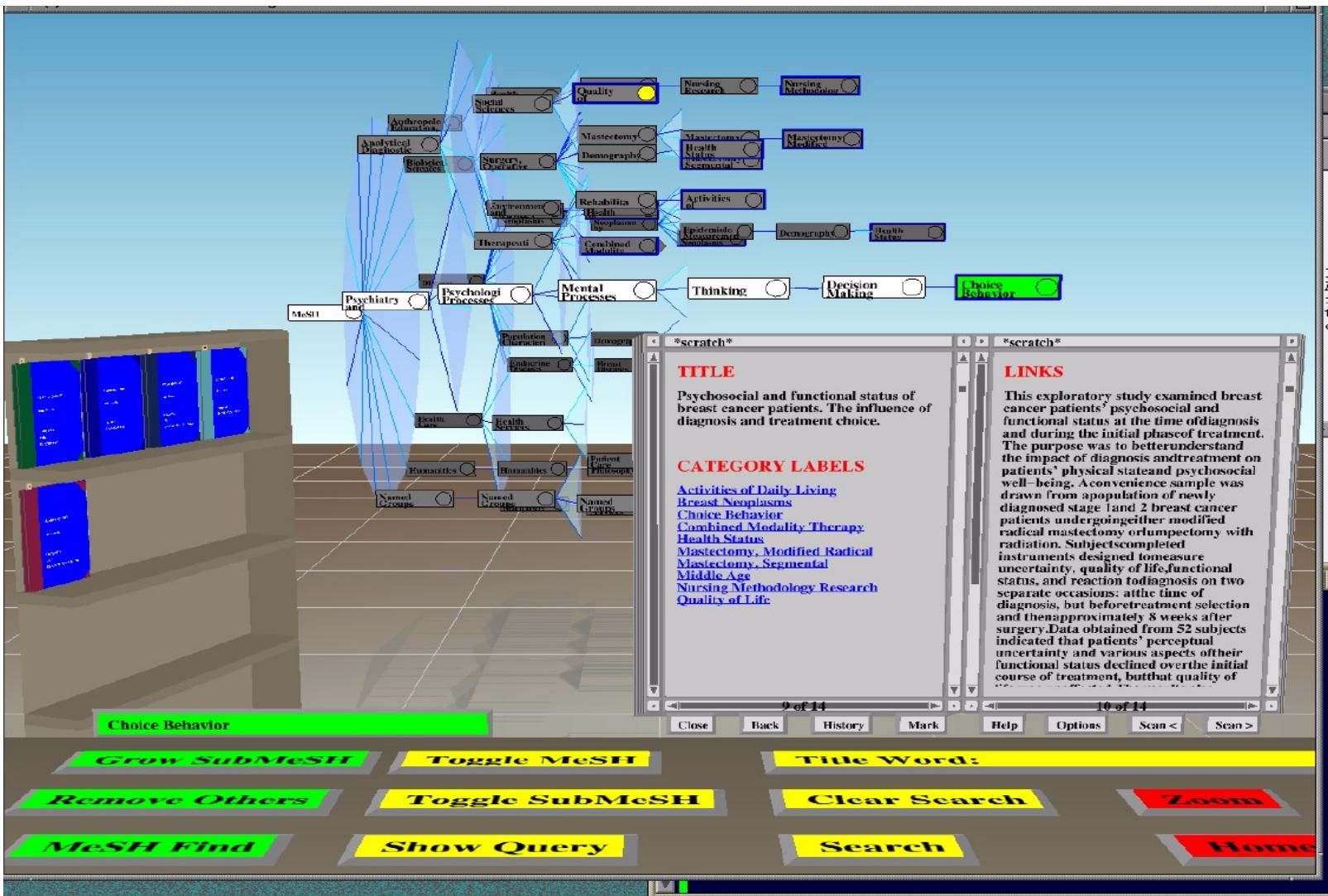
# Grouping Search Results

- Categorizing search results
  - Key idea: utilizing existing categories to group documents into categories
  - Example: Cat-a-Cone
- Clustering search results
  - Based on inter-document similarity
    - Computed using the cosine measure, for example
  - Text; Vivisimo's clusty, scatter/gather
  - 1D: scatter/gather
  - 2D: Kartoo
  - 3D: ThemeView

# Cat-a-Cone

- Key Ideas:
  - Integrate searching and browsing of very large category hierarchies with their associated text collections
  - Separate documents from category labels
    - allows the display of multiple categories per document
  - display of *multiple* selected categories simultaneously, complete with their hierarchical context
- No online demo, but with many detail discussion at  
<http://people.ischool.berkeley.edu/~hearst/cac-overview.html>

# Cat-a-Cone Interface



# Scatter/Gather

Cutting, Pedersen, Tukey & Karger 92, 93, Hearst & Pedersen 95

- How it works
  - Cluster sets of documents into general “themes”, just like a table of contents
  - Display the contents of the clusters by showing topical terms and typical titles
  - User chooses subsets of the clusters and re-clusters the documents within
  - Resulting new groups have different “themes”
- Originally used to give collection overview
- Evidence suggests more appropriate for displaying retrieval results in context

# Example: query on “star”

Encyclopedia text

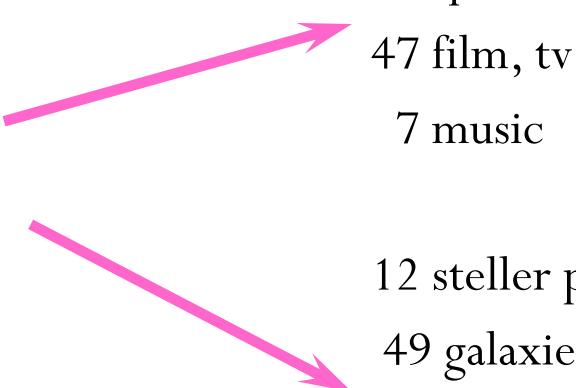
8 symbols

68 film, tv (p)

97 astrophysics

67 astronomy(p)

10 flora/fauna



Clustering and **re-clustering** is entirely automated

# Vivisimo's Clustering Results

The screenshot shows the Clusty search interface. At the top, there is a navigation bar with links for web, news, images, wikipedia, blogs, jobs, and more. Below the navigation bar is a search bar containing the query "information retrieval". To the right of the search bar are buttons for "Search" and "advanced preferences".

The main content area displays the search results for "information retrieval". It starts with a summary: "Top 243 results of at least 1,682,000 retrieved for the query information retrieval (details)".

On the left side, there is a sidebar titled "clusters" which lists various categories of results:

- All Results (247)
- + Resources (22)
- + Language (18)
- + Software (19)
- + Conference (10)
- + Modern (9)
- + Analysis (8)
- + Science of searching (8)
- + Information Retrieval Group (8)
- + Z39.50 (8)
- + Intelligent (7)
- [more | all clusters](#)

Below the sidebar, there is a "find in clusters:" input field and a "Find" button.

Font size:

The main results section has two tabs: "Sponsored Results" and "Search Results".

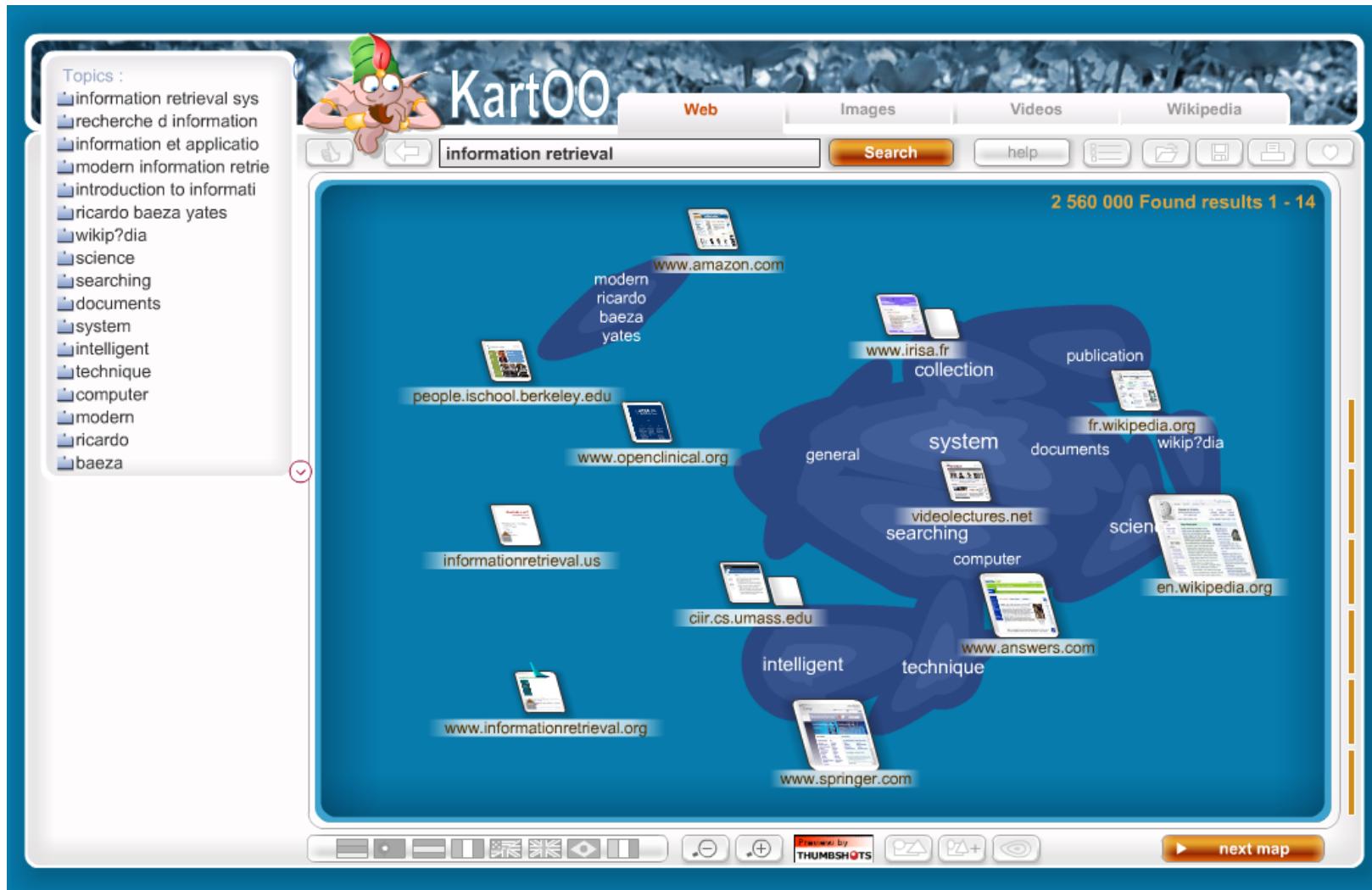
**Sponsored Results:**

- Information Retrieval**  
Search your intranet, file shares & more with solutions from Google. - [www.google.com/enterprise](http://www.google.com/enterprise)
- ISYS Information Retrieval Software**  
Bring the timesaving value of **Information Retrieval** to your business. Quickly locate content on PCs, networks, websites and intranets, regardless of file type or location. Try ISYS for free. - [www.isys-search.com](http://www.isys-search.com)

**Search Results:**

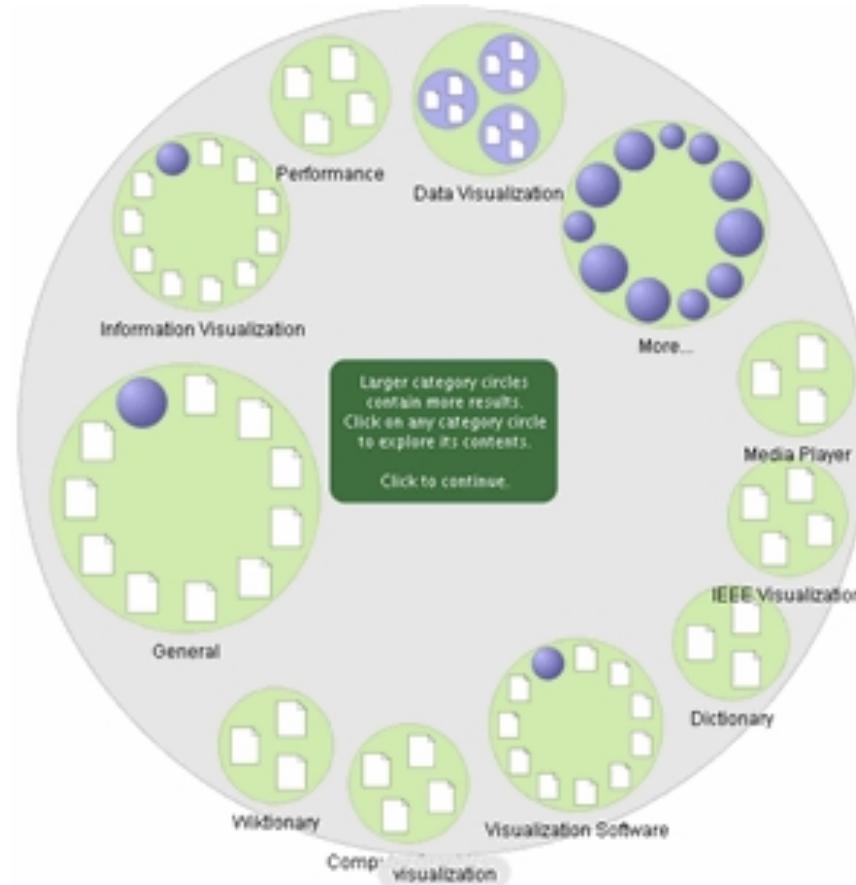
- Information retrieval**   
**Information retrieval (IR)** is the art and science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational stand alone databases or hypertext networked databases such as the Internet or intranets, for text, sound, images or data. There is a common confusion, however, between data retrieval, document retrieval, information retrieval, and text retrieval, and each of these have their own bodies of literature, theory, praxis and technologies.  
[en.wikipedia.org/wiki/Information\\_retrieval](http://en.wikipedia.org/wiki/Information_retrieval) - [cache] - Wikipedia, Live, Ask, Gigablast
- UMASS Amherst: Center for Intelligent Information Retrieval**   
The NSF funded CIIR at UMass Amherst carries out basic research and technology transfer in the area of text-based and multimedia **information** systems. Today, this Center represents an innovative and effective new model for more streamlined technology transfer, public/private partnership, and economic development  
[ciir.cs.umass.edu](http://ciir.cs.umass.edu) - [cache] - Gigablast, Open Directory, Ask
- Text REtrieval Conference (TREC) Home Page**   
An annual **information retrieval** conference and competition, the purpose of which is to support and further research within the **information retrieval** community.  
[trec.nist.gov](http://trec.nist.gov) - [cache] - Gigablast, Open Directory, Ask

# Kartoo's Cluster Visualization



# Grokker's Clustering Visualization

- clusters its results and presents them in a unique circular map
- [www.grokker.com](http://www.grokker.com) no longer work now



# Visualizing Named Entities in Results

NameSieve

pittsburgh

Search

User: dan444 logged in | Logout | Help

Carnegie Museum of Art

Matching 1000 records | Pages 1 2 3 4 5 6 7 8 9 10 Next» | Note selection

1 CMA-59085



Inscription on Kodak Royal Pan film box reads "May - 56." Charlene Foggie Barnett identified location (visit 7/16/2008). Published in the Pittsburgh Courier Newspaper, March 31, 1956, page A1.

[Group portrait including, left to right, guest speaker and director of the Washington D.C. NAACP office, Clarence Mitchell, Pittsburgh Mayor David L. Lawrence, executive secretary of the Pittsburgh NAACP affiliate, Marion Bond Jordon, Pittsburgh NAACP president, Rev. Charles H. Foglie, Pittsburgh councilman and chairman of the dinner, Paul F. Jones, and Pittsburgh NAACP executive board member, John Golightly, gathered for the Human Rights Dinner at Gateway Plaza, Downtown]

2 CMA-56466



Published in Pittsburgh Courier Newspaper, September 1954, page 1. Inscription on Kodak Super Pancho Press film box reads "Courier Aug 54"

[Pittsburgh Mayor David L. Lawrence signing the proclamation declaring September 5 - 10 Urban League Week, seated with him are Pittsburgh Urban League board members, Jessie Vann, and Henry Pearson, standing over are, Pittsburgh Urban League secretary, Mrs. Nathaniel Dandridge, and Councilman and Pittsburgh Urban League president, Patrick T. Fagan, gathered in the Office of the Mayor at the City County Building, Downtown]

3 CMA-60163



Inscription on Kodak Royal Pan film box reads "May 57". Published in the Pittsburgh Courier Newspaper, May 23, 1959, page A15.

[Group portrait of four women from the Pittsburgh Club of Negro Business and Professional Women, including, from left to right, Marion Sappington Bryant, Wilnette B. Price, Florence Allen Holmes, and Blanche Russell Crayton, Pittsburgh Club of the Negro Business and Professional Women, gathered for the Founders Day Breakfast in the Pittsburgh Room at the Penn Sheraton Hotel]

4 CMA-17297



Published in the Pittsburgh Courier Newspaper, January 12, 1946, page 1.

[Pittsburgh Mayor David L. Lawrence swearing in Robert E. "Pappy" Williams in Pittsburgh City Council Chambers at the City County Building, Downtown]

5 CMA-36776



Note on Pittsburgh Courier paper attached to negatives reads " Pitt Gad. [University of Pittsburgh Graduates], Feb. '52"

[Group portrait of four men wearing dark graduation gowns and mortar boards, possibly University of Pittsburgh graduates]

6 CMA-36779



Note on Pittsburgh Courier paper attached to negatives reads " Pitt Gad. [University of Pittsburgh Graduates], Feb. '52"

[Portrait of woman wearing dark graduation gown and mortar board, standing between two plaques, possibly University of Pittsburgh graduate]

7 CMA-36780

Note on Pittsburgh Courier paper attached to negatives reads " Pitt Gad. [University of Pittsburgh Gr

Query Terms  
**pittsburgh** (1000)

Named Entities

Who Where When What

A. H. Burchfield » Ani Howell » Aug 54 » Blanche Russell Crayton » Booker » Chairman » Charles H. Foglie » Clarence Mitchell » Councilman » Courier » Courier Newspaper » Dapper Dan » **David L. Lawrence** » David W. Walker » Delilah Court » Eugene Geller » Executive Board Member » Executive Secretary » Florence Allen Holmes » Garland » Graduates » Guest Speaker » Harry Kenny » Henry Pearson » J. H. Campbell » James Hamlett » John Golightly » Leonard Wright » Man » Marion Sappington

Total 0 NE selected

Apply Filter Clear Filters

Task Model Notes

Note something to put it here

# Result Examination Interfaces

- Present result documents in details
  - Text window to show the content
  - KWIC in documents
- To help presenting long documents
  - “Best passage” function helps users get started
    - Overlapping 300 word passages work well
  - Often visualization techniques are used

# Overview + Details in ICDL

<http://www.icdlbooks.org/>

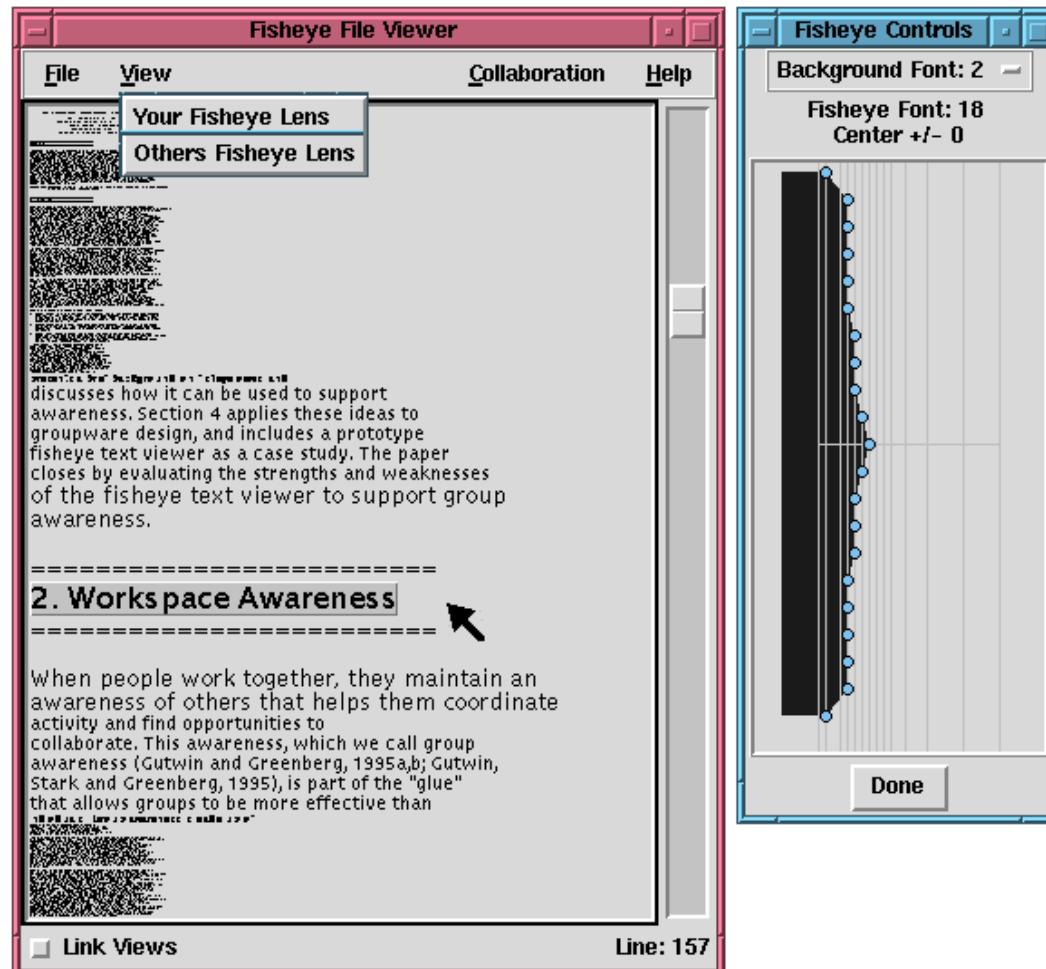
Book Overview

Language English  Library Account [Register / Sign in](#)

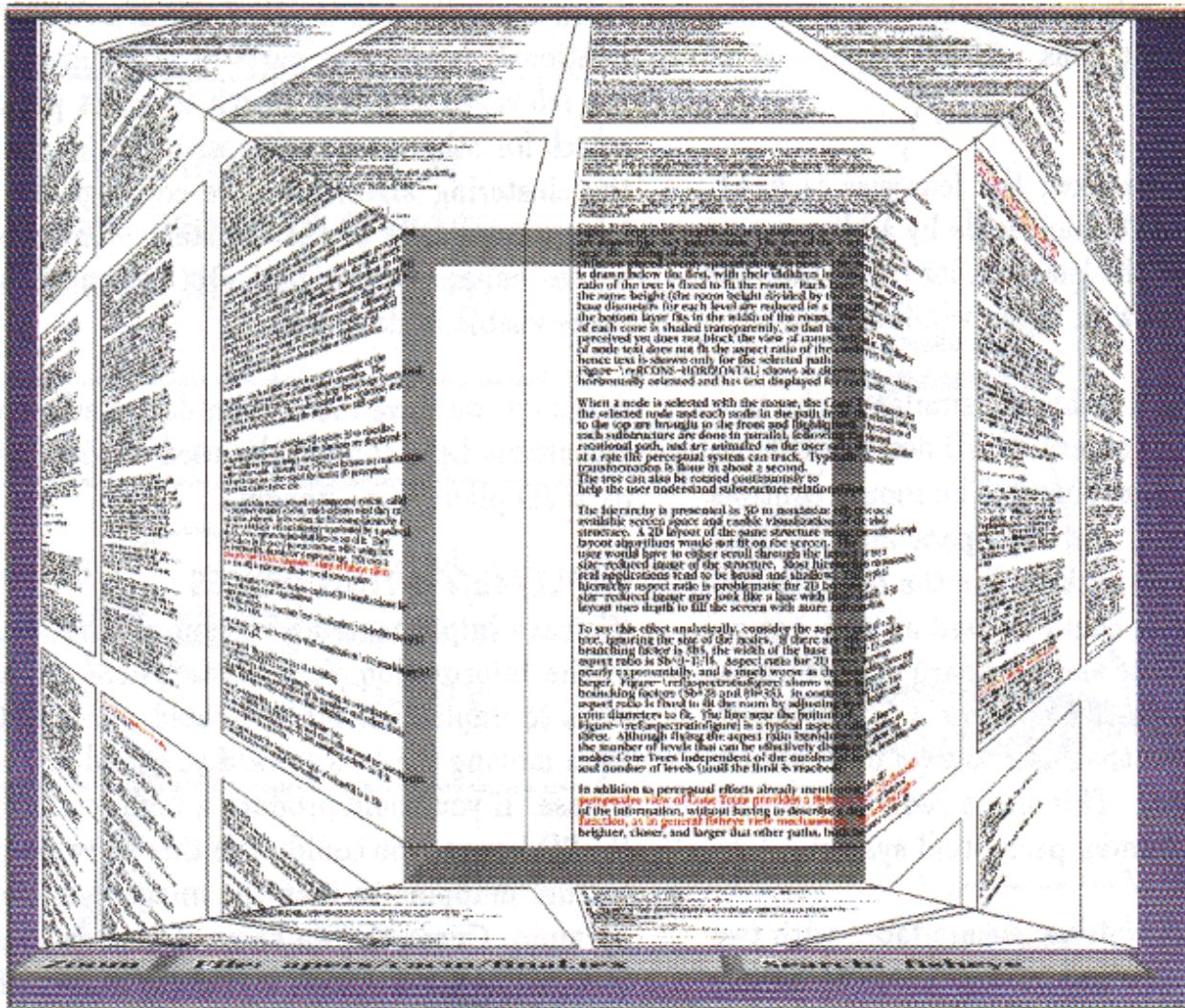
**Bigwig zero**  
Click pages to read

More book viewers

# Fisheye Views of Documents



# The Document Lens



# Mobile Search Interface

# Mobile search interfaces

- Increasingly popular
  - Predict by 2020, most people will use mobile system to access internet
- Just emerge
  - Only in the past 4-5 years, relative large, high resolution screen become popular, such as iphone, ipad, etc.
- A testbed for new technologies
  - Android mobile operating system
  - Siri speech recognition system

# Mobile Search Usages

- Unique information needs
  - Information: Temporal and geolocation information can be the focus
  - Needs: route-finding, planning, related to current location
    - Answer questions that came up in current activity, location, time and conversation
- More focused, more repeated queries
  - Most frequent 1000 mobile queries account 22% of all mobile queries
    - In contrast, most frequent 1000 desktop queries account 6%
- Shallower interactions with results
  - Fewer page views per query
  - But more rely on current context

# Old Google Interface

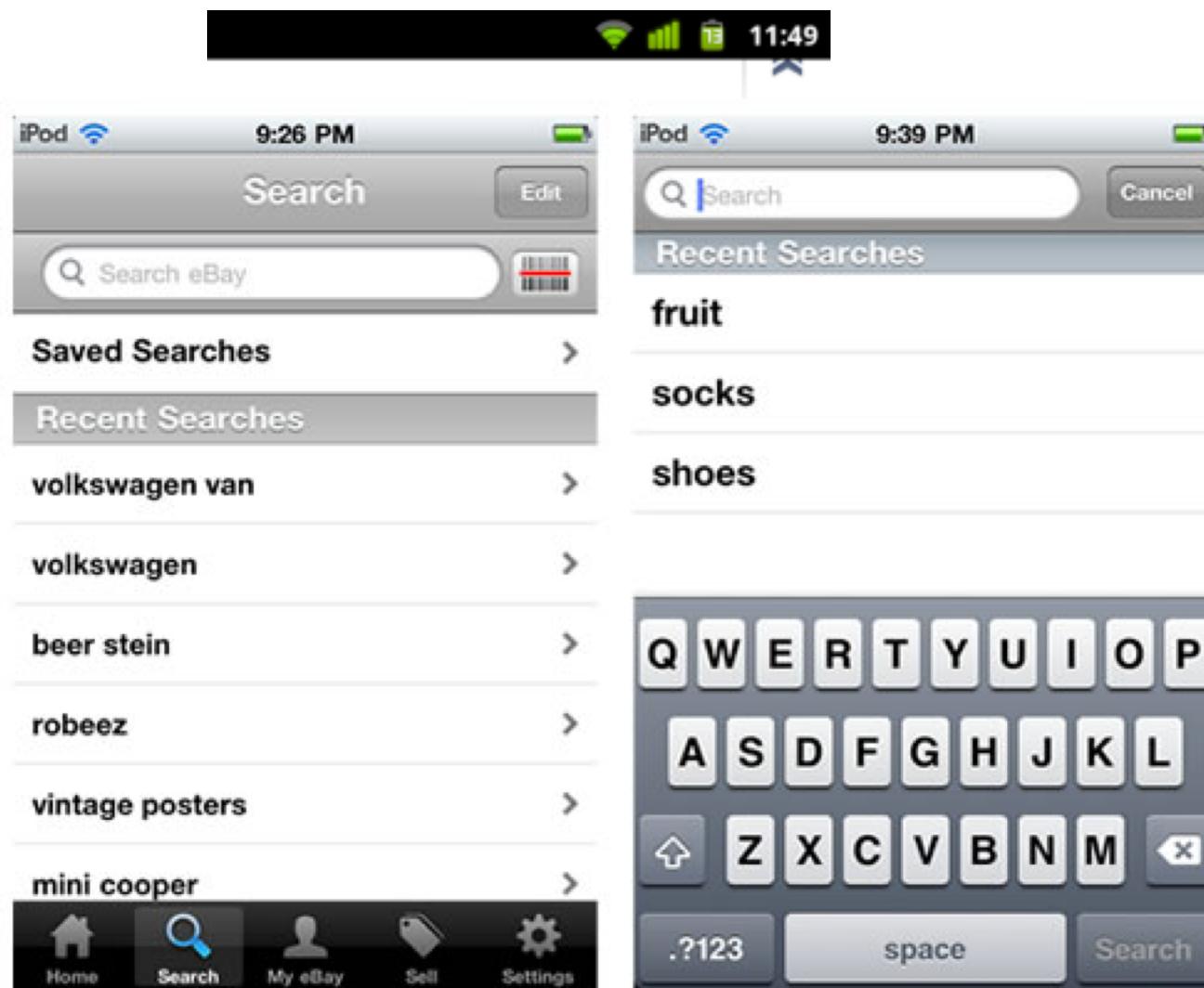
The screenshot shows the classic Google search interface. At the top is the Google logo. Below it is a search bar with a placeholder text that is mostly obscured by a large redacted area. To the right of the search bar is a blue "Search" button. Underneath the search bar are four radio buttons with labels: "Web" (selected), "Images", "Local", and "Mobile Web (Beta)".

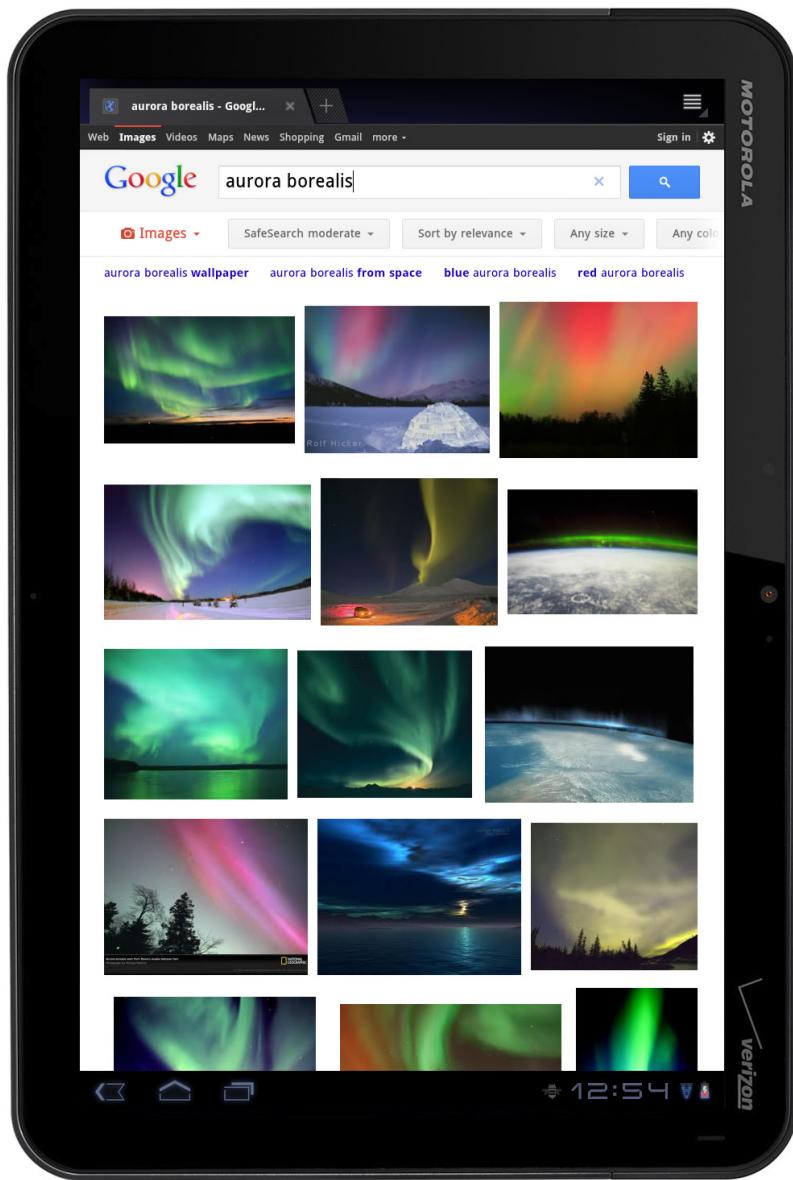
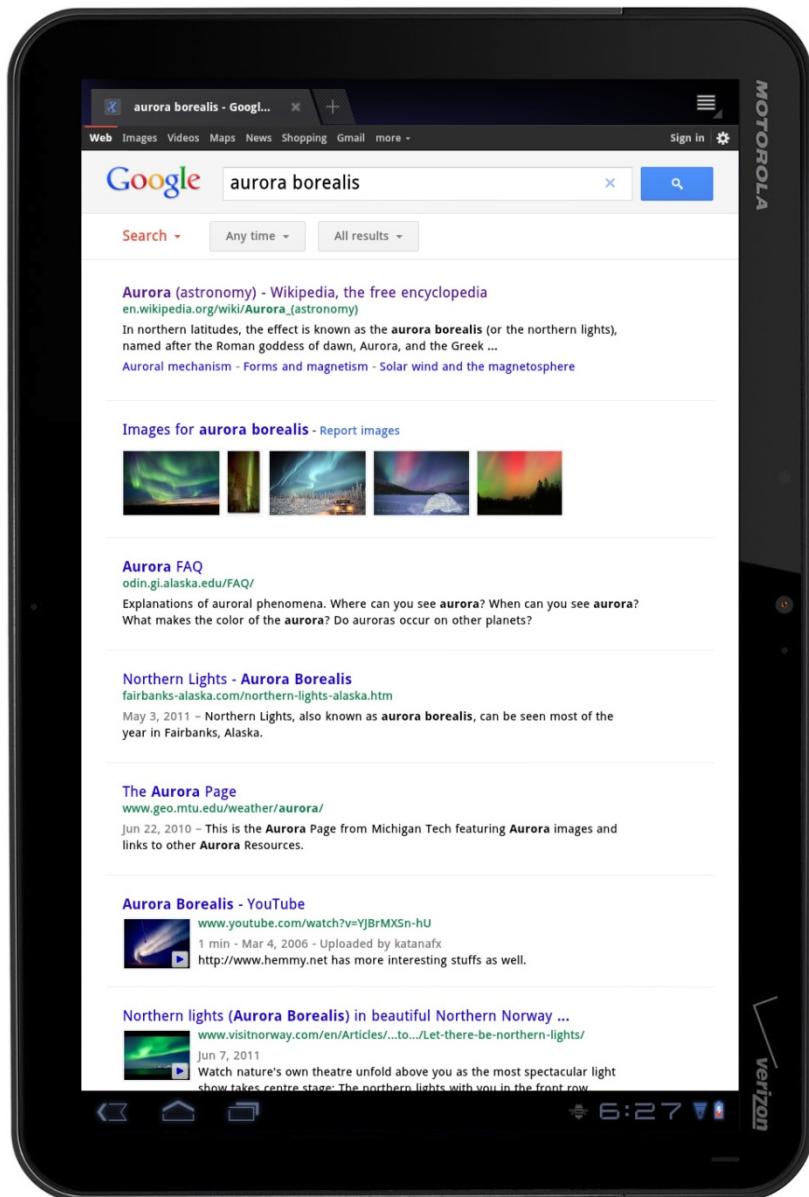
The screenshot shows the search results page for the query "cars". The title "Web results: 'cars'" is displayed. Below it, the text "Results 1 - 10 of about 86,800,000." is shown. The first result is a link to Cars.com: "1 [Buy new & used cars online, research prices & dealers, sell your ...](#) - Cars.com is your online source to buy new and used cars. Sell your used car, or - www.cars.com/".

The second result is a link to Jaguar: "2 [Jaguar](#) - Official worldwide web site of Jaguar Cars. -

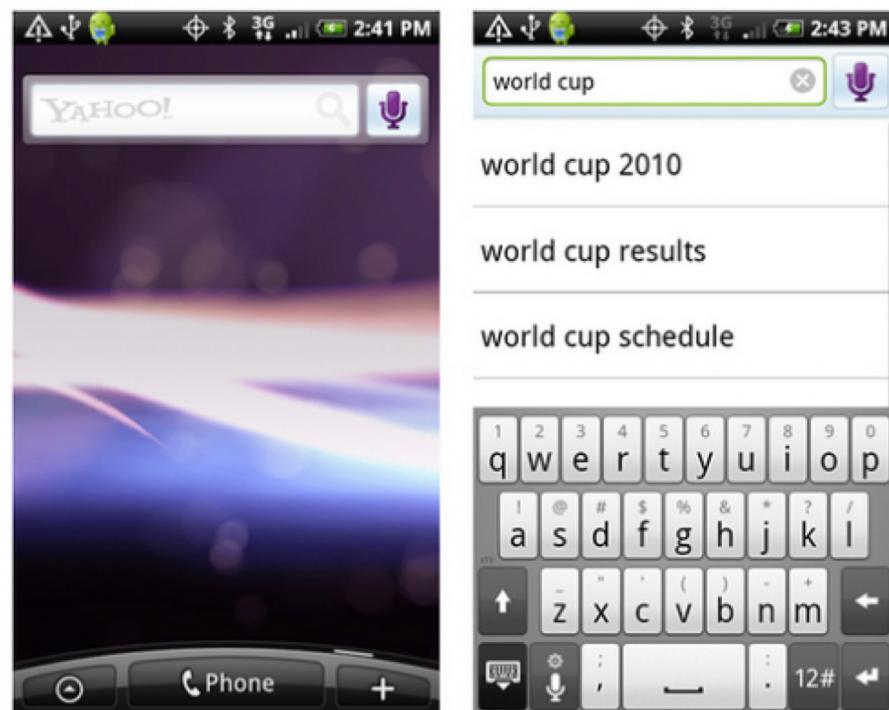
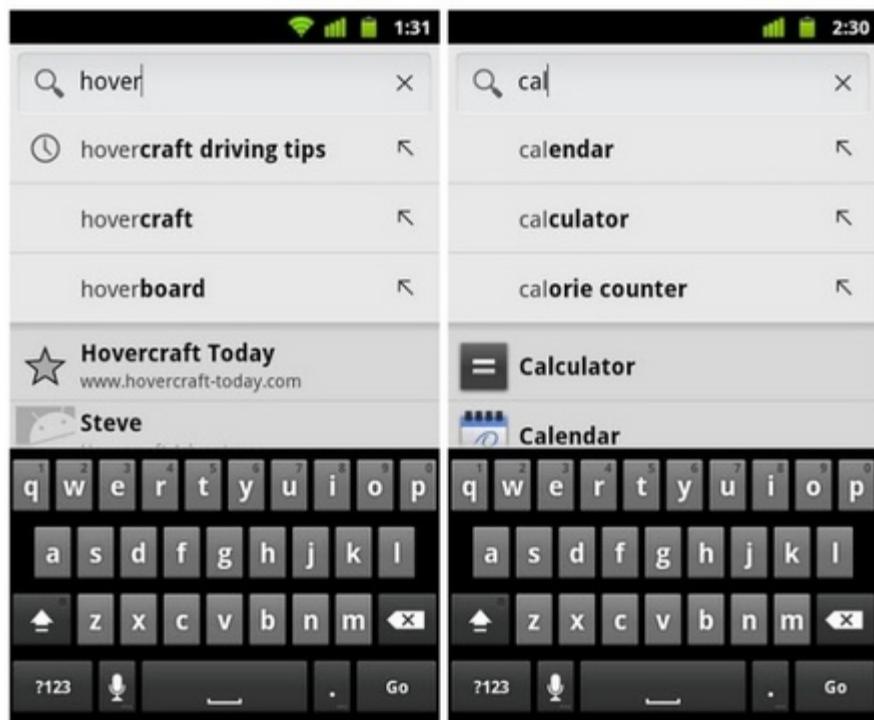
The screenshot shows the detailed view of the first search result for "cars". The URL "www.virgincars.com/" is at the top. Below it, the text "10 [globeandmail.com - Breaking Megawheels News](#) - Few cars will be ready for satellite radio in fall. Canadian consumers generally - www.globemegawheels.com/" is displayed. At the bottom of the page are "Prev" and "Next" links, followed by a search bar containing "cars" and a "Google Search" button, along with the same four navigation radio buttons as the top left panel.

# New Google Search Interface

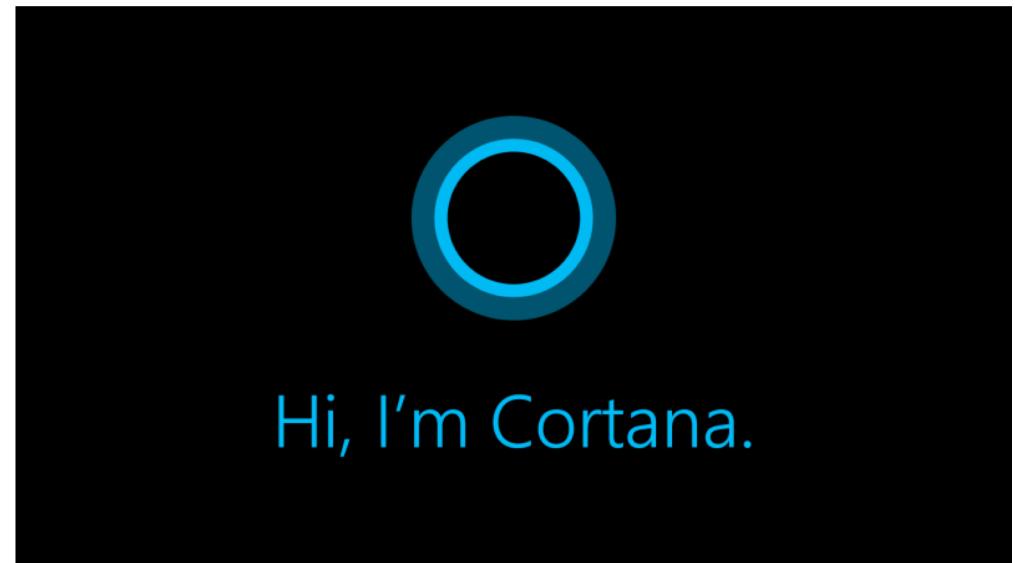
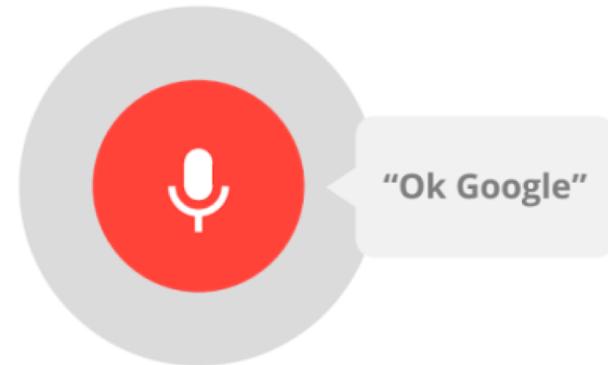




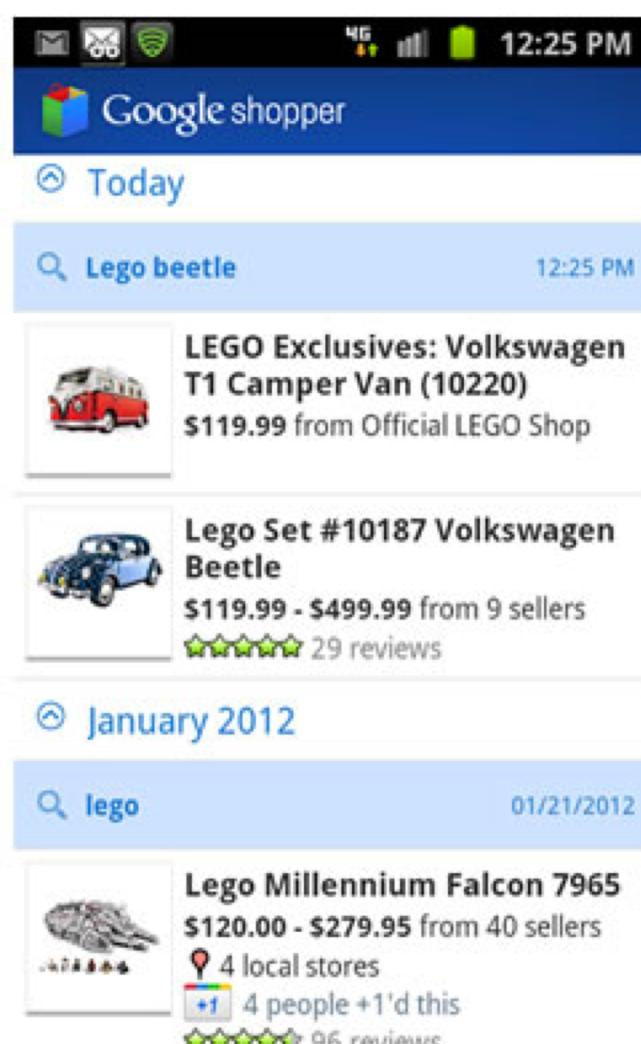
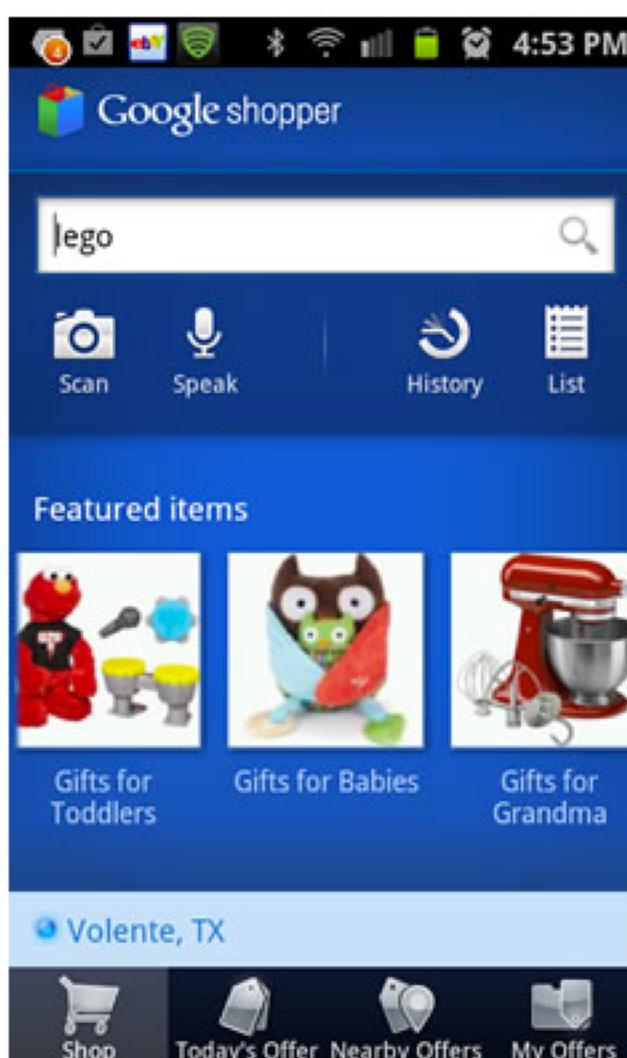
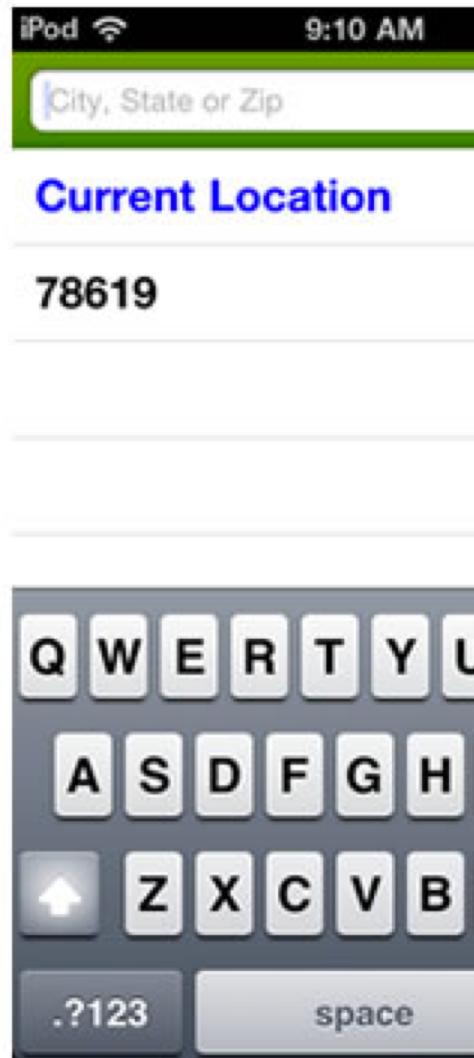
# Query Auto-Completion



# Speech based Search Input



# Context-based Search: Location



# Search Results

iPod WiFi  
Flights  
Austin to Rio de Janeiro, Brazil  
Dec 05 to Dec 26, 2013  
**\$2463**  
12/05 AUS 7:04 AM  
12/26 GIG 10:55 PM

**\$2463**  
12/05 AUS 7:04 AM  
12/26 GIG 10:55 PM

**\$2463**  
12/05 AUS 7:04 AM  
12/26 GIG 10:55 PM

**\$2470**  
Filter

airbnb  
copenhagen  
List  
20 of 757 results for "copenhagen"

Cop...  
Cop...  
Cop...  
Best...  
Cope...  
14 r...

'robbee boys': 92 found

SORT FILTER

**Robeez**  
Ballerina Bear Soft Sole...  
★★★★★ \$24.00

**Robeez**  
Fire Engine Soft Soles (Infant/Toddler)...  
★★★★★ \$24.00

**Robeez**  
Wave Crasher Mini Shoe...  
★★★★★ \$32.00 NEW

**Robeez**  
Happy Snowman (Infant/Toddler)...  
★★★★★ \$21.99 SALE

**Robeez**  
Skiing Penguin (Infant/Toddler)...  
★★★★★ \$21.99 SALE

Robeez

'robbee boys': 92 found

SORT FILTER



Robeez Surf Shack Soft Sol... \$21.99  
SKU: #7899100  
1 - 26



# Sorting Results

The image displays two screenshots of an iPod interface, likely running iOS 4.2, illustrating the process of sorting search results.

**Screenshot 1 (Left):** Shows a search for "santa monica" on June 4, 2011, from Austin to Rio de Janeiro. The results page lists three flight options:

- Intex 1 Groun**: \$299.0 Was: \$399.0. Multiple Airlines. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.
- Intex I Water**: \$2463. Multiple Airlines. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.
- Intex 1 Above**: \$2463. Continental. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.
- Backya Cedar**: \$399.0. Multiple Airlines. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.

**Screenshot 2 (Right):** Shows a search for "santa monica" on June 4, 2011, from Austin to Los Angeles. The results page lists two flight options:

- \$759**: Delta. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.
- \$764**: Delta. Departure: 12/05 AUS 7:04p, Arrival: 12/26 GIG 10:55p.

A modal dialog box is open, listing sorting options:

- Least expensive** (selected)
- Shortest duration**
- Leaving soonest**

**Bottom Bar:** Filter, Sort

# Mobile Touch Interactions

## TOUCH GESTURES



AT&T 3:33 PM 88% ACM Scholar Paper Testing 54.243.145.55:8080/ Reader Search

**Real-time recommendation of diverse related arti ...**  
News articles typically drive a lot of traffic in the form of comments posted by users on a news site. Such user-generated content tends to carry ad ...

**Multi-label learning with millions of labels: re ...**  
Recommending phrases from web pages for advertisers to bid on against search engine queries is an important research problem with direct commercial ...

**Hierarchical geographical modeling of user locat ...**  
With the availability of cheap location sensors, geotagging of messages in online social networks is proliferating. For instance, Twitter, Facebook, ...

**Distributed large-scale natural graph factorizat ...**  
Natural graphs, such as social networks, email graphs, or instant messaging patterns, have become pervasive through the internet. These graphs are in ...

**A CRM system for social media: challenges and ex ...**  
The social Customer Relationship Management (CRM) landscape is attracting significant attention from customers and enterprises alike as a sustainabl ...

**Here's my cert, so trust me, maybe?: understandi ...**  
When browsers report TLS errors, they cannot distinguish

90

15

IS2140

Jing He

# Further Topics on Mobile Search

- Mobile information seeking behaviors
  - Small screen/keyboard size make it difficult to type keywords. Studies found that people did not issue shorter queries, why ?
  - Spent much more time on reading a web page
  - Clicked more high-quality documents
- Cross-device web search
  - People often migrated tasks from one device to another
- Utilizing unique information in mobile web search
  - Contextual information
  - Mobile touch interactions

# Evaluating Interactive IR Systems

# Evaluation Dimensions

- Three dimensions of usability
  - Effectiveness: Accuracy and completeness with which users achieve specified goals.
  - Efficiency: Resources expended in relation to the accuracy and completeness with which users achieve goals (usually, how much time).
  - Satisfactory: Freedom from discomfort, and positive attitudes towards the use of the product.
- Evaluation Methods
  - Standard IR metrics – precision, recall, nDCG and etc.
  - Informal usability test – interview, field study in early stage
  - Formal usability test – controlled user experiments (user study)
  - Search log analysis and A/B Testing

# Controlled User Study

- Techniques in User Study
  - IRB (Institutional review board)
  - Consent form
  - Training task
  - Pilot study
- Balance condition ordering effects
  - Removing learning effects, fatigue effects and task effects
  - Latin Square ordering
- Obtaining participants
- Obtain participants subjective responses
  - Cognitive load

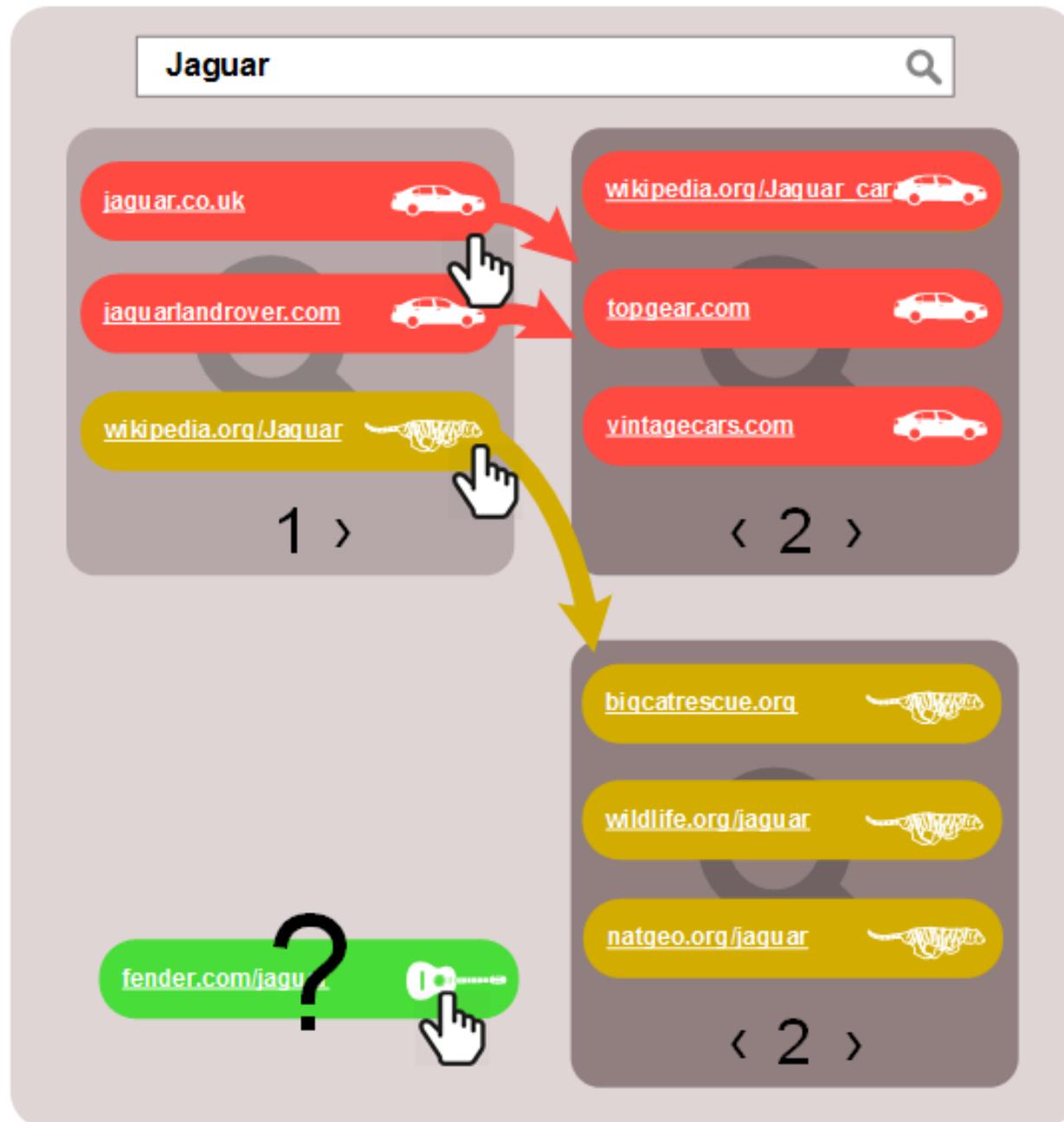
# Summary

# Synergy between Human and IR system

- The strength of one covers the weakness of the other
  - Humans do things that human are good at
  - Computers do things that computers are good at
- What human good at?
  - Sense low level stimuli, Recognize patterns, Reason inductively, Communicate with multiple channels, Apply multiple strategies, Adapt to changes or unexpected events
- What machine good at?
  - Sense stimuli outside human's range, Calculate fast and mechanical, Store large quantities and recall accurately, Response rapidly and consistently, Perform repetitive actions reliably, Maintain performance under heavy load and extended time

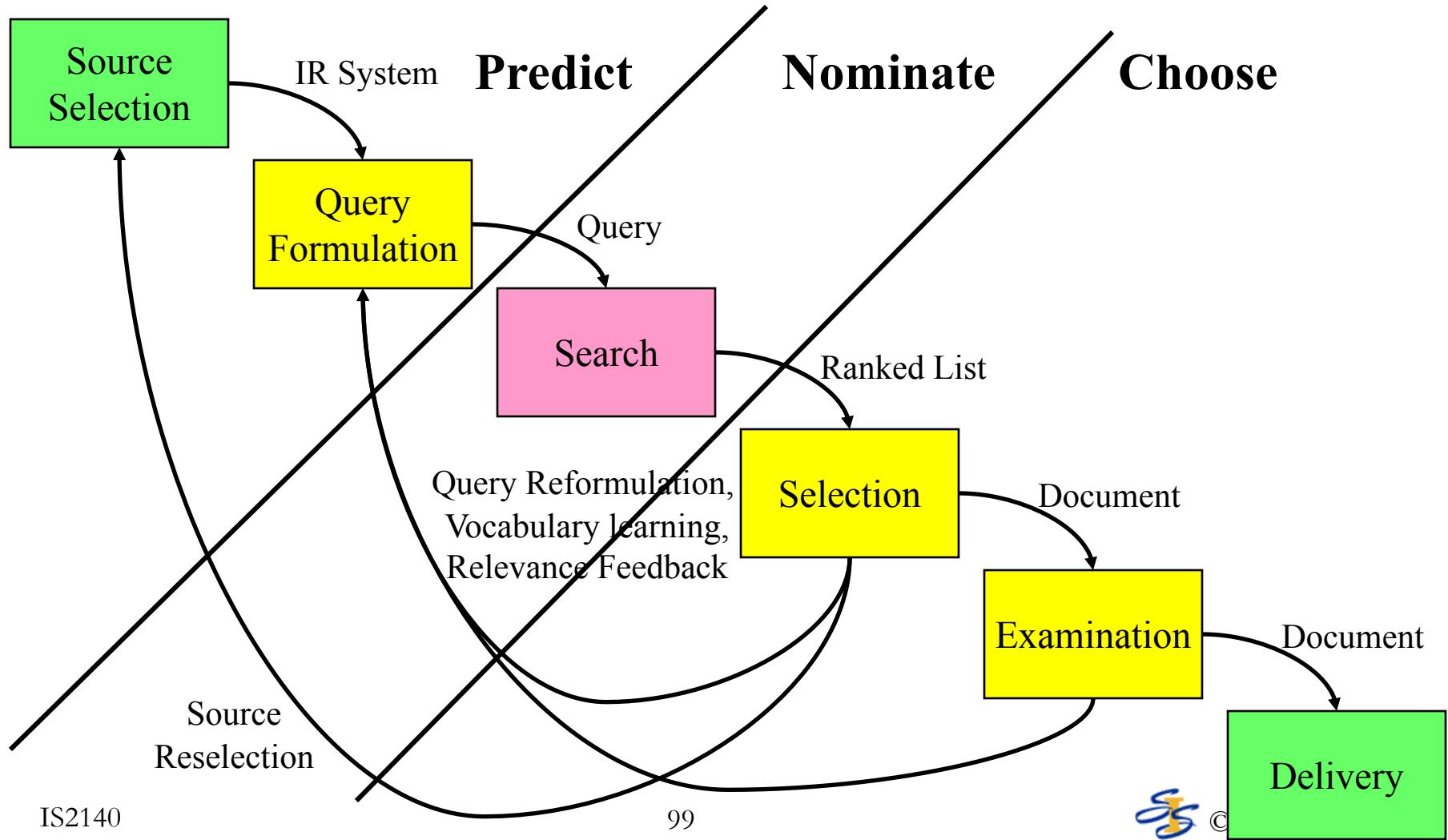
# The Importance of Interfaces

- 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



*Clicked webpages lead to the personalization of the second page of results based on the subtopic clicked, but not all of the subtopics are represented.*

# Information Retrieval Cycle



# An Example

Current query  
Results preview  
Subject: asset swap  
Body: asset swap  
From: To:  
With documents: doc5

Filter by: Frequency, Viewed, Saved, Selected, All docs, Date: All years

Recent queries:  
body:(asset swap) | 1 doc Yesterday  
body:(asset swap) Yesterday  
body:(chewco) Wednesday  
body:(jedi) Wednesday

Sort by: Rank T, Date T, Task score I, Recently seen I

1 2 3 4 5 6 7 8 9 10 >>

Subject: SEC VaR  
Sent: 08 Feb 2001, 09:37  
From: adarsh.vakharia@enron.com  
To: georgeanne.hedges@enron.com, jan.johnson@enron.com, sally.beck@enron.com, ... more  
Georgeanne told Eugenio about another Enron stock swap. This one is very the Phantom. As a result, the VaR for non-trading securities increased from an average \$8 to \$39 million.

We also recalculated Merchant... more

Subject: Eligible Swap Participant  
Sent: 21 Dec 1999, 10:37  
From: sara.shackleton@enron.com  
To: deborah.culver@enron.com  
Just to be certain that the EES entity that might enter into the swap is million or (2) has a net worth of \$1 million and enters into the swap agreement in connection with the conduct of its business or to manage... more

Subject: Raptor Questions  
Sent: 18 Aug 2000, 09:42  
From: sara.shackleton@enron.com  
To: julia.murray@enron.com, mary.cook@enron.com  
drafted the share swap confirm as a PRICE RETURN swap. Scott then requested a TOTAL RETURN swap (ie, dividends included). I met with accounting this morning and (Ryan believes... more

Subject: swap on Ponderosa shares of Enron International Brazil 2000 Ltd.  
Sent: 26 Sep 2000, 10:44  
From: sara.shackleton@enron.com  
To: scott.sefton@enron.com, ryan.siurek@enron.com, gordon.mckillop@enron.com, ... more  
Attached is a first draft of the swap for the Raptor 2 structure. Please accordingly:

1. I have ... that everyone is in agreement as to how this swap will terminate. [Do we want partial terminations based upon... more

Subject: Raptor Hedges  
Sent: 30 Aug 2000, 06:38  
From: scott.sefton@enron.com, gordon.mckillop@enron.com, ryan.siurek@enron.com, ... more  
that the swaps for assets in JEDL 1 and 2 will cover our proportionate economic interest in the assets within JEDL. Gordon believes the spreadsheet of proposed swaps reflects this. The swap will be between Harmer... more

Current Queries Documents Comments Notes Collaborators Tasks gene's research Gene | Logout

Chart Document

