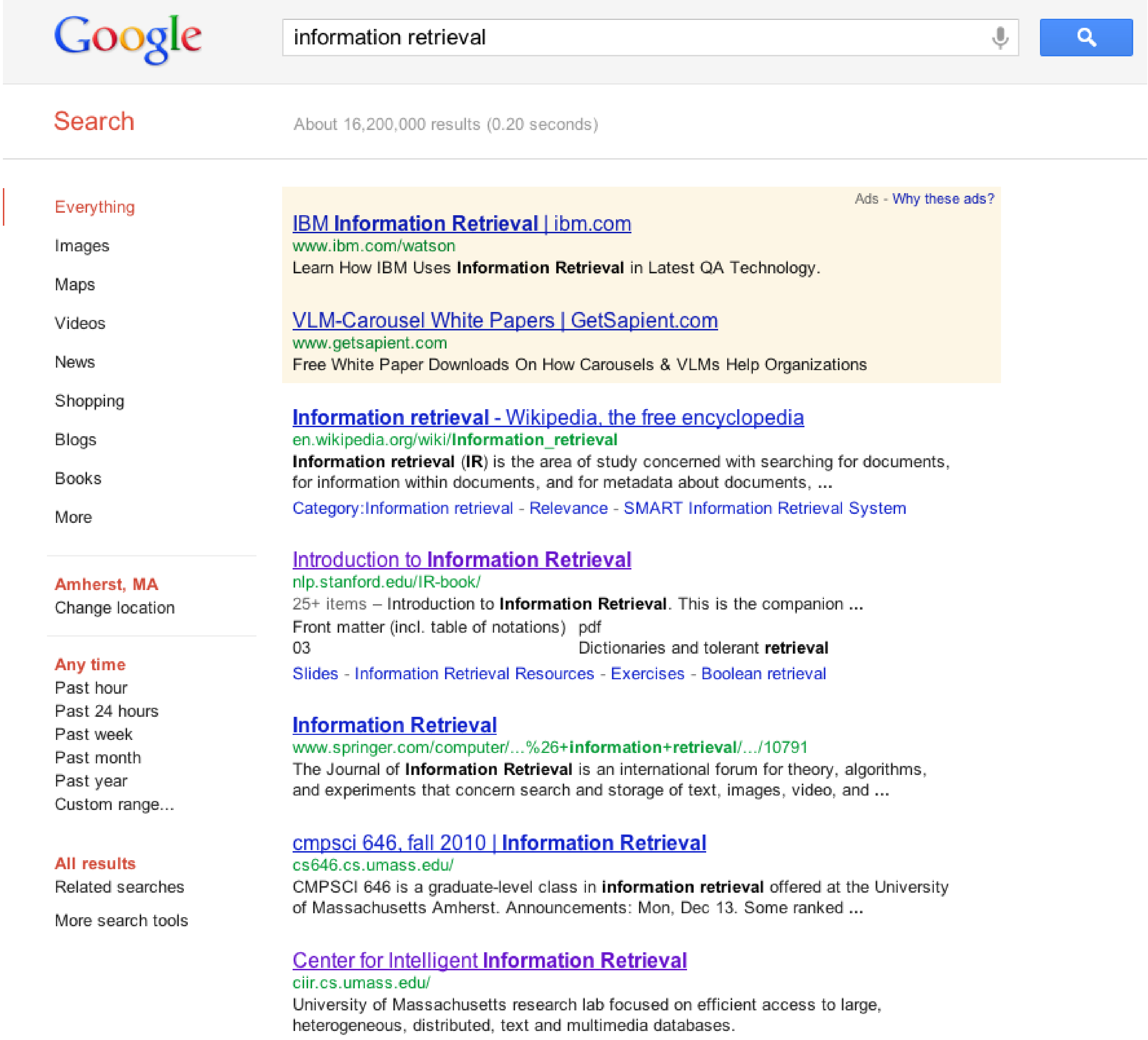
CMPSCI 446
Search Engines



information retrieval

About 16,200,000 results (0.20 seconds)

Ads - Why these ads?

IBM **Information Retrieval** | ibm.com
www.ibm.com/watson
Learn How IBM Uses **Information Retrieval** in Latest QA Technology.

VLM-Carousel White Papers | GetSapient.com
www.getsapient.com
Free White Paper Downloads On How Carousels & VLMs Help Organizations

Information retrieval - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**Information_retrieval**
Information retrieval (**IR**) is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, **...**
Category:Information retrieval - Relevance - SMART Information Retrieval System

Introduction to **Information Retrieval**
nlp.stanford.edu/IR-book/
25+ items – Introduction to **Information Retrieval**. This is the companion **...**

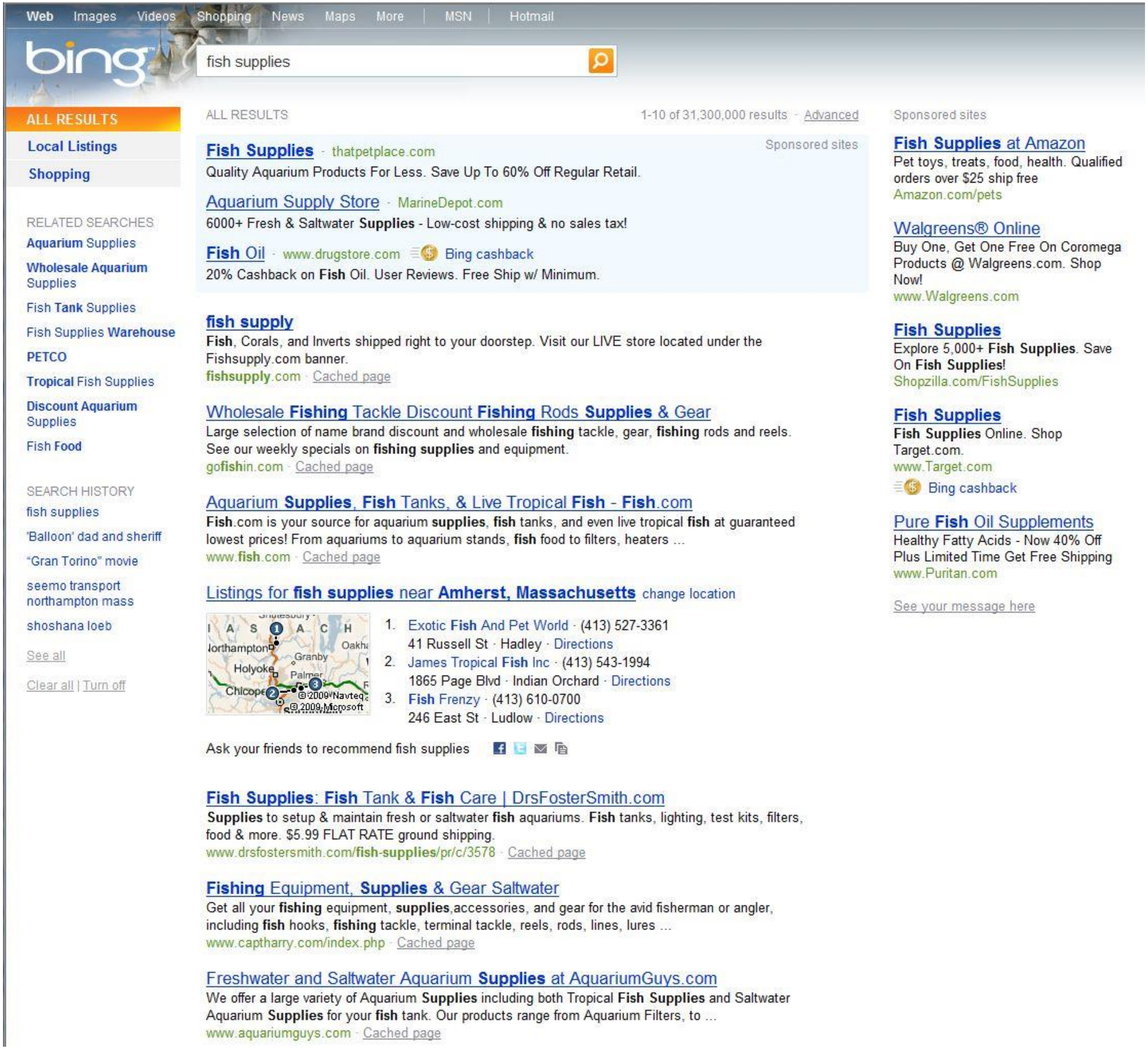
Front matter (incl. table of notations)	pdf
03	Dictionaries and tolerant **retrieval**

Slides - Information Retrieval Resources - Exercises - Boolean retrieval

Information Retrieval
www.springer.com/computer/...%26+**information+retrieval**/.../10791
The Journal of **Information Retrieval** is an international forum for theory, algorithms, and experiments that concern search and storage of text, images, video, and **...**

cmpsci 646, fall 2010 | **Information Retrieval**
cs646.cs.umass.edu/
CMPSCI 646 is a graduate-level class in **information retrieval** offered at the University of Massachusetts Amherst. Announcements: Mon, Dec 13. Some ranked **...**

Center for Intelligent **Information Retrieval**
ciir.cs.umass.edu/
University of Massachusetts research lab focused on efficient access to large, heterogeneous, distributed, text and multimedia databases.



ALL RESULTS

Local Listings

Shopping

RELATED SEARCHES

Aquarium Supplies

Wholesale Aquarium Supplies

Fish Tank Supplies

Fish Supplies Warehouse

PETCO

Tropical Fish Supplies

Discount Aquarium Supplies

Fish Food

SEARCH HISTORY

fish supplies

'Balloon' dad and sheriff

"Gran Torino" movie

seemo transport northampton mass

shoshana loeb

See all

Clear all | Turn off

ALL RESULTS 1-10 of 31,300,000 results · Advanced

Sponsored sites

Fish Supplies · thatpetplace.com
Quality Aquarium Products For Less. Save Up To 60% Off Regular Retail.

Aquarium Supply Store · MarineDepot.com
6000+ Fresh & Saltwater **Supplies** - Low-cost shipping & no sales tax!

Fish Oil · www.drugstore.com · Bing cashback
20% Cashback on **Fish** Oil. User Reviews. Free Ship w/ Minimum.

fish supply
Fish, Corals, and Inverts shipped right to your doorstep. Visit our LIVE store located under the Fishsupply.com banner.
fishsupply.com · Cached page

Wholesale **Fishing** Tackle Discount **Fishing** Rods **Supplies** & Gear
Large selection of name brand discount and wholesale **fishing** tackle, gear, **fishing** rods and reels. See our weekly specials on **fishing supplies** and equipment.
go**fish**in.com · Cached page

Aquarium **Supplies**, **Fish** Tanks, & Live Tropical **Fish** - **Fish**.com
Fish.com is your source for aquarium **supplies**, **fish** tanks, and even live tropical **fish** at guaranteed lowest prices! From aquariums to aquarium stands, **fish** food to filters, heaters ...
www.**fish**.com · Cached page

Listings for **fish supplies** near **Amherst, Massachusetts** change location

1. Exotic **Fish** And Pet World · (413) 527-3361
41 Russell St · Hadley · Directions
2. James Tropical **Fish** Inc · (413) 543-1994
1865 Page Blvd · Indian Orchard · Directions
3. **Fish** Frenzy · (413) 610-0700
246 East St · Ludlow · Directions

Ask your friends to recommend fish supplies

Fish Supplies: **Fish** Tank & **Fish** Care | DrsFosterSmith.com
Supplies to setup & maintain fresh or saltwater **fish** aquariums. **Fish** tanks, lighting, test kits, filters, food & more. $5.99 FLAT RATE ground shipping.
www.drsfostersmith.com/**fish-supplies**/pr/c/3578 · Cached page

Fishing Equipment, **Supplies** & Gear Saltwater
Get all your **fishing** equipment, **supplies**,accessories, and gear for the avid fisherman or angler, including **fish** hooks, **fishing** tackle, terminal tackle, reels, rods, lines, lures ...
www.captharry.com/index.php · Cached page

Freshwater and Saltwater Aquarium **Supplies** at AquariumGuys.com
We offer a large variety of Aquarium **Supplies** including both Tropical **Fish Supplies** and Saltwater Aquarium **Supplies** for your **fish** tank. Our products range from Aquarium Filters, to ...
www.aquariumguys.com · Cached page

Sponsored sites

Fish Supplies at Amazon
Pet toys, treats, food, health. Qualified orders over $25 ship free
Amazon.com/pets

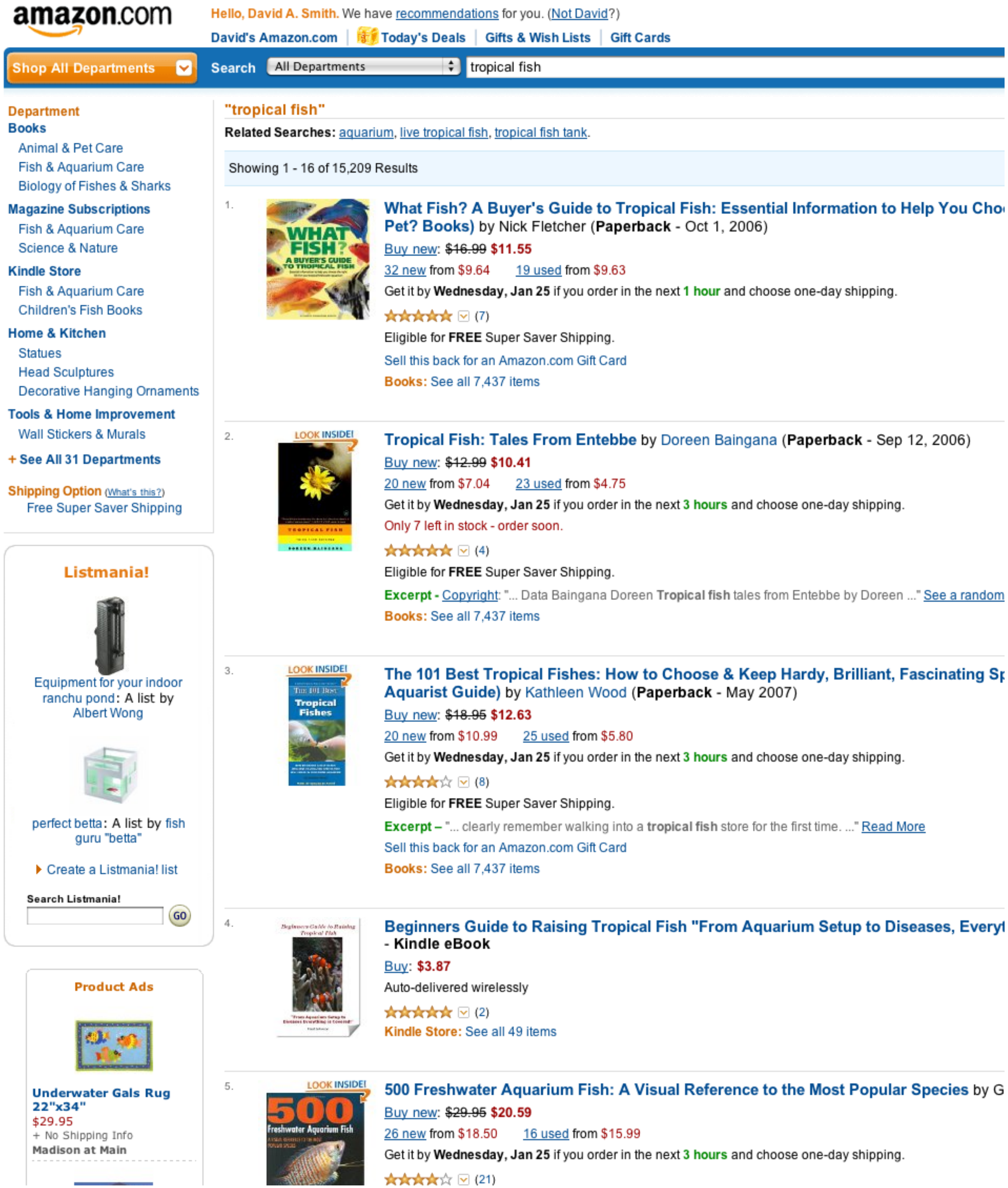
Walgreens® Online
Buy One, Get One Free On Coromega Products @ Walgreens.com. Shop Now!
www.Walgreens.com

Fish Supplies
Explore 5,000+ **Fish Supplies**. Save On **Fish Supplies**!
Shopzilla.com/FishSupplies

Fish Supplies
Fish Supplies Online. Shop Target.com.
www.Target.com
Bing cashback

Pure **Fish** Oil **Supplements**
Healthy Fatty Acids - Now 40% Off Plus Limited Time Get Free Shipping
www.Puritan.com

See your message here



amazon.com

Hello, David A. Smith. We have recommendations for you. (Not David?)

David's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Shop All Departments | Search | All Departments | tropical fish

Department

Books

- Animal & Pet Care
- Fish & Aquarium Care
- Biology of Fishes & Sharks

Magazine Subscriptions

- Fish & Aquarium Care
- Science & Nature

Kindle Store

- Fish & Aquarium Care
- Children's Fish Books

Home & Kitchen

- Statues
- Head Sculptures
- Decorative Hanging Ornaments

Tools & Home Improvement

- Wall Stickers & Murals

+ See All 31 Departments

Shipping Option (What's this?)

Free Super Saver Shipping

Listmania!

Equipment for your indoor ranchu pond: A list by Albert Wong

perfect betta: A list by fish guru "betta"

▸ Create a Listmania! list

Search Listmania! GO

Product Ads

Underwater Gals Rug 22"x34"
$29.95
+ No Shipping Info
Madison at Main

"tropical fish"

Related Searches: aquarium, live tropical fish, tropical fish tank.

Showing 1 - 16 of 15,209 Results

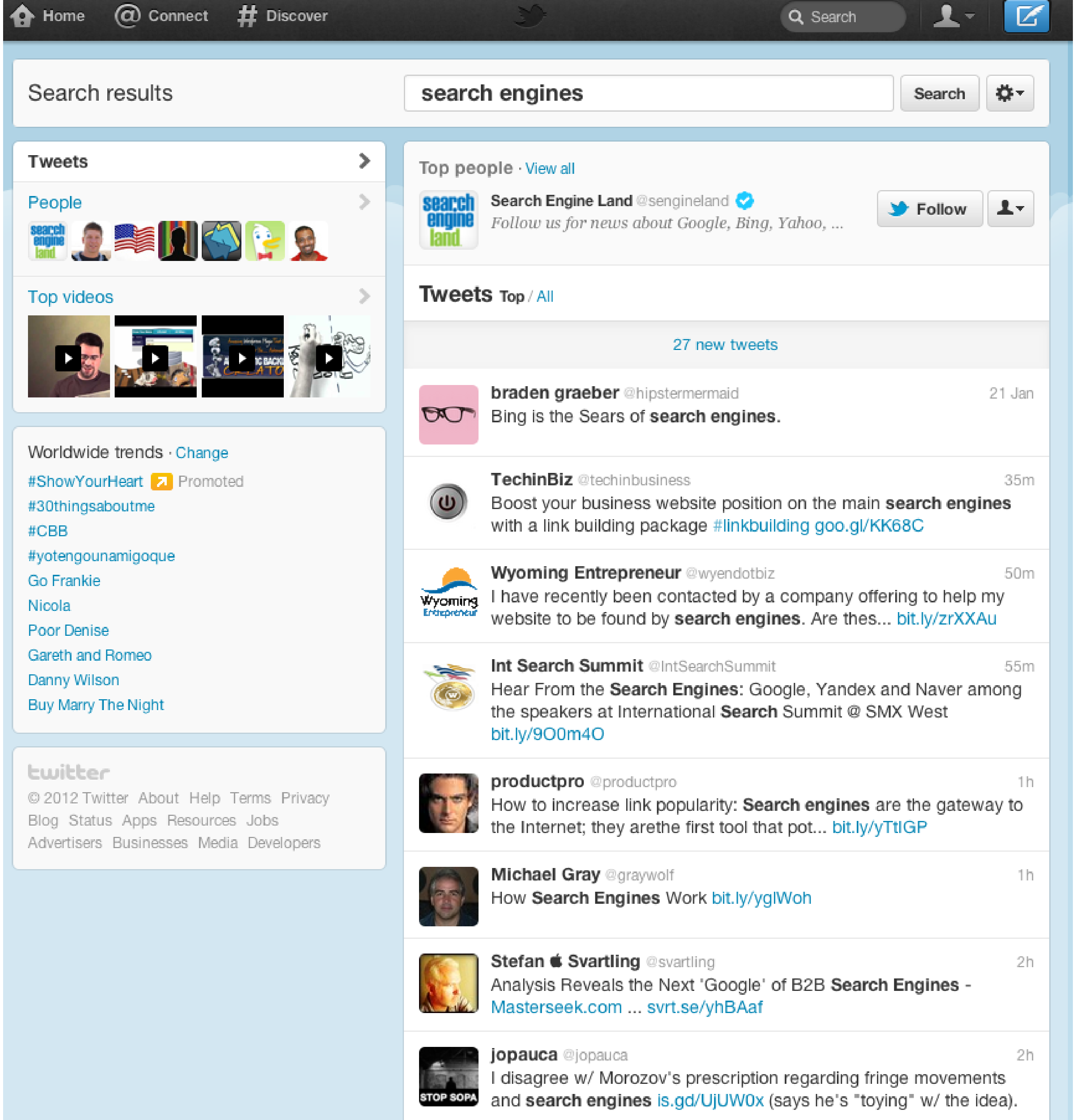
1. **What Fish? A Buyer's Guide to Tropical Fish: Essential Information to Help You Cho Pet? Books)** by Nick Fletcher (**Paperback** - Oct 1, 2006)
Buy new: ~~$16.99~~ $11.55
32 new from $9.64 19 used from $9.63
Get it by **Wednesday, Jan 25** if you order in the next **1 hour** and choose one-day shipping.
★★★★★ (7)
Eligible for **FREE** Super Saver Shipping.
Sell this back for an Amazon.com Gift Card
Books: See all 7,437 items

2. **Tropical Fish: Tales From Entebbe** by Doreen Baingana (**Paperback** - Sep 12, 2006)
Buy new: ~~$12.99~~ $10.41
20 new from $7.04 23 used from $4.75
Get it by **Wednesday, Jan 25** if you order in the next **3 hours** and choose one-day shipping.
Only 7 left in stock - order soon.
★★★★★ (4)
Eligible for **FREE** Super Saver Shipping.
Excerpt - Copyright: "... Data Baingana Doreen **Tropical fish** tales from Entebbe by Doreen ..." See a random
Books: See all 7,437 items

3. **The 101 Best Tropical Fishes: How to Choose & Keep Hardy, Brilliant, Fascinating Sp Aquarist Guide)** by Kathleen Wood (**Paperback** - May 2007)
Buy new: ~~$18.95~~ $12.63
20 new from $10.99 25 used from $5.80
Get it by **Wednesday, Jan 25** if you order in the next **3 hours** and choose one-day shipping.
★★★★☆ (8)
Eligible for **FREE** Super Saver Shipping.
Excerpt – "... clearly remember walking into a **tropical fish** store for the first time. ..." Read More
Sell this back for an Amazon.com Gift Card
Books: See all 7,437 items

4. **Beginners Guide to Raising Tropical Fish "From Aquarium Setup to Diseases, Everyt - Kindle eBook**
Buy: $3.87
Auto-delivered wirelessly
★★★★★ (2)
Kindle Store: See all 49 items

5. **500 Freshwater Aquarium Fish: A Visual Reference to the Most Popular Species** by G
Buy new: ~~$29.95~~ $20.59
26 new from $18.50 16 used from $15.99
Get it by **Wednesday, Jan 25** if you order in the next **3 hours** and choose one-day shipping.
★★★★☆ (21)



Search results

search engines

Tweets

People

Top videos

Worldwide trends · Change

#ShowYourHeart Promoted
#30thingsaboutme
#CBB
#yotengounamigoque
Go Frankie
Nicola
Poor Denise
Gareth and Romeo
Danny Wilson
Buy Marry The Night

twitter

© 2012 Twitter About Help Terms Privacy
Blog Status Apps Resources Jobs
Advertisers Businesses Media Developers

Top people · View all

Search Engine Land @sengineland
Follow us for news about Google, Bing, Yahoo, ...

Tweets Top / All

27 new tweets

braden graeber @hipstermermaid 21 Jan
Bing is the Sears of **search engines.**

TechinBiz @techinbusiness 35m
Boost your business website position on the main **search engines** with a link building package #linkbuilding goo.gl/KK68C

Wyoming Entrepreneur @wyendotbiz 50m
I have recently been contacted by a company offering to help my website to be found by **search engines**. Are thes... bit.ly/zrXXAu

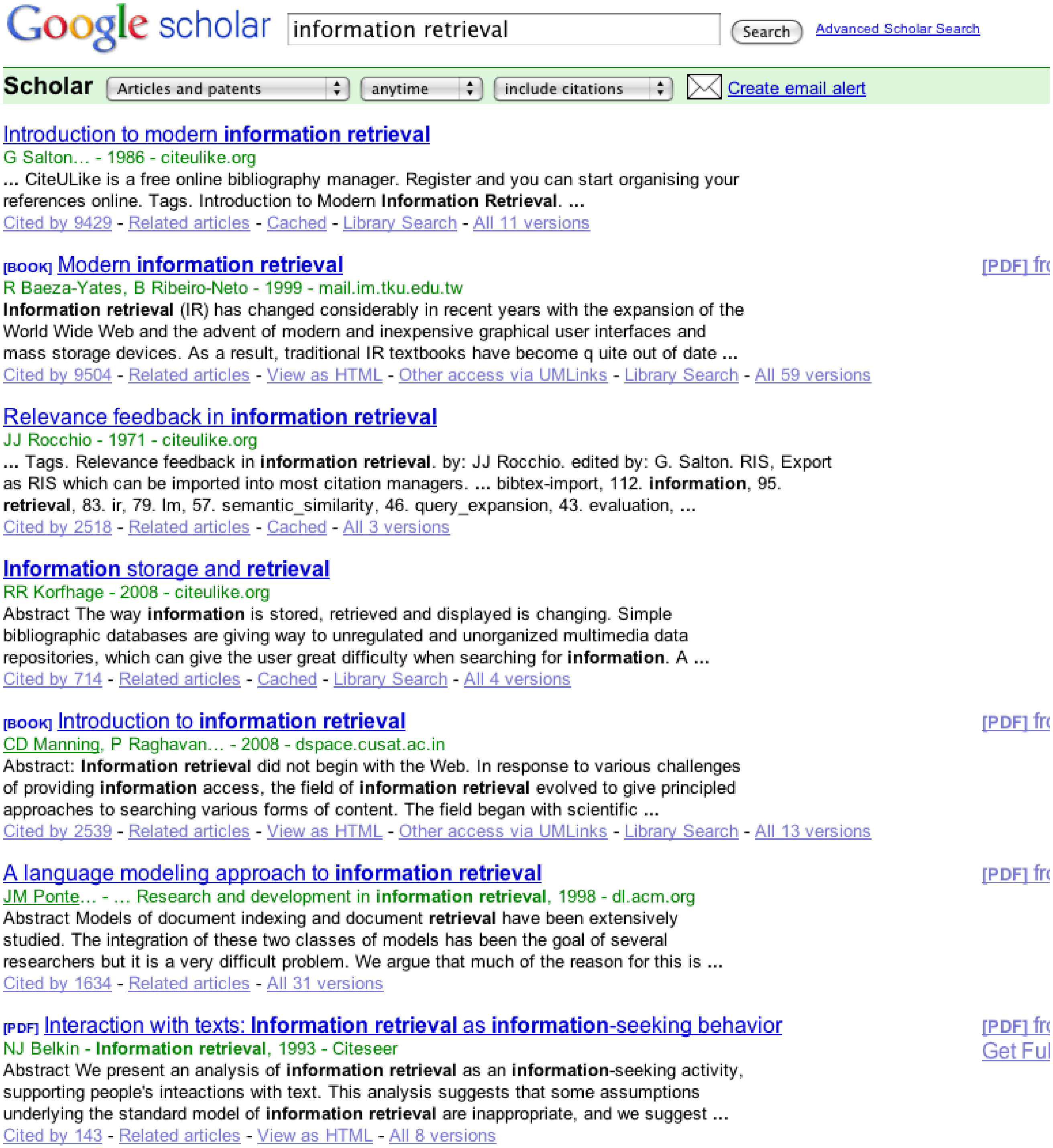
Int Search Summit @IntSearchSummit 55m
Hear From the **Search Engines:** Google, Yandex and Naver among the speakers at International **Search** Summit @ SMX West bit.ly/9O0m4O

productpro @productpro 1h
How to increase link popularity: **Search engines** are the gateway to the Internet; they arethe first tool that pot... bit.ly/yTtlGP

Michael Gray @graywolf 1h
How **Search Engines** Work bit.ly/yglWoh

Stefan Svartling @svartling 2h
Analysis Reveals the Next 'Google' of B2B **Search Engines** - Masterseek.com ... svrt.se/yhBAaf

jopauca @jopauca 2h
I disagree w/ Morozov's prescription regarding fringe movements and **search engines** is.gd/UjUW0x (says he's "toying" w/ the idea).



Google scholar information retrieval Search Advanced Scholar Search

Scholar Articles and patents anytime include citations Create email alert

Introduction to modern **information retrieval**
G Salton… - 1986 - citeulike.org
… CiteULike is a free online bibliography manager. Register and you can start organising your references online. Tags. Introduction to Modern **Information Retrieval**. **…**
Cited by 9429 - Related articles - Cached - Library Search - All 11 versions

[BOOK] Modern **information retrieval** [PDF] fr
R Baeza-Yates, B Ribeiro-Neto - 1999 - mail.im.tku.edu.tw
Information retrieval (IR) has changed considerably in recent years with the expansion of the World Wide Web and the advent of modern and inexpensive graphical user interfaces and mass storage devices. As a result, traditional IR textbooks have become q uite out of date **…**
Cited by 9504 - Related articles - View as HTML - Other access via UMLinks - Library Search - All 59 versions

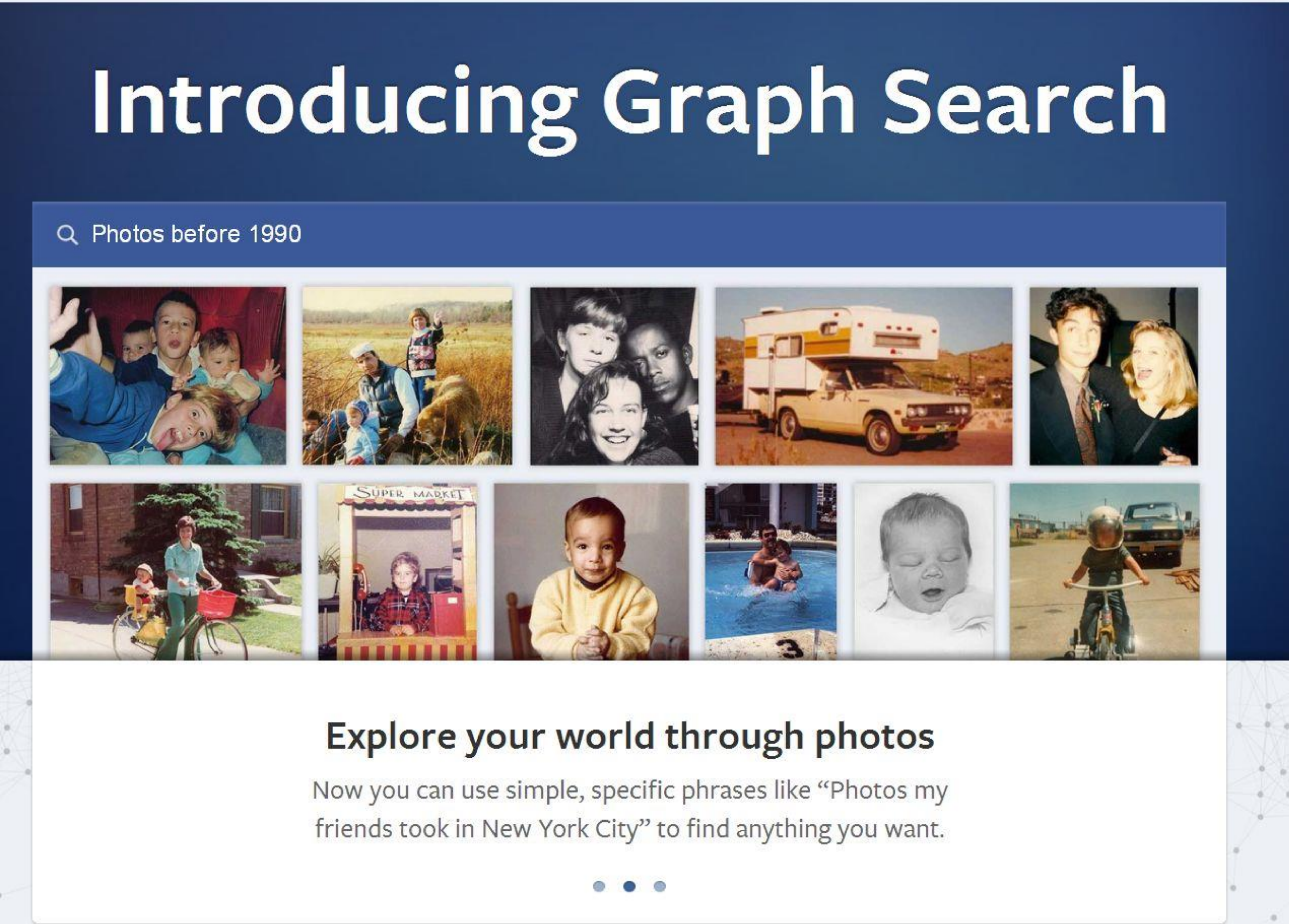
Relevance feedback in **information retrieval**
JJ Rocchio - 1971 - citeulike.org
… Tags. Relevance feedback in **information retrieval**. by: JJ Rocchio. edited by: G. Salton. RIS, Export as RIS which can be imported into most citation managers. **…** bibtex-import, 112. **information**, 95. **retrieval**, 83. ir, 79. lm, 57. semantic_similarity, 46. query_expansion, 43. evaluation, **…**
Cited by 2518 - Related articles - Cached - All 3 versions

Information storage and **retrieval**
RR Korfhage - 2008 - citeulike.org
Abstract The way **information** is stored, retrieved and displayed is changing. Simple bibliographic databases are giving way to unregulated and unorganized multimedia data repositories, which can give the user great difficulty when searching for **information**. A **…**
Cited by 714 - Related articles - Cached - Library Search - All 4 versions

[BOOK] Introduction to **information retrieval** [PDF] fr
CD Manning, P Raghavan… - 2008 - dspace.cusat.ac.in
Abstract: **Information retrieval** did not begin with the Web. In response to various challenges of providing **information** access, the field of **information retrieval** evolved to give principled approaches to searching various forms of content. The field began with scientific **…**
Cited by 2539 - Related articles - View as HTML - Other access via UMLinks - Library Search - All 13 versions

A language modeling approach to **information retrieval** [PDF] fr
JM Ponte… - … Research and development in **information retrieval**, 1998 - dl.acm.org
Abstract Models of document indexing and document **retrieval** have been extensively studied. The integration of these two classes of models has been the goal of several researchers but it is a very difficult problem. We argue that much of the reason for this is **…**
Cited by 1634 - Related articles - All 31 versions

[PDF] Interaction with texts: **Information retrieval** as **information**-seeking behavior [PDF] fr
Get Ful
NJ Belkin - **Information retrieval**, 1993 - Citeseer
Abstract We present an analysis of **information retrieval** as an **information**-seeking activity, supporting people's inteactions with text. This analysis suggests that some assumptions underlying the standard model of **information retrieval** are inappropriate, and we suggest **…**
Cited by 143 - Related articles - View as HTML - All 8 versions



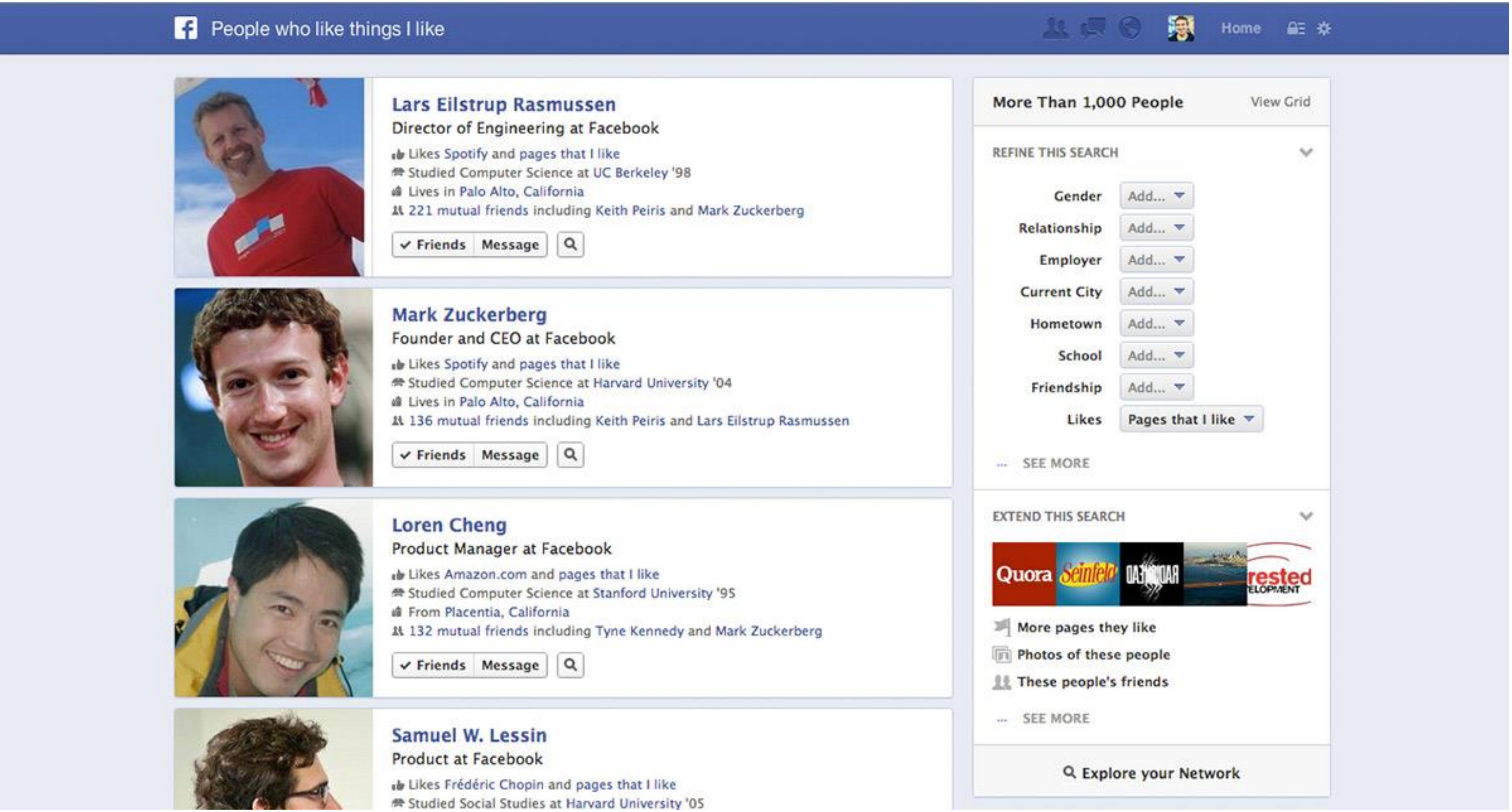
Introducing Graph Search

Photos before 1990

Explore your world through photos

Now you can use simple, specific phrases like “Photos my friends took in New York City” to find anything you want.

○ ● ○



People who like things I like

Lars Eilstrup Rasmussen
Director of Engineering at Facebook
Likes Spotify and pages that I like
Studied Computer Science at UC Berkeley '98
Lives in Palo Alto, California
221 mutual friends including Keith Peiris and Mark Zuckerberg

Mark Zuckerberg
Founder and CEO at Facebook
Likes Spotify and pages that I like
Studied Computer Science at Harvard University '04
Lives in Palo Alto, California
136 mutual friends including Keith Peiris and Lars Eilstrup Rasmussen

Loren Cheng
Product Manager at Facebook
Likes Amazon.com and pages that I like
Studied Computer Science at Stanford University '95
From Placentia, California
132 mutual friends including Tyne Kennedy and Mark Zuckerberg

Samuel W. Lessin
Product at Facebook
Likes Frédéric Chopin and pages that I like
Studied Social Studies at Harvard University '05

More Than 1,000 People

REFINE THIS SEARCH

Gender	Add...
Relationship	Add...
Employer	Add...
Current City	Add...
Hometown	Add...
School	Add...
Friendship	Add...
Likes	Pages that I like

EXTEND THIS SEARCH

More pages they like
Photos of these people
These people's friends

Course Goals

- To help you to understand search engines, evaluate and compare them, and modify them for specific applications
- Provide broad coverage of the important issues in information retrieval and search engines
- Book:
  - “*Search Engines*: Information Retrieval in Practice”
    - Textbook annex and ebook at www.coursesmart.com
    - Kindle version also available (about half price)

Topics

- *Overview*
- Architecture of a search engine
- Data acquisition
- Text representation
- Indexing
- Query processing
- Ranking
- Evaluation
- Classification and clustering

Course Evaluation

- Exams (2): 50%
- Homework assignments (8): 50%
  - Mostly programming
    - Building indexing and search software for sample collections of text
  - Deadlines will be mostly on Fridays
    - Submissions will be accepted over the weekend
      - 30% penalty (if no excuse)

Contact

- Bruce Croft (CS370) and Mostafa Keikha (CS207)
  - Office hours: Set appointment via e-mail or Moodle
  - Questions about course content and exams
  - croft@cs.umass.edu, keikham@cs.umass.edu
- TAs: Jae-Hyun Park, Youngho Kim
  - All programming and assignment grading questions
  - Office hours
    - Jae-Hyun: Mon, 14:00-16:00 (CS207)
    - Youngho: Wed, 14:00-16:00 (CS207)
  - yhkim@cs.umass.edu, jhpark@cs.umass.edu

Homework

Jae-Hyun

Youngho

- Submitted via Moodle
  - pointers to data, updates, etc.
- Textbook website (errata, Galago, etc.)
  - http://search-engines-book.com

Information Retrieval

- *"Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information."* (Salton, 1968)
  - General definition that can be applied to many types of information and search applications
  - Primary focus of IR since the 50s has been on *text* and *documents*

What is a Document?

- Examples:
  - web pages, email, books, news stories, scholarly papers, text messages, tweets, MS Office, PDF, facebook pages, blogs, forum postings, IM sessions, etc.
- Common properties
  - text content
  - some structure (e.g., title, author, date for papers; subject, sender, destination for email)

Documents vs. Database Records

- Database records (or *tuples* in relational databases) are typically made up of well-defined fields (or *attributes*)
  - e.g., bank records with account numbers, balances, names, addresses, social security numbers, dates of birth, etc.
- Easy to compare fields with well-defined semantics to queries in order to find matches
- Text is more difficult

Documents vs. Records

- Example bank database query
  - *Find records with balance > $50,000 in branches located in Amherst, MA.*
  - Matches easily found by comparison with field values of records
- Example search engine query
  - *bank scandals in western mass*
  - This text must be compared to the text of entire news stories

Comparing Text

- Comparing the query text to the document text and determining what is a good match is the <u>core issue</u> of information retrieval
- Exact matching of words is not enough
  - Many different ways to write the same thing in a "natural language" like English
  - e.g., does a news story containing the text *"bank director in Amherst steals funds"* match the query?
  - Some stories will be better matches than others

Dimensions of IR

- IR is more than just text, and more than just web search
  - although these are central
- People doing IR work with different media, different types of search applications, and different tasks

Dimensions of IR

Content	Applications	Tasks
Text	Web search	Ad hoc search
Images	Vertical search	Filtering
Video	Enterprise search	Classification
Scanned docs	Mobile search	Question answering
Audio	Social search	
Music	Desktop search	
	Literature search	

Big Issues in IR

- Relevance
  - What is it?
  - Simple (and simplistic) definition: A relevant document contains the information that a person was looking for when they submitted a query to the search engine
  - Many factors influence a person’s decision about what is relevant: e.g., task, context, novelty
  - *Topical relevance* (same topic) vs. *user relevance* (everything else)

Big Issues in IR

- Relevance
  - *Retrieval models* define a view of relevance
  - *Ranking algorithms* used in search engines are based on retrieval models
  - Most models based on statistical properties of text rather than linguistic
    - i.e. counting simple text features such as words instead of parsing and analyzing the sentences

Big Issues in IR

- Evaluation
  - Experimental procedures and measures for comparing system output with user expectations
  - IR evaluation methods now used in many fields
    - e.g. Netflix competition
  - Typically use *test collection* of documents, queries, and relevance judgments
    - Most commonly used are TREC collections
  - *Recall* and *precision* are two examples of <u>effectiveness</u> measures

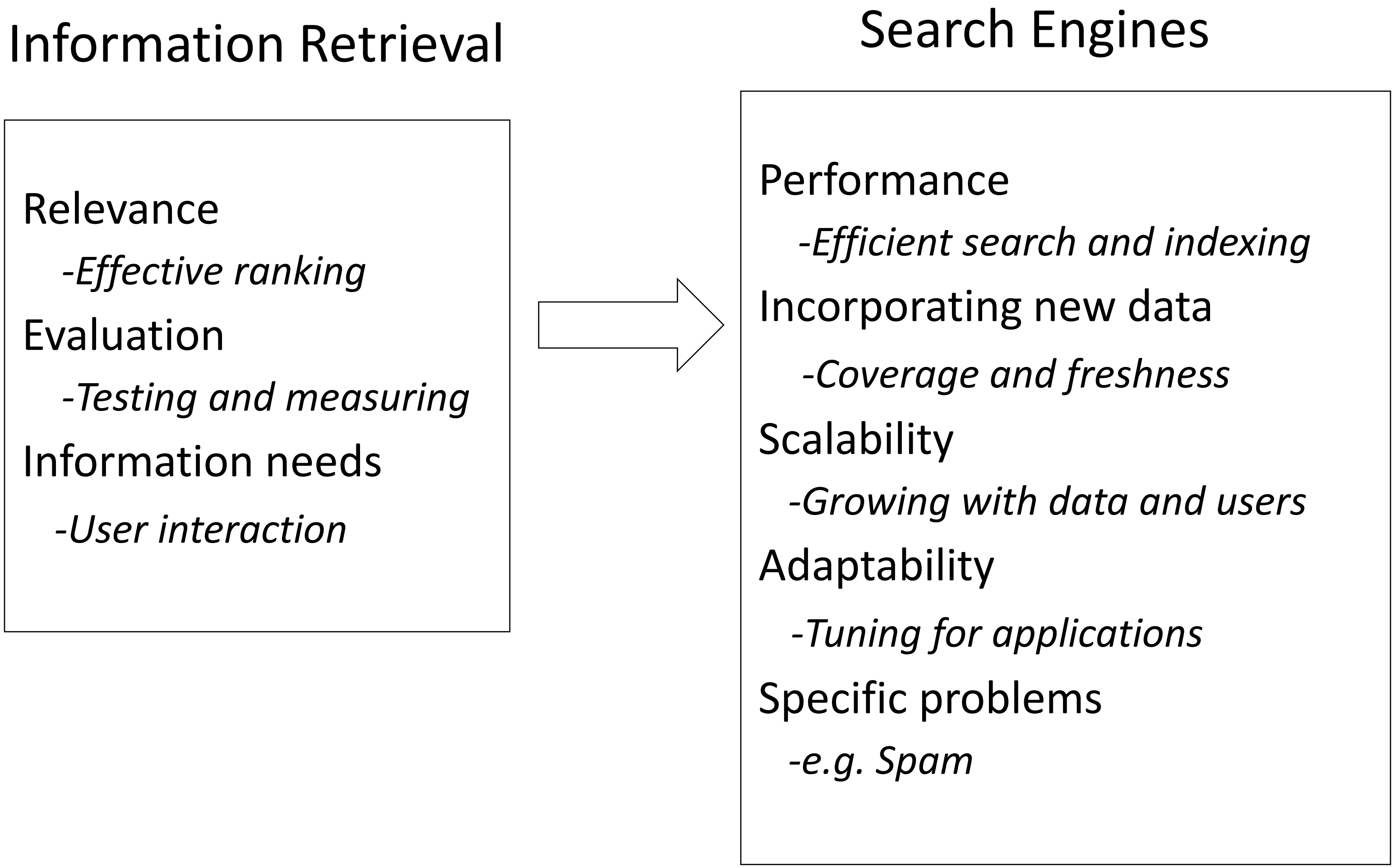
Big Issues in IR

- Users and Information Needs
  - Search evaluation is user-centered
  - Keyword queries are often poor descriptions of actual information needs
  - Interaction and context are important for understanding user intent
  - Query refinement techniques such as *query expansion*, *query suggestion*, *relevance feedback* improve ranking

IR and Search Engines

- A search engine is the practical application of information retrieval techniques to large scale text collections
- Web search engines are best-known examples, but many others
  - *Open source* search engines are important for research and development
    - e.g., Lucene, Lemur/Indri, Galago
- Big issues include main IR issues but also some others

IR and Search Engines



Information Retrieval

Relevance

-Effective ranking

Evaluation

-Testing and measuring

Information needs

-User interaction

Search Engines

Performance

-Efficient search and indexing

Incorporating new data

-Coverage and freshness

Scalability

-Growing with data and users

Adaptability

-Tuning for applications

Specific problems

-e.g. Spam

Search Engine Issues

- Performance
  - Measuring and improving the efficiency of search
    - e.g., reducing *response time*, increasing *query throughput,* increasing *indexing speed*
  - *Indexes* are data structures designed to improve search efficiency
    - designing and implementing them are major issues for search engines

Search Engine Issues

- Dynamic data
  - The “collection” for most real applications is constantly changing in terms of updates, additions, deletions
    - e.g., web pages
  - Acquiring or “crawling” the documents is a major task
    - Typical measures are *coverage* (how much has been indexed) and *freshness* (how recently was it indexed)
  - Updating the indexes while processing queries is also a design issue

Search Engine Issues

- Scalability
  - Making everything work with millions of users every day, and many terabytes of documents
  - Distributed processing is essential
- Adaptability
  - Changing and tuning search engine components such as ranking algorithm, indexing strategy, interface for different applications
  - Exploiting huge sources of user data such as query logs, locations, social networks

Search Engine Issues

- Spam
  - For web search, spam in all its forms is one of the major issues
  - Affects the efficiency of search engines and, more seriously, the effectiveness of the results
  - Many types of spam
  - New subfield called *adversarial IR*, since spammers are “adversaries” with different goals

Topics

- Overview
- *Architecture of a search engine*