CS6200
Information Retrieval

David Smith
College of Computer and Information Science
Northeastern University
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.
gofishin.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) 527-3381
   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.

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.capharry.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
**fish supplies**

- **Fish Supplies**
  - www.PetSolutions.com/Fish_Supplies
    - Wide Selection Of fish supplies: Great Prices & Fast Shipping Here!

- **Fish Supplies**
  - www.hoskingnursery.net
    - Professional Landscape Design With Plants, Pond & Fountains. Call Now!

- **Cabela's - Fishing**
  - www.Cabelas.com
    - Cabela's: Free Shipping When You Choose In-Store Pick Up At Checkout

- **Fish And Supplies**
  - BizRate.com
    - Bargain Prices. Smart Deals. Explore Fish And Supplies!

---

**fish supply**

- New-Hydror Performer Protein Skimmer Price starts at $349.99. Hydor's New PERFORMER skimmer range is high quality product designed by Hydor. The PERFORMER skimmers are equipped with ...

  - fishsupply.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 ...

  - www.fish.com
    - 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 ...

  - www.aquariumguys.com
    - Cached page

- **Fish Supplies - Fish Products and Fish Tanks from PETCO.com**
  - Fish Supplies - Fish Products and Fish Tanks Available Online from PETCO.com. Find the fish tanks and fish products you need to enhance the beauty of your aquarium. Check out ...

  - www.petco.com/Fish-Home.aspx
    - Cached page

- **Discount Online Fish, Aquarium, Supplies, Compare Prices**
  - Compare Prices: 50,000 pet, fish, aquarium, freshwater, saltwater, products, supply supplies, accessories, equipment, pumps, filters, tanks, food.

  - www.chiapetsstore.com/Fish-Aquariums
    - Cached page

---

**Shop for fish supplies**

- **Red Sea Fish pHarm Deluxe from $43.74**
  - bing.com/shopping

- **Red Sea Fish Pharm Ltd Sea from $5.80**
  - bing.com/shopping

- **TopDawg Pet Supply Discus from $74.19**
  - bing.com/shopping

- **Fish $14.00**
  - bing.com/shopping
Tropical Fish
Find complete details about Tropical Fish from Aqua-Plus Inc. You may also find variety of Tropical Fish from other Suppliers and Manufacturers at...
www.gongchang.com/Tropical_Fish-dp8 2012-8-12 - 百度快照

Tropical Fish
Tropical Fish 1 through 10 of 78 下一个 > 1 2 3 ... 8 ... Currently, we have 3 fish and my older brother mostly take care of them...
zh.englishbaby.com/forum/ ... /139467 2012-8-31 - 百度快照

tropical_fish CSDN 下载频道
正在浏览用户"tropical_fish"发布的资源查看tropical_fish的所有资源 有10人下载... Tag：ANSYS 用户: tropical_fish 发布日期：2009-02-25有28人下载 21天...
tropical_fish.download.csdn.net/ 2009-7-29 - 百度快照

Tropical Fish (film) - Wikipedia, the free encyclopedia
Tropical Fish (Chinese: 热带鱼; pinyin: Rèdài Yú; Pe h-œ-ji: Jia t-tài H ) is a 1995 Taiwanese comedy-drama film written and...
en.wikipedia.org/ ... _Fish_(film) 2012-8-19 繁体 - 百度快照

Tropical Fish - Peter Stadelmann - Google 图书
Aquarium owners will find valuable information on maintaining correct water chemistry and temperature, and creating an optimal environment for tropical fish ...
books.google.com.hk/ ... Sh5&hl=zh-CN 2012-7-20 - 百度快照

tropical_fish_百度词典
[词典释义] ph. 1. 热带鱼
[网络释义] 1.热带鱼
dict.baidu.com/s?wd=tropical fish 2012-9-6

AC Tropical Fish & Aquarium
AC Tropical fish & aquarium features a large number of aquarium articles, species profiles and other resources for all those who keep tropical fish in ...
www.aquaticcommunity.com/ 2012-8-31 - 百度快照
Search results

Tweets

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
Bing is the Sears of search engines.

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

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

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/BQ0m4O

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

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

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

jopauc@jopauc
2h
I disagree w/ Morozov’s prescription regarding fringe movements and search engines is.gd/UjJW6x (says he’s “toying” w/ the idea).
Who am I? Why am I tweeting from Wheat Thins? Find out. Do it.
youtu.be/DIFG34xDoI

@WheatThins

Promoted by Wheat Thins
View video

D&T Pet Supplies @dtpets

We now carry supplies for your dogs, cats, birds, fish, and small animals
dtpetsupplies.com/catalog/

58m

Dale Wall @asseenontxyguy
tinyurl.com/6leakx5 Before buying fish aquarium supplies, consider the size of the tank and the size of the aquarium.

5h

LiveAquaria @LiveAquaria

Nice assortment of Aquariums on sale right now! Looking for a new QT or new Nano? Maybe a nice little desk...

5h

tkquincy @SunnyFishTank

USED 55 GALLON FISH TANK | eBay: USED 55 GALLON FISH TANK in Pet Supplies, Aquarium Fish, Aquariums | eBay.

21h

Aaron Patterson @tenderlove

Picking up sausage supplies @ Pike Place Fish instagr.am/p/jqb9d/
View photo

22 Jan

Josh Toeb @tobeman

That-Fish-Place-That-Pet-Place: Discount Dog Supplies from the Largest Pet Store In the World: That Pet Place
dlvr.it/16bnjp

22 Jan

eBC Pets Northeast @eBC_Pets_NE

Dutch Country: 9 Aquariums(10-75 gal), fish, and supplies - Please Contact (Mount Wolf)
bit.ly/xCPHAK #eBC #Pets

21 Jan
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

Modern information retrieval
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 quite out of date ...
Cited by 6524 - 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. ... bibtx-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

Introduction to information retrieval
CD Manning, P Raghavan... - 2008 - dspace.cuset.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
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

Interaction with texts: Information retrieval as information-seeking behavior
NJ Belkin - Information retrieval, 1993 - Citeseer
Abstract: We present an analysis of information retrieval as an information-seeking activity, supporting people's interactions 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
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
• Books (Kindle editions OK):
  – *Search Engines: Information Retrieval in Practice*
    • Croft, Metzler, and Strohman
  – *Introduction to Information Retrieval*
    • Manning, Raghavan, and Schütze
  – Not required, but readings will be recommended
Topics

- Overview
- Architecture of a search engine
- Data acquisition
- Text representation
- Information extraction
- Indexing
- Query processing
- Ranking
- Evaluation
- Classification and clustering
- Social search
- More...
Course Evaluation

• Six assignments (50%)
  – All are worth equal points; lowest-scoring assignment dropped
  – Some require written answers and some short programming exercises.
  – Others involve more extensive programming and a short written report on your design choices and experimental results.
  – You may use any programming language installed on college Linux machines, so we can (compile and) run your code.
  – You may save work by using reasonable libraries that don’t simply implement the goal of the assignment
    • E.g., HTTP request libraries are OK; web-crawling libraries are not.

• Two exams
  – Midterm (20%) in class
  – Final (30%)
Late Policy

• Assignments are due at the beginning of class on the announced due date.
• You may take a single homework extension of four calendar days, to be used at your discretion.
• After the first late assignment, unexcused late assignments will be penalized 10% per calendar day late. We normally will not accept assignments after the date on which the following assignment is due or after the solutions have been handed out, whichever comes first.
• If you are in circumstances that would cause you to turn in an assignment late, please contact me in advance to ask for an extension for cause.
Academic Honesty

• All work submitted for credit must be your own.
• You may discuss the assignments with your classmates, the TA, and the instructor. You must acknowledge the people with whom you discussed your work, and you must write up your own solutions.
• Any written sources used (apart from the text) must also be acknowledged; however, you may not consult any solutions from previous years’ assignments whether they are student or faculty generated.
Contact

• Me
  – Office hours: Thursdays 3–5, Room 356
  – dasmith@ccs.neu.edu

• TAs
  – TBA

• Course website and Piazza list
  – http://www.ccs.neu.edu/course/cs6200f15/
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 Somerville, 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 core issue 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

<table>
<thead>
<tr>
<th>Content</th>
<th>Applications</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Web search</td>
<td>Ad hoc search</td>
</tr>
<tr>
<td>Images</td>
<td>Vertical search</td>
<td>Filtering</td>
</tr>
<tr>
<td>Video</td>
<td>Enterprise search</td>
<td>Classification</td>
</tr>
<tr>
<td>Scanned docs</td>
<td>Desktop search</td>
<td>Question answering</td>
</tr>
<tr>
<td>Audio</td>
<td>Forum search</td>
<td>Summarization</td>
</tr>
<tr>
<td>Music</td>
<td>P2P search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Literature search</td>
<td></td>
</tr>
</tbody>
</table>
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 users’ decision about what is relevant: e.g., task, context, novelty, style, other documents they’ve already read
  – *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 deep linguistic analysis
    • 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 *effectiveness* measures
• 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
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
  – Proliferation of spam varieties
    • e.g. spamdexing or term spam, link spam, “search engine optimization”
  – New subfield called adversarial IR, since spammers are “adversaries” with different goals
Topics

• Overview

• *Architecture of a search engine*

• For background, read chapters 1 and 2 of *Search Engines* by Croft, Metzler, and Strohman