CS6200/IS4200 Information Retrieval

David Smith
Khoury College of Computer Sciences
Northeastern University

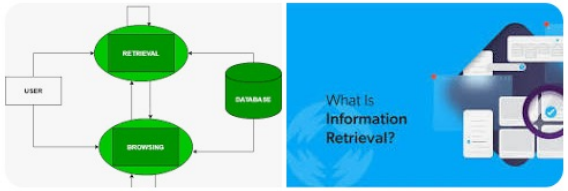
information retrieval

All Images Videos News Books Shopping Forums More Tools

Information retrieval is **the process of accessing data resources**. Usually documents or other unstructured data for the purpose of sharing knowledge. More specifically, an information retrieval system provides an interface between users and large data repositories – especially textual repositories. May 15, 2024

Coveo
https://www.coveo.com

What Is Information Retrieval? - Coveo



About featured snippets • Feedback

Wikipedia
https://en.wikipedia.org

Information retrieval

Information retrieval (IR) in computing and information science is the task of identifying and retrieving information system resources that are relevant to ...

Stanford University
https://nlp.stanford.edu

Introduction to Information Retrieval - Stanford NLP Group

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 ...

Things to know

Purpose

LinkedIn

What is Information Retrieval System? - LinkedIn

How Does the Information Retrieval System Work? Key components of an ... do in Excel, Sheets, or Docs. So, for search...

Alltius

What is Information Retrieval? | Alltius Glossary

Whether it's finding the answer to a simple question or conducting in-depth research, information retrieval systems

Information retrieval

Information retrieval in computing and information science is the task of identifying and retrieving information system resources that are relevant to an information need. The information need can be specified in the form of a search query. Wikipedia

People also search for


Artificial intelligence


Machine translation


Information

Machine learning

Feedback

Google

information retrieval

All | Books | News | Images | Videos | More | Settings | Tools

About 59,800,000 results (0.85 seconds)

Dictionary

Search for a word

in·for·ma·tion re·triev·al

/ˌinfərˈmāSHən rəˈˌtrēvəl/

noun COMPUTING

the tracing and recovery of specific information from stored data.
"an information retrieval system"

Translations, word origin, and more definitions

From Oxford *Feedback*

Information retrieval - Wikipedia

https://en.wikipedia.org › wiki › Information_retrieval

Information retrieval is the science of searching for **information** in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds.

Information retrieval · Evaluation measures · Applications

People also ask

- What is information retrieval services?
- How do you perform information retrieval?
- What is the goal of information retrieval?
- What is Information Retrieval System PDF?

Feedback

Introduction to Information Retrieval - Stanford NLP Group

https://nlp.stanford.edu › IR-book › html › htmledition › irbook

Website: http://**informationretrieval**.org/. Cambridge ... **informationretrieval** (at) yahoogroups.com ... Statistical properties of terms in **information retrieval**.

[PDF] Introduction to Information Retrieval - Stanford NLP Group

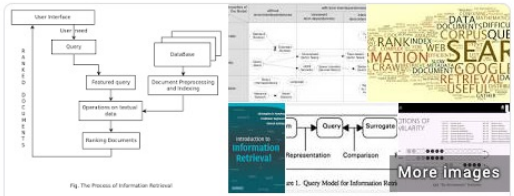
https://nlp.stanford.edu › IR-book › pdf

As defined in this way, **information retrieval** used to be an activity that only a few people engaged in: reference librarians, paralegals, and similar pro- fessional ...

Information Retrieval - an overview | ScienceDirect Topics

https://www.sciencedirect.com › topics › computer-science › information-ret...

Information retrieval theorists verified the importance of information-seeking research in informing **information retrieval** design. Marchionini (1995) observed ...



Information retrieval

Information retrieval is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Wikipedia

People also search for

View 10+ more

Research | Informati... | Psychology | Semantics | Video

Feedback

information retrieval

Privacy, simplified.

Web Images Videos News Maps | Definition Settings

All Regions Safe Search: Moderate Any Time

Information Retrieval and Mining Massive Data Sets (AD)
udemy.com/Online-Courses/Bing-Promotion | Report Ad
Udemy Helps You Gain The Skills You Need To Achieve Your Goals! Start Today

Udemy for Business
Prepare Your Team, Stay Ahead
3,000+ Curated High-Quality...

Teach the World Online
Share Your Knowledge, Make
Money Reach Students Across Th...

Design Courses
Discover Top Courses On Web
Design Graphic Design, UX And...

Browse All Courses
Find The Right Course For You.
Over 100,000 High-Quality...

Information retrieval - Wikipedia
https://en.wikipedia.org/wiki/Information_retrieval
Information retrieval is the science of searching for **information** in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds. Automated **information retrieval** systems are used to reduce what has been called **information** overload. An IR system is a ...

Information Retrieval | Definition of Information Retrieval ...
https://www.merriam-webster.com/dictionary/information retrieval
Information retrieval definition is - the techniques of storing and recovering and often disseminating recorded data especially through the use of a computerized system.

Information retrieval | computer and information science ...
https://www.britannica.com/technology/information-retrieval
Information retrieval, Recovery of **information**, especially in a database stored in a computer. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links.

Information Retrieval | Article about Information Retrieval ...
https://encyclopedia2.thefreedictionary.com/Information+Retrieval
Information retrieval must be distinguished from logical **information** processing, without which direct replies to the questions posed by a human being is impossible. In **information retrieval**, only the **information** that was input to the **information retrieval** system is sought—only that **information** can be found.

Introduction to Information Retrieval - nlp.stanford.edu
https://nlp.stanford.edu/IR-book/html/htmledition/irbook.html
Dictionaries and tolerant **retrieval**; Index construction; Index compression; Scoring, term weighting and the vector space model; Computing scores in a complete search system; Evaluation in **information retrieval**; Relevance feedback and query expansion; XML **retrieval**; Probabilistic **information retrieval**; Language models for **information retrieval**

Information Retrieval System - Library & Information Science ...
www.lisbdnet.com/information-retrieval-syste/
Information Retrieval system is a part and parcel of communication system. The main objectives of **Information retrieval** is to supply right **information**, to the hand of right user at a right time. Various materials and methods are used for retrieving our desired

Information retrieval

Information retrieval is the activity of obtaining information system resources relevant to an information need from a collection of information resources. Searches can be based on full-text or other content-based indexing. **Wikipedia**

Feedback

Job **Information Retrieval** System | 53 urgent openings. Apply now (AD)
us.jobrapido.com/Job Information Retrieval System/Jobs
Find the job you want! All latest vacancies in the US listed on Jobrapido™

tropical fish 百度一下 百度首页

网页 资讯 视频 图片 知道 文库 贴吧 采购 地图 更多»

百度为您找到相关结果约446,000个 搜索工具

tropical fish_百度翻译

tropical fish

英 [ˈtrɒpɪkl fɪʃ] 美 [ˈtrɑːpɪkl fɪʃ]

网络 热带鱼; 熱帶魚; 热带观赏鱼; 带鱼;

[例句] I ate all the tropical fish in the aquarium.
我把鱼缸里的热带鱼吃光了。

进行更多翻译

fanyi.baidu.com

tropical fish_视频大全_高清在线观看

tropical fish热带鱼类,海底世...
好看视频

少儿英语:Tropical Fish美丽可...

碰碰狐儿歌合集：第348集 tr...
爱奇艺

英语动物儿歌 第二季：18 Tro...

动物英语单词儿歌《Tropical F...
播视网

热带鱼 Tropical Fish 电影
腾讯视频

碰碰狐 英语动物儿歌 第2季 第...
爱奇艺

◆折纸大全◆怎样折纸热带鱼 T...

查看更多视频>>

tropical fish_百度图片

image.baidu.com - 查看全部6,309张图片

Tropical Fish

Tropical fish profiles, aquarium fish and reef care information, freshwater and saltwater fish discussion forums, and aquarium product reviews.

https://www.tropicalfishkeepin... - 百度快照 - 翻译此页

tropical fish在线试听_高音质歌曲_网易云音乐

网易云音乐 酷我音乐 酷狗音乐 QQ音乐 虾米音乐

全选	播放选中歌曲		播放	歌词
01	tropical fish	squaaks		
02	tropical fish	mystery dates ...		
03	tropical fish	gong global fa...		
04	tropical fish	かわさき みれい		
05	tropical fish	gong		

相关生物 展开


草虾


螃蟹


河马


仓鼠


刺猬


蜗牛


大象


郁金香

相关人物 展开


jellyfish


tank


rainbow


帕查拉·奇拉锡瓦特

搜索热点 换一换

1	津巴布韦总统去世 新	1282万
2	中国男篮对战韩国	1200万
3	中国男篮打排位赛	827万
4	蜘蛛侠离开漫威	816万
5	LPL夏季赛总决赛	813万
6	秦海璐回应怼黄晓明 新	800万
7	上海迪士尼松口 新	754万
8	FPX夺冠	726万
9	北京年平均工资	713万
10	法国打鸣公鸡胜诉 新	406万

查看更多>>

Microsoft Bing

Where can I grow black currants in the US?

dasmith@northeastern.edu 15

SEARCH CHAT SCHOOL IMAGES VIDEOS MAPS NEWS SHOPPING MORE

10,500,000 Results Any time

See Where Can I Grow Black Currants In the Us?

Ac

1 - 10 Consort Black Curra…
$19.95
Etsy
★★★★★ 5K+

Consort Black Currant - 1…
$49.79 ~~59.99~~
Nature Hills …

Black Currant (Ribes…
$6.86
Etsy

Wild Black Currant See…
$8.97
Etsy
★★★★½ 1K+

Black Currant Shrub - 8 Oz.
$13.99
Walmart

European Black Curra…
$25.00
Etsy
★★★★½ 54

Lot 5 Blackcurrant…
$19.00
Etsy
★★★★★ 4

25 Consort Black Curra…
$4.50
Etsy
★★★★½ 1K+

10+ Pre Stratified…
$5.50
Etsy
★★★★★ 1K+

Swi
$4.9
Etsy
★★★

Black currants can be grown in a variety of habitats, including stream banks, moist ravines, wet meadows, floodplains, and woodland edges [1]. They prefer a full sun position and grow best when given shelter from strong hot winds [2]. Black currants can be grown in New York, Connecticut, Oregon, and Vermont [3] [4]. They can be grown as a foundational hedge, in a shady or unused spot, in a wildlife hedge, in a poultry foraging area, at the edges of open woods or in dappled shade in the woods, in a pollination garden, in a medicinal garden, or in a jelly garden [5].

Learn more: 1. plants.usda.gov 2. gardenexpress.com.au 3. foodtolive.com 4. wikipedia.org 5. tenthacrefarm.com — see less

Feedback

Black Currants: Varieties, Growing Guide, Care, …

from morningchores.com

From this page

1 **Where are black currants native to?**

YOU.com

Chat | All | Images | Videos | News | Maps | Social | More

Where can I grow black currants in the US?

To grow black currants in the United States, you can consider the following areas:

1. Northeast region: Black currants are grown by farmers in the Northeast, including states like New York and Connecticut [1]. In fact, New York lifted the ban on black currants in 2006 [1], allowing farmers to cultivate them again.
2. Pacific Northwest region: Black currants are also grown in the Pacific Northwest, which includes states like Oregon. This region has favorable climate conditions for black currant cultivation.

It's worth noting that while the ban on black currants has been lifted at the federal level, there may still be regulations in certain states regarding the cultivation of black currants. Therefore, it's advisable to check the specific regulations in your state before growing black currants.

planetnatural [1]

Is it legal in Massachusetts?

According to the search results, the cultivation of black currants is currently prohibited in Massachusetts under plant quarantine regulations. **As of this time, black currants or any hybrids of black currants [1], including jostaberries, are not permitted to be grown in Massachusetts** [^1^]. The ban on black currants in Massachusetts is primarily due to their role as a host for the white pine blister rust [2], which is a threat to the timber industry [^5^].

It's important to note that regulations regarding the cultivation of black currants can change over time, so it's advisable to check with the relevant authorities or local agricultural extension offices for the most up-

+ Wikipedia

black currant growing laws | black currant cultivation legal

Ask a follow-up... (⌘ + K)

Standard

1-48 of over 10,000 results for "tropical fish"

Sort by: Featured

Amazon Prime
- [ ] prime

Eligible for Free Shipping
- [ ] Free Shipping by Amazon

All customers get FREE Shipping on orders over $25 shipped by Amazon

Pantry
- [ ] prime pantry

Local Stores
- [ ] Amazon Fresh

Department
- Kindle Store
  - Fish & Aquarium Care
  - Crafts, Hobbies & Home
- Pet Supplies
  - Fish Food
- Books
  - Animal & Pet Care
  - Fish & Aquarium Care
  - See more
- Toys & Games
  - Miniature Novelty Toys
- See All 33 Departments

Avg. Customer Review
- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up
- ★☆☆☆☆ & Up

Brand
- [ ] U.S. Toy
- [ ] TROPICAL FISH
- [ ] MoreBuyBuy
- [ ] Imperial Tropicals
- [ ] TROPICAL FISH LIQUID SOLAR BLANKET
- [ ] Ken's Fish
- [ ] LX
- [ ] 3dRose
- [ ] Fun Express
- [ ] Jolee's Boutique
- [ ] RJC
- [ ] Tervis
- [ ] Bandito Tropical Fish Collection
- [ ] Charlotte International
- [ ] Cool-Patches

Subscribe & Save
- [ ] Subscribe & Save Eligible

Fish Type
- [ ] Tropical
- [ ] Marine

Craft & Hobby Book Format

Search for tropical fish in:

Live Fish & Aquatic Pets

Fish Food

Live Aquarium Plants

Aquariums

Aquarium Décor Ornaments

Amazon's Choice — Customers also shopped Amazon's Choice for...

"tropical fish"

TetraMin Plus Tropical Flakes, Cleaner and Clearer Water Formula
★★★★½ 407
$6.49
Save 5% more with Subscribe & Save
prime

"tropical fish decorations"

Kristin Paradise 30Ct Tropical Fish Hanging Swirl Decorations, Under T...
★★★★★ 15
$12.99
prime

"tropical fish food"

API Fish Food Flakes, Formulated to Help Fish More readily use nutrients...
★★★☆☆ 316
$12.18 ~~$13.99~~
Save 5% more with Subscribe & Save
prime

"tropical fish supplies"

Tetra Blood Worms Freeze Dried Treat
★★★★½ 901
$3.33 ~~$6.99~~
Save 5% more with Subscribe & Save
prime

Google Scholar

information retrieval

Articles

About 8,470,000 results (0.60 sec)

Any time
Since 2024
Since 2023
Since 2020
Custom range...

Sort by relevance
Sort by date

Any type
Review articles

include patents
include citations

Create alert

[PDF] Modern **information retrieval**: A brief overview [PDF] academia.edu
A Singhal - IEEE Data Eng. Bull., 2001 - academia.edu
… the key advances in the field of **Information Retrieval**, and a description of … **information**-ridden world. With exponential growth in the amount of **information** available, **information retrieval** …
Save Cite Cited by 2687 Related articles All 29 versions

Information retrieval on the web [PDF] acm.org GetIt@Northeastern
M Kobayashi, K Takeda - ACM computing surveys (CSUR), 2000 - dl.acm.org
… historical development of **information retrieval** is … **information** available on the Internet, and the growth in users. In the second section we present tools for Web-based **information retrieval**…
Save Cite Cited by 1026 Related articles All 21 versions Web of Science: 258

[BOOK] Introduction to modern **information retrieval**
GG Chowdhury - 2010 - books.google.com
… **information retrieval**. While the original aim of this book – to provide a blend of traditional and new approaches to **information retrieval** … developments – in **information retrieval**. The book …
Save Cite Cited by 1121 Related articles All 6 versions

[BOOK] Readings in **information retrieval**
KS Jones, P Willett - 1997 - books.google.com
… searching bibliographic records, **information retrieval**, or IR, … basis for the design of **retrieval** methods for future systems not … , focusing on the written document **retrieval** task; but because …
Save Cite Cited by 1054 Related articles All 3 versions

Related searches

private information retrieval
semantic information retrieval
information retrieval **query**
relevance information retrieval
rank information retrieval
cross-language information retrieval
introduction to modern information retrieval
learning to rank information retrieval

[BOOK] **Information retrieval** evaluation
D Harman - 2011 - books.google.com
… played a major role in **information retrieval**, with the early … The **retrieval** community has been extremely fortunate to … of the early evaluation of **information retrieval** systems, starting with …
Save Cite Cited by 201 Related articles All 5 versions

Combining approaches to **information retrieval** [PDF] psu.edu
WB Croft - … from the Center for Intelligent **Information Retrieval**, 2000 - Springer
… **retrieval** models that have been proposed as formal frameworks for combination. We show that combining approaches for **information retrieval** … that a new approach to **retrieval** based on …
Save Cite Cited by 319 Related articles All 8 versions

The contributions of Melvil Dewey and Belle da Costa Greene represent two distinct but complementary aspects of early library cataloging in the United States. Dewey's systematic approach to general library collections provided a foundation for the modern organization of knowledge, while Greene's work emphasized the importance of detailed, contextual cataloging for rare and unique materials. Together, they helped shape the standards and practices that are still in use in libraries today.

Please list some sources where I can find out more about their work.

Certainly! Here are some sources where you can learn more about the work of Melvil Dewey and Belle da Costa Greene:

Sources on Belle da Costa Greene

1. **Books:**
   - **"An Illuminated Life: Belle da Costa Greene's Journey from Prejudice to Privilege"** by Heidi Ardizzone: This biography provides an in-depth look at Greene's life, her role at the Pierpont Morgan Library, and her contributions to the world of rare books and manuscripts.
   - **"The Personal Librarian"** by Marie Benedict and Victoria Christopher Murray: A historical novel based on Belle da Costa Greene's life, offering a blend of fact and fiction that illuminates her work and legacy.
2. **Articles and Journals:**
   - **"Belle da Costa Greene: The World's Librarian"** in *The Atlantic*: This article explores Greene's life, career, and impact on the Pierpont Morgan Library.
   - **"Cataloging the Morgan: Belle da Costa Greene and the Organization of a Collection"** in *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage*: This scholarly article discusses Greene's work in cataloging and organizing the Morgan Library's collection.

???

"Cataloging the Morgan: Belle da Costa Greene and the Organizatior

All Images Videos Shopping Forums News Web More Tools

No results found for **"Cataloging the Morgan: Belle da Costa Greene and the Organization of a Collection"**.

Results for **Cataloging the Morgan: Belle da Costa Greene and the Organization of a Collection** (without quotes):

Videos

A Librarian Like No Other: Belle da Costa Greene and Self ...

YouTube · Rare Book School
Jun 28, 2024

The Women Who Made the Morgan: Belle da Costa Greene ...

YouTube · The Morgan Library & Museum
Apr 23, 2021

10 key moments in this video

Belle da Costa Greene and the Women of the Morgan

YouTube · The Morgan Library & Museum
Sep 27, 2022

Belle da Costa Greene: a Librarian's Legacy

YouTube · National Collaborative for Women's History Sites
Mar 30, 2023

10 key moments in this video

Feedback

View all →

The Morgan Library & Museum
https://www.themorgan.org

Belle da Costa Greene: A Librarian's Legacy

The **Morgan** Library & Museum will present a major exhibition devoted to the life and career of its inaugural director, **Belle da Costa Greene** (1879–1950).

The Morgan Library & Museum
https://www.themorgan.org

Belle da Costa Greene, the Morgan's First Librarian and ...

Belle da Costa Greene (1879–1950) was one of the most prominent librarians in American history. She ran the **Morgan** Library for forty-three years.

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
- Readings from recommended books:
  - *Search Engines: Information Retrieval in Practice*
    - Croft, Metzler, and Strohman
  - *Introduction to Information Retrieval*
    - Manning, Raghavan, and Schütze
  - *Conversational Information Seeking*
    - Zamani, Trippas, Dalton, and Radlinski

Topics

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

Course Evaluation

- Five assignments (10% each of course grade)
  - Mostly programming and a short written report on your design choices and experimental results.
  - Written answers and code must be your individual work.
  - Programming assignments will use notebooks with python starter code, via GitHub Classroom
    - Get a GitHub main account (*not* a Khoury hosted GitHub account)
  - 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.
  - Some differences in questions for IS4200

Course Evaluation

- One course project (30%)
  - Working individually or in teams
  - Designing and evaluating a new IR task
  - Final report due December 13
- One exam in class probably on November 7
  - Midterm (20%) in class
  - Focus on evaluating and thinking critically about retrieval models
  - Some differences in questions for IS4200

Late Policy

- Assignments are due at the announced time (usually 11:59pm)
- You may take a single homework extension of four calendar days, no questions asked. Mention this when turning in.
- 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 the instructor *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 textbooks) must also be acknowledged; however, you may not consult any solutions from previous years' assignments whether they are student or faculty generated.

Contact

- Me
  - Virtual office hours:
    - TBA; or by appt.
  - dasmith@ccs.neu.edu
- TAs:
  - TBA
- See Canvas and Piazza for Zoom links and other announcements

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, social posts, slide decks, PDF, group pages, blogs, forum postings, chat messages, etc.
- Common properties
  - recorded content (text or other media)
  - some structure (e.g., title, author, date for papers; subject, sender, destination for email) relating content to other documents

Mlle SUZANNE BRIET

QU'EST-CE QUE LA DOCUMENTATION ?

1951
ÉDIT
ÉDITIONS DOCUMENTAIRES
INDUSTRIELLES ET TECHNIQUES
17, Rue de Grenelle, PARIS (7e).

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, images, video, etc., are 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 <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 Worcester 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	Desktop search	Question answering
Audio	Forum search	Summarization
Music	Social 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 users' decision about what is relevant: e.g., task, context, novelty, style, other documents they've already read (marginal relevance)
  - *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 structured data
    - Working directly with document contents rather than turning document collections into databases
    - Mapping documents to lower dimensional spaces using heuristic (e.g., IDF) and statistical models (e.g., language models)

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. Speech and translation, other ML leaderboards
  - 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 (models) of actual information needs
  - Belkin's *anomalous state of knowledge*
    - If you knew what you wanted, you wouldn't need to ask.
  - 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 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, Solr
- Big issues include main IR issues but also some others
  - Connections to systems, NLP, ML, vision, etc.

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 and retraining models while processing queries is also a design issue

Search Engine Issues

- Scalability
  - Making everything work with millions of users every day, and terabytes/petabybtes of data
  - 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 <u>the</u> major issues
  - Affects the efficiency of search engines and, more seriously, the <u>effectiveness</u> 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