

Information Retrieval and Web Search

Cornelia Caragea

Computer Science
University of Illinois at Chicago

Credits for slides: Mooney

IR System Architecture & Examples of Retrieval Models

Announcements

- Exam 1 date: October 11
- Exam 2 date: November 29
- Class project due: December 9

Required Reading

- “Information Retrieval” textbook
 - Chapter 1
 - Chapter 2

Information Retrieval

- The processing, indexing and retrieval of textual documents.
- Searching for pages on the World Wide Web is perhaps the most widely used IR application
- Concerned firstly with retrieving relevant documents to a query.
- Concerned secondly with retrieving from large sets of documents efficiently.

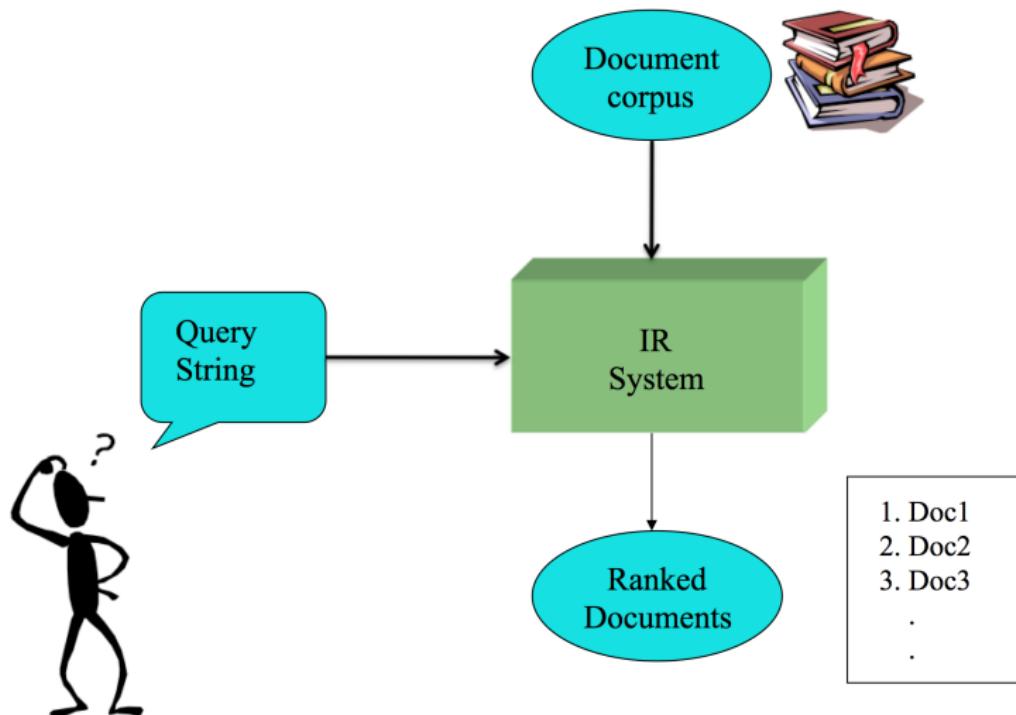
Typical IR Task

- Given:
 - A corpus of textual natural-language documents
 - A user query in the form of a textual string
- Find:
 - A ranked set of documents that are **relevant** to the query

Key Terms Used in IR

- **Query:** a representation of what the user is looking for - can be a list of words or a phrase.
- **Document:** an information entity that the user wants to retrieve
- **Collection or corpus:** a set of documents
- **Index:** a representation of information that makes querying easier
- **Term:** word or concept that appears in a document or a query

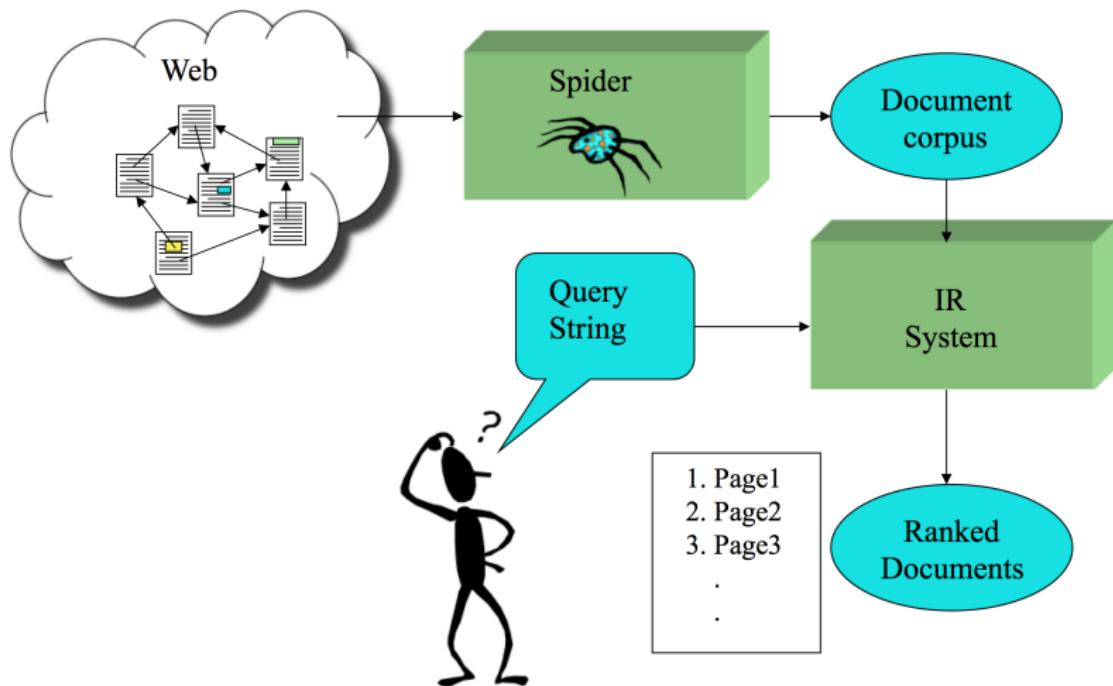
Typical IR System Architecture



Web Search

- Application of IR to HTML documents on the World Wide Web.
- Differences:
 - Must assemble a document corpus by *spidering* the Web.
 - Documents change uncontrollably (unlike closed collections).
 - Can exploit the structural layout information in HTML (or XML).
 - Can exploit the link structure of the Web.

Web Search System



Relevance

- Relevance is a subjective judgment and may include:
 - Being on the proper subject.
 - Being timely (recent information).
 - Being authoritative (from a trusted source).
 - Satisfying the goals of the user and his/her intended use of the information (**information need**)
- Main **relevance criterion**: an IR system should fulfill a **user's information need**

Basic IR Approach: Keyword Search

- The simplest notion of relevance is that the query string appears verbatim in the document.
 - Example: “unsupervised keyphrase extraction”
- Slightly less strict notion is that the words in the query appear frequently in the document, in any order - **bag of words** representation
 - Example: “keyphrase extraction unsupervised”

- How does this approach work?
 - Find words/concepts in documents
 - Compare them to words in a query
 - Very effective!



Problems with Keyword Search

- May not retrieve relevant documents that include synonymous terms.
 - “restaurant” vs. “café”
 - “PRC” vs. “China” (PRC = People’s Republic of China)
- May retrieve irrelevant documents that include ambiguous terms.
 - “bat” (baseball vs. mammal)
 - “bit” (unit of data vs. act of eating)
 - “Apple” (company vs. fruit)

“Apple” (company vs. fruit) Chicago, IL

Google apple

All News Images Shopping Maps More Settings Tools

About 4,530,000,000 results (0.94 seconds)

Apple
<https://www.apple.com/> •
Discover the innovative world of Apple and shop everything iPhone, iPad, Apple Watch, Mac, and Apple TV, plus explore accessories, entertainment, and expert ...
You've visited this page 2 times. Last visit: 11/15/18

Search apple.com

iPhone
For a limited time, get iPhone XR from \$449 or iPhone XS from ...

Apple Michigan Avenue
Apple Store Michigan Avenue store hours, contact information ...

iPad
Check out iPad Pro, available in two sizes, iPad, and iPad mini ...

Apple Support
Apple support is here to help. Learn more about popular ...

Lincoln Park
Apple Store Lincoln Park store hours, contact information, and ...

Chicago
Find out what's going on around Chicago with Today at Apple.

 Map data ©2019 Google

Hours ▾ Your past visits ▾ Sort by ▾

A Apple Michigan Avenue
1.6 mi · Chicago, IL · (312) 529-9500
Closed · Opens 9AM Thu

B Apple Lincoln Park
3.3 mi · Chicago, IL · (312) 777-4200
Closed · Opens 9AM Thu

C Apple Oakbrook
Oakbrook Center
17.1 mi · Oak Brook, IL · (630) 481-3470
Closed · Opens 10AM Thu

More locations

Apple
Technology company



 apple.com

Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, computer software, and online services. It is considered one of the Big Four of technology along with Amazon, Google and Facebook. [Wikipedia](#)

Stock price: AAPL (NASDAQ) \$154.94 +1.87 (+1.22%)
Jan 16, 4:00 PM EST · [Disclaimer](#)

Founded: April 1, 1976, Cupertino, CA
Headquarters: Cupertino, CA
CEO: Tim Cook (Aug 24, 2011–)
Founders: Steve Jobs, Steve Wozniak, Ronald Wayne
Products: iPhone, Apple Watch, Apple iPad Family, Macintosh, MORE
Did you know: Apple Inc. is the world's largest information technology company by revenue. [wikipedia.org](#)

Popular products

[Mobile Phones](#) [Tablet Computers](#) [Watches](#) [Laptops](#) C >

 iPhone XR

 AirPods

 iPhone X

★★★★★ (2k+) ★★★★★ (37k+) ★★★★★ (13k+)

Profiles

 Facebook  YouTube  Twitter  LinkedIn  Google+

“Apple” (company vs. fruit) Chicago, IL

Top stories



Apple's iPhone XS battery case actually works with the iPhone X

The Verge
8 hours ago



Apple's new battery cases are its most powerful ones yet

The Verge
7 hours ago



Apple spent \$310K on Tim Cook's personal security in 2018, here's how that compares to...

9to5Mac
1 hour ago

People also search for



Samsung Group



Google



Microsoft Corporation



Amazon.com



Qualcomm

View 15+ more

Feedback

Disclaimer

→ More for apple

Apple snubs Chicago in expansion plans - Crain's Chicago Business

<https://www.chicagobusiness.com/john-pletz-technology/choose-chicago-not-apple>
Dec 13, 2018 - Apple is doubling down on Austin and adding jobs in several cities, but not Chicago, confirming what local officials have known or suspected ...

Videos



How Qualcomm became a chip giant and why its business model is ...

CNBC.com - 5 hours ago



Top 5 Alternatives To The Apple MacBook Air 2018

News18.com - 18 hours ago



Cramer: Netflix's price bump means Amazon, Apple's stocks are ...

CNBC.com - 1 day ago

People also ask

Do students get discount at Apple?



How can I contact Apple support?



How do you make a reservation at Apple Store?



What is Apple known for?



Feedback

“Apple” (company vs. fruit) Chicago, IL

Apple - YouTube

<https://www.youtube.com/user/Apple> ▾

Apple revolutionized personal technology with the introduction of the Macintosh in 1984. Today, Apple leads the world in innovation with iPhone, iPad, Mac, A...

Apple Inc. - Wikipedia

https://en.wikipedia.org/wiki/Apple_Inc._ ▾

Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, ...

Number of employees: 132,000 (2018) Net income: US\$59.531 billion (2018)

Number of locations: 500 retail stores (2018) Revenue: US\$265.595 billion (2018)

Apple's new battery cases are its most powerful ones yet - The Verge

<https://www.theverge.com/circuitbreaker/2019/1/16/.../apple-iphone-battery-case-size> ▾

7 hours ago - Apple's newly launched battery cases for the iPhone XS and XS Max have more battery capacity than any of Apple's prior cases. The Verge ...

Tech companies



Organization software



Storage and organization



Searches related to apple

apple support	apple watch
apple sign in	apple store near me
apple store	apple number
apple iphone	apple refurbished

“Apple” (company vs. fruit) Manhattan, KS

Google search results for "apple" in Manhattan, KS.

Search Results:

- Apple - Official site**
Mac, iPad, iPhone, Apple Watch, and Apple TV. Learn more and shop now.
Apple Back to School 2017 · iPhone 7 · Mac · Carpool Karaoke
- Apple**
<https://www.apple.com/>
Discover the innovative world of Apple and shop everything iPhone, iPad, Apple Watch, Mac, and Apple TV, plus explore accessories, entertainment, and expert ...
- Apple Support**
Apple support is here to help. Learn more about popular ...
- Mac**
MacBook Pro - MacBook - Compare -
MacBook Air - iMac
- iPhone**
Explore iPhone, the world's most powerful personal device ...
- iPad**
Check out iPad Pro, available in two sizes, iPad, and iPad mini ...
- Find a Store**
Complete store list - Genius Bar - Today at Apple - ...
- Apple ID**
Your Apple ID is the account you use for all Apple services.

Map: A map showing locations in Manhattan, KS, including Wakefield, Alida, Fort Riley, Ogden, St George, Zeeland, and Pillbury Crossing. Points A, B, and C are marked: A is Little Apple Hearing Aid Center, B is TWISTED APPLE TATTOO INC., and C is Little Apple Toyota Honda.

Business Listings:

- A Little Apple Hearing Aid Center**
0.8 mi - Manhattan, KS - (785) 537-0252
Closed now
- B TWISTED APPLE TATTOO INC.**
14.4 mi - Junction City, KS - (785) 579-5144
Closed now
- C Little Apple Toyota Honda**
1.1 mi - Manhattan, KS - (785) 539-7441
Closed now

Profile: Apple Technology company

Apple Inc. is an American multinational technology company headquartered in Cupertino, California that designs, develops, and sells consumer electronics, computer software, and online services. [Wikipedia](#)

Stock price: AAPL (NASDAQ) \$159.98 +0.20 (+0.13%)
Aug 23, 4:00 PM EDT · Disclaimer

Founded: April 1, 1976, Cupertino, CA
Headquarters: Cupertino, CA
Founders: Steve Jobs, Steve Wozniak, Ronald Wayne
Products: iPhone, iPad, iPhone 7, iPod, Macintosh, Apple Watch, MORE

Did you know: Apple Inc. is the world's largest information technology company by revenue. [wikipedia.org](#)

Profiles:

Tim Cook (CEO), Angela Ahrendts (SVP, Retail), Eddy Cue (SVP, Internet Software, Services), Craig Federighi (SVP, Software Engineering), Jonathan Ive (Chief Design Officer)

Executives:

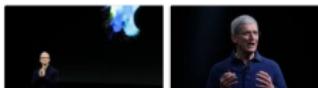
View 10+ more

People also search for:

Federal Bureau of Investigation, Google, Microsoft Corporation, Samsung Group, Dell

“Apple” (company vs. fruit) Manhattan, KS

Top stories



Apple Scales Back Its Ambitions for a Self-Driving Car

The New York Times · 1 day ...

Apple could announce the next iPhone on September 12

TechCrunch · 10 hours ago

Apple refreshes News Publisher with new dashboard, fonts, and more

9to5Mac · 2 hours ago

→ More for apple

Apple Macintosh and iPhone Stores in Manhattan, KS

apple-store.in/66503-manhattan/ ▾
RadioShack Store - 018889, RadioShack Store - 018889, Local 25c, Dorado, PR 646, phone: n/a; distance from Manhattan, KS: 3.3 miles. iPod Store: yes; iPad ...

Little Apple Brewing Company: Manhattan, Kansas

www.littleapplebrewery.com/ ▾
About Us · Menus · Libations · Reservations · Events/Calendar · Press · Gift Cards. Beet_carpaccio. Hours. MONDAY - THURSDAY: 11 AM - 11 PM* FRIDAY ...

Apple's New Campus: An Exclusive Look Inside the Mothership | WIRED

<https://www.wired.com/2017/05/apple-park-new-silicon-valley-campus/>
Flawless curves, milled aluminum, endless glass, walled garden—sounds like an Apple product. In fact, it's the company's biggest launch ever.

Life and death in Apple's forbidden city | Technology | The Guardian

<https://www.theguardian.com/.../foxconn-life-death-forbidden-city-longhua-suicide-apple...>
Foxconn's enormous Longhua plant is a major manufacturer of Apple products. It might be the best-known factory in the world; it might also ...

Apple's Next Big Thing: Augmented Reality - Bloomberg

<https://www.bloomberg.com/news/articles/2017-03-20/apples-next-big-thing>
Apple's Next Big Thing: Augmented Reality. CEO Tim Cook is betting on augmented reality, a cousin of VR that he believes will keep his ...

Green Apple Bikes, KS | Official Website

greenapplebikes.com/ ▾
Green Apple Bikes is not owned by any government or corporation. It is entirely run by a board of volunteers. Bicycle maintenance is provided by volunteers.

Manhattan, Kansas - Wikipedia

https://en.wikipedia.org/wiki/Manhattan,_Kansas ▾
Manhattan is a city in northeastern Kansas in the United States at the junction of the Kansas ...
Nicknamed "The Little Apple" as a play on New York City's "Big Apple", Manhattan is best known as the home of Kansas State University and has a ...

“Apple” (company vs. fruit) Denton, TX

Google search results for "apple":

All News Images Maps Shopping More Search tools

About 1,520,000,000 results (0.70 seconds)

Apple
www.apple.com/ • Apple •
Apple leads the world in innovation with iPhone, iPad, Mac, Apple Watch, iOS, OS X, watchOS and more. Visit the site to learn, buy, and get support.

Results from apple.com

Apple Support
Apple's Official tech support page with support information for both ...

Mac
MacBook Pro - MacBook - MacBook Air - iMac - Mac Pro

iPad
Compare iPad models - iPad Air 2 - iPad mini 4 - Apple Support - ...

iPhone
iPhone 6s. With the most powerful technology and most intuitive ...

Apple Watch
Apple Watch is the most personal product we've ever made ...

Music
Stream from the entire Apple Music library. Tune in to live, 24/7 ...

In the news

Apple falls close to a freak-out level
 USA TODAY - 48 mins ago
Apple (AAPL) fans are still defending the stock - calling it a bargain. The market agreed ...

Apple iOS 9.2.1 Has A Nasty Surprise
Forbes - 7 hours ago

Apple, Samsung and Sony face child labour claims - BBC News
BBC News - 5 hours ago

More news for apple

Apple Store in Denton, Texas with Reviews & Ratings - YP ...
www.yellowpages.com › Denton, TX • Yellowpages.com •
Results 1 - 30 of 133 - Find 133 listings related to Apple Store in Denton on YP.com. See reviews, photos, directions, phone numbers and more for Apple Store ...

Apple Macintosh and iPhone Stores in Denton, TX
apple-store.in/76207-denton/ •
Sams Club Store - 4905, Sams Club Store - 4905 2850 W University Dr Denton, TX 76201, phone: (940) 889-1529; distance from Denton, TX: 0.7 miles.

iPhone Killer: The Secret History of the Apple Watch | WIRED
www.wired.com/2015/04/the-apple-watch/
In early 2013, Kevin Lynch accepted a job offer from Apple. Funny thing about the offer: It didn't say what he would be doing. So intense is Apple's secrecy that ...

How Apple Is Giving Design A Bad Name - Co.Design
www.fastcodesign.com/3053406/how-apple-is-giving-design-a-bad-name
Once upon a time, Apple was known for designing easy-to-use, easy-to-understand products. It was a champion of the graphical user interface, ...

Jonathan Ive and the Future of Apple - The New Yorker
www.newyorker.com/magazine/2015/02/23/shape-things-come
Ian Parker visits the top-secret lab where the world's most powerful design team created the Apple Watch.

Orthodontics Denton TX | Apple Orthodontix
www.appleortho.com › Locations •
Denton Orthodontics Office Mares Y Del Mundo, DMD, MS Apple Orthodontix Servicing Denton, Little Elm, Aubrey, Corinth, Krum, Ponder, Sanger Offering ...

Apple Inc. - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Apple_Inc._ • Wikipedia •
Apple Inc. is an American multinational technology company headquartered in Cupertino, California, that designs, develops, and sells consumer electronics, ...
Founders: Steve Jobs; Steve Wozniak; ... Headquarters: Apple Campus, Cupertino, ...
Founded: April 1, 1976, in Cupertino, C... Products: Mac; iPod; iPhone; iPad; Apple ...

Searches related to apple

apple store apple tv
best buy apple news

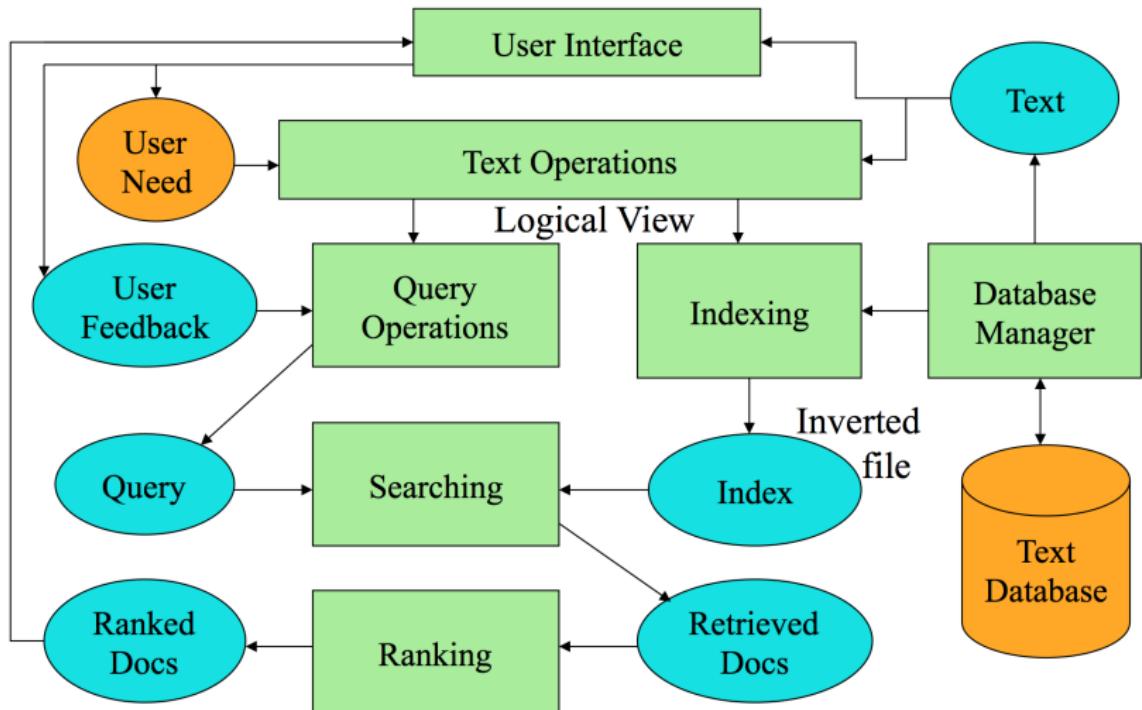
In this course

- We will cover the basics of keyword-based IR
- Also, address more complex techniques for “intelligent” IR

Techniques for Intelligent IR

- Take into account the **meaning** of the words used
- Take into account the **order** of words in the query
- Adapt to the user based on implicit or explicit **feedback**
- **Extend** search with related terms
- Perform automatic **spell checking / diacritics restoration**
- Take into account the **authority** of the source.

IR System Architecture



IR System Components

- **Text Operations** form index words (tokens)
 - Tokenization
 - Stop-word removal
 - Stemming
- **Indexing** constructs an **inverted index** of word to document pointers.
 - Mapping from keywords to document ids

Indexing - Inverted Index

Doc 1

I did enact Julius Caesar: I was killed
i' the Capitol; Brutus killed me.

term	docID	term	docID
I	1	ambitious	2
did	1	be	2
enact	1	brutus	1
julius	1	brutus	2
caesar	1	capitol	1
I	1	caesar	1
was	1	caesar	2
killed	1	caesar	2
i'	1	did	1
the	1	enact	1
capitol	1	hath	1
brutus	1	I	1
killed	1	I	1
me	1	i'	1
so	2	it	2
let	2	julius	1
it	2	killed	1
be	2	killed	1
with	2	let	2
caesar	2	me	1
the	2	noble	2
noble	2	so	2
brutus	2	the	1
hath	2	the	2
told	2	told	1
you	2	you	1
caesar	2	was	1
was	2	was	2
ambitious	2	with	2

Doc 2

So let it be with Caesar. The noble Brutus
hath told you Caesar was ambitious:

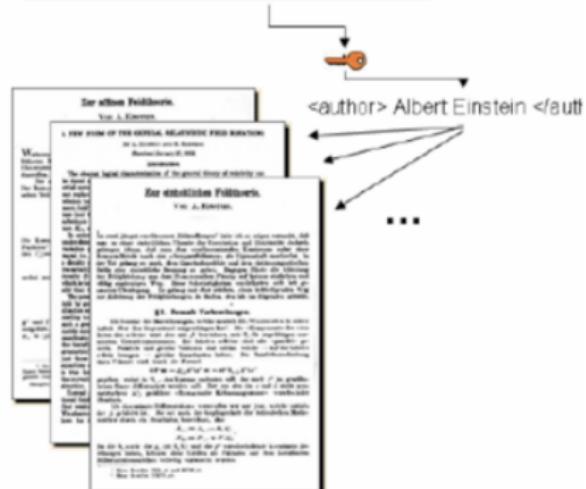
term	doc. freq.	→	postings lists
ambitious	1	→	2
be	1	→	2
brutus	2	→	1 → 2
capitol	1	→	1
caesar	2	→	1 → 2
did	1	→	1
enact	1	→	1
hath	1	→	2
I	1	→	1
i'	1	→	1
it	1	→	2
julius	1	→	1
killed	1	→	1
let	1	→	2
me	1	→	1
noble	1	→	2
so	1	→	2
the	2	→	1 → 2
told	1	→	2
you	1	→	2
was	2	→	1 → 2
with	1	→	2

Document Indexing

- **Index:** associates a document with one or more **keys (keywords)**
- Present key → identify documents that match key
- Efficiency is crucial, fast access

Author Search

Albert Einstein

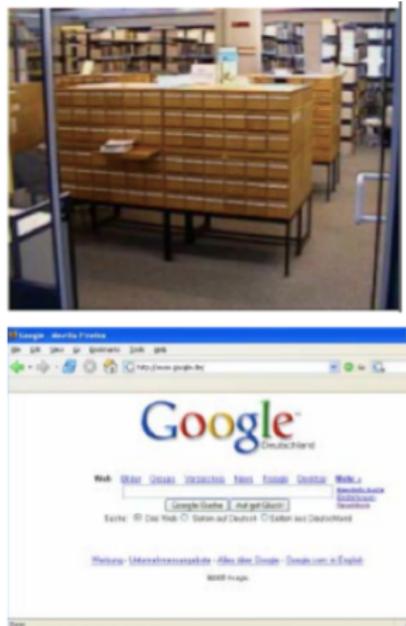


Document Indexing: Questions and Challenges

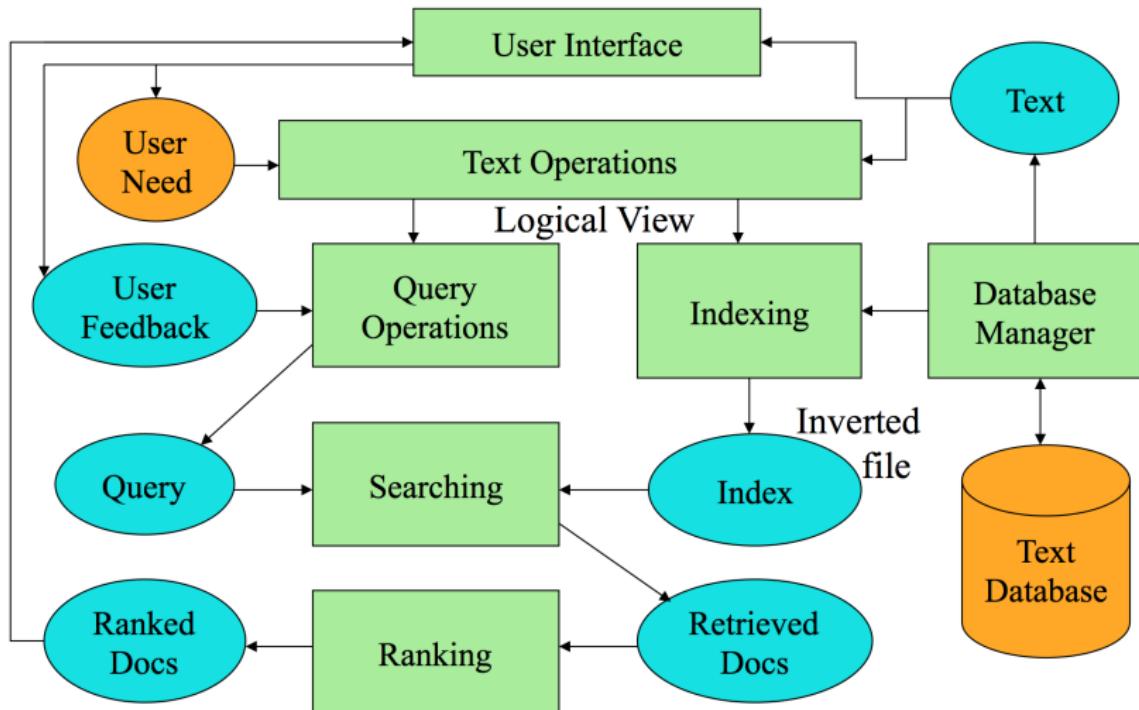
- Two questions:
 - Which **data structures** and **algorithms** should be used for creating and maintaining indices?
 - Which **attributes** of documents should be utilized as **keys**?
- Conceptual challenges:
 - **Usage:** specify keys that are useful in searching and retrieving of information
 - **Creation:** select attributes as keys that can be reliably determined from documents

Controlled Vocabulary or Full-Text Indexing

- Controlled vocabulary: list of terms whose meanings are specifically defined
 - terms need to be agreed upon
 - user needs to be aware of available vocabulary
 - approach pursued in the context of (digital) libraries / catalogues
- Full-text indexing: index terms are automatically selected from the document collection
 - documents need to be processed in order to extract index terms
 - selection heuristics for choosing appropriate terms
 - approach pursued by search engines



IR System Architecture



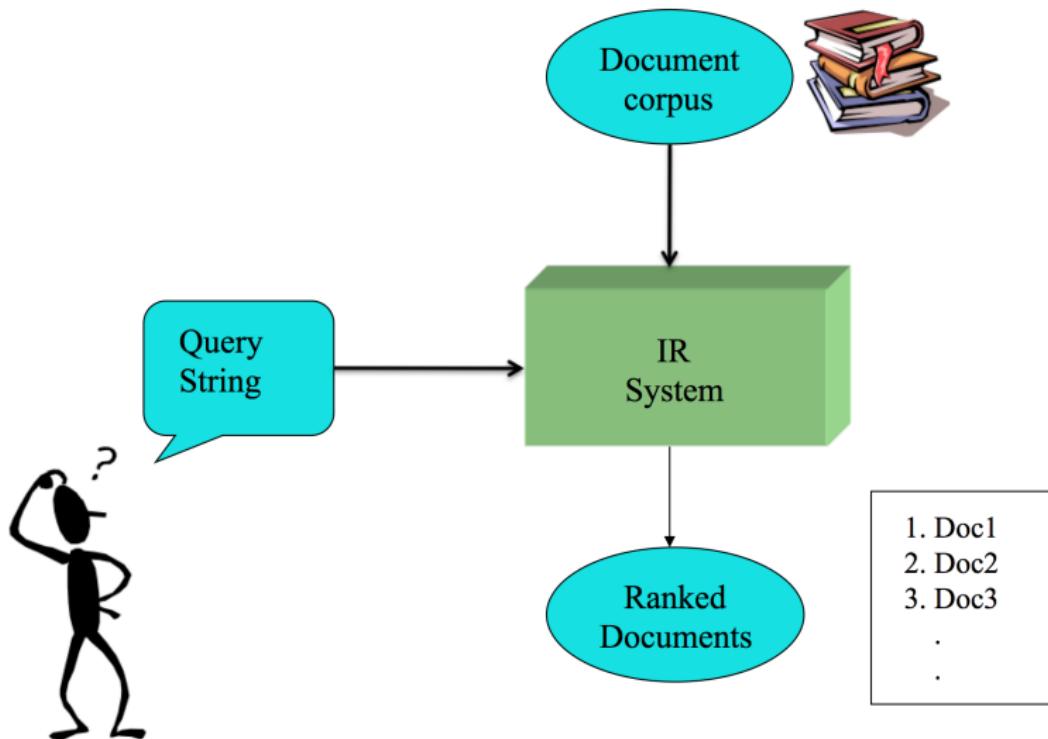
IR System Components

- **Searching** retrieves documents that contain a given query token from the inverted index.
- **Ranking** scores all retrieved documents according to a relevance metric.
- **User Interface** manages interaction with the user:
 - Query input and document output
 - Relevance feedback
 - Visualization of results
- **Query Operations** transform the query to improve retrieval:
 - Query expansion using a thesaurus
 - Query transformation using relevance feedback

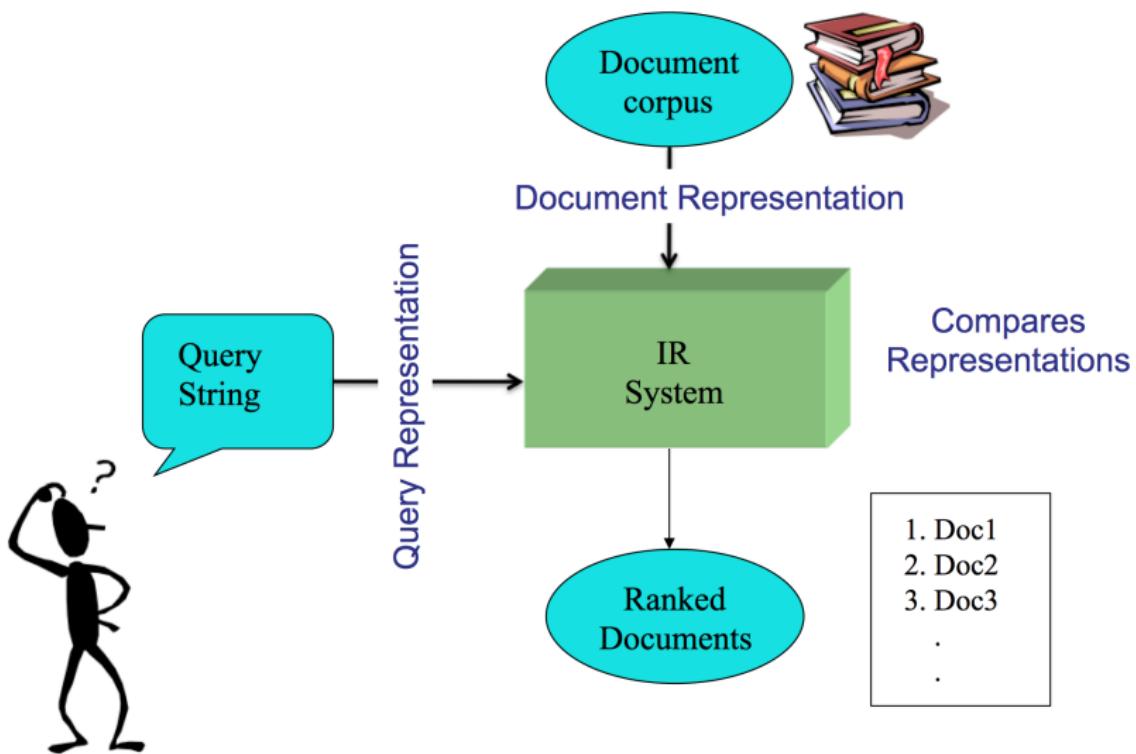
Retrieval Models

- A retrieval model specifies the details of:
 - Document representation
 - Query representation
 - How do we compare representations - retrieval function?
- Determines a notion of relevance.
- Notion of relevance can be binary or continuous (i.e. ranked retrieval).

Typical IR System Architecture



Typical IR System Architecture



Classes of Retrieval Models

- Boolean models (set theoretic)
 - Boolean Retrieval Models
- Vector space models (algebraic)
 - Generalized VS
 - Latent Semantic Indexing
- Probabilistic models

About IR and Machine Learning

- 12 IT skills that employers can't say no to
 - Machine Learning is #1

"It's not just the case for Google," says Kevin Scott, senior engineering manager at Google Inc. "There are lots of applications that have big, big, big data sizes, which creates a fundamental problem of how you organize the data and present it to users."

http://www.computerworld.com/s/article/9026623/12_IT_skills_that_employers_can_t_say_no_to_