BLG 540E
TEXT RETRIEVAL SYSTEMS

Introduction

Arzucan Özgür

Faculty of Computer and Informatics, İstanbul Techical University
February 11, 2011

Course Staff

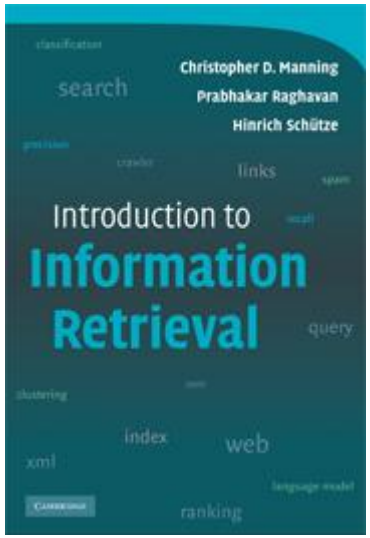
- Instructor: Arzucan Özgür
  - Office: EEB 2316
  - Phone: (0212) 285 3608
  - E-mail: ozgura@itu.edu.tr

(Please include BLG540E in your subjet when sending e-mail.)

  - Ofis hours: Wednesday, 14:00-16:00

References

- Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008.



Available online at the companion website of the book:
http://nlp.stanford.edu/IR-book/information-retrieval-book.html

- Course Web Site: TBA

Grading

- Paper presentation: 20%
- Project: 30%
- Final: 40%
- Class participation: 10%

Final project format

- Research paper - using the SIGIR format. Students will be in charge of problem formulation, literature survey, hypothesis formulation, experimental design, implementation, and possibly submission to a conference like SIGIR or WWW.

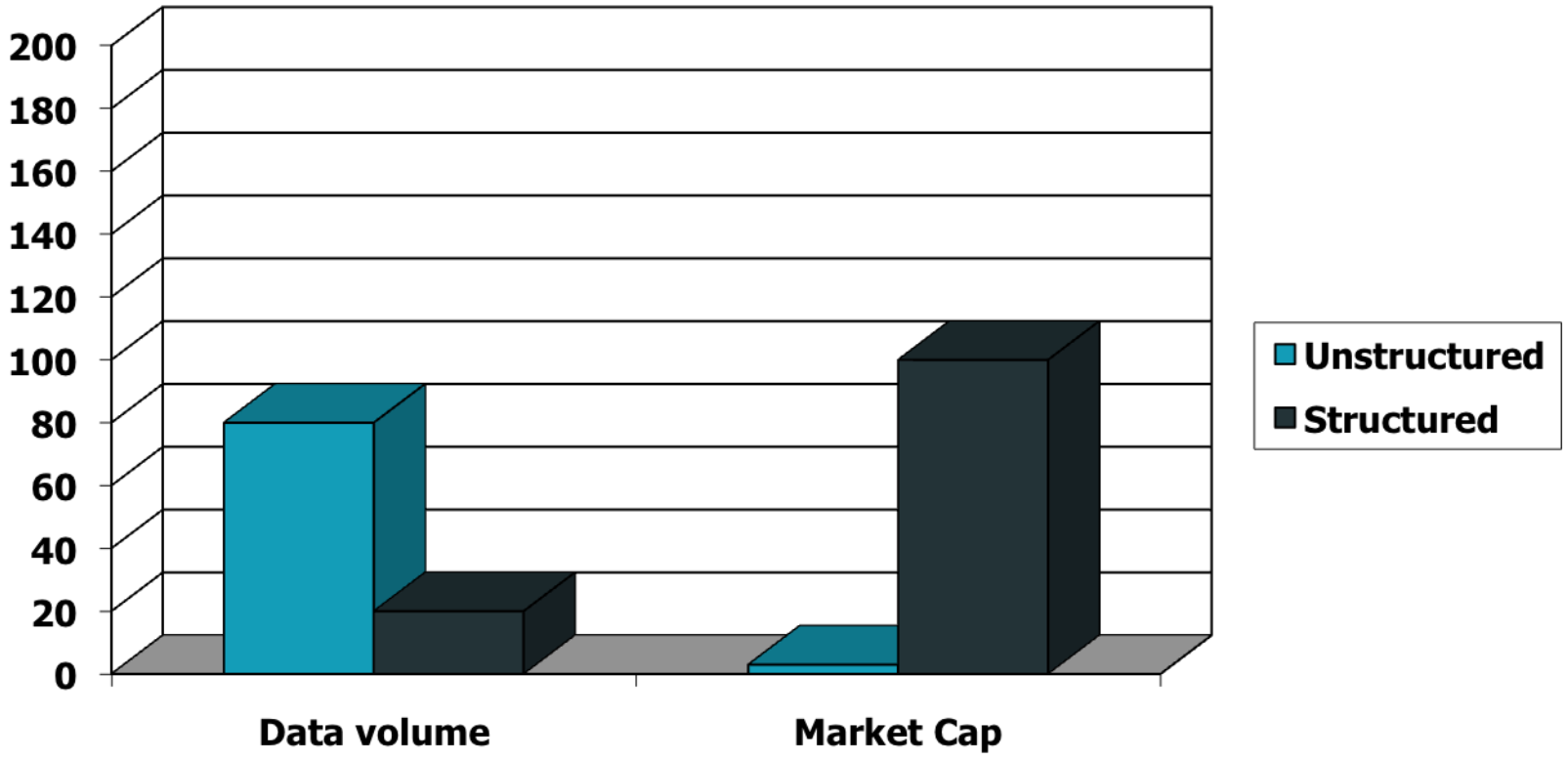
- Software system - develop a working system or API. Students will be responsible for identifying a problem, implementing it and deploying it, either on the Web or as an open-source downloadable tool. The system can be either stand alone or an extension to an existing one.

- Survey paper - identify a topic of research in IR and summarize 15-20 recent papers on it, along with an introduction that compares and contrasts the papers involved.

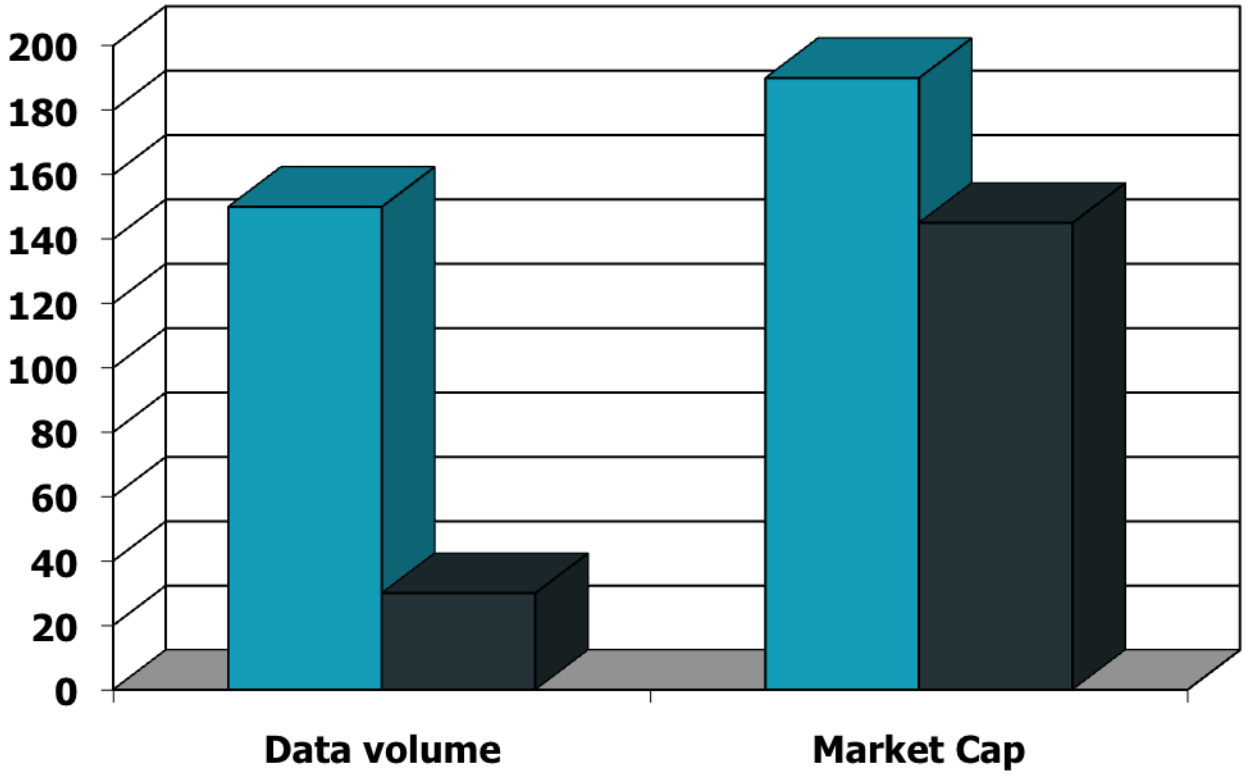
Information Retrieval

- Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).

Unstructured (text) vs. structured (database) data in 1996



Unstructured (text) vs. structured (database) data in 2009



Examples of search engines

- Conventional (library catalog).

  Search by keyword, title, author, etc.
- Text-based (Google, Yahoo!, Bing).

  Search by keywords. Limited search using queries in natural language.
- Multimedia (QBIC, WebSeek, SaFe)

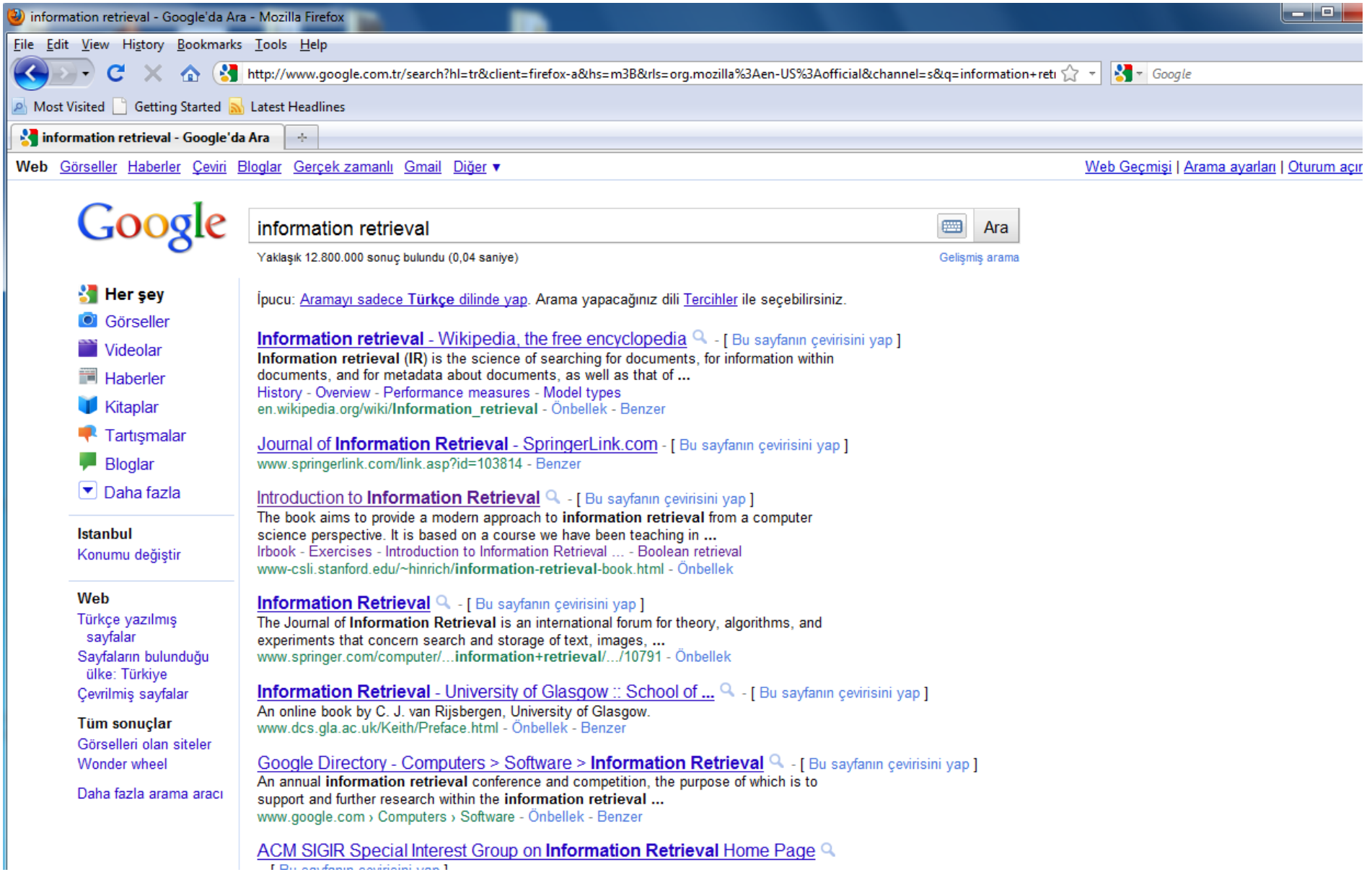
  Search by visual appearance (shapes, colors,…).
- Question answering systems (Ask, NSIR, Answerbus)

  Search in (restricted) natural language
- Other:

  cross language information retrieval, music retrieval

IR systems on the Web

- Search for Web pages http://www.google.com
- Search for images http://www.picsearch.com
- Search for image content http://wang14.ist.psu.edu/
- Search for answers to questions http://www.askjeeves.com
- Music retrieval http://www.rotorbrain.com/foote/musicr/


information retrieval - Google'da Ara - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com.tr/search?hl=tr&client=firefox-a&hs=m3B&rls=org.mozilla%3Aen-US%3Aofficial&channel=s&q=information+reti
Google
Most Visited Getting Started Latest Headlines
information retrieval - Google'da Ara
Web Görseller Haberler Çeviri Bloglar Gerçek zamanlı Gmail Diğer ▾
Web Geçmişi | Arama ayarları | Oturum açın
Google
information retrieval
Ara
Yaklaşık 12.800.000 sonuç bulundu (0,04 saniye)
Gelişmiş arama
Her şey
Görseller
Videolar
Haberler
Kitaplar
Tartışmalar
Bloglar
Daha fazla
Istanbul
Konumu değiştir
Web
Türkçe yazılmış sayfalar
Sayfaların bulunduğu ülke: Türkiye
Çevrilmiş sayfalar
Tüm sonuçlar
Görselleri olan siteler
Wonder wheel
Daha fazla arama aracı
İpucu: Aramayı sadece Türkçe dilinde yap. Arama yapacağınız dili Tercihler ile seçebilirsiniz.
Information retrieval - Wikipedia, the free encyclopedia - [Bu sayfanın çevirisini yap]
Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of ...
History - Overview - Performance measures - Model types
en.wikipedia.org/wiki/Information_retrieval - Önbellek - Benzer
Journal of Information Retrieval - SpringerLink.com - [Bu sayfanın çevirisini yap]
www.springerlink.com/link.asp?id=103814 - Benzer
Introduction to Information Retrieval - [Bu sayfanın çevirisini yap]
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 ...
Irbook - Exercises - Introduction to Information Retrieval ... - Boolean retrieval
www-csli.stanford.edu/~hinrich/information-retrieval-book.html - Önbellek
Information Retrieval - [Bu sayfanın çevirisini yap]
The Journal of Information Retrieval is an international forum for theory, algorithms, and experiments that concern search and storage of text, images, ...
www.springer.com/computer/...information+retrieval/.../10791 - Önbellek
Information Retrieval - University of Glasgow :: School of ... - [Bu sayfanın çevirisini yap]
An online book by C. J. van Rijsbergen, University of Glasgow.
www.dcs.gla.ac.uk/Keith/Preface.html - Önbellek - Benzer
Google Directory - Computers > Software > Information Retrieval - [Bu sayfanın çevirisini yap]
An annual information retrieval conference and competition, the purpose of which is to support and further research within the information retrieval ...
www.google.com › Computers › Software - Önbellek - Benzer
ACM SIGIR Special Interest Group on Information Retrieval Home Page

İTÜ Kütüphane Hizmetleri | İTÜ Mustafa İnan Kütüphanesi - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.kutuphane.itu.edu.tr/

Most Visited Getting Started Latest Headlines

İstanbul Teknik Üniversitesi | İTÜ Kütüphane Hizmetleri | İTÜ ...

İstanbul Teknik Üniversitesi
Kütüphane ve Dokümantasyon Daire Başkanlığı

Kütüphane Hizmetleri

Kütüphane Hakkında | Kütüphane Kataloğu | Elektronik Kaynaklar | Hizmetler | Kullanıcı İşlemleri | Bağlantılar | Site Haritası

HIZLI LİNKLER

- A-Z İndeks
- Sıkça Sorulan Sorular
- Kütüphaneciye Danışın
- Sanal Tur
- Kat Planı
- Yeni Gelen Yayınlar
- Yayın Uzatma
- Veritabanları
- Belgeler ve Formlar
- Tanıtım Filmi
- Bağışta Bulunanlar
- Çalışma Saatleri
- Kütüphane Albümü
- İTÜ Anasayfa

HIZLI TARAMA

Kelime | TARA

>> katalog tarama | detaylı tarama

ELEKTRONİK KAYNAKLARDAN TOPLU TARAMA

Başlık | TARA

HEPSİ

>> gelişmiş tarama | kampüs dışı erişim

ELEKTRONİK DERGİ LİSTESİ

kelimelerini içeren | GÖSTER

>> konuya göre | gelişmiş arayüz | kampüs dışı erişim

DUYURULAR

Yeni Alınan Veritabanları | Deneme Erişimleri | Diğer Duyurular

- İTÜ Mustafa İnan Kütüphansi 14 Şubat 2011 tarihinden itibaren 7/24 saat hizmet vermeye başlayacaktır. (10.02.2011)
- Derwent Innovation Index Online veritabanına 2011 yılı için abone olunmuştur. (26.01.2011)
- Westlaw International veritabanına 2011 yılı için abone olunmuştur. (26.01.2011)
- AGU (The American Geophysical Union) veritabanına 2011 yılı için abone olunmuştur. (24.01.2011)

KÜTÜPHANE BÜLTENİ

Kütüphane bültenlerine bu yolu kullanarak erişebilirsiniz.

ETKİNLİKLER

Kütüphane etkinliklerimize bu yolu kullanarak ulaşabilirsiniz.

ANI KİTABI

KAYNAK VE HİZMETLER

GOOGLE'DA ARAMA

Yahoo! Image Search Results for apple ipod - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help

http://images.search.yahoo.com/search/images?fr=lo&p=apple+ipod

pubmed million abstracts

Yahoo! My Yahoo! Mail Welcome, **Guest** [Sign In]

Web | **Images** | Video | Local | Shopping | more

YAHOO! SEARCH apple ipod Search

SafeSearch is ON Advanced Search Preferences

Image Results 1 - 20 of about 320,723 for **apple ipod** - 0.10 sec.

Show: All | Wallpaper - Large - Medium - Small | Color - Black & White

Also try: **apple ipod** shuffle, **apple ipod** downloads More...

Apple iPods on Yahoo! Shopping
Yahoo! Shortcut - About

Apple iPod Vide...er.jpg
111 x 175 pixels - 4.5kB
www.gizmania.com/archives/.../
apple_ipod_vide.html

apple_ipod_60_g...el.jpg
200 x 276 pixels - 21.1kB
www.mobilewhack.com/reviews/
apple_ipod_60gb_photo.html

apple_ipod_vide...er.jpg
250 x 201 pixels - 10.6kB
www.mobilewhack.com/reviews

apple_ipod_nano...b.jpg
Severe Weather Alert
www.cnet.com.au/mp3players/.../
0,39029137,40003685,00.htm

Find: million Next Previous Highlight all Match case

Done FoxyT GP Open Notebook

european union - Google News - Mozilla Firefox

File Edit View History Bookmarks ScrapBook Tools Help

www.google.com

Sign in

Google News

Web Images Video **News** Maps more »

european union | Search News | Search the Web | Advanced news search | Preferences

Results **1** - **10** of about **27,833** for **european union**. (**0.71** seconds)

Try your search on Yahoo News, Ask, AllTheWeb, MSN, Lycos, Sky News, CNN, Feedster, Daypop, Bloglines

Top Stories
World
U.S.
Business
Sci/Tech
Sports
Entertainment
Health
Most Popular

News Alerts

RSS | Atom
About Feeds

Mobile News

About Google News

Sorted by relevance Sort by date

Washington Post

EXTREME SOLIDARITY Far-Right Parties Form New Group in **European ...**

Spiegel Online, Germany - 42 minutes ago

European Union expansion is a topic typically supported by those on the left of the continent's political spectrum and opposed by those on the right. **...**

Far-right EU lawmakers form coalition Olberlin

all 88 news articles »

EnjoyFrance.com

Wild bird trade to be banned by **European Union**

EnjoyFrance.com, France - 1 hour ago

The **European Union** is going to ban the trade in wild birds starting in July, EU animal health officials have announced. Animal welfare campaigners are

Wild bird imports to end Green Consumer Guide

UN-Backed Body 'Disappointed' By Bird Trade Ban

Scoop.co.nz (press release)

EU To Ban Wild Birds Imports All Headline News

Earthtimes.org

all 8 news articles »

Russia, **European Union** A serious problem of trust

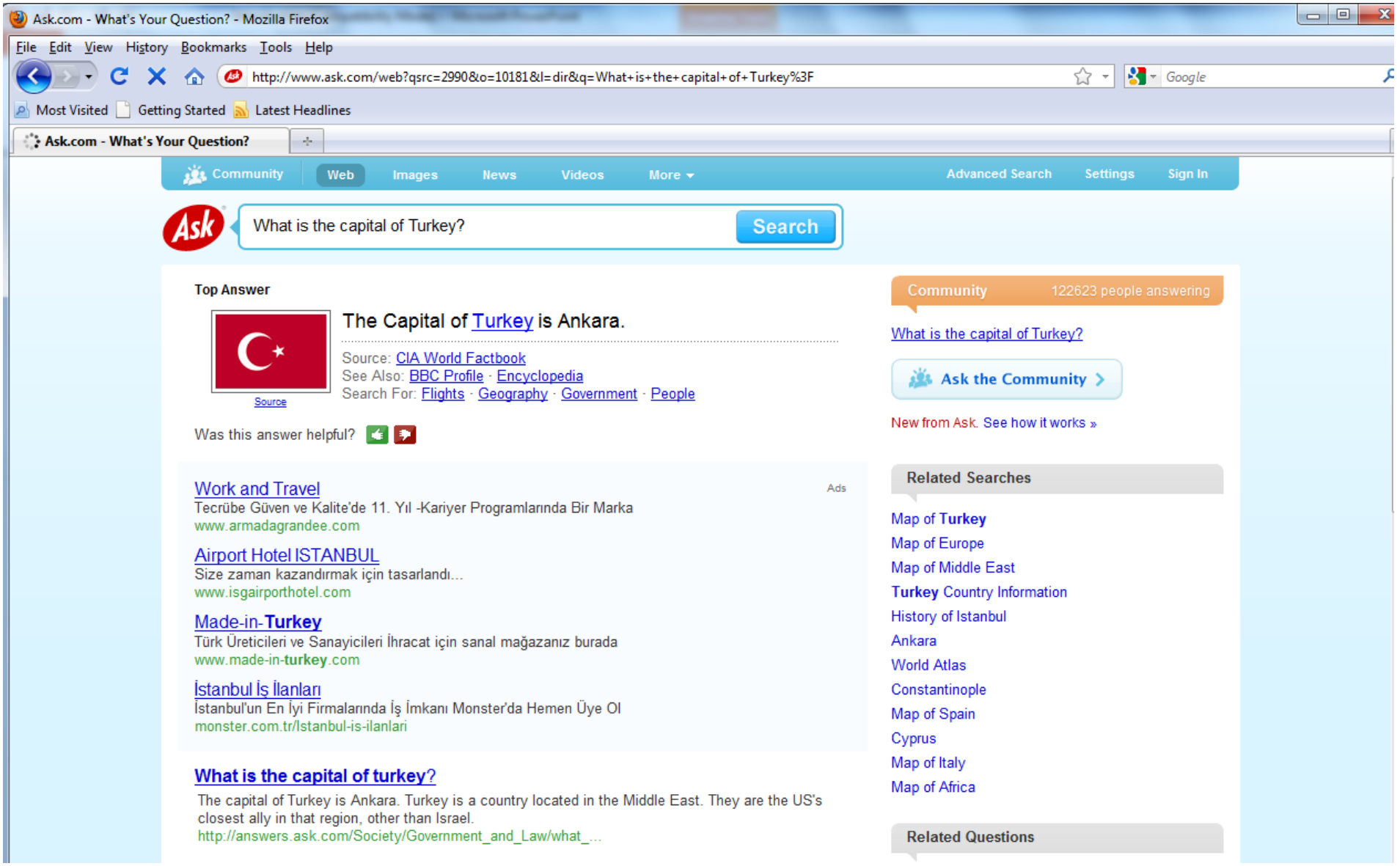
Monday Morning, Lebanon - 5 hours ago

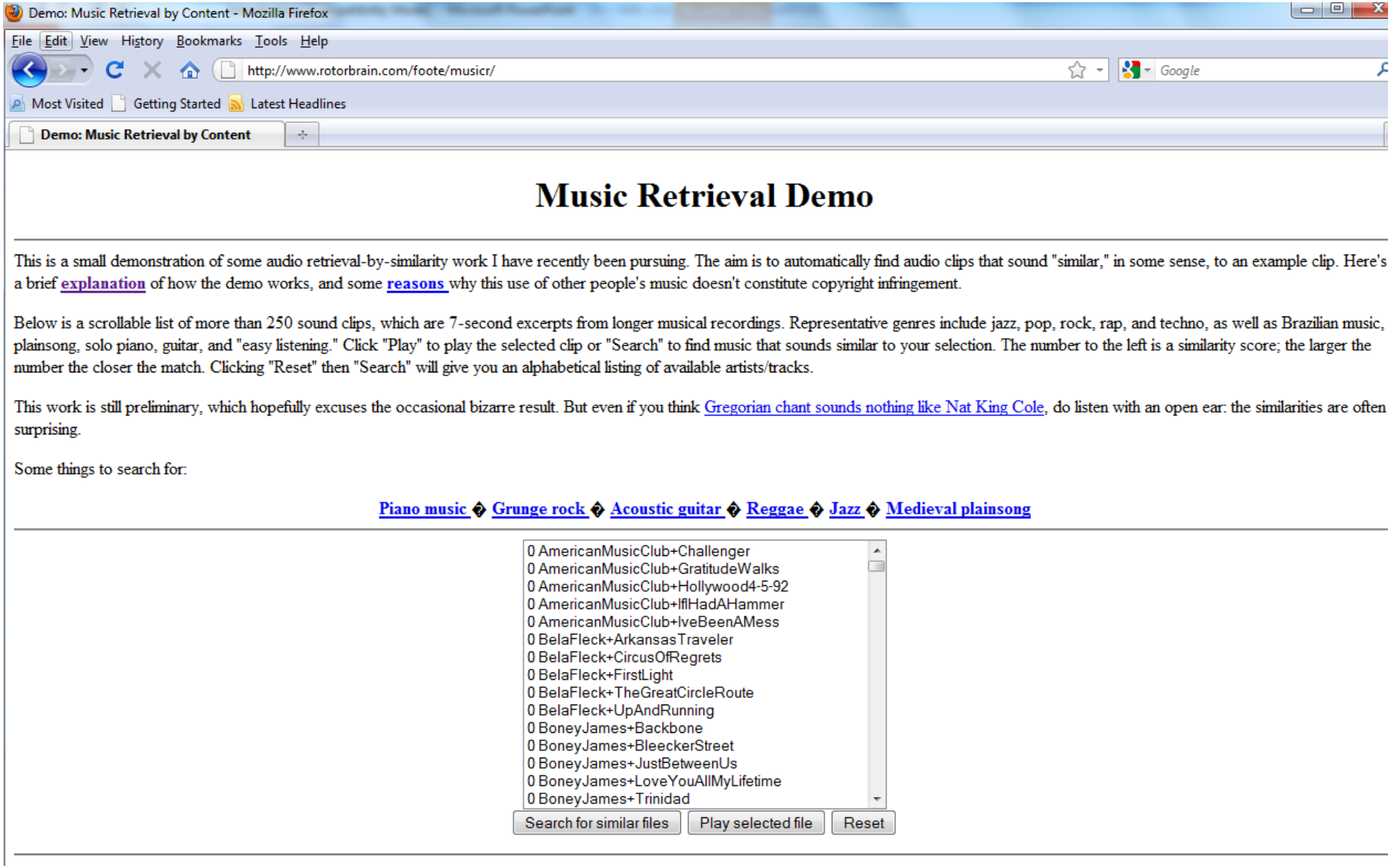
Merkel, in contrast, is wary of depending heavily on Russia for oil and gas and

Severe Weather Alert

Find: Next Previous Highlight all

Done FoxyT GP Open Notebook


Ask.com - What's Your Question? - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.ask.com/web?qsrc=2990&o=10181&l=dir&q=What+is+the+capital+of+Turkey%3F
Google
Most Visited Getting Started Latest Headlines
Ask.com - What's Your Question?
Community Web Images News Videos More Advanced Search Settings Sign In
Ask
What is the capital of Turkey?
Search
Top Answer
The Capital of Turkey is Ankara.
Source: CIA World Factbook
See Also: BBC Profile · Encyclopedia
Search For: Flights · Geography · Government · People
Source
Was this answer helpful?
Work and Travel
Ads
Tecrübe Güven ve Kalite'de 11. Yıl -Kariyer Programlarında Bir Marka
www.armadagrandee.com
Airport Hotel ISTANBUL
Size zaman kazandırmak için tasarlandı...
www.isgairporthotel.com
Made-in-Turkey
Türk Üreticileri ve Sanayicileri İhracat için sanal mağazanız burada
www.made-in-turkey.com
İstanbul İş İlanları
İstanbul'un En İyi Firmalarında İş İmkanı Monster'da Hemen Üye Ol
monster.com.tr/Istanbul-is-ilanlari
What is the capital of turkey?
The capital of Turkey is Ankara. Turkey is a country located in the Middle East. They are the US's closest ally in that region, other than Israel.
http://answers.ask.com/Society/Government_and_Law/what_...
Community 122623 people answering
What is the capital of Turkey?
Ask the Community
New from Ask. See how it works »
Related Searches
Map of Turkey
Map of Europe
Map of Middle East
Turkey Country Information
History of Istanbul
Ankara
World Atlas
Constantinople
Map of Spain
Cyprus
Map of Italy
Map of Africa
Related Questions


Demo: Music Retrieval by Content - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.rotorbrain.com/foote/musicr/
Google
Most Visited Getting Started Latest Headlines
Demo: Music Retrieval by Content
Music Retrieval Demo
This is a small demonstration of some audio retrieval-by-similarity work I have recently been pursuing. The aim is to automatically find audio clips that sound "similar," in some sense, to an example clip. Here's a brief explanation of how the demo works, and some reasons why this use of other people's music doesn't constitute copyright infringement.
Below is a scrollable list of more than 250 sound clips, which are 7-second excerpts from longer musical recordings. Representative genres include jazz, pop, rock, rap, and techno, as well as Brazilian music, plainsong, solo piano, guitar, and "easy listening." Click "Play" to play the selected clip or "Search" to find music that sounds similar to your selection. The number to the left is a similarity score; the larger the number the closer the match. Clicking "Reset" then "Search" will give you an alphabetical listing of available artists/tracks.
This work is still preliminary, which hopefully excuses the occasional bizarre result. But even if you think Gregorian chant sounds nothing like Nat King Cole, do listen with an open ear: the similarities are often surprising.
Some things to search for:
Piano music Grunge rock Acoustic guitar Reggae Jazz Medieval plainsong
0 AmericanMusicClub+Challenger
0 AmericanMusicClub+GratitudeWalks
0 AmericanMusicClub+Hollywood4-5-92
0 AmericanMusicClub+IflHadAHammer
0 AmericanMusicClub+IveBeenAMess
0 BelaFleck+ArkansasTraveler
0 BelaFleck+CircusOfRegrets
0 BelaFleck+FirstLight
0 BelaFleck+TheGreatCircleRoute
0 BelaFleck+UpAndRunning
0 BoneyJames+Backbone
0 BoneyJames+BleeckerStreet
0 BoneyJames+JustBetweenUs
0 BoneyJames+LoveYouAllMyLifetime
0 BoneyJames+Trinidad
Search for similar files
Play selected file
Reset

What does it take to build a search engine?

- Decide what to index
- Collect it
- Index it (efficiently)
- Keep the index up to date
- Provide user-friendly query facilities

What else?

- Understand the structure of the web for efficient crawling
- Understand user information needs
- Preprocess text and other unstructured data
- Cluster data
- Classify data
- Evaluate performance

Goals of the course

- Understand how search engines work
- Understand the limits of existing search technology
- Learn about the state of the art in IR research
- Learn to analyze textual and semi-structured data sets
- Learn to evaluate information retrieval
- Learn about standardized document collections
- Learn about text similarity measures
- Learn about semantic dimensionality reduction
- Learn about web crawling
- Learn to use existing software
- Understand the dynamics of the Web by building appropriate mathematical models
- Build working systems that assist users in finding useful information on the Web

References

- Some slides were adapted from:
  - Prof. Dragomir Radev's lectures at the University of Michigan:
    - http://clair.si.umich.edu/~radev/teaching.html
  - The book's companion website:
    - http://nlp.stanford.edu/IR-book/information-retrieval-book.html