

Information Retrieval

PRI 23/24 · Information Processing and Retrieval
M.EIC · Master in Informatics Engineering and Computation

Sérgio Nunes
Dept. Informatics Engineering
FEUP · U.Porto

Based on Chapter 1 from Modern Information Retrieval, Baeza-Yates, R. et al. (2011)
Based on Chapters 1 and 2 from Search Engines in Practice, Croft, B. et al. (2008)

Project Milestone 1 – Data Preparation

Project M1: Data Preparation

- M1 presentations are this week.
- Deadline for submitting your materials is the day before the lab class (until 18h).
- Presentation schedules are online.
 - Be there on time.
 - 8 min presentations + 5 min discussion.
 - Main elements: sources, pipeline highlights, documents, collection, and information needs.
 - See checklist.
- Don't forget the individual participation component (10%) — ask questions!

Project Overview

→ M1:

- **from...** selected data sources
- **to...** an organized (and studied) collection of documents
- **plus...** and a set of prospective search tasks

→ M2:

- **from...** a collection of documents
- **to...** the first version of a search system

Today's Lecture Outline

- Search systems
- Information retrieval concepts
 - Collections, documents, queries, results
 - Tasks, test collections, measures
 - Users, information needs, evaluation
- Architecture of information retrieval systems
 - Components of an information retrieval system
- History of information retrieval
 - Relation to text and data mining, machine learning, natural language processing

Search systems

Search and Information Retrieval

- Search is a daily activity for every computer, web, smartphone user.
- Search exists, *and is expected*, in every application.
- The web has brought the concept of keyword-based search to the center stage.
- Search contexts: web, site-specific, drive, computer, folders, email, messaging system, etc.
- Many different types: web search, maps search, image similarity search, social search, etc.

u porto

Google

All Images News Maps Videos More Tools

About 395,000,000 results (0.87 seconds)

<https://www.up.pt> · Translate this page

Universidade do Porto

A Universidade do Porto é uma das melhores instituições de ensino público da Europa. É reconhecida pela qualidade da formação e investigação, ...

Results from up.pt

Oferta Formativa
Cursos por Faculdade · Licenciatura em Arquitetura ...

Sigarra
O nível organizacional do Sistema de Informação SIGARRA ...

Master's Degrees - - - Courses
If you already have a Bachelor's Degree and would like to ...

International Students
Why should I choose U.Porto for a study period? U.Porto is the ...

People also ask

Quais são as melhores faculdades de Portugal?

Quanto custa a inscrição na Universidade do Porto?

Quantas faculdades têm a Universidade do Porto?

The screenshot shows a Google search results page for the query "u porto". The top result is the official website of the University of Porto (U.PORTO). To the right of the search results, there is a Knowledge Graph card for the University of Porto. The card features the university's logo, a map showing its location near Livraria Lello and Clérigos, and sections for "See photos" and "See outside". Below the map, the card displays the name "University of Porto (Universidade do Porto)" along with links for "Website", "Directions", and "Save". It also provides basic information such as being a "Public university in Porto" and a brief history. At the bottom of the card, it lists nearby landmarks: "A 1-min walk from Livraria Lello & Irmão".

U.PORTO

Livraria Lello
CLÉRIGOS
Jardim da

See photos See outside

University of Porto
(Universidade do Porto)

Website Directions Save

Public university in Porto

The University of Porto is a Portuguese public research university located in Porto, and founded on 22 March 1911. It is the second largest Portuguese university by number of enrolled students, after the University of Lisbon, and has one of the most noted research outputs in Portugal. [Wikipedia](#)

A 1-min walk from [Livraria Lello & Irmão](#)

Address: Praça de Gomes Teixeira, 4099-002 Porto

Hours: Open 24 hours

Phone: 22 040 8000

Total enrollment: 31,352 (2013)

google.com

fc porto

Images News Standings Players Results Games Score Website Maps All filters Tools SafeSearch

About 93,800,000 results (0.32 seconds)

FC Porto
4th in Primeira Liga

MATCHES NEWS STANDINGS PLAYERS

Primeira Liga · Today, 18:00
Porto vs Portimonense

Taça de Portugal · Third round
Vilar de Perdizes vs Porto

UEFA Champions League · Group stage · Matchday 3 of 6
Antwerp vs Porto

All times are in Western European Time

Feedback

See more

Display a menu

FC Porto (Futebol Clube do Porto)
Sports club

Players

Pepe	3
Defender	
Mehdi Taremi	9
Forward	
Alan Varela	22
Midfielder	
Iván Jaime	17
Midfielder	

[View 25+ more](#)

 fcporto.pt

Futebol Clube do Porto, MHIH, OM, commonly known as FC Porto or simply Porto, is a Portuguese professional sports club based in Porto. It is best

restaurante

Rating Hours All filters

Eleven Lab
4.7 ★★★★★ (22) ⓘ
Restaurant · R. de José Falcão 138
Closes soon · 7PM · Opens 11AM Sat
✓ Dine-in · ✗ Takeaway · ✗ Delivery

[RESERVE A TABLE](#)

Restaurante Solar Santa Catarina
4.6 ★★★★★ (70) ⓘ · €
Restaurant · Rua de Santa Catarina 842-980, 4000-455
Open · Closes 8PM
Dine-in · Takeaway · No delivery

17th Restaurant & Bar
4.5 ★★★★★ (1,377) ⓘ · \$\$\$
Restaurant · R. do Bolhão 223
Open · Closes 12AM
Dine-in · No takeaway · No delivery

[RESERVE A TABLE](#)

Restaurante Encontro Art Caffe
4.6 ★★★★★ (74) ⓘ
Restaurant · Rua de Santos Pousada 554
Open · Closes 11PM
Dine-in · Takeaway · Delivery

Update results when map moves

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

english | ajuda

Esqueceu-se da senha? | Outras formas de autenticação Utilizador Senha Iniciar sessão

Você está em: Início > Pesquisa de Ocorrências de UC

Pesquisa de Ocorrências de UC

Ano Letivo: 2022/2023

Unidades Curriculares

Código:

Sigla:

Nome:

Ano: Qualquer Ano

Período de Aulas: Todos os períodos

Curso/CE: ➔

Tipo de curso/CE:

Língua de trabalho:

Tipo de avaliação:

Docente

Código:

Sigla:

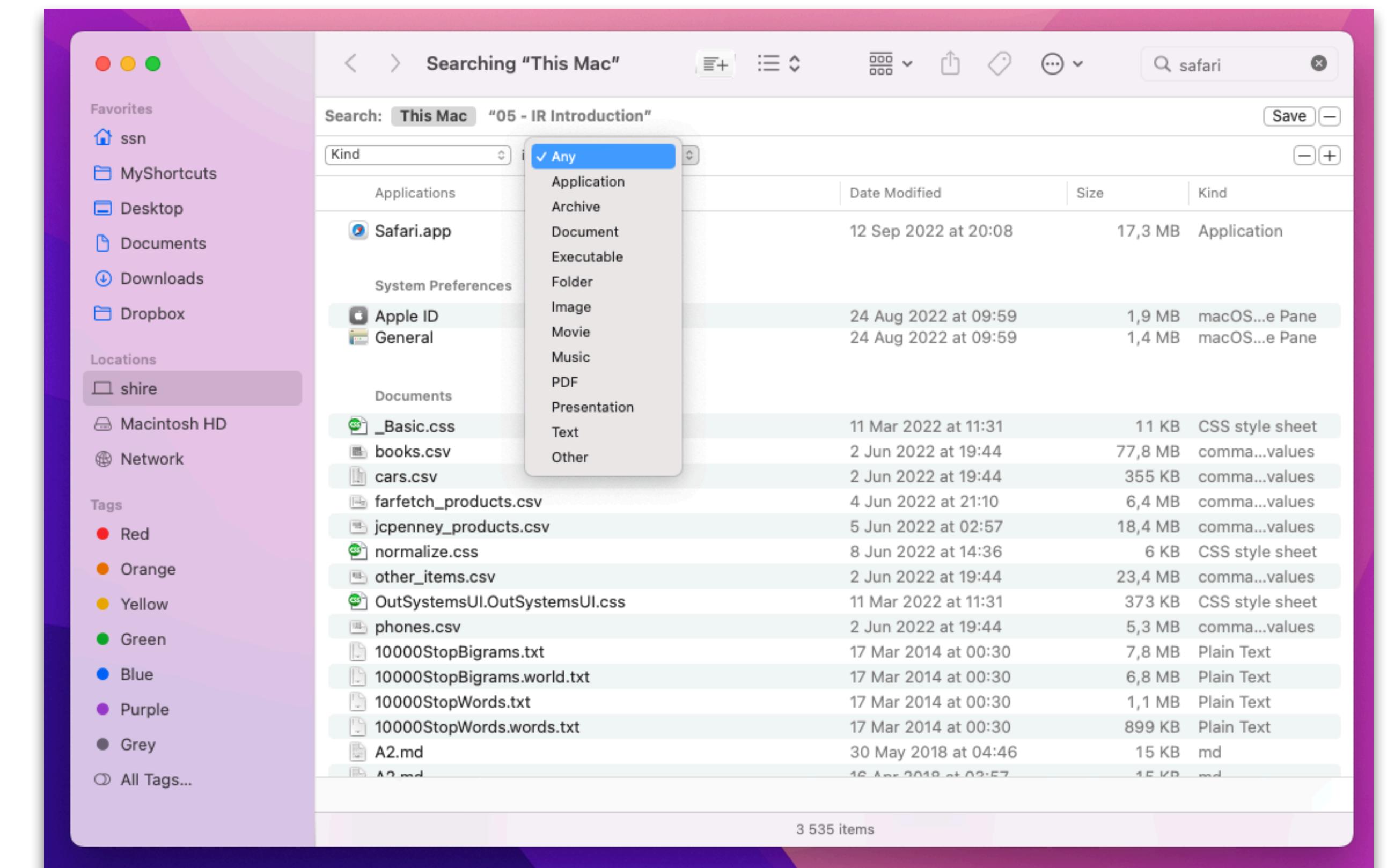
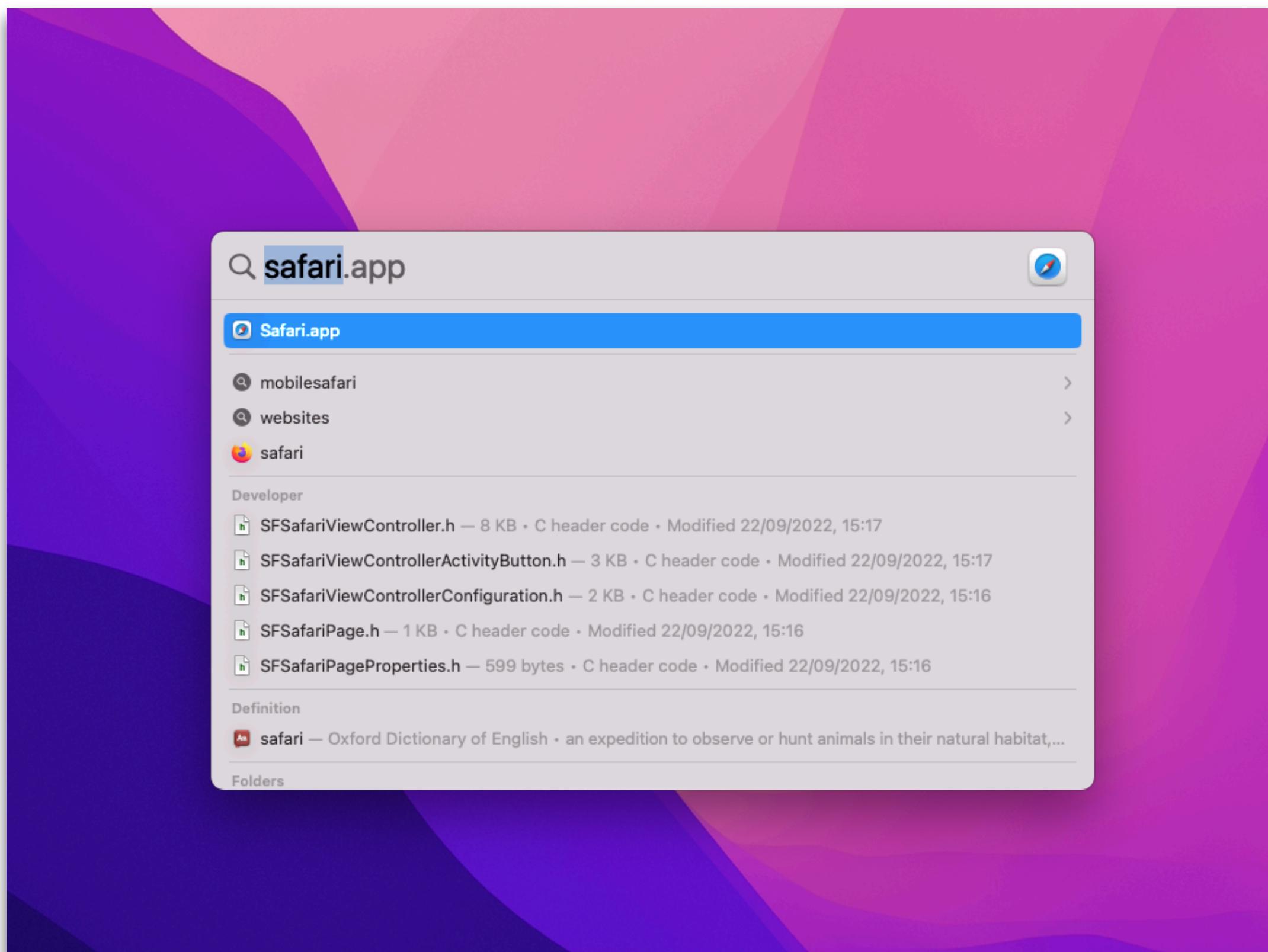
Nome:

Pesquisar

Mapa das Instalações

Centenário do Nascimento de Agustina Bessa-Luís

Clube de Leitura VAMOS À LER



amazon.com

All pencil

EN Hello, sign in Account & Lists Returns & Orders Cart 0

All Today's Deals Customer Service Registry Gift Cards Sell

1-48 of over 4,000 results for "pencil"

Sort by: Featured

Climate Pledge Friendly Climate Pledge Friendly

Department

Pencils

Woodcase Lead Pencils
Mechanical Pencils

Customer Reviews

★★★★★ & Up
★★★★★ & Up
★★★★★ & Up
★★★★★ & Up

Brands

BIC
 Amazon Basics
 Ticonderoga
 Paper Mate
 Pentel
 Dixon
 Dixon Ticonderoga
[See more](#)

Price

Under \$25
\$25 to \$50
\$50 to \$100
\$100 to \$200
\$200 & Above

Writing Instrument Point Type

Bold
 Broad
 Chisel
 Extra Fine
 Fine

RESULTS

Best Seller



Featured from our brands

Amazon Basics Woodcased #2 Pencils, Pre-sharpened, HB Lead, Box of 30

★★★★★ 74,896

\$5⁰³ (\$0.17/Count)

Ships to Portugal

Amazon brand

Amazon's Choice



Sponsored

Wood-Cased #2 HB Pencils, Yellow, Pre-sharpened, Class Pack, 320 pencils

★★★★★ 2,361

\$25⁴⁹ (\$0.08/Count) \$29.99

Ships to Portugal

Best Seller



Sponsored

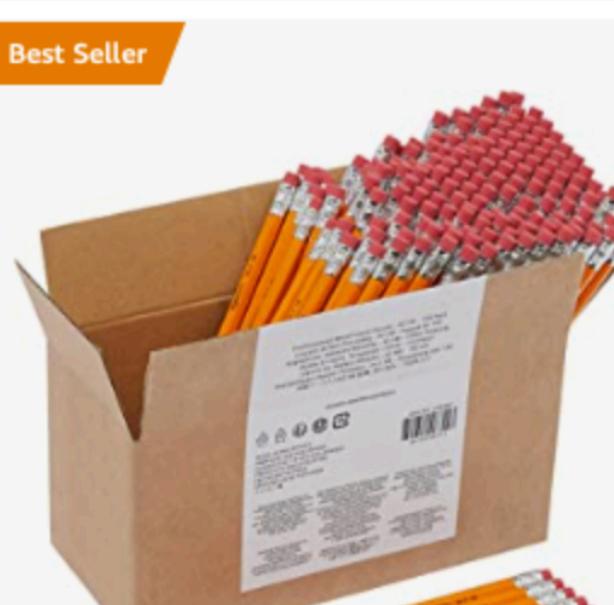
Wood-Cased #2 HB Pencils, Yellow, Pre-sharpened, Class Pack, 576 pencils in box by Madisi

★★★★★ 1,441

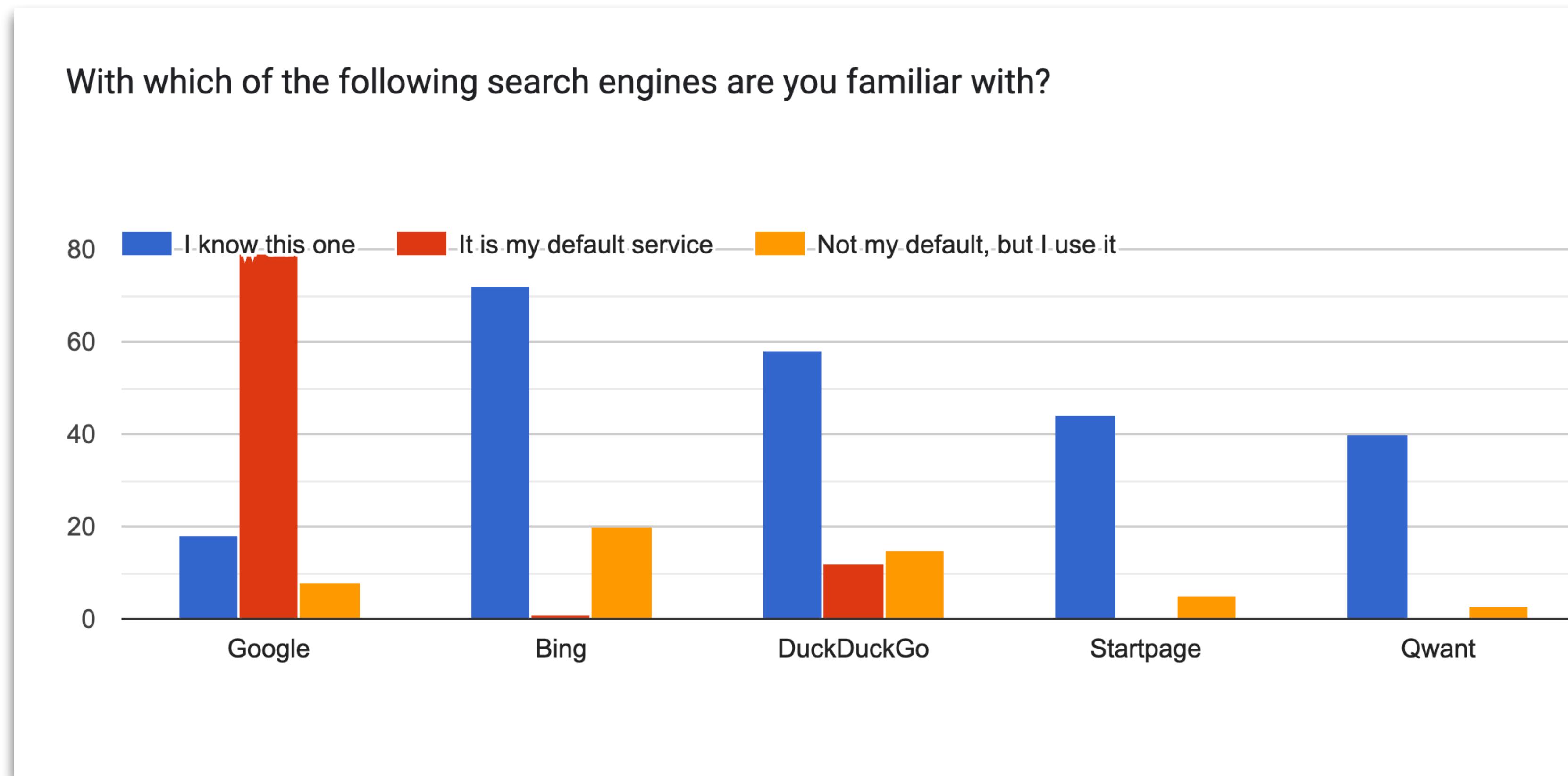
\$41⁹⁹ (\$0.07/Count)

Ships to Portugal

Best Seller



Student Survey 23/24





How Google Search Works

(in 5 minutes)

<https://www.youtube.com/watch?v=0eKVizvYSUQ>

Information retrieval concepts

Information Retrieval

- In computer science, the field that is most involved in search problems and developments is
 - [Information Systems > **Information Retrieval**], and includes sub-areas such as:
 - Document representation
 - Information retrieval query processing
 - Users and interactive retrieval
 - Retrieval models and ranking
 - Retrieval tasks and goals
 - Evaluation of retrieval results
 - Search engines architectures and scalability

Defining 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 search applications.
 - Primary focus since the 50s has been on text and documents.
- “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)”, Manning et al. (2008)

Documents

- Documents can be web pages, email messages, books, news articles, research papers, computer files (PDF, Word, text), text messages, etc.
- Common properties:
 - Significant text content.
 - Some structure, e.g. title, body, date, tags, author.
- From PRI projects:
 - linux packages, songs lyrics, cooking recipes, MSc dissertations, game plays, etc.

Different from Databases

- “A database management system is a software system that enables the creation, maintenance, and use of large amounts of data” (Abiteboul et al., 1995)
- Information retrieval and database systems have a lot in common.
- But emphasize different aspects of information management.
 - Databases contains highly structured data, queried using formal query languages, and the results are set-based – results are true or false for a given condition.
 - Information retrieval systems support interactive processes, support natural language queries, and results are rank-based – results are more or less relevant for a given information need. Not just text matching!

Information Retrieval and Databases

Table 1.1 Comparison of database systems and information retrieval, based on [40]

	Database systems	Information retrieval
Data type	Numbers, short strings	Text
Foundation	Algebraic/logic based	Probabilistic/statistics based
Search paradigm	Boolean retrieval	Ranked retrieval
Queries	Structured query languages	Free text queries
Evaluation criteria	Efficiency	Effectiveness (user satisfaction)
User	Programmer	Nontechnical person

Document vs. Record (tuple) Queries

- Example **database queries**
 - All students enrolled in course “ABC” in year 22/23.
 - All bank account transferences between account A and account B, within a given date interval, ordered by transferred amount.
- Example **search engine queries**
 - [feup enrolled students ABC 22/23]
 - [weather porto site:pt]

Information Retrieval Tasks

- Information retrieval applications and research are typically structured around tasks.
- Tasks address specific contexts, and specific information access problems.
- Examples of classical information retrieval tasks:
 - Ad-hoc search, find relevant documents for an arbitrary text query.
 - Filtering, identify relevant user profiles for a new document.
 - Classification, identify relevant labels for documents.
 - Question answering, give specific answer to a question.
- Yearly evaluation initiatives with classical and novel tasks at TREC
 - <https://trec.nist.gov/pubs/call2023.html>

Dimensions of Information Retrieval

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	
Music	P2P search	
	Literature search	

Central issues in information retrieval

Relevance

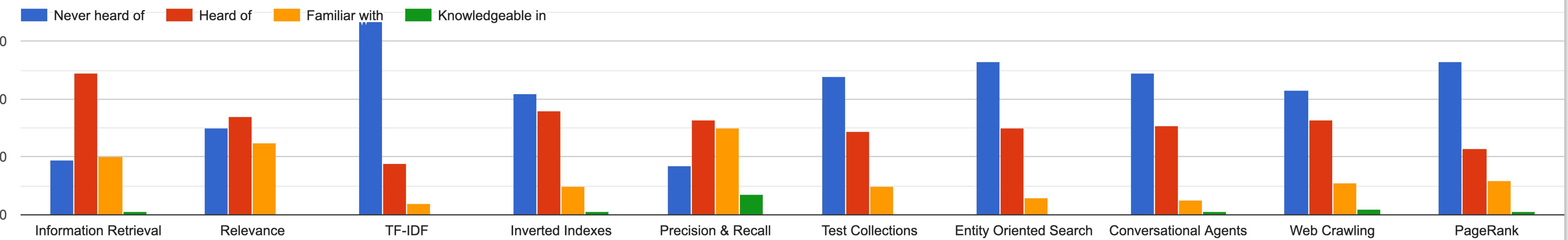
- The concept of relevance is a key issue in information retrieval.
- Many factors impact a person's decision of what is relevant,
e.g. task, context, previous knowledge, etc.
- Simple definition of relevance,
 - a relevant documents contains information that satisfies the information need of the user
that submitted a query to the search engine

Evaluation

- Evaluation setups are based on the use of annotated test collections of documents, queries, and relevance judgements.
- Evaluation in information retrieval dates back to the 1960s (Cranfield paradigm).
- IR evaluation methodologies have been adopted by many fields.
 - Recall and precision are two examples of widely used measures.
- Information needs drive the search process and result from the underlying task users' have.

Student Survey - IR Concepts

With which of the following search concepts are you familiar with ?



Information Retrieval and Search Engines

- Search engines are practical applications of information retrieval techniques.
- Existing open-source and commercial solutions that support development and research:
 - Apache Solr, Elasticsearch, Open Search, Terrier.

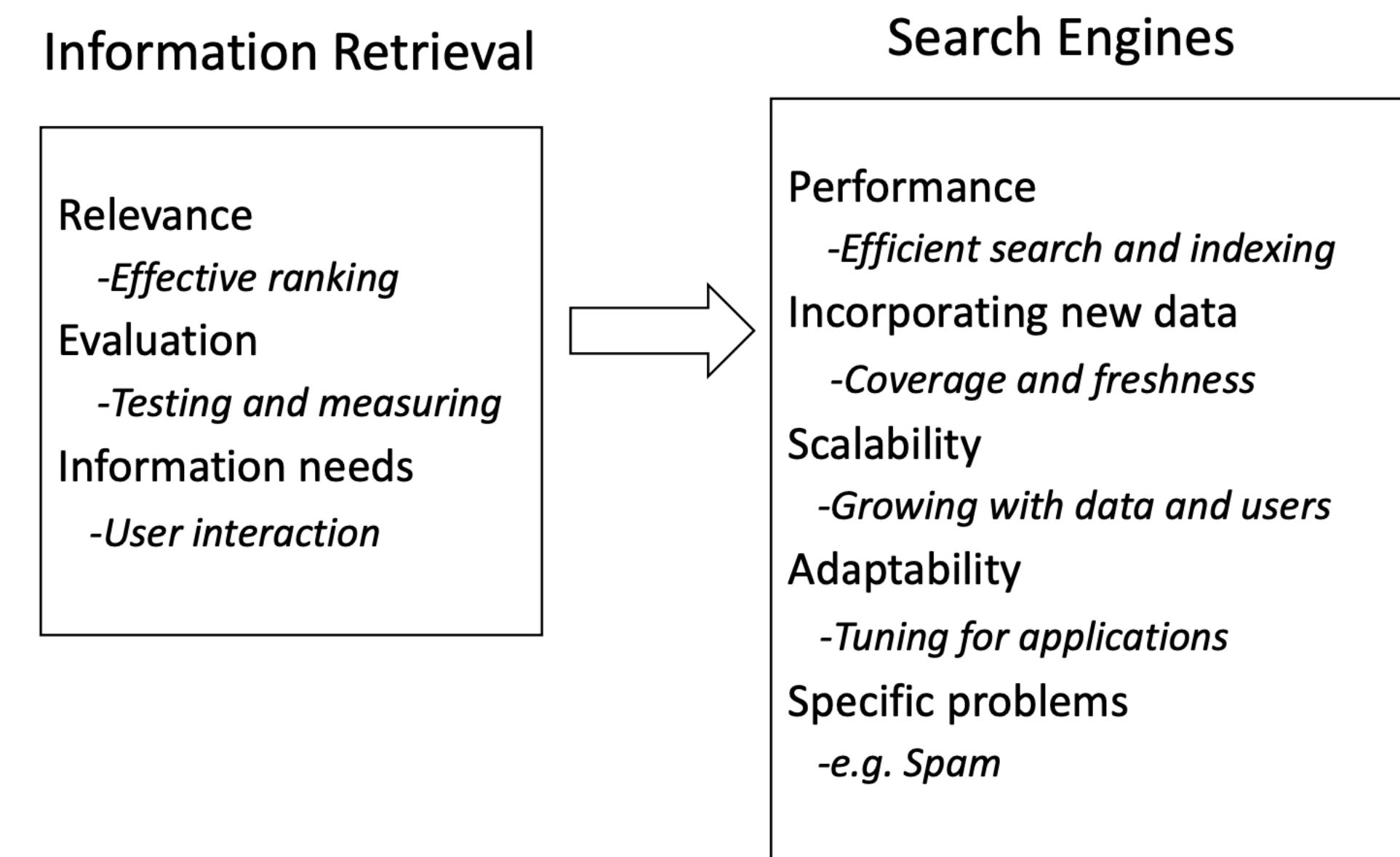
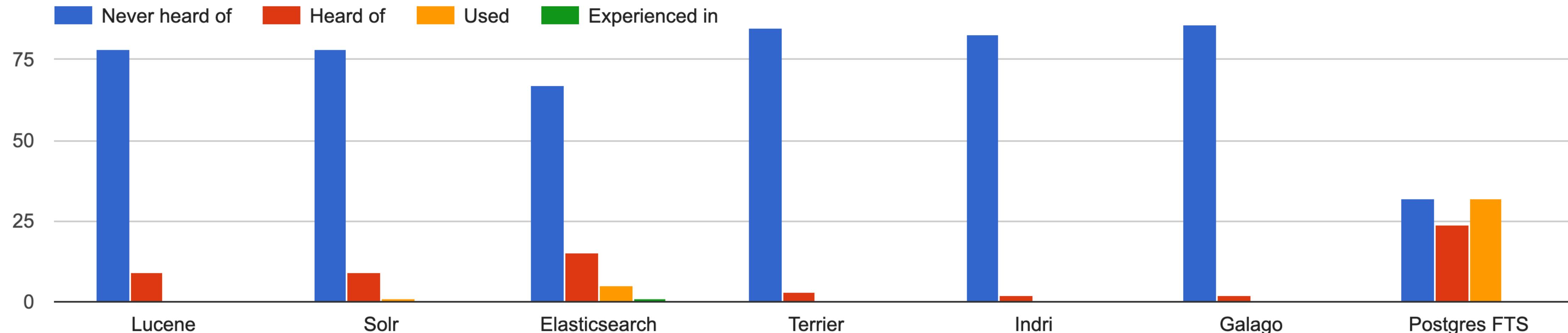


Image from *Search Engines: Information Retrieval in Practice*, Croft et al. (2009)

Student Survey - IR Technologies

With which of the following search tools are you familiar with ?

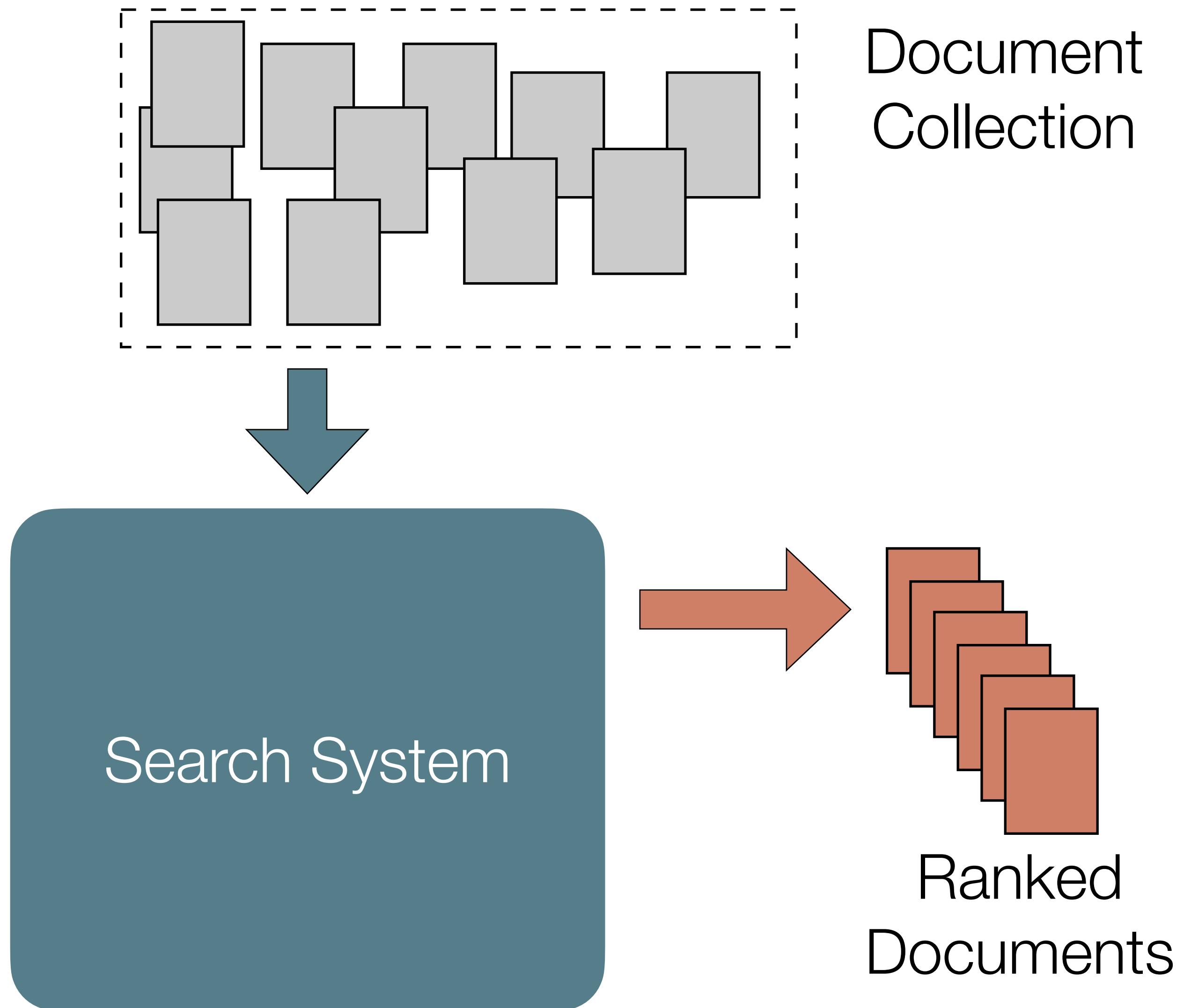


Architecture of information retrieval systems

Classic Information Retrieval

Information Need
When did the latest lunar
eclipse occurred?

Query
[latest eclipse]



Architecture of an Information Retrieval System

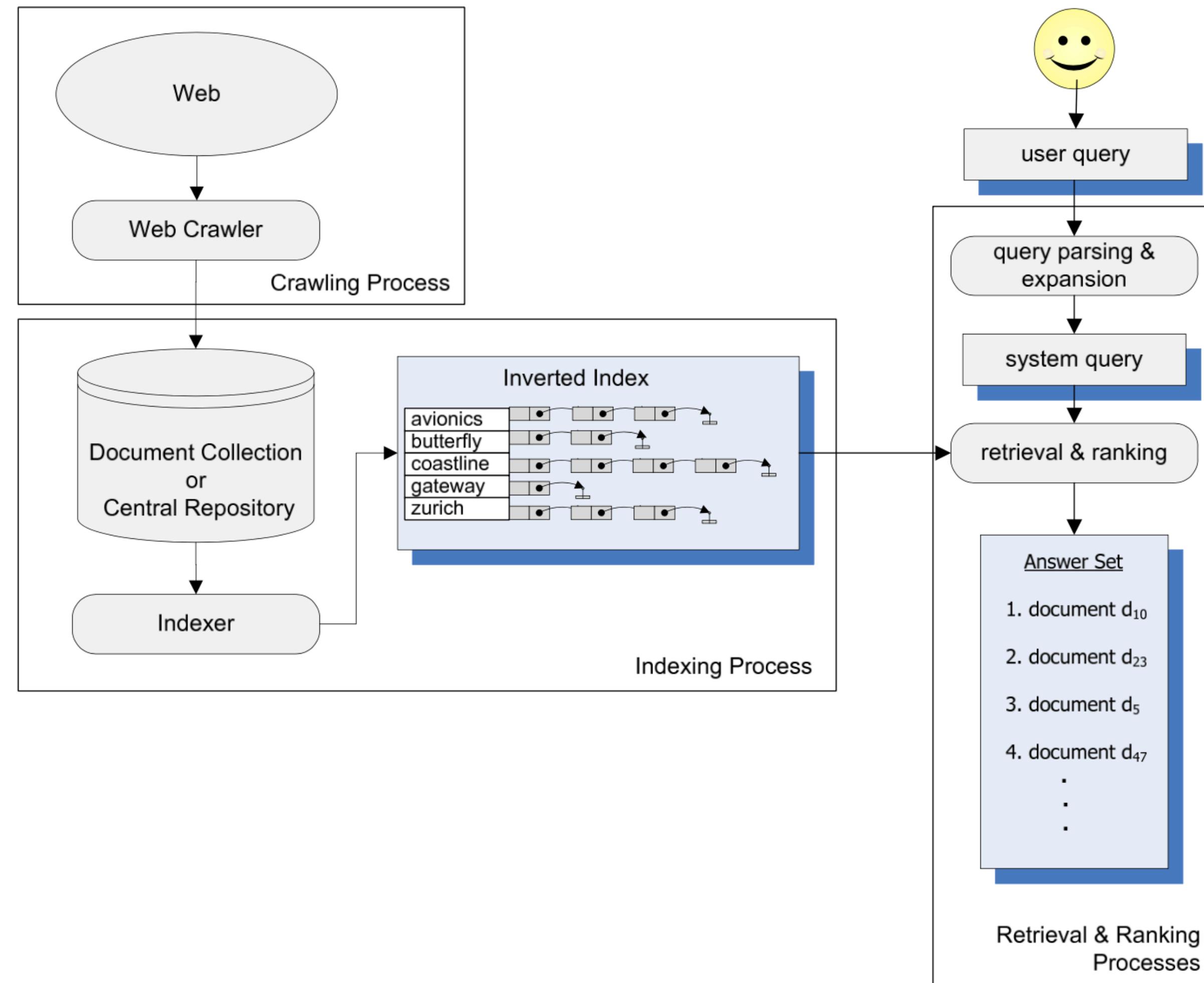


Image from Modern Information Retrieval, Baeza-Yates et al. (2011)

Indexing Process

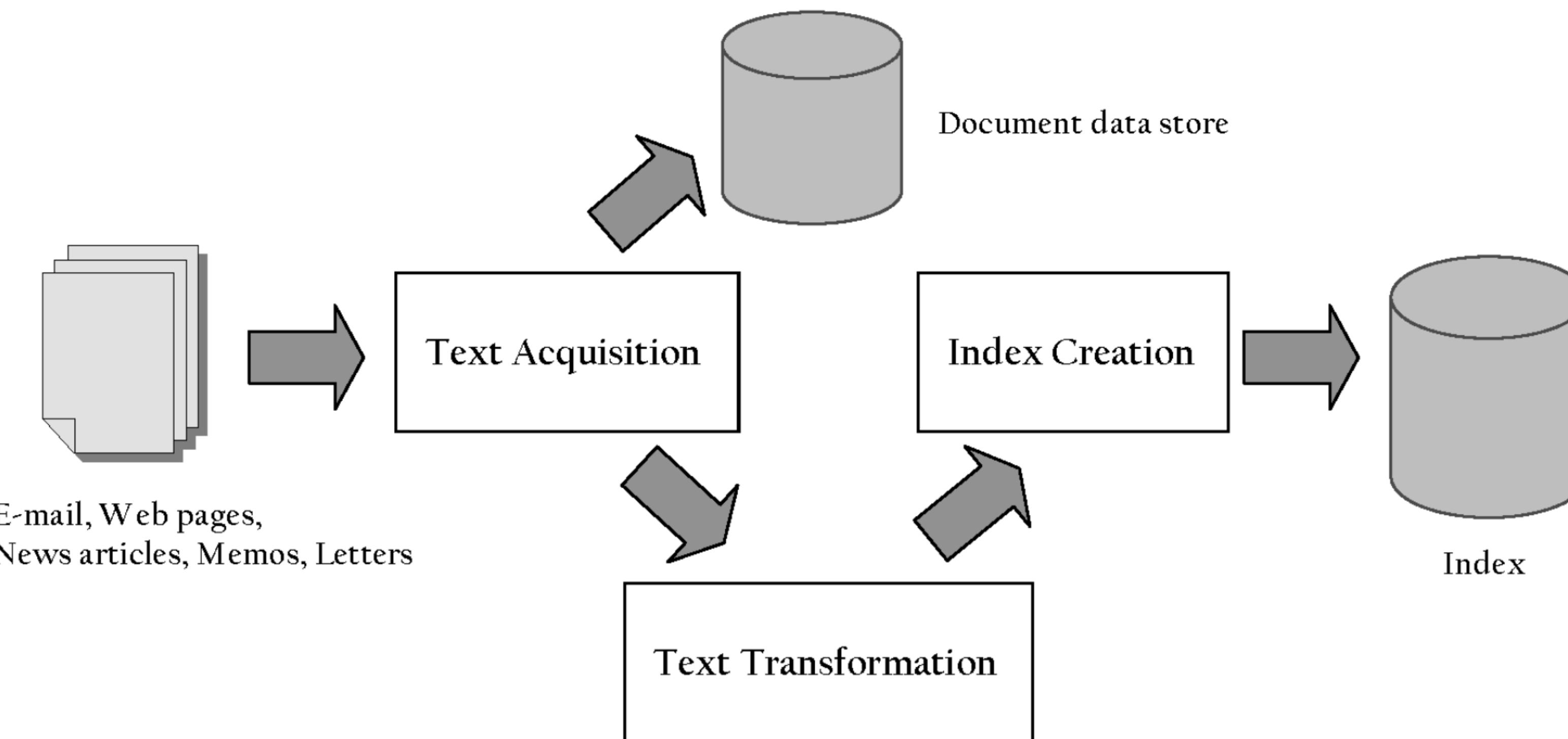


Image from Search Engines: Information Retrieval in Practice, Croft et al. (2009)

Query Process

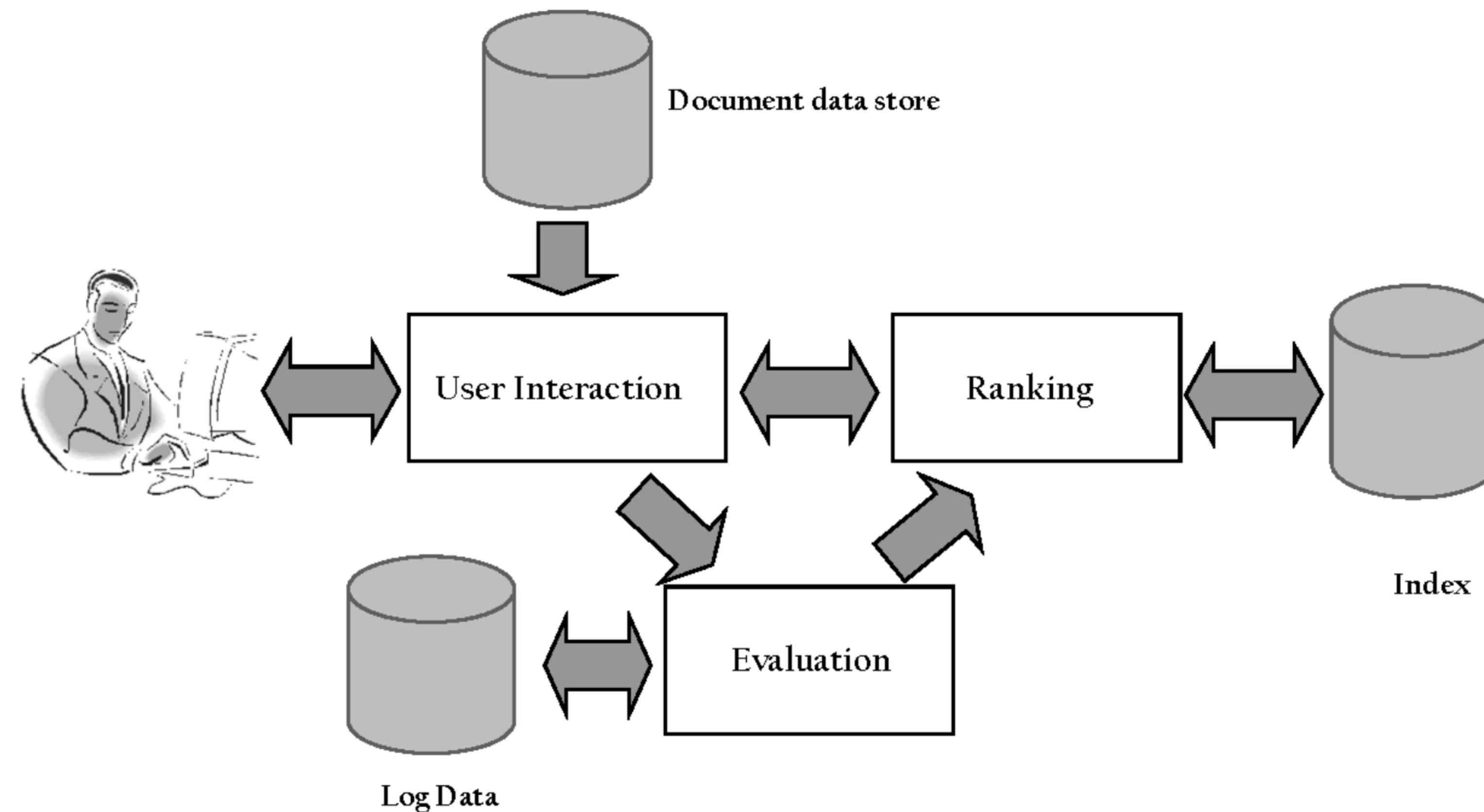


Image from Search Engines: Information Retrieval in Practice, Croft et al. (2009)

History of information retrieval

Historic Context

- At the end of World War II, Vannevar Bush, who headed the R&D office during the war, reflected on applications of new technologies in peace times.
- He wrote a landmark essay entitled “As We May Think” in 1945, where a new device, the Memex, capable of managing and interlinking knowledge is described.
 - <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>
- This essay greatly influenced the next generation of innovators.
 - Douglas Engelbart and the development of the computer mouse (“the mother of demos”), 1968.
 - Ted Nelson and the development of the concept of hypertext (Project Xanadu), 1960s.
 - Tim Berners-Lee and the World Wide Web, in 1989.
 - Web search engines, such as Lycos (94), Altavista (95), and Google (98).

Searching the Web

- The web represented a unique search challenge.
- A unique document collection, millions of documents, connected through hyperlinks, and distributed through multiple repositories (required crawling).
- The size of the collection and the user base, millions of documents and users.
Performance and scalability became a key issue.
- Predicting relevance in such a large and heterogeneous collection is complex. New sources of evidence appeared (e.g. hyperlinks).
- The web is also a business platform, thus users needs are not limited to text search.
- Web spam is a real problem on the web, sometimes referred to as “adversarial information retrieval”.

Evolution in Web Search

Google mp3 pure 10 results All Languages

Google Search I'm Feeling Lucky

Search Tips Language Options Email These Results

Sponsored Links

[Click here to find MP3 files. Listen, Share and Store them for free!](#)

<http://www.myplay.com> Sign up Today for a myplay Locker and get 3GB of Free Storage!

[Music - 100's of songs to choose from! Stop searching. Start listening.](#)

www.firstlook.com Where tomorrow's music is heard first! Firstlook.com

Google results 1-10 of about 23,997 for mp3 pure. Search took 0.34 seconds.

Category : [Arts > Music > Sound Files > MP3 > Link Lists](#)

[PureMP3 - If you want pr0n, go to a pr0n site!](#)

...Welcome to Pure MP3 Pure 1 : Separate from all...

Description: Organization of over 1,500 MP3 sites that pledge to keep their sites free of adult...

Category: [Arts > Music > Sound Files > MP3 > Link Lists](#)

www.puremp3.org/ - Show matches - 17k - Similar Pages

www.mtvhits.de/pureoldiemp3z

[Similar Pages](#)

[The MP3 Master List | Pure downloads](#)

Reliable mp3 downloads for over 100 popular new songs. Comprehensive instructions

Description: Download current, popular music in CD-quality mp3 audio. Multiple download locations...

Category: [Arts > Music > Sound Files > MP3](#)

www.mp3masterlist.com/ - Show matches - 4k - Similar Pages

[dmusic.com - MUSIC IS FREE.](#)

...News Media MP3 News, Music News, Newsletters ... Pure MP3...

Description: MP3 portal with info, news, reviews, a search engine and legal MP3 downloads.

Category: [Arts > Music > Sound Files > MP3](#)

www.dmusic.com/ - Show matches - 27k - Similar Pages

Google Fox News

All News Books Videos Shopping More Settings Tools

Axios > People also search for > Fox News

POLITICO CNN MSNBC NBC NEWS FOX NEWS Bloomberg BuzzFeed PROPUBLICA NBC NEWS CNBC BUSINESS INSIDER

Fox News

Top stories

Michelle Obama says Lean In is '---' that doesn't work during book tour stop in New... Fox News 1 hour ago

Joy Behar Invites Kid Rock on 'The View' After Fox News Remarks Variety 2 days ago

Former Kansas City Chiefs running back Kareem Hunt admits to lying about assault Fox News 1 hour ago

More for Fox News

Fox News - Breaking News Updates | Latest News Headlines | Photos ... https://www.foxnews.com/ Breaking News, Latest News and Current News from FOXNews.com. Breaking news and video. Latest Current News: U.S., World, Entertainment, Health, ...

US Fox news features local news and national news articles and videos.

Politics Presidential politics and political news from foxnews.com. News ... More results from foxnews.com

World Latest world news, international news, world video, international ...

Watch Live Watch Fox News Channel, Fox Business Network, and ...

Videos

Watch Fox News Channel and Fox Business Network Online Fox News - Dec 30, 2017

US Fifth Fleet commander found dead in Bahrain Fox News - 4 hours ago

Lewandowski rips media, Trump aides | On Air Videos Fox News - 3 hours ago

Fox News (@FoxNews) · Twitter https://twitter.com/FoxNews

@cathymcmorris: "We didn't need to be in this situation. @SenSchumer and @TheDemocrats in the Senate decided to take a different course." #shutdown pic.twitter.com/1z4P9cE...

Jan 20, 2018 Twitter

Fox News Live 1996 - 2008 The Fire Since 2011 Fox & Friends Since 1998

Profiles Facebook Twitter YouTube

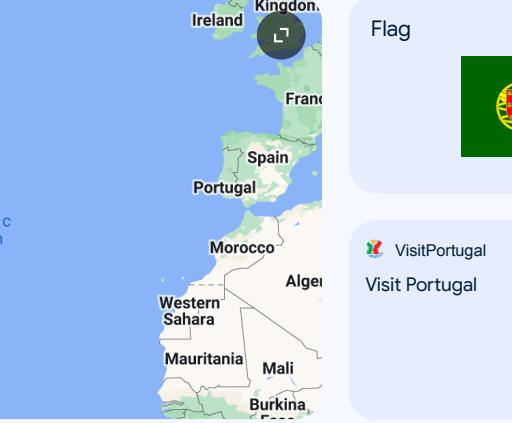
Web Search

Google portugal

About 5,420,000,000 results (0.40 seconds)

Portugal

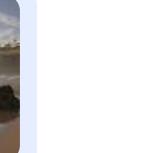
Country :



Flag



VisitPortugal Visit Portugal



[Wikipedia](#) https://pt.wikipedia.org/wiki/Por... · [Translate this page](#)

Portugal – Wikipédia, a encyclopédia livre

Portugal é a nação mais a oeste do continente europeu. O nome do país provém da sua segunda maior cidade, Porto, cujo nome latino-celta era Portus Cale.

[Etimologia](#) · [Geografia](#)

[Wikipedia](#) https://en.wikipedia.org/wiki/Portugal · [Translate this page](#)

Portugal

Portugal officially the Portuguese Republic is a country located on the Iberian Peninsula, in Southwestern Europe, and whose territory also includes the ...

[Kingdom of Portugal](#) · [Portugal \(disambiguation\)](#) · [Continental Portugal](#) · [Lisbon](#)

Related questions

- Quem tem mais poder em Portugal?
- Como era chamado Portugal antigamente?
- O que caracteriza Portugal?
- Qual é o tipo de turismo mais praticado em Portugal?

Places to visit

Lisbon

Alfama, São Jorge Castle & fado music
€83 1h

Porto

Ribeira district & port wine cellars
1 hr

Madeira

Volcanic islands, Funchal & Madeira wine
€38 2h

Feedback

People also search for



Lisbon Spain Porto Madeira

See more →

Things to do



National Palace of Pena
4.4 ★ (80K)
Castle
€13.30



Belém Tower
4.6 ★ (84K)
Tourist attraction
€6.00



Castelo de S. Jorge
4.4 ★ (81K)
Castle
€15.92

More things to do →

Top stories

 Negócios
Precisamos de falar sobre o compadrio em Portugal



SAPODESPORTO
Portugal defronta Fiji (apenas) em busca da primeira vitória no Mundial...

1 hour ago

 SIC Notícias
"Quero agradecer a Portugal. Estamos unidos contra o terrorismo"



RTP
Woody Allen diz que gostava de filmar em Portugal

1 hour ago

 Portugal.com
https://www.portugal.com · [Translate this page](#)

Portugal.com - #1 Guide to Portugal

Portugal.com brings you to Portugal. Read our city guides, food & drink recommendations, learn about Portuguese history and more.



 Tripadvisor
https://www.tripadvisor.com/Tourism-g189100-Port... · [Translate this page](#)

Portugal 2023: Best Places to Visit

Portugal Tourism: TripAdvisor has 7723075 reviews of Portugal Hotels, Attractions, and Restaurants making it your best Portugal resource.



 VisitPortugal
https://www.visitportugal.com/pt-pt · [Translate this page](#)

Visit Portugal

Conhecer Portugal através dos seus vinhos é uma forma autêntica e original de abraçar a cultura do país. Para os apreciar e conhecer, nada como visitar as ...



 XXIII Governo
https://www.portugal.gov.pt/... · [Translate this page](#)

República Portuguesa

Portugal bate recorde de estudantes no Ensino Superior. Número de inscritos ultrapassou os 446 mil, atingindo o valor mais alto de sempre. 2023-10-02 às 8h00.

 Portal Diplomático
https://portaldiplomatico.mne.gov.pt/... · [Translate this page](#)

Sobre Portugal

Portugal, oficialmente República Portuguesa, é um Estado da Europa Meridional, fundado em 1143, que ocupa uma área total de 92.212 Km². A parte continental ...



 europa.eu
https://european-union.europa.eu/... · [Translate this page](#)

Portugal – perfil do país membro da UE| União Europeia

Capital: Lisboa. Língua oficial da UE: português. Adesão à UE: 1 de janeiro de 1986. Moeda: euro. Membro da zona euro desde 1 de janeiro de 1999.



 Turismo de Portugal
https://turismodeportugal.pt · [Translate this page](#)

Turismo de Portugal

Bem-vindo ao Turismo de Portugal. Agenda do Turismo para o Interior. Agenda para a Valorização dos Profissionais do Turismo. Plano Turismo +Sustentável ...



 Instagram
https://www.instagram.com/portugal · [Translate this page](#)

portugal - Instagram photos and videos

Beyond Web Search

- Web Search is an important area for Information Retrieval, but not all.
- Information Retrieval is also includes:
 - Recommendations;
 - Question answering;
 - Ad retrieval;
 - Enterprise search;
 - Many more...

Information Retrieval Summary

- Information Retrieval dates back to 1940s.
- Well established field in Computer Science.
- Deals with a wide range of information access problems — from systems design to user interaction; from theoretical to applications.
- Organized in tasks: web search, filtering, question answering, podcast search, conversational, enterprise / intranet search, entity-oriented search
- Strong links with information science, NLP, machine learning, databases, etc.
- Core concepts: information need, index, query, relevance.

Course Plan

Course Plan

- Information Retrieval Overview (**today**)
- Evaluation
- Models and Indexing
- Retrieval and Web Retrieval
- Query Processing
- Entity-Oriented Search
- Neural IR / Semantic Search

Lab Classes

→ Tutorials on Solr and IR Evaluation

→ Develop M2: Information Retrieval

→ Index

→ Retrieve

→ Evaluate

References and Further Reading

- Modern Information Retrieval (2nd Edition) (2011)
Ricardo Baeza-Yates, Berthier Ribeiro-Neto. Addison-Wesley Professional.
<http://www.mir2ed.org>
- Introduction to Information Retrieval (2008)
Christopher Manning, Prabhakar Raghavan, Hinrich Schütze. Cambridge University Press.
<http://nlp.stanford.edu/IR-book>
- Search Engines: Information Retrieval in Practice (2009)
W. Bruce Croft, Donald Metzler, Trevor Strohman. Pearson.
<http://ciir.cs.umass.edu/downloads/SEIRiP.pdf>
- Entity-Oriented Search (2018)
Krisztian Balog. Springer One.
<https://link.springer.com/book/10.1007/978-3-319-93935-3>