UFMG UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Information Retrieval

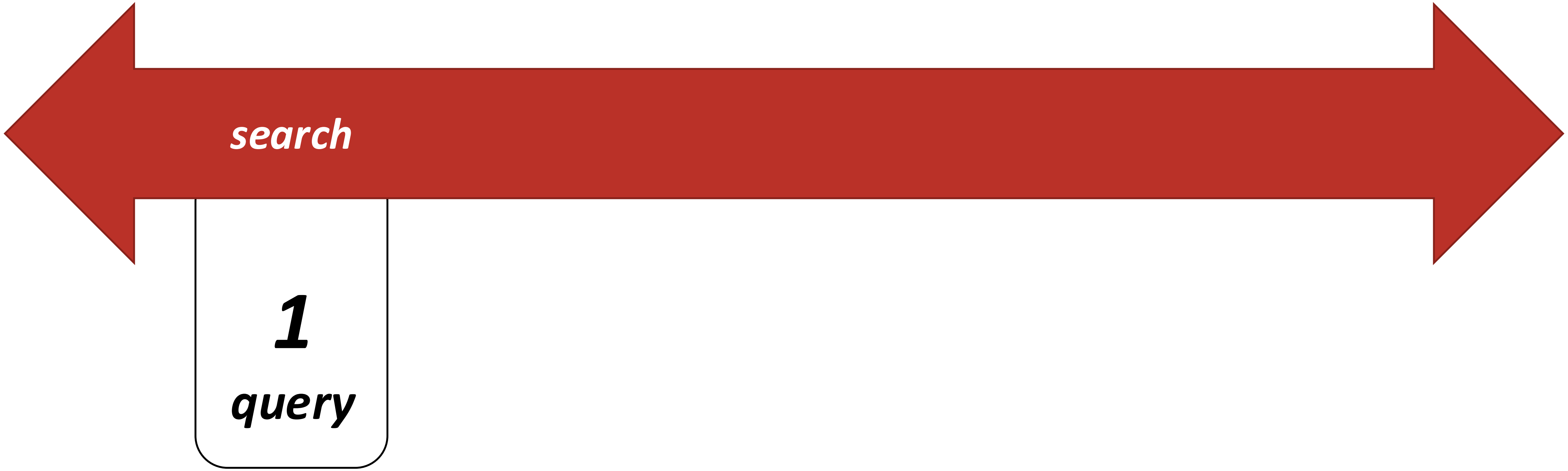
Introduction

Rodrygo L. T. Santos
rodrygo@dcc.ufmg.br

Information retrieval

> *Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information.*
>
> - Gerard Salton, 1968

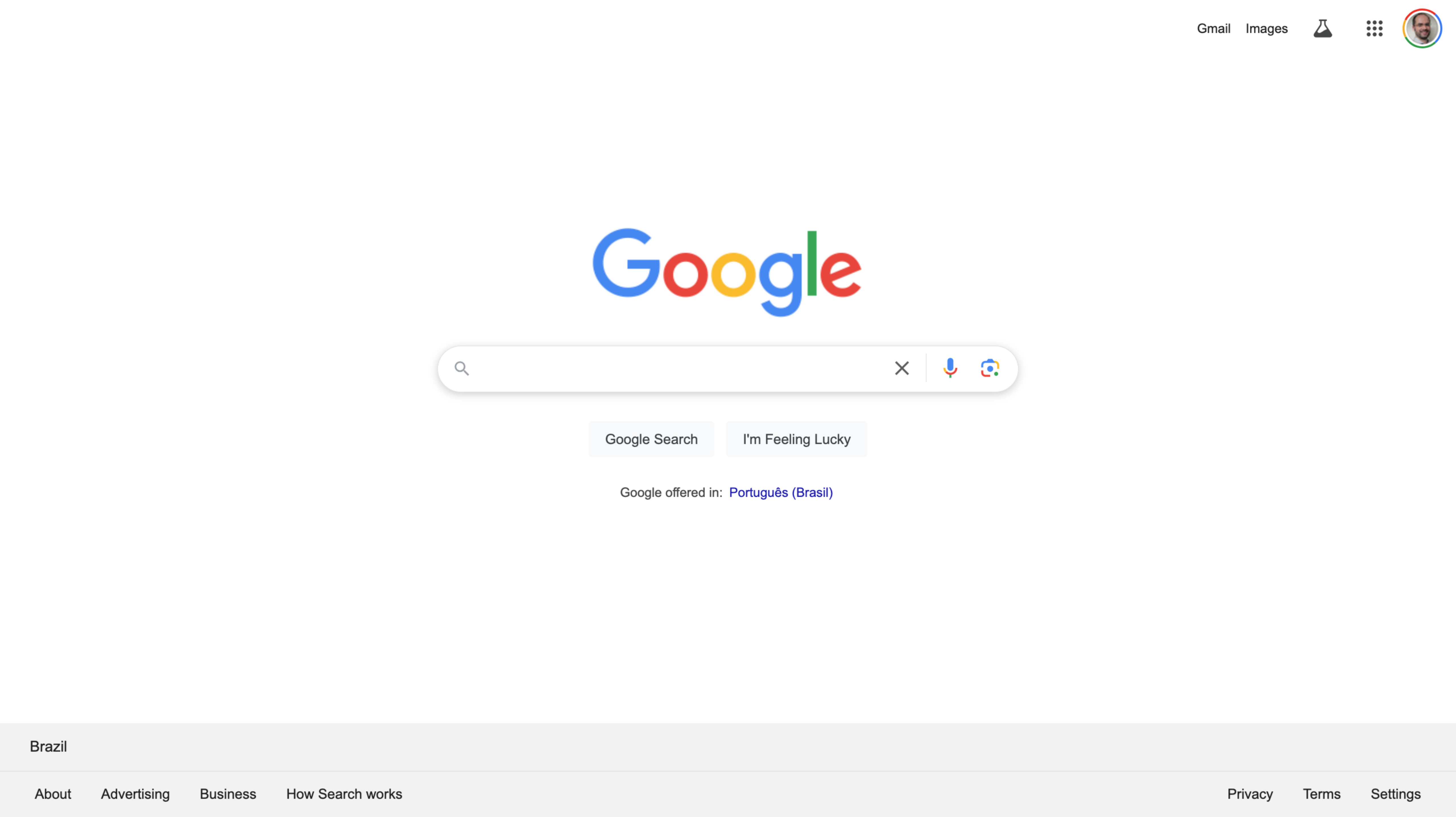
Retrieval tasks



search

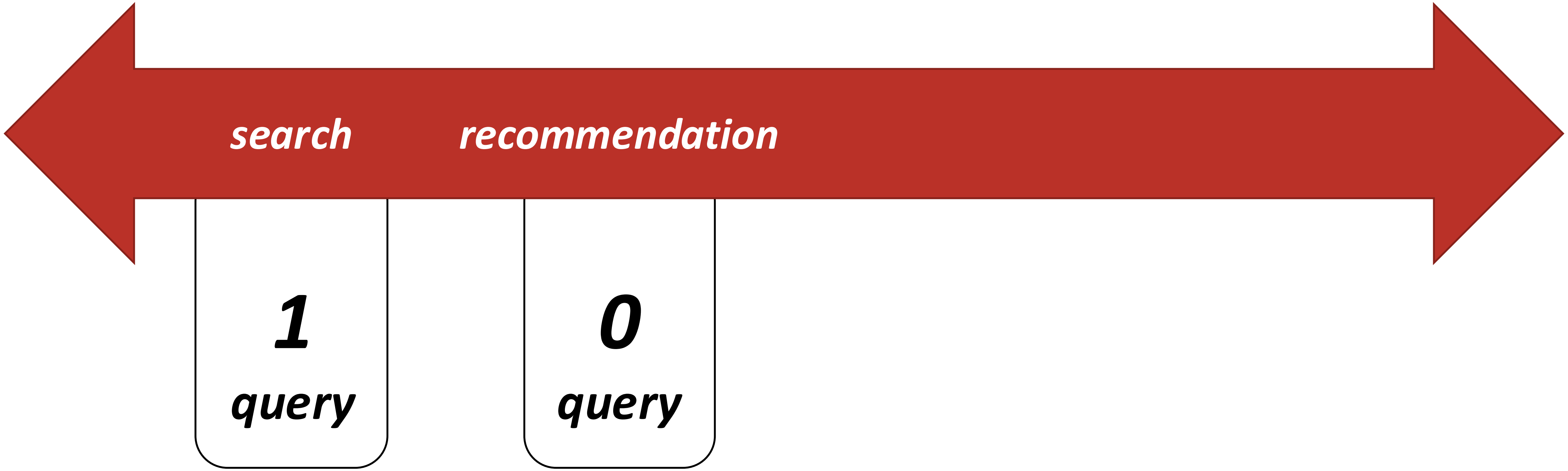
1

query



Google offered in: Português (Brasil)

Retrieval tasks



search

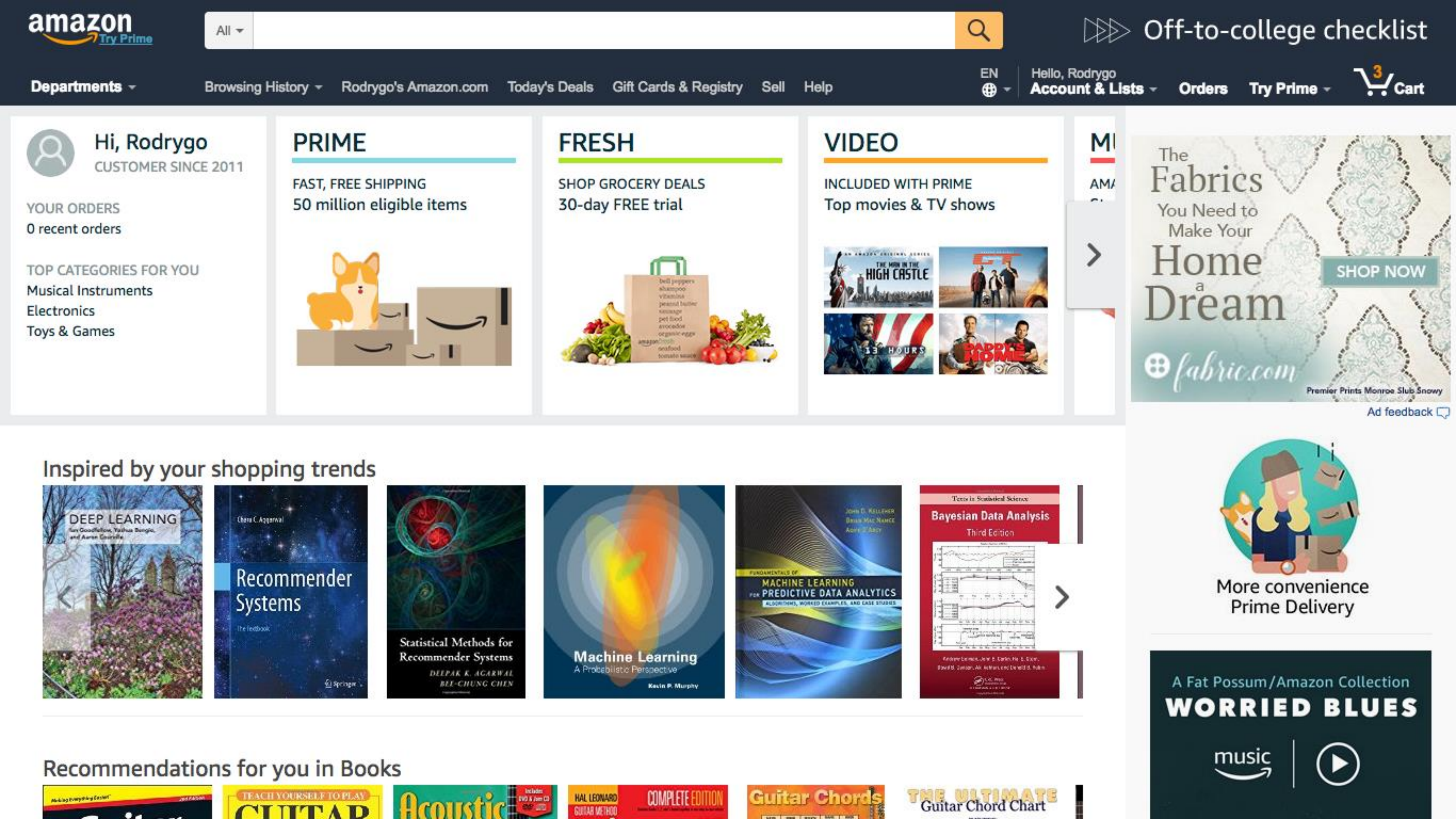
1

query

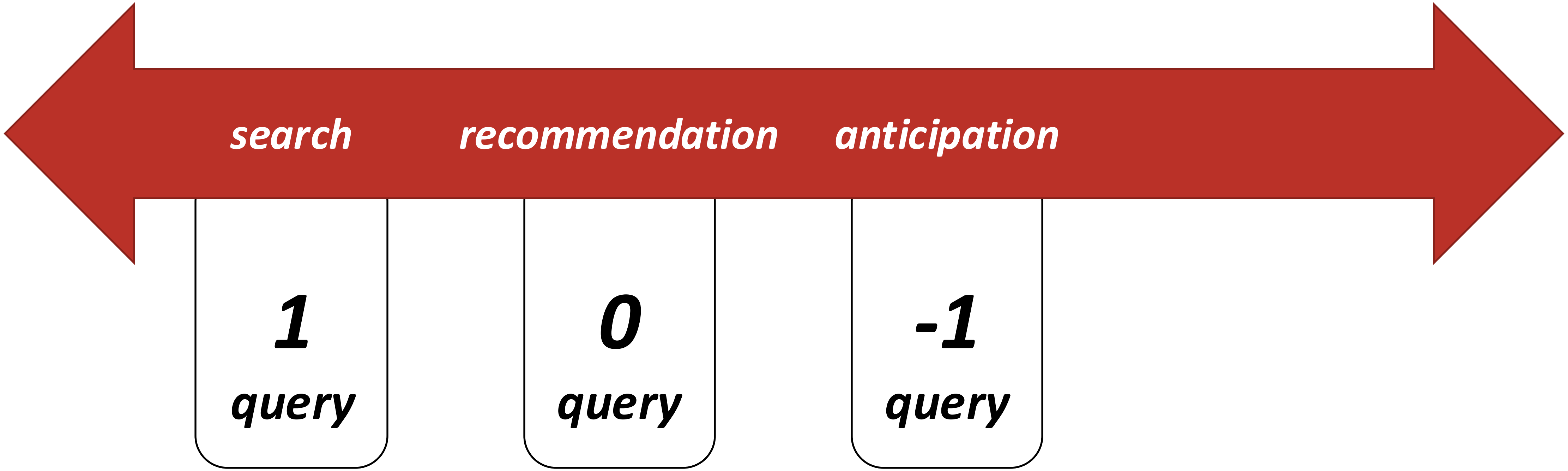
recommendation

0

query


amazon
Try Prime
All
Off-to-college checklist
Departments
Browsing History
Rodrygo's Amazon.com
Today's Deals
Gift Cards & Registry
Sell
Help
EN
Hello, Rodrygo
Account & Lists
Orders
Try Prime
3
Cart
Hi, Rodrygo
CUSTOMER SINCE 2011
YOUR ORDERS
0 recent orders
TOP CATEGORIES FOR YOU
Musical Instruments
Electronics
Toys & Games
PRIME
FAST, FREE SHIPPING
50 million eligible items
FRESH
SHOP GROCERY DEALS
30-day FREE trial
VIDEO
INCLUDED WITH PRIME
Top movies & TV shows
THE MAN IN THE HIGH CASTLE
13 HOURS
DADDY'S HOME
The Fabrics You Need to Make Your Home a Dream
SHOP NOW
fabric.com
Premier Prints Monroe Slub Snowy
Ad feedback
Inspired by your shopping trends
DEEP LEARNING
Ian Goodfellow, Yoshua Bengio, and Aaron Courville
Charu C. Aggarwal
Recommender Systems
The Textbook
Springer
Statistical Methods for Recommender Systems
DEEPAK K. AGARWAL
BEE-CHUNG CHEN
Machine Learning
A Probabilistic Perspective
Kevin P. Murphy
John D. Kelleher
Brian Mac Namee
Aoife D'Arcy
FUNDAMENTALS OF MACHINE LEARNING FOR PREDICTIVE DATA ANALYTICS
ALGORITHMS, WORKED EXAMPLES, AND CASE STUDIES
Texts in Statistical Science
Bayesian Data Analysis
Third Edition
More convenience
Prime Delivery
A Fat Possum/Amazon Collection
WORRIED BLUES
music
Recommendations for you in Books
TEACH YOURSELF TO PLAY GUITAR
Acoustic
HAL LEONARD
GUITAR METHOD
COMPLETE EDITION
Guitar Chords
THE ULTIMATE Guitar Chord Chart

Retrieval tasks



search

1

query

recommendation

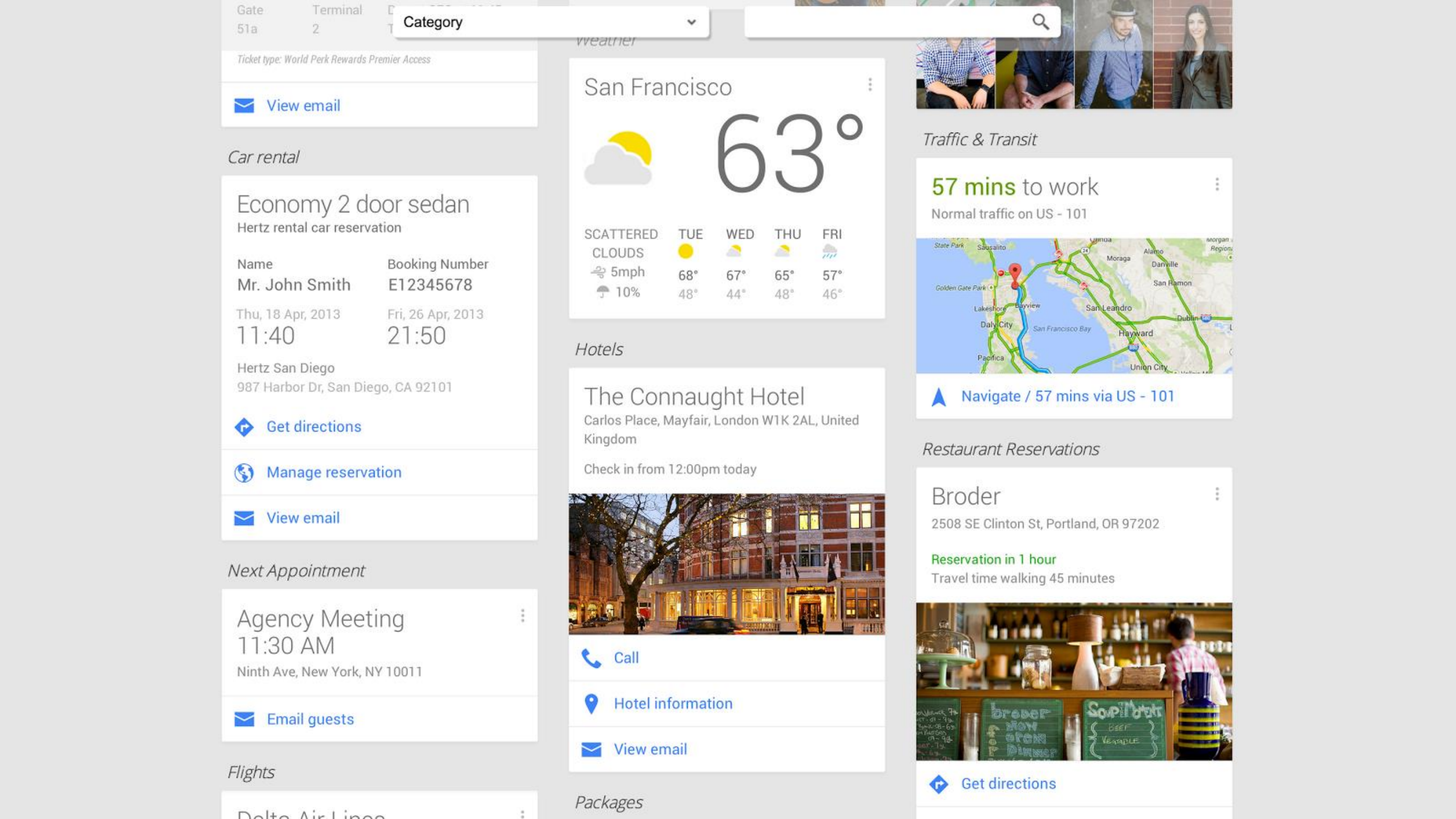
0

query

anticipation

-1

query



Ticket type: World Perk Rewards Premier Access

✉ View email

Car rental

Economy 2 door sedan
Hertz rental car reservation

Name	Booking Number
Mr. John Smith	E12345678
Thu, 18 Apr, 2013 11:40	Fri, 26 Apr, 2013 21:50

Hertz San Diego
987 Harbor Dr, San Diego, CA 92101

Get directions

Manage reservation

✉ View email

Next Appointment

Agency Meeting
11:30 AM
Ninth Ave, New York, NY 10011

✉ Email guests

Flights

Delta Air Lines

Weather

San Francisco

63°

SCATTERED CLOUDS	TUE	WED	THU	FRI
5mph	68°	67°	65°	57°
10%	48°	44°	48°	46°

Hotels

The Connaught Hotel
Carlos Place, Mayfair, London W1K 2AL, United Kingdom

Check in from 12:00pm today

☎ Call

Hotel information

✉ View email

Packages

Traffic & Transit

57 mins to work
Normal traffic on US - 101

Navigate / 57 mins via US - 101

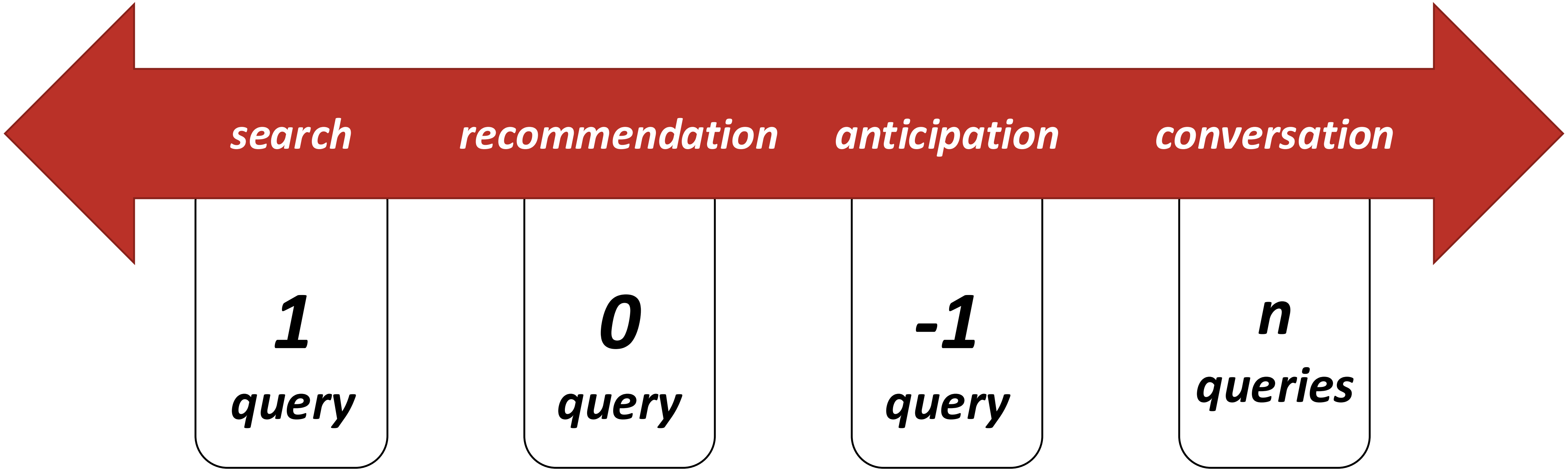
Restaurant Reservations

Broder
2508 SE Clinton St, Portland, OR 97202

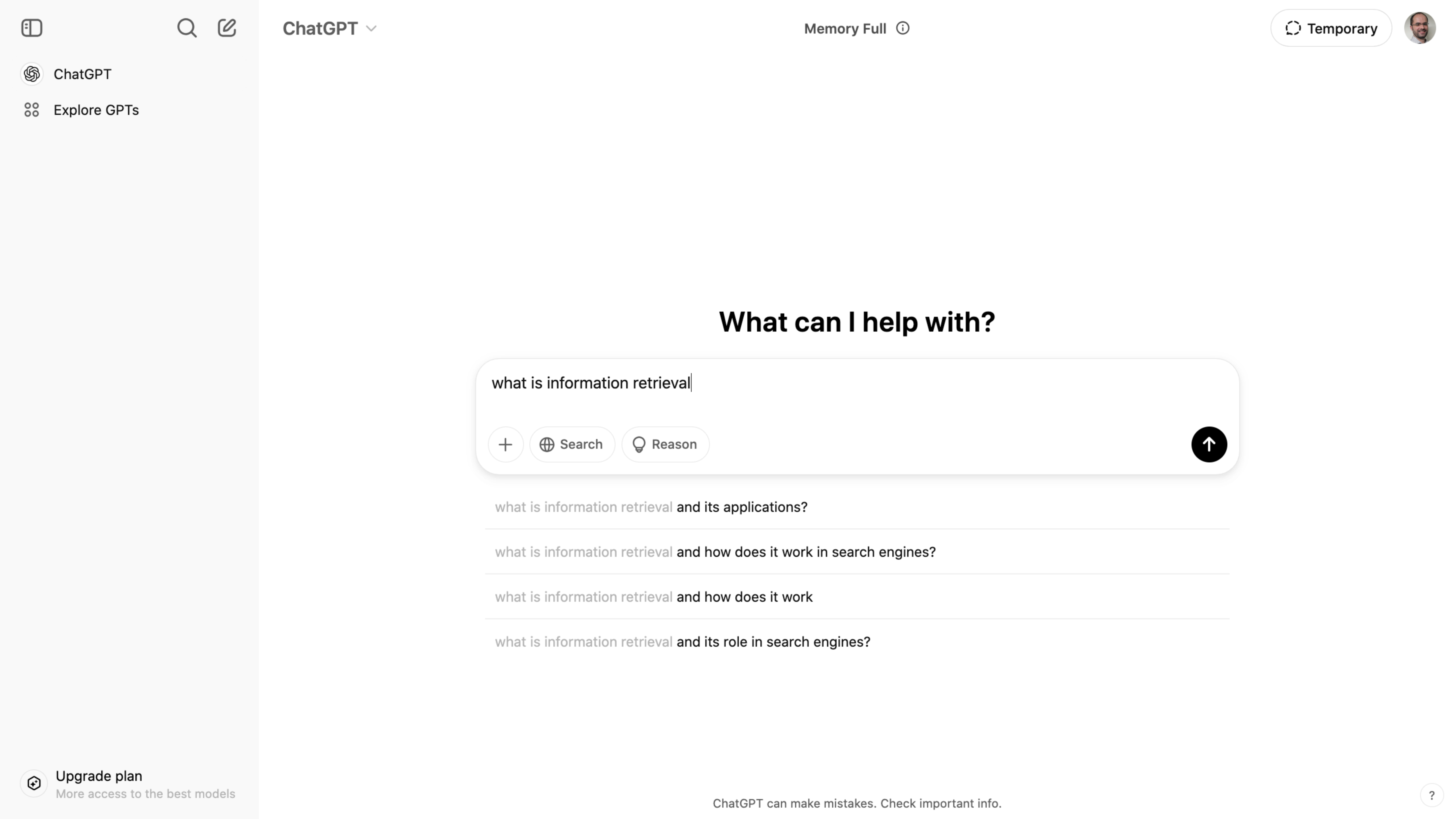
Reservation in 1 hour
Travel time walking 45 minutes

Get directions

Retrieval tasks



search	*recommendation*	*anticipation*	*conversation*
1 ***query***	***0*** ***query***	***-1*** ***query***	***n*** ***queries***



What can I help with?

what is information retrieval

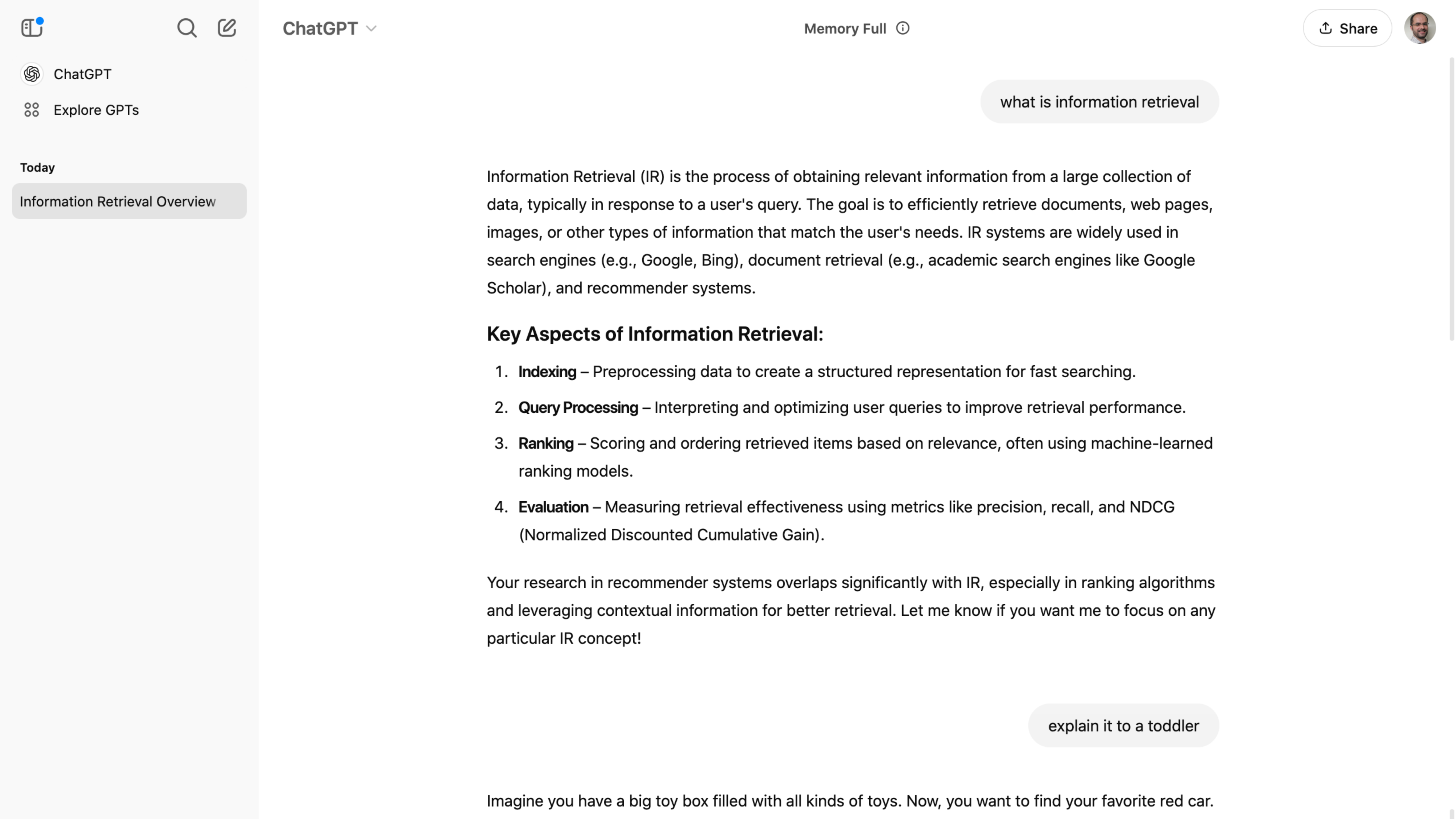
what is information retrieval and its applications?

what is information retrieval and how does it work in search engines?

what is information retrieval and how does it work

what is information retrieval and its role in search engines?

ChatGPT can make mistakes. Check important info.



what is information retrieval

Information Retrieval (IR) is the process of obtaining relevant information from a large collection of data, typically in response to a user's query. The goal is to efficiently retrieve documents, web pages, images, or other types of information that match the user's needs. IR systems are widely used in search engines (e.g., Google, Bing), document retrieval (e.g., academic search engines like Google Scholar), and recommender systems.

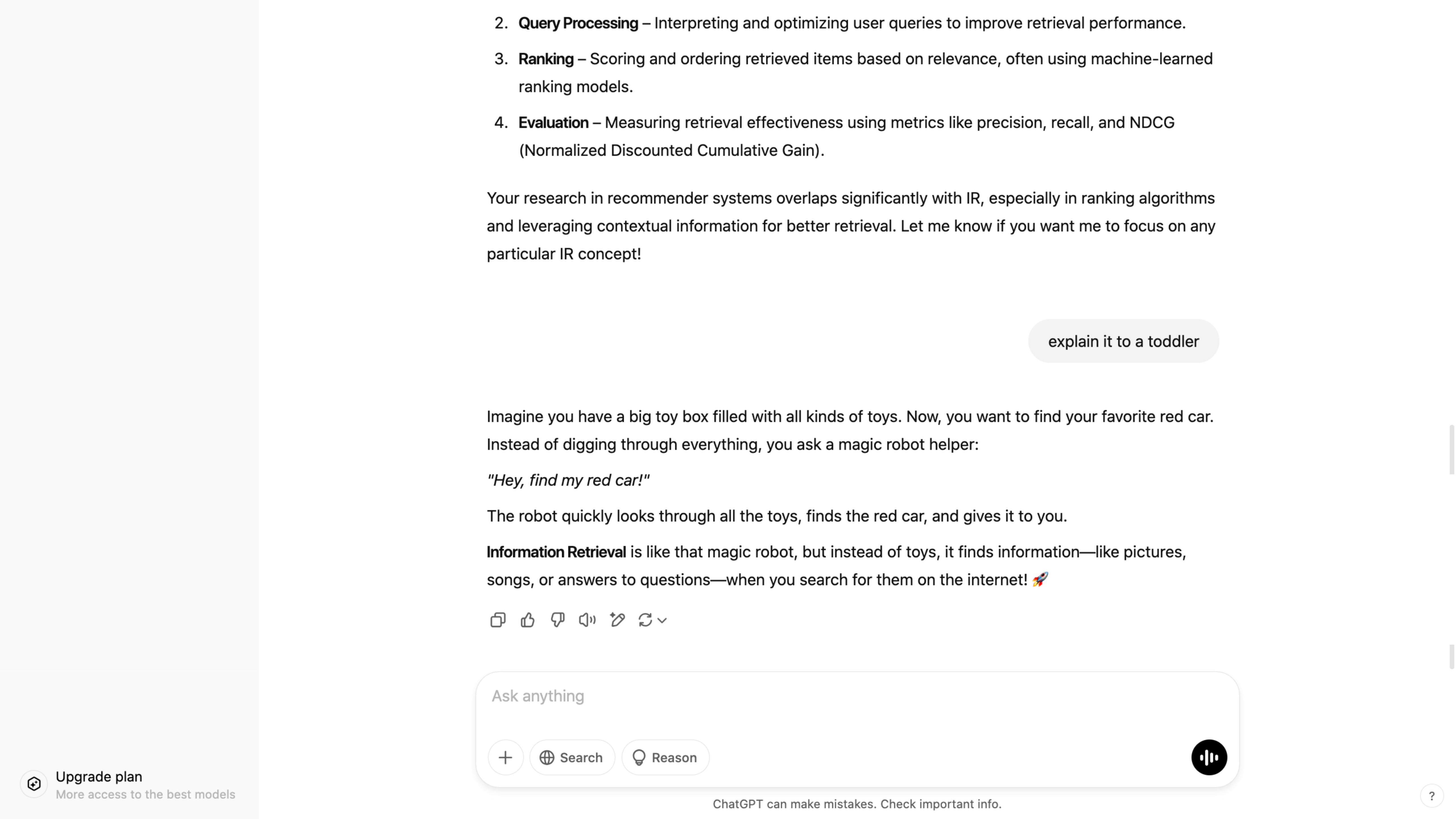
Key Aspects of Information Retrieval:

1. **Indexing** – Preprocessing data to create a structured representation for fast searching.
2. **Query Processing** – Interpreting and optimizing user queries to improve retrieval performance.
3. **Ranking** – Scoring and ordering retrieved items based on relevance, often using machine-learned ranking models.
4. **Evaluation** – Measuring retrieval effectiveness using metrics like precision, recall, and NDCG (Normalized Discounted Cumulative Gain).

Your research in recommender systems overlaps significantly with IR, especially in ranking algorithms and leveraging contextual information for better retrieval. Let me know if you want me to focus on any particular IR concept!

explain it to a toddler

Imagine you have a big toy box filled with all kinds of toys. Now, you want to find your favorite red car.



2. **Query Processing** – Interpreting and optimizing user queries to improve retrieval performance.
3. **Ranking** – Scoring and ordering retrieved items based on relevance, often using machine-learned ranking models.
4. **Evaluation** – Measuring retrieval effectiveness using metrics like precision, recall, and NDCG (Normalized Discounted Cumulative Gain).

Your research in recommender systems overlaps significantly with IR, especially in ranking algorithms and leveraging contextual information for better retrieval. Let me know if you want me to focus on any particular IR concept!

explain it to a toddler

Imagine you have a big toy box filled with all kinds of toys. Now, you want to find your favorite red car. Instead of digging through everything, you ask a magic robot helper:

"Hey, find my red car!"

The robot quickly looks through all the toys, finds the red car, and gives it to you.

Information Retrieval is like that magic robot, but instead of toys, it finds information—like pictures, songs, or answers to questions—when you search for them on the internet! 🚀

Ask anything
Search
Reason
Upgrade plan
More access to the best models
ChatGPT can make mistakes. Check important info.
?

Our focus: search

User’s perspective: accessing information

- Relevance, speed

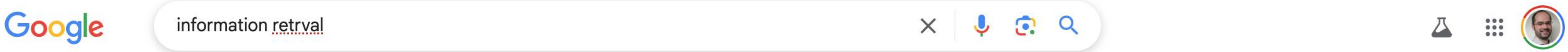
Advertiser’s perspective: publicity

- More attention, less ad costs

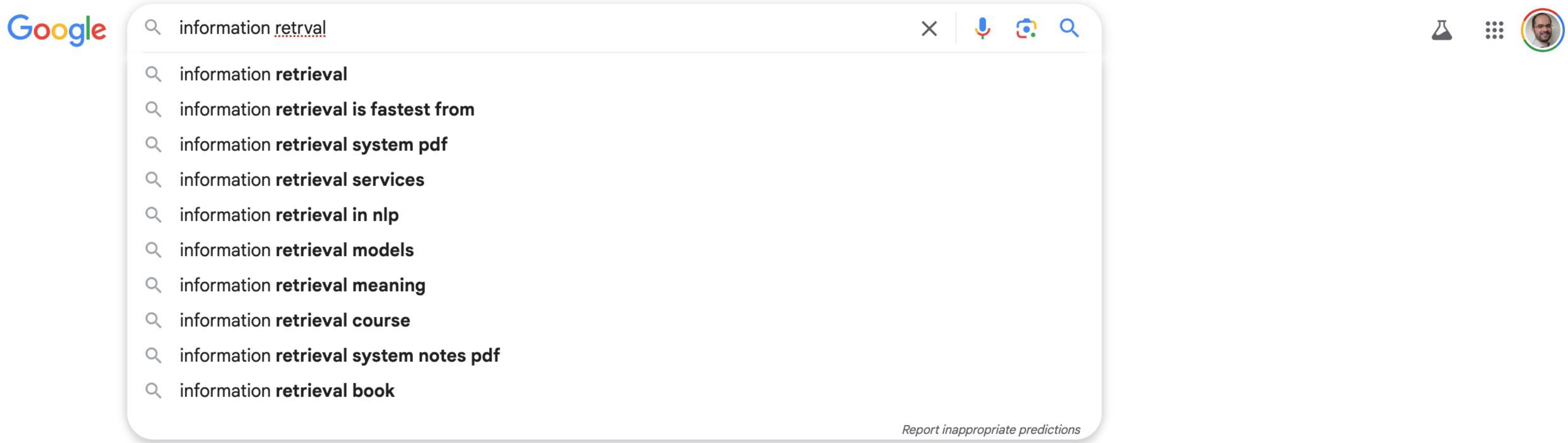
Search engine’s perspective: monetization

- More users, more ad revenue, less op costs

What does a search engine do?


information retrval

Spelling correction



information retrval

information **retrieval**

information **retrieval is fastest from**

information **retrieval system pdf**

information **retrieval services**

information **retrieval in nlp**

information **retrieval models**

information **retrieval meaning**

information **retrieval course**

information **retrieval system notes pdf**

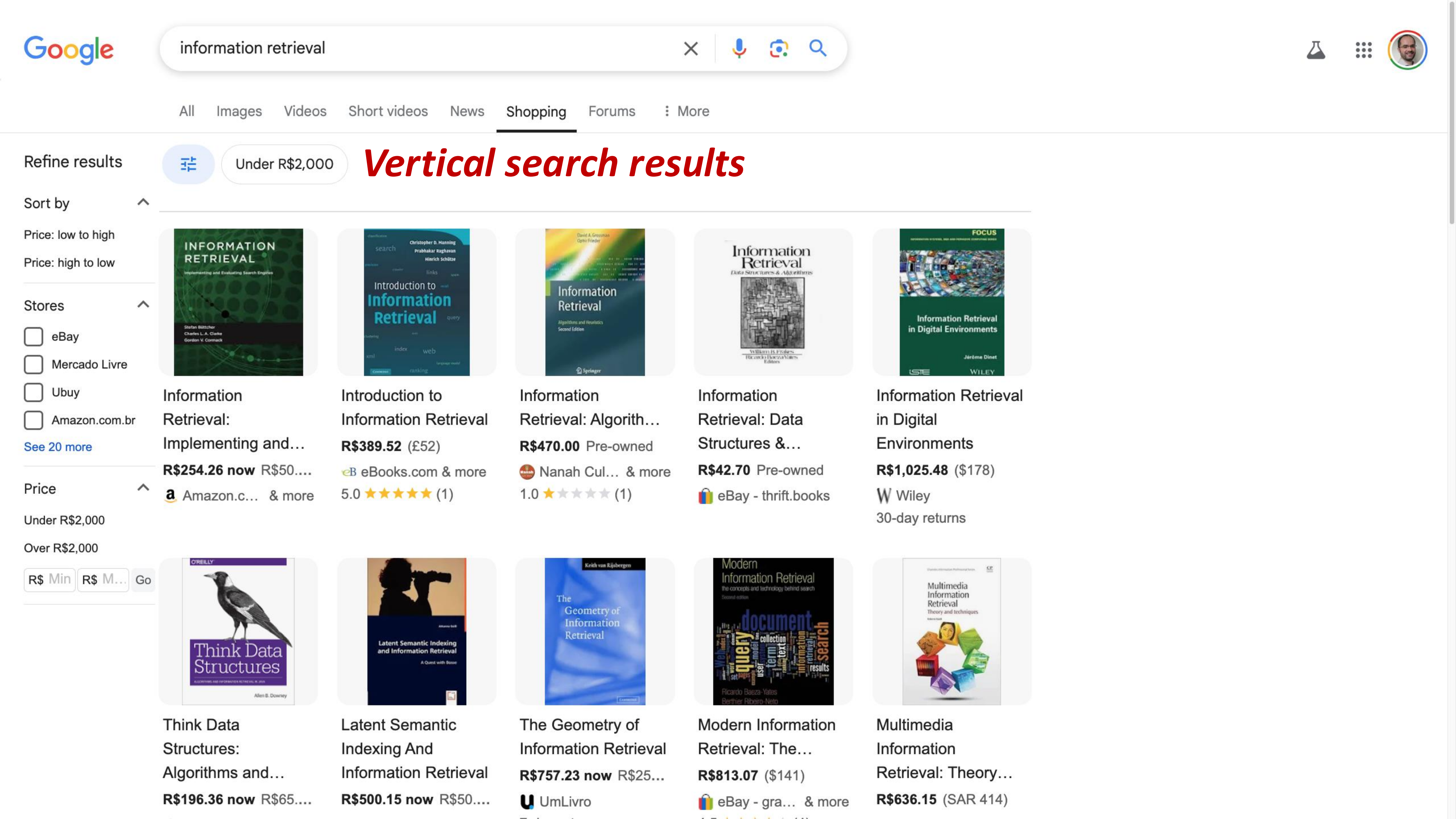
information **retrieval book**

Report inappropriate predictions

Query autocompletion

Query suggestions

Vertical search results



Refine results

Sort by
- Price: low to high
- Price: high to low

Stores
- ☐ eBay
- ☐ Mercado Livre
- ☐ Ubuy
- ☐ Amazon.com.br
- See 20 more

Price
- Under R$2,000
- Over R$2,000
- R$ Min R$ M… Go

Information Retrieval: Implementing and…
R$254.26 now R$50.…
Amazon.c… & more

Introduction to Information Retrieval
R$389.52 (£52)
eBooks.com & more
5.0 ★★★★★ (1)

Information Retrieval: Algorith…
R$470.00 Pre-owned
Nanah Cul… & more
1.0 ★☆☆☆☆ (1)

Information Retrieval: Data Structures &…
R$42.70 Pre-owned
eBay - thrift.books

Information Retrieval in Digital Environments
R$1,025.48 ($178)
Wiley
30-day returns

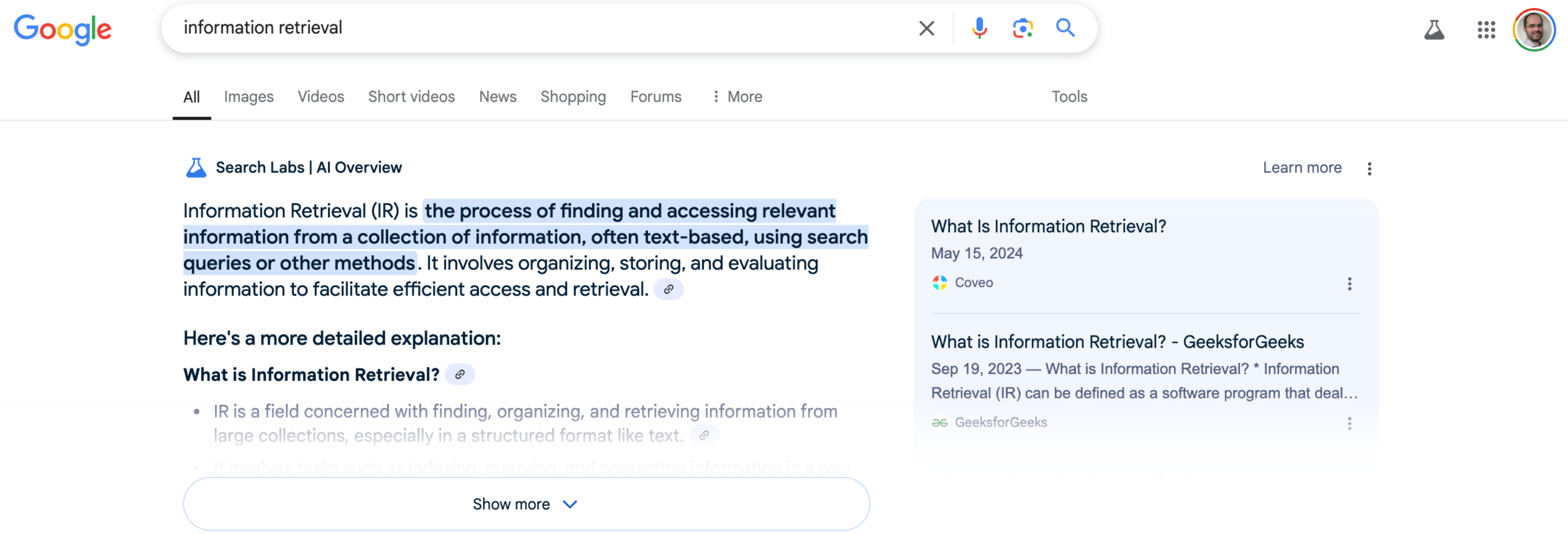
Think Data Structures: Algorithms and…
R$196.36 now R$65.…

Latent Semantic Indexing And Information Retrieval
R$500.15 now R$50.…

The Geometry of Information Retrieval
R$757.23 now R$25…
UmLivro

Modern Information Retrieval: The…
R$813.07 ($141)
eBay - gra… & more

Multimedia Information Retrieval: Theory…
R$636.15 (SAR 414)



Information Retrieval (IR) is **the process of finding and accessing relevant information from a collection of information, often text-based, using search queries or other methods**. It involves organizing, storing, and evaluating information to facilitate efficient access and retrieval.

Here's a more detailed explanation:

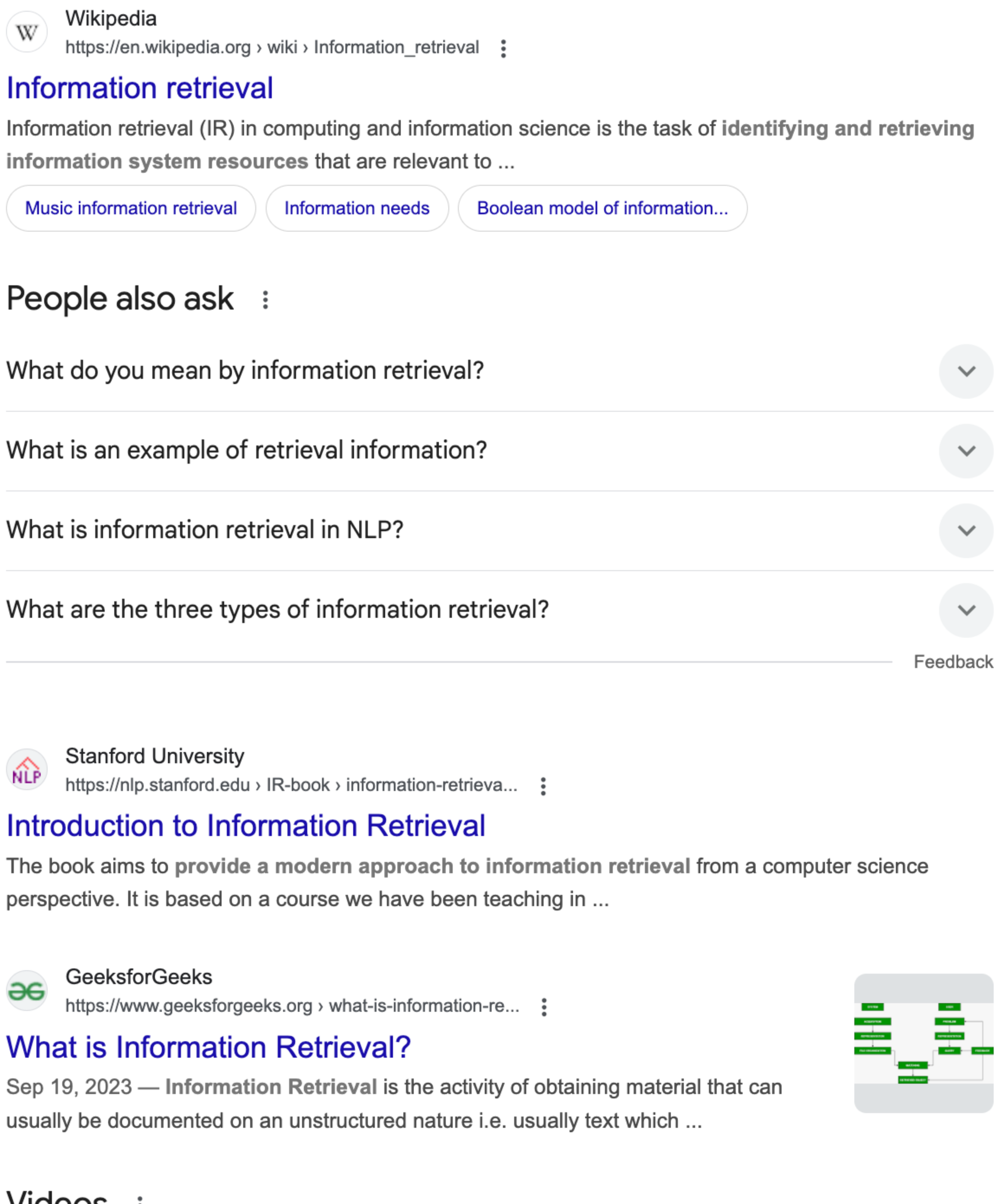
What is Information Retrieval?

- IR is a field concerned with finding, organizing, and retrieving information from large collections, especially in a structured format like text.
- It involves tasks such as indexing, querying, and presenting information in a way

What Is Information Retrieval?
May 15, 2024
Coveo

What is Information Retrieval? - GeeksforGeeks
Sep 19, 2023 — What is Information Retrieval? * Information Retrieval (IR) can be defined as a software program that deal…
GeeksforGeeks

Generated answers



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

Information retrieval

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

People also ask

What do you mean by information retrieval?

What is an example of retrieval information?

What is information retrieval in NLP?

What are the three types of information retrieval?

Stanford University
https://nlp.stanford.edu › IR-book › information-retrieva...

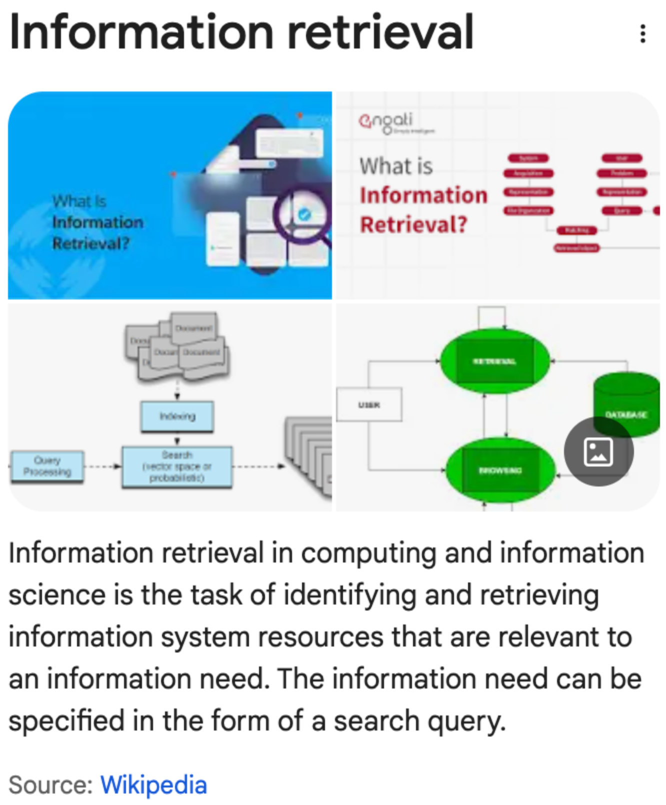
Introduction to Information Retrieval

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 ...

GeeksforGeeks
https://www.geeksforgeeks.org › what-is-information-re...

What is Information Retrieval?

Sep 19, 2023 — **Information Retrieval** is the activity of obtaining material that can usually be documented on an unstructured nature i.e. usually text which ...



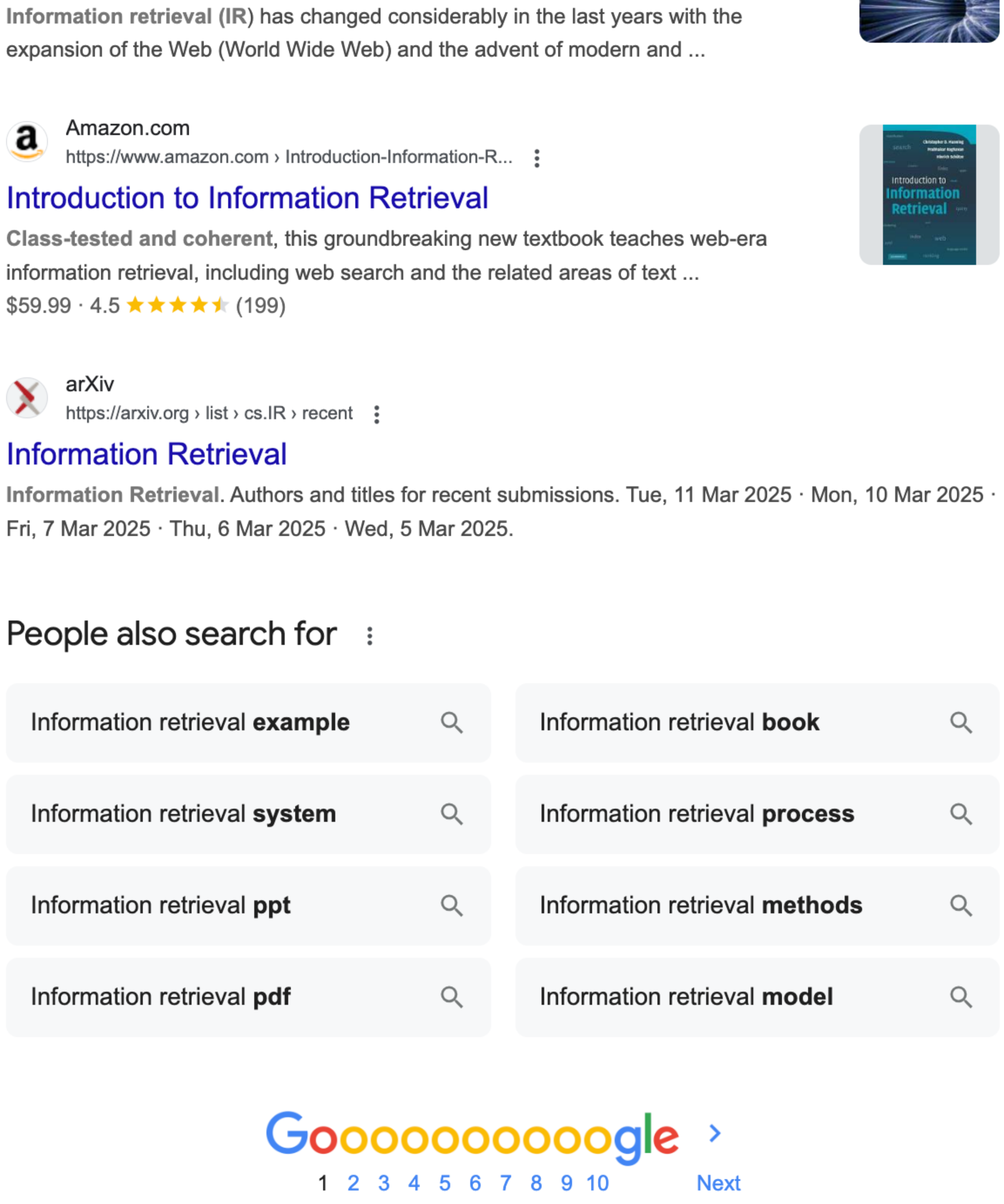
Information retrieval

What Is Information Retrieval?
What is Information Retrieval?
Indexing
Query Processing
Search (vector space or probabilistic)
USER
RETRIEVAL
DATABASE
BROWSING

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.

Source: Wikipedia

Knowledge snippets



Information retrieval (IR) has changed considerably in the last years with the expansion of the Web (World Wide Web) and the advent of modern and ...

Amazon.com
https://www.amazon.com › Introduction-Information-R...

Introduction to Information Retrieval

Class-tested and coherent, this groundbreaking new textbook teaches web-era information retrieval, including web search and the related areas of text ...
$59.99 · 4.5 ★★★★★ (199)

arXiv
https://arxiv.org › list › cs.IR › recent

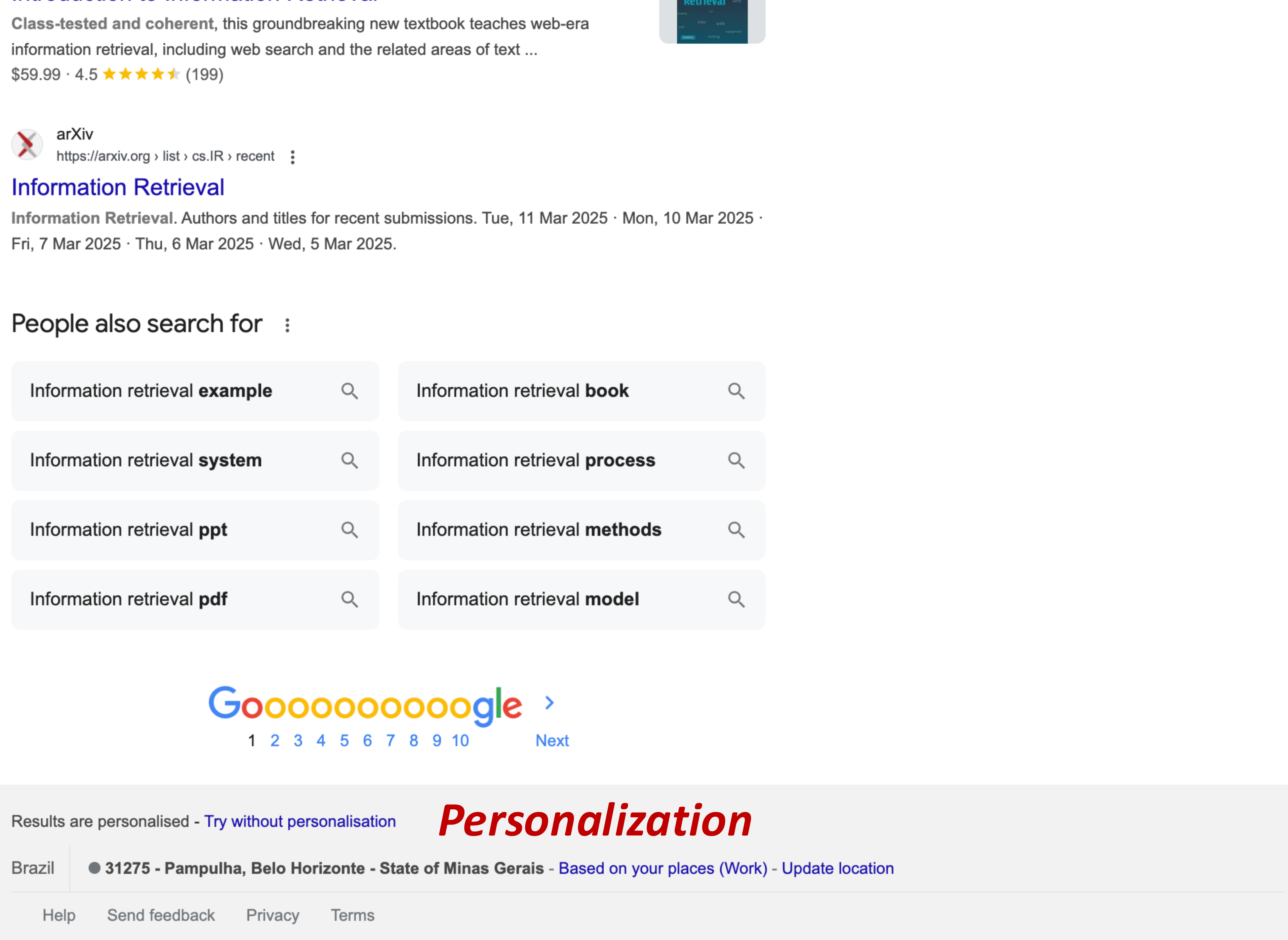
Information Retrieval

Information Retrieval. Authors and titles for recent submissions. Tue, 11 Mar 2025 · Mon, 10 Mar 2025 · Fri, 7 Mar 2025 · Thu, 6 Mar 2025 · Wed, 5 Mar 2025.

People also search for

Information retrieval **example**	Information retrieval **book**
Information retrieval **system**	Information retrieval **process**
Information retrieval **ppt**	Information retrieval **methods**
Information retrieval **pdf**	Information retrieval **model**

Related searches



Class-tested and coherent, this groundbreaking new textbook teaches web-era information retrieval, including web search and the related areas of text ...

$59.99 · 4.5 ★★★★★ (199)

arXiv

https://arxiv.org › list › cs.IR › recent

Information Retrieval

Information Retrieval. Authors and titles for recent submissions. Tue, 11 Mar 2025 · Mon, 10 Mar 2025 · Fri, 7 Mar 2025 · Thu, 6 Mar 2025 · Wed, 5 Mar 2025.

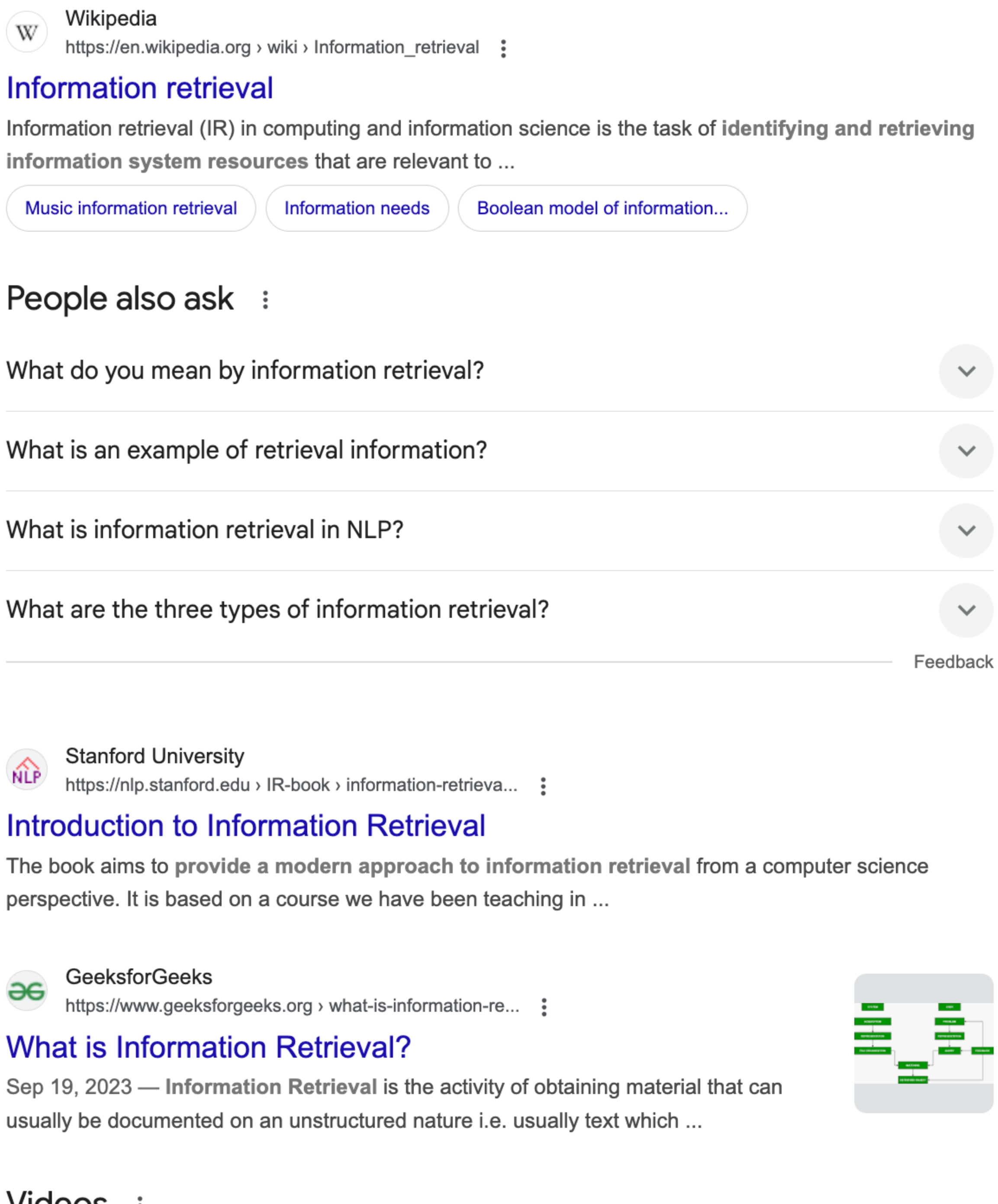
People also search for

Information retrieval **example**	Information retrieval **book**
Information retrieval **system**	Information retrieval **process**
Information retrieval **ppt**	Information retrieval **methods**
Information retrieval **pdf**	Information retrieval **model**

Results are personalised - Try without personalisation

Personalization

Brazil ● **31275 - Pampulha, Belo Horizonte - State of Minas Gerais** - Based on your places (Work) - Update location


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

Information retrieval

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

People also ask

What do you mean by information retrieval?

What is an example of retrieval information?

What is information retrieval in NLP?

What are the three types of information retrieval?

Stanford University
https://nlp.stanford.edu › IR-book › information-retrieva...

Introduction to Information Retrieval

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 ...

GeeksforGeeks
https://www.geeksforgeeks.org › what-is-information-re...

What is Information Retrieval?

Sep 19, 2023 — **Information Retrieval** is the activity of obtaining material that can usually be documented on an unstructured nature i.e. usually text which ...

Videos

ten blue links

The search problem

Given

- Some evidence of the user’s need

Produce

- Relevant information

The search problem

Given

- Some evidence of the user’s need

Produce

- A list of matching information items
- In decreasing order of relevance

The search problem

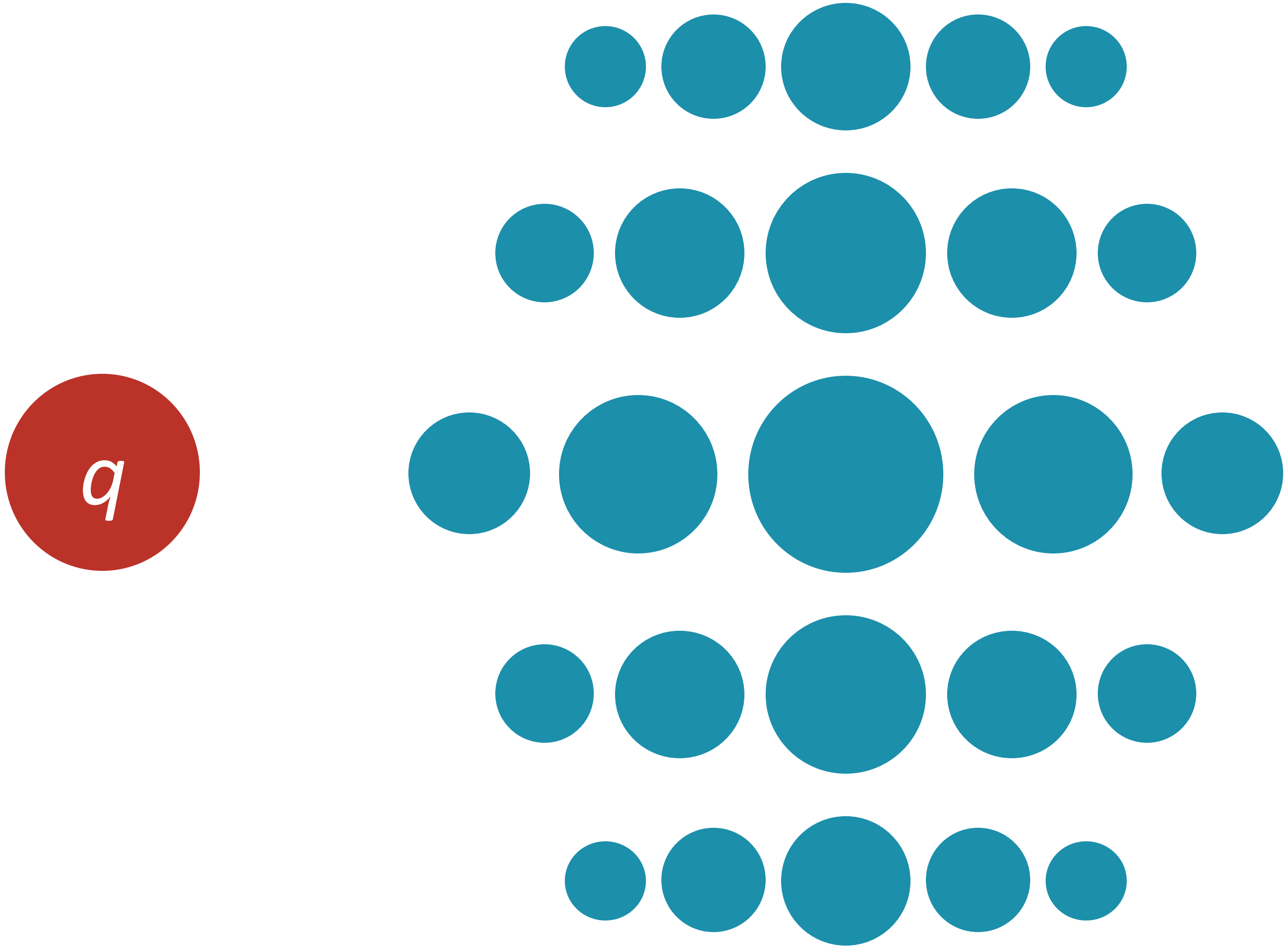
Given

- Some ~~evidence of the user's need~~ ***query***

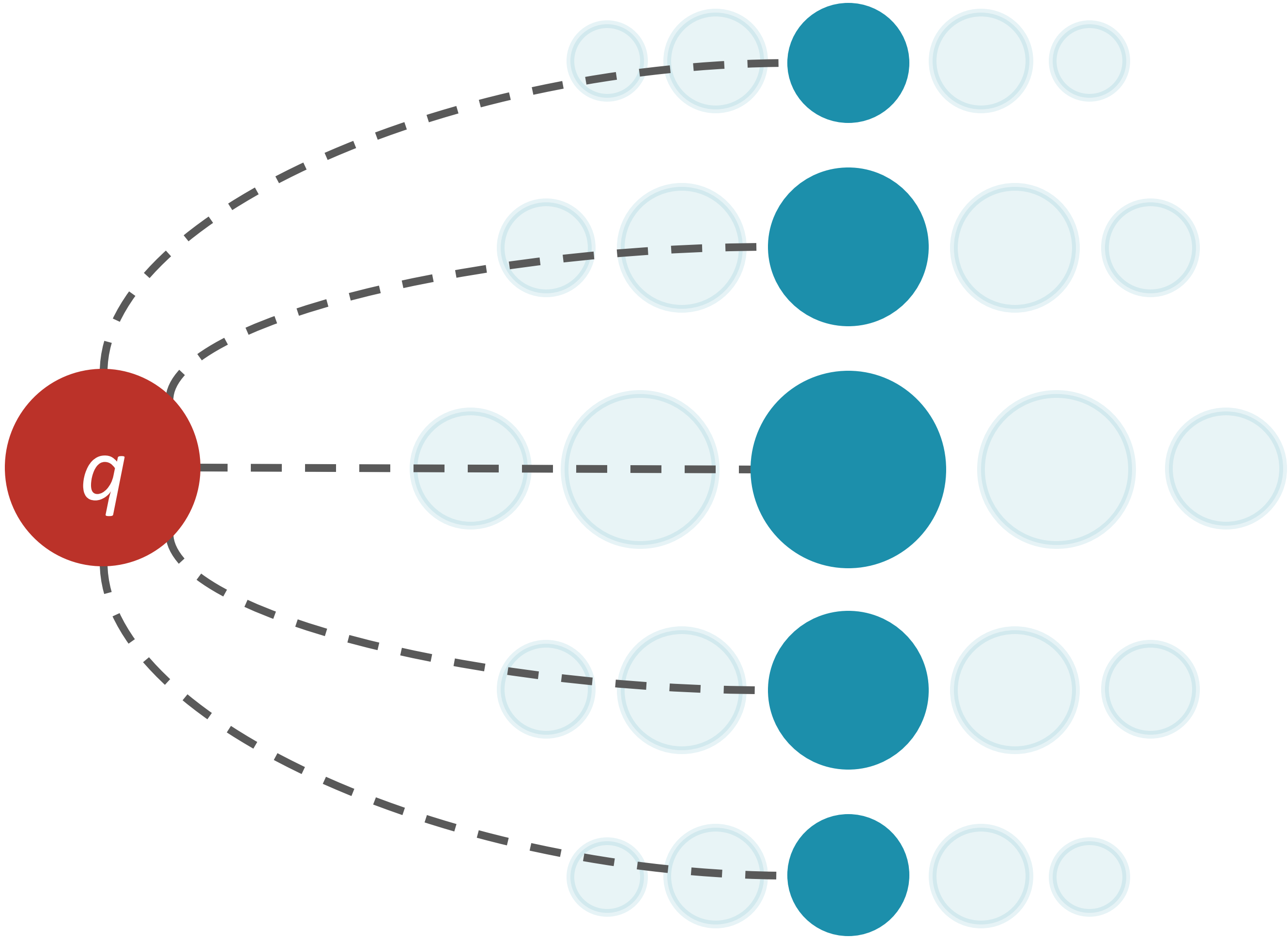
Produce

- A list of matching ~~information items~~ ***documents***
- In decreasing order of relevance

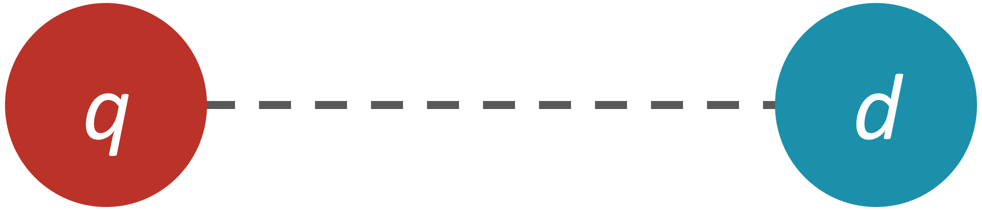
1) What documents do we show?



2) What order do we show them in?

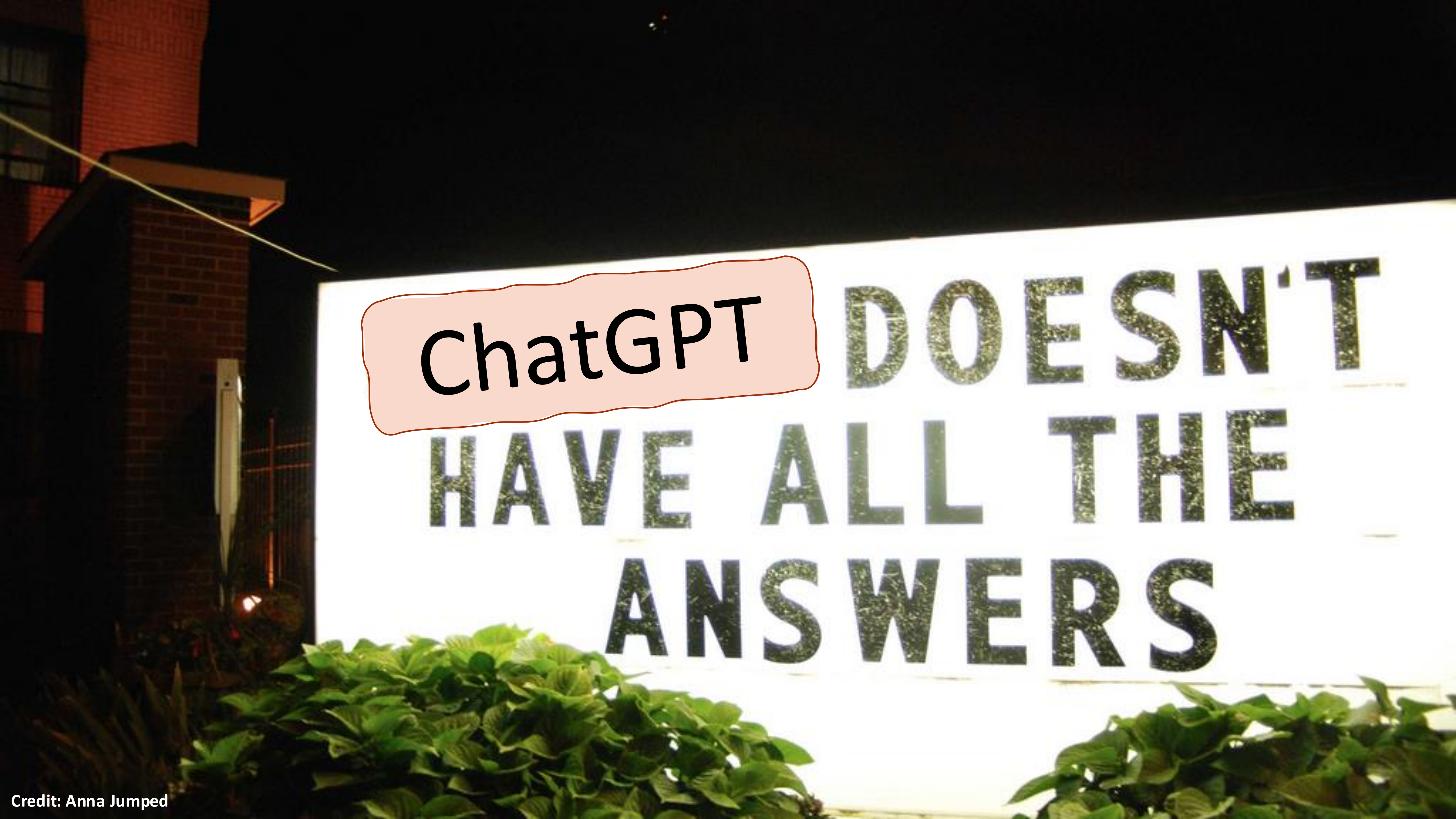


2) What order do we show them in?



$$f(q, d)$$

Isn’t it a solved problem?



Credit: Anna Jumped

Search in numbers

A lot of people **.......................... *10^4 queries per second***

From a lot of places **.......... *whole planet (and beyond?)***

Using a lot of devices **...................... *smart-you-name-it***

Looking for a lot of info **........................ *10^{11} documents***

Spread all over the Internet **......................... *10^7 servers***

Efficiency

Efficiency is about doing something (good or bad) in an optimal way (i.e., faster or with fewer resources)

Key performance indicators

- *Query latency:* searching billions of documents
- *Query throughput:* serving thousands of users
- *Document latency:* serving freshly published content

Effectiveness

Effectiveness is about doing the right thing; it’s about finding documents that are relevant to the user

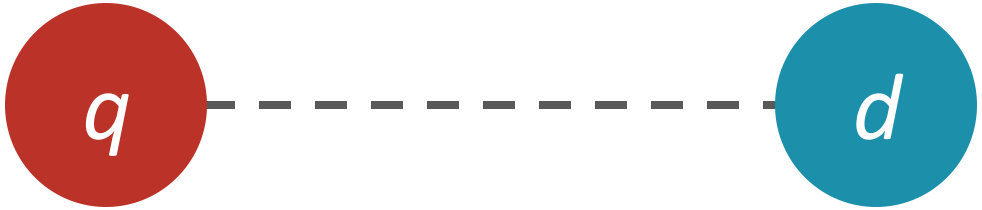
Relevance is influenced by many factors

- Topical relevance vs. user relevance
- Task, context, novelty, style

Ranking models define ***a view of*** relevance

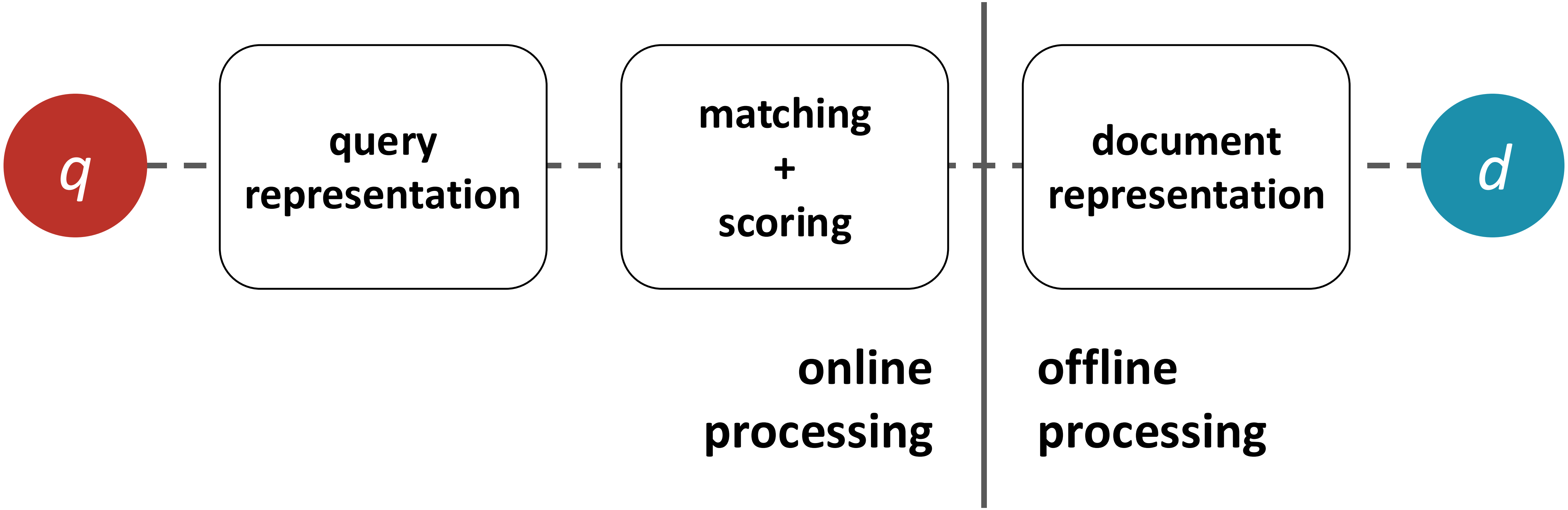
What do search engineers do?

The search problem



$$f(q, d)$$

Search pipeline



q

query representation

matching + scoring

document representation

d

online processing

offline processing

(Continuous) offline processing

Document acquisition

Document understanding

Document indexing

Document acquisition

The Web is huge

- Trillions of known URLs, billions fetched

The Web is constantly evolving

- Updates, additions, deletions

Efficient crawling is key

- Must aim for coverage, but also freshness

Document understanding

Documents carry meaning

- Term-based matching as a first approximation
- Several techniques to leverage semantics

Documents vary in quality

- *Genuinely:* accessibility, readability, authority, depth
- *Maliciously:* content / link farms, misinformation

Document indexing

Efficient retrieval through indexes

- Like the index of a book
    - For each word, a list of documents it appears on
- Broken up into shards of millions of documents
    - 1000s of shards for the web index
- Plus per-document metadata
- Plus document embeddings

Online processing

Query understanding

Matching and scoring

Post-processing

Query understanding

Keywords are poor descriptions of the user’s need

- Interaction and context also matter

Query understanding techniques can help

- Query segmentation, query scoping
- Query relaxation, query expansion
- Query embedding

Query understanding

Query scoping through semantic annotation

- [**san jose** convention center]
- [**matt cutts**]

Query expansion through acronym expansion

- [**gm** trucks] → [**general motors** trucks]
- [**gm** corn] → [**genetically modified** corn]

Matching and scoring

Send the query to all the shards

Each shard

- Finds matching documents
- Scores each query-document pair
- Sends back the top *n* documents

Combine all the top documents and sort by score

Ranking evaluation

Relevance is a user’s prerogative

- We can observe changes in user behavior
- Or directly ask the user how we’re doing

Evaluation is an empirical science

- It must be scientifically rigorous
- It must be economically viable

Course goals

Provide an introductory account of methods for building and evaluating search engines

Provide an exploration of recent advances and current research directions in the field

Course scope

System view

- Crawling, indexing, retrieval

Modeling view

- Ranking models

Behavioral view

- Ranking evaluation

Out-of-scope

We have dedicated courses for:

- Recommender systems
- Natural language processing
- Machine learning
- Data mining

Course grading (tentative)

Exams: 50%

Assignments: 40%

Seminars: 10%

Course attendance

> ***O que é necessário para ser aprovado em uma dada atividade acadêmica curricular?***
>
> *É necessário obter nota final igual ou superior a 60, em uma escala de 0 a 100, bem como a indicação de assiduidade, a qual deve ser igual ou superior a 75% (art. 12 das NGG).*

Course materials: textbooks

Search Engines: Information Retrieval in Practice

by B. Croft, D. Metzler, and T. Strohman

Introduction to Information Retrieval

by C. Manning, P. Raghavan, and H. Schütze

Modern Information Retrieval

by R. Baeza-Yates and B. Ribeiro-Neto

Course materials: textbooks

Information Retrieval: Implementing and Evaluating Search Engines
by S. Büttcher, C. Clarke, and G. Cormack

Text Data Management: A Practical Introduction to Information Retrieval and Text Mining
by C. Zhai and S. Massung

Course materials: surveys

[Foundations and Trends in Information Retrieval]

by several authors

[Synthesis Lectures on Information Concepts, Retrieval, and Services]

by several authors

Other relevant material

General background

- Algorithms and data structures
- Basic statistics
- Basic linear algebra

Advanced readings

- Google Scholar is your friend

References

Search Engines: Information Retrieval in Practice, Ch. 1
Croft et al., 2009

How Google Works: A Google Ranking Engineer’s Story
Haahr, SMX West 2016

Ten blue links on Mars
Clarke et al., WWW 2017

Pre-course survey

Fill in a short survey describing your past experience and expectations related to the course

- https://forms.gle/7mcatGc5LtAFM2ta7

UFMG

UNIVERSIDADE*FEDERAL
DE*MINAS*GERAIS

Coming next…

Search Architecture

Rodrygo L. T. Santos
rodrygo@dcc.ufmg.br