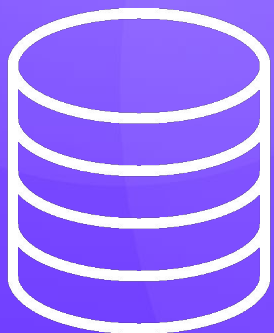


# Privacy vs Relevancy

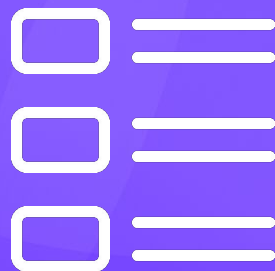
Web Crawler



Simple Storage



Computation



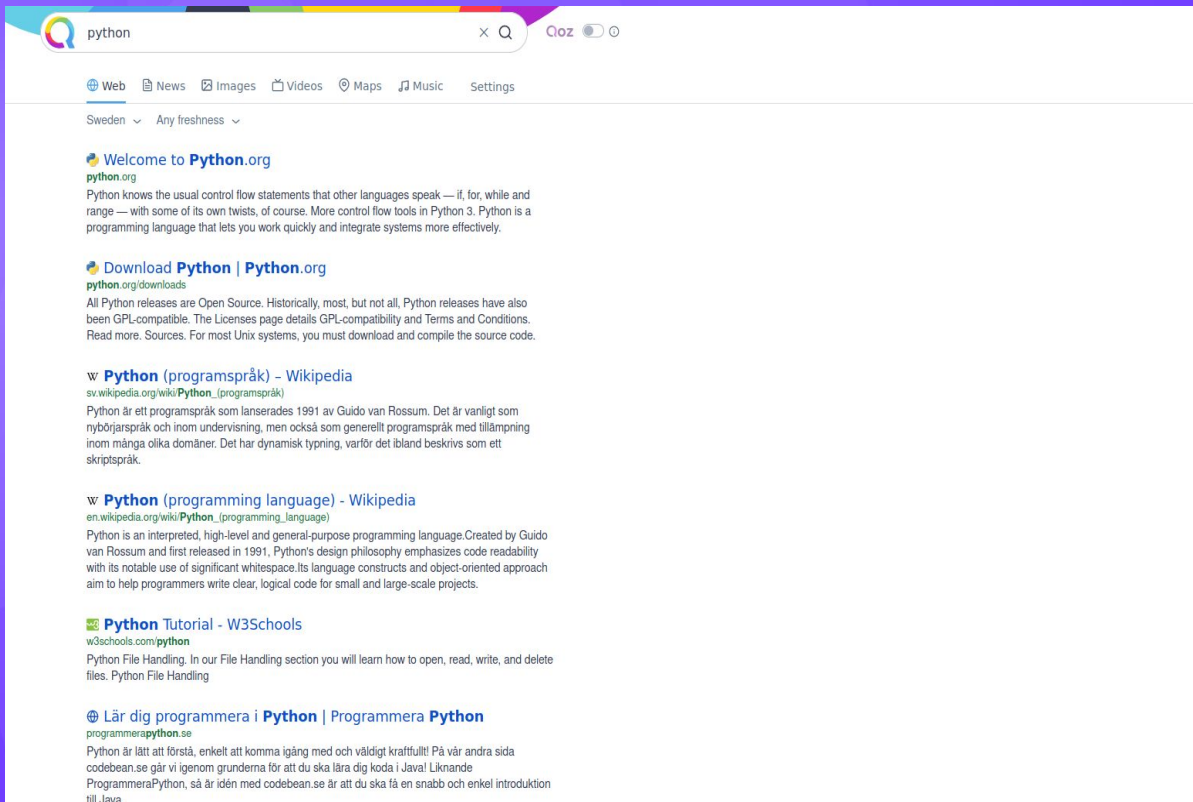
Indexed Data



# Query Intent: Python



# Query Intent: Python



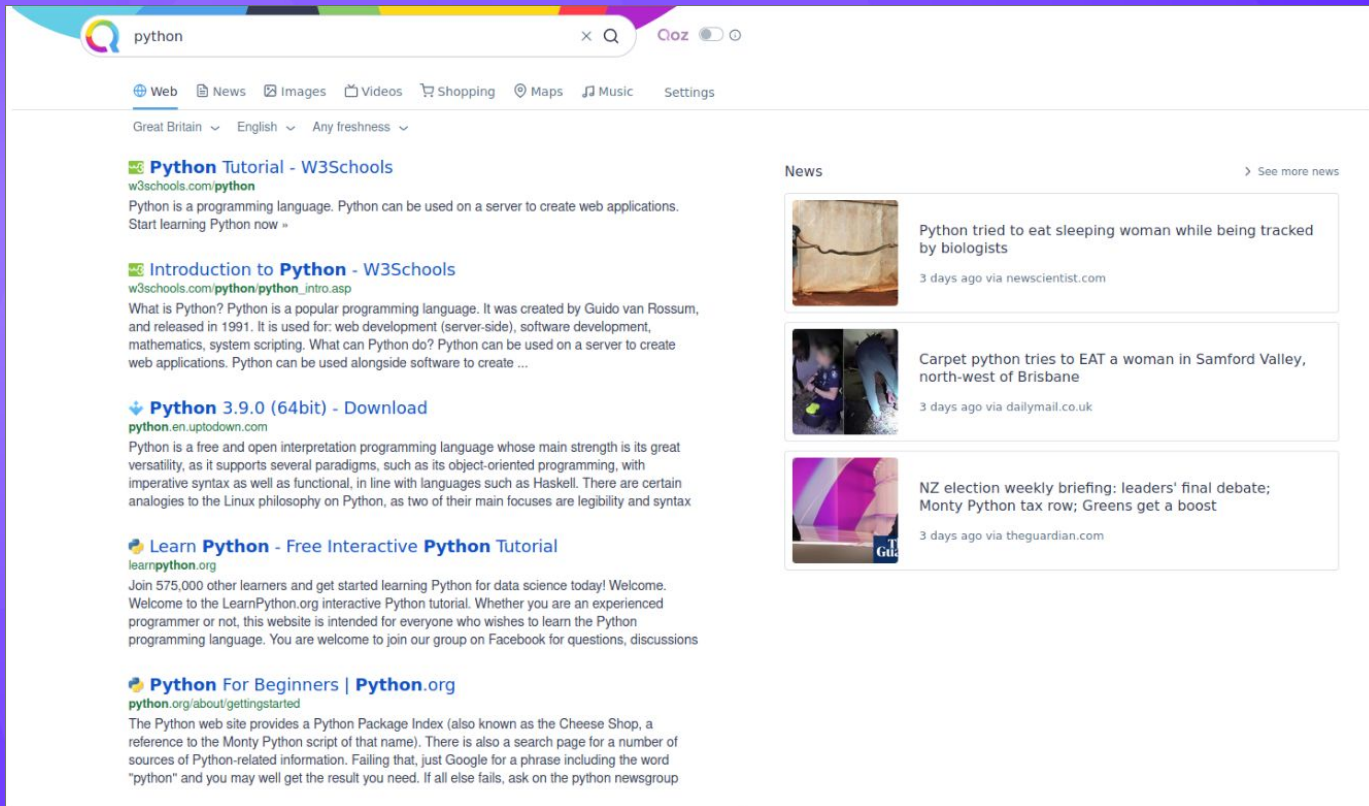
The screenshot shows a search engine interface with the query 'python' entered in the search bar. The search results are displayed below the search bar, showing several relevant links and snippets of text.

Search results for 'python':

- Welcome to Python.org**  
python.org  
Python knows the usual control flow statements that other languages speak — if, for, while and range — with some of its own twists, of course. More control flow tools in Python 3. Python is a programming language that lets you work quickly and integrate systems more effectively.
- Download Python | Python.org**  
python.org/downloads  
All Python releases are Open Source. Historically, most, but not all, Python releases have also been GPL-compatible. The Licenses page details GPL-compatibility and Terms and Conditions. Read more. Sources. For most Unix systems, you must download and compile the source code.
- Python (programspråk) - Wikipedia**  
sv.wikipedia.org/wiki/Python\_(programspråk)  
Python är ett programspråk som lanserades 1991 av Guido van Rossum. Det är vanligt som rybbjarspråk och inom undervisning, men också som generellt programspråk med tillämpning inom många olika domäner. Det har dynamisk typning, varför det ibland beskrivs som ett skriptspråk.
- Python (programming language) - Wikipedia**  
en.wikipedia.org/wiki/Python\_(programming\_language)  
Python is an interpreted, high-level and general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.
- Python Tutorial - W3Schools**  
w3schools.com/python  
Python File Handling. In our File Handling section you will learn how to open, read, write, and delete files. Python File Handling
- Lär dig programmera i Python | Programmera Python**  
programmerapython.se  
Python är lätt att förstå, enkelt att komma igång med och väldigt kraftfullt! På vår andra sida codebean.se går vi igenom grunderna för att du ska lära dig koda i Java! Liknande ProgrammeraPython, så är idén med codebean.se är att du ska få en snabb och enkel introduktion till Java



# Query Intent: Python



The screenshot shows a Google search for "python". The search bar at the top contains "python" and the Google logo. Below the search bar, there are navigation links for Web, News, Images, Videos, Shopping, Maps, Music, and Settings. The location is set to Great Britain and the language to English. The search results are displayed in a list format, with each result including a logo, title, URL, and a brief description. On the right side, there is a "News" section with three news items, each featuring a small image, a headline, and the source.

**python** × Q Coz

Web News Images Videos Shopping Maps Music Settings

Great Britain English Any freshness

**Python Tutorial - W3Schools**  
w3schools.com/python  
Python is a programming language. Python can be used on a server to create web applications. Start learning Python now -


**Introduction to Python - W3Schools**  
w3schools.com/python/python\_intro.asp  
What is Python? Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. It is used for: web development (server-side), software development, mathematics, system scripting. What can Python do? Python can be used on a server to create web applications. Python can be used alongside software to create ...


**Python 3.9.0 (64bit) - Download**  
python.en.uptodown.com  
Python is a free and open interpretation programming language whose main strength is its great versatility, as it supports several paradigms, such as its object-oriented programming, with imperative syntax as well as functional, in line with languages such as Haskell. There are certain analogies to the Linux philosophy on Python, as two of their main focuses are legibility and syntax


**Learn Python - Free Interactive Python Tutorial**  
learnpython.org  
Join 575,000 other learners and get started learning Python for data science today! Welcome. Welcome to the LearnPython.org interactive Python tutorial. Whether you are an experienced programmer or not, this website is intended for everyone who wishes to learn the Python programming language. You are welcome to join our group on Facebook for questions, discussions

**Python For Beginners | Python.org**  
python.org/about/gettingstarted  
The Python web site provides a Python Package Index (also known as the Cheese Shop, a reference to the Monty Python script of that name). There is also a search page for a number of sources of Python-related information. Failing that, just Google for a phrase including the word "python" and you may well get the result you need. If all else fails, ask on the python newsgroup

**News** > See more news

 Python tried to eat sleeping woman while being tracked by biologists  
3 days ago via newscientist.com

 Carpet python tries to EAT a woman in Samford Valley, north-west of Brisbane  
3 days ago via dailymail.co.uk

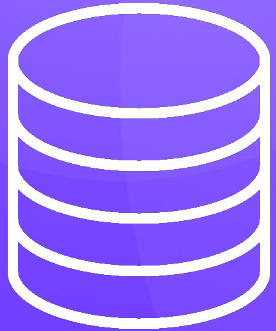
 NZ election weekly briefing: leaders' final debate; Monty Python tax row; Greens get a boost  
3 days ago via theguardian.com



# Query Intent: Python



Web Crawler

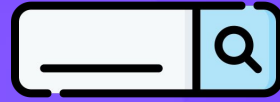


Simple Storage

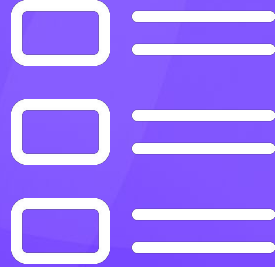
1. Language Classification



Computation



Language Identification



Indexed Data



# Query Intent: Apple



# Query Intent: Apple

The image shows a Google search for "apple". The search bar at the top contains "apple" and shows "About 4 860 000 000 results (0,88 seconds)". Below the search bar are navigation tabs for "All", "Images", "Maps", "News", "Videos", and "More".

The main search results section includes:

- www.apple.com** - **Apple**  
Discover the innovative world of **Apple** and shop everything iPhone, iPad, **Apple Watch**, Mac, and **Apple TV**, plus explore accessories, entertainment, and expert ...
- Apple Store**  
Shop apple.com with the services you love: personal shopping ...
- iPhone**  
Compare - iPhone 11 - iPhone SE - iPhone XR - AirPods - iOS 14
- Mac**  
MacBook Air - iMac - Accessories - Compare Mac models - Mac Pro
- iPad**  
Featuring an all-new iPad Air, an even faster iPad, iPad Pro in ...
- Choose your country or region**  
Choose Your Country or Region. Africa, Middle East, and India ...
- Support**  
Apple support is here to help. Learn more about popular ...

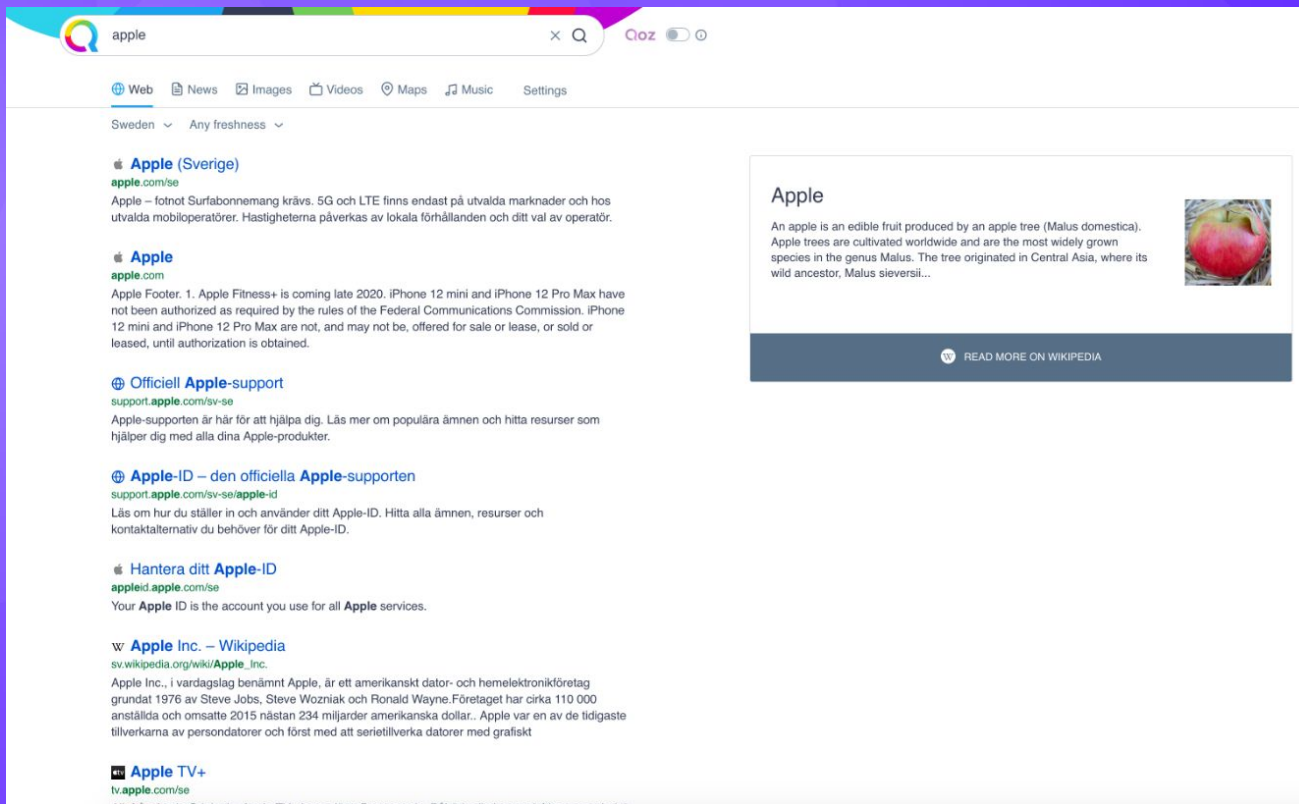
A map shows the location of an Apple store in Täby, Sweden. Below the map, the listing for "Apple Täby Centrum" is visible, including the address "Täby · In the Täby Centrum" and the status "Closes soon · 8PM · 08-446 93 00".

The right-hand side of the page features a detailed knowledge panel for Apple:

- 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 Tech technology companies, alongside Amazon, Google, Microsoft, and Facebook. [Wikipedia](#)
- CEO:** **Tim Cook** (Aug 24, 2011–) *Trending*
- Stock price:** **AAPL** (NASDAQ) **118,70** **US\$** **-0,32** **(-0,27 %)**  
19 Oct, 13:03 GMT-4 - Disclaimer
- Founded:** **April 1, 1976**, Cupertino, California, United States
- Technical support:** 020-10 05 29
- Headquarters:** Cupertino, California, United States
- Founders:** Steve Jobs, Steve Wozniak, Ronald Wayne
- Subsidiaries:** Beats Electronics, Apple Store, Mobeewave Inc., [MORE](#)
- Replacement battery
- Mission statement
- Operating system
- Slogan



# Query Intent: Apple



The screenshot shows a search engine results page for the query "apple". The search bar at the top contains the text "apple" and a magnifying glass icon. Below the search bar, there are navigation tabs for "Web", "News", "Images", "Videos", "Maps", "Music", and "Settings". The location is set to "Sweden" and the freshness is set to "Any freshness".

The search results are as follows:

- Apple (Sverige)**  
apple.com/se  
Apple – fotnot Surlabonnemang krävs. 5G och LTE finns endast på utvalda marknader och hos utvalda mobiloperatörer. Hastigheterna påverkas av lokala förhållanden och ditt val av operatör.
- Apple**  
apple.com  
Apple Footer. 1. Apple Fitness+ is coming late 2020. iPhone 12 mini and iPhone 12 Pro Max have not been authorized as required by the rules of the Federal Communications Commission. iPhone 12 mini and iPhone 12 Pro Max are not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained.
- Officiell Apple-support**  
support.apple.com/sv-se  
Apple-supporten är här för att hjälpa dig. Läs mer om populära ämnen och hitta resurser som hjälper dig med alla dina Apple-produkter.
- Apple-ID – den officiella Apple-supporten**  
support.apple.com/sv-se/apple-id  
Läs om hur du ställer in och använder ditt Apple-ID. Hitta alla ämnen, resurser och kontaktalternativ du behöver för ditt Apple-ID.
- Hantera ditt Apple-ID**  
appleid.apple.com/se  
Your Apple ID is the account you use for all Apple services.
- Apple Inc. – Wikipedia**  
sv.wikipedia.org/wiki/Apple\_Inc.  
Apple Inc., i vardagslag benämnt Apple, är ett amerikanskt dator- och hemelektronikföretag grundat 1976 av Steve Jobs, Steve Wozniak och Ronald Wayne. Företaget har cirka 110 000 anställda och omsatte 2015 nästan 234 miljarder amerikanska dollar.. Apple var en av de tidigaste tillverkarna av persondatorer och först med att serietilverka datorer med grafiskt
- Apple TV+**  
tv.apple.com/se  
Nåväl, Apple Östasiatiska Apple TV+ baserat på Disney+ till. Bättre än de bästa försvaret 17

On the right side of the page, there is a featured snippet for "Apple". It includes a title "Apple", a small image of a red apple, and a text description: "An apple is an edible fruit produced by an apple tree (Malus domestica). Apple trees are cultivated worldwide and are the most widely grown species in the genus Malus. The tree originated in Central Asia, where its wild ancestor, Malus sieversii...". Below the text is a button that says "READ MORE ON WIKIPEDIA".



# Query Intent: Apple



# Query Intent

Summer dress in Rome

Summer dress

in

Rome



# Query Intent

Summer dress in Rome

**E-shops or Wiki?**

**In or For Rome?**

**Or -> <https://summerdressesinrome.com>**



# Query Intent

Heuristics

NER

Knowledge Graph (KG)



# Query Intent - Heuristics

- Navigational:
  - Query length  $< 3$  ('netflix signup')
  - Domains suffixes ('translate', 'twitter')
- Transactional:
  - Movies, songs, lyrics, etc.. ('The Social Dilemma', 'Thriller')
  - Terms with: 'obtaining', 'download', etc..
- Informational:
  - 'How to', 'Ways of', 'Why is'
  - Query length  $> 2$

Jansen, Bernard J., et al. "Determining the Informational, Navigational, and Transactional Intent of Web Queries." Information Processing and Management, vol. 44, no. 3, 2008, pp. 1251-1266.



# Query Intent - NER

“Named entity recognition (NER) — sometimes referred to as entity chunking, extraction, or identification — is the task of identifying and categorizing key information (entities) in text.”

**New York Times**      newspaper

**New York**      city

<https://medium.com/mysuperaai/what-is-named-entity-recognition-ner-and-how-can-i-use-it-2b68cf6f545d>

<https://nlp.stanford.edu/software/CRF-NER.shtml>

*Matteo Ghetti, nonoRank*



# Query Intent - Knowledge Graph

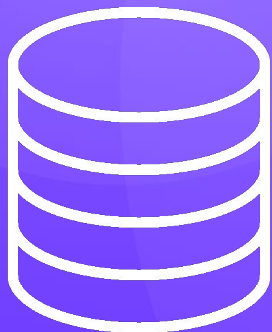
“Things not Strings” - Google

Query: “Capital of Sweden”

The screenshot shows a Google search page for the query "capital of sweden". The search results include a map of Stockholm, Sweden, and a list of associated searches such as "Suède", "Oslo", "Comté de Stockholm", "Copenha...", "Helsinki", "Göteborg", and "Malmö". Below the map, there is a section for "Recherches associées" with a link to "Voir d'autres éléments (plus de 10)". At the bottom, there is a snippet from Wikipedia titled "Sweden - Wikipedia" which states: "The capital city is Stockholm. Sweden has a total population of 10.3 million of which 2.6 million have a foreign background. Persons with foreign ...".



Web Crawler

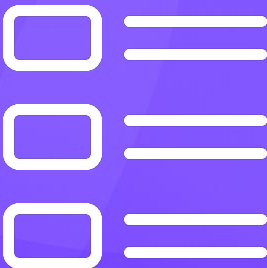


Simple Storage

1. Language Classification
2. Knowledge Graph Structure



Computation



Indexed Data



Language Identification



Heuristics + NER



Knowledge Graph



# Relevancy Algorithms

**Katz's index** -> measures degree of a network

**PageRank** -> Google's ranking system 1998

**HITS** -> mutual reinforcement

**SALSA** -> similar to HITS but  $\lambda$ -matrix-normalization

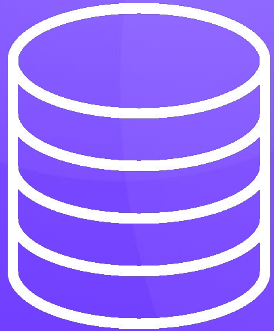
**BM25** -> query terms appearing in each document (bag of words)

... and many more: **Left dominant eigenvector**, **Seeley's index**, **Harmonic Centrality**, **Closeness**, **Lin-centrality**.

Boldi, Paolo, and Sebastiano Vigna. "Axioms for Centrality." Internet Mathematics, vol. 10, no. 3, 2014, pp. 222–262.



Web Crawler

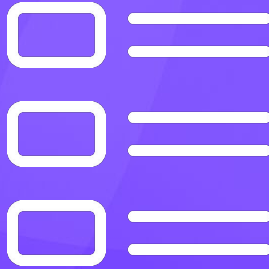


Simple Storage

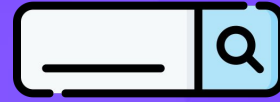
1. Language Classification
2. Knowledge Graph Structure



Computation



Indexed Data



Language Identification



Heuristics + NER



**BM25**

Knowledge Graph



# “Capital of Sweden”

Is this what we are looking for?



The screenshot shows the GIGABLAST search engine interface. The search bar contains the text "Capital of sweden" and a red "GO" button. Below the search bar, there are options for "Try your search on:" with links for "startpage (uses google)" and "duckduckgo (uses bing)". The search results are displayed in a list format. The first result is "Sweden: Rape Capital of the West" from the website "www.gatestoneinstitute.org/5195/sweden/rape". This result is circled in red. Below it, there is another result from the same website, "Sweden: Rape Capital of the West", dated Feb 14 2015. The second result is also circled in red. The third result is "Sweden :: Writings by Gatestone Institute" from "https://www.gatestoneinstitute.org/topics/36/sweden". The fourth result is "This is Malmö: Football capital of Sweden - The Local" from "https://www.thelocal.se/20161017/this-is-malm-footba...". This result is also circled in red. The search results are sorted by relevance and are in English, with 25 results found in 3.19 seconds.

**GIGABLAST**

Capital of sweden **GO**

Try your search on: startpage (uses google) duckduckgo (uses bing)

United States (English) ▼ Any time ▼ Sorted by relevance ▼

Results 1 to 25 for *Capital of sweden* in 3.19 seconds

**Sweden: Rape Capital of the West**  
<https://gatestoneinstitute.org/5195/sweden/rape>  
Sweden: Rape Capital of the West. Denigration of ethnic groups.

**Sweden: Rape Capital of the West**  
[www.gatestoneinstitute.org/5195/...](http://www.gatestoneinstitute.org/5195/...)  
Feb 14 2015 - Sweden: Rape Capital of the West. It only takes a little effort to publicize Sweden's views on rape.

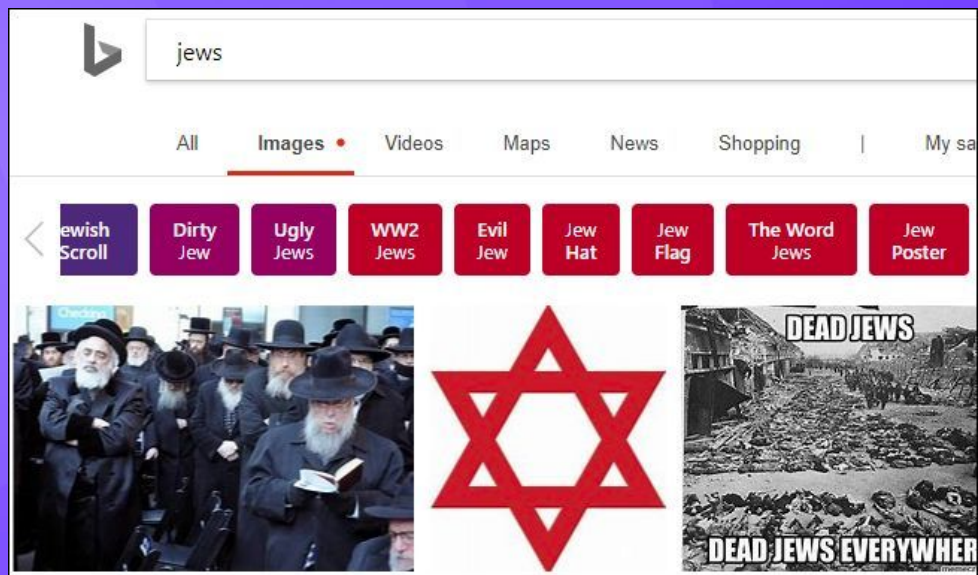
**Sweden :: Writings by Gatestone Institute**  
<https://www.gatestoneinstitute.org/topics/36/sweden>  
Sweden: Rape Capital of the West. Anti-Semitism in Sweden.

more from this site

**This is Malmö: Football capital of Sweden - The Local**  
<https://www.thelocal.se/20161017/this-is-malm-footba...>  
Oct 17 2016 - This is Malmö: Football capital of Sweden. Sweden's news in English.



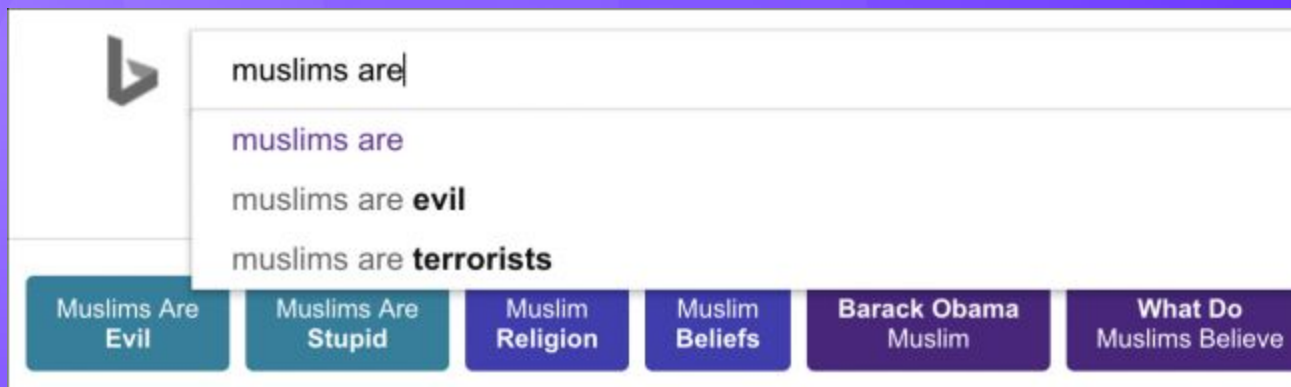
# We are Bing 10th October 2018



<https://www.howtogeek.com/367878/bing-is-suggesting-the-worst-things-you-can-imagine/>



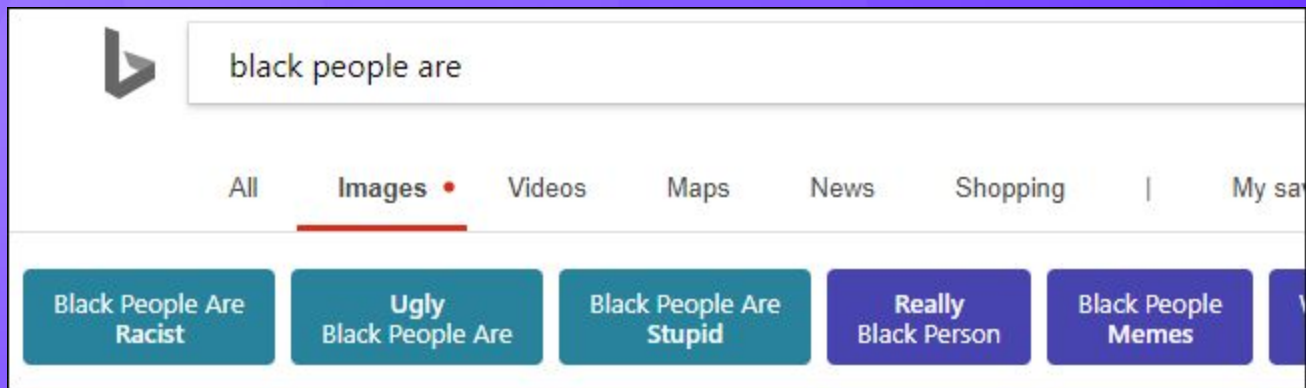
# We are Bing 10th October 2018



<https://www.howtogeek.com/367878/bing-is-suggesting-the-worst-things-you-can-imagine/>



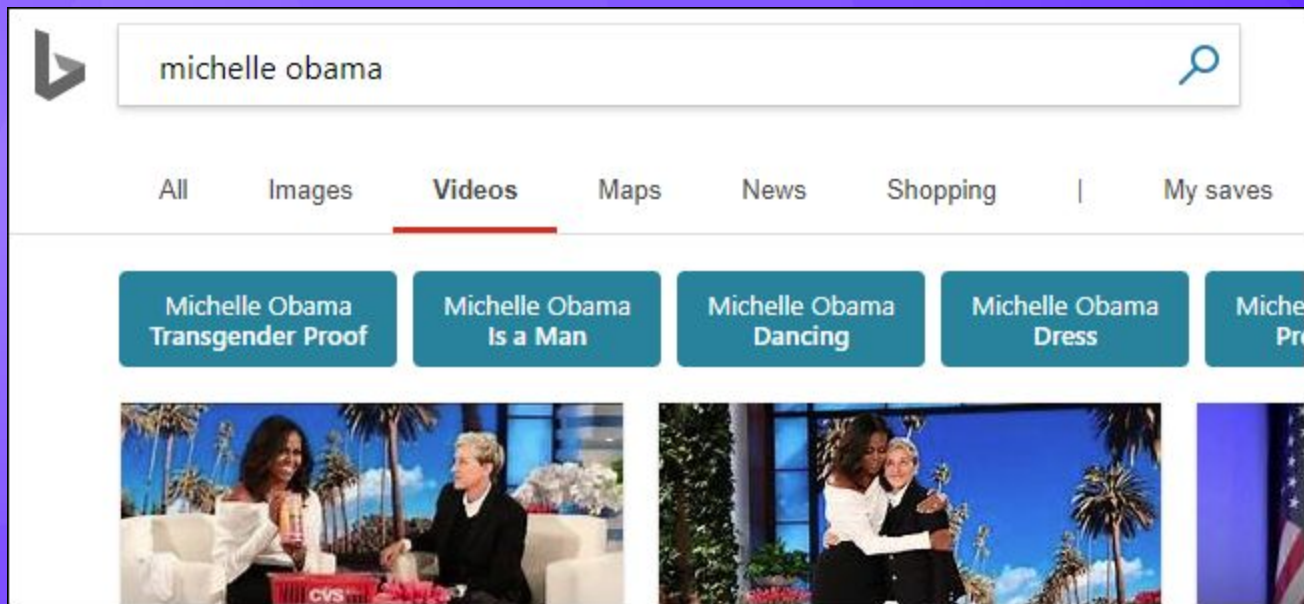
# We are Bing 10th October 2018



<https://www.howtogeek.com/367878/bing-is-suggesting-the-worst-things-you-can-imagine/>



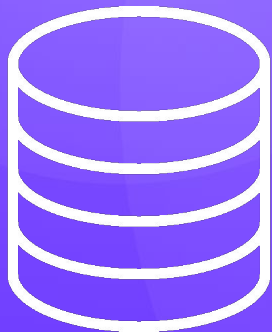
# We are Bing 10th October 2018



<https://www.howtogeek.com/367878/bing-is-suggesting-the-worst-things-you-can-imagine/>



Web Crawler



Simple Storage

- 1. Language Classification
- 2. Knowledge Graph Structure



- 3. Categorization and Filtering:  
ML + AI

Computation



Language Identification



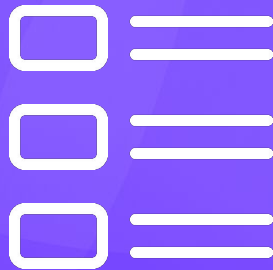
Heuristics + NER



Knowledge Graph



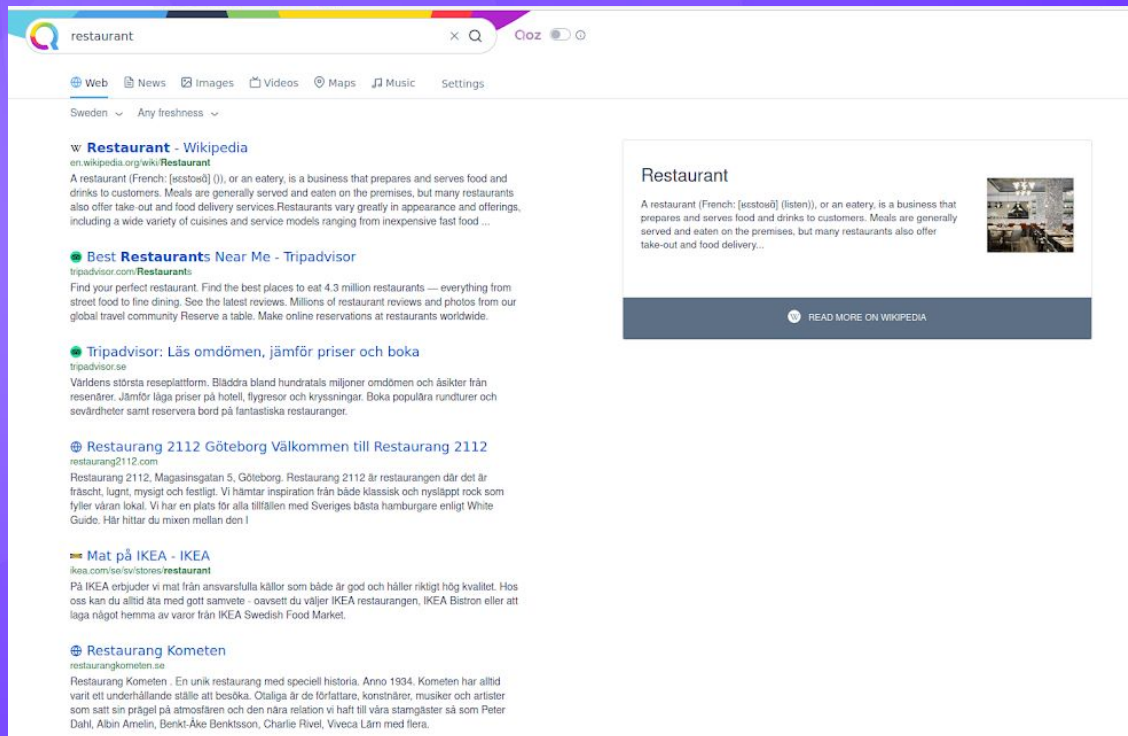
**BM25**



Indexed Data



# “Restaurant”



The screenshot shows a search engine interface with the search term "restaurant". The search bar includes a magnifying glass icon, a close button (X), and a "Go" button. Below the search bar are navigation tabs for "Web", "News", "Images", "Videos", "Maps", "Music", and "Settings". The location is set to "Sweden" and the freshness filter is "Any freshness".

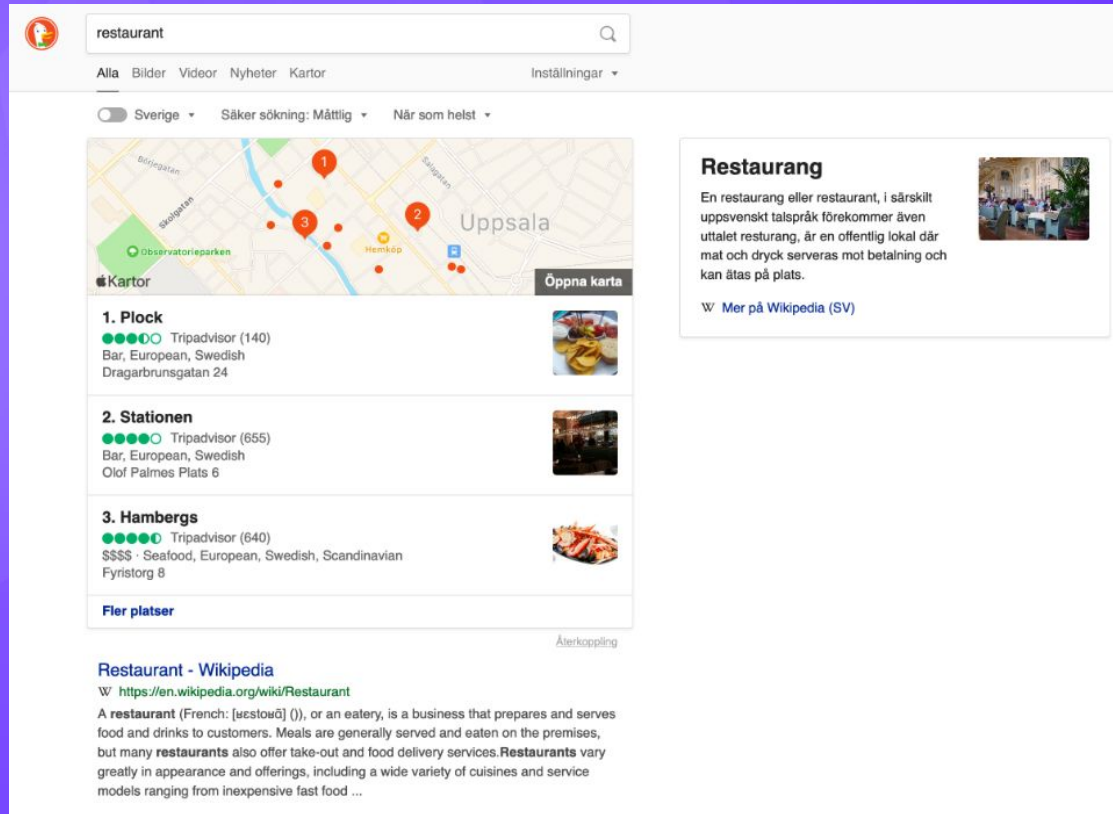
The search results are as follows:

- Restaurant - Wikipedia**  
en.wikipedia.org/wiki/Restaurant  
A restaurant (French: [ʁestotʁɑ̃] ()), or an eatery, is a business that prepares and serves food and drinks to customers. Meals are generally served and eaten on the premises, but many restaurants also offer take-out and food delivery services. Restaurants vary greatly in appearance and offerings, including a wide variety of cuisines and service models ranging from inexpensive fast food ...
- Best Restaurants Near Me - Tripadvisor**  
tripadvisor.com/Restaurants  
Find your perfect restaurant. Find the best places to eat 4.3 million restaurants — everything from street food to fine dining. See the latest reviews. Millions of restaurant reviews and photos from our global travel community. Reserve a table. Make online reservations at restaurants worldwide.
- Tripadvisor: Läs omdömen, jämför priser och boka**  
tripadvisor.se  
Världens största reseplattform. Bläddra bland hundratals miljoner omdömen och åsikter från resenärer. Jämför låga priser på hotell, flygresor och kryssningar. Boka populära rundturer och sovårdheter samt reservera bord på fantastiska restauranger.
- Restaurang 2112 Göteborg Välkommen till Restaurang 2112**  
restaurang2112.com  
Restaurang 2112, Magasinsgatan 5, Göteborg. Restaurang 2112 är restaurangen där det är fräscht, lugnt, mysigt och festligt. Vi hämtar inspiration från både klassisk och nysläppt rock som fyller våran lokal. Vi har en plats för alla tillfällen med Sveriges bästa hamburgare enligt White Guide. Här hittar du mixen mellan den!
- Mat på IKEA - IKEA**  
ikea.com/se/sv/stores/restaurant  
På IKEA erbjuder vi mat från ansvarstulla källor som både är god och håller riktigt hög kvalitet. Hos oss kan du alltid äta med gott samvete - oavsett du väljer IKEA restaurangen, IKEA Bistron eller att laga något hemma av varor från IKEA Swedish Food Market.
- Restaurang Kometen**  
restaurangkometen.se  
Restaurang Kometen - En unik restaurang med speciell historia. Anno 1934. Kometen har alltid varit ett underhållande ställe att besöka. Otaliga är de författare, konstnärer, musiker och artister som satt sin prägel på atmosfären och den nära relation vi haft till våra stamgäster så som Peter Dahl, Albin Amelin, Benkt:Åke Benktsson, Charlie Rivel, Viveca Lärm med flera.

On the right side of the page, there is a featured snippet for "Restaurant" with a small image of a restaurant interior and a "READ MORE ON WIKIPEDIA" button.



# “Restaurant”



restaurant

Alla Bilder Videor Nyheter Kartor Inställningar

Sverige Säker sökning: Mättlig När som helst

Uppsala

1. **Plock**  
●●●○ TripAdvisor (140)  
Bar, European, Swedish  
Dragarbrungatan 24

2. **Stationen**  
●●●○ TripAdvisor (655)  
Bar, European, Swedish  
Olof Palmes Plats 6

3. **Hambergs**  
●●●● TripAdvisor (640)  
\$\$\$\$ - Seafood, European, Swedish, Scandinavian  
Fyristorg 8

Fler platser

Återkoppling

**Restaurang**

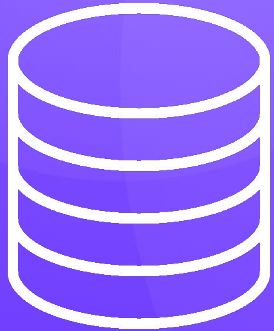
En restaurang eller restaurant, i särskilt uppsvenskt talspråk förekommer även uttalet resturang, är en offentlig lokal där mat och dryck serveras mot betalning och kan ätas på plats.

Mer på Wikipedia (SV)

**Restaurant - Wikipedia**  
W <https://en.wikipedia.org/wiki/Restaurant>  
A **restaurant** (French: [ʁɛstouʁ] ()), or an eatery, is a business that prepares and serves food and drinks to customers. Meals are generally served and eaten on the premises, but many **restaurants** also offer take-out and food delivery services. **Restaurants** vary greatly in appearance and offerings, including a wide variety of cuisines and service models ranging from inexpensive fast food ...



Web Crawler



Simple Storage

1. Language Classification
2. Knowledge Graph Structure



3. Categorization and Filtering:  
ML + AI

Computation



Language Identification/  
Geolocation

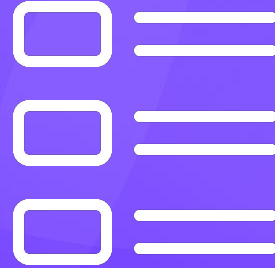


Heuristics + NER



**BM25**

Knowledge Graph



Indexed Data



# Personalizations and Recommendations

“If you and I Google the same thing at the same time, we may get very different results.”

Predicting Models

Recommendation Systems

Previous queries, links, location, language, age, gender, etc...



# Personalizations and Recommendations

“Every search should be pure and democratic. The Pure Web.”

1. What about **racism**?
2. What about **language/geolocalization**?
3. What about **conspiracy theories**?



# Which of the two is the tool?

Relevancy is not **the** factor, is **a** factor



# Which of the two is the tool?

Relevancy is not **the** factor, is **a** factor

**Revenue:** Advertisement, Clicks, Engagement-time, etc...

## **Ads:**

- Anti-vax.
- Extreme political: left-right.
- Flat earth
- Sexism
- Racism



# What about privacy search engines?



# What about privacy search engines?



# Privacy Trap

“When the web was born 30 years or so ago, there were no algorithms. Everyone had access to the same internet.[...] A privacy-focused search engine such as Ecosia will not only protect you from surveillance, it will also give you neutral search results.”

Privacy is **not** burning your data on the fly, but *“the storing, repurposing, provision to third parties, and displaying of information pertaining to oneself via the internet”*



# Generalized Data Protection Regulation (GDPR)

The GDPR's primary aim is to give control to individuals over their personal data and to simplify the regulatory environment.

Think about your hospital

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679>



# Search Engine Gold - Chicken and the Egg

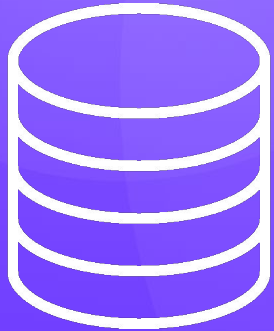
“I would use [random search engine name] if it would be better!”

1. Previous Queries
2. First-page-click-rate (FPCR)
3. Learning to Rank (LTR)
4. A/B Testing

“Then use it more!”



Web Crawler



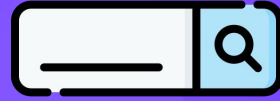
Simple Storage

- 1. Language Classification
- 2. Knowledge Graph Structure



- 3. Categorization and Filtering:  
ML + AI

Computation



Language Identification/  
Geolocation

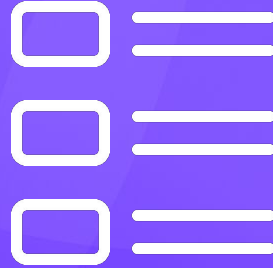


Heuristics + NER



**BM25**

Knowledge Graph



Indexed Data



# Relevancy and Privacy

“Ecological food costs too much, I’ll buy the normal one”

The risk is the **Bubble Effect**

Change search engine often

You know better than anyone else what you are looking for

“You are big and vaccinated” - italian way of saying



# Privacy vs Relevancy

Thanks for your patience!

