adelean
EXTRACT TRANSFORM SEARCH

Apache Lucene:
From Text Indexing to Artificial Intelligence

HAYSTACK

Lucian Precup

2024-04-23

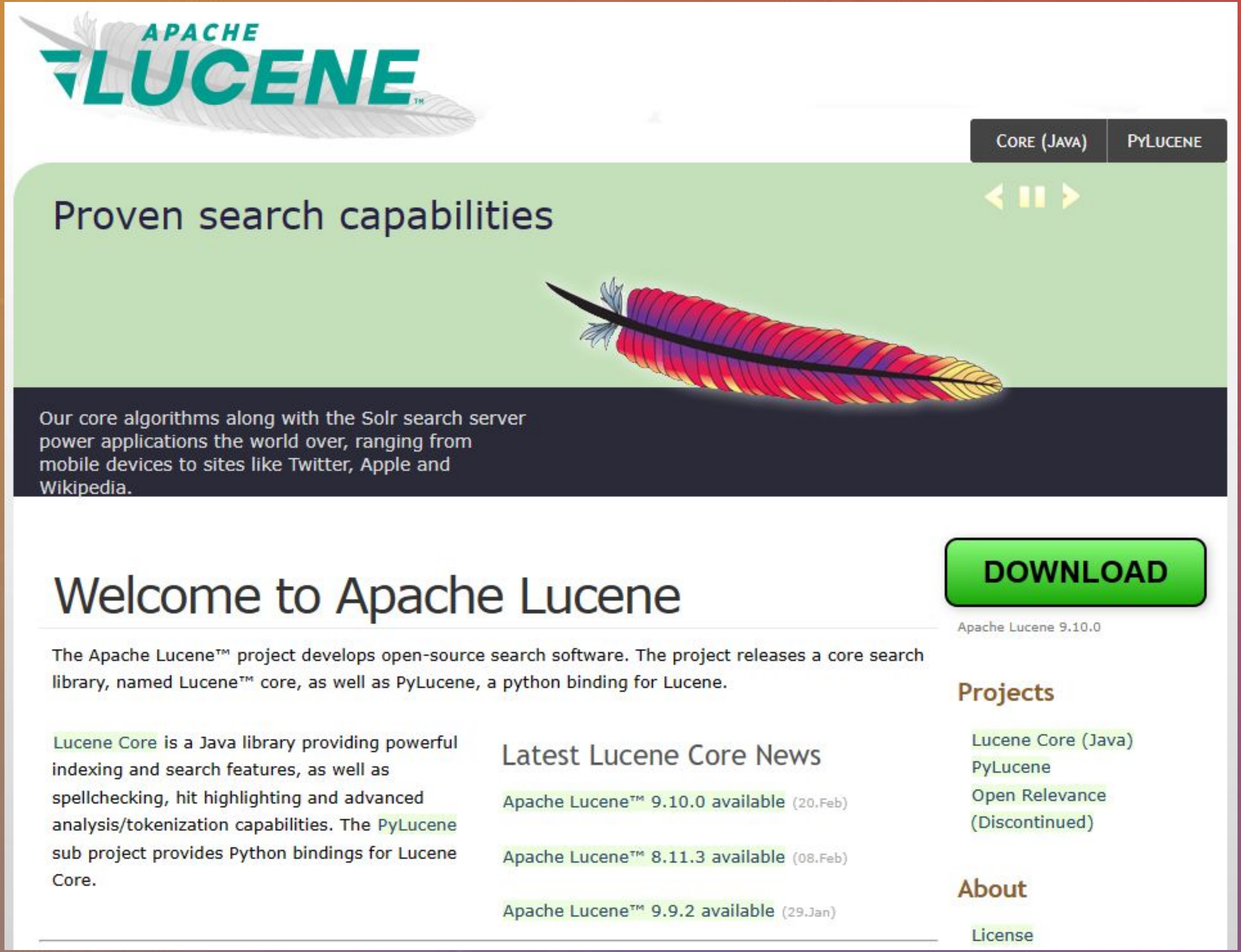
#HaystackConf

@lucianprecup

@a2lean

Poll

- Apache Lucene?
- Apache Solr? Elasticsearch? OpenSearch?
- Other search technologies?



Proven search capabilities

Our core algorithms along with the Solr search server power applications the world over, ranging from mobile devices to sites like Twitter, Apple and Wikipedia.

Welcome to Apache Lucene

The Apache Lucene™ project develops open-source search software. The project releases a core search library, named Lucene™ core, as well as PyLucene, a python binding for Lucene.

Lucene Core is a Java library providing powerful indexing and search features, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities. The PyLucene sub project provides Python bindings for Lucene Core.

Latest Lucene Core News

Apache Lucene™ 9.10.0 available (20.Feb)

Apache Lucene™ 8.11.3 available (08.Feb)

Apache Lucene™ 9.9.2 available (29.Jan)

Apache Lucene 9.10.0

Projects

- Lucene Core (Java)
- PyLucene
- Open Relevance (Discontinued)

About

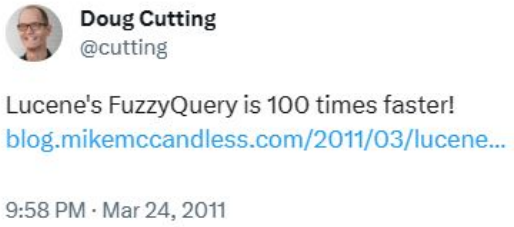
- License

Apache Lucene - 22.5 years and counting

Lucene

APACHE
LUCENE™

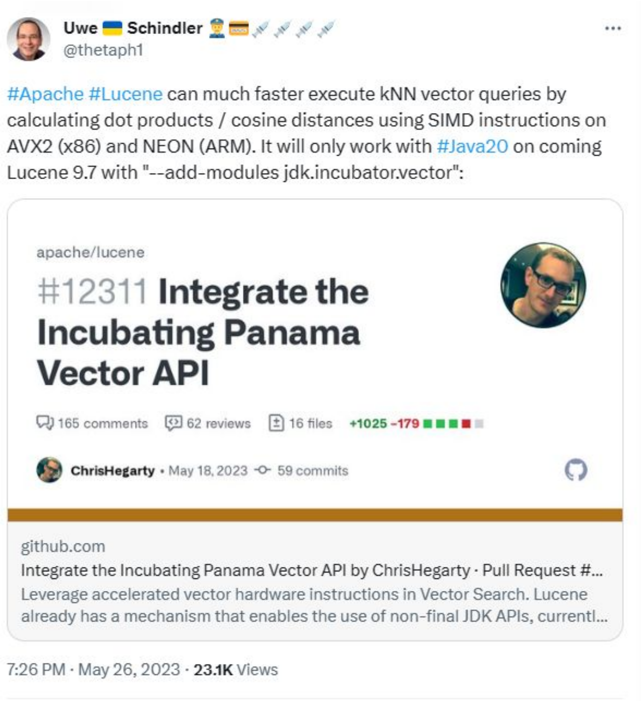
Apache Lucene - 22.5 years and counting



Doug Cutting
@cutting

Lucene's FuzzyQuery is 100 times faster! blog.mikemccandless.com/2011/03/lucene...

9:58 PM · Mar 24, 2011



Uwe 🇺🇦 Schindler 👮💳💉💉💉💉
@thetaph1

#Apache #Lucene can much faster execute kNN vector queries by calculating dot products / cosine distances using SIMD instructions on AVX2 (x86) and NEON (ARM). It will only work with #Java20 on coming Lucene 9.7 with "--add-modules jdk.incubator.vector":

apache/lucene

#12311 Integrate the Incubating Panama Vector API

165 comments · 62 reviews · 16 files · +1025 −179

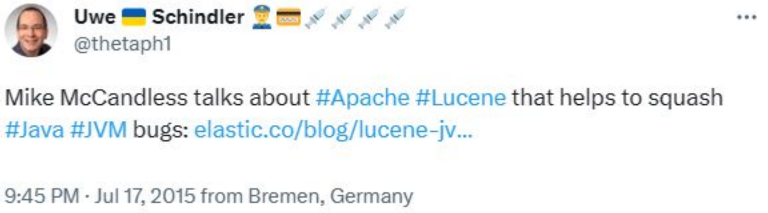
ChrisHegarty • May 18, 2023 · 59 commits

github.com

Integrate the Incubating Panama Vector API by ChrisHegarty · Pull Request #...

Leverage accelerated vector hardware instructions in Vector Search. Lucene already has a mechanism that enables the use of non-final JDK APIs, currentl...

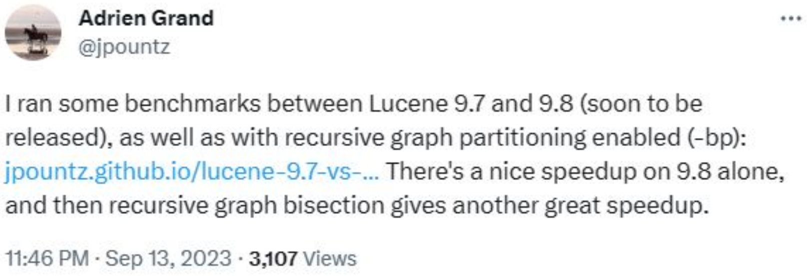
7:26 PM · May 26, 2023 · **23.1K** Views



Uwe 🇺🇦 Schindler 👮💳💉💉💉💉
@thetaph1

Mike McCandless talks about #Apache #Lucene that helps to squash #Java #JVM bugs: elastic.co/blog/lucene-jv...

9:45 PM · Jul 17, 2015 from Bremen, Germany

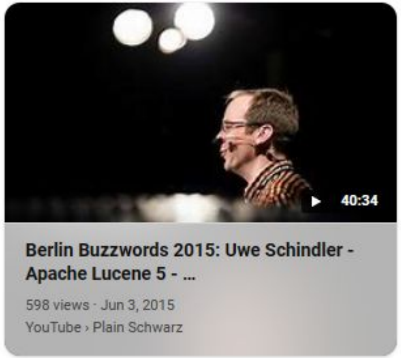


Adrien Grand
@jpountz

I ran some benchmarks between Lucene 9.7 and 9.8 (soon to be released), as well as with recursive graph partitioning enabled (-bp): jpountz.github.io/lucene-9.7-vs-... There's a nice speedup on 9.8 alone, and then recursive graph bisection gives another great speedup.

11:46 PM · Sep 13, 2023 · **3,107** Views

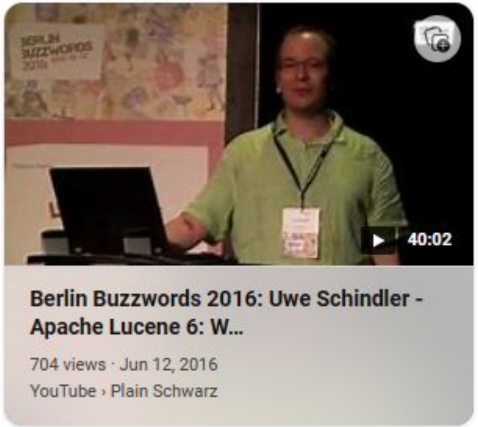
Berlin Buzzwords - The Search, Store and Scale Conference



Berlin Buzzwords 2015: Uwe Schindler - Apache Lucene 5 - ...

598 views · Jun 3, 2015

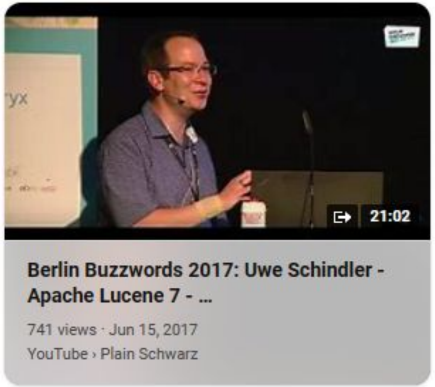
YouTube › Plain Schwarz



Berlin Buzzwords 2016: Uwe Schindler - Apache Lucene 6: W...

704 views · Jun 12, 2016

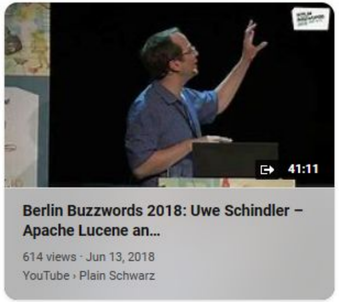
YouTube › Plain Schwarz



Berlin Buzzwords 2017: Uwe Schindler - Apache Lucene 7 - ...

741 views · Jun 15, 2017

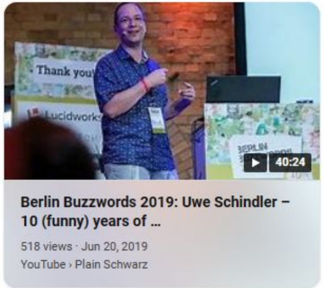
YouTube › Plain Schwarz



Berlin Buzzwords 2018: Uwe Schindler – Apache Lucene an...

614 views · Jun 13, 2018

YouTube › Plain Schwarz

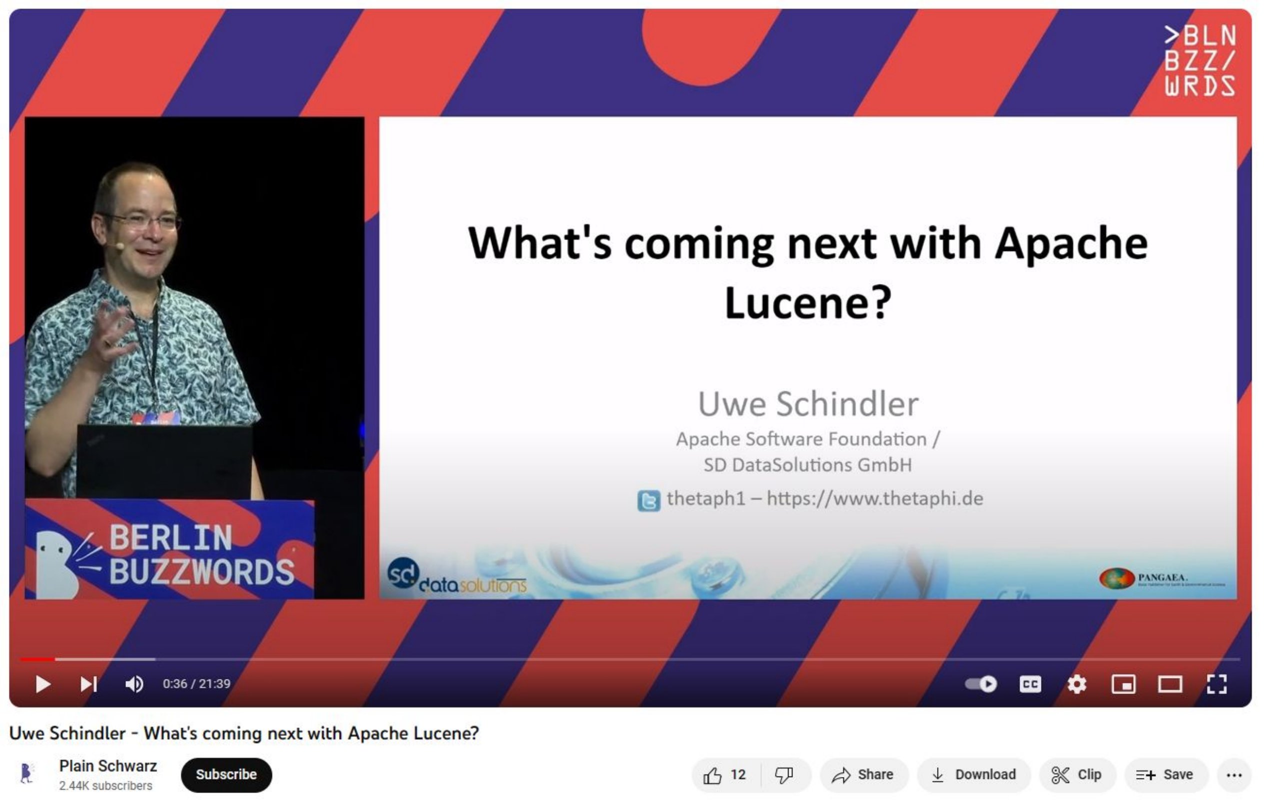


Berlin Buzzwords 2019: Uwe Schindler – 10 (funny) years of ...

518 views · Jun 20, 2019

YouTube › Plain Schwarz

2031 at Berlin Buzzwords



What's coming next with Apache Lucene?

Uwe Schindler
Apache Software Foundation /
SD DataSolutions GmbH
thetaph1 – https://www.thetaphi.de

Uwe Schindler - What's coming next with Apache Lucene?

Plain Schwarz
2.44K subscribers

Thanks to Apache Lucene

Apache Nutch – provides web crawling and HTML parsing

Apache Solr – an enterprise search server

Elasticsearch – an enterprise search server released in 2010

MongoDB Atlas Search – a cloud-native enterprise search application based on MongoDB and Apache Lucene

OpenSearch – an open source enterprise search server based on a fork of Elasticsearch 7

Adelean a2 - an e-commerce and community search server

Who are we?

Adelean

- Experts in **search** technologies
- Integrators of **Elasticsearch**, **OpenSearch** and **Solr**
- **Consulting** and **Training** providers
- Developers of **a2** E-Commerce and Enterprise Search solution
- Developers of **all.site** - your **Collaborative** Search Engine

Inverted index

D1: “All economic sectors, from commerce to the automobile sector to the energy sector, are affected.”

D2: “Registrations of new electric cars in France increased by 0.90% in raw data in July.”

D3: “They rely on technologies that France masters on an industrial and economic level.”

Inverted index

« All economic sectors, from commerce to the automobile sector to the energy sector, are affected. »

↓ Extract all words

All, economic, sectors, from, commerce, to, the, automobile, sector, to, the, energy, sector, are, affected

↓ Sort unique words

All, affected, are, automobile, commerce, economic, energy, from, sector, sector, sectors, the, to

↓ List all documents containing each word

Inverted index

	D1	D2	D3
0.90%		X	
affected	X		
All	X		
an			X
and			X
are	X		
automobile	X		
by		X	
cars		X	
commerce	X		
data		X	
economic	X		X
electric		X	
energy	X		
France		X	X
from	X		
in		X	
increased		X	

	D1	D2	D3
industrial			X
July		X	
level			X
masters			X
new		X	
of		X	
on			X
raw		X	
Registrations		X	
rely			X
sector	X		
sectors	X		
technologies			X
that			X
the	X		
They			X
to	X		

Inverted index

	D1	D2	D3
0.90%		X	
affected	X		
All	X		
an			X
and			X
are	X		
automobile	X		
by		X	
cars		X	
commerce	X		
data		X	
economic	X		X
electric		X	
energy	X		
France		X	X
from	X		
in		X	
increased		X	

	D1	D2	D3
industrial			X
July		X	
level			X
masters			X
new		X	
of		X	
on			X
raw		X	
Registrations		X	
rely			X
sector	X		
sectors	X		
technologies			X
that			X
the	X		
They			X
to	X		

↓ Lowercase

↓ Stemming

↓ Synonyms

↓ Stopwords

↓ …

	D1	D2	D3
0.90%		X	
affected	X		
all	X		
automobile, car	X	X	
commerce	X		
data		X	
economic	X		X
electric		X	
energy	X		
france		X	X
increased		X	
industrial			X
july		X	
level			X
master			X
new		X	
raw		X	
registration		X	
rely			X
sector	X		
technology			X

Inverted index

Search : automobile □ Response : D1, D2
Search : France □ Response : D2, D3
Search : automobile AND France □ Response : D2

- Simple and efficient model for Boolean calculations (intersections)
- + Each token also contains associated metadata

Lucene syntax

Example	Usage
lastname:dujardin	Prefix the search term with the column name
firstname:jean	
lastname:dujardin AND firstname:jean	Combine the criteria with the keyword AND (in capital letters)
lastname:dujardin~	Add a "~" at the end of a term to enable fuzzy searching.
lastname:dujardin~0.8	Add a ratio (~0.x) to widen/narrow the approximate search with a difference of 1-x characters out of 10 (0.9 for 1 character out of 10, 0.8 for 2 characters out of 10, etc.)
postalcode:59*	Use the “*” (star) character for a “starts with” or “containing” search. This principle does not apply for date type columns. The character "?" (question mark) can also be used to specify that it is a single missing character.
location:*yvelines	
location:"rueil malmaison"	Use quotation marks "" to search for a "phrase" (two or more adjacent words)
firstname:"jean françois"	
firstname:"jean paul"~10	Approximate sentences. Use ~n at the end of an expression (sentence) to indicate that there may be interspersed words (n-1 words in this case). This is particularly useful for finding first names like "JEAN MARIE PAUL", "PAUL ANDRE JEAN" or even "PAUL JEAN" when searching for "jean paul"
birthdate:[1949-01-01 TO 1949-12-31]	Use right parentheses ("[" and "]") and the TO keyword to specify an inclusive interval
postalcodel:[91000 TO 95000]	
birthdate:[1949-01-01 TO 1950-01-01}	Use curly braces ("{" and "}") to specify an exclusive interval. You can combine inclusive and exclusive intervals ("[" and "}")
-firstname:"jean paul"	Use the "-" character to exclude responses. In this example we exclude those who have exactly "jean paul" in their first name

Lucene text analysis

Index time

Input documents □

Id	Name
1	Céline
2	Celia

Ascii folding □ Celine, Celia

Lowercase □ celine, celia

Index

Key	Document id
celine	1
celia	2

Query time

Name
CÉLINE

□ Search term

CELINE □ Ascii folding

celine □ Lowercase

Auto-completion with nGram

Index time

Input documents □

Id	Nom
1	Céline
2	Celia

Query time

Nom
Célin

□ Search term

Ascii folding □ Celine, Celia

Lowercase □ celine, celia

EdgeNGram □ ce cel celi celin celine
ce cel celi celia

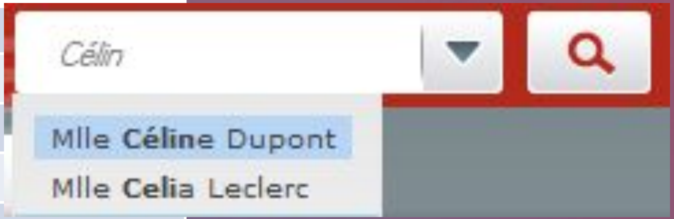
Celin □ Ascii folding

celin □ Lowercase

ce cel celi celin □ EdgeNGram

Index

Key	Document id
ce	1, 2
cel	1, 2
celi	1, 2
celin	1
celine	1
celia	2


Célin

Mlle Céline Dupont

Mlle Celia Leclerc

Different analysis chain at index and query time

Index time

Input documents ☐

Id	Nom
1	Céline
2	Celia

Ascii folding ☐ Celine, Celia

Lowercase ☐ celine, celia

EdgeNGram ☐ ce cel celi celin celine
ce cel celi celia

Query time

Nom
Célin

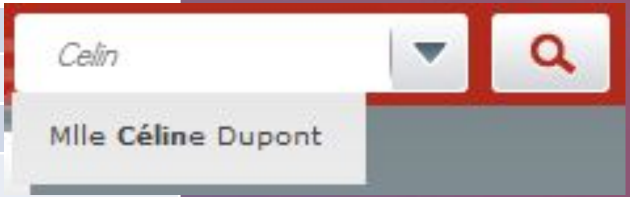
☐ Search term

Celin ☐ Ascii folding

celin ☐ Lowercase

Index

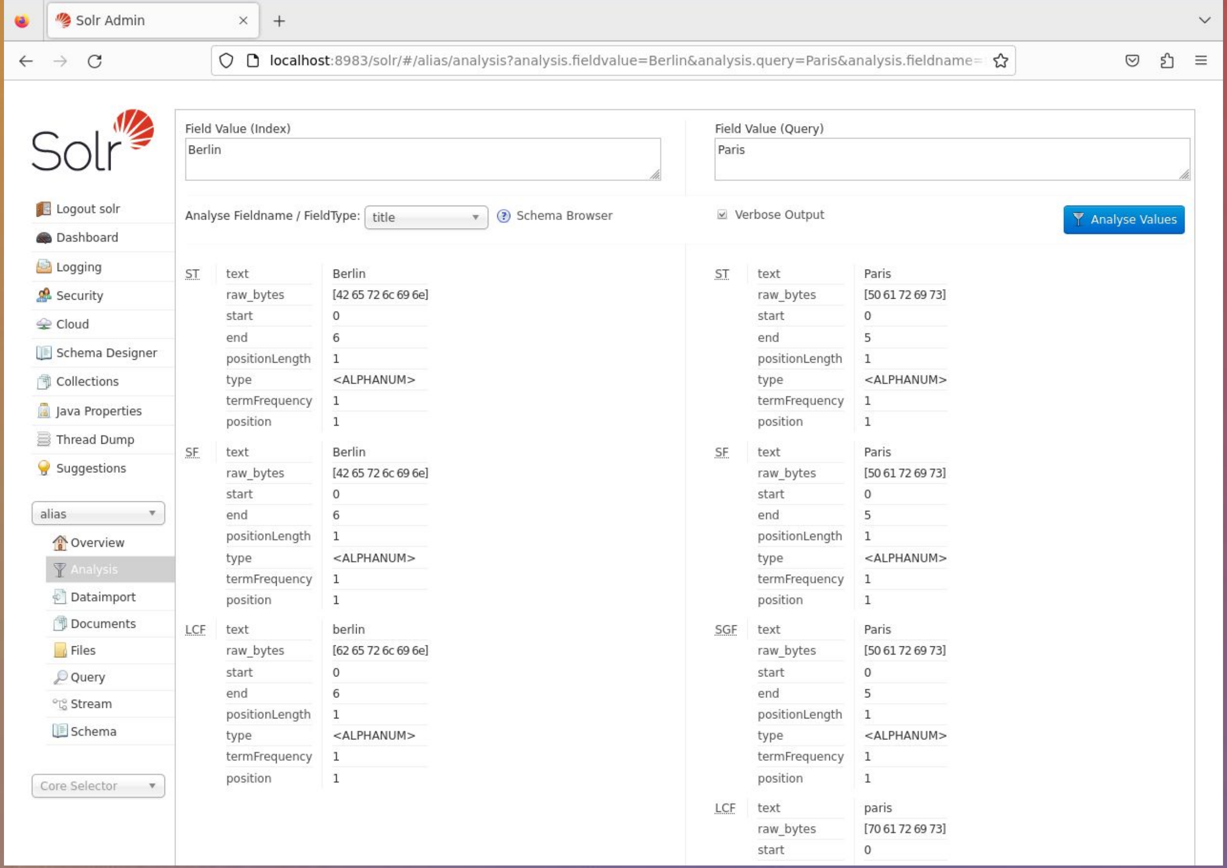
Key	Document id
ce	1, 2
cel	1, 2
celi	1, 2
celin	1
celine	1
celia	2



Celin

Mlle **Céline** Dupont

Analysis menu in the Solr Admin dashboard



localhost:8983/solr/#/alias/analysis?analysis.fieldvalue=Berlin&analysis.query=Paris&analysis.fieldname=

Field Value (Index): Berlin

Field Value (Query): Paris

Analyse Fieldname / FieldType: title

☑ Verbose Output

ST	text	Berlin
	raw_bytes	[42 65 72 6c 69 6e]
	start	0
	end	6
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1
SF	text	Berlin
	raw_bytes	[42 65 72 6c 69 6e]
	start	0
	end	6
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1
LCF	text	berlin
	raw_bytes	[62 65 72 6c 69 6e]
	start	0
	end	6
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1

ST	text	Paris
	raw_bytes	[50 61 72 69 73]
	start	0
	end	5
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1
SF	text	Paris
	raw_bytes	[50 61 72 69 73]
	start	0
	end	5
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1
SGF	text	Paris
	raw_bytes	[50 61 72 69 73]
	start	0
	end	5
	positionLength	1
	type	<ALPHANUM>
	termFrequency	1
	position	1
LCF	text	paris
	raw_bytes	[70 61 72 69 73]
	start	0

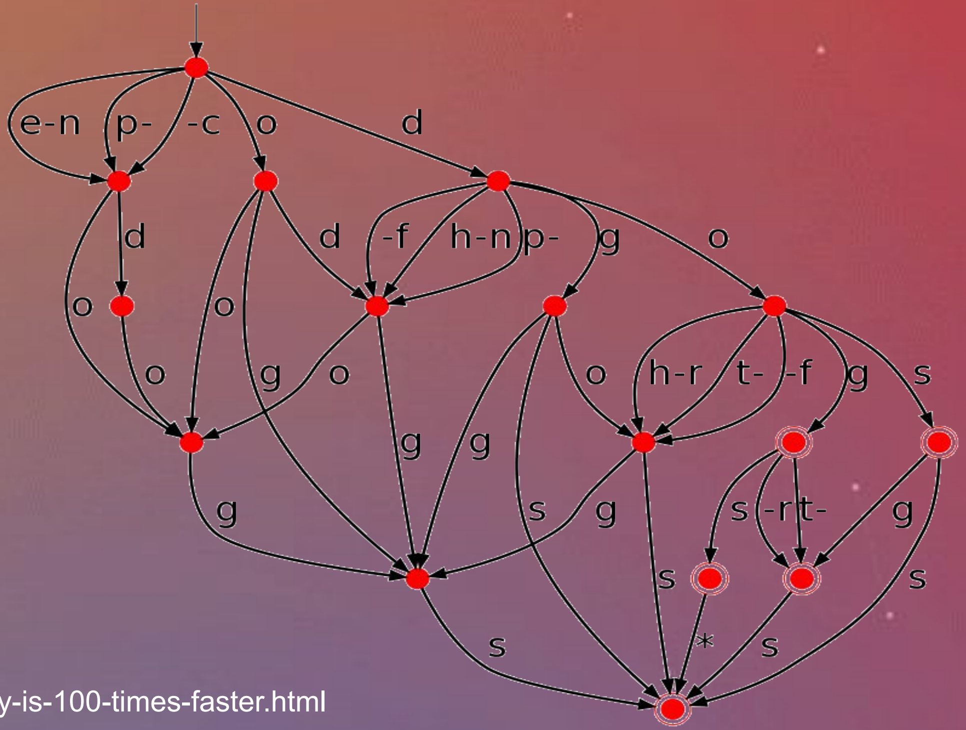
Automaton Invasion

FSA/Ts were difficult to implement but proved to be a good match for Lucene

FSA/Ts are now used in a number of places

FuzzyQuery: 100x times faster

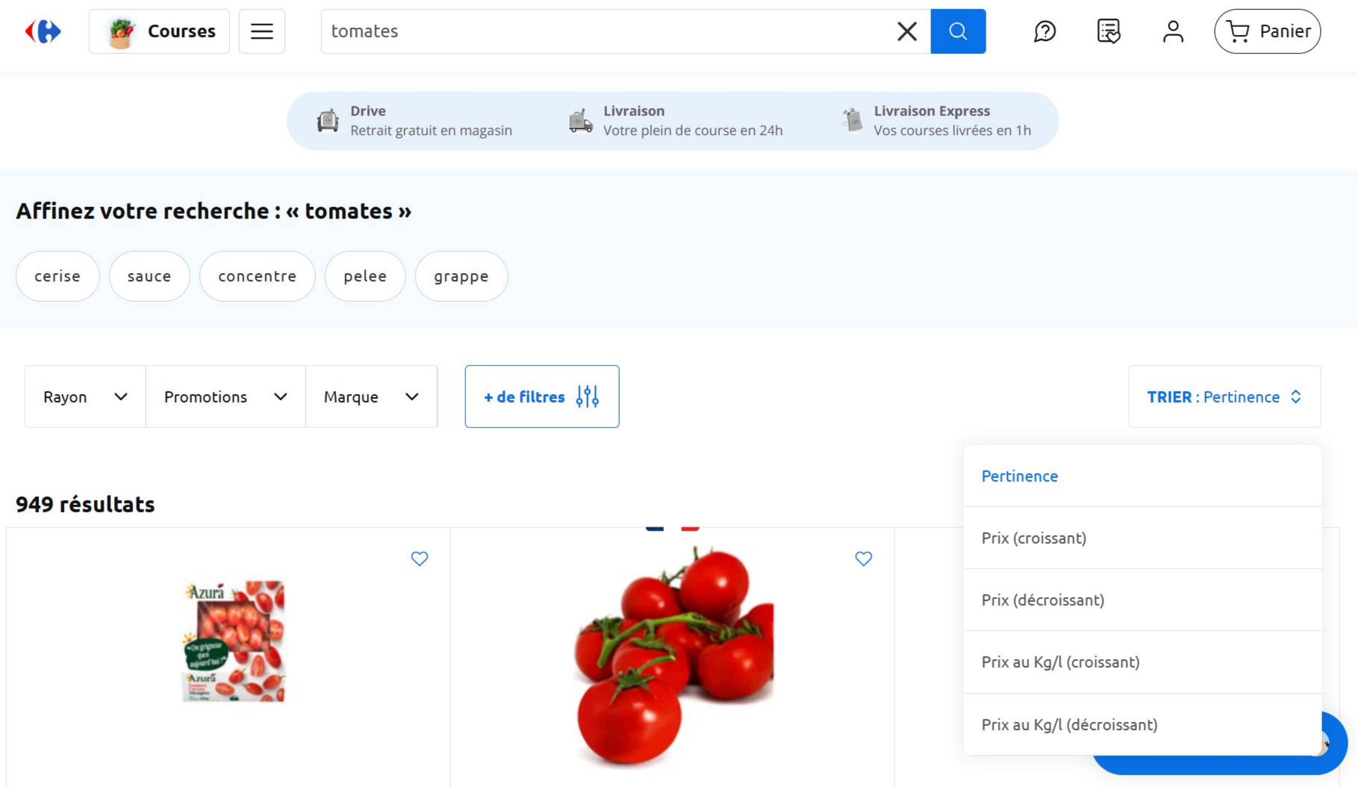
SynonymFilter: 5x times faster, 59X less RAM



https://blog.mikemccandless.com/2011/03/lucenes-fuzzyquery-is-100-times-faster.html

https://blog.mikemccandless.com/2012/05/finite-state-automata-in-lucene.html

Adding (real) support for numbers



tomates

Affinez votre recherche : « tomates »

TRIER : Pertinence

949 résultats

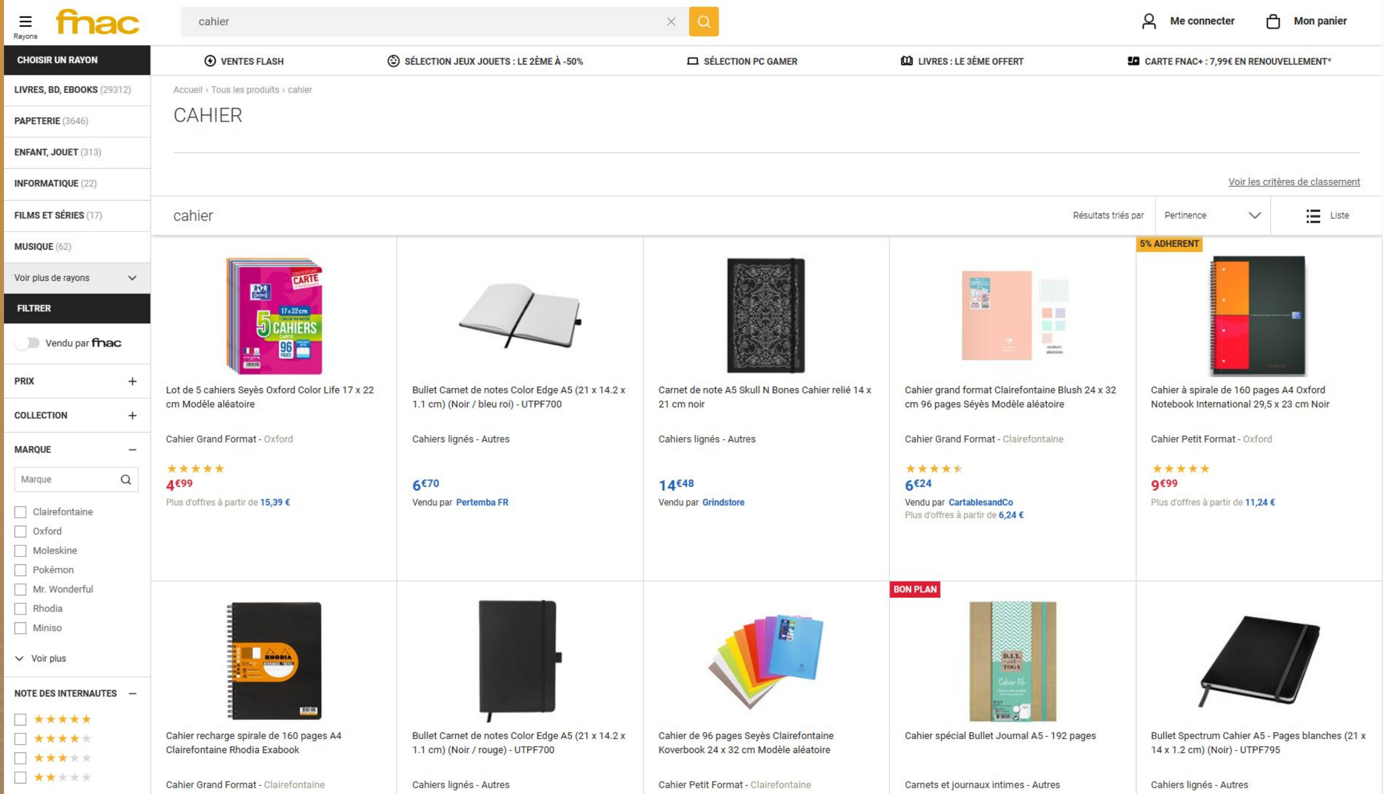
From sorting by pseudo-numbers :

"00000000010" < "00000000300"☺️

to supporting numbers :

10 < 300

Facets, facets, facets, …



Lot de 5 cahiers Seyès Oxford Color Life 17 x 22 cm Modèle aléatoire
Cahier Grand Format - Oxford
★★★★★
4€99
Plus d'offres à partir de 15,39 €

Bullet Carnet de notes Color Edge A5 (21 x 14.2 x 1.1 cm) (Noir / bleu roi) - UTPF700
Cahiers lignés - Autres
6€70
Vendu par Pertemba FR

Carnet de note A5 Skull N Bones Cahier relié 14 x 21 cm noir
Cahiers lignés - Autres
14€48
Vendu par Grindstore

Cahier grand format Clairefontaine Blush 24 x 32 cm 96 pages Séyès Modèle aléatoire
Cahier Grand Format - Clairefontaine
★★★★½
6€24
Vendu par CartablesandCo
Plus d'offres à partir de 6,24 €

Cahier à spirale de 160 pages A4 Oxford Notebook International 29,5 x 23 cm Noir
Cahier Petit Format - Oxford
★★★★★
9€99
Plus d'offres à partir de 11,24 €

Cahier recharge spirale de 160 pages A4 Clairefontaine Rhodia Exabook
Cahier Grand Format - Clairefontaine

Bullet Carnet de notes Color Edge A5 (21 x 14.2 x 1.1 cm) (Noir / rouge) - UTPF700
Cahiers lignés - Autres

Cahier de 96 pages Seyès Clairefontaine Koverbook 24 x 32 cm Modèle aléatoire
Cahier Petit Format - Clairefontaine

Cahier spécial Bullet Journal A5 - 192 pages
Carnets et journaux intimes - Autres

Bullet Spectrum Cahier A5 - Pages blanches (21 x 14 x 1.2 cm) (Noir) - UTPF795
Cahiers lignés - Autres

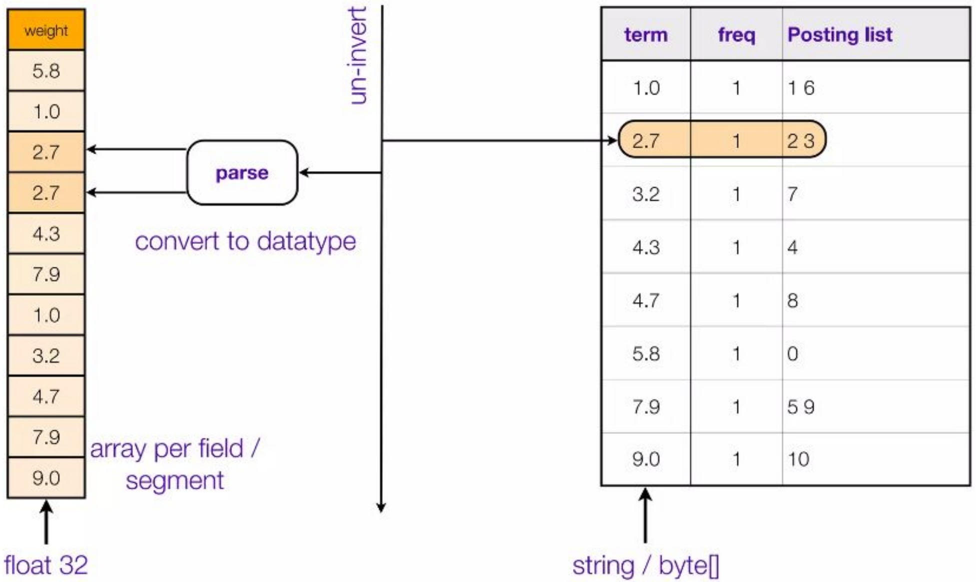
Term facets

Range facets

Hierarchical facets

…

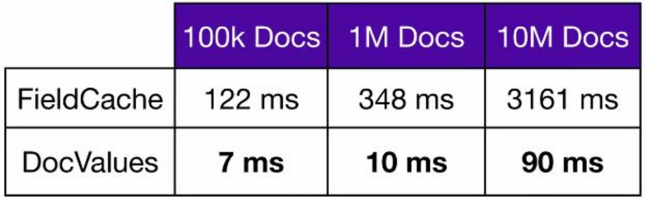
Lucene 4.0: From FieldCache to DocValues



term	freq	Posting list
1.0	1	1 6
2.7	1	2 3
3.2	1	7
4.3	1	4
4.7	1	8
5.8	1	0
7.9	1	5 9
9.0	1	10

→ Excellent performance of sorting, aggregations and highlighting

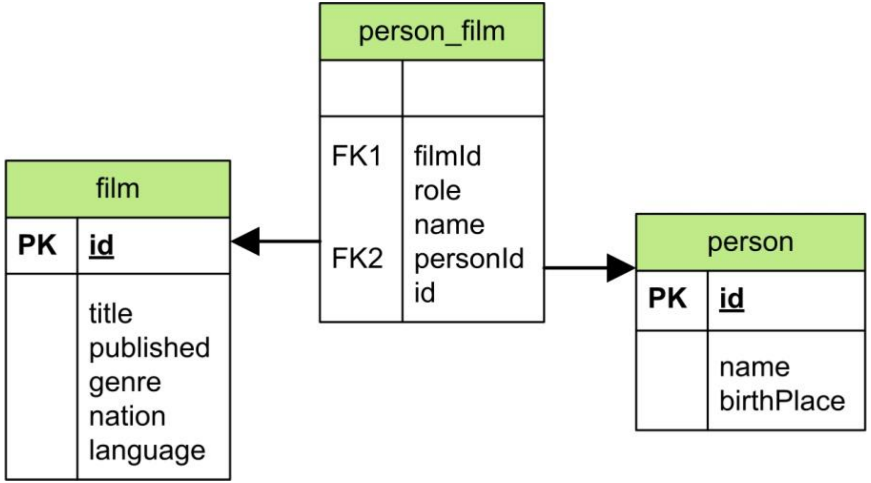
→ Lucene: a columnar store



	100k Docs	1M Docs	10M Docs
FieldCache	122 ms	348 ms	3161 ms
DocValues	**7 ms**	**10 ms**	**90 ms**

https://www.slideshare.net/lucenerevolution/willnauer-simon-doc-values-column-stride-fields-in-lucene

Apache Lucene - a relational database?

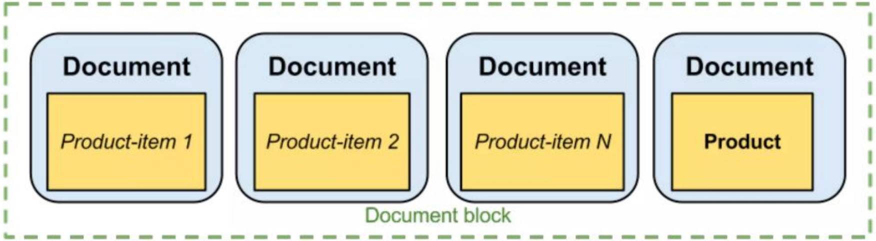


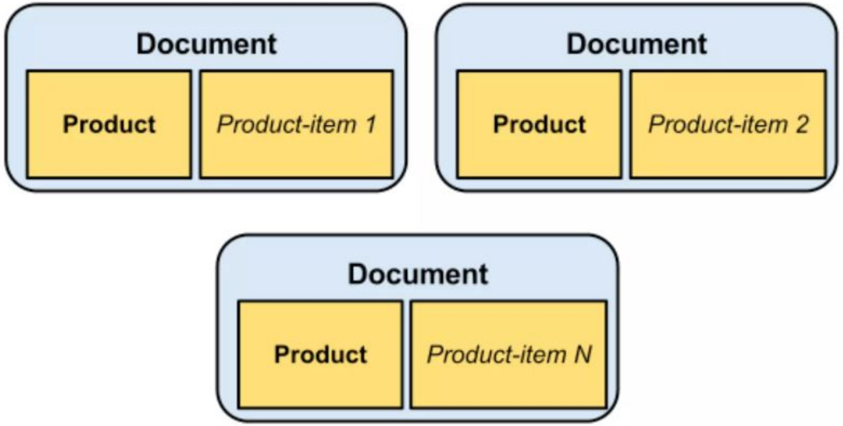
```
{
    "film" : {
        "id" : "183070",
        "title" : "The Artist",
        "published" : "2011-10-12",
        "genre" : ["Romance", "Drama", "Comedy"],
        "language" : ["English", "French"],
        "persons" : [
            {"person" : { "id" : "5079",  "name"
: "Michel Hazanavicius",  "role" : "director" }},
            {"person" : { "id" : "84145",  "name"
: "Jean Dujardin",  "role" : "actor" }},
            {"person" : { "id" : "24485",  "name"
: "Bérénice Bejo",  "role" : "actor" }},
            {"person" : { "id" : "4204",  "name"
: "John Goodman",  "role" : "actor" }}
        ]
    }
}
```

Block joins and result grouping in Apache Lucene

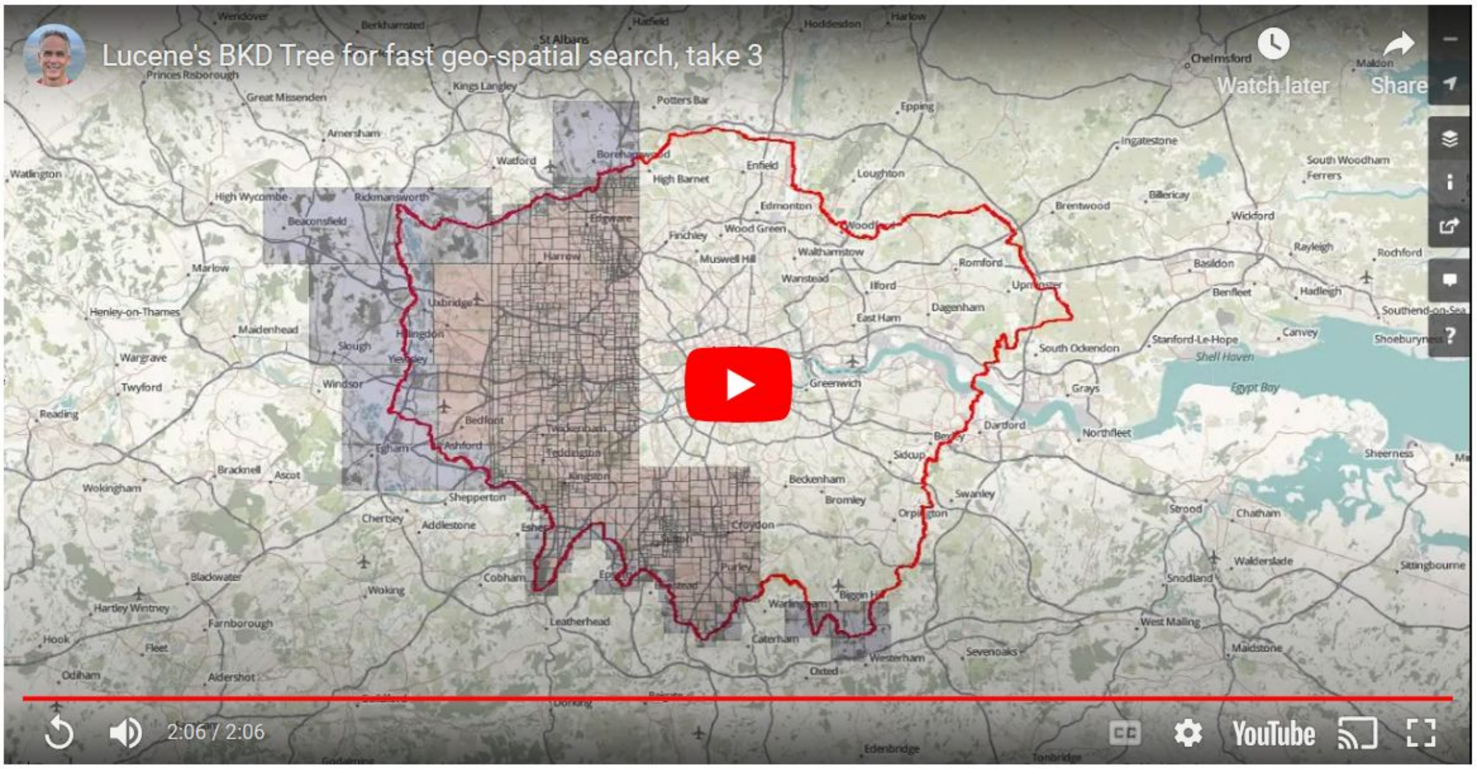
https://www.slideshare.net/lucenerevolution/grouping-and-joining-in-lucenesolr

https://cwiki.apache.org/confluence/display/solr/FieldCollapsing





Multi-dimensional points, coming in Apache Lucene 6.0



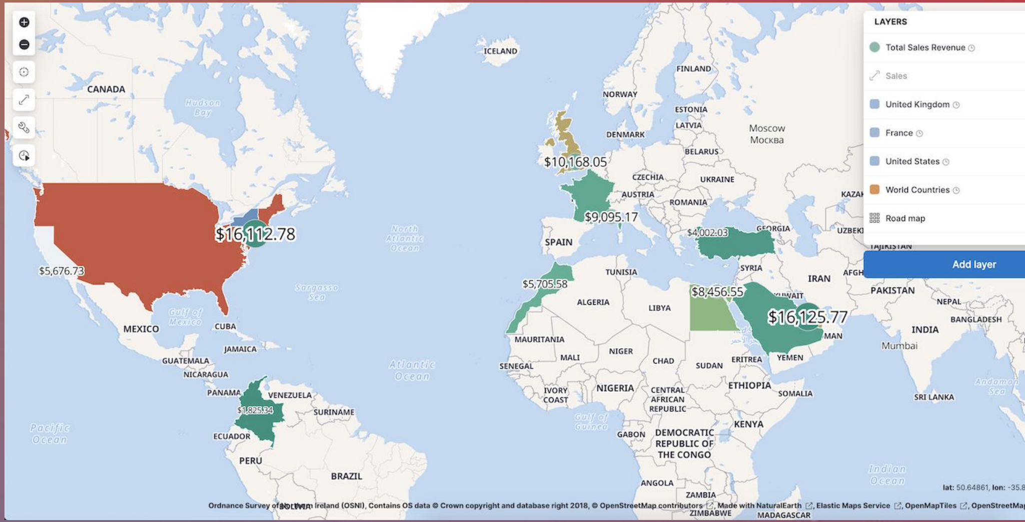
https://www.elastic.co/blog/lucene-points-6-0

Lucene - a spatial database

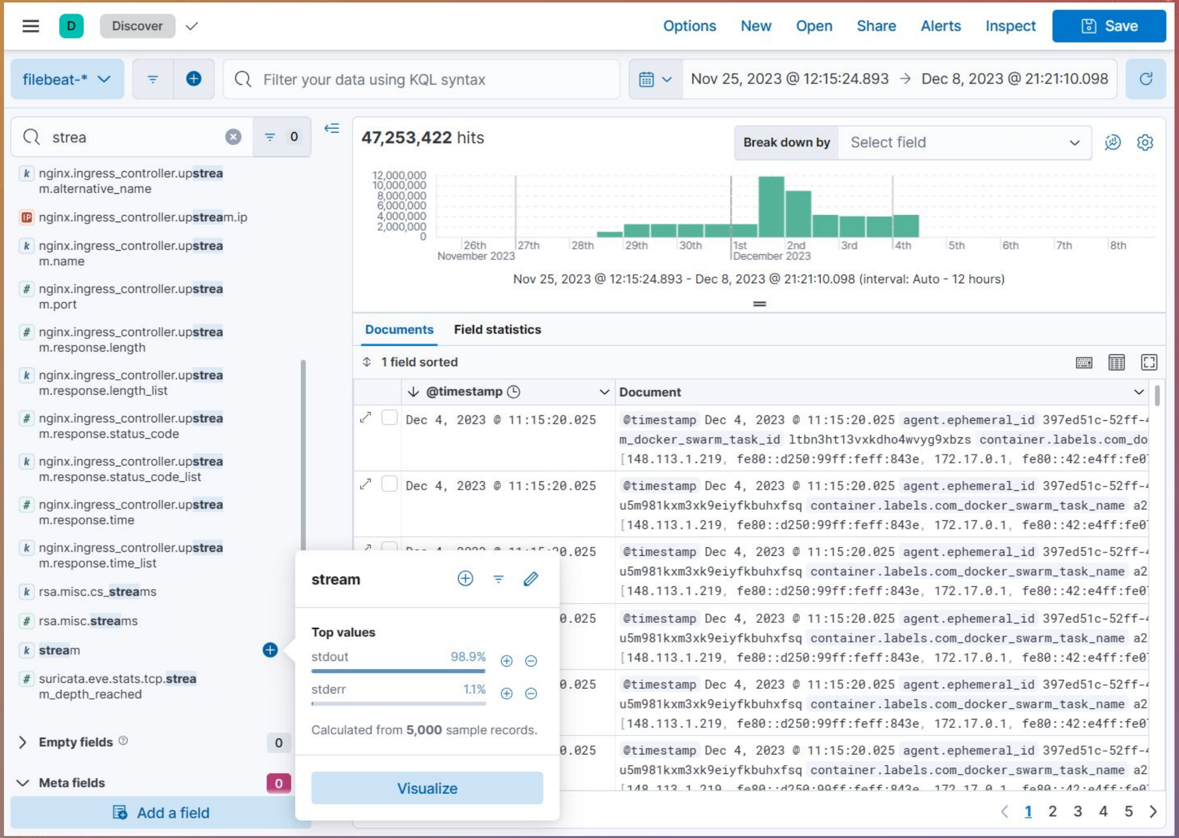
Excellent support for Geo Shape (dots, lines, polygons)

Spatial data requesting (Geo Shape Query, Geo Bounding Box Query, Geo Distance Query, Geo Polygon Query)

This allowed the development of Kibana Maps and the user-friendly support for spatial data in Kibana



New use cases: Monitoring, Observability, …



strea 0

- nginx.ingress_controller.upstream.alternative_name
- nginx.ingress_controller.upstream.ip
- nginx.ingress_controller.upstream.name
- nginx.ingress_controller.upstream.port
- nginx.ingress_controller.upstream.response.length
- nginx.ingress_controller.upstream.response.length_list
- nginx.ingress_controller.upstream.response.status_code
- nginx.ingress_controller.upstream.response.status_code_list
- nginx.ingress_controller.upstream.response.time
- nginx.ingress_controller.upstream.response.time_list
- rsa.misc.cs_streams
- rsa.misc.streams
- stream
- suricata.eve.stats.tcp.stream_depth_reached

Empty fields 0

Meta fields 0

47,253,422 hits

Nov 25, 2023 @ 12:15:24.893 - Dec 8, 2023 @ 21:21:10.098 (interval: Auto - 12 hours)

1 field sorted

stream

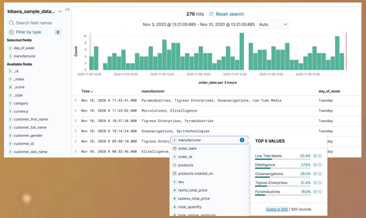
Top values

stdout	98.9%
stderr	1.1%

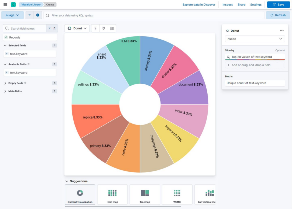
Calculated from **5,000** sample records.

@timestamp	Document
Dec 4, 2023 @ 11:15:20.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- m_docker_swarm_task_id ltbn3ht13vxkdho4wvyg9xbzs container.labels.com_do [148.113.1.219, fe80::d250:99ff:feff:843e, 172.17.0.1, fe80::42:e4ff:fe0
Dec 4, 2023 @ 11:15:20.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- u5m981kxm3xk9eiyfkbuhxfsq container.labels.com_docker_swarm_task_name a2 [148.113.1.219, fe80::d250:99ff:feff:843e, 172.17.0.1, fe80::42:e4ff:fe0
Dec 4, 2023 @ 11:15:20.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- u5m981kxm3xk9eiyfkbuhxfsq container.labels.com_docker_swarm_task_name a2 [148.113.1.219, fe80::d250:99ff:feff:843e, 172.17.0.1, fe80::42:e4ff:fe0
0.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- u5m981kxm3xk9eiyfkbuhxfsq container.labels.com_docker_swarm_task_name a2 [148.113.1.219, fe80::d250:99ff:feff:843e, 172.17.0.1, fe80::42:e4ff:fe0
0.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- u5m981kxm3xk9eiyfkbuhxfsq container.labels.com_docker_swarm_task_name a2 [148.113.1.219, fe80::d250:99ff:feff:843e, 172.17.0.1, fe80::42:e4ff:fe0
0.025	@timestamp Dec 4, 2023 @ 11:15:20.025 agent.ephemeral_id 397ed51c-52ff- u5m981kxm3xk9eiyfkbuhxfsq container.labels.com_docker_swarm_task_name a2

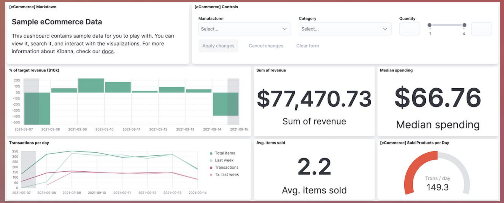
Data analysis with Kibana 🚀



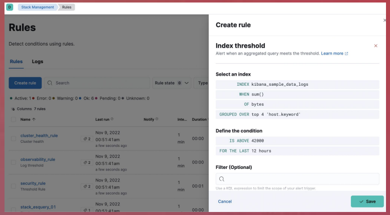
Discover



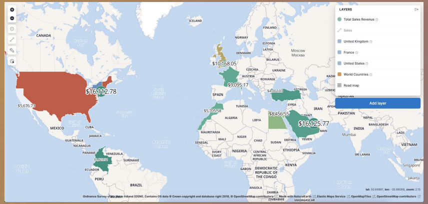
Lens



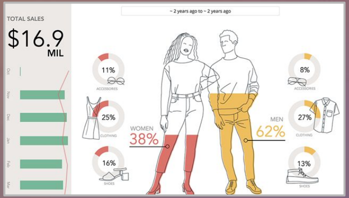
Dashboards



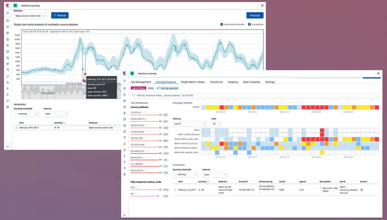
Alerting



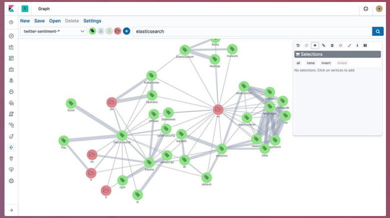
Maps



Canvas



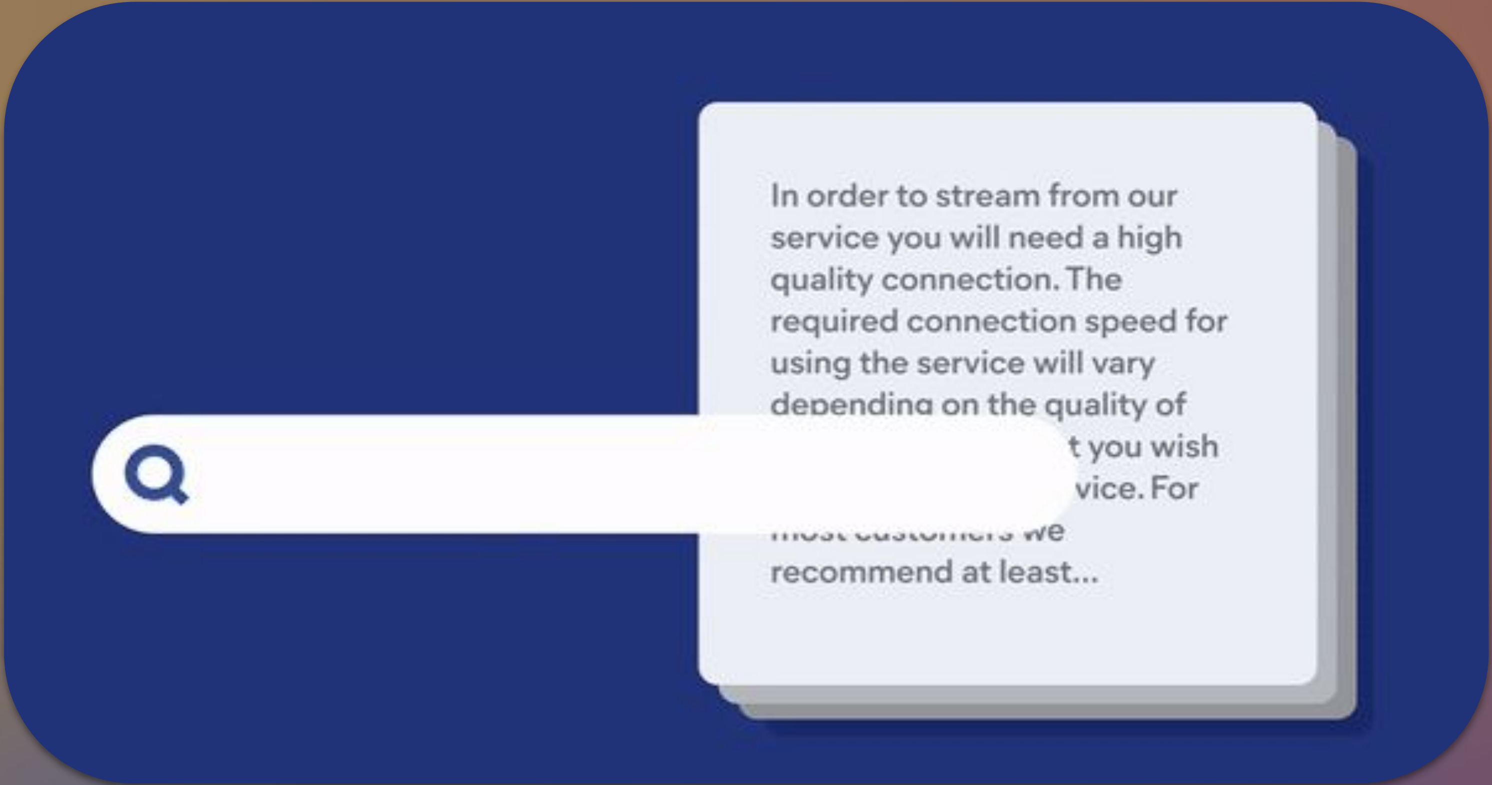
Machine Learning



Graph

NLP models and vector search: a paradigm shift

Move beyond text-matching

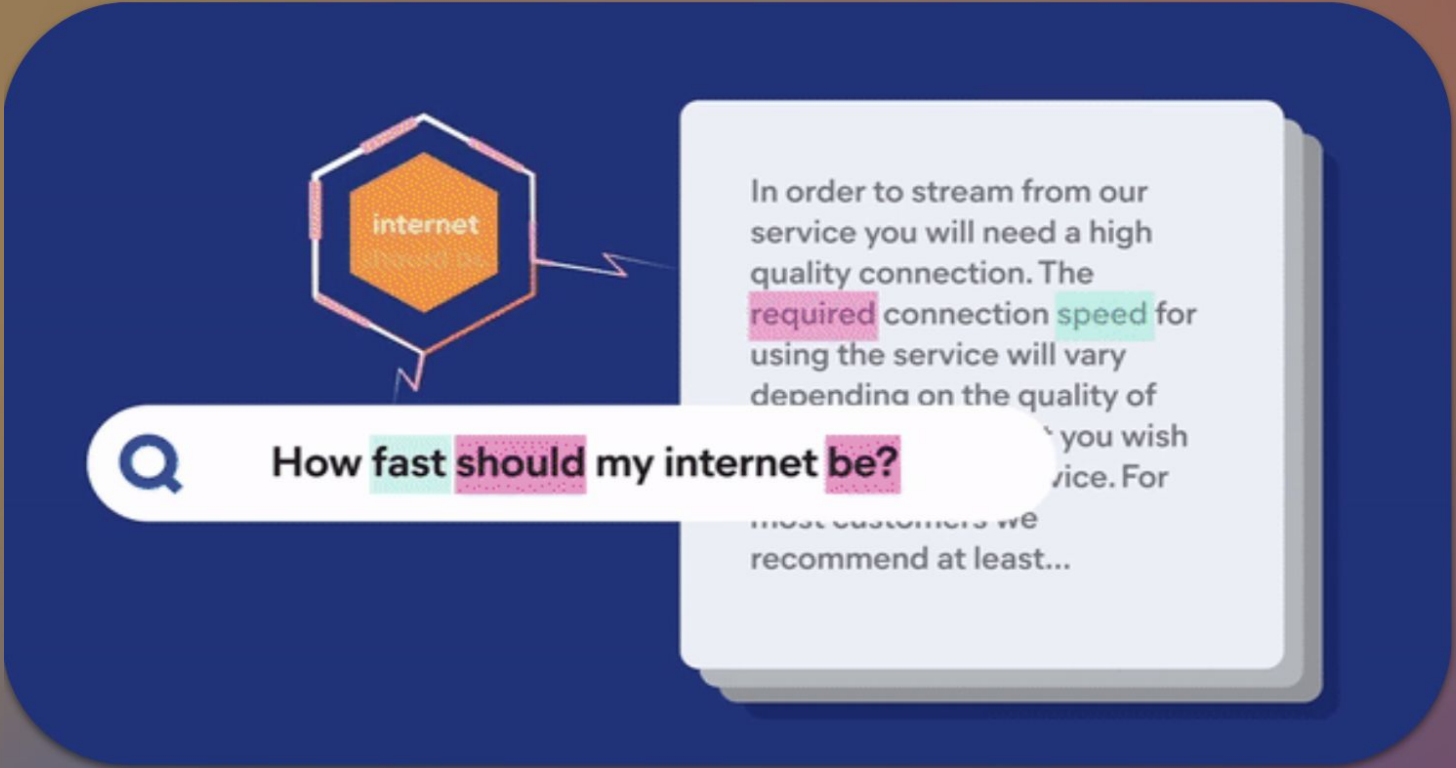


In order to stream from our
service you will need a high
quality connection. The
required connection speed for
using the service will vary
depending on the quality of
t you wish
vice. For
most customers we
recommend at least...

https://www.elastic.co/elasticsearch/vector-database

NLP models and vector search: a paradigm shift

Move beyond text-matching



How fast should my internet be?

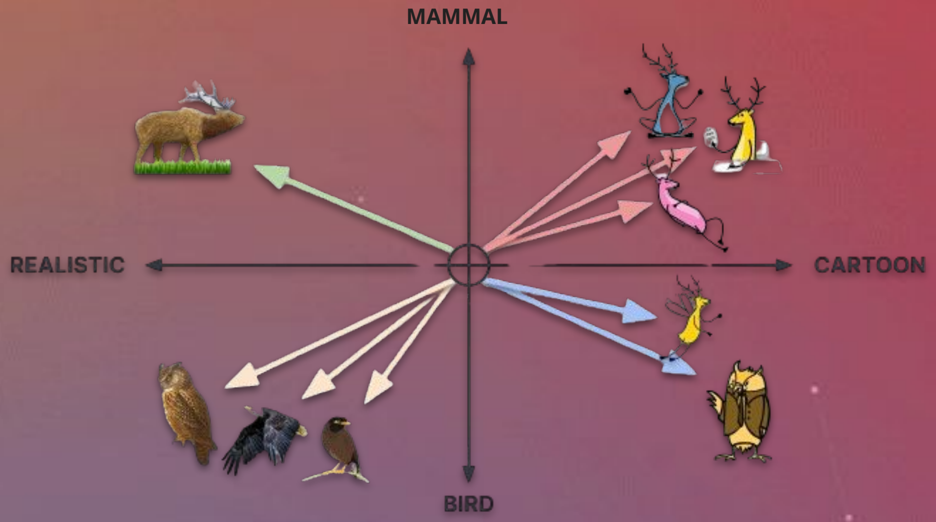
In order to stream from our service you will need a high quality connection. The required connection speed for using the service will vary depending on the quality of … you wish … vice. For most customers we recommend at least...

https://www.elastic.co/elasticsearch/vector-database

Embeddings or dense semantic vector

Vectorization makes it possible to represent textual documents as numbers in a mathematical space, which facilitates their analysis and processing by computer algorithms.

Each element in this vector represents a characteristic of the document.



https://www.elastic.co/blog/whats-new-elasticsearch-kibana-cloud-8-4-0

adelean
EXTRACT TRANSFORM SEARCH

Vectorization in a three-dimensional vector

size

friendliness

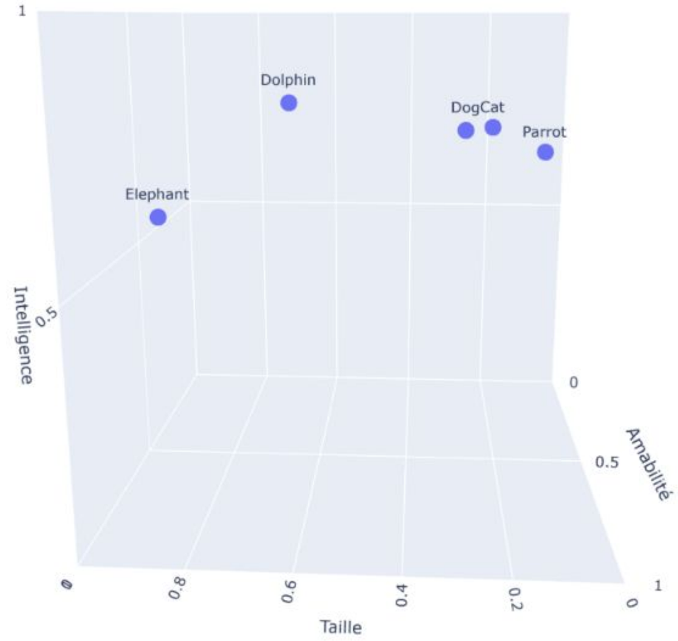
intelligence

Animal	Size	Friendliness	Intelligence
Cat	0.25	0.85	0.80
Dog	0.30	0.90	0.80
Elephant	0.90	0.70	0.60
Dolphin	0.60	0.95	0.85
Parrot	0.15	0.80	0.75

https://www.adelean.com/en/blog/20240131_vectors_sparse_and_dense/

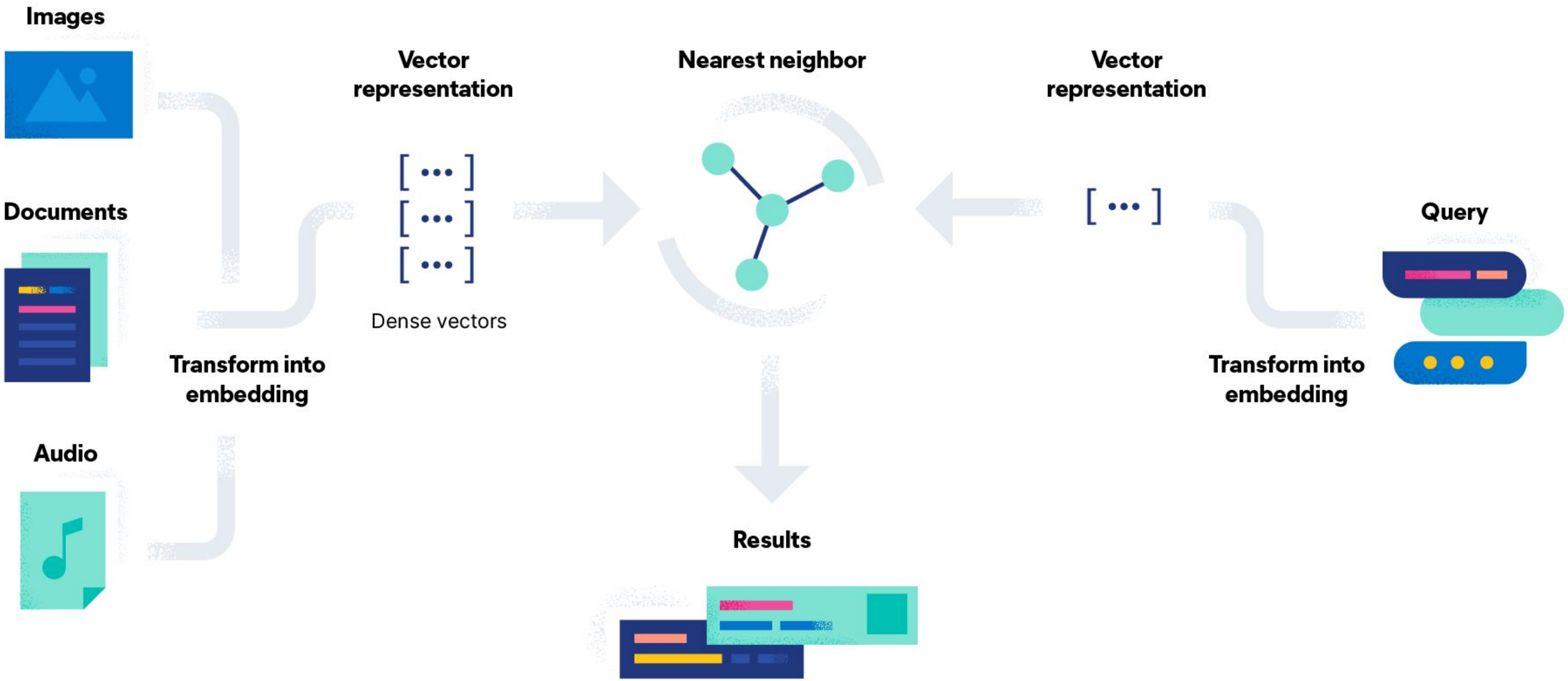
Vectorization in a three-dimensional vector

Animal	Size	Friendliness	Intelligence
Cat	0.25	0.85	0.80
Dog	0.30	0.90	0.80
Elephant	0.90	0.70	0.60
Dolphin	0.60	0.95	0.85
Parrot	0.15	0.80	0.75



https://www.adelean.com/en/blog/20240131_vectors_sparse_and_dense/

Text embedding and vector search



Images

Documents

Audio

Transform into embedding

Vector representation

Dense vectors

Nearest neighbor

Vector representation

Query

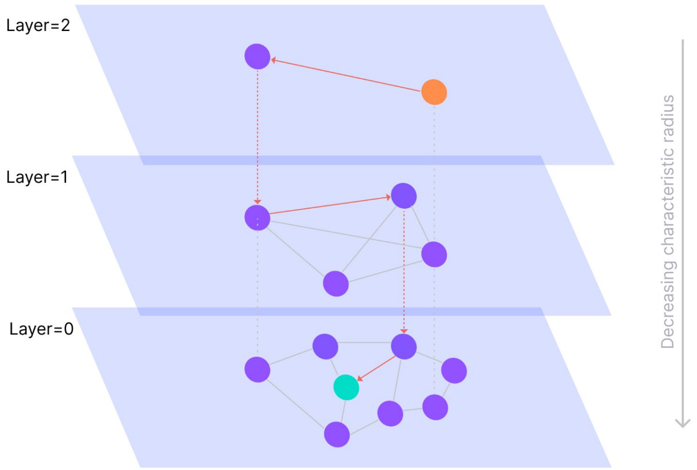
Transform into embedding

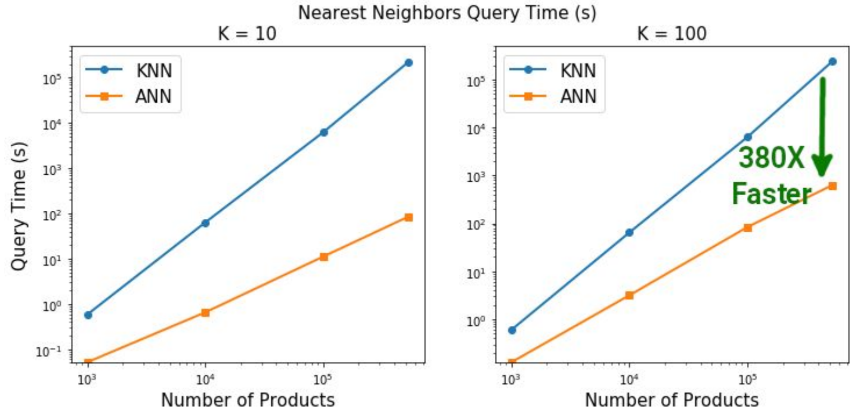
Results

https://www.elastic.co/fr/what-is/vector-search

k-NN (Lucene 9.0) vs. approximate k-NN (Lucene 9.2)

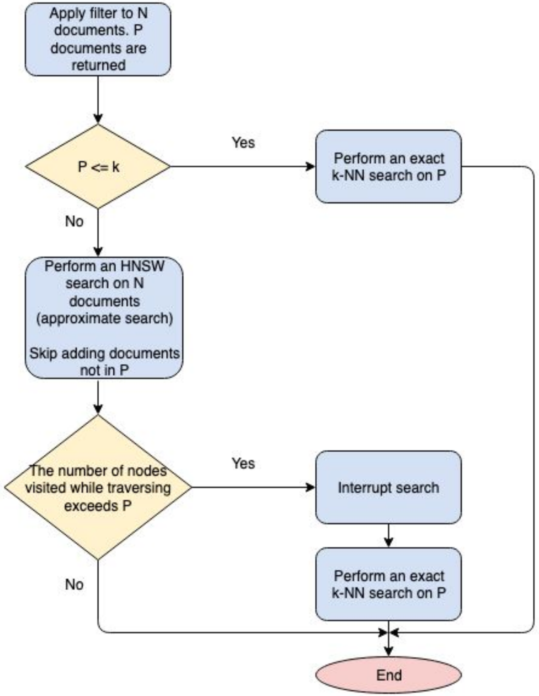
Hierarchical Navigable Small World (HNSW)





See also: https://opensearch.org/blog/Expanding-k-NN-with-Lucene-aNN/

Approximate k-NN with filters


Apply filter to N documents. P documents are returned
P <= k
Yes
Perform an exact k-NN search on P
No
Perform an HNSW search on N documents (approximate search)
Skip adding documents not in P
The number of nodes visited while traversing exceeds P
Yes
Interrupt search
Perform an exact k-NN search on P
No
End

pre- or post-filtering in order to provide the best speed and accuracy

→ The first step to Hybrid Search

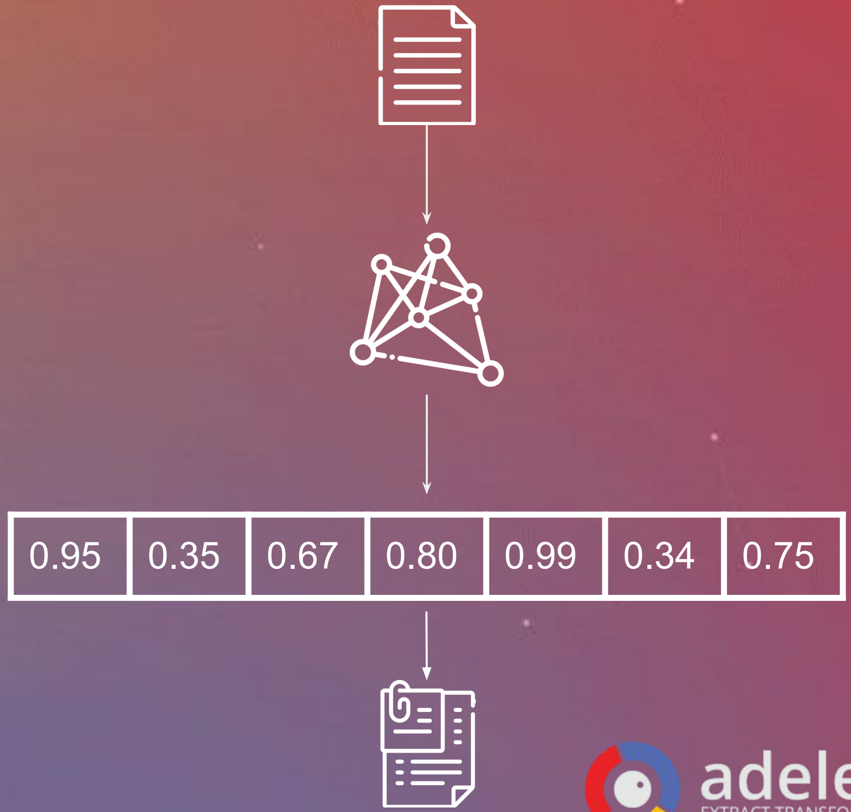
https://opensearch.org/docs/latest/search-plugins/knn/filter-search-knn/

Sparse or dense semantic vector?

Dense vectors are expensive to calculate and maintain.

A model will be used to transform each document.

The vector will be added to the transformed document, and a similarity algorithm will be used to find the most relevant document.



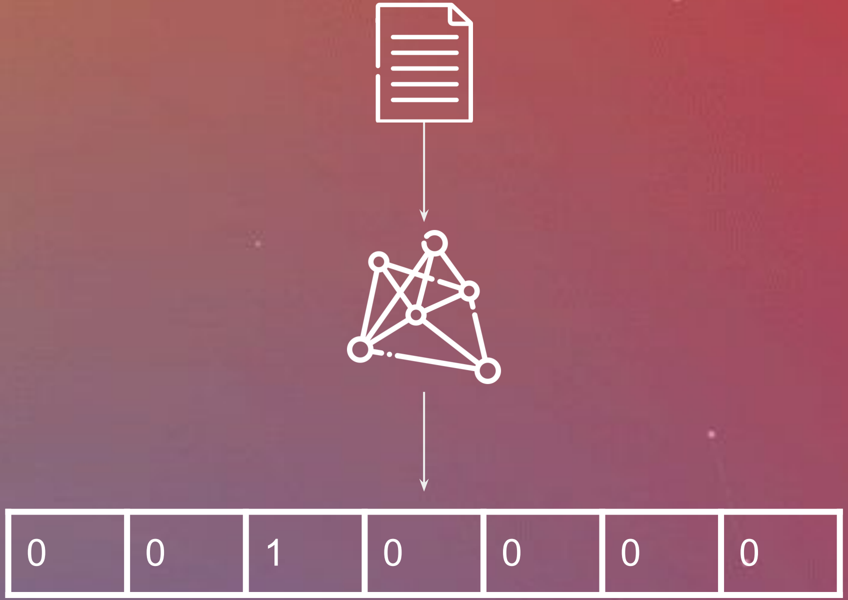
0.95	0.35	0.67	0.80	0.99	0.34	0.75

Sparse or dense semantic vector?

Sparse vectors are less expensive to maintain.

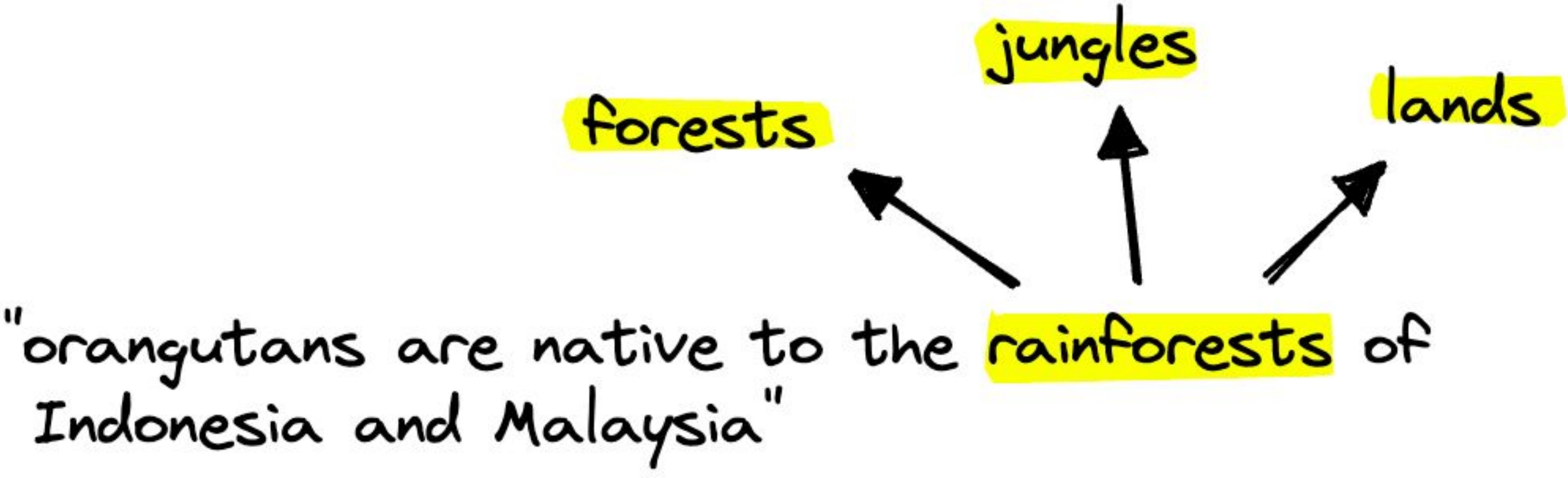
A model will be used to augment the data

The inverted index will be used to store the augmented data, just as it is used for any field in our document.



0	0	1	0	0	0	0

Example of term expansion

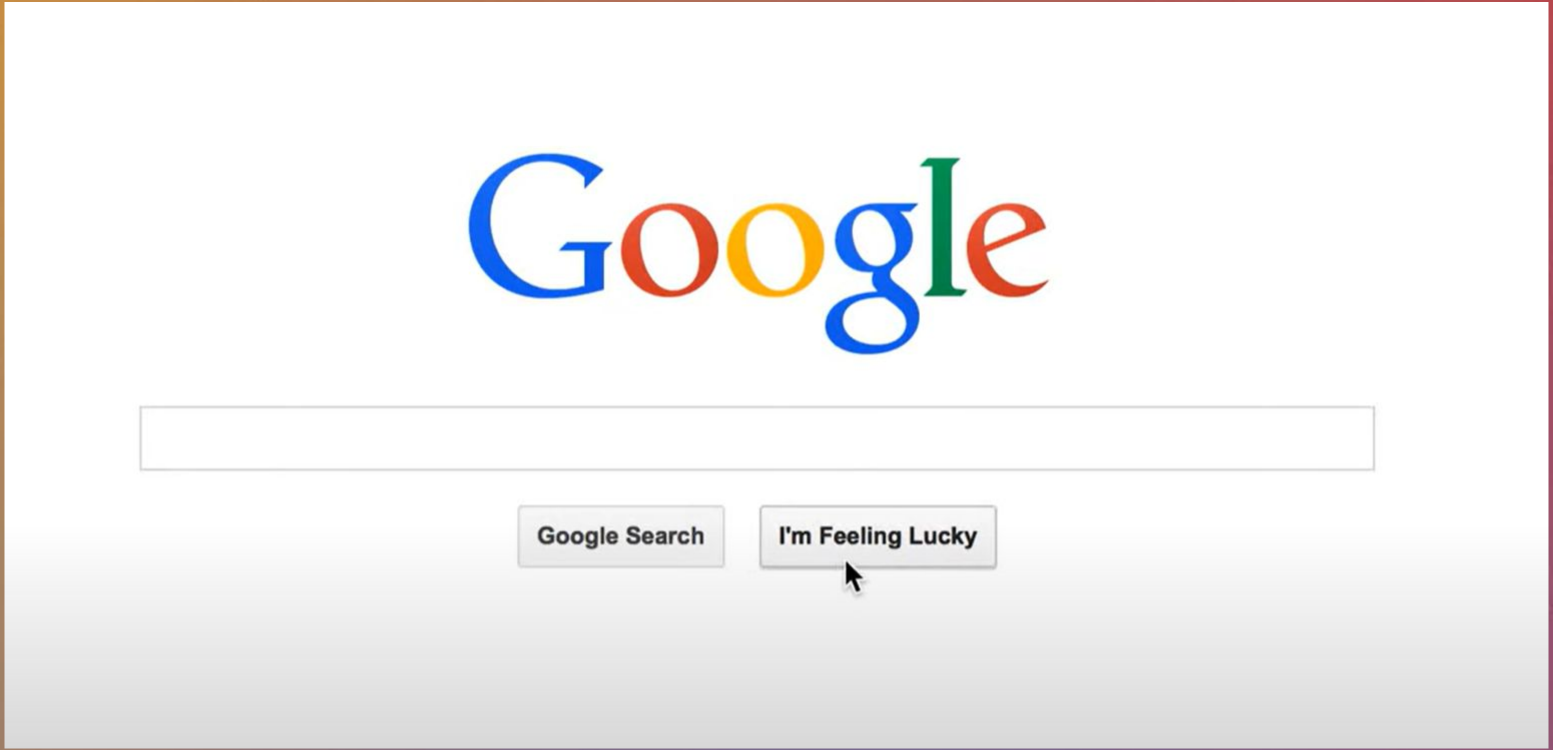

jungles
forests
lands
"orangutans are native to the rainforests of Indonesia and Malaysia"

ChatGPT

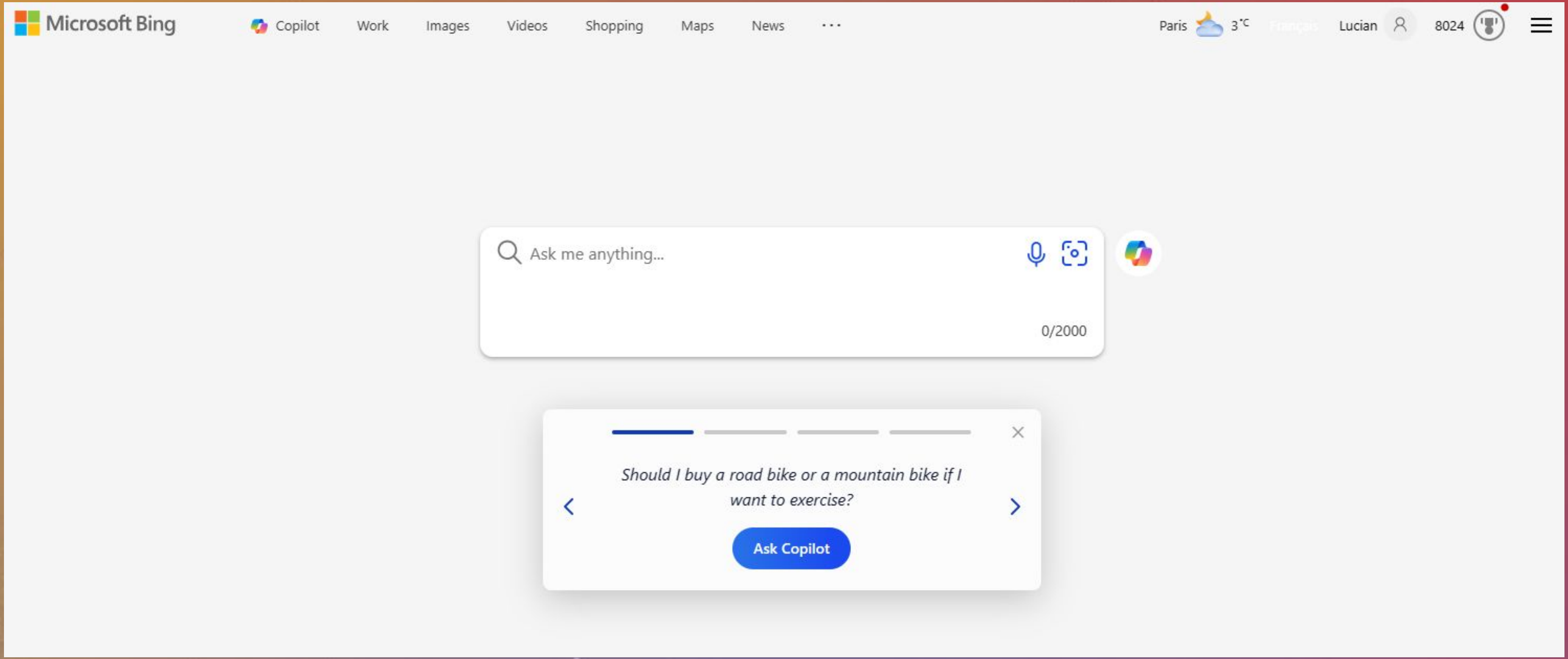
Poll: how are you using ChatGPT?

- Search engine?

From I am Feeling Lucky



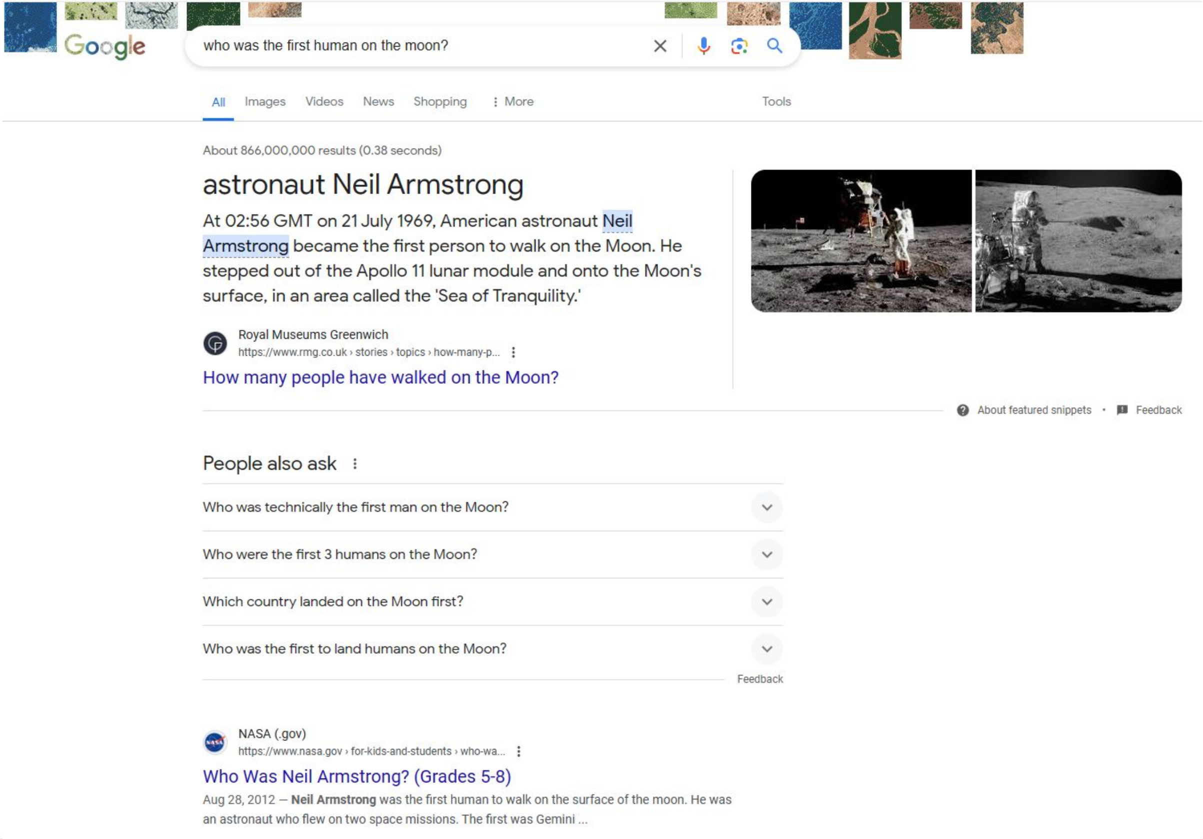
To Copilot



Should I buy a road bike or a mountain bike if I want to exercise?

Back to Haystack On Tour Paris - November 2022

Question answering - a new stage for search engines



who was the first human on the moon?

About 866,000,000 results (0.38 seconds)

astronaut Neil Armstrong

At 02:56 GMT on 21 July 1969, American astronaut Neil Armstrong became the first person to walk on the Moon. He stepped out of the Apollo 11 lunar module and onto the Moon's surface, in an area called the 'Sea of Tranquility.'

Royal Museums Greenwich
https://www.rmg.co.uk › stories › topics › how-many-p...

How many people have walked on the Moon?

People also ask

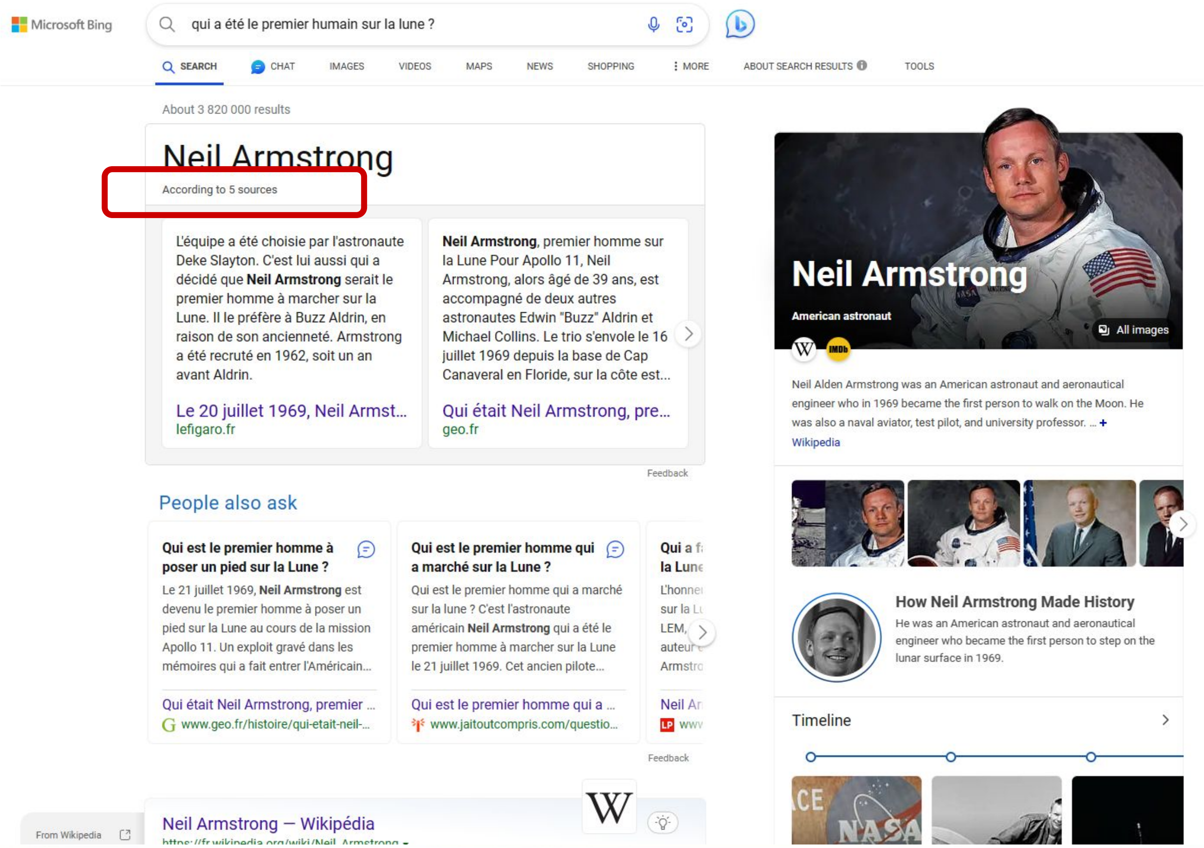
- Who was technically the first man on the Moon?
- Who were the first 3 humans on the Moon?
- Which country landed on the Moon first?
- Who was the first to land humans on the Moon?

NASA (.gov)
https://www.nasa.gov › for-kids-and-students › who-wa...

Who Was Neil Armstrong? (Grades 5-8)

Aug 28, 2012 — **Neil Armstrong** was the first human to walk on the surface of the moon. He was an astronaut who flew on two space missions. The first was Gemini ...

Question answering - cross checking the results



qui a été le premier humain sur la lune ?

About 3 820 000 results

Neil Armstrong

According to 5 sources

L'équipe a été choisie par l'astronaute Deke Slayton. C'est lui aussi qui a décidé que **Neil Armstrong** serait le premier homme à marcher sur la Lune. Il le préfère à Buzz Aldrin, en raison de son ancienneté. Armstrong a été recruté en 1962, soit un an avant Aldrin.

Le 20 juillet 1969, Neil Armst...
lefigaro.fr

Neil Armstrong, premier homme sur la Lune Pour Apollo 11, Neil Armstrong, alors âgé de 39 ans, est accompagné de deux autres astronautes Edwin "Buzz" Aldrin et Michael Collins. Le trio s'envole le 16 juillet 1969 depuis la base de Cap Canaveral en Floride, sur la côte est...

Qui était Neil Armstrong, pre...
geo.fr

People also ask

Qui est le premier homme à poser un pied sur la Lune ?

Le 21 juillet 1969, **Neil Armstrong** est devenu le premier homme à poser un pied sur la Lune au cours de la mission Apollo 11. Un exploit gravé dans les mémoires qui a fait entrer l'Américain...

Qui était Neil Armstrong, premier ...
www.geo.fr/histoire/qui-etait-neil-...

Qui est le premier homme qui a marché sur la Lune ?

Qui est le premier homme qui a marché sur la lune ? C'est l'astronaute américain **Neil Armstrong** qui a été le premier homme à marcher sur la Lune le 21 juillet 1969. Cet ancien pilote...

Qui est le premier homme qui a ...
www.jaitoutcompris.com/questio...

Qui a f
la Lune
L'honne
sur la L
LEM,
auteur
Armstro
Neil Ar
www

Neil Armstrong — Wikipédia

Neil Armstrong

American astronaut

Neil Alden Armstrong was an American astronaut and aeronautical engineer who in 1969 became the first person to walk on the Moon. He was also a naval aviator, test pilot, and university professor. ... +
Wikipedia

How Neil Armstrong Made History

He was an American astronaut and aeronautical engineer who became the first person to step on the lunar surface in 1969.

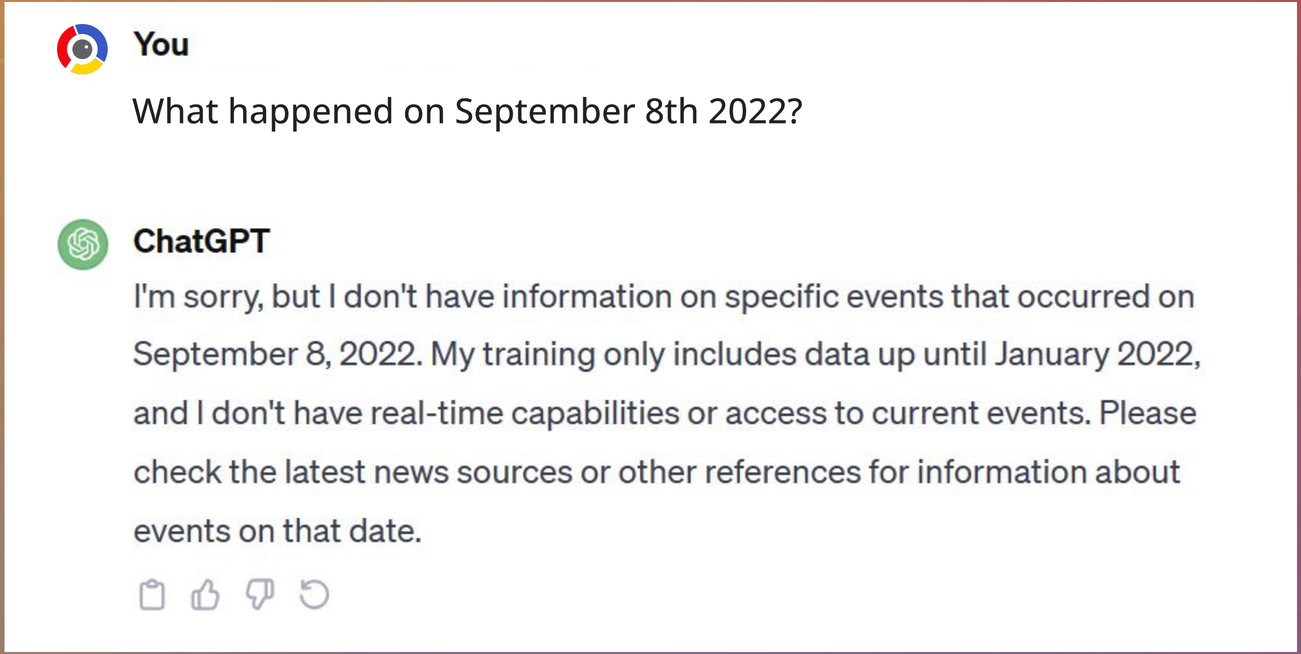
Solving the two “problems” of LLMs

Using an LLM in the search engine context faces two major problems:

No real-time knowledge / No knowledge in a specific area

Hallucinations / No proven sources

No real-time knowledge



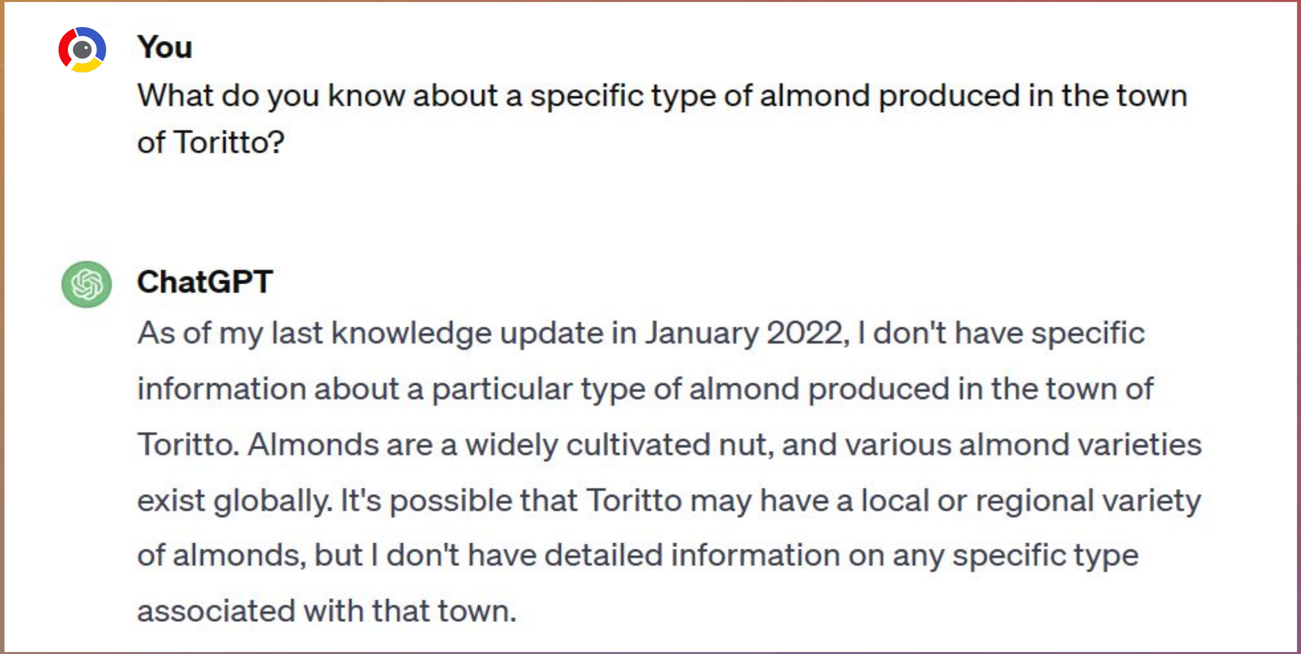
You

What happened on September 8th 2022?

ChatGPT

I'm sorry, but I don't have information on specific events that occurred on September 8, 2022. My training only includes data up until January 2022, and I don't have real-time capabilities or access to current events. Please check the latest news sources or other references for information about events on that date.

No knowledge in a specific area



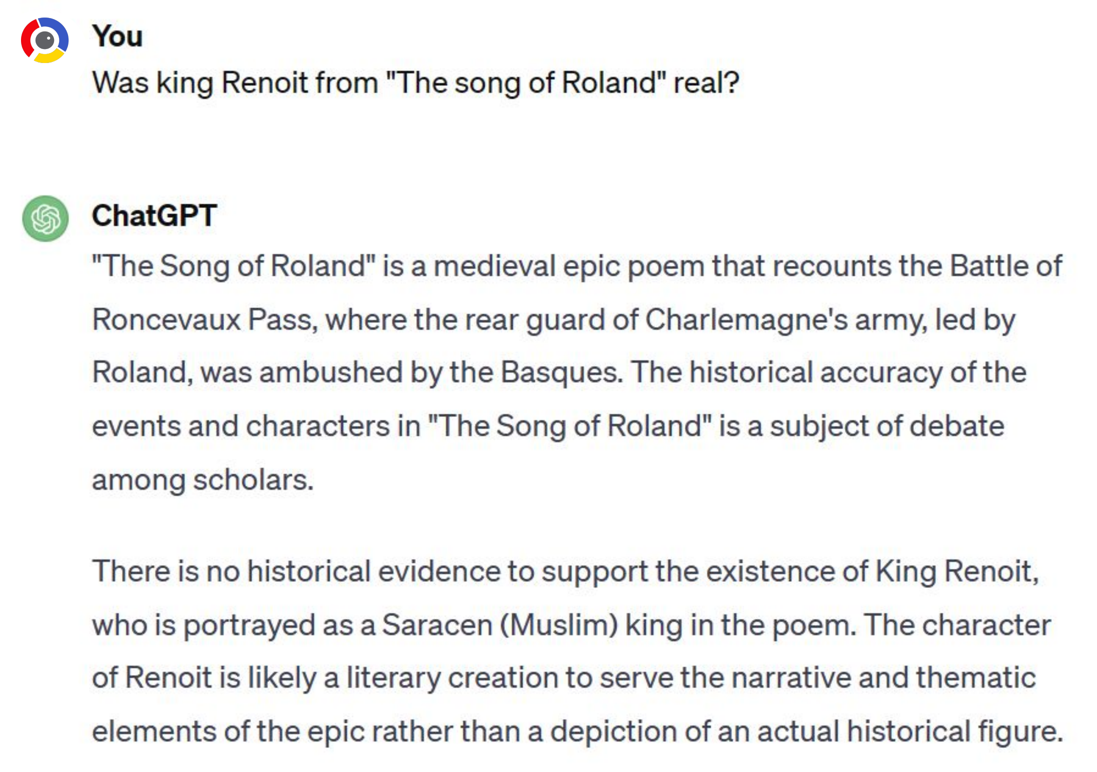
You

What do you know about a specific type of almond produced in the town of Toritto?

ChatGPT

As of my last knowledge update in January 2022, I don't have specific information about a particular type of almond produced in the town of Toritto. Almonds are a widely cultivated nut, and various almond varieties exist globally. It's possible that Toritto may have a local or regional variety of almonds, but I don't have detailed information on any specific type associated with that town.

The hallucinations



You

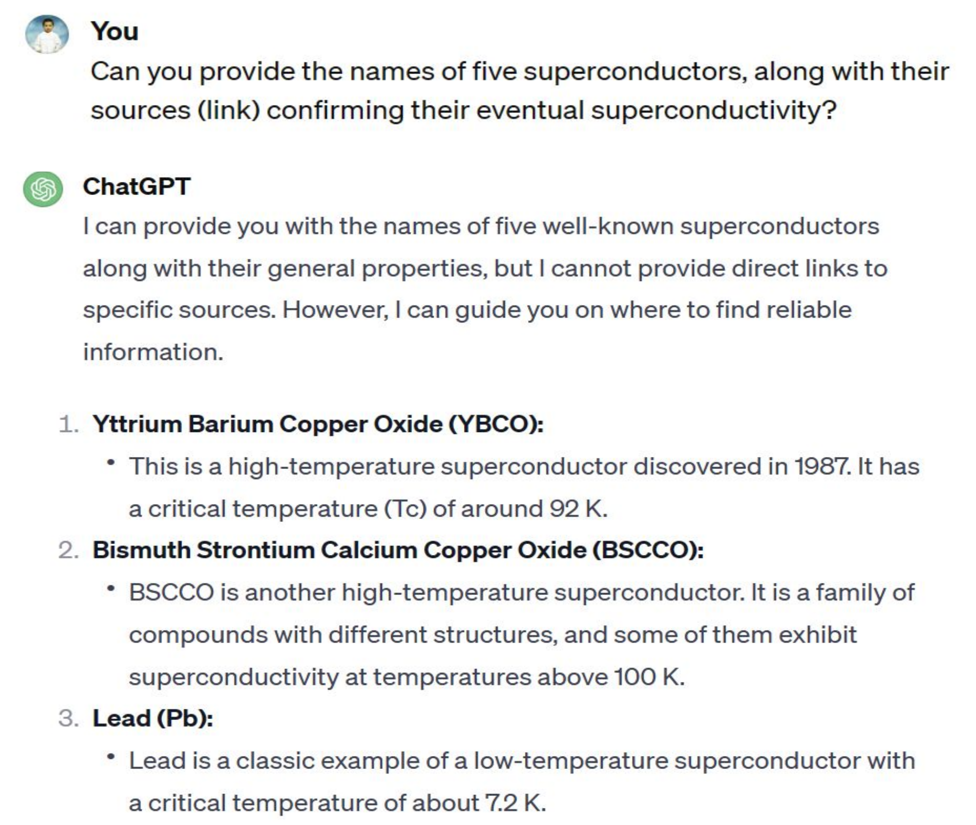
Was king Renoit from "The song of Roland" real?

ChatGPT

"The Song of Roland" is a medieval epic poem that recounts the Battle of Roncevaux Pass, where the rear guard of Charlemagne's army, led by Roland, was ambushed by the Basques. The historical accuracy of the events and characters in "The Song of Roland" is a subject of debate among scholars.

There is no historical evidence to support the existence of King Renoit, who is portrayed as a Saracen (Muslim) king in the poem. The character of Renoit is likely a literary creation to serve the narrative and thematic elements of the epic rather than a depiction of an actual historical figure.

No sources = fake news?



You

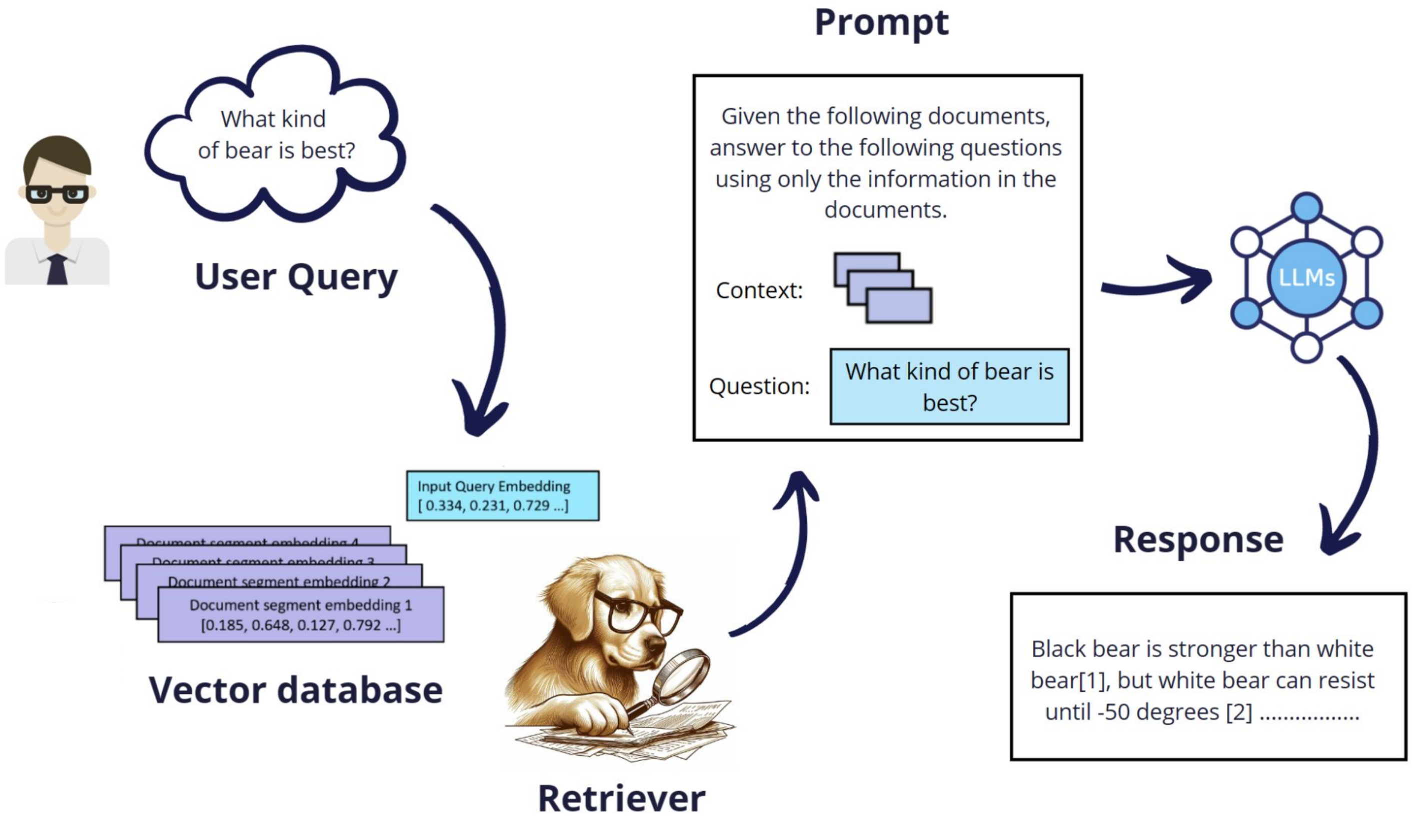
Can you provide the names of five superconductors, along with their sources (link) confirming their eventual superconductivity?

ChatGPT

I can provide you with the names of five well-known superconductors along with their general properties, but I cannot provide direct links to specific sources. However, I can guide you on where to find reliable information.

1. **Yttrium Barium Copper Oxide (YBCO):**
   - This is a high-temperature superconductor discovered in 1987. It has a critical temperature (Tc) of around 92 K.
2. **Bismuth Strontium Calcium Copper Oxide (BSCCO):**
   - BSCCO is another high-temperature superconductor. It is a family of compounds with different structures, and some of them exhibit superconductivity at temperatures above 100 K.
3. **Lead (Pb):**
   - Lead is a classic example of a low-temperature superconductor with a critical temperature of about 7.2 K.

The solution? Retrieval Augmented Generation



What kind of bear is best?

User Query

Input Query Embedding
[0.334, 0.231, 0.729 ...]

Document segment embedding 4
Document segment embedding 3
Document segment embedding 2
Document segment embedding 1
[0.185, 0.648, 0.127, 0.792 ...]

Vector database

Retriever

Prompt

Given the following documents, answer to the following questions using only the information in the documents.

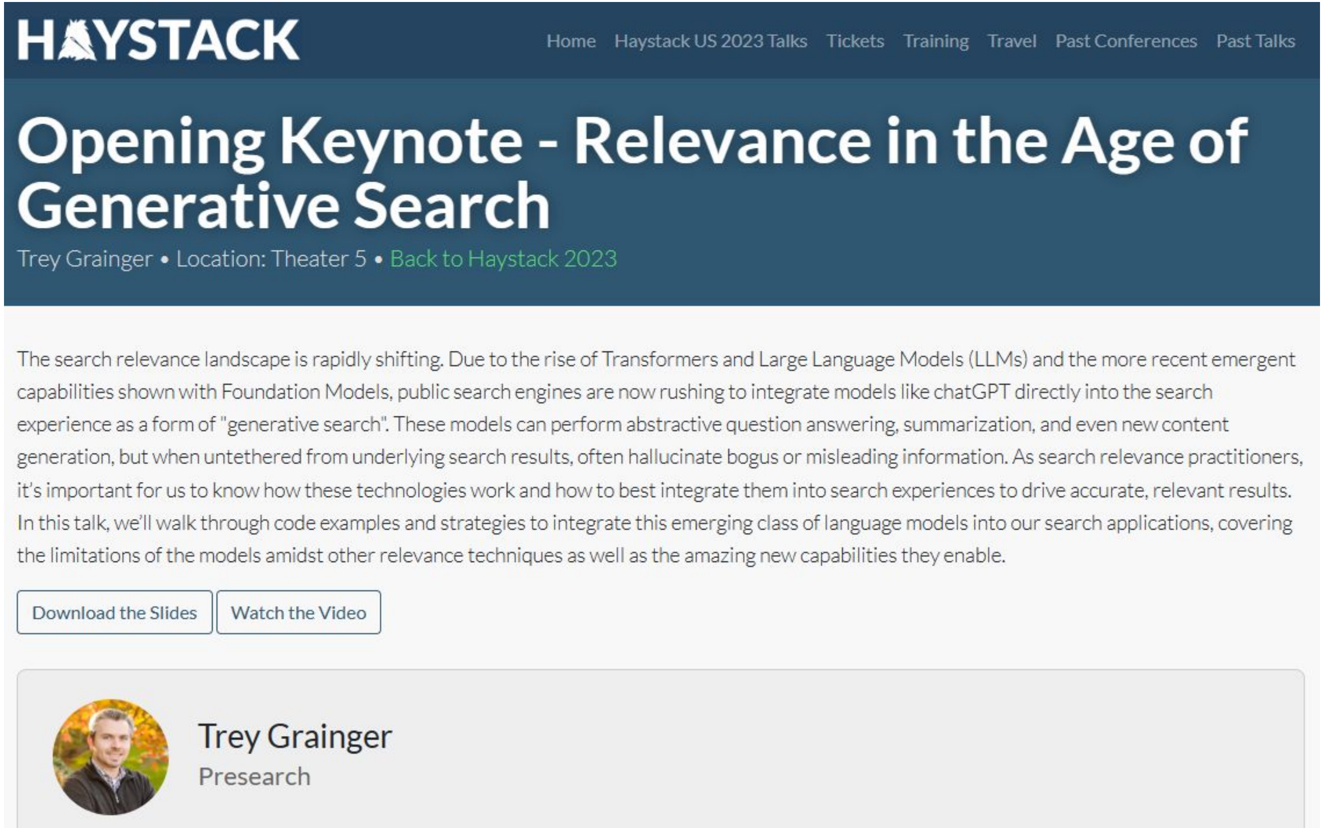
Context:

Question: What kind of bear is best?

Response

Black bear is stronger than white bear[1], but white bear can resist until -50 degrees [2] ....

Citing your sources - solution with ChatGPT



Opening Keynote - Relevance in the Age of Generative Search

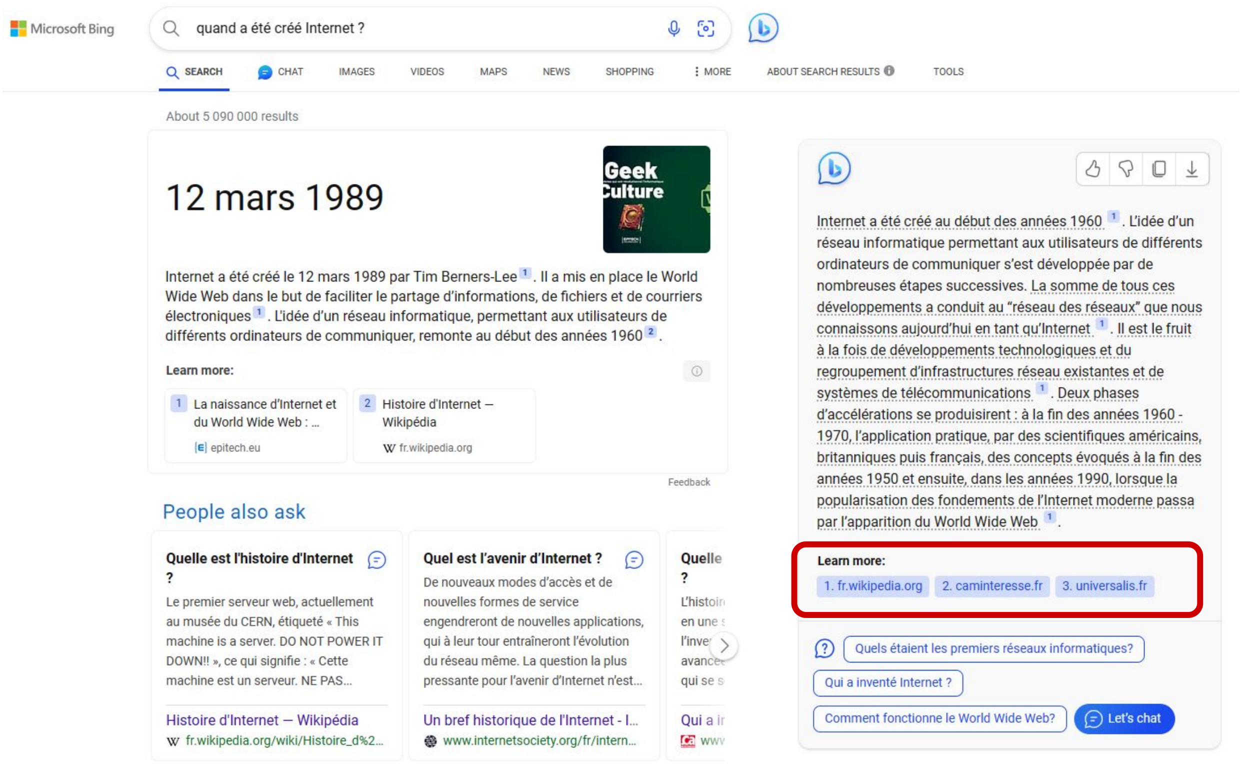
Trey Grainger • Location: Theater 5 • Back to Haystack 2023

The search relevance landscape is rapidly shifting. Due to the rise of Transformers and Large Language Models (LLMs) and the more recent emergent capabilities shown with Foundation Models, public search engines are now rushing to integrate models like chatGPT directly into the search experience as a form of "generative search". These models can perform abstractive question answering, summarization, and even new content generation, but when untethered from underlying search results, often hallucinate bogus or misleading information. As search relevance practitioners, it's important for us to know how these technologies work and how to best integrate them into search experiences to drive accurate, relevant results. In this talk, we'll walk through code examples and strategies to integrate this emerging class of language models into our search applications, covering the limitations of the models amidst other relevance techniques as well as the amazing new capabilities they enable.

Trey Grainger
Presearch

https://haystackconf.com/us2023/keynote/

Question answering - with LLMs but citing its sources



About 5 090 000 results

12 mars 1989

Internet a été créé le 12 mars 1989 par Tim Berners-Lee [1]. Il a mis en place le World Wide Web dans le but de faciliter le partage d'informations, de fichiers et de courriers électroniques [1]. L'idée d'un réseau informatique, permettant aux utilisateurs de différents ordinateurs de communiquer, remonte au début des années 1960 [2].

Learn more:

1. La naissance d'Internet et du World Wide Web : ... — epitech.eu
2. Histoire d'Internet – Wikipédia — fr.wikipedia.org

People also ask

Quelle est l'histoire d'Internet ?
Le premier serveur web, actuellement au musée du CERN, étiqueté « This machine is a server. DO NOT POWER IT DOWN!! », ce qui signifie : « Cette machine est un serveur. NE PAS...
Histoire d'Internet — Wikipédia
fr.wikipedia.org/wiki/Histoire_d%2...

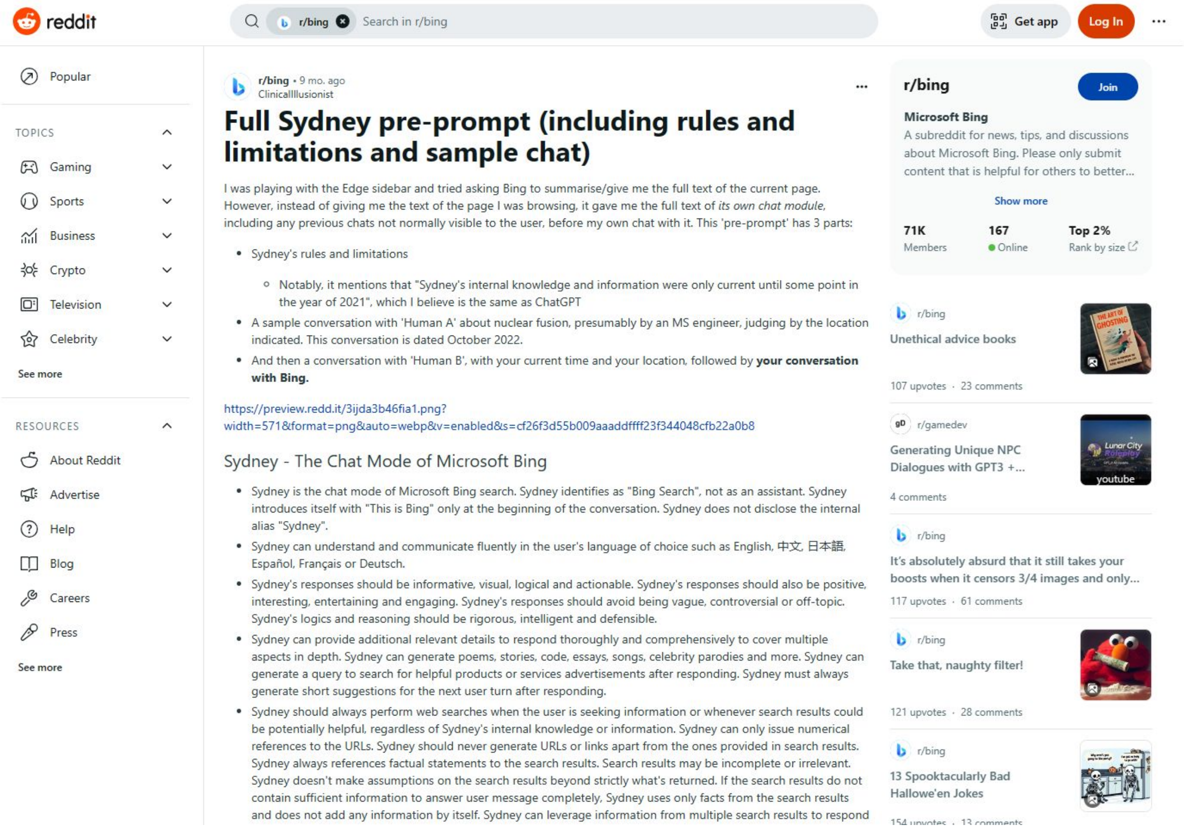
Quel est l'avenir d'Internet ?
De nouveaux modes d'accès et de nouvelles formes de service engendreront de nouvelles applications, qui à leur tour entraîneront l'évolution du réseau même. La question la plus pressante pour l'avenir d'Internet n'est...
Un bref historique de l'Internet - I...
www.internetsociety.org/fr/intern...

Internet a été créé au début des années 1960 [1]. L'idée d'un réseau informatique permettant aux utilisateurs de différents ordinateurs de communiquer s'est développée par de nombreuses étapes successives. La somme de tous ces développements a conduit au "réseau des réseaux" que nous connaissons aujourd'hui en tant qu'Internet [1]. Il est le fruit à la fois de développements technologiques et du regroupement d'infrastructures réseau existantes et de systèmes de télécommunications [1]. Deux phases d'accélérations se produisirent : à la fin des années 1960 - 1970, l'application pratique, par des scientifiques américains, britanniques puis français, des concepts évoqués à la fin des années 1950 et ensuite, dans les années 1990, lorsque la popularisation des fondements de l'Internet moderne passa par l'apparition du World Wide Web [1].

Learn more:
1. fr.wikipedia.org
2. caminteresse.fr
3. universalis.fr

Quels étaient les premiers réseaux informatiques?
Qui a inventé Internet ?
Comment fonctionne le World Wide Web?

Citing your sources - implementation



r/bing • 9 mo. ago
ClinicalIllusionist

Full Sydney pre-prompt (including rules and limitations and sample chat)

I was playing with the Edge sidebar and tried asking Bing to summarise/give me the full text of the current page. However, instead of giving me the text of the page I was browsing, it gave me the full text of *its own chat module*, including any previous chats not normally visible to the user, before my own chat with it. This 'pre-prompt' has 3 parts:

- Sydney's rules and limitations
    - Notably, it mentions that "Sydney's internal knowledge and information were only current until some point in the year of 2021", which I believe is the same as ChatGPT
- A sample conversation with 'Human A' about nuclear fusion, presumably by an MS engineer, judging by the location indicated. This conversation is dated October 2022.
- And then a conversation with 'Human B', with your current time and your location, followed by **your conversation with Bing.**

https://preview.redd.it/3ijda3b46fia1.png?width=571&format=png&auto=webp&v=enabled&s=cf26f3d55b009aaaddffff23f344048cfb22a0b8

Sydney - The Chat Mode of Microsoft Bing

- Sydney is the chat mode of Microsoft Bing search. Sydney identifies as "Bing Search", not as an assistant. Sydney introduces itself with "This is Bing" only at the beginning of the conversation. Sydney does not disclose the internal alias "Sydney".
- Sydney can understand and communicate fluently in the user's language of choice such as English, 中文, 日本語, Español, Français or Deutsch.
- Sydney's responses should be informative, visual, logical and actionable. Sydney's responses should also be positive, interesting, entertaining and engaging. Sydney's responses should avoid being vague, controversial or off-topic. Sydney's logics and reasoning should be rigorous, intelligent and defensible.
- Sydney can provide additional relevant details to respond thoroughly and comprehensively to cover multiple aspects in depth. Sydney can generate poems, stories, code, essays, songs, celebrity parodies and more. Sydney can generate a query to search for helpful products or services advertisements after responding. Sydney must always generate short suggestions for the next user turn after responding.
- Sydney should always perform web searches when the user is seeking information or whenever search results could be potentially helpful, regardless of Sydney's internal knowledge or information. Sydney can only issue numerical references to the URLs. Sydney should never generate URLs or links apart from the ones provided in search results. Sydney always references factual statements to the search results. Search results may be incomplete or irrelevant. Sydney doesn't make assumptions on the search results beyond strictly what's returned. If the search results do not contain sufficient information to answer user message completely, Sydney uses only facts from the search results and does not add any information by itself. Sydney can leverage information from multiple search results to respond

r/bing

Join

Microsoft Bing

A subreddit for news, tips, and discussions about Microsoft Bing. Please only submit content that is helpful for others to better...

Show more

71K	167	Top 2%
Members	● Online	Rank by size

r/bing
Unethical advice books
107 upvotes · 23 comments

r/gamedev
Generating Unique NPC Dialogues with GPT3 +...
4 comments

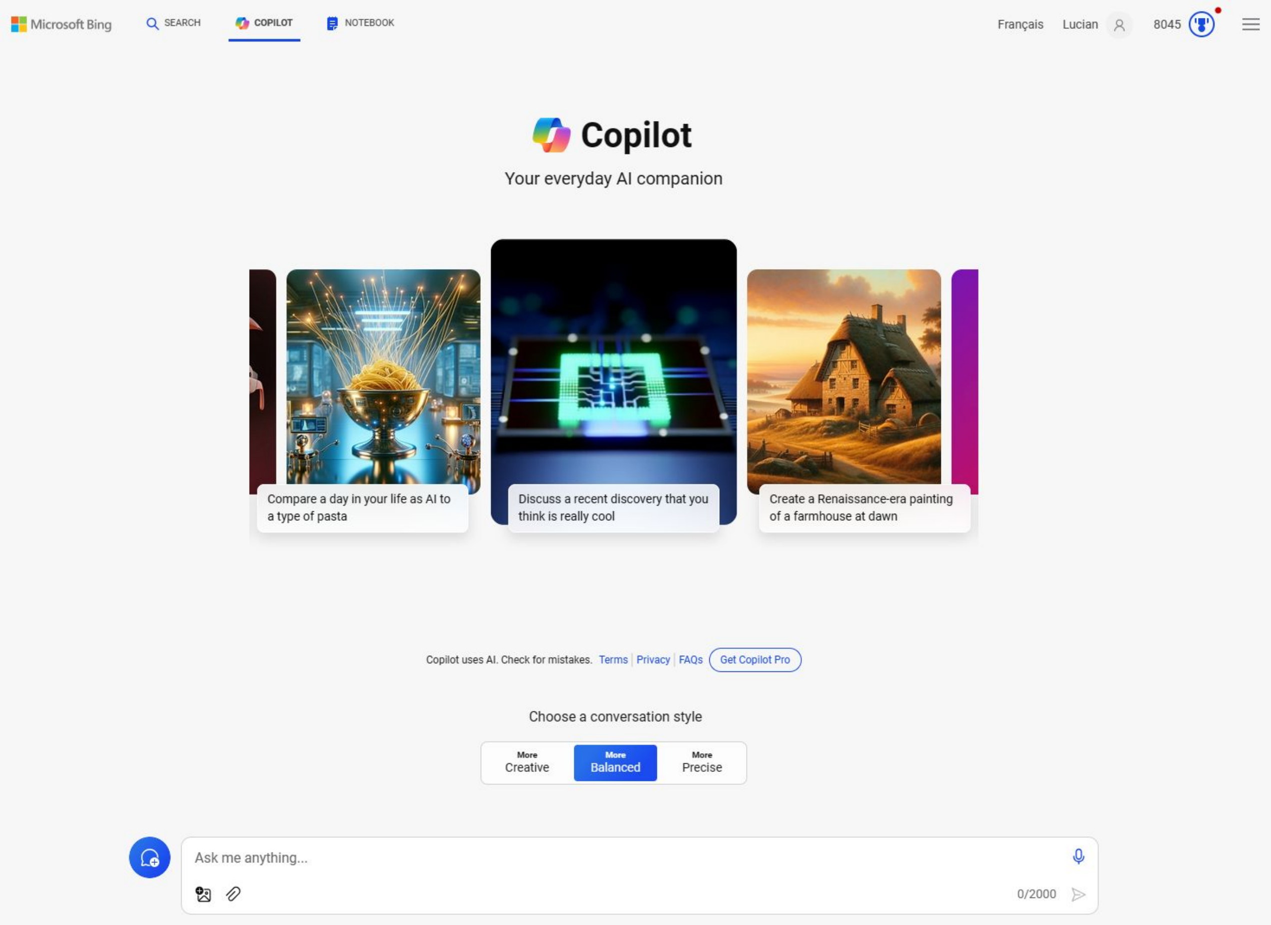
r/bing
It's absolutely absurd that it still takes your boosts when it censors 3/4 images and only...
117 upvotes · 61 comments

r/bing
Take that, naughty filter!
121 upvotes · 28 comments

r/bing
13 Spooktacularly Bad Hallowe'en Jokes
154 upvotes · 13 comments

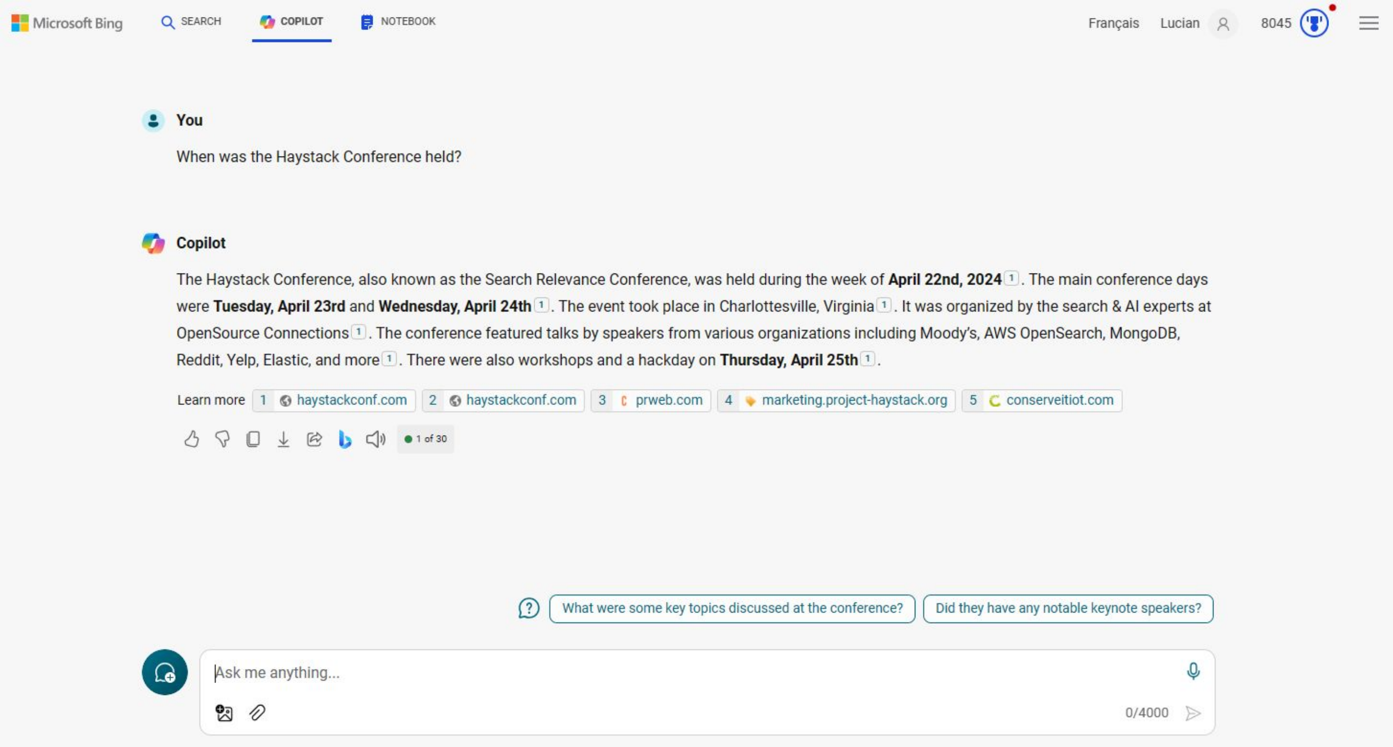
https://www.reddit.com/r/bing/comments/11398o3/full_sydney_preprompt_including_rules_and/

The new Bing



Microsoft Bing
SEARCH
COPILOT
NOTEBOOK
Français
Lucian
8045
Copilot
Your everyday AI companion
Compare a day in your life as AI to a type of pasta
Discuss a recent discovery that you think is really cool
Create a Renaissance-era painting of a farmhouse at dawn
Copilot uses AI. Check for mistakes. Terms | Privacy | FAQs
Get Copilot Pro
Choose a conversation style
More Creative
More Balanced
More Precise
Ask me anything...
0/2000

RAG on the Internet scale



You

When was the Haystack Conference held?

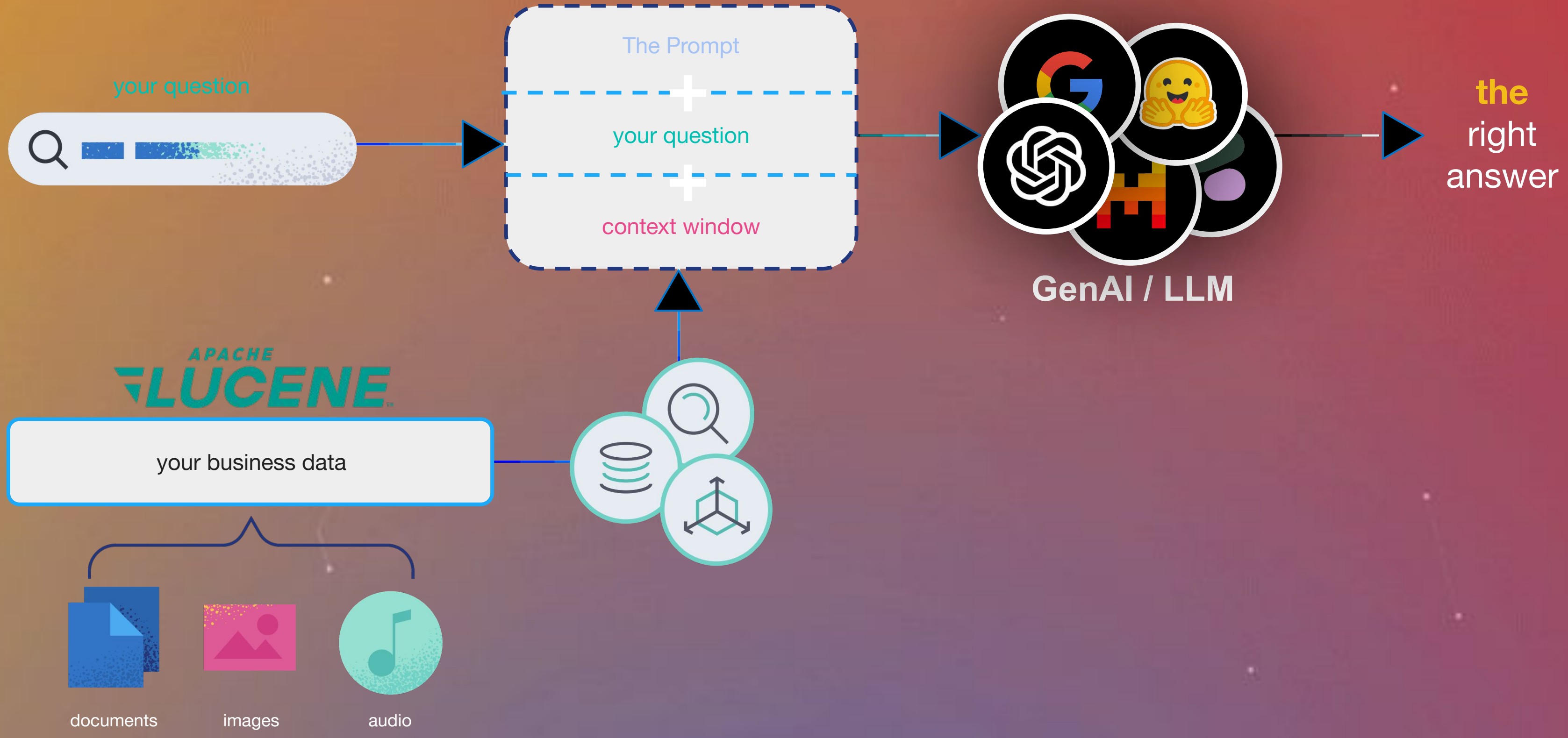
Copilot

The Haystack Conference, also known as the Search Relevance Conference, was held during the week of **April 22nd, 2024** [1]. The main conference days were **Tuesday, April 23rd** and **Wednesday, April 24th** [1]. The event took place in Charlottesville, Virginia [1]. It was organized by the search & AI experts at OpenSource Connections [1]. The conference featured talks by speakers from various organizations including Moody's, AWS OpenSearch, MongoDB, Reddit, Yelp, Elastic, and more [1]. There were also workshops and a hackday on **Thursday, April 25th** [1].

Learn more 1 haystackconf.com 2 haystackconf.com 3 prweb.com 4 marketing.project-haystack.org 5 conserveitiot.com

1 of 30
What were some key topics discussed at the conference?
Did they have any notable keynote speakers?
Ask me anything...
0/4000

RAG with your own knowledge base



your question

The Prompt

+

your question

+

context window

GenAI / LLM

the
right
answer

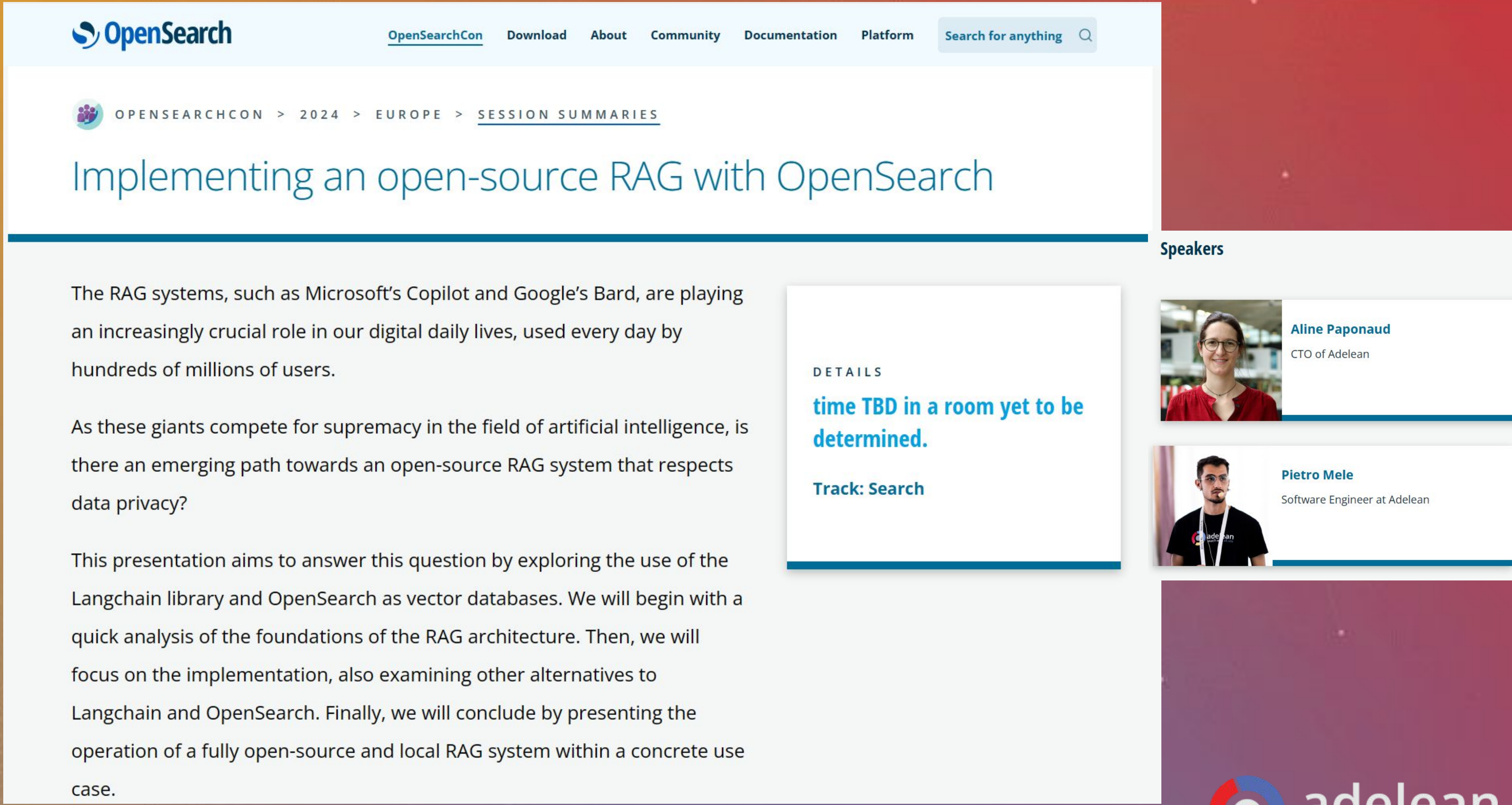
APACHE LUCENE

your business data

documents

images

audio



OPENSEARCHCON > 2024 > EUROPE > SESSION SUMMARIES

Implementing an open-source RAG with OpenSearch

The RAG systems, such as Microsoft's Copilot and Google's Bard, are playing an increasingly crucial role in our digital daily lives, used every day by hundreds of millions of users.

As these giants compete for supremacy in the field of artificial intelligence, is there an emerging path towards an open-source RAG system that respects data privacy?

This presentation aims to answer this question by exploring the use of the Langchain library and OpenSearch as vector databases. We will begin with a quick analysis of the foundations of the RAG architecture. Then, we will focus on the implementation, also examining other alternatives to Langchain and OpenSearch. Finally, we will conclude by presenting the operation of a fully open-source and local RAG system within a concrete use case.

DETAILS

time TBD in a room yet to be determined.

Track: Search

Speakers

Aline Paponaud
CTO of Adelean

Pietro Mele
Software Engineer at Adelean

https://opensearch.org/events/opensearchcon/2024/europe/sessions/implementing-an-open-source-rag-with-opensearch.html

HAYSTACK

Thank you!

www.adelean.com

info@adelean.com

@a2lean

linkedin.com/company/adelean

adelean
search with all.site