IN[34]120 Søketeknologi

2024-10-01 14:15 @ Chill 🎃🎃

Agenda:
- Introdusere oblig C
- Case study av søk i praksis
- Oblighjelp

Oblig A: Fasit ute

- Kom ut natt til mandag😎👍
- Ligger på Github
- Pull filene!!
- Kan få følgefeil i obligene

Oblig B

- Hadde frist på fredag
- Vanskelig👏
- Blir rettet fortløpende

TIDEN FLYR 🕌

Oblig 3/5: C

Frist: 2024-10-11

Assignment C-1

Deadline: 2024-10-11

The purpose of this assignment is to implement a simple query evaluator that efficiently performs *n*-of-*m* matching over a simple memory-based inverted index. I.e., if the query contains *m* unique query terms, each document in the result set should contain at least *n* of these *m* terms. For example, 2-of-3 matching over the query *orange apple banana* would be logically equivalent to the following predicate:

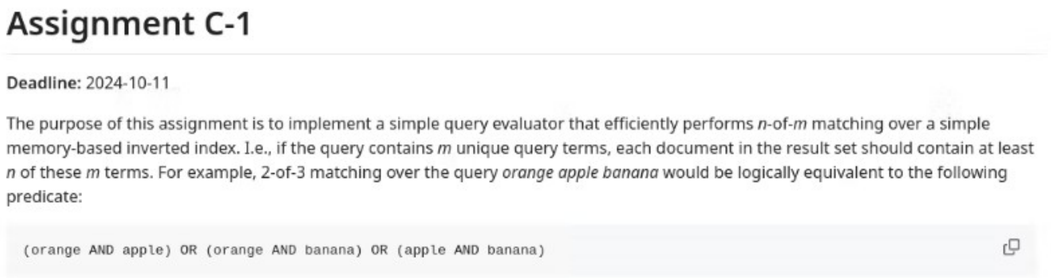
```
(orange AND apple) OR (orange AND banana) OR (apple AND banana)
```

Oblig C-1

- → Søkemotor, n / m terms
- → Litt som postingsmerger i A

Oblig C-1: Tips

- → WAND, "weak and" -> paper
- → NB: DAAT, ikke TAAT
- → Lecture 7, slide #5, #13



Assignment C-1

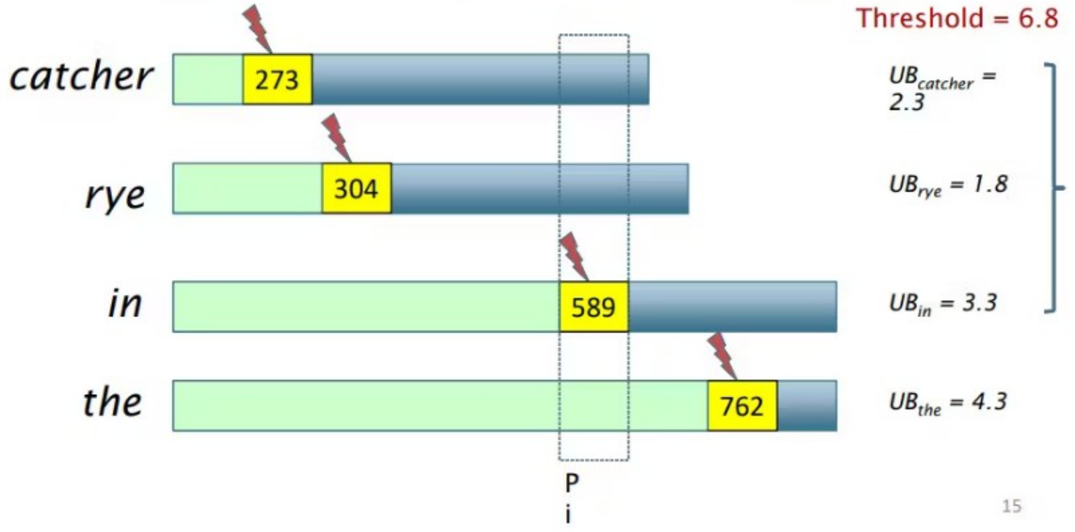
Deadline: 2024-10-11

The purpose of this assignment is to implement a simple query evaluator that efficiently performs *n*-of-*m* matching over a simple memory-based inverted index. I.e., if the query contains *m* unique query terms, each document in the result set should contain at least *n* of these *m* terms. For example, 2-of-3 matching over the query *orange apple banana* would be logically equivalent to the following predicate:

```
(orange AND apple) OR (orange AND banana) OR (apple AND banana)
```

Pivoting

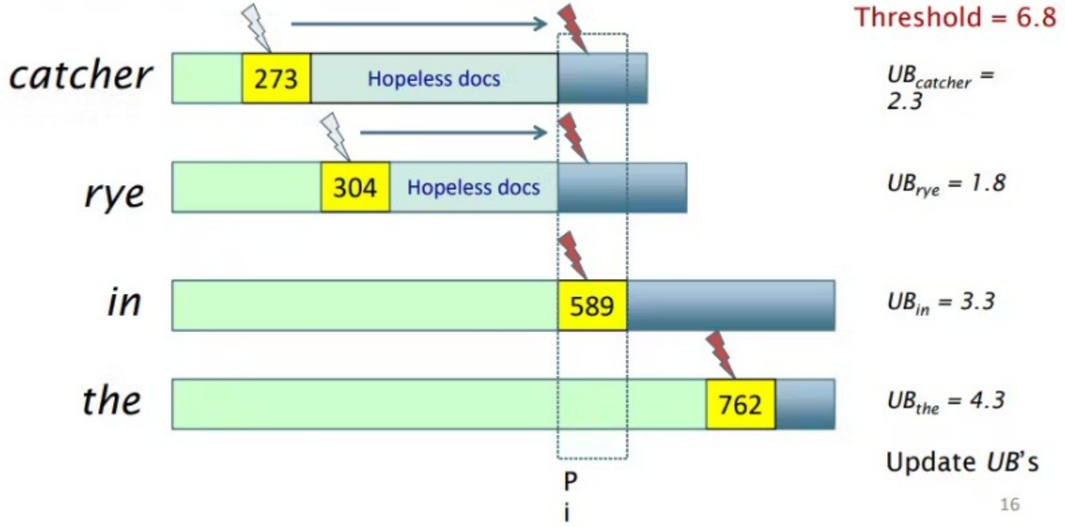
- Query: *catcher in the rye*
- Let's say the current finger positions are as below



C-1: Konseptet "fingers"

Prune docs that have no hope

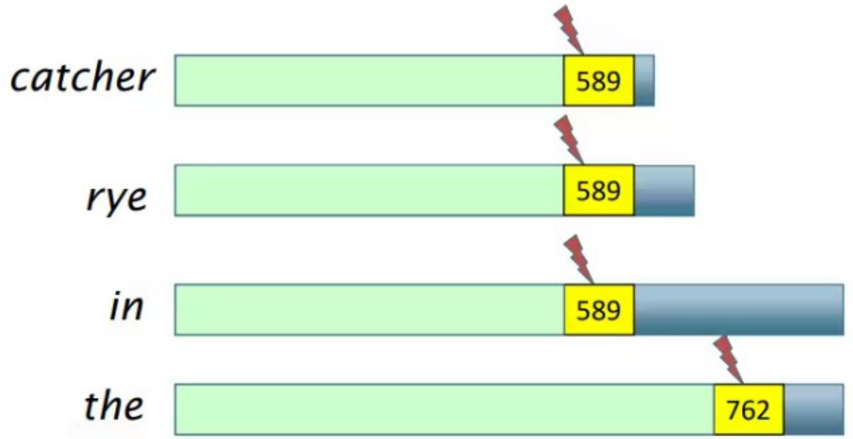
- Terms sorted in order of finger positions
- Move fingers to 589 or right



C-1: Algoritme for å gå videre (1/2)

Compute 589's score if need be

- If 589 is present in enough postings, compute its full cosine score – else some fingers to right of 589
- Pivot again …



C-1: Algoritme for å gå videre (2/2)

Oblig C-2

- Søkemotor, legge til nye operators
- Prekode: motor med AND, OR, ANDNOT
- Legge til WILDCARD, LOOKSLIKE, SOUNDSLIKE, SYNONYM

Assignment C-2

Deadline: 2024-10-11

The purpose of this assignment is to extend the `BooleanSearchEngine` class with a handful of new linguistic operators. The existing implementation already provides support for the logical operators `AND`, `OR`, and `ANDNOT`. Your task is to add support for the following new operators:

- `WILDCARD`: Support for simple wildcard matching. E.g., the query expression `WILDCARD("fi*er")` would match documents that contain any of the indexed terms {*fishmonger*, *filibuster*, *fisher*, *finder*, ...} that start with *fi* and end with *er*. The types of wildcards supported should be the same as what you can do with a permuterm index.
- `LOOKSLIKE`: Support for simple approximate matching based on edit distance. E.g., the query expression `LOOKSLIKE("adverb")` would match documents that contain any of the indexed terms {*adverb*, *advert*, *avderb*, *advverb*, ...} that are a single edit away from the given term. Edit distance should be measured according to the Damerau-Levenshtein metric.
- `SOUNDSLIKE`: Support for simple phonetic matching. E.g., the query expression `SOUNDSLIKE("smith")` would match documents that contain any of the indexed terms {*smith*, *smyth*, *smitt*, ...} that have similar pronounciations as the given term. Phonetic similarity should be assessed using the Soundex algorithm.
- `SYNONYM`: Support for simple equivalence-classing of terms using a provided dictionary. E.g., the query expression `SYNONYM("car")` would match documents that contain any of the indexed terms {*car*, *automobile*, ...}, assuming that this mapping is contained in the provided dictionary.

Assignment C-2

Deadline: 2024-10-11

The purpose of this assignment is to extend the `BooleanSearchEngine` class with a handful of new linguistic operators. The existing implementation already provides support for the logical operators `AND`, `OR`, and `ANDNOT`. Your task is to add support for the following new operators:

- `WILDCARD`: Support for simple wildcard matching. E.g., the query expression `WILDCARD("fi*er")` would match documents that contain any of the indexed terms {*fishmonger*, *filibuster*, *fisher*, *finder*, ...} that start with *fi* and end with *er*. The types of wildcards supported should be the same as what you can do with a permuterm index.
- `LOOKSLIKE`: Support for simple approximate matching based on edit distance. E.g., the query expression `LOOKSLIKE("adverb")` would match documents that contain any of the indexed terms {*adverb*, *advert*, *avderb*, *advverb*, ...} that are a single edit away from the given term. Edit distance should be measured according to the Damerau-Levenshtein metric.
- `SOUNDSLIKE`: Support for simple phonetic matching. E.g., the query expression `SOUNDSLIKE("smith")` would match documents that contain any of the indexed terms {*smith*, *smyth*, *smitt*, ...} that have similar pronounciations as the given term. Phonetic similarity should be assessed using the Soundex algorithm.
- `SYNONYM`: Support for simple equivalence-classing of terms using a provided dictionary. E.g., the query expression `SYNONYM("car")` would match documents that contain any of the indexed terms {*car*, *automobile*, ...}, assuming that this mapping is contained in the provided dictionary.

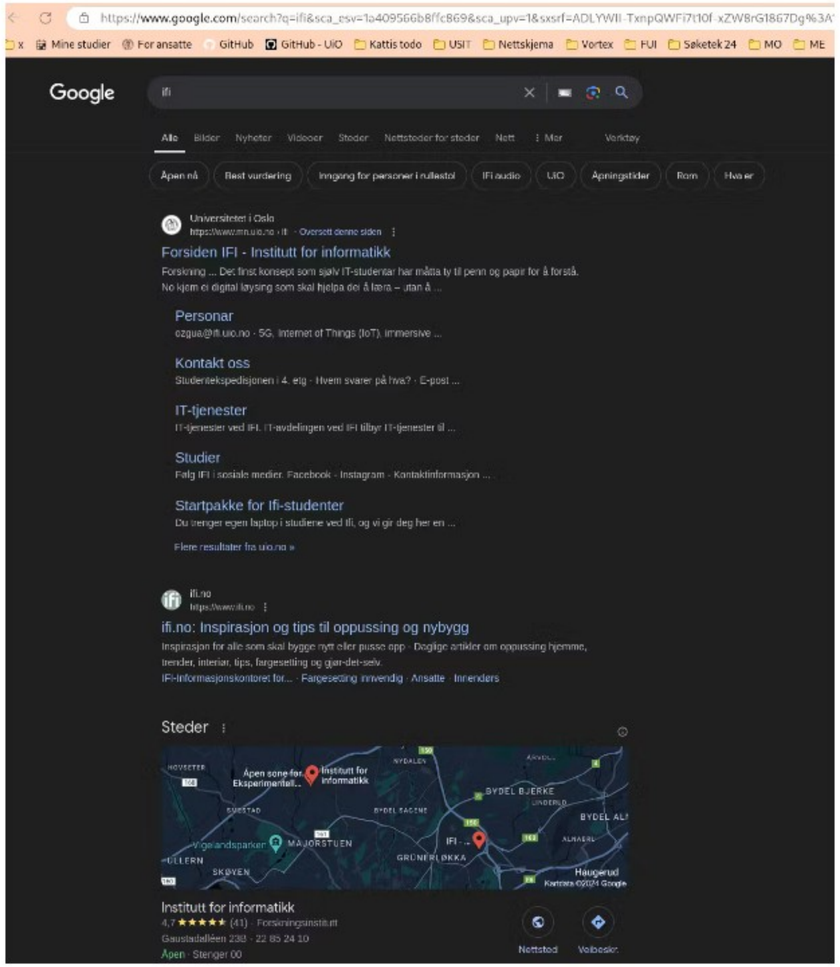
Oblig C-2: Tips

→ Bruker *din* kode fra A -> pull!

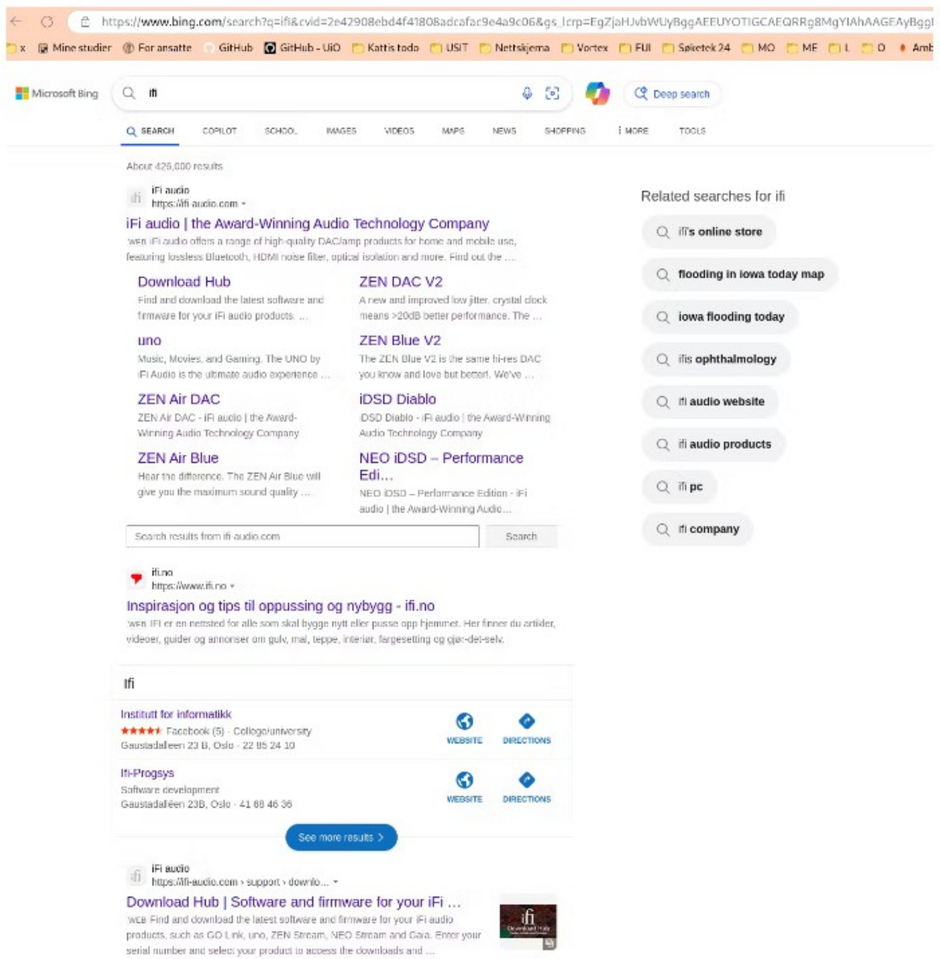
→ Ny oblig -> meld fra om feil

Case study av websøk

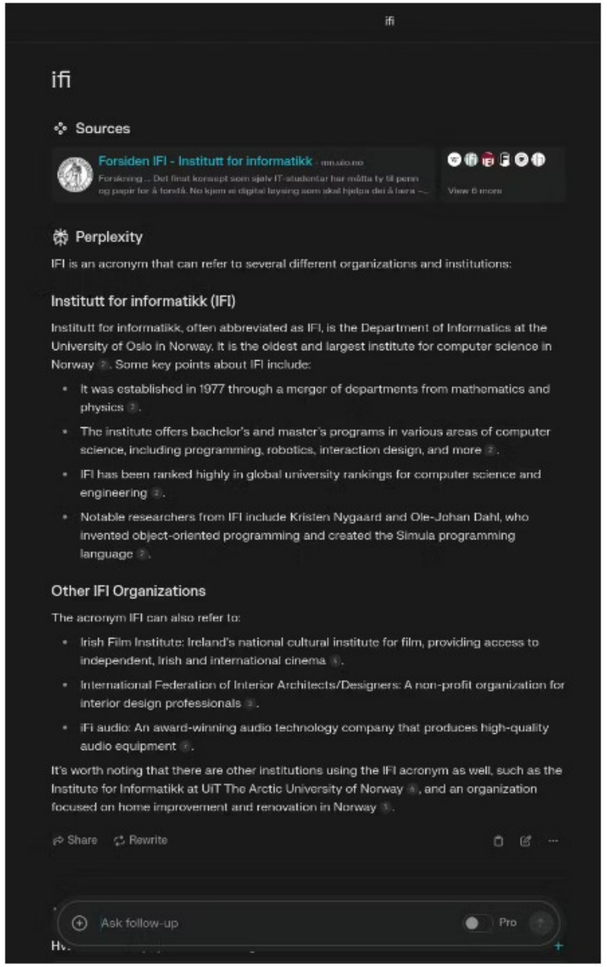
(Ikke pensum i seg selv, praktisk eksempel av hva vi har lært)


https://www.google.com/search?q=ifi&sca_esv=1a409566b8ffc869&sca_upv=1&sxsrf=ADLYWII-TxnpQWFi7t10f-xZW8rG1867Dg%3A
Mine studier
For ansatte
GitHub
GitHub - UiO
Kattis todo
USIT
Nettskjema
Vortex
FUI
Søketek 24
MO
ME
Google
ifi
Alle
Bilder
Nyheter
Videoer
Steder
Nettsteder for steder
Nett
Mer
Verktøy
Åpen nå
Best vurdering
Inngang for personer i rullestol
IFI audio
UiO
Åpningstider
Rom
Hva er
Universitetet i Oslo
https://www.mn.uio.no › it · Oversett denne siden
Forsiden IFI - Institutt for informatikk
Forskning ... Det finst konsept som sjølv IT-studentar har måtta ty til penn og papir for å forstå. No kjem ei digital løysing som skal hjelpa dei å lære – utan å ...
Personar
ozgua@ifi.uio.no · 5G, Internet of Things (IoT), immersive ...
Kontakt oss
Studentekspedisjonen i 4. etg · Hvem svarer på hva? · E-post ...
IT-tjenester
IT-tjenester ved IFI. IT-avdelingen ved IFI tilbyr IT-tjenester til ...
Studier
Følg IFI i sosiale medier: Facebook · Instagram · Kontaktinformasjon ...
Startpakke for Ifi-studenter
Du trenger egen laptop i studiene ved Ifi, og vi gir deg her en ...
Flere resultater fra uio.no »
ifi.no
https://www.ifi.no
ifi.no: Inspirasjon og tips til oppussing og nybygg
Inspirasjon for alle som skal bygge nytt eller pusse opp · Daglige artikler om oppussing hjemme, trender, interiør, tips, fargesetting og gjør-det-selv.
IFI-informasjonskontoret for... · Fargesetting innvendig · Ansatte · Innenders
Steder
Åpen sone for Eksperimentell...
Institutt for informatikk
NYDALEN
HOVSETER
SVESTAD
BYDEL BJERKE
BYDEL SAGENE
BYDEL ALN
Vigelandsparken
MAJORSTUEN
IFI - ...
GRØNLØKKA
ULLERN
SKØYEN
Haugerud
Kartdata ©2024 Google
Institutt for informatikk
4,7 ★★★★★ (41) · Forskningsinstitutt
Gaustadalléen 238 · 22 85 24 10
Åpen · Stenger 00
Nettsted
Veibeskr.

Google: Ifi

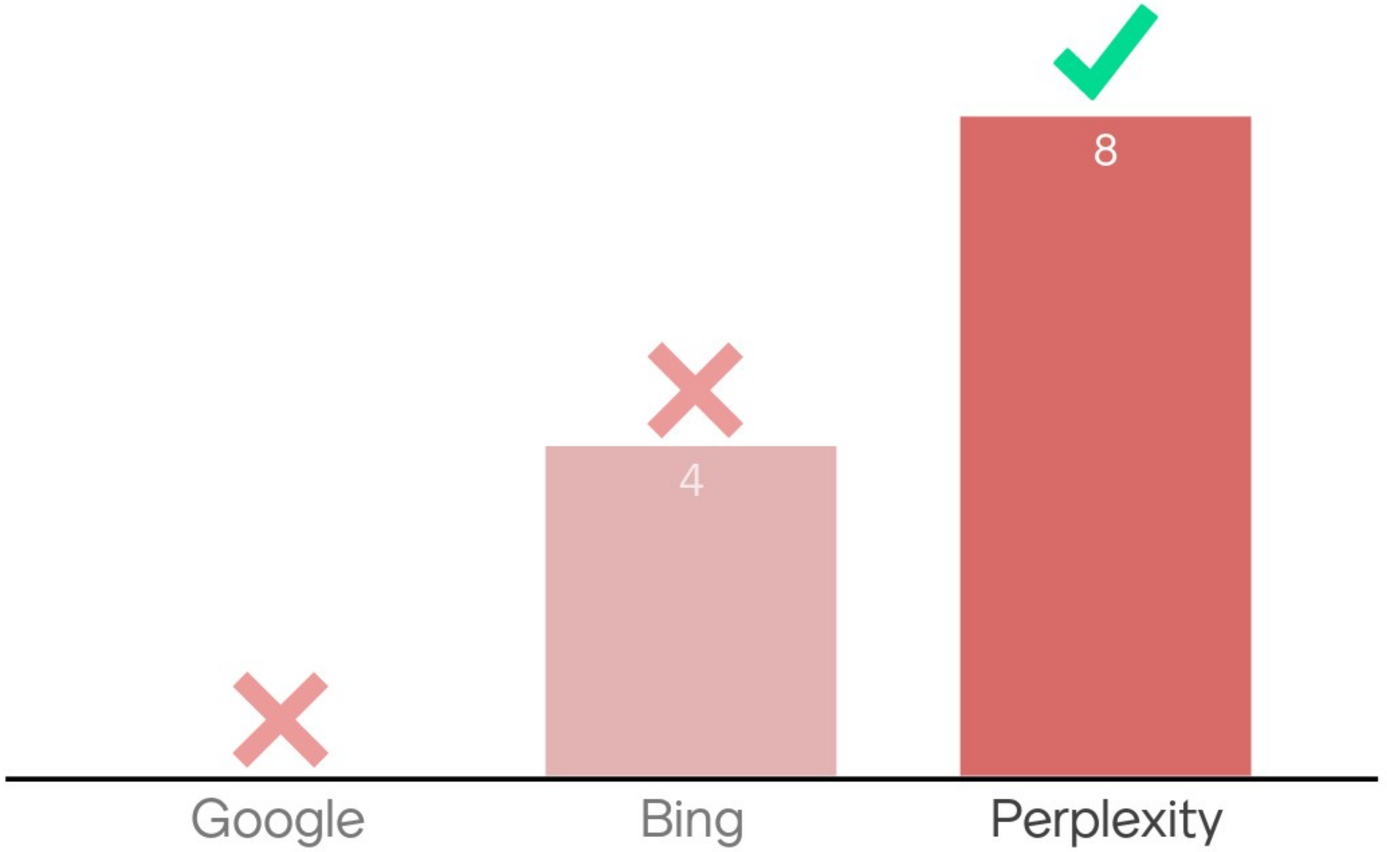

https://www.bing.com/search?q=ifi&cvid=2e42908ebd4f4180adcafac9e4a9c06&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIGCAEQRRg8MgYIAhAAGEAyBggl
Microsoft Bing
ifi
Deep search
SEARCH
COPILOT
SCHOOL
IMAGES
VIDEOS
MAPS
NEWS
SHOPPING
MORE
TOOLS
About 426,000 results
iFi audio
https://ifi-audio.com
iFi audio | the Award-Winning Audio Technology Company
Download Hub
Find and download the latest software and firmware for your iFi audio products. ...
ZEN DAC V2
A new and improved low jitter, crystal clock means >20dB better performance. The ...
uno
Music, Movies, and Gaming. The UNO by iFi Audio is the ultimate audio experience ...
ZEN Blue V2
The ZEN Blue V2 is the same hi-res DAC you know and love but better. We've ...
ZEN Air DAC
ZEN Air DAC - iFi audio | the Award-Winning Audio Technology Company
iDSD Diablo
iDSD Diablo - iFi audio | the Award-Winning Audio Technology Company
ZEN Air Blue
Hear the difference. The ZEN Air Blue will give you the maximum sound quality ...
NEO iDSD – Performance Edi...
NEO iDSD – Performance Edition - iFi audio | the Award-Winning Audio...
Search results from ifi-audio.com
Search
Related searches for ifi
ifi's online store
flooding in iowa today map
iowa flooding today
ifis ophthalmology
ifi audio website
ifi audio products
ifi pc
ifi company
ifi.no
https://www.ifi.no
Inspirasjon og tips til oppussing og nybygg - ifi.no
Ifi
Institutt for informatikk
Facebook (5) · College/university
Gaustadalleen 23 B, Oslo · 22 85 24 10
WEBSITE
DIRECTIONS
Ifi-Progsys
Software development
Gaustadalléen 23B, Oslo · 41 69 46 36
WEBSITE
DIRECTIONS
See more results
iFi audio
https://ifi-audio.com › support › downlo...
Download Hub | Software and firmware for your iFi ...

Bing: Ifi...?



Perplexity: Ifi. Nice.

Hvem skal ut?



Hvorfor skal Perplexity ut?

Fordi den er AI-basert

Fordi den "søker" på en helt annen måte

Fordi det er basert på LLMs. Mulig den bruker RAG?

perplexity er ikke en tradisjonel søkemotor, men en large language model

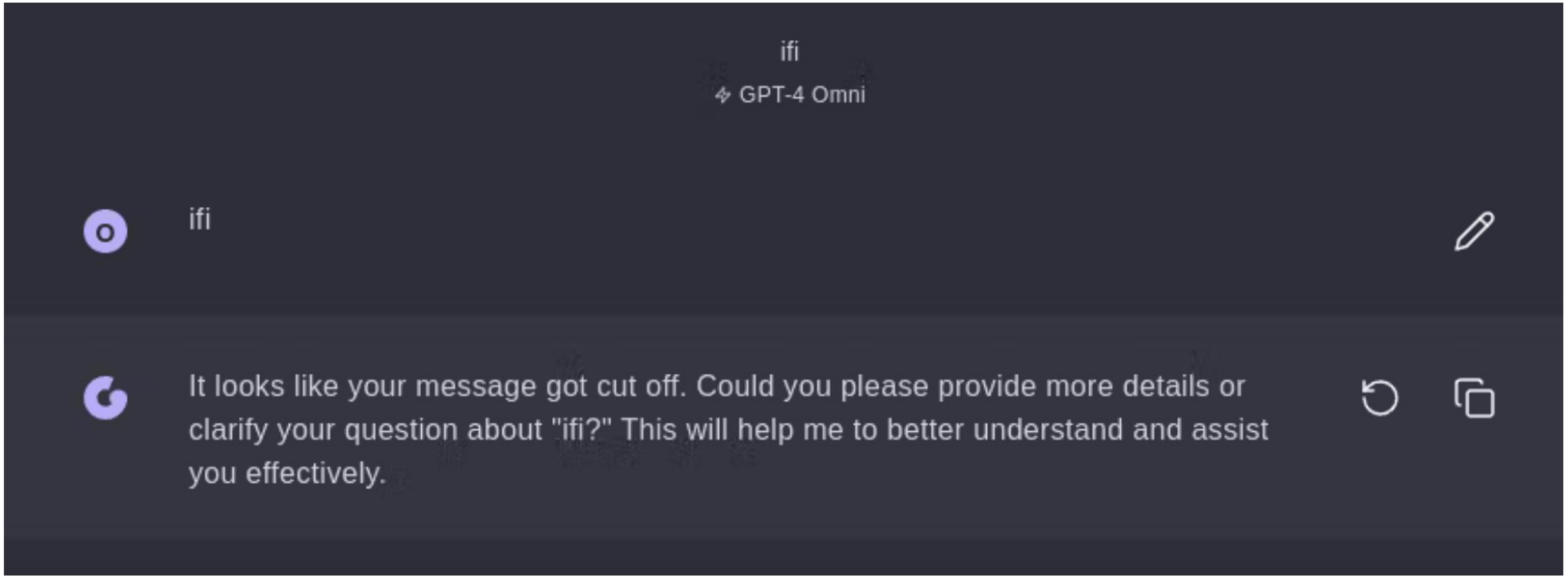
Baserer seg ikke på lenkene vi ser etter når vi bruker en søkemotor

Hvorfor bytte til edge???

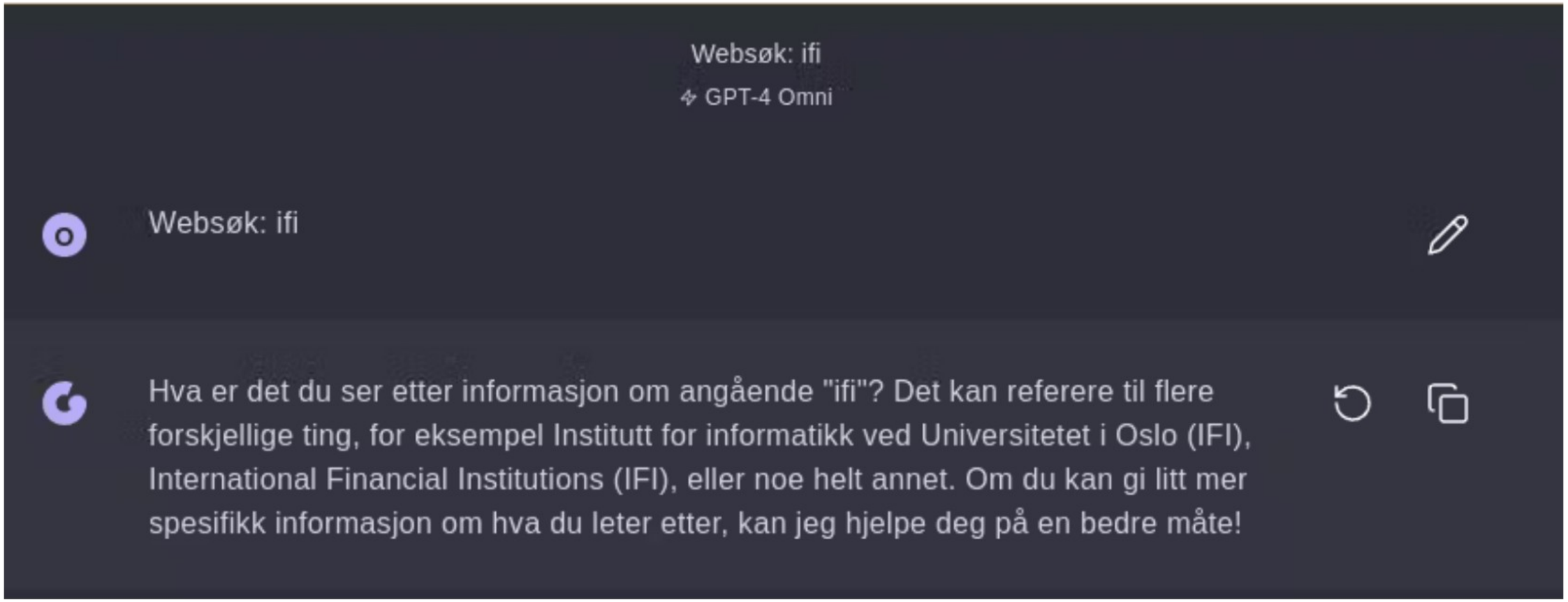
llm

Perplexity er et mål på en språkmodells prediksjonsevne

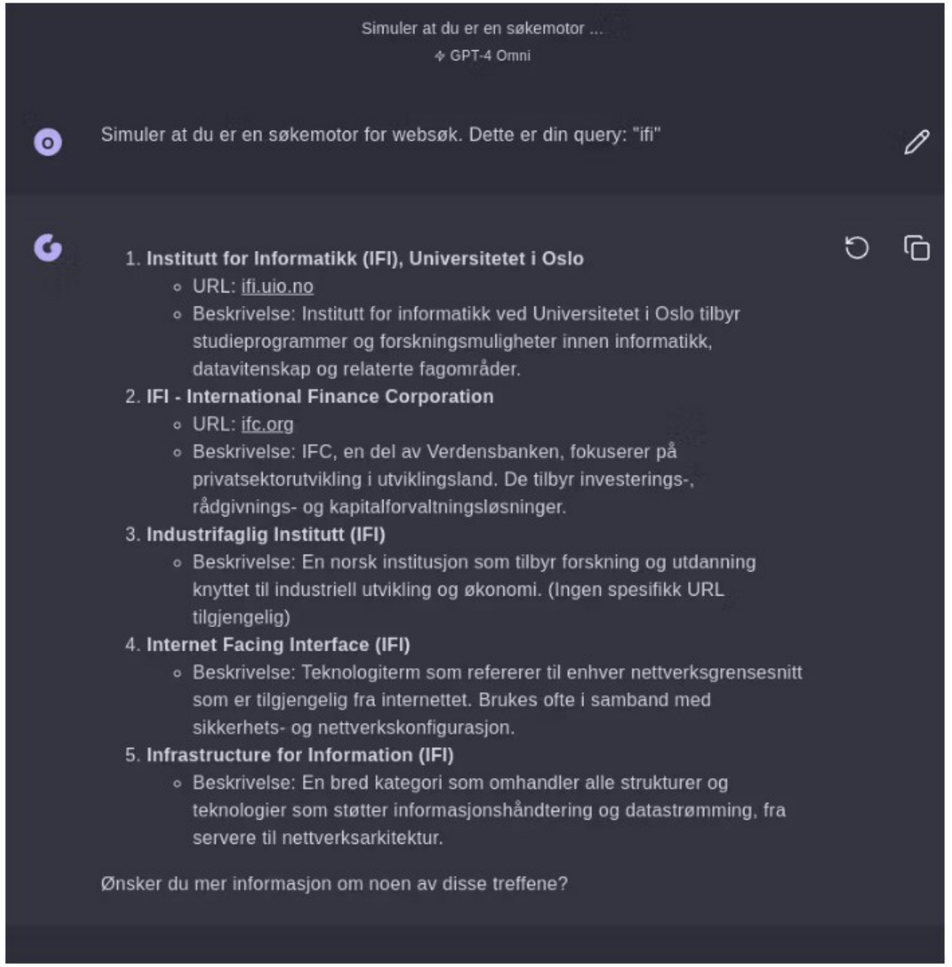

Friendship ended with
INVERTED
INDEX
Now
LLMER
is my
best friend
imgflip.com


ifi
GPT-4 Omni
ifi
It looks like your message got cut off. Could you please provide more details or clarify your question about "ifi?" This will help me to better understand and assist you effectively.

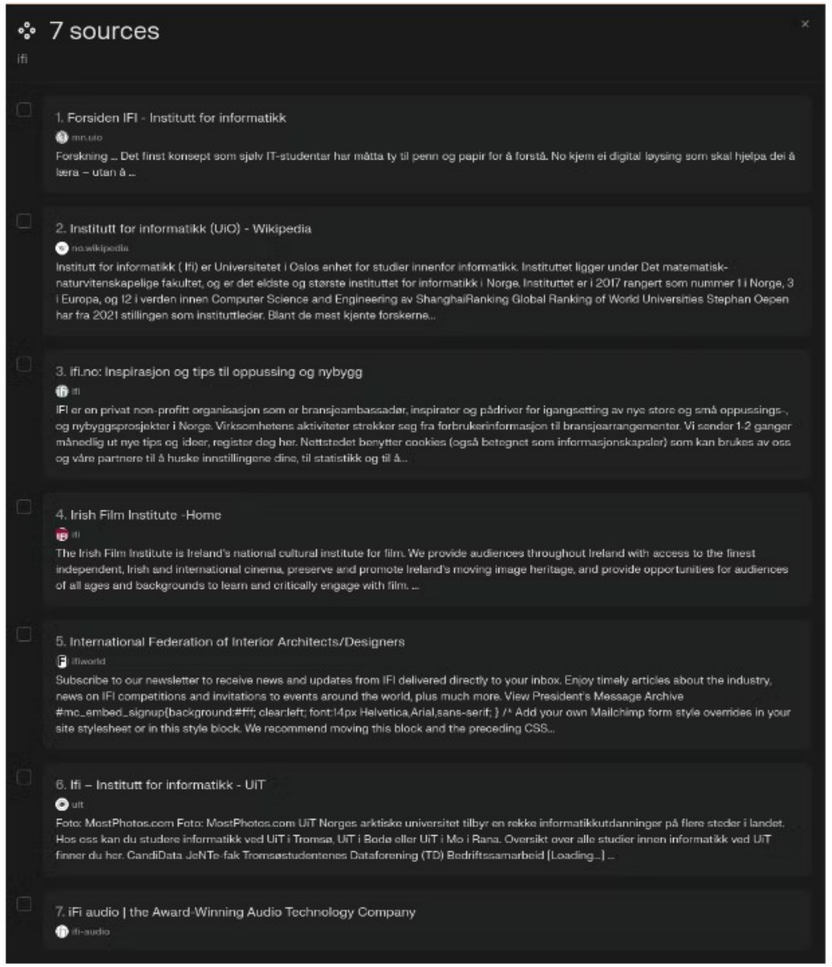
UiO GPT ? ? ? ?



UiO GPT...bedre?



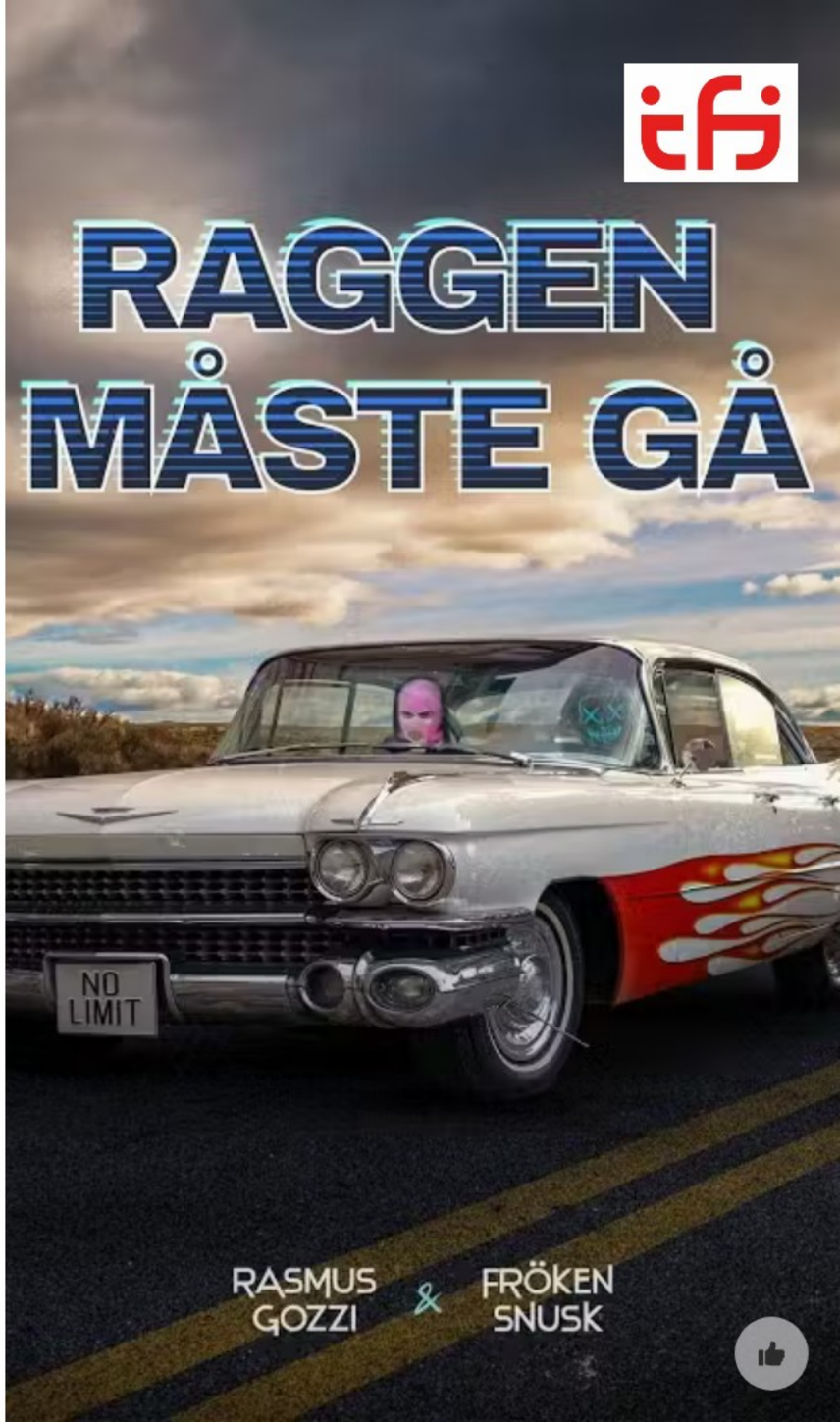
Enorm tangent: Prompt engineering

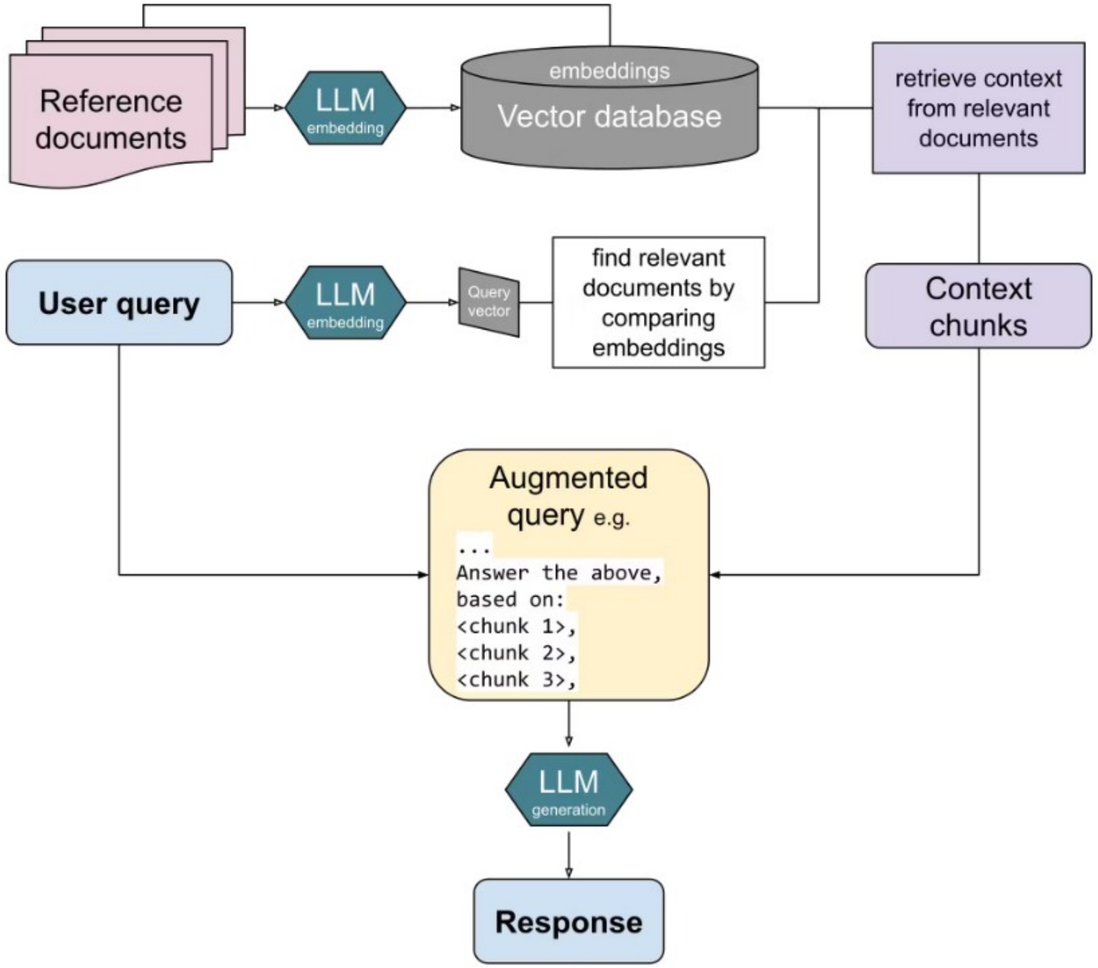


Perplexity: Kilder!

Further reading: RAG

- "Retrieval augmented generation"
- Virkelig **ikke** pensum
- (Men veldig kult)





RAG-arkitektur

Grønn informatikk: søk

- Tradisjonelt søk: Lite vannforbruk
- LLM-prompt: *Mye* Vannforbruk
- Ikke bra!

ChatGPT-3 vs. ChatGPT-4:
Training Power Consumption and Costs in U.S. Dollars

Model	GPT-3	GPT-4
Days of training	34	100
Training energy consumption	1,287,000 kWh	62,318,800 kWh
Cost of traning*	$168,597	$8,163,763

Data sources: U.S. Energy Information Administration, Electric Power Research Institute (EPRI)
*Calculated based on the average U.S. commercial electricity rate of $0.131 per kWh
as of June 2024

BestBrokers.com

The Brussels Times

SUBSCRIBE

MAGAZINE | BRUSSELS | BELGIUM | BUSINESS | ART & CULTURE | EU AFFAIRS | WORLD | BELGIUM UNLOCKED | JOBS

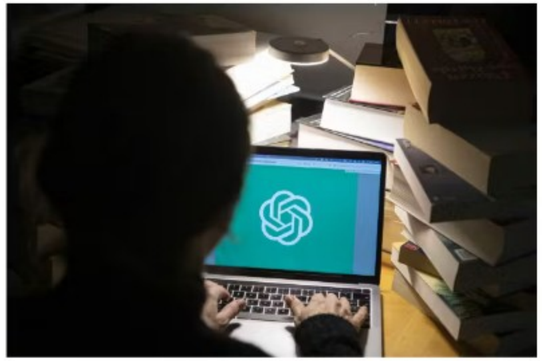
Most Read

1. National strike: Major disruption of flights departing from Belgium on Tuesday
2. TBT CEOs earn almost 200% more than average employee, Brussels has highest wages
3. Can Philippe Close speak for the City of Brussels?
4. Two people linked to fatal home invasion in Ixelles arrested in Lisbon
5. Brussels municipality launches free babysitting service for single parents

ChatGPT consumes 25 times more energy than Google

Sunday 12 May 2024

By The Brussels Times Newsroom



Credit: Belga

Energy consumption by Artificial Intelligence (AI) is rising rapidly: AI is predicted to consume twice as much energy as the whole of France by 2030, according to some calculations.

Not only does AI already have a solid climate impact now, it also complicates the fight against climate change. All this extra demand for electricity makes it even more difficult to electrify transport, for example, as soon as possible. The question arises: what do we best use AI for?

A year and a half after its launch, ChatGPT has 180 million users. The AI chatbot is used for all sorts of things, from creating a travel itinerary to writing a paper or just asking a silly question.

Pause til 15:15 😄

OBLIGHJELP TIL 16

Obligverksted. Skriv noe og det vil bli snakket om i plenum

Krever C-2 at man har gjort B-2 siden den bruker EditSearchEngine?