






•  \_\_\_\_\_

•  \_\_\_\_\_



 A look at search engines with their own indexes



  Rohan Kumar




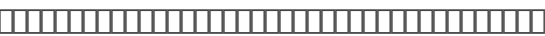
 2025-01-30 (commit hash: 15ddf9743664d7b1b6401ba6d49b1bbace9c06ed)










   Google  Bing  Yandex  GBY  GBY  
 GBY 





  


 “”  
  


 / RDFa  
 JSON-LD 











 “”  
  


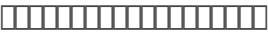














Google



-  [Startpage](#)  [Google](#)  [Startpage](#)  [Bing](#)<sup>[2]</sup>
- [GMX Search](#)
-  [Runnaroo](#)
- [Mullvad Leta](#)
- [SAPO](#)
- [DSearch](#)
- [13TABS](#)
- [Zarebin](#)
-   \_\_\_\_\_ 

Bing



- [Yahoo](#) [OneSearch](#)
- [DuckDuckGo](#)<sup>[3]</sup> [Tor](#)  [JS](#)  [TUI](#)  “  ”  

- [AOL](#)
- [Qwant](#) <sup>[4]</sup>
- [Ecosia](#)
- [Ekoru](#)
- [Privado](#)
- [Findx](#)
- [Disconnect Search](#)<sup>[5]</sup>
- [PrivacyWall](#)
- [Lilo](#)
- [SearchScene](#)
- [Peekier](#) [Peekr](#) 
- [Oscobo](#)
- [Million Short](#)

- Yippy search[6]
- Lycos
- Givero
- Swisscows
- Fireball
- Netzzappen
- You.com[7]
- Vuhuv
- Metager [ ]
- ChatGPT Search[8]
- [ ] Bing [ ]

## Yandex

[ ]

IndexNow API [ ] Bing [ ] Seznam [ ] IndexNow [ ]

- Epic Search[ ] 2021 [ ] 6 [ ]
- [ ] DuckDuckGo [ ] Bing[ ] [ ] *DuckDuckgo* [ ] Yandex “[ ] ”
- [ ] [ ] “[ ] ” [ ] [ ]
- Petal[ ]

## Mojeek

[ ]

GBY[ ]

Mojeek [ ] [eTools.ch](#) [ ] [ ]  
*Mojeek* [ ] *GBY* [ ] [ ]

Google[ ] Bing [ ] Yandex [ ] microformats1[ ] microdata[ ] RDFa[ ] Open Graph  
 markup [ ] JSON-LD[ ] Yandex [ ] microformats1 [ ] H-Card  
 [ ] Open Graph [ ] Schema.org [ ] Mojeek  
 [ ] Open Graph [ ] Schema.org [ ]

[ ]

[ ] “[ ] ”  
 [ ]  
 [ ]

## Stract

[ ] Stract [ ] “optics”[ ] Brave [ ] “  
 [ ] ”[ ] Stract [ ] [ ] AGPL-3.0  
 [ ] Stract [ ] Common Crawl





Secret Search Engine Labs

SEO
“”
CashRank

Gabanza

Jambo

2006

search.dxhub.de

Gigablast
Gigablast

Fynd

URL
/

Yessle

Bloopish

YaCy

/

Scopia

MetaGer Bing 2024 9  
10 MetaGer

Artado Search

Plumb JS Google Bing  
Yahoo Petal MetaGer  
"twitter"  
"wikipedia" "reddit"

Active Search Results

Crawlson

10 URL seirdy.one  
Crawlson

Anoox

Yioop!

UseNet Yioop API  
WARC Meorca Yioop

Spyda

James Mills So I'm a Knucklehead eh? Go  
MIT Spyda

Slzii.com



seirdy.one 

--	--	--	--	--	--	--

Qwant

Qwant

Bing

Bing

Neeva

Kagi Search

Neeva

TeclisTeclis

GoogleBing

Teclis

Kagi

Kagi.ai

TinyGem

Kagi.com

BraveAPI

PriEco

Google“”

GBY

Marginallia Search

/

GBY

SEO

SERPs

Teclis

Kagi2022-

05-28

Marginallia.nu

Ichido

Marginalia

CAPTCHA

SEO

Ichido

Teclis

```

graph TD
    KagiSearch[Kagi search] --> uBlockOrigin[uBlock Origin]
    KagiSearch --> ReadabilityJS[Readability.js]
    KagiSearch --> Trafilatura[Trafilatura]
    uBlockOrigin --> ReadabilityJS
    ReadabilityJS --> Trafilatura
    Trafilatura --> HTML[HTML]
    Trafilatura --> Web[Web]
    Trafilatura --> Marginalia[Marginalia]
    Trafilatura --> API[API]
    Trafilatura --> Kagi[Kagi]
    HTML --> Web
    Web --> Marginalia
    Web --> API
    Web --> Kagi
    Marginalia --> API
    API --> Kagi
    Kagi --> Kagi
  
```

Clew


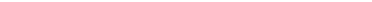
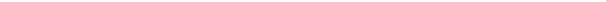



FOSS  
seirdy.one

## Lixia Labs Search

The diagram illustrates the memory layout for JavaScript. It consists of two horizontal rows of small rectangular cells, representing memory slots. The top row contains 16 cells, and the bottom row contains 16 cells. To the right of the top row, the text "JavaScript" is written, indicating the language associated with this memory layout.

--	--	--	--	--

## Kozmonavt

 800   
 Kozmonavt  
  
 /  seirdy.one 

search.tl

The diagram shows a URL structure with the following components highlighted in boxes:

- TLD** (Top Level Domain)
- .com** (Domain extension)
- TLD** (Top Level Domain)
- [13]** (Index number)
- UI** (User Interface)
- TLD** (Top Level Domain)
- /** (Slash separator)
- tld** (Top Level Domain)
- URL** (Uniform Resource Locator)
- .org** (Domain extension)
- &tld=org** (Query parameter)
- URL** (Uniform Resource Locator)
- Amidalla** (Domain name)
- Amidalla** (Domain name)
- URL** (Uniform Resource Locator)
- search.tl** (Domain name)

## Thunderstone

Diagram illustrating the structure of a URL:

- Protocol: http
- Subdomain: www
- Domain Name: billybob
- Top-Level Domain: com
- Second-Level Domain: net
- Third-Level Domain: org
- Path Segment: web
- Path Segment: yahoo
- Path Segment: dogpile
- Path Segment: aol

The full URL is: http://www.billybob.com/net/org/web/yahoo/dogpile/aol/

sengine.info

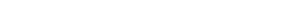



netEstate GmbH

## Gnomit

2009  
"IRC"  IRC

--	--

## High Browse


 SEO 
 "  "  "

 " 

## Keybot

[illegible]

## Quor


 2021  6 
 www dot quor dot com 

## Semantic Scholar

Allen Institute for AI PDF  
[Redacted]








## Bonzamate

searchcode

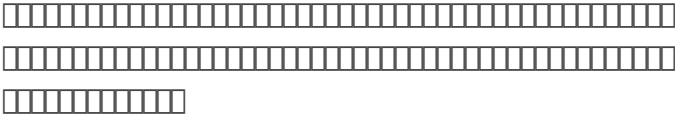
[illegible]

- Baidu GBY
- 360 Qihoo 360
- Toutiao
- Sogou
- Yisou
- Naver Searx
- Daum
- Seznam seirdy.one
- IndexNow Bing Yandex
- Cốc Cốc
- go.mail.ru
- LetSearch.ru URL

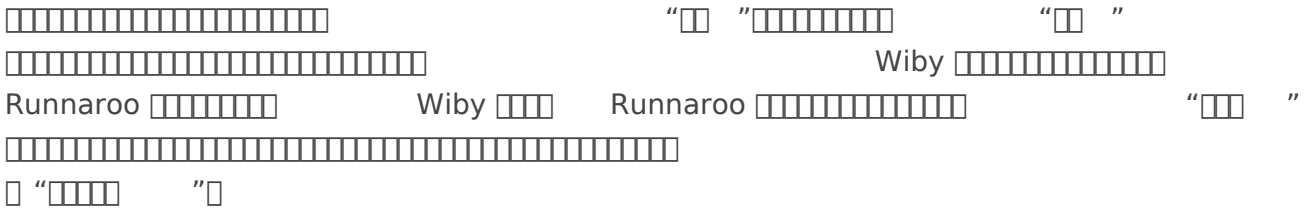
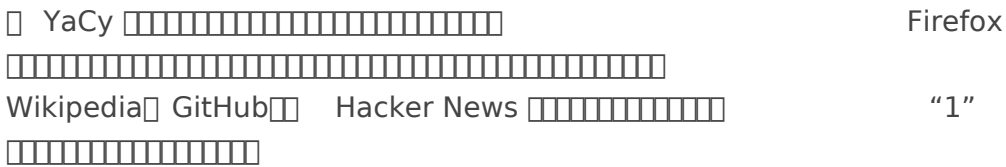


- ALibw.com 
- Vuhuv   
- search.ch 
- fastbot 
- SOLOFIELD 

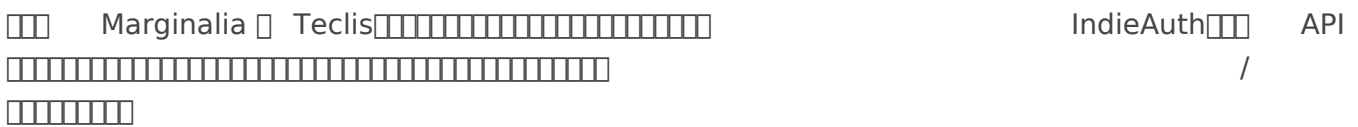
- kaz.kz [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] " [ ] [ ] [ ] [ ] "



wiby.me □ wiby.org

Mwmbi

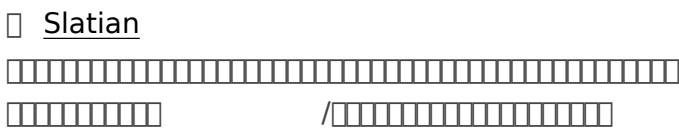
## Search My Site



Kukei.eu



## Unobtainium Search





Ask.com

info.com kensaq.com

Google Bing Yandex  
"ask.com " directhit.com

## Infinity Search

Infinity De  
  
 "  "  URL

# Infinity Decentralized



- 
- 



# Petal Search

# JavaScript

Yandex ☐ Qwant ☐ 2023 ☐ 6

## Neeva

“” Bing Bing Neeva  
Bing Bing  
OAuth  
Snowflake 2023 5

Gigablast

Web  
Private.sh Gigablast Right Dao 2023

wbsrch

Gowiki

seirdy.one  
2022

Meorca

“” “”  
seirdy.one

Ninfex

“” “”

Marlo

Marlo Haskell  
marlo.sandymaguire.me

websearchengine.org tuxdex.com

inetdex.com  
1000 Cookies



Entfer

[redacted]  
[redacted]

Siik

[redacted] ToS [redacted] PHP  
[redacted]  
[redacted]

Blog Surf

[redacted] RSS/Atom  
[redacted]  
“MarketRank” [redacted] “Hacker News” [redacted]  
  
[redacted]

- Parsijoo [redacted]
- Moose.at [redacted] Brave [redacted]



[redacted]  
  
[redacted]  
  
[redacted] (POC)  
[redacted]



[redacted]



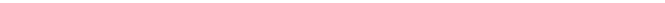



Google [redacted] Bing [redacted] Yandex [redacted] Google [redacted]  
YouTube [redacted] LinkedIn  
[redacted]  
[redacted]  
[redacted]





[illegible]


 Kagi  Ask.com  
  




ToS 

--	--	--	--	--	--

Diagram illustrating the distribution of 100% across different categories:


- Top bar: 100% (represented by 100 small squares)
- Second bar: 10% (represented by 10 small squares) labeled "matza gebrent" and "matzo brei"
- Third bar: 10% (represented by 10 small squares) labeled Mojeek
- Fourth bar: 20% (represented by 20 small squares) labeled 20
- Fifth bar: 30-40% (represented by 30-40 small squares) labeled 30-40

--	--








--	--	--	--

Diagram illustrating memory allocation for Google and Chromium/Firefox. The diagram shows three rows of memory blocks, each represented by a horizontal bar divided into segments. The top row is labeled 'Google' and contains 4 blocks. The middle row is labeled 'Google' and contains 8 blocks. The bottom row is labeled 'Chromium' and 'Firefox' and contains 1 block for Chromium and 19 blocks for Firefox.

Diagram illustrating the distribution of bot traffic across different bots. The diagram shows four horizontal bars representing the relative volume of traffic from various bots. The bots listed are Cloudflare, Googlebot, BingBot, and TwitterBot. Cloudflare has the highest volume, followed by Googlebot, BingBot, and TwitterBot. The bars are color-coded: Cloudflare is light blue, Googlebot is light green, BingBot is light orange, and TwitterBot is light purple.

Google JavaScript " " GBY JavaScript

Google



11

Qwant  Bing  “” 



13.

--	--	--	--	--	--	--	--

site:

TLD [ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ][ ]

site:.one

--	--	--	--	--	--	--

“.one” TLD ☐☐☐☐

14.

Gigablast ☐ Matt ☐☐☐☐☐☐

Google ☐ Microsoft☐☐☐

YouTube  LinkedIn



[TheCoffeMaker/shm: SelfHoster Manifesto, coz Arrr! Rulez - Codeberg.org](#)

"

"

"

19

"




[illegible]

15

14 