






•  _____

•  _____



 A look at search engines with their own indexes









  Rohan Kumar

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
















 [1] Google Bing Yandex GBY GBY
 GBY 








 “”



 / RDFa
 JSON-LD 



 “”











  



Google



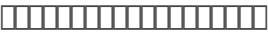
 Websub 









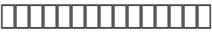



-  Startpage Google  Startpage  Bing[2]
- GMX Search
-  Runnaroo
- Mullvad Leta
- SAPO
- DSearch
- 13TABS
- Zarebin
-   _____ 

Bing

 IndexNow API 

Yandex  Seznam  IndexNow



- Yahoo OneSearch
- DuckDuckGo[3] Tor  JS  TUI  “  ”

- AOL
- Qwant [4]
- Ecosia
- Ekoru
- Privado
- Findx
- Disconnect Search[5]
- 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

WARC

Yioop

API

Meorca

Yioop

Spyda

James Mills

So I'm a Knucklehead eh?

Go

MIT

Spyda

Slzii.com

seirdy.one

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

[illegible]

Brave Search

[11]

URL

Brave Search ☐ Tor ☐☐☐☐☐☐☐☐☐☐ JS ☐ ☐☐☐☐☐

- Ghostery Private Search [12]
- Kagi 2023 2024 5 1

The diagram illustrates the layout of a web page with various elements and their bounding boxes. The elements are arranged in a grid-like structure, with each element represented by a rectangle. The elements are as follows:

- Google**: A large rectangle at the top left.
- cse.google.com**: A small rectangle at the top right.
- Google**: A small rectangle at the top right, next to the **cse.google.com** rectangle.
- Plumb**: A small rectangle in the middle left.
- URL**: A small rectangle in the middle right.
- hCaptcha**: A small rectangle in the middle right, next to the **URL** rectangle.
- Gibiru**: A small rectangle in the bottom right.

The diagram shows the relative positions and sizes of these elements, with the **Google** element being the largest and the **Gibiru** element being the smallest.

Qwant

Qwant

Bing

Bing

Neeva

Kagi Search

Neeva

TeclisTeclis

GoogleBing

Teclis

KagiKagi.aiTinyGemKagi.com

BraveAPI

PriEco

Google“”

GBY

Marginallia Search

/

GBY

SEOSERPsTeclisKagi2022-

05-28Marginallia.nu

Ichido

Marginalia

CAPTCHASEO

Ichido

Teclis

```

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

Clew


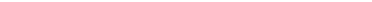
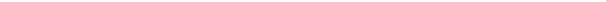



FOSS
seirdy.one

Lixia Labs Search

[illegible]

--	--	--	--	--

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

The diagram illustrates the structure of a packet, likely a DNS query, with the following fields and sizes:

- Header:** 10 bytes
- Question:** 10 bytes
- Answer:** 10 bytes
- Authority:** 10 bytes
- Additional:** 10 bytes
- Web:** 10 bytes
- COM:** 1 byte
- NET:** 1 byte
- ORG:** 1 byte
- “BillyBob”:** 10 bytes
- AOL:** 1 byte
- “:** 1 byte
- Yahoo:** 1 byte
- Dogpile:** 1 byte
- “:** 1 byte
- ”:** 1 byte

sengine.info

netEstate GmbH

Gnomit

2009
"IRC" IRC

High Browse

SEO

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

 “

--	--	--	--

 ”

--	--	--	--

 ”

Keybot

Quor

□□□□□□□□□□
2021 □ 6 □□□□□□□□□□
www dot quor dot com □□□

Semantic Scholar

PDF

Bonzamate








 Boyter
 Bonzamante

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

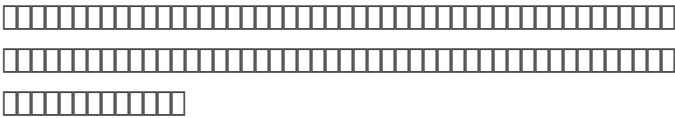
--	--	--	--	--	--	--	--	--	--

 “

















































































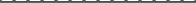




















--	--	--	--	--	--	--	--

 ”

--	--	--



wiby.me □ wiby.org

Mwmbi

Yacy ██████████ Firefox
██
Wikipedia ██████████ GitHub ██████████ Hacker News ██████████ "1"
████████████████████

Search My Site

Marginalia Teclis IndieAuth API

Kukei.eu

Web

Unobtainium Search

[illegible]



Ask.com

info.com kensag.com

Google Bing Yandex
"ask.com " directhit.com

Infinity Search

Infinity De

 " " URL

Infinity Decentralized



-
- The diagram illustrates two data structures for storing information:
- Array:** A single long row of 25 slots. It contains the following labels from left to right:
 - Slot 1: Siri
 - Slot 2: Applebot
 - Slot 3: (empty)
 - Slot 4: (empty)
 - Slot 5: (empty)
 - Slot 6: (empty)
 - Slot 7: (empty)
 - Slot 8: (empty)
 - Slot 9: (empty)
 - Slot 10: (empty)
 - Slot 11: (empty)
 - Slot 12: (empty)
 - Slot 13: (empty)
 - Slot 14: (empty)
 - Slot 15: (empty)
 - Slot 16: (empty)
 - Slot 17: (empty)
 - Slot 18: (empty)
 - Slot 19: (empty)
 - Slot 20: (empty)
 - Slot 21: (empty)
 - Slot 22: (empty)
 - Slot 23: (empty)
 - Slot 24: (empty)
 - Slot 25: (empty)
 - Hash Table:** A structure with 5 slots, each containing a pointer to an array of slots.
 - Slot 1: Points to an array of 25 slots. The first slot contains 'Alexa Internet', and the rest are empty.
 - Slot 2: Points to an array of 25 slots. The first slot contains 'Amazonbot', and the rest are empty.
 - Slot 3: Points to an array of 25 slots. The first slot contains 'Applebot', and the rest are empty.
 - Slot 4: Points to an array of 25 slots. The first slot contains 'Siri', and the rest are empty.
 - Slot 5: Points to an array of 25 slots. The first slot contains 'Alexa', and the rest are empty.



Petal Search

The diagram illustrates a memory stack with seven horizontal bars of decreasing length from top to bottom. The top bar is labeled 'Android'.

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]



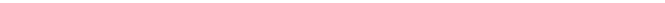

[redacted] (POC)
[redacted]



[redacted]



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


 Kagi  Ask.com



ToS

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

matza gebrent

matzo brei

Mojeek


20

30-40

--	--








--	--	--	--

Cloudflare

Googlebot

BingBot

TwitterBot

Google JavaScript " " GBY

Google

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][illegible]

15

14