

Search The Internet

Search . . .

Search

## About

This is an independent DIY search engine that focuses on non-commercial content, and attempts to show you sites you perhaps weren't aware of in favor of the sort of sites you probably already knew existed.

The software for this search engine is all custom-built, and all crawling and indexing is done in-house. The project is open source. Feel free to poke about in the [source code](#) or contribute to the development!

The search engine is currently serving about 56 queries/minute.

Consider [supporting the project!](#)

[Read More](#)

## Tips

This search engine isn't particularly well equipped to answering queries posed like questions, instead try to imagine some text that might appear in the website you are looking for, and search for that.

Where this search engine really shines is finding small, old and obscure websites about some given topic, perhaps [old video games](#), [a mystery](#), [theology](#), [the occult](#), [knitting](#), [computer science](#), or [art](#).

## Publicity, Discussion and Events

[Marginalia: A New Lens for the Internet](#)

2024-04-03 Perephoneia

[Folk Search Engines](#)

2024-02-20 Escape The Algorithm

[How bad are search results? Let's compare](#)

[Google, Bing, Marginalia, Kagi, Mwmbli, and ChatGPT](#)

2023-12-31 Dan Luu

[Interview with Viktor Löfgren from](#)

[Marginalia.nu](#)

2023-12-30 YouTube, Rickard

Andersson

[FreeWebSearchDay Interview](#)

2023-09-29 NLnet

[Marginalia Search receives FUTO Grant](#)

2023-09-15

[Marginalia \[...\] is currently on the HN front page and handling the traffic with one \\$5k commodity server](#)

2023-04-19 🐦 Twitter @vboykis

[Marginalia: DIY search engine that focuses on non-commercial content](#)

2023-04-18 Hacker News

[Marginalia Receives NLnet grant](#)

2023-02-26

[Marginalia's Index Reaches 100,000,000](#)

[Documents](#) 🧑🏻🧑🏻🧑🏻

2022-10-21

[Google ei enää tideä](#)

2022-10-19 Helsing Sanomat 🇫🇮

[Kritik an Googles Suche - Platzhirsch auf dem Nebenschauplatz](#)

2022-08-18 Deutschlandfunk Kultur 🇩🇪

[Marginalia Goes Open Source](#)

2022-05-28 Hacker News

[You Should Check Out the Indie Web](#) 🏠

2022-03-15 YouTube, You've Got Kat

[What Google Search Isn't Showing You](#)

2022-03-10 The New Yorker 🇺🇸

2022-03-09 MetaFilter

 [First anniversary!](#) 

2022-02-26

[A search engine that favors text-heavy sites and punishes modern web design](#)

2021-09-16 Hacker News

[A Search Engine Designed To Surprise You](#)

2021-09-16 Clive Thompson OneZero

 [RSS Feed](#)

## Syntax

---

This is a keyword-based search engine. When entering multiple search terms, the search engine will attempt to match them against documents where the terms occur in close proximity.

Search terms can be excluded with a hyphen.

While the search engine at present does not allow full text search, quotes can be used to specifically search for names or terms in the title. Using quotes will also cause the search engine to be as literal as possible in interpreting the query.

Parentheses can be used to add terms to the query without giving weight to the terms when ranking the search results.

## Samples

---

soup -chicken

Look for keywords that contain soup, but not chicken.

"keyboard"

Look for pages containing the exact word keyboard, not keyboards or the like.

## Special Keywords

---

Several special keywords are supported by the search engine.

Keyword	Meaning
<code>site:example.com</code>	Display site information about <i>example.com</i>
<code>site:example.com keyword</code>	Search <i>example.com</i> for <i>keyword</i>
<code>browse:example.com</code>	Show similar websites to <i>example.com</i>
<code>ip:127.0.0.1</code>	Search documents hosted at <i>127.0.0.1</i>
<code>links:example.com</code>	Search documents linking to <i>example.com</i>
<code>tld:edu keyword</code>	Search documents with the top level domain <i>edu</i> .
<code>?tld:edu keyword</code>	Prefer but do not require results with the top level domain <i>edu</i> . This syntax is also possible for <code>links:...</code> , <code>ip:...</code> and <code>site:...</code>

	Keyword	Meaning
"steve mcqueen"		
Look for pages containing the exact words steve mcqueen in that order, with no words in between.	q>5	The amount of javascript and modern features is at least 5 (on a scale 0 to 25)
apology (plato)		
Look for pages containing apology and plato, but only rank them based on their relevance to apology	q<5	The amount of javascript and modern features is at most 5 (on a scale 0 to 25)
	year>2005	(beta) The document was ostensibly published in or after 2005
	year=2005	(beta) The document was ostensibly published in 2005
	year<2005	(beta) The document was ostensibly published in or before 2005
	rank>50	The ranking of the website is at least 50 in a span of 1 - 255
	rank<50	The ranking of the website is at most 50 in a span of 1 - 255
	count>10	The search term must appear in at least 10 results form the domain
	count<10	The search term must appear in at most 10 results from the domain
	format:html5	Filter documents using the HTML5 standard. This is typically modern websites.
	format:xhtml	Filter documents using the XHTML standard
	format:html123	Filter documents using the HTML standards 1, 2, and 3. This is typically very old websites.
	generator:wordpress	Filter documents with the specified generator, in this case wordpress

## Keyword

## Meaning


file:zip	Filter documents containing a link to a zip file (most file-endings work)
file:audio	Filter documents containing a link to an audio file
file:video	Filter documents containing a link to a video file
file:archive	Filter documents containing a link to a compressed archive
file:document	Filter documents containing a link to a document
-special:media	Filter out documents with audio or video tags
-special:scripts	Filter out documents with javascript
-special:affiliate	Filter out documents with likely Amazon affiliate links
-special:tracking	Filter out documents with analytics or tracking code
-special:cookies	Filter out documents with cookies


## Results Legend

---

The estimated relevance of the search result is indicated using the color saturation of the color of the search result, as well as the order the results are presented.

Information about the position of the match is indicated using a dot matrix in the bottom bar of each search result. Each dot represents four sentences, and are presented in an order of top-to-bottom, left-to-right.

 — The terms occur heavily toward the beginning of the document.

 — The terms occur sparsely throughout

## Policies

---

This website complies with the GDPR by *not collecting any personal information*, and with the EU Cookie Directive by *not using cookies*. [More Information](#).

## Contact

---


Reach me at [kontakt@marginalia.nu](mailto:kontakt@marginalia.nu), [@MarginaliaNu](https://twitter.com/MarginaliaNu) on twitter.


## Open Source

---

The search engine is open source with an AGPL license. The sources can be perused at

the document.

 — The terms occur only in a single sentence.

Potentially problems with the document are presented with a warning triangle, e.g.  3. Desktop users can mouse-over this to get a detailed breakdown.

<https://git.marginalia.nu/>.

## Data Sources

---

IP geolocation is sourced from the IP2Location LITE data available from <https://lite.ip2location.com/> under [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).