

# Goodbye, Google covertly sabotaged FINDX – Findx is shutting down

It is over – before it really started. Findx is shutting down.

The last few years with Findx have been incredible, educational, challenging, fun and frustrating, sometimes even all at once.

Findx was a dream about building a private and independent search engine – a true alternative to Google designed with respect for people’s privacy at its core.

It wasn’t just a search engine – we were committed to protecting people’s online privacy and we built a private search engine, a desktop browser, browser extensions and mobile apps for enhanced private browsing – all with full transparency, mostly based on open source projects.

Maybe we were a little ahead of our time with our strong focus on privacy, and it turned out to be a lot harder than we imagined to make people realize that privacy is important - but we were right! The focus on privacy has now been named as consumers’ most important topic by Gartner.

We’ve learned that people say one thing about privacy, and do something else when it comes to privacy vs. convenience. Another thing that surprised us a bit was the resistance shown

by tin-foil-hat privacy activists... could any tool ever be good enough for them?

# Thank you

We want to take this opportunity to thank you – all of you! The people who used our services and the supporters who actively participated, the developers involved, the few people who actually got it, everyone who listened, cheered, gave critiques and also those of you who shook your heads.

I, as the founder of the company, want to thank the team for their efforts the last 3 years. Thank you to the backend developers [Ivan Skytte Jørgensen](#) and [Ai Lin Chia](#) who worked tirelessly to improve search results and make the backend fast and stable.

Thank you to our outreach guy, [Brian Schildt Laursen](#), whose enthusiasm about the project was amazing and showed at the meetups and conferences where he preached privacy to all that would listen to him.

Thank you to our freelancers who are too many to mention, but you know who you are! None mentioned, none forgotten.

We learned a lot about crawling, indexing and scaling along the way – thanks to [Greg Lindahl](#) (formerly Blekko) for sparring with us and giving us advice. We wish you lots of luck with your endeavours and hope they will see more success than our attempt.

Our front-end was cutting edge (maybe a bit too cutting edge) – thanks to [Denis Izmaylov](#) and his team for pushing it. Together we learned a lot that can be used in future projects.

Apps and browsers were made possible thanks to Mozilla, Raymond Hill (uBlock Origin) and all the open source developers involved in those projects.

## A word of warning

The search engine was based on the [Gigablast](#) open-source search engine by Matt Wells. Normally I endorse the *“If you don’t have something nice to say, don’t say anything at all”* saying, but... That was one of the biggest mistakes we made. Looking back, that codebase was outdated and we ended up putting way too much time and money into bug-fixing and optimizing it instead of adding new features. It was not as flexible or feature rich as it needed to be, to be the foundation of a modern search engine. We came close on a number of occasions to making a deal with Matt to develop it further, but every time he ended up disappearing. We will leave our version available for the foreseeable future, but our recommendation is clear – **DO NOT USE IT** – not ours, and especially not the original. Start somewhere else if you want to build a new search engine. Do not base your project on a one-man project, no matter how confidently it is “sold” to you.

# The long read

Here is an explanation of some of the successes and the challenges that we faced as we crafted our apps and search engine.

## **Privacy is important, and people are starting to get it**

Findx was ahead of its time, and apart from the technical development, one of the greatest challenges was to gain traction and interest around our strong focus on data privacy. People tend to be annoyed by intrusive ads, and you can solve it by visually removing them with an adblocker – but when it comes to changing a core service, like people’s standard search engine, it is very hard to break away from the monopoly and convenience of Google Search – and the tolerance for alternative search results is very low.

# **The many technological challenges faced by search engines**

As an independent search engine, there were several focus areas for our programmers:

1. The finite server resources.
2. The crawler that trawled the web to index pages.
3. The index itself that pages were added to.
4. The ranking algorithms that displays the most relevant pages at the top.

## **Search engines eat resources like crazy**

All of these are resource-intensive tasks. They need a huge amount of disk space and processing power, and they need to scale up as user numbers increase. We invested in our own servers, customized for speed and scaling – an expensive decision that never really paid off, but it was the right decision due to the amount of data we stored in our 2+ billion page index and the amount of traffic we received. A cloud-based service couldn't have done the same, at least not with the software we used and for the same cost.

We know that speed is important for people and we spent a lot of time trying to optimize the code to minimize resource use and maximize performance.

## **Search engine crawlers**

Crawling the internet is not as simple as it may sound, and we ran into several bumps along the way.

## **Badly behaved crawlers are blacklisted**

Because of bugs in the original Gigablast spidering code, the Findx crawler ended up on a blacklist in Project Honeypot as being “badly behaved” (fixed in our fork). That meant quite a bit of trouble for us because CDN providers, which are a very powerful hubs for internet traffic, put a lot of weight on this blacklist. Some of the most popular websites and services on the internet run through services like Cloudflare and other CDNs – so if you are in bad standing with them, suddenly a large part of the internet is not available, and we weren’t able index it.

Unfortunately, sites with diligent webmasters blocked the Findxbot on their sites. As these blocks were often manually set up, getting them removed was near impossible, even though we got our bot whitelisted and proved that it was very well behaved after its rocky start.

## **Big sites forbid independent indexing**

Many large websites like LinkedIn, Yelp, Quora, Github, Facebook and others only allow certain specific crawlers like Google and Bing to include their webpages in a search engine index (*maybe something for European Commissioner for Competition Margrethe Vestager to look into?*) Other sites put their content behind a paywall.

That meant that the Findx search index was incomplete and was not able to return results that were likely both relevant and good

quality. When you compare any independent search engine's results to Google for example, they have no chance to be as relevant or complete because many large websites refuse to allow any other search engine to include their pages.

## **Building a quality index**

The number one problem for all search engine indexes is spam and malicious web pages. It is programmatically impossible for a crawler to know whether a web page is good quality or not. This is a problem that even Google struggles with daily, employing thousands of quality checkers.

Our approach was to build in a quality rating tool, letting searchers rate their search results. We still believe that crowd-sourced quality control is a good thing, and we are pleased to see that such a feature has recently been added to a couple of the major search engines.

## **Spam is never ending and always increasing**

We spent a lot of time fine-tuning our rules to reduce the amount of spam, malicious pages and fake web shops in our index. As part of that initiative, we made the decision to not index known porn sites – too many of such pages contain malicious code like computer viruses or worse. Together with e-mærket and SØIK we participated actively to fight the increasing number of fake web shops that are constantly appearing.

## **Returning good results**

There are a number of technically difficult challenges in understanding a search phrase and matching it to appropriate (relevant) results. As Findx was a European search engine, we made it harder for ourselves – Europe has many languages! It would have been much easier to get a search engine to return good results in one language first, and we did quite well in Danish towards the end. A special thanks to Sussi Olsen from the Centre for Language Technology at the University of Copenhagen, and their STO database, which made it possible for our developers to improve the Danish search results.

## Monetizing - the chicken and the egg

Most advertisers won't work with you unless you either give them data about your users, so they can effectively target them, or unless you have a lot of users already.

Being a new and independent search engine that was doing the time-consuming work of growing its index from scratch, and being unwilling to compromise on our user's privacy, Findx was unable to attract such partners.

Earlier this year we entered into search partnership discussions with Bing and Yahoo, in order to get access to both search results and ads – like our competitors from DuckDuckGo and StartPage and Qwant do. We simply asked for same terms as these three services already work with, and we could have become the “European DuckDuckGo” - using a third-party search feed while working on our own index, but they were not willing to work with us. This pretty much killed the idea of ever monetizing our search engine moving forward.

We could not retain users because our results were not good enough, and search feed providers that could improve our results refused to work with us before we had a large userbase ... the chicken and the egg problem.

## **Small independent services are actively discouraged**

From forbidding crawlers to index popular and useful websites and refusing to enter into advertising partnerships without large user numbers, to stacking the requirements for search extension behaviour in browsers, the big players actively squash small and independent search providers out of their market.

Unfortunately, the reality is that the Findx search engine was unsustainable.

# Please support our supporters

During our journey, we received enthusiastic support from a number of privacy focused organisations and tech companies.

Findx was one of the first search engines to be a verified publisher in the [Brave browser](#). Brave is an exciting and worthwhile project where users can contribute directly to content producers as they spend time browsing their favourite websites.

[Waterfox](#), a desktop browser with focus on speed and privacy, developed by [Alex Kontos](#) – thank you for being one of the first to add Findx as a search option.

[SnowHaze](#) is a lightweight privacy-focused mobile browser for iOS, with a built-in tracking protector and adblocker, as well as device fingerprinting protection.

[DataEthics](#) is a politically independent organisation based in Denmark with a European (and global) outreach. We were delighted to be invited to participate in every DataEthics Forum conference since the organisation was founded in 2016.

We have also participated in and sponsored several important yearly events, including the [Data Privacy Day](#), the [National Cybersecurity Awareness Month](#) by Stay Safe Online, and the [Safer Internet Day](#). These events help raise awareness about the importance of data privacy both at home, at work and throughout society.

Please continue to support those who are actively working to raise awareness about data privacy and those who protect your data privacy.

We hosted our backend servers with [Netgroup](#) in Denmark, and always received wonderful service from especially the site manager [Johnni Andersen](#), and our former account manager [Peter Trautner](#). Thanks guys! I would be happy to work with you again sometime.

# That's all folks

On behalf of the team – goodbye.

[Brian Rasmusson](#), Founder of Findx (br@findx.com)