

Rote Karte für den Suchgiganten

Egg, 14. September 2020: Über unsere Linux-Distribution AVMultimedia, welche im Hauptspeicher (RAM) läuft, wurde bereits vielfach berichtet. Neben vielen Vorteilen (z.B. schnelles Starten von Programmen) führt ein Neustart des Systems dazu, dass alle gespeicherten Cookies gelöscht werden. Das passt leider nicht allen, und davon handelt dieser Blog. Wem dieser zu lange ist, findet in nachfolgendem Video auf gute zwei Minuten verkürzt den wesentlichen Inhalt ebenfalls. Viel Spass!

<https://archivista.ch/cms/wp-content/uploads/file/avmultimedia-searx.mp4>

Wenn Google jede Arbeit verweigert

Seit längerer Zeit erfragt der Suchgigant beim ersten Aufruf seiner Homepage nach der Zustimmung zu Cookies. Das sind kleine Datenfragmente im eigenen Browser, die meist dazu verwendet werden, eine eindeutige Identität des Webbrowsers und damit der Nutzer/innen festzuhalten. Das Cookie dient dazu einzig als Ortungsnummer. Die für die Auswertung der Suchabfragen und des Surfverhaltens interessanten Informationen werden dabei auf den Seiten der Anbieter, hier Google gespeichert.

Bislang war es möglich, diese Zustimmung erst nach einer gewissen Nutzung zu erteilen. Das reichte bei der Arbeit mit AVMultimedia meist aus, um einige Tage die Suchmaschine sporadisch und sehr selektiv zu nutzen oder verlinkte YouTube-Videos aufzurufen. Die voreingestellte Suchmaschine bei AVMultimedia verweist auf DuckDuckGo. Dabei werden keine Cookies gespeichert. Und ja, selbstverständlich können auch andere Anbieter zum Zuge kommen.

Namentlich erwähnt seien hier die französische Suchmaschine [qwant.com](https://www.qwant.com) sowie die Schweizer Lösung [swisscows.ch](https://www.swisscows.ch). Beide liefern gute Treffer.



Leider erfordert der Suchgigant seit einigen Tagen immer eine Zustimmung, nur schon um überhaupt die Seite aufrufen zu dürfen. Wer dabei z.B. keine personenbezogene Werbung eingeblendet erhalten möchte, muss sich mühsam durch viele Seiten klicken, das einleitende Video veranschaulicht dies sehr gut. Kurz und gut, für ein paar gelegentliche YouTube-Videos, die sich kaum beim Surfen vermeiden lassen, wird der Klickaufwand mittlerweile unerträglich.

SearX, die dezentrale Alternative

Bei der Suche nach einer Lösung, diese «Klickorgien» zu vermeiden, trat die dezentrale Open Source Meta-Suchmaschine SearX in Erscheinung. Dezentral bedeutet, es gibt viele Instanzen von SearX, eine Liste findet sich unter searx.me. Dort kann aus einer Liste von über 100 Servern ein Anbieter

herausgesucht werden, um mit der Suchmaschine zu arbeiten. Nun ist SearX Open Source, d.h. der Quellcode ist offengelegt, und so wurde versucht SearX lokal zu installieren. Unter Linux kann dabei der Befehl **'pip3 install searx'** verwendet werden. Anschliessend muss **'searx-run'** gestartet werden. Wow, so einfach steht damit eine lokale Instanz von SearX auf dem eigenen Linux-Rechner zur Verfügung.



Geklärt muss an dieser Stelle noch der Begriff Meta-Suchmaschine. Würde SearX eigene Such-Verzeichnisse anlegen, müsste sehr viel Speicher und Zeit aufgewendet werden, um diese zu pflegen. Meta-Suchmaschinen wie SearX erstellen keine eigenen Verzeichnisse, vielmehr suchen sie über Dritte (z.B. Bing, DuckDuckGo und viele mehr, selbst Google ist dabei).

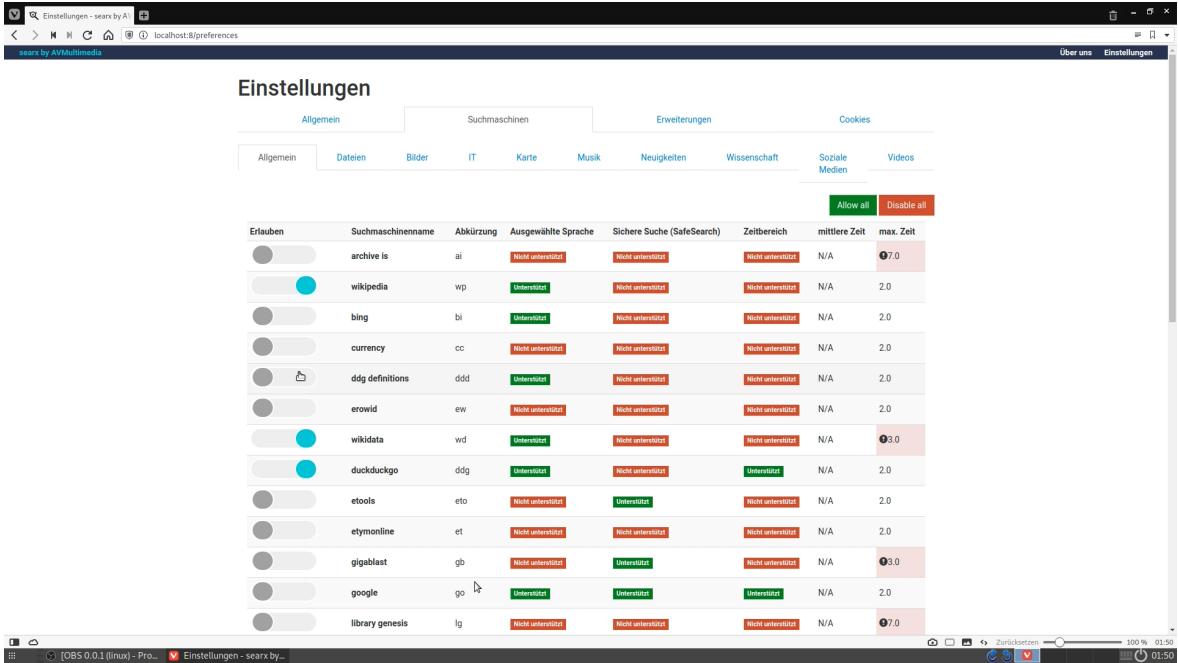
SearX lokal bei AVMultimedia und der ArchivistaBox

Auf den ersten Blick erscheint der Nutzen von SearX marginal, was sollte ein Dienst bringen, der in anderen Suchmaschinen sucht und Ergebnisse präsentiert? Wer jedoch einmal mit SearX arbeitet, wird SearX sehr schnell ins Herz schliessen. Dies vor allem deshalb, weil SearX sehr einfach aufgebaut ist und die Resultate sehr übersichtlich darstellt. Zwei Punkte, die bei vielen Suchmaschinen heute eher ein Fremdwort zu sein scheint.

SearX erledigt die Arbeit derart gut, dass sofort und ultimativ der Entschluss reifte, AVMultimedia eine lokale SearX-Instanz zu spendieren. Gesagt getan, ab heutigem Release lässt sich SearX auf AVMultimedia wie auf der ArchivistaBox mit „**localhost:8**“ aufrufen. Ebenfalls direkt SearX wird aufgerufen, wer beim Vivaldi-Browser eine Suchabfrage direkt in der Adresszeile eingibt. Bei der ArchivistaBox besteht zusätzlich die Möglichkeit, SearX im Intranet zu betreiben oder SearX zu deaktivieren. Dazu gibt es auf dem Desktop bei „**ArchivistaSetup**“ den neuen Eintrag „**SearX konfigurieren**“. Wird bei der ArchivistaBox SearX als Web-Dienst betrieben, kann die Abfrage direkt mit der IP-Adresse oder dem DNS-Namen (sofern verfügbar) und dem Zusatz `/search` oder `/searx` erfolgen. **Beispiel: <http://192.168.0.177/search>**

Mit SearX Suchmaschinen vergleichen

Bei AVMultimedia bzw. der ArchivistaBox sind im Standard DuckDuckGo und Wikipedia voreingestellt. Über Einstellungen und Suchmaschinen kann sehr effizient festgelegt werden, welche Dienste zum Einsatz kommen sollen. Dabei lassen sich mehrere Anbieter miteinander verknüpfen. Wer z.B. gleichzeitig mit Bing und DuckDuckGo arbeiten möchte, kein Problem.



The screenshot shows the 'Einstellungen' (Settings) page of the SearX application. The main title is 'Einstellungen' (Settings). Below it, there are several tabs: 'Allgemein' (General), 'Suchmaschinen' (Search Engines), 'Erweiterungen' (Extensions), and 'Cookies'. The 'Allgemein' tab is selected. Under 'Allgemein', there are sub-tabs: 'Allgemein', 'Dateien', 'Bilder', 'IT', 'Karte', 'Musik', 'Neuigkeiten', 'Wissenschaft', 'Soziale Medien', and 'Videos'. The 'IT' tab is selected. The main content area displays a table of search engines with their names, abbreviations, and various configuration options. The table includes columns for 'Erlauben' (Allow), 'Suchmaschinenname' (Search Engine Name), 'Abkürzung' (Abbreviation), 'Ausgewählte Sprache' (Selected Language), 'Sichere Suche (SafeSearch)', 'Zeitbereich' (Time Range), 'mittlere Zeit' (Average Time), and 'max. Zeit' (Max Time). Buttons for 'Allow all' and 'Disable all' are located at the top right of the table. The table lists the following search engines:

Erlauben	Suchmaschinenname	Abkürzung	Ausgewählte Sprache	Sichere Suche (SafeSearch)	Zeitbereich	mittlere Zeit	max. Zeit
<input checked="" type="checkbox"/>	archive is	ai	Nicht unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	7.0
<input checked="" type="checkbox"/>	wikipedia	wp	Unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	bing	bi	Unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	currency	cc	Nicht unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	ddg definitions	ddd	Unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	erowid	ew	Nicht unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	wikidata	wd	Unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	3.0
<input checked="" type="checkbox"/>	duckduckgo	ddg	Unterstützt	Nicht unterstützt	Unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	etools	eto	Nicht unterstützt	Unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	etymonline	et	Nicht unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	gigablest	gb	Nicht unterstützt	Unterstützt	Nicht unterstützt	N/A	3.0
<input checked="" type="checkbox"/>	google	go	Unterstützt	Unterstützt	Unterstützt	N/A	2.0
<input checked="" type="checkbox"/>	library genesis	lg	Nicht unterstützt	Nicht unterstützt	Nicht unterstützt	N/A	7.0

Ebenso einfach ist es herauszufinden, wie gut die verschiedenen Suchmaschinen arbeiten. Und dabei gab es Überraschungen. Der Suchprimat schneidet längst nicht immer am besten ab. Letztlich spuckt der Gigant wohl primär jene Homepages aus, welche die Optimierung der Schlüsselwörter (Keywords) am besten implementiert haben.

Warum scheitert Google bei ‚Island einmal um die Insel‘?

An einem konkreten Beispiel soll die Problematik veranschaulicht werden. Einleitend sei dazu gesagt, dass die Familie des Geschäftsführers im Sommer über drei Wochen die Insel Island umrundete. Daraus ist ein nicht ganz kleiner Film entstanden. Natürlich darf über Inhalt und Qualität des Werkes gestritten werden, wer sich seine eigene Meinung bilden möchte, findet die Seite unter <https://azurgo.ch/aktuell/island-einmal-um-die-insel> (Film mit 72 Minuten Länge, mit oder ohne Untertitel).

Spannend wurde es, als der Island-Film auf einer Drittseite empfohlen und bei Google gelistet wurde, nicht aber die Seite mit dem Island-Film selber. Bei den anderen Anbietern erschien die Seite mit gleicher Abfrage längst auf der ersten Seite.



Ebenfalls erhellend ist, wenn beim «Suchmeister» die Abfrage mit Video verfeinert wird. Es tauchen viele Videos auf, die mit „Island, einmal um die Insel“ oft nicht viel gemein haben. Meist geht es darum, (meist teure) Reise anzupreisen. Oft sind es Angebote zu Island, nicht selten aber gibt es auch Angebote all over the world. Zwar lassen sich interessante Inhalte finden, nur gibt es dabei viel Werbung und einen hohen Klickaufwand.

Schon erstaunlich, da liefern kleine Anbieter wie DuckDuckGo, swisscows.ch oder quant.com erstaunlich gute Resultate. Warum ist dies beim Suchriesen oft seltener der Fall? Könnte es nicht sein, dass es bei 90% Marktanteil ganz einfach zu lukrativ ist, die Suchenden etwas länger bei Stange zu halten als notwendig? Was hat Google davon, wenn bei guten Treffern schnell weitergeklickt wird?



Die Macht der Konformität durchbrechen

Weit einschneidender ist die Macht des Giganten für jene, die Inhalte anbieten. Nicht offengelegte Mechanismen eines Anbieters entscheiden zu 90% über Sein oder Nicht Sein. Zwar bietet der Suchgigant an, die Inhalte spezifisch

aufzubereiten, doch ohne Fehl und Tadel ist dies nicht.

Wenn es nur noch darum geht, etwas suchmaschinengerecht für einen Anbieter abzufassen, dann bleiben die Inhalte auf der Strecke. Das geht mittlerweile soweit, dass bei Webseiten-Lösungen Tools mahnen, noch ein paar Sätze zu schreiben, weil die Inhalte ansonsten nicht gelistet würden.

Umso spezifischer die Inhalte aber für einen Anbieter aufbereitet werden, umso weniger kann dabei Vielfalt entstehen. Und daher ist es Zeit, dem Suchgiganten die rote Karte zu zeigen. Für diese Homepage heisst dies, dass die Inhalte nicht länger «googlekonform» aufbereitet werden. Für AVMultimedia und die ArchivistaBox bedeutet dies, dass mit der neuen lokalen Suchmaschine SearX ein neues Zeitalter beim Suchen im Netz beginnt.