

Vom Umgang mit Suchmaschinen

4

Ein Großteil der Kollegen verwechselt tatsächlich die digitale Recherche mit dem Anwerfen einer Suchmaschine. Und bei den Suchmaschinen wiederum meint eine weitere Mehrheit der Kollegen, dass die Nutzung von Google nicht nur angesagt, sondern geradezu alternativlos sei. Dieser Trend hat sich noch verstärkt, seit Google in Medienhäusern und Einrichtungen der journalistischen Aus- und Fortbildung kostenlos Recherche-Seminare und entsprechende Webinare durchführt.

Dort werden natürlich die von Google bereitgestellten digitalen Recherche-Tools vorgestellt. Und hier steht die Suchmaschine im Mittelpunkt. Denn für Google laufen solche „Medienseminare“ natürlich unter der Kostenstelle „Marketing“. Sie wollen möglichst viele Journalisten auf ihre Seiten mit den hübsch aufgemachten Recherche-Tools ziehen. Das bringt Traffic und Daten. Und mit der Weiterverarbeitung und -vermarktung dieser Daten verdient Google natürlich Geld.

Dabei gibt es Alternativen zu Google Und das sollten nicht unbedingt die Suchmaschinen von Microsoft namens Bing oder Yahoo als Teil von Oath sein. Yahoo verlangt sogar für die Nutzung der Suchmaschine Zugriff auf das Gerät und erhebt Standortdaten. Bing von Microsoft verfährt ähnlich wie Google und erstellt aus den Suchdaten Profile. Diese Profile sind der Werbeindustrie lieb und teuer.

Dagegen verzichten Suchmaschinen wie Fireball, Metacrawler, Meta-Spinner, Search oder Duckduckgo auf eben eine solche Profilbildung. Sie löschen die Anfragedaten einfach nach Beendigung der Sitzung. Das bedeutet allerdings keineswegs, dass man mit diesen Suchmaschinen anonym arbeiten kann. Spätestens wenn ein Link zu einem Suchergebnis aufgerufen wird, werden

personenbezogene Daten wie die Internet-Protokolladresse weitergeleitet. Lediglich die beiden Suchmaschinen Startpage (früher: ixquick) und Metager verzichten darauf und erlauben einen anonymen Aufruf der Treffer.

Neben diesen Universal-Suchmaschinen sind auch noch einige Maschinen für Sondersuchen verfügbar. Nicht nur Hörfunk-Journalisten schätzen spaactor.com, eine Suchmaschine für das gesprochene Wort. Sie bietet schon mit der Suchmaske einen gewissen Recherchekomfort, lassen sich doch hier Sprache, Suchzeitraum und die zu durchsuchenden Kanäle direkt angeben und einstellen. Bei den Universal-Suchmaschinen müssen dafür Operatoren eingesetzt werden.

Unter 4chansearch.com lässt sich das Bulletinboard 4chan durchsuchen Das ist ganz hilfreich, weil die auf 4chan geführten politischen und technischen Diskussionen meist sehr ausufernd sind, und die Suchmaschinen so das Lesen vieler Threads erspart. Bei der Recherche, was nun eigentlich genau hinter dem Entzug der Akkreditierung von 32 Journalisten auf dem G20-Gipfel im Jahr 2017 in Hamburg steckt, hat die zielführende Suche auf dem Board 4chan sehr geholfen. Denn dort haben sehr viele extrem gut informierte Aktivisten ihr Wissen über politische und Verwaltungshintergründe rund um den G20-Gipfel geteilt. Das hatte auch damit zu tun, dass Absprachen über Demonstrationen, Aktionen einzelner zivilgesellschaftlicher Gruppen hier geteilt und teilweise über das Board gesteuert wurden.

Und so war, kurz nachdem 32 Journalisten die Akkreditierung entzogen war, auch sofort eine sehr lebhaft debattierte Diskussion über Namensverwechslungen, falsche und illegale Einträge in Sicherheitsdateien und massive methodische Fehler bei der Übertragung von polizeilichen Feststellungen vor Ort in einzelne Sicherheitsdateien auf dem Board zu beobachten.

Solche Fehler haben letztlich dazu geführt, dass Polizisten insgesamt 32 Journalisten am Eingang zum Medienzentrum in Hamburg unzulässig die Akkreditierung abgenommen haben. So wurde dem NDR-Journalisten Christian Wolf die Akkreditierung aufgrund einer sogenannten „Open-Source-Intelligence“ entzogen. Die Verfassungsschützer durchsuchen ständig alle möglichen sozialen Netze. In verschiedenen Blogbeiträgen und auch Facebook haben sie daraufhin den Namen Christian Wolf mit Bezügen zu den sog. „Reichsbürgern“ identifiziert. Dieses Material aus Social Media Plattformen ist dann einfach ohne weitere Quellenkritik in eine Sicherheitsdatei eingespeist worden. Und so wurde durch einen technisch bedingten Fehler bei der Dateneinspeisung aus dem NDR-Journalisten Christian Wolf der Reichsbürger Wolf. Das BKA hat sich dann später für diese „Namensverwechslung“ entschuldigt.

Nicht selten landen in dieser Straftäterdaten die Namen von völlig unbescholtenen Bürgern, die niemals wegen einer Straftat vor einem Gericht standen und gegen die auch kein Ermittlungsverfahren läuft. Dem liegt ein methodischer Fehler zugrunde.

Zum einen werden sogar Anmelder von Demonstrationen und sogenannte „Prüffälle“ in diese Straftäterdatei aufgenommen. Zum anderen werden auch Namen aus Projektdateien übernommen, die das Bundesamt für Verfassungsschutz gemeinsam mit dem Bundeskriminalamt führt.

Ausgangspunkt für sehr umfassende Recherchen waren die auf 4chan geführten Diskussionen darüber. Die wurden dann durch Auswertungen von Bundestagsdrucksachen, Prüfberichten und anderen Verwaltungsdokumenten ergänzt, die über die Parlamentarischen Informationssysteme der Landtage und des Bundestages getätigt wurden.

Den Streit, ob diese Parlamentarischen Informationssysteme nun schon zum sogenannten Deep Web gezählt werden, weil die Crawler und Robots der konventionellen Suchmaschinen sie nicht oder nur bruchstückhaft auswerten, will ich hier ganz bewusst nicht führen. Es sei nur darauf verwiesen, dass jeder Landtag hier einen Dokumentenschatz bereitstellt, der für die journalistische Arbeit unentbehrlich ist, aber von zu vielen Kollegen zu selten genutzt wird.

Bei der Faktenprüfung müssen oft ergänzende Videos nach Geodaten gesucht werden, um bestätigendes oder widerlegendes Material zu erhalten. Dabei hilft die von Google bereitgestellte Suchmaschine für Youtube-Videos, die nach Geodaten aufschlüsselt namens montage.meedan.com. Nutzer müssen sich dort mit ihrem Google-Account dort anmelden.

Die meisten Suchmaschinen arbeiten mit Boolescher Algebra Das hört sich komplizierter an, als es im Alltag ist. Werden Suchbegriffe mit dem Operator „and“ verbunden, so suchen die Algorithmen nach allen diesen Wörtern. Alle Suchbegriffe müssen also in einem Dokument oder auf einer Seite enthalten sein, damit sie als Treffer angezeigt wird. Dadurch lässt sich vermeiden, von einer zu großen Trefferzahl regelrecht „erschlagen“ zu werden. Werden Suchbegriffe mit dem Operator „or“ verbunden, reicht es, wenn das eine oder das andere Suchwort gefunden werden, um einen Treffer anzuzeigen.

Mitunter möchte man auch einen Begriff auf einer bestimmten Website suchen, die aber keine eigene Suchroutine vorhält. Auch das machen Suchmaschinen. Die zusätzliche Angabe „site:welchering.de“ sucht einen Begriff nur auf der Seite www.welchering.de. Ähnlich lässt sich die Suche auf bestimmte Dateiformate (nur pdf oder nur mp4 usw.) einschränken, wenn der Filetype-Operator gesetzt wird. Filetype:xls würde demzufolge nur Dateien mit