

Art Education Research No. 23/2023

Shusha Niederberger

Frag das Internet - Digitale Wissensordnungen anhand einer kurzen Geschichte der Suchmaschinen.

Wie Fragen gestellt werden, beeinflusst was passiert: Kommen die Gedanken in Fluss wie bei einer Inspiration, oder bricht Panik aus, wie etwa bei einer Prüfung? Die Art der Frage und ihren Kontext bestimmt mit, wie Wissen in Erscheinung tritt. In der Vermittlung wissen wir von den komplexen Wegen, auf denen Wissen sich artikuliert, entsteht, aufblitzt, Eindruck macht, begeistert oder aber auch umkämpft ist. Wissen ist nichts, was man auf Knopfdruck abfragen oder übertragen kann. Dies ist aber die Vorstellung von Suchen im Internet: Wir fragen Google und Google liefert uns die relevanten Infos. Wir haben uns so an diesen automatisierten Prozess gewöhnt, dass er uns selbstverständlich und unproblematisch erscheint. Aber auch technologische Systeme haben ihre Bedingtheiten. Dieser Artikel erläutert die Wissensordnungen von Suchen im Internet, stellt Grundbegriffe des kritischen Technologiediskurses vor und stellt diese in ein Verhältnis zu einem institutionskritischen Vermittlungsbegriff.

Suchen und Finden: Datenordnungen

Beginnen wir mit einem Vergleich: Genauso normal wie die Suche im Internet heute war vor 30 Jahren die telefonische Auskunft. Interessant ist dieser Vergleich, weil es sich bei der telefonischen Auskunft auch um ein technologisches System handelte, das aber ungleich einfacher war als die Suche im Internet. Bei der Auskunft konnte man anrufen und nach der Nummer von irgendjemandem von irgendwo fragen. Dafür musste man nur einen Namen und einen Ort nennen, denn die Person hinter der Auskunft verfügte über das schweizweite Telefonbuch und den Zugang zum Telefonsystem selber, um ,falls gewünscht, gleich eine Verbindung herzustellen. Im Prinzip ist es dasselbe mit Suchmaschinen heute: Wir haben mit dem Internet einen Bestand verfügbarer Informationen (das Telefonbuch), und mit der Suchmaschine eine Schnittstelle (die Auskunft), über die wir ihn abfragen können, und bekommen mit der Liste an Ergebnissen eine Ausgabe (die Verbindung), die uns zur gewünschten Information weiterleitet.

Natürlich ist das nicht so simpel und es war auch beim Telefonbuch nicht simpel. Der Bestand an Information ist eine Datensammlung, die bestimmte (und nur bestimmte) Informationen in einer spe-

zifischen Ordnung beinhaltet. Diese Ordnung ist nur teilweise durch die enthaltenen Informationen bestimmt, sondern zu einem erheblichen Teil auch von Gegebenheiten ausserhalb des Systems geprägt. Das Telefonbuch von vor 30 Jahren verzeichnete nur Festnetzanschlüsse (das Handy war noch nicht erfunden), und so waren die Anschlüsse nicht an Personen, sondern an Institutionen vergeben: Firmen und Haushalte. So war es zwar üblich, zur Differenzierung gleicher Namen den Beruf der Inhaber*innen der Telefonnummer anzugeben, aber geordnet waren die Anschlüsse nach Namen. Bei Verheirateten konnte der Namen der Frau hintenangestellt werden, Kinder wurden gar nicht aufgeführt. Diese Logik der Daten bildet eine soziale Ordnung ab, sie ist nicht neutral, und darum kann sie auch nicht den Bedürfnissen aller potentiellen Suchenden entsprechen. So war es mittels Telefonbuch unmöglich, die Nummer einer alten Bekannten zu finden, wenn diese in der Zwischenzeit geheiratet hatte – wer jemals eine Klassenzusammenkunft organisiert hat, kennt dieses Problem. Dies ist ein einfaches Beispiel dessen, was wir heute unter Bias verstehen: eine Art Verformung der Daten, die gewisse Aspekte bevorzugt, und andere benachteiligt. Es gibt unterschiedliche Biases. Das hier genannte ist ein social bias, das heisst, die Verformungen entstehen über die Reproduktion von sozialen Normen in Datenordnungen. Denn es ist ja nicht so, dass die Mädchennamen der verheirateten Frauen grundsätzlich unterschlagen würden, sie sind tatsächlich Teil der Daten, aber die Ordnung der Daten macht sie für eine Suche unzugänglich. Deshalb ist es wichtig zu unterscheiden zwischen Daten (die Informationen an und für sich) und Datenstrukturen (die Art, wie Daten in konkreten Systemen geordnet sind). Dieses einfache Beispiel des Telefonbuches zeigt, wie sehr sogar relativ einfache technologische Systeme von sozialen Ordnungen geprägt sind, und wie diese Ordnungen die Möglichkeiten von Wissen prägen.

Heute haben sich die Möglichkeiten von Auskunft wesentlich verbreitert: Wir können das Internet nach allen möglichen Auskünften durchsuchen, die nichts mit Telefonnummern zu tun haben brauchen. Auch wenn das Prinzip dasselbe geblieben ist, so ist es trotzdem einiges komplizierter geworden. Schauen wir uns die Geschichte der Suchmaschinen an.

Eine kurze Geschichte der Suchmaschinen

Suchmaschinen gibt es noch nicht sehr lange. Die ersten tauchten nach der Popularisierung des Internets um die Mitte der 1990er Jahre auf. Sie waren langsam und fanden nur das, was buchstäblich (und korrekt geschrieben) gesucht wurde. Dies bedeutete, dass User*innen sehr genau wissen mussten, wonach sie suchten – am besten gleich den Namen der Webseite. Eine inhaltliche Suche war noch nicht möglich. Und doch waren diese ersten Suchmaschinen ein grosser Durchbruch, denn davor war die Orientierung im Internet als Katalog organisiert, ähnlich wie ein Bibliothekskatalog – eine thematisch und hierarchisch geordnete Liste von Webseiten. Wer eine neue Seite gebaut hatte, schickte den Link an das Verzeichnis mit den Angaben, in welchem thematischen Ordner sie abgelegt werden sollte. Aber da das Internet ab Mitte 1990er Jahre sehr schnell wuchs, so wuchs auch die Menge an Webseiten in den Verzeichnissen, und das Ganze wurde sehr schnell sehr unübersichtlich. Um dieser Unübersichtlichkeit zu begegnen, fingen die Betreiber der Verzeichnisse an, von ihnen als unwichtig gehaltene Webseiten aus den Verzeichnissen zu entfernen. Der Abdeckungsgrad der Verzeichnisse war deshalb auch eher klein (unter 20%), was bedeutete, dass grosse Bereiche des Internets gar nicht gefunden werden konnten, wenn man nicht schon von ihnen wusste, oder über eine direkte Verlinkung dahin kam.

Wissen was da ist: Spider

Die erste Suchmaschine, die diesem kontinuierlich wachsenden Korpus gerecht wurde, war Altavista. Sie war schnell, hatte einen hohen Abdeckungsgrad, ermöglichte inhaltliches Suchen und sogar die logische Verknüpfung von Stichworten in der Abfrage. Die grosse Erfindung hinter dieser neuen Generation von Suchmaschinen war die automatisierte Erfassung von Webseiten und deren Inhalten durch sogenannte Spiders. Spiders sind Programme, die im Web herumreisen (sie folgen einfach allen Links) und alles, was sie finden, in einer Datenbank ablegen, die dann die Ausgangslage für die Suchmaschine bildet. Das ist in heutigen Suchmaschinen wie Google noch immer so: Wenn wir das Internet fragen, fragen wir also eigentlich die zusammengespidernten Informationen in der Datenbank der Suchmaschine. Das Spidern, also die Erfassung der Inhalte des Internets, und die Suchanfragen sind zwei verschiedene Prozesse, die unabhängig voneinander stattfinden. Darum kann es sein, dass

¹ Um vergangene Daten im Internet zu finden kann man eine Art digitaler Zeitreise mit der way-back-machine versuchen: <http://web.archive.org/>

man über die Suche auf Inhalte verwiesen wird, die schon nicht mehr online sind¹.

Wissen was wichtig ist: Page Ranking

Aber das Problem der immer grösser werdenden Datenmengen wird mit Spiders nicht gelöst. Suchen im Internet war bis weit nach 2000 eine heute unvorstellbar anstrengende Sache, weil seitenweise Links geprüft werden mussten, bis man die für sich relevanten Informationen fand. Die Ordnung der ausgegebenen Daten war undurchsichtig und schwierig zu navigieren.

Die Grösse und uferlose Vielfältigkeit des Bestandes an verfügbarer Information im Internet bringt neue Probleme: Wie kann in dieser unfassbar grossen Datenmenge von den heute ungefähr 2,000,000,000 Webseiten (von denen einige eigene Universen von Informationen sind, z.B. Social Media-Plattformen oder Zeitungen) etwas von Relevanz gefunden werden? Während bei der telefonischen Auskunft die Anzahl gefundener Anschlüsse sogar für grosse Städte überschaubar war, produziert die Suche im Internet heute eine unabsehbare Menge von gültigen Treffern. Neben der verfügbaren Information kommt bei der Suche im Internet also noch etwas Neues hinzu: die Sortierung gefundener Treffer nach Relevanz in der Ausgabe. Datenordnung ist nun etwas, das nicht nur im Bestand stattfindet (wie beim Telefonbuch), sondern auch in der Ausgabe.

Aber wie kann Relevanz technologisch bestimmt werden? Ist Relevanz denn nicht etwas Persönliches, in der Person des Fragenden zu Verortendes? Google hat die Sortierung von Ergebnissen nach Relevanz nicht erfunden, aber früh konsequent entwickelt. Inzwischen haben sie das Prinzip zu einem unglaublich einträglichen Geschäftsfeld entwickelt und dominieren das Feld der Suchmaschinen mit einem Marktanteil von 90%. Die Idee des berühmten Page Rank-Algorithmus von Google wurde von den zwei Gründern vor der Firmengründung in 1996 entwickelt und 1997 patentiert. Das Prinzip ist: Je mehr Links auf eine Webseite zeigen, desto relevanter ist sie. Und je höher die Relevanz einer Seite ist, desto mehr Gewicht gibt sie an die von ihr verlinkte Seite weiter. Aufgrund dieser Berechnungen wird die Ausgabe der Suche dann geordnet. Dies bedeutet nun, dass beim Spidern des Internets nicht nur Inhalte erfasst werden, sondern auch die Verlinkungsstruktur analysiert und die Gewichtung abgelegt wird. Die Datenordnung des Bestandes ist damit nochmal etwas komplexer geworden. Dieser Algorithmus wurde

seither kontinuierlich weiterentwickelt (ausserdem ist das Patent für den ursprünglichen Page Rank-Algorithmus 2012 ausgelaufen) und heute berücksichtigt Google nach eigener Auskunft rund 200 Faktoren, um die Ausgabeordnung der Suchanfrage zu bestimmen. Da Algorithmen wie diese zu den Geschäftsgeheimnissen zählen, ist es nicht möglich, diese im Detail nachzuvollziehen.

Die Berechnung von Relevanz kompliziert nun die Wissensordnung nochmal. Wir haben nicht nur eine Ordnung der Daten, die über ihre Erfassung (das Spidern) strukturiert ist, wir haben dazu eine neue Ordnung auch auf der Seite der Ausgabe. Und auch diese Ordnung von Information stellt eine sehr spezifische Ordnung her, die nicht neutral ist, also ebenfalls von Verformungen unterschiedlicher Art geprägt ist. Das Geschäftsfeld von Relevanz ist es, Orientierung in einem unüberschaubaren Masse von Information zu schaffen. Es ist wichtig zu verstehen, dass dies heute absolut notwendig ist: Ohne Algorithmen, die Information für uns ordnen, wären wir verloren im Internet (Stalder: 2016). Allerdings ist es genauso wichtig zu verstehen, dass algorithmische Systeme Relevanz nicht nur abbilden (indem sie sie modellieren versuchen), sondern sie gleichzeitig auch herstellen über die Gewährung bzw. Verweigerung von Sichtbarkeit (Noble: 2018). Diese Politik von Sichtbarkeit durch Relevanzordnungen betrifft nicht nur Suchmaschinen, sondern auch Social Media und alle anderen Plattformen, auf denen Empfehlungen und Angebote algorithmisch geordnet werden, wie z.B Spotify oder AirBnB. Man spricht von digitaler Diskriminierung, wenn z.B. Angebote und Geschäfte von Schwarzen Menschen grundsätzlich tiefere Gewichtungen zugewiesen bekommen als die ihrer Weissen Kolleg*innen, oder weiblich gelesene Personen in stereotypisierender Weise dargestellt sind. Diese Diskriminierung ist nicht nur symbolisch, sie hat konkrete Auswirkungen: Wenn z.B. Angebote nicht in den ersten 10 Suchergebnissen erscheinen, werden sie kaum gefunden, sie existieren darum de facto nicht.

Wissen wer wissen will: Tracking

Nun wird es aber nochmal komplizierter, denn Google erfasst und analysiert nicht nur die Inhalte, also den Bestand an Information im Internet und seine Struktur, sondern auch die User*innen mittels sogenannter Tracker. Tracking passiert auf vielen Wegen (und wird nicht nur von Google gemacht): über Cookies im Browser, welche das Surfverhalten aufzeichnen und weitergeben, aber auch über Mobile-Apps, die den Standort der User*innen erfassen (z.B. Goo-

gle Maps) und sogar Dienstleistungen wie die Suche selber ermöglichen es über das Aufzeichnen von Suchanfragen, Informationen über Interessen und die gegenwärtigen Umstände der User*innen zu bekommen. Was genau in welcher Form erfasst und an wen weitergeleitet wird und wozu diese Informationen verwendet werden ist unklar, extrem schwierig herauszufinden und wird erst langsam, aber zunehmend Gegenstand von Regulierungsversuchen. Gegenwärtig ist bei der Suche also neben dem Informationsbestand selber und den Ergebnissen noch ein dritter Bestand an Informationen aktiv: Information über den*die User*in selbst. Diese Informationen kommen z.T. über die Suche selber zustande, aber werden auch aus anderen Bereichen gewonnen, z.B. anderen Services desselben Konzerns. G-Mail z.B. ermöglicht mindestens das Erfassen von sozialen Verbindungen über Kommunikationsstrukturen, ungeklärt ist, ob die Inhalte der Nachrichten ebenfalls ausgewertet werden.

Im Internet eine Frage zu stellen, bedeutet also immer auch, Auskunft über sich selbst zu geben. Diese Informationen werden in die Berechnung der Antworten miteinbezogen. So bekommen wir, wenn wir nach etwas googeln, nicht einfach Information aus dem Internet, sondern wir bekommen diese eine Antwort, weil wir es sind, die fragen. Eine Antwort, die auf uns zugeschnitten ist, nicht nur in Bezug auf Relevanz der Information an und für sich, sondern auch in Bezug auf unsere vermuteten Lebensumstände wie z.B. unserer Kaufkraft beim dynamic pricing, oder den vermuteten Interessen und Gesinnungen bei der Empfehlung von Youtube-Videos. Die Verschaltung dieser drei Wissensbestände führt also nicht unbedingt zu einer präziseren Antwort, sondern vor allem zu einer Verengung des Ausschnittes an verfügbarer Information. Diese Verengung ist von undeckelten Interessen geprägt, die oft die Interessen der User:innen verletzen (z.B. teurere Preise und Verletzung der Privatsphäre), aber auch gesamtgesellschaftliche Effekte haben, wie die Verschärfung sozialer Ungleichheit, Desinformation und Polarisierung (O,Neil: 2017). Der Grund dafür liegt in den Geschäftsmodellen der Suchmaschinen, welche sich über den Werbemarkt finanzieren.

Wissen als Währung

Googles Mutterkonzern Alphabet betont immer wieder, dass die Erfassung von User*innendaten nur zur Verbesserung ihrer Produkte verwendet werde. Und hier genau liegt ein wichtiger Hinweis. Was ist denn eigentlich das Produkt von Google? Alphabet bietet

die Google Search Dienstleistung gratis an, neben vielen anderen Services wie Youtube, Gmail und die GoogleDocs. All diese Dienste erfordern komplexe und umfassende, global verteilte und vernetzte technologische Infrastrukturen, die unterhalten werden müssen. Wie kann so etwas gratis sein? Die Antwort ist Werbung. Denn das Businessmodell von Google ist das Geschäft mit Werbung. Google schaltet auf allen Plattformen und Diensten Werbung und diese Werbung ist ganz anders als eine Plakatkampagne im öffentlichen Raum oder eine Anzeige in einer Zeitung: Sie ist personalisiert. Personalisierte Werbung bedeutet, dass die Werbung dir angezeigt wird aufgrund dessen, was die Plattform über dich weiss. Und hier kommen den Informationen aus dem User-Tracking überragende Bedeutung zu. Aufgrund dieser Informationen werden User*innen bestimmten Gruppen zugeordnet: Von demographischen Angaben wie Alter, Geschlecht, geografischer Raum bis zu persönlichen Vorlieben, Bildungsstand, Gesundheitszustand, psychischer Verfassung, politischer Ausrichtung und so weiter und so fort. Diese Klassifizierungen sind für die User:innen unsichtbar, es gibt keine Möglichkeit zu wissen, in welche Schubladen man gesteckt wurde und es gibt keine Möglichkeiten, sich gegen unvoreilhaftige Zuweisungen zu wehren. Auf der anderen Seite aber können die Werbekunden von Google diese Kategorien (oder Schnittmengen unterschiedlicher Kategorien) als Zielgruppen kaufen. Dies nennt sich targeted advertising. Die Probleme mit diesem Geschäftsfeld sind seit dem Cambridge Analytica Skandal bekannt, denn es kann genauso gut für Desinformation und gezielte politische Propaganda benutzt werden, wie um die richtige Käuferschicht für den neuesten Turnschuh zu finden.

Für unsere Frage nach Wissenszusammenhängen in technologischen Systemen ist aber wichtig zu verstehen, dass Google seine Dienste nicht einfach mit Werbung querfinanziert. Im breiten medialen Diskurs hören wir oft das Argument, dass wir die Dienstleistungen mit unseren Daten bezahlen. Aber so einfach ist das nicht: Das Verhältnis von Datenwirtschaft und Dienstleistungen ist nicht einfach eine Transaktion zwischen getrennten Bereichen. Ganz im Gegenteil, die Dienstleistungen von Google sind auf die Erfassung und Analyse von User:innen ausgerichtet, um das Werbegeschäft immer profitabler zu machen. Neben dem schon beschriebenen Tracking von User*innen versuchen alle Plattformen in der Regel, die Aufenthaltsdauer von User*innen möglichst lange zu halten. Youtube macht das, indem es immer krassere Videos vorschlägt, Google in-

dem es immer mehr direkten Content anzeigt, um User*innen nicht über Links auf andere Seite zu verlieren (woher Google diesen Content hat, ist unklar und umstritten). Das Kalkül dahinter: Je länger User*innen sich auf einer Plattform aufhalten, desto mehr Werbung kann man ihnen zeigen, desto einträglicher fürs Geschäft.

Dies bedeutet, dass wir es hier mit einer neuen Ebene von Information zu tun haben, die nichts mehr mit der ursprünglichen Suche im Internet zu tun hat. Aus der Sichtbarkeit von User*innen, die für eine sinnvolle Organisation von Suchergebnissen verwendet wurde, ist ein eigenes Produkt geworden, das auf dem Werbemarkt verkauft wird. Mit diesem Schritt ändern sich die Verhältnisse elementar. Die Suchmaschine ist keine Dienstleistung mehr, die Zugang und Orientierung in Wissensbeständen anbietet, sondern ein Mechanismus, der Informationen für den Werbemarkt bereitstellt, um Aufmerksamkeit für gekaufte Kommunikation aller Art zu optimieren. Dies ist, was die Wirtschaftswissenschaftlerin Shoshana Zuboff als „Überwachungskapitalismus“ bezeichnet (Zuboff: 2018), nämlich eine relativ neue und umfassende ökonomische Ordnung, die nicht nur im digitalen Raum wirkmächtig ist, sondern auf einer elementaren Ebene die Möglichkeiten für unser Zusammenleben neu strukturiert, indem sie Zugang zu Ressourcen zunehmend über die Vermessung und Überwachung von Menschen organisiert. Davon sind auch Kultur- und Bildungsinstitutionen auf vielfältige Arten betroffen, von denen auf einer unmittelbaren Ebene die immer grösseren Aufwände für Evaluierung genannt werden können.

Kritische Wissenspraxen

Dieser Artikel hat mit einer Sensibilität aus der Vermittlung für Wissenszusammenhänge und ihre Performativität begonnen. Natürlich lässt sich die Analyse von Suchmaschinen nicht einfach so auf Vermittlungszusammenhänge übertragen und natürlich sind Kulturinstitutionen nicht mit Alphabet vergleichbar. Aus der Analyse der Wissensordnungen im Internet anhand der Geschichte von Suchmaschinen lassen sich aber Perspektiven auf die Wissensordnungen in der Vermittlung formulieren, die zu einer institutionskritischen Vermittlungsarbeit beitragen können. Denn auch in Kulturinstitutionen existieren Bias, die normative soziale Vorstellungen reproduzieren. Auch hier werden bestimmte Wissensbestände, -formen und -subjekte priorisiert auf Kosten anderer, die unsichtbar bleiben. Durch eine institutionskritische

Vermittlungsarbeit kann Wissen anders, inklusiver und vielfältiger in Bewegung gesetzt werden. Und dies schliesst die digitale Ebene mit ein: Eine institutionskritische Vermittlung muss auch die Verwendung digitaler Technologien miteinbeziehen. Welche Tools und Plattformen werden in der Vermittlung normalisiert? Welche Verhältnisse dadurch gestützt und aufrechterhalten? Wissensordnungen können performt werden (mit allen Freuden und Risiken, die das birgt) und ich kann mir keinen besseren Ort vorstellen dafür als die Vermittlung.

Literatur

Noble, Safiya Umoja (2018): *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York, NYU Press.

O,Neil, Cathie (2017): *Angriff der Algorithmen: Wie sie Wahlen manipulieren, Berufschancen zerstören und unsere Gesundheit gefährden*. München, Hanser.

Stalder, Felix (2016): *Kultur der Digitalität*. Berlin, Suhrkamp.

Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt / New York, Campus.

Weitere Informationen:

Alternativen zu Google: <https://www.qwant.com>, <https://duckduckgo.com/>

Digitale Diskriminierung:

<https://www.zdf.de/dokumentation/digital-empire/algorithmus-ki-programmierung-vorurteile-diskriminierung-rassismus-social-media-100.html>

<https://www.faz.net/aktuell/feuilleton/aus-dem-maschinenraum/digitale-diskriminierung-wenn-daten-buerger-verdaechtig-machen-15500391.html>

<https://www.bpb.de/shop/zeitschriften/apuz/digitale-gesellschaft-2022/505681/von-software-beton-falschen-vorhersagen-und-intelligenter-diskriminierung/>

Datenwirtschaft:

Data Brokers: Last Week Tonight with John Oliver (HBO) <https://www.youtube.com/watch?v=wqn3gR1WTcA>

Tech Monopolies:

Last Week Tonight with John Oliver (HBO) <https://www.youtube.com/watch?v=jXf04bhcjbg>

CV

Shusha Niederberger arbeitet an der Schnittstelle von Technologie, Kunst und Kultur. Sie studierte Bildende und Digitale Kunst in Zürich und Wien, und hat fürs HEK (Haus der Elektronischen Künste) Basel die Kunstvermittlung aufgebaut und geleitet (2014-2021). Seit 2015 ist sie Dozentin für Zeitgenössische Netzkulturen an der F+F Schule für Kunst und Medien in Zürich. Sie war Teil des Forschungsprojekts Creating Commons (2017-2019, IFCAR, ZHDK), das digitale künstlerische Praktiken als Commons untersuchte. Gegenwärtig forscht sie zu User Praxis in datenintensiven Umgebungen im Rahmen des Forschungsprojekts „Latent Spaces – Performing Ambiguous Data“ (IFCAR, ZHDK).