

Digitalgespräch Folge 60

Weit mehr als Technik: Suchmaschinen verstehen

Mit Dirk Lewandowski von der Hochschule für Angewandte Wissenschaften

Hamburg, 28. Januar 2025

<https://zevedi.de/digitalgespraech-060-dirk-lewandowski/>

[Der Vorspann mit Musik und Ausschnitten aus dem Gespräch beginnt.]

Marlene Görger [mg]: Herr Lewandowski, Sie sind Professor für Information Research and Information Retrieval an der Hochschule für Angewandte Wissenschaften Hamburg. Ihr Standardwerk *Suchmaschinen verstehen* ist 2023 auch in englischer Sprache erschienen.

Dirk Lewandowski [Lewandowski]: Also ich bin alt genug und mache das Spiel lang genug mit, dass ich schon alle sogenannten Google-Killer, wie es früher immer in der Presse hieß, gesehen habe, wie die alle wieder eingegangen sind. Also nehmen Sie mal eine Suchanfrage wie 'Kredit ohne Schufa' und schauen Sie sich mal die Ergebnisse an. Das sind sicher nicht die besten Ergebnisse zum Thema, sondern die, die am besten optimiert haben.

Petra Gehring [pgg]: Zu dem Punkt, dass wir die Möglichkeiten der Suchmaschinen so krass unterfordern, gibt es da irgendwie Wissensvermittlung, Suchkurse?

[Lewandowski]: Generative Sprachmodelle spielen natürlich eine große Rolle jetzt auch für die Websuche. Sie werden Suchmaschinen nicht obsolet machen, sie werden Suchmaschinen verändern und das sehen wir auch schon. Wir sollten uns Sorgen machen. Es ist eine Basistechnologie letztlich für das Web. Es ist schon sicher sinnvoll, sowas auch in Europa zu haben.

[Der Vorspann endet, das Gespräch beginnt.]

[mg]: Seit die Forschungseinrichtung CERN zu Beginn der 90er Jahre das World Wide Web der Öffentlichkeit übergeben hat, ist die Menge der Inhalte, die darin abrufbar sind, ins Unfassliche gewachsen. Ohne die Verfügbarkeit von Suchmaschinen, die für uns große Datenmengen durchforsten und relevante Webadressen zusammenstellen, wäre diese monströse Ansammlung von Informationen absurd. Die übersichtlichen Startseiten von Internetsuchmaschinen vermitteln den Eindruck von Durchschaubarkeit des World Wide Web. Und die schnelle Suche im WWW ist so unmittelbar im Browser und Betriebssysteme unserer Smartphones, Laptops und PCs integriert, dass sie fast wie der natürliche Zugang zum Netz wirkt. Wir verlassen uns darauf, dass die Qualität der Suchergebnisse zumindest akzeptabel ist. Wir haben uns an die Werbung gewöhnt und wir nehmen hin, wenn Anbieter von der Erhebung unserer Nutzerdaten profitieren. Immerhin bezahlen wir für diese Dienste, ohne die wir das Internet nicht so nutzen könnten, wie wir es tun, selbst kein Geld. Trotzdem gibt es

einen Markt der Suchmaschinen, auch wenn er von Google dominiert wird. Wie unterscheiden sich die anderen Anbieter von diesem Quasimonopolisten? Welche Geschäftsmodelle und Werte stehen jeweils dahinter und welche Funktionsweisen, Rahmenbedingungen und Abhängigkeiten sollten wir uns klar machen, wenn wir Suchmaschinen als Zugänge zum Internet nutzen? Das ist unser Thema heute im Digitalgespräch. Mein Name ist Marlene Görger. Ich bin Physikerin und Technikphilosophin und arbeite für das Zentrum verantwortungsbewusste Digitalisierung.

[pgg]: Und ich bin Petra Gehring, Professorin für Philosophie an der Technischen Universität Darmstadt. Bei uns im Podcast dürfen wir heute Prof. Dirk Lewandowski begrüßen, der Experte für unser Thema ist und uns aus Hamburg zugeschaltet. Herzlich willkommen im Digitalgespräch, Herr Lewandowski. Wir freuen uns sehr.

[Lewandowski]: Vielen Dank. Ich freue mich auch sehr. Hallo.

[mg]: Herr Lewandowski, Sie sind Professor für Information Research and Information Retrieval an der Hochschule für Angewandte Wissenschaften Hamburg. Und Sie bringen in Ihrer Arbeit einen breiten fachlichen Hintergrund ein. Sie haben Bibliothekswesen, Philosophie, Medienwissenschaften und auch Psychologie studiert, in Informationswissenschaften promoviert. Und Sie waren als ausgewiesener Experte für Suchmaschinen auf internationaler Ebene auch als Sprecher und Berater tätig. Etwa für den High Court of Justice in Großbritannien und den Deutschen Bundestag. Ihr Standardwerk *Suchmaschinen verstehen* ist 2023 auch in englischer Sprache erschienen. Und *Verstehen* bezieht sich dabei nicht nur auf technische Details und Funktionsweisen, sondern auch auf deren Bedeutung für die pluralistische Demokratie. Damit wir diese Zusammenhänge besprechen können, erklären Sie doch bitte zunächst, was Suchmaschinen im engeren Sinne sind und wie sie funktionieren.

[Lewandowski]: Das ist ja schon mal eine sehr weitreichende Frage. Ich versuche es mal mit einer kurzen Antwort. Also vielleicht fange ich mal an mit der Idee, was es bedeutet, Suchmaschinen zu verstehen und dass es eben nicht nur ein technisches Problem ist. Natürlich muss man den technischen Aufbau von Suchmaschinen verstehen, um das Gesamtbild zu verstehen. Aber es sind eben unterschiedliche Perspektiven, die wir sicher nachher auch alle besprechen werden. Die der Nutzenden, die der Suchmaschinenanbieter, von Werbetreibenden, von Inhalteanbietern und so weiter. Und ich glaube daran, dass man nur, wenn man dieses Gesamtbild versteht, wirklich verstehen kann, wie Suchmaschinen funktionieren. Das als kurze Vorrede. Jetzt vielleicht ganz kurz zum Technischen, wie das funktioniert. Also, Sie hatten es ja in der Anmoderation schon gesagt. Als allererstes ist es ja mal so, dass wir eine riesige Masse von Inhalten im Web haben. Und die sind erst mal nicht strukturiert. Jeder kann Inhalte hinzufügen. Und das machen die Leute auch fleißig. Gute Inhalte, schlechte Inhalte, vollkommenen Schrott, Lügen, was auch immer. Also, wir haben ein Qualitätslevel von unterirdisch bis ganz toll. Und die Aufgabe von Suchmaschinen ist es jetzt also erst mal, diese Inhalte alle aufzuspüren, die ganz schlechten schon mal rauszuschmeißen, und

dann die guten Inhalte oder die, die zumindest ein gewisses Qualitätslevel erreichen, verfügbar zu machen. Das passiert in erster Linie über das sogenannte Crawling. Das heißt, die Suchmaschinen folgen im Prinzip im Web den Links, die zu finden sind. Klicken die sich quasi durch, von einem bekannten Dokument oder bekannten Dokumenten aus, und finden so neue Dokumente. Und im Idealfall ist es so: Allein durch das Verfolgen der Links entdeckt man alle Dokumente im World Wide Web. Das ist jetzt sehr vereinfacht, dass das so klappt. Aber erst mal ist das das Prinzip und neue Dokumente werden eben dadurch gefunden, dass irgendwo ein Link gesetzt wird auf einer Webseite, die schon bekannt ist. Dann findet die Suchmaschine das Dokument. Und so entsteht eben die Datenbasis. Und jetzt kommt so das erste Problem, das erste große Problem ins Spiel: dass eben wahnsinnig viele Inhalte im Web sind. Wir kennen Zahlen von Google, die haben gesagt, dass sie 2020 in ihrem Datenbestand etwa 400 Milliarden Dokumente aus dem Web hatten. Also eine gigantische Zahl. Und das ist ja schon so weit gefiltert, dass der ganze Spam-Inhalt, die ganzen wirklich schlechten Sachen schon raus sind. Also, man kann sich vorstellen, dass noch weit mehr im Web vorhanden ist. Das ist die erste große Aufgabe von Suchmaschinen, bevor es überhaupt daran geht, diese Dokumente zu erschließen. Also irgendwie durchsuchbar zu machen, nach ihrer Qualität zu bewerten etc. und dann schließlich in ein Ranking zu bringen, also mit Suchanfragen abzugleichen, und daraus eine Trefferdarstellung, die in irgendeiner Weise geordnet ist, zu machen. Also das Zusammenstellen der Suchergebnisseiten, wie wir sie kennen. Das ist also der Schritt nur bis zu dem Punkt, wo eine Suchanfrage eingegeben wurde, der Antwort gegeben wurde. Also, Sie sehen schon den riesigen Aufwand hier.

[mg]: Sind denn alle Suchmaschinen so aufgebaut? Funktionieren die gleich?

[Lewandowski]: Ja, das ist grundsätzlich im Prinzip, weil es gäbe eigentlich keine andere Möglichkeit, an die Inhalte des Web überhaupt zu kommen. Stellen Sie sich vor, man würde jetzt sagen, die Leute müssen ihre Websites anmelden oder sowas. Es wäre nicht möglich, dass jeder die Websites immer anmeldet bei allen möglichen Suchmaschinen. Da würde man den Überblick verlieren, würde vielleicht wichtige Inhalte verlieren. Und der andere Weg wäre ja: Es gibt irgendwie ein zentrales Verzeichnis, aber das haben wir eben nicht. Also ein Verzeichnis, aus dem die Suchmaschinen sich bedienen könnten. Und das ist ja gerade das Kernproblem dieser Web-Inhalte, dass sie eben verstreut vorliegen und sich eben permanent ändern.

[mg]: Wie wird denn die Qualität festgestellt bei so einem Dokument, das passiert ja sicher automatisiert?

[Lewandowski]: Es muss automatisiert passieren, denn allein aufgrund der Masse wäre es überhaupt nicht möglich, da irgendwie von Hand was zu machen. Also Qualität kann ich jetzt ja nicht direkt messen von einem Dokument. Ich kann nur quasi so einen Stellvertreter einsetzen und oft ist der Stellvertreter eine Form von Popularität. Also man kann beispielsweise messen: Was wird wie häufig angeklickt? Man kann messen: Welche Dokumente werden besonders lange gelesen oder intensiv gelesen? Wir

können aber auch, und das ist so ein Klassiker bei den Suchmaschinen, wir können auf die Verlinkungsstruktur schauen: Also auf welche Dokumente wird besonders häufig verwiesen, von besonders guten Seiten aus? Also die Links werden quasi als eine Stimme für das Dokument gewertet, auf das verlinkt wird. Aber nicht jede Stimme zählt gleich viel. Also wenn Sie einen Link von meiner privaten Homepage bekommen, zählt es nichts. Wenn Sie einen Link bekommen von Zeit Online meinetwegen, dann zählt es halt sehr viel, weil es wiederum als eine sehr wichtig erkannte Website über diese Verlinkungsstrukturen erkannt wurde.

[pgg]: Das heißt, es gibt so eine Art Reputationshierarchie? Bestimmte Linkgeber haben eine sehr, sehr hohe Reputationswerte und reichen dann gewissermaßen den Wert auch weiter an die Linknehmer?

[Lewandowski]: Genau, das ist der Klassiker, quasi, der von Google in der PageRank-Verfahren erfunden wurde, dass genau so gemacht wird. Es gibt viele Varianten davon, das nochmal zu verfeinern, weil es so in der Reinform so nicht ganz funktioniert. Aber es ist so, dass eben, wie Sie es gesagt haben, Linkgeber eine unterschiedliche Wertigkeit haben. Die haben sie sich aber auch quasi durch Meriten erworben, indem sie selber viele gute Links gesammelt haben. Also, das ist quasi so ein, wie soll man sagen, ein iteratives System, was sich gegenseitig bedingt dann immer wieder.

[pgg]: Nochmal nachgefragt: Die aktuelle Nutzungsfrequenz spielt dann gar nicht so eine große Rolle, sondern wirklich die Verlinkung? Denn man könnte sich ja vorstellen, dass gerade bei Sachen, die schnell wichtig werden, vielleicht auch gut sind und schnell wichtig werden, auch die Zugriffszahl selber irgendwas aussagt und nicht nur die Verlinkung.

[Lewandowski]: Natürlich können auch die Zugriffszahlen was aussagen. Die Verlinkung, und da ist noch ein kleiner Trick drin, die muss sich nicht nur auf das einzelne Dokument beziehen, sondern die kann sich auch auf eine Quelle beziehen. Und wenn ich mir jetzt als Suchmaschine überlegen würde: Es kommt ein aktuelles Ereignis, dann würde ich mir überlegen, wenn ich jetzt Dokumente ranken muss: Naja, ich habe diese Dokumente vielleicht noch nicht gut genug analysiert, ich habe noch nicht genug Links, die auf ein bestimmtes Dokument zeigen, ich kann es eigentlich noch nicht gut bewerten. Dann würde ich auf die Quellenbewertung gehen. Also ich würde dann sagen, um bei dem Beispiel zu bleiben: Zu irgendeinem Thema, über das ich noch nichts analysieren kann, zeige ich lieber ein Dokument von Zeit Online an als irgendeins von irgendeiner anderen Website. Oder also generell von einer bekannten Nachrichtenquelle oder einer Quelle, die ich generell als vertrauenswürdig eingeschätzt habe, um eben nicht in die Gefahr zu geraten, da irgendeinen totalen Schrott anzuzeigen, was durchaus vorkommen kann bei Rankingverfahren, keine Frage. Aber eben indem man sich auf zuverlässige Quellen fokussiert, kann man die Gefahr so ein bisschen runterbiegen.

[mg]: Jetzt habe ich ja als Nutzer die Möglichkeit, einfach einen Suchbegriff in dieses Fenster einzugeben und dann bekomme ich eine Liste von allen möglichen Ergebnissen. Wenn ich jetzt sage, ich möchte gerne nur diese verlässlichen Quellen haben, habe ich auch die Möglichkeit, das einzuschränken?

[Lewandowski]: Also es gibt immer die Möglichkeit, auf eine bestimmte Quelle, auf eine bestimmte Domain einzuschränken. So eine Einschränkung, die direkt sagt 'nur verlässliche Quellen durchsuchen' gibt es nicht. Also, es ist ja auch eine der oft vorgebrachten Kritiken an den Suchmaschinen, dass sie uns eigentlich zu wenig Steuerung überlassen, wenn wir sie denn haben wollen. Was für ein Quellenspektrum wir eigentlich im Moment durchsuchen wollen.

[mg]: Jetzt sagten wir schon auch am Anfang, Google ist wirklich so der Marktführer, wenn Sie so den Markt beschreiben würden oder auch die Akteure, Anbieter von Suchmaschinen, die sonst noch eine Rolle spielen, wer ist denn da besonders zu nennen?

[Lewandowski]: Es sind eigentlich relativ wenige, und das hat damit zu tun, dass, das was ich vorhin sagte, der Aufbau des Datenbestands, also des Index, eine so große Aufgabe ist, die einfach so viele technische Mittel und auch finanzielle Mittel erfordert, dass sich kaum eine Firma das Ganze leisten kann. Also wir haben weltweit, wenn wir jetzt auf nennenswert große Suchmaschinendatenbanken gucken, haben wir eigentlich vier Anbieter. Das ist Google, das ist Microsoft mit der Suchmaschine Bing, das ist Yandex in Russland und Baidu in China. Und dann haben wir, also wenn man sich anguckt, Listen von Suchmaschinen, da gibt es dann Hunderte von verschiedenen Suchmaschinen. Da haben wir ja eigentlich eine Einteilung, wo wir sagen können, es gibt also kleine Suchmaschinen, die haben in der Regel das Problem, dass die einen Datenbestand haben, der nicht besonders groß ist und damit nicht konkurrenzfähig zu den großen vier ist. Und wir haben Suchmaschinen, die nur vermeintliche Suchmaschinen sind, die nämlich Ergebnisse von einer der großen Suchmaschinen anzeigen. Das ist bei uns in der Regel sind das Ergebnisse entweder von Google oder von Bing. Da haben wir auf der einen Seite die ganzen berühmten Namen von früher, die nur noch Suchportale sind, ob Sie Altavista, Yahoo oder wen auch immer nehmen, die zeigen alle Ergebnisse von anderen an. Also Yahoo beispielsweise hat eine Kooperation mit Microsoft und zeigt Ergebnisse von Bing an. Wenn sie zum Beispiel bei T-Online suchen oder bei Web.de, dann kriegen sie Ergebnisse von Google. Und so geht es weiter, es sind keine echten Suchmaschinen mehr. Und das betrifft auch viele Suchmaschinen, die als Alternativen auftreten und das in gewisser Weise auch sind, die Ihnen zum Beispiel besseren Datenschutz versprechen, wie Startpage, die aber dann auch Google-Ergebnisse anzeigen, oder Ecosia, die ein Geschäftsmodell haben, wo sie von ihren Gewinnen Bäume pflanzen und für die Umwelt was Gutes tun, die aber auch Ergebnisse von Bing anzeigen. Also was ich damit sagen will, ist, wir haben eigentlich die beiden großen Suchmaschinen, Google und Bing, und dann kleinere, die interessant sein können, aber nicht unbedingt konkurrenzfähig sind, und dann diese ganzen

Suchportale, die uns suggerieren, dass es wahnsinnig viel Auswahl auf dem Suchmaschinenmarkt geben würde, was aber de facto nicht der Fall ist.

[pgg]: *Das heißt, im Grunde sind es so zwei Bausteine? Das Dahinterliegende, also der eigentliche Zugriff auf die Masse der Information und dann so ein Portalbaustein, der die Präsentation und vielleicht auch noch Details in der Darstellung anbietet?*

[Lewandowski]: *Genau, also wir sprechen immer von quasi dem Index, also dem Datenbestand und den Services, die da drauf aufgebaut werden, und ein Service kann eben so ein Suchportal sein. Da geht es natürlich auch um Nutzerinteraktion, aber das ist bei Suche eigentlich auch relativ klar, wie das abläuft, zumindest standardmäßig. Also mit relativ wenig Aufwand können Sie so ein Suchportal generieren, und Sie können das auch gut finanzieren oder monetarisieren, weil Sie letztlich, wenn Sie mit Google oder Bing so eine Kooperation eingehen, dann läuft es so, dass Sie die Ergebnisse von dieser Suchmaschine angezeigt bekommen, aber auch die Werbeanzeigen, und dann teilt man den Umsatz einfach auf nach einem Schlüssel. Sowohl die gebende Suchmaschine als auch das nehmende Suchportal profitieren davon.*

[pgg]: *Jetzt haben Sie die Werbung erwähnt. Das ist ja ein wichtiges Stück der Realität. Also eine Suchmaschine ist nicht einfach nur ein Superkatalog fürs gesamte Internet sozusagen, sondern wir kriegen am Ende des Tages auch durch Werbung gepushte Mitteilungen oder auch Treffer, und die sind dann nicht einfach allein nur auf der Basis sozusagen dieser Rankingverfahren erstellt, sondern da kommt dann noch so ein Hochpunkten dazu, für das sich die Suchmaschine bezahlen lässt. Habe ich das richtig formuliert?*

[Lewandowski]: *Jein. Man muss unterscheiden zwischen zwei Formen. Das eine ist, dass wir tatsächlich die Möglichkeit haben, Anzeigen zu buchen in den Suchmaschinen, das heißt, man bucht eine Anzeige zu einem bestimmten Keyword, und wenn dieses Wort oder dieser Suchbegriff gesucht wird, dann wird die Anzeige angezeigt. Und die andere Möglichkeit ist, dass Sie Suchmaschinenoptimierung betreiben. Das klingt ein bisschen nach: Man würde die Suchmaschine optimieren. Das ist nicht der Fall, sondern man optimiert seine eigene Website dahingehend, dass sie besser in den Suchmaschinen gerankt wird, und damit können Sie, ohne, dass Sie den Suchmaschinenbetreiber bezahlen, in den Ergebnissen nach oben kommen. Wenn Sie die Anzeigen kaufen, müssen Sie halt für jeden Klick auf die Anzeige bezahlen. Und der Kern des Problems mit den Anzeigen ist eigentlich, dass viele Nutzende einfach nicht auseinanderhalten können, was Anzeigen sind und was organische Suchergebnisse sind, also die, die wirklich durch den Ranking-Algorithmus erstellt werden, und da beginnen eigentlich die Probleme.*

[pgg]: *Die sehen ja auch ganz ähnlich aus in der Darstellung. Da gibt es irgendwie so oben gelistete Sachen, die sehen aus wie ein Treffer, und dann gibt es so eine ganz zarte*

Linie, die so kenntlich macht: Bis hierhin waren es Anzeigen, und darunter kommen dann, in Anführungsstrichen, echte Treffer.

[Lewandowski]: *Wenn es überhaupt eine Linie gibt, also bei Google beispielsweise, gibt es gar keine Linie, und es ist in der Tat so, dass die Anzeigen sehr ähnlich aussehen, was sicher auch kein Zufall ist. Wir haben hier in Hamburg viele Studien dazu gemacht und haben unter anderem auch Personen mit Eye-Tracking beobachtet. Da können Sie die Blickbewegungen auf dem Bildschirm nachvollziehen, und man sieht eben deutlich, dass Leute, die nicht wissen, was die Werbung ist und was die organischen Ergebnisse sind, die beiden Trefferlisten einfach wie eine Liste von oben nach unten durchlesen und gar nicht zwischen diesen beiden Formen unterscheiden.*

[mg]: *Jetzt bin ich als Nutzerin ja schon darauf angewiesen, dass Suchmaschinen ein Stück weit gut funktionieren, in meinem Interesse funktionieren. Es sei denn, ich suche ganz gezielt nach einem ganz bestimmten Dokument oder so. Dann interessiert mich jetzt natürlich nicht so sehr, wie Google funktioniert. Aber wenn ich mir irgendwie die Inhalte im Web erschließen will, läuft ja eigentlich diese ganze Realität und Praxis der Geschäftsmodelle meinen Interessen als Suchende zuwider, oder? Ist das zu drastisch?*

[Lewandowski]: *Ja, es ist zu drastisch, würde ich sagen. Und der Grund ist, man muss sich erstmal überlegen, dass kommerzielle Suchmaschinen es ja nicht nur geschafft haben, für uns die Inhalte des Web zu erschließen und irgendwie verfügbar zu machen. Was nicht-kommerzielle Akteure nicht geschafft haben. Sondern Suchmaschinen haben es auch geschafft, dadurch, dass sie eine Möglichkeit bieten für Inhalteanbieter, dass sie da mit ihren Inhalten Geld verdienen können durch die Vermittlung über Suchmaschinen, also dass sie Traffic-Besucher bekommen und das zum Beispiel durch Werbung monetarisieren können wieder, dass ja auch Inhalte überhaupt produziert werden. Also wenn man sich den Online-Journalismus beispielsweise anguckt, der lebt ja nicht davon, dass sie Abos verkaufen. Der lebt davon, dass Werbung platziert wird. Kann man jetzt auch sagen, viele schlechte Inhalte würden nicht produziert werden, wenn es die Werbemöglichkeit und den Traffic über Suchmaschinen nicht geben würde. Aber viele gute Inhalte könnte man schlicht auch nicht monetarisieren, ohne die Besucher, die über Suchmaschinen vermittelt werden.*

[pgg]: *Diese Broker-Funktion, sagen wir jetzt mal, die eigentlich für alle wichtig ist, die überhaupt an einem gut benutzbaren und üppigen Internet interessiert sind. Sie haben das schon angedeutet: Es ist wahnsinnig aufwendig, das mit diesen unheimlichen Informationsmengen zu leisten, diese Rolle wirklich gut zu spielen. Wäre es nicht doch denkbar, dass dieser Markt sich noch anreichert, dass es irgendwie mehr werden als nur diese vier, die Sie erwähnt haben? Und wäre es nicht doch denkbar, dass vielleicht auch, ja, so eine öffentliche Suchmaschine oder eine durch Staaten finanzierte oder wie auch immer von der UNO, UNESCO, ich spinn jetzt mal so rum, also irgendwie eine werbefreie, und sei es Basissuchmaschine, mit der man abgleichen kann oder so, noch parallel läuft?*

[Lewandowski]: Also, ich glaube, die Frage braucht eigentlich zwei Antworten oder muss in zwei Teile geteilt werden. Das eine ist die Frage: Warum hat es denn bisher nicht funktioniert? Also, wir haben ja die Situation auf dem Suchmaschinenmarkt, dass wir im Prinzip seit 20 Jahren Google mit einem Marktanteil bei 90 Prozent aller Suchanfragen haben und da ändert sich nichts. Also, ich bin alt genug und mache das Spiel lang genug mit, dass ich schon alle sogenannten Google-Killer, wie es früher immer in der Presse hieß, die ganzen Ankündigungen gehört habe und gesehen habe, wie die alle wieder eingegangen sind. Und wenn man sich überlegt, Microsoft, ein Unternehmen, was sicher sehr viel Geld hat und sehr viel Geld investiert hat in seine Suchmaschine, und auch ein gutes Produkt gemacht hat, bei einem Marktanteil von knapp fünf Prozent liegt, dann ist es schon deutlich, dass da irgendwas anderes auch eine Rolle spielt. Also, wenn man mit Google spricht oder liest, was Google an Statements rausgibt, dann heißt es immer: Naja, die Leute verwenden Google, weil die Qualität so überragend ist. Und das wäre der einzige Grund. Das ist aber was, was man jetzt eigentlich so in Vergleichsstudien nicht direkt sehen kann. Also, die haben keinen Qualitätsvorteil, der diese Marktverhältnisse rechtfertigen würde. Und der zweite Punkt ist, und das ist jetzt im Rahmen von diversen Untersuchungen, Department of Justice in den USA hat gegen Google gearbeitet, in den Anhörungen nochmal deutlich zutage getreten. Man wusste es schon vorher, aber da ist es nochmal ein bisschen populärer geworden, dass Suchmaschinen wie Google enormes Geld zahlen, was als sogenannte Traffic Acquisition Cost bezeichnet wird. Das heißt, sie bezahlen Geld dafür, dass die Besucher auf ihre Suchmaschine, auf ihre Website kommen. Das sehen wir daran, dass beispielsweise Google, es gilt für andere Suchmaschinen auch, dass Google in Infrastrukturen integriert ist. Also, wenn Sie zum Beispiel einen Browser haben wie Firefox oder Browser wie Safari auf Apple-Geräten, ist Google als Suchmaschine voreingestellt. Und dafür bezahlt Google eine ganze Menge Geld. Also, man kann sagen, es geht um Milliardenbeträge hier, die Google beispielsweise an Apple bezahlt. Die Mozilla Foundation hinter Firefox wäre pleite, die würde es nicht geben, wenn Google sie nicht finanzieren würde, wenn nicht Google als Standard-Suchmaschine eingestellt wäre. Android-Mobiltelefone haben Google voreingestellt als Suchmaschine und so weiter. Also, Sie sehen, es ist nicht so leicht, dem zu entkommen, auf der einen Seite. Und auf der anderen Seite entkräftet es natürlich das Argument, das allein mit der Qualität der Suchmaschine zu tun hat, dass die Nutzenden eine Suchmaschine auswählen für ihre Suchen. Sonst bräuchte Google ja nicht diese Milliarden ausgeben. So, und das ist die Erklärung für die Marktsituation. Und jetzt zum zweiten Teil der Frage. Tut mir leid, dass das eine lange Antwort ist. Wäre es nicht gut irgendwie, wenn es eine öffentliche Suchmaschine geben würde, wer auch immer die machen würde? Grundsätzlich würde ich sagen: Ja, wäre eine schöne Sache. Das Problem ist nur, wir haben diesen enormen Aufwand, das Web zu erschließen. Und jetzt stellen Sie sich vor, Sie machen eine Suchmaschine, Sie investieren Hunderte von Millionen bis in die Milliarden rein, um so einen Suchindex aufzubauen, bauen Ihre Suchmaschine auf und die ist auch gut, aber Sie machen schlechtes Marketing. Niemand benutzt die Suchmaschine, die wird wieder eingehen. Und deswegen glaube ich nicht daran, an die eine alternative Suchmaschine. Also, es gab ja verschiedene Vorschläge, so etwas zu machen. Eine staatliche Suchmaschine, eine öffentlich-rechtliche Suchmaschine

orientiert an dem Prinzip vom öffentlich-rechtlichen Rundfunk, solche Dinge. Aber immer entsteht eine Suchmaschine daraus. Mein Vorschlag, den ich schon vor vielen Jahren gemacht habe, geht in eine andere Richtung. Und da ist nochmal diese Trennung zwischen dem Index und den Services wichtig. Mein Vorschlag war zu sagen, man macht eine öffentliche Infrastruktur für den Index und erlaubt allen, ihre Services darauf aufzusetzen. Also nicht eine Suchmaschine, sondern tausend Suchmaschinen. Jeder kann mitmachen und durch so einen offenen Webindex, also dass quasi die Infrastruktur öffentlich finanziert ist und darauf aufsetzend dann die Services kommerziell/nichtkommerziell gemacht werden können, dadurch könnte wieder Wettbewerb entstehen.

[p99]: Also an guter Werbung alleine kann es ja schon deswegen nicht hängen, weil Sie ja eben auch geschildert haben, da sind handfeste Verträge, da wird gezahlt, sozusagen für den Einbau der Suchmaschine in Infrastrukturen. Das müsste ja ein öffentliches Angebot auch irgendwie machen. Also um diesen Mechanismus zu beseitigen, dass in den Geräten gewissermaßen immer schon Google drin ist.

[Lewandowski]: Genau, das wäre eine Möglichkeit. Ich meine, es wäre auch eine Möglichkeit, wenn man an öffentliche Einrichtungen denkt, dass natürlich öffentliche Einrichtungen sich auch zusammenschließen können und das entsprechend gemeinsam promoten könnten. Also wenn Sie dann an die Bibliotheken denken, an die Unis und so weiter.

[mg]: Wir setzen jetzt so ein bisschen voraus, dass das ein Problem ist, dass alle Google benutzen. Warum ist es denn ein Problem oder ist es denn eins?

[Lewandowski]: Es ist aus meiner Sicht in der Hinsicht ein Problem, ohne Google anzugreifen, dass sie jetzt irgendwie einen schlechten Job machen. Also das eine ist die Kritik, aber das hat mit der Kernsuche nicht direkt etwas zu tun. Die Kritik nochmal, dass eben Werbung und organische Ergebnisse zumindest für die Nutzenden vermischt werden. Das zweite Problem ist, dass wir diese ganze Szene der Suchmaschinenoptimierer haben, was grundsätzlich erstmal auch eine gute Sache ist, dass sich Menschen damit beschäftigen, wie Websites aufbereitet werden können, dass sie in Suchmaschinen möglichst optimal gefunden werden. Aber natürlich gibt es ein Wettrennen um die ersten Plätze in Google, die sehr häufig gesehen werden und die auch sehr häufig angeklickt werden. Und das führt dazu, dass Manipulationsversuche gemacht werden, um eben bestimmte Dokumente nach vorne zu bringen. Also nehmen Sie mal eine Suchanfrage wie 'Kredit ohne Schufa' und schauen Sie sich mal die Ergebnisse an. Das sind sicher nicht die besten Ergebnisse zum Thema, sondern die, die am besten optimiert haben, die es am besten geschafft haben, aus kommerziellen Motiven da nach vorne zu kommen. Da finden Sie jetzt nicht unbedingt Informationen, wie sich Leute helfen können, die so verzweifelt sind, dass sie einen Kredit ohne Schufa-Auskunft haben wollen. Grundsätzlich ist es so, dass jede Suchmaschine ja die Inhalte des Web in irgendeiner Weise interpretiert. Also, man kann so von algorithmischen Interpretationen sprechen. Und ein Irrtum ist, zu glauben, dass

es eben eine richtige Interpretation der Inhalte geben würde. Das ist nur sehr eingeschränkt. Also wenn Sie so Suchanfragen haben, wir sprechen von navigationsorientierten Suchanfragen 'Ich möchte die Website der Uni Darmstadt finden', gebe ich Uni Darmstadt ein, da gibt es ein richtiges Ergebnis dafür. Da ist es eindeutig. Wenn Sie sich informieren wollen über irgendein Thema, 'welche Bedeutung hat Goethes Faust für die deutsche Literatur?', gibt es viele Antworten darauf. Jede Suchmaschine kann andere Antworten geben. Und das sehen wir auch in so Vergleichsstudien, die wir gemacht haben. Sie können Suchmaschinen haben, also zwei Suchmaschinen, die beide relevante Ergebnisse ausgeben. Und diese Ergebnisse können trotzdem komplett verschieden sein. Und das ist aus meiner Sicht der wesentliche Grund, warum wir eine Vielfalt von Suchmaschinen brauchen. Dass wir nicht einem Anbieter die Interpretation der Webinhalte überlassen. Wo wir auch messen können, welche Tendenzen da drin sind. Also welche Quellen besonders häufig auftauchen etc., sondern dass wir eben eine Vielfalt dieser Interpretationen haben und dadurch natürlich auch Innovation in dem Markt entsteht. Wir glauben immer, Suche muss so sein, wie Google uns das anbietet. Suche könnte auch ganz anders sein.

[pgg]: Auf der anderen Seite, wenn ich mir das jetzt so ausmale, Tausende von zumindest Portalen, die dann aufsetzen auf die Möglichkeit eines ganz breiten Zugriffs, aber alle ein bisschen verschieden angelegt und gemacht. Da ist dann wieder eine Menge zusätzliches Wissen erforderlich. Also, mit welcher Maschine frage ich denn jetzt mal dieses Thema Goethe und Deutschliteratur ab? Vielleicht gibt es eine andere Maschine, die mir viel besser über Hobbysportarten Auskunft erteilen kann. Oder wenn ich auch bestimmte Weltregionen oder Sprachen interessant finde, wäre ja denkbar, dass man dann sozusagen zweistufig sucht. Also erst mal suche ich mir die richtige Suchmaschine. Und dann suche ich mir mit dieser Suchmaschine zu dieser Frage passend das optimale Ergebnis. Das ist ja auch eine Aufgabenstellung. Ich habe die Chance des Vergleichens. Aber ich muss dann doch sozusagen zerteilen den Zugriff aufs Wissen in so Vorfragen und Überlegungen, die fast so was wie eine Methode dann erzwingen.

[Lewandowski]: Das müsste nicht sein. Denn man kann natürlich auf so einem Index, und das können Sie mit Googles Index genauso sehen, die haben ja auch Spezialsuchmaschinen, so wie Sie das jetzt sagen. Also wenn Sie jetzt sagen würden, Sie suchen nur nach wissenschaftlichen Inhalten, dann wissen Sie, es gibt Google Scholar, dann gehen Sie zu Google Scholar und durchsuchen nur diese Kollektion. Man hat halt gesehen, dass Nutzende in der Regel diesen Schritt nicht gehen und halt alles in die Standard-Suchmaschine eingeben. Und das ist der Grund, warum wir diese Ergebnisseiten bei Google und auch bei anderen Suchmaschinen sehen, die eben Inhalte aus diesen verschiedenen Kollektionen – Nachrichten, Videos, wissenschaftliche Inhalte, Bücher etc. – zusammenfassen und auf einer Ergebnisseite produzieren. Und auch wenn wir Suchmaschinen hätten, die auf so einem offenen Index aufbauen, würde es nicht bedeuten, dass es notwendigerweise Spezialsuchmaschinen wären. Das könnte man machen. Man könnte sagen, ich will nur eine Spezialsuchmaschine für wissenschaftliche Inhalte aufbauen. Sie könnten aber auch eine Suchmaschine

aufbauen, die genauso wie Google das komplette Web abdeckt. Also das ist nicht eine Beschränkung, die da drinstecken würde.

[mg]: Es gibt ja die Möglichkeit, auch jetzt eine Suchmaschine wie Google viel spezieller anzufragen. Also, ich kann ja ein Wort einfach eingeben oder eine Frage eingeben, so wie Sie das eben formuliert haben. Welche Bedeutung hat Faust für die deutsche Literatur? Aber es gibt ja, und viele wissen das vielleicht gar nicht, auch die Möglichkeit, die Suchanfrage einzuschränken auf bestimmte Gebiete oder auf bestimmte Aspekte. Wie funktioniert das genau? Also mache ich das als Nutzerin? Das ist meine erste Frage. Und die zweite wäre dann: Warum ist das so versteckt, obwohl es ja eigentlich die Suche so viel besser und präziser machen würde?

[Lewandowski]: Genau, es gibt ein ganzes Suchvokabular, das man verwenden kann. Das sind auf der einen Seite logische Operatoren, auf der anderen Seite spezielle Befehle, zum Beispiel für bestimmte Dateitypen, um das Datum einzuschränken. Aber auch in allen Spezialsuchen gibt es nochmal extra Möglichkeiten, die Suchen einzuschränken. Also beispielsweise bei Bildern können Sie nach Farben einschränken, nach Rechten, ob Sie die Bilder weiterverwenden dürfen, und so weiter. Also Möglichkeiten gibt es relativ viele. Und genutzt werden Sie so gut wie nicht. Da sind wir eigentlich an einem Punkt, wo sich so ein bisschen die Katze in den Schwanz beißt. Also, wir haben am Anfang in der Entwicklung der Suchmaschinen es so gehabt, dass Suchmaschinen eigentlich so klassische Information Retriever-Systeme nachgebildet haben, wo man eben mit sehr viel Vokabular suchen konnte, sehr ausführlich das machen konnte. Und dann hat man festgestellt: Naja, das macht halt niemand. Und wenn es die Leute machen, dann machen sie es leider auch häufig falsch. Also sprich, sie verbessern ihre Suchergebnisse nicht, sondern sie verschlechtern sie eigentlich. Und aus diesem Grund haben die Suchmaschinen angefangen, die Anfragen auch sehr stark zu interpretieren, bis hin zu dem Thema Personalisierung. Also, Sie geben eine Suchanfrage ein, die aus einem Wort, aus zwei Wörtern besteht, und Sie kriegen die Ergebnisse direkt angepasst. Wenn Sie mit dem Suchvokabular suchen würden und Ihre Suchanfragen sehr präzise formulieren würden, wäre das ja gar nicht nötig. Sondern Sie würden genauer sagen, was Sie haben wollen. Da aber Nutzende das nicht machen oder auch nicht in der Lage dazu sind häufig, müssen die Suchmaschinen eben interpretieren. Und das ist das Problem: Man wirft den Suchmaschinen vor, dass sie so stark interpretieren. Auf der anderen Seite bleibt ihnen eigentlich fast nichts anderes über. Wenn wir alle besser suchen würden, wäre das nicht so nötig.

[mg]: Jetzt gerade dieses Beispiel, ich finde das klasse, dass wir dieses Faust- und deutsche Literaturbeispiel haben. Weil diese Frage, die Sie formuliert haben, die klingt fast ein bisschen wie eine Hausaufgabe in der Oberstufe. Da ist man ja relativ schnell, jetzt heutzutage, bei ChatGPT. Welche Rolle spielen denn Sprachmodelle? Verändern die die Funktionsweise von Suchmaschinen einerseits? Oder, sind die Konkurrenz zu Suchmaschinen? Oder machen sie Suchmaschinen vielleicht sogar obsolet?

[Lewandowski]: Also generative Sprachmodelle spielen natürlich eine große Rolle jetzt auch für die Websuche. Ich kann gleich den letzten Teil der Frage beantworten: Sie werden Suchmaschinen nicht obsolet machen, sie werden Suchmaschinen verändern. Und das sehen wir auch schon. Was aber im weiten Teil eine Veränderung im Interaktionsverhalten zwischen den Nutzenden und der Suchmaschine ist, also dass wir jetzt Dialoge führen können mit der Suchmaschine und eben das Ganze interaktiver haben, verfeinern können und so weiter. Das ist auf der Nutzerseite, das sehe ich erstmal sehr positiv. Warum die generativen Sprachmodelle, die chatbasierten Systeme, nicht die Suchmaschinen ersetzen werden, ich glaube, der Hauptgrund dafür ist, dass wir eine Entwicklung sehen, dass auf der einen Seite die Suchmaschinen solche Funktionen mit einbauen und auf der anderen Seite sehen wir das System, wie ChatGPT eine Websuche mit einbauen. Also das Ganze wächst zusammen und vielleicht wird es in drei Jahren gar nicht mehr eine Rolle spielen, aus welcher Richtung das jetzt kam, ob das ursprünglich mal ein Chat-System war oder eine Suchmaschine war. Ich weiß nicht, ob es sich irgendwie in der Mitte trifft und so ziemlich das Gleiche ist. Die große Veränderung, die wir sehen, das hat wieder mit dem Datenbestand zu tun. Wir haben ja bisher darüber gesprochen, dass der Datenbestand aus Dokumenten besteht. Das heißt, wenn ich in der heutigen Version von Google etwas suche, dann bekomme ich eine Liste von Dokumenten erstmal und kann auf ein Dokument klicken und sehe die Quelle dafür. Und wenn ich bei Google eine direkte Antwort bekomme auf eine Frage, dann ist es ein Satz, der aus einem Dokument extrahiert wurde. Also auch da kann ich wieder genau die Quelle sehen, kann draufklicken und sehe das. Jetzt haben wir es bei generativen Sprachmodellen anders, nämlich der Datenbestand, also der Index mit den Dokumenten, fällt erstmal weg, kommt gleich wieder ins Spiel, aber der fällt erstmal weg und die Dokumente werden nur zum Lernen verwendet. Und wenn das System gelernt hat, generiert es automatisch Antworten. Und die Antworten können richtig sein, die können falsch sein, ich kann es nicht mehr nachprüfen. Das klingt ja oft plausibel, aber inhaltlich stimmt es halt nicht. Und da kommt jetzt ins Spiel, dass man eben versucht, diese beiden Welten zu verbinden und zu sagen, entweder man holt sich aus dem Dokumentenindex eine bestimmte Anzahl von Dokumenten, sagen wir mal die Top 5 Dokumente, die gerankt werden, und lässt die einfach von einem generativen Sprachmodell zusammenfassen zu einer konsistenten Antwort, sodass also nur das, was wirklich in diesen Dokumenten auch inhaltlich vorkommt, in der Antwort mit drin ist. Oder man generiert eine Antwort durch das Sprachmodell und prüft diese Antwort nachher nochmal über die Dokumente aus dem Dokumentenindex, ob das plausibel ist und ob es stimmt. Und damit ist halt eine Frage verbunden, und das ist die nach Korrektheit, Wahrheit, Vertrauenswürdigkeit von den Antworten, wenn das Ganze nicht mehr direkt auf den Dokumenten beruht.

[mg]: Das heißt, wir landen am Ende dann aber trotzdem wieder bei einer Sammlung über die Dokumente, die irgendwann verfügbar sind, die im Sprachmodell eben die Trainingsdaten sind? Also wäre sozusagen, wenn ich jetzt anschau, was sind das für Dokumente und wo liegen die, wäre das dann identisch?

[Lewandowski]: *Ne, es muss nicht identisch sein, weil die Sache ist ja die, Sie wollen ja die Dokumente, die im Dokumentenindex sind, irgendwie auch an Nutzende vermitteln. Wenn Sie Dokumente haben, die Sie zum Lernen verwenden, und das sehen wir im Moment mit diesen ganzen Lizenzvereinbarungen, die geschlossen werden, sind es ja unter Umständen Kollektionen, die gar nicht über das Web öffentlich zugänglich sind. Also wenn wir zum Beispiel sehen, vielleicht haben Sie es auch bekommen von den wissenschaftlichen Verlagen, die jetzt alle Ihre Wissenschaftler anfragen, wie es denn so aussieht, dass Sie gerne die Inhalte auch verwerten wollen für generative Sprachmodelle, da kriegen Sie ja hinterher die Dokumente nicht. Und wenn jetzt zum Beispiel aus einem meiner Aufsätze irgendwas rausgenommen wird und es wird durch generatives Sprachmodell als Antwort irgendwie ausgegeben, ist es ja nicht mehr nachvollziehbar, woher das kommt, wenn dieses Dokument nicht im Webindex ist und nochmal überprüft wird an der Stelle. Also der Index oder der Datenbestand, um die generativen Sprachmodelle anzulernen, kann unterschiedlich sein von dem Webindex und das kann auch zu Problemen führen.*

[pgg]: *Ist es nicht denkbar, dass die Verbindung von generativer KI und Suchmaschinen auch den Suchmaschinen die Möglichkeit gibt, sozusagen noch viel deutlicher zu interpretieren, worauf sie zugreifen? Das schlichte Vermitteln von Treffern ist ja eine relativ bescheidene Aufgabe, sage ich jetzt mal. Also bescheiden im Sinne von Ansammeln von Wissen. Aber in dem Moment, wo eine generative KI gewissermaßen mit durch die KI generierten eigenen Informationsmengen interagiert, mit der Suchmaschine, wird ja auch eine bestimmte Sorte von nachhaltigem Wissen mit der Dokumentensammlung verknüpft. Nachhaltig nicht als Bewertung, sondern einfach nur verfügbar, vielleicht nicht identisch reproduzierbar, aber da ist gewissermaßen immer eine interpretierende Maschine auch schon am Werk. Und das macht ja die Suchmaschine nicht.*

[Lewandowski]: *Genau, die Suchmaschine interpretiert die Kollektion nur in der Hinsicht, dass sie quasi eine Bewertung vornimmt und dann eine Liste ausgibt. Das ist, glaube ich, der Schritt, wo wir sehen, dass eine Suchmaschine oder ein KI-System, was also im Rahmen der Suchmaschine agiert, einen Schritt weitergeht. Wir haben immer diese Unterscheidung zwischen Information Seeking und Information Use. Und eigentlich hat man bisher immer das Information Seeking unterstützt und hatte die Schwierigkeit, dass die Informationsnutzung am Ende, also da, wo wir dann selber dransitzen und Informationen zusammenfassen, die Berge von Dokumenten durchlesen müssen und so weiter, uns da auch weiterhilft. Und die beiden Bereiche wachsen jetzt zusammen. Das ist, glaube ich, die Stärke und das, was Sie beschrieben haben, wie sich die Kollektion quasi und die Interpretation der Kollektion verändert.*

[mg]: *Trotzdem bleibt es ja dabei. Ich komme jetzt nochmal zurück zu diesen unterschiedlichen Indizien, die es so gibt auf der Welt. Wir hatten jetzt zwei amerikanische genannt. Es gibt einen russischen, einen chinesischen. Da war jetzt kein europäischer dabei.*

[Lewandowski]: Genau.

[mg]: Ist das ein Problem? Sollten wir uns darüber Sorgen machen oder gibt es Bestrebungen, daran was zu ändern?

[Lewandowski]: Genau, ich denke, wir sollten uns Sorgen machen. Es ist schon sicher sinnvoll, sowas auch in Europa zu haben. Das ist eine Basistechnologie letztlich für das Web. Es gibt tatsächlich Bestrebungen und es gibt zwei Projekte, die sich damit beschäftigen. Das eine ist ein von der EU gefördertes, eher wissenschaftliches Projekt. Das ist die Open Search Foundation, die da federführend ist und versucht, so einen Index aufzubauen, der eben für andere nachnutzbar ist. Und dann haben jüngst zwei Unternehmen, nämlich Ecosia und Quant, angekündigt, dass sie sowas auf kommerzieller Basis machen möchten. Also sprich, anderen Unternehmen dann Zugang zu so einem Index zur Verfügung zu stellen, das aber gegen Bezahlung eben zu machen, um daraus ein Geschäftsmodell zu entwickeln. Das sind im Prinzip zwei Ansätze, die die gleiche Idee verfolgen oder versuchen, die gleiche Idee umzusetzen, eben auf unterschiedliche Weise, mit ein bisschen unterschiedlichem Fokus.

[pgg]: Wann wird man wissen, ob daraus was wird?

[Lewandowski]: Also, soweit ich weiß, ist das EU-Projekt noch bis Ende des Jahres, glaube ich, gefördert. Die kommen jetzt bald mit einer ersten Version raus. Über das Ecosia-Quant-Projekt weiß ich jetzt nichts im Detail, aber die wollen auch noch in diesem Jahr direkt starten. Also da wird man nicht vielleicht direkt sehen, ob es erfolgreich ist, aber man wird erstmal einen Eindruck von den beiden bekommen können.

[mg]: Sie sagten ja schon, eigentlich wäre der Weg aus Ihrer Sicht, dass es den Index gibt oder die Indizes gibt und man verschiedene Anwendungen dann draufsetzen kann. Würde sich das unterscheiden, je nachdem: Das öffentlich geförderte Projekt oder das private Projekt von Ecosia und Quant?

[Lewandowski]: Also, die werden sicher unterschiedliche Bedingungen haben für die Zugänglichkeit, aber letztlich ist die Idee ja die gleiche, dass es letztlich für alle Beteiligten möglich ist. Ich denke, jetzt bei einem kommerziellen Projekt wird es natürlich sein: Es muss sich monetarisieren. Da werden sie wahrscheinlich höhere Zugangskosten erstmal haben oder Verarbeitungskosten für die Sachen. Das ist dann schlicht eine Businessfrage, wie sie es durchrechnen, mit welchem Modell sie da gut zu Rande kommen. Es kann natürlich auch Qualitätsunterschiede geben. Das kann ich jetzt aus heutiger Perspektive nicht beurteilen.

[pgg]: Ich habe nochmal eine Frage zu dem Punkt, dass wir die Möglichkeiten der Suchmaschinen so krass unterfordern. Gibt es da irgendwie Wissensvermittlung, gibt es sowas wie Suchkurse?

[Lewandowski]: Also ja, es ist ganz witzig, wenn man sich so die Literatur anguckt, dann hatten wir, als es mit Suche losging im World Wide Web Mitte/Ende 90er Jahre bis in die frühen 2000er rein, gab es eigentlich eine relativ reiche Literatur. Gab es Bücher – Wie kann ich besser suchen im Internet? – Anleitungen und die Leute haben sich damit beschäftigt, was auch mit der schlechten Qualität der Ergebnisse zu tun hatte. Das ist eigentlich inzwischen vollkommen ausgestorben. Also nicht mal mehr im Profibereich, da ist, glaube ich, das letzte Buch, soweit ich weiß, auch vor 7, 8 Jahren erschienen, so Professional Internet Search. Also, das scheint kaum mehr ein Thema zu sein. Es hat auch damit zu tun, dass die Leute mit Suchmaschinen ja auch sehr zufrieden sind. Es gibt Kurse, wenn Sie an Ihre Uni gucken, da gibt es garantiert von der Bibliothek auch Informationskompetenzschulungen. Wo sowas behandelt wird, da ist jetzt vielleicht nicht der Schwerpunkt immer genau auf Suchmaschinen, aber generell auf die Frage, wie man besser suchen kann. Was ja auch eine Frage ist: Wann verwende ich eigentlich eine Suchmaschine und wann sollte ich besser andere Tools verwenden?

[pgg]: Kataloge zum Beispiel.

[Lewandowski]: Zum Beispiel oder Fachdatenbanken. Also ich glaube, da ist noch viel zu tun. Was wir halt sehen, ist, dass die Leute in aller Regel wenig Bewusstsein dafür haben, dass sie nicht gut suchen können. Also, wir haben verschiedene Befragungen gemacht und haben auch abgefragt, die Leute ihre Suchkompetenz einschätzen. Und da können Sie sehen, dass die weit meisten Leute ihre Suchkompetenz als sehr gut oder gut einschätzen. Was sich dann in den Tests, die wir erstellen, nicht unbedingt bewahrheitet.

[mg]: Das heißt, die Einschätzung kommt daher, dass man halt was findet, was irgendwie passt?

[Lewandowski]: Ja, genau. Also ich meine, es ist ja auch schön. Also ich meine, man muss sich das nochmal klar machen: die Leistung der Suchmaschinen. Also, wenn ich zurückdenke, bevor es Suchmaschinen gab, wie schwierig Suche war. Die Suchmaschinen haben sich also auch positiv auf alle anderen Suchsysteme ausgewirkt, die dann nachziehen mussten und ihre Sachen vereinfachen mussten. Suche ist so viel einfacher geworden. Jeder kann was finden. Das ist erstmal absolut toll. Es vernebelt halt so ein bisschen, dass man an irgendeinem Punkt, wo es wichtig wird, auch mal vielleicht woanders gucken müsste.

[Der Abspann mit Musik beginnt.]

[mg]: Und damit ist dieses Digitalgespräch zu Ende. Und wir bedanken uns bei Dirk Lewandowski von der Hochschule für Angewandte Wissenschaft in Hamburg für die spannenden Einblicke und die interessante Diskussion. Viele Grüße. Und wie immer auch vielen Dank an Sie, liebe Zuhörerinnen und Zuhörer, für das Interesse und die Aufmerksamkeit. Wenn Sie mögen, hören wir uns wieder in drei Wochen. Zur nächsten

Folge des Digitalgesprächs, einem Podcast von ZEVEDI, dem Zentrum verantwortungsbewusste Digitalisierung.



This work is licensed under CC BY-NC-ND 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>