

Wer und Wen

Deutsche Forscher arbeiten an Wissensmaschinen, die verstehen, wonach der Nutzer fragt.

Die Eingabe „Wer hat Steuern hinterzogen?“ fördert bei Google allerlei zutage, nur keinen Steuerhinterzieher unter den ersten zehn Treffern – stattdessen etwa einen Zeitungsartikel über Tarnfirmen und den Paragraphen 235 der Abgabenordnung, der regelt, wie hinterzogene Steuern verzinst werden.

Das war aber nicht die Frage. Die Suchmaschine Sempria macht es besser: Sie liefert eine hübsche Liste einschlägiger Sünder, angeführt vom vormaligen Post-Chef Klaus Zumwinkel.

Auch hier gibt es noch Fehltreffer, doch Sempria versteht offenbar, dass man tunlichst nach Personen suchen sollte, wenn die Frage mit „Wer“ beginnt. „Google wirft Fragewörter gleich weg, wir aber nutzen sie“, sagt Hermann Helbig, pensionierter Professor der Fernuni Hagen. Im vergan-

Auf manche Fragen kommen da schon Antworten, die verständlich erscheinen: Wo hat Oliver Bierhoff studiert? „An der Fernuniversität in Hagen.“ Sempria versucht sogar, zwischen „Wer“ und „Wen“ zu unterscheiden, zwischen Akteur und Objekt: Wen hat Lee Harvey Oswald ermordet? „John F. Kennedy.“ Wer hat Lee Harvey Oswald ermordet? „Jack Ruby.“

In Berlin ist unterdessen eine Wissensmaschine namens Alexandria in Arbeit, die noch einen Schritt weitergeht: Sie stellt Berechnungen und Vergleiche an. Man kann von ihr erfahren, wer in der Formel 1 mehr Rennen gewonnen hat als Niki Lauda. Und welcher Schauspieler in den meisten Filmen überhaupt mitgespielt hat.

Freilich muss man der Software alles einzeln beibringen: Menschen können

etwas gewinnen, Menschen können irgendwo mitspielen. Dass sie auch Hirnzellen haben, ist offenbar noch nicht im Programm. Die Frage, wie viele das sein mögen, weist Alexandria zurück: „Ein Mensch hat keine Gehirnzellen.“

Daran wird derzeit gearbeitet.

Die Software gehört zum weitläufigen Forschungsprogramm Theseus, wo intelligente Suchtechnik entstehen soll. „Wir sind auf dem Weg zur Antwortmaschine“, sagt der Informatiker Wolfgang Wahlster, einer der Koordinatoren.

Die größte Dynamik sieht Wahlster bei den mobilen Anwendungen. Der Mensch unterwegs hat immer mal wieder dringende Bedürfnisse, tippt nicht gern auf den winzigen Geräten und hätte wohl am liebsten ein Handy, das ihn einfach versteht, was immer er sagt: „Wann fährt die letzte S-Bahn retour?“ Oder vielleicht auch nur: „Bring mich hier weg!“

„In drei Jahren haben wir Frage-Antwort-Systeme, die für den Hausgebrauch taugen“, sagt Wahlster. „Bei philosophischen Fragen wäre ich mir aber nicht so sicher – das schaffen wir vielleicht in 50 Jahren nicht.“



Theseus-Innovationszentrum in Berlin
„Bring mich hier weg“

genen Herbst gründete er mit ein paar langjährigen Mitarbeitern die Firma Sempria. Die gleichnamige Suchmaschine steht vor der Markteinführung.

Anders als Google weiß Sempria auch, dass Zumwinkel eine Person ist, die Leine hingegen ein Fluss, wenn eine Stadt an ihr liegt, nicht aber, wenn ein Hund an ihr zerrt.

Das ganze Internet könnte Sempria nicht bewältigen – weil die Software versucht, die Texte auch inhaltlich im Ansatz zu verstehen, taugt sie vorerst eher für begrenzte Bestände, etwa zum Durchsuchen von Firmen-Web-Seiten oder Archiven. Eine Testversion läuft auf Basis der Wikipedia und etlicher Nachrichtenquellen.

ge Weltliteratur, die Theaterstücke Shakespeares, etliche Jahrgänge der „New York Times“ und, für alle Fälle, ein paar Übersetzungen der Bibel.

„Zusammen sind das wohl an die hundert Gigabyte Text“, sagt David Ferrucci. „Aber das garantiert uns gar nichts.“ Jederzeit kann eine Quizfrage daherkommen wie diese hier: „Chile teilt seine längste Grenze mit diesem Land.“ Watson lag immerhin mit Bolivien nur knapp daneben; auf Argentinien kam er nicht. „Ein Laie meint vielleicht, da schlägt man eben in einer Datenbank nach“, sagt Ferrucci. „Aber wir haben diese Datenbank nicht. Wer hätte denn mit einer solchen Frage rechnen können?“

Eine Verbindung zum Internet gibt es ebenfalls nicht. Der Computer ist, wie seine menschlichen Gegner, auf sich gestellt.

Am Anfang tat Watson sich denn auch schwer. Obwohl er die Rechenleistung des Vorgängers Deep Blue um etwa das Zehnfache übertrifft, verstand er oft schon die Frage kaum. Doch im Lauf der Zeit wurde der Computer immer besser. Heute besteht er auch heikle Prüfungen: „In dieser Sprache, gesprochen von 120 Millionen weltweit, enden alle Wochentage außer einem mit den gleichen drei Buchstaben.“ Watson musste nicht lange überlegen: „Deutsch.“

Der Zauber ist nicht ganz neu; erste Antwortmaschinen gab es bereits in den Sechzigern. Das System „Lunar“ betörte 1971 einen Mondforscherkongress mit vollständigen Auskünften über das Gestein, das die „Apollo“-Astronauten mitgebracht hatten. Ein Quiz zur Geologie des Erdtrabanten hätte es spielend gewonnen – aber schon nach Neil Armstrong hätte man es nicht mehr fragen dürfen.

Solche Computer waren früher auf winzige Fachgebiete beschränkt. Denn die Forscher mussten ihnen noch alles Weltwissen mühsam von Hand einlöffeln. Damit kamen sie nicht weit.

Vor gut einem Jahr ging dann noch einmal ein Großvorhaben in diesem Geiste in Betrieb: die Suchmaschine „Wolfram Alpha“. Dutzende Mitarbeiter haben jahrelang Faktenwissen und Formeln dafür zusammengetragen und in Datenbanken gefüllt. „Wolfram Alpha“ eignet sich denn auch am besten für Anfragen, die etwas ausgerechnet haben wollen. Der Suchende kann hier erfahren, dass 125 Stück Apfelkuchen den Tagesbedarf an Vitamin B12 zu 146 Prozent stillen – nach Rezepten aber stöbert er hier vergebens.

Watson dagegen soll mit der ganzen Welt des Wissens umgehen können. Das geht nur dank einer Methode, die in letzter Zeit große Fortschritte gemacht hat: Der Computer lernt selbst, wie es in der Welt zugeht, und er braucht dafür vor allem Statistik. Seine Hauptfrage an einen Text ist: Welche Wörter treten häufig zusammen auf, welche sehr selten?