

Quelle	DER SPIEGEL vom 26.04.2010
Seite	156 - 157
Jahrgang	2010
Nummer	17
Ressort	Wissenschaft
Rubrik	Wissenschaft
Seitentitel	Wissenschaft
Kurztitel	Internet: Ein Deutscher schuf bei Google das bislang beste Übersetzungsprogramm
Autor	Bethge, Philip



INTERNET

Googeln in fremden Zungen

Ein deutscher Forscher hat eines der ersten alltagstauglichen Übersetzungsprogramme entwickelt. Brachiale Rechenkraft verhilft der Google-Software zu verblüffend guten Ergebnissen.

Es ist ein gutes Zeichen, wenn der Schöpfer einer Software diese auch endlich selbst benutzt. In Japan war Franz Och jüngst auf Reisen. Im Restaurant zückte er sein Handy und entschlüsselte die Speisekarte. Auch Lokalnachrichten las er ohne Mühe. Denn die Übersetzungen erhielt er in Sekunden.

Seit sechs Jahren entwickelt der 38-Jährige im Google-Hauptquartier im kalifornischen Mountain View das Übersetzungsprogramm "Google Translate", "und bislang habe ich es eigentlich nie wirklich selbst benutzt", räumt Och ein. Doch nun der Sinneswandel: "Ich bin sehr glücklich über das, was wir erreicht haben."

Der Deutsche ist der heimliche Star einer Software-Branche, die kein geringeres Ziel hat, als die globalen Sprachbarrieren einzureißen. Och ringt für Google mit Schachtelsätzen, Konjunktionen und Hilfsverben. Das Resultat ist ein Affront für jeden Linguisten. Denn nicht Sprach-Know-how, sondern brachiale Rechenkraft erlaubt die maschinelle Dolmetscherei.

52 Sprachen beherrscht das System bereits. Die Datenbanken für 296 weitere sind gleichsam im Rohbau. Unter ihnen finden sich Exoten wie Sardisch, Westfriesisch oder Zulu.

Komplette Internetseiten, Diplomarbeiten, selbst Liebesbriefe übersetzt "Google Translate" im Handumdrehen - und liefert dabei oftmals verblüffend brauchbare Ergebnisse. Für Google liegt der Nutzen auf der Hand: Noch mehr Internetsurfer als bislang lassen sich mit einer derart nützlichen und zudem kostenlosen Anwendung auf die Web-

site der Firma locken.

"Maschinelle Übersetzung ist auf einer neuen Stufe angekommen", schwärmt Och, "die Leute setzen unser Programm inzwischen massiv ein; die Software hat die reale Welt erreicht."

"Was Google hier macht, ist sehr beeindruckend", bestätigt Alon Lavie von der Carnegie Mellon University in Pittsburgh. Der Informatiker sieht die gesamte Branche in Bewegung. Der Markt für Übersetzungssoftware wachse rapide: "Es sind sehr aufregende Zeiten."

Die Epoche der maschinellen Übersetzung ist angebrochen. Programme wie "Google Translate" weisen den Weg in eine Zukunft, in der jedermann auf Knopfdruck in fremden Zungen sprechen kann. Das ultimative Ziel der Sprachtütfler ist eine elektronische Version jenes Babelfischs, den der britische Autor Douglas Adams in seinem Science-Fiction-Klassiker "Per Anhalter durch die Galaxis" erfand: Ins Ohr eingesetzt, übersetzt die blutegelartige Kreatur jede Sprache simultan. Selbst die krude Poesie der Vogonen kann Romanheld Arthur Dent verstehen.

So weit haben es die Entwickler in der Realität zwar noch nicht gebracht. Doch schon gibt es iPhone-Apps wie "Jibbig", das gesprochenes Englisch wieselflink ins Spanische übersetzt. Schöpfer der Software ist [Alex Waibel, Informatiker an der Universität Karlsruhe](#) und an der Carnegie Mellon University. Waibel lässt bereits viele seiner Vorlesungen von Rechnern simultan übersetzen. Auch an Parlamentsdebatten testete er die Technik schon.

Dabei war der Dolmetscher aus dem Computerlabor lange ein kühner Traum. Woher etwa soll die Maschine wissen, dass beim Englischen "breaking records" keine Schallplatten zerdeppert werden? In dem Satz "wir treffen uns im Schloss" wiederum muss die Software die Vokabel "Schloss" mit "castle" übersetzen und keinesfalls mit "lock" (Türschloss), was offensichtlich ziemlicher Blödsinn wäre.

Lange Zeit versuchten die Informatiker, den Programmen derlei Weltwissen über ein komplexes Regelwerk einzubimsen. Doch selbst bei tadellos formulierten Texten geriet die Software oftmals in fröhliches Delirieren. Einen "Alptraum" nennt etwa Swamy Viswanathan von der US-Firma Language Weaver den Versuch, beispielsweise die englische Sprache mit all ihren Nuancen in Regeln zu pressen. "Wörter haben häufig mehrere Bedeutungen, und die Zahl der Kombinationen ist schier unendlich", klagt Viswanathan.

Die Experten von Language Weaver verfolgten daher schon früh ein anderes Konzept. Sie fütterten ihre Systeme mit unzähligen Texten aus dem Internet, die bereits mehrsprachig vorlagen. Der Ansatz der Spezialisten: Fast jeder Satz, jede Phrase ist längst mehrfach übersetzt worden. Reine Statistik reicht daher aus, um ein Sprachkonstrukt zu entschlüsseln.

Um etwa das Beispiel "wir treffen uns im Schloss" zu deuten, fahndet das Programm in seiner Datenbank nach Texten, in denen "treffen" und "Schloss" nah beieinanderstehen. Dann durchforstet es die Übersetzungen dieser Texte

und findet dort häufig das Wort "castle". Daher gibt die Maschine "we meet in the castle" und nicht "we meet in the lock" aus.

Och hat dieses statistische Verfahren für Google inzwischen perfektioniert. Schon während seiner Promotion spezialisierte sich der Franke auf Spracherkennung. Danach ging er an die University of Southern California. Bald interessierte sich das Pentagon für seine Arbeit. Nach 9/11 wollten die US-Geheimdienstler arabische Zeitungen, Chatrooms und Websites verschärft kontrollieren.

Doch 2004 lockte Google den Sprachbezwinger nach Mountain View. Dort kommt Och seither die gewaltige Rechenleistung des Internetimperiums zugute. Zahlen mag Och nicht nennen. Für viele Sprachpaare lagern jedoch Billionen Einträge in den Google-Datenbanken. Wichtige Ressourcen für das Wortarchiv sind etwa die in zahlreiche Sprachen übersetzte Bibel, Transkripte der Vereinten Nationen oder die 23-sprachigen EU-Dokumente.

Solcherlei "Paralleltexte" sind so etwas wie der Stein von Rosette des Digitalzeitalters: Das antike Vorbild trägt dieselbe Inschrift in Griechisch, Demotisch und in Hieroglyphen. 1822 lüftete der Ägyptologe Jean-Francois Champollion mit ihrer Hilfe das Geheimnis der Hieroglyphen.

Genauso macht es nun Ochs Software. Die Stärken des Systems: Ein und derselbe Programmcode funktioniert für alle Sprachen. Es muss nur genug übersetzter Text vorliegen.

Ein Buchstabenfresser als Universaldolmetscher? Viele Linguisten halten derlei Rechenschiebertricks für Tinnef. "Die statistische Übersetzung stößt

schnell an ihre Grenzen", sagt etwa der Sprachwissenschaftler Martin Kay von der Stanford University, "der Ansatz ignoriert die komplexe Struktur von Sprache." Bei der im Deutschen üblichen Stellung von Vollverb und Hilfsverb etwa versage die Technik. Auch an der Unterscheidung von Subjekt und Objekt habe sie zu knapsen.

"Für wirklich gute Ergebnisse müssen wir tiefer in die Sprache eintauchen", sagt auch Hassan Sawaf, Chefentwickler der US-Software-Schmiede Apptek. Die Firma geht einen Mittelweg. Neben statistischen Algorithmen nutzt Sawaf auch klassische Grammatikregeln: "Das verbessert den Satzbau und die Verständlichkeit erheblich."

Sawaf kritisiert zudem, dass Ochs System nur online funktioniert: "Wer offline arbeitet, kann 'Google Translate' vergessen." Auch Waibel ist skeptisch. "Stellen Sie sich vor, Sie sind im Ausland, wollen sich mit einem Verkäufer unterhalten und müssen erst mal ein Netz suchen und dann auch noch hohe Roaming-Gebühren bezahlen - praktikal ist das nicht."

Tatsächlich stellt die Netzhängigkeit eine der größten Schwächen der Google-Übersetzungsmaschine dar. Doch die Kalifornier weichen nicht von ihrem Kurs ab. Schon entwickeln sie eine spezielle Programmversion mit integrierter Spracherkennung für das hauseigene Handy-Betriebssystem "Android". Und bald soll es auch möglich sein, Texte auf Fotos blitzschnell übersetzen zu lassen. So könnte der ortsunkundige Reisende künftig in China das Schild mit der Aufschrift ablichten - und wüsste umgehend, dass er sich auf dem Weg nach Peking befindet.

Eine weitere Gelddruckmaschine für

den Internetgiganten scheint also heranzureifen. Doch Och winkt ab. Wie viele Google-Mitarbeiter wähnt er sich lieber auf einem Feldzug für Freiheit und Gleichheit im Netz. "Wer kein Englisch spricht, kann nur einen Bruchteil des Internets nutzen", sagt er. Es gelte, die Vielfalt allen zugänglich zu machen.

Ein Indiz für die hehren Absichten des Programmierers gibt es immerhin. Och und sein Team haben eine Spezialsoftware entwickelt, mit deren Hilfe Dolmetscher auf eigene Faust Übersetzungen in das System einspeisen können - und zwar auch für äußerst exotische Idiome wie die Bantusprache Xhosa, die Sprache der in Japan lebenden Ainu oder die Inuit-Sprache Inuktitut. Auf diese Weise wollen die Software-Entwickler auch jenen Zungen Gehör verschaffen, die bereits in Vergessenheit zu geraten drohen. Der neuseeländische Computeringenieur Te Taka Keegan von der University of Waikato testete das Programm bereits für die Sprache der Maori. Sechs Monate verbrachte Keegan kürzlich bei Google, um auszuloten, ob das digitale Sprachenmirakel aus Mountain View das Idiom der neuseeländischen Ureinwohner vor dem Verschwinden bewahren könnte. Seine Erfahrungen sind durchweg positiv.

"Die Zahl und Qualität der Maori-Übersetzungen wächst mit Hilfe dieses Werkzeugs ständig", berichtet Keegan. Ein digitales Archiv entstehe, das der Sprache erheblichen Auftrieb verleihe. "Unsere Kinder wachsen in eine digitale Welt hinein", sagt Keegan. "Nur wenn wir es schaffen, Maori zum Teil dieser Welt zu machen, wird die Sprache überleben."

Abbildung Informatiker Och: Ein Affront für jeden Linguisten, CLAY MCLACHLAN / AURORA