

INTERNET

Schatztruhe des Wissens

Der Google-Konzern unterstützt Gelehrte bei der Auswertung seiner 15 Millionen Bücher umfassenden Digitalbibliothek. Geisteswissenschaftler sind begeistert.

An Weitsicht fehlte es dem Griechen nicht. Um die „bewundernswerten Taten“ der Menschen „nicht ruhmlos vorübergehen“ zu lassen, verfasste Herodot im 5. Jahrhundert vor Christus seine „Historien“. Detailliert schilderte der antike Gelehrte darin, wie und warum sich Perser und Griechen wiederholt gegenseitig die Köpfe einschlugen.

Rund 2500 Jahre später können Historiker das antike Meucheln auf eine Weise erforschen, die sich Herodot nicht hätte ausmalen können: Teile des rund 700 Seiten umfassenden Werks gehören zu jenen Klassikern, die Google für seine Digitalbibliothek „Google Books“ eingescannt hat. In einer Initiative namens „Digital Humanities“ bereiten die kalifornischen Datensammler die Bücher derzeit für Forschungszwecke auf.

„Seit es Computer gibt, haben Menschen die Vision, das Wissen der Welt zu verknüpfen“, sagt Projektmanager Jon Orwant. Bis heute jedoch sei historische und philosophische Weisheit aus der Zeit vor Erfindung des Internets online kaum zugänglich. Diese Schatztruhe des Wissens will Google der Gelehrtenwelt öffnen.

Geschichts- und Kulturwissenschaftler sind begeistert. „Google Books ist eine Art Genom der Weltkultur“, urteilt Eric Kansa von der University of California, Berkeley. „Wir haben es mit einer Ressource von außergewöhnlicher Bedeutung zu tun.“

Rund 15 Millionen Bücher in etwa 400 Sprachen hat Google bislang digitalisiert – etwa zwölf Prozent aller Bücher, die seit der Gutenberg-Bibel veröffentlicht wurden. Doch die Datensammlung

hat dasselbe Problem wie andere große Archive auch: Wie lässt sich das umfangreiche Material bündig und sinnvoll nutzen?

Orwant und seine Kollegen glauben eine Antwort zu haben: Sie entwickeln Software-Systeme, um die Digitalbibliothek zu analysieren. Zu diesem Zweck kooperieren sie eng mit Forschungsgruppen. Rund eine Million Dollar hat Google für bislang 24 Projekte bereitgestellt.

Der Historiker Dan Cohen von der US-amerikanischen George Mason University gehört zum Beispiel zu den Google-Partnern. Cohen interessiert sich für das Viktorianische Zeitalter. Bislang mühte sich der Professor durch einzelne Bücher der Epoche. Dank Google hat er auf einmal Zugriff auf nahezu 1,7 Millionen Bände aus der Zeit von 1789 bis 1914.

Cohen ergründet den Gemütszustand der viktorianischen Gesellschaft. Mit Hilfe der Google-Daten hat er herausgefunden, dass in den Buchtiteln der viktorianischen Literatur die Wörter „Religion“ und „heilig“ mit der Zeit seltener auftauchen. Das Interesse an „Wissenschaft“ nahm dagegen zu. Für Cohen bestätigt dies, dass die Untertanen Königin Victorias langsam vom Glauben abfielen.

Der Historiker verwendet sogenannte Ngramm-Statistiken, Häufigkeiten einzelner Wörter. Jeder kann die Methode bei Google selbst ausprobieren*. Andere Ge-

lehrte sind noch kreativer. Das Projekt „Google Ancient Places“ etwa untersucht, welche Stätten des klassischen Altertums in Werken aus der oder über die Antike eine Rolle spielen. „Wer sich für irgendeine römische oder griechische Stätte interessiert, kann erfahren, welche Autoren diesen Ort beschreiben“, erläutert Teammitglied Leif Isaksen von der University of Southampton. Auch die entsprechenden Textstellen lassen sich nachlesen**.

Der Gewinn, den die Geisteswissenschaften aus der Computeranalyse ziehen könnten, scheint enorm. Welcher Altphilologe würde sich nicht freuen, die Übersetzungen von Homers „Ilias“ per Mausklick miteinander vergleichen zu können? Und wären nicht Linguisten hoch erfreut, Details darüber zu erfahren, wie sich das mitteldeutsche „nuwelinges“ in der Literatur zum „neulich“ wandelte?

Doch es gibt auch Kritiker. So mancher Geisteswissenschaftler sucht sein Glück lieber zwischen echten Buchdeckeln. Viele Gelehrte, erstmals mit dem Konzept konfrontiert, betrachteten es als „die Ankunft des Antichrists“ persönlich, sagt der US-Historiker Anthony Grafton.

Orwant versichert, dass die digitale Analyse keinesfalls herkömmliche Forschung ersetzen soll. Ohnehin können die Wissenschaftler bislang viele neuere Werke nicht im Volltext durchstöbern. Denn seit Jahren streitet Google mit heutigen Autoren und Verlagen um Nutzungsrechte. Kritiker halten die Digitalbibliothek für Teufelszeug, weil völlig unklar ist, wie ein Vergütungsmodell für die Urheber der Texte aussehen könnte.

So kann Googles „Digital Humanities“-Initiative auch als Versuch gedeutet werden, die Gelehrten aus den Universitäten als Verbündete im Rechtstreit zu gewinnen.

Und auch bei anderen Kulturschaffenden macht Google gute Stimmung. Gerade kündigte die Firma eine Kooperation mit dem belgischen „Mundaneum“ an, das eine einzigartige lexikalische Sammlung beherbergt. Auch sie könnte künftig Teil der Google-Bibliothek werden.

Datenexperte Orwant jedenfalls hat das Ziel klar vor Augen: „Wir wollen, dass der Kosmos der Ideen und Gedanken uneingeschränkt erforscht werden kann“, sagt er. Das Potential der Digitalbibliothek sei überwältigend – „und wir können noch gar nicht absehen, welche Schätze es dort zu heben gibt“.

PHILIP BETHGE



Google-Datenexperte Orwant: „Kosmos der Ideen“

* <http://books.google.com/ngrams/>

** <http://gap.alexandriaarchive.org/gapvis/index.html>