

sensbasis Cyc will Computern Selbstverständlichkeiten beibrin

Die Zeit, Hamburg, Germany
Die Zeit, Hamburg, Germany

DIE ZEIT

Künstliche Intelligenz: Die Wissensbasis Cyc will Computern Selbstverständlichkeiten beibringen. Jetzt kann sie über das Internet genutzt werden

Global banal

Thomas J. Schult

Tony im australischen Ballarat liebt die Rolling Stones. Hidetaro aus Osaka schwingt sich nach Feierabend gern aufs Mountainbike. Und Harald aus Hannover nennt zwei Nymphensittiche sein eigen. Das Internet ist voll von banalem Zeug. Jetzt kommen auf einen Schlag Tausende Banalitäten hinzu, die es aber in sich haben: Die Wissensbasis Cyc, Paradeprojekt der künstlichen Intelligenz und monumentales Kompendium der Trivialitäten (ZEIT Nr. 48/1994), ist nun veröffentlicht, zumindest in wichtigen Teilen. Sie steht im Internet unter <http://www.cyc.com>.

Cyc (gesprochen "baik") soll Computer schlauer machen. Menschen wissen sowieso schon alles, was in dieser Datenbank verzeichnet ist: daß Lebewesen mit der Haut tasten können, daß medizinische Instrumente vor einer Operation steril sind, daß Enttäuschung mit Trauer einhergeht. Zwölf Jahre lang haben zwanzig menschliche "Cyclists", wie sich die Cyc-Forscher nannten, Hunderttausende von solchen Zusammenhängen gesammelt. Zusätzlich programmierten sie Schlußverfahren, um neue Sachverhalte aus den bekannten abzuleiten. Daß Helmut Kohl berühmt ist, muß beispielsweise nicht eigens aufgeführt sein – es ergibt sich aus der Tatsache, daß er Regierungschef eines großen Staates ist.

Die üblichen Computer sind bekanntlich dumm wie Bohnenstroh, und unsere Sprache verstehen sie sowieso nicht. Der US-Informatiker Douglas Lenat trat vor zwölf Jahren an, dies zu ändern. Das Alltagswissen der Menschen wollte er Stück für Stück in Softwareform gießen. Andere Programme sollten diesen Wissensschatz dann nutzen können, um den Anwender vor dem Bildschirm mit einer Art gesundem Menschenverstand zu beeindrucken. Cyc wurde das größte Projekt der Forschungsdisziplin künstliche Intelligenz. Technikmultis wie Microsoft und Kodak finanzierten Lenat und seine Cyclists.

Noch ist gar nicht klar, was man mit einer Enzyklopädie der Selbstverständlichkeiten alles anfangen kann. Vieles wünscht sich der geplagte Computernutzer: daß sich der Rechner beschwert, wenn ein zwanzigjähriger Mitarbeiter mit zehn Jahren Berufserfahrung in einer Datenbank verzeichnet ist. Daß die Textverarbeitung ihm nicht nur offensichtliche Rechtschreibfehler ankreidet, sondern auch Sätze wie "Hase und Igel macht ein Rennen". Vielleicht, daß das Eintippen von "Alles auf Diskette, was in der letzten Woche geändert wurde" im Nu zur Datensicherung führt.

Auch wenn Cys Alltagsweisheiten nun offenliegen, ist noch viel zu tun, um sie sinnvoll zu nutzen. Entwickler mit einer Idee und einem Internetanschluß können nun selbst anstreben, ihren Programmen einen Hauch Menschenverstand zu vermitteln. Aber auch Philosophen werden sich dafür interessieren. Unter dem Titel Ontologie mühen sie sich schließlich schon seit Jahrtausenden, das in Begriffe zu fassen, was ist.

Wer Cyc auf dem Web besichtigt, klickt sich durch 3000 vielfach vernetzte und penibel erläuterte Begriffe. Es sind nur ein paar Prozent der Konzeptwelt des Systems, die Spitze des Wissensberges, beginnend mit dem allumfassenden "Thing" und seinen Teilmengen "Individual Object" und "Intangible". Als "Intangible" gilt alles, was keine körperliche Gestalt hat. In der kompletten Cyc-Wissensbasis hängen allein 5000 speziellere Begriffe direkt unter Intangible. Nur ein Teil davon ist veröffentlicht, etwa "Process", "Holiday" und "Mathematical Object".

Die Sponsoren machen wegen der investierten Millionen Cyc nicht als Ganzes öffentlich. Neben der großen Masse an Detailkonzepten fehlen zum Beispiel die sprachverstehende Komponente und der Mechanismus für

: Künstliche Intelligenz: Die Wissensbasis Cyc will Computern Selbstverständlichkeiten beibringen. Jetzt kann sie über das

logische Schlüsse. Der veröffentlichte Kern des Systems soll Softwareentwicklern genügen, um mit Cyc-Anwendungen zu experimentieren. Douglas Lenat will mit der Cyc-Beschreibungssprache CycL einen Standard dafür setzen, wie Wissen überhaupt repräsentiert, also formal beschrieben werden kann.

Wenn es eine solche Norm gäbe, dann ließen sich Wissensbestände zwischen verschiedenen Anwendungen leichter austauschen: Ein Expertensystem zur Diagnose von Automotoren könnte etwa sein Wissen über die Maschinen an ein Programm übergeben, das technische Geräte automatisch dokumentiert, also ein Handbuch für Bedienung und Wartung erstellt. Für Texte, Bilder und Vektorgraphiken sind solche Standards alltäglich, für die Beschreibung von Wissen fehlt ein gemeinsames Format.

Breiteres Interesse und neue Einnahmen könnte Cyc gut gebrauchen, denn die alten Sponsoren meinen, genug investiert zu haben. Im vergangenen Jahr mußte Lenat mit seiner Truppe lernen, auf eigenen Füßen zu stehen. Unter dem Firmennamen Cycorp ist das Projekt jetzt im Besitz der Mitarbeiter und muß selbst Geld einbringen.

Zwei Firmen haben Cyc bisher für interne Aufgaben eingesetzt: Die eine verwaltet ein Archiv von mehreren hunderttausend Photos samt Bildunterschriften. Mit Cycs Hilfe vermag das Archivsystem auch Aufnahmen zu finden, die mit völlig anderen Wörtern beschrieben sind, als die Anfrage nennt. Cycorp nennt das Beispiel eines Bildes mit der Beschreibung "Soldat hält einer Frau seine Pistole an die Schläfe"; man würde es auch finden, wenn man nach "Ein Mensch in Gefahr" oder "Ein Mann bedroht jemanden" sucht.

Der zweite Anwender ist ein Versicherungskonzern, der seine Daten auf viele verschiedene Computersysteme und Datenbankstrukturen verteilt hatte. In dem entstandenen Dschungel waren Informationen kaum mehr konsistent zu halten: Stand in der Datenbank der Personalentwicklung noch geschrieben, daß Mitarbeiterin Müller Sachbearbeiterin ist, so war sie in den Tabellen der Buchhaltung schon korrekt als Abteilungsleiterin geführt. Cyc konnte diese verteilten Daten mit Hilfe seines Wissens über die verwendeten Begriffe in eine einzige Datenbank überführen.

Die Cyclists arbeiten auch mit Hochschulen zusammen, um Prototypen für neue Anwendungen zu bauen. In Deutschland ist das Ulmer Forschungsinstitut für Anwendungsorientierte Wissensverarbeitung (FAW) dabei. Die Schwaben versuchen mit Unterstützung von Cyc, Dokumentationen von technischen Produkten automatisch zu erzeugen. "Cyc hat ein großes kommerzielles Potential, aber es kann noch einige Jahre dauern, bis es soweit ist", sagt der FAW-Chef Franz Josef Radermacher. Er zählt zum Lager der Cyc-Optimisten, im Gegensatz etwa zu seinem Kollegen Wolfgang Wahlster vom Deutschen Forschungszentrum für Künstliche Intelligenz in Saarbrücken, der just die formale Struktur von Cyc für überholt und unzureichend hält.

Douglas Lenats ursprüngliche Hoffnung, daß spätestens Windows 99 dank Cyc-Technik dem PC-Benutzer das Leben leichter macht, wird sich auch kaum erfüllen. Den menschlichen Common sense wird Cyc mancherorts überzeugend vertreten, anderswo jedoch kaum ersetzen können, wie Lenat mittlerweile zugab. Viel versprochen, nur manches gehalten – ein Motto, das in der künstlichen Intelligenz schon öfter galt.

(c) *DIE ZEIT* 1996

49/1996