



Sprachförderung für Computer

TU Darmstadt: Adolf-Messer-Preis für Sprachtechnologien Chris Biemann

Bitte beachten Sie die Sperrfrist: Montag, 25. November, 16 Uhr

Darmstadt, 25. November 2013. Wie lehrt man Computer, natürliche Sprache zu verstehen, um so auch große Textmengen zu strukturieren und erschließbar zu machen? Diesem Problem hat sich an der TU Darmstadt Professor Dr. Chris Biemann mit seiner Forschergruppe am Fachbereich Informatik verschrieben. Für seine Arbeit wurde er heute mit dem Adolf Messer Preis ausgezeichnet, der mit 50.000 Euro dotiert ist.

Computer haben ihre eigene Sprache, die nach klaren, logischen Regeln funktioniert und stets eindeutig ist. „Die menschliche Sprache ist dagegen gewachsen und nicht immer logisch und eindeutig“, erklärt Chris Biemann. Soll ein Computer einen von Menschen geschriebenen Text „verstehen“, also interpretieren, muss man ihm beibringen, Wörter nicht nur als Ansammlung von Buchstaben zu sehen. Ein Rechner muss auch erschließen können, welches Konzept hinter einem Wort steht. Wichtig wird das zum Beispiel bei Polysemen – hier hat ein Wort mehrere Bedeutungen – oder Synonymen – mehrere Wörter, die alle ungefähr dieselbe Bedeutung haben. Chris Biemann erklärt das Problem, vor dem Computer dabei stehen, am Beispielsatz „Sie ging zur Bank und hob Geld ab“. Hinter dem Polysem „Bank“ könnten zwei verschiedene theoretische Konzepte stehen: eine Sitzgelegenheit oder ein Geldinstitut.

Menschen erschließen sich die jeweilige Bedeutung des Wortes „Bank“ aus dem Zusammenhang. Computer können das bislang kaum und wenn nur in kleinem Rahmen. „Diese Zuordnung automatisch zu bewerkstelligen und so Computern Wissen über unsere Welt beizubringen, ist nicht trivial“, sagt Biemann. Gemeinsam mit seiner Forschungsgruppe arbeitet er daran, große Datenmengen algorithmisch zu analysieren. Vereinfacht gesagt: Computerprogramme durchforsten riesige Textmengen, zum Beispiel 20 Jahrgänge Tageszeitungen, und entdecken dabei, dass manche Wörter typischerweise ähnlich verwendet werden. Nach und nach entstehen so zu allen Wörtern Listen mit ähnlich verwendeten Wörtern. Die „zweite Dimension“ eines Textes bildet sich ab. Der Computer lernt, Synonyme zu „verstehen“ und Polyseme zu erkennen und richtig zu deuten. Die Methode, zu einem schlichten Text einen „zweidimensionalen Text“ generieren zu lassen, ist eine Entwicklung der Forscher an der TU Darmstadt.

Kommunikation und Medien
Corporate Communications

Karolinenplatz 5
64289 Darmstadt

Ihre Ansprechpartnerin:
Silke Paradowski
Tel. 06151 16 - 32 29
Fax 06151 16 - 41 28
paradowski.si@pvw.tu-darmstadt.de

www.tu-darmstadt.de/presse
presse@tu-darmstadt.de



In weiteren Analyseschritten, an denen die Forscher der Arbeitsgruppe Sprachtechnologie derzeit arbeiten, erschließen die Auswertelgorithmen dann auch Oberkategorien für Wörter – so fallen Begriffe wie „Hemd, Rock, Hose“ alle in die Kategorie „Kleidung“ – und Verbindungen zwischen Wörtern: in der Formulierung „das blaue Kleid“ beschreibt ein Wort das andere. Später sollen Computer auch in der Lage sein, Kausalzusammenhänge zwischen Satzteilen und Aussagen, kurz, große übergeordnete Strukturen zu erkennen, die weit über das einzelne Wort hinausreichen.

Der Darmstädter Forschungsansatz lässt Computer eigenständig lernen. Das funktioniert auch in fremden Sprachen, ohne dass die Programmierer sie sprächen oder zuvor Wörterbücher einspeisten, sagt Biemann – vorausgesetzt, der Computer hat genug Ausgangsmaterial, das er auf Ähnlichkeiten und Strukturen abklopfen kann. „Es geht darum, Computer Strukturen aus dem vorhandenen Material – zum Beispiel großen Textmengen – heraus entwickeln zu lassen, nicht eine Struktur zu entwerfen, die dann mehr oder minder gut auf Texte angewendet wird“, erklärt Biemann. „Das ist das Besondere an unserer Forschung hier in Darmstadt.“

In der Praxis werden die an der TU Darmstadt entwickelten Methoden bereits eingesetzt, und zwar immer dann, wenn es ums Wissensmanagement geht. Suchanfragen in Internet-Suchmaschinen sind ein Beispiel. Da nicht genau absehbar ist, welche Formulierung und welche Suchbegriffe der Suchende wählen wird, muss die Suchmaschine in der Lage sein, mit einer Anfrage „intelligent“ umzugehen und sie zu interpretieren. „Immer, wenn ein Thema noch nicht gut erschlossen ist, hilft zweidimensionaler Text weiter“, sagt Chris Biemann.

Allerdings ist es nicht nur die praktische Verwertbarkeit, die ihn an seinem Forschungsgebiet Computerlinguistik und Sprachtechnologie fasziniert: „Sprache hat so viel mit menschlichem Denken zu tun – man kann sich die Frage stellen, ob Sprache und Denken nicht dasselbe ist“, sagt er. Einerseits gehe es darum, Intelligenz so zu formalisieren, dass Maschinen „intelligent“ agieren könnten. „Aber letztlich geht es auch darum zu verstehen, was den Menschen zum Menschen macht.“

Für seine Arbeit wird Chris Biemann in diesem Jahr mit dem Adolf-Messer-Preis ausgezeichnet. Das Preisgeld fließt zurück in die Forschung: Verschiedene weiterführende Teilprojekte, teils geleitet von renommierten Gastwissenschaftlern, werden ebenso finanziert wie ein neuer Rechner, der von seiner Architektur her darauf ausgelegt ist, große Datenmengen zu verarbeiten und damit die Forschung voranzutreiben.



Chris Biemann studierte Diplominformatik an der Universität Leipzig, wo er von 2003 bis 2007 auch promovierte. In seiner Dissertation brach er mit den Traditionen von Computerlinguistik und Sprachtechnologie, indem er zeigte, dass Maschinen computerlinguistische Aufgaben lösen können, ohne auf explizit gegebenes Wissen zurückzugreifen. Die Maschine lernt „von selbst“. Nach der Promotion arbeitete Biemann in San Francisco bei der semantischen Suchmaschine Powerset und später bei Microsoft Bing. 2011 folgte er dem Ruf auf die Juniorprofessur am Fachbereich Informatik der TU Darmstadt. Er leitet die Fachgruppe Sprachtechnologie. In den vergangenen beiden Jahren entstanden hier zahlreiche Publikationen sowie eine Reihe von Forschungsprojekten und Kooperationen.

Der mit 50.000 Euro dotierte Preis der Adolf Messer Stiftung wird jährlich verliehen. Er fördert die Forschung und Lehre von Nachwuchswissenschaftlerinnen und -wissenschaftlern an der TU Darmstadt. Gewürdigt werden herausragende Leistungen in den Naturwissenschaften, Ingenieurwissenschaften sowie Wirtschafts-, Sozial- und Geisteswissenschaften.

MI-Nr. 108/2013, sip