

Pressemitteilung

Universität des Saarlandes

Claudia Ehrlich

25.02.2015

<http://idw-online.de/de/news626293>

Forschungsprojekte
Informationstechnik, Sprache / Literatur
überregional



Sprachforscher bringen dem Computer bei, zwischen den Zeilen zu lesen

Michael Wiegand von der Universität des Saarlandes arbeitet daran, dass der Computer in Zukunft wie ein Mensch Meinungen, Stimmungen und Gefühle aus Äußerungen herauslesen kann. Gemeinsam mit Forschern der Universität Hildesheim will der Sprachtechnologe dazu beitragen, dass Computer den Sinn von Texten automatisch besser verstehen und nicht von vornherein an Wortwitz oder Sprachspiel scheitern. Hierzu kombinieren sie verschiedene Sprachanalyseverfahren, um den Rechner mit Informationen zu füttern, die hinter Wörtern und Sätzen stecken. Die Forschung wird von der Deutschen Forschungsgemeinschaft (DFG) in Höhe von 500.000 Euro gefördert, rund die Hälfte davon fließt an die Saar-Universität.

Schrecklich, dumm, schön, super – hinter diesen Wörtern steckt mehr: eine Wertung. Und sie können dazu noch überaus doppeldeutig sein: Schrecklich gut! Ganz schön schwer! Das hat ja super geklappt! – wenn in Wahrheit alles schiefgegangen ist. Sollen heutige Computer verstehen, was hier zwischen den Zeilen steht, sind sie „raus“. Fangfragen, Wortspiel, Ironie, mit denen Menschen Meinung, Stimmung oder Gefühl rüberbringen, werden für den Rechner zum Stolperstein – was sich bei automatischer Sprachübersetzung ebenso zeigt wie bei der Spracherkennung am Telefon. „Wörter und linguistische Regeln zu kennen, ist eben nicht genug, um hinter den Sinn zu schauen. Hier setzen wir an: Wir wollen dem Computer die fehlende Information geben, damit er Äußerungen, die Meinungen enthalten, in Texten automatisch identifizieren und analysieren kann“, sagt Michael Wiegand. Der promovierte Computerlinguist forscht an der Saar-Uni am Lehrstuhl für Sprach- und Signalverarbeitung von Professor Dietrich Klakow.

Die Forscher arbeiten mit zwei Sprachen: Englisch und Deutsch. „Für die deutsche Sprache besteht großer Bedarf, weil auf diesem Gebiet noch kaum etwas existiert. Im Englischen gibt es bereits Forschungsergebnisse, die wir mit unserer Arbeit ergänzen wollen“, sagt Wiegand. So gewann Supercomputer „Watson“ 2011 im US-amerikanischen Fernsehquiz „Jeopardy“ gegen die menschlichen Kandidaten auch bei Fragen, bei denen quergedacht werden musste. Beim Quizlösen half Watson damals übrigens auch eine Computerlinguistin der Saar-Uni. „Aber davon, Meinung oder gar Ironie zu erkennen, ist Watson noch weit entfernt“, sagt Wiegand.

Bislang wird in der Forschung bei der so genannten „Sentimentanalyse“ vor allem auf große Textmengen gesetzt: Zum Beispiel zählen Analyseprogramme Wörter, die mehr verraten: Je nachdem wie häufig etwa „schlecht“ oder „traurig“ vorkommen, kann der Rechner schätzen, ob es im Text eher positiv oder negativ zugeht. Demgegenüber begeben sich Wiegand und seine Forscherkollegen in ihrem Projekt sozusagen auf „molekulare“ Ebene: „Wir schauen auf den einzelnen Satz und die Wörter, konzentrieren uns also auf die feingranulare Analyse. Dort bestimmen wir, wer im Text etwas sagt, das mit Meinung zu tun hat“, erklärt Wiegand. So ermitteln die Forscher, ob zum Beispiel mit den gewählten Worten in Wahrheit das Gegenteil gemeint ist und übersetzen es dann so für den Computer, dass er diese Information erkennen und verarbeiten kann.

Um den Sinn hinter den Wörtern für den Computer sichtbar zu machen, kombinieren sie verschiedene Techniken der Sprachanalyse, unter anderem auch solche, die die Wörter in Beziehung zu ihrem gewöhnlichen Kontext setzen, sowie ontologische Verfahren, also solche die dem Rechner verraten, wie was womit im Text zusammenhängt. Die

Computerlinguisten erstellen zunächst von Hand eine Textsammlung, eine Ressource: Sie schreiben Wörtern ihre Bedeutungen zu, zum Beispiel werden dem Wort „dumm“ Informationen hinterlegt, die deutlich machen, dass es hier um eine Wertung geht. Diese manuelle „Annotation“, wie die Sprachwissenschaftler es nennen, werden sie automatisieren. Um zu überprüfen, ob die automatisierten Verfahren den Sinn richtig erfassen, vergleichen sie die Ergebnisse dann mit Hilfe ihrer manuell erfassten Textsammlung.

Ihre Ergebnisse wollen die Sprachtechnologien als lexikalische Ressource – also als eine Art „Wörterbuch“ für künftige computerlinguistische Forschung – zur Verfügung stellen. Ihr Projekt, das von der DFG gefördert wird, ist auf drei Jahre angelegt.

Hintergrund:

Der Campus der Universität des Saarlandes ist eine der Hochburgen der Sprachtechnologie. Die Fäden vieler internationaler Projekte laufen hier zusammen. Es wird erforscht, wie Mensch und Computer einander besser verstehen, wie Lügen im Internet entlarvt werden, ein neuer Sonderforschungsbereich untersucht die Informationsdichte sprachlicher Äußerungen. Ein Ziel des Exzellenzclusters „Multimodal Computing and Interaction“ ist es, Computersysteme zu bauen, die eine ähnliche Interaktion mit dem Benutzer ermöglichen, wie von Mensch zu Mensch. Die Forscher der Saar-Uni arbeiten eng mit dem Deutschen Forschungszentrum für Künstliche Intelligenz und den beiden Max-Planck-Instituten für Informatik und Software-Systeme zusammen, die auf dem Campus ansässig sind.

Kontakt: Dr. Michael Wiegand: Tel.: 0681 302 - 58 133,
E-Mail: michael.wiegand@lsv.uni-saarland.de

Ein Pressefoto für den kostenlosen Gebrauch finden Sie unter www.uni-saarland.de/pressefotos. Bitte beachten Sie die Nutzungsbedingungen.

Hinweis für Hörfunk-Journalisten: Telefoninterviews in Studioqualität sind möglich über Rundfunk-Codec (IP-Verbindung mit Direktanwahl oder über ARD-Sternpunkt 106813020001). Kontakt: 0681/302-2601, oder -64091.