



URL: [http://www.uni-jena.de/PM090407\\_HahnWM2.pdf](http://www.uni-jena.de/PM090407_HahnWM2.pdf)

## Computerlinguisten sind "Vize-Weltmeister"

### Team von Udo Hahn belegt Spitzenplatz bei internationalem Software-Wettbewerb

Man kennt Weltmeister im Fußball, Eiskunstenlaufen oder Schach - nun ist ein Forschungsteam der Universität Jena "Vize-Weltmeister in einer nicht alltäglichen Disziplin geworden. *"Man könnte es eine Weltmeisterschaft im Darstellen wissenschaftlich relevanter Textinhalte nennen"*, erklärt Prof. Dr. Udo Hahn von der Friedrich-Schiller-Universität Jena. Seine Arbeitsgruppe entwickelt seit mehreren Jahren Computersysteme, die in der Lage sind, inhaltliche Aussagen vollautomatisiert aus Millionen von Dokumenten herauszuziehen und sie in Datenbanken einzuspeisen, die dann weltweit von Forschern und Entwicklern angefragt werden.

In dem internationalen Software-Wettbewerb BioNLP'09 Shared Task, in dem die Jenaer Arbeitsgruppe mit ihrem System jetzt den zweiten Platz belegt hat, ging es um das Auffinden von wichtigen Forschungsergebnissen in biologischen Fachtexten. Damit sind etwa Regulations- und Bindungsprozesse zwischen Proteinen oder Kopiervorgänge von Genen gemeint, wie sie von Biologen täglich in Tausenden von Fachartikeln beschrieben werden.

#### Hilfe für Computer bei der Sinnfindung

*"Bislang scheitern Computer genau daran, womit Menschen beim Lesen solcher Texte keine Probleme haben", so Hahn. "Für einen Rechner ist ein englischsprachiger Text eine sinnfreie Abfolge einzelner Buchstaben, für die es, anders als für Zahlen, keine im Rechner eingebaute Interpretationshilfe gibt", skizziert der Jenaer Computerlinguist das grundlegende Problem. "Wir stellen Rechnern solche Hilfen bereit. Die Frage war bislang nur: Wie gut sind diese Hilfen?"*

Diese Frage wurde nun durch den internationalen Wettbewerb entschieden. Dabei lief alles wie bei einer "richtigen" Weltmeisterschaft ab. Die vereinbarte Disziplin war das Auffinden biologischer Prozesse, es gab eine öffentliche Bekanntgabe der Wettbewerbsregeln und eine Meldefrist für die Teilnehmer. Anhand gleicher "Spiel"-Daten wurde an vielen computerlinguistischen Instituten der Welt "trainiert", bis schließlich der Wettbewerb, überwacht von einem internationalen Schiedsrichtergremium, begann. *"Wir hatten im Wesentlichen sechs Wochen Zeit, unser System an die Anforderungen der Organisatoren von der Universität Tokio anzupassen, bevor es in die entscheidende Endphase ging", so Hahn. "Das war eine gewaltige Herausforderung für uns und nur durch das große Engagement meines gesamten Teams zu bewältigen, das sich teilweise bis tief in die Nächte und an den Wochenenden mit viel Elan eingesetzt hat"*, freut sich Prof. Hahn über den "Sportsgeist" seiner Arbeitsgruppe.

#### Das Jenaer System hat die meisten Herausforderungen bravourös gemeistert

Die Ausgangslage, also die bislang geheim gehaltenen Wettbewerbstexte, denen die Systeme weltweit ausgesetzt waren, war für alle Teilnehmer identisch. Nur so konnte bestimmt werden, wer am Ende die besten, aber auch wer die schlechtesten Ergebnisse des Wettbewerbs erzielt hatte. *"Unser System hat die meisten Herausforderungen jedoch bravourös gemeistert und uns damit sozusagen zum ‚Vize-Weltmeister‘ gemacht"*, so der Computerlinguist von der Universität Jena.

Zu dem weltweiten Wettbewerb hatten sich ursprünglich mehr als 40 wissenschaftliche Institutionen angemeldet. 24 Teams haben sich letztendlich den strengen Bewertungsmaßstäben ausgesetzt. *"Dass wir 22 der weltweit besten Arbeitsgruppen unseres Fachgebiets schlagen konnten und nur ein einziges Team besser als unseres ist, empfinde ich als eindrucksvolle Bestätigung der Qualität unserer Arbeit, die in der Fachwelt aufmerksam registriert und gewürdigt wird"*, so Hahn. Die Methoden der Gewinnerteams werden demnächst bei einer Fachtagung und in der einschlägigen Fachpresse veröffentlicht. Damit bilden sie auch die Basis für neue Entwicklungen auf dem Gebiet. *"Dass wir etwas zu diesem Fortschritt beitragen können, macht uns alle sehr stolz"*, berichtet Prof. Hahn.

*"Wir wollen letztlich dazu beitragen, dass das wissenschaftliche Know-how, das weltweit mit großem Aufwand an öffentlichen und privaten Forschungsgeldern geschaffen wird, nicht ungelesen in Zeitschriftenaufsätzen verstaubt, sondern aktiv zum Einsatz kommt - sei es in der Krebs- oder Altersforschung, der Arzneimittelentwicklung oder über öffentlich zugängliche Gesundheitsportale."*

**Kontakt:**

Prof. Dr. Udo Hahn

Institut für Germanistische Sprachwissenschaft der Universität Jena

Fürstengraben 30, 07743 Jena

Tel.: 03641 / 944320

E-Mail:

Meldung vom: 07.04.2009 12:31 Uhr