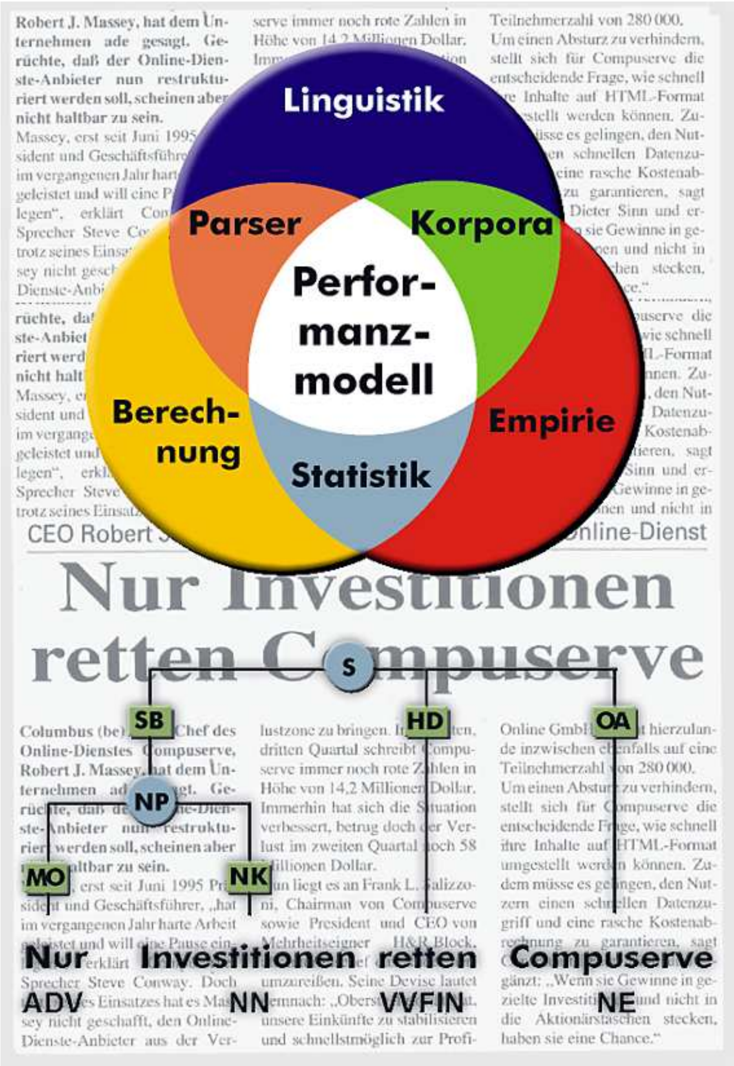
NEGRA: Nebenläufige grammatische Verarbeitung



NEGRA ist ein Gemeinschaftsprojekt von Wissenschaftlern aus der Computerlinguistik und der Informatik. Ziel ist Entwicklung neuer Technologien zur grammatischen Verarbeitung. Dies geschieht durch die Entwicklung neuer Berechnungsverfahren sowie die Kombination und Weiterentwicklung von Wissenschaftsbereichen, die bisher getrennte Ansätze verfolgt haben.

NEGRA verbindet linguistische Forschung, empirische Untersuchungen von Sprachdaten und die Entwicklung von Berechnungsverfahren. In dem Projekt werden aktuell gesammelte Sprachdaten mit linguistischen Interpretationen verbunden. Die so erzeugten Korpora bieten somit adäquate Beschreibungen tagtäglich auftretender Sprachphänomene sowie Angaben über deren Häufigkeit. Statistische Verarbeitungsverfahren nutzen diese Häufigkeitsverteilungen, um neue, bisher ungesehene Sprachdaten zu analysieren. Sie lernen anhand voranalysierter Daten und liefern anschließend eine sehr robuste und fehlertolerante Verarbeitungsmethode. Constraintbasierte Parser hingegen nutzen explizit, in der Regel manuell kodiertes linguistisches Wissen, um ebenfalls Analysen neuer Texte zu erstellen. Schwerpunkt sind nebenläufige Berechnungsverfahren, die durch quasi-parallele Prozesse große Effizenz erlauben.

Wie die constraintbasierte und die statistikbasierte Verarbeitung unter Beibehaltung ihrer jeweiligen Vorteile zusammengeführt werden können, ist ein ungelöstes Problem, an dem das Projekt NEGRA arbeitet. Genau diese Zusammenführung verspricht entscheidende Fortschritte für die Sprachverarbeitung.

Als wichtiges Teilergebnis wurde im Projekt NEGRA das erste deutsche linguistisch interpretierte Textkorpus aufgebaut. Es besteht zur Zeit aus aus ca. 20.000 Sätzen Zeitungstext der Frankfurter Rundschau und wächst täglich. Die Analysen wurde unter Verwendung der neu entwickelten Anätze semiautomatisch erstellt und bilden so in einem Bootstrapping-Prozess die Grundlage für Untersuchungen über die Verwendung von Sprache, das automatische Lernen und die Weiterentwicklung der Parsingtechnologie auch außerhalb des Sonderforschungsbereiches.

Prof. Dr. Werner H. Tack
SFB 378 – der Sprecher
Universität des Saarlandes
Postfach 151150
D-66041 Saarbrücken

URL: http://www.coli.uni-sb.de/info/projects/negra.html
Kontakt: Prof. Dr. Hans Uszkoreit
Telefon: +49-681-302-4115
Telefax: +49-681-302-4700
email: hansu@coli.uni-sb.de