



SEMPRIA GmbH, Grafenberger Allee 277–287, 40237 Düsseldorf

SEMPRIA GmbH
Grafenberger Allee 277–287
40237 Düsseldorf

Telefon: 0211 566693-57
Telefax: 0211 566693-58
info@sempria.de
<http://www.sempria.de/>

Pressemitteilung vom 10.02.2011

David und Goliath - Während der Weltkonzern IBM sein Frage-Antwort-System WATSON in einen spektakulären Wettstreit mit menschlichen Quiz-Gegnern schickt, erneuert ein deutsches Startup mit ähnlichen Technologien des Sprachverstehens die Archivsuche und die lokale Suche in Websites (Site-Search).

Zwei ganz unterschiedliche Institutionen, ein Weltkonzern und eine Forschergruppe aus Hagen, hatten zunächst einen ähnlichen Ausgangspunkt für die Entwicklung Ihrer Software-Systeme: ein Frage-Antwort-System zu schaffen, das sich bei internationalen, universitär geprägten Wettbewerben für Suchsysteme gut schlägt.

Der Weltkonzern (IBM) hat mit großem Aufwand sein System in eine Spezialmaschine verwandelt, die nun bei der bekannten US-Quiz-Show Jeopardy antritt. Und wohl gewinnen wird. Dazu wird jedoch ein Supercomputer aus mehreren Tausend Rechnern benötigt.

Die Forscher aus Deutschland sind mit ihrer Suchmaschine (SEMPRIA) bei kleineren Rechnerclustern geblieben, um eine Nutzung der neuen, logisch und linguistisch fundierten Technologie auch interessierten Anwendern wie Verlagen und Archiven zu ermöglichen, die nicht über Supercomputer verfügen. Im Kontakt mit diesen Anwendern haben sie ermittelt, was Nutzer wirklich weiterbringt. Ausformulierte Fragen wie bei WATSON werden von SEMPRIA zwar auch beherrscht, sind aber nicht so häufig, wie man denken könnte. Eher gesucht wird mit zwei bis drei Begriffen, die in einem bestimmten Sinnzusammenhang stehen sollen, wie "Niederlage des FCK" oder "Kritik von Obama". Da dieser Sinnzusammenhang von traditionellen Suchmaschinen wie Google und Bing nicht verstanden wird, hängen Nutzer oft einfach nur Substantive aneinander: "niederlage fck" oder "kritik obama". Interessanterweise führen die gegensätzlichen Suchbedürfnisse "Niederlagen gegen den FCK" und "Kritik an Obama"

zu denselben Anfragen in der Google-Suchsprache. Bei der deutschen Suchmaschine SEMPRIA werden die vermeintlich kleinen Wörter wie "an" und "von" im Obama-Beispiel hingegen genutzt, um die Genauigkeit der Suche zu steigern. Zudem können die vorhandenen Informationen durch ganz unterschiedliche Formulierungen erfragt werden, die dasselbe Suchinteresse ausdrücken. Dass SEMPRIA die vorhandenen Texte auch bei abweichender Wortwahl findet, bedeutet gerade für die Archivsuche einen Riesenvorteil.

"Diese bedeutungsorientierte Suche für das Deutsche gelingt uns durch Technologien, die wir ursprünglich für die Fragebeantwortung geschaffen haben", erklärt Prof. Hermann Helbig, unter dessen Leitung SEMPRIA entstanden ist. "SEMPRIA analysiert jedes Textdokument und jede Suchanfrage inhaltlich. Es ermittelt die Zusammenhänge zwischen den Begriffen und findet auf dieser Basis passende Textstellen durch logischen Vergleich zwischen Suchanfrage und Text." Und warum macht dies Google nicht? "Der sprachverstehende Ansatz ist um mehrere Größenordnungen aufwändiger als die Verfahren von Google - und damit in den nächsten Jahren als Websuche nicht rentabel zu betreiben", so Helbig. Wohl aber als Archivsuche und sogenannte Site-Search, wie es momentan in SEMPRIA-Testsystemen bei einigen Verlagen nachgewiesen wird.

Für Rückfragen:

Dr. Sven Hartrumpf, E-Mail: s.hartrumpf@sempria.de, Telefon: 0211/566693-57

Belegexemplar erbeten