

Mit Bison gegen den Datenberg

Prof. Michael Berthold sucht nach einem Informationsnetzwerk zur Unterstützung menschlicher Kreativität

Prof. Michael Berthold hat den idealen Mitarbeiter im Auge, wenn er das neue EU-Projekt erklärt: Dieser ist schon seit Jahrzehnten in der Firma, kennt jeden, hat alles Relevante gelesen, jedes Experiment gesehen, jeden Artikel durchgelesen, mit jedem geredet. Zu guter Letzt kann er all dies auch jederzeit aus dem Gedächtnis abrufen und interessante Verbindungen aufzeigen. Ein Kollege mit einem Problem schildert diesem Alleswisser sein Anliegen, und der zieht zwei Artikel aus dem Ordner, die möglicherweise interessante Lösungsansätze beschreiben.

Der ideale Mitarbeiter hat nicht die Lösung des Problems parat, sondern gibt eine Anregung, wie die Lösung möglicherweise zu finden ist. Genau das soll das Software-System leisten, das unter der Überschrift „Bisociation Networks for Creative Information Discovery“, kurz „Bison“, ab kommenden Juni von einem EU-Konsortium mit acht assoziierten Mitgliedern entwickelt werden soll. Von dem Informationsnetzwerk wird erwartet, dass es die menschliche Kreativität unterstützt angesichts einer überwältigenden Masse an Informationen.

Bisher muss man die Frage schon kennen, wenn man sich an eine Datenbank wendet, um Informationen zu erhalten. Die Datenbanken arbeiten, wie die meisten semantischen Netzwerken, innerhalb einer Domäne: Auf drei bestimmte Gene würden sie ein viertes Gen assoziieren. Berthold, Lehrstuhlinhaber für Bioinformatik und Information Mining an der Universität Konstanz und Koordinator des Projekts, und seine acht Partner wollen, dass ihr System bissoziiert: d.h. es assoziiert über Domänengrenzen hinweg. Die drei Gene werden hier nicht nur mit einem vierten Gen in Verbindung gebracht, sondern mit etwas ganz anderem, mit etwas, das auf den ersten Blick vielleicht nichts damit zu tun hat.

„Wir wollen verschiedenste Informationsquellen in einem losen Netzwerk zusammenfassen“, erklärt Michael Berthold. Die Netzwerke funktionieren über automatische Datenanalysen. Dabei muss die Datenbank zum Beispiel einen Text nicht notwendigerweise „verstehen“, sondern wertet nur statistisch aus, wie häufig ein bestimmtes Wort oder bestimmte Wortkombinationen vorkommen. Je mehr Informationsquellen zusammenkommen und je öfter Gemeinsamkeiten auftauchen, desto wahrscheinlicher ist es, dass beides etwas miteinander zu tun hat. Das kann für

Texte gelten, für Experimente, Bildmaterial usw. Der Benutzer soll am Ende sagen, das ist interessant, und möglicherweise eine neue Fragestellung erreichen.

Der Prototyp am Ende der drei Förderjahre, die im Rahmen des EU-Programms „Future and Emerging Technologies“ finanziert werden, käme einem Paradigmenwechsel in der Wissens- und Informationsmanagement-Forschung gleich. Der neue Weg läuft auf Systeme hinaus, die keine klaren Antworten mehr geben, sondern versuchen, stimulierende, interessante Verknüpfungen zu finden. Vom klaren Frage-Antwort-Spiel muss man sich ohnehin verabschieden, meint der Informatiker Berthold. „Wir ertrinken in Daten und können gar nicht immer nur sauber gepflegte Informationsnetze aufbauen. Wir können es uns aber mittlerweile auch leisten, nur anhand von Korrelationen, von statistischen Zusammenhängen, Verbindungen herzustellen“, so Michael Berthold. Angesichts des Datenbergs, der immer schneller wächst, komme man nicht mehr hinterher mit der Katalogisierung von Wissen.

Anstatt nur lange Listen von ähnlichen Dokumenten (oder Bildern) zu liefern, müssen Informationsmanagementsysteme in Zukunft auch helfen, den Riesenberg von möglicherweise verwandten Informationen schnell und effizient zu explorieren – ähnlich wie das menschliche Hirn ja auch keine lange Liste von „hits“ liefert, sondern sehr schnell aussortieren kann, was im Moment nicht interessant ist, und sich auf das Wesentliche und gerade Interessante konzentriert.



Prof. Michael Berthold hat seit 2003 den Nycomed-Stiftungslehrstuhl, vormals ALTANA-Stiftungslehrstuhl, für Angewandte Informatik mit Schwerpunkt Bioinformatik und Information Mining an der Universität Konstanz inne.

msp.

