idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

Pressemitteilung

Hasso-Plattner-Institut für Softwaresystemtechnik (HPI)
Hans-Joachim Allgaier

15.06.2012
http://idw-online.de/de/news483607

Forschungsprojekte, Wissenschaftliche Tagungen
Gesellschaft, Informationstechnik, Politik, Recht, Wirtschaft
überregional

Data Mining: Hasso-Plattner-Institut erläutert Ansätze der Grundlagenforschung

"Forschung missbraucht keine Daten, sondern analysiert lediglich, wie man gesammelte Informationen nutzen kann." In so deutlicher Weise äußerte sich Prof. Hasso Plattner zum Thema Data Mining. Seine Keynote hielt er auf einem Symposium der HPI Research School, dem internationalen Forschungskolleg des Hasso-Plattner-Instituts, an der Universität Potsdam. Dabei bezog sich Plattner auch auf die jüngsten Ereignisse rund um das vom HPI gestoppte Schufa-Grundlagenforschungsprojekt.

"Es ist klar, dass in Deutschland personenbezogene Daten geschützt sind und bleiben müssen", sagte Plattner. Allerdings sei es heute schon ebenso selbstverständlich, dass von Anwendern frei veröffentlichte Informationen von Unternehmen gesammelt und ausgewertet würden.

Beispielsweise interessierten sich heute bereits Unternehmen dafür, wie oft Anwender über ein bestimmtes Produkt kommunizieren. Die Zahl der Erwähnungen sind eine verlässliche Größe bei der Bewertung einer Produkteinführung.

Allerdings müsse es auch beim Data Mining "klare rechtliche und ethische Grundlagen" geben, sagte Plattner und ergänzte: "Für den Schutz der Persönlichkeitsrechte ist der Gesetzgeber verantwortlich." Damit der Schutz wirksam sei, müsse erforscht werden, wie sich strukturierte und unstrukturierte Daten verknüpfen lassen.

Die Berichterstattung zum vorzeitig beendeten Forschungsvorhaben der Schufa in Kooperation mit dem Hasso-Plattner-Institut hat den Begriff "Data Mining" in das Blickfeld der Öffentlichkeit gerückt. Was sich dahinter verbirgt, blieb in den Medien oft aber unklar.

Was ist Data Mining?

Beim Data Mining geht es um das „Schürfen“ nach Informationen im World Wide Web. Eingesetzt werden dabei mathematische und statistische Methoden der Mustererkennung in großen „Bergen“ von frei zugänglichen Daten. Die Vorgehensweise soll Forschern dabei helfen, bestimmte Schlüsse aus den riesigen Mengen strukturierter und unstrukturierter Daten zu ziehen.

Gesucht wird beim Data Mining zum Beispiel nach einzigartigen Besonderheiten in riesigen Mengen von frei zugänglichen Daten im Internet oder nach Ähnlichkeiten, um sie in Beziehung zu setzen, Verbindungen herzustellen und Gruppen und Objektklassen zu bilden. Das hilft bei der Anlage und sachgerechten Interpretation von Käufer-, Partner- oder Job-Profilen.

Gefragt wird aber auch danach, ob Daten überhaupt einen Zusammenhang mit anderen haben. Die Ergebnisse solch logischer Verknüpfungen von Informationen findet man etwa auf Plattformen wie Amazon oder im iTunes Store: Zu Produkten erhalten die Käufer dort neben der Beschreibung auch einen Hinweis darauf, was andere Käufer dieses

Produkts darüber hinaus erworben oder zumindest angesehen haben. Amazon generiert daraus beispielsweise Angebotspakete nützlicher Produktzusammenstellungen.

Aber nicht nur "seelenlose" Software sucht nach Verknüpfungen. Nach Angaben der Zeitschrift "Computerwoche" gaben in einer Umfrage 2010 bereits 22 Prozent von 230 befragten Personalentscheider an, Informationen über Bewerber im Internet und hier vor allem in sozialen Netzwerken zu suchen und zu interpretieren.

Data Mining - ein Forschungsfeld am HPI

Auch das HPI forscht zum Thema Data Mining unter dem Aspekt "Wie lassen sich sinnvolle Verknüpfungen zwischen einzelnen frei zugänglichen Daten herstellen, um eine Interpretation zu ermöglichen?" Bereits 2010 präsentierte das HPI das Projekt "GovWild" zur Transparenz von Regierungsdaten und erhielt dafür den "Scalable Data Analytics for a Smarter Planet Innovation Award" von IBM. In einem ähnlichen Data Mining-Zusammenhang hätte auch das gemeinsame Grundlagenforschungsprojekt mit der Schufa zum Tragen kommen sollen.

Wie jedes Forschungsobjekt hat aber auch Data Mining seine Tücken. Funktioniert eine sinnvolle Verknüpfung der Daten überhaupt und wie verhindert man eine falsche Interpretation? Und welche Schwierigkeiten es zudem mit automatisierter Textanalyse geben kann, lässt sich gut am Beispiel von Übersetzungssoftware erklären. Wer bereits einmal versucht hat, einen deutschen Text zunächst in eine beliebige Fremdsprache und anschließend zurück ins Deutsche zu übersetzen, der wird als Ergebnis einen Text erhalten, der sehr an eine verunglückte Bedienungsanleitung von Geräten aus Fernost erinnert.

Das Ziel der HPI-Informatikwissenschaftler war es daher nie, Grundlagenforschung rund ums Text Mining zur Bewertung der Zahlungsfähigkeit einer Person heranzuziehen. Beide, die Schufa und das Hasso-Plattner-Institut, hatten bereits in den ersten Gesprächen klargestellt, dass sich solche explorative Grundlagenforschung über den Einsatz von Textanalyseprogrammen nicht für Bonitätsprüfungen eignen würde oder verwenden ließe.

Da Data Mining ausschließlich frei zugängliche Internet-Informationen verwenden kann, ist der Schutz von persönlichen Daten enorm wichtig. Auch wenn Technik und Gesetzgeber für ein hohes Maß an Datenschutz sorgen, enthebt das den Einzelnen nicht der Aufgabe, mit den eigenen Daten verantwortungsvoll umzugehen. Das HPI lehrt seine Studierenden deshalb seit Langem, besonders sparsam und vorsichtig im Umgang mit Internetdaten zu sein.

Das HPI kümmert sich, wie jüngst beim Thema neuer Internetstandard IPv6, zusammen mit dem Bundesdatenschützer aber auch um den Schutz der Privatsphäre bei neuen Technologien (www.hpi.uni-potsdam.de/presse/mitteilung/beitrag/datenschutz-schaar-und-ipv6-rat-zum-internet-der-neuen-generation.html). Betont wird auch die Internetsicherheit: So lehrt das HPI in einem Tele-Lab (www.tele-lab.org) den Schutz vor Attacken auf den eigenen Rechner und stellt unter anderem Internetsicherheitsvorlesungen kostenfrei und offen für jeden ins Netz (www.tele-task.de). Mit dem „Lock-Keeper“ hat das HPI eine Hochsicherheitsschleuse gegen das Online-Hacken im Internet entwickelt (www.lock-keeper.org).

Den mit der SCHUFA Holding AG geschlossenen Vertrag über eine Zusammenarbeit bei der Grundlagenforschung rund um technische Verarbeitung öffentlicher Web-Daten hatte das HPI noch gekündigt, bevor die Forschungsarbeiten überhaupt begonnen hatten. Grund war, dass angesichts mancher Missverständnisse in der Öffentlichkeit über den vereinbarten Forschungsansatz und darauf aufbauender Reaktionen ein solches wissenschaftliches Projekt nicht unbelastet und mit der nötigen Ruhe hätte durchgeführt werden können.

URL zur Pressemitteilung: http://www.hpi.uni-potsdam.de/news/beitrag/stets-zu-diensten-super-computing-modular-hpi-forschungssymposium-beendet.html - Bericht über das Future SOC Symposium der HPI Research School

Prof. Hasso Plattner äußerte sich in Potsdam zur Erforschung des Data Mining
HPI/K. Herschelmann