

Aktuelle Meldung

Fehlerquelle Dubletten: HPI-Studierende entwickeln Benchmark zur effizienten Datenbereinigung

8. Juli 2021

Potsdam. „Eine Studie zeigte, dass bis zu 20% der Patientendaten in US Krankenhäusern Duplikate sind. Daraus resultieren fehlerhafte Behandlungen und hohe Kosten“, sagte Florian Sold, einer von vier Bachelorstudierenden des Hasso-Plattner-Instituts (HPI), welche die Vergleichsplattform „Snowman“ für die Duplikaterkennung in Datensätzen (Data Matching) entwickelten. Die Forschungsergebnisse wurden sowohl auf der internationalen Fachkonferenz „VLDB“ eingereicht als auch beim digitalen Bachelorpodium des HPI am 8. Juli vorgestellt.

Die in Kooperation mit SAP SE entwickelte Vergleichsplattform "Snowman" erlaubt es, Data Matching Lösungen auf Basis verschiedener mathematischer Funktionen (Metriken) gegeneinander zu vergleichen und so die qualitativ beste Lösung für ein gegebenes Matching-Problem zu finden.

Gleichzeitig kann „Snowman“ eingesetzt werden, um die Entwicklung von Data Matching Lösungen zu unterstützen: Neben klar messbaren Ergebnissen weisen sorgfältig ausgewählte Daten-Beispiele sowie Visualisierungen auf vorhandene Stärken und Schwächen hin und ermöglichen somit lösungsorientierte Weiterentwicklungen. Zudem können Data Matching Lösungen innerhalb ihres gesamten Lebenszyklus analysiert und somit direkt im wirtschaftlichen Kontext evaluiert werden.

Somit kann der Projektpartner SAP SE neue Verfahren aus der Forschung und Entwicklung evaluieren, einschätzen in welcher Form sich aktuelle SAP Matchingsysteme für neue Anwendungsgebiete eignen und nicht zuletzt aktuelle Verfahren optimal konfigurieren. "Snowman schafft eine höhere Transparenz bei unseren Algorithmen und macht unsere Prozesse noch effizienter. Dadurch verkürzen sich unsere Feedbackzyklen und wir können schneller unsere Produkte verbessern", lobt Ashish Shrivastava, Data Matching Professional bei SAP SE. Die Vergleichsplattform wird bereits in mehreren internationalen Teams des Projektpartners aktiv einsetzen.

Betreut wurde die Bachelorprojektgruppe von Prof. Dr. Felix Naumann, Leiter des Fachgebiets Informationssysteme.

Bachelorpodium – Ausweis der praxisnahen Ausbildung am HPI

Das „Bachelorpodium“ des Hasso-Plattner-Instituts gibt es bereits seit dem Jahr 2005. Seitdem präsentieren Bachelorstudierende des HPI einmal im Jahr öffentlich die Ergebnisse ihrer Praxis-Projekte, die sie in Teams von vier bis



acht Studierenden am Ende des Bachelorstudiums absolviert haben. Sie zeigen, wie sie zwei Semester lang - von ihren Professorinnen und Professoren angeleitet - größere praktische Aufgaben der Informationstechnologie eigenverantwortlich angepackt und welche innovativen Lösungen für Wirtschaft und Gesellschaft sie dabei entwickelt haben. Projektgeber sind renommierte Unternehmen und Institutionen aus ganz Deutschland. Eine Übersicht über die laufenden Projekte gibt die HPI-Internetseite <http://hpi.de/studium/it-systems-engineering/bachelor/bachelorprojekte.html>

Kurzprofil Hasso-Plattner-Institut

Das Hasso-Plattner-Institut (HPI) in Potsdam ist Deutschlands universitäres Exzellenz-Zentrum für Digital Engineering (<https://hpi.de>). Mit dem Bachelorstudiengang „IT-Systems Engineering“ bietet die gemeinsame Digital-Engineering-Fakultät des HPI und der Universität Potsdam ein deutschlandweit einmaliges und besonders praxisnahes ingenieurwissenschaftliches Informatikstudium an, das von derzeit rund 700 Studierenden genutzt wird. In den vier Masterstudiengängen „IT-Systems Engineering“, „Digital Health“, „Data Engineering“ und „Cybersecurity“ können darauf aufbauend eigene Forschungsschwerpunkte gesetzt werden. Bei den CHE-Hochschulrankings belegt das HPI stets Spitzenplätze. Die HPI School of Design Thinking, Europas erste Innovationsschule für Studenten nach dem Vorbild der Stanforder d.school, bietet jährlich 240 Plätze für ein Zusatzstudium an. Derzeit sind am HPI 21 Professorinnen und Professoren sowie über 50 weitere Gastprofessoren, Lehrbeauftragte und Dozenten tätig. Es betreibt exzellente universitäre Forschung – in seinen IT-Fachgebieten, aber auch in der HPI Research School für Doktoranden mit ihren Forschungsaußenstellen in Kapstadt, Irvine, Haifa und Nanjing. Schwerpunkt der HPI-Lehre und -Forschung sind die Grundlagen und Anwendungen großer, hoch komplexer und vernetzter IT-Systeme. Hinzu kommt das Entwickeln und Erforschen nutzerorientierter Innovationen für alle Lebensbereiche.

Pressekontakt: presse@hpi.de

Christiane Rosenbach, Tel. 0331 5509-119, christiane.rosenbach@hpi.de und
Carina Kretzschmar, Tel. 0331 5509-177, carina.kretzschmar@hpi.de