

08.09.2026

Vom Preisvergleich in die KI-Forschung

billiger.de und Universität Mannheim entwickeln gemeinsamen Entity-Matching-Benchmark

Karlsruhe/Mannheim. Welches Angebot gehört zu welchem Produkt? Was für Verbraucherinnen und Verbraucher beim Online-Shopping selbstverständlich erscheint, ist für Preisvergleichsplattformen eine komplexe Aufgabe. Produktbezeichnungen, Beschreibungen und Attribute können sich je nach Anbieter teilweise erheblich unterscheiden. Damit Nutzerinnen und Nutzer auf billiger.de dennoch zuverlässig Preise vergleichen können, müssen Millionen Händlerangebote den richtigen Produkten zugeordnet werden. Im Online-Handel wird dieser Prozess als „Entity Matching“ bezeichnet.

Die Forschungsgruppe „Web-based Systems“ um Professor Dr. Christian Bizer an der Universität Mannheim hat nun in Zusammenarbeit mit billiger.de einen frei verfügbaren Benchmark für dieses Forschungsfeld veröffentlicht. „Billiger.de Products“ umfasst 13.730 Angebote, die 2.168 unterschiedlichen Produkten zugeordnet sind. Nach Angaben der Projektgruppe handelt es sich, nach aktuellem Kenntnisstand, um den größten öffentlich verfügbaren Entity-Matching-Benchmark für deutschsprachige Produktdaten.

Eine Besonderheit liegt in der Mehrsprachigkeit des Benchmarks: Neben den deutschen Produktdaten stehen auch automatisch erzeugte englische Übersetzungen zur Verfügung. Damit zählt „billiger.de Products“ zu den wenigen öffentlich verfügbaren mehrsprachigen Benchmarks in diesem Forschungsfeld.

„Der billiger.de Products Benchmark ist besonders anspruchsvoll, da er ein breites Spektrum an Produktkategorien abdeckt und sowohl das Matching strukturierter als auch unstrukturierter Produktbeschreibungen erfordert. Die Mehrsprachigkeit des Benchmarks ermöglicht es zudem zu untersuchen, wie gut sich in einer Sprache erlerntes Wissen auf Produktangebote in anderen Sprachen übertragen lässt“, so Prof. Dr. Christian Bizer.

Thilo Gans, Geschäftsführer der solute GmbH, sieht den engen Austausch zwischen Wissenschaft und Wirtschaft als wichtigen Baustein für die Entwicklung neuer KI-Anwendungen:

„Unternehmen können reale Fragestellungen und Daten aus der Praxis einbringen, während die Wissenschaft neue Ansätze entwickelt und unabhängig untersucht. Genau aus dieser Verbindung können Lösungen entstehen, die am Ende auch einen konkreten Mehrwert für Verbraucherinnen und Verbraucher schaffen.“

Reale Produktdaten aus dem Preisvergleich

Die Grundlage des Benchmarks bilden reale Produktdaten aus dem Umfeld von billiger.de. Die Angebote wurden zunächst von Expertinnen und Experten bei solute/billiger.de manuell den jeweiligen Produkten zugeordnet. Anschließend wurden die Daten der Forschungsgruppe der Universität Mannheim zur Verfügung gestellt. Dort erfolgten die Vereinheitlichung und weitere Aufbereitung der Daten sowie eine KI-gestützte Übersetzung ins Englische. Darüber hinaus wurden die Daten überprüft und mögliche Fehler korrigiert.

„Wir arbeiten bei billiger.de täglich daran, Angebote aus unterschiedlichsten Shops zuverlässig denselben Produkten zuzuordnen. Dass unsere von Expertinnen und Experten kuratierten Daten nun auch der Forschung zur Verfügung stehen, schafft einen wertvollen Austausch zwischen Praxis und Wissenschaft. Langfristig profitieren davon vor allem Nutzerinnen und Nutzer von E-Commerce- und Preisvergleichsplattformen“, sagt Johannes Knopp, Lead Developer bei solute und verantwortlich für das Thema Matching.

KI-Verfahren unter realistischen Bedingungen testen

Auf Grundlage der Produktdaten entwickelte die Forschungsgruppe verschiedene Trainings-, Validierungs- und Testdatensätze. Damit lässt sich untersuchen, wie zuverlässig unterschiedliche Entity-Matching-Verfahren erkennen, ob zwei Händlerangebote tatsächlich dasselbe Produkt beschreiben. Eindeutige Produktkennzeichnungen wurden dafür bewusst entfernt. Eine besondere Herausforderung bilden Produkte, die sich in Bezeichnung und Beschreibung stark ähneln, sich aber beispielsweise in Größe, Farbe, Ausstattung oder

Packungsmenge unterscheiden. Gleichzeitig wurde untersucht, wie gut die Verfahren mit Produkten umgehen können, die zuvor nicht Teil der Trainingsdaten waren. Die entwickelten Datensätze ermöglichen es so, unterschiedliche Matching-Verfahren unter vergleichbaren und realitätsnahen Bedingungen zu testen.

Vom Smartphone bis zum Werkzeug

Der Benchmark deckt insgesamt **13 unterschiedliche Produktkategorien** ab und bildet damit verschiedene Bereiche des Online-Handels ab. Besonders stark vertreten sind Bekleidung und Accessoires mit 25,7 Prozent, Möbel und Wohnen mit 24,4 Prozent sowie Elektronik und Computer mit 15,7 Prozent. Weitere Produktdaten stammen unter anderem aus den Bereichen Werkzeuge und Heimwerken, Sport und Freizeit, Spielwaren, Automotive, Kosmetik, Gesundheit, Bürobedarf und Lebensmittel.

Die große Bandbreite ist für die Forschung besonders relevant: Je nach Händler können identische Produkte unterschiedlich bezeichnet und beschrieben werden. Entity-Matching-Systeme müssen diese Unterschiede erkennen und dennoch zuverlässig bestimmen, welche Angebote tatsächlich zum selben Produkt gehören.

Wenn Wissenschaft auf Wirtschaft trifft

Mit „Billiger.de Products“ verbinden die Universität Mannheim und billiger.de wissenschaftliche Forschung mit konkreten Herausforderungen aus der E-Commerce-Praxis. Während billiger.de reale und von Expertinnen und Experten kuratierte Produktdaten zur Verfügung stellt, schafft die wissenschaftliche Aufbereitung einen frei verfügbaren Benchmark, mit dem neue Verfahren entwickelt und bestehende Ansätze miteinander verglichen werden können. Von diesem Austausch können beide Seiten profitieren: Die Wissenschaft erhält Zugang zu realitätsnahen Daten, während neue Forschungsergebnisse langfristig dazu beitragen können, technologische Lösungen im E-Commerce weiterzuentwickeln.

Der frei verfügbare Benchmark schafft damit nicht nur eine neue Grundlage für die internationale Forschung zum Entity Matching. Das gemeinsame Projekt zeigt zugleich, wie der Wissenstransfer zwischen Wissenschaft und Wirtschaft dazu beitragen kann, KI-

PRESSEINFORMATION



Forschung näher an reale Anwendungen zu bringen und Innovationen im E-Commerce voranzutreiben.

Website des Benchmarks: <https://wbsg-uni-mannheim.github.io/billiger-de-products/>

Preis-Nachfrage-Index: <https://www.billiger.de/data>

Einkaufskalender: <https://www.billiger.de/einkaufskalender>

Medienkontakt: Diskurs Communication // Jürgen Scheurer // T +49 30 2023536-11 // presse@billiger.de //

Weitere Informationen unter: <https://company.billiger.de/presse/>

Über billiger.de

Seit 2004 verschafft billiger.de Online-Shoppern einen Überblick über die besten Angebote aus über 22.500 Shops. Der Preisvergleich von billiger.de ist neutral und unabhängig. Die Angebote der Shops werden mehrmals täglich aktualisiert – dies gewährleistet Internetnutzern und auch den Händlern Transparenz und Übersicht im Online-Markt. Neben dem Preisvergleich stellt billiger.de umfassende Produktinformationen, wie Produktdaten, -bilder, -videos sowie Testberichte, Nutzerbewertungen oder Ratgeber bereit. So werden die Nutzer mit zusätzlicher Beratung unterstützt und können sich für das richtige Produkt entscheiden.