

Datenintegration im Web mit vernetzten Daten

Andreas Harth and Denny Vrandečić

Karlsruher Institut für Technologie

harth@kit.edu

denny.vrandecic@kit.edu

Abstract: Das Tutorial gibt einen Überblick über Inhalte und Techniken von vernetzten Daten (Linked Data). Das World Wide Web entwickelt sich von einem Medium der Dokumente zu einem Medium der Daten. In den letzten Jahren werden vermehrt Daten nach den von Tim Berners-Lee postulierten "Linked Data" Prinzipien veröffentlicht, von Organisationen wie der Deutschen Nationalbibliothek oder der New York Times bis hin zu kleinen Gruppen und Privatpersonen. In der sogenannten "Linking Open Data Cloud" sind dutzende Datensätze verzeichnet; der gesamte Datenbestand beläuft sich auf Milliarden von Datenelementen und wächst ständig. Die Prinzipien sind begründet in der Architektur des Web, und basieren auf Standards wie dem Hypertext Transfer Protocol (HTTP), Universal Resource Identifiers (URI), dem Resource Description Framework (RDF) und (optional) der Anfragesprache SPARQL. Die Prinzipien stellen einen einheitlichen Rahmen zur Publikation von und zum Zugriff auf Daten dar. Durch die Vernetzung der Daten wird es möglich, dezentral Daten zu publizieren, und die derart publizierten Daten miteinander zu verknüpfen. Das Tutorial behandelt die Grundlagen vernetzter Daten, geht auf etablierte Praktiken zum Veröffentlichen dieser Daten ein, und zeigt anhand von Fallstudien, wie diese Daten in Anwendungen integriert, angefragt und visualisiert werden können.