

Von Mediatoren über Ontologien zu Linked Open Data - Zum Stand der Informationsintegration und -fusion aus Sicht der Life Sciences

Kai-Uwe Sattler

Fakultät für Informatik und Automatisierung
Technische Universität Ilmenau
Postfach 100 565
D-98684 Ilmenau
kus@tu-ilmenau.de

Abstract: Die Verknüpfung von Daten aus unterschiedlichen Quellen ist in vielen Anwendungsbereichen eine zentrale Aufgabe. In den vergangenen 25 Jahren wurde daher eine Vielzahl von Ansätzen zur Lösung dieses Problems entwickelt. Während jedoch im Business-Bereich Data-Warehouse-Systeme als Integrationsplattform fest etabliert sind, stellen die Lebenswissenschaften besondere Anforderungen. So wurden in den letzten Jahren unzählige wissenschaftliche Datensammlungen veröffentlicht und über das Internet zugänglich gemacht, die in verschiedenster Weise genutzt werden können. Eine Verknüpfung und übergreifende Analyse der Daten wird jedoch durch Systemgrenzen und Heterogenitäten erschwert. Benötigt werden daher Techniken zur Informationsintegration, die nicht nur einen transparenten Zugriff auf Daten aus anderen Systemen ermöglichen, sondern auch komplexe Extraktions-, Transformations- und Analyseschritte automatisieren können und dabei die Charakteristika der Daten wie Qualität, Kontext und Vertraulichkeit berücksichtigen.

Der Beitrag gibt einen Überblick zum Stand der Forschung im Bereich Informationsintegration und betrachtet neben klassischen Datenbanktechniken auch aktuelle Ansätze zur semantischen Integration und Fusion. Im Mittelpunkt stehen dabei die Eignung für Life-Science-Anwendungen sowie die speziellen Herausforderungen dieser Domäne.