

Schema- und Metadatenmanagement in Peer Data Management Systemen

Felix Naumann

HPI Potsdam

(eingeladener Vortrag)

Abstract:

Peer Data Management Systeme (PDMS) sind die natürliche Erweiterung föderierter Informationssysteme. An Stelle einer zentralen Komponente, die ein integriertes Schema hält, treten einzelne Peers, die sowohl die Rolle von Datenquellen als auch die Rolle einer integrierten Komponente annehmen. Aufgrund der Verteilung und Autonomie der einzelnen Peers sind Architektur und Anfrageplanung in PDMS stark von Methoden des Schemamanagement und des Metadatenmanagement abhängig.

Der Aufbau eines PDMS ist geprägt durch die Erstellung von Schema Mappings zwischen den Schemata einzelner Peers. Nur sie erlauben die Umschreibung von Anfragen eines Peers zu Anfragen an benachbarte Peers. Techniken des Schema Matching, also der automatischen Erstellung von Schema Mappings, stehen in diesem Zusammenhang besonderen Herausforderungen gegenüber, da nur deren ad-hoc Erstellung und Veränderung der Dynamik eines PDMS gerecht wird.

Relevante Metadaten betreffen allerdings nicht nur die Schemata der Peers, sondern auch deren Daten. Um, wie in herkömmlichen DBMS, Abschätzungen über die Kardinalitäten von Anfrageergebnissen und somit auch über Anfragekosten treffen zu können, müssen entsprechende Statistiken über benachbarte Peers erhoben und aktualisiert werden. Eine besondere Schwierigkeit ergibt sich durch die Tatsache, dass die von einem benachbarten Peer erhaltenen Anfrageergebnisse nicht unbedingt von jenem Peer selbst stammen, sondern von dahinter liegenden Peers bereitgestellt worden sein können.