

„P2P based RDF Querying and Reasoning for Grid Resource Description and Matching“

Felix Heine

Zusammenfassung

Diese Dissertation behandelt das Problem der **Ressourcenauffindung** in großen, heterogenen Grids. Weltweite Grids bringen zwei Aspekte dieses Problems besonders in den Vordergrund: **Heterogenität** und **Größe**.

Heterogenität bedeutet, dass viele unterschiedliche Typen von Ressourcen im Grid vorkommen. Neben den traditionellen Ressourcen wie Hochleistungs-Clustern sowie Speichergeräten, werden zusätzlich beliebige Typen von Diensten, Anwendungen sowie teuren physikalische Spezialinstrumente als Ressourcen behandelt. Daher kann kein Standard jeden Ressourcentyp beinhalten, den es zu Beschreiben gilt. Sobald mehrere Standards verwendet werden, wird zusätzliches Wissen benötigt, um zwischen den unterschiedlichen Standards zu vermitteln.

Größe bedeutet dass eine skalierbare Lösung zum Problem der Ressourcenauffindung benötigt wird. Die heute schon existierenden Reasoning-Systeme erwarten typischerweise, dass alles Wissen an einer Stelle gesammelt wird. Dies funktioniert nicht für beliebig große Informationssammlungen.

Diese Arbeit beschreibt ein System, das die ersten Schritte zur Lösung des geschilderten Problems beiträgt. Der Umfang der Arbeit geht allerdings über das Grid Computing hinaus, auch wenn dies die motivierende Anwendung darstellt. Das Ziel der Arbeit ist die Entwicklung eines **skalierbaren Systems, welches große Sammlungen von maschinenlesbaren Informationen kombiniert und abfragt**.

Die Hauptelemente der Arbeit sind die Beschreibung einer Systemarchitektur basierend auf einem strukturierten P2P Netzwerk, ein Verteilungsalgorithmus der Informationen an wohldefinierten Knoten platziert, ein Reasoning-Mechanismus der neues Wissen aus existierendem ableitet und dabei Informationen von unterschiedlichen Knoten kombiniert, sowie zwei Auswertungsstrategien für Anfragen.

Die erste Auswertungsstrategie ermittelt alle Treffer für eine gegebene Anfrage. Dazu nutzt sie verschiedene Strategien, um die Netzbelastung zu minimieren. Bei Anfragen mit sehr vielen Treffern ist es allerdings unrealistisch, alle Treffer zu ermitteln. Daher ergänzt eine zweite Strategie für solche Fälle das System. Diese Strategie ermittelt nur eine begrenzte Menge von Treffern, in Abhängigkeit von einem Sortierkriterium. Der Algorithmus nutzt Caching und Vorausschau-Methoden, um eine effizient Ausführung zu gewährleisten.

Das System wurde prototypisch implementiert. Der Prototyp belegt mit zahlreichen Resultaten die Effizienz des Systems, sowohl durch Simulationen als auch durch echte Benchmarks.

Schlüsselwörter

Grids, Ressourcenbeschreibung, Ressourcenauffindung, Semantic Grid, Informations-Integration, Semantic Web, RDF, RDFS, Reasoning, SPARQL, Anfrageauswertung, Top k Anfrageauswertung, Skalierbarkeit, Datenorganisation, P2P, DHT, Strukturierte P2P Netzwerke