

Das MOMIS Schema-Merging Verfahren

Rainer Dobiasch

1 Einleitung

Bei der Zusammenführung von Daten, wie zum Beispiel in Integrationsplattformen großer Firmen wo die Daten zwischen Customer-Relationship-Management Systemen und der Abrechnung abgeglichen werden müssen, stellt sich immer wieder die Aufgabe des Schemen-Abgleichs. Es gibt hier die Aufgabe der reinen Transformation (Anpassung von einlangenden Quelldaten an das Schema einer Zieldatenbank, z.B. Webdatenextraktion) sowie die Herausforderung der Integration heterogener Datenhaltungssysteme (z.B. bei einem Webportal).

In der Literatur finden sich bereits zahlreiche Ansätze zu diesen Themen. Dieser Vortrag soll das Verfahren, welche im MOMIS¹ Projekt an der Universität von Modena entwickelt wurde vorstellen.

2 MOMIS Architektur

Wenn hier von MOMIS gesprochen wird, ist dies leider stark vereinfacht. In Wirklichkeit handelt es sich hier um ein Zusammenspiel vieler Komponenten verschiedenster Projekte (SI-Designer, MOMIS, ARTEMIS, ODB-Tools,...), wobei die Grenzen oft nicht ganz klar sind.

MOMIS erlaubt die Integration verschiedenster Datenquellen zu einem Gesamtsystem. Im Zentrum dieses Systems sitzt der Mediator. Dieser benötigt als Basis das sogenannte *globale Schema*, eine Integration der Schemata aller zu integrierenden Datenquellen. Anfragen an das Gesamtsystem können von dem Mediator zerlegt, an die entsprechenden Datenquellen weitergeleitet und die Ergebnisse wieder zusammengefasst werden. Da aber die Datenquellen sehr unterschiedlicher Natur sein können (Relationale Datenbanken, XML, ...) ist es notwendig jeder Datenquelle eine Wrapper vorzuschalten, der für die notwendigen Umsetzungen sorgt.

Im Vortrag soll diese Architektur kurz vorgestellt werden, das Hauptthema ist allerdings das Verfahren mit welchem das *globale Schema* erstellt wird.

2.1 Das Globale Schema

Bei der Erstellung des *globalen Schemas* gilt es ein Schema zu finden, in welchem sich die Schema-

ta der zu integrierenden Datenquellen wiederfinden können. Eine solche Aufgabe wird als *Schema Merging* bezeichnet, und die Kunst ist, sie möglichst automatisiert (mit minimaler Benutzerinteraktion) zu lösen.

MOMIS löst diese Aufgabe in drei Schritten:

1. Aufbau eines *Common Thesaurus*

Der *Common Thesaurus* beinhaltet dann alle Begriffe der zu integrierenden Schemata und die Beziehungen zwischen ihnen (BT/RT/SYN). Hier bedient sich MOMIS der Hilfe der WordNet Datenbank, einem Semantischen Netz, welches von der Princeton University entwickelt wurde.

2. Clustering der Elemente

Anhand der Beziehungen im Common Thesaurus werden die Elemente (Klassen/Relationen der Schemata) in einem hierarchischen Clusteringverfahren zu Gruppen (Clustern) zusammengefasst.

3. Aufbau des globalen Schemas

Für jeden Cluster wird schließlich ein Element (Klasse) im *globalen Schema* erstellt. In diesem Schritt gilt es, diese Element mit ihren Attributen zu definieren und auch die Namen für die Elemente und Attribute im Globalen Schema festzulegen.

Im Rahmen des Vortrages sollen diese drei Schritte im Detail betrachtet werden.

2.2 Zu meiner Person

Ich habe an der TU-Wien Industrielle Elektronik und Regelungstechnik studiert und arbeite seit 1988 bei Siemens im Bereich Programm und Systementwicklung. Ich arbeitete an Kraftwerkseinsatzoptimierung, Zuflussprognosen und Expertensystemprototypen für Nutzlastüberwachung in der Raumfahrt. Zur Zeit betreue ich Product Lifecycle Management Systeme. Privat beschäftige ich mit dem Thema *Schema Matching* im Rahmen einer Dissertation.

¹Mediator environment for Multiple Information Sources