

**Flexible und konfigurierbare
Software-Architekturen für datenintensive
ontologiebasierte Informationssysteme**

Vom Promotionsausschuss der
Technischen Universität Hamburg-Harburg
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

von

Michael Wessel

aus Hamburg

2008

Gutachter:

Prof. Dr. Ralf Möller

Prof. Dr. Friedrich Mayer-Lindenberg

Prof. Dr. Dieter Gollmann (Drittgutachter)

Tag der mündlichen Prüfung:

14.4.2008

Kurzfassung

Ontologien werden voraussichtlich eine wichtige Rolle für die nächste Generation von Informationssystemen spielen. Heutzutage erfährt insbesondere die Vision des Semantischen Webs (Semantic Web) einige Beachtung - soll doch durch seine Realisierung nicht nur die "intelligente" Beantwortung inhaltsbasierter Suchanfragen ermöglicht werden, sondern beispielsweise auch Dienste im Web automatisch anhand funktionaler Dienstbeschreibungen ermittelt, miteinander kombiniert, und entspr. invoziert werden. Ontologien sind jedoch nicht nur im Semantischen Web relevant, sondern können in einer Vielzahl anderer Informationssystem-Domänen durch die Bereitstellung eines reichhaltigen logisch-formalen Begriffsrahmens der Domäne, einer sog. Konzeptualisierung, neuartige und innovative Zusatzdienste für Informationssysteme erbringen.

In dieser Arbeit wird ein formaler und softwaretechnischer Architekturrahmen zur Konstruktion ontologiebasierter Informationssysteme unter formalen und softwaretechnischen Gesichtspunkten entworfen und realisiert. Zur Sicherstellung einer hohen Flexibilität und Anwendbarkeit in mehreren Informationssystem-Domänen wird der Rahmen so entworfen, dass nicht nur einzelne Punkte im Informationssystem-Entwurfsraum, sondern ganze Regionen abgedeckt und adressiert werden können. Aktuelle Semantic-Web-Sprachen, wie die Web Ontology Language (OWL), können hier ebenfalls Anwendung finden.

Durch durchgängige Betrachtung zweier sog. Leitdomänen für Informationssysteme werden domänenspezifische Aspekte von der Analyse über den Entwurf bis hin zur Realisierung prototypischer ontologiebasierter Informationssysteme in diesen Leitdomänen in den Rahmen einfließen. Zum einen wird die Domäne des Semantischen Webs betrachtet, und zum anderen die Domäne der geografischen bzw. räumlichen Informationssysteme. Ein wesentlicher Schwerpunkt der Arbeit liegt auf der Realisierung einer Maschine zur ontologiebasierten Anfragebeantwortung. Die Arbeit adressiert alle hierfür zu lösenden Teilproblemgruppen, insbesondere das Repräsentations-, das Anfragebeantwortungs-, und das Inferenzproblem. Die sog. Beschreibungslogiken bilden ein weiteres wichtiges Standbein, stellen sie doch die theoretische Fundierung für aktuelle Ontologie-Sprachen (wie OWL) bereit. Die Arbeit beschreibt zudem den Entwurf und die Implementierung der Beschreibungslogik-Anfragesprache nRQL, die von vielen Nutzern des bekannten Beschreibungslogiksystems RacerPro verwendet wird. nRQL ist ein wichtiger Bestandteil des in dieser Arbeit entworfenen und realisierten Gesamtrahmens.

Danksagungen

Ich danke allen Personen und Institutionen, die mich auf dem langen Weg bis zur Vollendung dieser Arbeit unterstützt haben. Insbesondere danke ich meiner Familie, und natürlich meinen Lehrern: Prof. Ralf Möller, Prof. Volker Haarslev, Prof. Bernd Neumann, und auch Prof. Peter Schefe, dem ich die ersten Kontakte mit der funktionale Programmierung im Rahmen der Informatik-Anfängervorlesung der Universität Hamburg im Wintersemester 1991/1992 zu verdanken habe, und dessen KI-Vorlesung später mein Interesse an Fragestellungen der Wissensrepräsentation geweckt hat. Vielen Dank an Prof. Mayer-Lindenberg für die Bereitschaft, die Rolle des Zweitgutachters zu übernehmen und diese doch recht umfangreiche Arbeit zu bewerten. Prof. ter Haseborg und Prof. Gollmann möchte ich für ihr Engagement in meinem Promotionsverfahren danken. Bei Prof. Leonie Dreschler-Fischer und Prof. Christian Freksa möchte ich mich für die Hilfe bei der Beantragung eines Grundstipendiums der Universität Hamburg, und bei der Universität Hamburg für das bewilligte Grundstipendium in den Jahren 1999 - 2001 bedanken. Bei den Kollegen und Studenten der Arbeitsbereiche KOGS (Fachbereichs Informatik, Universität Hamburg) und STS (Technische Universität Hamburg-Harburg) möchte ich mich für unzählige interessante Diskussionen und Anregungen bedanken, ebenso wie bei Kay Hidde von Racer Systems und Mike Fischer. Meinem Freundeskreis danke ich für motivierende Gespräche.

Hamburg, im Juli 2007

Inhaltsverzeichnis

1	Einleitung und Übersicht	7
1.1	Zum Vorgehen in dieser Arbeit	19
1.2	Ergebnisse der Arbeit	20
1.3	Struktur der Arbeit und weitere Anmerkungen	21
2	Benötigte Grundlagen aus Theorie und Praxis	23
2.1	Logikbasierte Wissensrepräsentation	25
2.2	Beschreibungslogiken	58
2.3	Zusammenfassung	121
3	Ontologiebasierte Informationssysteme – Zwei Leitdomänen	123
3.1	Leitdomäne 1: Geographische Informationssysteme	129
3.2	Leitdomäne 2: Information Retrieval im SEMANTIC WEB	159
3.3	Zusammenfassung	183
4	Ein formales Rahmenwerk für ontologiebasierte Informationssysteme	187
4.1	Das Substrat-Datenmodell	189
4.2	Eine generische Anfragesprache für Substrate	199
4.3	Zusammenfassung	227
5	Softwaretechnische Realisierung des formalen Rahmenwerks	229
5.1	Realisierung des Substrat-Datenmodells	233
5.2	Realisierung der Substrat-Anfragebeantwortungsmaschine	276
5.3	Vier konkrete Substrat-Anfragesprachen	316
5.4	Zusammenfassung	396
6	Bereitstellung offener und flexibler DL-Beweiser im Rahmenwerk	397
6.1	Motivation und Anforderungen	399
6.2	Entwurf von MIDELORA	410
6.3	Implementation von MIDELORA	486
6.4	Vorstellung einiger MIDELORA-Instantiierungen	497
6.5	Zusammenfassung	520
7	Evaluation und Ausblick	521

Inhaltsverzeichnis

7.1	Evaluation der Anfragebeantwortungsmaschine	523
7.2	Evaluation von MIDELORA	540
7.3	Schlusswort und Ausblick	550
A	Anhang	553
A.1	LUBM-TBox	553
A.2	LUBM-Taxonomie	556
A.3	DISK 1	557
A.4	DISK 2	558
A.5	Verzeichnis wichtiger Akronyme, Abkürzungen und Namen	559
	Literaturverzeichnis	553