

Seminar

Semantische Datenintegration

18.4.2005

Joachim Hammer <hammer@tzi.de>

Sebastian Hübner <huebner@tzi.de>

TZI - Technologie-Zentrum Informatik
Universität Bremen

Veranstalter und Webseite

- ▶ Joachim Hammer <hammer@tzi.de>
Büro: TZI/Seekamp, Raum 11, 1 . Stock, Universitätsallee 21-23
- ▶ Sebastian Hübner <huebner@tzi.de>
Büro: TZI/Seekamp, Raum 16, 1 . Stock, Universitätsallee 21-23
- ▶ Sprechzeiten: nach Vereinbarung
- ▶ Homepage der Veranstaltung:
[http://www.tzi.de/~cyco/lehre/semdatint.html](http://www.tzi.de/~cyco/lehre/semmdatint.html)

Semantische Datenintegration

- ▶ Situation: mehr und mehr Daten werden elektronisch erfasst, verarbeitet und gespeichert
 - Kosten für Speicherplatz sinken kontinuierlich (2005: 1 GB ≈ US \$ 1; 1995: 1GB ≈ US \$ 1,000)
 - Beispiele: Firmeneigene Intranets, WWW, Digital Libraries, Human Genome Project
- ▶ Wunsch: einheitlicher Zugriff auf verschiedene Datenquellen in verteilten Umgebungen
- ▶ Problem: Datenquellen unabhängig voneinander entwickelt
 - Unterschiede in Struktur und Bedeutung (Semantik)
- ▶ Ziele: Beseitigung von strukturellen und semantischen Heterogenitätskonflikten, Konsolidierung von Daten, einheitliche Repräsentation
- ▶ Meist manueller und daher sehr fehlerbehafteter Prozess
 - Flaschenhals in der Anwendung moderner Informationssysteme

Die Fragestellungen

- Was sind die wissenschaftlichen Grundlagen der Datenintegration?
- Welche verschiedenen Ansätze der semantischen Datenintegration gibt es?
- Worin unterscheiden sie sich und wann werden sie am sinnvollsten eingesetzt?
 - Welche Annahmen werden gemacht, wie sinnvoll sind diese?
 - Nach welchen Kriterien kann man die Ansätze klassifizieren?
- Welche Vor- und Nachteile haben die einzelnen Ansätze?
 - Wie gut funktionieren sie in der Praxis?

Organisation und Ablauf des Seminars

- ▶ Joachim und Sebastian halten die Einführungs- und (vielleicht) Abschlussvorlesung
 - Hängt von der Anzahl der Studenten und Themen ab
 - Moderieren die Präsentationen der Studenten und stehen als Ansprechpartner während der Recherche zur Verfügung
- ▶ Studenten erarbeiten sich den Stoff zu einem bestimmten Thema und präsentieren ihre Ergebnisse in einem Vortrag
 - Referenzpapier als Startpunkt der Recherche
 - Auf Fragestellungen Rücksicht nehmen - Gliederung und Inhalt mit Veranstalter vorher besprechen
 - Schriftliche Ausarbeitung zur Vertiefung der Materie
- ▶ Ergebnisse des Seminars sind am Ende des Semesters in Form eines Sammelbandes aller Ausarbeitungen verfügbar

Scheinbedingungen

1. Nachweis, wissenschaftlich arbeiten zu können
 - a) **Vortrag:** 45 Minuten (mit anschließender Diskussion)
 - b) **Ausarbeitung:** 10 Seiten
 - c) **Einreichen** der technischen Ausarbeitung **innerhalb von 4 Wochen**

Falls Vortrag kürzer (mehrere Vorträge pro Seminar), Ausarbeitung bis zu 15 Seiten

2. Zur Verfügung gestelltes Material ist Ausgangsbasis für eigene Recherchen
3. Mitarbeit im Seminar (Diskussion) – das setzt Anwesenheit voraus!
 - ✓ Bitte pünktlich erscheinen

Hinweise zum Vortrag

- ▶ Inhalt
 - eigenständige Wiedergabe der Fakten
 - Berücksichtigung der Fragestellungen und des bisherigen Inhalts des Seminars

- ▶ Darstellung
 - sinnvoll gegliedert, verständlich aufbereitet
 - eine Woche vorher mit einem der Veranstalter besprechen (Termin ausmachen)

- ▶ Präsentation
 - ansprechend präsentiert
 - Verwendung von Folien - wenn kein eigener Laptop, PowerPoint Präsentation bitte einen Tag vorher zuschicken

Wissenschaftliches Arbeiten umfasst auch eine adäquate Präsentation der Ergebnisse!

Hinweise zur Ausarbeitung

▶ Inhalt

- eigenständige Wiedergabe der Fakten
- Berücksichtigung der Fragestellungen und des bisherigen Inhalts des Seminars

▶ Darstellung

- sinnvoll gegliedert
- verständlich aufbereitet

▶ Technisches

- Bis zu 15 Seiten LaTeX, Standard `\documentclass{article}`
 - Titelseite, Abbildungs- oder Tabellenverzeichnis nicht Teil der Mindestseitenzahl!
- Ordentliches Literaturverzeichnis, keine Link-Sammlung!
 - Nur Originalquellen, z.B. Wikipedia nicht zulässig
- Abgabe bis spätestens 4 Wochen nach dem Vortrag **als PDF**, inkl. Vortragsfolien

Themen (Einstiegsblock)

Termin	Thema (1-3)	Referenzpapier	Name
25. April	Klassifikation von Integrationskonflikten	Wache 2002, Kap. 2.2 (deutsch)	
2. Mai	Strategien zur Datenintegration inkl. Mediator	Wache 2002, Kap. 2.3-2.4 (deutsch)	
9. Mai	Ontologien und Ontologiesprachen		
16. Mai	<i>Pfingsten - Seminar entfällt</i>		

Die Themen (Vertiefungsblock)

Termin	Thema (4-9)	Referenzpapier	Name
23. Mai	(Semantische) Beschreibung von Datenquellen		
30. Mai	Schema Matching	Bernstein & Rahm 2000, eng.	
6. Juni	Integrationsformalisten		
6. Juni	Datenintegration mit SIMS		
13. Juni	Datenintegration mit Observer		
13. Juni	Datenintegration mit COIN		

Die Themen (Vertiefungsblock)

Termin	Thema (10-14)	Referenzpapier	Name
20. Juni	Datenintegration mit <i>Piazza</i>		
20. Juni	Datenintegration mit <i>Buster</i>		
27. Juni	Datenintegration mit IBM's Information Integrator		
27. Juni	Datenintegration mit BEA's Liquid Data		
4. Juli	Wohin geht die Forschung? Z.B. Integration von Sensordaten, "On-the-fly" Datenintegration, usw.		