

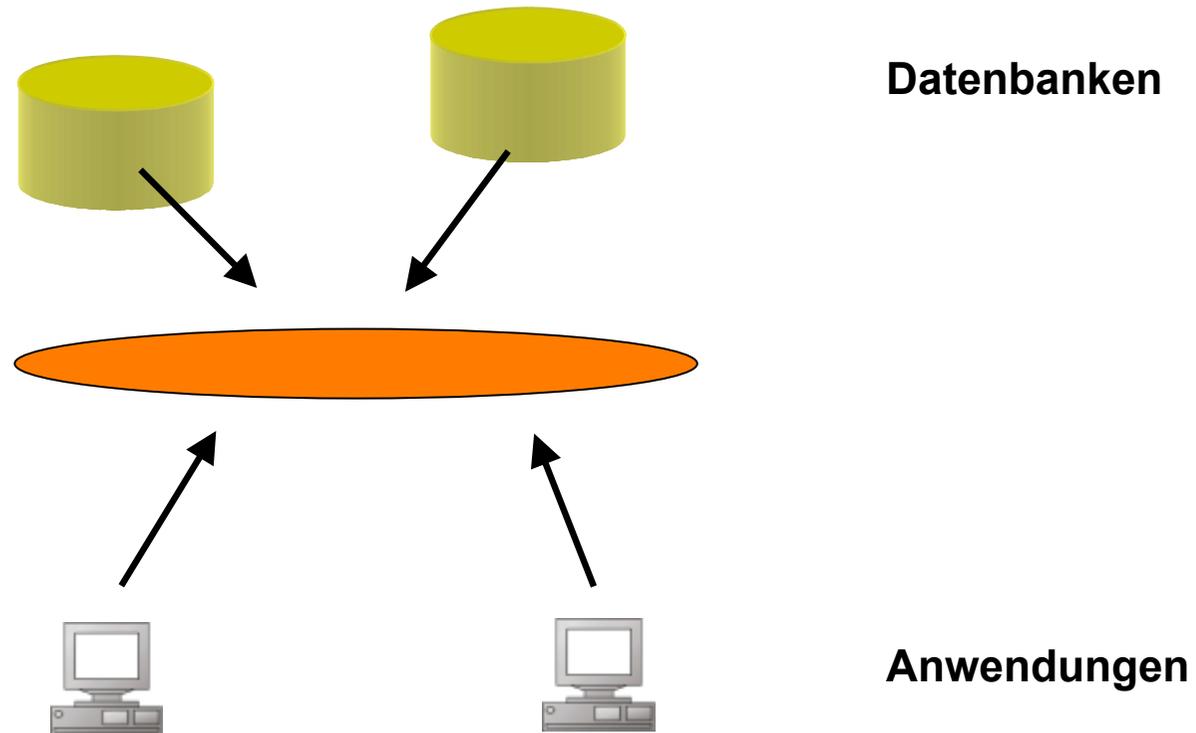
Semantische Datenintegration:

Strategien zur Integration von Datenbanken

Inhalt

1. Wiederholung Integrationskonflikte
2. Klassische Strategien zur Integration
 1. Eng gekoppelte Ansätze
 2. Lose gekoppelte Ansätze
3. Der Mediator
 - Anforderungen an die Beschreibung der Daten
 - Anforderungen an den semantischen Mediator

Problemstellung



1. Wiederholung: Integrationskonflikte

- Strukturelle Heterogenitätsk.
- Semantischen Heterogenitätsk.
- Inkonsistenz- und Redundanzprobleme

Strukturelle Heterogenitätskonflikte

- Bilateral (1:1)
 - Bezeichnerkonflikte
 - Datentypkonflikte
 - ...
- Multilaterale (n:n)
 - Verteilte Informationen
- Metaebene-Diskrepanzen

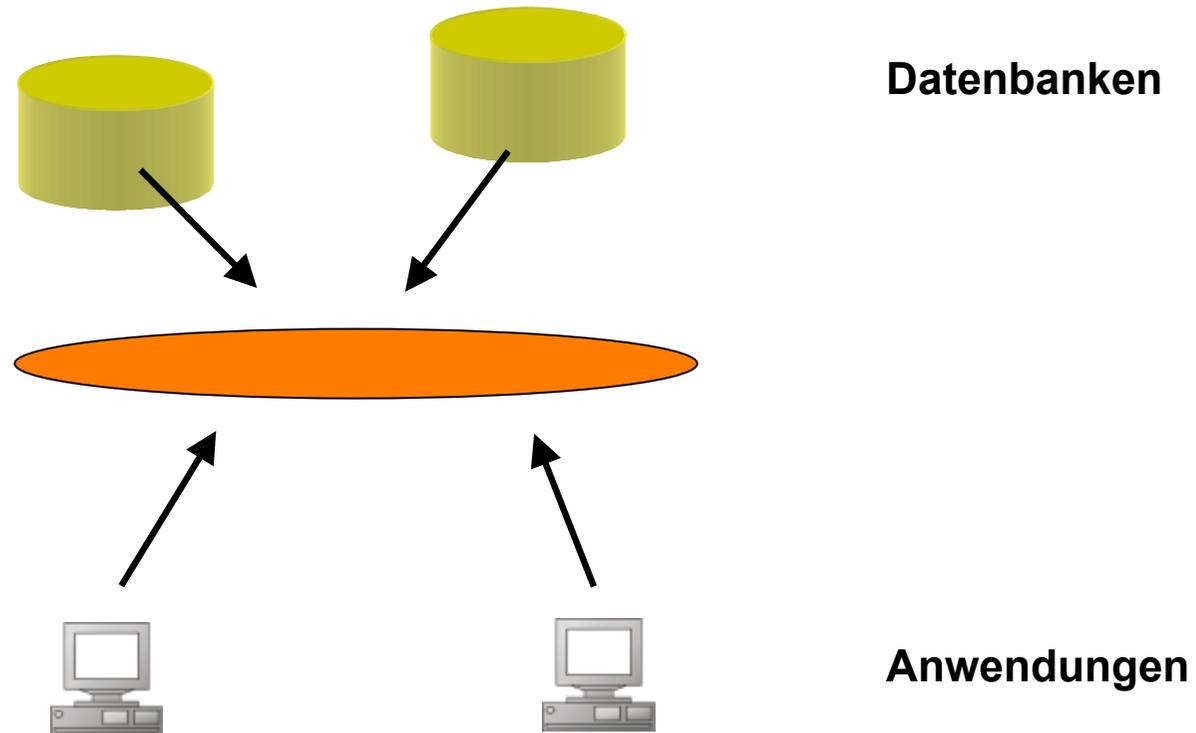
Semantische Heterogenitätskonflikte

- Datenheterogenität
 - Skalierungs- und Einheitenkonflikte
 - Repräsentationskonflikte
 - ...
- Domänenheterogenität
 - Subsumptionskonflikte
 - Überlappungskonflikte
 - ...

Inkonsistenz- und Redundanzkonflikte

- Datenungenauigkeit
- Temporale Inkonsistenz
- Konflikte durch fehlende Daten

Problemstellung



2. Klassische Strategien

- 2.1. Eng gekoppelte Ansätze
- 2.2. Lose gekoppelte Ansätze
- 2.3. Objektorientierte Informationssysteme

2.1. Eng gekoppelte Ansätze

- Lösung: a priori -> Föderiertes Schema

Bank1 || Nachname | Vorname | Ktostand (EUR) | ...

Bank2 || Name | Ktostand (USD) | ...

FS || Name | Ktostand | Währung | ...



2.1. Eng gekoppelte Ansätze

- Umsetzung von Anfragen in Subanfragen
- Ergebnis: Schemaintegration

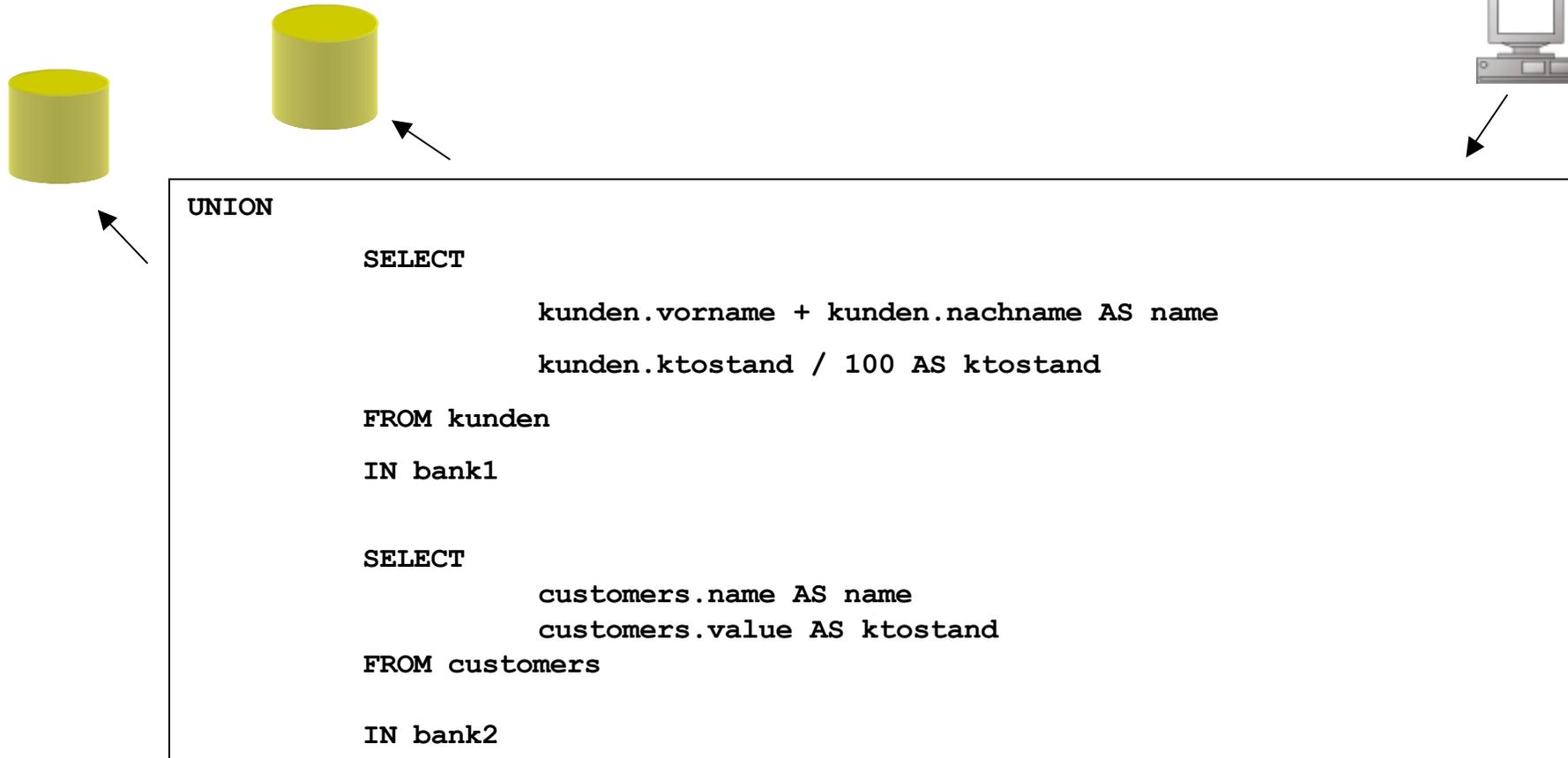
- Probleme:
 - Anwendungsheterogenitäten
 - Effizienzprobleme (2x konvertieren)
 - Föderierte DBS: Geringe Skalierbarkeit und Flexibilität

2.2. Lose gekoppelte Systeme

- Kein globales Schema
- A posteriori
- Anfrageformalismus
- Multi-Datenbank-Abfragesprache

Beispiel ->

2.2. Lose gekoppelte Systeme

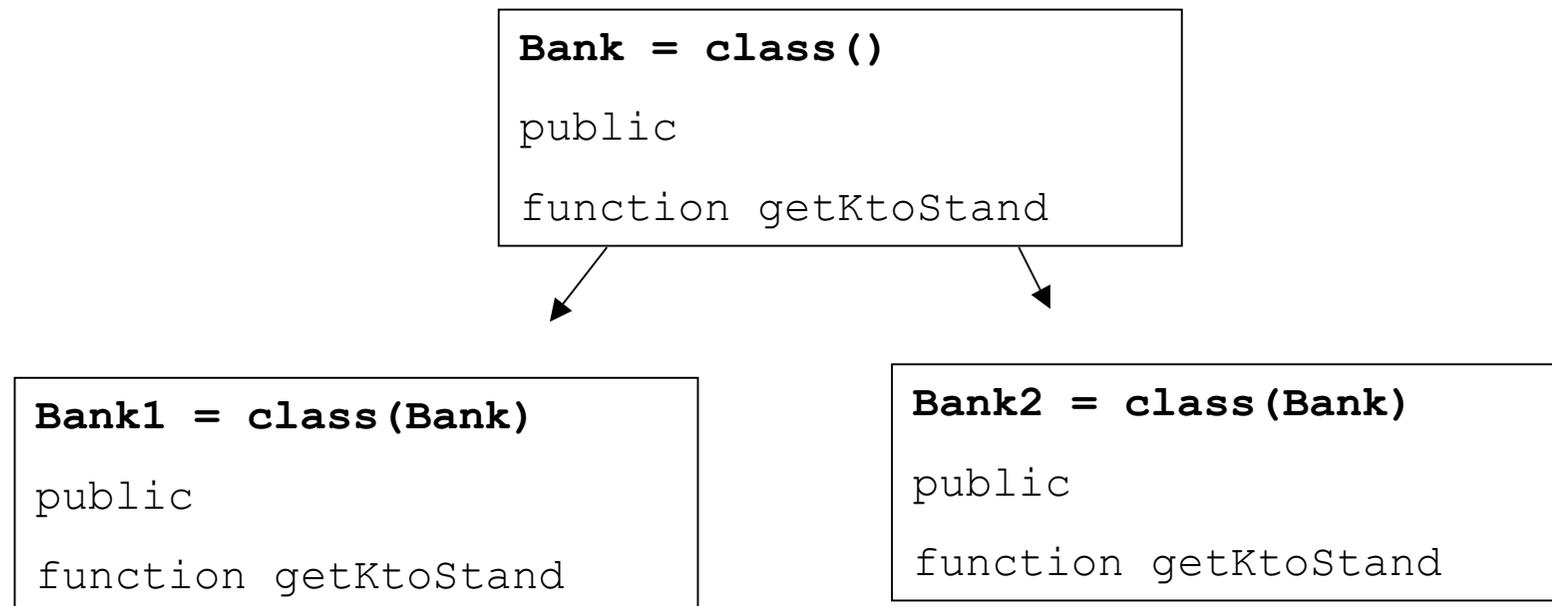


2.2. Lose gekoppelte Systeme

- Probleme
 - Sehr komplexe Abfragen
 - Schlechter Überblick
 - Genauer Kenntnisse der Datenbanken erforderlich
 - Schlechte Skalierbarkeit und Flexibilität

2.3. Objektorientierte Informationssysteme

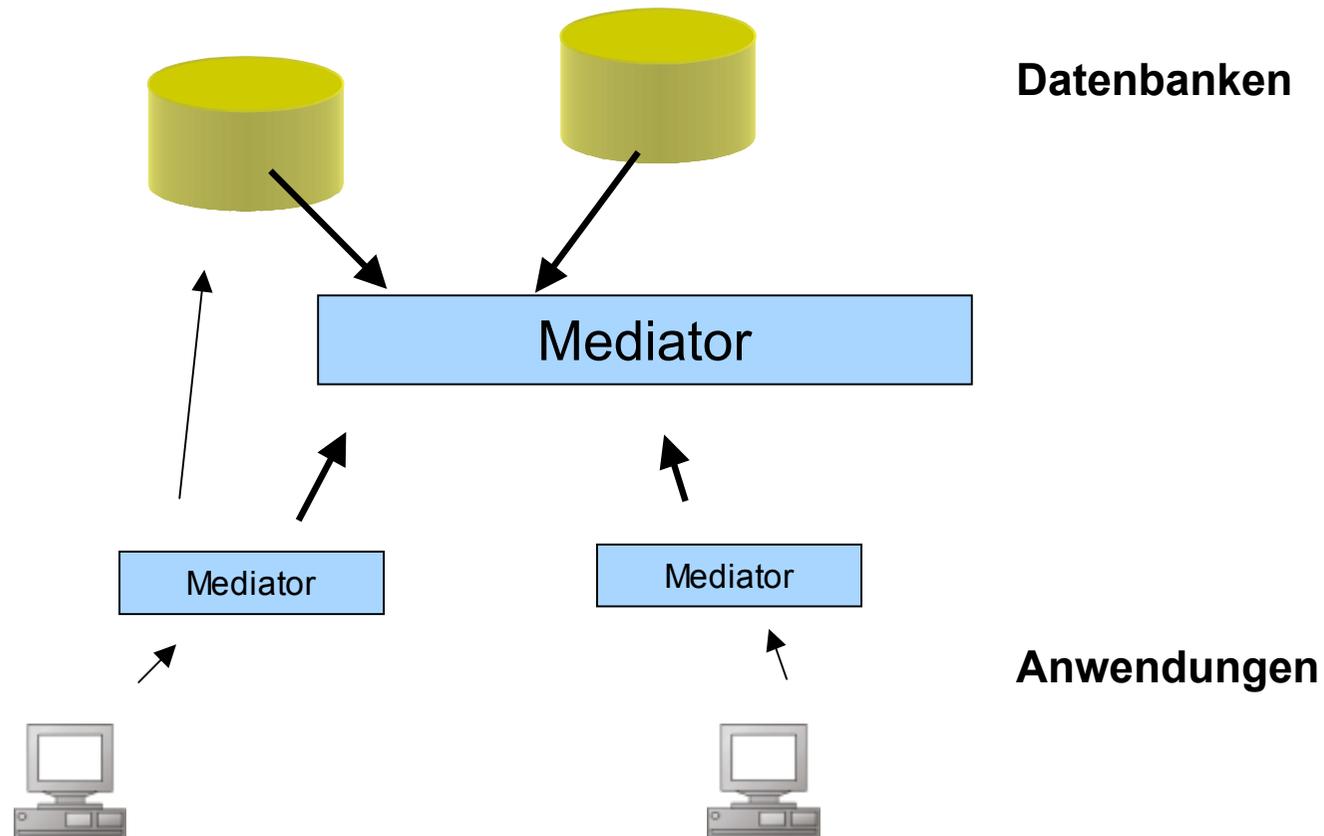
- Ersetzung des Föderierten Schemas durch objektorientierte Form



3. Der Mediator

- Beschreibung
- Anforderungen an die vollständige Beschreibung
 - Anforderungen an die Bedeutung
 - Anforderungen an den Kontext
- Anforderungen an den semantischen Mediator
 - Anforderungen an den Integrationsformalismus
 - Anforderungen an die Inferenzen

Der Mediator



Der Mediator

- Modulares Architekturkonzept
- Selbstständige Komponente
- Sowohl eng gekoppelt als auch lose gekoppelt
- Mischformen

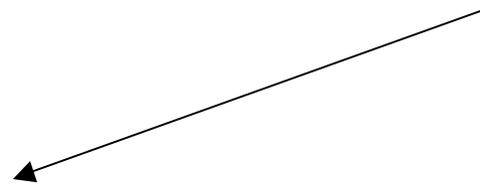
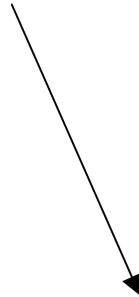
3.1. Anforderungen an vollständige Beschreibung

- Vollständige Beschreibung:
 - Syntaktisch
 - explizite Repräsentation der Semantik
- Strukturelle Konflikte durch syntaktische Beschr.
- Semantische Konflikte und Kontexttransformationen durch expl. Reprä. d. S.

Bedeutung vs Kontext

Bank1 || Name | Ktostand (EUR) | ...

Bank2 || Name | Ktostand (USD) | ...



Gleiche **Bedeutung**: Guthaben des Kunden

Ungleicher **Kontext**: Euro vs USD

Bedeutung vs Kontext

- Bedeutung: Konstanter Kern des Informationsgehalts
- Kontext: von IS abhängiger Informationsgehalt

=> Semantische Beschreibung braucht Unterscheidung zwischen Bedeutung und Kontext

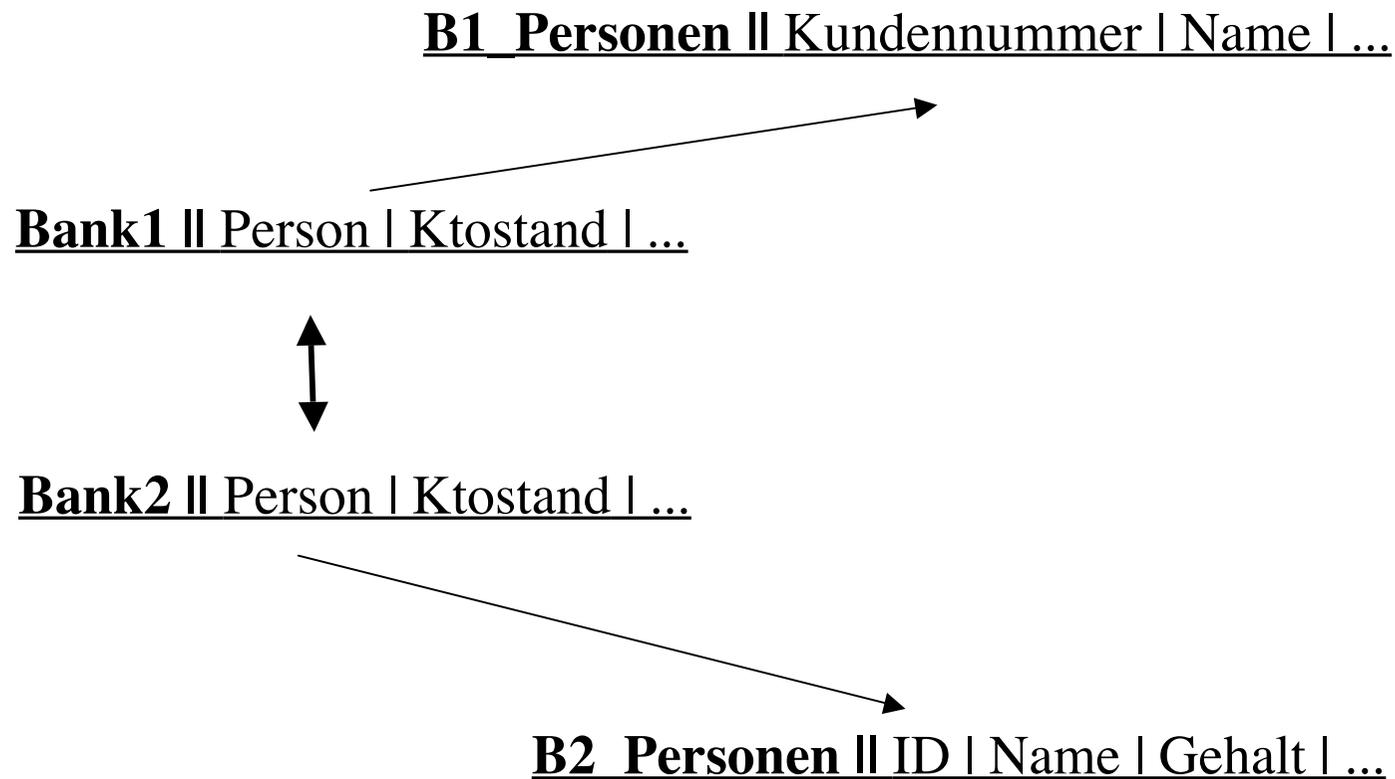
3.1.1. Anforderungen an die Bedeutung

- Wichtig zur Beseitigung von semantischen **Domänenheterogenitäten**
- Entitäten semantisch äquivalent, wenn:
 - Keine Subsumptionskonflikte
 - Keine Überlappungskonflikte
 - Keine Aggregationskonflikte
 - Keine Inkompatibilitäten
- Betrachtung von „Bedeutung“ ausreichend

Problem: Semantik von Entitäten

!= Semantik der enthaltenen Daten

Vergleich von Entitäten



3.1.2. Anforderungen an den Kontext

- Wichtig zur Beseitigung von semantischen **Datenheterogenitäten**
- Vollständige Kontextbeschreibung
 - Relevanter Teil (unterschiedliche Interpretationen)
 - Unrelevanter Teil (gleiche Informationen)
- Minimalität: Nur Relevanter Informationen
- Aber: Neues Informationssystem?!

Anforderungen an den Kontext

- Beschreibung sollte:
 - Ädequat
 - Deklarativ
 - => Angemessen, verständlich und gleichartig
 - Skalierbar und Flexibel
 - => Einfach zu erweitern

3.2. Anforderungen an den semantischen Mediator

- Semantische Mediator
 - Kernmethode
 - Kontexttransformation
 - Integrationsformalismus
 - Spezifiziert Integrationsabbildung
 - Dadurch konfigurierbar

3.2.1. Anforderungen an den Integritätsformalismus

- Muss so mächtig sein, dass strukturelle und semantische Heterogenitätskonflikte beseitigt werden.
- Informationen selektieren und konvertieren
- Fehlende Daten: Möglichkeiten der Fehlerbehandlung und autom. Berechnung

3.2.1. Anforderungen an den Integritätsformalismus

- Hohe Modularität
 - Wirkt Komplexität der Integrationskonflikte entgegen
 - Schnelle und kostengünstige Einpflege von Änderungen
- Getrennte Repräsentationen
 - Konzept der Kontexttransformation
- Deklarativ und adäquat
- Skalierbar und flexibel

3.2.2. Anforderungen an die Inferenzen

- Eigentliche Transformation auch ohne Kontexttransformation
- Selbstständiges Aufdecken von Semantischen Heterogenitätskonflikten
- Einbringung von Bedingungen für Transformation

Zusammenfassung

- Mediator nicht perfekt aber bester Ansatz
- Komplexität durch Modularisierung reduzierbar
- Hohe Anforderungen an Daten und Implementierung

Literatur

- Klassifikation von Integrationskonflikten; Wache, 2003, S. 13-39; Akademische Verlagsgesellschaft Aka GmbH; Berlin
- Strategien zur Datenintegration inkl. Mediator; Wache, 2003, S. 39-58; Akademische Verlagsgesellschaft Aka GmbH; Berlin
- Einführung in die KI; Kreißl, 2002, im Eigenverlag; Chemnitz
- Datenbanken: Konzepte und Sprachen; Heuer & Saake, 2000; mitp-Verlag; Bonn
- Grundlagen von Datenbank-Systemen; Elmasri & Navathe, 2002; Pearson Studium; München

Fragen?

Vielen Dank!