

Autonomie & Heterogenität

Komplexe Informationssysteme

Fabian Panse

panse@informatik.uni-hamburg.de

Universität Hamburg



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG



Agenda

- 1 Einführung
- 2 Autonomie
 - Design-Autonomie
 - Kommunikations-Autonomie
 - Ausführungs-Autonomie
- 3 Heterogenität
 - Syntaktische Heterogenität
 - Strukturelle Heterogenität
 - Semantische Heterogenität

Probleme

Probleme

- Daten sind auf mehrere Quellen verteilt

Probleme

- Daten sind auf mehrere Quellen verteilt
- Verteilung führt zu Autonomie
 - Intra-Organisation: Historisch
 - Inter-Organisation: Naturgemäß (Internet & WWW)

Probleme

- Daten sind auf mehrere Quellen verteilt
- Verteilung führt zu Autonomie
 - Intra-Organisation: Historisch
 - Inter-Organisation: Naturgemäß (Internet & WWW)
- ... und Autonomie führt zu Heterogenität
 - Daten-Verantwortung bzw. -Eigentümer meist lokal
 - interessiert an lokaler Nutzbarkeit
 - lokale Design-Entscheidungen

Agenda

- 1 Einführung
- 2 **Autonomie**
 - Design-Autonomie
 - Kommunikations-Autonomie
 - Ausführungs-Autonomie
- 3 Heterogenität
 - Syntaktische Heterogenität
 - Strukturelle Heterogenität
 - Semantische Heterogenität

Übersicht Autonomie

Übersicht Autonomie

- Grad zu dem verschiedene Informationssysteme unabhängig operieren können

Übersicht Autonomie

- Grad zu dem verschiedene Informationssysteme unabhängig operieren können
- Bezieht sich auf Kontrolle, nicht auf Daten

Übersicht Autonomie

- Grad zu dem verschiedene Informationssysteme unabhängig operieren können
- Bezieht sich auf Kontrolle, nicht auf Daten
- Klassen nach [OV11]:
 - Design-Autonomie
 - Kommunikations-Autonomie
 - Ausführungs-Autonomie

Design-Autonomie

- Freiheit des lokalen Informationssystems bezüglich

Design-Autonomie

- Freiheit des lokalen Informationssystems bezüglich
 - Datenmodell (relational, hierarchisch, XML)

Design-Autonomie

- Freiheit des lokalen Informationssystems bezüglich
 - Datenmodell (relational, hierarchisch, XML)
 - Schema:
 - Abdeckung der Domäne (*universe of discourse*)
 - Grad der Normalisierung
 - Benennung

Design-Autonomie

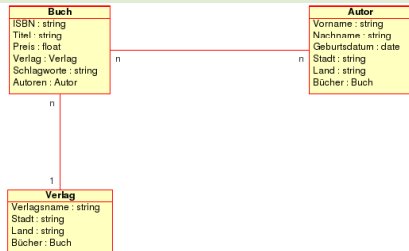
- Freiheit des lokalen Informationssystems bezüglich
 - Datenmodell (relational, hierarchisch, XML)
 - Schema:
 - Abdeckung der Domäne (*universe of discourse*)
 - Grad der Normalisierung
 - Benennung
 - Transaktionsmanagement (z.B. Isolationslevel)

Design-Autonomie

- Freiheit des lokalen Informationssystems bezüglich
 - Datenmodell (relational, hierarchisch, XML)
 - Schema:
 - Abdeckung der Domäne (*universe of discourse*)
 - Grad der Normalisierung
 - Benennung
 - Transaktionsmanagement (z.B. Isolationslevel)
- Freiheit dies jederzeit zu ändern (besonders problematisch!)

Design-Autonomie (Beispiel)

Flaches, (fast) relationales Schema



Hierarchisches XML Schema

```

<xs:element name="author" minOccurs="0" maxOccurs="unbounded">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="name" type="xs:string"/>
      <xs:element name="publication" minOccurs="0" maxOccurs="unbounded">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="title" type="xs:string"/>
            <xs:element name="year" type="xs:string"/>
            <xs:element name="booktitle" type="xs:string" minOccurs="0"/>
            <xs:element name="journal" type="xs:string" minOccurs="0"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>
  
```

Quelle: Melanie Herschel, Universität Stuttgart

Kommunikations-Autonomie

Lokale Informationssysteme frei bezüglich:

Kommunikations-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl mit *welchen* Systemen kommuniziert wird

Kommunikations-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl mit *welchen* Systemen kommuniziert wird
- Wahl *wann* mit anderen Systemen kommuniziert wird:
Jederzeit Eintritt/Austritt aus integriertem System

Kommunikations-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl mit *welchen* Systemen kommuniziert wird
- Wahl *wann* mit anderen Systemen kommuniziert wird:
Jederzeit Eintritt/Austritt aus integriertem System
- Wahl *was* (welcher Teil der Information) kommuniziert wird

Kommunikations-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl mit *welchen* Systemen kommuniziert wird
- Wahl *wann* mit anderen Systemen kommuniziert wird:
Jederzeit Eintritt/Austritt aus integriertem System
- Wahl *was* (welcher Teil der Information) kommuniziert wird
- Wahl *wie* mit anderen Systemen kommuniziert wird
(Anfragesprache: z.B. Prädikate, Projektionen)

Kommunikations-Autonomie (Beispiel)

Extrem 1: Voller SQL Zugang

- z.B. via JDBC
- Transaktionen
- Optimierung
- Lesend (und schreibend?)
- Schemaveränderungen?
- Antwort als Ergebnisrelation

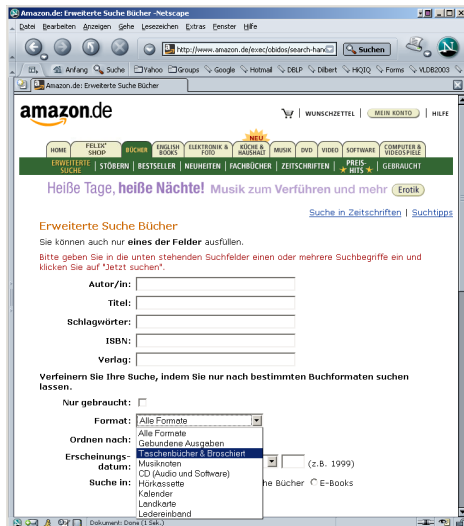
Kommunikations-Autonomie (Beispiel)

Extrem 1: Voller SQL Zugang

- z.B. via JDBC
- Transaktionen
- Optimierung
- Lesend (und schreibend?)
- Schemaveränderungen?
- Antwort als Ergebnisrelation

Extrem 2: HTML Formular

- Nur einzelne Suchfelder
- Antwort als HTML Text
- Nur Teile der Daten
(*public area*)



Quelle: Melanie Herschel, Universität Stuttgart

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl *wann* Anfragen ausgeführt werden

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl *wann* Anfragen ausgeführt werden
- Wahl *wie* Anfragen ausgeführt werden

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl *wann* Anfragen ausgeführt werden
- Wahl *wie* Anfragen ausgeführt werden
- Wahl der Scheduling-Strategien

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl *wann* Anfragen ausgeführt werden
- Wahl *wie* Anfragen ausgeführt werden
- Wahl der Scheduling-Strategien
- Wahl Optimierungsstrategien

Ausführungs-Autonomie

Lokale Informationssysteme frei bezüglich:

- Wahl *wann* Anfragen ausgeführt werden
- Wahl *wie* Anfragen ausgeführt werden
- Wahl der Scheduling-Strategien
- Wahl Optimierungsstrategien
- Wahl ob globale Transaktionen unterstützt werden

Ausführungs-Autonomie (Beispiel)

Ausführungs-Autonomie (Beispiel)

Optimierung und Scheduling

- Behandlung externer vs. lokaler Anfragen
- *Golden customers*
- Garantierte Antwortzeiten

Ausführungs-Autonomie (Beispiel)

Optimierung und Scheduling

- Behandlung externer vs. lokaler Anfragen
- *Golden customers*
- Garantierte Antwortzeiten

Transaktionen

- Ist *Dirty Read* egal?

Ausführungs-Autonomie (Beispiel)

Optimierung und Scheduling

- Behandlung externer vs. lokaler Anfragen
- *Golden customers*
- Garantierte Antwortzeiten

Transaktionen

- Ist *Dirty Read* egal?

Verteilung

- Welcher Grad an Konsistenz wird durch Quelle garantiert?
(CAP-Theorem)
 - *read-your-writes*
 - *monotonic reads*
 - ...

Agenda

- ① Einführung
- ② Autonomie
 - Design-Autonomie
 - Kommunikations-Autonomie
 - Ausführungs-Autonomie
- ③ Heterogenität
 - Syntaktische Heterogenität
 - Strukturelle Heterogenität
 - Semantische Heterogenität

Übersicht Heterogenität

Heterogenität herrscht, wenn sich zwei miteinander verbundene Informationssysteme syntaktisch, strukturell oder inhaltlich unterscheiden.

Übersicht Heterogenität

Heterogenität herrscht, wenn sich zwei miteinander verbundene Informationssysteme syntaktisch, strukturell oder inhaltlich unterscheiden.

Typen von Heterogenität:

- **Syntaktische** Heterogenität
- **Strukturelle** Heterogenität
- **Semantische** Heterogenität

Übersicht Heterogenität

Heterogenität herrscht, wenn sich zwei miteinander verbundene Informationssysteme syntaktisch, strukturell oder inhaltlich unterscheiden.

Typen von Heterogenität:

- Syntaktische Heterogenität
- Strukturelle Heterogenität
- Semantische Heterogenität

Heterogenitäten zu überbrücken ist die Kernaufgabe der Informationsintegration.

Syntaktische Heterogenität

- Hardware-Heterogenität (Bandbreite, CPU, Hauptspeicher)
- Software-Heterogenität
(z.B. Betriebssystem, Protokolle, Sicherheit)

Syntaktische Heterogenität

- Hardware-Heterogenität (Bandbreite, CPU, Hauptspeicher)
- Software-Heterogenität
(z.B. Betriebssystem, Protokolle, Sicherheit)
- Schnittstellen-Heterogenität: Anfragesprachen unterschiedlich
 - Negation?
 - Ungleichheit?
 - Prädikate nur mit Konstanten oder über mehrere Variablen?
 - Prädikate nur mit bestimmten Konstanten

Syntaktische Heterogenität

- Hardware-Heterogenität (Bandbreite, CPU, Hauptspeicher)
- Software-Heterogenität (z.B. Betriebssystem, Protokolle, Sicherheit)
- Schnittstellen-Heterogenität: Anfragesprachen unterschiedlich
 - Negation?
 - Ungleichheit?
 - Prädikate nur mit Konstanten oder über mehrere Variablen?
 - Prädikate nur mit bestimmten Konstanten
- Gebundene (Werte erforderlich) und freie Variablen in Anfrageschnittstellen

Schnittstellen-Heterogenität (Anfragesprache)

Probleme für integrierte Systeme

- Globale Anfragesprache ist mächtiger als lokale Anfragesprache
 - Anfragen eventuell nicht ausführbar
 - Oder globales System muss kompensieren

Schnittstellen-Heterogenität (Anfragesprache)

Probleme für integrierte Systeme

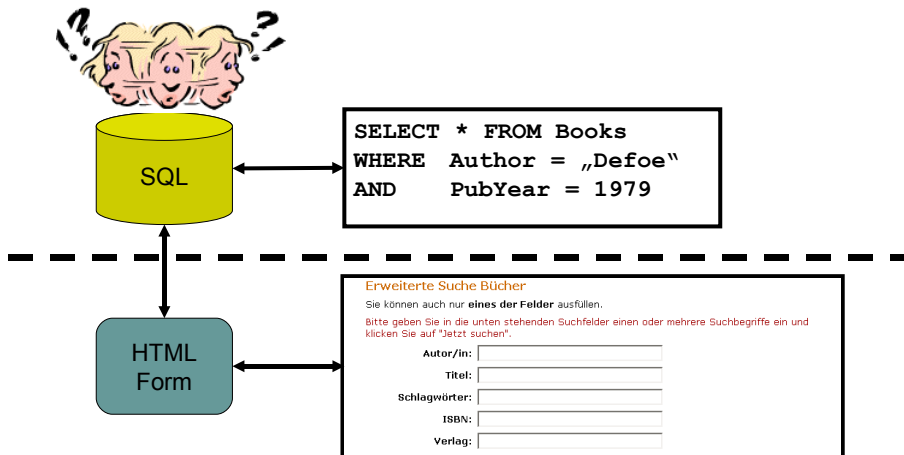
- Globale Anfragesprache ist mächtiger als lokale Anfragesprache
 - Anfragen eventuell nicht ausführbar
 - Oder globales System muss kompensieren
- Lokale Anfragesprache ist mächtiger als globale Anfragesprache:
Verpasste Chance, lokale (effiziente) Ausführung auszunutzen

Schnittstellen-Heterogenität (Anfragesprache)

Probleme für integrierte Systeme

- Globale Anfragesprache ist mächtiger als lokale Anfragesprache
 - Anfragen eventuell nicht ausführbar
 - Oder globales System muss kompensieren
- Lokale Anfragesprache ist mächtiger als globale Anfragesprache:
Verpasste Chance, lokale (effiziente) Ausführung auszunutzen
- Gebundene und freie Variablen sind inkompatibel:
Anfragen eventuell nicht ausführbar

Mächtige globale Anfragesprache



Kompensation von Anfragefragmenten

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear = 1979
```



Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Suchbegriffe ein und klicken Sie auf "Jetzt suchen".

Autor/in:

Titel:

Schlagwörter:

ISBN:

Verlag:

Kompensation von Anfragefragmenten

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear = 1979
```



Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Suchbegriffe ein und klicken Sie auf "Jetzt suchen".

Autor/in:

Titel:

Schlagwörter:

ISBN:

Verlag:

Kompensation von Anfragefragmenten

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear = 1979
```



Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Suchbegriffe ein und klicken Sie auf "Jetzt suchen".

Autor/in:

Titel:

Schlagwörter:

ISBN:

Verlag:

Kompensation von Anfragefragmenten

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear = 1979
```



```
Daniel Defoe, Robinson Crusoe, 1986
Daniel Defoe, Robinson Crusoe, 1979
Daniel Defoe, Moll Flanders, 1933
```

Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Begriffe ein und klicken Sie auf "Jetzt suchen".

Autor/in:

Titel:

Schlagwörter:

ISBN:

Verlag:



Kompensation von Anfragefragmenten

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear = 1979
```

Daniel Defoe, Robinson Crusoe, 1979

PubYear = 1979

Daniel Defoe, Robinson Crusoe, 1986
Daniel Defoe, Robinson Crusoe, 1979
Daniel Defoe, Moll Flanders, 1933

Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Begriffe ein und klicken Sie auf "Jetzt suchen".

Autor/in:

Titel:

Schlagwörter:

ISBN:

Verlag:

Mächtige globale Anfragesprache (2)

```
SELECT * FROM Books
WHERE Author = „Defoe“
AND PubYear > 1979
```

Daniel Defoe, Robinson Crusoe, 1986

PubYear > 1979

Daniel Defoe, Robinson Crusoe, 1986
 Daniel Defoe, Robinson Crusoe, 1979
 Daniel Defoe, Moll Flanders, 1933

Erweiterte Suche Bücher

Sie können auch nur **eines der Felder** ausfüllen.

Bitte geben Sie in die unten stehenden Suchfelder einen oder mehrere Begriffe ein. Wenn Sie fertig sind, klicken Sie auf "Jetzt suchen".

Autor/in:

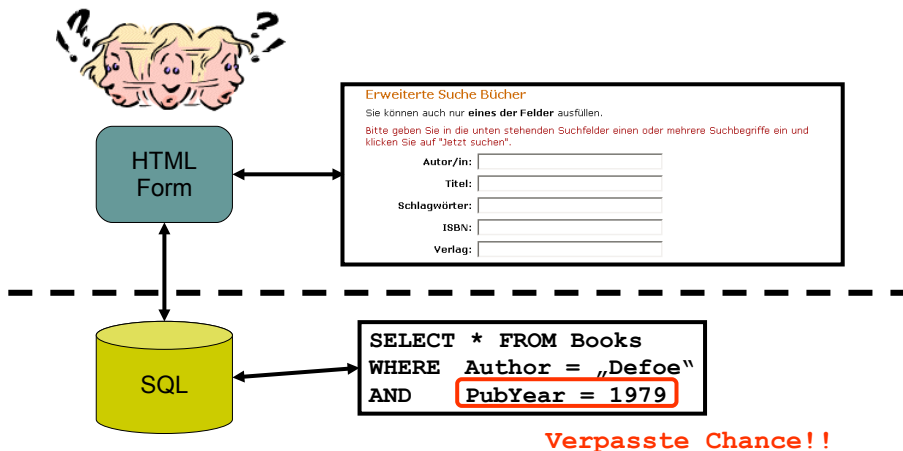
Titel:

Schlagwörter:

ISBN:

Jahr:

Mächtige lokale Anfragesprache



Gebundene & Freie Variablen

- Wie teuer ist die billigste CD mit einem Song namens 'Friends'?
- Szenario 1: Keine gebundenen Variablen

SONGS	Song	CD
	Friends	Life
	Friends	Love

CDs	CD	Künstler	Preis
	Love	Lucy	15
	Story	Snoopy	14

Künstler	CD	Künstler	Preis
	Story	Lucy	13
	Love	Snoopy	10
	Life	Charlie	8

Gebundene & Freie Variablen

- Wie teuer ist die billigste CD mit einem Song namens 'Friends'?
- Szenario 1: Keine gebundenen Variablen

SONGS	Song	CD
	Friends	Life
	Friends	Love

CDs	CD	Künstler	Preis
	Love	Lucy	15
	Story	Snoopy	14

Künstler	CD	Künstler	Preis
	Story	Lucy	13
	Love	Snoopy	10
	Life	Charlie	8

- Antwort: 8

Gebundene & Freie Variablen

- Wie teuer ist die billigste CD mit einem Song namens 'Friends', die Sie anfragen können?
- Szenario 2: Gebundene Variablen (unterstrichen)

SONGS	<u>Song</u>	CD
	Friends	Life
	Friends	Love

CDs	<u>CD</u>	Künstler	Preis
	Love	Lucy	15
	Story	Snoopy	14

Künstler	CD	<u>Künstler</u>	Preis
	Story	Lucy	13
	Love	Snoopy	10
	Life	Charlie	8

Gebundene & Freie Variablen

- Wie teuer ist die billigste CD mit einem Song namens 'Friends', die Sie anfragen können?
- Szenario 2: Gebundene Variablen (unterstrichen)

SONGS	<u>Song</u>	CD
	Friends	Life
	Friends	Love

CDs	<u>CD</u>	Künstler	Preis
	Love	Lucy	15
	Story	Snoopy	14

Künstler	CD	<u>Künstler</u>	Preis
	Story	Lucy	13
	Love	Snoopy	10
	Life	Charlie	8

- Antwort: 15 (Quelle Künstler kann nicht angefragt werden, da der Künstlernamen gebunden aber nicht bekannt ist)

Strukturelle Heterogenität

- Datenmodell¹-Heterogenität
 - Unterschiedliche Semantik
 - Unterschiedliche Struktur

¹Ein Datenmodell ist ein Modell von Schemata.

Strukturelle Heterogenität

- Datenmodell¹-Heterogenität
 - Unterschiedliche Semantik
 - Unterschiedliche Struktur
- Schematische Heterogenität
 - Integritätsbedingungen, Schlüssel, Fremdschlüssel, etc.
 - Schema (Attribut vs. Relation, etc.)
 - Struktur (Gruppierung in Tabellen)

¹Ein Datenmodell ist ein Modell von Schemata.

Schematische Heterogenität

- Modellierung:
 - Relation vs. Attribut
 - Attribut vs. Wert
 - Relation vs. Wert

Schematische Heterogenität

- Modellierung:
 - Relation vs. Attribut
 - Attribut vs. Wert
 - Relation vs. Wert
- Benennung:
 - Relation
 - Attribute
 - Homonyme und Synonyme

Schematische Heterogenität

- Modellierung:
 - Relation vs. Attribut
 - Attribut vs. Wert
 - Relation vs. Wert
- Benennung:
 - Relation
 - Attribute
 - Homonyme und Synonyme
- Normalisiert vs. denormalisiert

Schematische Heterogenität

- Modellierung:
 - Relation vs. Attribut
 - Attribut vs. Wert
 - Relation vs. Wert
- Benennung:
 - Relation
 - Attribute
 - Homonyme und Synonyme
- Normalisiert vs. denormalisiert
- Geschachtelt vs. Fremdschlüssel

Schematische Heterogenität

- Modellierung:
 - Relation vs. Attribut
 - Attribut vs. Wert
 - Relation vs. Wert
- Benennung:
 - Relation
 - Attribute
 - Homonyme und Synonyme
- Normalisiert vs. denormalisiert
- Geschachtelt vs. Fremdschlüssel

Diese Probleme sogar bei gleichem Datenmodell!

Schematische Heterogenität (Beispiel)



```
Male(Id, firstName, lastName)
Female(Id, firstName, lastName)
```

Relation vs.
Attribut

```
Person(Id, firstName,  
lastName, isMale, isFemale)
```

Relation vs.
Attributwert

```
Person(Id, firstName,  
lastName, gender)
```

Attribut vs.
Attributwert

Schematische Heterogenität

Tabellen-Tabellen Konflikte:

- Namenskonflikte
 - Semantisch gleiche Tabellen mit verschiedenen Namen (Synonym)
 - Verschiedene Tabellen mit gleichem Namen (Homonym)
 - Strukturkonflikte:
 - fehlende Attribute
 - fehlende, aber ableitbare Attribute (z.B. Alter von Geburtsdatum)

Schematische Heterogenität

Tabellen-Tabellen Konflikte:

- Namenskonflikte
 - Semantisch gleiche Tabellen mit verschiedenen Namen (Synonym)
 - Verschiedene Tabellen mit gleichem Namen (Homonym)
 - Strukturkonflikte:
 - fehlende Attribute
 - fehlende, aber ableitbare Attribute (z.B. Alter von Geburtsdatum)
- Konflikte in Integritätsbedingungen

Schematische Heterogenität (Beispiel)

Beispiele für Tabellen-Tabellen Konflikte

Mitarbeiter

P_ID	Vorname	Name	Funktion
1	Peter	Müller	Sachbearb.
5	Petra	Schmidt	AB Leitung

Mitarbeiter (leitend)

P_ID	Vorname	Name
2	Stefanie	Meier
2	Petra	Schmidt

Homonym

IC Konflikt
(Eindeutigkeit)

Quelle: Melanie Herschel, Universität Stuttgart

Schematische Heterogenität

Attribut-Attribut Konflikte:

- Namenskonflikte
 - Verschiedene Namen für gleiche Attribute (Synonyme)
 - Gleiche Namen für verschiedene Attribute (Homonyme)

Schematische Heterogenität

Attribut-Attribut Konflikte:

- Namenskonflikte
 - Verschiedene Namen für gleiche Attribute (Synonyme)
 - Gleiche Namen für verschiedene Attribute (Homonyme)
- Default-Wert-Konflikte

Schematische Heterogenität

Attribut-Attribut Konflikte:

- Namenskonflikte
 - Verschiedene Namen für gleiche Attribute (Synonyme)
 - Gleiche Namen für verschiedene Attribute (Homonyme)
- Default-Wert-Konflikte
- Konflikte in Integritätsbedingungen
 - Datentypkonflikte
 - Bedingungskonflikte
- Verschiedene Abstraktionsebenen ('Hamburg' vs. 'Altona')

Schematische Heterogenität (Beispiel)

Beispiele für Attribut-Attribut Konflikte

Mitarbeiter IC: alter > 18

p_id	Vorname VARCHAR(35)	nachname	alter
1	Wolfgang	Meyer	33
5	Klaus	Schmidt	NULL
...

Mitarbeiter

p_id	Vorname VARCHAR(20)	name	alter
1	Peter	Müller	0
5	Petra	Weger	17
...

Datentypkonflikt Synonym Defaultwert

Quelle: Melanie Herschel, Universität Stuttgart

Schematische Heterogenität (Beispiel)

Normalisiert vs. Denormalisiert

1:n Assoziationen werden unterschiedlich dargestellt:

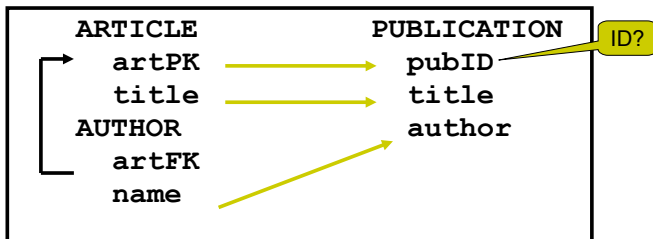
- durch Schlüssel-Fremdschlüssel-Beziehung
- durch Vorkommen im gleichen Tupel

Schematische Heterogenität (Beispiel)

Normalisiert vs. Denormalisiert

1:n Assoziationen werden unterschiedlich dargestellt:

- durch Schlüssel-Fremdschlüssel-Beziehung
- durch Vorkommen im gleichen Tupel



Schematische Heterogenität (Beispiel)

Geschachtelt vs. Flach

1:n Assoziationen werden unterschiedlich dargestellt:

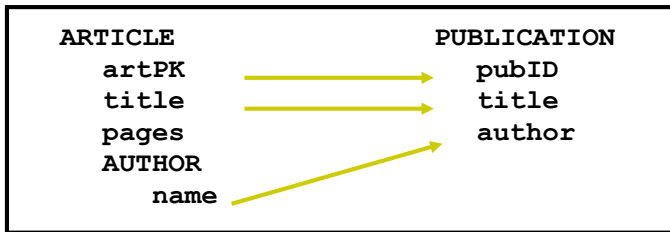
- als geschachtelte Elemente (z.B. in XML)
- durch Vorkommen im gleichen Tupel

Schematische Heterogenität (Beispiel)

Geschachtelt vs. Flach

1:n Assoziationen werden unterschiedlich dargestellt:

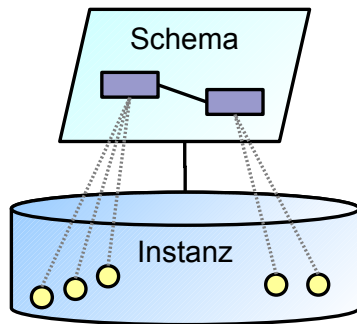
- als geschachtelte Elemente (z.B. in XML)
- durch Vorkommen im gleichen Tupel



Semantische Heterogenität

Semantik von Modellen

- Bedeutung der Konzepte des Modells?
- **Semantik** eines Konzepts :=
Zugeordnete Entitäten der realen Welt
(bzw. deren Repräsentanten in der
Datenbank-Instanz)



Semantische Heterogenität (Namenskonflikte)

- Definition Konzept in Modell?
(Wieviele Mitarbeiter hat IBM?)
- Korrespondenzarten zwischen Semantik unterschiedlicher Konzepte A und B :
 - $A = B$ Äquivalenz
 - $A \subseteq B$ Inklusion
 - $A \cap B$ Überlappung
 - $A \neq B$ Disjunktion
- Synonyme (z.B. surname vs. last name)
- Homonyme
- Einheiten
- Werte

Semantische Heterogenität (Identität)

Identität von *Real-world Entities*

- Zentrale Fragen:
 - Was ist ein (Geschäfts-)Objekt?
 - XML: Über mehrere Schachtelungsebenen hinweg
 - Relationales Modell: Über mehrere Relationen hinweg
 - Repräsentiert Objekt *a* die gleiche *real-world* Entität wie Objekt *b*?
 - Wie finde ich effizient gleiche Repräsentationen (d.h. ohne quadratische Laufzeit, Ähnlichkeitsmaße)?

Semantische Heterogenität (Identität)

Identität von *Real-world Entities*

- Zentrale Fragen:
 - Was ist ein (Geschäfts-)Objekt?
 - XML: Über mehrere Schachtelungsebenen hinweg
 - Relationales Modell: Über mehrere Relationen hinweg
 - Repräsentiert Objekt *a* die gleiche *real-world* Entität wie Objekt *b*?
 - Wie finde ich effizient gleiche Repräsentationen (d.h. ohne quadratische Laufzeit, Ähnlichkeitsmaße)?
- Synonyme für Problem:
 - Duplikaterkennung
 - Objektidentifikation
 - *Record Linkage*
 - *Entity Resolution*
 - ...

Semantische Heterogenität (Identität)

Identität von *Real-world Entities*

- Zentrale Fragen:
 - Was ist ein (Geschäfts-)Objekt?
 - XML: Über mehrere Schachtelungsebenen hinweg
 - Relationales Modell: Über mehrere Relationen hinweg
 - Repräsentiert Objekt *a* die gleiche *real-world* Entität wie Objekt *b*?
 - Wie finde ich effizient gleiche Repräsentationen (d.h. ohne quadratische Laufzeit, Ähnlichkeitsmaße)?
- Synonyme für Problem:
 - Duplikaterkennung
 - Objektidentifikation
 - *Record Linkage*
 - *Entity Resolution*
 - ...
- Auf Datenebene

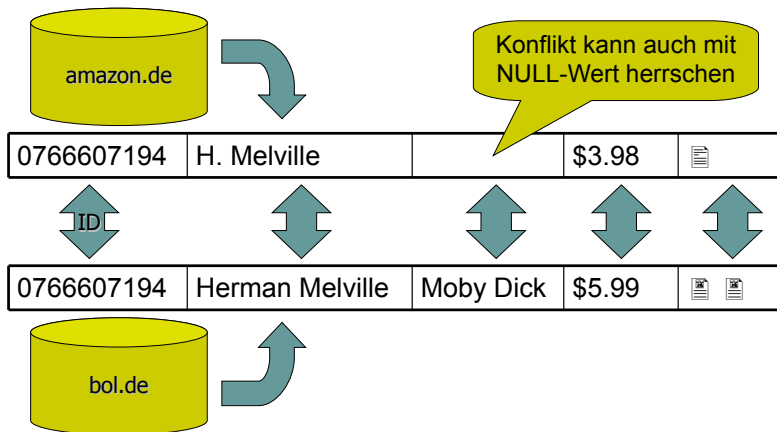
Semantische Heterogenität (Datenkonflikte)

- Datenkonflikt:
 - Zwei Duplikate haben unterschiedliche Attributwerte für ein semantisch gleiches Attribut.
 - Im Gegensatz zu Konflikten mit Integritätsbedingungen

Semantische Heterogenität (Datenkonflikte)

- Datenkonflikt:
 - Zwei Duplikate haben unterschiedliche Attributwerte für ein semantisch gleiches Attribut.
 - Im Gegensatz zu Konflikten mit Integritätsbedingungen
- Datenkonflikte entstehen
 - innerhalb eines Informationssystems (*intra-source*)
 - bei Integration mehrerer Informationssysteme (*inter-source*).

Datenkonflikte (Beispiel)



Herausforderungen (Zusammenfassung)

Drei orthogonale Dimensionen (siehe auch [BKLW99])

Herausforderungen (Zusammenfassung)

Drei orthogonale Dimensionen (siehe auch [BKLW99])

- **Verteilung**: physikalische und logische Verteilung

Herausforderungen (Zusammenfassung)

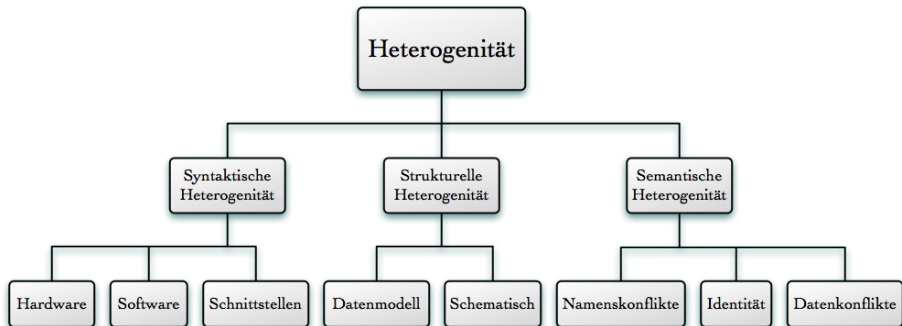
Drei orthogonale Dimensionen (siehe auch [BKLW99])

- **Verteilung**: physikalische und logische Verteilung
- **Autonomie**: Design-, Kommunikations- und Ausführungs-Autonomie

Herausforderungen (Zusammenfassung)

Drei orthogonale Dimensionen (siehe auch [BKLW99])

- **Verteilung**: physikalische und logische Verteilung
- **Autonomie**: Design-, Kommunikations- und Ausführungs-Autonomie
- **Heterogenität**



Literatur

- [BKLW99] S. Busse, R.-D. Kutsche, U. Leser, and H. Weber.
Federated information systems: Concepts, terminology and architecture.
Technical Report 99-9, Technische Universität Berlin, 1999.
citeseer.nj.nec.com/busse99federated.html.
- [Len02] Maurizio Lenzerini.
Data integration: A theoretical perspective.
In *Proc. of the Symposium on Principles of Database Systems (PODS)*, 2002.
(optional, sehr theoretisch).
- [OV11] M. T. Özsu and P. Valduriez.
Principles of distributed database systems.
Prentice Hall, 3rd edition, 2011.