

Informatik ist eine
Strukturwissenschaft

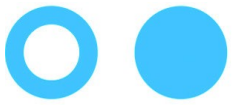
Teil II

Von Daten und ihren Modellen

Robert Hartmann (SoSe 2024)

basierend auf Folien von
Prof. Dr. Harm Knolle

Fachbereich Informatik
Hochschule Bonn-Rhein-Sieg



- Kapitel 2 - Prozess des Datenbankentwurfs -

Inhalt

0 - Vorbemerkungen

Teil I - Von EDV-Anwendungen und Ihren Anforderungen

1 - Einführung

Teil II - Von Daten und ihren Modellen

2 - Prozess des Datenbankentwurfs

3 - Semantische Datenmodelle

4 - Logische Datenmodelle

5 - Datenbankmodelle

6 - Datenanfrage und Datenänderung

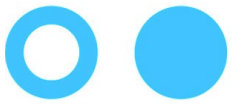
Teil III - Von Datenbanken und ihren Systemen

7 - Datenbanksysteme

8 - Speicherstrukturen

Überblick

- ♦ Datenbankentwurf
- ♦ Datenmodell



- Prozess des Datenbankentwurfs -

Ziel

- ♦ Wie wird eine Datenbank erstellt?
- ♦ Ingenieurmäßiger Prozess des Datenbankentwurfs
- ♦ Strukturierung / Modellierung / Ordnung von Informationen
- ♦ Beschreibung der Semantik der Anwendung ("Miniwelt")
- ♦ Bedeutung der unterschiedlichen Datenmodelle
- ♦ Ergebnisse der unterschiedlichen Entwurfsphasen

Hilfsmittel

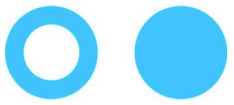
- ♦ Vorgehensmodell des Datenbankentwurfs
- ♦ Datenmodelle als Werkzeuge zur Beschreibung der Ergebnisse einzelner Phasen

Inhalt

- ♦ Datenbankentwurf
- ♦ Datenmodell

Literatur

- ♦ KeEi15, Kapitel 2 „Datenbankentwurf“ bis Abschnitt 2.4.2
- ♦ Ku15, Kapitel 2: „Informationsmodellierung“ bis Abschnitt 2.1
- ♦ SSH18, Kapitel 6: „Phasen des Datenbankentwurfs“ bis Abschnitt 6.2.9



- Datenbankentwurf -

Inhalt

- ♦ **Datenbankentwurf**
- ♦ Modellierung im Allgemeinen
- ♦ Datenmodell

Überblick

- ♦ Datenunabhängigkeit
- ♦ Schemaarchitektur einer Datenbank
- ♦ Beschreibung des Speicherschemas
- ♦ Phasen des Datenbankentwurfs
- ♦ Vom Fachproblem zur Datenbank
- ♦ Zusammenfassung

- Datenunabhängigkeit -

Ziel des Datenbankentwurfs

- ♦ Trennung der semantischen Datensicht von der physischen Speicherung (Speicherstrukturen) der Daten (**physische Datenunabhängigkeit**)
- ♦ Kostenreduktion
 - Fehler lassen sich häufig bereits in frühen Entwurfsschritten erkennen
 - eine Behebung im laufenden Betrieb ist dagegen mit hohen Kosten verbunden
 - geschätzt Faktor 10 pro Entwurfsebene
- ♦ Vereinfachung der Wartung
 - ausführliche Dokumentation aller Entwurfsschritte auf allen Entwurfsebenen
- ♦ Stabilität der Anwendung
 - einerseits gegen Änderungen der Datenbank
 - andererseits gegen Änderungen der semantischen Datensicht (**logische Datenunabhängigkeit**)

Physische Datenunabhängigkeit

- ♦ Vermeidung von Abhängigkeiten zwischen Anwendungsprogramm und physischer Datenspeicherung
- ♦ Veränderung bzw. Anpassung der Speicherorganisation haben keinen Einfluss auf die Anwendungen

Logische Datenunabhängigkeit

- ♦ Veränderungen bzw. Anpassungen der semantischen Datensicht haben keine Auswirkungen auf die Anwendungen

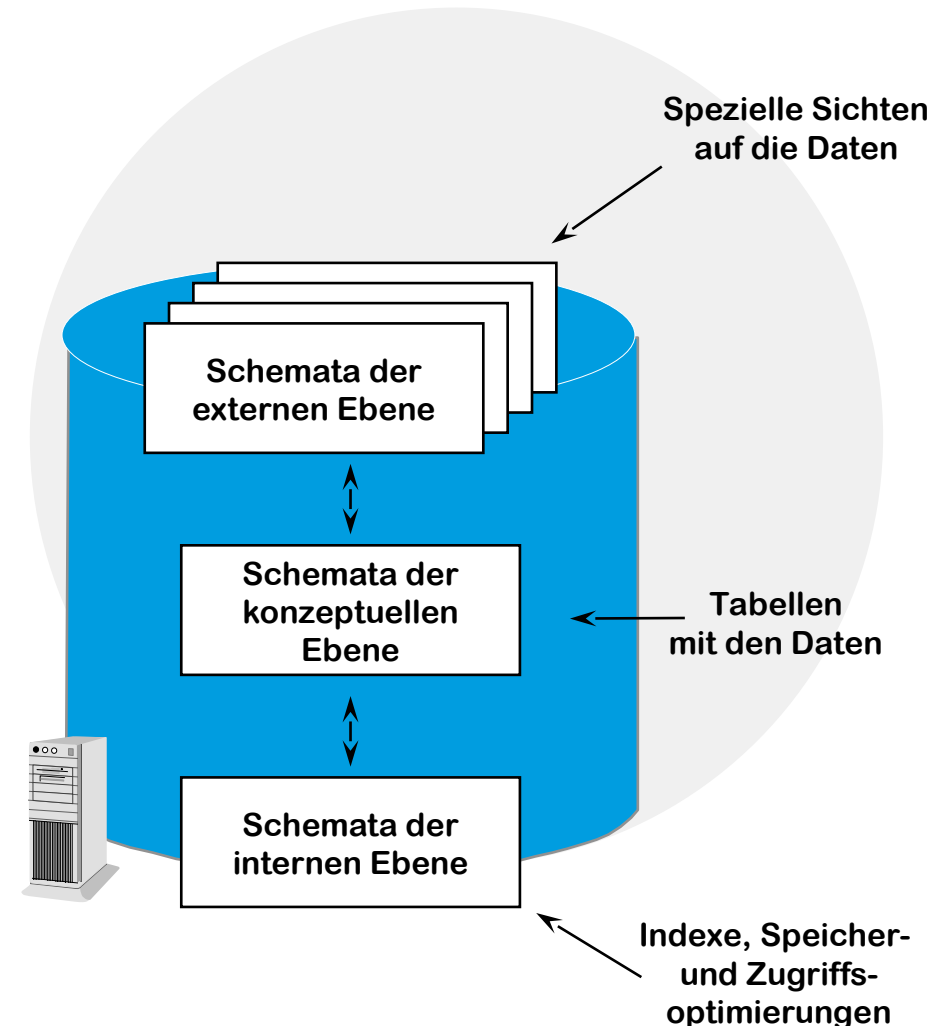
- Schemaarchitektur einer Datenbank -

Idee: Trennung von Schema und Instanz

- ♦ Schema
 - lediglich Beschreibung der Daten, Metadaten, keine Werte bzw. Anwenderdaten
- ♦ Instanz
 - Anwenderdaten, Ausprägung der Datenbank mit Werten

3-Schema Architektur nach ANSI/X3/SPARC

- ♦ American National Standard Institute Study Group on Data Base Management Systems (1975)
- ♦ Ziel: Trennung der semantischen Datensicht von der physischen Speicherung
- ♦ stabiles und globales Referenzschema (Schemata der konzeptuellen Ebene)
- ♦ Spezielle Sichten bzw. Ausschnitte der Endanwender (Schemata der externen Ebene)
- ♦ Realisierung und Optimierung von Speicherstrukturen (Schema der internen Ebene)



- Beschreibung des Speicherschemas -

Idee: Trennung von Schema und Instanz

- ♦ **Schema**
 - lediglich Beschreibung der Daten, Metadaten, keine Werte bzw. Anwenderdaten
- ♦ **Instanz**
 - Anwenderdaten, Ausprägung der Datenbank mit Werten

Problem: wie und womit das Schema beschreiben?

- ♦ eine Beschreibung, wie die Daten am effizientesten gespeichert werden sollen,
 - ist in der Regel abhängig von den technologischen Möglichkeiten des Datenbanksystems
 - und somit semantisch nur wenig ausdrucksstark
- ♦ eine abstrakte und semantisch ausdrucksstarke Beschreibung des Fachproblems ist mit Hilfe der konzeptuellen Werkzeuge heutiger Datenbanksysteme allerdings nur sehr begrenzt möglich
- ♦ andererseits können abstrakte und semantisch ausdrucksstarke Beschreibungen von Modellierungswerkzeugen z.Z. nur von sehr wenigen Datenbanksystemen vollständig verstanden werden

- Phasen des Datenbankentwurfs -

Derzeitige Lösung des Problems

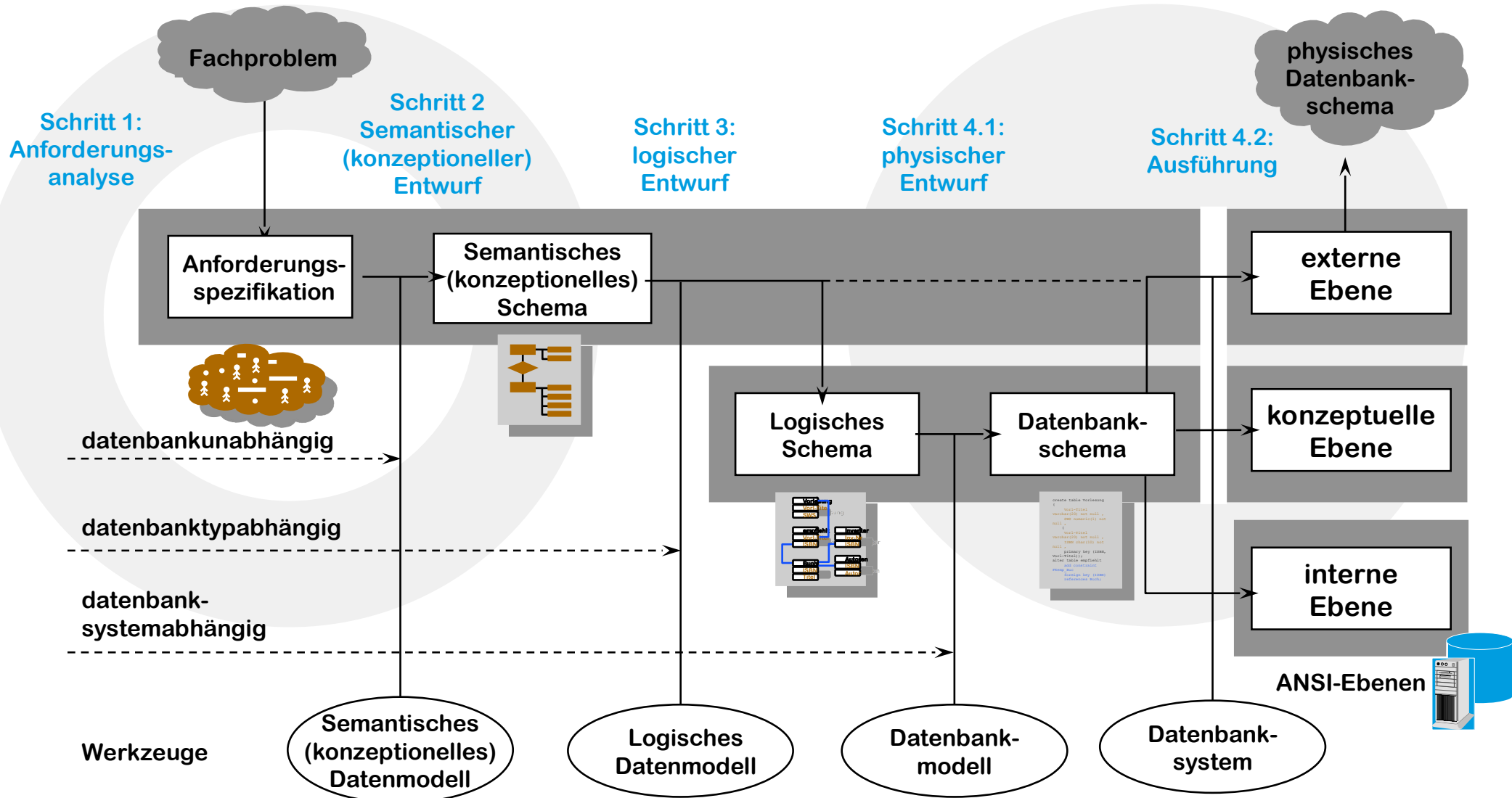
- ♦ Entwurf der Datenbank in mehreren Phasen
- ♦ Schritt 1: **Miniwelt (Diskursbereich)**
 - Analyse des Fachproblems, Abgrenzung, Abstraktion, Erstellung im Rahmen des Pflichten-/Lastenheftes
- ♦ Schritt 2: **konzeptionelle oder semantisch ausdrucksstarke Beschreibung**
 - natürliche, abstrakte und semantisch ausdrucksstarke Beschreibung des Fachproblems **unabhängig** von der Wahl eines Datenbanksystems und seinen Werkzeugen
- ♦ Schritt 3: **technologische oder logische Beschreibung**
 - schrittweise Übersetzung der abstrakten Beschreibung in konkrete **datenbanknahe** technische Beschreibungen, die von den Werkzeugen des Datenbanksystems verstanden werden

- ♦ Schritt 4: **Datenbankbeschreibung**
 - Implementierung der konkreten technischen Beschreibung in der Sprache der Datenbank
- ♦ starke Parallele zur Softwareentwicklung mit den Phasen
 - Grobkonzept
 - Feinkonzept
 - Implementierung

Alternative Lösung des Problems

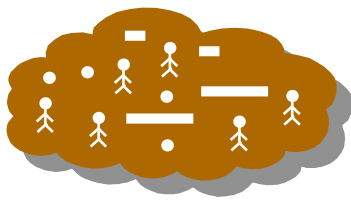
- ♦ etwas Abhilfe schaffen sogenannte objektorientierte bzw. objektrelationale Datenbanksysteme mit ihren Werkzeugen
- ♦ solche Datenbanksysteme sind vermehrt in der Lage, semantisch ausdrucksstarke Beschreibung des Fachproblems unmittelbar in interne Speicherstrukturen des Datenbanksystems umzusetzen

- Vom Fachproblem zur Datenbank -

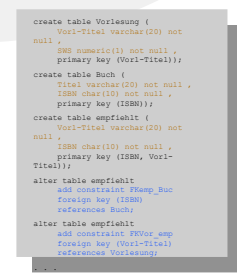
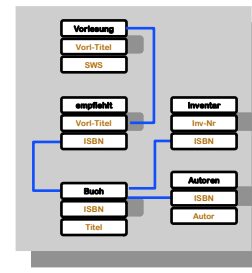
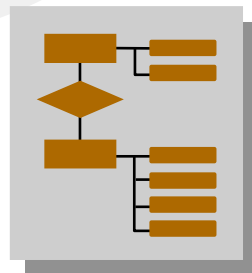


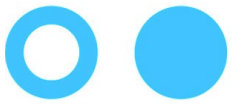
- Zusammenfassung -

Entwurfsprozess	Phasen des Datenbankentwurfs		
	Semantischer Entwurf	Logischer Entwurf	Physischer Entwurf
Entwurfsziel	erste formale Beschreibung des Fachproblems mit einem aussagekräftigen Datenmodell ohne Hinblick auf das Zieldatenbanksystem	konkretisierte und optimierte formalen Beschreibung des Fachproblems im Hinblick auf das Datenmodell des Zieldatenbanksystems	formale Beschreibung des Fachproblems in der Sprache des Zieldatenbanksystems
Resultat	semantisches Schema	logisches Schema	Datenbank Schema
Form	Diagramm	Diagramm	textuelle Skripte
Werkzeug	Semantische Datenmodelle	Logische Datenmodelle	Datenbankmodelle
Grad der Abstraktion	kein Bezug zum Datenbanksystem	Bezug zu typgleichen Datenbanksystem	direkter Bezug zu einem konkreten Datenbanksystem
ANSI-Ebene	externe Ebene, konzeptuelle Ebene	konzeptuelle Ebene	externe Ebene, konzeptuelle Ebene, interne Ebene



Ausgangssituation: Miniwelt





- Modellierung im Allgemeinen -

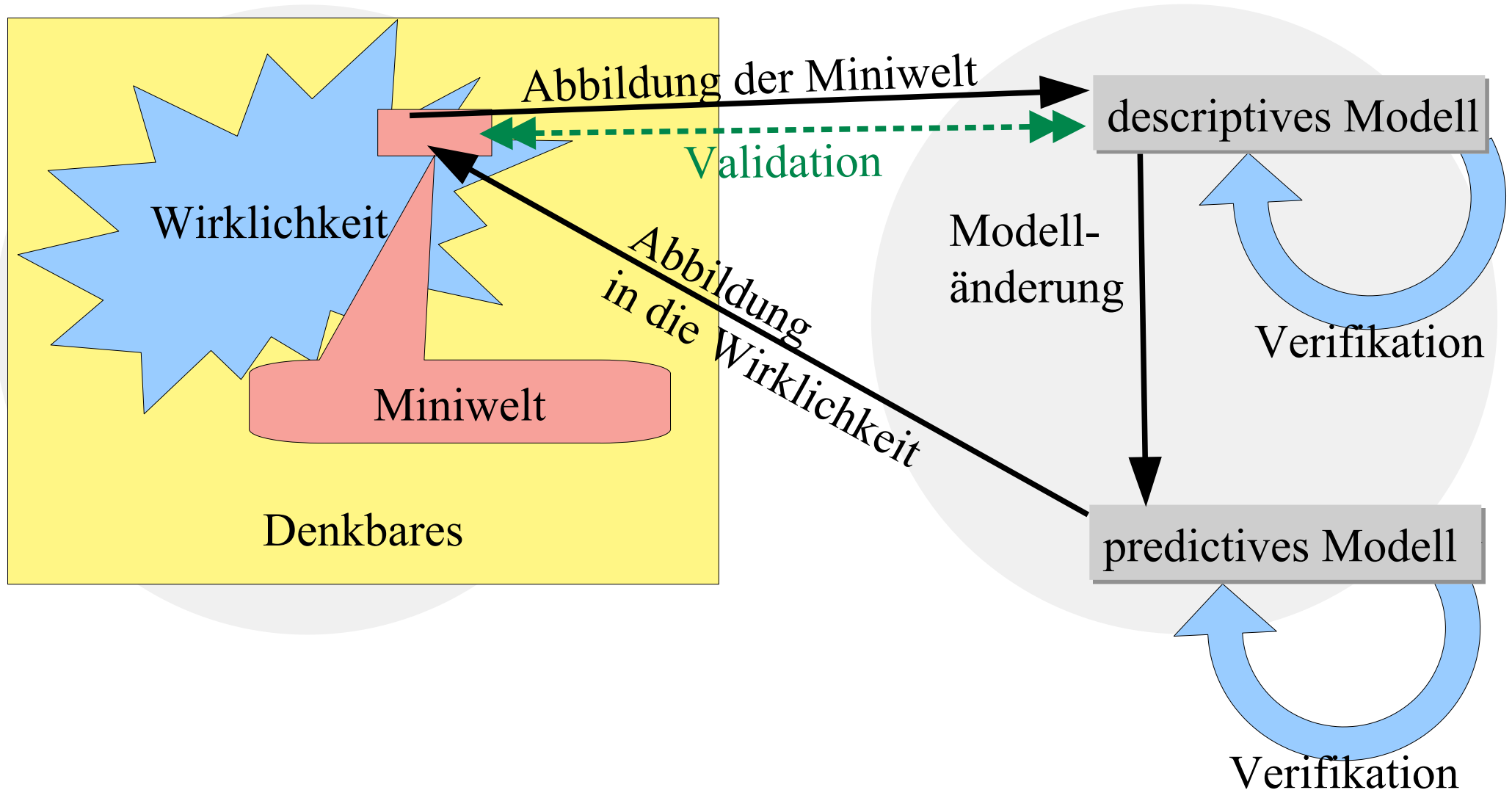
Inhalt

- ♦ Datenbankentwurf
- ♦ **Modellierung im Allgemeinen**
- ♦ Datenmodell

Überblick

- ♦ Modellierung: Ein iterativer Prozess
- ♦ Unterstützende Techniken
- ♦ Modell-Typen aus anderen Lehrveranstaltungen

- Modellierung ein iterativer Prozess -



- Unterstützende Techniken -

Beobachtung

- ♦ zielgerichtete und kontrollierte Wahrnehmung von Objekten, Ereignissen, Prozessen zur Gewinnung von Daten.
- ♦ z.B. physikalische Messungen

Befragung

- ♦ Der zielgerichtete, soziale Vorgang der Interaktion zwischen Individuen zur Gewinnung von Daten.
- ♦ **Problem: Beurteilung des Wahrheitswertes**

Dokumentenauswertung

- ♦ Die zielgerichtete systematische Analyse von Aufzeichnungen zur Gewinnung von Daten.

Merke:

Es soll nur das analysiert werden, was für Entwurfszwecke erforderlich ist.

Es soll nur das erhoben werden, was für Analysezwecke benötigt wird.

- Mit Kreativ-Techniken die Miniwelt begreifen -

Kreativ-Techniken

- ♦ Werkzeugkasten auch Anwendern zugänglich
 - Brainstorming
 - Mind Maps
 - Walt-Disney-Methode (4 Rollen)
 - Der Träumer: Ich wünsch mir was.
 - Der Realist: Ich betrachte Kosten und Machbarkeit unvoreingenommen. Ich setze alles um.
 - Der Kritiker: Worst-Case-Szenarios sind meine Spezialität. Ich suche Schwachstellen.
 - Der Neutrale: Ich betrachte und berate neutral von Außen. Ich bin Anwender, Kunde, Presse, Politiker, Jurist ...
- ♦ Die 6-Denkhüte-Methode (1986) von De Bono
 - Der Moderator (Hut Blau): Ich bin Protokollant, Schriftführer.
 - Der Optimist (Hut Gelb) : Best-Case-Szenarien sind meine Spezialität.
 - Der Denker (Hut Grün): Ich denke in die Breite und suche neue Ideen/Informationen. Den Tiefgang überlasse ich anderen.
 - Der Emotionale (Hut Rot): Mir sind Gefühle, Empfindungen wichtig. Im Vordergrund stehen meine persönliche Meinung und Geschmack
 - Der Kritiker (Hut Schwarz): Worst-Case-Szenarien sind meine Spezialität. Ich suche Schwachstellen
 - Der Objektive (Hut Weiß): Alle meine Entscheidungen sind beweisbar. Für mich zählen nur Zahlen, Fakten, Berechnungen und Beweise.

- Modell-Typen aus anderen Lehrveranstaltungen -

Dynamische Modelle

- ♦ Einf. i.d. Programmierung:
 - Flußdiagramme
 - Struktogramme
 - Java-Code
- ♦ Software Engineering:
 - UML – Aktivitätsdiagramme
 - UML – Sequenzdiagramme
 - ...

Statische Modelle

- ♦ Einf. i.d. Programmierung:
 - Klassendiagramme
 - Java-Code
- ♦ Software Engineering
 - Klassendiagramme
 - Paketdiagramme
 - ...

Nicht vergessen:
Mathematische, physikalische Modelle



- Datenmodell -

Inhalt

- ♦ Datenbankentwurf
- ♦ Modellierung im Allgemeinen
- ♦ **Datenmodell**

Überblick

- ♦ Zweck der Modellierung von Daten
- ♦ Prozess der Datenmodellierung
- ♦ Schema der Miniwelt, Diskursbereich
- ♦ Miniwelt vs. Semantisches Datenschema
- ♦ Semantisches vs. logisches Datenmodell
- ♦ Logisches Datenmodell vs. Datenbankmodell
- ♦ Transformation der Schemata

- Zweck der Modellierung von Daten (I) -

Forderung einer Datenbank

- ♦ Zu jedem Zeitpunkt ...
 - ein **korrektes Abbild**
 - ihrer zu beschreibenden **Realität**

Beschreibende Realität?

- ♦ Ausgangssituation: Schema der Miniwelt
 - Anforderungsanalyse, gedankliches Gebilde, Diskursbereich
- ♦ zweckdienliche Abstraktion
- ♦ zweckdienlicher Ausschnitt
- ♦ zweckdienliche Gesetzmäßigkeiten

Korrektes Abbild?

- ♦ technische Zugänglichkeit
über formale Beschreibung (Datenstruktur)
- ♦ korrektes Abbild, wenn sich das Schema der Miniwelt vollständig im Modell der formalen Beschreibung erfassen lässt

- Zweck der Modellierung von Daten (II) -

Forderung einer Datenbank

- ♦ Zu jedem Zeitpunkt ...
 - ein **korrektes Abbild**
 - ihrer zu beschreibenden **Realität**

Erfüllung dieser Forderung

- ♦ über den Prozess des **Datenbankentwurfs** (Datenmodellierung)
- ♦ für die gesamte Betriebszeit der Datenbank

Umsetzung dieser Forderung

- ♦ spezielle Methoden
- ♦ spezielle Verfahrensweisen
- ♦ spezielle Techniken
- ♦ Erfahrung erforderlich
 - Umfeld der Anwendung
 - Datenmodellierung

Zeitpunkt der Datenmodellierung

- ♦ Ziel: Vorschriften zur Beachtung
 - während des Betriebs der Datenbank
 - vor dem Betrieb der Datenbank

- Prozess der Datenmodellierung -

Datenmodellierung

- ♦ Ingenieurmäßiger Prozess mit dem Ziel des Datenbankentwurfs
- ♦ Modellierung der Miniwelt im Rahmen der Anforderungsanalyse
- ♦ Umsetzung des Schemas der Miniwelt und ihrer Gesetzmäßigkeiten in eine formale Beschreibung
- ♦ Zuhilfenahme von Datenmodellen als Werkzeuge für die unterschiedlichen Abstraktionsebenen
 - semantische Datenmodellierung
 - logische Datenmodellierung
 - Datenbankmodellierung

Datenmodell

- ♦ Datenmodelle als Sammelbegriff geeigneter Beschreibungsformalismen (Werkzeuge)
- ♦ Datenbankmodelle für unterschiedliche Abstraktionsebenen
 - semantische Datenmodelle
 - logische Datenmodelle
 - Datenbankmodelle
- ♦ Miniwelt der Anforderungsanalyse hingegen unterliegt keinem speziellen Datenmodell

- Schema der Miniwelt (Diskursbereich) (I) -

Anforderungsanalyse

- ♦ Abgrenzung und Abstraktion
- ♦ nicht alle Dinge müssen verwaltet werden
- ♦ von konkreten zu verwaltenden Dingen muss zu Typen abstrahiert werden

„Dinge“ im Beispiel Lehrbetrieb (Ausschnitt)

- ♦ beschrieben werden sollen die "Phänomene" bzw. "Dinge"
 - {„Informatik“, „Betriebswirtschaftslehre“, „Physik“, „Recht“, ...}
zu **Fachbereiche**
 - {„Müller“, „Meier“, „Lehmann“, ...}
zu **Professoren**
 - {„Schmidt“, „Schuzle“, „Müller“, ...}
zu **Studenten**
 - {„Mathematik I“, „Datenbanksysteme“, „Programmierung I“, ...}
zu **Lehrveranstaltungen**

Weitere typisierte "Dinge" des Lehrbetriebs

- ♦ Fachbereiche
- ♦ Lehrveranstaltungen
- ♦ Professoren
- ♦ **Fahrzeuge** (gehört nicht zum Diskursbereich)
- ♦ Dozenten
- ♦ Angestellte / Mitarbeiter (**Unterschied?**)
- ♦ **Hochschulen** (nicht erforderlich, da nicht mehrere sondern nur die aktuelle betrachtet wird)
- ♦ Studenten
- ♦ Bücher
- ♦ Prüfungen
- ♦ **Informatik** (ist kein typisiertes „Ding“, sondern eher ein konkretes „Ding“ oder Ausprägung eines Merkmals eines konkreten „Dings“)
- ♦ **Lehrplanung** (ist kein Ding, sondern ein Anwendungsfall)

- Schema der Miniwelt (Diskursbereich) (II) -

Merkmale der Dinge im Beispiel Lehrbetrieb (Ausschnitt)

- ♦ Beschrieben werden sollen zudem die **Merkmale** der "Phänomene" bzw. "Dinge"
 - Fachbereiche (**Name**, **Dekan**, **Studenten**, **Professoren**, ...)
 - Lehrveranstaltungen (**Name**, **Fachbereich**, **SWS**, ...)
 - Professoren (**Name**, **Gehalt**, **Fachbereich**, **Fachgebiet**, ...)
 - Dozenten (**Name**, **Fachgebiet**, ...)
 - Angestellte / Mitarbeiter (**Name**, **Fachbereich**, **Anschrift**, ...)
 - Studenten (**Name**, **Anschrift**, **Haarfarbe**, **Fachbereich**, ...)
 - Bücher (**ISBN**, **Titel**, ...)
 - Prüfungen (**Student**, **Lehrveranstaltung**, **Professor**, **Datum**, **Note**, ...)
 - ...
- ♦ **Gehalt** und **Haarfarbe** gehören nicht zum Diskursbereich des Lehrbetriebs
 - (vielleicht aber zu anderen Anwendungsszenarien)
- ♦ **Beziehungsmerkmale** (Zusammenspiel) zu anderen **eigenständigen Dingen** werden gesondert behandelt

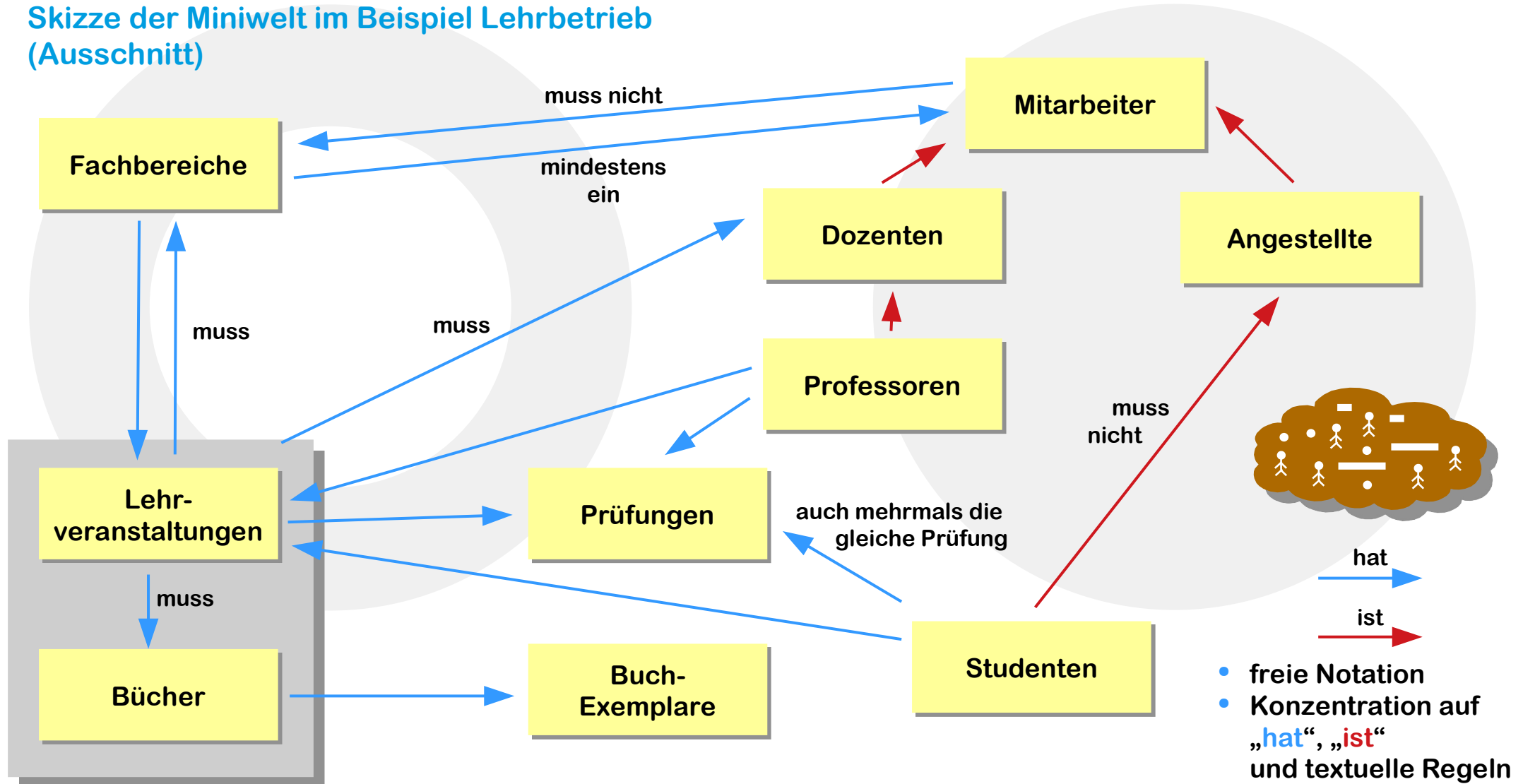
- Schema der Miniwelt (Diskursbereich) (III) -

Zusammenspiel der Dinge im Beispiel Lehrbetrieb (Ausschnitt)

- ♦ Beschrieben werden sollen zudem die **Regeln** des sinnvollen Zusammenspiels zwischen den "Phänomenen" bzw. "Dingen"
 - Fachbereiche bieten Lehrveranstaltungen an
 - jeder Fachbereich **muss** mindestens einen Mitarbeiter haben
 - Mitarbeiter sind Profs, Angestellte und Studenten
 - Professoren halten Lehrveranstaltungen
 - Studenten belegen Lehrveranstaltungen
 - Professoren prüfen Studenten über Lehrveranstaltungen
 - Studenten können **auch mehrmals** über die gleiche Lehrveranstaltung geprüft werden
- jede Lehrveranstaltung **muss** zu einem Fachbereich gehören, und **muss** von einem Dozenten gehalten werden
- Mitarbeiter **müssen nicht** unbedingt zu Fachbereichen gehören Studenten **können** gleichzeitig Angestellte sein
- zu **jeder** Lehrveranstaltung gibt es Buchempfehlungen
- Bücher können mehrere Exemplare haben
- ♦ Fragen
 - erfordert die Existenz eines Dings immer eine Beziehung zu anderen Dingen oder kann ein Ding auch alleine existieren?
 - kann ein Ding auch mehrere Beziehungen zu Dingen desselben Typs haben?
- Ziel: Erkennung und Vermeidung unsinniger Zustände in der späteren Datenbank

- Schema der Miniwelt (Diskursbereich) (IV) -

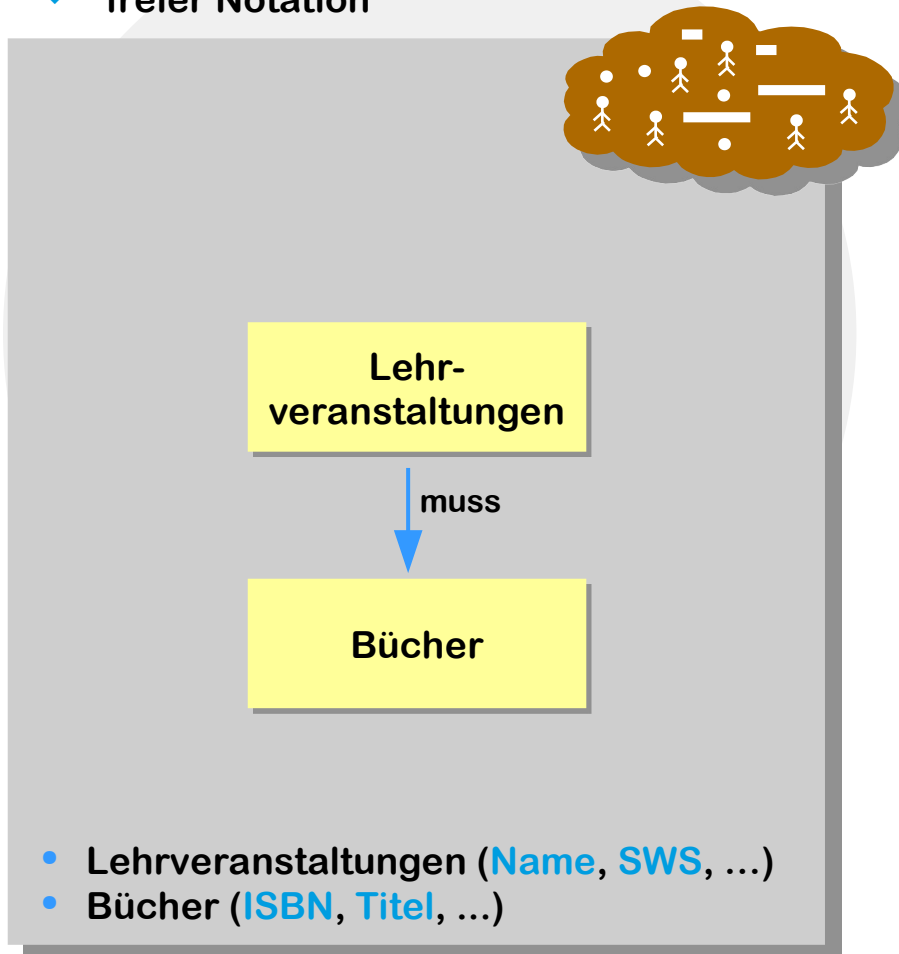
Skizze der Miniwelt im Beispiel Lehrbetrieb (Ausschnitt)



- Miniwelt vs. Semantisches Datenschema -

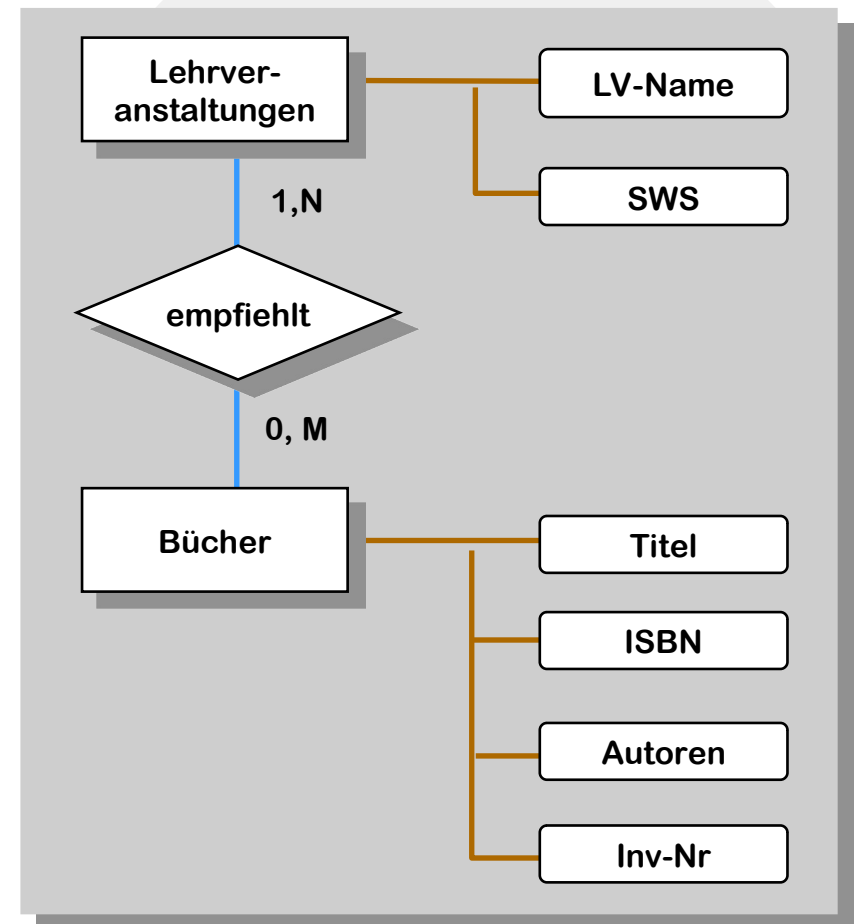
Miniwelt (Diskursbereich) mittels

- ♦ freier Notation



Semantisches (konzeptionelles) Schema mittels

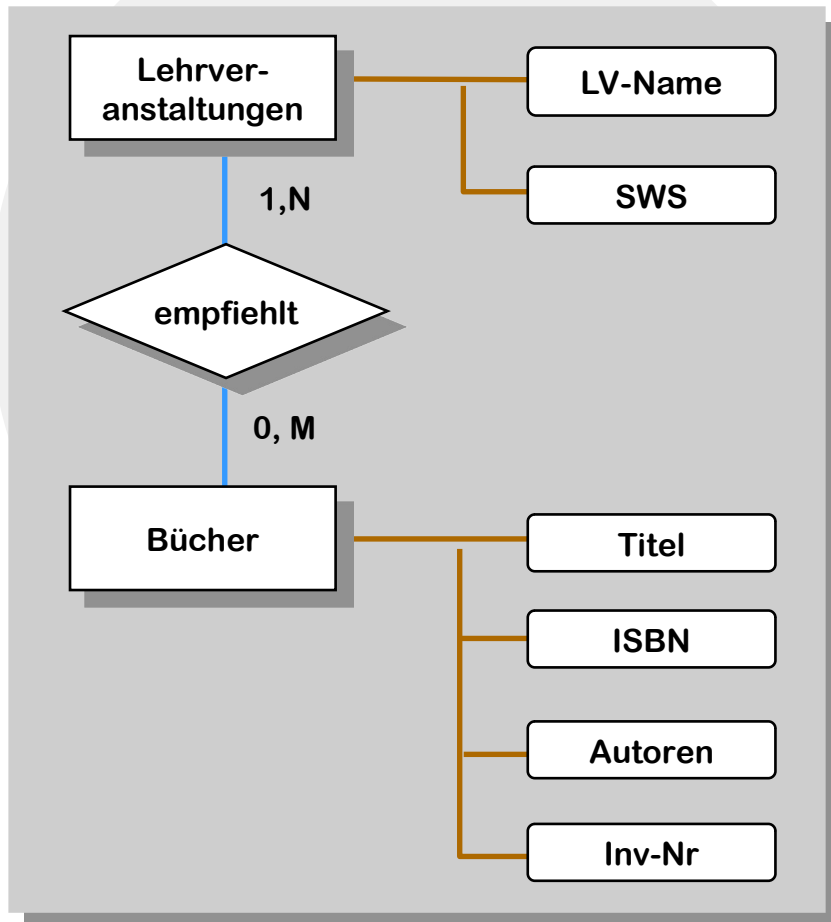
- ♦ „Entity Relationship“-Datenmodell



- Semantisches vs. logisches Datenschema -

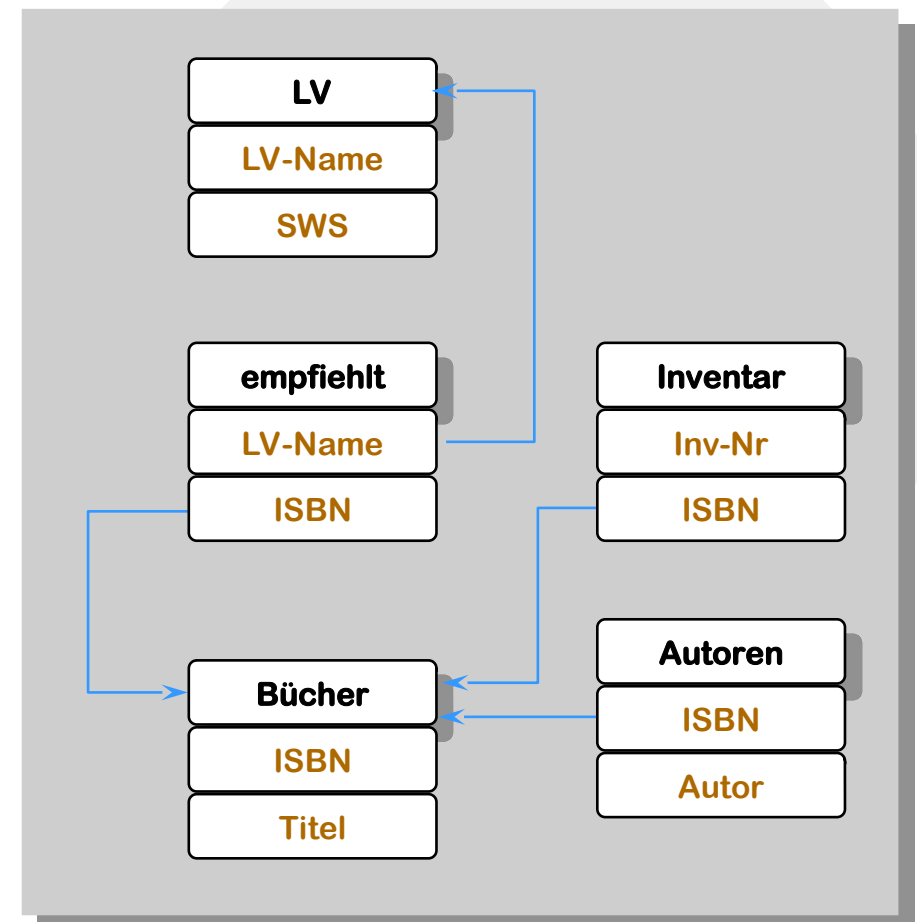
Semantisches (konzeptionelles) Schema mittels

♦ „Entity Relationship“-Datenmodell



Logisches Schema mittels

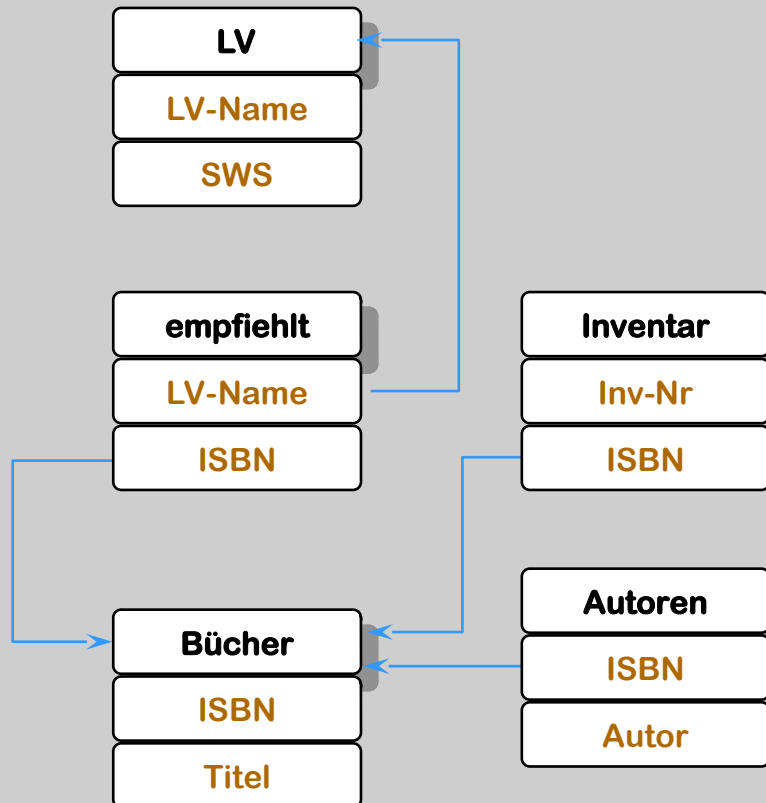
♦ Relationen-Datenmodell



- Logisches Datenschema vs. Datenbankschema -

Logisches Schema mittels

♦ Relationen Datenmodell



Datenbankschema mittels

♦ DDL der SQL

```

create table Lehrveranstaltung (
    LV-Name varchar(20) not null ,
    SWS numeric(1) not null ,
    primary key (Vorl-Titel));

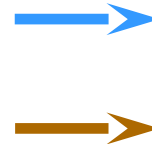
create table Buecher (
    Titel varchar(20) not null ,
    ISBN char(10) not null ,
    primary key (ISBN));

create table empfiehl (
    LV-Name varchar(20) not null ,
    ISBN char(10) not null ,
    primary key (ISBN, Vorl-Titel));

alter table empfiehl
    add constraint RK_emp_Buc
    foreign key (ISBN)
    references Buecher;

alter table empfiehl
    add constraint RK_Vor_emp
    foreign key (LV-Name)
    references Lehrveranstaltung;

. . .
  
```



- Transformation der Schemata -

