

- Sommersemester 2006 • Institut für Germanistik I

## Vorlesung Computerphilologie

# Themenfeld Datenbanken

„Welche Optionen hat man beim Aufbau einer Datenbank und was braucht man für was?“

# Datenbank - Definition

- Eine Datenbank ist eine Sammlung von Daten die zueinander in Beziehung stehen.
- Eine Datenbank ist die moderne (elektronische) Form eines oder mehrerer Karteikästen.
- Typischerweise gehören die Daten zu demselben Themenbereich.
- Um eine Datenbank zu erstellen und zu benutzen wird ein Management-Programm benötigt.

# Datenbanksystem

- Ein Datenbanksystem ist ein Programm, mit dessen Hilfe eine Datenbank :
  - erstellt
  - verändert
  - durchsucht

werden kann

- in der Literatur wird manchmal auch der Terminus Datenbank-Management-System (DBMS) benutzt
- Beispiele: MS-Access, FoxPro, Oracle, FileMaker

# Datenbank - Prinzipien (I)

- Vermeidung von Redundanz
  - Daten werden möglichst nur einmal gespeichert, auch wenn sie in mehreren Zusammenhängen benutzt werden.
- Konsistenz
  - Plausibilitätsbedingungen werden geprüft
  - z.B. In einer Autoren-DB: ein Jahr muß positiv sein, oder eine Autor-ID muß immer einem Autor in der Datenbank entsprechen
- Integrität
  - DB prüft Daten auf interne Widersprüchlichkeit

## Datenbank - Prinzipien (II)

- Zugriffsschutz:
  - Wer darf Daten lesen oder ändern ?
- Kontrolle des gleichzeitigen Zugriffs mehrerer Anwender
  - typischerweise können mehrere Anwender Daten lesen oder abfragen, aber niemals gleichzeitig sichern.

# Datenbanktypen

- Flache Datenbanken (Record-Datenbanken)
- Indexierte Datenbanken (z.B. Relationale Datenbanken)

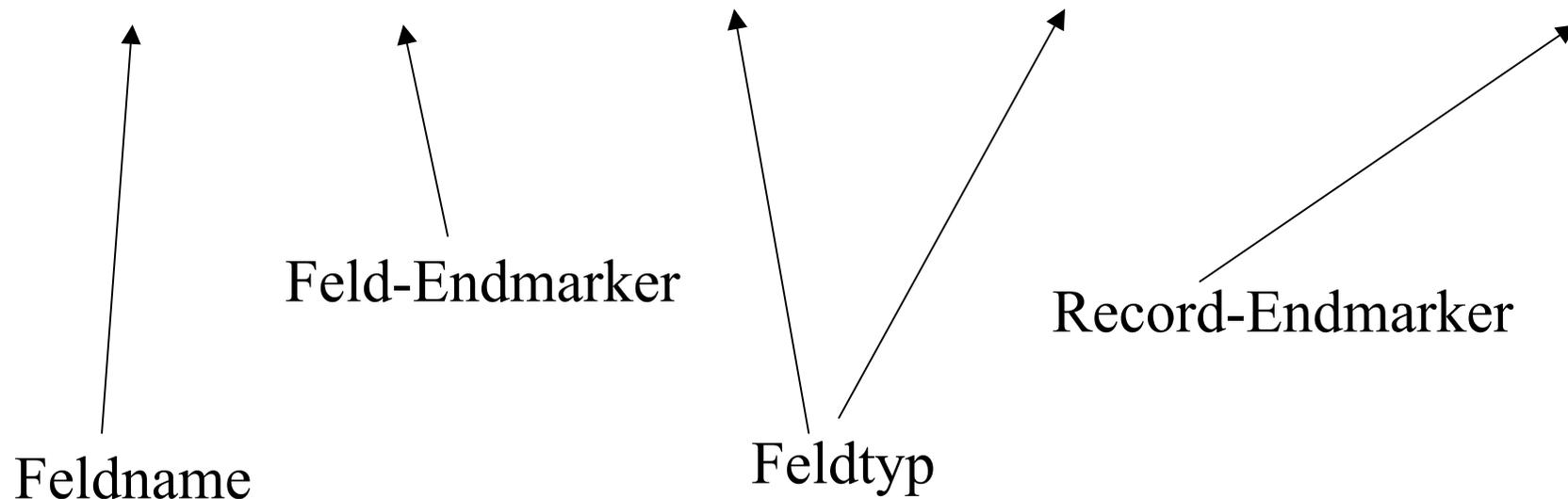
# Flache Datenbanken (I)

- Eine Flache (Flat)- Datenbank: enthält eine oder mehrere Dateien
- Terminologie:
  - Feld
  - Record (Eintrag)
  - Feld Endmarker
  - Record Endmarker

## Flache Datenbanken (II)

- Z. B. Ein Record in einer Autoren-Datenbank:

Name string \t Vorname string \t Jahr number \t Vol. Number \\\



## Flache Datenbanken (II)

- Vorteile
  - einfach zu erstellen und zu benutzen
- Nachteile
  - alle Programme, die die Datenbank benutzen, müssen die Felder genau in der Reihenfolge benutzen. Man kann die DB natürlich in anderer Reihenfolge exportieren)
  - normalerweise wird eine Sortierung nach einem Feld gemacht, um die Datenrecherche zu erleichtern
  - danach machen alle Operationen (einfügen, löschen, ändern) eine neue Sortierung nötig.

# Indexierte Datenbank

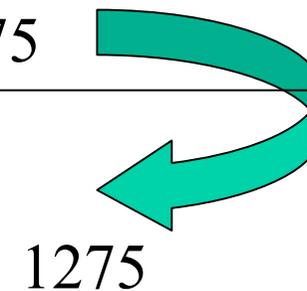
- Indizierung (Indexierung) :
  - Hier gibt es den Begriff eines festen Records nicht mehr,
  - ein Index speichert neben dem Feld die Position der zugehörigen Daten.
  - Ziel ist ein schneller Zugriff
  - Es werden spezielle (komplexe) Datenstrukturen benutzt

# Flache vs. Indexierte Datenbanken

Journal_Name	Vol_number	Serie	Serie -Titel	Preis
“Advances in ..”	12	1	“.....”	20
“Advances in ..”	12	2	“....”	35
“Advances in ...”	12	3	“....”	15

Journal_Name	Vol_number	Serie
“Advances in ...”	12	1275

Serie_Nr	Serie_Titel	Preis
1	“Titel 1”	20
2	“Titel 2”	35



# Datenbanken -Benutzung

- Transaktionen:
  - Folge von Aktionen, die vollständig (oder gar nicht) durchgeführt werden
  - z.B. “Autor wird aus Sammlung A gelöscht und in Sammlung B eingefügt.”.

# Datenbankabfragen - Möglichkeiten

- Vorbereitete Abfragen über Menü-Selektion:
  - einfach zu nutzen, aber starr
- vorbereitete Abfragen über Formular und Eingabe von Werten
  - einfach zu nutzen, aber beschränkte Flexibilität
- natürlich-sprachliche Abfragen
  - komfortabel, flexibel, aber unzuverlässig
- Abfragen mit Hilfe einer Anfragesprache (z.B. SQL):
  - flexibel, zuverlässig aber wenig komfortabel.

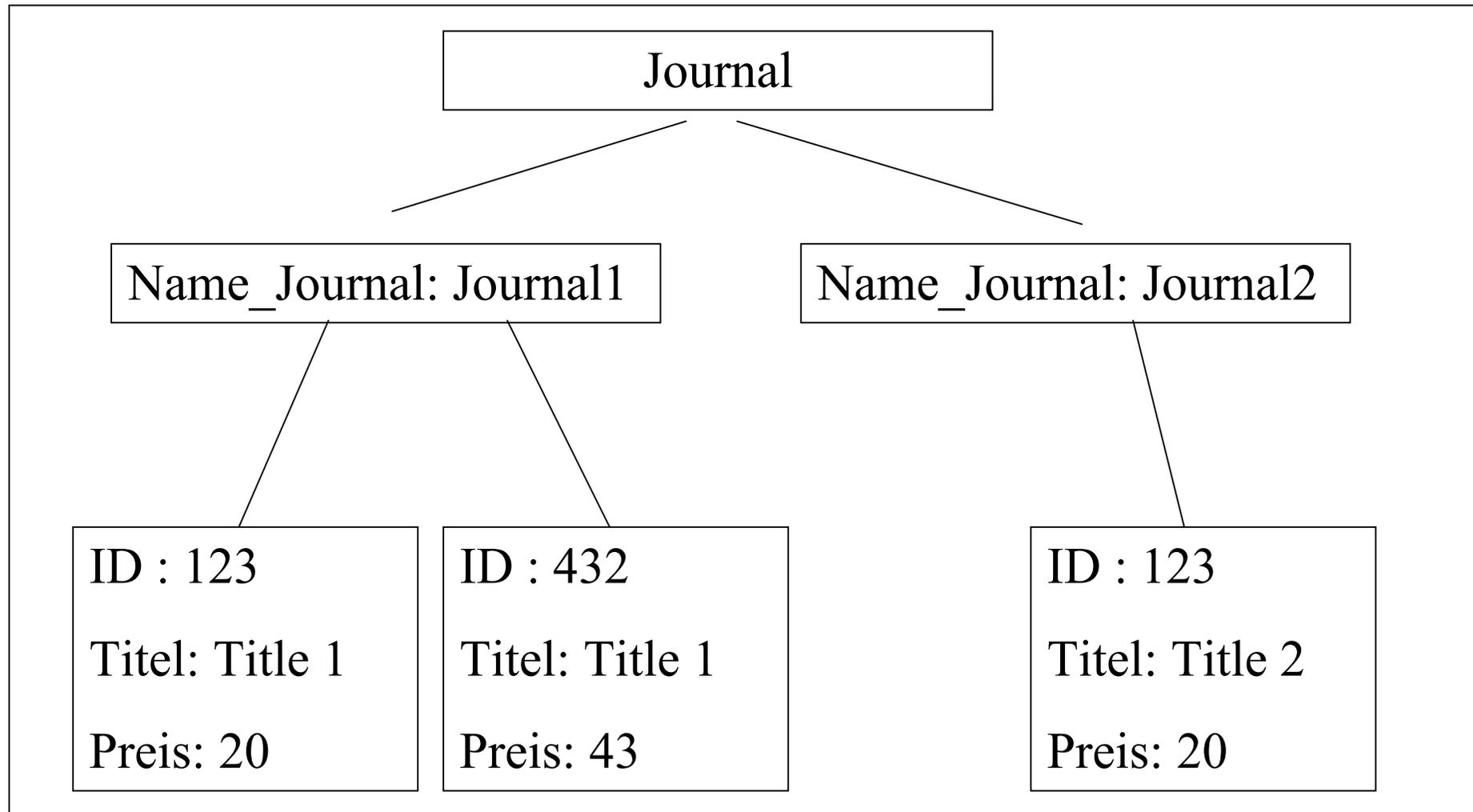
# Datenbanken - Architektur

- Internes Datenbankschema
  - Repräsentation der Daten im Computer (Speicherstruktur und Zugriffsmethoden)
- Konzeptuelles Datenbankschema
  - Abstraktion des Anwendungsbereichs
  - Datenmodell (s. nächste Folie)
- Externes Datenbankschema
  - Sicht des Benutzers (Tabellen, Formular, Bericht usw.)

# Datenbankmodelle

- Hierarchisches Datenbankmodell
  - Datenorganisation ist eine Baumstruktur
  - Problematisch bei Änderungen
- Netzwerk- Datenbankmodell
  - für komplexere Beziehungen zwischen Daten
- Relationales Datenbankmodell
  - wichtigstes Modell in der Praxis,
  - Daten in Tabellen und Untertabellen gespeichert.

# Hierarchische Datenbanken - Beispiel



# Relationale Datenbanken - Beispiel

A	Name_Journal	Exemplare
	Journal 1	3
	Journal 2	2

B	ID.	Titel	Preis
	123	Titel1	20
	222	Titel2	15
	432	Titel1	34

Attribute

Jeweilige primäre Schlüssel

C	Name_Journal	ID.
	Journal 1	123
	Journal 1	432
	Journal 2	222

**Beziehungen:**

Attribute in C:

- Primärer Schlüssel von A
- Primärer Schlüssel von B

# Relationale Datenbank - Begriffe (I)

- Schlüssel
  - primärer Suchbegriff
  - identifiziert einen Datensatz eindeutig
  - dient zur Verknüpfung von Tabellen
- Die Daten werden typischerweise über mehrere Tabellen verteilt. Die Tabellen werden über Schlüsselfelder wieder miteinander verknüpft

## Relationale Datenbank - Begriffe (II)

- Datensatz (Tupel) = Daten in einer Tabellenzeile
- Datenfeld: Spalte in einer Tabelle
  - hat eine feste Länge
  - darf nur Einträge eines Typs enthalten

# Beispiel : Terminologische Datenbanken

- Eine Datenbank für Fachbegriffe die folgende Datenfelder enthält:
  - Benennung (in verschiedene Sprachen), Orthographische Varianten
  - Fachgebiet
  - Definition, Kontextbeispiele
  - Quellenangabe
  - Synonyme, Antonyme, Hyponyme (Thesaurus-Relationen)
  - Symbole, Graphiken, Bilder
  - grammatische Informationen
- <http://www.ttt.org/clsframe>



Quellen:  
Th.Mann  
.....

Sprachen:  
Italienisch  
.....

Gebiete:  
Medizin  
.....

# Was brauchen Computerphilologen?

- Record-Datenbanken haben eine flache Datenstruktur,
- Sie sind einfach zu handhaben und reichen für eine "elektronische Modellierung des Zettelkastens",
- Sie haben aber ggf. eine viel höhere Redundanz (Schreibaufwand!): Sie schreiben auf 22.000 Zetteln: "Verein für die Pflege des niederdeutschen Kulturguts in Hamburg",
- Sie haben meist eine einfache graphisch-interaktive Oberfläche,
- Sie werden meist für interaktive "Einmal-Anfragen" eines Datensatzes benutzt.
- Indexierte, meistens Relationale Datenbanken werden mit einem Datenbankschema in Tabellenform definiert. Hier legt man Objekte, Attribute und Relationen fest,
- Dadurch kann man komplexe Sachverhalte modellieren,
- Relationale Datenbanken werden mit einer komplizierten formalen Anfragesprache benutzt
- Jede Abfrage stellt genau die Daten neu zusammen, die in der Abfrage vorkommen,
- Relationale Datenbanken werden meist von anderen Anwendungssystemen aus benutzt.  
Wenn man solche später benutzt, können RelDBs sinnvoll sein.