
Non-Standard-Datenbanken

Prof. Dr. Ralf Möller

Universität zu Lübeck

Institut für Informationssysteme

Marc Stelzner (Übungen)

Organisatorisches: Übungen

- **Start:** Montag, 19. Oktober 2014
- **Zwei Übungen:** Fr. 8-9 Uhr, 9-10 Uhr , IFIS, Geb. 64, Raum 2035
(Anmeldung über Moodle nach dieser Veranstaltung)
- **Übungsaufgaben** stehen jeweils nach der Vorlesung ca. ab 16 Uhr über Moodle bereit
- **Abgabe der Lösungen** erfolgt bis **Montag 14 Uhr** in der IFIS-Teeküche (jeweils in der Woche nach der Ausgabe, 1 Kasten pro Übungsgruppe)
- Aufgaben können in einer **2-er Gruppe** bearbeitet werden (also bitte Name(n), Matrikelnummer(n) und Übungsgruppennummer vermerken)
- In den Übungen am Freitag wird der Übungszettel besprochen, dessen Lösungen bis zum jeweils vorigen Montag abgegeben werden, und auch Fragen zum jeweils neuen Übungszettel geklärt (ggf. mit Präsenzaufgaben als Hilfestellung)



Vorlesung, Übung und Eigeninitiative

Organisatorisches: Prüfung

- Die **Eintragung in den Kurs** und in eine Übungsgruppe ist **Voraussetzung**, um an dem Modul Non-Standard-Datenbanken teilnehmen zu können
- Am Ende des Semesters findet eine **Klausur** statt
- **Voraussetzung** zur Teilnahme an der Klausur sind mindestens **50% der gesamtöglichen Punkte aller Übungszettel**

Teilnehmerkreis und Voraussetzungen

Studiengänge

- Bachelor **Medizinische Informatik**
- Bachelor **Medieninformatik**
- Bachelor **Informatik**
- Bachelor **Mathematik in Medizin und Lebenswissenschaften**
- Master **Informatik**
- Master **MML**

Voraussetzungen

- Algorithmen und Datenstrukturen
- Lineare Algebra und Diskrete Strukturen 1
- Datenbanken
- Theoretische Informatik (kontextfreie Grammatiken)

Vorteilhaft

- Einführung in die Logik

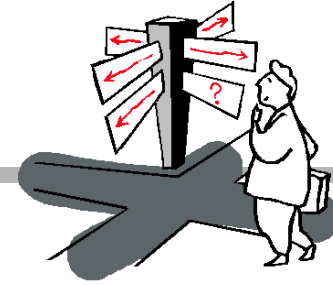
Merkmale von
Standard-Datenbanken?



Merkmale von Standard-Datenbanken

- **Datenmodell: relational** („Tabellen“ und „Tupel“)
- Annahmen:
 - **Strukturen fix**
 - Verwerfen d. A. führt zu **semistrukturierten Datenbanken**
 - **Datenelemente Tupel mit abstrakten Assoziationen**
 - Verwerfen führt zu **temporalen, sequenzorientierten, räumlichen, und multimodalen Datenbanken**
 - **Daten persistent, stabil**
 - Verwerfen führt zu **stromorientierten Datenbanken**
 - **Daten enthalten feste Werte bzw. Referenzen**
 - Verwerfen führt zu
 - Datenbanken für **unsichere Information**
 - Datenbanken für **unvollständige Information**

Inhalt der Vorlesung



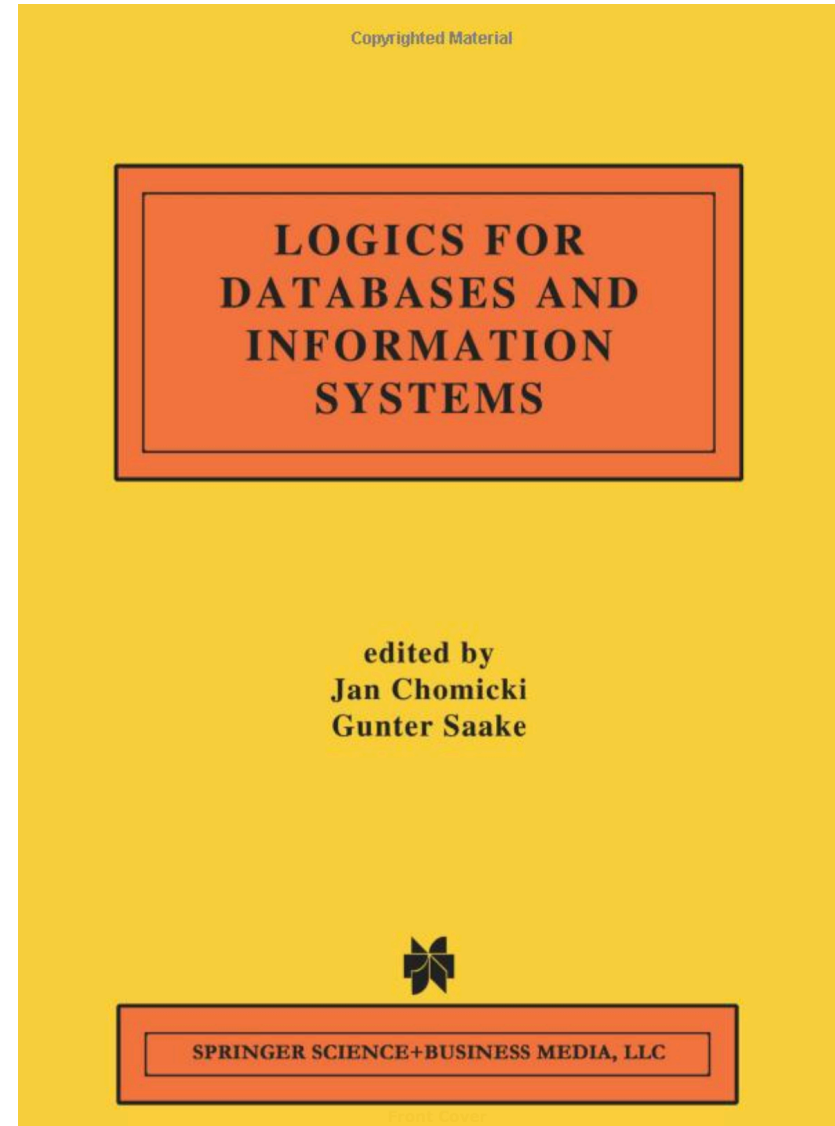
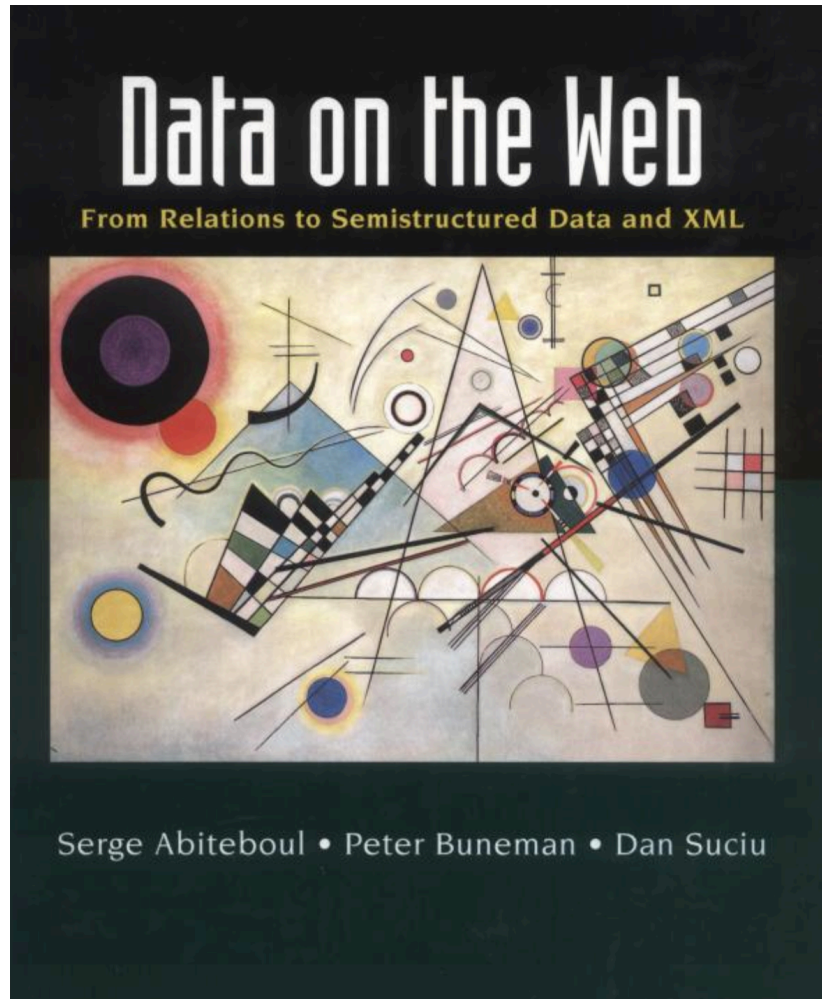
- **Semistrukturierte** Datenbanken (XML)
- **Räumliche und multimodale** Datenbanken
 - lineare und mehrdimensionale Strukturen
- **Temporale Datenbanken**
 - zeitlich beschränkte Gültigkeiten
- Datenbanken für **Datenströme** (Fensterkonzept)
- **Probabilistische** Datenbanken zur Repräsentation **unsicherer Information**
- **Bewertung** von Antworten (Top-k-Anfragen)
- **Ausdrucksstärke:** Rekursive Anfragen

Behandlung **unvollständiger Information** im Master

Literatur

1999

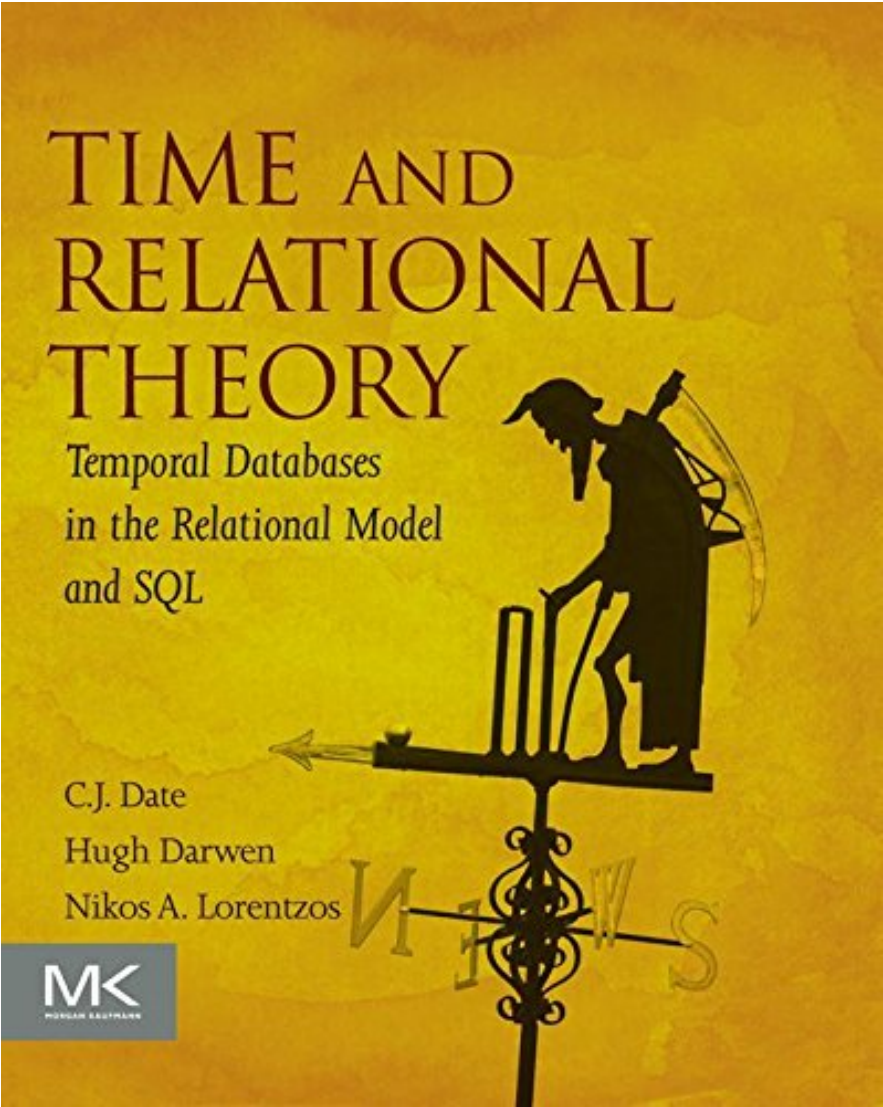
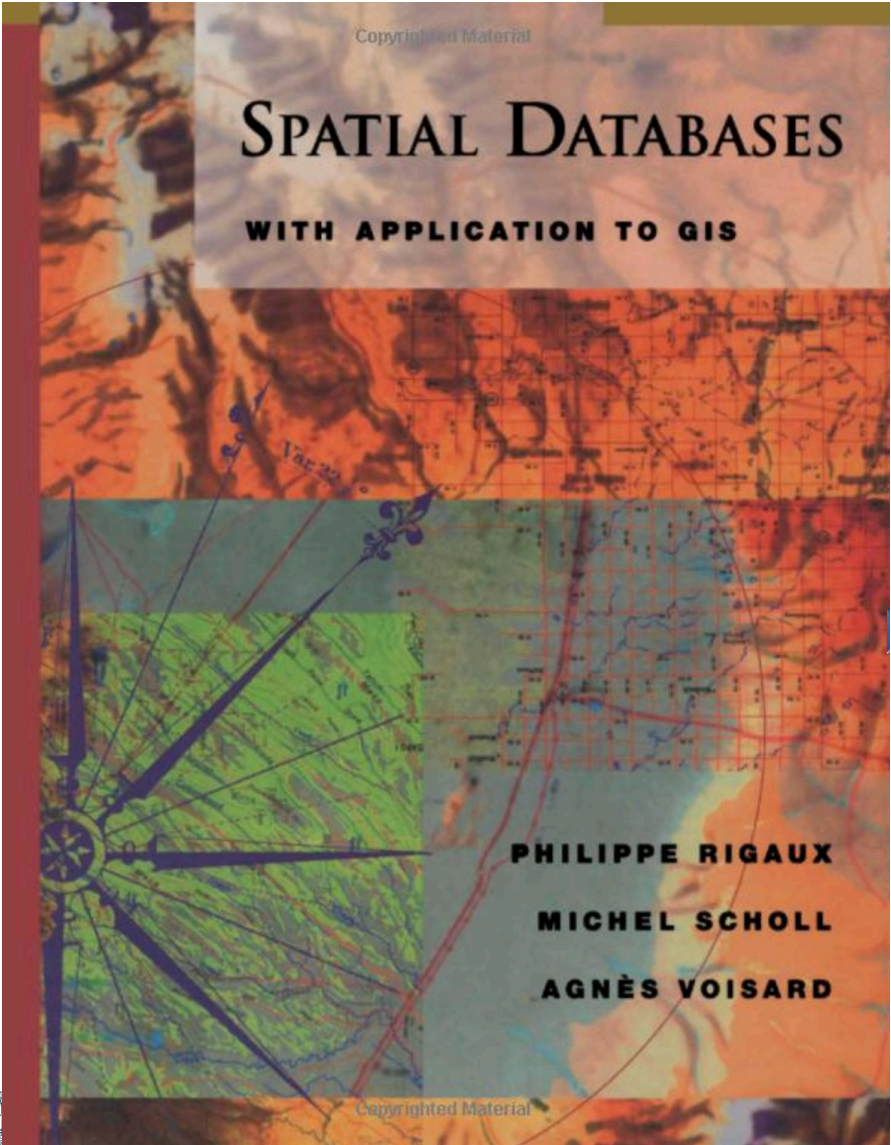
1998



Literatur

2001

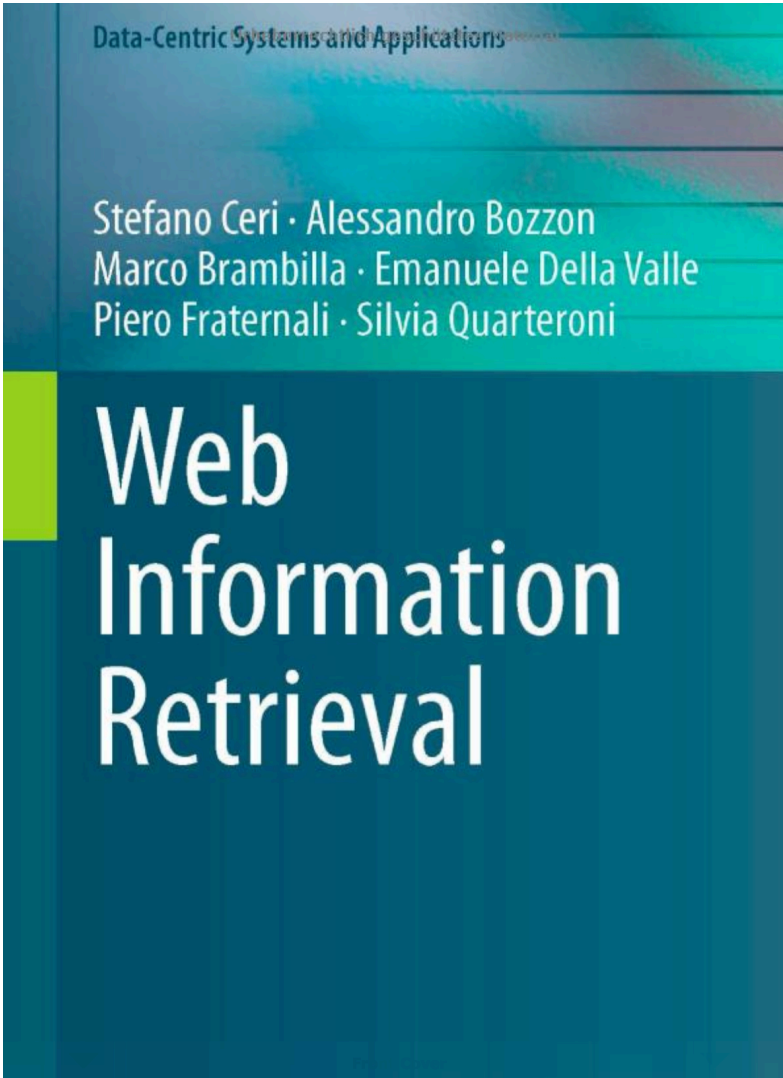
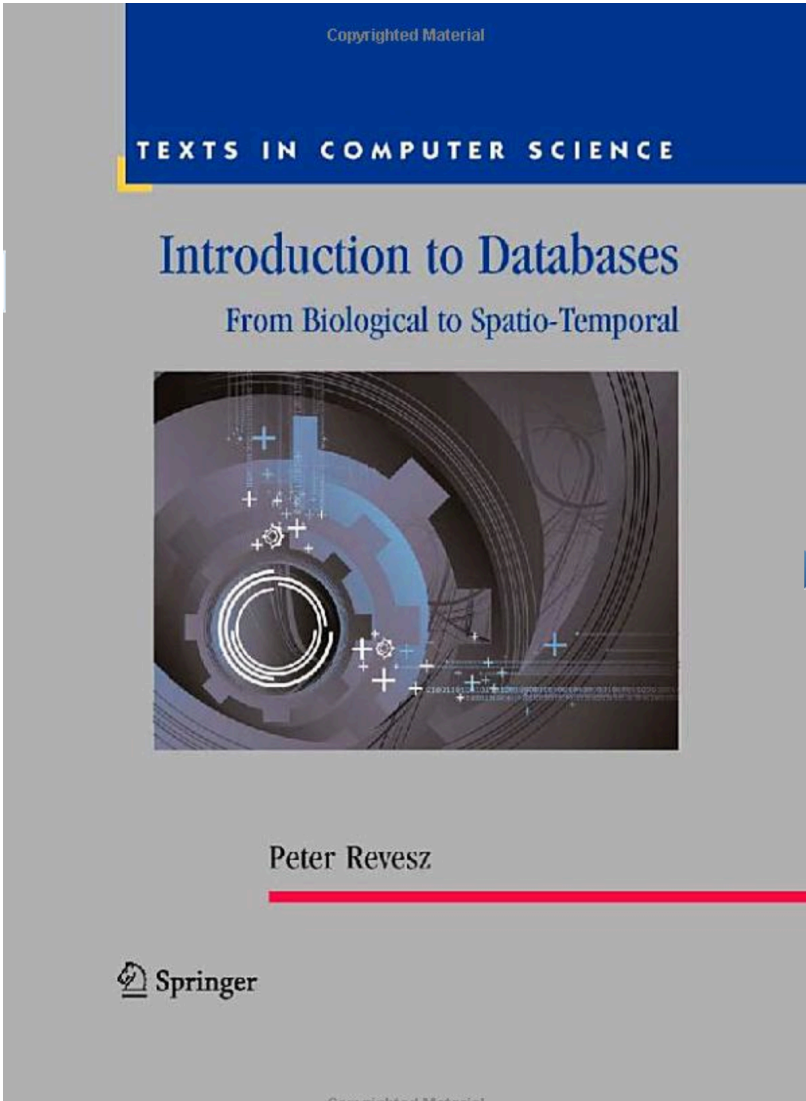
2014



Literatur

2010

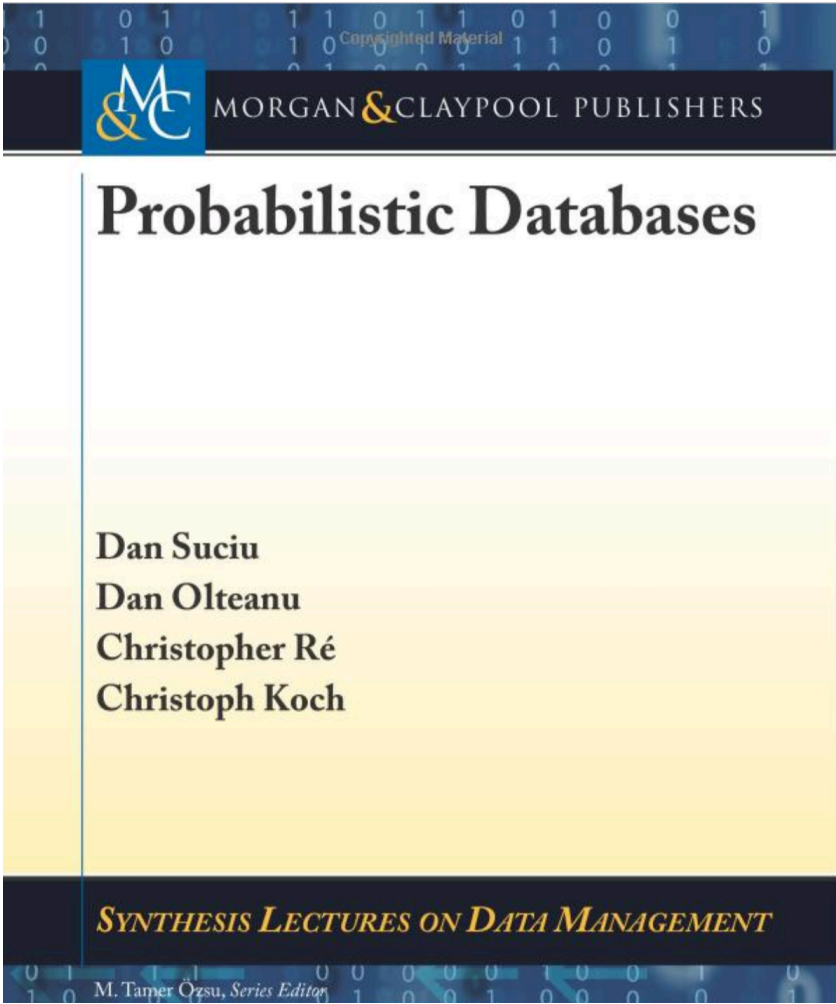
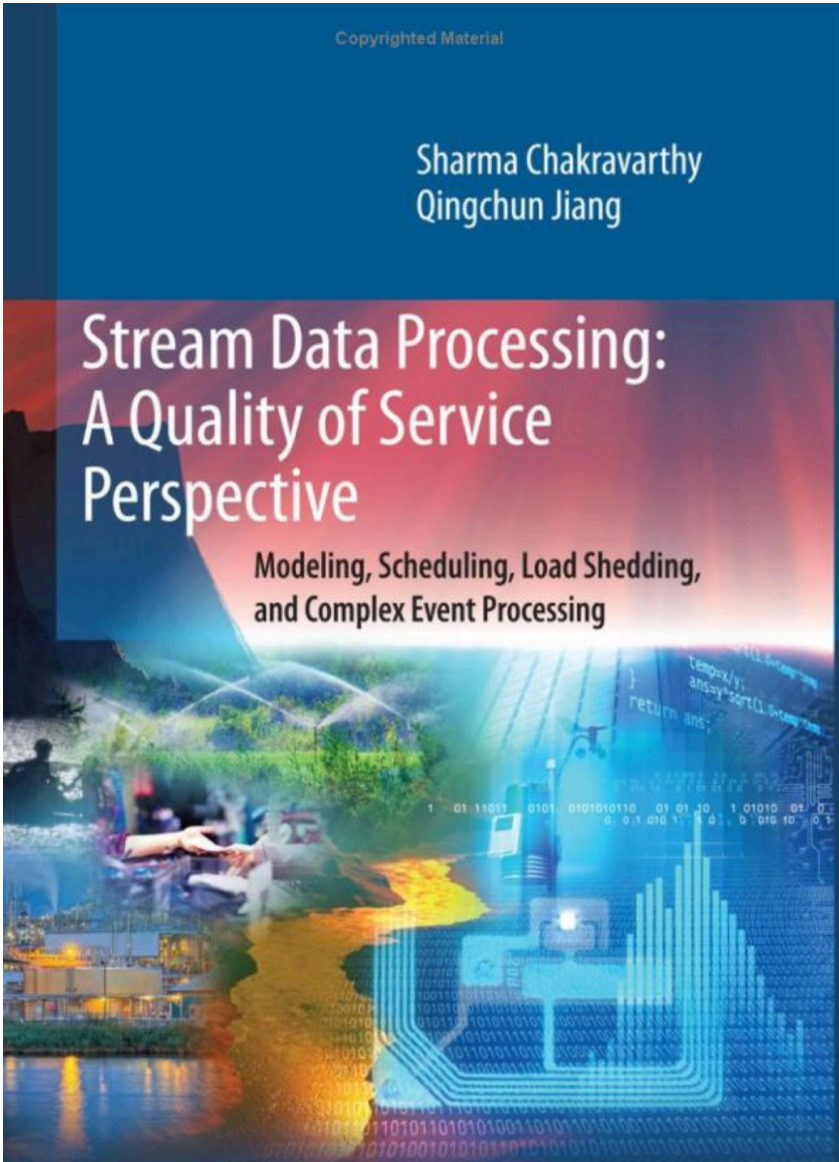
2013



Literatur

2009

2011



Literatur

S. Abiteboul, P. Buneman, D. Suciu, Data on the Web - From Relations to Semistructured Data and XML, Morgan Kaufmann, 1999

J. Chomicki, G. Saake (Eds.), Logics for Databases and Information Systems, Springer, 1998

P. Rigaux, M. Scholl, A. Voisard, Spatial Databases With Applications to GIS, Morgan Kaufmann, 2001

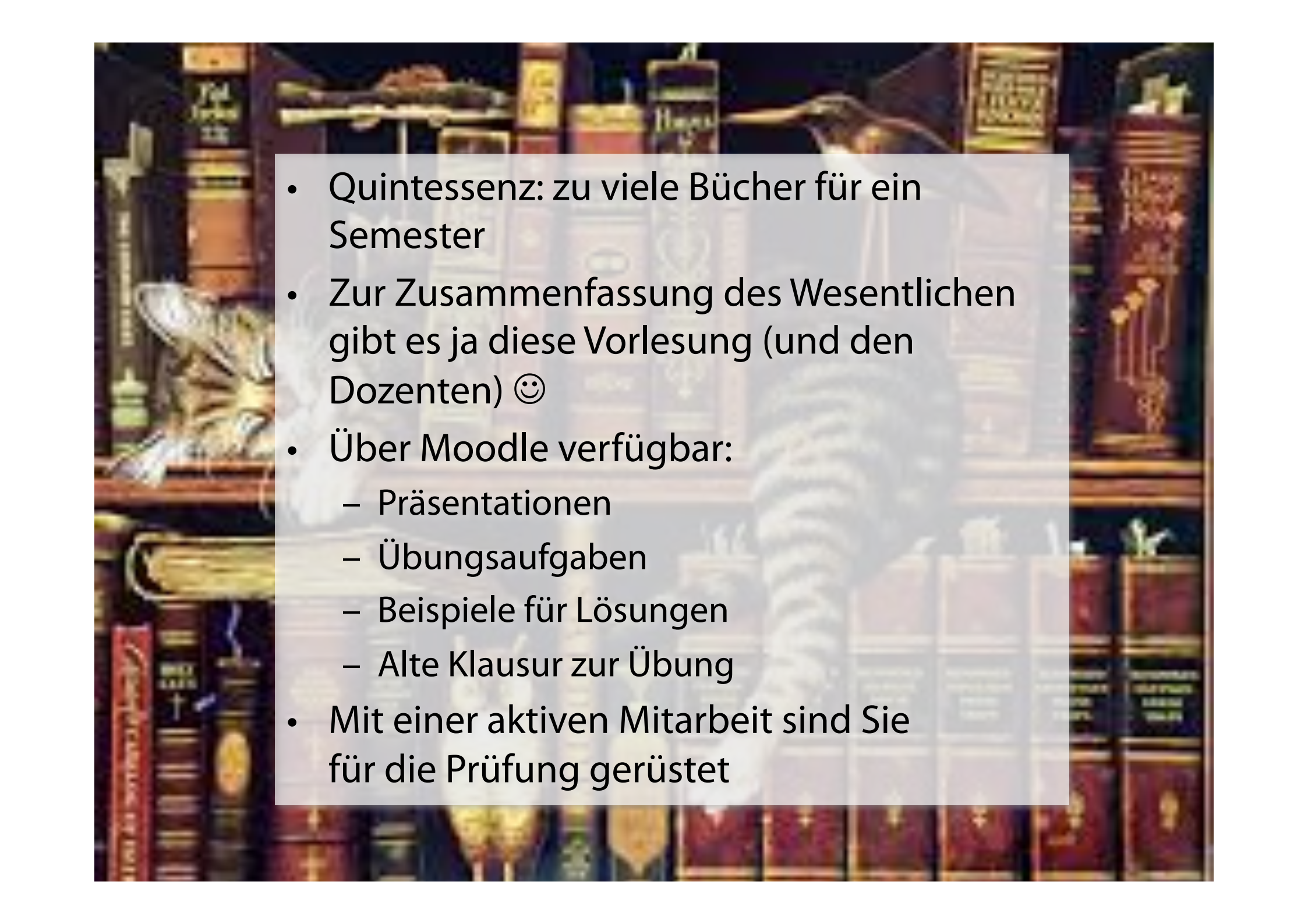
C. J. Date, H. Darwen, N.A. Lorentzos, Time and Relational Theory: Temporal Databases in the Relational Model and SQL, Morgan Kaufmann, 2014

P. Revesz, Introduction to Databases- From Biological to Spatio-Temporal, Springer, 2010

S. Ceri, A. Bozzon, M. Brambilla, E. Della Valle, P. Fraternali, S. Quarteroni, Web Information Retrieval, Springer, 2013

S. Chakravarthy, Q. Jiang, Stream Data Processing A Quality of Service Perspective, Springer, 2009

D. Suciu, D. Olteanu, Chr. Re, Chr. Koch, Probabilistic Databases, Morgan & Claypool, 2011

- 
- Quintessenz: zu viele Bücher für ein Semester
 - Zur Zusammenfassung des Wesentlichen gibt es ja diese Vorlesung (und den Dozenten) 😊
 - Über Moodle verfügbar:
 - Präsentationen
 - Übungsaufgaben
 - Beispiele für Lösungen
 - Alte Klausur zur Übung
 - Mit einer aktiven Mitarbeit sind Sie für die Prüfung gerüstet

Merkmale von Standard-Datenbanken

- **Relationale Anfragesprachen**
 - Relationale Algebra (SQL mit DISTINCT)
 - Tupelorientierter Relationenkalkül (SQL)
 - Domänenorientierter Relationenkalkül (Prädikatenlogik)
- **Deklarative Semantik:**
 - Active-Domain-Semantik (Quantifizierte Variablen laufen über Werte/Referenzen in der Datenbank)
 - Bereichsunabhängigkeit (Anfrageergebnis gleich, auch wenn Variablen über eine beliebige, ggf. unendliche Grundmenge laufen, Safe-Range Normalform für Anfragen)

Aufgabe

Addiere 17 zu den
kleinsten A-Werten in
Tabelle T



UPDATE T SET A = A+17
WHERE A <= ALL (SELECT A FROM T)

Merkmale von Standard-Datenbanken

- Deklarativ: **Semantik** von Anfragen **unabhängig vom Ausführungsmodell** (was gut ist!)

```
UPDATE T SET A = A+17  
WHERE A <= ALL (SELECT A FROM T)
```

- Viele Nutzer verstehen eine Anfragesprache *nur* mit Bezugnahme auf eine suggerierte Ausführung
- Scheinbar werden hier alle A-Werte betrachtet (→ „Performanzbauchschmerzen“)

Merkmale von Standard-Datenbanken

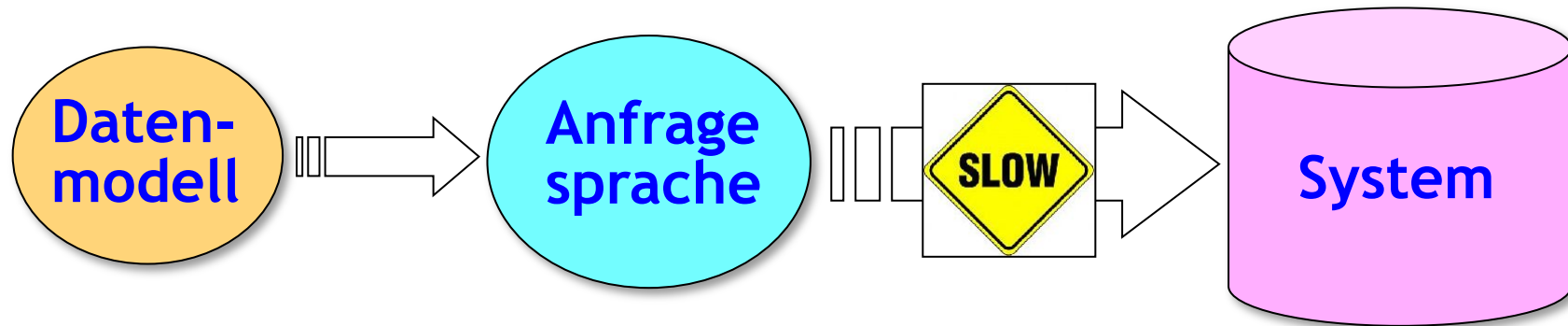
- **Ausdrucksmächtigkeit** von SQL (SQL-92)
 - Prädikatenlogik für Anfragen
 - Domänenmodell (z.B. ER, UML) nicht berücksichtigt
 - Nicht ausdrückbar: Erreichbarkeit von Knoten in einem beliebigen Graphen (Rekursion oder Fixpunkt notwendig)
 - **Korrekte Ergebnisse, keine Approximation**
- **Pragmatik** der Praxis
 - **Daten durch (komplexe) Anwendungsprogramme weiterverarbeitet**
 - Zugriff auf große Datenmengen dabei meist langsam
 - Korrektheit (Serialisierbarkeit) nicht (immer) betrachtet

Merkmale von Standard-Datenbanken

- **Skalierbarkeit, Ausdruckstärke**
- Komplexität des Anfragebeantwortungsproblems für eine gegebene Anfragesprache
 - **Datenkomplexität**
 - Wie wirkt sich eine **Verdopplung des Datenbestandes** bei **fixer Anfrage** auf die **Worst-Case-Laufzeit** der besten Anfragebeantwortungsalgorithmen aus?
 - Kombinierte Komplexität (Anfragelänge ist auch relevant, selten betrachtet)
- **These:** Bei Standard-Datenbanken ist die Ausdruckstärke durch betriebswirtschaftliche Anwendungen beeinflusst

Datenbanken als Wissenschaft

- Motivation Ausdrucksstärke
- Motivation Benutzbarkeit
- Motivation Skalierbarkeit



Ausblick über IFIS Module

- **Bachelor-Programm**
 - AuD, Datenbanken
 - *Non-Standard-Datenbanken*
- **Master-Programm**
 - Webbasierte Informationssysteme
 - Datenmanagement
 - Mobile und verteilte Datenbanken
 - Semantic Web
 - Web and Data Science
 - Ontology-based Data Access
 - Web and Database Mining Agents

