



# Informationsmanagement

Sommersemester 2019

## Kapitel 1: **Grundlagen des Informationsmanagements**

# Informationsmanagement

?

*Fugenlaut  
(dazu später mehr)*

*organisieren  
steuern  
bewirtschaften  
verwalten*

...

## 1.1 | Daten, Information und Wissen

## 1.2 | Informationsmanagement

## 1.3 | Struktur der Lehrveranstaltung

\$?+?⊕⊖%F \$%\$?▷⊖▷↑ ◆▷%¥ ⊕↑%6↑ -- \$▷.:∇6▷ ↑◆▷} #%‡  
↑%F\*\* ▷ \* ▷‡ \$▷.:∇6▷ ↑◆▷} #%‡ \$%\$\$F▷ % ⊕?⊕F▷ □∇6¥,  
\$%\$?▷⊕ ?‡ \$?+?⊕⊖%F ◆∇⊖⊕▷◆∇F¥⊕ =%⊖▷↑ % ◆▷%¥ ⊕↑%6↑  
?‡ F?.:▷, %##∇6¥?⊕ ↑∇ % ↑▷%= ∇.: ⊕#?▷‡↑?⊕⊕⊕?‡ ?↑%F}.  
6%↑◆▷6 ↑◆%‡ #∇‡.:⊖⊕?⊕ \$%\$?▷⊕, ◆▷%6?⊕ =∇6▷ ↑◆%‡ ∇‡▷  
F%⊕⊖%⊖? \* ▷⊕ ‡▷□\$∇6‡⊕ % =▷‡↑%F \$∇∇⊕↑, %##∇6¥?⊕ ↑∇  
↑◆▷ ‡▷□ ⊕↑⊖¥], □◆?#◆ ↑▷⊕↑▷¥ ⊕▷ \* ▷‡=∇‡↑◆-∇F¥ ?‡.:%‡↑⊕ "◆‡  
=%‡] ▷⊖∇▷%‡ #∇⊖‡↑6?▷⊕, °%6▷‡↑⊕ %6▷ □%6] ∇.⊖? \*?⊕  
% \$?+?⊕⊖%F ▷¥⊖#%↑?∇‡ ↑∇ ↑◆▷?6 \*?¥⊕ %‡¥ ↑6] ↑∇ ⊕▷%\*\*  
∇‡F} ∇‡▷ F%⊕⊖%⊖, " ⊕%?¥ ⊕↑⊖¥] %⊖↑◆∇6 ↗%#□⊖▷⊕ =▷◆F▷6  
∇.: ↑◆▷ F%⊕⊖%⊖, #∇‡?↑?∇‡, %‡¥ ¥▷ \* ▷F∇◦=▷‡↑ F%\$ %↑ ↑◆▷  
?‡↑▷6‡%↑?∇‡%F ⊕#◆∇∇F .:∇6 %¥ \* %‡#▷¥ ⊕↑⊖¥?▷⊕ ?‡  
↑6?▷⊕↑▷, ?↑%F}. "↑◆▷} %6▷ %.:6%?¥ [↑◆▷?6 #◆?F¥6▷‡] =?⊖↑  
⊕⊖.:.:▷6 □◆▷‡ ↑◆▷⊖▷↑ ↑∇ ⊕#◆∇∇F %‡¥ ⊕∇ ∇‡, " =▷◆F▷6 ⊕%?¥.  
"\$▷#%⊖⊕▷ ∇.: ∇⊖6 6▷⊕⊖F↑⊕, ? ¥∇⊖\$↑ ↑◆%↑ \* ▷6] =⊖#◆."

- Gegeben sei ein **Alphabet**  $\Sigma = \{A, B, C, \dots, a, b, c, \dots, 0, 1, \dots\}$
- Ein Element  $s' \in \Sigma$  heißt **Symbol** oder **Zeichen**
- Eine Folge von Symbolen  $s = s_1 s_2 \dots s_n \in \Sigma^*$  heißt **Zeichenkette** oder **Zeichenfolge** (*String*)
- Eine Zeichenkette, die nach bestimmten Regeln (**Grammatik**) gebildet wird, nennt man auch **Nachricht**
- Eine Definition für **Daten** lautet: ausgetauschte Nachrichten; d.h. noch nicht interpretierte Zeichenfolgen, die nach gewissen Regeln hergestellt wurden
  - *Beispiel: Die Zeichenkette der vorigen Folie, die durch das Medium Folienpräsentation vom mir (Sender) zu Ihnen (Empfänger) übertragen wurde*

(Heyer et al. 2006)

- Daten, die nicht interpretiert sind oder gar nicht interpretiert werden können, sind nicht sehr nützlich!
- Erst durch Herstellen eines **Interpretationsbezugs** werden Daten zu Information
- Eine **Information** ist folglich eine Nachricht mit Bedeutung für den Empfänger
- *Beispiel: Für viele Empfänger dürften die Daten der vorletzten Folie nicht interpretierbar sein.  
Daher: keine Information!*

- Selbst wenn ein Empfänger Daten interpretieren kann, also Informationen erlangt, müssen diese nicht nützlich sein
  - *Beispiel: Die Daten „Der Mond ist kleiner als die Erde“ kann (vermutlich) von Ihnen interpretiert werden. Die daraus abgeleitete Information ist aber (vermutlich) nicht nützlich, z.B. da bereits bekannt*
- Damit Informationen für Empfänger nützlich sind, müssen diese mit aktuellen oder früheren Informationen vernetzt werden
  - „Die Informationen werden mit unserem **Wissen** vernetzt“
- **Wissen**: Gesamtheit der Kenntnisse, Fähigkeiten, Fertigkeiten, die Personen zur Lösung von Problemen einsetzen

# Beispiel

**Wissen**

*+ Vernetzung*

**Information**

*+ Interpretation*

**Daten**

*+ Syntax*

**Zeichenkette**

Bis dahin muss ich  
vorbereitet sein!

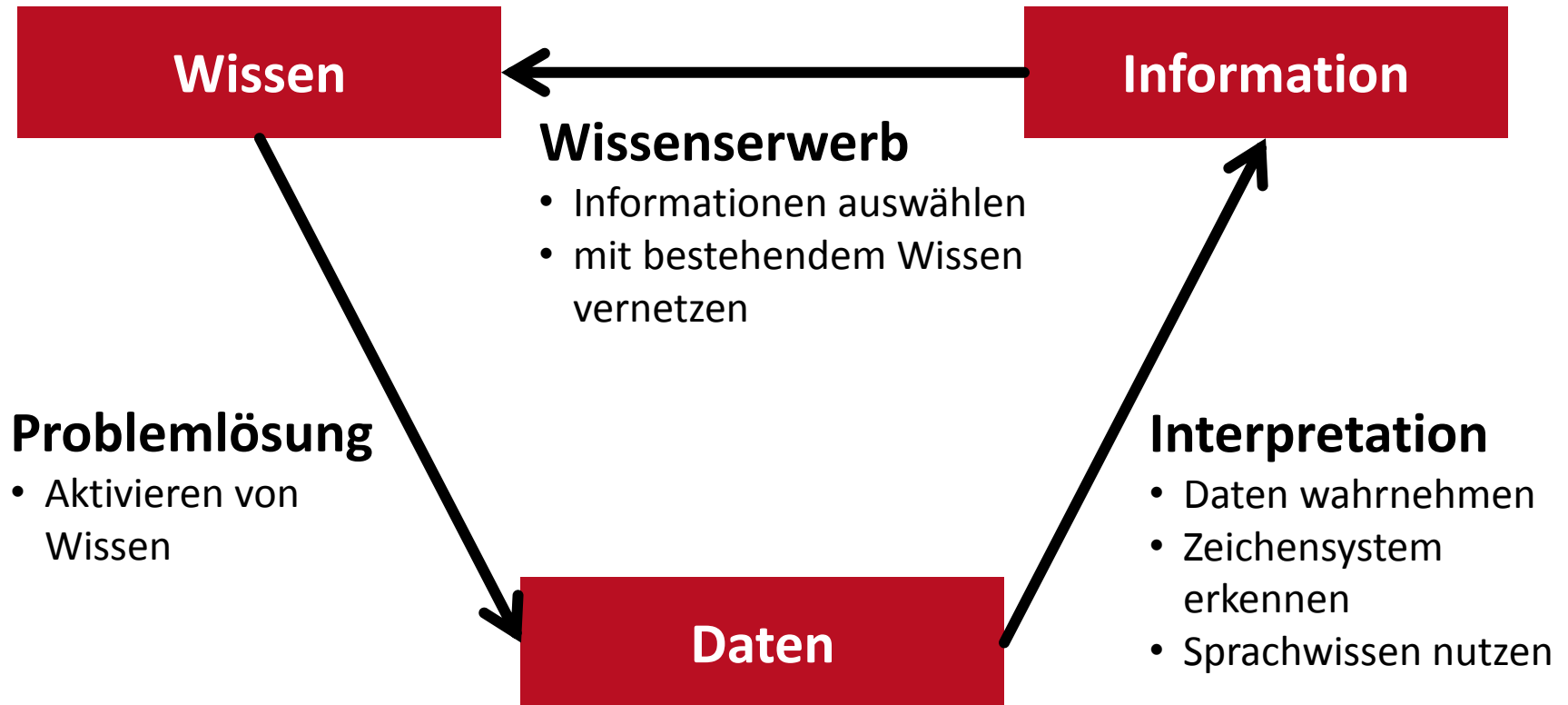
Am 09.09.2019 ist die  
InfMan-Klausur.

09.09.2019

2▲0↕1◆.

(in Anlehnung an M. Ziesak)





(nach Lehner, 2000)

## Strukturierte Daten

- Daten mit gleichartiger Datenstruktur, einem Datenmodell folgend
- *Beispiele:* **Tabelle**, Diagramm, Wort-/Namensliste, Excel-Dokument, Messwerte, Instanzen einer Java-Klasse,...

## Semistrukturierte Daten

- Daten ohne festes Datenmodell, die Strukturen implizieren oder ein erweiterbares Datenmodell enthalten
- *Beispiele:* XML-Dokumente (mit und ohne DTD), Object Exchange Model

## „Unstrukturierte“ Daten

- Daten ohne formalisierte (inhaltliche) Struktur
- *Beispiele:* **Text**, Schaubild, Fotografie, Musikstück, Video,...

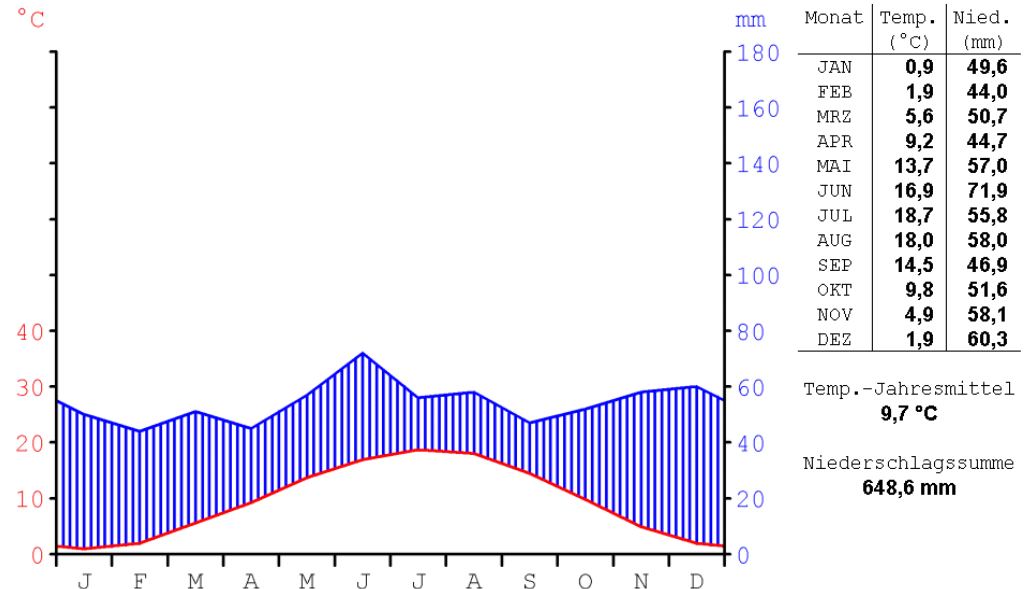
# Strukturierte Daten

## Veralteter Wortschatz

### B...

- Backfischaquarium
- Baffer
- barbieren
- baulich<sub>2</sub>
- beizu
- Bettgänger
- beweiben
- Bibax
- Bierhobel
- ...

**Frankfurt am Main (Hessen)/Deutschland**  
**50°8'N/8°40'E**  
**125m**



**Studierende:**

Matrikel#	Vorname	Nachname
12345678	Max	Mustermann
12345679	Erika	Mustermann
12345681	John	Doe

# Semistrukturierte Daten

```
<entry xml:id="JB04816" n="2.0923.48">
  <form type="artkopf">
    <ref type="DWB" target="B2S0466Z79"></ref>
    <form type="lemma">Buch</form>
  </form>
  <sense rend="artkoerper">
    <sense n="I" level="2" value="20923.48">
      <def rend="leitbem">das zu einer Werkeinheit gebundene
od geheftete, geschriebene od gedruckte Literaturwerk</def>
      <sense n="1" level="3" value="20923.49">
        <def rend="leitbem">konkret, bes im Hinblick auf
Format u Bindeart, Typographie u Ausstattung</def>
        <sense n="a" level="4" value="20923.50">
          <def rend="leitbem">als handschriftl Buch</def>
          <cit>
            <quote>Das schöngeschriebene B.</quote>
            <bibl>6,38 Vs 4 DivHafis</bibl>
          </cit>
```

[...]

*TEI-kodierter Artikel „Buch“  
des Goethe-Wörterbuchs  
(edd. BBAW/AdW-Gö/HAdW  
seit 1978)*

## Buch

*I das zu einer Werkeinheit gebundene od  
geheftete, geschriebene od gedruckte  
Literaturwerk*

*1 konkret, bes im Hinblick auf Format u Bindeart,  
Typographie u Ausstattung*

*a als handschriftl Buch Das schöngeschriebne B.  
6,38 Vs 4 DivHafis B25,184,8 Lorsbach 31.1.15  
B21,14,21 Christiane 28.7.09 uö*

(Hildenbrandt, 2011)

**Natürlichsprachlicher Text** gilt als „unstrukturiert“

- für **Menschen** sind Texte oft intuitiv verständlich
  - seit Jahrhunderten präferierte Form, um Wissen festzuhalten („in Daten zu kodieren“)
- für **Rechner** sind die Inhalte dagegen nur schwer erfassbar
  - Herausforderungen: Mehrsprachigkeit, Mehrdeutigkeit, lexikalische Variation, implizites Wissen, Fehler, u.v.m.

**Natürlichsprachlicher Text** gilt als „unstrukturiert“

**Aber:** meist lassen sich strukturierte Daten aus Texten ableiten!

- Forschungsgebiet **Text Mining** bzw. allgemein **Natural Language Processing** (NLP)
- Interpretation dieser extrahierten Daten liefert Informationen, die zu neuem Wissen führen können
- Text als „**Wissensrohstoff**“ (Heyer et al., 2006)
- WWW ist die größte bekannte, digital vorliegende Datenquelle

Teil 2 der  
Lehrveranstaltung

**1.1 | Daten, Information und Wissen**

**1.2 | Informationsmanagement**

**1.3 | Struktur der Lehrveranstaltung**

## Daten

- organisieren und strukturieren
- speichern
- abfragen
- neu erstellen, verändern, löschen
- pflegen
- interpretieren (Informationen schaffen)
- vernetzen (Wissenserwerb unterstützen)
- analysieren
- verifizieren
- ...





### Anforderungen:

- Für Anwendung nötige Daten sollen aus den gespeicherten Daten **vollständig abgeleitet** werden können (**Informationserhalt**)
- Wiedergewinnung der Daten soll möglichst **effizient** sein Kapitel 9
- Nur vernünftige Daten sollen gespeichert werden (je nach **Informationsbedarf**) bzw. gespeichert werden können (**Konsistenzerhaltung**)
- Anwendungsdaten sollen **möglichst redundanzfrei** gespeichert werden, um Speicherplatz zu sparen und **Anomalien** zu vermeiden

Kapitel 6

# Typischer Entwurfsprozess (1)

## 1. Anforderungsanalyse (*requirements analysis*)

- Fachwissen, Terminologie, Geschäftsfälle, Kosten-/Nutzen-Analyse, Informationsbedarf
- Informelle Dokumentation, z.B. Interviews, Texte, Formblätter

## 2. Konzeptioneller Entwurf (*conceptual design*)

- Abstrakte Modellierung der Anwendungsdomäne
- Welche Objekte spielen eine Rolle, wie hängen diese zusammen?
- Formale Beschreibung, z.B. ER-Modell, UML-Strukturmodell

## 3. Verteilungsentwurf (*distributed system design*)

- Fragmentierung der Daten, Synchronisation und Replikation

# Typischer Entwurfsprozess (2)

## 4. Logischer Entwurf (*logical design*)

- Abbildung des konzeptionellen Modells auf die Konzepte des eingesetzten Informationssystems, implementierungsspezifisch, aber geräteunabhängig
- Logische Datenmodelle, z.B. relationales Modell

## 5. Datendefinition (*data definition*)

- Deklaration/Programmierung des Datenmodells im Informationssystem
- Definitionssprachen, z.B. SQL

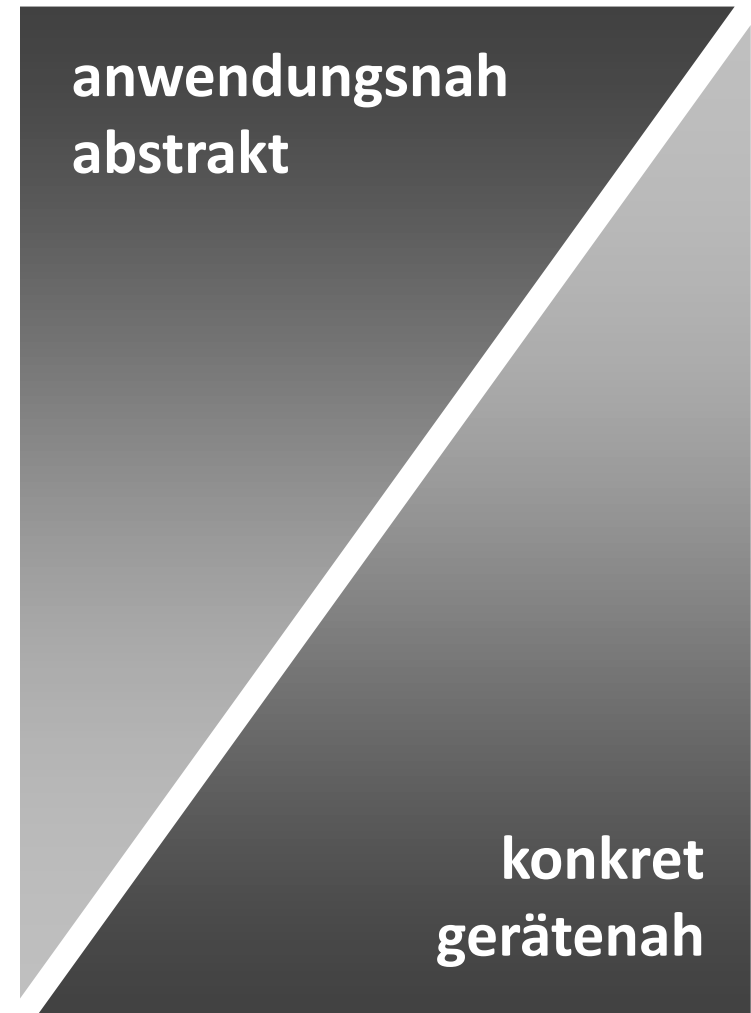
## 6. Physischer Entwurf (*physical design*)

- Definition von Zugriffs- und Speicherstrukturen
- Wie werden die Daten auf der Festplatte abgelegt?

## 7. Implementierung und Wartung

# Typischer Entwurfsprozess (3)

1. Anforderungsanalyse Übung 1
2. Konzeptioneller Entwurf Kapitel 3
3. Verteilungsentwurf
4. Logischer Entwurf Kapitel 4–6
5. Datendefinition Kapitel 7
6. Physischer Entwurf Kapitel 8–10
7. Implementierung und Wartung



Speichern von Daten ist sinnlos, wenn diese nicht **abgefragt** werden können

### **Deklarativer Zugang (Suche)**

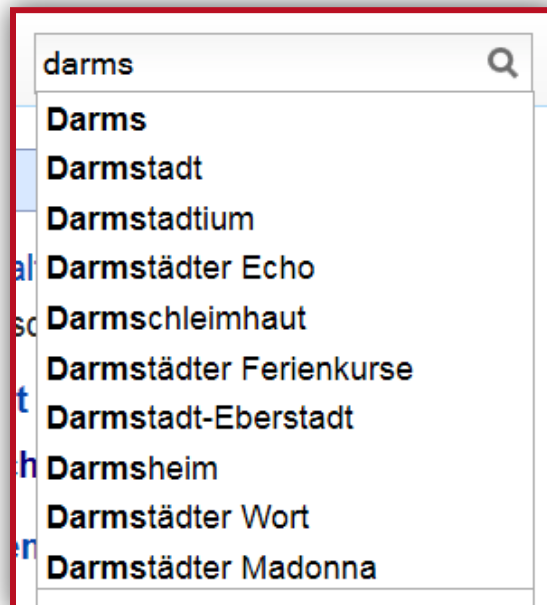
- Spezifiziert durch Prädikate
- *Beispiel: Abfrage aller TUDA-Studierenden mit 1.0 in InfMan*

### **Navigierender Zugang (Blättern)**

- Anfängliche Positionierung und Verfolgen von Zeigern (Pointern)
- *Beispiel: gehe alle Studierenden durch, summiere deren Noten und zähle sie, drucke das arithmetische Mittel, gehe als nächstes zur Liste der Professoren, etc.*

# Datenzugang am Beispiel Wikipedia

## Deklarativer Zugang



## Navigierender Zugang

**Darmstadt** ist eine [kreisfreie Großstadt](#) im Süden [Hessens](#), Verwaltungssitz des [Regierungsbezirks Darmstadt](#) und des [Landkreises Darmstadt-Dieburg](#). Die Stadt gehört zum [Rhein-Main-Gebiet](#) und ist eines der zehn [Oberzentren](#) des Landes Hessen. Darmstadt hat etwa 152.000 Einwohner, die [Stadtregion](#) (*Langer Urban Zone*) Einwohner.

Das **Rhein-Main-Gebiet**, auch **Metropolregion Frankfurt/Rhein-Main**, ist eine der elf europäischen [Metropolregionen](#) in Deutschland, die von der [Ministerkonferenz für Raumordnung](#) definiert wurden. Es ist benannt nach den beiden Flüssen [Main](#) und

Eine **Metropolregion** (in der [Schweiz](#) meist **Metropolitanraum**) ist ein stark verdichteter Ballungsraum einer [Metropole](#). Metropolregionen werden als Motoren der sozialen, gesellschaftlichen und wirtschaftlichen Entwicklung eines Landes betrachtet.

s sowie Teilen der [Pfalz](#) (*Rheinhausen*) der Metropolregion ist [Frankfurt/Rhein-Main](#). Die Einwohner, der in etwa 2,2 Millionen.

# Beispielhafte Fragestellungen

- *Der Präsident der TU Darmstadt möchte wissen, wie viele Studierende ihren Informatik-Master 2017 begonnen haben.*
- *Eine Buchhandlung verkauft 3 Datenbankbücher und ein Notizbuch. Die Verkäuferin muss den Preis wissen und überlegen, ob eine Nachbestellung sinnvoll ist.*
- *Warren Buffett möchte eine Software bauen, die automatisch Börsenpapiere kauft oder verkauft mit dem Ziel der Gewinnmaximierung.*
- *Der Nachrichtendienst X möchte wissen, was Sie in Ihrer Freizeit machen.*
- *IBM möchte ein Informationssystem schaffen, das natürlichsprachliche Fragen beantwortet.*

# IBM Watson



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



**Für Interessierte:**

<https://www.youtube.com/watch?v=Xcmh1LQB9I>

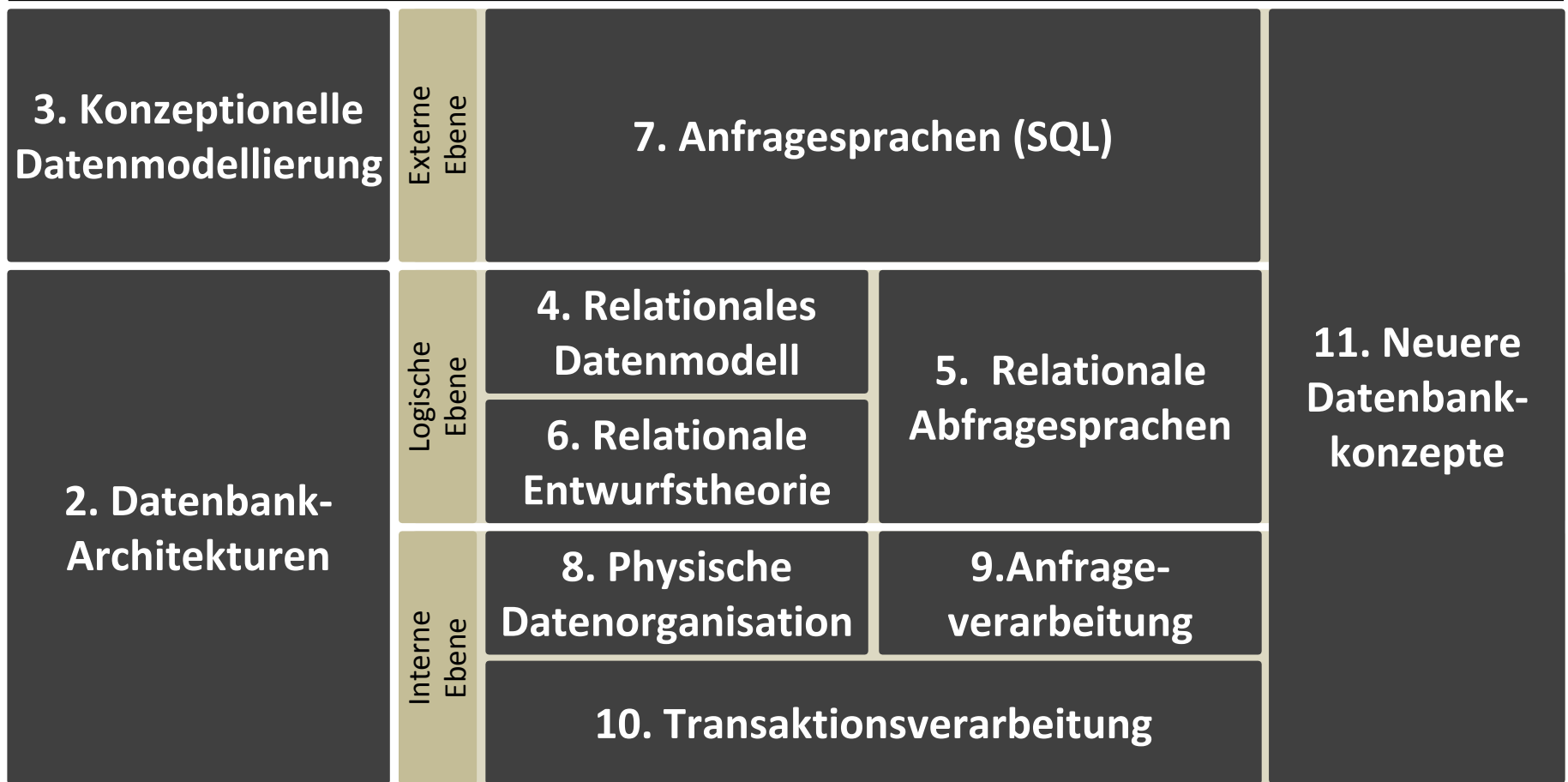


**1.1 | Daten, Information und Wissen**

**1.2 | Informationsmanagement**

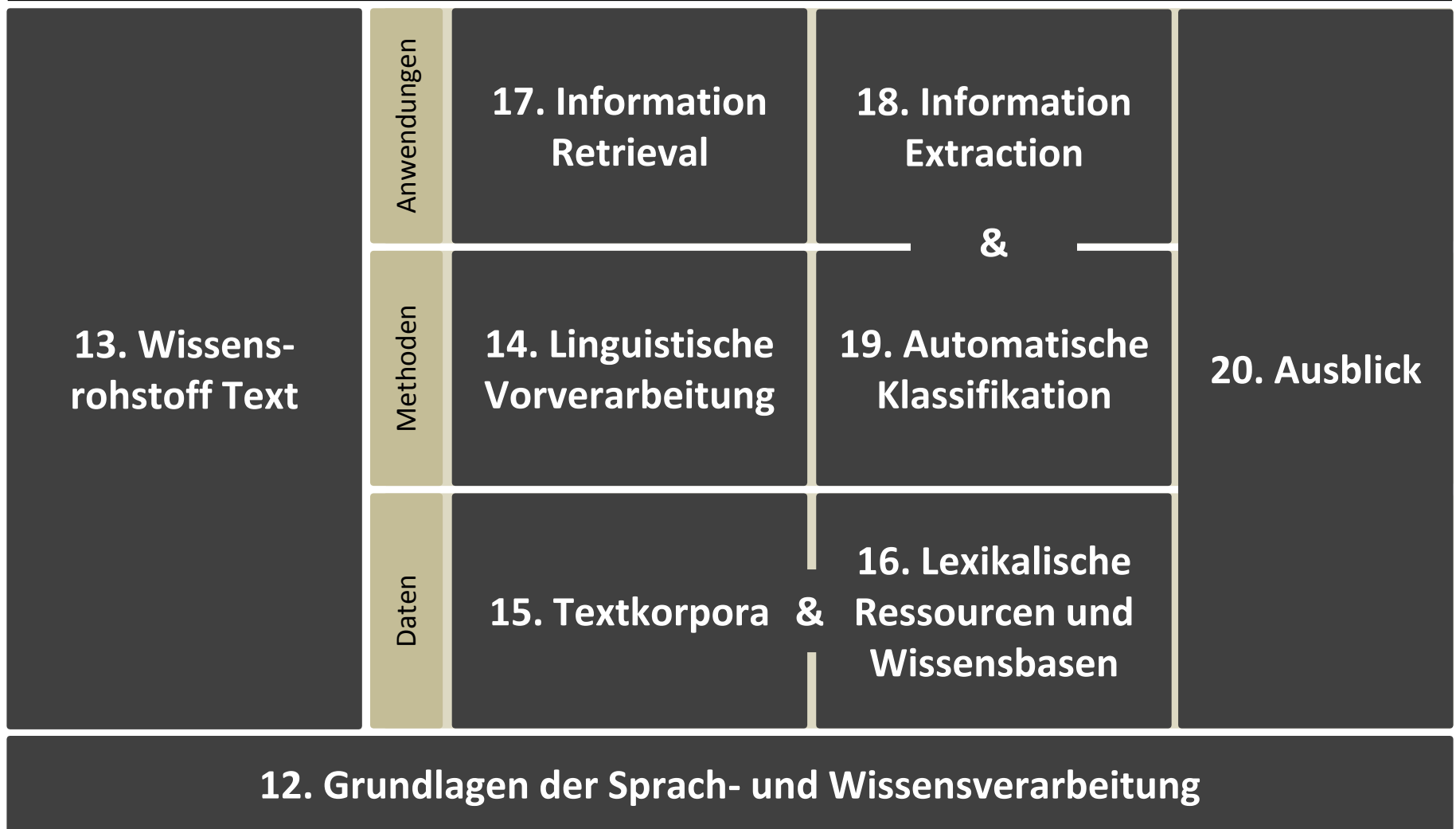
**1.3 | Struktur der Lehrveranstaltung**

# Themenüberblick (Teil 1)



## 1. Grundlagen des Informationsmanagements

# Themenüberblick (Teil 2)





# Fragen?

- Gerhard Heyer, Uwe Quasthoff, Thomas Wittig: *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke/Bochum: W3L, 2006.
- Vera Hildenbrandt: TEI-basierte Modellierung von Retrodigitalisaten (am Beispiel des Trierer Wörterbuchnetzes). In: *Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“* (= OPAL – Online publizierte Arbeiten zur Linguistik 2/2011). S. 21–35, Mannheim: Institut für Deutsche Sprache, 2011.
- Christoph Lehner: Beitrag zu einer holistischen Theorie für die Informationswissenschaften. In: *Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme. Proceedings der 6. Deutschen ISKO-Tagung, Hamburg 1999*, S. 49–66, Würzburg: Ergon, 2000.
- Gunter Saake, Kai-Uwe Sattler, Andreas Heuer: *Datenbanken: Konzepte und Sprachen*. 4. Auflage, Heidelberg: mitp, 2010.

**Weitere Literaturtipps zur Veranstaltung finden sich im Moodle-Kurs!**