

DAB1 – Datenbanken 1

Dr. Daniel Aebi (aebd@zhaw.ch)

Lektion 1: Einführung

Modul DAB1

Lernziele:

- Sie kennen das **Relationenmodell** als theoretische Grundlage der heutigen relationalen Datenbanksysteme.
- Sie kennen die Grundlagen der **relationalen Algebra** und der **relationalen Bags**.
- Sie können das **konzeptionelle** und **logische Schema** eines einfachen Datenbanksystems entwerfen und implementieren.
- Sie können **Datenbankanfragen** in SQL formulieren.

Modul DAB1 – Inhalt

Selbststudium (siehe Dokument «Buff_Logik_Grundlagen.pdf» auf OLAT):

- Logische Grundlagen: Mengenlehre, Boole'sche Algebra und Aussagenlogik, Prädikatenlogik

Vorlesung:

- Relationenmodell
- Relationale Algebra
- Datenmodellierung
- Entity-Relationship Model (ERM)
- Datenbank-Entwurf
- (Normalformen)
- Abfragesprache SQL (Structured Query Language)

Unterlagen

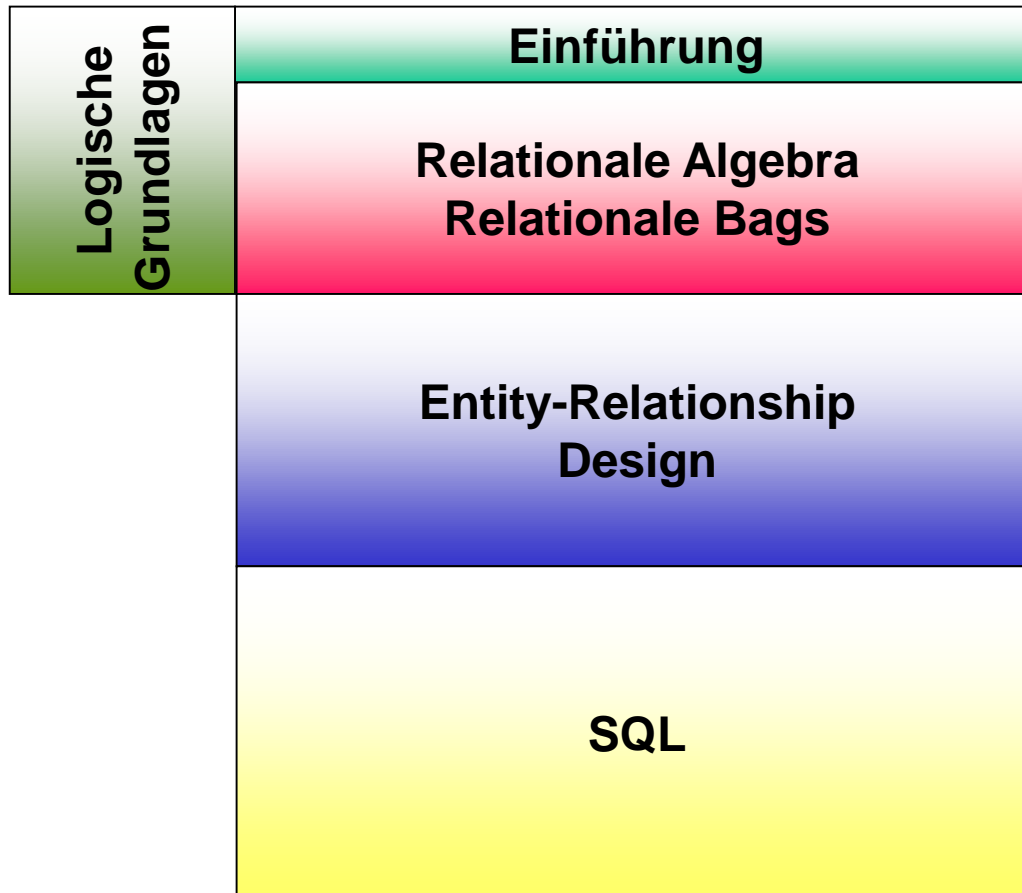
Foliensatz (OLAT):

- Referenzfoliensatz
- Folien des Dozenten pro Lektion

Literatur (optional):

- H. Buff: Datenbanktheorie, ISBN 3-0344-0201-5
- ... es gibt sehr viele Bücher zum Thema Datenbanken

Wo stehen wir?



← "You are here"

Begriff: «Information»



"We have lots of information technology. We just don't have any information."

Begriff: «Information»

- Tja, guter Rat ist teuer... fragen wir doch einfach Wikipedia:

„Information“ wird zum Beispiel auch **synonym** für Nachricht, Auskunft, Belehrung, Aufklärung verwendet,^[5] zum Teil auch für Medien wie Zeitungsartikel, Internet-Seiten, E-Mails, Telefonate, Berichte (Quartals-, Projekt-, Geschäftsbericht), Prospekte und Broschüren, Fahrpläne, Wetterberichte u. v. a. m. – die aber i. d. S. nur die „Träger von Informationen“, nicht die Information selbst sind. Diese Beispiele zeigen die weite Verbreitung und grundlegende Bedeutung des Begriffs 'Information' in nahezu allen (Lebens-) Bereichen.

Ist der Empfänger ein Benutzer eines (elektronischen) Informationssystems, das gewissermaßen als Sender und Kanal fungiert, so gibt es die Effekte des **Browsing** und der **Serendipity**. Unter *Browsing* versteht man das zufällige Aufstöbern von Information während der eigentlichen Suche nach einer bestimmten Information, wobei diese nicht aus den Augen verloren wird. Es können also möglicherweise handlungsrelevante Problemstellungen auch erst durch zufällig gefundene, passende Informationen bewusst werden. *Serendipity* bezeichnet den (zunehmenden) Verlust der Zielorientierung bei der Suche nach bestimmter Information durch Ablenkung, wobei das eigentliche Ziel aus den Augen verloren wird.^[3]

Definitionen [\[Bearbeiten\]](#)

Da der Begriff der Information häufig definiert worden ist, werden einige klassische Definitionsansätze vorgestellt, die zugleich den verschiedenen Bedeutungen der Information entsprechen.^[3]

- Die Definition „Information ist die Teilmenge von Wissen, die von einer bestimmten Person oder Gruppe in einer konkreten Situation benötigt wird und häufig nicht explizit vorhanden ist“ stellt besonders den Bedarf und den Neuigkeitswert aus Sicht des Empfängers (Nutzers) in den Mittelpunkt.
- „Information ist die Verringerung von Ungewissheit aufgrund von fachlichen Informationsprozessen“ ist primär auf den Vermittlungsprozess, also die Tätigkeit des Senders bezogen.
- Zimmermann vertritt einen nutzerorientierten Ansatz, der die handlungsrelevante Wissensveränderung in den Mittelpunkt stellt: „Information ist der (geglückte) Transfer von Wissen,“ ist also das (neue) Wissen, das beim Rezipienten zu einer Veränderung des bisherigen Wissens führt. Im engeren Sinne ist es das Wissen, das einem Menschen (oder einer Institution) zuvor fehlte, um bei einem aktuellen Problem eine sachgerechte Entscheidung zu treffen.
- Mit „Information ist Wissen in Aktion“ macht Kuhlens den Handlungsaspekt von Information deutlich.

Der Begriff in verschiedenen Wissenschaften [\[Bearbeiten\]](#)

Information ist ein weitläufig verwendeter und schwer abzugrenzender Begriff. Verschiedene **Wissenschaften (Struktur- und Geisteswissenschaften)** betrachten die Information als ihr Arbeitsgebiet, namentlich die **Informatik**, die **Informationstheorie** und die **Informationswissenschaft**, die **Nachrichtentechnik**, die **Informationsökonomik** und die **Semiotik**, sie kann ein **mathematischer**, **philosophischer** oder **empirischer** (etwa soziologischer) Begriff sein.

Erst in jüngster Zeit gibt es Bestrebungen, die einzelnen Ansätze zu verbinden und zu einem allgemeingültigen Informationsbegriff zu kommen. Entsprechende Literatur findet sich derzeit meist unter dem Stichwort **Philosophie** (etwa im Bereich **Erkenntnistheorie**). Von einer vereinheitlichten, allgemein akzeptierten **Theorie** der Information kann vorläufig noch nicht gesprochen werden.

Im allgemeinen Sprachgebrauch sowie in einigen Wissenschaften (Semiotik, Informationswissenschaften) wird „Information“ mit „Bedeutung“ oder „übertragenem Wissen“ gleichgesetzt. Eine andere Sichtweise des Begriffes, die heute beispielsweise in der **Computertechnik** von großer praktischer Bedeutung ist, stammt aus der **Nachrichtentechnik**. Die wegweisende Theorie dort ist die von **Claude Shannon**; er betrachtet die statistischen Aspekte der Zeichen in einem Code, der Information repräsentiert. Die Bedeutung der Information geht bei Shannon nur implizit in den Wahrscheinlichkeiten der verwendeten Zeichen ein, die letztlich nur unter Zuhilfenahme eines Menschen bestimmt werden könne, da nur der Mensch in der Lage sei, die Bedeutung eines Codes bewusst zu erfassen und dabei sinnvollen von nicht sinnvollem Code unterscheiden könne. Das unmittelbare Ziel seiner Überlegungen ist die optimale Übertragung von Information in einem Nachrichtenkanal (**Telefonie**, **Funktechnik**).

Der Begriff Information und andere Begriffe aus der Informationstheorie werden oftmals im alltäglichen Sprachgebrauch und auch in den **Naturwissenschaften** in einer **metaphorischen** Weise benutzt. Eine direkte Übernahme des Begriffes Information in naturwissenschaftliche Theorien, so wie er in den **Ingenieurwissenschaften** benutzt wird, wird jedoch von einigen **Wissenschaftstheoretikern** als unzulässig abgelehnt.^[6] So warnte beispielsweise der Wissenschaftsphilosoph **Wolfgang Stegmüller** vor einem Wiederaufleben des **Neovitalismus** durch unangemessenen Gebrauch informationstheoretischer Begriffe in der **Biologie**. Es kann jedoch nicht ausgeschlossen werden, dass in Zukunft der naturwissenschaftliche Strukturbeginn und der Informationsbegriff aufeinander

Von einer vereinheitlichten, allgemein akzeptierten Theorie der Information kann vorläufig noch nicht gesprochen werden.

Begriff: «Daten»

- Lat. datum: „gegebenes“ (Singular oft: Datenelement)
- Zeichen oder kontinuierliche Funktionen, die zum Zweck der maschinellen Verarbeitung Informationen aufgrund bekannter oder unterstellter Abmachungen darstellen.
[DIN 44300]
- Meist numerische oder alphanumerische Angaben über verschiedene Dinge und Sachverhalte.
[Lexikon der Informatik 91]
- Angaben über Dinge und Sachverhalte, die - entsprechend codiert -
elektronisch gespeichert und verarbeitet werden können.

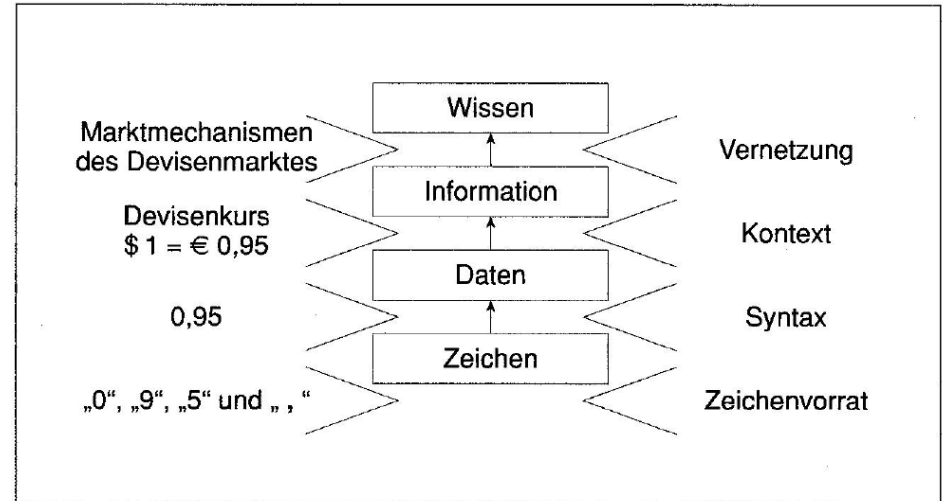
[Aebi 96]

Begriffshierarchie

- Zeichen entstammen einem **Zeichenvorrat**, der beliebig angeordnet sein kann.
- Bei Daten handelt es sich um bereits **strukturierte Zeichen**. Das können alle Zeichen jeglicher Alphabete, Zahlen, Satzzeichen sowie Piktogramme sein. Sie befinden sich immer auf einem Datenträger, wie z.B. auf einem Blatt Papier.
- Wenn diese Daten in einen **Kontext** gestellt werden, erhält man Information.
- Kommt es letztlich zur **Verknüpfung** von Informationen und zur **intellektuellen Einbettung**, so spricht man von Wissen.

Begriffshierarchie

- Diese Begriffshierarchie ist allgemein etabliert
- Es gibt aber **keine präzisen Definitionen**
- Die Übergänge sind fließend (Modell ist eher als Kontinuum zu verstehen)
- Folgende Begriffe und Disziplinen werden unterschieden:
 - Datenmanagement
 - Informationsmanagement
 - Wissensmanagement



Begriffshierarchie: Beispiel

- **Zeichen:**
 - Beispiele: «0», «1», ... «9», «.»
- **Daten:**
 - Beispiele: «293.01», «Hans»
- **Information:**
 - Daten + Bedeutung/Interpretation
 - Information reduziert Unsicherheit, gesprochene und geschriebene Sprache (bestehend aus Symbolen: Worte, Buchstaben, Zahlen)
 - Beispiele: «293.01 Stundenkilometer», «Vorname: Hans»
- **Wissen:**
 - Information + „sinnvolle“ Anwendung
 - «Dieses Auto fährt maximal 293.01 Stundenkilometer»

Begriff: «Datenmanagement»

- Datenmanagement ist die Menge aller methodischen, konzeptionellen, organisatorischen und technischen Massnahmen und Verfahren zur Behandlung der **Ressource Daten** mit dem Ziel, die Daten mit ihrem maximalen Nutzungspotenzial in die Geschäftsprozesse einzubringen und im laufenden Betrieb die optimale Nutzung der Daten zu gewährleisten.

[Wikipedia]

- Wichtige IT-Themen im Bereich Datenmanagement:
 - **Datenbanktechnologien**
 - **Datendefinitions- und Manipulationssprachen**
 - Datenschutz und Datensicherheit
 - Big data, NoSQL, ...
 -

Datenverwaltung – Dateisysteme

- Speicherung von Daten in **Dateien** (verwaltet vom Betriebssystem).
- Anwendungen/Programme lesen/schreiben Daten direkt.
- Vorteile:
 - Einfach, auf Anwendung angepasst, effizient implementierbar.
 - Anwendung muss keine Rücksicht nehmen auf „andere“.
 - Proprietäre (nicht allgemein anerkannten Standards entsprechend) Formate möglich.
 - ...
- Nach wie vor sehr weit verbreitet
 - Office-Anwendungen (Word, Excel, ...).
 - Buchhaltung.
 - ...

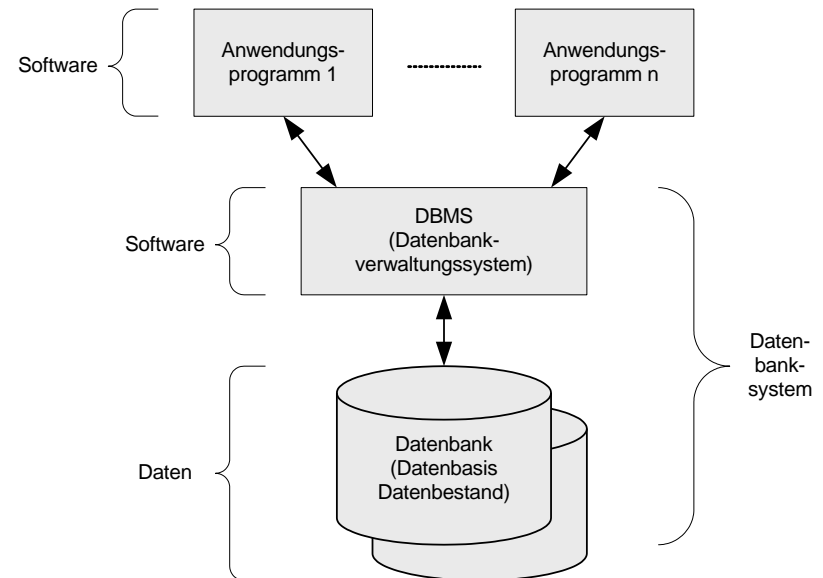
Datenverwaltung – Dateisysteme

- Nachteile:
 - Probleme bei **Mehrfachverwendung** der Daten für unterschiedliche Zwecke.
 - Datenstrukturänderung bedeutet i.d.R. Programmänderung.
 - **Gleichzeitiger Zugriff** aufwändig zu realisieren.
 - Abgestufte Zugriffsrechte aufwändig zu realisieren.
 - Daten werden **oft mehrfach gespeichert**.
 - Datenaustausch, -integration komplex.
 - ...

→ Logische und physische Datenabhängigkeit!
- Deshalb: Einsatz von **Datenbanksystemen**

Begriff: Datenbanksystem

- **Datenbanksystem** (DBS)
 - **DBMS** (Datenbankverwaltungssystem)
 - **Datenbank**
(Datenbasis, Datenbestand)
- Anwendungsprogramme (AP)
- DBS + AP = **Informationssystem** (IS)



Beispiele für Informationssysteme

Hochschulinformationssysteme:

- Verwaltung von Departementen, denen sowohl die Studierenden als auch die Dozenten zugeordnet sind. Studierende belegen Vorlesungen von Dozenten und legen bei ihnen Prüfungen ab.
- Anwendungen sind z.B.: Immatrikulation, Rückmeldung, Exmatrikulationen, Stundenplanerstellung und Planung der Raumbellegung, Ausstellen von Zeugnissen, Statistiken über Hörerzahlen / Prüfungsergebnisse, etc.

Beispiele für Informationssysteme

Informationssysteme einer Fluggesellschaft:

- Verwaltung von Flugstrecken / Flugzeugen / Personal (Piloten, Bord- und Bodenpersonal). Auf diesen Flugstrecken werden Flugzeuge bestimmter Typen mit dafür ausgebildetem Personal eingesetzt.
- Typische Anwendungen sind z.B.: Flugbuchungen von Passagieren, Erstellung von Passagierlisten, Personaleinsatzplanung, Materialeinsatzplanung, Flugplanerstellung, Überwachung der Wartefristen, Gehaltsabrechnung.

Beispiele für Informationssysteme

Bankinformationssysteme:

- Eine Bank gliedert sich gewöhnlich in mehrere Zweigstellen, denen die Angestellten und Bankkunden zugeordnet sind. Verwaltung verschiedenartiger Konten der Bankkunden, z.B. Girokonten, Sparkonten, Hypothekarkonten, Kleinkreditkonten, Wertpapierkonten, etc.
- Anwendungen sind z.B.: Kontoeinrichtung/-auflösung, Buchung von Zahlungsvorgängen, Kreditgewährung, Zinsberechnung/-verbuchung, Personalverwaltung (Gehaltsabrechnung etc.)

Daten > Software > Hardware

In sehr vielen Anwendungsszenarien sind die Daten mit Abstand der **wertvollste** Bestandteil eines Informationssystems.

Faustregel:

- $\text{Wert}(\text{Daten}) = 10 * \text{Wert}(\text{Software})$
- $\text{Wert}(\text{Software}) = 10 * \text{Wert}(\text{Hardware})$

Datenbanksysteme – Ziele

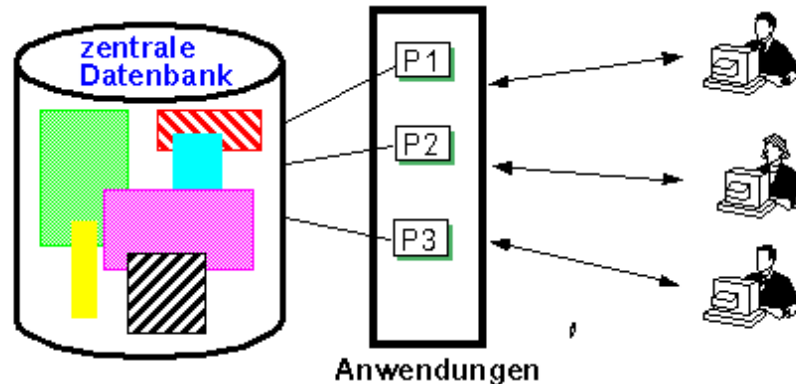
- Ziele, die mit dem Einsatz von Datenbanksystemen verfolgt werden:
 - **Langfristige, dauerhafte** „Aufbewahrung“ von Daten
 - **Datenunabhängigkeit** (Trennung von Speicherung und Verarbeitung)
 - Zentrale Datenhaltung & Konsistenzkontrolle
 - Zentrale **Zugriffskontrolle**
 - Zentrale Redundanzkontrolle
(Redundanz = Mehrfachspeicherung)
 - Viele, **gleichzeitig** arbeitende Benutzer
 - Vereinfachte Anwendungsentwicklung
 - ...

Aufgaben/Eigenschaften eines DBMS

1. Zentrale Kontrolle über die operationalen Daten
 2. Hoher Grad an Datenunabhängigkeit
 3. Hohe Leistung und Skalierbarkeit
 4. Mächtige Datenmodelle und Abfragesprachen / leichte Handhabbarkeit
 5. Transaktionskonzept (ACID), Datenkontrolle
 6. Ständige Betriebsbereitschaft (hohe Verfügbarkeit und Fehlertoleranz)
 - 7x24h-Betrieb
 - keine Offline-Zeiten für DB-Reorganisation u.ä.
- Effiziente und flexible Verwaltung grosser Mengen persistenter Daten (z.B. GBytes - TBytes)

1. Zentrale Kontrolle über die Daten

- Alle (operationalen) Daten können/müssen **gemeinsam benutzt** werden:
Keine verstreuten Dateien → Querauswertungen aufgrund inhaltlicher Zusammenhänge möglich
- Eliminierung der Redundanz: Vermeidung von **Inkonsistenzen** → keine unterschiedlichen Änderungsstände
- Datenbankadministrator (DBA) hat **zentrale Verantwortung** für Daten
- **Einfache Entwicklung** neuer Anwendungen auf der existierenden DB;
Erweiterung/Anpassung der DB (Änderung des Informationsbedarfs)



2. Datenunabhängigkeit

- **Datenunabhängigkeit** = Mass für die Isolation zwischen Anwendungsprogrammen und Daten
- Konventionelle Anwendungsprogramme mit Dateizugriff:
 - Nutzung von Kenntnissen der Datenorganisation und Zugriffstechnik kann gutes Leistungsverhalten ermöglichen, aber datenabhängige Anwendungen sind äusserst unerwünscht (Daten leben länger!)
- Möglichst starke **Isolation** der Anwendungsprogramme von den Daten
- **Physische Datenunabhängigkeit** (Minimalziel):
 - Unabhängigkeit gegenüber Geräteeigenschaften, Speicherungsstrukturen, Indexstrukturen, etc.
- **Logische Datenunabhängigkeit**:
 - Unabhängigkeit gegenüber logischer Strukturierung der Daten
 - i.a. nur teilweise erreichbar

3. Hohe Leistung und Skalierbarkeit

- Hoher Durchsatz / kurze Antwortzeiten für DB-Operationen auf grossen Datenmengen
- Leistungsverhalten:
 - **DBS-Problem**, nicht Anwendungsproblem
 - **Zugriffsoptimierung** für DB-Anfragen durch das **DBS** (Query-Optimierung)
 - Festlegung von Zugriffspfaden (Indexstrukturen), Datenallokation etc. durch den DBA (Idealerweise durch das DBS)
 - Automatische Nutzung von Mehrprozessorsystemen, etc.
- Hohe Skalierbarkeit:
 - Anpassung an gewachsene Leistungsanforderungen (wachsende Datenmengen und Anzahl der Benutzer)
 - Nutzung zusätzlicher/schnellerer Hardware-Ressourcen

→ Modul DAB2

4. Mächtige Datenmodelle

- Operationen zur Definition von Datenstrukturen (Data Definition Language, DDL)
- Definition von Integritätsbedingungen und Zugriffskontrollbedingungen (Datenschutz)
- Operationen zum Aufsuchen und Verändern von Daten (Data Manipulation Language, DML)
- Beschreibung der logischen Aspekte der Daten, neutral gegenüber Anwendungen
- Anwendung erhält **logische**, auf ihren Bedarf ausgerichtete **Sicht** auf die Daten

4. Mächtige Abfragesprachen

Art der Abfragesprache (query language):

- Formale Sprache
- Navigierend oder deskriptiv, abhängig von Datenmodell
- Satz- („record“) oder mengenorientiert

Wünschenswert:

- Deskriptive Problemformulierung
- Leicht erlernbar \leftrightarrow Mächtigkeit
- DB-Zugriff im Dialog und von Programmen aus
- Standardisierung (SQL)
- Geeignet für verschiedene Benutzerklassen
 - Systempersonal, Anwendungsprogrammierer, anspruchsvolle Laien, gelegentliche Benutzer etc.

5. Transaktionskonzept

- **Transaktion**
 - Folge von Lese- und Schreibzugriffen.
 - In sich geschlossene „Arbeitseinheit“.
- Anforderungen an ein DBMS:
 - **Gleichzeitiger** Zugriff **mehrerer** Benutzer ermöglichen.
 - **Automatische** Behandlung von **Ausnahmesituationen** (Fehlern).
 - Schneller (möglichst automatischer) Wiederanlauf nach schwerwiegenden Fehlern.

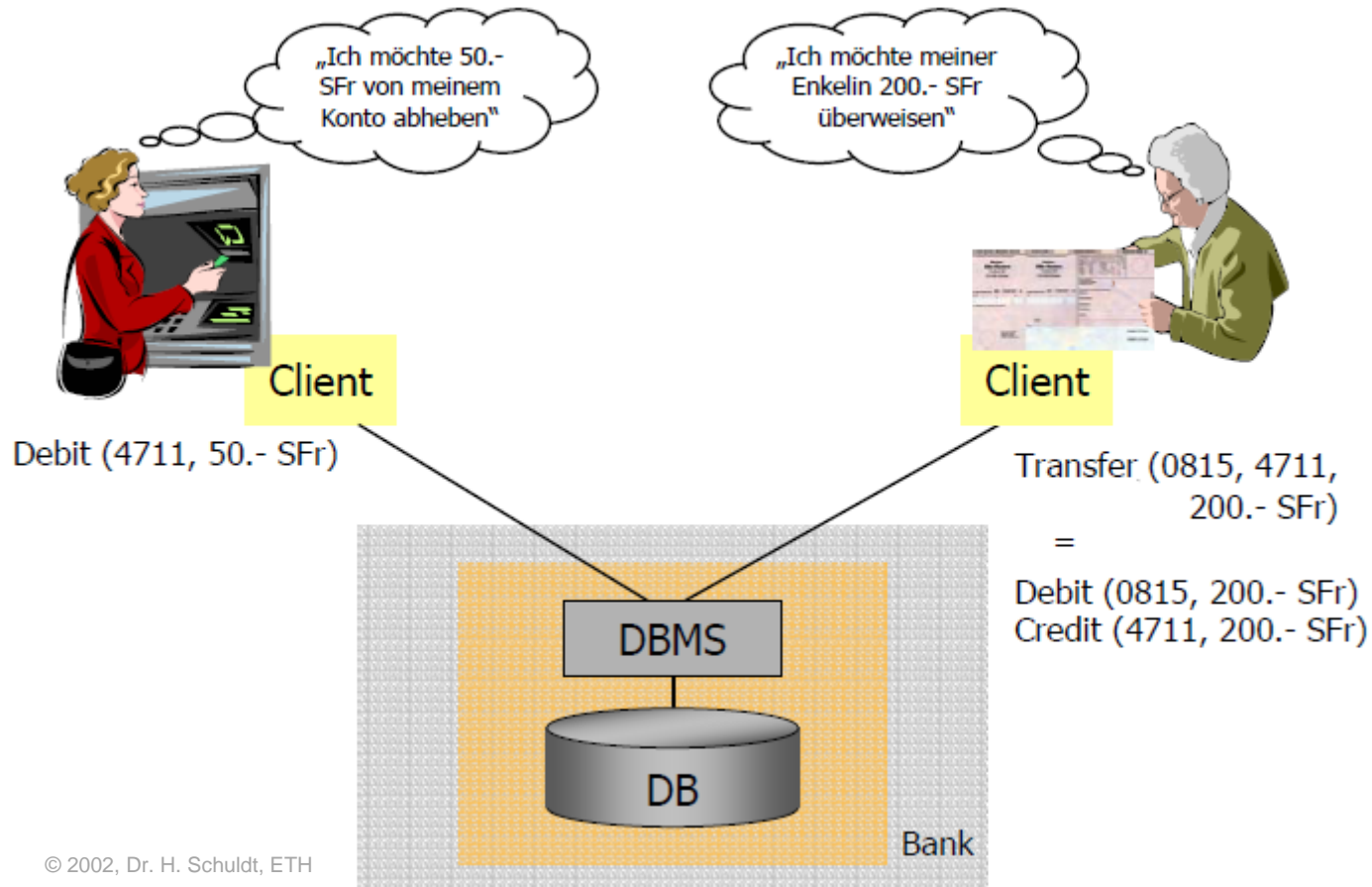
5. Transaktionskonzept

- Anforderungen an Transaktionen, ACID-Prinzip:
 - **Atomicity** - Atomarität
 - Zusammengehörige Aktion werden entweder ganz oder gar nicht ausgeführt.
 - **Consistency** (preservation) - Konsistenzerhaltung
 - Alle Aktionen hinterlassen die Datenbank in einem konsistenten Zustand (alle Integritätsbedingungen bleiben eingehalten).
 - **Isolation** - Isolation
 - Die Aktionen einer Transaktion laufen unabhängig von Aktionen anderer Transaktionen ab.
 - **Durability** - Dauerhaftigkeit
 - Alle Resultate bleiben nach Durchführung der Transaktion dauerhaft.

Beispiel: Banküberweisung

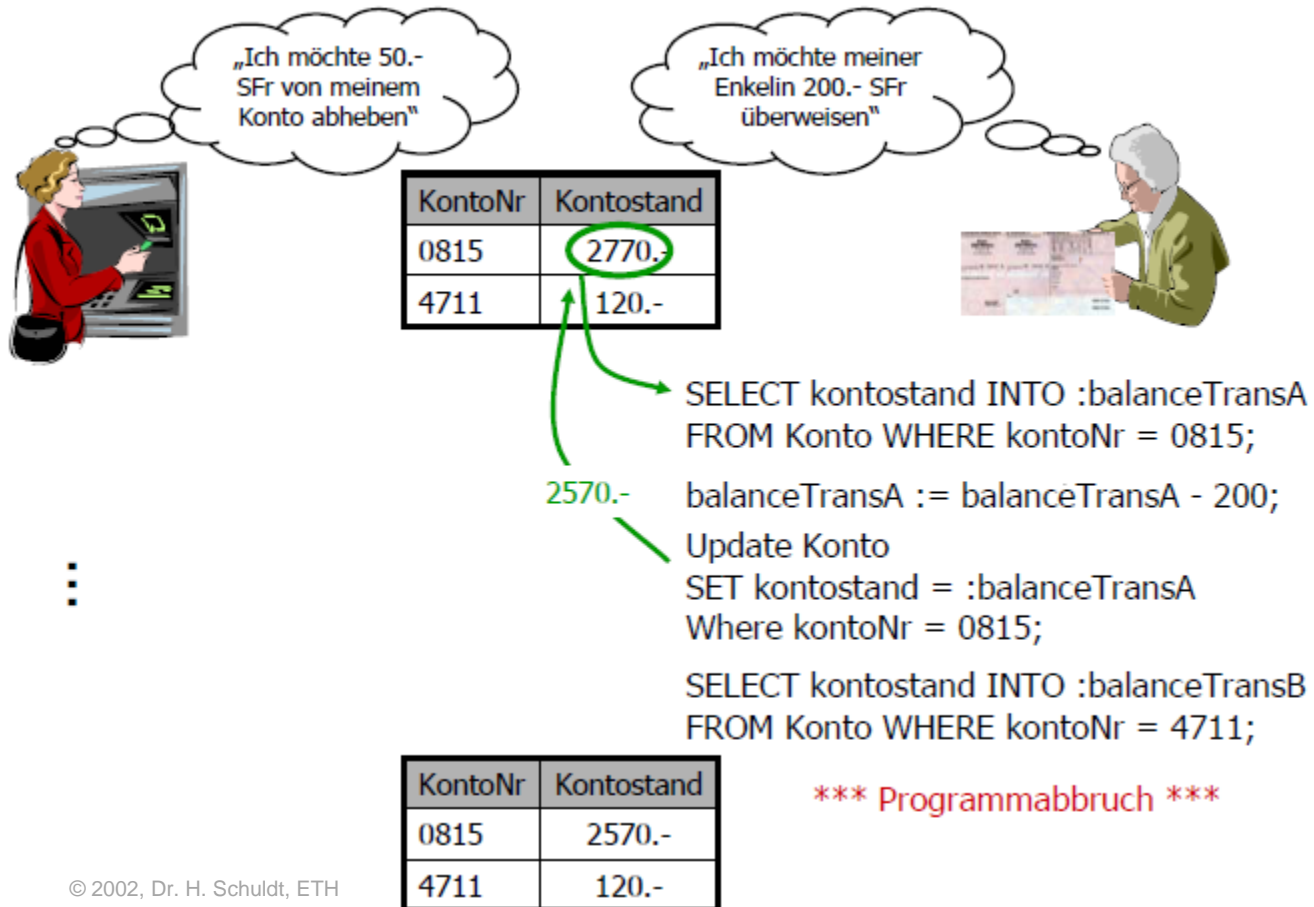
- Klassischer Vertreter von Datenbanktransaktionen.
- Kurzlebige Transaktionen, operieren auf wenigen Datenobjekten.
- Parallele Zugriffe auf Konto-Relationen:
Konto (KontoNr, KundenNr, Kontostand, ...)
- Typische Operationen:
 - Debit (KontoNr, Betrag) bucht Betrag von Konto KontoNr ab.
 - Credit (KontoNr, Betrag) bucht Betrag auf Konto KontoNr.
 - Transfer (KontoNrA, KontoNrB, Betrag), das aus einem Debit(KontoNrA, Betrag) und einem Credit(KontoNrB, Betrag) besteht.

Beispiel: Banküberweisung



© 2002, Dr. H. Schuldt, ETH

Beispiel: Banküberweisung

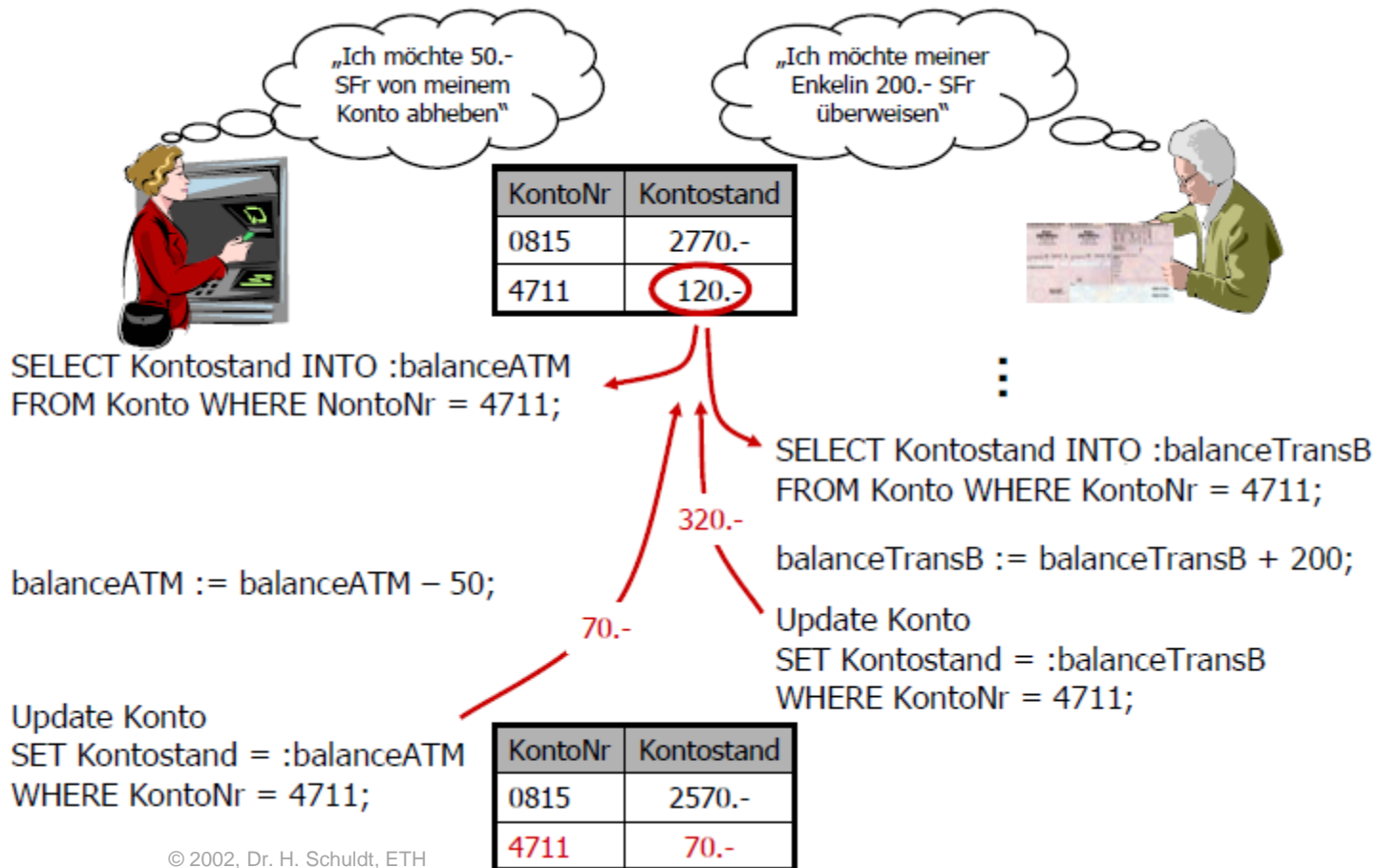


© 2002, Dr. H. Schuldt, ETH

Beispiel: Banküberweisung

- Atomarität: Im vorigen Ablauf ist die «Alles-oder-Nichts-Semantik» der Atomarität verletzt. Ein inkonsistenter Zwischenzustand bleibt erhalten (bei der fehlgeschlagenen Überweisung ging Geld verloren!)
 - Abhilfe: Für die durch den Systemabsturz abgebrochene Überweisungstransaktion sollte das bereits erfolgte Debit **automatisch** durch das Datenbanksystem **rückgängig** gemacht werden.
- Gefragt sind also geeignete Recovery-Protokolle, die beispielsweise den Zustand einspielen, der direkt vor Beginn der Überweisungstransaktion vorlag.

Beispiel: Banküberweisung



Beispiel: Banküberweisung

- Isolation: Im vorigen Ablauf überschneiden sich die beiden Transaktionen in inkorrekt Weise. Die Änderungsoperation der Überweisungstransaktion geht komplett verloren (und damit auch das überwiesene Geld).
- Der korrekte Endzustand nach kompletter Ausführung beider Transaktionen wäre:

KontoNr	Kontostand
0815	2570.-
4711	270.-

→ Das System muss daher die beiden Transaktionen durch geeignete Concurrency-Control-Protokolle voneinander **isolieren**, indem z.B. die Überweisungstransaktion den Kontostand von 4711 erst lesen und verändern kann, nachdem die ATM-Transaktion beendet ist.

Beispiel: Banküberweisung

- Weitere Anforderungen:
 - Konsistenz: Z.B. Einschränkung, dass der Kontostand niemals negativ werden kann. Dies wird in der Regel durch Integritätsbedingungen zugesichert (z.B. einer CHECK-Klausel). Auch die Konsistenz wird vom System garantiert und muss nicht vom Anwendungsentwickler berücksichtigt werden.
 - Dauerhaftigkeit: Änderungen am Kontostand durch korrekt (mit Commit) beendete Transaktionen müssen **dauerhaft** sein, also z.B. auch Systemabstürze überstehen.
 - Performanz: Natürlich muss das System mehr als zwei Benutzer parallel unterstützen, ohne dass dabei die Synchronisation konkurrierender Transaktionen zu signifikanten Performance-Einbussen führt.

Beispiele für Informationsbedürfnisse

- Gegeben sei folgende Tabelle mit Studenteninformationen:

Name	Klasse	Sprache
Andermatten	IT1B	DE
Andermatten	IT1B	FR
Bühler	IT1B	DE
Conrad	IT1A	DE
Dubach	KI1A	DE
Dubach	KI1A	EN
Erni	IT1B	DE
Erni	IT1B	FR
Funès	IT1A	FR

- Studenten der Klasse IT1B, welche Deutsch sprechen:

Name	Klasse	Sprache
Andermatten	IT1B	DE
Bühler	IT1B	DE
Erni	IT1B	DE

- Studenten, welche Französisch sprechen:

Name	Klasse	Sprache
Andermatten	IT1B	FR
Erni	IT1B	FR
Funès	IT1A	FR

Beispiele für Informationsbedürfnisse

- Versuch: Studenten, welche nicht Französisch sprechen?

Name	Klasse	Sprache
Andermatten	IT1B	DE
Bühler	IT1B	DE
Conrad	IT1A	DE
Dubach	KI1A	DE
Dubach	KI1A	EN
Erni	IT1B	DE

- Studenten tauchen auf, welche Französisch können
 - Wir haben alle Tabelleneinträge erhalten, welche aussagen, dass jemand eine andere Sprache als Französisch spricht.
 - Beispiel: „Andermatten spricht Deutsch“ (erste Zeile) ist eine Information, welche auf unsere Anfrage passt.
- Gewollt war: «Alle Studenten, für welche kein Eintrag existiert, dass sie Französisch sprechen»

Und weiter...

Das nächste Mal: Relationenmodell, Relationale Algebra

