

Geographic Information Technology Training Alliance (GITTA) presents:

Einführung in Datenbanksysteme

Verantwortliche Personen: Stephan Nebiker, Susanne Bleisch

Inhaltsverzeichnis

1. Einführung in Datenbanksysteme	2
1.1. Begriffsdefinitionen	3
1.1.1. Daten versus Informationen	3
1.1.2. Die Komponenten eines Informationssystems	4
1.2. Eigenschaften des Datenbankansatzes	6
1.2.1. Mehrfachnutzung	6
1.2.2. Strukturierte und beschriebene Daten	7
1.2.3. Trennung von Daten und Anwendungen	7
1.2.4. Datenintegrität	8
1.2.5. Transaktionen	9
1.2.6. Datenpersistenz	9
1.2.7. Datensichten	9
1.3. Anwendungsgebiete	11
1.3.1. Verwaltung von Bankkonten	11
1.3.2. Fahrplan-Informationssystem	11
1.3.3. Bibliothekskatalog	12
1.3.4. Zentrales Geodata-Warehouse	14
1.3.5. Übung	15
1.4. Vorteile und Nachteile	16
1.4.1. DBS versus dateibasiert	16
1.4.2. Vorteile DBMS	16
1.4.3. Nachteile DBMS	17
1.5. Aufgaben	18
1.6. Zusammenfassung	19
1.7. Literaturempfehlungen	20
1.8. Glossar	21
1.9. Bibliographie	23

1. Einführung in Datenbanksysteme

Datenverwaltung und insbesondere die Verwaltung von *Geodaten*¹ ist grundsätzlich nicht an eine bestimmte Technologie gebunden. Denkbar wären somit analoge Planarchive oder auch dateibasierte Ablagesysteme. Der Begriff Geoinformationssystem impliziert jedoch gewisse Anforderungen, die weit über das Speichern, Archivieren und Abrufen von Daten hinausgehen und nur mit Datenbanksystemen sinnvoll befriedigt werden können.

Der Fokus dieser Lektion und des gesamten Moduls liegt deshalb auf datenbankbasierten Konzepten und Architekturen. Nach der Einführung und Definition einiger wichtiger Begriffe in der Unit „**Begriffsdefinitionen**“ wenden wir uns den spezifischen Eigenschaften des Datenbankansatzes (Unit „**Eigenschaften des Datenbankansatzes**“) zu. Die Betrachtung verschiedener Anwendungsgebiete in der Unit „**Anwendungsgebiete**“ soll ermöglichen, verschiedene Einsatzgebiete von Datenbanken kennen zu lernen und die Datenbankeigenschaften zu vertiefen. Ein Vergleich des Datenbankansatzes mit dateibasierten Lösungsansätzen erfolgt in der Unit „**Vorteile und Nachteile**“.

Lernziele

- Sie sind mit der Begriffswelt im Umfeld Daten, Information und Datenverwaltung vertraut und können die wichtigsten Begriffe erklären.
- Sie kennen und verstehen die charakteristischen Eigenschaften von Datenbanksystemen und können diese auf Anwendungen im eigenen Umfeld übertragen.

¹ Unter Geodaten oder raumbezogenen Daten versteht man Datenobjekte, die durch eine Position im Raum direkt oder indirekt referenzierbar sind. Der Raum ist dabei definiert durch ein Koordinatensystem, das den Bezug zur Erdoberfläche herstellt. Geodaten werden in der Regel grafisch in Papierform oder an grafikfähigen Bildschirmen präsentiert. Aus informationstechnischer Sicht kann man die Daten, die wiederum zu Geodaten gehören, einteilen in: · Geometrie (Position und geometrische Ausprägung) · Topologie (explizit gespeicherte geometrisch-topologische Beziehungen) · Präsentation (graphische Ausprägungen wie Signaturen, Farbe, Typographie) · Sachdaten (alphanumerische Daten zur Beschreibung der Semantik). Geodaten stellen in der Informationsverarbeitung eine besondere Herausforderung dar. Die Gründe dafür sind: · der hohe Aufwand für die Erfassung · die große Menge der anfallenden Daten · die geforderten Antwortzeiten beim Zugriff auf Geodaten · die Verarbeitung nach räumlichen Kriterien sowie · die Komplexität der Beziehungen.

1.1. Begriffsdefinitionen

Von Daten zu nutzbaren Informationen

Bevor wir uns mit dem Einsatz und der Architektur von Datenverwaltungslösungen beschäftigen, wollen wir uns mit Begriffen wie Information, Daten und Datenbanksystemen auseinandersetzen – dem Basisvokabular für unser Modul. Viele dieser Begriffe werden zwar täglich verwendet, oft aber in einem falschen Zusammenhang.

1.1.1. Daten versus Informationen

Daten (Computerdaten):

Die Darstellung von Tatsachen (zum Beispiel Messwerte) oder Konzepten, die in einer durch den Computer lesbaren Form erzeugt oder in eine entsprechende Form gebracht werden.

Information:

Information ist nutzbare Antwort auf eine konkrete Fragestellung (ZEHNDER 1998). Information wird meistens aus Daten extrahiert oder abgeleitet.

Von Information wird gesprochen, wenn auf eine spezifische Frage eine Antwort gegeben wird, die das Verständnisniveau der Fragenden erhöht und sie befähigt, einem bestimmten Ziel näher zu kommen. (ZEHNDER 1998)

Information hat die folgenden Aspekte:

- *strukturelle*⁴ und *syntaktische*⁵
- *semantische*⁶ (inhaltliche)
- *pragmatische* (anwendungsrelevante)

Das Verhältnis von Daten und Information

Die Begriffe Daten und Information werden sehr oft vermischt und im falschen Zusammenhang eingesetzt. Daher sollen nachfolgend ein paar Unterscheidungsmerkmale aufgeführt werden obwohl es nicht möglich ist strikt zwischen Daten und Information zu trennen:

- Semantische Aspekte (die Bedeutung) werden in Daten oft codiert, wobei diese Codes nach vorgängig vereinbarten Konventionen definiert und interpretiert werden müssen (z. B. Notenskala von 1 bis 6 mit der Konvention 6 = sehr gut).
- Information muss in der Regel zuerst aus Daten rekonstruiert oder abgeleitet werden. Dies geschieht in Bezug auf eine bestimmte Fragestellung. Ein Beispiel: Die Daten enthalten Angaben zu den Niederschlagsmengen in jedem Monat während der letzten zehn Jahre. Jemand möchte wissen, wieviel

⁴ Die Struktur ist der abstrakte innere Aufbau.

⁵ Syntax ist id Lehre vom Satz, Lehre von den Regeln, wonach in Sprache(n) aus den Wörtern zusammengehörige Wortgruppen gebildet werden, und die Lehre von der Kombination von Wörtern zu Sätzen. „Syntaxis“ heisst auf Griechisch Zusammenstellung, Anordnung, Aneinanderreihung.

⁶ Die Informationstheorie unterscheidet drei Dimensionen einer Information, die syntaktische, die semantische und die pragmatische. Nehmen wir als Beispiel eine Verkehrsampel. In der syntaktischen Dimension unterscheiden wir die drei Farben Rot, Gelb und Grün. Ihren Sinn bekommt die Ampel aber erst in der semantischen Dimension. Hier wird den Farben ihre Bedeutung zugeordnet. Rot hat die Bedeutung Halt, grün bedeutet freie Fahrt. Erst in der pragmatischen Dimension aber wird das Ampelsignal für die Verkehrsregelung brauchbar. Pragmatisch gesehen bedeutet das rote Licht, als Autofahrer muss ich anhalten.

es im Durchschnitt während dem Monat Juli geregnet hat (über die letzten zehn Jahre). Somit ist die durchschnittliche Niederschlagsmenge im Monat Juli der letzten zehn Jahre Information die die konkrete Fragestellung beantwortet. Aber die berechnete Durchschnittszahl ohne entsprechende Frage ist den Daten zuzuordnen.

- Daten enthalten in der Regel keine anwendungsrelevanten Aspekte (z. B. lassen sich aus den Koordinaten eines Grundstücks keine Informationen für unterschiedliche Anwendungen wie Steuern, Erschliessung, Überflutungsrisiken etc. ableiten).

我 方 去
讲真相劝善世
写下的标

2002年6月13日 星期

觉醒后的肺腑之

亲爱的先生，

四年多以前，当您来中国访问的时候，中华人民共和国广播
事业局和马里国家

电台签订了第一个广播合作协定，为中马两国在广播
方面进行友好合作创造了良好的条件，这对于增进两国
间的相互了解和发展两国人民之间的友谊是有益

谢谢你的注意。 问候！

Bei einem Brief in fremder Sprache erkennen wir zwar die strukturellen und syntaktischen Aspekte, d. h. die Abschnitte, Sätze und Wörter, nicht aber den Sinn des Geschriebenen.

Auch ein Schreiben in der eigenen Sprache kann nicht in jedem Fall als Information bezeichnet werden. Wir mögen zwar dessen Inhalt (*Semantik*) verstehen, wenn dieser für uns aber irrelevant oder uninteressant ist, dann fehlt der wichtige Aspekt des Nutzens.

Ein Brief in chinesischer Sprache

1.1.2. Die Komponenten eines Informationssystems

Aus *konzeptioneller*⁷ Sicht hat ein Informationssystem einen schalenförmigen Aufbau.

Fahren Sie mit der Maus über die Begriffe in der nachstehenden Interaktion und erkennen Sie, aus welchen Teilen ein Informationssystem aufgebaut ist.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [link]

Die Datenverwaltungs-Komponenten eines Informationssystems sind:

⁷ Nach Duden: die einem Werk zugrunde liegende Anschauung, Leitidee; für das Datenmodell gilt: unabhängig von einer bestimmten Technologie

Datenbasis und Datenbank:

Eine Datenbasis ist eine Menge von Daten, die aus Sicht der Systembetreiber in irgendeiner Weise als zusammengehörig betrachtet werden. Angereichert um weitere Daten, die das DBMS zur Erfüllung seiner Aufgaben benötigt, bilden sie eine Datenbank (DB).

Datenbankverwaltungssystem:

Datenbankverwaltungssysteme (engl. database management system oder DBMS) sind Softwareprodukte für die dauerhafte, integre und anwendungsunabhängige Speicherung und Verwaltung sowie die flexible und bequeme Verwendung von grossen, mehrfach benutzbaren integrierten Datenbasen.

Datenbanksystem:

Ein Datenbanksystem (DBS) besteht aus einem DBMS und einer oder mehreren Datenbanken.



Datenbankverwaltungssysteme und Datenbanksysteme stehen im Zentrum dieses Moduls.

Informationssystem:

Ein Informationssystem erweitert die Datenbank um eine Reihe von Werkzeugen (engl. software tools) zur Abfrage, Darstellung, Transformation und Analyse von Daten.



Gemäss der im ersten Teil der Unit behandelten **Unterscheidung zwischen Daten und Informationen**, bereichern die Werkzeuge eines Informationssystems die Daten um *semantische* und *pragmatische* bzw. anwendungsrelevante Aspekte.

Sicherlich haben Sie auch schon den Begriff Geoinformationssystem gehört und möglicherweise auch schon Definitionen dazu gelesen. Der nächste Abschnitt definiert den Begriff Geoinformationssystem nochmals zur Erinnerung und vergleicht diese Definition mit der des Informationssystems.

Geoinformationssystem:

"Ein Geoinformationssystem dient der Erfassung, Speicherung, Analyse und Darstellung aller Daten, die einen Teil der Erdoberfläche und die darauf befindlichen technischen und administrativen Einrichtungen sowie geowissenschaftliche, ökonomische und ökologische Gegebenheiten beschreiben." (BARTELME 1989)

Diese Definition enthält die wichtigsten Aspekte der obigen Definition eines Informationssystems, jedoch mit einer Fokussierung auf Daten mit einem Raumbezug.

1.2.2. Strukturierte und beschriebene Daten

Eine grundsätzliche Eigenschaft des Datenbankansatzes ist, dass ein Datenbanksystem nicht nur die Daten enthält, sondern auch eine komplette Definition oder Beschreibung der Daten. Diese Beschreibung besteht aus Angaben über den Umfang, die Struktur, die Art und das Format aller Daten sowie über die Beziehungen der Daten untereinander. Diese gespeicherten Informationen werden auch Metadaten („Daten-über-Daten“) genannt.

*Metadaten*¹⁴ werden von der DBMS-Software, aber auch von Anwendungsprogrammen (z. B. GIS) und Anwendern einer Datenbank verwendet. Da DBMS-Software nicht für eine spezielle Datenbankanwendung geschrieben ist, muss sie die Metadaten einer Datenbank verwenden, um deren Umfang, Struktur etc. zu kennen.

Hier nun ein einfaches Beispiel dafür, wie in einer Datenbank Daten beschrieben sein können.

Nachstehend ist eine Datenbanktabelle aufgeführt. Durch die Struktur dieser Tabelle (1. Spalte = Vorname, 2. Spalte = Name, 3. Spalte = PLZ, 4. Spalte = Wohnort) weiss man, dass ein Eintrag in der ersten Spalte ein Vorname als Zeichenkette (String) und ein Eintrag in der dritten Spalte eine Postleitzahl (Nummer) sein muss.

Vorname [string]	Name [string]	PLZ [number]	Wohnort [string]
Klaus	Meier	5369	Bühlwil
Sabine	Klett	4812	Moordorf
...

Beispiel einer Datenbanktabelle

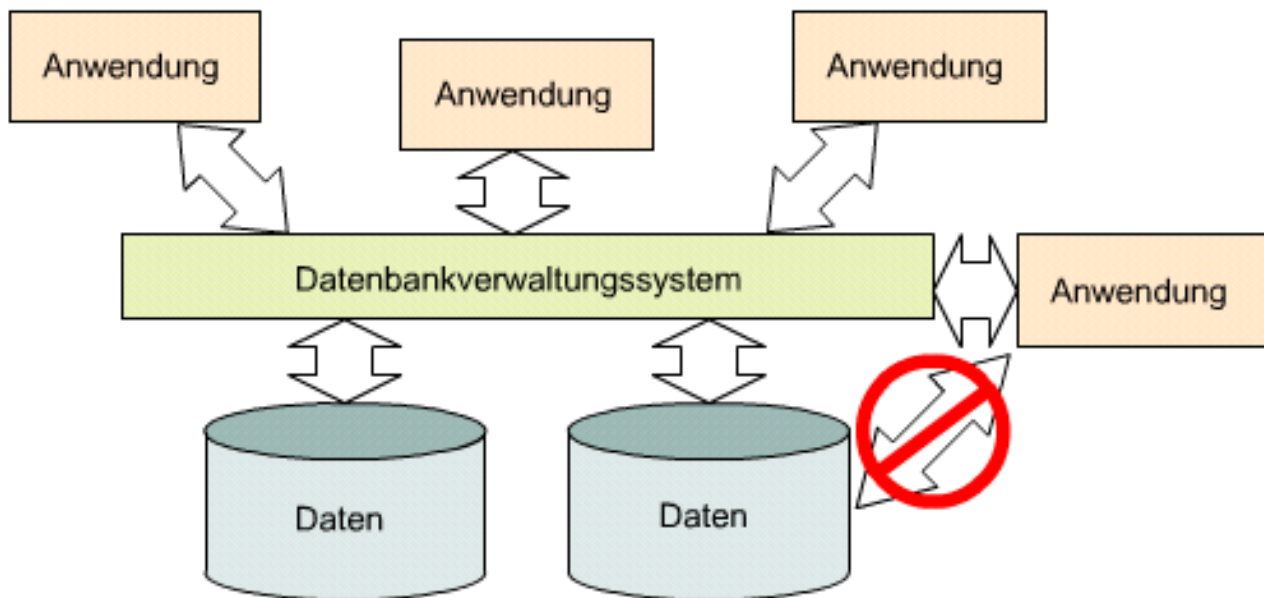
1.2.3. Trennung von Daten und Anwendungen

Wie in der Eigenschaft „**Strukturierte Daten**“ beschrieben, wird die Struktur einer Datenbank durch die *Metadaten*, die ebenfalls in der Datenbank abgelegt sind, beschrieben.

Das Anwendungsprogramm benötigt keine Kenntnis über die physikalische Datenspeicherung (Codierung, Format, Speicherort etc.). Es kommuniziert mit dem Verwaltungssystem einer Datenbank (DBMS) über eine normierte Schnittstelle mittels einer standardisierten Sprache (z. B. SQL). Den Zugriff auf die eigentlichen Daten und die *Metadaten* übernimmt dabei das DBMS.

Auf diese Weise können die Anwendungen völlig von den Daten getrennt werden, und datenbank-interne Effizienzverbesserungen und Reorganisationen haben keinen Einfluss auf die Anwenderprogramme.

¹⁴ Meta ist ein Präfix, das in den meisten informationstechnologischen Anwendungen „eine zugrundeliegende Definition oder Beschreibung“ meint. Metadaten sind also eine Definition oder Beschreibung von Daten. Man spricht auch von „Daten über Daten“.



Trennung der Anwendungen von den Daten

1.2.4. Datenintegrität

Datenintegrität ist ein Begriff für die Qualität und Zuverlässigkeit von Daten eines Datenbanksystems. Im weiteren Sinne zählt zur Integrität auch der Schutz der Datenbank vor unberechtigt Zugriff (Vertraulichkeit) und Veränderungen.

Daten widerspiegeln Sachverhalte der realen Welt. Logischerweise wird verlangt, dass sie dies korrekt tun. Ein DBMS soll Unterstützung bieten bei der Aufgabe, nur korrekte und widerspruchsfreie („konsistente“) Daten in die Datenbank gelangen zu lassen. Ausserdem wird mit korrekten **Transaktionen** die Konsistenz auch während des Systembetriebs aufrechterhalten.

In derselben Datenbank dürfen keine widersprüchlichen Aussagen stehen.

Wohnstrassen

Strassenname	Fahrverbot
Lindenstrasse	Nein
Rosenstrasse	Ja
Gartenstrasse	19.00–7.00
...	

Strassenplanung

Strassenname	Art	Einschränkungen
Hauptstrasse	Kantonsstrasse	keine
Gartenstrasse	Quartierstrasse	keine
Jägerweg	Quartierweg	Fussweg
...		

Zwei Datenbanktabellen mit widersprüchlichen Datensätzen

Diese Datenbank besteht aus den zwei Tabellen „Wohnstrassen“ und „Strassenplanung“. In der Tabelle „Wohnstrassen“ ist die Gartenstrasse von 19.00–7.00 mit einem Fahrverbot belegt. In der Tabelle „Strassenplanung“ jedoch bestehen für die Gartenstrasse keine Einschränkungen. Dies ist ein Beispiel für einen Widerspruch, wie er in einer Datenbank nicht vorkommen darf.

Solche Widersprüche können verhindert werden, indem die Datenstruktur sorgfältig modelliert wird. In diesem Fall müsste man schauen, wie man die Informationen über die Gartenstrasse in einer Tabelle zusammenfassen könnte. Dies erleichtert auch die Nachführung, da dann nur noch eine Tabelle geändert werden muss, wenn sich an den Informationen zur Gartenstrasse etwas ändert.

1.2.5. Transaktionen

Eine Transaktion ist ein Bündel von Aktionen, die in der Datenbank durchgeführt werden, um diese von einem konsistenten Zustand wieder in einen konsistenten (widerspruchsfreien) Zustand zu überführen. Dazwischen sind die Daten zum Teil zwangsläufig inkonsistent.

Eine Transaktion ist atomar, d. h. nicht weiter zerlegbar. Innerhalb einer Transaktion werden entweder alle Aktionen oder keine durchgeführt. Nur ein Teil der Aktionen würde zu einem inkonsistenten Datenbankzustand führen.

Ein Beispiel einer Transaktion ist das Verschieben einer bestimmten Summe Geld von einem Konto auf ein anderes. Die Abbuchung des Geldes von einem Konto und die Gutschrift auf dem anderen Konto machen zusammen eine konsistente Transaktion aus. Diese Transaktion ist ausserdem atomar. Die Abbuchung oder die Gutschrift alleine würde zu einem inkonsistenten Zustand führen. Nach Abschluss der Transaktion (Abbuchung und Gutschrift) wird die Änderung an beiden Konti dauerhaft, und der Geldgeber sieht nun einen kleineren Kontostand, während der Empfänger des Geldes sich über seinen höheren Kontostand freuen kann.

Probieren Sie es aus, indem Sie mit den Buttons unten links durch die einzelnen Schritte dieses Beispiels navigieren.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

1.2.6. Datenpersistenz

Datenpersistenz meint, dass in einem DBMS einzelne Daten solange aufbewahrt werden müssen, bis sie explizit gelöscht werden. Die Lebensdauer von Daten muss also von den Benutzern direkt oder indirekt bestimmbar sein und darf nicht von irgendwelchen Systemgegebenheiten abhängen. Ebenso wenig dürfen einmal in die Datenbank aufgenommene Daten verloren gehen.

Änderungen, die eine **Transaktion** in einer Datenbank vornimmt, sind dauerhaft. Wenn die Transaktion abgeschlossen ist, kann auch ein darauf folgender Systemabsturz die Daten nicht mehr gefährden.

1.2.7. Datensichten

Typischerweise hat eine Datenbank mehrere Benutzer, und jeder benötigt, je nach Zugangsrechten und Bedürfnissen, eine individuelle Ansicht der Daten (Inhalt und Form). Eine solche Datensicht kann aus einer Teilmenge der gespeicherten Daten oder aus daraus abgeleiteten (nicht explizit gespeicherten) Daten bestehen. Beispiel: An einer Hochschule werden Daten über die Studierenden verwaltet. Neben Matrikelnummer, Name, Adresse etc. werden auch andere Informationen erfasst, wie die Semesterzugehörigkeit oder die Angabe darüber, ob der oder die Studierende das Semester repetiert.

Diese umfassende Datenbank wird von Personen mit unterschiedlichen Bedürfnissen und Berechtigungen benützt.

Klicken Sie auf einen der vier Buttons, um die verschiedenen Sichten auf die Datenbank dargestellt zu bekommen.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Der Datenbankadministrator hat eine Sicht auf die gesamte Datenbank, während andere Nutzer in diesem Beispiel jeweils eine sehr eingeschränkte Sicht auf die Datenbank haben. Die Administration benötigt für die Erstellung einer Statistik über die Repetenten in den Abteilungen keine Namen oder Matrikelnummern. In einer Liste der Studierenden sollten heikle Daten wie der Repetentenstatus keinesfalls sichtbar sein. Der Dozent einer Klasse jedoch wünscht sehr detaillierte Auskunft über die Studierenden dieser Klasse.

1.3. Anwendungsgebiete

Aktuelle Informatiklösungen zeichnen sich aus durch viele verteilte Benutzer, die laufend aktualisierte Datenbestände gemeinsam nutzen möchten. Dafür werden in den verschiedensten Bereichen Datenbanksysteme eingesetzt, und es gibt kaum eine IT-Anwendung, die noch ohne sie auskommt.

Nachstehend sind einige Beispiele für den Einsatz von Datenbanksystemen aufgeführt.

Es wird jeweils kurz die Anwendung beschrieben und danach auf für diese Anwendung speziell wichtige Datenbankeigenschaften eingegangen. Die jeweiligen Eigenschaften sind mit den Beschreibungen derselben in der Unit „**Eigenschaften des Datenbankansatzes**“ verknüpft.

1.3.1. Verwaltung von Bankkonten

Die Verwaltung von Bankkonten ist eine anspruchsvolle Aufgabe, die schon längere Zeit Datenbanksysteme als Hilfsmittel verwendet. Heutzutage könnte sich vermutlich niemand mehr vorstellen, in der komplexen Finanzwelt zu operieren, ohne von Datenbanksystemen unterstützt zu werden.

Speziell wichtige Eigenschaften dieses Datenbanksystems:

Transaktionen	Der erfolgreiche und korrekte Ablauf von Transaktionen ist ein wichtiger Punkt in der Bankkontenverwaltung. Man stelle sich nur vor, dass Gutschriften nicht richtig verbucht oder Abzüge mehrfach getätigt würden.
Datenintegrität	Dabei spielt natürlich auch die Datenintegrität eine grosse Rolle. Es muss klar festgelegt sein, wie die Konsistenzbedingungen und -regeln aussehen und wie diese eingehalten werden können.
Datenpersistenz	Als Bankkontenbesitzer ist es beruhigend zu wissen, dass die Persistenz der Daten gewährleistet ist. Daten werden nicht willkürlich gelöscht und gehen nicht auf mysteriöse Weise verloren.

1.3.2. Fahrplan-Informationssystem

Der Online-Fahrplan der SBB (Schweizerische Bundesbahnen) ist ein Beispiel eines Datenbanksystems aus dem Bereich des öffentlichen Verkehrs und eines darauf aufbauenden webbasierten Informationssystems. Seine Aufgabe ist es, die Benutzer jederzeit mit aktuellen und korrekten Informationen über optimale Verbindungen und den Zugbetrieb der SBB zu versorgen.

Speziell wichtige Eigenschaften dieses Datenbanksystems:

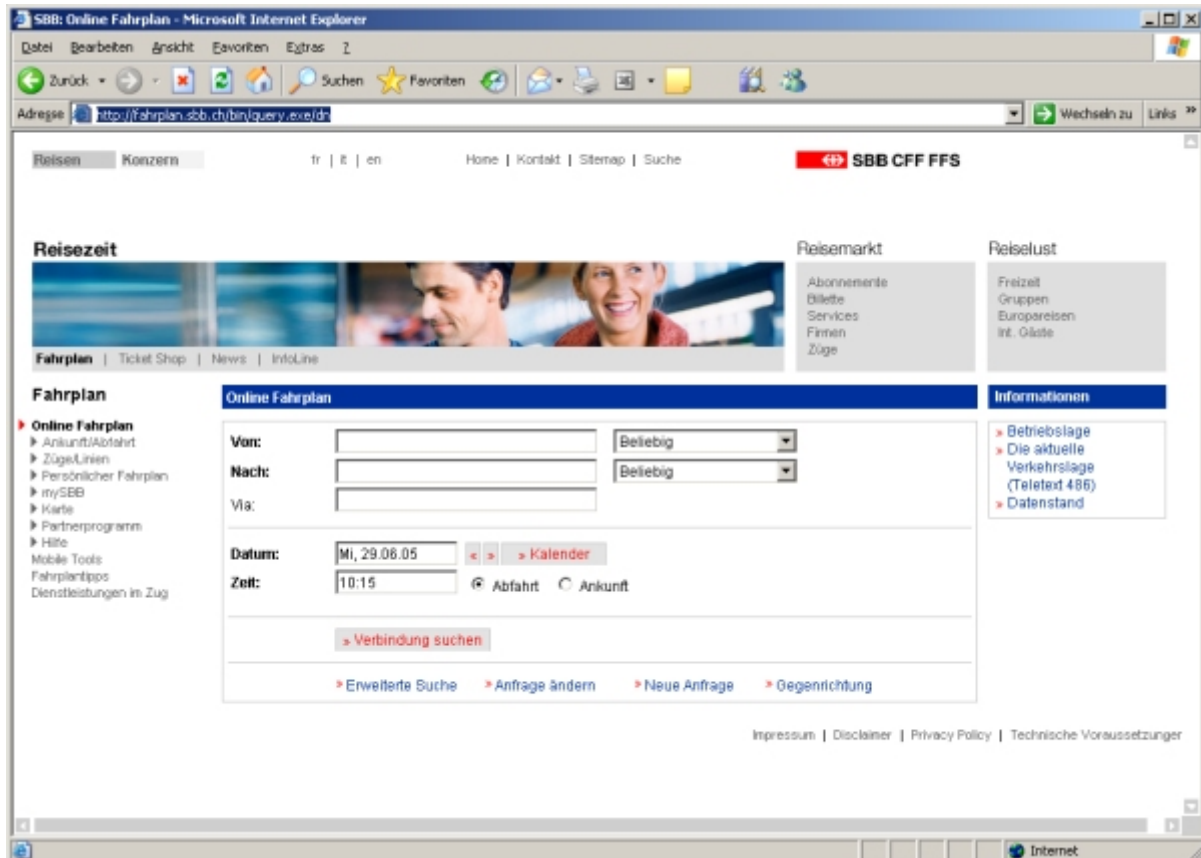
Mehrfachnutzung	Das Datenbanksystem des SBB-Fahrplans kann von verschiedenen Anwendungen und Benutzern verwendet werden. Die Fahrplandaten werden SBB-intern verwendet, aber auch die Öffentlichkeit kann via Internet oder WAP ¹⁵ auf die Zugfahrpläne und weitere Auskünfte zugreifen.
------------------------	---

¹⁵ Das Wireless Application Protocol – kurz: WAP – ist der wichtigste Standard, um Internetkommunikation und interaktive Dienste für Handys und andere mobile Endgeräte zu realisieren: Nachrichten aus dem Internet abrufen, Börsenkurse abfragen, Flugtickets buchen.

Datenintegrität

Die Benutzer des SBB-Fahrplans möchten immer eine aktuelle und korrekte Auskunft erhalten, dies stellt hohe Anforderungen an die Datenintegrität. Dafür müssen bei Fahrplan-, Gleis- oder anderen Änderungen die Daten des Datenbanksystems laufend aktualisiert werden.

Für ein solches Datenbanksystem, das in Spitzenzeiten deutlich mehr als eine Anfrage pro Sekunde erhält, ist neben den erwähnten Punkten auch die Performance (Leistung) eines Datenbanksystems eine speziell geforderte Eigenschaft.



Screenshot SBB-Fahrplan-Informationssystem (SBB AG 2005)

Internet-Zugriff auf den SBB-Fahrplan: <http://fahrplan.sbb.ch/>

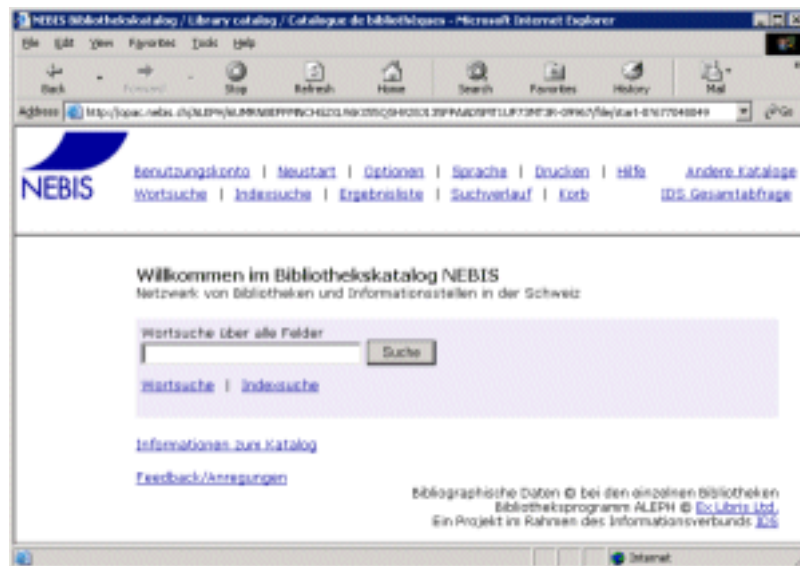
1.3.3. Bibliothekskatalog

Der Bibliothekskatalog NEBIS (Lierz et al. 2004) ist der Katalog eines Netzwerks von Bibliotheken und Informationsstellen in der Schweiz. Auf diese Weise kann schweizweit in Bibliotheken nach bestimmten Büchern und Veröffentlichungen gesucht werden.

Speziell wichtige Eigenschaften dieses Datenbanksystems:

Strukturierte und beschriebene Daten

Es ist mehr als hilfreich, wenn man sich bei der Erfassung und Nachführung von Tausenden von Büchern und Zeitschriften an eine klare und festgelegte Struktur halten kann. Die Beschreibung der Daten ermöglicht ausserdem eine gezielte Suche nach einzelnen Objekten.



Screenshot Bibliothekskatalog (Lierz et al. 2004)

Internet-Zugriff auf den Nebis-Bibliothekskatalog: <http://www.nebis.ch/>

1.3.4. Zentrales Geodata-Warehouse

In städtischen und kantonalen Verwaltungen werden immer häufiger die raumbezogenen Daten aller Amtsstellen in einem zentralen Datenpool, dem sogenannten Geodata-Warehouse (einem raumbezogenen Datenbankverwaltungssystem), verwaltet und nachgeführt. Dies bringt enorme Einsparungen mit sich, da die Daten nun redundanzfrei sind und nur einmal nachgeführt werden müssen. Im bisherigen Betrieb wurden Daten häufig in den verschiedenen Ämtern mehrfach geführt, was deren Aktualisierung enorm erschwerte. Ausserdem müssen bei ämterübergreifenden Projekten die Geodaten nicht mehr mühsam zusammengesucht werden, sondern können einfach aus dem Datenpool geholt werden.

Speziell wichtige Eigenschaften dieses Datenbanksystems:

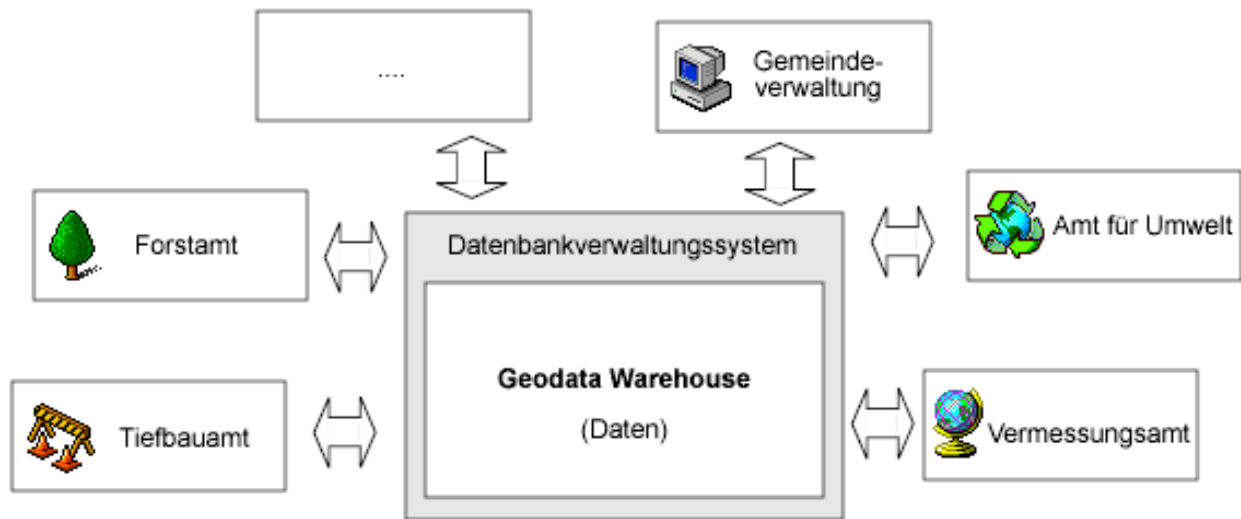
- | | |
|---|--|
| Mehrfachnutzung | Ein zentrales Geodata-Warehouse ist in mehrfacher Hinsicht ein schönes Beispiel für die Mehrfachnutzung. Einerseits greifen mehrere Anwender darauf zu – die diversen Angestellten der verschiedenen Ämter. Andererseits werden für diese Zugriffe auch unterschiedliche Anwendungsprogramme (z. B. GIS-Systeme) verwendet. So kann es sein, dass das Forstamt mit der GIS-Software A auf die Daten zugreift, während das Vermessungsamt die GIS-Software B verwendet. |
| Trennung von Daten und Anwendungen | Wie oben beschrieben, greifen verschiedene Benutzer mit unterschiedlichen Anwendungen auf die Daten zu. Dies ist nur möglich, wenn die Daten von den Anwendungen getrennt sind. Wenn nämlich die Daten mit einer Anwendung verbunden wären, gäbe es einen grossen Aufwand, diese Daten so aufzubereiten, dass auch andere Anwendungen sie lesen und verwenden könnten. Diese Unabhängigkeit wird besonders wichtig, wenn die DBMS-Software erneuert werden soll, ohne Dutzende von Anwenderprogrammen anpassen zu müssen. Beispiel: Ein Word-File ist sehr schwierig im Excel-Programm zu öffnen, und dies, obwohl beide Programme vom selben Hersteller kommen. |

Datenpersistenz

Die Erfassung von *Geodaten* und anderen Daten ist meistens mit einem grossen Aufwand und hohen Kosten verbunden. Deshalb ist die Datenpersistenz ein speziell geforderte Eigenschaft eines Geodata Warehouse. Damit wird sichergestellt, dass Daten nicht einfach verloren gehen und danach aufwendig und teuer wiederbeschafft werden müssen.

Datenintegrität

Da die Daten eines Amtes häufig auch Auskunft über rechtliche Verhältnisse geben müssen, wie zum Beispiel die Grundbuchplandaten der amtlichen Vermessung, wird von ihnen gefordert, dass sie absolut korrekt und zuverlässig, also integer, sind. Dies wird durch die Definition und Einhaltung von ganz bestimmten Konsistenzbedingungen und -regeln erreicht.



Schematische Darstellung eines Geodata-Warehouse und der möglichen Zugriffe aus verschiedenen Ämtern

1.3.5. Übung

Suchen Sie selber eine weitere Datenbank-Anwendung (in Ihrem Umfeld, im Internet, ...) und versuchen Sie abzuschätzen, welches speziell wichtige / speziell geforderte Eigenschaften für diese Anwendung sind.

Verfassen Sie eine kurze Zusammenstellung der Informationen (im gleichen Stil wie die Anwendungen in dieser Unit) und veröffentlichen Sie diese im Diskussionsforum unter dem Thema „Datenbank-Anwendungen“. Schauen Sie sich unbedingt auch weitere Zusammenstellungen ihrer Mitstudierenden an.

Ihre Zusammenstellung wird von einem Tutor angeschaut, und er wird ein kurzes Feedback dazu geben (ebenfalls im Forum).

1.4. Vorteile und Nachteile

Vom Karteikasten zur Datenbank

Die Art der Datenspeicherung und Verwaltung hat im Verlauf der Jahre eine grosse Entwicklung durchgemacht – vom manuellen Karteikartensystem über die erste (dateibasierte) Computerisierung desselben bis hin zu modernen Datenbanksystemen. Der erste Teil dieser Unit befasst sich mit dateibasierten Systemen im Vergleich zu Datenbanksystemen. Danach werden die Vor- und Nachteile von Datenbanksystemen aufgeführt.

1.4.1. DBS versus dateibasiert

Mit den Eigenschaften eines Datenbanksystems (Unit **Eigenschaften des Datenbankansatzes** und Unit **Anwendungsgebiete**) im Hinterkopf wenden wir uns den dateibasierten Systemen zu. Beim dateibasierten Ansatz definiert und kreiert jeder Anwender mit einem spezifischen Programm jeweils die Dateien, die er für eine spezielle Anwendung benötigt. Im Vergleich zum Datenbanksystemansatz ergeben sich daraus einige Einschränkungen.

Fahren Sie mit der Maus über die fetten Schlagwörter auf der linken Seite und sehen Sie die Einschränkungen gegenüber dem Datenbankansatz erklärt in Wort (rechts) und Bild (unten).

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Bei einem dateibasierten Ansatz kann es geschehen, dass die Administration einer Schule Daten pflegt (z. B. basierend auf einem Tabellenkalkulationssystem wie zum Beispiel Microsoft Excel), die Auskunft über die Registrierung und Gebührenbezahlung der Studenten geben. Währenddessen verwalten die Dozenten Daten über die Studierenden und ihre Noten. Obwohl beide Anwender an Studierendendaten interessiert sind, haben sie doch unterschiedliche Dateien und unterschiedliche Programme, um diese zu verwalten und zu ändern. Diese *Redundanz* im Definieren und Speichern von Daten verschwendet Speicherplatz. Ausserdem ist ein mehrfacher Aufwand nötig, um die Daten zu aktualisieren. Änderungen an den Studierendendaten müssen von der Administration und von den Dozenten unabhängig voneinander vorgenommen werden.

1.4.2. Vorteile DBMS

Grundsätzlich können hier alle in der Unit **Eigenschaften des Datenbankansatzes** erwähnten Punkte noch einmal als Vorteile aufgeführt werden.

- **Mehrfachnutzung**
- **Strukturierte und beschriebene Daten**
- **Trennung von Daten und Anwendungen**
- **Datenintegrität**
- **Transaktionen**
- **Datenpersistenz**
- **Datensichten**

Dazu kommen noch weitere, bis jetzt nicht explizit erwähnte Vorteile.

Verwenden Sie für die Navigation die blauen Buttons.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Klicken Sie auf den nachstehenden Link für eine Liste der weiteren Vorteile. ([PDF-File der weiteren Vorteile](#))

1.4.3. Nachteile DBMS

Neben den zahlreichen Vorteilen eines Datenbanksystems dürfen aber auch die Nachteile nicht unerwähnt bleiben.

Verwenden Sie für die Navigation die blauen Buttons.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Klicken Sie auf den nachstehenden Link für eine Liste der Nachteile. ([PDF-File der Nachteile](#))

1.5. Aufgaben

Die nachstehenden Aufgaben sollen Ihnen ermöglichen, zu testen, ob Sie die Inhalte dieser Lektion verstanden haben.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

Ziehen Sie in der nachstehenden Übung die Nummern neben den Begriffen auf die Kreise in der Grafik. Arbeiten Sie möglichst exakt, und wenn Sie alle Nummern platziert haben, können Sie mit „Check“ testen, ob alles richtig ist. Der Reset-Button setzt alle Nummern in die Ursprungsposition zurück.

Achtung: Der Check-Button funktioniert nur korrekt, wenn alle Kreise mit einer Nummer bedeckt sind.

Dieses Element (Animation, Video etc.) kann nicht dargestellt werden und ist nur in der Onlineversion sichtbar. [\[link\]](#)

1.6. Zusammenfassung

In dieser Lektion wurden die Bedeutung einer datenbankgestützten Datenverwaltung erläutert. Dazu wurden eingangs Begriffe wie Datenbanksystem und Geoinformationssystem definiert bzw. repetiert und im Kontext der Begriffe Daten und Informationen positioniert.

Anschliessend wurden wichtige Gründe für den Einsatz von Datenbanksystemen bei der Verwaltung von Daten im Allgemeinen und von Geodaten im Speziellen aufgeführt und anhand von Beispielen aus verschiedenen Anwendungsbereichen illustriert. Zu den massgebenden Eigenschaften des Datenbankansatzes gehören die Mehrfachnutzung sowie die Strukturierung und Beschreibung der Daten. Weiter wurden die Vorteile der Trennung von Daten und Anwendungen und die verschiedenen Merkmale zur Gewährleistung einer hohen Zuverlässigkeit und Datensicherheit behandelt, namentlich die Konzepte der Transaktionen und der Datensichten.

1.7. Literaturempfehlungen

- **ZEHNDER, C.A.**, 1998. *Informationssysteme und Datenbanken*. Zürich: vdf-Hochschulverlag AG.
Einführung ins Thema Informationssysteme und Datenbanken auf Deutsch

1.8. Glossar

Daten (Computerdaten):

Die Darstellung von Tatsachen (zum Beispiel Messwerte) oder Konzepten, die in einer durch den Computer lesbaren Form erzeugt oder in eine entsprechende Form gebracht werden.

Datenbanksystem:

Ein Datenbanksystem (DBS) besteht aus einem DBMS und einer oder mehreren Datenbanken.

Datenbankverwaltungssystem:

Datenbankverwaltungssysteme (engl. database management system oder DBMS) sind Softwareprodukte für die dauerhafte, integre und anwendungsunabhängige Speicherung und Verwaltung sowie die flexible und bequeme Verwendung von grossen, mehrfach benutzbaren integrierten Datenbasen.

Datenbasis und Datenbank:

Eine Datenbasis ist eine Menge von Daten, die aus Sicht der Systembetreiber in irgendeiner Weise als zusammengehörig betrachtet werden. Angereichert um weitere Daten, die das DBMS zur Erfüllung seiner Aufgaben benötigt, bilden sie eine Datenbank (DB).

Dimensions of Information:

Die Informationstheorie unterscheidet drei Dimensionen einer Information, die syntaktische, die semantische und die pragmatische. Nehmen wir als Beispiel eine Verkehrsampel. In der syntaktischen Dimension unterscheiden wir die drei Farben Rot, Gelb und Grün. Ihren Sinn bekommt die Ampel aber erst in der semantischen Dimension. Hier wird den Farben ihre Bedeutung zugeordnet. Rot hat die Bedeutung Halt, grün bedeutet freie Fahrt. Erst in der pragmatischen Dimension aber wird das Ampelsignal für die Verkehrsregelung brauchbar. Pragmatisch gesehen bedeutet das rote Licht, als Autofahrer muss ich anhalten. (KOCH)

Geodaten:

Unter Geodaten oder raumbezogenen Daten versteht man Datenobjekte, die durch eine Position im Raum direkt oder indirekt referenzierbar sind. Der Raum ist dabei definiert durch ein Koordinatensystem, das den Bezug zur Erdoberfläche herstellt. Geodaten werden in der Regel grafisch in Papierform oder an grafikfähigen Bildschirmen präsentiert. Aus informationstechnischer Sicht kann man die Daten, die wiederum zu Geodaten gehören, einteilen in:

- Geometrie (Position und geometrische Ausprägung)
- Topologie (explizit gespeicherte geometrisch-topologische Beziehungen)
- Präsentation (graphische Ausprägungen wie Signaturen, Farbe, Typographie)
- Sachdaten (alphanumerische Daten zur Beschreibung der Semantik).

Geodaten stellen in der Informationsverarbeitung eine besondere Herausforderung dar. Die Gründe dafür sind:

- der hohe Aufwand für die Erfassung
- die große Menge der anfallenden Daten
- die geforderten Antwortzeiten beim Zugriff auf Geodaten
- die Verarbeitung nach räumlichen Kriterien sowie
- die Komplexität der Beziehungen.

Geoinformationssystem:

"Ein Geoinformationssystem dient der Erfassung, Speicherung, Analyse und Darstellung aller Daten, die einen Teil der Erdoberfläche und die darauf befindlichen technischen und administrativen Einrichtungen sowie geowissenschaftliche, ökonomische und ökologische Gegebenheiten beschreiben." (BARTELME 1989)

Information:

Information ist nutzbare Antwort auf eine konkrete Fragestellung (ZEHNDER 1998). Information wird meistens aus Daten extrahiert oder abgeleitet.

Informationssystem:

Ein Informationssystem erweitert die Datenbank um eine Reihe von Werkzeugen (engl. software tools) zur Abfrage, Darstellung, Transformation und Analyse von Daten.

Konzeptionell:

Nach Duden: die einem Werk zugrunde liegende Anschauung, Leitidee; für das Datenmodell gilt: unabhängig von einer bestimmten Technologie

Metadaten:

Meta ist ein Präfix, das in den meisten informationstechnologischen Anwendungen „eine zugrundeliegende Definition oder Beschreibung“ meint. Metadaten sind also eine Definition oder Beschreibung von Daten. Man spricht auch von „Daten über Daten“.

Redundanz:

Bezeichnet die mehrfache Speicherung gleicher Daten an verschiedenen Stellen (einer Datenbank).

Struktur:

Die Struktur ist der abstrakte innere Aufbau.

Strukturierter Datenbestand:

Ein Datenbestand wird als strukturiert bezeichnet, wenn er systematische Untergliederungen und Verknüpfungen zulässt.

Syntax:

Syntax ist id Lehre vom Satz, Lehre von den Regeln, wonach in Sprache(n) aus den Wörtern zusammengehörige Wortgruppen gebildet werden, und die Lehre von der Kombination von Wörtern zu Sätzen. „Syntaxis“ heisst auf Griechisch Zusammenstellung, Anordnung, Aneinanderreihung. (WARTH)

WAP:

Das Wireless Application Protocol – kurz: WAP – ist der wichtigste Standard, um Internetkommunikation und interaktive Dienste für Handys und andere mobile Endgeräte zu realisieren: Nachrichten aus dem Internet abrufen, Börsenkurse abfragen, Flugtickets buchen.

1.9. Bibliographie

- **BARTELME, N.**, 2000. *Geoinformatik - Modelle, Strukturen, Funktionen*. 3rd. Berlin: Springer.
- **BARTELME, N.**, 1989. *GIS-Technologie. Geoinformationssysteme, Landinformationssysteme und ihre Grundlagen*. Springer.
- **DITTRICH, K.**, 1999. Datenbanksysteme. In: **RECHENBERG, P.; POMBERGER, G.**, ed. *Informatik-Handbuch*. Wien: Carl Hanser Verlag, 875-908.
- **KOCH, J.**. Available from: http://www.foraus.de/download/aktuelles/hl_ausb_m_lernaufg.pdf [Accessed 27. Januar 2010].
- **Lierz, W.; Scherer, E.** (2004). *Netzwerk von Bibliotheken und Informationsstellen in der Schweiz* [online]. Zürich. Available from: <http://www.nebis.ch/> [Accessed 29. Juni 2005].
- **SBB AG** (2005). *SBB: Online Fahrplan* [online]. Available from: <http://www.sbb.ch/> [Accessed 29. Juni 2005].
- **WARTH, D.**. *Einführung in die Sprachwissenschaft* [online]. Available from: <http://www.fask.uni-mainz.de/inst/iaspk/Linguistik/Syntax/Was.html> [Accessed 27. Januar 2010].
- **ZEHNDER, C.A.**, 1998. *Informationssysteme und Datenbanken*. Zürich: vdf-Hochschulverlag AG.