

Praktikum Datenbanken / DB2
Woche 2: Datenbankentwurf / Datenmodellierung

Raum: LF 230

Bearbeitung: 25.-29. April 2005

Datum	
Gruppe	
Vorbereitung	
Präsenz	

Aktuelle Informationen unter:

http://www.is.informatik.uni-duisburg.de/courses/dbp_ss03/index.html

Datenbankentwurf

Der Entwurf einer Datenbank beschreibt den Prozess der Umsetzung einer Mini-Welt in ein Datenbankschema, das in der Lage ist, die gewünschten Daten dieser Welt mit ihren Eigenschaften und Beziehungen darzustellen.

Der Entwurf, an den sich dann die Implementierung anschließt, besteht im Wesentlichen aus diesen Schritten:

- Modellierung
- Umsetzung in ein Relationenschema
- Normalisierung

Modellierung mit E/R-Diagrammen

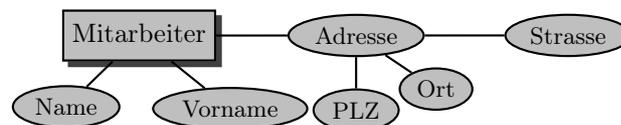
Die Modellierung dient der systematischen Darstellung einer Abstraktion der abzubildenden Miniwelt. Er dient auch zur Kommunikation mit Nichtfachleuten, die in der Regel nicht das Denken in Relationen oder Attributen gewohnt sind. Als eine Möglichkeit zur Modellierung einer Miniwelt ist das Entity-Relationship-Modell weit verbreitet. Zur Darstellung dieser Modelle werden E-R-Diagramme benutzt.

Entitäten und Attribute

Ein Objekt der realen Welt wird als **Entität** modelliert. Eine Entität ist normalerweise eine Person, ein Prozess oder ein Gegenstand der realen Welt, z.B. ein Mitarbeiter, eine Lieferung, ein Inventargegenstand oder ein Schriftstück. Gleichartige Entitäten bilden einen **Entitätstyp**, z.B. alle Wissenschaftlichen Angestellten, alle Autoteile oder alle Arbeitsverträge. In einem E-R-Diagramm wird ein Entitätstyp durch ein Rechteck dargestellt, das den Namen des Entitätstypen umschließt.

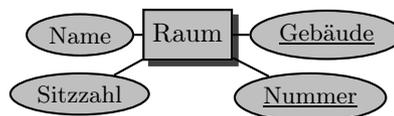
Mitarbeiter

Eine Entität (bzw. ein Entitätstyp) wird durch eine Menge von **Attributen** beschrieben. In einem E-R-Diagramm umschließt eine Ellipse den Namen des Attributes und eine Linie verbindet es mit dem zugehörigen Entitätstyp. Ein Attribut von Mitarbeiter ist z.B. der Name. Man spricht von einem **zusammengesetzten Attribut**, wenn ihm wiederum verschiedene Attribute zugeordnet sind. In einem E-R-Diagramm werden zusammengesetzte Attribute genau wie atomare Attribute dargestellt.

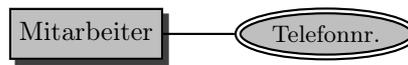


Attributkombinationen, die zur eindeutigen Identifikation einer Entität dienen, werden **Schlüssel** genannt. Beim Entitätstyp Mitarbeiter könnte z.B. (Name, Vorname, Adresse) oder (Mitarbeiternummer) als Schlüssel dienen. Ein Schlüssel identifiziert über die Attributwerte eine Entität eindeutig, insbesondere können keine zwei Entitäten in allen ihren Schlüsselattributen übereinstimmen. Unter allen Schlüsseln wird ein Schlüssel ausgezeichnet, dieser heißt **Primärschlüssel**.

In einem E-R-Diagramm werden die Namen aller Attribute, die zum Primärschlüssel gehören, unterstrichen dargestellt.



Mehrwertige Attribute sind Attribute, die als Wert eine Liste oder eine Menge annehmen können. In einem E-R-Diagramm werden mehrwertige Attribute durch den Attributnamen in einem doppelten Oval dargestellt.



Beziehungen (Relationships)

Die Objekte der realen Welt stehen zueinander in Beziehung. Durch **Beziehungen** werden im E-R-Modell Verknüpfungen, Zusammenhänge und Interaktionen zwischen Entitäten, bzw. den Entitätstypen, modelliert. Beispiele sind etwa ist-Teil-von, arbeitet-für, usw.

Beziehungen derselben Art werden zu sogenannten **Beziehungstypen** zusammengefaßt. In einem E-R-Diagramm werden Beziehungstypen durch eine Raute dargestellt, die den Namen des Beziehungstyps enthält und mit allen Entitätstypen verbunden ist, die an dieser Beziehung teilnehmen. Ein Entitätstyp kann mehrfach an einer Beziehung teilnehmen. Es ist auch möglich, dass ein Entitätstyp in Beziehung mit sich selbst steht (das nennt man dann *rekursive* Beziehung).

Verbindet ein Beziehungstyp nur zwei Entitätstypen jeweils einmal, so spricht man von einem binären Beziehungstyp, ebenso kennt man ternäre Beziehungen, usw.



Beziehungen können wie Entitätstypen Attribute besitzen. Diese werden im E-R-Diagramm genauso dargestellt wie bei Entitätstypen. Da Beziehungstypen nur Zusammenhänge und Beziehungen modellieren, können sie keine identifizierenden Schlüssel besitzen. Im E-R-Diagramm werde keine Schlüssel für Beziehungstypen ausgezeichnet.

Bei Beziehungstypen ist es sinnvoll, sich schon in der Modellierungsphase klarzumachen, wie oft die verschiedenen Entitäten an einer Beziehung teilnehmen können. Diese Überlegungen führen zum Konzept der **Funktionalitäten**:

Nimmt jede Entität eines Typs genau n -mal an einer Beziehung teil, so wird im E-R-Diagramm die Verbindungslinie mit dieser Zahl markiert. Nimmt jede Entität mindestens p -mal und höchstens k -mal an einer Beziehung teil, so wird in der (min,max) -Notation im E-R-Diagramm die Verbindungslinie mit (p,k) markiert. Das Zeichen N als Maximum zeigt normalerweise an, dass eine Entität dieses Typs beliebig oft an dieser Beziehung teilnehmen kann.



Im E-R-Modell ist es nicht möglich, dass Beziehungen selbst wieder an Bezie-

hungen teilnehmen. Das kann aber manchmal nötig und sinnvoll sein. Hierzu fasst man einen Beziehungstyp mit den an ihm beteiligten Entitätstypen zu einem aggregierten Entitätstyp zusammen. Dieser kann wiederum an Beziehungen teilnehmen. Im E-R-Diagramm wird ein aggregierter Entitätstyp dadurch dargestellt, dass die ihn definierende Entitätstypen- und Beziehungstypensymbole in ein Rechteck eingeschlossen werden.

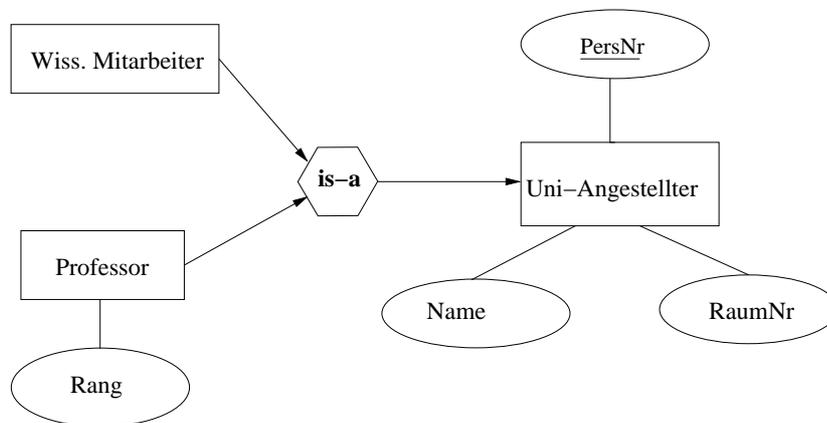
Generalisierung

Um die Entitätstypen besser zu strukturieren, kann man im Entwurf **Generalisierungen** einführen. Dieses objektorientierte Konzept abstrahiert Entitätstypen und bildet Obertypen. Die Entitätstypen, die zu einem Obertyp generalisiert wurden, heißen Kategorien oder Untertypen des Obertyps.

Ein Obertyp fasst allen Kategorien gemeinsame Eigenschaften zusammen. Diese werden vom Untertyp geerbt; zusätzlich kann ein Untertyp Eigenschaften haben, die nur für diesen charakterisierend sind. Die Entitätsmenge des Untertyps ist eine Teilmenge der Entitätsmenge des Obertyps, wobei üblicherweise zwei Fälle besonders interessant sind:

- *disjunkte* Spezialisierung: die Schnittmenge von je zwei Untertypen ist leer
- *vollständige* Spezialisierung: Obertyp ergibt sich als Vereinigung der Untertypen

In einem E-R-Diagramm modelliert man die Generalisierungsbeziehung als Beziehung **is-a** durch ein spezielles Symbol: ein Sechseck statt einer Raute.

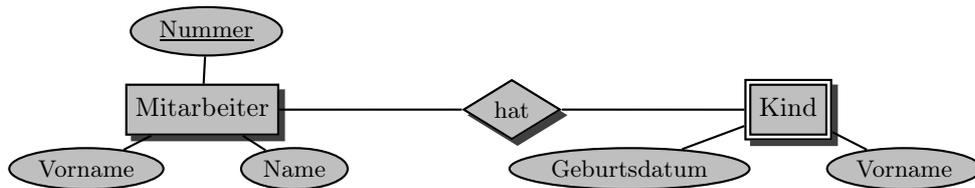


Schwache Entitätstypen

Wenn der Fall auftritt, dass ein Entitätstyp keinen eigenen Schlüsselkandidaten besitzt, spricht man von **schwachen Entitätstypen**. Sie nehmen an einer Beziehung mit einem anderen Entitätstyp teil, dem Eltern-Entitätstyp. Jedes Objekt des schwachen Entitätstyp nimmt mindestens an einer Beziehung zum

Eltern-Entitätstyp teil. Das Objekt wird über diese Beziehung identifiziert. Der Primärschlüssel des Eltern-Entitätstyps wird auch für den schwachen Entitätstyp verwendet (eventuell durch einen Teilschlüssel erweitert).

In der Regel sind die Objekte des schwachen Entitätstyps von den Objekten des Eltern-Entitätstyp abhängig, d.h. ohne diese sind sie nicht von Interesse bzw. besitzen keine eigene Existenz. Im E-R-Diagramm werden schwache Entitätstypen von normalen Entitätstypen dadurch unterschieden, dass ihr Name durch ein doppeltes Rechteck eingeschlossen ist. Die Teilschlüssel werden durch eine gestrichelte Unterstreichung ausgezeichnet.



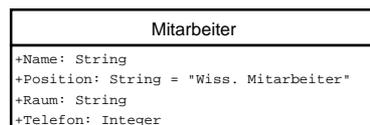
UML

Die *Unified Modelling Language* UML ist eine Standardmodellierungssprache für den objekt-orientierten Entwurf von Software, und findet auch beim Datenbankentwurf ihre Verwendung. Zentrales Objekt von UML ist die Klasse, mit der gleichartige Objekte hinsichtlich ihrer Struktur und ihres Verhaltens modelliert werden. Klassen entsprechen in etwa den Entitätstypen des E-R-Modells. Assoziationen zwischen Klassen entsprechen Beziehungstypen.

Ebenso können Generalisierung und Aggregation mit der UML modelliert werden. Die Angabe von Multiplizitäten entspricht den Funktionalitäten im ER-Modell. Zu Details der UML wird auf das Material der Vorlesung verwiesen, das aus Platzgründen hier nicht wiederholt wird.

Notation

Man stellt in UML-Klassendiagrammen eine Klasse durch ein Rechteck dar, das am oberen Rand den Namen der Klasse und darunter die Liste der Attribute und Operationen der Klasse enthält. Klassenname, Attribute und Operationen werden jeweils durch eine horizontale Linie voneinander getrennt. Dabei stehen alle Klassennamen im Singular und beginnen mit einem Großbuchstaben. Attribute können näher beschrieben sein, etwa durch Attributtyp oder Initialwerte (in der Form `Name: Typ = Initialwert`), besitzen aber mindestens einen Namen.



Assoziationsbeziehungen werden durch eine Verbindungslinie zwischen zwei Klassensymbolen dargestellt. Diese Linie wird mit dem Namen der Beziehung versehen. An den Enden der Verbindungslinie kann die Multiplizität der Beziehung angegeben werden. Sie wird als einzelne Zahl oder als Wertebereich auf jeder Seite der Assoziation notiert (z.B. 1 oder 0..*). Zusätzlich kann an den Enden der Verbindungslinie auch eine Rolle angegeben werden.



Eine Aggregation wird wie eine Assoziation als Linie zwischen zwei Klassen dargestellt, und zusätzlich auf einer Seite der Linie mit einer Raute versehen. Die Raute steht auf der Seite des Aggregats (des Ganzen) und symbolisiert das Behälterobjekt, in dem die Teile gesammelt sind. Auch hier können Multiplizitäten angegeben werden.



Aufgaben

Fallstudie

Im Laufe des Praktikums soll eine Datenbank zur Verwaltung geographischer und politischer Daten entworfen und genutzt werden. Betrachten wir daher zunächst die Miniwelt, die wir modellieren wollen.

- Es gibt Länder, die (zu einem bestimmten Anteil) auf einem oder mehreren Kontinenten liegen.
- Länder können in Landesteile oder Provinzen unterteilt sein. Es gibt Städte, die in einer dieser Provinzen und einem Land liegen. Außerdem kann eine Stadt Hauptstadt eines Landes und/oder einer Provinz sein.
- Länder haben gemeinsame Grenzen.
- Zu einem Land sollen folgende geographische Daten gegeben sein: Name, Ländercode (z.B. D für Deutschland), Größe des Gebietes, Bevölkerungszahl.
- Zu einem Land sollen außerdem folgende wirtschaftliche Daten gegeben sein: Bruttosozialprodukt, Anteile des primären, sekundären und tertiären Sektors an der Wirtschaftsleistung, Inflationsrate.
- Zu einem Land sollen außerdem folgende Bevölkerungsdaten gegeben sein: Bevölkerungswachstum und Kindersterblichkeit.
- Zu einem Land sollen schließlich auch folgende politische Daten gegeben sein: Datum der Unabhängigkeit, Regierungsart.
- Es gibt verschiedene Arten von Gewässern: Flüsse, Seen und Meere. Alle Gewässer haben einen Namen. Flüsse fließen nach einer bestimmten Länge in andere Gewässer: einen Fluß, einen See oder ein Meer. Seen haben eine Fläche und Meere eine maximale Tiefe.
- Es gibt Berge mit einer Gipelhöhe und einer geographischen Position nach Länge und Breite.
- Es gibt Inseln, die zu einer Inselgruppe gehören, eine Fläche besitzen und eine geographische Position Länge und Breite.
- Berge, Flüsse und Wüsten liegen in Provinzen. Meere und Seen grenzen an Provinzen. Inseln gehören zu Provinzen.
- Es gibt politische Organisationen. Diese haben außer einem Namen eine eindeutige Abkürzung (z.B. EU für Europäische Union), ein Gründungsdatum und einen Sitz in einer bestimmten Stadt der Provinz eines Landes.
- Länder sind Mitglieder in Organisationen. Es gibt unterschiedliche Arten von Mitgliedschaften, z.B. Vollmitgliedschaft, assoziierte Mitgliedschaft oder Kandidat.
- Sprachen werden in einem Land von einem bestimmten (Prozent-)Anteil der Bevölkerung gesprochen.

- Religionen werden in einem Land von einem bestimmten (Prozent-)Anteil der Bevölkerung ausgeübt.

- (b) **UML:** Betrachtet den Ausschnitt der Miniwelt, der nur aus Ländern, Provinzen und Städten und ihren Beziehungen miteinander besteht. Modelliert diesen mit `dia` als ein UML-Klassendiagramm.