

Praktikum Datenbanken / DB2
Woche 2: Datenbankentwurf

Raum: LF 230

Nächste Sitzung: 20./23. Oktober 2003

Aktuelle Informationen unter:

http://www.is.informatik.uni-duisburg.de/teaching/lectures/dbp_ws03/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.

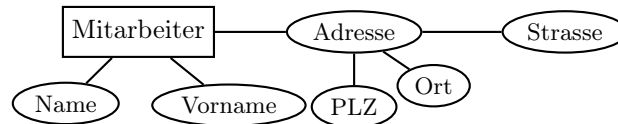
Entitäten

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

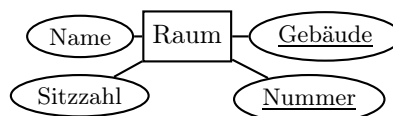
Eine Entität (bzw. ein Entitätstyp) wird durch eine Menge von **Attributen** beschrieben. In einem E-R-Diagramm umschließt eine Ellipse den Namen des

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.

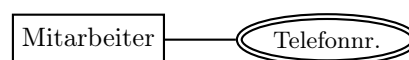


Attributkombinationen, die zur eindeutigen Identifikation einer Entität dienen, werden **Schlüssel** genannt. Beim Entitätstyp Mitarbeiter könnte z.B. (Name, Vorname, Adresse) oder (Mitarbeiternummer) als Schlüssel dienen. Ein Schlüssel identifiziert über die Attributwerte eine Entität eindeutig, insbesondere 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.

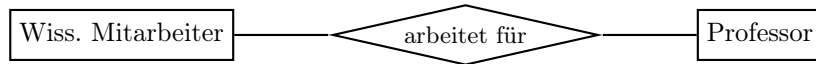


Relationen (Beziehungen)

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 (rekursive Beziehung).

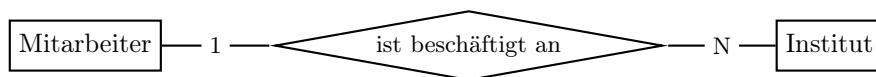
Verbindet ein Beziehungstyp nur zwei Entitätstypen jeweils einmal, so spricht man von einem binären Beziehungstyp, ebenso kennt man ternäre Beziehungen, usw.



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 Beziehungen 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

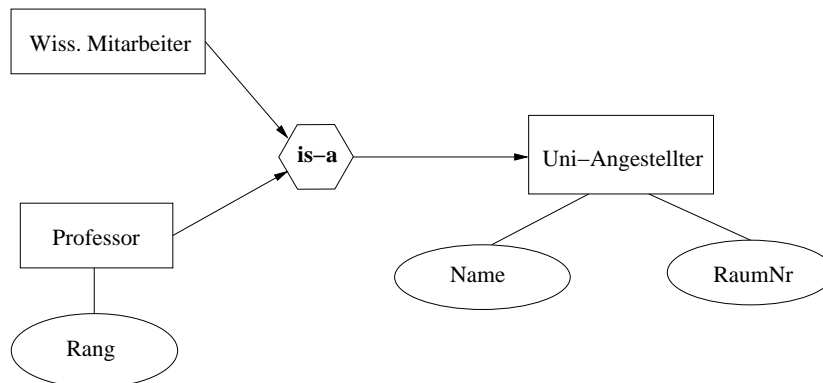
Um die Entitätstypen besser zu strukturieren, kann man im Entwurf **Generalisierungen** einführen. Dieses objektorientierte Konzept abstrahiert Entitätstypen und bildet Obertypen. Die Entitätstypen, die zu einem Obertyp generalisiert wurden, heißen Kategorien oder Untertypen des Obertyps.

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

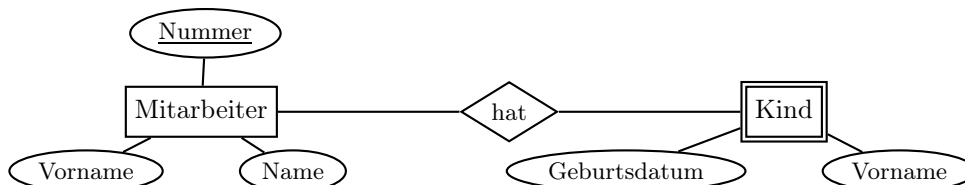
In einem E-R-Diagramm modelliert man die Generalisierungsbeziehung als Beziehung **is-a** durch ein spezielles Symbol: ein Rechteck statt einer Raute.



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.



Vorüberlegungen

- (a) Die Vermeidung von Redundanzen in der Modellierung ist wichtig. Angenommen wir modellieren eine Universität. In diesem Modell soll es unter anderen Entitäten Mitarbeiter und Büro geben. Zwischen diesen gibt es eine Beziehung "sitzt-in", außerdem hat die Entität Mitarbeiter noch ein Attribut "Raum-Nummer". Obwohl das erlaubt ist, ist es schlechtes Design. Warum?
- (b) Nicht jede interessante Beziehung zwischen zwei Entitäten sollte auch modelliert werden. Manche Beziehungen lassen sich aus bereits modellierten Beziehungen ableiten. Überlegt Euch ein nicht-triviales Beispiel für eine solche ableitbare Beziehung.
- (c) Die Entscheidung, ob bestimmte Elemente der zu modellierenden Welt besser als eigenständiger Entitätentyp oder als Attribut eines anderen Entitätentyps beschrieben werden sollen, ist nicht immer leicht zu treffen. Lässt man im Beispiel unter Vorüberlegung 1 die Modellierung von Büros als Entitäten fallen, so beseitigt man die Redundanz des Modells. Was spräche dafür, Büros weiterhin als Entitätstyp statt als Attribut zu modellieren?

Fallstudie

www.cooledvds.de ist ein Online-DVD-Verleih. Als zusätzlichen Service für ihre Kunden will die Videothek eine Filmdatenbank aufbauen, die über ihre Webseite zur Verfügung gestellt werden soll.

Nach ersten Überlegungen kommen die Verantwortlichen von www.cooledvds.de zum Schluss, dass die Filmdatenbank Information über Filme, Fernsehserien, Schauspieler und Regisseure, sowie die Beziehungen dieser untereinander enthalten soll. Das folgende Wissen über die zu modellierende Welt ist bekannt:

- Filme haben einen Titel und ein Drehjahr. Es gibt zahlreiche Filme, die den gleichen Titel haben, aber es soll angenommen werden, dass in einem Jahr keine zwei Filme mit dem gleichen Titel gedreht werden.
- Serien haben einen Titel und ein Drehjahr. Wie bei Filmen soll angenommen werden, dass der Titel allein eine Serie nicht eindeutig kennzeichnet.
- Eine Serie hat im Allgemeinen eine Laufzeit, deren Startjahr auch verschieden vom Drehjahr sein kann. Eine Serie kann entweder eine normale Serie oder eine Miniserie sein.
- Es gibt verschiedene Arten von Filmen: Kinofilm (M), Videofilm (V), Videospiel (VG) oder Fernsehspiel (TV). Manche Filme werden zwar gedreht, aber nie veröffentlicht, also muss auch diese Information in der Datenbank festgehalten werden.
- Serien und Filme sind Produktionen. Produktionen haben ein Budget und ein Studio, in dem sie gedreht werden.

- Genres sind Kategorien von Filmen oder Serie wie Horror, Science-Fiction, Dokumentation, etc. Jede Produktion gehört zu mindestens einem Genre.
- Personen seien durch ihren Künstlernamen eindeutig identifiziert. Ausserdem haben Personen einen Geburtsnamen, ein Geburtsjahr und eventuell ein Todesjahr.
- Schauspieler spielen in einer Produktion eine Rolle (einen Charakter).
- Andere Personen, die in eine Produktion involviert sind, sind Regisseure, Komponisten und Autoren.
- Autoren schreiben die Filmvorlage. Diese kann z.B. ein Drehbuch, aber auch ein Roman, ein Schauspiel, oder etwas anderes sein.
- Zwischen zwei verschiedenen Produktionen kann es eine besondere Verknüpfung geben, die von einer bestimmten Art ist: Fortsetzung, "Prequel", Parodie, etc.

Der DVD-Verleih benutzt natürlich auch eine Datenbank, um sein eigentliches Geschäft zu verwalten. In dieser werden DVDs, Kunden, Mitarbeiter und Verleihvorgänge modelliert.

Aufgaben

- (a) Diskutiert in Eurer Gruppe die oben beschriebene Miniwelt und modelliert sie mit Hilfe eines E-R-Diagramms. Gebt insbesondere auch Primärschlüssel und Funktionalitäten an. Verwendet auch Generalisierung zur Strukturierung der Entitätstypen. Ihr solltet Eure Designentscheidungen begründen können.
- (b) Modelliert die Grundlagen eines Verwaltungssystems für den DVD-Verleih. Kann man die beiden Modellierungen sinnvoll vereinen?