

# Methoden des Datenbankdesigns

## Teil II

In diesem Abschnitt sollen die gängigen Vorgehensweisen während der Entwicklung des konzeptionellen Schemas beschrieben werden.

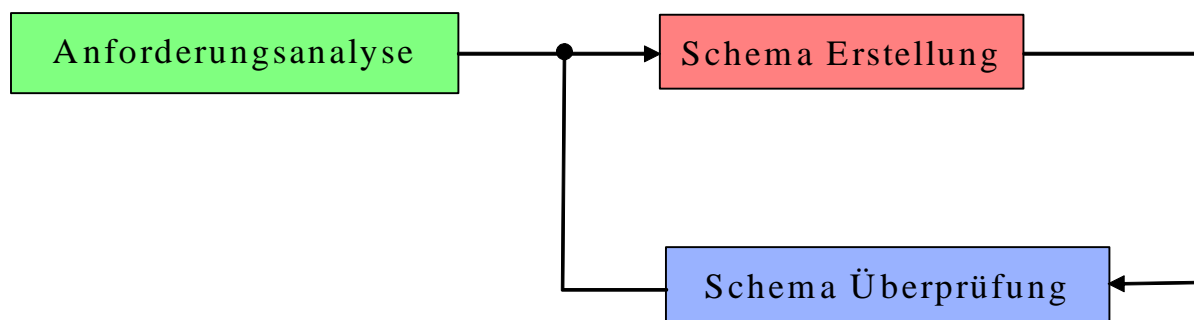
Vor der Erstellung der ersten Schemata steht die *Anforderungsanalyse* (Requirements collection and analysis).

**Ziel der Anforderungsanalyse** ist die Klärung folgender Fragen:

- Welche Daten werden benötigt?
- Was ist die genaue Bedeutung der einzelnen Daten?
- Wie stehen die Daten zueinander in Beziehung?
- Wie häufig/schnell benötige ich welche Daten?

## Anforderungsanalyse

*Anforderungsanalyse* = Sammlung und Analyse von Anforderungen



Die Konstruktion des endgültigen Schemas ist oft ein inkrementeller Vorgang. Während der *Schema Überprüfung* werden oft weitere Bedingungen und Anforderungen entdeckt, so dass die *Schema Erstellung* sich oft mit der Anforderungsanalyse überschneidet.

Die Anforderungsanalyse lässt sich nur schwer standardisieren, da sie stark von der gewünschten Anwendung abhängt.

Unter die **Sammlung von Anforderungen** fällt sowohl die Identifikation aller Probleme, die die gewünschte Anwendung lösen soll, als auch die Funktionalitäten, die sie charakterisieren und bieten soll.

Die Beschreibung der Anforderungen findet in natürlicher Sprache statt. Die **Analyse der Anforderungen** dient der Klärung und Organisation der Anforderungen, d.h. Beseitigung von Zweideutigkeiten und Sortierung nach logischen Zusammenhängen.

Die Anforderungen kommen aus verschiedenen **Quellen**.  
Diese sind im Allgemeinen:

- Die **Benutzer der Anwendung** – durch Interviews oder Fragebögen.
- **Existierende Dokumentation** – zu bisherigen Lösungen, gesetzlichen Grundlagen, der üblichen Arbeitsweise in Unternehmen, etc.
- Mögliche vorherige **Applikationen**, die **ersetzt** werden sollen.
- Mögliche vorherige **Applikationen**, mit denen **interagiert** werden muss.

## Beispiel: Schulungsunternehmen

1 We wish to create a database for a company that runs training courses.  
2 For this, we must store data about the trainees and the instructors. For  
3 each course participant (about 5000), identified by a code, we want to  
4 store the social security number, surname, age, sex, place of birth,  
5 employer's name, address and telephone number, previous employers  
6 (and period employed), the courses attended (there are about 200  
7 courses) and the final assessment of each course. We need also to  
8 represent the seminars that each participant is attending at present  
9 and, for each day, the places and times the classes are held. Each course  
10 has a code and a title and any course can be given any number of times.  
11 Each time a particular course is given, we will call it an 'edition' of the  
12 course. For each edition, we represent the start date, the end date, and  
13 the number of participants. If a trainee is a self-employed professional,  
14 we need to know his or her area of expertise, and, if appropriate, his or  
15 her title. For somebody who works for a company, we store the level  
16 and position held. For each instructor (about 300), we will show the  
17 surname, age, place of birth, the edition of the course taught, those  
18 taught in the past and the courses that the tutor is qualified to teach.  
19 All the instructors' telephone numbers are also stored. An instructor  
20 can be permanently employed by the training company or can be  
21 freelance.

## Regeln zur Spezifikation von Anforderungen

Um die Anforderungen möglichst präzise und eindeutig zu formulieren, sollten folgende Regeln beachtet werden:

- Die **Wahl der Abstraktionsebene**. Um das gewünschte Konzept klar zu beschreiben, sollte man weder zu abstrakt noch zu speziell formulieren.

In unserem Beispiel finden wir die Begriffe *period* (Zeile 6), *title* (Zeile 15) und *assessment* (Zeile 7). Sie können präziser beschrieben werden, beispielsweise durch:

- *start date* und *end date* für *period*
- *professional title* für *title*
- *marks out of ten* für *assessment*

- **Standardisierung der Satzstrukturen.**  
Die Anforderungen sollen vom Aufbau her möglichst einheitlich formuliert werden.  
Z.B. ‚Für <Konzept> wird benötigt <Eigenschaft>‘.

- **Vermeidung komplexer Formulierungen.**  
Definitionen sollten einfach und klar formuliert sein.

In unserem Beispiel schreibt man besser *employee* anstelle von *somebody who works for a company* (Zeile 15).



- Identifizierung von **Synonymen und Homonymen** und Standardisierung dieser Begriffe.

- **Synonyme** sind verschiedene Begriffe mit gleicher Bedeutung. Hier sollte man sich auf einen Begriff einigen.

Z.B. *tutor* (Zeile 18) und *instructor* (Zeile 16) oder *course participant* (Zeile 3) und *trainee* (Zeile 2)

- **Homonyme** sind Begriffe mit mehreren Bedeutungen. Hier sollte ein eindeutiger Begriff verwendet werden.

Z.B. *place* wird sowohl für *place of birth* (Zeile 4) als auch für den Unterrichtsort (Zeile 9) verwendet.

- Explizites **Benennen von Querverweisen**.

Es ist nicht klar, ob beispielsweise die Begriffe *address* und *telephone number* (Zeile 4) zum Teilnehmer oder dessen Arbeitgeber gehören.

Des Weiteren müssen Ausdrücke wie *somebody who works for ...* (Zeile 15) genau zugeordnet werden können (Kursleiter, Teilnehmer?), um Missverständnisse zu vermeiden.

- Erstellung eines **Glossars**, in dem die einzelnen verwendeten Begriffe beschrieben werden.

Für unser Beispiel kann das Glossar wie folgt aussehen:

<b>Term</b>	<b>Description</b>	<b>Synonym</b>	<b>Links</b>
Trainee	Participant in a course. Can be an employee or self-employed.	Participant	Course, Company
Instructor	Course tutor. Can be freelance.	Tutor	Course
Course	Course offered. Can have various editions.	Seminar	Instructor, Trainee
Company	Company by which participant is employed or has been employed.		Trainee

## Einteilung der Aussagen in logische Gruppen

- **Phrases of a general nature**

We wish to create a database for a company that runs training courses. We wish to hold the data for the trainees and the instructors.

- **Phrases relating to the trainees**

For each trainee (about 5000), identified by a code, we will hold the social security number, surname, age, sex, town of birth, current employer, previous employers (along with the start date and the end date of the period employed), the editions of the courses the trainee is attending at present and those he or she has attended in the past, with the final marks out of ten.

- **Phrases relating to the employers of the trainees**

For each employer of a trainee we will hold the name, address and telephone number.

- **Phrases relating to the courses**

For each course (about 200), we will hold the name and code. Each time a particular course is given, we will call it an 'edition' of the course. For each edition, we will hold the start date, the end date, and the number of participants. For the editions currently in progress, we will hold the dates, the classrooms and the times in which the classes are held.

- **Phrases relating to specific types of trainee**

For a trainee who is a self-employed professional, we will hold the area of expertise and, if appropriate, the professional title. For a trainee who is an employee, we will hold the level and position held.

- **Phrases relating to the instructors**

For each instructor (about 300), we will hold surname, age, town of birth, all telephone numbers, the edition of courses taught, those taught in the past and the courses the instructor is qualified to teach. The instructors can be permanently employed by the training company or can be freelance.

Neben der Spezifikation der benötigten Daten muss auch bestimmt werden, welche **Operationen** (O) auf diesen Daten nötig sind und wie oft diese ausgeführt werden müssen.

Beispiel:

- O 1:** Einfügen eines neuen Teilnehmers mit sämtlichen Daten.  
(ca. 40 mal pro Tag).
  
- O 2:** Zuordnung eines Teilnehmers zu einem Kurs.  
(ca. 50 mal pro Tag).
  
- O 3:** Einfügen eines neuen Lehrers mit sämtlichen Daten und den Kursen, die er unterrichten kann.  
(ca. 2 mal pro Tag).
  
- O 4:** Zuordnung eines qualifizierten Lehrers zu einem Kurs.  
(ca. 15 mal pro Tag).

- O 5:** Anzeigen aller Informationen über vergangene Kurse.  
(ca. 10 mal pro Tag).
  
- O 6:** Anzeigen aller angebotenen Kurse und der Lehrer, die qualifiziert sind, diese Kurse zu unterrichten.  
(ca. 20 mal pro Tag).
  
- O 7:** Anzeigen aller Teilnehmer, die von einem Lehrer unterrichtet wurden oder werden.  
(ca. 5 mal pro Woche).
  
- O 8:** Statistische Analyse über alle Teilnehmer bezüglich der Kurse, an denen sie teilnahmen und der entsprechenden Beurteilungen durch den Lehrer.

## Generelle Kriterien zur Datenrepräsentation

Die Modellierung gegebener Daten kann je nach Perspektive verschieden ausfallen. Es gibt einige ‚konzeptionelle Regeln‘, die befolgt werden sollten:

- Falls ein Konzept signifikante Eigenschaften aufweist und/oder eine Menge von Objekten beschreibt, die autonom existieren, ist es üblich, diese auch als **Objekte (Entitäten)** darzustellen.

Beispiel:

Ein Lehrer (instructor) in unserem Beispiel des Schulungsunternehmens würde als Objekt realisiert werden, da er verschiedene Attribute aufweist (Name, Alter, Geburtsort, ...) und seine Existenz unabhängig von anderen Konzepten ist.



- Falls ein Konzept eine einfache Struktur hat und keine relevanten Eigenschaften mit ihm verbunden sind, ist es üblich, das Konzept als **Attribut** eines anderen Konzeptes, mit dem es verbunden ist, darzustellen.

**Beispiel:**

Das Alter eines Lehrers oder Kurs-Teilnehmers.

- Wenn eine Anforderung ein Konzept enthält, welches eine logische Verbindung zwischen zwei oder mehr Objekten (Entitäten) darstellt, kann dieses durch eine Beziehung oder **Relation (relationship)** dargestellt werden.

**Beispiel:**

Im Schulungsunternehmen würde die Teilnahme an einem Kurs mit einer Relation veranschaulicht werden. Die Relation verbindet Kurs und Teilnehmer.

- Falls ein oder mehrere Konzepte bestimmte Teilmengen eines anderen Konzeptes sind, bietet sich die **Generalisierung** an, um diese zu repräsentieren.

Beispiel:

Ein Kurs-Teilnehmer kann ein Angestellter einer Firma oder selbständig sein.