



HOCHSCHULE KONSTANZ TECHNIK, WIRTSCHAFT UND GESTALTUNG
UNIVERSITY OF APPLIED SCIENCES

Modellierung von Testdaten

Nikolaus Moll

287336

Konstanz, 11. Oktober 2013

Master-Arbeit

Master-Arbeit

zur Erlangung des akademischen Grades

Master of Science

an der

Hochschule Konstanz

Technik, Wirtschaft und Gestaltung

Fakultät Informatik

Studiengang Master Informatik

Thema:	Modellierung von Testdaten
Verfasser:	Nikolaus Moll TODO TODO TODO
1. Prüfer:	TODO TODO TODO TODO TODO TODO
2. Prüfer:	PRUEFERBTITLE PRUEFERB TODO TODO TODO TODO
Abgabedatum:	11. Oktober 2013

Abstract

Thema: Modellierung von Testdaten

Verfasser: Nikolaus Moll

Betreuer: TODO TODO
PRUEFERBTITLE PRUEFERB

Abgabedatum: 11. Oktober 2013

Das Abstract befindet sich in `formal/abstract.tex`.

Ehrenwörtliche Erklärung

Hiermit erkläre ich *Nikolaus Moll*, geboren am *22.12.1981* in *TODO*, dass ich

- (1) meine Master-Arbeit mit dem Titel

Modellierung von Testdaten

selbständig und ohne fremde Hilfe angefertigt und keine anderen als die angeführten Hilfen benutzt habe;

- (2) die Übernahme wörtlicher Zitate, von Tabellen, Zeichnungen, Bildern und Programmen aus der Literatur oder anderen Quellen (Internet) sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet habe.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben kann.

Konstanz, 11. Oktober 2013

Nikolaus Moll

Inhaltsverzeichnis

Abstract	v
Ehrenwörtliche Erklärung	vii
1 Einleitung	1
2 Grundlegende Konzepte	3
2.1 Modellgetriebene Software-Entwicklung	3
2.2 Software-Tests	3
2.3 DbUnit	4
2.4 SB Testing DB	4
2.5 Konventionen	6
2.5.1 Datenbank-Diagramme	6
3 Anforderungsanalyse / Fragestellung	7
3.1 Allgemeine Anforderungen	7
3.2 Fortlaufendes Beispiel	8
3.2.1 Voraussetzungen	8
3.2.2 Gewählte Problemstellung	8
3.2.3 Beispiel-Use-Cases	11
3.3 Modellierungsvarianten der Testdaten für DbUnit	12
3.3.1 XML-Dataset	12
3.3.2 Default-Dataset	14
3.3.3 SB-Testing-DB-DataSet	15
4 Modellierung der Test-Daten	17
4.1 DSL-Entwürfe	17
4.1.1 Entwurf 1	17
4.1.2 Entwurf 2	18

INHALTSVERZEICHNIS

4.1.3	Entwurf 3	18
4.2	Wahl der DSL	19
4.3	Implementierungsvorbereitung	19
4.3.1	Implementierungsvarianten	20
4.3.2	Implementierungsentscheidung	22
4.4	Realisierung	22
4.4.1	Builder-Klassen für das Datenbankmodell	23
4.4.2	Spalten-Flags	23
4.4.3	Architektur der generierten Klassen	23
4.4.4	Allgemeiner Tabellenparser	24
4.4.5	Builder für DataSet	25
4.4.6	Builder für Tabellen	25
4.4.7	ColumnBinding	25
4.4.8	Referenzen	25
4.4.9	Scopes	25
4.4.10	Erweiterungen in generierter API	25
4.4.11	JavaDoc	25
5	Generieren von Testdaten	27
6	Proof of Concept	29
7	Zusammenfassung und Ausblick	31
	Abkürzungsverzeichnis	35
	Abbildungsverzeichnis	37
	Listings	39
	Literaturverzeichnis	41

Kapitel 1

Einleitung

Kapitel 2

Grundlegende Konzepte

noch zu erklären: - benutzte Terminologie - Modellgetriebene Software-Entwicklung - Tests, Testdatengenerierung - Literatur nutzen

2.1 Modellgetriebene Software-Entwicklung

- **M0:** Konkrete Information
- **M1:** Meta-Daten zum Beschreiben der Information. Auch als *Modell* bezeichnet.
- **M2:**
- **M3:**

- Modell, Meta-Modell, Modell-Ebenen - <http://www.omg.org/spec/MOF/ISO/19502/PDF/>
spricht von den klassischen 4 schichten... - DSL, intern vs. extern

2.2 Software-Tests

Das zu testende System wird im Rahmen von Software-Tests als *System Under Test* (abgekürzt SUT) bezeichnet. Dabei kann es sich je nach Test und Kontext auf Klassen, Objekte, Methoden, vollständige Anwendungen oder Teile davon beziehen. [8, 810f]

Alle Voraussetzungen und Vorbedingungen für einen Testlauf werden unter der Bezeichnung *Test Fixture* zusammengefasst. Es repräsentiert den Zustand des SUT vor den Tests. [8, S. 814] Es gibt verschiedene Arten von Test Fixtures. Die im Rahmen dieser Arbeit relevanten sind *Standard Fixture* und *Minimal Fixture*.

Ein Test Fixture wird als Standard Fixture bezeichnet, wenn es für alle bzw. fast alle Tests verwendet werden kann. Ein Standard Feature reduziert nicht nur den Aufwand zum Entwerfen von Testdaten für die einzelnen Tests, sondern verhindert darüber hinaus, dass der Test-Ingenieur sich bei verschiedenen Tests immer wieder in unterschiedliche Test-Daten hineinversetzen muss. Nur in Ausnahmefällen sollten Tests modifizierte oder eigene Testdaten verwenden. ([8, S. 305])

Minimal Fixtures stellen Test Fixtures dar, deren Umfang auf ein Minimum reduziert wurde. Dadurch lassen sich Minimal Fixtures im Allgemeinen leichter verstehen. Das Reduzieren der Daten kann auch zu Leistungsvorteilen bei der Ausführung der Tests führen. ([8, S. 302])

Eine übliche Vorgehensweise, Systeme in Verbindung mit Datenbanken zu testen, stellt *Back Door Manipulation* dar. Dabei wird die Datenbank über direkten Zugriff, vorbei am zu testenden System, in den Anfangszustand gebracht. Anschließend können die zu testenden Operationen am System durchgeführt werden. Um zu überprüfen, ob sich das System richtig verhalten hat, wird der Zustand der Datenbank mit dem erwarteten Zustand verglichen - ebenfalls am zu testenden System vorbei. [8, 327ff]

To do (1)

Es gibt mehrere Vorteile, die Datenbank nicht über das zu testende System in den Anfangszustand zu bringen. Einerseits können semantische Fehler im zu testenden System unter Umständen nur so gefunden werden. Andererseits kann der Zustand mitunter schneller in die Datenbank geschrieben werden, wenn nicht der Weg über das zu testende System gemacht wird. Außerdem bietet es in Bezug auf die Zustände eine höhere Flexibilität: Die Datenbank kann auch in Zustände gebracht werden, die über das System nicht erreicht werden können. Dafür leidet die Flexibilität an einer anderen Stelle: Die Tests sind abhängig vom konkret verwendeten Datenbank-System. Außerdem setzt der direkte Zugriff auf die Datenbank voraus, dass die Semantik der zu testenden Anwendung berücksichtigt wird. Aus Sicht der Anwendung dürfen sich von der Anwendung eingespielte Daten in ihrer Form nicht von den manuell in die Datenbank geschriebenen Daten unterscheiden.

To do (2)

2.3 DbUnit

To do (3)

2.4 SB Testing DB

Die Firma Seitenbau GmbH verwendet für die Java-basierten Datenbankanwendungen das Framework JUnit mit der Erweiterung DbUnit. Da die Modellierung der Testdaten mit DbUnit-eigenen Mitteln einige Nachteile hat (siehe Abschnitt 3.3), hat Seitenbau die Bibliothek SB Testing DB entwickelt. SB Testing DB verfolgt keinen generischen Ansatz, der eine Standard-API für die Modellierung von DataSets definieren würde. Stattdessen wird für jedes Datenbank-Modell ein individuelles API bzw. eine interne Java-DSL generiert.

Abbildung 2.1 stellt grafisch dar, wie aus einem Datenbank-Modell die Java-DSL erzeugt wird. Ausgangspunkt ist ein relationales Datenbankmodell. Üblicherweise liegt es bei Seitenbau als *Apache-Torque*-Modell im XML-Format vor. Dieses muss vor der eigentlichen Code-Erzeugung in ein für den Generator interpretierbares Modell (das SB-Testing-DB-Modell) transformiert werden, das die Meta-Informationen zur Datenbank enthält. Zu den notwendigen Meta-Informationen gehören Tabellennamen und Daten zu den Spalten, z.B. Name und Datentypen. Das SB-Testing-DB-Modell kann manuell gepflegt und auch vollständig manuell entwickelt werden. Mit Hilfe des SB-Testing-DB-Modells erzeugt der Generator die interne Java-DSL (DSL Model).

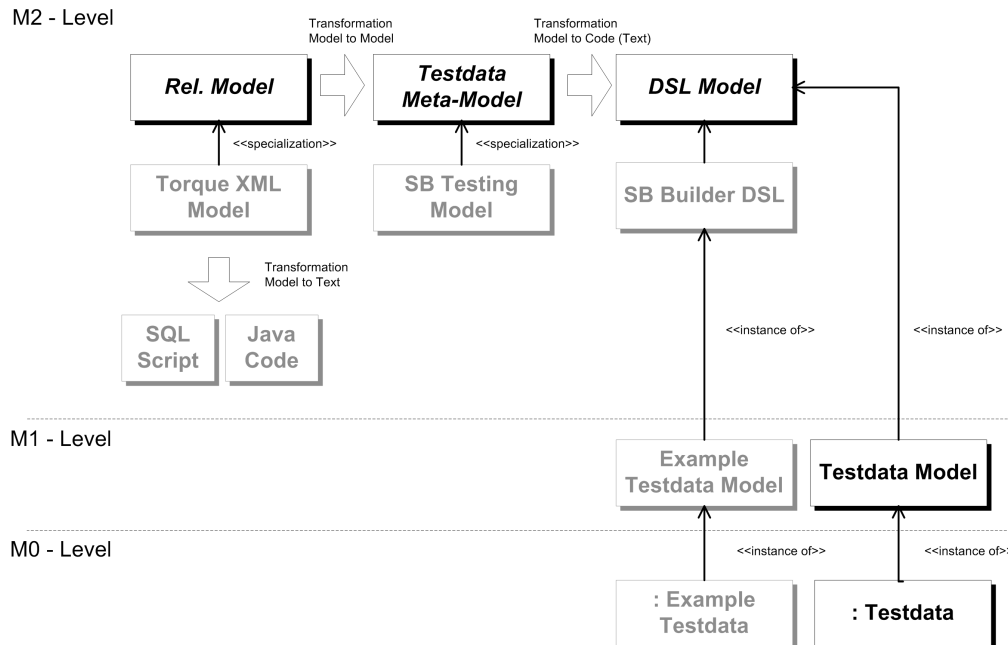


Abbildung 2.1: Modell-Beschreibung

Das SB-Testing-DB-Modell enthält keine Datenbank-Constraints. Eine Abbildung dieser würde keine wesentliche Vorteile bringen. Das API bzw. die erzeugten DataSets sind ausschließlich für den Einsatz Test-Umfeld gedacht. Sollte ein DataSet Daten enthalten, die gegen die in der Datenbank definierten Constraints verstoßen, scheitert das Einspielen des DataSets und eine Exception wird ausgelöst. Aus Sicht des Testers ist dieses Verhalten ausreichend, da die Exception zum Scheitern der Unit-Tests führen wird. Der Mehrwert, das ungültige DataSets schon vor dem Einspielen als solches zu erkennen, ist minimal im Vergleich zu dem Aufwand, Constraints verschiedener Datenbanken nachzubauen.

Der Generator erzeugt die Java-Klassen mit Hilfe von *Apache Velocity*. Velocity ist eine sogenannte Template-Engine, die aus Vorlagen Dokumente erzeugt. Die Vorlagen können Platzhalter enthalten, die von Velocity durch konkrete Werte ausgetauscht werden, und auch von Velocity interpretierte Steueranweisungen, z.B. Verzweigungen und Schleifen.

Die generierten Klassen setzen das Builder-Pattern mit einem Fluent API um. ^{To do (4)}

Unter anderem werden Klassen der folgenden Kategorien erzeugt:

- **DataSet:** Es wird eine abstrakte DataSet-Klasse generiert. Der Zugriff auf die Tabellen erfolgt über öffentliche Felder. Die Klasse enthält die Methode `createDBUnitDataSet`, um die für die Unit-Tests benötigten DbUnit-DataSets zu erzeugen. Dabei werden Template-Methoden definiert, die genutzt werden können, um in den Erzeugungsprozess von DataSets einzugreifen. Die Klasse enthält darüber hinaus einige Convenience-Methoden zum Hinzufügen von Zeilen in die entsprechende Tabellen.
- **Table:** Für jede Tabelle wird eine entsprechende Klasse generiert. Die Klasse stellt Methoden zum Hinzufügen und zum Löschen von Tabellenzeilen bereit. Für Tabellenzeilen gibt es eine passende RowBuilder-Klasse. Außerdem kann nach Zeilen gesucht werden. Dafür gibt es einerseits eine Klasse (*FindWhere*, Beschreibung siehe

unten) für einfache Werte-Suchen. Andererseits gibt es auch die Möglichkeit, eine Zeile in Form eines RowBuilders als Such-Parameter zu verwenden. Die gefundenen Zeilen müssen in allen auf dem RowBuilder gesetzten Werten in den jeweiligen Spalten übereinstimmen.

Da die Klasse das DbUnit-Interface ITable implementiert, kann sie direkt in DbUnit-DataSets verwendet werden. Dafür muss die Klasse auch Meta-Informationen zu den Spalten beinhalten.

- **RowBuilder:** Zu jeder Tabelle wird eine Klasse zur Repräsentation einer Tabellenzeile generiert. Die Klasse beinhaltet für jede Spalte mehrere Methoden zum Setzen und Abfragen des jeweiligen Wertes. Die Methodennamen setzen sich zusammen aus der Aufgabe (get bzw. set) und dem Spaltennamen. **To do** (5)
- **FindWhere:** Für einfache Suchanfragen gibt es die Klasse FindWhere. Sie ermöglicht die Suche nach einem Wert in einer Spalte. Es wird eine Liste von Tabellenzeilen zurückgeliefert. Nicht zu unrecht trägt sie den Ausdruck *find* im Namen. Sie dient dem Zweck, Zeilen zu finden, von denen bekannt ist, dass es sie gibt. Entsprechend wird eine Exception geworfen, wenn keine Zeile entsprechend dem Suchkriterium gefunden wird.

To do (6)

2.5 Konventionen

2.5.1 Datenbank-Diagramme

Für die Darstellung von Datenbank-Diagrammen wird ein einheitlicher Stil verwendet. Dieser orientiert sich an Ambler aus [1]. Auf die Angabe von Stereotypen wird sowohl bei den Tabellen, als auch bei den Beziehungen zwischen Tabellen verzichtet.

Erklären: - Spalten/PK/FK - Kardinalitäten

Abbildung 2.2 zeigt ein Diagramm mit zwei Tabellen. Tabelle 2 enthält einen Fremdschlüssel, der einem Primärschlüssel aus Tabelle 1

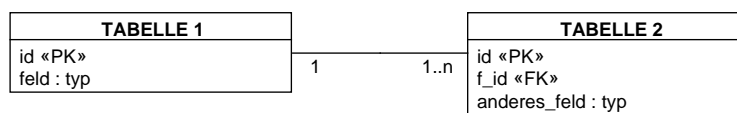


Abbildung 2.2: Datenbank-Diagramm-Stil nach Ambler

Kapitel 3

Anforderungsanalyse / Fragestellung

Einleiten: Zwei Fragestellungen, DSL und Generierung

3.1 Allgemeine Anforderungen

Die Firma *Seitenbau GmbH* verwendet für die Java-basierten Datenbank Anwendungen das Framework *JUnit* und die Erweiterung *DbUnit*. Zwar erlaubt die verwendete Bibliothek *SB Testing DB* (siehe 2.4) eine einfachere Modellierung als mit *DbUnit*-eigenen Mitteln,

Da die Modellierung der Testdaten mit *DbUnit*-eigenen Mitteln einige Nachteile hat, hat *Seitenbau* die Bibliothek *SB Testing DB* entwickelt. Allerdings kann *SB Testing DB* nicht alle Nachteile wett machen, so bleibt die Modellierung von Beziehungen unübersichtlich. Eine genauere Betrachtung der Vor- und Nachteile verschiedener Modellierungsvarianten in Zusammenhang mit *DbUnit* wird in Abschnitt 3.3 beschrieben.

Die allgemeinen Anforderungen an die zu entwickelnde Sprache sind wie folgt:

- **Integration in Werkzeugkette:** Eine der wichtigsten Anforderungen an die DSL ist, dass sie sich in die bestehende Werkzeugkette der Firma *Seitenbau* integrieren lassen muss. Daraus folgt die Anforderung, dass sie sich in Java nutzen lassen soll. Ähnlich wie bei *SB Testing DB* sollen Datasets auch nachträglich veränderbar sein.
- **Schlankheit:** Die Sprache soll auf syntaktischen Ballast verzichten und einen übersichtlichen Code zur Modellierung der Daten ermöglichen. Meta-Informationen sollten ausschließlich in Form von Sprachelementen auftauchen.
- **Beziehungen:** Beziehungen sollen sich einfach und typsicher modellieren lassen. Es soll nicht mehr notwendig sein, symbolische Java-Konstanten z.B. für die Definition von ID-Nummern zu verwenden.
- **Typ-Sicherheit:** Beim Test müssen falsche Typen (z.B. bei Beziehungen) erkannt werden und den Test scheitern lassen. Idealerweise sollten die Typen allerdings schon zur Compiler-Zeit überprüft werden können.
- **Funktionen als Werte:** - Datumswerte - Berechnungen - Einlesen von BLOBs
- **Gültigkeitsbereiche:**

- **Zielgruppe:** Die Zielgruppe für die DSL sind überwiegend Software-Entwickler. Anwender, die versiert sind im Umgang mit Datenbanken, sollten zumindest keine Probleme beim Lesen und Verstehen der DSL haben.
- **-Diskussion-**: Sollen sich auch „ungültige Daten“ modellieren lassen?

3.2 Fortlaufendes Beispiel

Eine einheitliche und fortlaufende Problemstellung soll der Arbeit als Grundlage dienen. Die Problemstellung besteht aus einem Modell und einem Satz von Testdaten. Alle im weiteren Verlauf diskutierten Modellierungsvarianten werden diese Problemstellung umsetzen und die Testdaten modellieren.

3.2.1 Voraussetzungen

Der Schwerpunkt der Modellierung liegt bei der Darstellung von Beziehungstypen zwischen Entitätstypen. Dabei soll die Problemstellung einerseits nicht zu komplex sein, damit sie überschaubar bleibt. Andererseits soll sie komplex genug sein, um möglichst alle Beziehungsarten zwischen Entitäten abzudecken. Die Testdaten sollten gleichzeitig ein *Standard Fixture* und ein *Minimal Fixture* darstellen (siehe Abschnitt 2.2).

3.2.2 Gewählte Problemstellung

Das gewählte Beispiel stellt eine starke Vereinfachung des Prüfungswesens an Hochschulen dar. Auf eine praxisnahe Umsetzung wird zugunsten der Komplexität verzichtet. Personenbezogene Begriffe werden in der maskulinen Form verwendet, ohne dabei Aussagen über das Geschlecht der repräsentierter Personen zu machen. Es beinhaltet die folgenden vier Entitätstypen:

- **Professor:** Ein Professor leitet Lehrveranstaltungen.
- **Lehrveranstaltung:** Eine Lehrveranstaltung wird von einem Professor geleitet. Es kann zu jeder Lehrveranstaltung eine Prüfung geben.
- **Prüfung:** Eine Prüfung ist einer Lehrveranstaltung zugeordnet. Außerdem hat mindestens ein Professor Aufsicht.
- **Student:** Studenten können an Lehrveranstaltungen und an Prüfungen teilnehmen. Studenten haben außerdem die Möglichkeit, Tutoren von Lehrveranstaltungen zu sein.
- **Raum:** Ein Professor kann einen Raum als Büro zugewiesen bekommen.

Die Beziehungen der Entitätstypen stellen sich wie folgt dar:

- Eine Lehrveranstaltung muss von genau einem Professor geleitet werden, ein Professor kann beliebig viele (also auch keine) Lehrveranstaltungen leiten.

3.2. FORTLAUFENDES BEISPIEL

- Eine Prüfung ist genau einer Lehrveranstaltung zugeordnet. Eine Lehrveranstaltung kann mehrere Prüfungen haben (z.B. Nachschreibprüfung).
- Eine Prüfung muss mindestens von einem Professor beaufsichtigt werden, ein Professor kann in beliebig vielen Prüfungen Aufsicht haben.
- Jeder Student kann beliebig vielen Lehrveranstaltungen besuchen. Lehrveranstaltungen benötigen jedoch mindestens drei Besucher um stattzufinden und sind aus Kapazitätsgründen auf 100 Teilnehmer begrenzt.
- Jeder Student kann bei beliebig vielen Lehrveranstaltungen Tutor sein und jede Lehrveranstaltung kann beliebig viele Tutoren haben.
- Jeder Student kann an beliebig vielen Prüfungen teilnehmen und umgekehrt eine Prüfung von einer beliebigen Anzahl von Studenten geschrieben werden.
- Ein Raum ist genau einem Professor zugewiesen. Ein Professor kann genau einen oder keinen Raum haben.

Abbildung 3.1 zeigt die Problemstellung grafisch in Form eines ER-Diagramms. **To do (7)**
To do (8) **To do (9)**

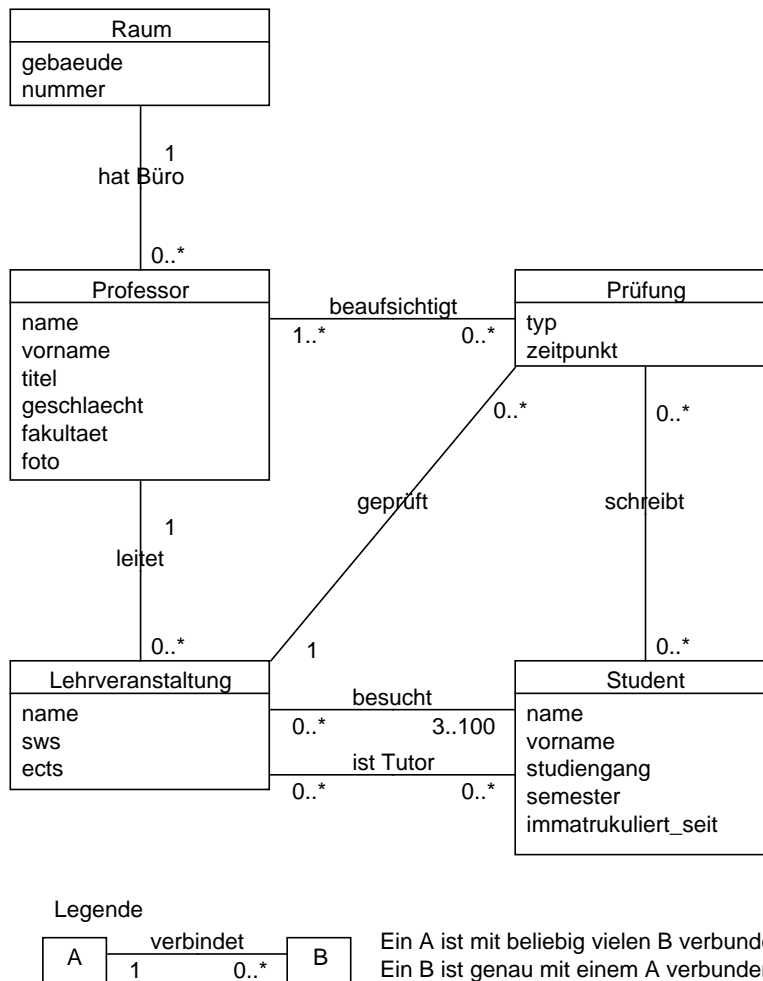


Abbildung 3.1: ER-Diagramm des fortlaufenden Beispiels

Das entsprechende Datenbank-Diagramm wird in Abbildung 3.2 dargestellt.

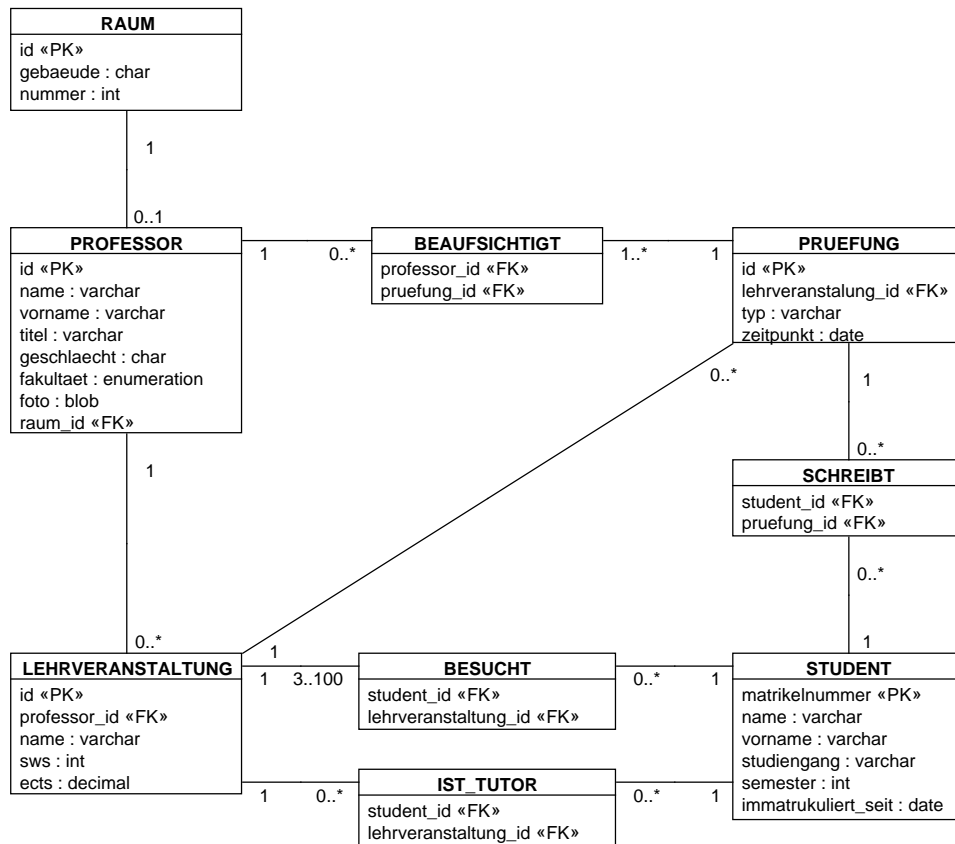


Abbildung 3.2: Datenbank-Diagramm des fortlaufenden Beispiels

To do (10)

Das Attribut „fakultaet“ in der Tabelle Professor soll als Aufzählungstyp (enumeration) realisiert werden. Mögliche Werte sind: Architektur, Bauingenieurwesen, Elektrotechnik, Informatik, Maschinenbau und Wirtschaftswesen. Das Foto des Professors wird als *Binary Large Object* (BLOB) dargestellt.

3.2.3 Beispiel-Use-Cases

Um den einen Kompromiss für die Komplexität der Testdaten zu finden, werden vier Fragestellungen definiert. Diese Fragen sollen dabei helfen, den Umfang der Testdaten bestimmen zu können. Die Fragen stellen sich wie folgt dar:

1. Welcher Professor unterrichtet die meisten Studenten?
2. Welcher Student nimmt an den meisten Prüfungen teil?
3. Welcher Student ist Tutor und nimmt gleichzeitig an der Prüfung teil?
4. Welcher Professor macht die wenigste Aufsicht in Fremdveranstaltungen (Lehrveranstaltungen eines anderen Professors)?

3.3 Modellierungsvarianten der Testdaten für DbUnit

In *DbUnit* werden die Datenbankzustände durch Datasets repräsentiert. Für einen Test werden gewöhnlich zwei Datasets benötigt: das erste für den Anfangszustand, das zweite für den erwarteten Zustand. Datasets aus *DbUnit* bieten allerdings nicht die Möglichkeit, aus einem bestehenden Dataset ein zweites zu erzeugen, das die Änderungen an der Datenbank beinhaltet.

Im Folgenden werden verschiedene Modellierungsarten für *DbUnit*-Datasets diskutiert. Die Erkenntnisse sollen in die Anforderungen an die DSL einfließen.

3.3.1 XML-Dataset

Eine Variante, ein Dataset für *DbUnit* zu modellieren, stellt XML dar. *DbUnit* bietet dazu die Klasse *XmlDataSet*, das eine XML-Datei nach einem vorgegebenen Dokumententyp einlesen kann. Das Listing 3.1 zeigt einen Ausschnitt einer solchen XML-Datei, in dem die beiden Tabellen *Professor* und *Lehrveranstaltung* definiert werden.

BLOBs können in XML problematisch werden. Werden sie direkt in die XML-Datei eingebettet, müssen sie in ein XML-kompatibles Format codiert werden. Die *DbUnit*-Datasets sehen keinen Mechanismus vor, BLOBs aus anderen Dateien einzulesen - eine solche Funktion müsste manuell implementiert werden.

```

1  <!DOCTYPE dataset SYSTEM "dataset.dtd">
2  <dataset>
3      <table name="PROFESSOR">
4          <column>id</column>
5          <column>name</column>
6          <column>vorname</column>
7          <column>titel</column>
8          <column>fakultaet</column>
9          <row>
10             <value>1</value>
11             <value>Wäsch</value>
12             <value>Jürgen</value>
13             <value>Prof. Dr.-Ing.</value>
14             <value>Informatik</value>
15          </row>
16          <row>
17             <value>2</value>
18             <value>Haase</value>
19             <value>Oliver</value>
20             <value>Prof. Dr.</value>
21             <value>Informatik</value>
22          </row>
23      </table>
24      <table name="LEHRVERANSTALTUNG">
25          <column>id</column>
26          <column>professor_id</column>
27          <column>name</column>
28          <column>sws</column>
29          <column>ects</column>
30          <row>
31             <value>1</value>
32             <value>2</value>
33             <value>Verteilte Systeme</value>
34             <value>4</value>
35             <value>5</value>
36          </row>
37          <row>
38             <value>2</value>
39             <value>2</value>
40             <value>Design Patterns</value>
41             <value>4</value>
42             <value>3</value>
43          </row>

```

3.3. MODELLIERUNGSVARIANTEN DER TESTDATEN FÜR DBUNIT

```
44     </table>
45     ...
46 </dataset>
```

Listing 3.1: XML-Dataset

Die positiven Eigenschaften bei der Modellierung in XML sind unter anderem, dass für XML ein breites Angebot an Werkzeugen zur Verfügung steht. Diese können über den Dokumententyp prüfen, ob die Datei den Regeln entspricht.

Leider können die Werkzeuge kaum erkennen, ob in den einzelnen Zellen die richtigen Typen verwendet werden. Die in der XML-Datei enthaltenen Meta-Informationen (Beschreibung der Spalten, Zeilen 4-8 und 25-29) reichen dafür nicht aus. Die Meta-Informationen sind redundant und erschweren die Pflege.

Das Modellieren von Referenzen findet auf einer niedrigen Abstraktionsebene statt und ist damit unübersichtlich und fehleranfällig. Primär- und Fremdschlüssel müssen von Hand gepflegt werden. In umfangreicheren Datasets sind unkommentierte Beziehungen für Betrachter nur schwer nach zu vollziehen, da ein Schlüsselwert üblicherweise keinen unmittelbaren Rückschluss auf den referenzierten Datensatz erlaubt.

Ein großer Nachteil von XML-Datasets ist, dass der erwartete Datenbankzustand selbst wieder den kompletten Datenbankbestand umfassen muss. DbUnit erlaubt zwar mehrere Datasets zu einem zusammenzufassen, das Entfernen von Datensätzen ist darüber aber nicht möglich. Mehrere XML-Dateien mit ähnlichen, überwiegend sogar gleichen Daten, sorgen für ein hohes Maß an Redundanz.

Datasets in XML wachsen schnell in vertikaler Richtung und enthalten unter Umständen auch viel syntaktischen Overhead. Von den rund 30 gezeigten Zeilen enthalten nur zehn Zeilen wirkliche Daten bzw. drücken Beziehungen aus (Zeilen 21 und 26).

Es gibt noch eine zweite Möglichkeit, Datasets via XML zu erzeugen: Das *FlatXmlDataSet*. Hierbei gibt es keine vordefinierte DTD, da die Tags den Tabellen-Namen entsprechen¹. Eine solche XML-Datei kommt ohne explizite Meta-Informationen zu den Tabellen aus. Stattdessen stellen sie eine Art Sprachelement dar und werden für die Zuweisung der Werte verwendet. In diesem Punkt ist das *FlatXmlDataSet* übersichtlicher als das *XmlDataSet* (siehe Listing 3.2).

```
1  <?xml version='1.0' encoding='UTF-8'?>
2  <dataset>
3      <PROFESSOR id="1"
4          name="Wäsch"
5          vorname="Jürgen"
6          titel="Prof._Dr.-Ing."
7          fakultaet="Informatik" />
8      <PROFESSOR id="2"
9          name="Haase"
10         vorname="Oliver"
11         titel="Prof._Dr."
12         fakultaet="Informatik" />
13      <LEHRVERANSTALTUNG id="1"
14         professor_id="2"
15         name="Verteilte_Systeme"
16         sws="4"
17         ects="5" />
18      <LEHRVERANSTALTUNG id="2"
19         professor_id="2"
20         name="Design_Patterns"
21         sws="4"
22         ects="3" />
23      ...
```

¹Es ist möglich, eine eigene DTD zu definieren.

24 | </dataset>

Listing 3.2: Flat-XML-Dataset

Wie auch beim XmlDataSet sollte der Übersicht wegen für jeden Wert eine Zeile verwendet werden. Durch die fehlende Hierarchie wirkt das FlatXmlDataSet etwas unübersichtlich.

3.3.2 Default-Dataset

Um einige der Probleme zu vermeiden, die in Verbindung mit XML-Datasets auftreten, kann das Default-Dataset verwendet werden. Dieses lässt sich programmatisch, also dynamisch zur Laufzeit, erstellen. Durch die Nutzung von symbolischen Konstanten als Schlüsselwerte können Beziehungen ausdrucksstärker modelliert werden. Das Erzeugen des Datasets, das den nach einem Test erwarteten Datenbankzustand repräsentiert, bleibt umständlich, ist aber auf Java-Ebene mit weniger Redundanz lösbar.

```

1  DefaultTable professor = new DefaultTable(
2      "professor",
3      new Column[] {
4          new Column("id", DataType.INTEGER),
5          new Column("name", DataType.VARCHAR),
6          new Column("vorname", DataType.VARCHAR),
7          new Column("titel", DataType.VARCHAR),
8          new Column("fakultaet", DataType.VARCHAR),
9      }
10 );
11 professor.addRow(new Object[] {
12     Parameters.Professor.WAESCH_ID,
13     "Wäsch",
14     "Jürgen",
15     "Prof._Dr.-Ing.",
16     "Informatik",
17 });
18 professor.addRow(new Object[] {
19     Parameters.Professor.HAASE_ID,
20     "Haase",
21     "Oliver",
22     "Prof._Dr.",
23     "Informatik",
24 });
25 dataSet.addTable(professor);
26
27 DefaultTable lehrveranstaltung = new DefaultTable(
28     "lehrveranstaltung",
29     new Column[] {
30         new Column("id", DataType.INTEGER),
31         new Column("professor_id", DataType.INTEGER),
32         new Column("name", DataType.VARCHAR),
33         new Column("sws", DataType.INTEGER),
34         new Column("ects", DataType.INTEGER),
35     }
36 );
37 lehrveranstaltung.addRow(new Object[] {
38     Parameters.Lehrveranstaltung.VSYSTEME_ID,
39     Parameters.Professor.HAASE_ID,
40     "Verteilte_Systeme",
41     4,
42     5,
43 });
44 lehrveranstaltung.addRow(new Object[] {
45     Parameters.Lehrveranstaltung.DESIGN_PATTERNS_ID,
46     Parameters.Professor.HAASE_ID,
47     "Design_Patterns",
48     4,
49     3,
50 });
51 dataSet.addTable(lehrveranstaltung);

```

Listing 3.3: Default-Dataset

3.3. MODELLIERUNGSVARIANTEN DER TESTDATEN FÜR DBUNIT

Diese Umsetzung löst allerdings nicht alle Probleme. So müssen immer noch Meta-Informationen über die Tabellen modelliert werden (Zeilen 3-9 und 29-36). Obwohl diese sogar Typinformationen beinhalten, werden Typ-Fehler erst zur Laufzeit erkannt. Der Einsatz von symbolischen Konstanten erleichtert zwar die Pflege des Datasets, dennoch lassen sich Konstanten doppelt belegen oder auch Primärschlüssel einer falschen Datenbank als Fremdschlüssel angegeben werden.

Ähnlich wie für die Modellierung über XML-Dateien sind für eine übersichtliche Formatierung viele Zeilen notwendig und umfangreiche Datensets werden schnell unübersichtlich. Insgesamt bietet die Nutzung der Java-Datasets in dieser Art nur wenig Vorteile gegenüber den XML-Datasets.

3.3.3 SB-Testing-DB-DataSet

Die Bibliothek *SB Testing DB* der Firma *Seitenbau GmbH* versucht Nachteile der Xml- und Default-Datasets aufzufangen. Ein Generator erzeugt aus Meta-Informationen zu den Tabellen eine einfache Java-DSL. Über diese DSL lassen sich die Testdaten modellieren. Im Gegensatz zu DbUnit-Datasets unterliegt dieses Modell weniger strikten Einschränkungen in Bezug auf Modifikationen, und erlaubt auch das Löschen von Datensätzen. Um die modellierten Daten in Verbindung mit DbUnit zu verwenden, kann aus dem Modell ein DbUnit-Dataset erzeugt werden. Der Vorteil dieses zusätzlichen Modells ist, dass sich daraus verhältnismäßig einfach Varianten von DbUnit-Datasets erzeugen lassen, z.B. ein Dataset mit dem Ausgangszustand, und ein Dataset mit dem erwarteten Zustand am Ende des Tests. Die Java-DSL sorgt für Typsicherheit zur Compilierzeit². Die Syntax ist kompakter und dennoch ausdrucksstärker als bei beiden vorherigen Varianten.

```
1 table_Professor
2   .insertRow()
3     .setId(Parameters.Professor.HAASE_ID)
4     .setName("Haase")
5     .setVorname("Oliver")
6     .setTitel("Prof. Dr.")
7     .setFakultaet("Informatik")
8   .insertRow()
9     .setId(Parameters.Professor.WAESCH_ID)
10    .setName("Wäsch")
11    .setVorname("Jürgen")
12    .setTitel("Prof. Dr.-Ing.")
13    .setFakultaet("Informatik");
14
15 table_Lehrveranstaltung
16   .insertRow()
17     .setId(Parameters.Lehrveranstaltung.VSYSTEME_ID)
18     .setProfessorId(Parameters.Professor.HAASE_ID)
19     .setName("Verteilte_Systeme")
20     .setSws(4)
21     .setEcts(5)
22   .insertRow()
23     .setId(Parameters.Lehrveranstaltung.DESIGN_PATTERNS_ID)
24     .setProfessorId(Parameters.Professor.HAASE_ID)
25     .setName("Design_Patterns")
26     .setSws(4)
27     .setEcts(3);
```

Listing 3.4: SB Testing Dataset (1)

Die Modellierung von Referenzen stellt sich als ähnlich problematisch wie bei den bisherigen Java-Datasets dar (siehe Abschnitt 3.3.2). Nach wie vor wächst das Dataset vertikal in der Datei.

²Gängige Entwicklungsumgebungen wie Eclipse zeigen falsche Typen bereits während der Entwicklung an.

Zumindest das Problem mit den Referenzen kann durch eine Erweiterung auf M2-Ebene etwas entschärft werden. Ein um Beziehungen erweitertes Modell ermöglicht typischere Referenzen auf andere Entitäten (siehe Listing 3.5, Zeilen 20 und 27). Bei dieser Variante kann unter Umständen darauf verzichtet werden, Primärschlüssel manuell zu vergeben.

```

1 RowBuilder_Professor haase =
2   table_Professor
3     .insertRow()
4     .setName("Haase")
5     .setVorname("Oliver")
6     .setTitel("Prof._Dr.")
7     .setFakultaet("Informatik");
8 RowBuilder_Professor waesch =
9   table_Professor
10    .insertRow()
11    .setName("Wäsch")
12    .setVorname("Jürgen")
13    .setTitel("Prof._Dr.-Ing.")
14    .setFakultaet("Informatik");
15
16 RowBuilder_Lehrveranstaltung vsys =
17   table_Lehrveranstaltung
18     .insertRow()
19     .setName("Verteilte_Systeme")
20     .refProfessorId(haase)
21     .setSws(4)
22     .setEcts(5);
23 RowBuilder_Lehrveranstaltung design_patterns =
24   table_Lehrveranstaltung
25     .insertRow()
26     .setName("Design_Patterns")
27     .refProfessorId(haase)
28     .setSws(4)
29     .setEcts(3);

```

Listing 3.5: SB Testing Dataset (2)

Kapitel 4

Modellierung der Test-Daten

4.1 DSL-Entwürfe

4.1.1 Entwurf 1

Eine DSL, die sich stark an *SB Testing DB* orientiert, könnte wie folgt aussehen:

```
1 HAASE = professor {
2   name      "Haase"
3   vorname   "Oliver"
4   titel     "Prof._Dr."
5   fakultaet "Informatik"
6 }
7
8 WAESCH = professor {
9   name      "Wäsch"
10  vorname   "Jürgen"
11  titel     "Prof._Dr.-Ing."
12  fakultaet "Informatik"
13 }
14
15 VSYS = lehrveranstaltung {
16   name      "Verteilte_Systeme"
17   sws       4
18   ects      5
19 }
20
21 DPATTERNS = lehrveranstaltung {
22   name      "Design_Patterns"
23   sws       4
24   ects      3
25 }
26
27 ...
28
29 HAASE leitet VSYS
30 HAASE leitet DPATTERNS
31 HAASE beaufsichtigt P_DPATTERNS
32 WAESCH beaufsichtigt P_VSYS
33 ...
```

Listing 4.1: Mögliche DSL (1)

Diese DSL kommt ohne manuell vergebene ID-Nummern aus und verwendet Variablennamen für die Modellierung von Beziehungen. Da für jeden Wert eine eigene Zeile verwendet wird, werden umfangreiche Daten schnell unübersichtlich. Die Beschreibung der Beziehungen abseits der Definition der Daten erschwert den Umgang mit den Daten und die Übersicht ebenfalls.

4.1.2 Entwurf 2

Ein leicht abgewandelter Entwurf zeigt, wie sich die Beziehungen näher an den eigentlichen Daten beschreiben lassen könnten. An dem Problem, dass die Daten relativ schnell in vertikaler Richtung wachsen, ändert das jedoch nichts.

```

1 HAASE = professor {
2   name      "Haase"
3   vorname   "Oliver"
4   titel     "Prof._Dr."
5   fakultaet "Informatik"
6   leitet    VSYS, DPATTERNS
7   beaufsichtigt P_DPATTERNS
8 }
9
10 WAESCH = professor {
11   name      "Wäsch"
12   vorname   "Jürgen"
13   titel     "Prof._Dr.-Ing."
14   fakultaet "Informatik"
15   beaufsichtigt P_VSYS
16 }
17
18 VSYS = lehrveranstaltung {
19   name      "Verteilte_Systeme"
20   sws       4
21   ects      5
22 }
23
24 DPATTERNS = lehrveranstaltung {
25   name      "Design_Patterns"
26   sws       4
27   ects      3
28 }
29
30 ...

```

Listing 4.2: Mögliche DSL (2)

4.1.3 Entwurf 3

Der dritte Entwurf versucht die Daten durch eine tabellarische Struktur übersichtlich zu gestalten. Sie kommt mit wenig syntaktischem Ballast aus. Ein Label vor einer Tabelle drückt aus, welche Daten folgen (Zeilen 1 und 6). Die Tabelle selbst beginnt mit einer Kopfzeile, die die Spaltenreihenfolge beschreibt (Zeilen 2 und 7).

```

1 professor:
2 REF  || name | vorname | titel | fakultaet | leitet | beaufsichtigt
3 HAASE || "Haase" | "Oliver" | "Prof._Dr." | "Informatik" | VSYS, DPATTERNS | P_DPATTERNS
4 WAESCH || "Wäsch" | "Jürgen" | "Prof._Dr.-Ing." | "Informatik" | | P_VSYS
5
6 lehrveranstaltung:
7 REF  || name | sws | ects
8 VSYS || "Verteilte_Systeme" | 4 | 5
9 DPATTERNS || "Design_Patterns" | 4 | 3
10
11 ...

```

Listing 4.3: Mögliche DSL (3)

Der Entwurf sieht vor, dass Beziehungen innerhalb beider Entitätstypen ausgedrückt werden können. So kann eine Tabelle um Spalten für Beziehungen ergänzt werden, die in dieser Form nicht Teil des relationalen Modells (siehe Abb. 3.2) sind. Dazu gehören die Spalten „leitet“ und „beaufsichtigt“ der Professor-Tabelle. Erstere drückt die 1:n-Beziehung zu einer Lehrveranstaltung aus, letztere die m:n-Beziehung zu Prüfungen.

Probleme bzw. Nachteile in der Darstellung können auftreten, wenn die Länge der Werte in einer Spalte stark variiert. Die Spaltenbreite wird vom längsten Element bestimmt.

Der Entwickler ist selbst dafür verantwortlich, die übersichtliche Darstellung einzuhalten. Auf Tabulatoren sollte unter Umständen verzichtet werden, da sie von verschiedenen Editoren unterschiedlich dargestellt werden können. Bei vielen Spalten wächst diese Darstellung horizontal. Bei optionalen Spalten bzw. kaum genutzte Spalten kann die tabellarische Darstellung unübersichtlich werden.

Einige Entwicklungsumgebungen wie Eclipse bieten spezielle Block-Bearbeitungsfunktionen an, die beim Arbeiten an einer Tabellen-DSL hilfreich sein kann. So können beispielsweise in einer Spalte über mehrere Zeilen hinweg Leerzeichen eingefügt oder entfernt werden.

Zur besseren Übersicht kann es bei größeren Tabellen sinnvoll sein, den Tabellenkopf zu wiederholen.

Der Double-Pipe-Operator (`||`) soll die Spalte mit dem Entitätsidentifikatoren visuell von den Datenspalten trennen.

4.2 Wahl der DSL

Der dritte Entwurf zeigt, dass eine tabellarische Schreibweise viele Schwächen der anderen Varianten ausmerzt. Die Darstellung wirkt übersichtlich, da Tabellen ... **To do** (11)

4.3 Implementierungsvorbereitung

Da sich die DSL in die bisherige Werkzeug-Kette von Seitenbau integrieren lassen soll (siehe Abschnitt 3.1), sollte die DSL in Java nutzbar sein. Zwar kann eine DSL grundsätzlich auch in Java realisiert werden, doch die Möglichkeiten diesbezüglich sind relativ eingeschränkt und die DSL sieht immer noch nach Java aus. Es lassen sich allerdings auch andere Sprachen im Java-Umfeld nutzen. Eine davon ist *Groovy*. Groovy ist eine dynamisch typisierte Sprache¹, die direkt in Java-Bytecode übersetzt wird und damit auch in einer Java Virtual Machine ausgeführt wird. Sie teilt sich das Objekt-Modell mit Java, so dass aus Groovy heraus instantiierte Objekte auch in der Host-Anwendung nutzbar sind (und umgekehrt). Auch wenn Java-Code bis auf wenige Ausnahmen gültiger Groovy-Code und sich dort gleich verhält, enthält Groovy Techniken, die den Code mehr wie eine natürliche Sprache aussehen lassen. So kann oftmals auf die Semikolons am Ende einer Anweisung verzichtet werden, und auch auf das Einklammern von Parametern kann bei Methoden aufrufen verzichtet werden (wenn die Methode genau einen Parameter erwartet). Außerdem kann statt dem Punkt zwischen Objekt und Methode beim Aufruf verzichtet werden.

Listing 4.4 zeigt einen Befehl einmal in typischer Java-Syntax und einmal mit den Syntax-Vereinfachungen von Groovy:

```
1 myList.append("value_1").append("value_2");
2 myList append "value_1" append "value_2"
```

Listing 4.4: Vereinfachung von Ausdrücken in Groovy

Groovy hebt sich ferner durch die Möglichkeit Operatoren zu überladen und durch Closures (Funktionsabschlüsse) von Java ab. Ein Closure ist ein Codeblock, der wie eine Funktion

¹Im Gegensatz zu statisch typisierten Sprachen finden bei dynamisch typisierten Typ-Überprüfungen überwiegend zur Laufzeit statt.

aufgerufen und genutzt werden kann. In Java lassen sich Closures mit syntaktisch umfangreicheren Methoden-Objekten nachbilden. Ein Methoden-Objekt stellt eine Instanz einer (möglicherweise anonymen) Klasse dar, die nur eine Methode implementiert. [7, S. 40]
To do ⁽¹²⁾ Die Unterstützung zur Meta-Programmierung stellt sich beim Implementieren einer DSL ebenfalls als nützlich heraus. Dadurch ist es z.B. möglich, abgeschlossene Klassen innerhalb von Groovy um Methoden zu erweitern oder auf den Zugriff von nicht definierten Klassenelementen zu reagieren.

Aus diesen Gründen empfiehlt Ghosh in [5, S. 148] Groovy als Host für DSLs in Verbindung mit Java-Anwendungen.

4.3.1 Implementierungsvarianten

Eine DSL kann auf unterschiedliche Arten implementiert werden. Groovy bietet dafür zwei Möglichkeiten der Meta-Programmierung an: Laufzeit-Meta-Programmierung und Compiler-Zeit-Meta-Programmierung, letzteres in Form von AST-Transformationen. Beide Ansätze bieten individuelle Vorteile, die im folgenden diskutiert werden.

Laufzeit-Meta-Programmierung

Eine Möglichkeit, die DSL mit Hilfe von Laufzeit-Meta-Programmierung zu implementieren sieht eine Klasse zum Parsen von Closures vor, die eine Tabelle beinhalten. Diese Klasse, *TableParser*, enthält dafür die Methode `parseTableClosure`. Die Methode soll als Ergebnis eine Liste von Tabellenzeilen zurückliefern. Da an dieser Stelle noch keinerlei Interpretation der Tabellenwerte durchgeführt wird, stellt eine Tabellenzeile selbst ebenfalls eine Liste dar - aus den Objekten der Spalten.

Der Ansatz ist, Operator-Überladen für das Parsen zu verwenden. Soll ein binärer Operator² implementiert werden, ist die übliche Vorgehensweise in Groovy, die Klasse des linken Operanden um eine entsprechende Methode für den Operator zu erweitern. Diese Methode trägt einen vorgegebenen Namen und erwartet als binärer Operator den rechten Operanden als Parameter (eine Übersicht findet sich beispielsweise in [7, S. 58]).

Auch wenn sich dank der Möglichkeiten der Meta-Programmierung Klassen in Groovy zur Laufzeit um Methoden ergänzen lassen, ist dieses Vorgehen nicht empfehlenswert für um eine Tabelle zu parsen. Dieser wenig generische Ansatz müsste jeden in den Tabellen mögliche Datentyp berücksichtigen - kommen neue Datentypen hinzu, müsste der Code erweitert werden. **To do** ⁽¹³⁾

Groovy bietet allerdings auch eine zweite Möglichkeit für das Operator-Überladen an. Anstatt den Operator als Methode dem linken Operand (bzw. der Klasse) hinzuzufügen, wird er als statische Methode (in einer beliebigen Klasse) realisiert. Da eine statische Methode ohne Kontext ausgeführt wird, benötigt sie alle beteiligten Operanden als Parameter. Eine solche Methode wird als Kategoriemethode bezeichnet. Über das Schlüsselwort *use*³ können die Kategoriemethoden in einem Closure verwendet werden. [7, S. 192]

Listing 4.5 zeigt das Grundgerüst des Tabellenparsers:

²Binär bezogen auf die Verknüpfung zweier Werte und nicht auf das Zahlensystem

³*use* wird in der Literatur meistens als Schlüsselwort bezeichnet, tatsächlich handelt es sich jedoch um eine Groovy-Methode in `java.lang.Object`

```

1 class TableParser {
2
3     static or(self, arg) {
4         // ...
5     }
6
7     def parseTableClosure(Closure tableData){
8         use(TableParser) {
9             tableData()
10        }
11    }
12
13 }

```

Listing 4.5: Tabellen-Parser Grundgerüst mit Operator-Überladen

Die Methode `or` erwartet zwei Parameter vom Typ *Object*. Obwohl in Groovy alle Typen von *Object* abgeleitet sind, gibt es Oder-Ausdrücke, bei denen diese Methode nicht aufgerufen wird. Ein in der Klasse definierter Operator mit passenden Datentypen wird dieser allgemeinen Methode bevorzugt, z.B. bei zwei *Integer*-Werten. Doch auch solche Operationen lassen sich überschreiben, wenn für die Datentypen passende Kategoriemethoden definiert werden.

Der Parser in der Form kann noch nicht mit selbst definierten Variablennamen für die Abbildung von Referenzen umgehen. Aus diesem Grund wird eine Methode `getProperty` definiert, die für jeden Variablennamen in der Tabelle aufgerufen werden soll. Dazu muss der Ausführungskontext des Closures auf die Instanz des Tabellenparsers geändert werden. Die Änderungen sind in Listing 4.6 dargestellt.

```

1 class TableParser {
2
3     static or(self, arg) {
4         // ...
5     }
6
7     static or(Integer self, Integer arg) {
8         // ...
9     }
10
11    static or(Boolean self, Boolean arg) {
12        // ...
13    }
14
15    def getProperty(String property) {
16        // ...
17    }
18
19
20    def parseTableClosure(Closure tableData){
21        use(TableParser) {
22            tableData.delegate = this // Change closure's context
23            tableData.resolveStrategy = Closure.DELEGATE_FIRST
24            tableData()
25        }
26    }
27
28 }

```

Listing 4.6: Tabellen-Parser Grundgerüst mit Operator-Überladen

Die statischen Methoden haben keinen Zugriff auf Instanz-Variablen der Klasse *TableParser*. Ihre Ergebnisse können sie demnach auch nur in statische Elementen aufbewahren. Um die Klasse Thread-sicher zu machen, d.h. das gleichzeitige Parsen von Tabellen aus verschiedenen Threads heraus, wird für die Ergebnisse eine threadlokale Liste verwendet. To do (14) [6]

Die Laufzeit-Meta-Programmierung kann die Syntax der Sprache nicht beliebig erweitern. Groovy kennt keinen Double-Pipe-Operator. Deshalb kann dieser weder überladen noch über Laufzeit-Meta-Programmierung eingeführt werden. Folglich ist es nicht möglich, den dritten Entwurf über reine Laufzeit-Meta-Programmierung zu realisieren. Allerdings kann eine Syntax erreicht werden, die dem Entwurf sehr nahe kommt (siehe Listing 4.7). Ein Platzhalter (Unterstrich) verhindert Syntax-Fehler, wenn in einer Spalte kein Wert vorkommt (siehe Zeile 4, Spalte „leitet“). Der Platzhalter könnte auch verwendet werden, um einem Datensatz keinen Bezeichner für Referenzen zu zu weisen. Aus Sicht des Parsers stellt der Unterstrich eine Variable dar.

```

1  def fixture = [
2    professor: {
3      REF | name | vorname | titel | fakultaet | leitet | beaufsichtigt
4      WAESCH | "Wäsch" | "Jürgen" | "Prof._Dr.-Ing." | "Informatik" | _ | P_VSYS
5      HAASE | "Haase" | "Oliver" | "Prof._Dr." | "Informatik" | VSYS & DPATTERNS | P_DPATTERNS
6    },
7
8    lehrveranstaltung: {
9      REF | name | sws | ects
10     VSYS | "Verteilte_Systeme" | 4 | 5
11     DPATTERNS | "Design_Patterns" | 4 | 3
12   },
13
14   ...
15 ]

```

Listing 4.7: DSL-Entwurf 3 für Laufzeit-Meta-Programmierung angepasst

AST-Transformation

Die AST-Transformationen stellen ein mächtiges Werkzeug zur Erweiterung der Syntax der Sprache dar. Mit Hilfe der Transformationen ist es möglich, Änderungen am AST durchzuführen, bevor er in Java-Bytecode übersetzt wird.

Dass AST-Transformationen mehr syntaktische Möglichkeiten bieten, zeigt sich auch daran, dass hier der Double-Pipe-Operator verwendet werden kann. Außerdem können Labels erkannt werden und Daten einer Tabelle müssen nicht zwangsläufig in einem eigenen Block definiert werden.

Allerdings muss zum Auswerten einer Tabelle bei AST-Transformationen ein relativ großer Aufwand betrieben werden. Der Zugriff auf den AST erfolgt dabei über das Visitor-Pattern ([4, 331ff]).

4.3.2 Implementierungsentscheidung

Der Vergleich zwischen Laufzeit-Meta-Programmierung und AST-Transformation zeigt, dass sich Groovy als Host-Sprache für die DSL eignet. Die Laufzeit-Meta-Programmierung erlaubt zwar weniger Anpassungen an die Sprache, ist aber für die gewünschte DSL ausreichend und die Umsetzung einfacher.

4.4 Realisierung

Erweiterungen und andere Verbesserungen fließen nicht in die bisher genutzte Bibliothek SB Testing DB ein. Stattdessen wird der Quellcode dieser Bibliothek als Ausgangspunkt für das neue Projekt *STU* (Simple Test Utils, <https://github.com/Seitenbau/stu>) verwendet und unter Open-Source gestellt. Diese Vorgehensweise stellt gleichzeitig sicher, dass bisherige Schnittstellen weitergenutzt werden können, erlaubt aber auch punktuelle

Änderungen an diesen. Auf bisherige Projekte haben diese Anpassungen keinen Einfluss, neue Projekte können von den Verbesserungen profitieren. Änderungen an Schnittstellen, bei denen Abwärtskompatibilität verzichtet wird, betreffen vor allem die Builder-Klassen zum Beschreiben des Datenbankmodells für den Generator.

- <http://martinfowler.com/eaCatalog/gateway.html>
- <http://martinfowler.com/eaCatalog/repository.html>
- <http://martinfowler.com/eaCatalog/registry.html>

4.4.1 Builder-Klassen für das Datenbankmodell

Bei den Builder-Klassen für die Modellierung des zu Grunde liegenden Datenbankmodells wird auf Abwärtskompatibilität verzichtet. Während das alte API auf überladene Methoden mit vielen Parametern setzt, ist das neue API entsprechend dem Builder-Pattern umgesetzt (Quelle). So enthält das alte API neun Methoden zum Hinzufügen einer Spalte in einer Tabelle enthält, wovon eine als *deprecated* eingestuft ist. Dieses Design ist unübersichtlich und nur schwer erweiterbar. Jeder weitere optionale Parameter könnte die Anzahl der Methoden verdoppeln. Demgegenüber gibt es beim Builder-Pattern für jeden optionalen Parameter eine einzelne Set-Methode.

Die neuen Builder-Klassen decken den Funktionsumfang der alten API ab. So werden Flags für Spalten nicht mehr über ein *EnumSet* festgelegt, sondern über Methoden für die vordefinierten Flags. Abschnitt 4.4.2 geht weiter auf das Thema Flag ein. Darüber hinaus bieten die neuen Klassen die Möglichkeit, Beschreibungen zu Tabellen und Spalten hinzu zu fügen. Diese werden bei der Code-Generierung für die Erstellung von JavaDoc-Kommentaren verwendet (siehe 4.4.11).

To do (15)

4.4.2 Spalten-Flags

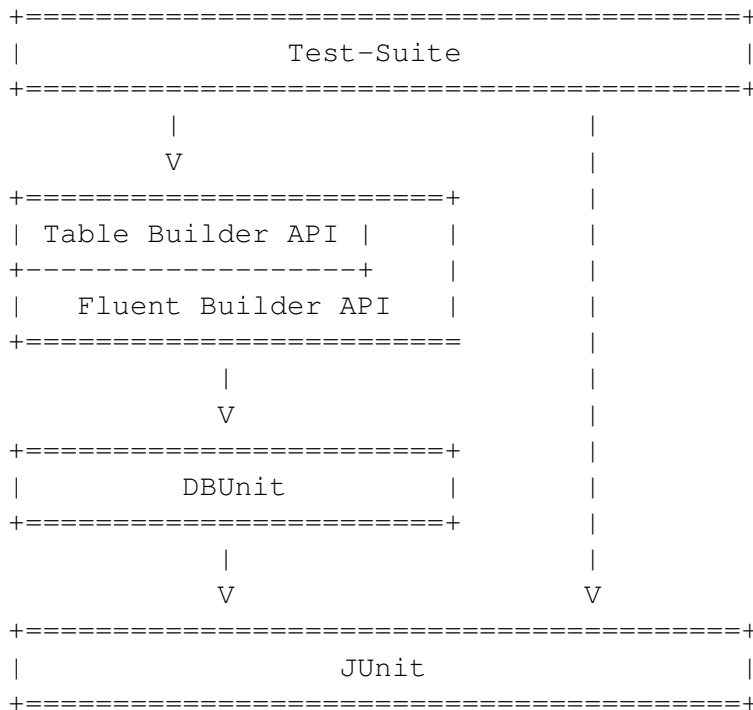
SB Testing DB sieht verschiedene Flags für Spalten vor, die in einem *Enum* zusammengefasst sind. Alle für eine Spalte gesetzten Flags müssen über ein *EnumSet* übergeben werden.

4.4.3 Architektur der generierten Klassen

Grundsätzlich erzeugt der Generator aus STU zwei APIs für die Modellierung von DataSets:

- **Fluent Builder API** (Java basiert)
- **Table Builder API** (Groovy basiert)

Die zusätzliche Table Builder API stellt eine zusätzliche Schicht über der bisherigen Fluent Builder API dar und nutzt diese selbst. Auf diese Weise können neue Funktionen ohne Berücksichtigung des Fluent Builder APIs hinzugefügt werden. Unter Umständen ist es allerdings sinnvoll oder sogar notwendig, Erweiterungen direkt in das Fluent Builder API zu integrieren. So müssen ursprünglich nur für das Table Builder API vorgesehene Features wie Referenzen und Scopes dem Fluent Builder API nicht vorenthalten werden.



Das Integrieren neuer Funktionen in die Fluent Builder API hat verschiedene Vorteile:

- **Code-Qualität:** Es gibt verschiedene Ansätze, Klassen um neue Funktionen zu erweitern oder ihr Verhalten zu ändern. Unabhängig davon, ob auf Vererbung oder Delegation gesetzt wird, werden neue Datentypen benötigt.

Soll die Schicht der Fluent Builder API nicht verändert werden, stellt Vererbung keine Option zur Erweiterungen von den Klassen dar, die von der Fluent-Builder-API-Schicht selbst instantiiert werden. Delegierende Klassen bestehen aus einer Vielzahl von Methoden, die nichts anderes machen, als die Aufgabe dem Delegat zu übergeben.

Wenn die Fluent-Builder-API-Schicht selbst geändert wird, gibt es keinen Grund, Vererbung zu verwenden. Die neuen Funktionen können direkt in die bisherigen Klassen integriert werden, vor allem wenn sie die bisherige Schnittstelle der Klasse nicht beeinflussen.

- **Mehrwert gegenüber SB Testing DB:** Auch wenn auf das neue Table Builder API verzichtet wird, bietet das Fluent Builder API einen Mehrwert gegenüber der bisherigen SB-Testing-DB-Implementierung.
- **Einheitliches Verhalten:** Beide APIs zeigen auf diese Weise ein einheitlicheres Verhalten.

4.4.4 Allgemeiner Tabellenparser

Der Tabellenparser basiert auf dem in Abschnitt 4.3.1 gezeigten Quellcode. - TableParser (Groovy) - TableParserContext - TableParserCallback - zeilenweise wegen Exceptions - Groovy-Anteil minimiert - Diagramm :-)

4.4.5 Builder für DataSet

4.4.6 Builder für Tabellen

4.4.7 ColumnBinding

4.4.8 Referenzen

- Referenzen (getter, Methoden für Beziehungen)

4.4.9 Scopes

- ScopeRegistry - IScope Interface (DataSetIdentifier)

4.4.10 Erweiterungen in generierter API

Zu den Erweiterungen in der Fluent-Builder-API-Schicht gehören:

- **RowBuilder:** - Beziehungen über REFs -
- **findWhere:** - REFs anstelle von konkreten Werten
- **getWhere:** - Zusätzlich zu findWhere
- **find:**
- **foreach:**
- **LazyValues:**

4.4.11 JavaDoc

To do (16)

Kapitel 5

Generieren von Testdaten

Fragen: - Wie muss das Modell „angereichert“ werden? - Wie können Daten sinnvoll generiert werden?

Kapitel 6

Proof of Concept

Kapitel 7

Zusammenfassung und Ausblick

Titel

Untertitel

Stichpunkte

- **Item 1:** Text
- **Item 2:** Text

Aufzählung

1. **Item 1:** Text
2. **Item 2:** Text

Abkürzung

Quellcode

1 Code

Listing 7.1: Der Titel

Verweise

1. siehe 7.1
2. siehe Listing 7.1
3. siehe Abb. 7.1
4. siehe Abschnitt 7
5. siehe Kapitel 7

Zitate

1. [9]
2. [11, 20ff]
3. [3]
4. [10]
5. [2]

Bild



Abbildung 7.1: Der Titel

Bildgruppe

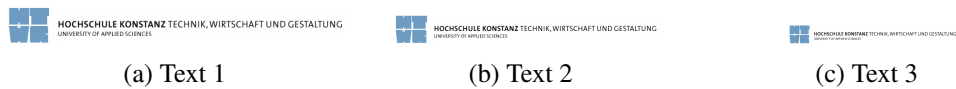


Abbildung 7.2: Gemeinsamer Titel

Abkürzungsverzeichnis

AST	Abstract Syntax Tree
BLOB	Binary Large Object)
GUI	Graphical User Interface (Grafische Benutzeroberfläche)
SUT	System Under Test (siehe Abschnitt 2.2)

Abbildungsverzeichnis

2.1	Modell-Beschreibung	5
2.2	Datenbank-Diagramm-Stil nach Ambler	6
3.1	ER-Diagramm des fortlaufenden Beispiels	10
3.2	Datenbank-Diagramm des fortlaufenden Beispiels	11
7.1	Der Titel	34
7.2	Gemeinsamer Titel	34

Listings

3.1	XML-Dataset	12
3.2	Flat-XML-Dataset	13
3.3	Default-Dataset	14
3.4	SB Testing Dataset (1)	15
3.5	SB Testing Dataset (2)	16
4.1	Mögliche DSL (1)	17
4.2	Mögliche DSL (2)	18
4.3	Mögliche DSL (3)	18
4.4	Vereinfachung von Ausdrücken in Groovy	19
4.5	Tabellen-Parser Grundgerüst mit Operator-Überladen	21
4.6	Tabellen-Parser Grundgerüst mit Operator-Überladen	21
4.7	DSL-Entwurf 3 für Laufzeit-Meta-Programmierung angepasst	22
7.1	Der Titel	33

Literatur

- [1] Scott W. Ambler und Pramod J. Sadalage. *Refactoring Databases, Evolutionary Database Design*. The Addison-Wesley Signature Series. Addison-Wesley, 2006. ISBN: 978-0-3212-9353-4. URL: <http://www.addison-wesley.de/main/main.asp?page=aktionen/bookdetails&ProductID=108888>.
- [2] Eric Evans. *Domain-driven Design: Tackling Complexity in the Heart of Software*. Addison Wesley Professional, 2004. ISBN: 978-0-321-12521-7. URL: <http://books.google.de/books?id=7dlaMs0SECsC>.
- [3] Martin Fowler. *Domain-Specific Languages*. The Addison-Wesley Signature Series. Addison-Wesley, 2010. ISBN: 978-0321712943. URL: <http://martinfowler.com/books/dsl.html>.
- [4] Erich Gamma u. a. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley, 1995. ISBN: 978-0-201-63361-0. URL: <http://books.google.de/books?id=6oHuKQe3TjQC>.
- [5] Debasish Ghosh. *DSLs in Action*. Manning, 2010. ISBN: 978-1-935182-45-0. URL: <http://www.manning.com/ghosh/>.
- [6] Brian Goetz. *Java concurrency in practice*. 7. print. Addison-Wesley, 2009. ISBN: 978-0-321-34960-6. URL: <http://www.gbv.de/dms/ilmeneau/toc/601225643.PDF>.
- [7] Dierk König. *Groovy im Einsatz*. Fachbuchverl. Leipzig im Carl-Hanser-Verl., 2007. ISBN: 978-3-446-41238-5. URL: http://deposit.d-nb.de/cgi-bin/dokserv?id=2948820&prov=M&dok_var=1&dok_ext=htm.
- [8] Gerard Meszaros. *XUnit Test Patterns: Refactoring Test Code*. The Addison-Wesley Signature Series. Addison-Wesley, 2007. ISBN: 978-0-13-149505-0. URL: <http://xunitpatterns.com/index.html>.
- [9] Andreas Spillner und Tilo Linz. *Basiswissen Softwaretest*. dpunkt.verlag, 2010. ISBN: 978-3-89864-642-0. URL: <http://www.dpunkt.de/buecher/4075.html>.
- [10] Thomas Stahl. *Modellgetriebene Softwareentwicklung: Techniken, Engineering, Management*. 2., aktualisierte und erw. Aufl. dpunkt-Verl., 2007. ISBN: 978-3-89864-448-8. URL: <http://www.gbv.de/dms/ilmeneau/toc/528370707.PDF>.
- [11] Mario Winter u. a. *Der Integrationstest*. Hanser, 2012. ISBN: 978-3-446-42564-4. URL: <http://www.dpunkt.de/buecher/4075.html>.

To do...

- ☐ 1 (p. 4): Grafik Back Door Manipulation
- ☐ 2 (p. 4): Layer test erklären
- ☐ 3 (p. 4): DataSets erklären
- ☐ 4 (p. 5): weiter erklären
- ☐ 5 (p. 6): Raw-Setter erklären?
- ☐ 6 (p. 6): einfaches Klassendiagramm für SB Testing DB generierte Klassen
- ☐ 7 (p. 9): Beispiel erweitern für 1:1-Beziehungen
- ☐ 8 (p. 9): Attribute einführen
- ☐ 9 (p. 9): Diagramm evtl in Chen-Notation
- ☐ 10 (p. 11): Legende
- ☐ 11 (p. 19): Hier wäre eine Quelle super, dass Menschen vertraut mit Tabellen sind
- ☐ 12 (p. 20): Quelle Kent Beck Smalltalk Best Practice Patterns
- ☐ 13 (p. 20): Mögliche ungewollte Seiteneffekte
- ☐ 14 (p. 21): thread local erklären mit quelle
- ☐ 15 (p. 23): Abhängigkeitsdiagramm der neuen Builder-Klassen?
- ☐ 16 (p. 25): „Muster“ für 1:1, 1:n und m:n