



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

Hochschule für Technik, Wirtschaft und Gestaltung

Fakultät Informatik

Studiengang 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

Ehrenwörtliche Erklärung	vii
1 Einleitung	1
2 Grundlagen	3
2.1 Fortlaufendes Beispiel	3
2.1.1 Voraussetzungen	3
2.1.2 Gewählte Problemstellung	3
2.1.3 Wahl der Testdaten	5
2.2 Modellierung der Testdaten in DbUnit	5
2.2.1 XML-Dataset	6
2.2.2 Java-Dataset	7
2.2.3 SB Testing DSL	8
2.3 Die DSL	9
2.3.1 Anforderungen an die DSL	9
2.3.2 Zielgruppe	10
2.3.3 DSL-Entwürfe	10
3 Zusammenfassung und Ausblick	11
4 Titel	13
4.1 Untertitel	13
4.1.1 Stichpunkte	13
4.1.2 Aufzählung	13
4.1.3 Abkürzung	13
4.1.4 Quellcode	13
4.1.5 Verweise	13
4.1.6 Zitate	14
4.1.7 Bild	14
4.1.8 Bildgruppe	14

Kapitel 1

Einleitung

Kapitel 2

Grundlagen

2.1 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.

2.1.1 Voraussetzungen

Der Schwerpunkt der Modellierung liegt bei der Darstellung von Beziehungen zwischen Entitäten. 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. *Standard Fixture* bedeutet, dass die selben Testdaten für alle Tests verwendet werden ([4, S. 317]). Nur in Ausnahmefällen sollten Tests modifizierte oder eigene Testdaten verwenden. Dadurch muss sich der Tester für verschiedene Tests nicht in unterschiedliche Daten hineinversetzen. Ein *Minimal Fixture* kann die Laufzeit der Tests verbessern, in dem das Fixture nur Daten enthält, die für den Test notwendig sind ([4, S. 302]). Da auf unnötige Testdaten verzichtet wird, kann der Tester die Daten leichter überblicken.

2.1.2 Gewählte Problemstellung

Das gewählte Beispiel stellt eine starke Vereinfachung des Prüfungswesens an Hochschulen dar. Auf eine praxisnahe Umsetzung wird zugunsten der Komplexität verzichtet. Es beinhaltet die folgenden vier Entitäten:

- **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.

Die Beziehungen der Entitäten stellen sich wie folgt dar: Eine Lehrveranstaltung muss von genau einem Professor geleitet werden, ein Professor kann beliebig viele (also auch keine) Lehrveranstaltungen leiten. 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 und Lehrveranstaltungen von beliebig vielen Studenten besucht werden. Die gleiche Beziehung gilt für Tutoren: Jeder Student kann bei beliebig vielen Lehrveranstaltungen Tutor sein und jede Lehrveranstaltung kann beliebig viele Tutoren haben. Schließlich kann jeder Student auch an beliebig vielen Prüfungen teilnehmen und umgekehrt eine Prüfung von einer beliebigen Anzahl von Studenten geschrieben werden.

Abbildung 2.1 zeigt die Problemstellung grafisch in Form eines ER-Diagramms. Sie veranschaulicht, dass es keine 1:1 Beziehung gibt. Eine 1:1-Beziehung kann jedoch als Spezialfall einer 1:n-Beziehung angesehen werden.

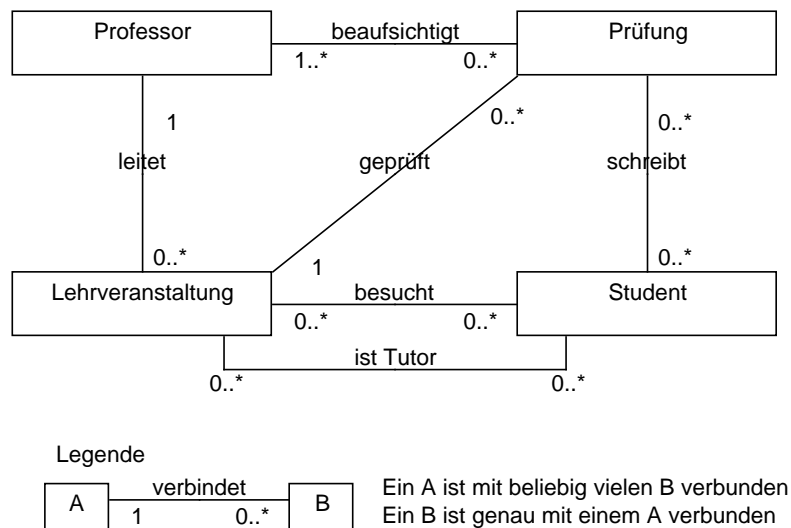


Abbildung 2.1: ER-Diagramm des fortlaufenden Beispiels

Das entsprechende relationale Modell sieht folgendermaßen aus:

2.2. MODELLIERUNG DER TESTDATEN IN DBUNIT

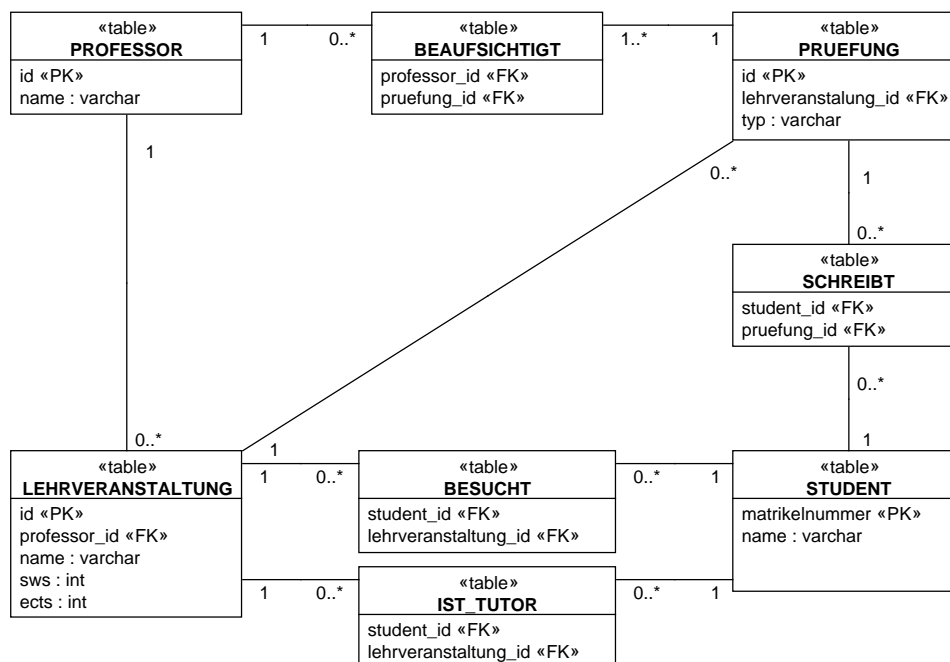


Abbildung 2.2: Relationales Modell des fortlaufenden Beispiels

2.1.3 Wahl der Testdaten

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)?

To do (1)

2.2 Modellierung der Testdaten in DbUnit

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. [4, 327ff]

To do (2)

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 von der Plattform.

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.

2.2.1 XML-Dataset

Eine Variante, ein Dataset für *DbUnit* zu modellieren, stellt XML dar. *DbUnit* bietet dazu die Klasse *XmlDataSet*, die eine XML-Datei nach einem vorgegebenen Dokumententyp einlesen kann. Das Listing 2.1 zeigt einen Ausschnitt einer solchen XML-Datei, in dem die beiden Tabellen *professor* und *lehrveranstaltung* definiert werden.

```

1  <!DOCTYPE dataset SYSTEM "dataset.dtd">
2  <dataset>
3    <table name="PROFESSOR">
4      <column>id</column>
5      <column>name</column>
6      <row>
7        <value>1</value>
8        <value>Jürgen Wäsch</value>
9      </row>
10     <row>
11       <value>2</value>
12       <value>Oliver Haase</value>
13     </row>
14   </table>
15   <table name="LEHRVERANSTALTUNG">
16     <column>id</column>
17     <column>professor_id</column>
18     <column>name</column>
19     <row>
20       <value>1</value>
21       <value>2</value>
22       <value>Verteilte Systeme</value>
23     </row>
24     <row>
25       <value>2</value>
26       <value>2</value>
27       <value>Design Patterns</value>
28     </row>
29   </table>
30   ...
31 </dataset>

```


Listing 2.1: 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 4-5 und 16-18) 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 angegeben werden, die Werte tauchen redundant auf und sind ohne Kommentare in umfangreicheren Datasets für Betrachter nur schwer nachzuvollziehen. Die Beziehungen werden auf einer sehr niedrigen Abstraktionsebene ausgedrückt und

Ein großer Nachteil von XML-Datasets ist, dass der erwartete Datenbankzustand selbst wieder den kompletten Datenbankbestand umfassen muss. Eine inkrementelle bzw. differentielle Definition der erwarteten Änderungen ist über XML-Datasets 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).

2.2.2 Java-Dataset

Um einige der Probleme zu vermeiden, die in Verbindung mit XML-Datasets auftreten, können Java-Datasets verwendet werden. Diese lassen sich dynamisch erstellen. Java-Datasets lösen einen Teil der Redundanz-Probleme, da symbolische Konstanten anstelle von numerischen verwendet werden können. Das Erzeugen des Datasets, das den 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     }
7 );
8 professor.addRow(new Object[] {
9     Parameters.Professor.WAESCH_ID,
10    "Jürgen_Wäsch"
11 });
12 professor.addRow(new Object[] {
13     Parameters.Professor.HAASE_ID,
14     "Oliver_Haase"
15 });
16 dataSet.addTable(professor);
17
18 DefaultTable lehrveranstaltung = new DefaultTable(
19     "lehrveranstaltung",

```

```

20     new Column[] {
21         new Column("id", DataType.INTEGER),
22         new Column("professorid", DataType.INTEGER),
23         new Column("name", DataType.VARCHAR),
24     }
25 );
26 lehrveranstaltung.addRow(new Object[] {
27     Parameters.Lehrveranstaltung.VERTEILTE_SYSTEME_ID,
28     Parameters.Professor.HAASE_ID,
29     "Verteilte_Systeme"
30 });
31 lehrveranstaltung.addRow(new Object[] {
32     Parameters.Lehrveranstaltung.DESIGN_PATTERNS_ID,
33     Parameters.Professor.HAASE_ID,
34     "Concurrency_and_Design_Patterns"
35 });
36 dataSet.addTable(lehrveranstaltung);

```

Listing 2.2: Java Dataset

Diese Umsetzung löst allerdings nicht alle Probleme. So müssen immer noch Meta-Informationen über die Tabellen modelliert werden. Obwohl diese sogar Typinformationen beinhalten, werden Typ-Fehler erst zur Laufzeit erkannt. Der Einsatz von symbolischen Konstanten verringert die Redundanz zwar und erleichtert damit die Pflege des Datasets, dennoch lassen sich Konstanten doppelt belegen und es können 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.

2.2.3 SB Testing DSL

To do (3) Generator erzeugt DSL auf Basis von Meta-Daten (M2).

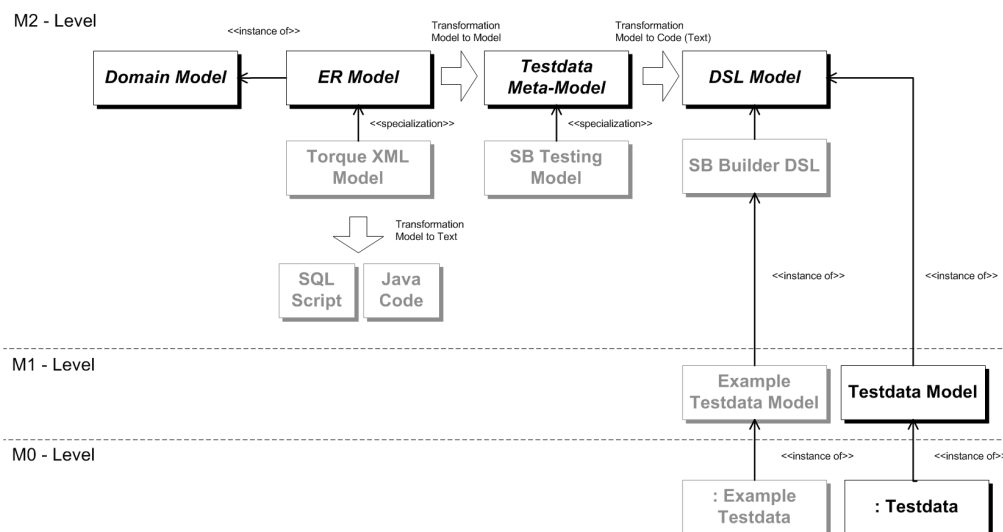


Abbildung 2.3: Modell-Beschreibung

```

1 table_Professor
2   .insertRow()
3     .setId(Parameters.Professor.HAASE_ID)
4     .setName("Oliver_Haase")
5   .insertRow()
6     .setId(Parameters.Professor.WAESCH_ID)
7     .setName("Jürgen_Wäsch");
8
9 table_Lehrveranstaltung
10  .insertRow()
11    .setId(Parameters.Lehrveranstaltung.
12           VERTEILTE_SYSTEME_ID)
13    .setProfessorId(Parameters.Professor.HAASE_ID)
14    .setName("Verteilte_Systeme")
15  .insertRow()
16    .setId(Parameters.Lehrveranstaltung.
17           DESIGN_PATTERNS_ID)
18    .setProfessorId(Parameters.Professor.HAASE_ID)
19    .setName("Design_Patterns");

```

Listing 2.3: SB Testing Dataset (1)

Die SB Testing DSL versucht die Nachteile der Java-Datasets zu umgehen. Die Syntax ist kompakter und dennoch ausdrucksstärker als die beiden vorherigen Varianten. Außerdem können falsche Typen bereits zur Compilierzeit erkannt werden.

Die Modellierung von Referenzen stellt sich als ähnlich problematisch wie bei den bisherigen Java-Datasets dar (siehe Abschnitt 2.2.2). Nach wie vor wächst das Dataset vertikal in der Datei.

Zumindest das Problem mit den Referenzen kann durch eine Erweiterung auf M2-Ebene etwas entschärft werden (siehe Listing 2.4).

```

1 RowBuilder_Professor haase =
2   table_Professor
3     .insertRow()
4     .setName("Oliver_Haase");
5 RowBuilder_Professor waesch =
6   table_Professor
7     .insertRow()
8     .setName("Jürgen_Wäsch");
9
10 RowBuilder_Lehrveranstaltung vsys =
11   table_Lehrveranstaltung
12     .insertRow()
13     .setName("Verteilte_Systeme")
14     .refProfessorId(haase);
15 RowBuilder_Lehrveranstaltung design_patterns =
16   table_Lehrveranstaltung
17     .insertRow()
18     .setName("Design_Patterns")
19     .refProfessorId(haase);

```

Listing 2.4: SB Testing Dataset (2)

2.3 Die DSL

2.3.1 Anforderungen an die DSL

Die Testdaten sollen in einer *Domain Specific Language* (DSL) beschrieben werden.

To do (4)

- Integration in bestehende Tool-Kette
- In Java nutzbar (interne DSL), Dataset soll zur Laufzeit erweiterbar bleiben in selber Syntax
- Zugriff auf Daten aus Java heraus
- „Meta-Informationen als Sprachelement“
- Symbolische Konstanten überflüssig machen
- „Dekorieren“ von Werten (before(date))
- wenig syntaktischer Ballast
- Typ-Prüfungen zu Compilierzeit
- übersichtliche Darstellung ermöglichen (v.a. vertikal)
- Namensräume/Scopes
- Beziehungen sollen sich einfach und typsicher modellieren lassen

2.3.2 Zielgruppe

2.3.3 DSL-Entwürfe

Kapitel 3

Zusammenfassung und Ausblick

Kapitel 4

Titel

4.1 Untertitel

4.1.1 Stichpunkte

- Item 1: Text
- Item 2: Text

4.1.2 Aufzählung

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

4.1.3 Abkürzung

4.1.4 Quellcode

```
1 Code
```

Listing 4.1: Der Titel

4.1.5 Verweise

1. siehe 4.1
2. siehe Listing 4.1
3. siehe Abb. 4.1
4. siehe Abschnitt 4.1
5. siehe Kapitel 4

4.1.6 Zitate

1. [2]
2. [3]
3. [1]
4. [5]
5. [6, 20ff]
6. [4]

4.1.7 Bild



Abbildung 4.1: Der Titel

4.1.8 Bildgruppe

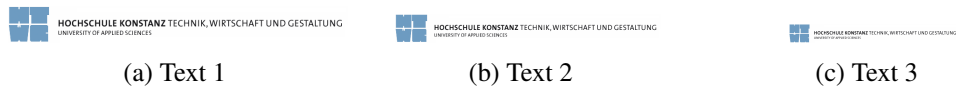


Abbildung 4.2: Gemeinsamer Titel

Abbildungsverzeichnis

2.1	ER-Diagramm des fortlaufenden Beispiels	4
2.2	Relationales Modell des fortlaufenden Beispiels	5
2.3	Modell-Beschreibung	8
4.1	Der Titel	14
4.2	Gemeinsamer Titel	14

Listings

2.1	XML Dataset	6
2.2	Java Dataset	7
2.3	SB Testing Dataset (1)	9
2.4	SB Testing Dataset (2)	9
4.1	Der Titel	13

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] Author. *Title*. Organisation. 2009. URL: <http://www.gidf.de> (besucht am 05.06.2011).
- [3] Debasish Ghosh. *DSLs in Action*. Manning, 2010. ISBN: 978-1-935182-45-0. URL: <http://www.manning.com/ghosh/>.
- [4] 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>.
- [5] Andreas Spillner und Tilo Linz. *Basiswissen Softwaretest*. dpunkt.verlag, 2010. ISBN: 978-3-89864-642-0. URL: <http://www.dpunkt.de/buecher/4075.html>.
- [6] 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. 5): Daten beschreiben
- ☐ 2 (p. 5): Grafik Back Door Manipulation
- ☐ 3 (p. 8): SB Testing DSL
- ☐ 4 (p. 9): Anforderungen