

Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt
Fakultät Informatik und Wirtschaftsinformatik

Bachelorarbeit

Über die Automatisierung der Entwicklung von Software Generatoren

**vorgelegt an der Hochschule für angewandte Wissenschaften
Würzburg-Schweinfurt in der Fakultät Informatik und Wirtschaftsinformatik zum
Abschluss eines Studiums im Studiengang Informatik**

René Ziegler

Eingereicht am: 22. Februar 2018

Erstprüfer: Prof. Dr. Peter Braun
Zweitprüfer: M.Sc. Tobias Fertig

Zusammenfassung

TODO

Abstract

TODO

Danksagung

Inhaltsverzeichnis

| | |
|--|----------|
| 1. Einführung | 1 |
| 1.1. Motivation | 2 |
| 1.2. Zielsetzung | 3 |
| 1.3. Aufbau der Arbeit | 3 |
| 2. Grundlagen | 5 |
| 2.1. Modellgetriebene Softwareentwicklung (MDSD) | 5 |
| 2.1.1. Domäne | 6 |
| 2.1.1.1. Definition | 7 |
| 2.1.1.2. Domänenanalyse | 7 |
| 2.1.1.3. Feature Modelling | 7 |
| 2.1.2. Metamodell | 8 |
| 2.1.2.1. Definition | 8 |
| 2.1.2.2. Abstrakte Syntax und Konkrete Syntax | 9 |
| 2.1.3. Domänenspezifische Sprache (DSL) | 9 |
| 2.1.3.1. Definition | 9 |
| 2.1.3.2. General Purpose Language (GPL) | 10 |
| 2.1.3.3. Interne DSLs | 10 |
| 2.1.3.4. Externe DSLs | 11 |
| 2.1.4. Parser | 11 |
| 2.1.4.1. Einlesen des Quelltextes | 12 |
| 2.1.4.2. Abstract Syntax Tree (AST) | 12 |
| 2.1.4.3. Weitere Verarbeitung des eingelesenen Quelltextes | 13 |
| 2.1.5. Code Generator | 15 |
| 2.1.5.1. Definition | 15 |
| 2.1.5.2. Techniken zur Generierung von Code | 15 |
| 2.1.5.3. Zusammenhang zu Transformatoren | 15 |
| 2.1.5.4. Abgrenzung zum Compilerbau | 15 |
| 2.2. Software Architektur | 15 |
| 2.2.1. Prinzipien der Softwaretechnik | 15 |
| 2.2.1.1. Abstraktion | 15 |
| 2.2.1.2. Bindung und Kopplung | 15 |
| 2.2.1.3. Modularisierung | 15 |
| 2.2.1.4. Weitere Prinzipien | 15 |
| 2.2.1.5. Abhängigkeiten | 15 |

| | | |
|-----------|--|-----------|
| 2.2.2. | Trennung durch Modelle | 15 |
| 2.2.3. | Modell-Transformations-Pipeline | 15 |
| 2.3. | Design Pattern objektorientierter Programmierung | 15 |
| 2.3.1. | Visitor Pattern | 15 |
| 2.3.1.1. | Zweck | 15 |
| 2.3.1.2. | Beschreibung | 15 |
| 2.3.1.3. | Anwendbarkeit | 15 |
| 2.3.2. | Builder Pattern und Fluent Interfaces | 15 |
| 2.3.2.1. | Zweck | 15 |
| 2.3.2.2. | Beschreibung | 15 |
| 2.3.2.3. | Anwendbarkeit | 15 |
| 2.3.3. | Factory Pattern | 15 |
| 2.3.3.1. | Zweck | 15 |
| 2.3.3.2. | Beschreibung | 15 |
| 2.3.3.3. | Anwendbarkeit | 15 |
| 3. | Analyse | 16 |
| 3.0. | Fragestellung | 16 |
| 3.1. | Aufwand eines konventionellen Softwareprojektes im Vergleich zu MDSD | 16 |
| 3.2. | Automatisierung der Entwicklung eines Code Generators | 16 |
| 3.2.1. | Java Code als Ausgangsmodell | 16 |
| 3.2.1.1. | Anreicherung des Java Codes mit Informationen | 16 |
| 3.2.1.2. | Parsen des Java Codes | 16 |
| 3.2.2. | Abstrakte Darstellung von Java Code als Modell | 16 |
| 3.2.2.1. | Anforderungen an das Modell | 16 |
| 3.2.2.2. | Schnittstellen zur Befüllung des Modells | 16 |
| 3.2.3. | Generierung von Java Code | 17 |
| 3.2.3.1. | Techniken zur Code Generierung | 17 |
| 3.2.3.2. | Vorhandene Frameworks zur Java Code Generierung | 17 |
| 4. | Konzept | 18 |
| 4.1. | Einsparung der Implementation des Code Generators durch einen Meta-Generator | 19 |
| 4.2. | Funktionsweise des Meta-Generators | 19 |
| 4.2.1. | Von Java Code zum Annotations-Modell | 19 |
| 4.2.1.1. | Annotationen als Mittel zur Informationsanreicherung | 19 |
| 4.2.1.2. | Parsen des annotierten Codes | 19 |
| 4.2.1.3. | Zweck des Annotations-Modells | 19 |
| 4.2.2. | Das CodeUnit-Modell | 19 |
| 4.2.2.1. | Baumstruktur des Modells | 19 |
| 4.2.2.2. | Spezialisierung durch ein Typ-Feld | 19 |
| 4.2.2.3. | Parametrisierung durch generische Datenstruktur | 19 |

| | | |
|-----------|--|-----------|
| 4.2.3. | Generierung von Buildern als interne DSL aus dem Annotations-Modell | 19 |
| 4.2.3.1. | Komposition der Builder aus benötigten Builder-Methoden | 19 |
| 4.2.3.2. | Übertagung vorgegebener Informationen in einen Builder | 19 |
| 4.2.3.3. | Verwendung von Plattform Code zur Generierung von vordefinierten CodeUnits | 19 |
| 4.2.3.4. | Auflösung von Referenzen in vordefinierten CodeUnits als nachgelagerter Verarbeitungsschritt | 19 |
| 4.2.4. | Erzeugung von Java Code aus einem befüllten CodeUnit-Modell . | 19 |
| 4.2.4.1. | Transformation des CodeUnit-Modells zum JavaFile-Modell | 19 |
| 4.2.4.2. | Erzeugung von Quelldateien | 19 |
| 5. | Lösung: Spectrum (Proof of Concept) | 20 |
| 5.1. | Verwendete Bibliotheken | 21 |
| 5.2. | Verwendeter Glossar | 21 |
| 5.2.1. | JavaParser mit JavaSymbolSolver | 21 |
| 5.2.2. | JavaPoet | 21 |
| 5.3. | Architekturübersicht | 21 |
| 5.3.1. | Amber | 21 |
| 5.3.1.1. | Annotationen | 21 |
| 5.3.1.2. | Parser | 21 |
| 5.3.1.3. | Modell | 21 |
| 5.3.2. | Cherry | 21 |
| 5.3.2.1. | Generator | 21 |
| 5.3.2.2. | Modell | 21 |
| 5.3.2.3. | Plattform | 21 |
| 5.3.2.4. | Generierte Builder | 21 |
| 5.3.3. | Jade | 21 |
| 5.3.3.1. | Transformator | 21 |
| 5.3.4. | Scarlet | 21 |
| 5.3.4.1. | Generator | 21 |
| 5.3.4.2. | Modell | 21 |
| 5.3.5. | Violet | 21 |
| 6. | Evaluierung | 23 |
| 6.1. | Kozept & Implementation | 23 |
| 6.1.1. | Amber | 23 |
| 6.1.2. | Cherry | 23 |
| 6.1.3. | Jade | 23 |
| 6.1.4. | Scarlet | 23 |
| 6.2. | Softwarequalität | 23 |
| 6.2.1. | Functionality | 23 |
| 6.2.2. | Maintainability | 23 |
| 6.2.3. | Performance | 23 |

Inhaltsverzeichnis

| | |
|--|-----------|
| 6.2.4. Usability | 23 |
| 6.3. Grenzen des Lösungsansatzes | 23 |
| 7. Abschluss | 24 |
| 7.1. Zusammenfassung | 24 |
| 7.2. Ausblick | 24 |
| A. Dokumentation | 25 |
| A.1. Verwendung der Annotationen | 25 |
| A.2. Verwendung der generierten CodeUnit-Builder | 25 |
| A.3. Klassendokumentation | 25 |
| A.3.1. Amber | 25 |
| A.3.2. Cherry | 25 |
| A.3.3. Jade | 25 |
| A.3.4. Scarlet | 25 |
| Verzeichnisse | 26 |
| Literatur | 29 |
| Eidesstattliche Erklärung | 30 |
| Zustimmung zur Plagiatsüberprüfung | 31 |

1. Einführung

Als Henry Ford 1913 die Produktion des Modell T, umgangssprachlich auch Tin Lizzie genannt, auf Fließbandfertigung umstellte, revolutionierte er die Automobilindustrie. Ford war nicht der erste, der diese Form der Automatisierung verwendete. Bereits 1830 kam in den Schlachthöfen von Chicago eine Maschine zum Einsatz, die an Fleischerhaken aufgehängte Tierkörper durch die Schlachtereie transportierte. Bei der Produktion des Oldsmobile Curved Dash lies Ransom Eli Olds 1910 erstmals die verschiedenen Arbeitsschritte an unterschiedlichen Arbeitsstationen durchführen. Fords Revolution war die Kombination beider Ideen. Er entwickelte eine Produktionsstraße, auf welcher die Karossen auf einem Fließband von Arbeitsstation zu Arbeitsstation befördert wurden. An jeder Haltestelle wurden nur wenige Handgriffe von spezialisierten Arbeitern durchgeführt [14].

Fords Vision war es, ein Auto herzustellen, welches sich Menschen aller Gesellschaftsschichten leisten konnten. Durch die Reduktion der Produktionszeit der Tin Lizzie von 12,5 Stunden auf etwa 6 Stunden konnte Ford den Preis senken. Kostete ein Auto des Model T vor der Einführung der Produktionsstraße 825\$, erreichte der Preis in den Jahren danach einen Tiefststand von 259\$ [13]. Setzt man diesen Preis in ein Verhältnis mit dem durchschnittlichen Einkommen in den USA, das 1910 bei jährlich 438\$ lag, kann man sagen, dass Fords Traum durch die eingesetzten Techniken Realität wurde [10].

Im Zuge der weiteren Entwicklung der Robotik wurden immer mehr Aufgaben, die bisher von Menschen am Fließband durchgeführt wurden, von Automaten übernommen. In der Automobil-Industrie war General Motors der erste Hersteller, bei welchem die Produktionsstraßen im Jahr 1961 mit 66 Robotern des Typs Unimation ausgestattet wurden. Bis zur Erfindung des integrierten Schaltkreises in den 1970ern waren die Roboter ineffizient. Der Markt für industrielle Roboter explodierte jedoch in den Folgejahren. Im Jahr 1984 waren weltweit ungefähr 100.000 Roboter im Einsatz [2, 18].

Die industrielle Revolution prägte die Autoindustrie: von der Erfindung auswechselbarer Teile 1910 bei Ransom Olds, über die Weiterentwicklung des Konzepts unter der Verwendung von Fließbändern bei Ford im Jahr 1913, bis hin zur abschließenden Automatisierung mit Industriellen Robotern in den frühen 1980ern [2].

1.1. Motivation

„If you can compose components manually, you can also automate this process.“

Das hervorgehobene Zitat nennen Czarnecki und Eisenecker die Automation Assumption. Diese allgemein gehaltene Aussage lässt die Parallelen, die die beiden Autoren zwischen der Automatisierung der Automobilindustrie und der automatischen Code Generierung sehen, erkennen. Dafür müssten die einzelnen Komponenten einer Softwarefamilie derart gestaltet werden, dass diese austauschbar in eine gemeinsame Struktur integriert werden können. Des weiteren müsste klar definiert sein, welche Teile eines Programms konfigurierbar seien und welche der einzelnen Komponenten in welcher Konfiguration benötigt werden. Setzt man dieses definierte Wissen in Programmcode um, könnte ein solches Programm eine Software in einer entsprechenden Konfiguration generieren [2].

Konkret bedeutet dies, dass entweder eine vorhandene Implementierung in Komponenten zerlegt werden muss oder eine für die Zwecke der Codegenerierung vorgesehene Referenzimplementierung geschrieben wird. Codeabschnitte, die in Ihrer Struktur gleich sind, sich jedoch inhaltlich unterscheiden, müssen formal beschrieben werden [16]. Ein solches abstraktes Modell wird dann mit Daten befüllt. Schlussendlich wird ein Generator implementiert, der den Quellcode für unterschiedliche Ausprägungen eines Programms einer Software-Familie, auf Basis des konkreten Modells, generieren kann [3].

Sowohl bei der Umsetzung von einzigartigen Anwendungen, als auch bei der Verwirklichung von Software mit mehreren Varianten, kann die Verwendung von bereits verfügbaren Code Generatoren oder die Entwicklung eigener Code Generatoren vorteilhaft sein. Die Entwicklungsgeschwindigkeit könnte erhöht, die Softwarequalität gesteigert und Komplexität durch Abstraktion reduziert werden [16]. Allgemein wird weniger Zeit benötigt, um eine größere Vielfalt an ähnlichen Programmen zu entwickeln [2].

Bisher müssen fast alle Teilaufgaben bei der Umsetzung eines Code Generators manuell durchgeführt werden. Werkzeuge wie Language Workbenches können Code bis zu einem gewissen Grad automatisiert generieren oder interpretieren. Sie haben aber in erster Linie die Aufgabe, den Entwickler beim Design von externen domänenspezifischen Sprachen zu unterstützen und dienen als Entwicklungsumgebung für die Arbeit mit der Sprache [3].

Soll ein Projekt Modellgetrieben entwickelt werden, so lohnt sich dies wirtschaftlich gesehen erst, wenn auf Basis des entwickelten Modells mehrere Programme entwickelt wurden []. Einer der Teilschritte dieses Entwicklungsprozesses ist die Planung und Implementation des Code Generators. Durch ihre hohe Komplexität, ist diese Aufgabe sehr zeitaufwendig [].

ÜBERARBEITEN

1.2. Zielsetzung

In dieser Arbeit soll untersucht werden, ob und wie die Entwicklung eines Codegenerators automatisiert werden kann. Eine zusätzliche Ebene der Indirektion könnte das komplexe Thema der Modellgetriebenen Softwareentwicklung weiter vereinfachen und somit Code Generierung für mehr Entwickler zugänglicher und somit auch wirtschaftlicher machen.

Im speziellen wird analysiert, wie ein Meta-Generator zur Erhöhung der Wirtschaftlichkeit modellgetriebener Softwareentwicklung umgesetzt werden könnte. Zu diesem Zweck wird eine beispielhafte Java Anwendung erarbeitet, welche es ermöglichen soll, vorhandenen Java Quelltext so mit Informationen anzureichern, dass aus diesem ein Meta-Generator abgeleitet werden kann.

1.3. Aufbau der Arbeit

Die sieben Kapitel dieser Bachelorarbeit versuchen den Leser Stück für Stück an das komplexe Thema der Meta-Generierung heranzuführen. Da es nicht möglich ist im Rahmen einer vergleichsweise kurzen Thesis wie dieser sämtliche Grundlagen der Informatik zu beschreiben, wird ein solides Fundament aus Vorwissen, wie man es zum Beispiel in einem Bachelorstudium erwerben kann, vorausgesetzt.

Eingeleitet wird diese Arbeit mit einem Kapitel zur Motivation und Zielsetzung, in welchem aufgezeigt werden soll warum es sinnvoll ist, sich mit dem Thema der modellgetriebene Softwareentwicklung zu beschäftigen.

Das zweite Kapitel behandelt die erweiterten, grundlegenden Kenntnisse, die zum Verständnis des Textes notwendig sind. Hier wird sowohl auf die modellgetriebene Softwareentwicklung, als auch auf die Software Architektur und die Verwendung von Design Pattern objektorientierter Programmierung eingegangen.

Aufbauend auf dem vorhergehenden Abschnitt sollte das dritte Kapitel, die Analyse, gut verständlich sein. Hier wird zuerst der wirtschaftliche Aufwand konventioneller Softwareprojekte mit dem Aufwand von Projekten welche ein modellgetriebenen Ansatz verfolgen verglichen. Danach werden die Probleme der einzelnen Teilschritte bei der Automatisierung der Entwicklung des Codegenerators untersucht.

1. Einführung

Das Konzept Kapitel erläutert nun den im Proof-of-Concept verfolgten Lösungsansatz der analysierten Probleme. Zur Veranschaulichung kommen hier lediglich Diagramme zum Einsatz, dadurch sollte es möglich sein die allgemeine Idee hinter der Implementation leichter zu verstehen und den Ansatz losgelöst von der Umsetzung weiter zu verfolgen.

Im fünften Kapitel werden jetzt zum einen die verwendeten externen Bibliotheken kurz vorgestellt, zum anderen wird genau auf die Architektur der entwickelten Anwendung eingegangen. Die Funktionsweise und der Aufbau jeder Programmkomponente wird anhand von Quelltext Auszügen genau erläutert, mithilfe von Beispielen wird die Verwendung der einzelnen Einheiten demonstriert.

Das vorletzte Kapitel der Arbeit evaluiert sowohl das Konzept als auch die Umsetzung aller Module im Detail. Außerdem wird die allgemeine Softwarequalität des entwickelten Generators nach bewährten Kriterien untersucht. Besonders umfangreich werden die Grenzen des Lösungsansatzes diskutiert und mögliche Antworten auf die hieraus entstehenden Fragen angesprochen.

Mit dem siebten Kapitel werden Inhalt und Erkenntnisse der Arbeit noch einmal in gebündelter Form zur Verfügung gestellt. Ein ausführlicher Ausblick soll die vielen möglichen Anknüpfungspunkte dieser Thesis aufzeigen.

2. Grundlagen

Um die grundlegenden Zusammenhänge zu verstehen wird im folgenden auf die verschiedenen Aspekte und Teilschritte der modellgetriebenen Softwareentwicklung eingegangen. Begriffe und Konzepte die in der Arbeit zur Anwendung kommen, werden eingeführt und definiert. Danach wird auf einige Grundlagen der Softwarearchitektur, vor allem auf die Prinzipien der Softwaretechnik eingegangen. Da in der Implementation auf einige Design Pattern objektorientierter Programmierung zurückgegriffen wird, werden deren Zweck, Aufbau und Anwendbarkeit abschließend in diesem Kapitel erläutert.

2.1. Modellgetriebene Softwareentwicklung (MDSD)

„Modellgetriebene Softwareentwicklung (Model Driven Software Development, MDSD) ist ein Oberbegriff für Techniken, die aus formalen Modellen automatisiert lauffähige Software erzeugen.“

Die obige Definition stammt aus dem Buch modellgetriebene Softwareentwicklung von Thomas Stahl und Markus Völter [16, S. 11]. Sie lässt sich gut erläutern wenn man sie in drei Teile zerlegt.

Zunächst einmal wäre dort der Ausdruck "formales Modell". Damit ist ein Modell gemeint welches einen Teil einer Software vollständig beschreibt. Jedoch soll das nicht heißen, dass dieses Modell allumfassend ist, sondern was genau von diesem Modell beschrieben wird muss eindeutig reguliert sein [16, S. 11f.]. Weiterhin bezieht sich die Definition darauf dass lauffähige Software erzeugt wird. D. h. wird das formale Modell nur zur Dokumentation verwendet oder dient es als Information zur händischen Umsetzung, so kann man das laut Stahl und Völter nicht als modellgetriebene Softwareentwicklung bezeichnen [16, S. 12]. Der letzte Teil der Definition den die Autoren explizit erläutern ist, dass die Umwandlung von Modell zu ausführbar Software automatisiert erfolgen soll. Insbesondere soll der Quelltext nicht nur einmal generiert und dann manuell verändert und weiterentwickelt werden, sondern das Modell soll anstelle des Quelltextes treten. Der Quelltext wird aus den geänderten Modellen generiert, dadurch kann aktueller und einheitlicher Quellcode gewährleistet werden [16, S. 13].

2. Grundlagen

In der Literatur findet sich auch die alternative Bezeichnung Model Driven Development (MDD) [15, 1]. Diese Thesis wird jedoch durchgängig die Bezeichnung MDSD verwenden.

Auch der Model Driven Architecture (MDA) Ansatz der Object Management Group (OMG) beschäftigt sich mit modellgetriebener Softwareentwicklung [8]. Dieser Ansatz beschreibt detailliert und umfassend den Gesamtprozess von der Analyse bis hin zur Implementation und führt eigene Standards ein.

Czarnecki und Eisenecker verwenden den Ausdruck Generative Programming, definieren ihn jedoch in den wesentlichen Punkten vergleichbar zu MDSD. Wobei Generative Programming nach Czarnecki und Eisenecker nach vollständiger Automation strebt und ein vollständiges Zwischen- oder Endprodukt erzeugen soll [2, S. 5]. Für MDSD merken Stahl und Völter an, dass die verwendeten Modelle nicht unbedingt das vollständige System abbilden, ein komplettes System enthalte sowohl manuell implementierte als auch automatisch generierte Anteile [16, S. 13].

Durch MDSD soll, wie bereits in der Einleitung dieser Arbeit beschrieben, die Qualität der entstandenen Software gesteigert werden. Dies wird durch den resultierenden einheitlichen Code und die erhöhte Wiederverwertbarkeit erreicht. Außerdem kann potentiell mithilfe der zusätzlichen Abstraktion eine erhöhte Entwicklungsgeschwindigkeit erzielt werden. Ein Bonus von MDSD ist es, dass die Software immer durch aktuelle Modelle beschrieben und somit zumindest in Teilen dokumentiert wird [16, S. 13ff.]. Die Vorteile von Generative Programming [2, S. 13ff.] und MDA [8] werden sehr Ähnlich beschrieben.

2.1.1. Domäne

Bei der modellgetriebenen Softwareentwicklung sind, wie bereits erwähnt, Modelle Dreh und Angelpunkt des Entwicklungsprozesses. Um jedoch ein Modell bilden zu können muss zuerst untersucht werden was mit diesem Modell abgebildet wird. Diese Untersuchung ist Teil des Domain Engineerings.

Domain Engineering umfasst die Analyse das Design und die Implementation einer Domäne. Design und Implementation beziehen sich bereits auf die Zusammensetzung des Systems und dessen technische Umsetzung [2, S. 21f.]. Zur Modellbildung ist für diese Arbeit vor allem die Domänenanalyse interessant.

2.1.1.1. Definition

Eine Domäne beinhaltet laut Czarnecki und Eisenecker das fachliche Wissen über ein Problem-Themengebiet. Jedoch geht die Domäne im Domain Engineering noch darüber hinaus. Sie umfasst nicht nur das fachliche Wissen, sondern auch Informationen darüber wie Anwendungen für diesen Themenbereich aufgebaut sind. Wichtig ist hierbei, dass die Domäne sich aus dem übereinstimmenden Wissen der beteiligten Stakeholder ergibt. Wobei hier alle Personen die ein Interesse an einer bestimmten Domäne haben als Stakeholder bezeichnet werden [2, S. 33].

2.1.1.2. Domänenanalyse

Bereits im 1990 erschienenen Paper Domain Analysis: An Introduction zählt der Autor Rubén Prieto-Díaz die drei Grundschrte zur Analyse einer Domäne auf [12]:

1. identification of reusable entities
2. abstraction or generalization
3. classification and cataloging for further reuse

Diese drei Punkte sind auch heute noch zutreffend. In der Analyse wird zunächst beim Domain Scoping die Domäne eingegrenzt. Dies ist notwendig um eine konsistente Umsetzung möglich zu machen und alle Anforderungen konkret und klar zu definieren [16, S. 239]. Danach werden im allgemeinen Entitäten, Operationen und Beziehungen zwischen diesen identifiziert. Sobald dies geschehen ist, können diese auf Ähnlichkeiten und Unterschiede versucht werden. Auf Basis der gefundenen Gemeinsamkeit bzw. Variabilitäten kann nun abstrahiert und generalisiert werden. Festgehalten werden diese Informationen in einem umfassenden Domänenmodell [2, S. 24ff.].

Als Domänenmodell bezeichnen Czarnecki und Eisenecker eine ausdrückliche Darstellung aller Gemeinsamkeiten und Variabilitäten eines Systems in einer Domäne, der Bedeutung dieser Domänenkonzepte und den Abhängigkeiten zwischen den unterschiedlichen Eigenschaften [2, S. 23f.].

2.1.1.3. Feature Modelling

Neben der Domänendefinition, dem Domänenlexikon welches den verwendeten Glossar auflistet und konzeptionellen Modellen können auch Featuremodelle Teil des Domänenmodells sein.

2. Grundlagen

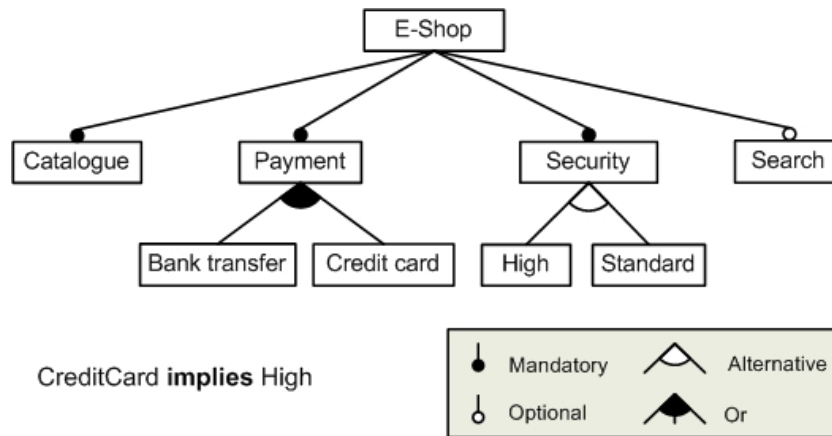


Abbildung 2.1.: Darstellung eines Featurediagramms für ein konfigurierbares E-Shop System von Segura09 in der Wikipedia auf Englisch (Transferred from en.wikipedia) [Public domain], via Wikimedia Commons

Das Feature Modelling erlaubt es Gemeinsamkeiten und Variabilitäten von Systemen zu dokumentieren. Neben einer genauen Beschreibung aller Features in schriftlicher Form kann man Zusammenhänge und Regeln zur Komposition der Features in einem Featu-rediagramm darstellen.

Ursprünglich wurde das Featurediagramm als Teil der Feature-Oriented Domain Analysis (FODA) definiert [5]. Es ermöglicht Merkmale hierarchisch darzustellen und erlaubt es Beziehungen zwischen ihnen genauer festzulegen. Weiterhin kann unterschieden werden ob genau eine oder mehrere Eigenschaften der Alternativen darf bzw. dürfen. Zudem können noch weitere Einschränkungen definiert werden welche es erlauben auch weiter voneinander entfernte Knotenpunkte zueinander in Beziehung zu setzen [16, S. 240f.]. Die Abbildung 2.1 stellt oben genannte Features beispielhaft dar.

2.1.2. Metamodell

Die Analyse des Problemfelds und Modellierung der Merkmale soll in erster Linie die Grundlage für die Definition des Metamodells der Domäne liefern [16, S. 200].

2.1.2.1. Definition

Zur Beschreibung welche Mittel zur Definition eines bestimmten Modells zur Verfügung stehen kann man wiederum ein Modell definieren. Dieses, über dem beschriebenen Modell stehende Modell, nennt man Metamodell. Stahl und Voelter bezeichnen es als Beschreibung der möglichen Struktur von Modellen. Dazu zählen sie die Konstrukte der

2. Grundlagen

Modellierungssprache, auf welche Weise diese zueinander in Beziehung stehen, sowie vorhandene Regeln bezüglich der Gültigkeit bzw. Modellierung [16, S. 59].

Ein Unified Modeling Language 2.0 (UML) Diagramm welches alle Java Sprachkonzepte, wie beispielsweise die Existenz von Klassen, Attributen und Instanzen, beschreibt, ist eine Darstellung für das Java Metamodell. Geschriebener Java Quelltext wäre somit eine Instanz dieses Modells, kann aber wiederum instanziiert werden. Handelt es sich nämlich um eine Definition einer Java Klasse, so können hiervon Objekte mit konkreten Werten existieren.

Es ist noch hervorzuheben, dass mit dem Begriff Metamodell nicht einfach nur eine Abstraktion gemeint ist. Die Silbe Meta kann hier als „die Definition von“ verstanden werden. Ein Modell kann ein anderes Modell abstrahieren in dem es beispielsweise Informationen auslässt, falls diese für ein Anwendungsfall nicht relevant oder implizit sind. Das ist im obigen Sinne dann jedoch noch kein Metamodell. Zu sagen ein Metamodell ist das Modell eines Modells, ist also nur bedingt sinnvoll [17, S. 27].

2.1.2.2. Abstrakte Syntax und Konkrete Syntax

Die in diesem Beispiel definierten Sprachkonzepte sind die abstrakte Syntax für Java. In Abgrenzung hierzu dient eine konkrete Syntax zur Beschreibung eines Modells. Möchte man einen Datentypen definieren, könnte man das mit einem UML Klassendiagramm. Da man den Datentypen auch mit Java-Code beschreiben könnte, sieht man, dass es für dasselbe Metamodell mehr als eine konkrete Syntax geben kann. [16, S. 59f.].

2.1.3. Domänenspezifische Sprache (DSL)

Wurde das Metamodell für eine Domäne definiert, also ist die abstrakte Syntax bekannt, wird eine Möglichkeit benötigt eine Instanz des Metamodells zu bilden. Durch anreichern des Modells mit Informationen wird eine konkrete Ausprägung des Metamodells geschaffen. Das Mittel hierfür ist eine domänenspezifische Programmiersprache.

2.1.3.1. Definition

Stahl und Völter definieren eine solche Programmiersprache mit: [16, S. 30]. „Eine DSL ist nichts anderes als eine Programmiersprache für eine Domäne.“ Martin Fowler hat in seinem Buch zu diesem Thema eine umfangreichere Definition geschaffen und erläutert diese mit vier Schlüsselementen [3, S. 27f.].

2. Grundlagen

„Domain-specific Language (noun): a computer programming language of limited expressiveness focused on a particular domain.“

Der Ausdruck „Computer programming language“ bedeutet, dass eine DSL vom Menschen verwendet wird um einem Computer Befehle zu geben. Also sollte sie wie jede moderne Programmiersprache durch ihre Struktur sowohl vom Menschen leicht verständlich als auch vom Computer ausführbar sein.

Laut Fowler ist ein weiteres Element die „Language nature“. Als Programmiersprache sollte die Ausdrucksstärke einer DSL nicht nur von den einzelnen Ausdrücke kommen, sondern auch wie die Ausdrücke zusammengesetzt werden.

Mit „Limited expressiveness“ meint Fowler, dass eine DSL im Vergleich zu allgemeinen Programmiersprachen nur das für die Beschreibung der Domäne notwendige Minimum an Funktionalität aufweist.

Im letzten Teil seiner Ausführung zu dieser Definition geht Fowler auf den „Domain focus“ ein. Eine beschränkte Sprache sei nur sinnvoll wenn sie klar auf eine Domäne eingegrenzt ist.

DSLs können in zwei Hauptkategorien unterteilt werden, bezeichnet als interne und externe DSLs.

2.1.3.2. General Purpose Language (GPL)

Die von Fowler angesprochenen allgemeinen Programmiersprachen bezeichnet man als General Purpose Language oder auch abgekürzt als GPL. Im Gegensatz zu DSLs sind GPLs nicht auf eine bestimmte Domäne zugeschnitten, sondern können breiter eingesetzt werden. Moderne Programmiersprachen wie Java und C# sind solche Universal-sprachen. Durch ihre Turing Vollständigkeit, kann man sie untereinander austauschen [4, S. 111]. Mehr als eine GPL gibt es, da sich einzelne Sprachen durch zusätzliche besondere Sprachfeatures voneinander abheben [17, S. 27]. Durch die Unterstützung von Pointern in C ist es zum Beispiel potenziell möglich Datenstrukturen besonders speicheroptimiert anzulegen [6, S. 93ff.].

2.1.3.3. Interne DSLs

Bettet man die Befehle zur Beschreibung einer Instanz des Metamodells auf geeignete Weise in eine Universal-sprache ein, kann man dies interne DSL nennen. Der Aufbau und

2. Grundlagen

die Folge von Befehlen soll sich bei einer internen DSL, soweit möglich, wie eine eigene Sprache anfühlen [3, S. 28].

Für die Implementation von internen DSLs eignen sich dynamisch zu beziehende Sprachen wie Ruby und Lisp gut, da sie hilfreiche Features wie die Definition von Makros unterstützen. Je mehr in die normale Syntax einer Sprache eingegriffen werden kann, desto eigenständiger lässt sich die Syntax der internen DSL formen [16, S. 98]. Hierdurch grenzt sich diese immer weiter von einem einfachen application programming Interface (API) ab. Gerade wenn das Metamodell bei objektorientierter Programmierung in Form von Klassen vorliegt, kann eine interne DSL eher wie eine einfache Schnittstelle wirken. Die Abgrenzung von internal DSL zu API ist nicht eindeutig und stellt eine Grauzone dar [3, S. 67].

Eine erwähnenswerte Sonderform der internen DSLs sind fragmentierte DSLs. Bei diesem Spezialfall wird die Hostsprache, also die Programmiersprache mit der die interne DSL ausgedrückt wird, an einzelnen Stellen mit Informationen angereichert. Reguläre Ausdrücke die als Bruchstücke in einer GPL verwendet werden oder auch Annotationen in Java kann man hier als Beispiel anführen [3, S. 32].

2.1.3.4. Externe DSLs

Eine externe DSL hat genau wie eine GPL eine eigene Syntax, laut Fowler ist es jedoch auch nicht unüblich dass die Syntax einer anderen Sprache, beispielsweise XML, verwendet wird. Sprachen wie die Structured Query Language (SQL) oder Cascading Style Sheets (CSS) sind externe DSLs [3, S. 28]. SQL erlaubt es mit syntaktisch einfachen Mitteln Anfragen an eine Datenbank zu schicken und mit CSS kann die grafische Repräsentation von HTML-Seiten manipuliert werden.

Im Gegensatz zu internen DSLs, welche auf der Struktur der zugrunde liegenden GPL beruhen, erlauben externe DSLs ihre Syntax weitestgehend frei zu gestalten. Für ihre Implementation kann grundsätzlich auf einen reichen Schatz an Techniken, die zum Verarbeiten von Programmiersprachen seit Jahren verwendet werden, zugegriffen werden [3, S. 89].

2.1.4. Parser

Ein großer Unterschied zwischen internen und externen DSLs besteht in ihrer Verarbeitung. Während interne DSLs eine Datenstruktur direkt mit Informationen befüllen, muss ein Ausdruck einer externen DSL, genau wie bei einem Ausdruck einer GPL, zuerst auf seinen semantischen Informationsgehalt analysiert werden.

2.1.4.1. Einlesen des Quelltextes

Normalerweise liegt der Quelltext, wie der Name schon sagt, in Textform vor. Wenn im folgenden von Quelltext die Rede ist so ist damit der Source Code von externen DSLs sowie der von GPLs gemeint.

Zur Verarbeitung liest ein Reader den Quelltext zuerst ein und meist wird dieser dann in einem ersten Schritt zeilenweise von einem Lexer in seine Bestandteile (Token) zerlegt. Diese aufgespaltenen Code Zeilen können dann auf syntaktische Korrektheit geprüft werden und in eine Übergangsrepräsentation überführt werden[11, S. 29f.]. Eine Form hierfür kann ein Abstract Syntax Tree (AST) sein.

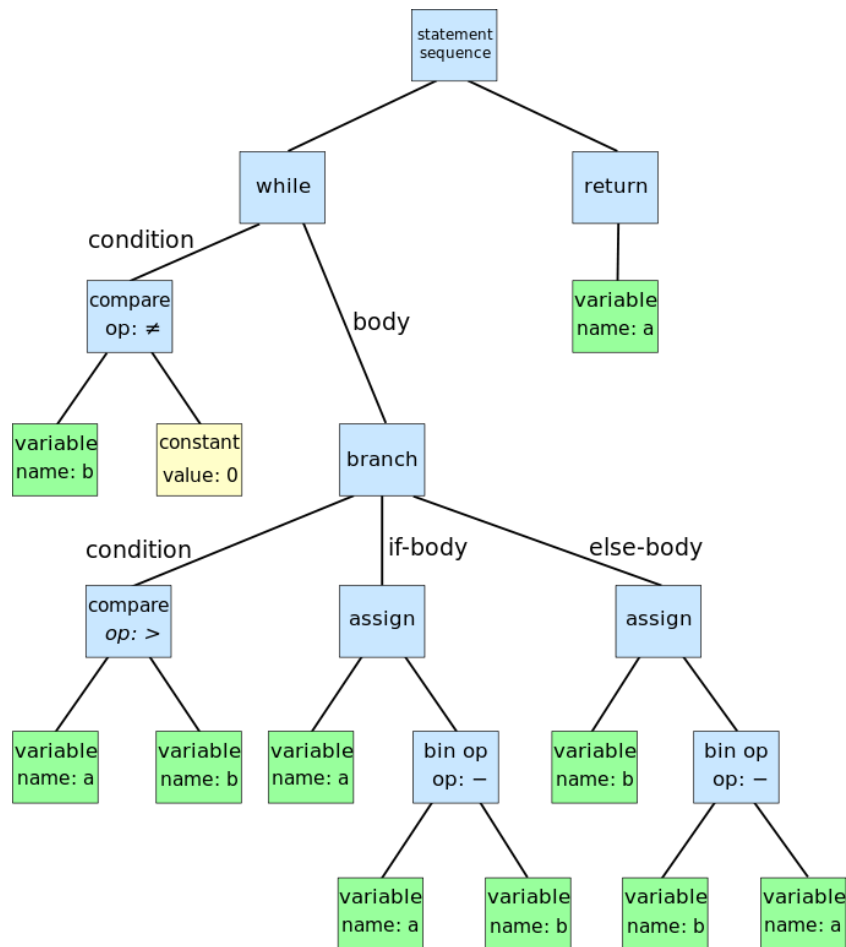


Abbildung 2.2.: Darstellung des euklidischen Algorithmus als AST von Dcoetzee (Eigenes Werk) [CC0], via Wikimedia Commons

2.1.4.2. Abstract Syntax Tree (AST)

Für jeden wichtigen Token wird im AST ein Knotenpunkt, mit für den weiteren Prozess wichtigen Daten, angelegt [11, S. 23]. Die grafische Darstellung des AST für den euklidischen Algorithmus auf Abbildung 2.2 zeigt diese Knoten-basierte Strukturierung des eingelesenen Codes. Die Knoten eines AST kennen ihre eigene Position im Baum und somit kann dieser mit geeigneten Strategien durchlaufen werden [11, S. 24].

Eine Möglichkeit ist die Anwendung des Visitor Patterns, welches in dieser Arbeit in einem späteren Unterkapitel der Grundlagen behandelt wird.

2.1.4.3. Weitere Verarbeitung des eingelesenen Quelltextes

Nachdem die syntaktische Korrektheit des Quelltextes sichergestellt ist und dieser in einer geeigneten Datenstruktur vorliegt, stehen jetzt mehrere Möglichkeiten zur Verfügung. Die jetzt in Baumstruktur vorliegenden Anweisungen sind jetzt durch Interpretation von einer Plattform ausführbar. Alternativ könnte auch eine Übersetzung in eine andere Sprache vorgenommen oder direkt zu ausführbaren Maschinencode kompiliert werden.

Bei Java wird der Quelltext beispielsweise eingelesen und zu Byte-Code transformiert, welcher dann von der Java Virtual Machine (JVM) interpretiert werden kann [7]. Die Übersetzung in eine andere Sprache bietet sich sowohl bei externen DSLs als auch bei GPLs an. Allgemein könnte diese Transformation sinnvoll sein, wenn man seinen Code für eine bestimmte Plattform kompatibel zu einer anderen Plattform machen möchte. Ein Beispiel hierfür ist der Kompilierprozess einer in C# geschriebenen Unity3D Anwendung zu einer WebGL App. Hierbei wird der C# Code zuerst in C++ und dann in JavaScript umgewandelt [9]. Native GPLs wie C werden direkt zu Maschinensprache kompiliert [6]. Externe DSLs werden von einem Generator häufig zu GPL Quelltext transformiert [17, S. 26].

2. Grundlagen

2.1.5. Code Generator

2.1.5.1. Definition

2.1.5.2. Techniken zur Generierung von Code

2.1.5.3. Zusammenhang zu Transformatoren

2.1.5.4. Abgrenzung zum Compilerbau

2.2. Software Architektur

2.2.1. Prinzipien der Softwaretechnik

2.2.1.1. Abstraktion

2.2.1.2. Bindung und Kopplung

2.2.1.3. Modularisierung

2.2.1.4. Weitere Prinzipien

2.2.1.5. Abhängigkeiten

2.2.2. Trennung durch Modelle

2.2.3. Modell-Transformations-Pipeline

2.3. Design Pattern objektorientierter Programmierung

2.3.1. Visitor Pattern

2.3.1.1. Zweck

2.3.1.2. Beschreibung

2.3.1.3. Anwendbarkeit

2.3.2. Builder Pattern und Fluent Interfaces

3. Analyse

3.0. Fragestellung: Wie kann, ausgehend von bestehendem Java-Code, ein Meta-Generator zur Erhöhung der Wirtschaftlichkeit Modellgetriebener Softwareentwicklung umgesetzt werden?

3.1. Aufwand eines konventionellen Softwareprojektes im Vergleich zu MDSD

3.2. Automatisierung der Entwicklung eines Code Generators

3.2.1. Java Code als Ausgangsmodell

3.2.1.1. Anreicherung des Java Codes mit Informationen

3.2.1.2. Parsen des Java Codes

3.2.2. Abstrakte Darstellung von Java Code als Modell

3.2.2.1. Anforderungen an das Modell

3.2.2.2. Schnittstellen zur Befüllung des Modells

(DSLs intern / extern)

3.2.3. Generierung von Java Code

3.2.3.1. Techniken zur Code Generierung

3.2.3.2. Vorhandene Frameworks zur Java Code Generierung

4. Konzept

4.1. Einsparung der Implementation des Code Generators durch einen Meta-Generator

4.2. Funktionsweise des Meta-Generators

4.2.1. Von Java Code zum Annotations-Modell

4.2.1.1. Annotationen als Mittel zur Informationsanreicherung

4.2.1.2. Parsen des annotierten Codes

4.2.1.3. Zweck des Annotations-Modells

4.2.2. Das CodeUnit-Modell

4.2.2.1. Baumstruktur des Modells

4.2.2.2. Spezialisierung durch ein Typ-Feld

4.2.2.3. Parametrisierung durch generische Datenstruktur

4.2.3. Generierung von Buildern als interne DSL aus dem Annotations-Modell

4.2.3.1. Komposition der Builder aus benötigten Builder-Methoden

4.2.3.2. Übertagung vorgegebener Informationen in einen Builder

4.2.3.3. Verwendung von Plattform Code zur Generierung von vordefinierten CodeUnits

4.2.3.4. Auflösung von Referenzen in vordefinierten CodeUnits als nachgelagerter Verarbeitungsschritt

4.2.4. Erzeugung von Java Code aus einem befüllten CodeUnit-Modell

4.2.4.1. Transformation des CodeUnit-Modells in ein File-Modell

5. Lösung: Spectrum (Proof of Concept)

5.1. Verwendete Bibliotheken

5.2. Verwendeter Glossar

5.2.1. JavaParser mit JavaSymbolSolver

5.2.2. JavaPoet

5.3. Architekturübersicht

5.3.1. Amber

5.3.1.1. Annotationen

5.3.1.2. Parser

5.3.1.3. Modell

5.3.2. Cherry

5.3.2.1. Generator

5.3.2.2. Modell

5.3.2.3. Plattform

5.3.2.4. Generierte Builder

5.3.3. Jade

21

5.3.3.1. Transformator

5.3.4. Scarlet

5.3.4.1. Compiler

5. Lösung: Spectrum (Proof of Concept)

Listing 5.1: Beispiel für einen Quelltext

```
1  
2 public void foo() {  
3     // Kommentar  
4 }
```

6. Evaluierung

6.1. Kozept & Implementation

6.1.1. Amber

6.1.2. Cherry

6.1.3. Jade

6.1.4. Scarlet

6.2. Softwarequalität

Nach Balzert S.111

6.2.1. Functionality

6.2.2. Maintainability

6.2.3. Performance

6.2.4. Usability

6.3. Grenzen des Lösungsansatzes

7. Abschluss

7.1. Zusammenfassung

7.2. Ausblick

A. Dokumentation

A.1. Verwendung der Annotationen

A.2. Verwendung der generierten CodeUnit-Builder

A.3. Klassendokumentation

A.3.1. Amber

A.3.2. Cherry

A.3.3. Jade

A.3.4. Scarlet

Abbildungsverzeichnis

| | | |
|------|---|----|
| 2.1. | Darstellung eines Featurediagramms für ein konfigurierbares E-Shop System von Segura09 in der Wikipedia auf Englisch (Transferred from en.wikipedia) [Public domain], via Wikimedia Commons | 8 |
| 2.2. | Darstellung des euklidischen Algorithmus als AST von Dcoetzee (Eigenes Werk) [CC0], via Wikimedia Commons | 12 |

Tabellenverzeichnis

Literatur

- [1] C. Atkinson und T. Kühne. „Model-Driven Development: A Metamodeling Foundation“. In: *IEEE Softw.* 20.5 (Sep. 2003), S. 36–41. ISSN: 0740-7459. URL: <http://dx.doi.org/10.1109/MS.2003.1231149>.
- [2] K. Czarnecki und U. Eisenecker. *Generative Programming: Methods, Tools and Applications*. Addison-Wesley, 2000.
- [3] M. Fowler. *Domain Specific Languages*. Addison-Wesley, 2011.
- [4] J. Hromkovič. *Theoretische Informatik: Formale Sprachen, Berechenbarkeit, Komplexitätstheorie, Algorithmik, Kommunikation und Kryptographie*. Springer Fachmedien, 2014. ISBN: 978-3-658-06433-4.
- [5] K. Kang u. a. *Feature-Oriented Domain Analysis (FODA) Feasibility Study*. Techn. Ber. CMU/SEI-90-TR-021. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University, 1990. URL: <http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=11231>.
- [6] B. Kernighan und D. Ritchie. *The C Programming Language*. 2nd. Prentice Hall Professional Technical Reference, 1988. ISBN: 0131103709.
- [7] T. Lindholm u. a. *The Java Virtual Machine Specification, Java SE 8 Edition*. 1st. Addison-Wesley Professional, 2014. ISBN: 013390590X, 9780133905908.
- [8] o.V. *OMG MDA Guide rev. 2.0*. 2014. URL: <https://www.omg.org/cgi-bin/doc?ormsc/14-06-01>.
- [9] o.V. *Unity User Manual (2017.3)*. Version 2017.3b-001X. Unity Technologies. Vienna, Austria, 2018. URL: <https://docs.unity3d.com/Manual/index.html>.
- [10] o.V. *Zahlen & Fakten: Einkommen und Preise 1900 - 1999*. o.D. URL: https://usa.usembassy.de/etexts/his/e_g_prices1.htm.
- [11] T. Parr. *Language Implementation Patterns: Create Your Own Domain-Specific and General Programming Languages*. 1st. Pragmatic Bookshelf, 2009. ISBN: 193435645X, 9781934356456.
- [12] R. Prieto-Díaz. „Domain Analysis: An Introduction“. In: *SIGSOFT Softw. Eng. Notes* 15.2 (Apr. 1990), S. 47–54. ISSN: 0163-5948. URL: <http://doi.acm.org/10.1145/382296.382703>.
- [13] J. Reichle. *100 Jahre Ford T-Modell: Schwarze Magie*. 2010. URL: <http://www.sueddeutsche.de/auto/jahre-ford-t-modell-schwarze-magie-1.702183>.

- [14] G. Sager. *Erfindung des Ford Modell T: Der kleine Schwarze*. 2008. URL: <http://www.spiegel.de/einestages/100-jahre-ford-modell-t-a-947930.html>.
- [15] B. Selic. „The Pragmatics of Model-Driven Development“. In: *IEEE Softw.* 20:5 (Sep. 2003), S. 19–25. ISSN: 0740-7459. URL: <http://dx.doi.org/10.1109/MS.2003.1231146>.
- [16] T. Stahl und M. Völter. *Modellgetriebene Softwareentwicklung: Techniken, Engineering, Management*. 2. Aufl. dpunkt.verlag, 2007.
- [17] M. Völter u. a. *DSL Engineering - Designing, Implementing and Using Domain-Specific Languages*. dslbook.org, 2013. ISBN: 978-1-4812-1858-0.
- [18] W. Wallén. *The history of the industrial robot*. Techn. Ber. Division of Automatic Control at Linköpings universitet, 2008. URL: <http://liu.diva-portal.org/smash/get/diva2:316930/FULLTEXT01.pdf>.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorgelegte Bachelorarbeit selbstständig verfasst und noch nicht anderweitig zu Prüfungszwecken vorgelegt habe. Alle benutzten Quellen und Hilfsmittel sind angegeben, wörtliche und sinngemäße Zitate wurden als solche gekennzeichnet.

René Ziegler, am 22. Februar 2018

Zustimmung zur Plagiatsüberprüfung

Hiermit willige ich ein, dass zum Zwecke der Überprüfung auf Plagiate meine vorgelegte Arbeit in digitaler Form an PlagScan (www.plagscan.com) übermittelt und diese vorrübergehend (max. 5 Jahre) in der von PlagScan geführten Datenbank gespeichert wird sowie persönliche Daten, die Teil dieser Arbeit sind, dort hinterlegt werden.

Die Einwilligung ist freiwillig. Ohne diese Einwilligung kann unter Entfernung aller persönlichen Angaben und Wahrung der urheberrechtlichen Vorgaben die Plagiatsüberprüfung nicht verhindert werden. Die Einwilligung zur Speicherung und Verwendung der persönlichen Daten kann jederzeit durch Erklärung gegenüber der Fakultät widerrufen werden.

René Ziegler, am 22. Februar 2018