technische universität
dortmund

# Entwicklung eines flexiblen Frameworks zur generierung von Datenmodellen

Alexander Brand

24. Februar 2025

Begutachtung:
Prof. Dr. Erich Schubert
Name des Zweitgutachters

Technische Universität Dortmund
Fakultät für Informatik
Data Mining Arbeitsgruppe
https://dm.cs.tu-dortmund.de/

# Inhaltsverzeichnis

# 1 Einleitung

## 1.1 Motivation

Die zentrale Modellierung von Domainmodellen ist sehr verbreitet in der Entwicklung von großen Software Projekten und zentraler Bestandteil von Product Line Engineering. Dabei stellt die Modellierung des Domain Modells einen Kompromiss zwischen der kompletten Modellierung einer Software und der klassischen Entwicklung ohne Modelle dar.
Ziel dieses Kompromisses ist die Effizienz und Sicherheit der Codegenerierung für das Datenmodell einzusetzen, um die Entwicklung der restlichen Software zu vereinfachen.

### 1.1.1 Product Line Engineering

Product Line Engineering befasst sich mit der Entwicklung von mehreren verwandten Softwareprodukten. Dabei handelt es sich häufig um Software für Teilaufgaben und angepasste Kundenversionen der Standardsoftware.
Hierbei besteht für eine Organisation die Gefahr, viele Komponenten mehrfach zu entwickeln und zu verwalten. Durch gemeinsam genutzte Komponenten (Assets) wird die Entwicklung vereinfacht und die Software verhält sich beim Kunden einheitlich. Datenmodelle stellen im Product Line Engineering wichtige Assets dar. Einheitliche Modelle verhindern das Übersetzen zwischen verschiedenen Produkte.

### 1.1.2 EMF

EMF(Eclipse Modelling Framework) ist ein häufig Framework zur Modellierung von Modellen in Java. Es lassen sich große Modelle darstellen und mithilfe von Maven Workflows können diese durch das Build Tool übersetzt werden.
EMF bietet dabei jedoch keine Wahl bei der IDE oder der Programmiersprache. Dies führt dazu, dass Projekte und ganze Firmen bei ihren bisherigen Technologien stehen bleiben. Es wird bei Neuentwicklungen nicht mehr die gefragt, was wären die besten Technologien um das Problem zu lösen, sondern es wird gefragt, wie lösen wir das mit unserer bisherigen Architektur.

### 1.1.3 Effekte eines unflexiblen Frameworks

Dieser fehlerhafte Ansatz schädigt das Projekt auf mehreren Ebenen:

1. Wissen und Erfahrung Konzentrierung
   Da nur eine Architektur in Betracht gezogen wird, hat jedes Mitglied des Teams nur Erfahrung mit der aktuellen Architektur und jegliche Erfahrung mit anderen Technologien verfällt mit der Zeit. Dies schränkt die die Perspektiven auf Probleme sehr stark ein und macht einen Wechsel sehr aufwendig.
2. Schrinkender Bewerberzahl
   Da nur Bewerber für die gewählten Technologien in Betracht gezogen werden, verringert sich die Anzahl stark. Der Effekt wird verstärkt, wenn die Technologien als veraltet gelten. Eine kleinere Bewerberanzahl zwingt Unternehmen auch Bewerber, die andernfalls nicht beachtet worden wären, in Betracht zu ziehen. Dies führt zu weiteren Negativen Effekten, da einige schlechte Angestellte die Produktivität vieler guter Angestellter stark senken können.
3. Anfälligkeit gegenüber Sicherheitslücken
   Eine starke Festlegung auf Technologien führt dazu, dass Sicherheitslücken gleich jedes Projekt betreffen. So könnten bei einer Zero-Day-Lücke direkt mehere Schicht im SSchweizer Käse Modell"(Source suchen) wegfallen. Diese Anfälligkeit wird stark erhöht sobald eine Technologie nicht mehr aktiv weiterentwickelt wird. Dies führt häufig dazu, dass andere Updates auch nicht genutzt werden können.

## 1.2 Aufgabenstellung

Das Domain Modell Framework(DMF) soll es ermöglichen Datenmodelle zentral zu modellieren, sodass diese von verschiedenen Software Projekten genutzt werden können. Dabei soll die Flexibilität besonders beachtet werden, um die bisher bestehenden Nachteile zu vermeiden. Zur Flexibilität gehört die (möglichst) freie Wahl der Programmiersprache und die freie Wahl der Entwicklungsumgebung.
Ziel ist es das DMF für Java und Typescript zu implementieren. Es sollen primär Intellij und Visual Studio Code unterstützt werden.

# 2 Entwicklung

## 2.1 Auswahl der verwendeten Technologien

Ein zentraler Teil einer Architektur ist die Auswahl der verwendeten Technologien. Diese Technologien sollen die Lösung der Aufgaben einer Software vereinfachen.
Im DMF müssen folgende Aufgaben gelöst werden:

1. Modelldatei Parsen und AST generieren
2. AST auslesen und verarbeiten
3. Kommunikation mit verschiedenen IDE's
4. Generieren von Codedateien in verschieden Sprachen
5. Integration mit verschiedenen Build Tools

### 2.1.1 Parser

Der Parser für das DMF muss große Dateien wiederholt mit kleinen Änderungen parsen. Diese Anforderung stammt aus der Notwendigkeit des AST's um Syntaktische und Semantische Fehler, sowie die verschiedenen Tokens(siehe Abschnitt LSP) nach jeder Eingabe an die IDE zu übermitteln. Hierbei ist Latenz die höchste Priorität, denn die Reaktionsfähigkeit der IDE beeinflusst die Geschwindigkeit mit der Entwickelt werden kann.
Zusätzlich muss der Parser auch von jeder anderen Komponente des DMFs verwendet werden. Deshalb ist hier die Einschränkung der anderen Technologien unerwünscht.

#### XText

XText ist ein Framework der Eclipse Foundation.
Es bietet die Möglichkeit eine DSL mit verschiedenen Modellen zu modellieren und Regeln automatisch zu überprüfen. XText setzt auf Modellierung vieler Bestandteile und generiert andere Komponenten komplett. Dies ermöglicht eine schnelle Entwicklung wenn die Anforderungen perfekt zu XText passen. XText schränkt stark ein, wo Anpassungen vorgenommen werden können. So ist es nicht vorgesehen die LSP-Server Implementierung anzupassen, obwohl XText nicht alle Features des LSP-Protokolls unterstützt. Dateigeneration und die Verarbeitung des AST's müssen auch mit dem Java-Interfaces von XText vorgenommen werden. Dies setzt immer die Verwendung von JVM basierten Sprachen vorraus. Jede JVM-Implementierung benötigt beachtliche Zeit zum Starten weshalb Code Generation immer auf den Start Warten muss.
Abschließend waren an XText die nicht funktionierenden Beispiel Projekte sowie zwingende Entwicklung in Eclipse seht abweisend. Ein Framework welche eine einfache und

flexible Entwicklung ermöglichen soll, sollte nicht schwer und nur in einer IDE zu entwickeln sein.

### Treesitter

Treesitter ist ein Open Source Framework zur Generierung von Parsern.
Dabei wird die Grammatik mithilfe einer Javascript API definiert. Mithilfe der Treesitter CLI wird aus der Javascript Datei der Parser generiert. Der generierte Parser nutzt C. C eignet sich hier sehr gut, da es die höchste Performance und die Möglichkeit es in jeder anderen Sprache zu nutzen bietet. Das Nutzen von C ist für jede Sprache eine Voraussetzung, um mit dem Betriebssystem zu kommunizieren. C's größter Nachteil, die manuelle Speicher Verwaltung, wird durch die Generation des Parsers gelöst. Die bereitgestellten Schnittstellen übergeben Strukturen welche vom Aufrufer verwaltet werden.

**Iteratives Parsen** Ein großes Unterscheidungsmerkmal von Treesitter ist die Möglichkeit iterativ zu parsen.

*With intelligent [node] reuse, changes match the user's in tuition; the size of the development record is decreased; and the performance of further analyses (such as semantics) improves.[?]*

Beim iterativen Parsen ist das Ziel den AST nicht bei jedem Parse Durchlauf neu zu erstellen, sondern möglichst viel des AST's wiederzuverwenden. Für das Iterative Parsen muss der AST sowie die bearbeiteten Textstellen an Treesitter übergeben werden. Die Durchlaufszeit des iterativen Parsedurchlaufs hängt nicht mehr der Länge der kompletten Modelldatei ab, sondern nur von den neuen Terminalsymbolen und Modifikationen im AST:

*Our incremental parsing algorithm runs in O(t + slgN) time for t new terminal symbols and s modification sites in a tree containing N nodes [?]*

### ANTLR

ANTLR ist sehr ähnlich zu Treesitter. Die größten Unterschiede sind die API's zum Schreiben der Grammatiken und die Möglichkeit iterativ zu Parsen. Zusätzlich unterstützt ANTRL nur Java, C# und C++. Dies zwingt einen in der Wahl der Implementierungssprache ein.

### Auswahl Parser

Für das DMF Framework wurde Treesitter verwendet. Die Exellente Performance sowie die Flexibilität bei der Implementierung der restlichen Komponenten hoben Treesitter von den restlichen Technologien ab.

h

### 2.1.2 AST Verarbeiten

Bei der Verarbeitung des AST's müssen verschiedene Regeln abgearbeitet werden und der Inhalt des AST's in einem Modell vorbereitet werden. Essentiell für die Verarbeitung ist die Zusammenarbeit mit den folgenden Komponenten.

Die Auswahl der Technologie für diesen Schritt basiert auf der Auswahl für die folgenden Schritte.

### 2.1.3 Kommunikation mit verschiedenen IDE's

Damit ein Framework die Entwicklung nicht einschränkt muss es in verschiedenen Ïntegrated Development Environments"(IDE's) genutzt werden können. Viele IDE's stellen Schnittstellen für Plugins bereit. Dazu zählen Intellij, Eclipse, NeoVim und VSCode. Jede Schnittstellen ist jedoch unterschiedlich, wodurch die Entwicklung von vielen Verschiedenen Plugins nötig wäre.

**Language Server Protokoll** Eine einfachere Möglichkeit bietet das "Language Server Protokoll"(LSP). Dieses Protokoll bietet die Möglichkeit, dass viele verschiedene IDE's eine Server Implementierung nutzen. Im Fall von Zed und Eclipse lassen sich LSP-Server sogar ohne jegliche Plugins einbinden. Wobei hier auf die schlechte Unterstützung des LSP-Protokolls in Eclipse hingewiesen werden muss. Intellij und NeoVim nutzen Plugins, um LSP-Server anzubinden. VSCode bietet eine API und einen einfachen LSP-Client in ein kleines Plugin zu implementieren. Im LSP-Server können gebündelt Logik und Prokoll implementiert werden.

LSP wird hauptsächlich über die Standard-Eingabe und -Ausgabe oder über einen Server Socket transportiert. Es wird ein JSON-RPC Format genutzt. Der LSP-Server muss somit JSON, Std-In und Std-Out, sowie Server Sockets unterstützen.

#### Typescript

Von der VSCode Dokumentation wird die Implementierung eines LSP-Servers in Typescript empfohlen. Dafür werden Bibliotheken bereit gestellt. Typescript eignet sich gut, für die JSON Parsing und für die Verwendung von Server Sockets. Probleme entstehen bei Typescript bei den Themen Performance, Anbindung an den Parser und bei der Fehlerbehandlung.

#### Golang

Golang ist eine Sprache welche für die Entwicklung von Backends ausgelegt wurde.
Es werden die Anforderungen für JSON-Parsing, Std-IO und Server Sockets erfüllt, durch

die große Standard Bibliothek erfüllt. Es gibt keine Bibliothek welche das komplette Protokoll beinhaltet. Dieses kann jedoch durch die Unterstützung von LLMs schnelle generiert werden.
Golang bietet zusätzlich eine simple Anbindung an den Parser und die Möglichkeit sehr einfach Parallelität einzubauen. Besonders erwähnenswert ist die Geschwindigkeit eines Golang Programmes und die Startgeschwindigkeit ohne auf Speichersicherheit zu verzichten.

#### Java

Java bietet eine bietet mit "lsp4jëine Bibliothek zur Einfachen Entwickung. Bei der Einbindung des Parsers gestallten sich jedoch zusätzliche Herrausforderungen da der Java Code Plattform unabhängig kompiliert wird und Plattform abhängigen Code aufrufen muss. Java benötigt für die Ausführung eine Installierte Instanz der JRE. Die JRE muss nicht nur zusätzlich zum LSP-Server verwaltet werden, sondern benötigt zusätzlich Zeit zum Starten. So muss der Entwickler länger warten bis seine Entwicklungsumgebung bereit steht.

### 2.1.4 Generation von Codedateien in verschiedenen Sprachen

Ziel des DMFs ist es große Mengen an Sourcecode zu Generieren. Dabei soll den Entwicklern die Wahl zwischen mehreren Zielsprachen gegeben werden. Diese Generation wird beim Build und damit sehr häufig ausgeführt. Eine Langsame Generation wird jeder Organisation viel Geld kosten.

Die Generation muss somit schnell und Zielsprachen unabhängig sein. Sie muss auch aus von den Build Tools gestartet werden.

#### Golang Templates

Golang Standardbibliothek bietet die Möglichkeit Templates zu definieren. Diese Templates werden hauptsächlich für die Generierung von HTML genutzt. Da sie Golang die Templates nicht nur für HTML sondern auch für generelle Texte anbietet, können diese auch für jede Zielsprache genutzt werden.
Die Anforderungen an einen Webserver (Geschwindigkeit, Resourcen schonend, Simpel) komplementieren die Anforderungen an einen Codegenerator sehr gut.
Golang Templates stechen besonders für ihre Integration in IDE's wie z.B. in Intellij heraus.

#### Java

Es gibt mehrere Template Engines für Java. Einige Beispiele wäre FreeMaker oder Apache Velocity. Beide sind gut unterstützt und bieten alle nötigen Features für die Generierung von Code.

#### Typescript

Für Typescript gibt es viele Template Engines. Zu den bekannten gehören Eta, liquidjs und squirrelly. Sie bieten alle die Möglichkeit verschiedene Zielsprachen zu generieren und können mit nodejs ausgeführt werden.

#### Auswahl

Da Golang eine exzellente Unterstützung in Intellij hat und keine Zusätzliche Installtion wie NodeJs oder JRE verwalten muss, fiehl meine Wahl auf Golang.
Mit der Wahl für Golang für den Generator, ist auch die Wahl für die Verarbeitung des AST's und für den LSP-Server.

### 2.1.5 Integration mit verschiedenen Build Tools

Damit eine Generation während des Buildvorgangs ist essentiell, um sicherzustellen das der generierte Code aktuell ist. Damit der Neugeneration werden auch alle eventuelle Anpassungen in den Dateien überschrieben, wodurch Fehler vermieden werden.

#### Maven

Maven ist ein sehr verbreitetes Build Tool für Java. Maven unterstützt Plugins welche während des Builds ausgeführt werden und in der Maven Konfiguration konfiguriert werden können. Die API für Maven Plugins ist in Java geschrieben. Dieses Plugin muss den Generator aufrufen. Dies ist möglich indem die Datei des Generators ausgeführt wird.

#### NPM

NPM ist das führende Build Tool für Typescript Projekte. NPM unterstützt die Ausführung von Terminal Befehlen. Der Generator kann somit über das Terminal ausgeführt werden.

## 2.2 Abstraktion des DMF

Das DMF basiert auf einer Abstraktion der Datenstrukturen aus mehreren Sprachen. Diese Abstraktion wurde nach einer Analyse entwickelt.

### 2.2.1 Analyse

Für die Analyse wurden die Sprachen Java, Typescript,Python, Golang, Rust und C analysiert. Die Sprachen wurden spezifisch ausgewählt. Java ist weitverbreitet in Enterprise Software. Typescript ist die Standardsprache für jegliche Websiten und viele Backends. Python ist in der Datenanalyse weit verbreitet. Durch die Popularität in Umfragen wurde Python miteinbezogen. Golang ist eine moderne Alternative für Backends und die Implementierungsprache des DMFs. Rust ist die moderne Wahl für 'low level' Programmierung. C ist die Standardsprache für jede 'Foreign-Function-Interfaces' und ist weit verbreitet für ältere 'low level' Software.

#### Analyse der Typen

Es wurde analysiert, welche Typen als Referenz oder als Wert als Variablentyp genutzt werden können.

| Typen | Java | Typescript | Python | Golang | Rust | C |
|---|---|---|---|---|---|---|
| Wert | Primitive Typen | Primitive Typen | Primitive Typen | Alle Typen | Alle Typen | Alle Typen |
| Referenz | Objekte | Objekte, Arrays, Funktionen, Klassen | Alles außer primitive Typen | Explizit | Explizit | Explizit |

Bei den Sprachen Java, Typescript und Python werden nur Primitive Typen als Wert Variablen gespeichert. Deshalb wurden die Pritiven Typen dieser Sprachen genauer verglichen:

| Primitive Typen | Java | Typescript | Python |
|---|---|---|---|
| | byte, short, int, long, float, double, char, boolean | number, bigint, string, boolean | int, float, bool, str |

Auffällig sind hierbei die Zusammenfassung der Typen byte, short, int, long in Java in den Typen int in Python, sowie die Zusammenfassung aller Zahlentypen, bis auf long, in number in Typescript. Java besitzt als einzige Sprache String nicht als primitiven Datentyp.

#### Analyse von Nullwerten

Nullwerte sind besonders aus Java bekannt und stellen das Fehlen eines Wertes dar. Es zählt zu der Definition eines Types dazu, zu definieren, ob der Typ Nullwerte erlaubt. Dies muss auch für Werte und Referenzen evaluiert werden.

| Nullwerte | Java | Typescript | Python | Golang | Rust | C |
|---|---|---|---|---|---|---|
| Wert | nein | nein | ja | nein | Explizit | nein |
| Referenz | ja | Explizit | ja | ja | Explizit | ja |

Es ist klar zu erkennen, dass bis auf Python jede Sprache Wert-Variablen ohne Nullwerte darstellen kann. Referenzen können auch in jeder Sprache Nullwerte beinhalten. In Typescript und Rust muss dies bloß explizit definiert werden. Aus diesen Ergebnissen ergibt sich, dass die Unterteilung in Wert- und Referenz-Variablen auch die Unterteilung in Nullbare und nicht Nullbare Variablen abbildet.

### Collectiontypen

Um 1:n- oder n:m-Beziehungen im Datenmodell modellieren zu können wurden drei Collection-Typen aus Java ausgewählt und passende Äquivalente zu finden.

| Collectiontypes | Java | Typescript | Python | Golang | Rust | C |
|---|---|---|---|---|---|---|
| List | ja | ja (Array) | ja | ja (slice) | ja | ja (Array) |
| Set | ja | ja | ja | nein | ja | nein |
| Map | ja | ja | ja (dictionary) | ja | ja | nein |

Die gewählten Typen sind die am häufigsten verwendeten Collection-Typen. Eine Map beinhaltet eine 1:n-Beziehung und ermöglicht einen schnellen Zugriff. Eine Liste bildet eine n:m-Beziehung zwischen den modelliertem Element und dem Inhalt der Liste. Ein Set bildet eine n:m.Beziehung mit der Garantie, dass jeder enthaltener Wert einzigartig ist.

In der Analyse der Liste gab es feine Unterschiede in der Implementierung. Typescript und C nutzen einen Array, jedoch verhält sich der Typescript Array wie eine Liste. In C sind Arrays in ihrer Größe bei ihrer Initialisierung festgelegt. Golang nutzt ein Konstrukt namens 'slice'. Es kommt mit bestimmten Eigenschaften, kann jedoch für eine Liste genutzt werden.

Ein Set findet sich nur in Golang und C nicht. Hier kann es durch eine Liste ersetzt werden. Die Garantien müssten selber verwaltet werden.

Bei der Analyse der Map wurde nur in C keine Implementierung gefunden. Python nutzt für die Map den Namen 'dictionary'.

### 2.2.2 Elemente eines Modells

Um mit dem DMF Daten in Strukturen verschiedener Programmiersprachen darstellen zu können, müssen auch diese Abstrahiert werden. Dieser Abschnitt beschriebt wie aus den Analysen der Programmiersprachen die Abstraktion des DMFs gebildet wurde.

#### Primitive Typen

Grundvoraussetzung sind die Primitiven Typen und Referenzen zu anderen Elementen. Bei der Analyse wurde ein unterschiedlichen Maß in der Feinheit der Zahlentypen festgestellt. Es gibt in SQL Datenbanksystem generell eine Unterscheidung zwischen ganzen Zahlen und rationalen Zahlen. Somit muss es eine Unterscheidung zwischen int und double geben. Es wird jedoch auch unterschieden wie groß ganze Zahlen werden, weshalb ein long Typ sinnvoll ist. Dieser kann auch mithilfe von bigint in Typescript abgebildet werden. Für die Verarbeitung von unbekannten Daten werden häufig Bytes genutzt. Von den drei verglichen Sprachen, beinhaltet nur Java den primitiven Typ. Die int-Typen der jeweiligen Sprachen ermöglichen jedoch ähnliche Operationen. Deshalb wurde auch Byte aufgenommen. Eine Unterscheidung zwischen float und double wurde nicht vorgenommen, da diese Unterscheidung in den Systemen die sie enthalten sehr wenig verwendet wird.
String ist vorallem in Scriptsprachen ein primitiver Typ und wird auch von Datenbanken unterstützt. Deshalb wurde auch String als primitiver Typ ins DMF aufgenommen.
Im Gegensatz zu allen verglichenen Programmiersprachen besitzen SQL-Datenbanken Unterstützung für Datum- und Zeitstempel-Werte. Damit die Generation diese Werte in das Datenbankmodell übernehmen kann, wurden 'date' und 'datetime' als primitive Typen hinzugefügt.
Abschließend gehört noch 'boolean' zu den primtiven Typen. Wahrheitswerte werden sowohl in allen Programmiersprachen als auch in allen Datenbankensystem unterstützt.

Somit beinhaltet das DMF die folgenden primitiven Typen:

| Typ | ganze Zahlen | rationale Zahlen | Text | Zeit | Wahrheitswert |
|---|---|---|---|---|---|
| | byte, int, long | double | string | date, datetime | boolean |

Diese Primitiven Typen werden im DMF in Argumenten abgebildet. Argumente bestehen aus einem Primitiven Typen und einem Namen. Alle anderen Datentypen werden als Referenzen abgebildet. Vorgesehen ist nur Referenzen explizit als Nullbar generiert werden.

#### Funktionen

Funktionen gehören zu den Elementen die sich in jeder Programmiersprache wiederfinden. Im DMF werden Funktionen nur im Zurückgabewert eingeschränkt. Statt mehreren Werten wie z.B. in Golang kannn im DMF nur ein einzelner Wert modelliert werden. Diese Einschränkung stammt aus vielen Sprachen, welche nur einen Wert unterstützen.

n

### Komplexe Datentypen

*Structured constructors. (+), types can be built up from these basic types by means of type. The type constructors in our language include function spaces Cartesian products (x), record types (also called labeled Cartesian products), and variant types (also called labeled disjoint sums).[?]*

*In nahezu allen Programmiersprachen gibt es die Möglichkeit, mit so genannten zusammengesetzten oder komplexen Datentypen zu arbeiten. Ihnen ist gemeinsam, dass wir mehrere Werte nebeneinander dort abspeichern können.[?]*

Das DMF muss diese Datentypen auch abbilden können. Deshalb beinhaltet es Structs. Der Name wurde von der Programmiersprache C übernommen, da diese Syntaktische Grundlage für fast alle Programmiersprachen dient.
Im DMF können Structs Argumente, Referenzen zu anderen Structs, Entities, Enums und Interfaces (siehe folgende Abschnitte) und Funktionen beinhalten. Funktionen gehören nicht zur Definition eines komplexen Datentyps, sondern stammen aus der Objekt Orientierten Programmierung. Da jedoch Funktionen auch ohne Objektorientierung für Datentypen generiert werden können, kann das DMF diese Abstraktion unterstützen.
Für die Modellierung wird auch die Abstraktion von Datentypen essentiell sein. Dafür müssen Structs von anderen Structs erben und Funktionen von Interfaces implementieren können.
Abstraktion funktioniert in jeder Sprache ein wenig unterschiedlich, weshalb das DMF nur garantieren kann, dass die Variablen und Funktionen die von einem Struct geerbt werden im Generat vorhanden sind. Zum Beispiel in C könnte ein DMF Generat keine Abstraktion generieren, sondern nur die Elemente kombinieren.

### Identität einer Instanz in der Datenbank

Ein Modell im DMF Framework soll in einer Datenbank gespeichert werden können. Dafür müssen Datenbankschlüssel definiert werden. Ein Schlüssel definiert die Identität eines Zeile in einer Tabelle. Diese Identität muss auch im Modell abgebildet werden. Das DMF fügt deshalb den Typen 'Entity' hinzu, welcher eine Identität besitzt. Er basiert auf dem Struct und kann somit Argumente, Referenzen und Funktionen beinhalten. Eine Entity muss die Definition eines Identifiers beinhalten.

Ein weitere Besonderheit ist die Vererbung bei Entities. Eine Entity darf sowohl von einem Struct als auch von einer Entity erben. Ein Struct darf nur von einem Struct erben.

#### Aufzählungen

Aufzählungen sind Bestandteil vieler Programmiersprachen. Häufig existieren sie als reine Liste aus Codesymbolen. Aus Sprachen wie Rust sind jedoch auch Aufzählungstypen dessen Einträge konstante Werte beinhalten können bekannt. TODO: Beispiel Rust Enum Code

Diese Funktion kann auch in Sprachen dessen Enums diese Möglichkeit nicht beinhalten, durch Funktionen die für den Enumeintrag den modellierten Wert zurückgeben, emuliert werden.

Im DMF lassen sich diese Werte mithilfe von Argumenten modellieren. Bei der Definition eines Enumeintrags müssen die Konstanten mit angeben werden.

#### Interfaces

Wichtig für die Abstraktion sind Interfaces. Sie stellen Funktionen bereit und können zusammen mit anderen Interfaces in Structs und Enities implementiert werden.

#### Organisation der Elemente

In großen Softwareprojekten werden Datentypen generell in Gruppen organisiert. Diese Gruppierung erfolgt meistens über das Dateisystem. Dabei repräsentiert ein Ordner eine Gruppe. Diese Gruppe wird meistens 'Package' genannt.
Das DMF beinhaltet auch Packages. Diese werden jedoch nicht im Dateisystem modelliert, sondern sollen als Elemente im Modell enthalten sein.

### 2.2.3 Zuweisungen der Abstraktionen

Damit diese Abstraktion genutzt werden kann, müssen für jeden abstrakten Typen im DMF eine Zuweisung in jeder Sprache festgelegt werden.

#### Java

| Element | Java |
|---|---|
| package | Java Package |
| struct | Java Klasse |
| entity | Java Klasse |
| interface | Java Interface |
| enum | Java Enum |

Die DMF Elemente können sehr gut in Java übersetzt werden. Für die Entity kann sogar die Identität mithilfe der Implementation von der Methoden 'hashCode' und 'equals' übernommen werden. Die Enums unterstützen auch die zusätzlichen Argumente.

p

| Datentyp | Java |
|---|---|
| byte | byte |
| int | int |
| long | long |
| double | double |
| boolean | boolean |
| string | java.lang.String |
| date | java.time.LocalDate |
| datetime | java.time.LocalDateTime |

Das DMF kann bei den Zahlen und Wahrheitswerten genau auf Java übersetzt werden. Für Text- und Zeitwerte werden Klassen der Standardbibliothek verwendet.

#### Typescript

| Element | Typescript |
|---|---|
| package | Ordner |
| struct | Typescript Klasse |
| entity | Typescript Klasse |
| interface | Typescript Interface |
| enum | Typescript Enum |

TODO Nach Entwicklung des Typescript Generators

| Datentyp | Typescript |
|---|---|
| byte | number |
| int | number |
| long | bigint |
| double | number |
| boolean | boolean |
| string | string |
| date | Date |
| datetime | Date |

TODO Nach Entwicklung des Typescript Generators

## 2.3 Aufbau der DMF-DSL

Im Kern des DMFs sind die Modelldateien. Sie enthalten alle Informationen und die Schnittstelle zwischen DMF und Entwickler*innen. Deswegen ist das Format essentiell. EMF setzt auf XML als Format für die Modelldateien und erleichtert die Bearbeitung durch Grafische

Editoren in Eclipse. Grafische Editoren funktionieren für das DMF nicht, da DMF mithilfe des LSP-Protokolls nur Text-Editoren unterstützt. Deshalb muss sich auch das Format der Modelldateien anpassen.

### 2.3.1 Die DSL und die EBNF

Die DMF-DSL besteht aus vielen Elementen die in einer EBNF dargestellt werden können. In diesem Abschnitt werden die verschiedenen Elemente und ausgewählte Regeln der EBNF Grammatik erläutert.

Die EBNF Grammatik nutzt "[]"für Optionale Elemente, *"für optionale Wiederholungen (mindestens 0 Elemente) und "+érforderliche Wiederholungen (mindestens 1 Element).
Sie beginnt mit der Regel ßource_file:

⟨*source_file*⟩ ::= ⟨*dmf_declaration*⟩ ⟨*new_line*⟩ ⟨*model_declaration*⟩ ⟨*new_line*⟩ [ ⟨*import_block*⟩] ⟨*model_content*⟩

#### Header

Jede DMF Datei beginnt mit einem Header. Dieser Header wird genutzt die Version des Formats und Metadaten über das Modell zu dokumentieren.

Listing 2.1: Header einer DMF-Modelldatei

```
1 dmf 1.0.0
2 model "beispiel" version 0.0.1
```

Der Header besteht aus 2 wichtigen Grammatik-Regeln: "dmf_declarationünd "model_-declaration". Beide bilden jeweils eine Zeile ab.

Listing 2.1: Die EBNF Regeln

⟨*dmf_declaration*⟩ ::= 'dmf' ⟨*version_number*⟩

⟨*model_declaration*⟩ ::= 'model' ⟨*string_value*⟩ 'version' ⟨*version_number*⟩

#### Imports

Es gibt zwei Gründe für den Import von vorhandenen Modellen:

1. Um die enthaltenen Klassen in das eigene Modell zu kopieren.
2. Um die enthaltenen Klassen zu erweitern.

Um ein DMF Modell zu Importieren wird die Datei und das Package das vom Modell übernommen werden soll. Es muss ein Package angegeben werden damit immer klar ist, welche Pfade schon belegt sind.

r

Listing 2.2: Import des Package de.base

```
1 import de.base from "./base.dmf"
```

Diese Importstatements werden alle im Import-Block zusammengefasst. Sollten keine Imports genutzt werden, so wird das komplette Import-Block aus dem AST entfernt, statt ein leeres Import-Block Element zu enthalten.

⟨*import_block*⟩ ::= ⟨*import_statement*⟩+

⟨*import_statement*⟩ ::= 'import' ⟨*package_string*⟩ 'from' ⟨*string_value*⟩

### Packages

Ein DMF Modell besteht aus Packages. Diese Unterteilen die restlichen Elementen und representieren Ordner im späteren Generat. Innerhalb eines Package können weitere Packages, Structs, Entities, Enums und Interfaces enthalten sein. Diese verschiedenen Elemente werden im weiterem als PackageElemente bezeichnet.

```
1 // Das ist ein Packages mit ddem Pfad "de.beispiel"
2 package de.beispiel {
3
4 }
```

Jedes PackageElement kann mit einem Kommentar, dem Keyword ëxpandünd einem Override-Block versehen werden. Die Auslagerung dieser AST-Elemente in das "package_content-Element verinfacht das Verarbeiten des ASTs, da ein allgemeines Vorgehen für alle PackageElemente genutzt werden kann.
Der Kommentar dient zur Dokumentation des jeweiliegen Elements.
Das ëxpandKeyword muss genutzt werden, wenn ein importiertes Element erweitert werden soll.
Der Override-Block lässt den Entwickler Zielsprachen Spezifische Anpassungen vornehmen. Dadurch können Spezielle Details die in der Syntax für alle Sprachen nicht abgebildet werden können, trotzdem generiert werden. Es können auch Bibliotheken in der Zielsprache eingebunden werden.
Alle Elemente folgen dem Schema das die *_content"Regel immer den Inhalt, der in unbestimmter Reihenfolge enthalten sein kann, enthält und die *_block"Regel die Elemente enthält, die den Inhalt einschließen.

⟨*model_content*⟩ ::= ⟨*package_content*⟩+

⟨*package_content*⟩ ::= [⟨*comment_block*⟩] [ 'expand'] ⟨*package_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*struct_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*enum_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*entity_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*interface_block*⟩ [⟨*override_block*⟩]

⟨*package_block*⟩ ::= 'package' '{' ⟨*package_content*⟩* '}'

### Structs

Ein Struct fasst Variablen zusammen ohne dabei eine Identität zu garantieren. Structs können Argumente, (Multi-)Referenzen und Funktionen beinhalten. Zusätzlich unterstützen Structs die Implementierung von mehreren Interfaces und sie können von einem anderen Struct erben.

```
1 struct Beispiel extends SuperBeispiel implements IBeispiel, IAnderesBeispiel {
2
3 }
```

Wie bei den PackageElementen gibt auch beim ßtruct_content"wieder einen optionalen Kommentar und Override-Block. Dieses Schema wird auch bei allen anderen PackageElementen fortgeführt.

⟨*struct_block*⟩ ::= 'struct' ⟨*identifier*⟩ [⟨*extends_block*⟩] [⟨*implements_block*⟩] '{' ⟨*struct_content*⟩* '}'

⟨*struct_content*⟩ ::= [⟨*comment_block*⟩] ⟨*arg_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*ref_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*multi_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*func_block*⟩ [⟨*override_block*⟩]

### Entity

Eine Entity ist ein spezializiertes Struct. Es besitzt eine Identität in der Datenbank. Die restlichen Elemente werden aus dem Struct übernommen. Jedoch kann eine Entity von Structs und Entities erben.

```
1 entity BeispielEntity extends AndereEntity implements IBeispiel, IAnderesBeisp
2     arg int i;
3     arg double d;
4     identifier(i,d);
5 }
```

Besonders bei der ëntity_blockRegel ist, dass sie direkt die ßtruct_contentRegel verwendet. Das Identifier-Statement schließt immer den Inhalt einer Entity ab. Hierbei ist zu beachten, dass ïdentifieräls Keyword und als Regelname verwendet wird.

⟨*entity_block*⟩ ::= 'entity' ⟨*identifier*⟩ [⟨*extends_block*⟩] [⟨*implements_block*⟩] '{' ⟨*struct_content*⟩* ⟨*identifier_statement*⟩ '}'

⟨*identifier_statement*⟩ ::= 'identifier' '(' ⟨*identifier*⟩ (',' ⟨*identifier*⟩)* ')' ';'

t

### Interface

Interfaces beinhalten Funktionen und können von Structs und Entities implementiert werden. Interfaces können auch andere Interfaces implementieren. Dabei werden die Methoden ins Interface übernommen.

```
1 interface IBeispiel implements IAnderesBeispiel {
2
3 }
```

⟨*interface_block*⟩ ::= 'interface' ⟨*identifier*⟩ [⟨*implements_block*⟩] '{' ⟨*interface_content*⟩* '}'

⟨*interface_content*⟩ ::= [⟨*comment_block*⟩] ⟨*func_block*⟩ [⟨*override_block*⟩]

### Enum

Enum sind Aufzählungen. Jedem Eintrag können mithilfe von Argumenten weitere Werte zugeordnet werden.

```
1 enum EBeispiel {
2     arg int i;
3
4     KONSTANTE(_,1);
5 }
```

⟨*enum_block*⟩ ::= 'enum' ⟨*identifier*⟩ '{' ⟨*enum_content*⟩* '}'

⟨*enum_content*⟩ ::= [⟨*comment_block*⟩] ⟨*arg_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*enum_constant*⟩ [⟨*override_block*⟩]

⟨*enum_constant*⟩ ::= ⟨*identifier*⟩ '(' ⟨*enum_index*⟩ ( ',' ⟨*primitive_value*⟩)* ')' ';'

⟨*enum_index*⟩ ::= '_' | ⟨*integerValue*⟩

# 3 Anhang

## 3.1 EBNF Grammatik für DMF

⟨*source_file*⟩ ::= ⟨*dmf_declaration*⟩ ⟨*new_line*⟩ ⟨*model_declaration*⟩ ⟨*new_line*⟩ [⟨*import_block*⟩] ⟨*model_content*⟩

⟨*dmf_declaration*⟩ ::= 'dmf' ⟨*version_number*⟩

⟨*model_declaration*⟩ ::= 'model' ⟨*string_value*⟩ 'version' ⟨*version_number*⟩

⟨*import_block*⟩ ::= ⟨*import_statement*⟩+

⟨*import_statement*⟩ ::= 'import' ⟨*package_string*⟩ 'from' ⟨*string_value*⟩ ⟨*new_line*⟩

⟨*model_content*⟩ ::= ⟨*package_content*⟩+

⟨*package_content*⟩ ::= [⟨*comment_block*⟩] ['expand'] ⟨*package_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*struct_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*enum_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*entity_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ['expand'] ⟨*interface_block*⟩ [⟨*override_block*⟩]

⟨*comment_block*⟩ ::= ⟨*comment*⟩+

⟨*comment*⟩ ::= R'//.*\n'

⟨*package_block*⟩ ::= 'package' '{' ⟨*package_content*⟩* '}'

⟨*struct_block*⟩ ::= 'struct' ⟨*identifier*⟩ [⟨*extends_block*⟩] [⟨*implements_block*⟩] '{' ⟨*struct_content*⟩* '}'

⟨*extends_block*⟩ ::= 'extends' ⟨*reftype*⟩

⟨*implements_block*⟩ ::= 'implements' ⟨*reftype*⟩ (',' ⟨*reftype*⟩)+

⟨*struct_content*⟩ ::= [⟨*comment_block*⟩] ⟨*arg_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*ref_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*multi_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*func_block*⟩ [⟨*override_block*⟩]

⟨*arg_block*⟩ ::= 'arg' ⟨*primitive_type*⟩ ⟨*identifier*⟩ ';'

⟨*ref_block*⟩ ::= 'ref' ⟨*reftype*⟩ ⟨*identifier*⟩ ';'

⟨*multi_block*⟩ ::= 'ref' ⟨*multi_name*⟩ '<' ⟨*primitive_type*⟩ [',' ⟨*primitive_type*⟩] '>' ⟨*identifier*⟩ ';'
| 'ref' ⟨*multi_name*⟩ '<' ⟨*reftype*⟩ [',' ⟨*primitive_type*⟩] '>' ⟨*identifier*⟩ ';'
| 'ref' ⟨*multi_name*⟩ '<' ⟨*primitive_type*⟩ [',' ⟨*reftype*⟩] '>' ⟨*identifier*⟩ ';'
| 'ref' ⟨*multi_name*⟩ '<' ⟨*reftype*⟩ [',' ⟨*reftype*⟩] '>' ⟨*identifier*⟩ ';'

⟨*func_block*⟩ ::= 'func' ⟨*reftype*⟩ ⟨*identifier*⟩ '(' [⟨*param_definition*⟩ ( ',' ⟨*param_definition*⟩)*] ')' ';'
| 'func' ⟨*primitive_type*⟩ ⟨*identifier*⟩ '(' [⟨*param_definition*⟩ ( ',' ⟨*param_definition*⟩)*] ')' ';'
| 'func' 'void' ⟨*identifier*⟩ '(' [⟨*param_definition*⟩ ( ',' ⟨*param_definition*⟩)*] ')' ';'

⟨*param_definition*⟩ ::= ⟨*reftype*⟩ ⟨*identifier*⟩
| ⟨*primitive_type*⟩ ⟨*identifier*⟩

⟨*enum_block*⟩ ::= 'enum' ⟨*identifier*⟩ '{' ⟨*enum_content*⟩* '}'

⟨*enum_content*⟩ ::= [⟨*comment_block*⟩] ⟨*arg_block*⟩ [⟨*override_block*⟩]
| [⟨*comment_block*⟩] ⟨*enum_constant*⟩ [⟨*override_block*⟩]

⟨*enum_constant*⟩ ::= ⟨*identifier*⟩ '(' ⟨*enum_index*⟩ ( ',' ⟨*primitive_value*⟩)* ')' ';'

⟨*enum_index*⟩ ::= '_' | ⟨*integerValue*⟩

⟨*entity_block*⟩ ::= 'entity' ⟨*identifier*⟩ [⟨*extends_block*⟩] [⟨*implements_block*⟩] '{' ⟨*struct_content*⟩* ⟨*identifier_statement*⟩ '}'

⟨*identifier_statement*⟩ ::= 'identifier' '(' ⟨*identifier*⟩ (',' ⟨*identifier*⟩)* ')' ';'

⟨*interface_block*⟩ ::= 'interface' ⟨*identifier*⟩ [⟨*implements_block*⟩] '{' ⟨*interface_content*⟩* '}'

⟨*interface_content*⟩ ::= [⟨*comment_block*⟩] ⟨*func_block*⟩ [⟨*override_block*⟩]

## 3.2 Beispiel

⟨*statement*⟩ ::= ⟨*ident*⟩ '=' ⟨*expr*⟩
| 'for' ⟨*ident*⟩ '=' ⟨*expr*⟩ 'to' ⟨*expr*⟩ 'do' ⟨*statement*⟩
| '{' ⟨*stat-list*⟩ '}'
| ⟨*empty*⟩

⟨*stat-list*⟩ ::= ⟨*statement*⟩ ';' ⟨*stat-list*⟩ | ⟨*statement*⟩