Chapter 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 "Schweizer 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.

Chapter 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.