

Konzepte zur Bachelorarbeit

Template-basierte Synthese von
Verzweigungsstrukturen mittels L-Systemen

Adrian Helberg

9. Februar 2021

Inhaltsverzeichnis

1	Literaturübersicht	2
1.1	Softwaretechnik	2
1.2	Grundlagen	2
1.2.1	L-Systeme	2
1.2.2	Logo-Turtle-Algorithmus	4
1.3	Basisquelle	5
2	Eigenleistung	5
3	Dokumentation	6
3.1	Gliederung	6
4	Softwareprojekt	7
4.1	Vorgehen	7
4.2	Technnologien	8
5	Softwarearchitektur	9
5.1	Einführung und Ziele	9
5.2	Kontextabgrenzung	10
5.3	Lösungsstrategie	11
5.4	Bausteinsicht	14
5.5	Laufzeitsicht	15
5.6	Verteilungsicht	15
5.7	Konzepte	16
5.7.1	Testbarkeit	16
5.7.2	Validierung	16

5.7.3	Fehlerbehandlung	16
5.7.4	Datenstrukturen	17
5.7.5	Workflows & Algorithmen	17
5.7.6	Entscheidungen	20
Quellen		21

1 Literaturübersicht

1.1 Softwaretechnik

Eine Fallstudie der Universität Karlsruhe [?] untersucht den Einsatz der Softwaretechnik **Extreme Programming** (XP) im Kontext der Erstellung von Abschlussarbeiten im Universitätsumfeld. Hierzu werden folgende Schlüsselpraktiken untersucht:

- XP als Softwaretechnik zur schrittweisen Annäherung an die Anforderungen eines Systems
- Änderung der Anforderungen an das Systems
- Funktionalitäten (**Features**) werden als Tätigkeiten des Benutzers (**User Stories**) definiert
- Zuerst werden Komponententests (Modultests) geschrieben und anschließend die Features (Test-driven Design)
- Keine separaten Testing-Phasen
- Keine formalen Reviews oder Inspektionen
- Regelmäßige Integration von Änderungen
- Gemeinsame Implementierung (Pair Programming) in Zweiergruppen

Aus der Fallstudie geht hervor, dass Extreme Programming einige Vorteile bei der Bearbeitung eines Softwareprojektes einer Bachelorarbeit bietet. Zum einen können sich Anforderungen an das zu erstellende System durch parallele Literaturrecherche ändern, zum anderen können Arbeitspakete durch Releases abgebildet werden.

1.2 Grundlagen

Im Folgenden wird auf relevante, grundlegende Themen eingegangen und zugehörige, wissenschaftliche Quellen vorgestellt.

1.2.1 L-Systeme

Lindenmayer [?] führt eine mathematische Beschreibung zum Wachstum fadenförmiger Organismen ein. Sie zeigt, wie sich der Status von Zellen infolge ein oder mehrerer Einflüsse verhält. L-Systeme sind Ersetzungssysteme, die

zur formalen Beschreibung von Zellteilung eingeführt werden, und atomare Teile mithilfe von Produktionsregeln ersetzen. Weiter werden Symbole zur formalen Beschreibung von Verzweigungen, die von Filamenten abgehen, genutzt. Die bekanntesten L-Systeme sind zeichenketten-basiert und werden von *Noam Chomsky* eingeführt. Sie ersetzen parallel Buchstaben eines Wortes, die von einer Grammatik über eine Sprache akzeptiert werden. L-Systeme können unter anderem parametrisiert oder nicht-parametrisiert und kontextfrei oder kontextsensitiv sein.

Formalismen

Ein L-System ist ein Tupel und hat folgende Form:

$$\mathcal{L} = \langle M, \omega, R \rangle, \text{ mit}$$

- M als Alphabet, das alle Symbole enthält, die in der Grammatik vorkommen können,
- ω als Axiom oder „Startwort“ und
- R als Menge aller Produktionsregeln, die für \mathcal{L} gelten

Das Alphabet eines parametrischen Systems enthält Module (Symbole mit Parametern) anstatt Symbole:

$$M = \{A(P), B(P), \dots\} \text{ mit}$$

- $P = p_1, p_2 \dots$ als Modulparameter

Zeichen des Alphabets, die Ziel einer Produktionsregel sind, heißen Variablen. Alle anderen Zeichen aus M sind die Konstanten.

Das Axiom ω ist eine nicht-leere Sequenz an Modulen aus M^+ mit

- M^+ als Menge aller möglichen Zeichenketten aus Modulen aus M

Produktionsregeln sind geordnete Paare aus Wörtern über dem Alphabet, die bestimmte Ersetzungsregeln umsetzen. Hierbei werden Symbole aus einem Wort, die einer rechten Seite (*engl. right hand side (RHS)*) einer Produktionsregel entsprechen, durch die linke Seite des Paares (*engl. left hand side (LHS)*) ersetzt. Sie sind folgendermaßen aufgebaut:

$$A(P) \rightarrow x, x \in M^*$$

M^* ist die Menge aller möglichen Zeichenketten von M inklusive der leeren Zeichenkette ε . Ist die RHS jeder Produktionsregel ein einzelnes Symbol und gibt es zu jeder Variablen eine Regel, spricht man von einem kontextfreien, andernfalls von einem kontextsensitiven L-System.

1.2.2 Logo-Turtle-Algorithmus

Der Logo-Turtle-Algorithmus [?] beschreibt ein Vorgehen zur graphischen Beschreibung von L-Systemen, bei dem jeder Buchstabe in einem Wort einer bestimmten Zeichenoperation zugewiesen wird. So kann aus einem L-System ein grafisches Muster generiert werden, das mit einer Abfolge von Zeichenbefehlen an eine „Schildkröte“ gezeichnet wird. Das Triplet (x, y, θ) definiert den Status (**State**) der Schildkröte. Dieser setzt sich aus der aktuellen Position $\begin{pmatrix} x \\ y \end{pmatrix}$ und dem aktuellen Rotationswinkel θ , der die Blickrichtung bestimmt, zusammen.

Der Algorithmus kann als Komprimierung eines geometrischen Musters gesehen werden. Folgende Symbole mit zugehörigen Steuerungsbefehlen und Statusveränderung sind definiert:

Symbol	Steuerung	Statusveränderung
$F(d)$	Gehe vom derzeitigen Punkt p_1 d Einheiten in die Blickrichtung zu dem Punkt p_2 . Zeichne ein Liniensegment zwischen p_1 und p_2	ja, neue Position p_2
$+(\alpha)$	Setzt neuen Rotationswinkel $\theta = \theta + \alpha$	ja, neuer Rotationswinkel θ
$-(\alpha)$	Setzt neuen Rotationswinkel $\theta = \theta - \alpha$	ja, neuer Rotationswinkel θ
[Lege den aktuellen State auf einen Stack	nein
]	Hole den State vom Stack und überschreibe den aktuellen mit diesem	nein

Alles zwischen den Symbolen [und] wird als Verzweigung interpretiert.

Bsp. $FF[FF]F$ mit Verzweigung $[FF]$

1.3 Basisquelle

Bei *Inverse Procedural Modeling of Branching Structures by Inferring L-Systems*[?] geht es um ein Modell zum Lernen von L-Systemen von Verzweigungsstrukturen mithilfe maschinellen Lernens (*Deep Learning*) anhand beliebiger Grafiken. Hierzu werden atomare Strukturen mit einem neuronalen Netz erkannt, eine hierarchische Topologie (Baumstruktur) aufgebaut, aus der ein L-System inferiert und mit einem **Greedy** Algorithmus optimiert wird. Ausgabe des Systems ist ein generalisiertes L-System, aus dem ähnliche Strukturen, wie die der Inputgrafik, erstellt werden können.

Aus dieser Quelle werden folgende Konzepte genutzt:

- Nutzer einer Baumstruktur zur Organisation von genutzten atomaren Verzweigungsstrukturen (**Templates**) mit Knoten für Templates und Kanten für geometrische Transformationen
- Untersuchen der Baumstruktur auf Wiederholungen
- Parametrisierte L-Systeme (L-System mit **Modulen**) zur Abbildung von Transformationsparametern
- Kostenfunktion zur Bewertung eines L-Systems

2 Eigenleistung

Die Eigenleistung der Bachelorarbeit besteht aus:

- Algorithmus Baumstruktur \rightarrow L-System
- Leere Symbole (ε) als Dummy-Knoten für Terminale
- Parameterverteilung als Histogramm
- Gewichtetes Randomisieren von Parametern
- Vergleich String-Matching Algorithmen
- ...

3 Dokumentation

3.1 Gliederung

Im Folgenden wird eine vorläufige Gliederung der schriftlichen Ausarbeitung gezeigt

- i. Abbildungs- und Tabellenverzeichnis
- ii. Abkürzungsverzeichnis
- 1. Einleitung
 - 1.1. Problemstellung
 - 1.2. Ziele
 - 1.3. Methodik
 - 1.4. Aufbau
- 2. Grundlagen
 - 2.1. Grundbegriffe
 - 2.2. Grundlegende Arbeiten
 - 2.3. Verwandte Arbeiten
- 3. Konzepte
 - 3.1. Probleme & Lösungsansätze
 - 3.2. Architektur
 - 3.3. Algorithmen
- 4. Implementierung
- 5. Evaluierung
 - 5.1. Testumgebung
 - 5.2. Beobachtungen & Ergebnisse
 - 5.3. Diskussion und Bewertung
- 6. Ausblick
- iii. Literaturverzeichnis
- iv. Eidesstattliche Erklärung

4 Softwareprojekt

4.1 Vorgehen

Das Programm zu dieser Arbeit wird mit einem XP-basierten Ansatz erarbeitet. Hierbei beinhaltet ein **Release** Funktionen, die insgesamt für eine neue Version des Systems ausreichen; also ein vollständig funktionsfähiges Programm liefern. **User Stories** sind innerhalb der Iterationen umzusetzende Teilaufgaben und deren Aufwandseinschätzung gibt Auskunft über den Entwicklungsaufwand einer Umsetzung.

Umsetzung des Softwareprojektes in Iterationen mit folgenden Phasen:

- Planung:
 - Release-Planung:
„*Welche Features werden in diesem Release umgesetzt?*“,
User Stories, Aufwandsschätzung, Anforderungsmanagement
 - Iterationsplanung:
Umwandlung der User Stories in kleine Arbeitsschritte,
Festlegen der Dauer einer Implementierung
- Entwurf: Architektur, Klassendiagramme, Schnittstellen
- Testing: (Automatisierte) Modul- und Regressionstests
- Programmierung: Umsetzung der Features, Implementierung, Modularisierung

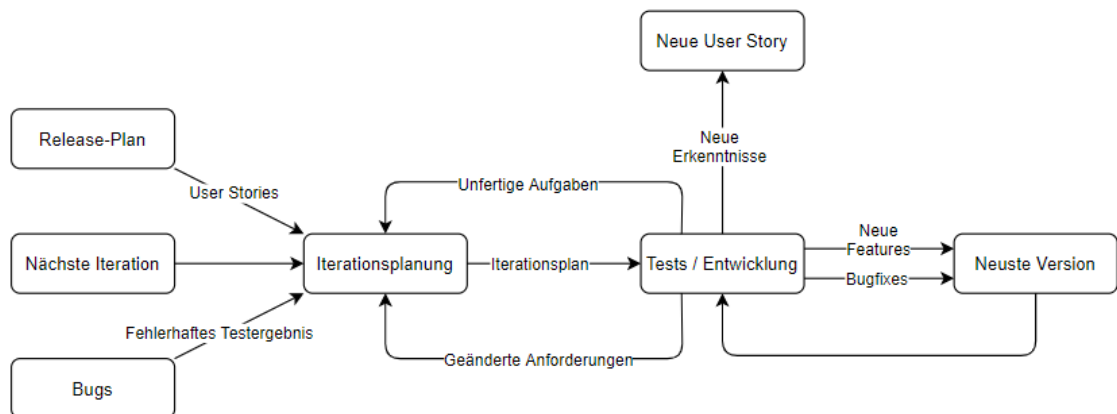


Abbildung 1: Ablaufdiagramm

4.2 Technnologien

- Programmiersprache: Java Version 11 mit
 - JavaFX Version 15 (openjfx)
- Build-Management-Tool: Gradle[?] Version 6.7
- Versionskontrolle: Github Repository[?] via Git[?]
- IDE: JetBrains IntelliJ IDEA[?] 2020.2.2 (Ultimate Edition)
- Betriebssystem: Microsoft Windows 10 Pro 64 Bit
- Prozessor: Intel Core i5-3570K CPU @ 3.40GHz
- User Story Map: Trello Board[?]

5 Softwarearchitektur

Die Gliederung der Inhalte für die Softwarearchitektur erfolgt nach der arc42-Vorlage [?]

5.1 Einführung und Ziele

Ziel ist die Erstellung eines Programms zur Synthetisierung von Ähnlichkeitsabbildungen einer vom Benutzer erstellten Verzweigungsstruktur mittels Inferieren und Optimieren von L-Systemen. Die wesentlichen Features des Programms sind:

- Erstellung einer Verzweigungsstruktur über eine grafische Benutzeroberfläche (**GUI**)
- Einbindung atomarer Strukturen über externe Dateien
- Synthetisierung von ähnlichen Verzweigungsstrukturen anhand der erstellten Struktur
- Anzeigen von Verzweigungsstrukturen

Priorisierte (absteigend) Qualitätsziele, die bei der Erstellung des Systems umgesetzt werden sollten:

- **Funktionalität** durch Umsetzung aller Teilsysteme
- **Interoperabilität** durch Nutzen einer allgemeinen Repräsentation von L-Systemen, damit diese auch in anderen Programmen oder Algorithmen verwendet werden kann
- **Erweiterbarkeit** durch offene Entwurfsmuster (Design Pattern)
- **Modulare** Implementierung für effiziente Wartung und Erweiterung
- **Effizienz** durch effiziente Programmierung
- **Attraktivität** durch intuitive Benutzung (Benutzerfreundlichkeit)
- **Plattformunabhängigkeit** durch Verwenden des Java-Frameworks

5.2 Kontextabgrenzung

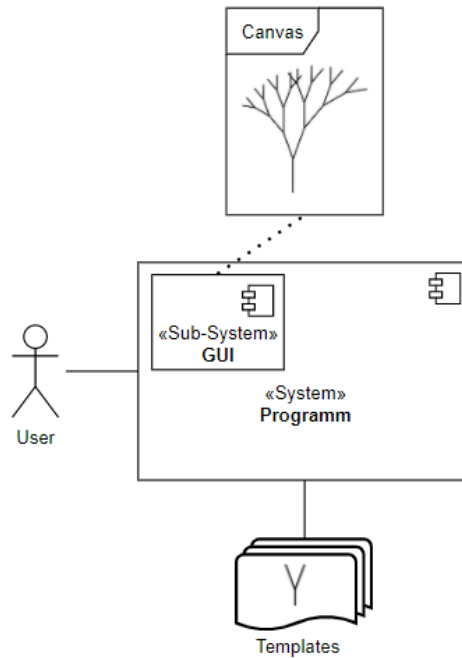


Abbildung 2: System und Systemumgebung

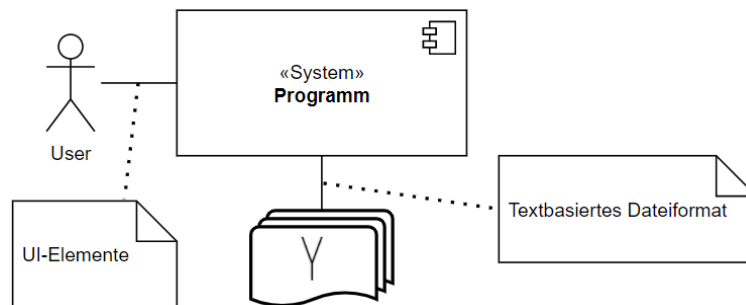


Abbildung 3: Technische Interaktion zwischen System und Systemumgebung

5.3 Lösungsstrategie

Gewählte Architekturansätze zur Erreichung der Qualitätsziele:

Qualitätsziel	Architekturansatz
Funktionalität	<ul style="list-style-type: none">• Grafische Benutzerschnittstelle: <i>GUI</i>• Generieren der Baumstruktur: <i>TreeGenerator</i>• Ableiten von L-Systemen aus Baumstrukturen: <i>Inferer</i>• Generalisieren von L-Systemen: <i>Generalizer</i>• Randomisieren von L-System Parametern: <i>Randomizer</i>
Interoperabilität	Durch das Nutzen allgemeingültiger mathematischer Beschreibungen sollen erstellte L-Systeme in Fremdsystemen, wie Online Visualisierer, genutzt werden können
Erweiterbarkeit	Das Nutzen des Pipeline Design Patterns soll das Erweitern des Systems durch Hinzufügen weiterer Teilschritte (Pipes) erleichtern. Trennung der grafischen Oberfläche und der Logik durch Aufbauen des Szenengraphen über ein XML-Dateiformat
Modularität	Sowohl eine sinnvolle Aufteilung von Funktionalitäten auf Dateien und Pakete (Packages), als auch effiziente Datenkapselung und geschlossene Informationskontexte sorgen für Modularität des Programms

Die Implementierung des Programms setzt sich aus folgenden Teilschritten zusammen:

- Erstellung der *GUI* (Paket *gui*, *tree*) mit
 - UI-Elementen
 - Render-Canvas
 - Dateianbindung der Templates
 - Erstellung der repräsentativen Baumstruktur (*treeGenerator*, Paket *tree*)
- Implementierung der Subsysteme als Pipes
 - *Inferer* (Paket *grammar*): Ableiten eines kompakten L-Systems aus einer Baumstruktur
 - *Generalizer* (Paket *grammar*): Generieren eines generalisierten L-Systems anhand eines „kleinen“ L-Systems
 - *Randomizer* (Paket *grammar*): Erzeugung von L-Systemen, die der erstellen Verzweigungsstruktur „ähnlich“ sind
- Komponenten- und Systemtests

Subsystem	Umsetzung
GUI	JavaFX als Framework zur Erstellung von grafischen und interaktiven Inhalten. Erstellung der Baumstruktur über dynamisches Erzeugen von Konten während der Strukturierung der Verzweigungsstruktur
Inferer	Algorithmus zum Iterieren maximaler Sub-Bäume und deren Reduzierung mittels Ersetzung durch Symbole und der zugehörigen Produktionsregel, bis eine Kostengrenze, die durch eine Kostenfunktion abgebildet werden kann, erreicht ist
Generalizer	Algorithmus zum Erweitern eines L-Systems um nicht-deterministischer Regeln und Erkennen rekursiver Strukturen
Randomizer	ddd sfevdhbnreaydydbfsdxc

5.4 Bausteinsicht

Ebene 1

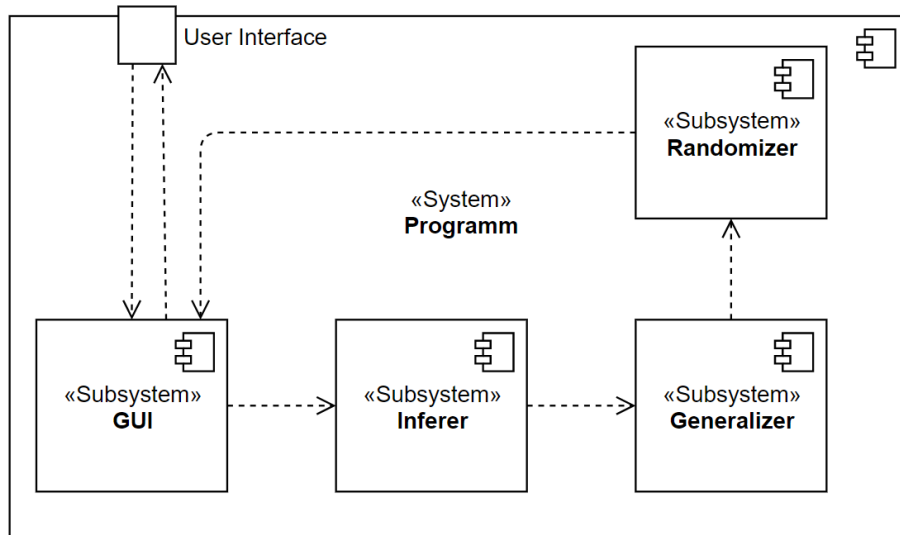


Abbildung 4: Subsysteme mit fachlichen Abhängigkeiten

Ebene 2

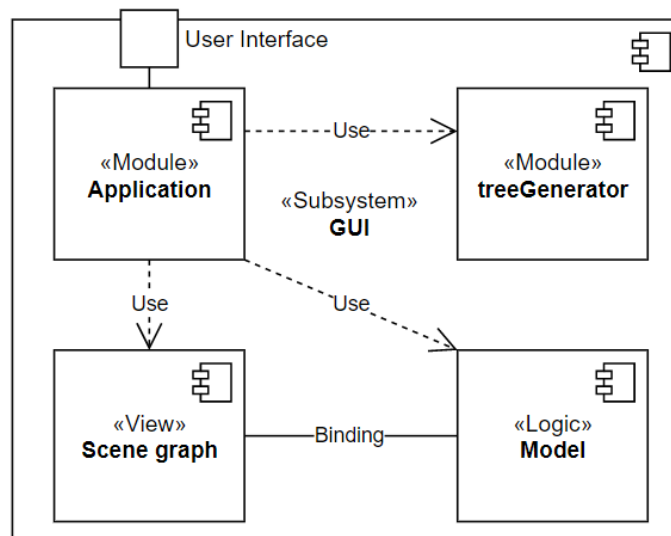


Abbildung 5: Subsystem GUI

5.5 Laufzeitsicht

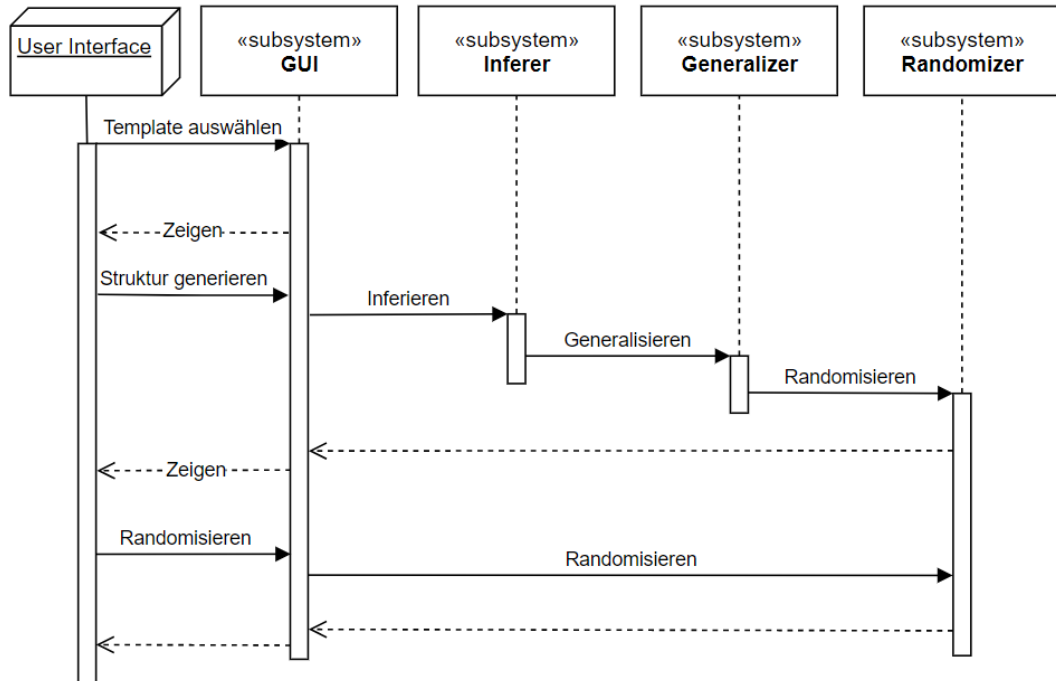


Abbildung 6: Laufzeitsicht

5.6 Verteilungssicht

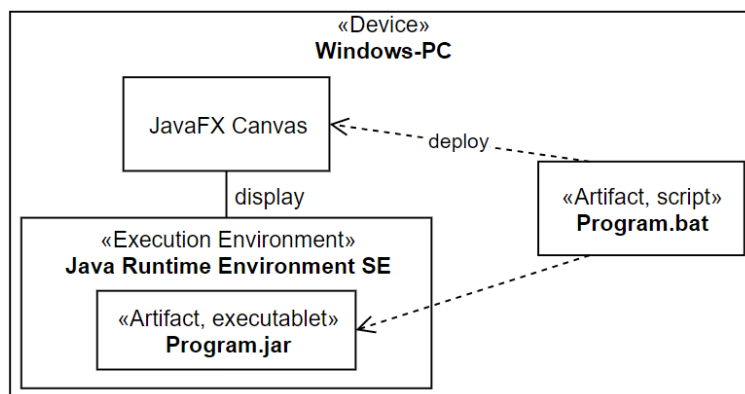


Abbildung 7: Infrastruktur Windows-PC

5.7 Konzepte

5.7.1 Testbarkeit

Um eine ausreichende Testabdeckung zu erreichen, werden Klassen als kleinstmögliche zu testende Einheit definiert und durch Komponententests geprüft. Der Name eines Tests setzt sich aus dem Präfix als Name der zu testenden Klasse oder Releases und dem Suffix „Test“ zusammen. Testsubjekte werden als **Blackbox** behandelt, also anhand der Spezifikation getestet.

Bsp.: Klasse *Inferer* mit Komponententest *InfererTest*

Da jedes Release der Implementierung ein funktionsfähiges System beinhaltet, kann auf Integrationstests verzichtet werden. Weiter wird ein Release anhand von funktionalen und nicht-funktionalen Anforderungen getestet. Anhand dieser Systemtests wird geprüft, ob Gesamtspezifikationen umgesetzt worden sind.

Bsp.: Release 2 mit Systemtest *Release2Test*

Zum Schluss der Implementierung wird ein Akzeptanztest durchgeführt.

5.7.2 Validierung

Der Benutzer des Systems nutzt eine grafische Schnittstelle. Somit kann sichergestellt werden, dass dieser keine ungültigen Eingaben tätigt. Werden Template-Dateien unter einem bestimmten Dateipfad nicht gefunden, wird eine *NotFound-Exception* protokolliert und dem Benutzer eine Nachricht über ein Pop-up mitgeteilt. Eine *IllegalArgumentException* wird verzeichnet und ausgegeben, wenn eingelesene Template-Dateien strukturelle Fehler aufweisen.

5.7.3 Fehlerbehandlung

Zur Fehlersuche und -behandlung biete sich eine Protokollierung über Vorgänge, Fehler und Ausnahmen an. Bestimmte Fehler werden dem Benutzer weitergegeben und grafisch angezeigt. Folgende Subsysteme werden in das Programm integriert:

- Ausnahmebehandlung (Exception Handling) und
- Protokollierung (Logging)

Erwartete Exceptions werden jeweils mit eigenen Klassen abgebildet, die von einer entsprechenden Klasse aus der Exception-Hierarchie abgeleitet sind. Das Logging ist statisch und überall im System zugreifbar, um eine einheitliche Protokollierung zu gewährleisten.

5.7.4 Datenstrukturen

TODO: Klassendiagramm aus IntelliJ generieren

5.7.5 Workflows & Algorithmen

Um als Benutzer des Systems eine Verzweigungsstruktur zu erstellen, wird folgender Arbeitsablauf umgesetzt:

Algorithmus 1: Erstellen einer Verzweigungsstruktur

```

1  Erster Anker ist vorselektiert
2  Wiederhole, bis Struktur fertiggestellt ist:
3      Selektiere ein Template aus der Liste
4      Setzt Parameter
5      Bestätige Auswahl und Parameter
6      Zeichne ausgewähltes Template mit Parametern
7      Wähle nächsten Anker aus

```

Aus der Verzweigungsstruktur kann nun ein L-System erzeugt werden:

Für $\mathcal{L} = \langle M, \omega, R \rangle$:

Algorithmus 2: Inferieren eines L-Systems aus einer Baumstruktur

```

1  Initialisierung:
2       $M = \{F, S\}$ 
3       $\omega = S$ 
4       $R \leftarrow \{\alpha: S \rightarrow A\}$ 
5       $\beta = \text{nächster Knoten*}$ 
6       $M \leftarrow \gamma \in \{A, B, \dots, Z\}, \text{ mit } \gamma \notin M$ 
7
8  Schleife:
9       $\delta = \text{Wort von } \beta$ 
10      $\forall \{X, Y, Z\} \in \delta:$ 
11         Ersetze mit  $\zeta \in \{A, B, \dots, Z\}, \text{ mit } \zeta \notin M$ 
12          $M \leftarrow \zeta$ 
13      $R \leftarrow \{\gamma \rightarrow \delta\}$ 
14     Wenn es ein Symbol  $\eta$  in  $M \setminus \{F, S\}$  gibt mit  $\{\eta \rightarrow \text{bel.}\} \notin R:$ 
15          $\gamma = \eta$ 
16     Sonst:
17         Breche Schleife ab
18      $\beta = \text{nächster Knoten*}$ 

```

* nach Breitensuche, beginnend bei Wurzelknoten S

Um ein kompakteres, gewichtetes L-System zu erzeugen, werden sich wiederholende Unterbäume gesucht und ersetzt. Die Gewichtung wird angewendet, um das erzeugte L-System in ein System mit kleiner Regelmenge ($w_l = 1$) oder mit großer Regelmenge ($w_l = 0$) umzuwandeln:

Algorithmus 3: Erstellen eines kompakten L-Systems mit Gewichtung w_l

```

1 Initialisierung :
2    $\mathcal{L}^+ \leftarrow L_s^*$ 
3    $\mathcal{L} = \emptyset$ 
4   Setze Gewichtsparameter  $w_l \in [0, 1]$ 
5   Finde maximalen Unterbaum  $T'$  aus  $T^{**}$  mit Wiederholungen  $n$ 
6
7 Reduzierung :
8 if  $n > 1$ 
9   Ersetze alle Vorkommen von  $T'$  mit dem selben Symbol  $\gamma \in \{A, B, \dots, Z\}$ 
10   $R \leftarrow \{\gamma \rightarrow L_s\}$  mit  $L_s$  aus  $T'$ ,  $R$  aus  $\mathcal{L}$ 
11  if  $C_i(\mathcal{L}) \geq C_i(\mathcal{L}^+)$ 
12    break
13   $T \leftarrow T'$ 
14   $\mathcal{L}^+ \leftarrow \mathcal{L}$ 
15  Finde maximalen Unterbaum  $T'$  aus  $T$  mit Wiederholungen  $n$ 

```

Die Kostenfunktion stellt die Anzahl Symbole aller *RHS* der Produktionsregeln mit der Menge an Anwendungen der *LHS* gegenüber:

Algorithmus 4: Kostenfunktion C_i mit Gewichtung w_l

```

1  $C_i(\mathcal{L}) = \sum_{A(P) \rightarrow M^* \in \mathcal{L}} w_l * |M^*| + (1 - w_l) * N(A(P) \rightarrow M^*)^{***}$ 

```

Da das kompakte L-System eine Repräsentation der vom Benutzer erzeugten Verzweigungsstruktur darstellt, werden nun ähnliche Regeln miteinander verbunden und mit einer Wahrscheinlichkeit versehen, um nicht-deterministische Regeln hinzuzufügen.

Algorithmus 5: Längenfunktion L für Grammatiken

```

1  $L(\mathcal{L}) = |M| + \sum_{A(P) \rightarrow M^* \in \mathcal{L}} |M^*|$ 

```

Algorithmus 6: Kostenfunktion C_g mit Gewichtung w_0

```

1  $C_g(\mathcal{L}^*, \mathcal{L}^+) = w_0 * (L(\mathcal{L}^*) - L(\mathcal{L}^+)) + (1 - w_0) + D_g(\mathcal{L}^+, \mathcal{L}^*)$ 

```

* L_s als Zeichenkette des ausgeführten L-Systems $\mathcal{L} = \langle M, \omega, R \rangle$

** T als Baumstruktur des L-Systems

*** $N(\cdot)$ als Zählfunktion für die Anzahl Wiederholungen einer *LHS* einer Regel in einem ausgeführten L-System

Algorithmus 7: Generalisieren eines L-Systems mit Gewichtung w_0

```

1 Initialisierung :
2   Regelpaar  $p^* = \emptyset$ 
3    $\mathcal{L}^* = \mathcal{L}^+$ 
4    $C_g^{old} = C_g(\mathcal{L}^* + \{p^*\}, \mathcal{L}^*)$ 
5
6 Schleife :
7 do
8   Finde Regelpaar  $p^*$  mit minimalen Kosten  $C_g(\mathcal{L}^* + \{p_i\}, \mathcal{L}^*), \forall p_i \in \mathcal{P}^*$ 
9   if  $C_g(\mathcal{L}^* + \{p^*\}, \mathcal{L}^*) \geq 0$ 
10    break
11    $c^* = C_g(\mathcal{L}^* + \{p^*\}, \mathcal{L}^*) - C_g^{old}$ 
12    $C_g^{old} = C_g(\mathcal{L}^* + \{p^*\}, \mathcal{L}^*)$ 
13    $\mathcal{L}^* = \mathcal{L}^* + \{p^*\}$ 
14 while  $c^* \leq 0$ 

```

5.7.6 Entscheidungen

Mutable or Immutable Objects?

Risiken

Qualitätsmerkmale

Alternativen

Aufwand der Implementierung

* \mathcal{P} als Menge aller möglichen Regelpaaren aus \mathcal{L}^*

