

Megamodel-driven Traceability Recovery of Correspondence & Conformance Links among O/R/X-Mapping Artifacts

Bachelorarbeit

zur Erlangung des Grades eines Bachelor of Science
im Studiengang Informatik

vorgelegt von

Maximilian Meffert

Erstgutachter: Prof. Dr. Ralf Lämmel
Institut für Informatik

Zweitgutachter: Msc. Johannes Härtel
Institut für Informatik

Koblenz, im November 2017

Erklärung

Hiermit bestätige ich, dass die vorliegende Arbeit von mir selbständig verfasst wurde und ich keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe und die Arbeit von mir vorher nicht in einem anderen Prüfungsverfahren eingereicht wurde. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD-Rom).

Ja Nein

Mit der Einstellung der Arbeit in die Bibliothek bin ich ein- ☐ ☐
verstanden.

Der Veröffentlichung dieser Arbeit im Internet stimme ich ☐ ☐
zu.

.....
(Ort, Datum) (Maximilian Meffert)

Zusammenfassung

TBD.

Abstract

TBD.

Acknowledgements

TBD.

Contents

1	Introduction	1
2	Background	2
2.1	Mereology	2
2.2	Formal Languages & Grammars	2
2.2.1	Context-Free Languages & Grammars	2
2.3	Traceability	2
2.3.1	Traceability Relationship	3
2.3.2	Traceability Link	3
2.3.3	Traceability Recovery	3
2.3.4	Traceability Exploration	3
2.4	Ontologies	3
2.5	Megamodeling	3
2.5.1	MegaL	3
2.5.2	MegaL/Xtext	3
2.6	Java Architecture for XML Binding (JAXB)	3
2.7	Hibernate	3
2.8	Another Tool For Language Recognition (ANTLR)	3
3	Related Work	4
4	Hypotheses	5
4.1	Fragments	5
4.1.1	Fragment Languages	8
4.2	Correspondence	8
4.3	Conformance	8

5	Methodology	9
5.1	Example Driven Development Process	9
5.2	Example Corpus	10
5.2.1	The 101HRMS Model	11
5.2.2	Linguistic Domains of the Example Corpus	11
6	Requirements	13
7	Design	14
7.1	Recovery System Design	14
7.1.1	Recovery Process	14
7.1.2	Recovery API	15
7.1.3	Recovery System	21
7.2	Megal/Xtext Integration Design	22
8	Implementation	25
8.1	Recovering Fragments	25
8.1.1	Concrete Fragment Models	26
8.1.2	Megamodeling Fragments	29
8.2	Recovering Parthood Links	31
8.3	Recovering Correspondence & Conformance Links	31
9	Results	32
10	Conclusion	33
	Glossary	34

List of Figures

5.1	The Example Driven Development Process	10
5.2	The 101 Human Resource Management System Model	11
5.3	Example Corpus Domains: Java O/R/X	12
7.1	The Recovery Process	14
7.2	The Recovery API	15
7.3	The Abstract Fragment Model	17
7.4	The Syntax Analysis API	18
7.5	IParserFactory Usage Example	18
7.6	AntlrParser Parse Tree Creation	18
7.7	AntlrParser Fragment Creation	19
7.8	Mereological Fragment Analysis API	20
7.9	Comparative Fragment Analysis API	21
7.10	The Recovery System	21
7.11	Integration of the Recovery System int MegaL/Xtext	23
7.12	MegaL Plug-In Instantiation	23
8.1	Idealized Recovery of Java Fragments	27
8.2	Java Fragment AST Model	28
8.3	Construction of Java Fragment ASTs	29
8.4	A Megamodel for Java Fragments	30

Chapter 1

Introduction

TBD.

Contributions

Non-Contributions

Chapter 2

Background

This chapter summarizes the necessary background topics of the thesis. Each topic is introduced as independent as possible, interrelation is done during synthesis of hypotheses for this thesis in chapter 4. However, some of the following sections are sorted in a way that one may be based on its predecessor. In particular sections on the necessary mathematical background share definitions. Introducing them repeatedly would be redundant.

2.1 Mereology

2.2 Formal Languages & Grammars

[14]

2.2.1 Context-Free Languages & Grammars

2.3 Traceability

[17] [5] [8] TBD.

2.3.1 Traceability Relationship

2.3.2 Traceability Link

2.3.3 Traceability Recovery

2.3.4 Traceability Exploration

2.4 Ontologies

TBD. [6]

2.5 Megamodeling

TBD. [1] [3]

2.5.1 MegaL

[10] [2] [9]

2.5.2 MegaL/Xtext

[7]

2.6 Java Architecture for XML Binding (JAXB)

2.7 Hibernate

2.8 Another Tool For Language Recognition (ANTLR)

[11]

Chapter 3

Related Work

TBD.

Chapter 4

Hypotheses

This chapter introduces the hypotheses this thesis is based upon. TBD.

4.1 Fragments

The Oxford English Dictionary defines fragment¹ as “[...] *isolated or incomplete part of something* [...]”. Regarding natural languages, sentence fragments are incomplete sentences lacking one or more grammatical elements. **ToDo: Lacks citation for "sentence fragment"!**

The intuitive notion of a fragment when working with and/or talking over computer programs is, that it is a piece of code containing a functionality of momentary interest. In that regard, we exclusively use the word fragment referring to isolated or incomplete parts of things having a textual representation, e.g. computer programs. However, computer science knows different levels of organization in text, which effects the semantics of textual fragments.

The most basic level of textual organization is no organization at all. Random text is only restricted by its alphabet, which allows us to specify fragmentation solely in terms of concatenation.

¹<https://en.oxforddictionaries.com/definition/fragment> (retrieved 08/28/2017)

Definition 1 (Fragments over Alphabets) Let Σ be an alphabet and $x, y \in \Sigma^*$ be two words over that alphabet. x is a fragment of y , if and only if there is a concatenation containing x and resulting in y :

$$\text{fragmentOf} : \Sigma^* \times \Sigma^* \quad (4.1)$$

$$x \text{ fragmentOf } y :\Leftrightarrow \exists u, v \in \Sigma^* (uxv = y) \quad (4.2)$$

We can also observe from definition 1 that textual fragmentation shares similarities with mereological parthood, i.e. the `fragmentOf` definition is also reflexive, antisymmetric and transitive:

1. Reflexivity: $x \text{ fragmentOf } x$
 $x \text{ fragmentOf } x$ holds for all x since $u = v = \epsilon \in \Sigma^*$.
2. Antisymmetry: $x \text{ fragmentOf } y \wedge y \text{ fragmentOf } x \rightarrow x = y$
 $x \text{ fragmentOf } y \wedge y \text{ fragmentOf } x$ is equivalent to $axb = y \wedge cyd = x$, from which follows $caxbd = x$ and then $a = b = c = d = \epsilon$, thus $x = y$ follows eventually
3. Transitivity: $x \text{ fragmentOf } y \wedge y \text{ fragmentOf } z \rightarrow x \text{ fragmentOf } z$
 $x \text{ fragmentOf } y \wedge y \text{ fragmentOf } z$ is equivalent to $axb = y \wedge cyd = z$, from which follows $caxbd = z$, which is in turn equivalent to $uxv = z$ with $u = ca$ and $v = bd$, thus $x \text{ fragmentOf } z$ follows eventually

There is a universal bottom element, which is fragment of all words:

$$\forall x \in \Sigma^* (\epsilon \text{ fragmentOf } x) \quad (4.3)$$

and there is a universal top element, all words are fragments of:

$$\begin{aligned} \forall x \in \Sigma^* (x \text{ fragmentOf } W) \\ W := w_1 w_2 w_3 \dots \quad w_i \in \Sigma^* \setminus \{\epsilon\}, i \in \mathbb{N} \end{aligned} \quad (4.4)$$

where W is the concatenation of all words in Σ^* and thus is an infinite word. $W \in \Sigma^\omega$ is interpreted as mapping $\mathbb{N} \rightarrow \Sigma$, giving a symbol from the alphabet for its position in the word. **ToDo: citation for infinite words, Handbook on Formal languages**

We can now easily agree upon, that `fragmentOf` is `partOf` for texts. Thus both are logically equivalent relations:

$$\text{fragmentOf} \equiv \text{partOf} \quad (4.5)$$

ToDo: refer to philosophical problems from SEP

A higher level of textual organization are formal languages, i.e. subsets of Σ^* for any alphabet Σ [14]. Given any language specified with set-builder notation:

$$L := \{w \in \Sigma^* \mid \Phi(w)\} \subseteq \Sigma^* \quad (4.6)$$

where $\Phi(w)$ is an arbitrary restriction on w , we can still observe the same behavior regarding fragments, e.g.:

$$L = \{a^n b^n \mid n \in \mathbb{N} : n \geq 2\} \subseteq \{a, b\}^* = \{aabb, aaabbb, aaaabbbb, \dots\}$$

...

ab `fragmentOf` $aabb$

$aabb$ `fragmentOf` $aaabbb$

$aaaabbbb$ `fragmentOf` $aaaabbbb$

...

However, we can only observe the behavior partly for `fragmentOf` : $\Sigma^* \times \Sigma^*$ and `fragmentOf` : $\Sigma^* \times L$, but not for `fragmentOf` : $L \times L$. In other words, the fragments of words in formal languages are not necessarily words of that language itself. This becomes even more apparent if we consider fragments and languages generated by formal grammars.

Definition 2 (Fragments generated by Grammars) Let $G = (V, \Sigma, P, S)$ be a grammar and $x, y \in \Sigma^*$ be two words over its terminal symbols. x is a fragment of y generated by G , if and only if there is a derivation from a concatenation containing x to y :

$$x \text{ fragmentOf}_G y \Leftrightarrow \exists u, v \in (V \cup \Sigma)^* (S \Rightarrow_G^* u x v \Rightarrow_G^* y) \quad (4.7)$$

4.1.1 Fragment Languages

4.2 Correspondence

The Oxford English Dictionary defines correspondence² as “[...] *close similarity, connection, or equivalence* [...]”. [9]

Definition 3 (Correspondence) Let $R \subseteq L_1 \times L_2$ be a relation between two languages L_1 and L_2 . Two artifacts a_1 and a_2 recursively correspond to each other in the sense of R if both artifacts contain parts, which correspond in the same fashion to exactly one part of the other artifact:

$$\begin{aligned}
 & (a_1, a_2) \in R \\
 & \wedge \forall b_1 \in L_1 \exists! b_2 \in L_2 (b_1 \text{ partOf } a_1 \rightarrow (b_2 \text{ partOf } a_2 \wedge b_1 \text{ correspondsTo}_R b_2)) \\
 & \wedge \forall b_2 \in L_2 \exists! b_1 \in L_1 (b_2 \text{ partOf } a_2 \rightarrow (b_1 \text{ partOf } a_1 \wedge b_2 \text{ correspondsTo}_R b_1)) \\
 & \Rightarrow a_1 \text{ correspondsTo}_R a_2
 \end{aligned} \tag{4.8}$$

4.3 Conformance

The Oxford English Dictionary defines conformance³ as synonym for conformity⁴ and thus in turn as “*Compliance with standards, rules, or laws*”.

[9]

Definition 4 (Conformance) Let $L \subseteq \text{Any}$ be a language, $\text{Def}_L \subseteq \text{Any}$ the definition language of L and $\text{def}_L \in \text{Def}_L$ the actual definition artifact of L . An arbitrary artifact conforms to the definition artifact of L , if and only if it is an element of L :

$$\forall x \in \text{Any} (x \text{ conformsTo } \text{def}_L \Leftrightarrow x \in L) \tag{4.9}$$

²<https://en.oxforddictionaries.com/definition/correspondence> (retrieved 08/28/2017)

³<https://en.oxforddictionaries.com/definition/conformance> (retrieved 08/28/2017)

⁴<https://en.oxforddictionaries.com/definition/conformity> (retrieved 08/28/2017)

Chapter 5

Methodology

This chapter summarizes the methodology used to develop the recovery system for this thesis. In short, the recovery system is developed example-driven using the model of a fictional Human Resource Management System (HRMS) used by the 101wiki¹ for its contributions.

5.1 Example Driven Development Process

The Example Driven Development Process used to develop the recovery system for this thesis is shown in Figure 5.1. Given we acquired a suitable example corpus, which is outlined in detail in §5.2, we have to define conditions for acceptable results. These conditions can be seen as functional requirements, for instance, we want a Java class artifact to be linked via correspondence to its XML (Extensible Markup Language)-serialized form:

```
public class Company {...} correspondsTo <company>...</company>
```

Then we implement a feature which produces the specified results and run it against the example corpus. If the result meets all conditions we are done. If not, we inspect the results to check whether our predefined conditions are too vague or wrong. If so, we refine our conditions, if not, we refine our feature implementation. Either way the process starts anew.

¹<https://101wiki.softlang.org/> (retrieved 12th November, 2017)

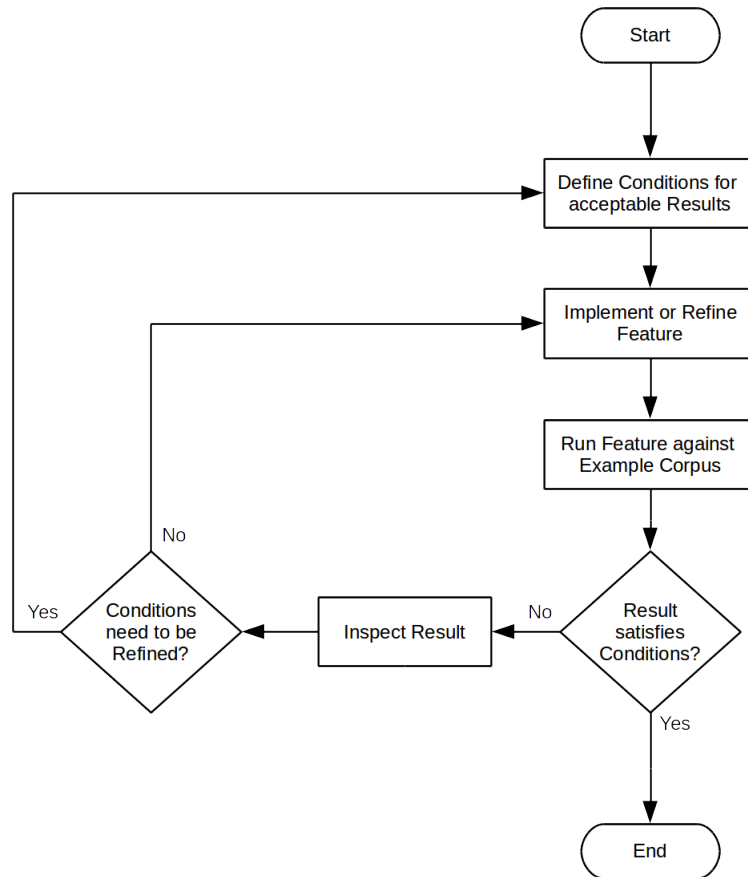


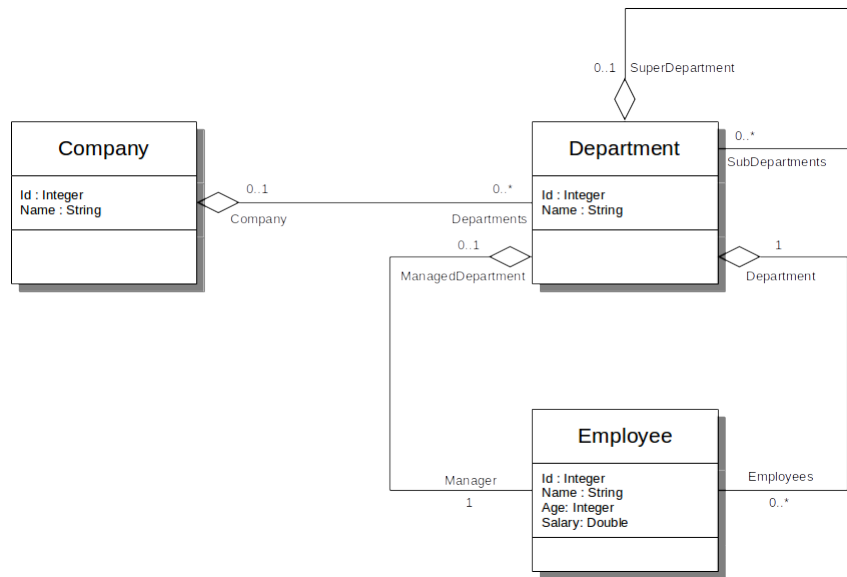
Figure 5.1 The Example Driven Development Process

5.2 Example Corpus

The example corpus used to develop the recovery system for this thesis consists of artifacts implementing a fictional HRMS within an Object-Relational- and XML-Mapping (O/R/X-Mapping) scenario using Java technologies. The model is implemented using plain Java. It is then mapped to plain XML/XSD (XML Schema Definition) with Java Architecture for XML Binding (JAXB) and to SQL/DDl statements using Hibernate mapping files and/or annotations.

5.2.1 The 101HRMS Model

101wiki Human Resource Management System (101HRMS)² provides a simple model of a company with many departments and employees. Figure 5.2 shows an UML (Unified Modeling Language) class diagram of a variant of this model.



This UML class diagram depicts the model of the 101HRMS. It consists of simple companies with nested departments and employees mapped to the latter.

Figure 5.2 The 101 Human Resource Management System Model

The 101HRMS model consists of companies attributed with a name. Each company accumulates departments. Each department is also attributed with a name, aggregates employees and has one employee acting as manager. Departments can further be refined into sub-departments. Each employee is attributed with a name, an age and a salary. Each entity is also attributed with an ID.

5.2.2 Linguistic Domains of the Example Corpus

The example corpus used to develop the recovery system contains artifacts implementing the 101HRMS model generated or used by Java technologies for O/R/X-Mapping, i.e. a Java model is mapped to plain XML/XSD with JAXB, to a Hi-

²<https://101wiki.softlang.org/101:@system> (retrieved 12th November, 2017)

berate mapping file and to SQL (Structured Query Language)/Data Definition Language (DDL) statements. Figure 5.3 shows a schematic illustration of the linguistic domains involved:

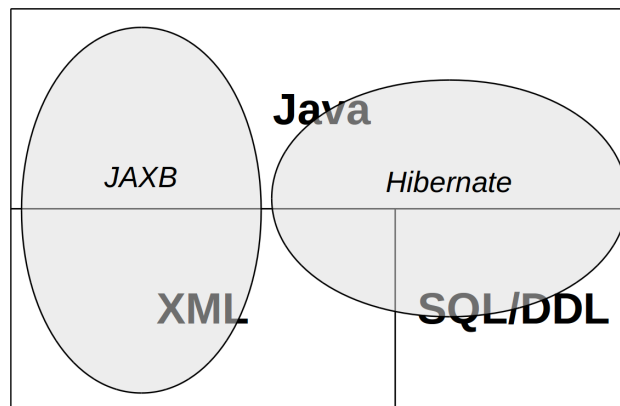
Java The language and technology used to implement the 101HRMS model.

XML The language used to serialize the 101HRMS model.

SQL/DDL The language used to persist the 101HRMS model.

JAXB The technology used to implement Object-XML-Mapping (O/X-Mapping) of the 101HRMS model.

Hibernate The technology used to implement Object-Relational-Mapping (O/R-Mapping) of the 101HRMS model.



This schematic illustration depicts the interrelation among linguistic domains of example corpus used. It depicts languages and technologies for O/R/X-Mapping with Java.

Figure 5.3 Example Corpus Domains: Java O/R/X

The languages (Java, XML & SQL) in Figure 5.3 are displayed as disjoint square sets. Technologies (JAXB & Hibernate) are displayed as oval sets intersecting languages. This is due to their linguistic nature, e.g. JAXB produces specific Java- and XML-Code which does not necessarily intersect with code produced by other technologies. Hibernate intersects all three languages. It uses XML files or Java-Annotations for describing O/R-Mapping of a data-model and generates SQL artifacts according to that mapping. In this sense, technologies create technology-specific subsets of a languages.

Chapter 6

Requirements

This chapter summarizes functional requirements for the recovery system developed as part of this thesis.

Requirement 1 The recovery system has to recover parthood links, i.e. links capturing the relation between syntactically well-formed software artifacts and their fragments.

Requirement 2 The recovery system has to recover correspondences links, i.e. links capturing the relation between fragments of software artifacts denoting a predefined similarity among them and their fragments.

ToDo: refer to a correspondence definition

Requirement 3 The recovery system has to recover conformance links, i.e. links capturing the relation between a software artifacts or fragments denoting that one defines the other.

ToDo: refer to a conformance definition

Requirement 4 The recovery system has to run within MegaL/Xtext, i.e. the implementations of requirements Requirement 1, Requirement 2 and Requirement 3 must be compatible with the MegaL/Xtext plug-in-system.

Chapter 7

Design

This chapter summarizes the design of the recovery system (§7.1) developed for this thesis and its integration in the MegaL/Xtext environment (§7.2).

7.1 Recovery System Design

This section summarizes the design of the recovery system developed for this thesis. §7.1.1 will describe the all over process of the recovery system. §7.1.2 will describe the design core API (Application Programming Interface) and its components developed for the recovery system. §7.1.3 will describe the design of the actual system for recovering links among O/R/X-Mapping artifacts.

7.1.1 Recovery Process

The recovery process is a straight forward analysis of two artifacts. Figure 7.1 shows a flowchart depicting this. Given two artifacts as input, the recovery pro-

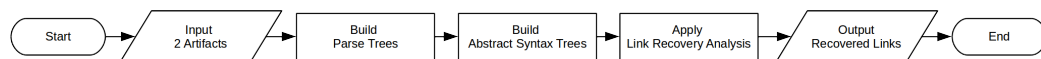


Figure 7.1 The Recovery Process

cess works as follows:

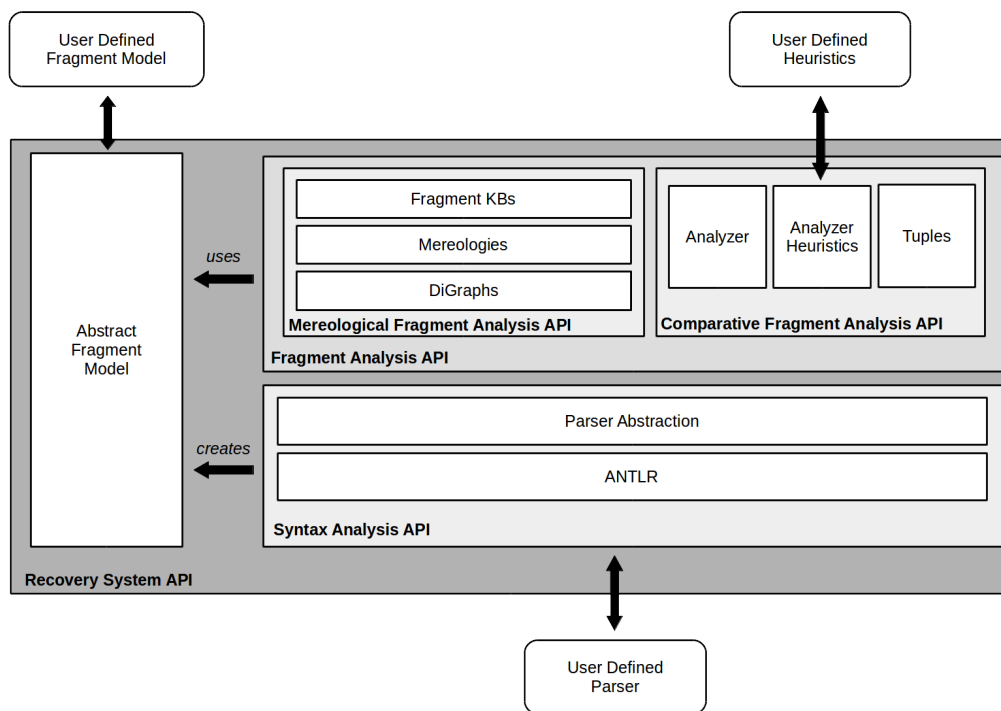
1. From each artifact a Concrete Syntax Tree (CST) or Parse Tree is constructed,

2. each Parse Tree is further refined into an Abstract Syntax Tree (AST),
3. both ASTs are compared with each other, i.e. both trees are traversed in a Depth-First Search (DFS) fashion and each pair of nodes is checked whether it can be recovered as link.

Eventually, the set of recovered links serves as output of the process.

7.1.2 Recovery API

The Recovery API is the core of the developed recovery system. It provides generalized data structures and methods for syntactic analysis of artifacts and their fragments. Figure 7.2 depicts the API as block diagram. The components of the



This block diagram depicts the functional outline of the Recovery System API (note, that this does not necessarily correspond to the package outline of its actual implementation).

Figure 7.2 The Recovery API

API are:

Abstract Fragment Model The Abstract Fragment Model serves as base model for all ASTs. It can be thought of as Data Transfer Object (DTO) for the analysis components. As user of the API, I have to derive a specific AST from this model. A detailed description follows in §7.1.2.1.

Syntax Analysis API The Syntax Analysis API is the abstraction layer for parsing and AST construction. It is currently backed by ANTLR (Another Tool For Language Recognition), but its internal design is loosely coupled, so other parser libraries can be used. As user of the API, I have to implement a parser constructing an AST deriving the Abstract Fragment Model. However, if one uses ANTLR, only AST construction from a CST is required. A detailed description follows in §7.1.2.2.

Fragment Analysis API The Fragment Analysis API consists of two components:

Mereological Fragment Analysis API The Mereological Fragment Analysis API provides components for deriving parthood links between syntactically well-formed fragments. As user of the API, I only have to apply its components to constructed ASTs. A detailed description follows in §7.1.2.3.

Comparative Fragment Analysis API The Comparative Fragment Analysis API provides components for deriving links between different artifacts and their fragments. As user of the API, I have to implement one or more specialized heuristics for deciding which links can be recovered. A detailed description follows in §7.1.2.4.

7.1.2.1 Abstract Fragment Model

The Abstract Fragment Model describes a tree in which each node represents an syntactically well-formed fragment. Figure 7.3 shows an UML class diagram of the model. The resulting tree data structure is doubly-linked, i.e. an `IFragment` node aggregates references to its children and to its parent, given it is not the root node. Each fragment `IFragment` contains the text it represents. In order to distinguish fragments which represent the same text, each node also carries the text's position in the artifact through `IFragmentPosition` instances. Positions inside the text are determined by the start and ending line number as well as the corre-

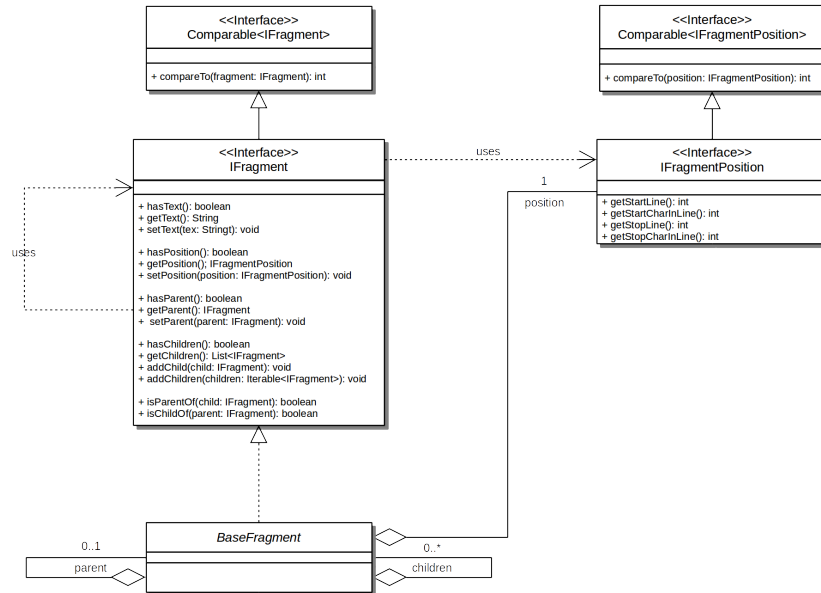


Figure 7.3 The Abstract Fragment Model

sponding first and last character inside the line. Most of `IFragment`'s relevant code for trees is pre-implemented in the `BaseFragment` abstract class.

7.1.2.2 Syntax Analysis API

The Syntax Analysis API provides abstraction for AST construction. The API itself is really small, it only consists of the `IParser` and `IParserFactory` interfaces. However, its implementation for ANTLR is designed to reduce ANTLR-specific boilerplate code. Figure 7.4 shows the UML class diagram for the ANTLR backed implementation. One can see, that this implementation makes heavy use of the Abstract Factory Pattern [4]. This allows for a quick and easy definition of new parsers as shown in Figure 7.5.

The other advantage of the Abstract Factory Pattern is that the instantiation of all relevant classes necessary for the creation of ANTLR Parse Trees takes place in the predefined `AntlrParser` class. This is exemplified by Figure 7.6.

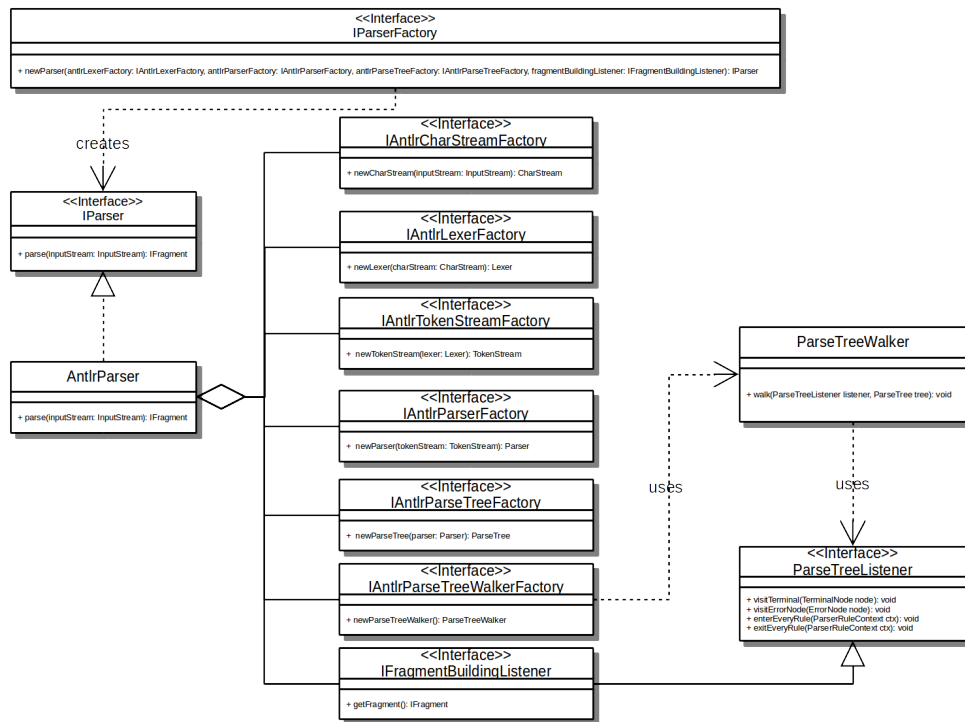


Figure 7.4 The Syntax Analysis API

```

1 ...
2 IParser java8Parser = parserFactory.newParser(
3     java8Lexer::new,
4     java8Parser::new,
5     java8Parser::compilationUnit,
6     java8FragmentBuildingListener::new);
7 ...

```

This example demonstrates the creation of a new parser using an `IParserFactory` instance with Java8 features. Lexer and Parser classes are generated by ANTLR. A suitable listener has to be implemented.

Figure 7.5 `IParserFactory` Usage Example

```

1 ...
2 CharStream charStream = antlrCharStreamFactory.newCharStream(inputStream);
3 Lexer lexer = antlrLexerFactory.newLexer(charStream);
4 TokenStream tokenStream = antlrTokenStreamFactory.newTokenStream(lexer);
5 Parser parser = antlrParserFactory.newParser(tokenStream);
6 ParseTree parseTree = antlrParseTreeFactory.newParseTree(parser);
7 ...

```

This example demonstrates the creation of an ANTLR `ParseTree` instance as implemented by the `AntlrParser` class.

Figure 7.6 `AntlrParser` Parse Tree Creation

Creation of ASTs or fragment trees is done using the `ParseTreeWalker` and `ParseTreeListener` infrastructure provided by ANTLR [11]. This is an variation of the Observer Pattern [4]. Listeners in conjunction with walkers are an alternative to the Visitor Pattern [4] provided by ANTLR. An instance of `ParseTreeWalker` traverses a Parse Tree using DFS. During traversal, a designated method of `ParseTreeListener` is executed. For AST creation, an API user has to implement the `IFragmentBuildingListener` interface, which is an extension of the `ParseTreeListener` interface. The creation of a fragment tree by `AntlrParser` is exemplified in Figure 7.7.

```

1 ...
2 ParseTreeWalker parseTreeWalker = antlrParseTreeWalkerFactory.newParseTreeWalker();
3 parseTreeWalker.walk(fragmentBuildingListener, parseTree);
4 IFragment fragment = fragmentBuildingListener.getFragment();
5 ...

```

This example demonstrates the creation of an `IFragment` AST instance as implemented by the `AntlrParser` class.

Figure 7.7 `AntlrParser` Fragment Creation

7.1.2.3 Mereological Fragment Analysis API

The Mereological Fragment Analysis API provides components for deriving parthood links between syntactically well-formed fragments. Direct parthood links are recovered from parent/child relationships within an `IFragment` AST. However, because parthood is transitive, we also need to recover these links. This is done using a digraph data structure upon which we can compute its reflexive transitive closure.

Figure 7.8 shows an UML class diagram depicting the relevant classes and interfaces of the Mereological Fragment Analysis API. At its heart, the API utilizes the generic `IDiGraph` interface whose implementations provide means for interrelating comparable objects, i.e. `IFragment` instances.

This digraph is wrapped by generic `IMereology` implementations, a data structure providing query methods on the topic of mereology, i.e. parthood relations. Mereologies are constructed using the Builder Pattern [4]. An `IMereology-Builder` instance adds nodes and edges into its digraph with semantically named methods allowing a descriptive programming style.

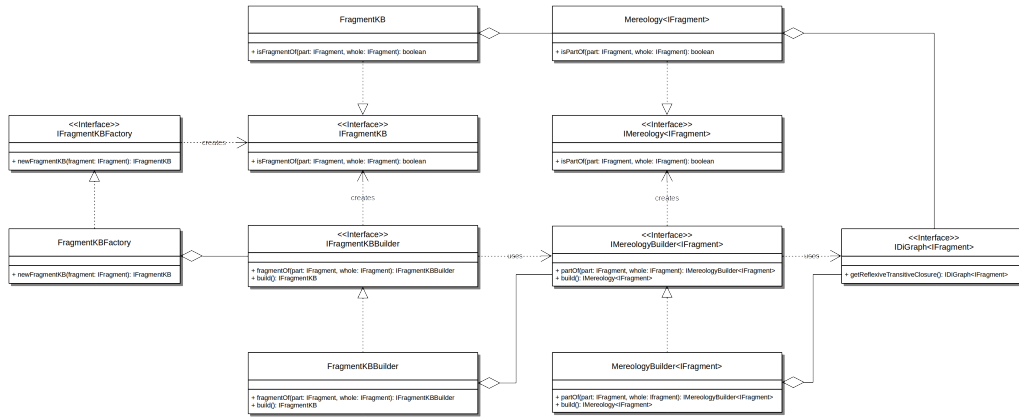


Figure 7.8 Mereological Fragment Analysis API

IMereology's are then further wrapped by IFragmentKB (a Knowledge Base over IFragment) implementations in order to avoid dealing with generics throughout the system. This is also done utilizing the Builder Pattern.

The recovery of parthood links is encapsulated through IFragmentKBFactory using the Abstract Factory Pattern, which computes IFragmentKB instances from an IFragment AST as input.

7.1.2.4 Comparative Fragment Analysis API

The Comparative Fragment Analysis API provides components for deriving links from two IFragment ASTs generated from different artifacts. Figure 7.9 shows an UML class diagram depicting the relevant interfaces and classes of the API. It utilizes the Strategy Pattern [4] through the IFragmentAnalyzerHeuristic interface since concrete behavior of the recovery analysis is for the user to implement. In this context, strategies are called heuristics, because recovery of links is not necessarily based on optimal, i.e. absolutely correct, solutions. Instead strategies may also implement practical solutions with reasonably sufficient results.

IFragmentAnalyzer objects allow one to apply multiple heuristics, since there may be more than one practical approach to achieve a goal. Also this allows strategies to keep a relatively small implementation footprint.

Recovered links are captured and stored in an IBinaryRelation instance which works like a set of pairs.

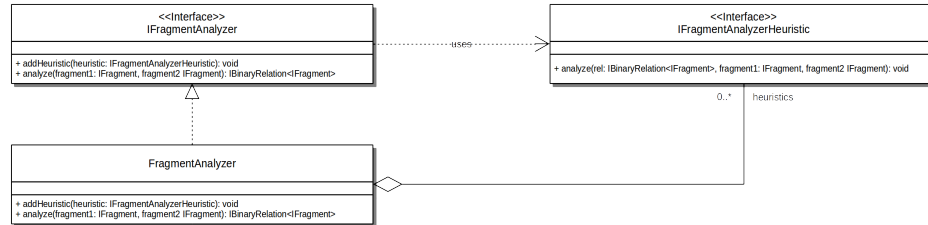
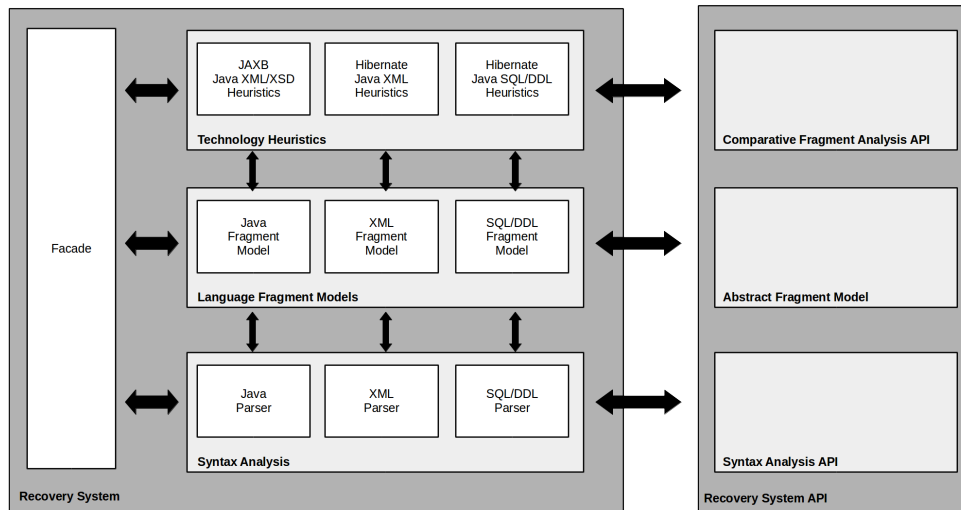


Figure 7.9 Comparative Fragment Analysis API

7.1.3 Recovery System

The Recovery System is the actual implementation of parthood, correspondence and conformance link recovery for Java based O/R/X-Mapping artifacts. Figure 7.10 shows a block diagram outlining the functional components of the system.



This block diagram depicts the functional outline of the Recovery System and its dependencies to the Recovery System API (note, that this does not necessarily correspond to the package outline of its actual implementation).

Figure 7.10 The Recovery System

The Recovery System utilizes the Syntactic Analysis API described in §7.1.2.2 for Java, XML and SQL/DDL. For this, the Abstract Fragment Model described in §7.1.2.1 is implemented. Note, fragment models do not implement AST suitable for compilation of the targeted language. Syntactic features unnecessary for the intended analysis, like local variable declarations, arithmetic expressions or invocations, are omitted. Fragment models rather focuses on structural features of the languages at hand.

The Comparative Fragment Analysis API described in §7.1.2.4 is implemented with focus on artifacts of technologies, namely JAXB and Hibernate. It implements heuristics for link recovery between:

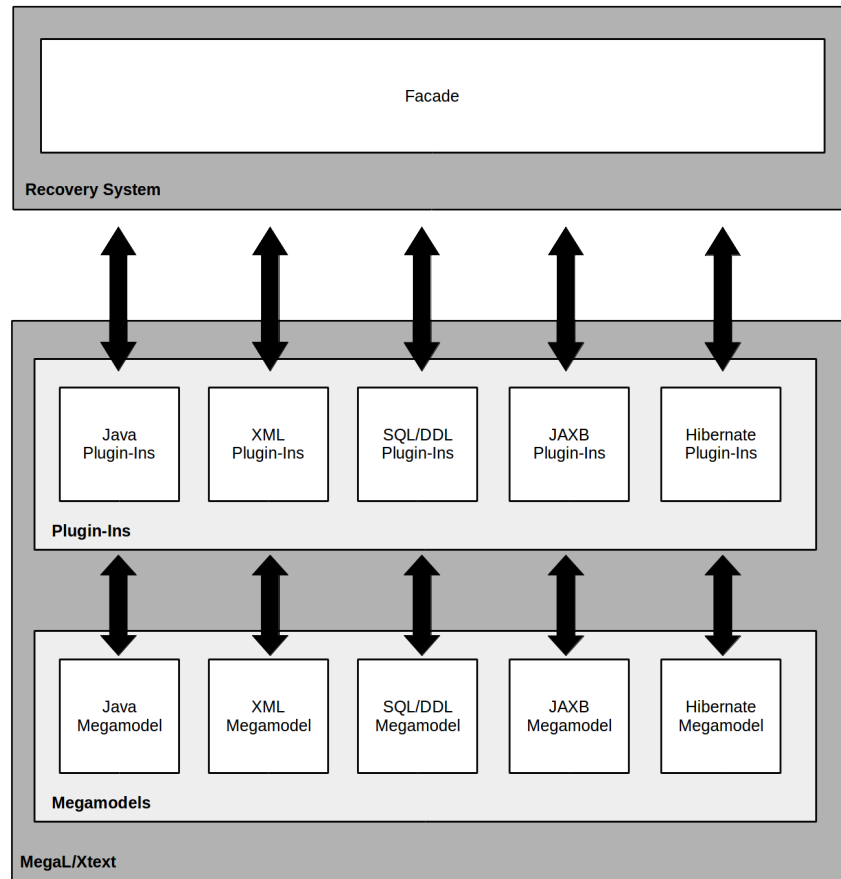
- Java and XML for JAXB artifacts, i.e. Java models are serialized as XML
- Java and XSD for JAXB artifacts, i.e. Java models are serialized as XSD
- XML and XSD for JAXB artifacts, i.e. the two previous scenarios occurred
- Java and XML for Hibernate mapping artifacts, i.e. Hibernate uses XML meta-data for O/R-Mapping.
- Java and SQL/DDL for Hibernate generated SQL artifacts, i.e. Hibernate uses Java annotations for O/R-Mapping

On top of the actual implementation, the Recovery System also utilizes the Facade Pattern [4]. This is to provide a single access point for all implemented analysis features.

7.2 Megal/Xtext Integration Design

This section summarizes the design of the recovery system's integration in the MegaL/Xtext environment. Figure 7.11 shows a block diagram depicting the outline this integration. Note, the Recovery System is implemented as separate project, which is then included as reference (through a `.jar` file) in a MegaL/Xtext project.

MegaL/Xtext provides a plug-in system, which allows to bind special Java classes to `Plugin` entities. Code of such classes is then executed during the evaluation of a megamodel. **ToDo: needs reference** Figure 7.12 exemplifies the declaration of Plug-in withing MegaL.



This block diagram depicts the functional outline of the Recovery System's integration into the MegaL/Xtext environment (note, that this does not necessarily correspond to the outline of its actual implementation).

Figure 7.11 Integration of the Recovery System into MegaL/Xtext

```

1 ...
2 JavaFragmentRecoveryPlugin : Plugin
3 JavaFragmentRecoveryPlugin realizationOf Java
4 JavaFragmentRecoveryPlugin partOf FileFragmentRecoveryReasonerPlugin
5 JavaFragmentRecoveryPlugin = 'classpath:org.softlang.megal.plugins.impl.java.JavaFragmentRecoveryPlugin'
6 ...

```

This snippet demonstrates the instantiation of `JavaFragmentRecoveryPlugin` as part of `FileFragmentRecoveryReasonerPlugin` in MegaL.

Figure 7.12 MegaL Plug-In Instantiation

The Recovery System or rather its facade (see §7.1.3) is used to implement MegaL plug-ins. These Plug-ins are then in turn bound within megamodels, which

are divided by language and technology, i.e. there are separate MegaL modules for Java, XML, SQL, Hibernate and JAXB.

Chapter 8

Implementation

This chapter summarizes the implementation of crucial parts of recovery system implemented for this thesis. §8.1 covers the implementation of fragment recovery. §8.2 covers the implementation of parthood link recovery. §8.3 covers the implementation of correspondence and conformance link recovery

8.1 Recovering Fragments

This section summarizes the implementation of fragment recovery. Fragments are syntactically well-formed pieces of code. For instance, consider the following Java class:

```
public class Company {  
    ...  
    private String name;  
    ...  
    public String getName() {  
        return name;  
    }  
    ...  
}
```

It consists, among others, of the following fragments:

- a field declaration fragment for `name`:

```
private String name;
```

- a method declaration fragment for an accessor-method of `name`:

```
public String getName() {  
    return name;  
}
```


- the return statement of the accessor-method:

```
return name;
```

- the class declaration itself can also be considered a fragment.

The task of recovering fragments is to add entities for each of such code pieces into a megamodel. Figure 8.1 exemplifies the recovery of Java¹ fragments in an idealized form.

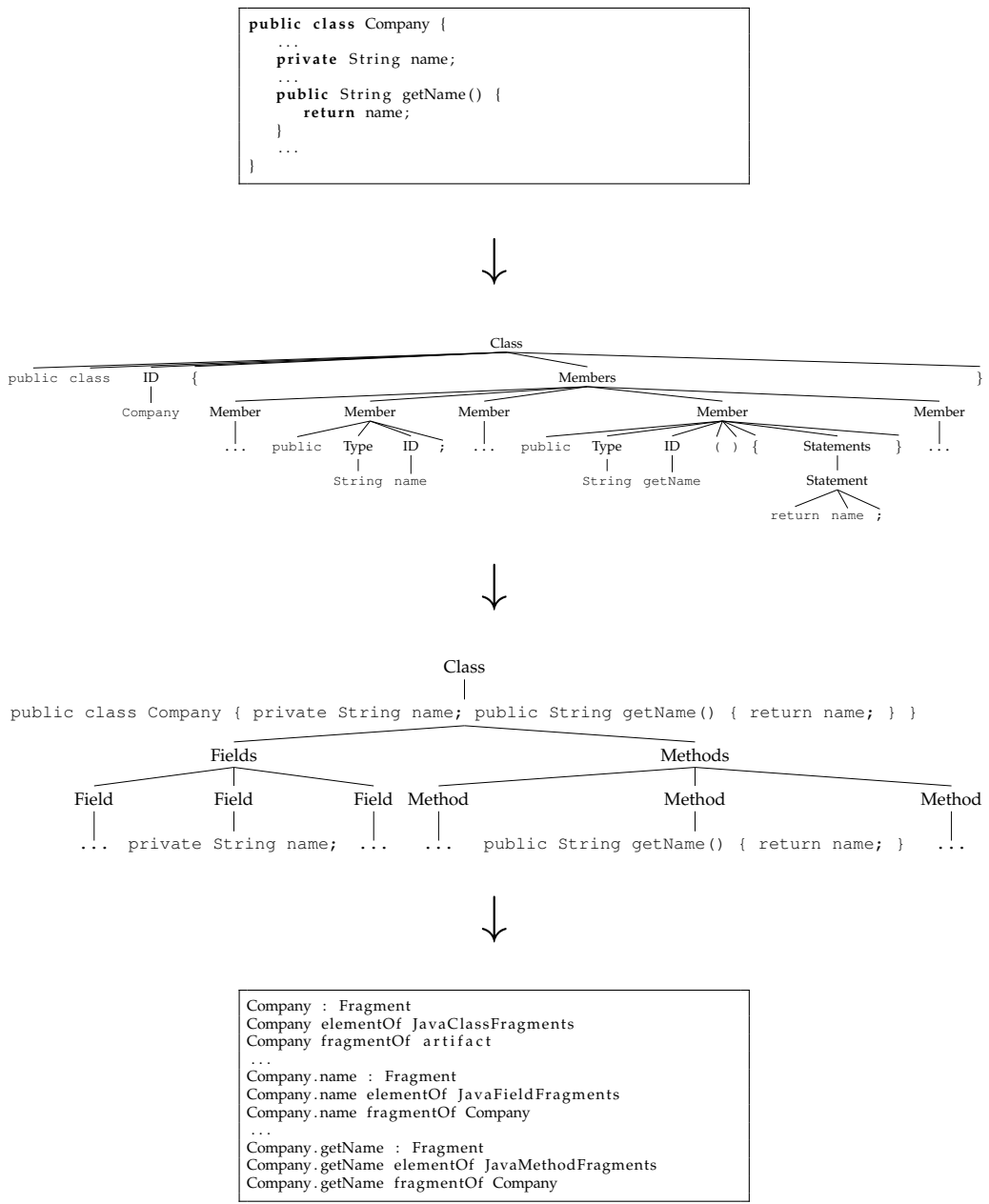
1. a Parse Tree is generated from an artifact; this is done via ANTLR as described in §7.1.2.2 and does not require further explanation, since it is only an application of the most common ANTLR use case
2. then the Parse Tree is transformed into an concrete fragment AST model; this is also done relying on ANTLR tree traversal utilities and described in detail in §8.1.1
3. eventually, nodes of the generated fragment AST are added to a megamodel as entities; the details are described in §8.1.2.

8.1.1 Concrete Fragment Models

As mentioned in §7.1.2.1, Concrete Fragment Models are derivations of the Abstract Fragment Model of the Recovery System API. Figure 8.2 shows an UML class diagram of the implemented Fragment AST for Java. All fragment classes derive from `IFragment` through the base class `JavaFragment`. From here, several specializations are introduced:

- `IdentifiedJavaFragment` for constructs with identifiers
- `ModifiedJavaFragment` for constructs with modifiers like `private`, `public`, `final`, `abstract`, `static`, etc. or annotation meta-data
- `TypedJavaFragment` for constructs with distinct (return-) type like fields, methods or variables

¹Recovery for XML and SQL/DDI is implemented in a similar fashion. If there is a noteworthy difference for other languages it will be explored, otherwise we keep using Java as example domain for the remaining sections of §8.1.



This picture shows an idealized recovery of Java fragments:

code artifact → Parse Tree → fragment AST → megamodel

Both Parse Tree and AST are depicted in a simplified, schematic form.

Figure 8.1 Idealized Recovery of Java Fragments

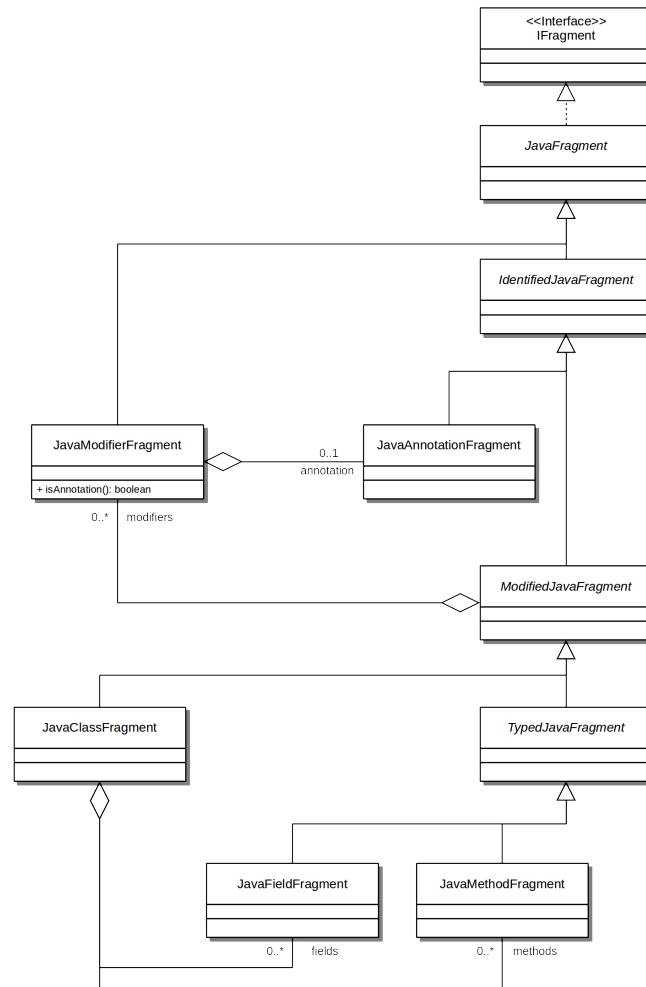


Figure 8.2 Java Fragment AST Model

The fragment model for Java only focuses on structural syntactical features of the language. Everything below method signature level is omitted, because fragments of this level are not important for the scope of this thesis. On the other hand, annotations are captured, since they provide significant information for further recovery analysis. Hibernate and JAXB rely on annotations for O/R/X-Mapping.

The AST is constructed using ANTLR `ParseTreeListener` approach as described in §7.1.2.2. Figure 8.3 shows an excerpt from the listener implementation used to construct the the AST for Java fragments. Here, fragment instances are

```

1  ...
2  @Override
3  public void exitMethodDeclaration (Java8Parser.MethodDeclarationContext ctx) {
4      javaMethodFragments.push(java8FragmentFactory.newJavaMethodFragment(ctx, javaMethodModifierFragments));
5  }
6
7  @Override
8  public void exitNormalClassDeclaration (Java8Parser.NormalClassDeclarationContext ctx) {
9      javaClassFragments.push(java8FragmentFactory.newJavaClassFragment(ctx, javaClassModifierFragments,
10         javaFieldFragments, javaMethodFragments, declaredPackage));
11 }
12 ...

```

This snippet shows an excerpt of the listener used to construct Java fragment ASTs.

Figure 8.3 Construction of Java Fragment ASTs

created when the listener leaves a certain node of the Parse Tree. Creation is delegated to a factory. The listener pushes fragments onto stacks, making the overall implementation work like a pushdown automaton or stack machine.

8.1.2 Megamodeling Fragments

In order to add fragments into a megamodel a suitable linguistic model has to be declared first. Consider the following instance level megamodel of recovered fragments:

```

File < Artifact
Fragment < Artifact
...
elementOf < Set * Set
fragmentOf < Fragment * Artifact
...
aJavaFile : File
aJavaFile = 'Company.java'
...
aJavaFile.Company : Fragment
aJavaFile.Company elementOf JavaClassFragments
aJavaFile.Company fragmentOf aJavaFile
...
aJavaFile.Company.name : Fragment
aJavaFile.Company.name elementOf JavaFieldFragments
aJavaFile.Company.name fragmentOf aJavaFile
aJavaFile.Company.name fragmentOf Company
...
aJavaFile.Company.getName : Fragment
aJavaFile.Company.getName elementOf JavaMethodFragments
aJavaFile.Company.getName fragmentOf aJavaFile
aJavaFile.Company.getName fragmentOf Company

```

This megamodel shows fragments of a Java file `Company.java`. It contains entities denoting a class declaration, a field declaration and a method declaration

fragment. All are declared instances of the entity type `Fragment`. Entity names use dot-notation for implying parthood, i.e.:

$$a.b \Rightarrow b \text{ partOf } a$$

Further details of naming-scheme employed for fragment entities are explored in §8.1.2.2.

We use `elementOf` for further specialization and/or generalization of the entitie’s kind². However, the right-hand side of these relationships used to specialize fragments are not simply sets, they are actually modeled as languages as we will see in §8.1.2.1.

8.1.2.1 Fragment Languages

As mentioned in §4.1, a fragment alone cannot be element of the language the artifact it originated from belongs to, e.g. a Java method alone cannot be accepted by the Java grammar, nor can XML attribute alone be accepted by the XML grammar. Therefore, in order to cleanly add fragments into a megamodel, we need to model fragment languages. Figure 8.4 shows the megamodel for Java fragments. It introduces the special relationship type `fragmentLanguageOf`, denoting the

```

1  ...
2  Language < Set
3  ...
4  fragmentLanguageOf < Language * Language
5  ...
6  Java : Language
7
8  JavaFragments : Language
9  JavaFragments fragmentLanguageOf Java
10
11 JavaClassFragments : Language
12 JavaClassFragments subsetOf JavaFragments
13
14 JavaFieldFragments : Language
15 JavaFieldFragments subsetOf JavaFragments
16
17 JavaMethodFragments : Language
18 JavaMethodFragments subsetOf JavaFragments
19 ...

```

Figure 8.4 A Megamodel for Java Fragments

left-hand side is the language containing all possible fragments the right-hand

²We use the term *kind* to avoid confusion with entity types, although we use it to refer to fragment types.

side's language elements can be deconstructed in. This also implies a super-set relation between the two languages as `fragmentOf` is reflexive like `partOf`, so:

$$A \text{ fragmentLanguageOf } B \Rightarrow B \subseteq A$$

8.1.2.2 Fragment Entity Identifiers

8.2 Recovering Parthood Links

This section summarizes the implementation of parthood link recovery.

8.3 Recovering Correspondence & Conformance Links

This section the implementation of correspondence and conformance link recovery

Chapter 9

Results

TBD.

Chapter 10

Conclusion

TBD.

Glossary

101HRMS 101wiki¹ Human Resource Management System². The model used by the 101wiki for its contributions. 31, 32, *see also* HRMS

Abstract Factory Pattern A creational GoF (Gang of Four) pattern used in software design to decouple instantiation from usage of objects. Hides the concrete nature of created instances. 38, 41

ANTLR Another Tool For Language Recognition. 37–40

API Application Programming Interface. 35–38, 40–43, 51

AST Abstract Syntax Tree: A tree data structure representing the abstract syntax of a parsed text. This tree omits syntactic features like parentheses for grouping or semicolons for sequencing. 36–38, 40, 41, 43

Builder Pattern A creational GoF pattern used in software design to prevent constructor parameters from piling up. 40, 41

conformance The relation between . 42, 46, 47

correspondence The relation between . 42, 46, 47

CST Concrete Syntax Tree: A tree data structure representing the concrete syntax of a parsed text.. 35, 37

DDL Data Definition Language. Language or subset of a language used to describe structure and content of data. 32

¹<https://101wiki.softlang.org/> (retrieved 12th November, 2017)

²<https://101wiki.softlang.org/101:@system> (retrieved 12th November, 2017)

- DFS** The algorithmic concept of traversing a tree or graph data structure 'top-down' until reaching the end of a path before backtracking and traversing another path. 36, 40
- DTO** Data Transfer Object. Objects with no relevant (business) logic of their own. Their sole purpose is to carry data between layers of a software system. 37
- Facade Pattern** A structural GoF pattern used in software design to simplify the usage of complex systems or APIs. It provides single access point for such system. Such access points are called facades. 43
- fragment** A syntactically well-formed piece of a possibly larger text. v, 46
- GoF** Gang of Four. A group of authors (Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides) publishing on the subject of object-oriented software design. The term may also refer to design patterns described in their book *Design Patterns: Elements of Reusable Object-Oriented Software* [4]. 50–52
- Hibernate** The Hibernate ORM Framework. 30–33, 43, 45
- HRMS** Human Resource Management System. 29, 30
- Java** The Java Programming Language and Platform. v, 29–33, 39, 43, 45, 46
- JAXB** Java Architecture for XML Binding. 30–32, 43, 45
- MegaL** The megamodeling language developed by the Softlang Team at the University of Koblenz-Landau for descriptively and prescriptively modeling linguistic architectures of software systems. 43–45, 51
- MegaL/Xtext** The Xtext implementation and eclipse IDE integration of MegaL. 34, 35, 43, 44
- megamodel** . v, 43, 44, 46
- mereology** . 40
- O/R-Mapping** Object-Relational-Mapping. 32, 33, 43

O/R/X-Mapping Object-Relational- and XML-Mapping. 30–32, 35, 42

O/X-Mapping Object-XML-Mapping. 32

Observer Pattern A behavioral GoF pattern used in software design to propagate state changes from one object to many dependent objects. 40

ORM . 51, *see* O/R-Mapping

Parse Tree . iv, 35, 36, 38–40, *see* CST

parthood The relation between an entity and its constituent parts. 40–42, 46, 47

SQL Structured Query Language. 32, 33, 43, 45

SQL/DDL The DDL subset of SQL. 30, 43

Strategy Pattern A behavioral GoF pattern used in software design to separate behavior from structure. It allows to encapsulate and reuse behavior as part of the configuration of larger constructs. 41

UML Unified Modeling Language. 31, 37, 38, 40, 41

Visitor Pattern A behavioral GoF pattern used in software design to separate behavior from structure. Visitors facilitate the extension of behavior without modifying structure. The Visitor Pattern can be used to traverse object graphs. 40

XML Extensible Markup Language. 29–33, 43, 45

XSD XML Schema Definition. 30, 31, 43

Bibliography

- [1] Anya Helene Bagge and Vadim Zaytsev. “Languages, Models and Megamodels”. In: *Post-proceedings of the Seventh Seminar on Advanced Techniques and Tools for Software Evolution, SATToSE 2014, L’Aquila, Italy, 9-11 July 2014*. 2014, pp. 132–143. URL: <http://ceur-ws.org/Vol-1354/paper-12.pdf>.
- [2] Jean-Marie Favre, Ralf Lämmel, and Andrei Varanovich. “Modeling the Linguistic Architecture of Software Products”. In: *Model Driven Engineering Languages and Systems - 15th International Conference, MODELS 2012, Innsbruck, Austria, September 30-October 5, 2012. Proceedings*. 2012, pp. 151–167. DOI: [10.1007/978-3-642-33666-9_11](https://doi.org/10.1007/978-3-642-33666-9_11). URL: http://dx.doi.org/10.1007/978-3-642-33666-9_11.
- [3] Jean-Marie Favre and Tam Nguyen. “Towards a Megamodel to Model Software Evolution Through Transformations”. In: *Electr. Notes Theor. Comput. Sci.* 127.3 (2005), pp. 59–74. DOI: [10.1016/j.entcs.2004.08.034](https://doi.org/10.1016/j.entcs.2004.08.034). URL: <http://dx.doi.org/10.1016/j.entcs.2004.08.034>.
- [4] Erich Gamma et al. *Design Patterns: Elements of Reusable Object-oriented Software*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1995. ISBN: 0-201-63361-2.
- [5] Orlena Gotel et al. “Traceability Fundamentals”. In: *Software and Systems Traceability*. 2012, pp. 3–22. DOI: [10.1007/978-1-4471-2239-5_1](https://doi.org/10.1007/978-1-4471-2239-5_1). URL: https://doi.org/10.1007/978-1-4471-2239-5_1.

- [6] Nicola Guarino, Daniel Oberle, and Steffen Staab. "What Is an Ontology?" In: *Handbook on Ontologies*. 2009, pp. 1–17. DOI: [10.1007/978-3-540-92673-3_0](https://doi.org/10.1007/978-3-540-92673-3_0). URL: https://doi.org/10.1007/978-3-540-92673-3_0.
- [7] Lukas Härtel. "Linguistic architecture on the workbench". Bachelor Thesis. University of Koblenz-Landau, Oct. 2015.
- [8] "IEEE Standard Glossary of Software Engineering Terminology". In: *IEEE Std 610.12-1990* (Dec. 1990), pp. 1–84. DOI: [10.1109/IEEESTD.1990.101064](https://doi.org/10.1109/IEEESTD.1990.101064). URL: <http://ieeexplore.ieee.org/servlet/opac?punumber=2238>.
- [9] Ralf Lämmel. "Coupled software transformations revisited". In: *Proceedings of the 2016 ACM SIGPLAN International Conference on Software Language Engineering, Amsterdam, The Netherlands, October 31 - November 1, 2016*. 2016, pp. 239–252. URL: <http://dl.acm.org/citation.cfm?id=2997366>.
- [10] Ralf Lämmel and Andrei Varanovich. "Interpretation of Linguistic Architecture". In: *Modelling Foundations and Applications - 10th European Conference, ECMFA 2014, Held as Part of STAF 2014, York, UK, July 21-25, 2014. Proceedings*. 2014, pp. 67–82. DOI: [10.1007/978-3-319-09195-2_5](https://doi.org/10.1007/978-3-319-09195-2_5). URL: http://dx.doi.org/10.1007/978-3-319-09195-2_5.
- [11] Terence Parr. *The Definitive ANTLR 4 Reference*. 2nd. Pragmatic Bookshelf, 2013. ISBN: 1934356999, 9781934356999.
- [12] Alan Rector et al. *Simple part-whole relations in OWL Ontologies*. retrieved 12. June 2017. Aug. 2005. URL: <http://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/simple-part-whole-relations-v1.5.html>.
- [13] Gunther Schmidt and Thomas Ströhlein. *Relationen und Graphen*. Mathematik für Informatiker. Springer, 1989. ISBN: 3-540-50304-8.

- [14] Uwe Schöning. *Theoretische Informatik - kurzgefaßt* (3. Aufl.) Hochschul-taschenbuch. Spektrum Akademischer Verlag, 1997. ISBN: 978-3-8274-0250-9.
- [15] Achille C. Varzi. "Mereology". In: *The Stanford encyclopedia of phi-losophy*. Ed. by Edward N. Zalta. Winter 2016 ed. retrieved 27. July 2017. 2016. URL: <https://plato.stanford.edu/archives/win2016/entries/mereology/>.
- [16] Achille C. Varzi. "Parts, Wholes, and Part-Whole Relations: The Prospects of Mereotopology". In: *Data Knowl. Eng.* 20.3 (1996), pp. 259–286. DOI: [10.1016/S0169-023X\(96\)00017-1](https://doi.org/10.1016/S0169-023X(96)00017-1). URL: [http://dx.doi.org/10.1016/S0169-023X\(96\)00017-1](http://dx.doi.org/10.1016/S0169-023X(96)00017-1).
- [17] Stefan Winkler and Jens Pilgrim. "A Survey of Traceability in Re-quirements Engineering and Model-driven Development". In: *Softw. Syst. Model.* 9.4 (Sept. 2010), pp. 529–565. ISSN: 1619-1366. DOI: [10.1007/s10270-009-0145-0](https://doi.org/10.1007/s10270-009-0145-0). URL: <http://dx.doi.org/10.1007/s10270-009-0145-0>.