UNIVERSITÄT
KOBLENZ · LANDAU

# Megamodel-driven Traceability Recovery & Exploration of Correspondence & Conformance Links

## Bachelorarbeit

zur Erlangung des Grades eines Bachelor of Science
im Studiengang Informatik

vorgelegt von

Maximilian Meffert

| | |
|---|---|
| Erstgutachter: | Prof. Dr. Ralf Lämmel<br>Institut für Informatik |
| Zweitgutachter: | Msc. Johannes Härtel<br>Institut für Informatik |

Koblenz, im Juni 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 einverstanden. | ☐ | ☐ |
| Der Veröffentlichung dieser Arbeit im Internet stimme ich zu. | ☐ | ☐ |

.......................................................................

(Ort, Datum) (Maximilian Meffert)

# Zusammenfassung

TBD.

# Abstract

TBD.

# Contents


**1 Introduction** **1**

**2 Related Work** **2**

**3 Background** **3**

3.1 Relations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

3.1.1 Closures . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.2 Mereology . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.3 Formal Languages & Grammars . . . . . . . . . . . . . . . . . 9

3.3.1 Context-Free Languages & Grammars . . . . . . . . . . . . 9

3.4 Traceability . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.4.1 Traceability Relationship . . . . . . . . . . . . . . . . . 9

3.4.2 Traceability Link . . . . . . . . . . . . . . . . . . . . . 9

3.4.3 Traceability Recovery . . . . . . . . . . . . . . . . . . . 9

3.4.4 Traceability Exploration . . . . . . . . . . . . . . . . . . 9

3.5 Megamodeling . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.5.1 MegaL . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.6 Ontologies . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.7 Program Analysis . . . . . . . . . . . . . . . . . . . . . . . . 9

3.8 XML Data Binding . . . . . . . . . . . . . . . . . . . . . . . . 9

3.8.1 Java Architecture for XML Binding (JAXB) . . . . . . . . . 10

3.9 Object Relational Mapping . . . . . . . . . . . . . . . . . . . 10

3.9.1 Java Persistence API (JPA) . . . . . . . . . . . . . . . . . 10

3.9.2 Hibernate . . . . . . . . . . . . . . . . . . . . . . . . . . 10

3.10 Another Tool For Language Recognition (ANTLR) . . . . . . . . . . 10

**4 Hypotheses** **11**
4.1 Fragments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
4.2 Correspondence . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
4.3 Conformance . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

**5 Methodology** **12**
5.1 The 101companies Human Resource Management System . . . . . 12
5.2 Link Proper Part Ratio . . . . . . . . . . . . . . . . . . . . . . . . 12

**6 Requirements** **13**

**7 Design** **14**

**8 Implementation** **15**
8.1 Context-Free Grammar Fragmentation . . . . . . . . . . . . . . . . 15
8.2 Name Correspondence Heuristic . . . . . . . . . . . . . . . . . . . 15

**9 Results** **16**

**10 Conclusion** **17**

# List of Figures

# Chapter 1

# Introduction

TBD.

# Chapter 2

# Related Work

TBD.

# Chapter 3

# Background

This section summarizes the necessary background topics of the thesis. Each topic is introduced independently, interrelation is done during synthesis of hypotheses for this thesis in chapter 4.

## 3.1 Relations

This section introduces the necessary aspects of mathematical relations for this thesis. The concept of relations is a generalization of semantic dependencies between two or more mathematical objects. This section is based on [1].

Relations are based on set-theory. We also introduce the necessary constructs of set-theory in order to clarify terminology and notation. A set is a collection of well distinguishable mathematical objects. Objects in a set are called elements of the set. A set does not contain two or more identical elements. The notation $x \in X$ denotes that $x$ is an element of the set $X$. The symbol $\emptyset$ denotes the *empty set*, which contains no elements. The symbol $\Omega$ denotes the universal set, which contains all elements.

**Definition 1 (Inclusion)** *Let $X$ and $Y$ be a sets. $Y$* includes *$X$ if and only if:*

$$X \subset Y :\Leftrightarrow \forall x[x \in X \Rightarrow x \in Y] \tag{3.1}$$

*Then $X$ is called* subset *of $Y$ and $Y$ is called* superset *of $X$. For an arbitrary set $Z \neq \emptyset$, the statement $\emptyset \subset Z$ is always true, respectively $Z \subset \emptyset$ is always false.*

**Definition 2 (Union)** *Let $X$ and $Y$ be a sets.*

$$X \cup Y := \{x | x \in X \vee x \in Y\} \tag{3.2}$$

*$X \cup Y$ is called* union *of $X$ and $Y$.*

**Definition 3 (Intersection)** *Let $X$ and $Y$ be a sets.*

$$X \cap Y := \{x | x \in X \wedge x \in Y\} \tag{3.3}$$

*$X \cap Y$ is called* intersection *of $X$ and $Y$.*

**Definition 4 (Power-Set)** *Let $X$ be a set, then the* power-set *of $X$ is defined as:*

$$\mathcal{P}(X) := \{Y | Y \subset X\} \tag{3.4}$$

The definition of inclusion provides an order for power-sets. So we may compare two sets $A$ and $B$ in the sense of *smaller* and *larger*, i.e.:

$$A \text{ is smaller than } B \Leftrightarrow A \subset B \tag{3.5}$$
$$A \text{ is larger than } B \Leftrightarrow B \subset A \tag{3.6}$$
$$A \text{ is the smallest subset of } B \Leftrightarrow \forall C \in \mathcal{P}(B) : A \subset C \tag{3.7}$$
$$A \text{ is the largest subset of } B \Leftrightarrow \forall C \in \mathcal{P}(B) : C \subset A \tag{3.8}$$

**Definition 5 (Upper & Lower Bound)** *Let $\Omega$ be a universe, $X \in \mathcal{P}(\Omega)$ be a set in the universe and $A \subset \mathcal{P}(U)$, $A \neq \emptyset$, non-empty subsets in the universe.*

$$X \textit{ is an } \text{upper bound} \textit{ for } A :\Leftrightarrow \forall Y \in A : Y \subset X \tag{3.9}$$
$$X \textit{ is a } \text{lower bound} \textit{ for } A :\Leftrightarrow \forall Y \in A : X \subset Y \tag{3.10}$$

*We also define:*

$$\mathbf{U}_A := \{U \in \mathcal{P}(\Omega) | \forall Y \in A : Y \subset U\} \tag{3.11}$$
$$\mathbf{L}_A := \{L \in \mathcal{P}(\Omega) | \forall Y \in A : L \subset Y\} \tag{3.12}$$

*as sets of all upper/lower bounds for $A$.*

Because our definition of upper and lower bounds is based on power-sets, existence is guaranteed: Given an arbitrary set $S$, the $S$ and $\emptyset$ are always elements of $\mathcal{P}(S)$. For each element $Y$ of a non-empty selection $A \subset \mathcal{P}(S)$ of the power-set, $Y \subset S$ and $\emptyset \subset Y$ holds. So $S$ is an upper and $\emptyset$ is a lower bound for $A$.

**Definition 6 (Supremum & Infimum)** *Let $\Omega$ be a universe, $X \in \mathcal{P}(\Omega)$ be sets in the universe and $A \subset \mathcal{P}(\Omega)$, $A \neq \emptyset$ a non-empty selection of sets in the universe. If*

$$X = \sup A := \bigcup_{Y \in A} Y :\Leftrightarrow X \in \mathbf{U}_A \wedge \forall U \in \mathbf{U}_A : X \subset U \tag{3.13}$$

$$X = \inf A := \bigcap_{Y \in A} Y :\Leftrightarrow X \in \mathbf{L}_A \wedge \forall L \in \mathbf{L}_A : L \subset X \tag{3.14}$$

*then $X$ is called* supremum/infimum *for $A$.*

Existence for supremum and infimum is guaranteed, because upper and lower bounds exist as shown above. Thus, for any non-empty selection $A \subset \mathcal{P}(S)$ of a power-set, $\mathbf{U}_A$ and $\mathbf{L}_A$ are not empty. So we need to proof, that $X = \bigcup_{Y \in A} Y$ respectively $X = \bigcap_{Y \in A} Y$ are in fact the smallest upper and the largest lower bound. Or in other words: Does another bound $X' \in \mathbf{U}_A$ respectively $X' \in \mathbf{L}_A$ exist with $X' \neq X$ and $X' \subset X$ respectively $X \subset X'$?

1. Supremum: We assume $X' \in \mathbf{U}_A$ with $X' \neq X$ and $X' \subset X$ for $X = \bigcup_{Y \in A} Y$ exists, then an element $x \in X$ exists, which is not element of $X'$. Because $X$ is the union of all sets in selection $A$, $x$ must be element of at least one of its sets. However, then $X'$ cannot include sets containing $x$. Thus, $X'$ cannot be an upper bound for $A$ and $X = \sup A$.

2. Infimum: We assume $X' \in \mathbf{L}_A$ with $X' \neq X$ and $X \subset X'$ for $X = \bigcap_{Y \in A} Y$ exists, then an element $x \in X'$ exists, which is not element of $X$. Because $X$ is the intersection of all sets in selection $A$, $x$ cannot be element of at least one of its sets. However, then $X'$ must include sets containing $x$. Thus, $X'$ cannot be a lower bound fo $A$ and $X = \inf A$.

Supremum and Infimum are unique for any non-empty selection of sets in a universe and can be obtained by its union respectively its intersection.

**Definition 7 (Cartesian Product)** *Let $U$ be a universe and $X_n \in \mathcal{P}(U)$ sets with $i = 1...n, n \in \mathbb{N}$, then:*

$$X_1 \times ... \times X_n := \{(x_1, ..., x_n)\} \tag{3.15}$$

*is called* Cartesian product*.*

**Definition 8 (Relation)** *A relation is a subset of a Cartesian product:*

$$R \subset X_1 \times ... \times X_n \tag{3.16}$$

*The relation of only two sets is called* binary relation*. Instead of writing $(x, y) \in R$ we may also use the shorter notation $xRy$.*

An arbitrary relation $R \subset A \times B$ is called *homogeneous* if $A = B$, otherwise it is called *heterogeneous*. However, an arbitrary relation $R \subset A \times B$ is also homogeneous in the sense of $R \subset A \times B \subset (A \cup B) \times (A \cup B)$.

In order to clarify our notation, when we are specifically working with relations instead of ordinary sets, we use the symbols $\sqsubset$ for inclusion, $\sqcup$ for union and $\sqcap$ for intersection, i.e.:

$$R \sqsubset S :\Leftrightarrow \forall x, y[(x, y) \in R \Rightarrow (x, y) \in S] \tag{3.17}$$

$$R \sqcup S := \{(x, y)|(x, y) \in R \vee (x, y) \in S\} \tag{3.18}$$

$$R \sqcap S := \{(x, y)|(x, y) \in R \wedge (x, y) \in S\} \tag{3.19}$$

**Definition 9 (Properties of Relations)** *Let $R \subset A \times A$ and $S \subset A \times B$ be homogeneous or arbitrary relations, they may satisfy one or more of the following properties:*

$$\begin{aligned}
\text{bijective} &:\Leftrightarrow \forall b \in B \exists! a \in A : (a,b) \in S && (3.20)\\
\text{function} &:\Leftrightarrow \forall a \in A \exists! b \in B : (a,b) \in S && (3.21)\\
\text{reflexive} &:\Leftrightarrow \forall a \in A : (a,a) \in R && (3.22)\\
\text{irreflexive} &:\Leftrightarrow \forall a \in A : (a,a) \notin R && (3.23)\\
\text{transitive} &:\Leftrightarrow \forall a,b,c \in A : (a,b) \in R \wedge (b,c) \in R \Rightarrow (a,c) \in R && (3.24)\\
\text{intransitive} &:\Leftrightarrow \forall a,b,c \in A : (a,b) \in R \wedge (b,c) \in R \Rightarrow (a,c) \notin R && (3.25)\\
\text{symmetric} &:\Leftrightarrow \forall a,b \in A : (a,b) \in R \Rightarrow (b,a) \in R && (3.26)\\
\text{asymmetric} &:\Leftrightarrow \forall a,b \in A : (a,b) \in R \Rightarrow (b,a) \notin R && (3.27)\\
\text{antisymmetric} &:\Leftrightarrow \forall a,b \in A : (a,b) \in R \wedge (b,a) \in R \Rightarrow a = b && (3.28)
\end{aligned}$$

**Definition 10 (Composition of Binary Relations)** *The composition $S \circ R \subseteq A \times D$ of two binary relations $R \subseteq A \times B$ and $S \subseteq C \times D$ is defined as:*

$$S \circ R := \{(a,d) \in A \times D | \exists x \in B \cap C : (a,x) \in R \wedge (x,d) \in S\} \quad (3.29)$$

**Definition 11 (Identity Relation)** *The relation $I$ over a set $A$ with:*

$$\mathbf{I} := \{(a,b) \in A \times A | a = b\} = \{(a,a) | a \in A\} \subseteq A \times A \quad (3.30)$$

*is called identity relation.*

**Definition 12 (Exponentiation of Relations)** *The $n$-th power $R^n$ of a relation $R$ with $n \in \mathbb{N}$ is recursively defined with:*

$$\begin{aligned}
R^0 &:= \mathbf{I} && (3.31)\\
R^n &:= R \circ R^{n-1} && (3.32)
\end{aligned}$$

**Definition 13 (Reflexive Closure)** *Let $R$ be a homogeneous relation. Then*

$$R^\circ := \inf\{S | R \sqsubset S \wedge S \textit{ reflexive }\} = R \sqcup \mathbf{I} \quad (3.33)$$

*is called* reflexive closure *of R.*

The reflexive closure of a homogeneous relation $R$ is the infimum or largest lower bound of the set $A = \{S|R \sqsubset S \wedge S$ reflexive $\}$ containing all reflexive relations, which include $R$. The smallest reflexive relation is $\mathbf{I}$. The smallest relation including $R$ is $R$ itself. So, for an arbitrary relation $R' \in A$ the inclusions $R \sqsubset R'$, $\mathbf{I} \sqsubset R'$ and $(R \sqcup \mathbf{I}) \sqsubset R'$ hold. Thus $R \sqcup \mathbf{I}$ is a lower bound for $A$ and $(R \sqcup \mathbf{I}) \sqsubset \inf A$ holds. Vice versa, $R \sqcup \mathbf{I}$ is an element of $A$ and any relation $R'' \sqsubset (R \sqcup \mathbf{I})$ does either not include $R$ or is not reflexive. Thus $R \sqcup \mathbf{I}$ is also the smallest relation in $A$ and $\inf A \sqsubset (R \sqcup \mathbf{I})$. From $(R \sqcup \mathbf{I}) \sqsubset \inf A$ and $\inf A \sqsubset (R \sqcup \mathbf{I})$ follows $\inf A = (R \sqcup \mathbf{I})$.

### 3.1.1 Closures

For a relation $R$ over a set $A$ its closure:

- $R \cup I$ is called *reflexive closure*
- $R^+ := R^1 \cup R^2 \cup R^3 \cup ...$ is called *transitive closure*

$$
\begin{aligned}
(a,b) \in R^+ &\Leftrightarrow (a,b) \in R^1 \vee (a,b) \in R^2 \vee (a,b) \in R^3 \vee ... && (3.34)\\
&\Leftrightarrow \exists n \in \mathbb{N} : n \geq 1 \wedge (a,b) \in R^n && (3.35)\\
&\Leftrightarrow \exists n \in \mathbb{N} : n \geq 1 \wedge (a,b) \in R \circ R^{n-1} && (3.36)\\
&\Leftrightarrow \exists n \in \mathbb{N} : n \geq 1 \wedge [\exists x \in A : (a,x) \in R \wedge (x,b) \in R^{n-1}] && (3.37)\\
&\Leftrightarrow \exists n \in \mathbb{N} \exists x \in A : n \geq 1 \wedge (a,x) \in R \wedge (x,b) \in R^{n-1} && (3.38)
\end{aligned}
$$

- $R^* := R^+ \cup I$ is called *reflexive-transitive closure*

$$
\begin{aligned}
(a,b) \in R^* &\Leftrightarrow (a,b) \in R^+ \vee (a,b) \in I && (3.39)\\
&\Leftrightarrow (a,b) \in R^+ \vee a = b && (3.40)
\end{aligned}
$$

## 3.2 Mereology

$$
\begin{aligned}
&x \ \mathsf{partOf}\ x && \text{(Reflexivity)} && (3.41)\\
&x \ \mathsf{partOf}\ y \wedge y \ \mathsf{partOf}\ x \Rightarrow x = y && \text{(Antisymmetry)} && (3.42)\\
&x \ \mathsf{partOf}\ y \wedge y \ \mathsf{partOf}\ z \Rightarrow x \ \mathsf{partOf}\ z && \text{(Transitivity)} && (3.43)
\end{aligned}
$$

TBD.

## 3.3 Formal Languages & Grammars

### 3.3.1 Context-Free Languages & Grammars

## 3.4 Traceability

[2] TBD.

### 3.4.1 Traceability Relationship

### 3.4.2 Traceability Link

### 3.4.3 Traceability Recovery

### 3.4.4 Traceability Exploration

## 3.5 Megamodeling

TBD.

### 3.5.1 MegaL

#### 3.5.1.1 MegaL/Xtext

## 3.6 Ontologies

TBD.

## 3.7 Program Analysis

TBD.

## 3.8 XML Data Binding

TBD.

### 3.8.1 Java Architecture for XML Binding (JAXB)

## 3.9 Object Relational Mapping

TBD.

### 3.9.1 Java Persistence API (JPA)

### 3.9.2 Hibernate

## 3.10 Another Tool For Language Recognition (ANTLR)

# Chapter 4

# Hypotheses

TBD.

## 4.1 Fragments

## 4.2 Correspondence

## 4.3 Conformance

# Chapter 5

# Methodology

TBD.

## 5.1 The 101companies Human Resource Management System

Description of the 101companies Human Resource Management System

## 5.2 Link Proper Part Ratio

The ratio between all proper parts of two artifacts and the proper parts of the same artifacts in a relationship.

$$\pi_{R,A_1,A_2} = \frac{|\{(p_1,p_2) \in R : p_1 \text{ properPartOf } A_1 \wedge p_2 \text{ properPartOf } A_2\}|}{|\{p : p \text{ properPartOf } A_1 \vee p \text{ properPartOf } A_2\}|}$$

# Chapter 6

# Requirements

TBD.

R1 asdf

R2 asdf

R3 asdf

R4 asdf

# Chapter 7

# Design

TBD.

# Chapter 8

# Implementation

TBD.

## 8.1 Context-Free Grammar Fragmentation

## 8.2 Name Correspondence Heuristic

Heuristics are quick and "simple" methods for finding good approximate solutions for complex problems. The Name Correspondence Heuristic determines correspondence between artifacts simply by finding similarities of names in those artifacts.

TBD.

# Chapter 9

# Results

TBD.

# Chapter 10

# Conclusion

TBD.

# Bibliography


[1] Gunther Schmidt and Thomas Ströhlein. *Relationen und Graphen*. Mathematik für Informatiker. Springer, 1989. ISBN: 3-540-50304-8.

[2] Stefan Winkler and Jens Pilgrim. "A Survey of Traceability in Requirements 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. URL: http://dx.doi.org/10.1007/s10270-009-0145-0.