

Masterarbeit

Entwurf und Aufbau einer semantischen Repräsentation von Quelltext

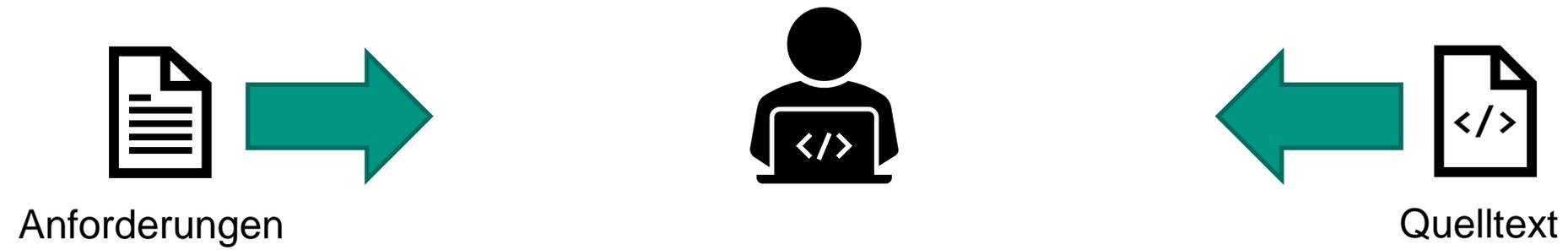
Felix Eurich

Betreut von Tobias Hey

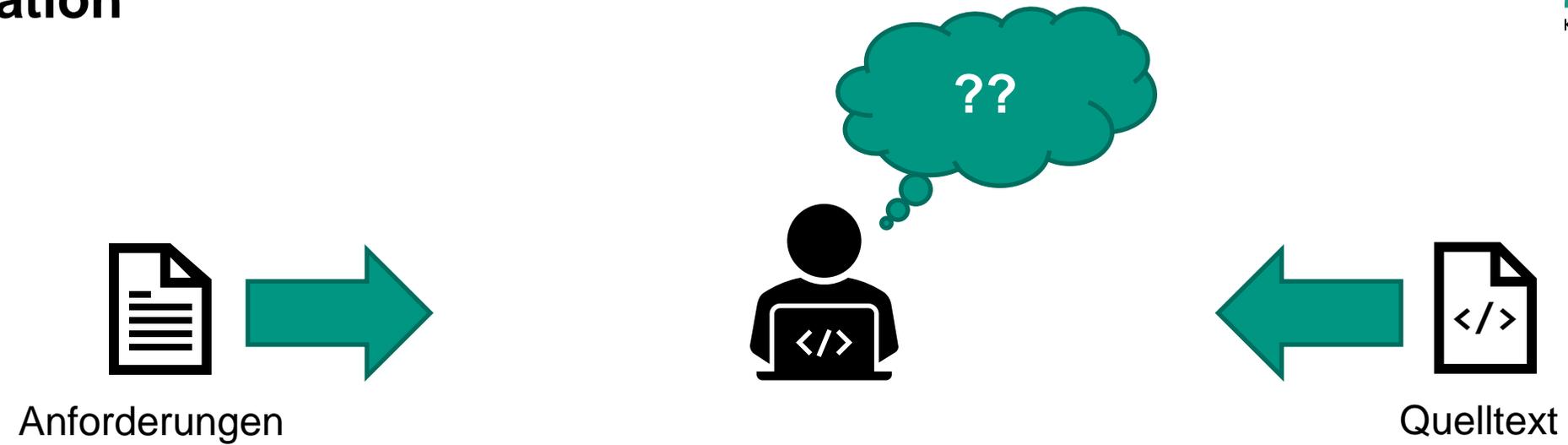
IPD Tichy, Fakultät für Informatik



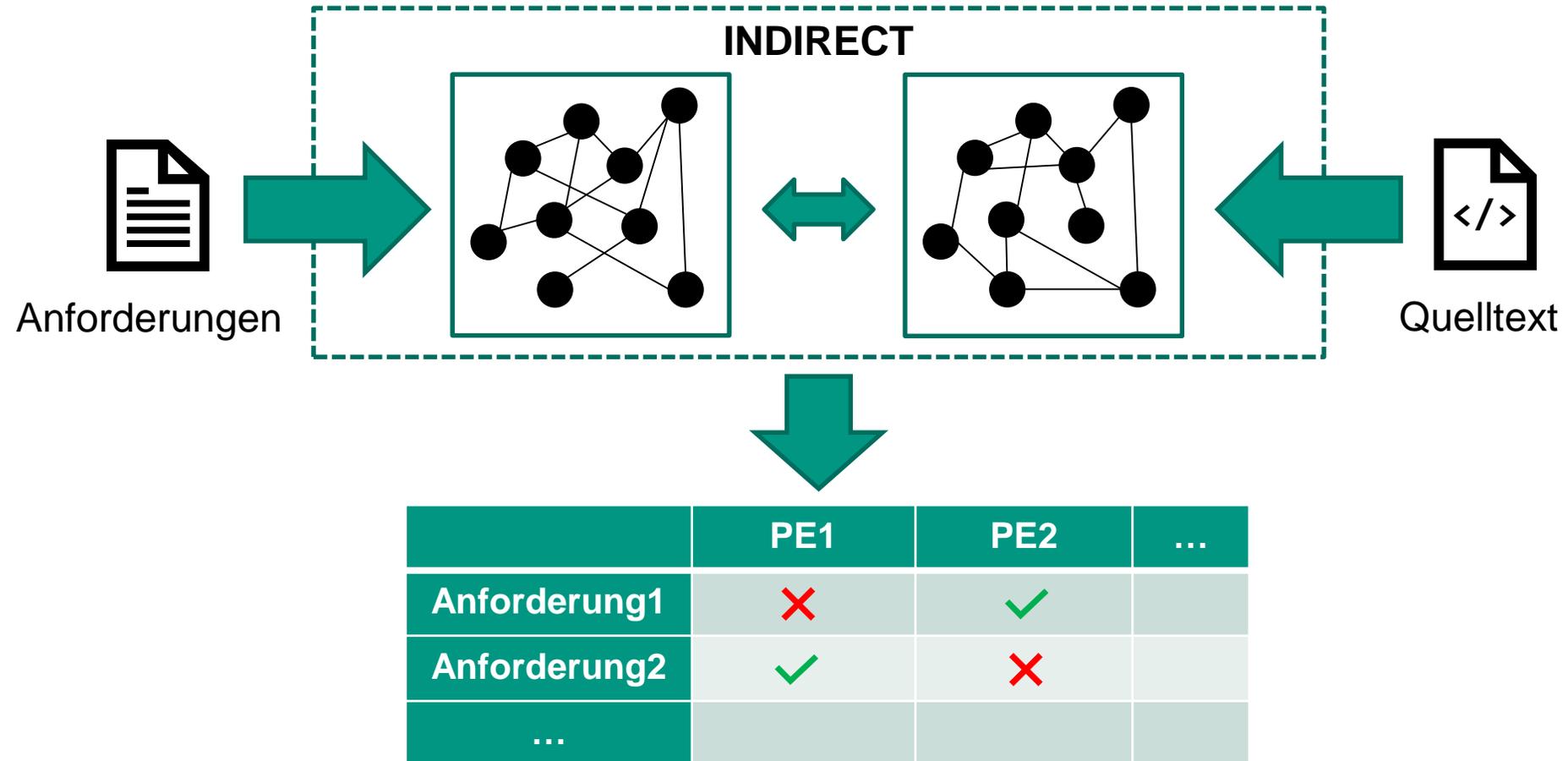
Motivation



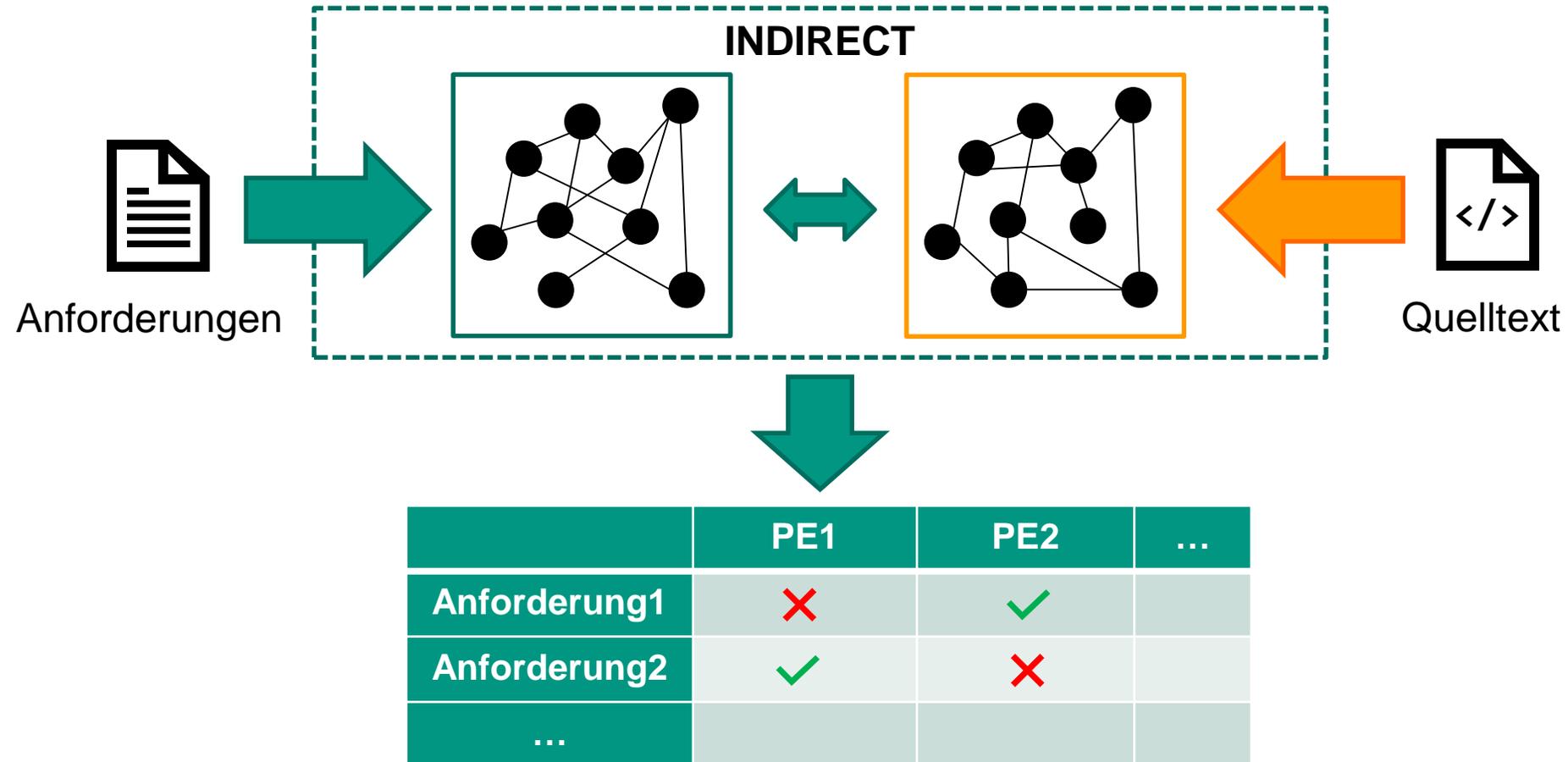
Motivation



Motivation

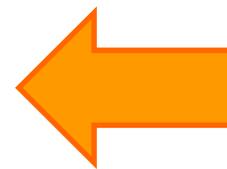
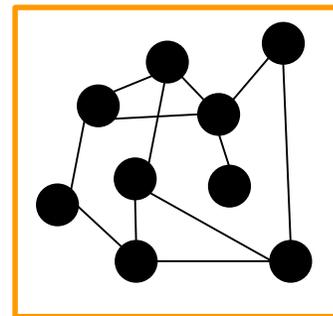
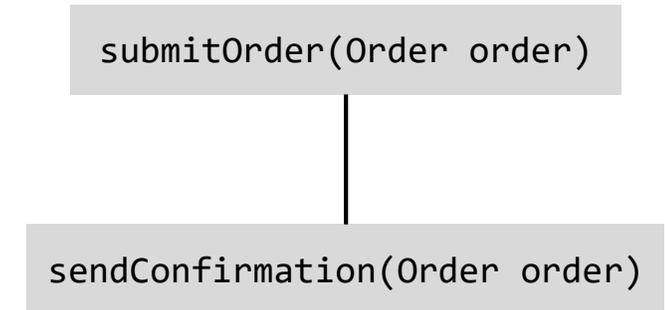
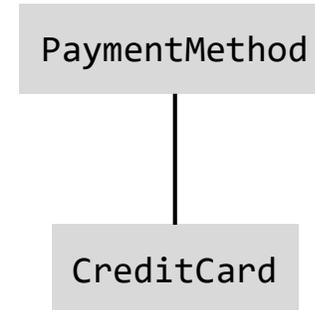


Motivation



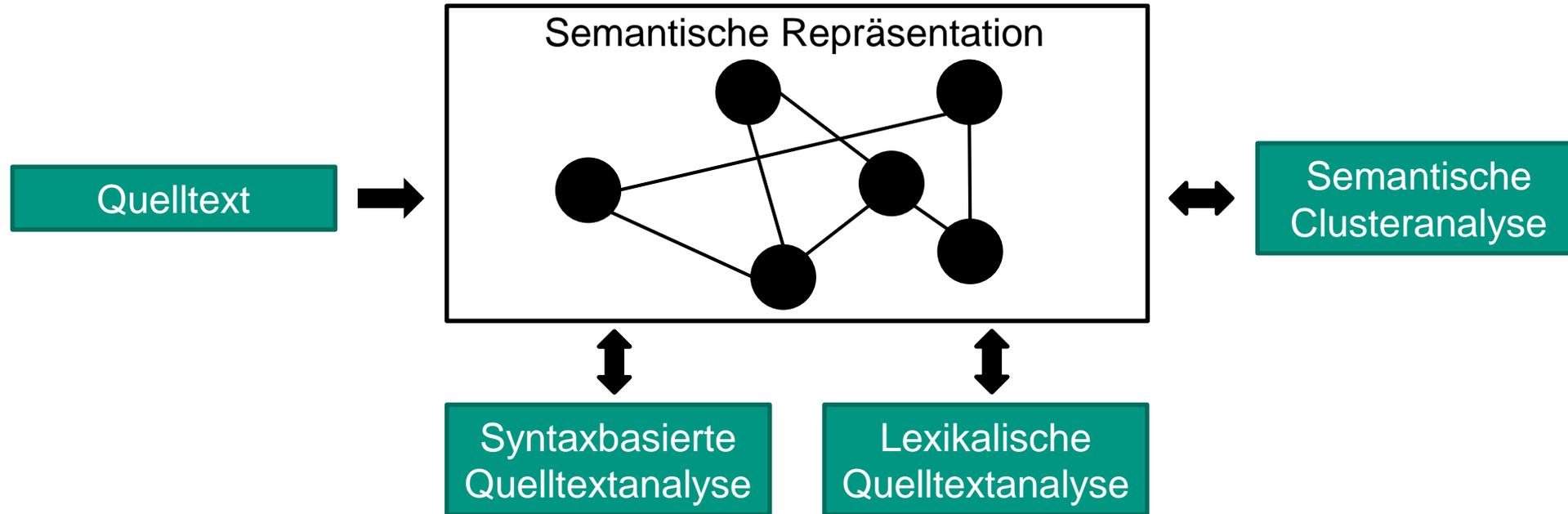
Zielsetzung

- Entwurf einer Graphrepräsentation
- Identifikation semantischer Zusammenhänge
 - Programmelemente mit gemeinsamer Absicht
- Einfache Erweiterbarkeit

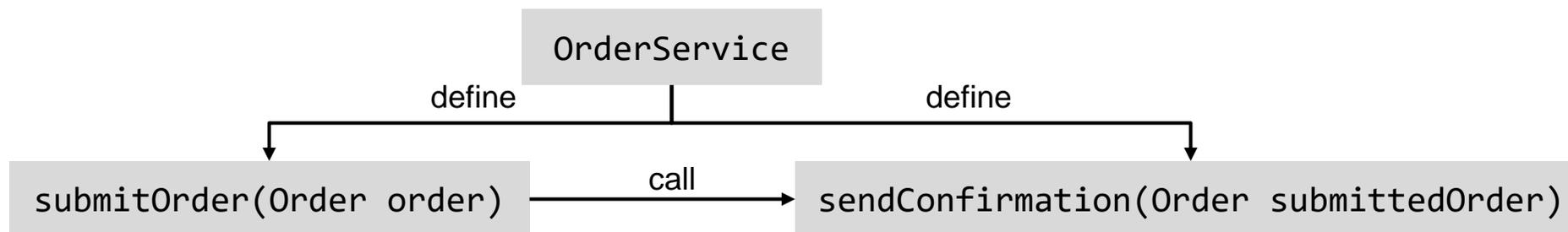
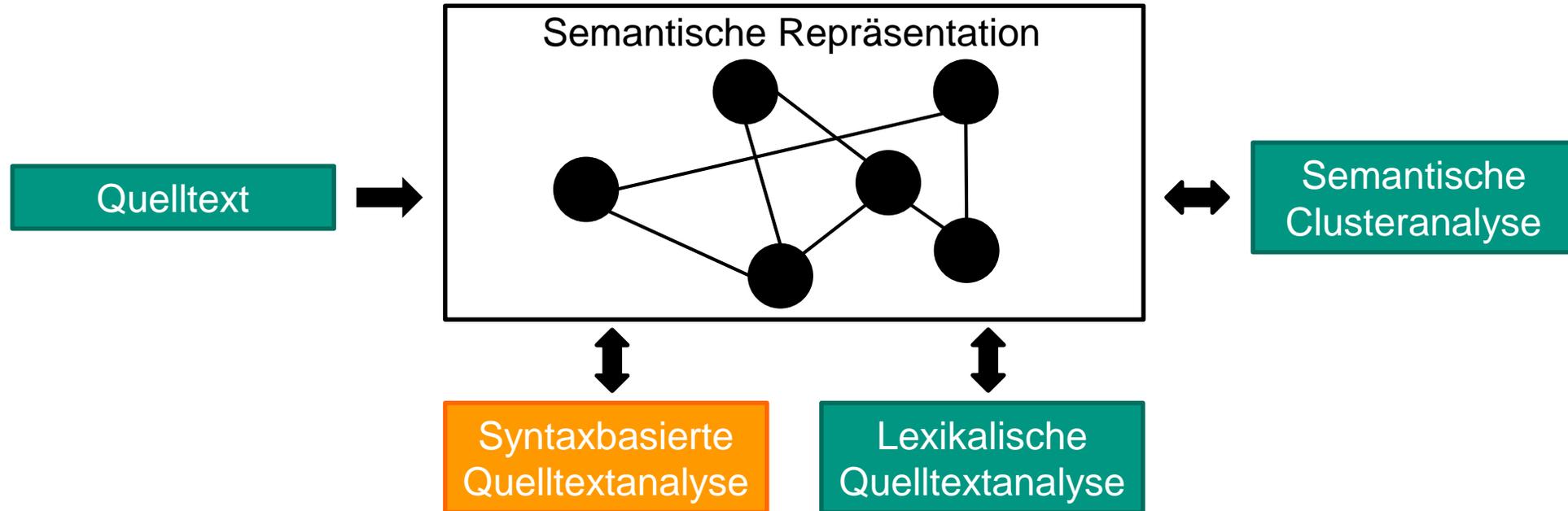


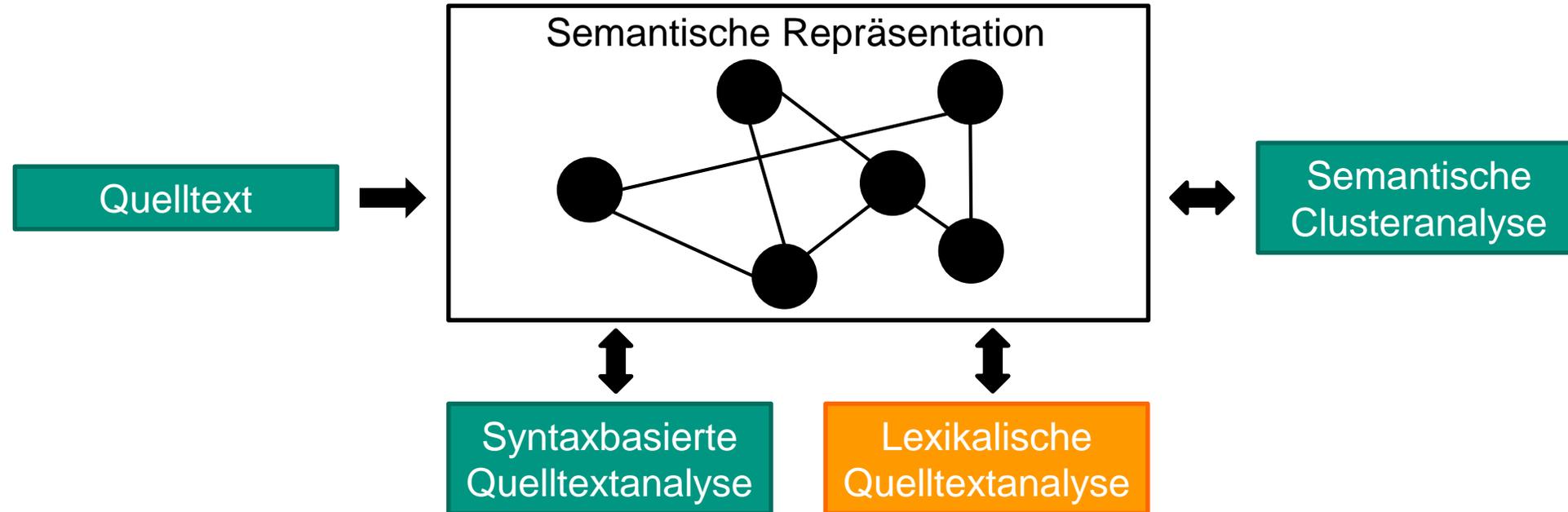
Quelltext

Ansatz



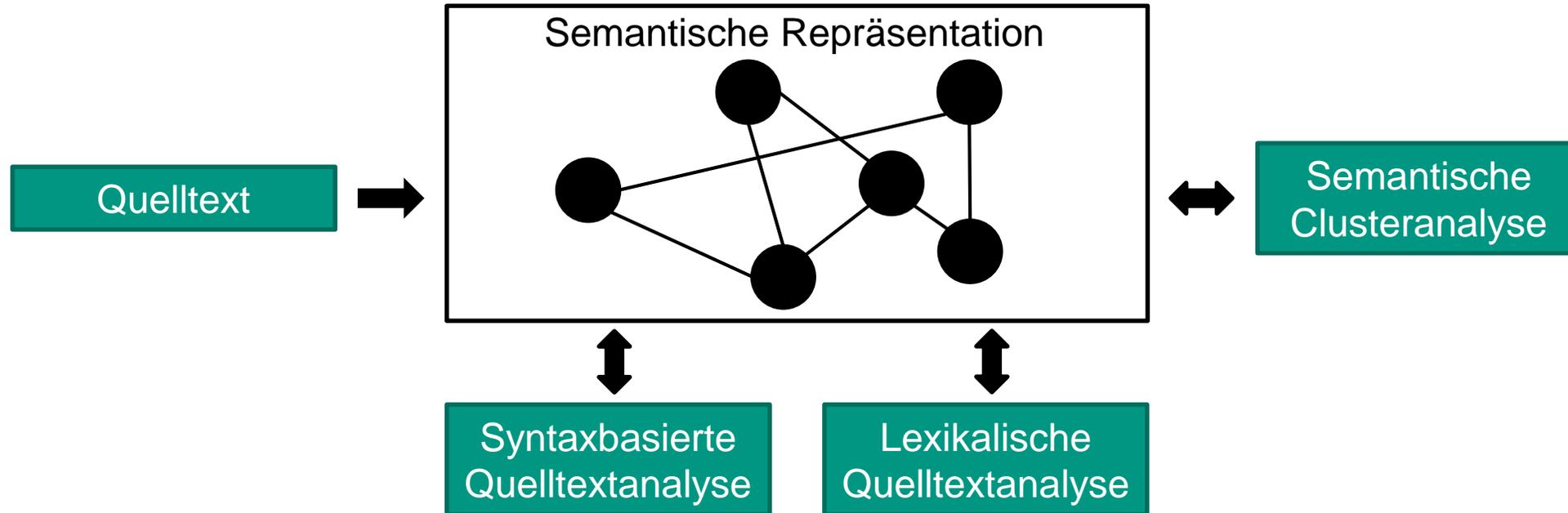
Ansatz





```
//Sends an order confirmation to the logged-in user  
void sendConfirmation(Order submittedOrder) {...}
```

Ansatz



Verwandte Arbeiten

- Anforderungsrückverfolgung [KNH+17]
- Konzeptlokalisierung [GAL18] [CDMMS16] [MAK+12] [SM11] [RM08] [KDG07]
- Software-Remodularisierung [AC17] [LOV08]

- Einschränkungen
 - Repräsentation von Zusammenhängen
 - Beschränkt auf Typen oder Methoden
 - Syntaxbasierte Analyse
 - Wenige Abhängigkeitsmerkmale
 - Lexikalische Analyse
 - Kein externes Wissen
 - Abkürzungen nicht berücksichtigt

Entwurf der semantischen Repräsentation

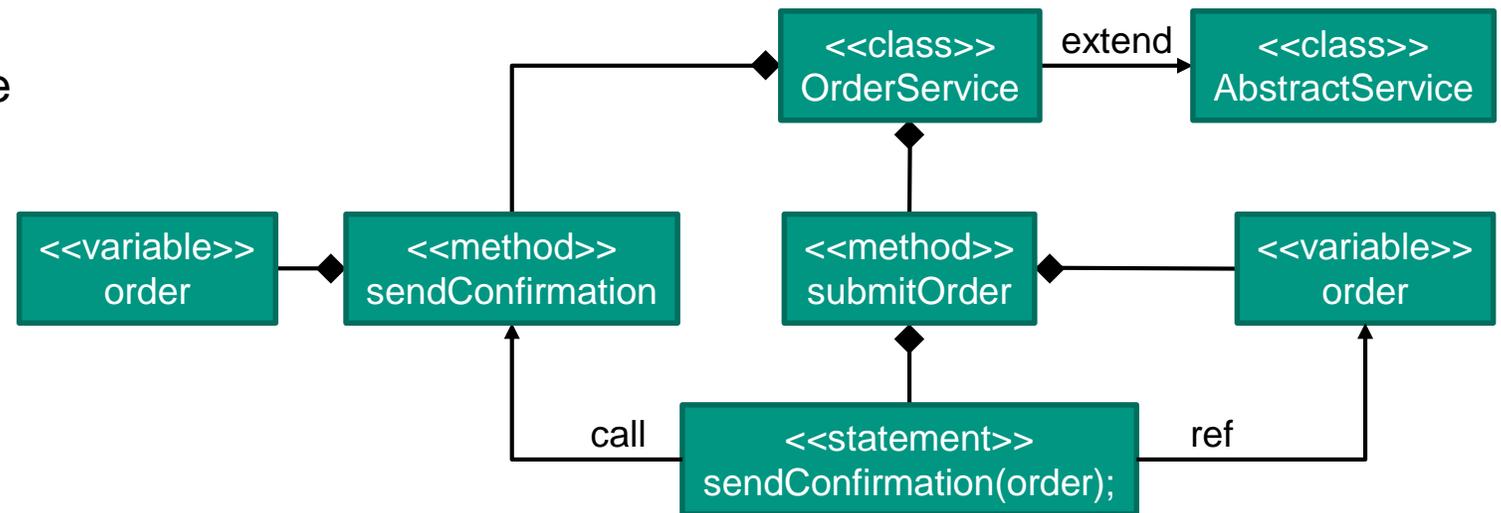
■ Quelltextelemente

- Knoten
 - Programmelemente
- Kanten
 - Definitionsursprung
 - Abhängigkeiten

```
public class OrderService extends AbstractService {
    ...
    public void submitOrder(Order order) {
        ...
        sendConfirmation(order);
    }
}
```

■ Semantische Elemente

- Semantische Zusammenhänge
- Absichtsknoten



Entwurf der semantischen Repräsentation

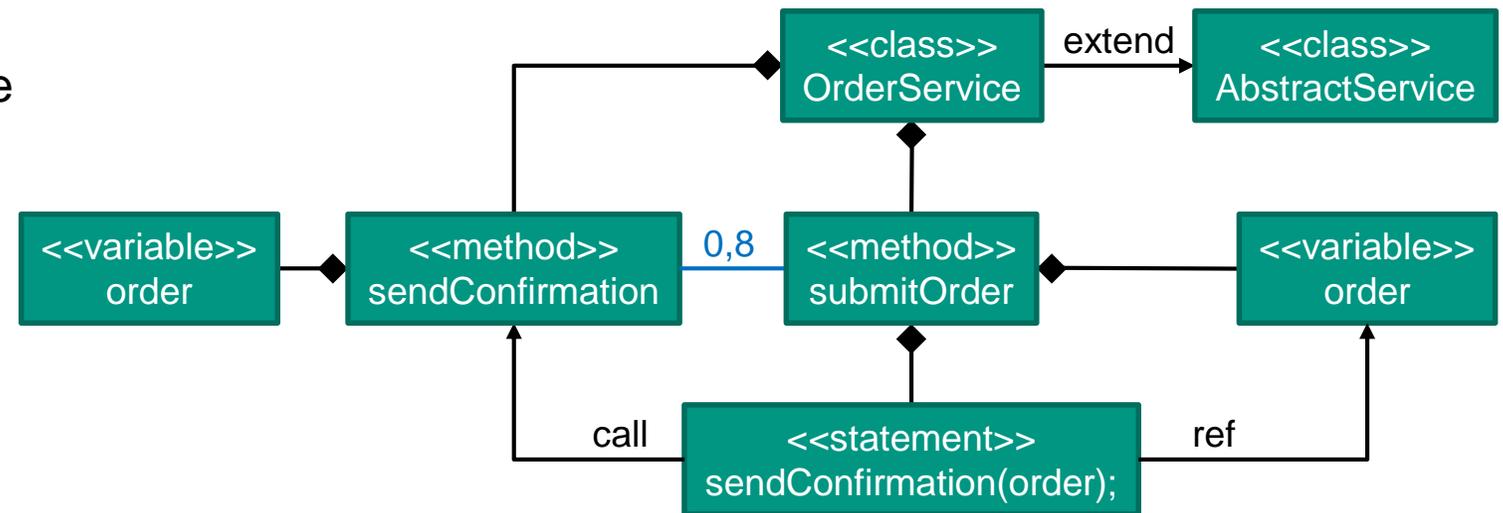
■ Quelltextelemente

- Knoten
 - Programmelemente
- Kanten
 - Definitionsursprung
 - Abhängigkeiten

```
public class OrderService extends AbstractService {
    ...
    public void submitOrder(Order order) {
        ...
        sendConfirmation(order);
    }
}
```

■ Semantische Elemente

- Semantische Zusammenhänge
- Absichtsknoten



Entwurf der semantischen Repräsentation

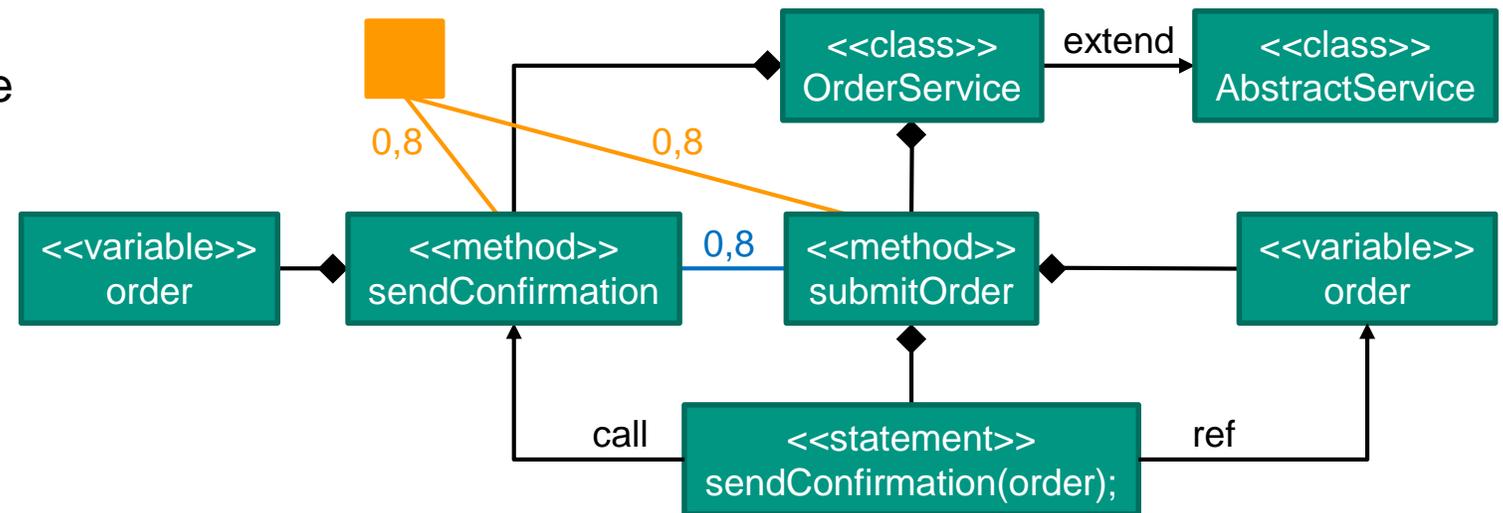
■ Quelltextelemente

- Knoten
 - Programmelemente
- Kanten
 - Definitionsursprung
 - Abhängigkeiten

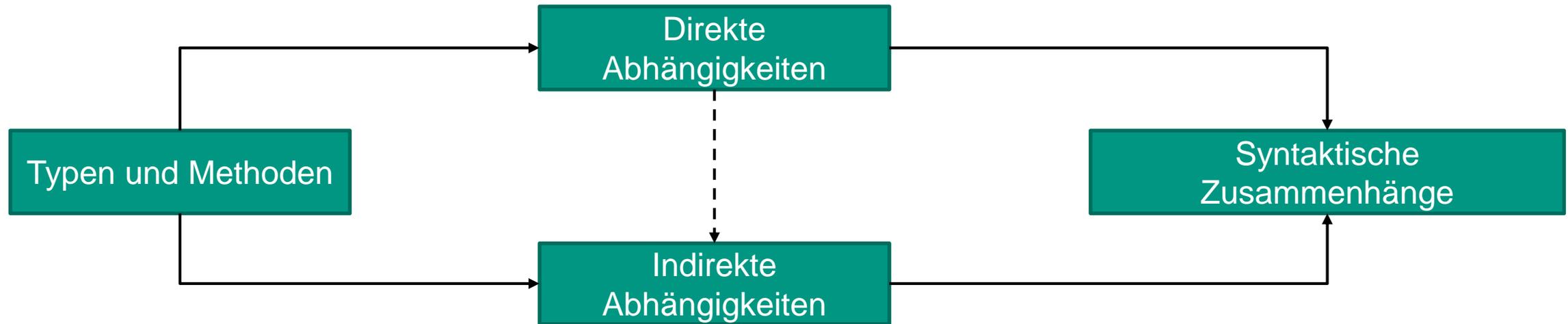
```
public class OrderService extends AbstractService {
    ...
    public void submitOrder(Order order) {
        ...
        sendConfirmation(order);
    }
}
```

■ Semantische Elemente

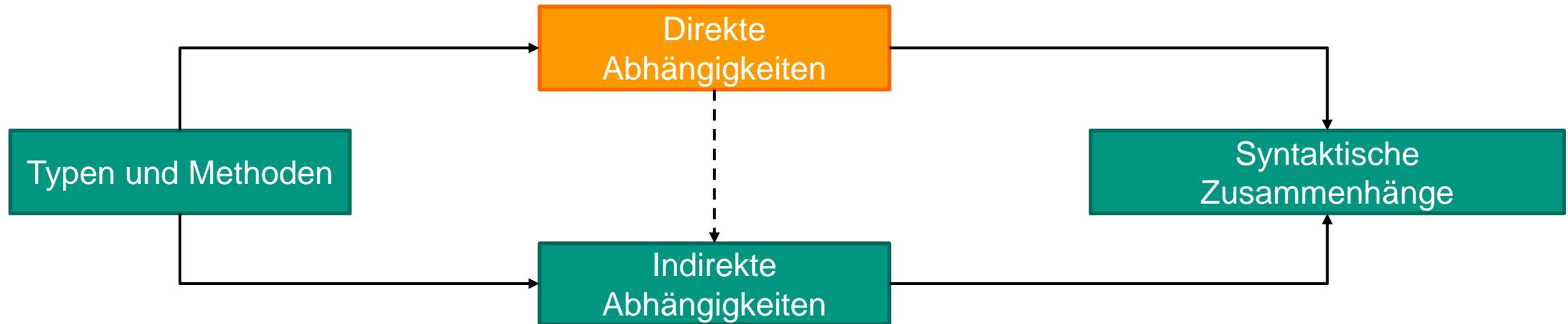
- Semantische Zusammenhänge
- Absichtsknoten



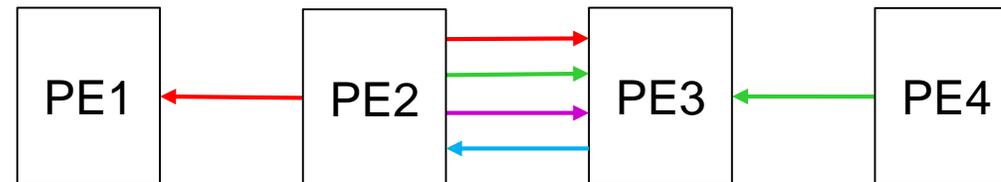
Syntaxbasierte Quelltextanalyse



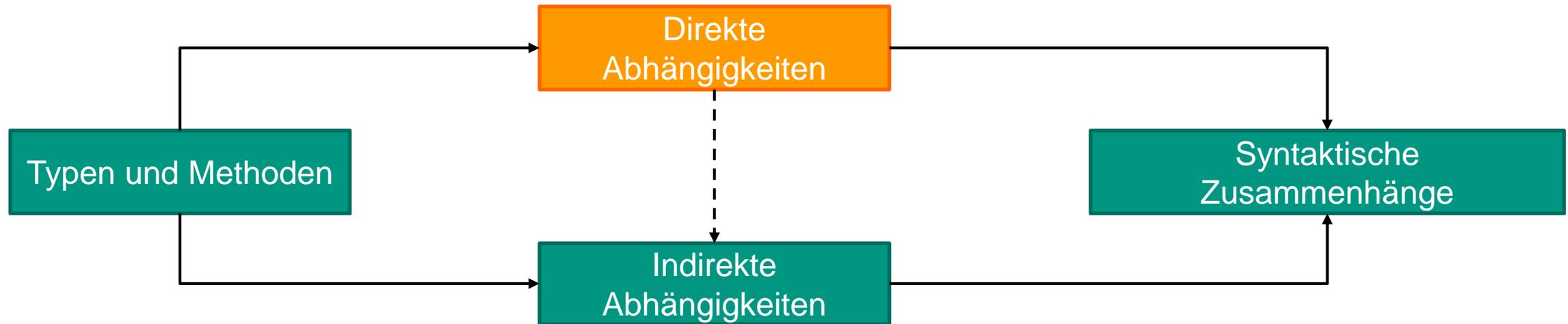
Syntaxbasierte Quelltextanalyse



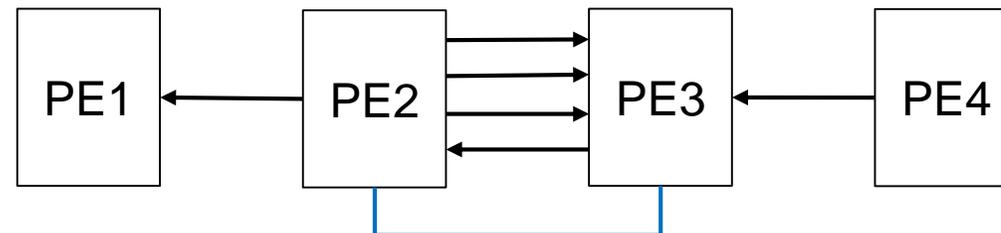
- 15 Abhängigkeitsmerkmale
- Anweisungsebene eingeschlossen
- Merkmale individuell gewichtet



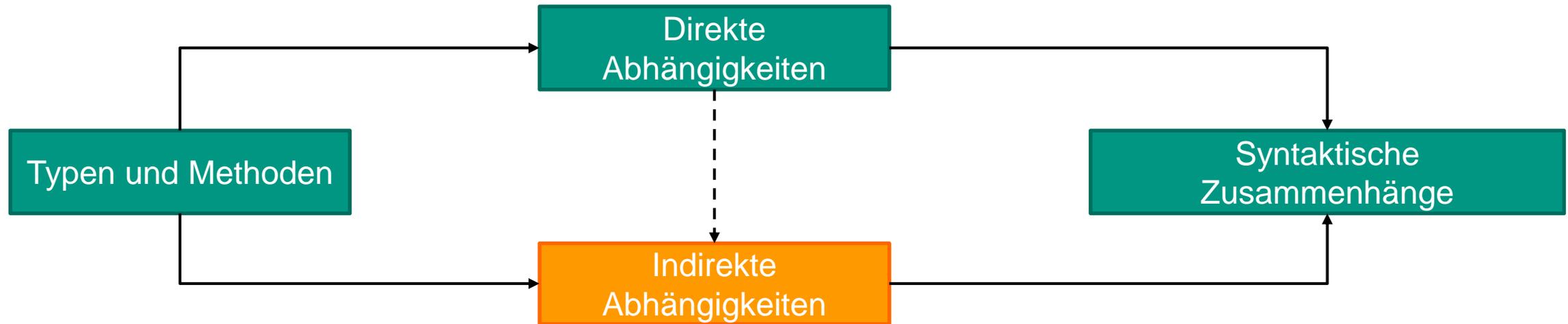
Syntaxbasierte Quelltextanalyse



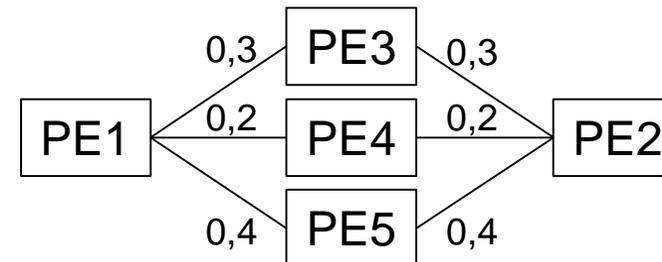
- 15 Abhängigkeitsmerkmale
- Anweisungsebene eingeschlossen
- Merkmale individuell gewichtet



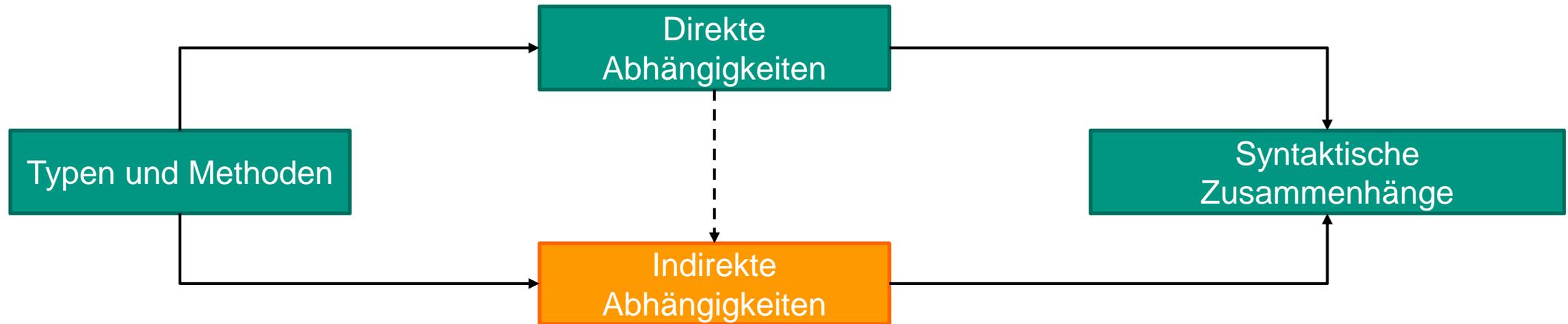
Syntaxbasierte Quelltextanalyse



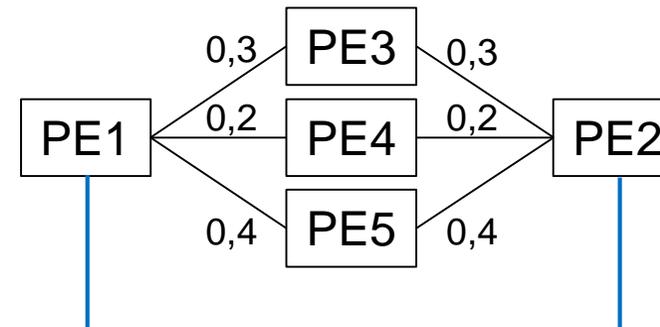
- Verwendungskontext
- Vektorabbildung
- Basiert auf direkten Abhängigkeiten
- Kosinus-Ähnlichkeit



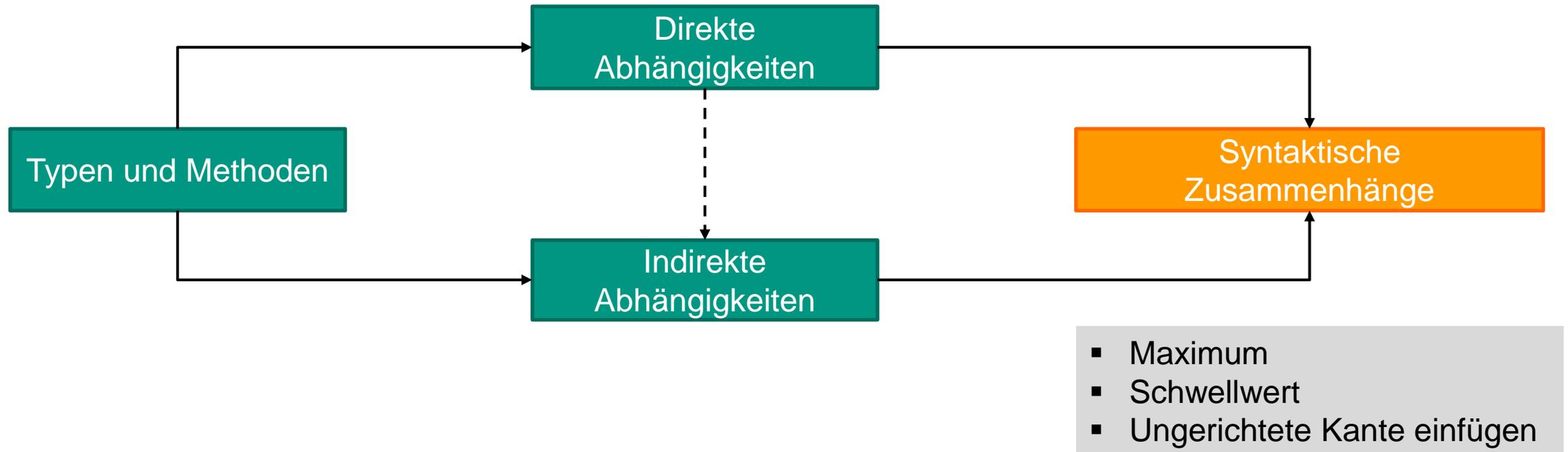
Syntaxbasierte Quelltextanalyse



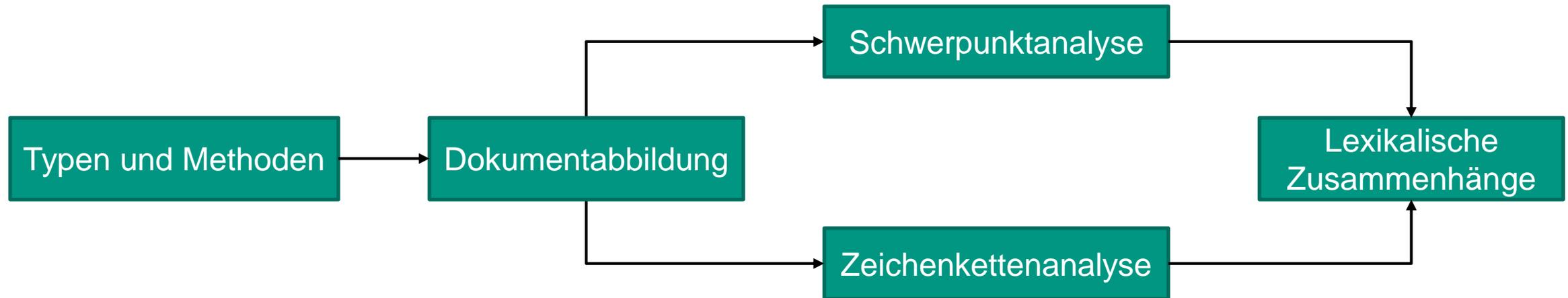
- Verwendungskontext
- Vektorabbildung
- Basiert auf direkten Abhängigkeiten
- Kosinus-Ähnlichkeit



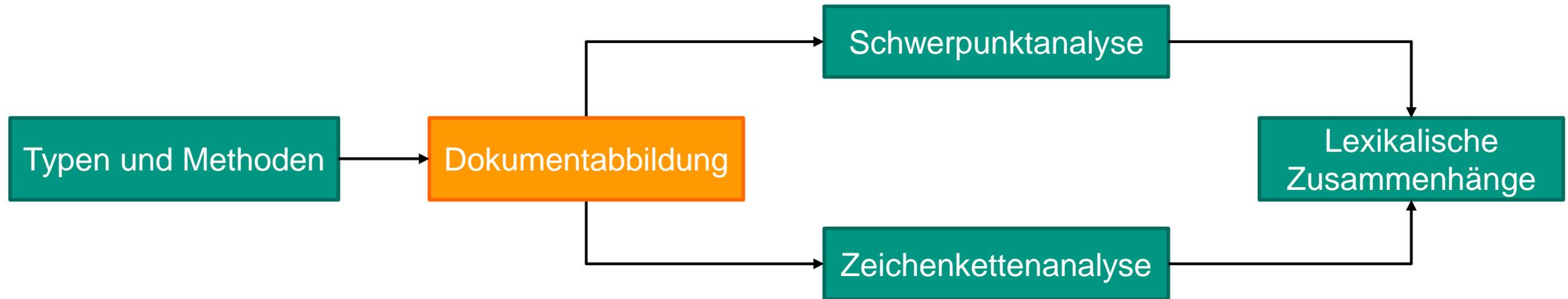
Syntaxbasierte Quelltextanalyse



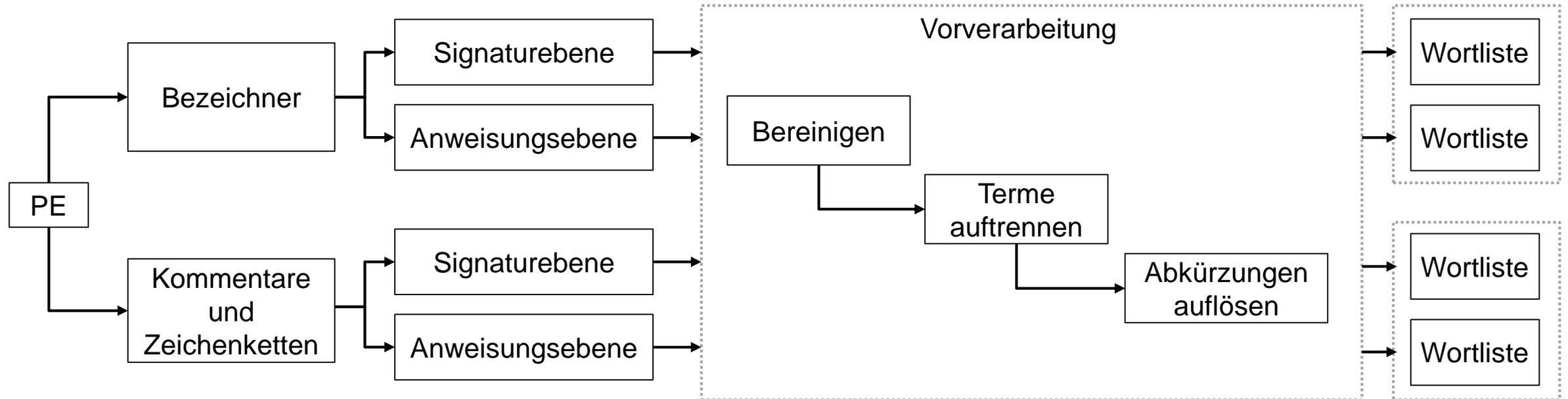
Lexikalische Quelltextanalyse



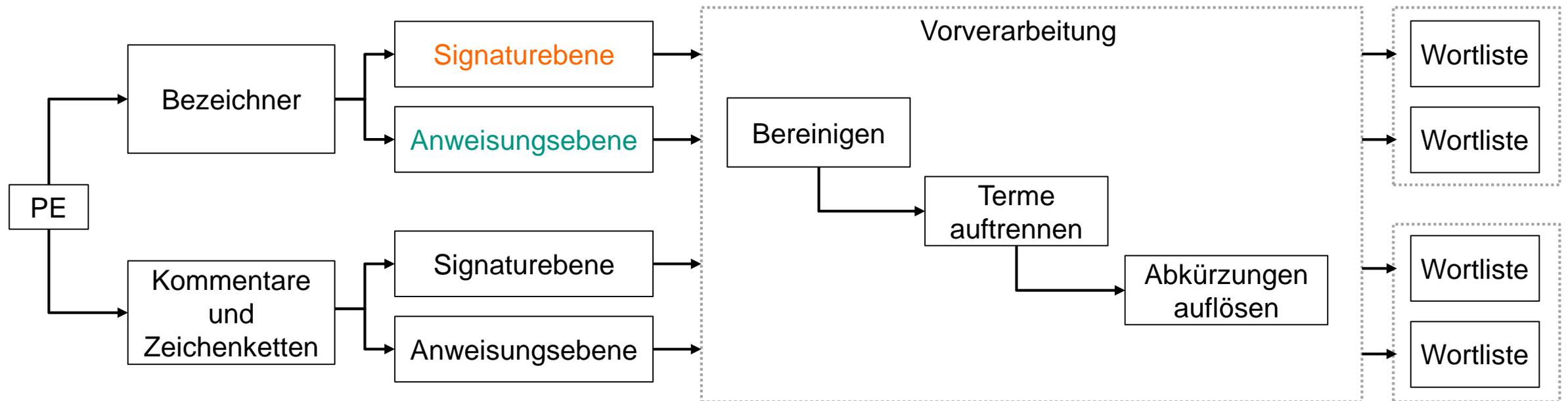
Lexikalische Quelltextanalyse



Lexikalische Quelltextanalyse



Lexikalische Quelltextanalyse

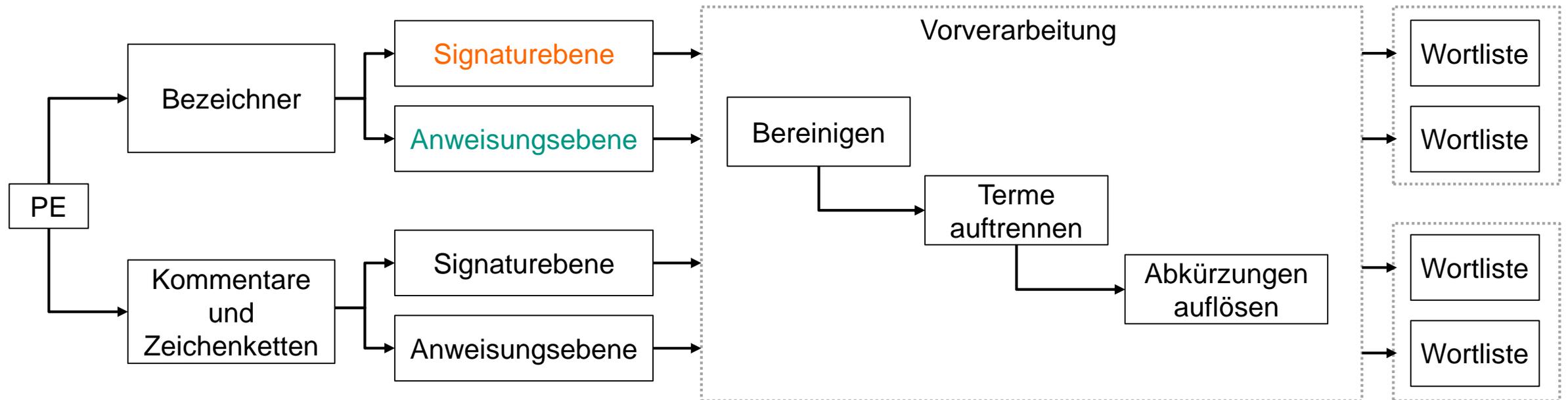


```

public double calcCompensation(Employee empl) {
    ...
    double workingHours = getWorkingHours(empl.getId());
    ...
}

```

Lexikalische Quelltextanalyse

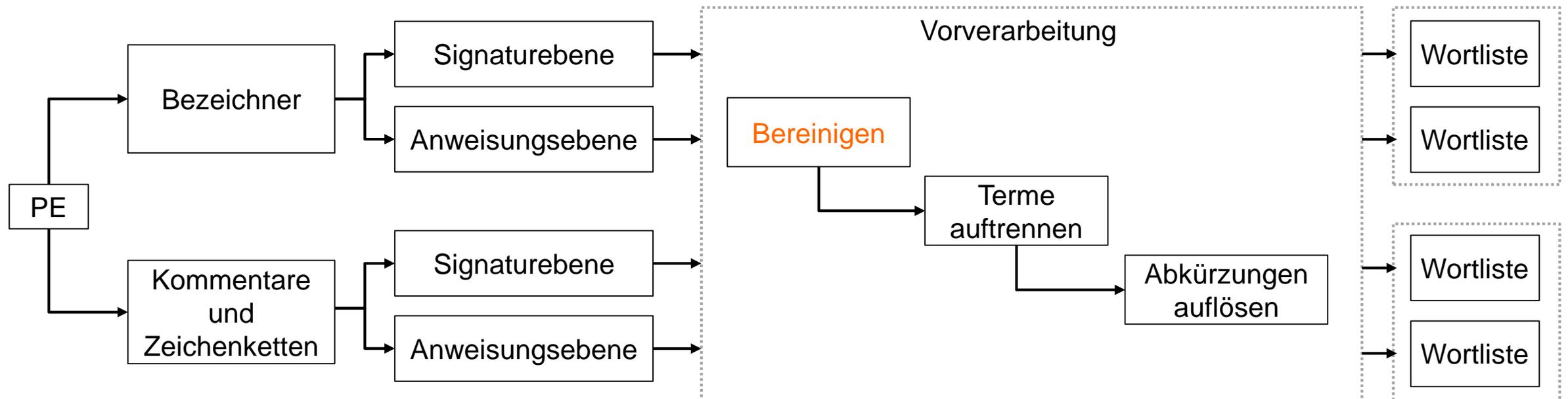


```
public double calcCompensation(Employee empl) {
  ...
  double workingHours = getWorkingHours(empl.getId());
  ...
}
```

public, double, calcCompensation,
employee, empl

double, workingHours,
getWorkingHours, empl, getId

Lexikalische Quelltextanalyse

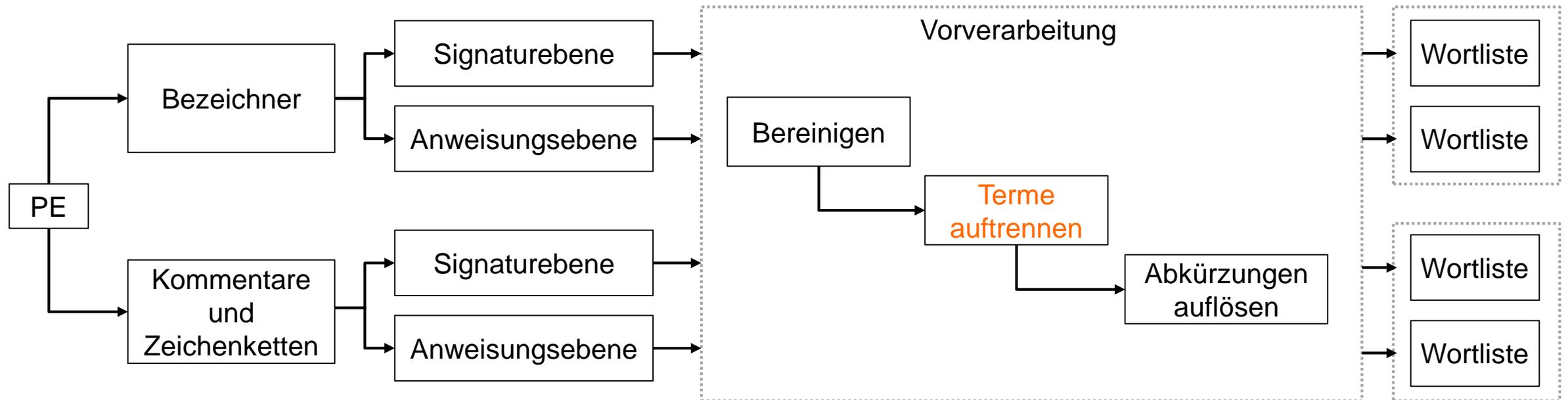


```
public double calcCompensation(Employee empl) {
  ...
  double workingHours = getWorkingHours(empl.getId());
  ...
}
```

calcCompensation, employee,
empl

workingHours, getWorkingHours,
empl, getId

Lexikalische Quelltextanalyse

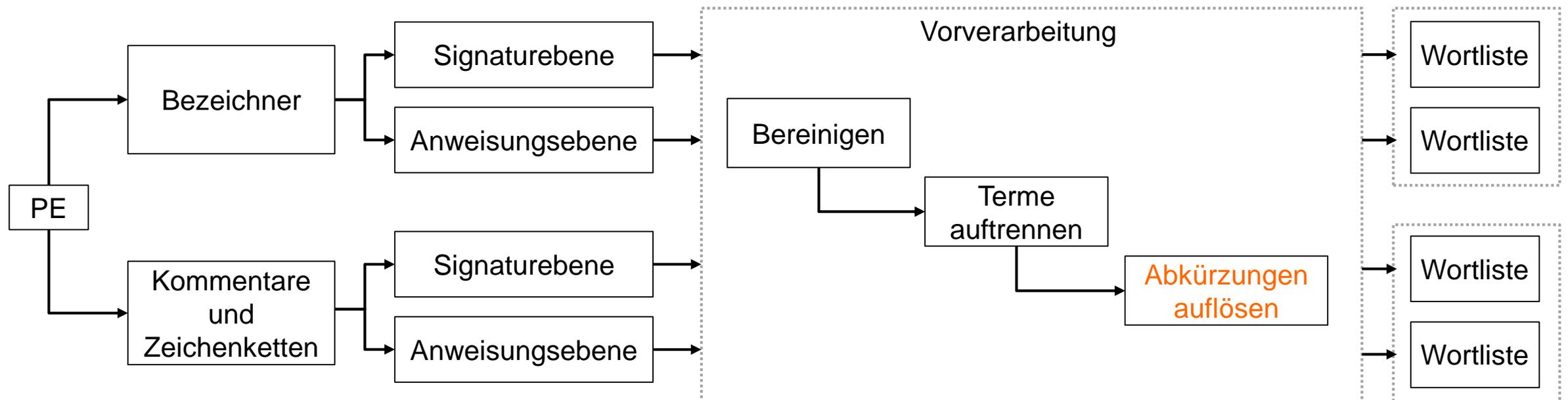


```
public double calcCompensation(Employee empl) {
  ...
  double workingHours = getWorkingHours(empl.getId());
  ...
}
```

calc, compensation, employee,
empl

working, hours, get, working,
hours, empl, get, id

Lexikalische Quelltextanalyse

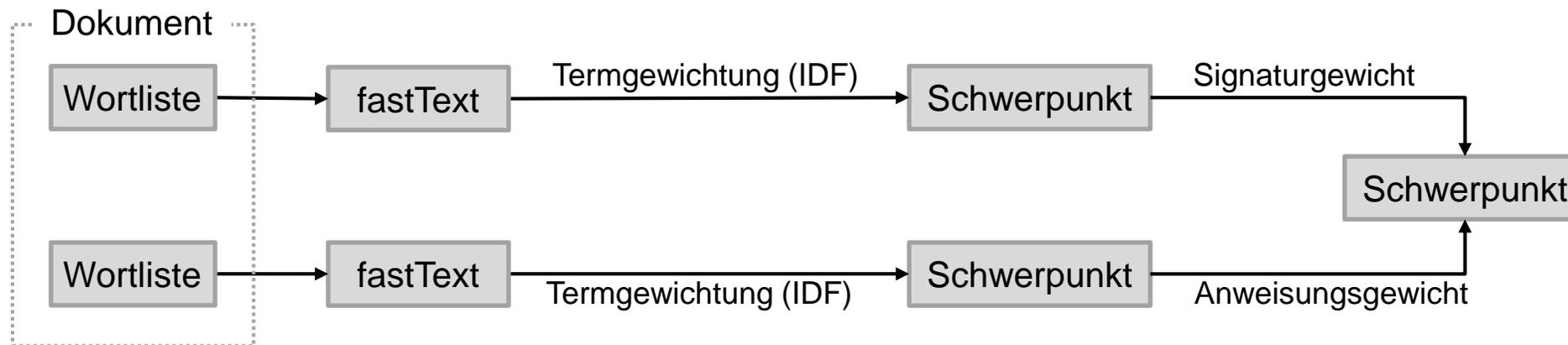
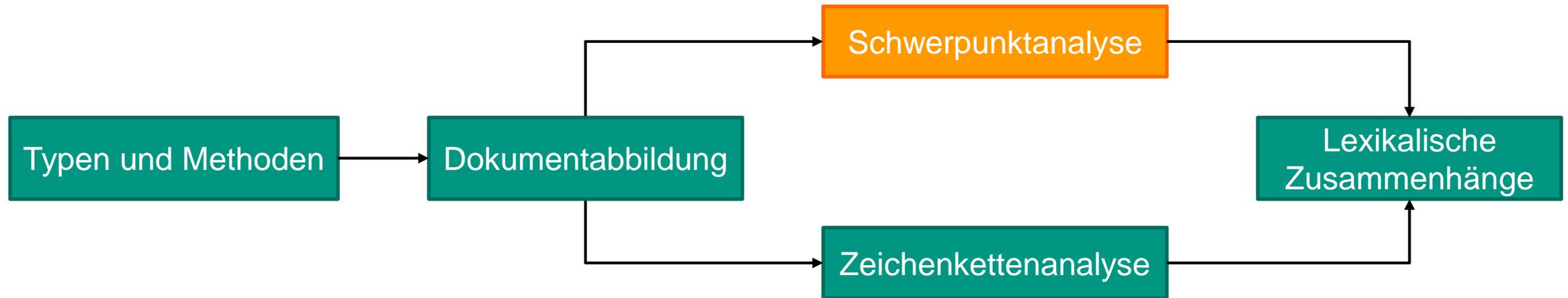


```
public double calcCompensation(Employee empl) {
  ...
  double workingHours = getWorkingHours(empl.getId());
  ...
}
```

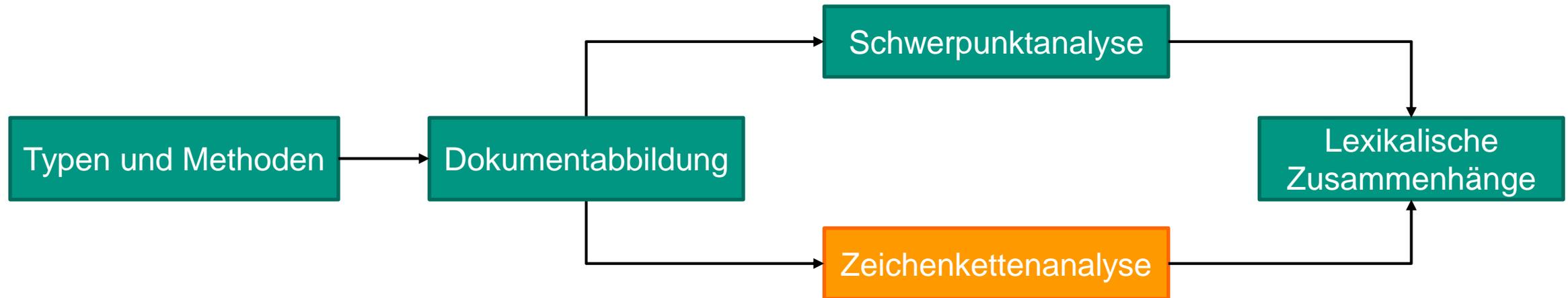
calculate, compensation,
employee, employee

working, hours, get, working,
hours, employee, get, identifier

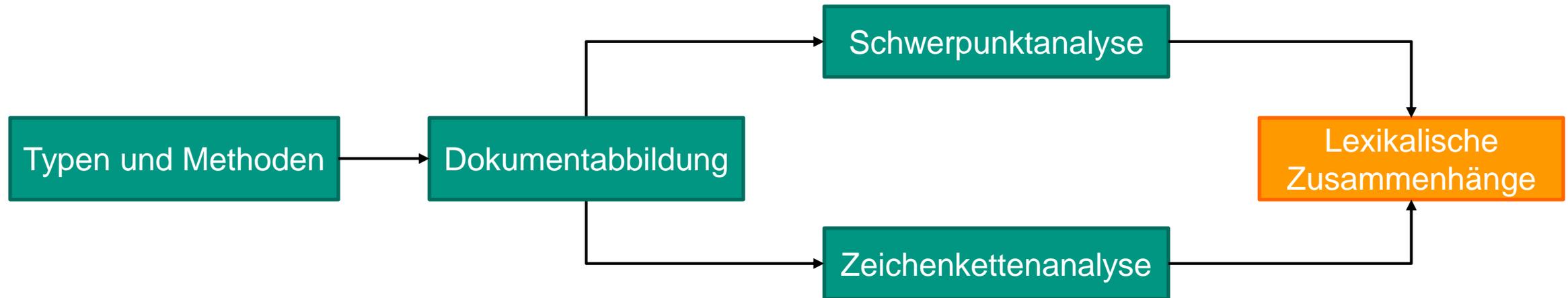
Lexikalische Quelltextanalyse



Lexikalische Quelltextanalyse



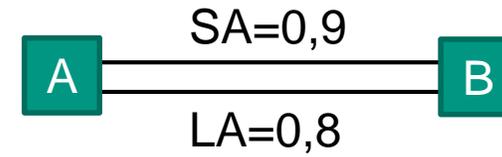
Lexikalische Quelltextanalyse



- Gewichteter Mittelwert
- Schwellwert
- Ungerichtete Kante einfügen

Semantische Clusteranalyse

- Semantischer Zusammenhang
 - Gewichteter Mittelwert



Semantische Clusteranalyse

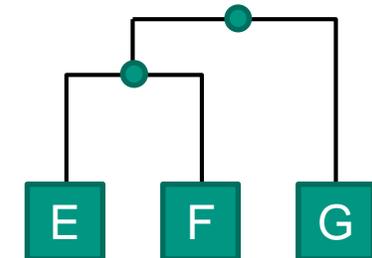
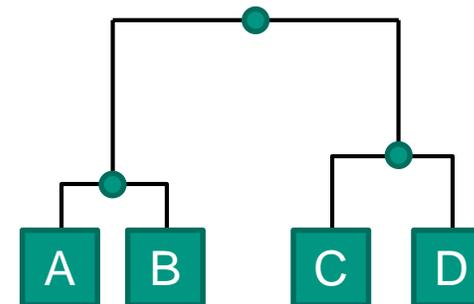
- Semantischer Zusammenhang
 - Gewichteter Mittelwert



Semantische Clusteranalyse

- Semantischer Zusammenhang
 - Gewichteter Mittelwert

- Hierarchische Clusteranalyse
 - Agglomerativ
 - Durchschnittlicher semantischer Zusammenhang
 - Schwellwert

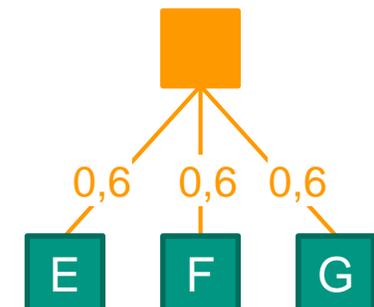
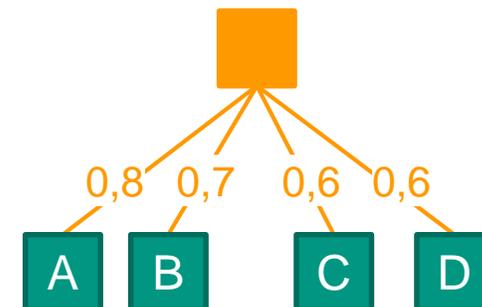
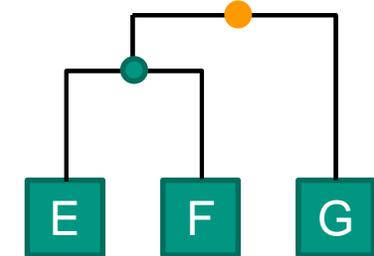
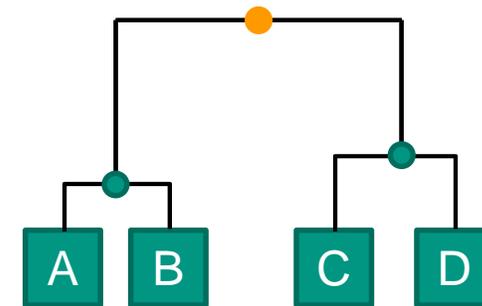


Semantische Clusteranalyse

- Semantischer Zusammenhang
 - Gewichteter Mittelwert

- Hierarchische Clusteranalyse
 - Agglomerativ
 - Durchschnittlicher semantischer Zusammenhang
 - Schwellwert

- Absichtsknoten
 - Cluster der höchsten Hierarchieebene



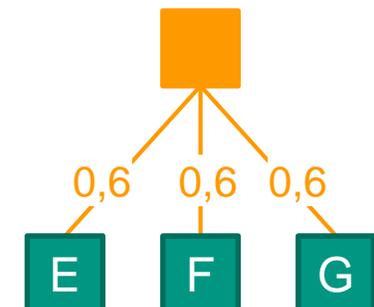
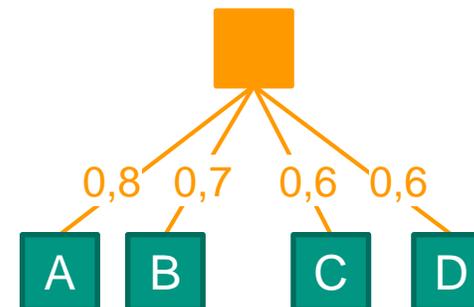
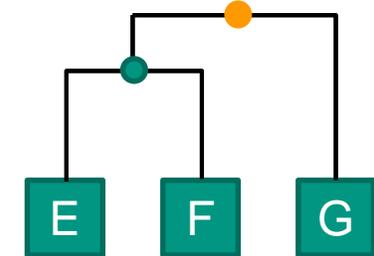
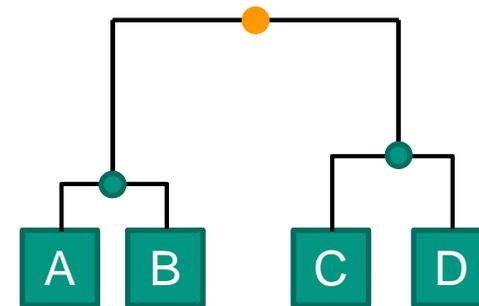
Semantische Clusteranalyse

- Semantischer Zusammenhang
 - Gewichteter Mittelwert

- Hierarchische Clusteranalyse
 - Agglomerativ
 - Durchschnittlicher semantischer Zusammenhang
 - Schwellwert

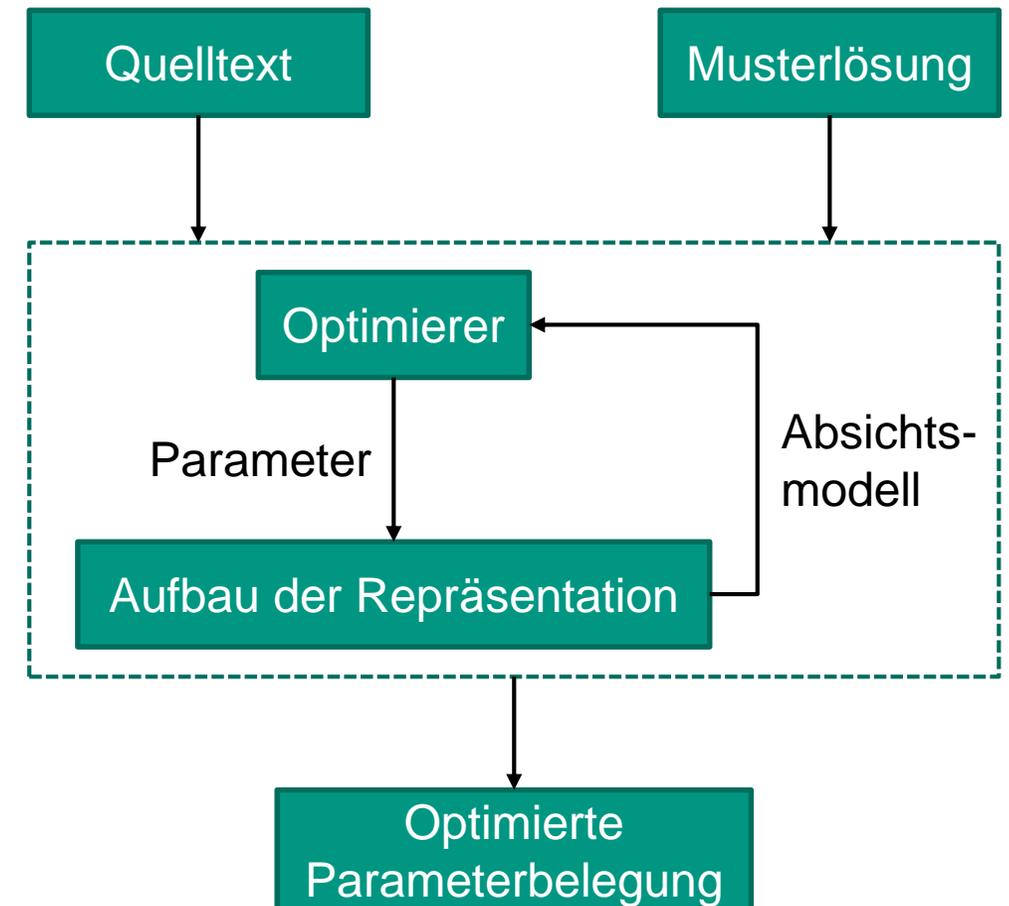
- Absichtsknoten
 - Cluster der höchsten Hierarchieebene

- Hinzunahme weiterer Quelltextanalysen einfach möglich



Parameteroptimierung

- 46 Parameter
 - Gewichte
 - Schwellwerte
- Evolutionäre Optimierung
 - 35 Parameterbelegungen pro Iteration
 - Max. 30 Iterationen
- Dronology
 - 59 Typen, 343 Methoden
 - Musterlösung
 - Rückverfolgbarkeitsinformationen auf Typebene
 - Erweitert um Methoden



Evaluation

- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)

Evaluation

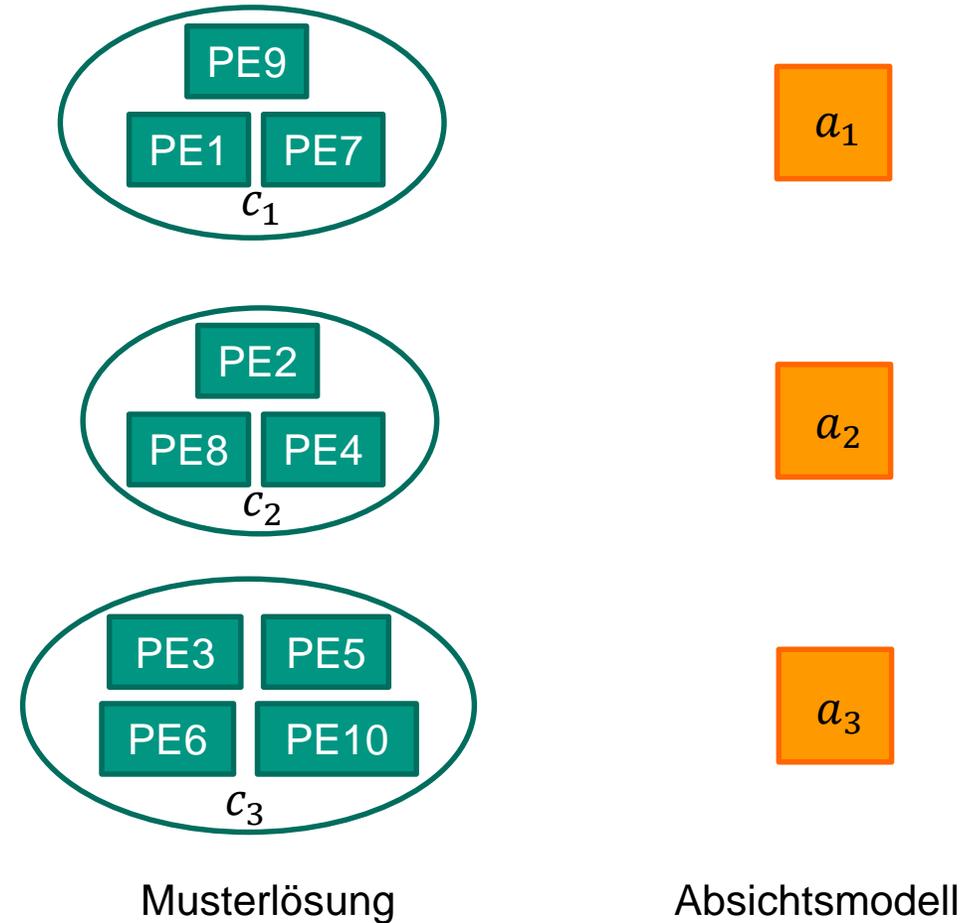
- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)
- Bewertung der Typ- und Methodenunterteilung
 - Keine offizielle Musterlösung
 - Vollständige Musterlösung nicht praktikabel
 - Systemausgabe in Musterlösung überführt

Evaluation

- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)
- Bewertung der Typ- und Methodenunterteilung
 - Keine offizielle Musterlösung
 - Vollständige Musterlösung nicht praktikabel
 - Systemausgabe in Musterlösung überführt
- Bewertung der Typunterteilung für iTrust
 - Offizielle Rückverfolgbarkeitsinformationen
 - Datensatz nicht vollständig

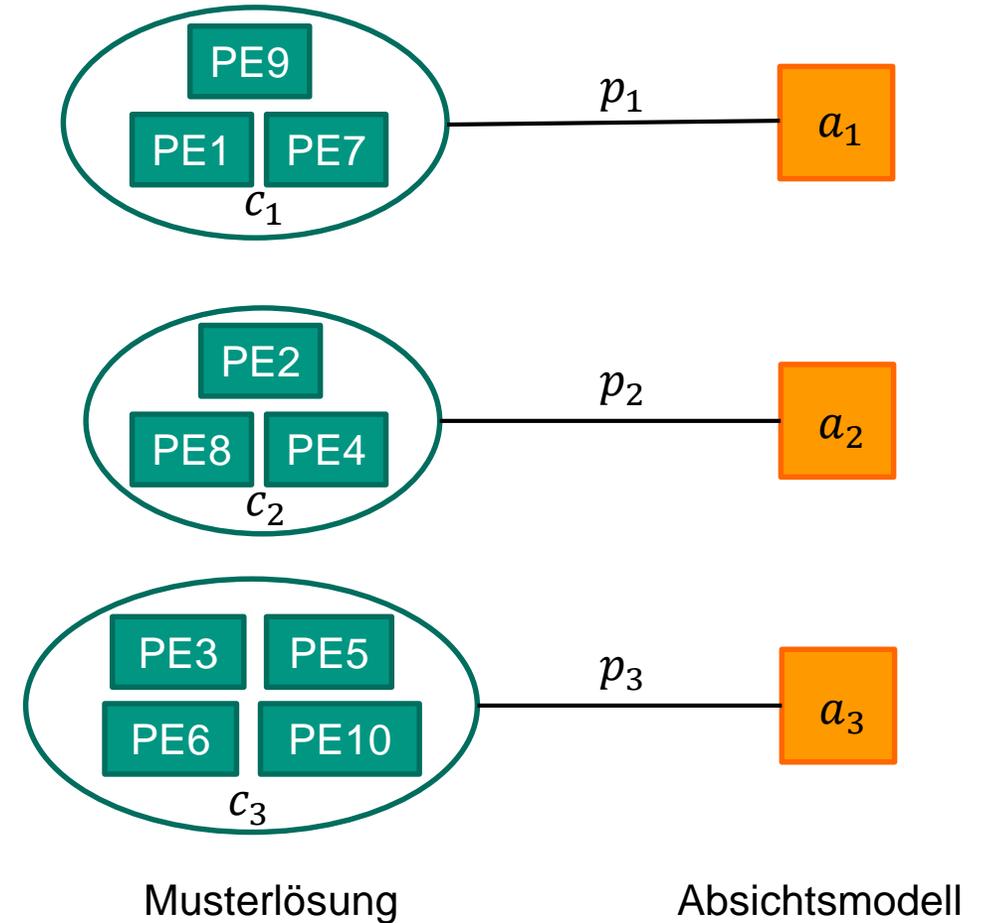
Evaluation

- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)
- Bewertung der Typ- und Methodenunterteilung
 - Keine offizielle Musterlösung
 - Vollständige Musterlösung nicht praktikabel
 - Systemausgabe in Musterlösung überführt
- Bewertung der Typunterteilung für iTrust
 - Offizielle Rückverfolgbarkeitsinformationen
 - Datensatz nicht vollständig



Evaluation

- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)
- Bewertung der Typ- und Methodenunterteilung
 - Keine offizielle Musterlösung
 - Vollständige Musterlösung nicht praktikabel
 - Systemausgabe in Musterlösung überführt
- Bewertung der Typunterteilung für iTrust
 - Offizielle Rückverfolgbarkeitsinformationen
 - Datensatz nicht vollständig



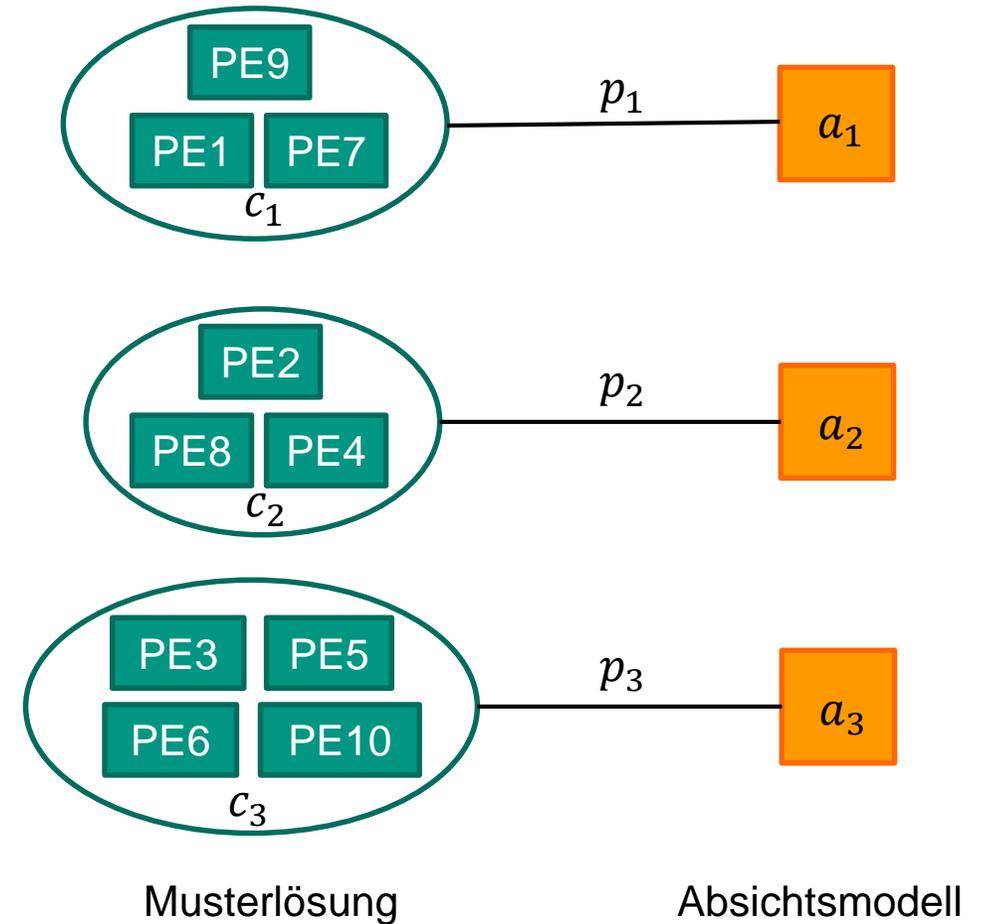
Evaluation

- iTrust (226 Typen, 1579 Methoden)
- Easymock (87 Typen, 753 Methoden)
- Bewertung der Typ- und Methodenunterteilung
 - Keine offizielle Musterlösung
 - Vollständige Musterlösung nicht praktikabel
 - Systemausgabe in Musterlösung überführt
- Bewertung der Typunterteilung für iTrust
 - Offizielle Rückverfolgbarkeitsinformationen
 - Datensatz nicht vollständig
- Metriken

- Cluster-Abdeckung: $CCov = \frac{\sum_{(c,a) \in P} |c \cap a|}{\sum_{(c,a) \in P} |c|}$

- Cluster-Präzision: $CPrec = \frac{\sum_{(c,a) \in P} |c \cap a|}{\sum_{(c,a) \in P} |a|}$

- Harmonisches Mittel (CHm)



Ergebnisse: Typen und Methoden

■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,44	0,66	0,53
<i>opt_{Dronology}</i>	(+0,11) 0,55	(+0,25) 0,91	(+0,16) 0,69
<i>opt_{iTrust}</i>	(+0,17) 0,61	(+0,25) 0,91	(+0,20) 0,73

Ergebnisse: Typen und Methoden

■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,44	0,66	0,53
<i>opt_{Dronology}</i>	(+0,11) 0,55	(+0,25) 0,91	(+0,16) 0,69
<i>opt_{iTrust}</i>	(+0,17) 0,61	(+0,25) 0,91	(+0,20) 0,73

- Schwellwerte unverändert
- Lexikalische Analyse “unverändert”
- Gewicht: Syntax=0,61 Sprache=0,39

Ergebnisse: Typen und Methoden

■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,44	0,66	0,53
<i>opt_{Dronology}</i>	(+0,11) 0,55	(+0,25) 0,91	(+0,16) 0,69
<i>opt_{iTrust}</i>	(+0,17) 0,61	(+0,25) 0,91	(+0,20) 0,73

- Schwellwerte unverändert
- Lexikalische Analyse “unverändert”
- Gewicht: Syntax=0,61 Sprache=0,39

■ Easymock

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,33	0,60	0,43
<i>opt_{Dronology}</i>	(+0,01) 0,34	(-0,06) 0,54	(-0,01) 0,42
<i>opt_{Easymock}</i>	(+0,18) 0,51	(-0,09) 0,51	(+0,08) 0,51

Ergebnisse: Typen und Methoden

■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,44	0,66	0,53
<i>opt_{Dronology}</i>	(+0,11) 0,55	(+0,25) 0,91	(+0,16) 0,69
<i>opt_{iTrust}</i>	(+0,17) 0,61	(+0,25) 0,91	(+0,20) 0,73

- Schwellwerte unverändert
- Lexikalische Analyse “unverändert”
- Gewicht: Syntax=0,61 Sprache=0,39

■ Easymock

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,33	0,60	0,43
<i>opt_{Dronology}</i>	(+0,01) 0,34	(-0,06) 0,54	(-0,01) 0,42
<i>opt_{Easymock}</i>	(+0,18) 0,51	(-0,09) 0,51	(+0,08) 0,51

- Schwellwert der Syntaxanalyse erhöht
- Lexikalische Analyse projektspezifisch
- Gewicht: Syntax=0,67 Sprache=0,33

Ergebnisse: Typebene

■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,23	0,27	0,25
<i>opt_{Dronology}</i>	(+0,00) 0,23	(+0,09) 0,36	(+0,03) 0,28
<i>type_opt_{iTrust}</i>	(+0,22) 0,45	(+0,20) 0,47	(+0,21) 0,46

Ergebnisse: Typebene

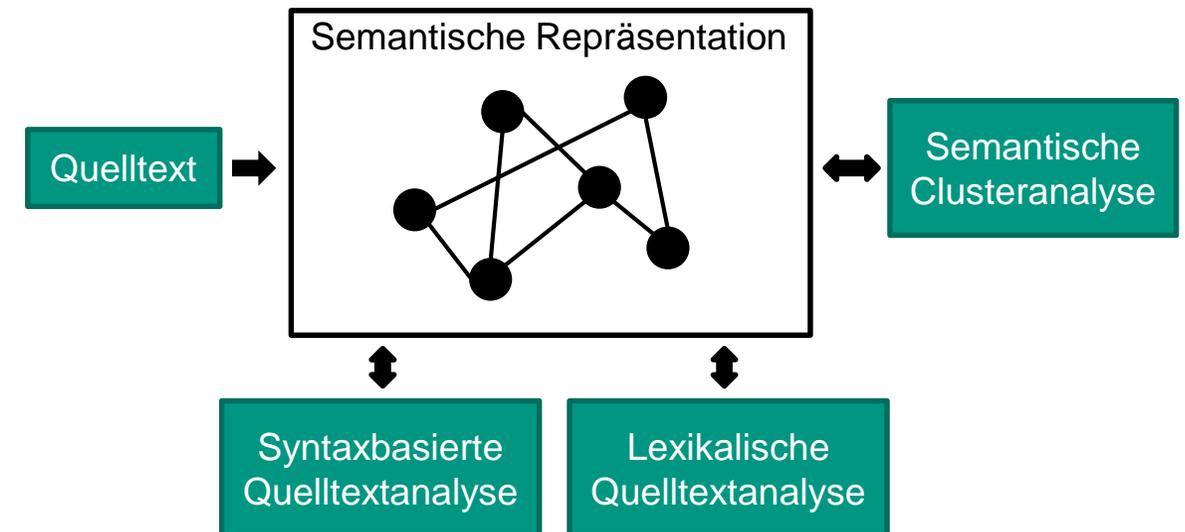
■ iTrust

Parameter	CPrec	CCov	CHm
<i>initial</i>	0,23	0,27	0,25
<i>opt_{Dronology}</i>	(+0,00) 0,23	(+0,09) 0,36	(+0,03) 0,28
<i>type_opt_{iTrust}</i>	(+0,22) 0,45	(+0,20) 0,47	(+0,21) 0,46

- Alle Schwellwerte erhöht
- Gewicht: Syntax=0,23 Sprache=0,77

Zusammenfassung

- Ziele
 - Graphrepräsentation von Quelltext
 - Semantische Zusammenhänge identifizieren
 - Einfache Erweiterbarkeit
- Umsetzung
 - Syntax und Sprachanalyse
 - Hierarchische Clusteranalyse
- Ergebnisse
 - Bis zu 0,73 (CHm) bei Typen und Methoden
 - Bis zu 0,46 (CHm) bei Typen
- Ausblick
 - Zusätzliche Quelltextanalysen
 - Erweiterte Abkürzungsauflösung
 - Erweiterte Clusteranalyse



Anhang

Abhängigkeitsmerkmale

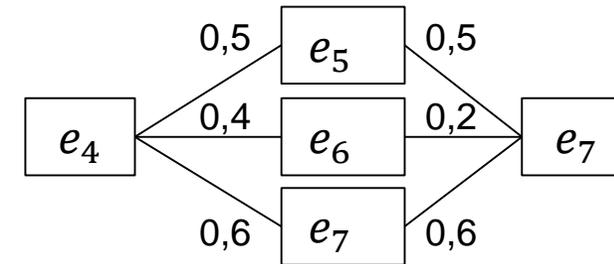
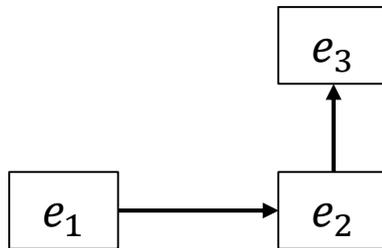
- Methodenaufruf
- Instanziierung
- Variablen- und Typzugriff
- Explizite Typreferenz
- Kontrollfluss
- Überschreiben
- Parametertyp
- Rückgabebetyp
- Ausnahmetyp
- Umschließender Typ
- Methodendefinition
- Innerer Typ
- Typhierarchie
- Attributtyp
- Paketstruktur

Abhängigkeitsmerkmale

- Methode → Methode
 - MM_controlFlow
 - MM_override
 - MM_methodCall
- Methode → Typ
 - MT_parameterType
 - MT_instantiationType
 - MT_controlFlow
 - MT_explicitTypeAccess
 - MT_throws
 - MT_variableOrTypeAccess
 - MT_methodParent
 - MT_returnType
- Typ → Methode
 - TM_methodDefinition
 - TM_controlFlow
 - TM_methodCall
- Typ → Typ
 - TT_parentType
 - TT_throws
 - TT_instantiationType
 - TT_controlFlow
 - TT_variableOrTypeAccess
 - TT_package
 - TT_parameterType
 - TT_typeHierarchy
 - TT_innerType
 - TT_explicitTypeAccess
 - TT_attributeType

Syntaxbasierte Quelltextanalyse

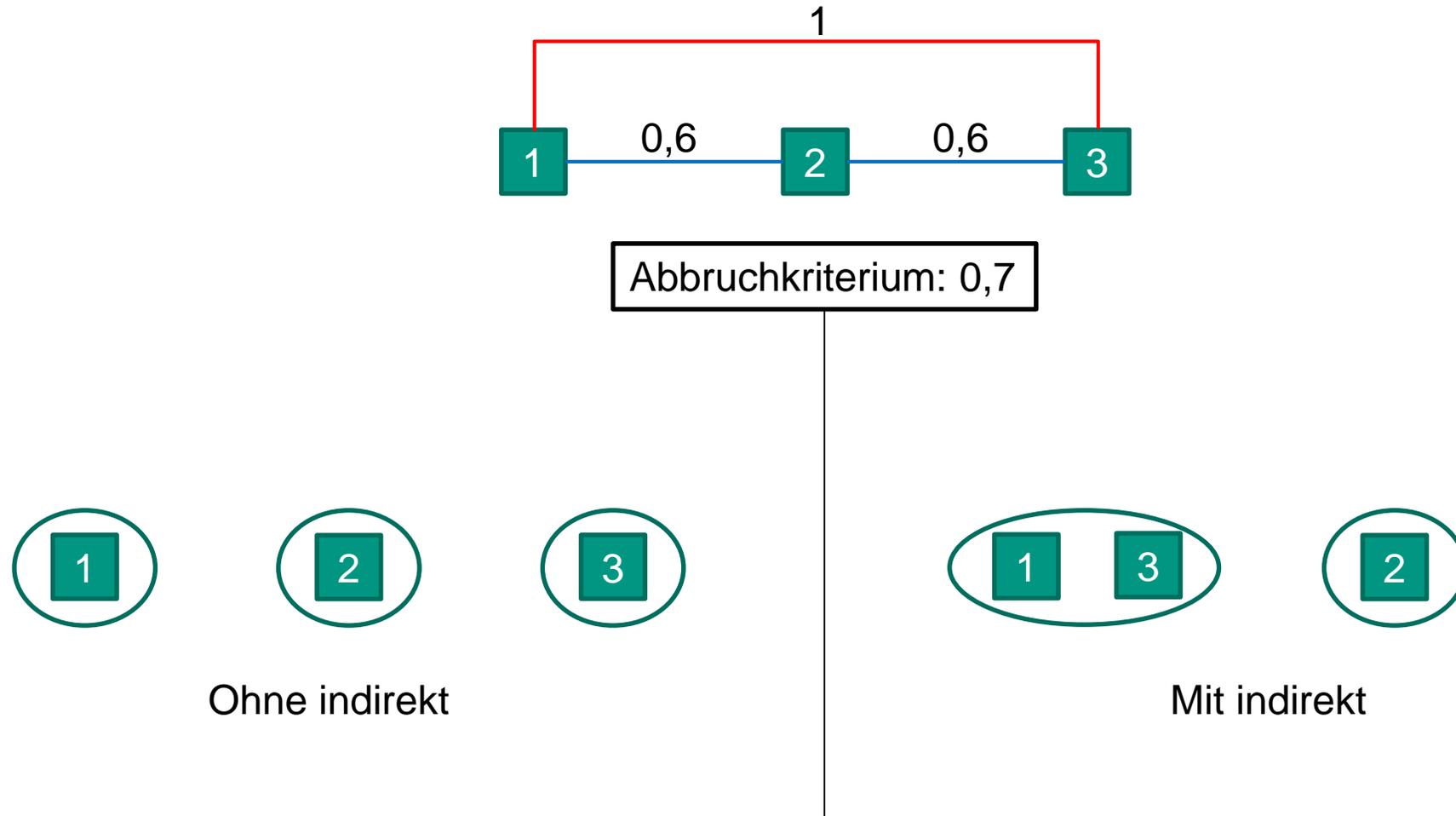
- $direct(e_1, e_2) = \sum_{r \in R(e_1, e_2)} w_r * \frac{2 * n_r(e_1, e_2)}{\max_{k \in E} n_r(e_1, k) + \max_{k \in E} n_r(k, e_2)}$
- $indirect(u, v) = \frac{\sum_{k=1}^n u_k * v_k}{\sqrt{\sum_{k=1}^n u_k^2} * \sqrt{\sum_{k=1}^n v_k^2}}$



$$\begin{aligned}
 dep_{dir}(e_1, e_2) &= 0,5 * direct(e_1, e_2) + direct(e_2, e_1) \\
 &= 0,5 * \left(\frac{2*1}{1+1} + \frac{2*0}{1+0} \right) = 0,5
 \end{aligned}$$

$$\begin{aligned}
 dep_{indir}(e_1, e_2) &= indirect(vec(e_1), vec(e_2)) \\
 &= cosineSim \left(\begin{pmatrix} 0,5 \\ 0,4 \\ 0,6 \end{pmatrix}, \begin{pmatrix} 0,5 \\ 0,2 \\ 0,6 \end{pmatrix} \right) = 0,975
 \end{aligned}$$

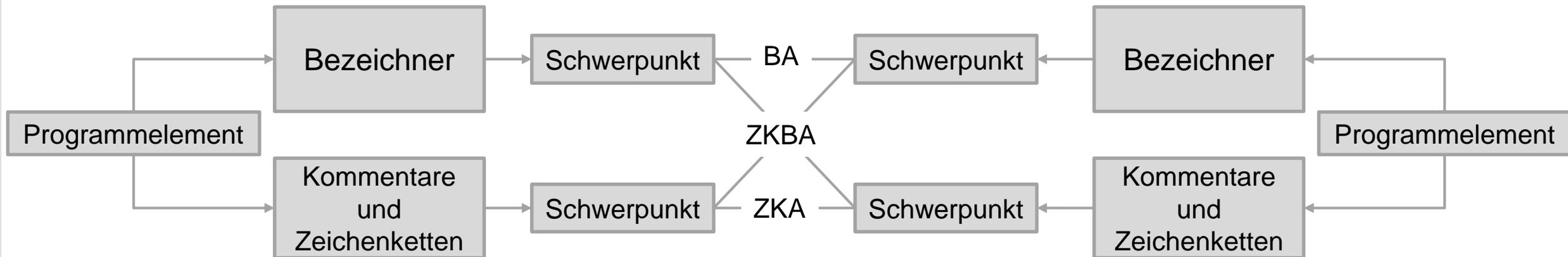
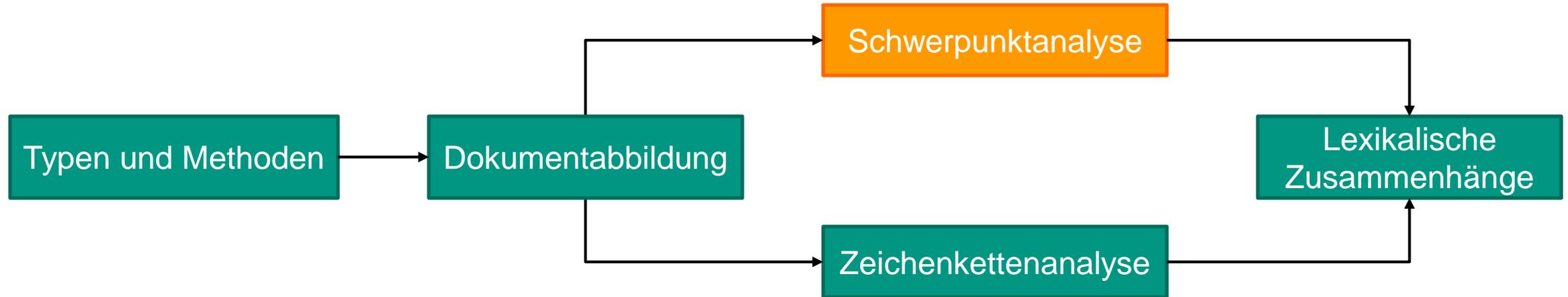
Einfluss indirekter Abhängigkeiten



Vorverarbeitung

- Kommentare filter
 - Todo/Fixme
 - Quelltext
 - Lizenzbeschreibungen
- Kommentare Bereinigen
 - Formatierung entfernen
 - JavaDoc markierungen entfernen (text beibehalten)
- Zerteilen an Leerstellen und Satzzeichen
- Auflösen zusammengesetzter Terme
- Abbildung auf Kleinbuchstaben
- Stoppwortentfernung
- Abkürzungsauflösung
 - Wörterbuch

Lexikalische Quelltextanalyse



Projekte und Musterlösungen

Projekt	Typen	Methoden	Quelltextzeilen	Kommentarzeilen
iTrust	226	1579	14.573	6.338
Easymock	87	753	4.952	4.998
Dronology	59	343		

Musterlösung	Projekt	Typen	Methoden	Cluster
Typen und Methoden	iTrust	172	104	40
	Easymock	46	211	43
	Dronology	28	61	15
Typen	iTrust	201	-	59
	Easymock	58	-	10

Ergebnisse: Typen und Methoden

Parameter	iTrust			Easymock		
	CPrec	CCov	CHm	CPrec	CCov	CHm
<i>init</i>	0,44	0,66	0,53	0,33	0,60	0,43
<i>opt_{dron}</i>	+0,11	+0,25	+0,16	+0,01	-0,06	-0,01
<i>opt_{iTrust}</i>	+0,17	+0,25	+0,20	+0,02	-0,14	-0,03
<i>opt_{Easymock}</i>	+0,16	+0,07	+0,13	+0,18	-0,09	+0,08

Ergebnisse: Einfluss verschiedener Komponenten

- Typen und Methoden
- iTrust

Konfiguration	CPrec	CCov	CHm
<i>opt_{iTrust}</i>	0,61	0,91	0,73
Opt. ohne Syntaxinformationen	-0,05	-0,45	-0,22
Opt. ohne lexikalische Informationen	±0,00	-0,12	-0,04
Opt. ohne indirekte Abhängigkeiten	-0,02	-0,45	-0,21
Opt. ohne direkte Abhängigkeiten	±0,00	±0,00	±0,00
Opt. ohne Zeichenkettenanalyse	-0,03	±0,00	-0,02

Ergebnisse: Typebene

Parameter	iTrust			Easymock		
	CPrec	CCov	CHm	CPrec	CCov	CHm
<i>init</i>	0,23	0,27	0,25	0,42	0,53	0,47
<i>opt_{dron}</i>	±0,00	+0,09	+0,03	+0,10	+0,26	+0,16
<i>type_opt_{iTrust}</i>	+0,22	+0,20	+0,21	+0,35	-0,13	+0,06
<i>type_opt_{easy}</i>	+0,09	+0,17	+0,12	+0,48	+0,07	+0,25

Ergebnisse: Worteinbettungsmodell

Modell	CPrec	CCov	CHm
Wikipedia	0,61	0,91	0,73
StackOverflow	-0,02	+0,01	-0,01
GitHub	-0,03	-0,07	-0,04

Übertragbarkeit Parameterbelegung

Parameter	Absichtsmodell	CPrec	CCov	CHm
<i>opt_{iTrust}</i>	Typen und Methoden	0,61	0,91	0,73
	Typen	0,19	0,29	0,23
<i>type_{-opt_{iTrust}}</i>	Typen und Methoden	0,45	0,41	0,43
	Typen	0,45	0,47	0,46

Parameterbelegung (Initial + Dronology)

Parameter	Initial	Optimiert
MM-Methodenaufruf	0,333	0,242
MM-Kontrollfluss	0,333	0,675
MM-Überschreiben	0,333	0,083
MT-Variablen- oder Typzugriff	0,125	0,078
MT-Explizite Typreferenz	0,125	0,072
MT-Instanziierung	0,125	0,235
MT-Ausnahme	0,125	0,039
MT-Parameter	0,125	0,244
MT-Umschließender Typ	0,125	0,042
MT-Rückgabe	0,125	0,006
MT-Kontrollfluss	0,125	0,284
TT-Typhierarchie	0,091	0,029
TT-Innerer Typ	0,091	0,180
TT-Attributtyp	0,091	0,130
TT-Umschließender Typ	0,091	0,005
TT-Paketstruktur	0,091	0,225
TT-Variablen- oder Typzugriff	0,091	0,026
TT-Explizite Typreferenz	0,091	0,201
TT-Instanziierung	0,091	0,004
TT-Ausnahme	0,091	0,148
TT-Parameter	0,091	0,041
TT-Kontrollfluss	0,091	0,011
TM-Methodendefinition	0,333	0,164
TM-Methodenaufruf	0,333	0,622
TM-Kontrollfluss	0,333	0,214
Schwellwert α_{sda}	0,0	0,066

Parameter	Initial	Optimiert
BA-Schwerpunkt	0,9	0,704
BA-Zeichenketten	0,1	0,296
BA-Signaturebene	0,6	0,594
BA-Anweisungsebene	0,4	0,406
ZKA-Schwerpunkt	0,9	0,081
ZKA-Zeichenketten	0,1	0,919
ZKA-Signaturebene	0,6	0,770
ZKA-Anweisungsebene	0,4	0,230
ZKBA-Schwerpunkt	0,9	0,211
ZKBA-Zeichenketten	0,1	0,789
ZKBA-Signaturebene	0,6	0,446
ZKBA-Anweisungsebene	0,4	0,554
BA-Gewicht	0,333	0,294
ZKA-Gewicht	0,333	0,260
ZKBA-Gewicht	0,333	0,446
Schwellwert α_{1a}	0,0	0,333

Parameter	Initial	Optimiert
LA-Gewicht	0,6	0,389
SDA-Gewicht	0,4	0,611
Schwellwert $\alpha_{semantic}$	0,0	0,384
Schwellwert $\alpha_{cluster}$	0,6	0,604

Parameterbelegung TM

Agent	Parameter	Wert	Agent	Parameter	Wert		
	MM_controlFlow	0,003		MM_controlFlow	0,761		
	MM_override	0,551		MM_override	0,027		
	MM_methodCall	0,446		MM_methodCall	0,212		
	MT_parameterType	0,217		MT_parameterType	0,201		
	MT_instantiationType	0,185		MT_instantiationType	0,224		
	MT_controlFlow	0,105		MT_controlFlow	0,303		
	MT_explicitTypeAccess	0,017		MT_explicitTypeAccess	0,077		
	MT_throws	0,06		MT_throws	0,052		
	MT_variableOrTypeAccess	0,018		MT_variableOrTypeAccess	0,056		
	MT_methodParent	0,184		MT_methodParent	0,046		
	MT_returnType	0,214		MT_returnType	0,04		
	SDA	TM_methodDefinition		0,184	SDA	TM_methodDefinition	0,024
		TM_controlFlow		0,118		TM_controlFlow	0,487
		TM_methodCall		0,698		TM_methodCall	0,489
		TT_parentType		0,003		TT_parentType	0,064
		TT_throws		0,091		TT_throws	0,124
		TT_instantiationType		0,002		TT_instantiationType	0,113
		TT_controlFlow		0,007		TT_controlFlow	0,09
		TT_variableOrTypeAccess		0,402		TT_variableOrTypeAccess	0,134
TT_package		0,138	TT_package	0,117			
TT_parameterType		0,025	TT_parameterType	0,035			
TT_typeHierarchy		0,018	TT_typeHierarchy	0,069			
TT_innerType		0,111	TT_innerType	0,114			
TT_explicitTypeAccess		0,123	TT_explicitTypeAccess	0,041			
TT_attributeType		0,08	TT_attributeType	0,1			
SDA_THRESHOLD		0,066	SDA_THRESHOLD	0,528			
LA		ANALYZER_COMMENT_WEIGHT	0,26	LA		ANALYZER_COMMENT_WEIGHT	0,388
		ANALYZER_IDENTIFIER_WEIGHT	0,294			ANALYZER_IDENTIFIER_WEIGHT	0,602
		ANALYZER_COMMENT_IDENTIFIER_WEIGHT	0,446			ANALYZER_COMMENT_IDENTIFIER_WEIGHT	0,01
		IDENTIFIER_SIGNATURE_WEIGHT	0,594			IDENTIFIER_SIGNATURE_WEIGHT	0,206
	IDENTIFIER_STATEMENT_WEIGHT	0,406	IDENTIFIER_STATEMENT_WEIGHT		0,794		
	IDENTIFIER_CENTROID_WEIGHT	0,704	IDENTIFIER_CENTROID_WEIGHT		0,601		
	IDENTIFIER_STRING_WEIGHT	0,296	IDENTIFIER_STRING_WEIGHT		0,399		
	COMMENT_SIGNATURE_WEIGHT	0,77	COMMENT_SIGNATURE_WEIGHT		0,663		
	COMMENT_STATEMENT_WEIGHT	0,23	COMMENT_STATEMENT_WEIGHT		0,334		
	COMMENT_CENTROID_WEIGHT	0,081	COMMENT_CENTROID_WEIGHT		0,479		
	COMMENT_STRING_WEIGHT	0,919	COMMENT_STRING_WEIGHT		0,521		
	CROSS_SIGNATURE_WEIGHT	0,446	CROSS_SIGNATURE_WEIGHT		0,504		
	CROSS_STATEMENT_WEIGHT	0,554	CROSS_STATEMENT_WEIGHT		0,496		
	CROSS_CENTROID_WEIGHT	0,211	CROSS_CENTROID_WEIGHT		0,458		
	CROSS_STRING_WEIGHT	0,789	CROSS_STRING_WEIGHT		0,542		
	LA_THRESHOLD	0,333	LA_THRESHOLD		0,313		
	Clusteranalyse	SDA_WEIGHT	0,611		Clusteranalyse	SDA_WEIGHT	0,669
		LA_WEIGHT	0,389			LA_WEIGHT	0,331
		SEMANTIC_RELATION_THRESHOLD	0,384			SEMANTIC_RELATION_THRESHOLD	0,38
CLUSTER_THRESHOLD		0,604	CLUSTER_THRESHOLD	0,64			

Parameterbelegung T

Agent	Parameter	Wert	Agent	Parameter	Wert
SDA	MM_controlFlow	0,022	SDA	MM_controlFlow	0,474
	MM_override	0,336		MM_override	0,394
	MM_methodCall	0,642		MM_methodCall	0,132
	MT_parameterType	0,122		MT_parameterType	0,017
	MT_instantiationType	0,179		MT_instantiationType	0,295
	MT_controlFlow	0,165		MT_controlFlow	0,147
	MT_explicitTypeAccess	0,093		MT_explicitTypeAccess	0,047
	MT_throws	0,173		MT_throws	0,032
	MT_variableOrTypeAccess	0,082		MT_variableOrTypeAccess	0,034
	MT_methodParent	0,097		MT_methodParent	0,028
	MT_returnType	0,089		MT_returnType	0,401
	TM_methodDefinition	0,376		TM_methodDefinition	0,02
	TM_controlFlow	0,509		TM_controlFlow	0,415
	TM_methodCall	0,115		TM_methodCall	0,564
	TT_parentType	0,214		TT_parentType	0,187
	TT_throws	0,011		TT_throws	0,034
	TT_instantiationType	0,066		TT_instantiationType	0,209
	TT_controlFlow	0,071		TT_controlFlow	0,261
	TT_variableOrTypeAccess	0,161		TT_variableOrTypeAccess	0,037
TT_package	0,023	TT_package	0,032		
TT_parameterType	0,034	TT_parameterType	0,01		
TT_typeHierarchy	0,1	TT_typeHierarchy	0,019		
TT_innerType	0,034	TT_innerType	0,031		
TT_explicitTypeAccess	0,059	TT_explicitTypeAccess	0,152		
TT_attributeType	0,227	TT_attributeType	0,027		
SDA_THRESHOLD	0,503	SDA_THRESHOLD	0,528		
LA	ANALYZER_COMMENT_WEIGHT	0,529	LA	ANALYZER_COMMENT_WEIGHT	0,56
	ANALYZER_IDENTIFIER_WEIGHT	0,151		ANALYZER_IDENTIFIER_WEIGHT	0,262
	ANALYZER_COMMENT_IDENTIFIER_WEIGHT	0,32		ANALYZER_COMMENT_IDENTIFIER_WEIGHT	0,179
	IDENTIFIER_SIGNATURE_WEIGHT	0,768		IDENTIFIER_SIGNATURE_WEIGHT	0,212
	IDENTIFIER_STATEMENT_WEIGHT	0,232		IDENTIFIER_STATEMENT_WEIGHT	0,788
	IDENTIFIER_CENTROID_WEIGHT	0,457		IDENTIFIER_CENTROID_WEIGHT	0,194
	IDENTIFIER_STRING_WEIGHT	0,543		IDENTIFIER_STRING_WEIGHT	0,806
	COMMENT_SIGNATURE_WEIGHT	0,315		COMMENT_SIGNATURE_WEIGHT	0,408
	COMMENT_STATEMENT_WEIGHT	0,685		COMMENT_STATEMENT_WEIGHT	0,592
	COMMENT_CENTROID_WEIGHT	0,701		COMMENT_CENTROID_WEIGHT	0,469
	COMMENT_STRING_WEIGHT	0,299		COMMENT_STRING_WEIGHT	0,531
	CROSS_SIGNATURE_WEIGHT	0,523		CROSS_SIGNATURE_WEIGHT	0,805
	CROSS_STATEMENT_WEIGHT	0,477		CROSS_STATEMENT_WEIGHT	0,195
	CROSS_CENTROID_WEIGHT	0,666		CROSS_CENTROID_WEIGHT	0,051
CROSS_STRING_WEIGHT	0,334	CROSS_STRING_WEIGHT	0,949		
LA_THRESHOLD	0,878	LA_THRESHOLD	0,003		
Clusteranalyse	SDA_WEIGHT	0,232	Clusteranalyse	SDA_WEIGHT	0,669
	LA_WEIGHT	0,768		LA_WEIGHT	0,331
	SEMANTIC_RELATION_THRESHOLD	0,635		SEMANTIC_RELATION_THRESHOLD	0,38
	CLUSTER_THRESHOLD	0,639		CLUSTER_THRESHOLD	0,64

Literatur

- [AC17] “Improving modular structure of software system using structural and lexical dependency”, Information and Software Technology, 2017.
- [CDMMS16] “Weighing lexical information for software clustering in the context of architecture recovery”, Empirical Software Engineering 21, 2016.
- [GAL18] “Automatic Hierarchical Clustering of Static Call Graphs for Program Comprehension”, IEEE International Conference on Big Data, 2018.
- [KDG07] “Semantic clustering: Identifying topics in source code” Information and Software Technology 49, 2007.
- [KNH17] “Analyzing closeness of code dependencies for improving IR-based Traceability Recovery”, IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER), 2017.
- [LOV08] “Using structural and semantic metrics to improve class cohesion”, 2008 IEEE International Conference on Software Maintenance, 2008.
- [MAK+12] “Software Clustering: Unifying Syntactic and Semantic Features”, 19th Working Conference on Reverse Engineering, 2012.
- [RM08] “Topology analysis of software dependencies”, ACM Transactions on Software Engineering and Methodology (TOSEM) 17, 2008.
- [SM11] “Clustering Support for Static Concept Location in Source Code”, 2011 IEEE 19th International Conference on Program Comprehension, 2011.