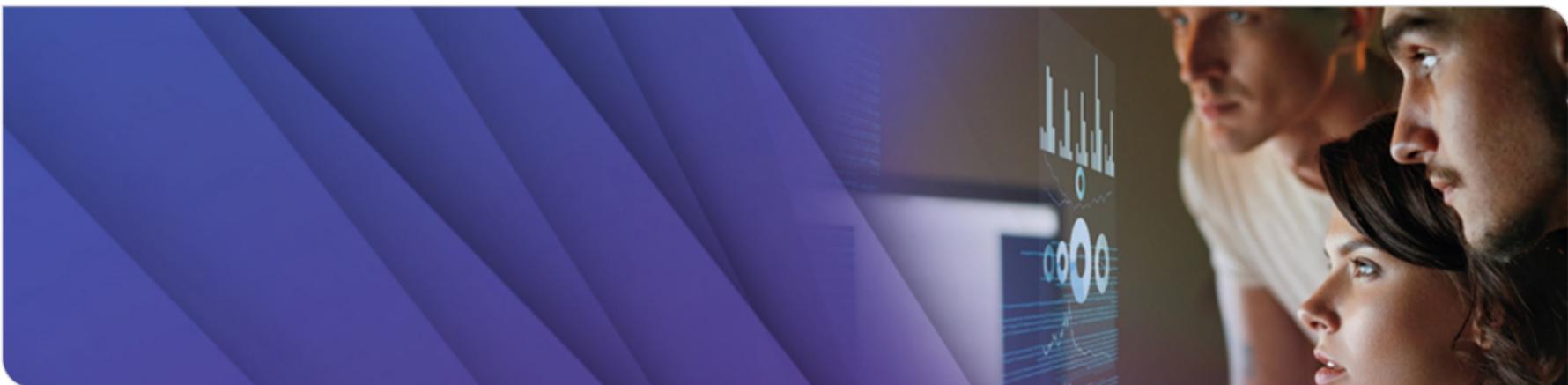


Automatisches Auflösen von Abkürzungen in Quelltext

Masterarbeit, betreut von Tobias Hey

Gilbert Groten | 7. Mai 2021



Motivation

- Problem: Abkürzungen

```
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Motivation

- Problem: Abkürzungen
- Auflösung durch Menschen

```
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Motivation

- Problem: Abkürzungen
- Auflösung durch Menschen

```
//how much money someone has  
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Motivation

- Problem: Abkürzungen
- Auflösung durch Menschen
- Auflösung durch Maschine

```
//how much money someone has  
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Motivation

- Problem: Abkürzungen
- Auflösung durch Menschen
- Auflösung durch Maschine

```
//how much money someone has  
static float cstmrbInc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Motivation

- Problem: Abkürzungen
- Auflösung durch Menschen
- Auflösung durch Maschine
- Beispiel: Rückverfolgung von Anforderungen

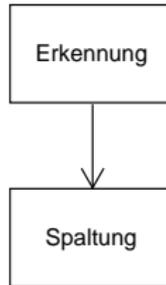
```
//how much money someone has  
static float cstmrbInc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Typisches Vorgehen

Erkennung

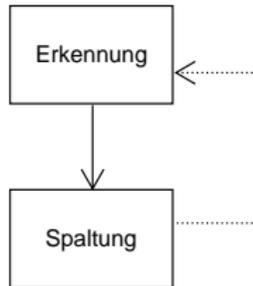
```
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Typisches Vorgehen



```
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

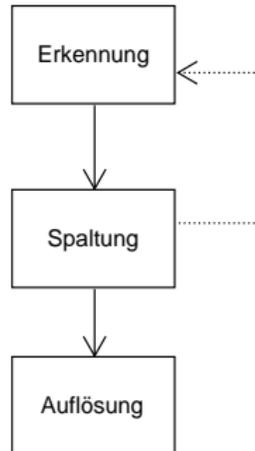
Typisches Vorgehen



```

static float cstmrblnc(String nm){
    int id = getCstmrID(nm);
    return accounts.get(id).blnc;
}
  
```

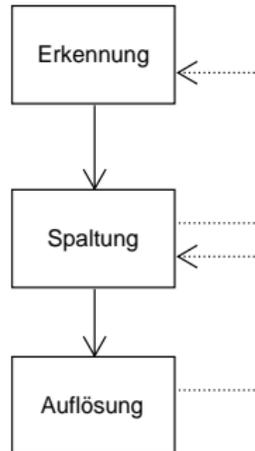
Typisches Vorgehen



```

static float cstmrblnc(String nm){
    int id = getCstmrID(nm);
    return accounts.get(id).blnc;
}
  
```

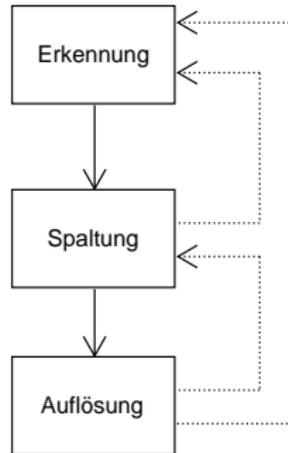
Typisches Vorgehen



```

static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Typisches Vorgehen



```

static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Verwandte Arbeiten



- Newman et al. [11]

Verwandte Arbeiten



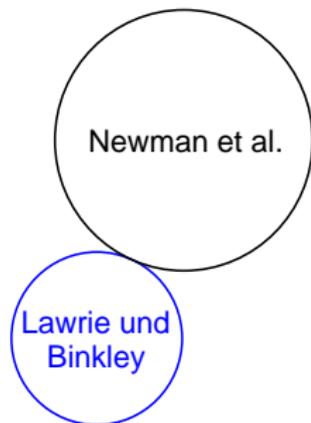
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung

Verwandte Arbeiten



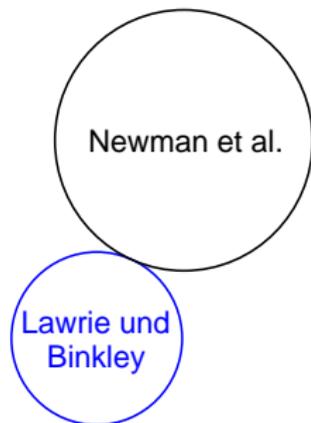
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen

Verwandte Arbeiten



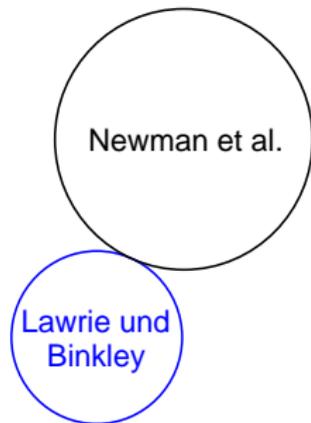
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]

Verwandte Arbeiten



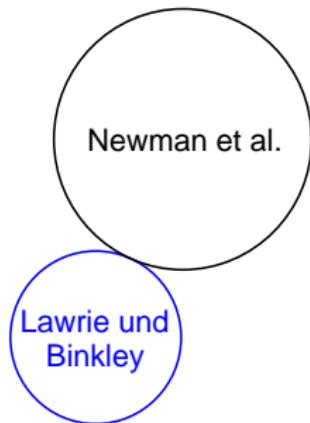
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken

Verwandte Arbeiten



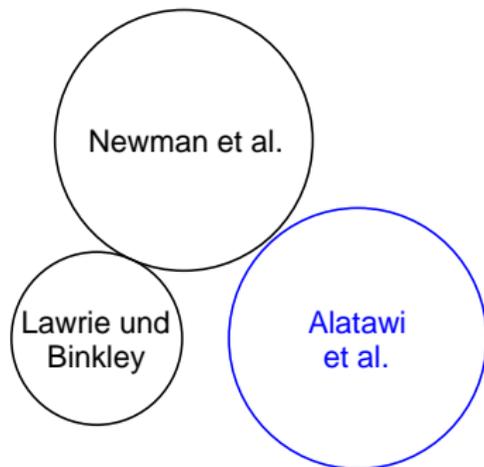
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung

Verwandte Arbeiten



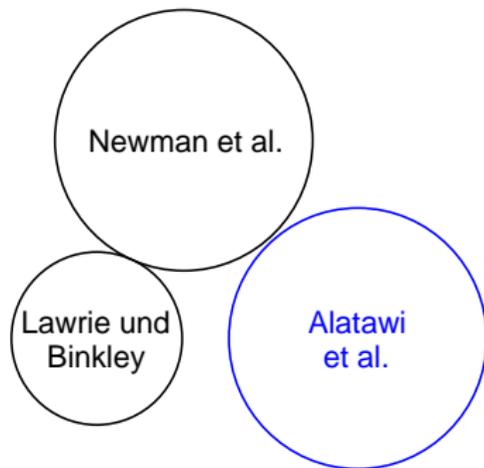
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*

Verwandte Arbeiten



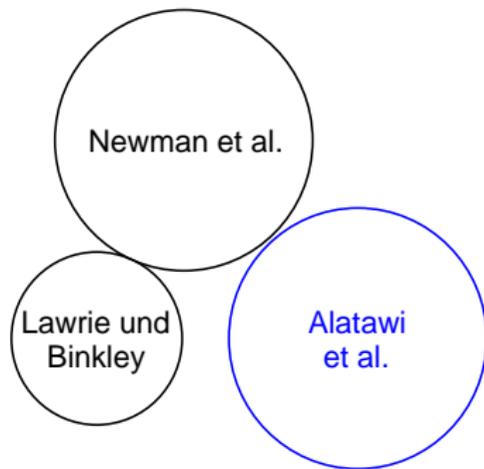
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]

Verwandte Arbeiten



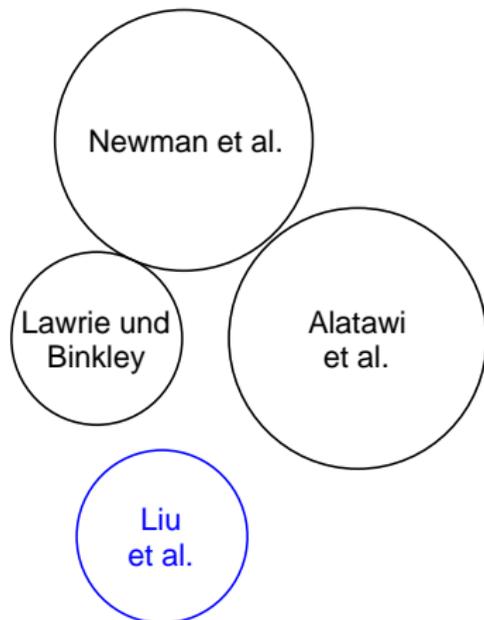
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]
 - Auflösung mittels statistischen Eigenschaften von Abkürzungen

Verwandte Arbeiten



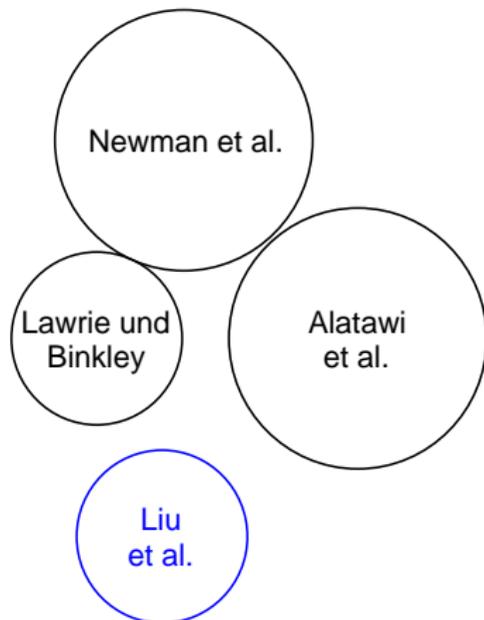
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]
 - Auflösung mittels statistischen Eigenschaften von Abkürzungen
 - Uni- bzw. Bigramm-basierte Auftretenswahrscheinlichkeit

Verwandte Arbeiten



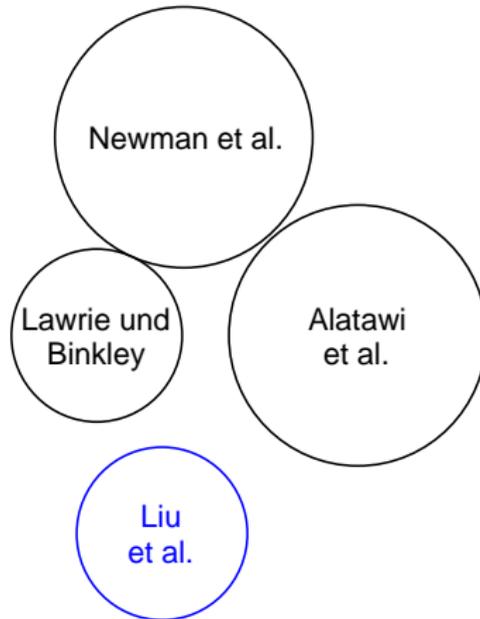
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]
 - Auflösung mittels statistischen Eigenschaften von Abkürzungen
 - Uni- bzw. Bigramm-basierte Auftretenswahrscheinlichkeit
- Liu et al. [10]

Verwandte Arbeiten



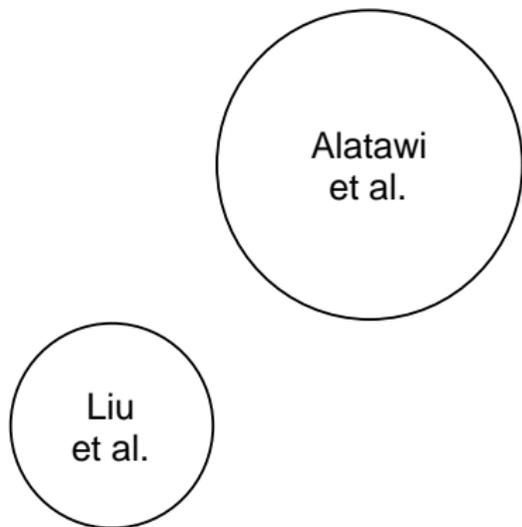
- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]
 - Auflösung mittels statistischen Eigenschaften von Abkürzungen
 - Uni- bzw. Bigramm-basierte Auftretenswahrscheinlichkeit
- Liu et al. [10]
 - Abkürzungen in medizinischen Notizen

Verwandte Arbeiten



- Newman et al. [11]
 - Metastudie über Abkürzungsauflösung
 - Eigenschaften und Statistiken zu Abkürzungen
- Lawrie und Binkley [9]
 - Spaltung mittels Wort-statistiken
 - Stellvertretersymbol-Auflösung
 - Musterlösungen für *a2ps* und *which*
- Alatawi et al. [1] [2]
 - Auflösung mittels statistischen Eigenschaften von Abkürzungen
 - Uni- bzw. Bigramm-basierte Auftretenswahrscheinlichkeit
- Liu et al. [10]
 - Abkürzungen in medizinischen Notizen
 - Worteinbettungen

Ansätze



Motivation

oo

Verwandte Arbeiten

o

Ansätze

●oooooooo

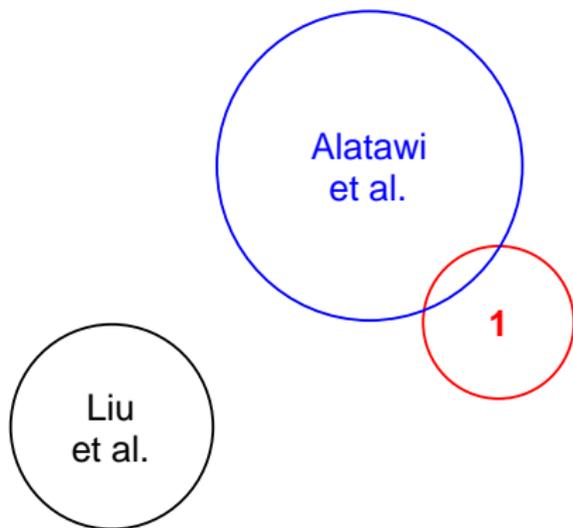
Evaluation

oooooooooooooooooooo

Zusammenfassung und Ausblick

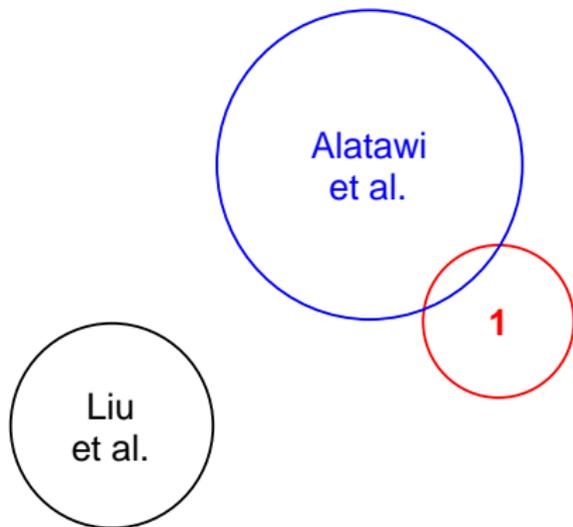
ooo

Ansätze



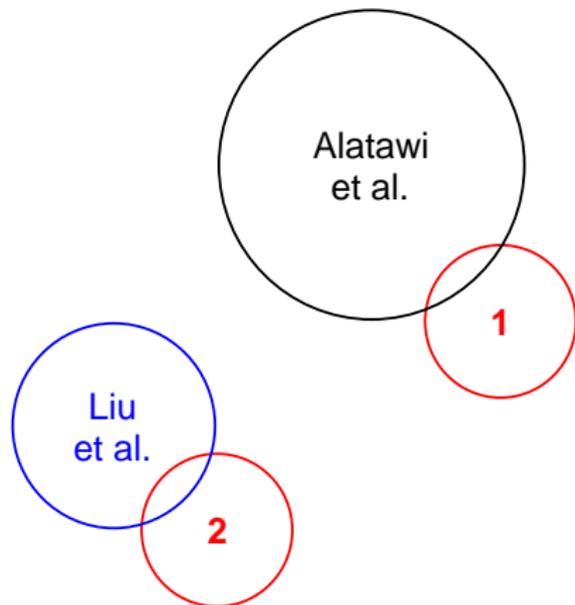
- Trigramm-basierte Abkürzungsauflösung

Ansätze



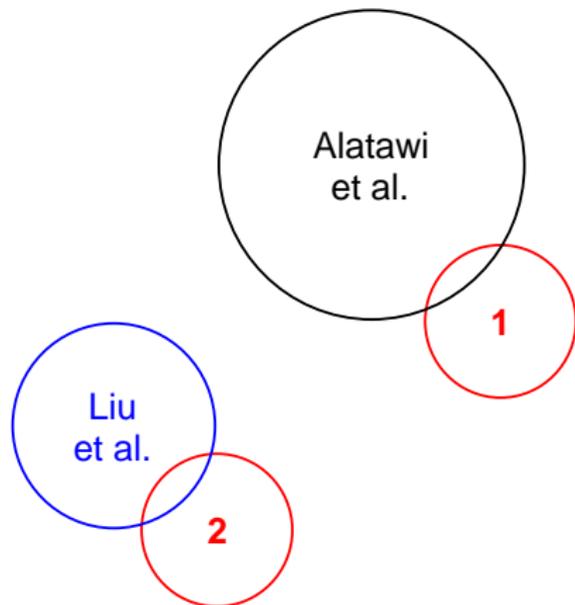
- Trigramm-basierte Abkürzungsauflösung

Ansätze



- 1 Trigramm-basierte Abkürzungsauflösung
- 2 Worteinbettungsbasierte Abkürzungsauflösung

Ansätze



- ① Trigramm-basierte Abkürzungsauflösung
- ② Worteinbettungsbasierte Abkürzungsauflösung

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○○

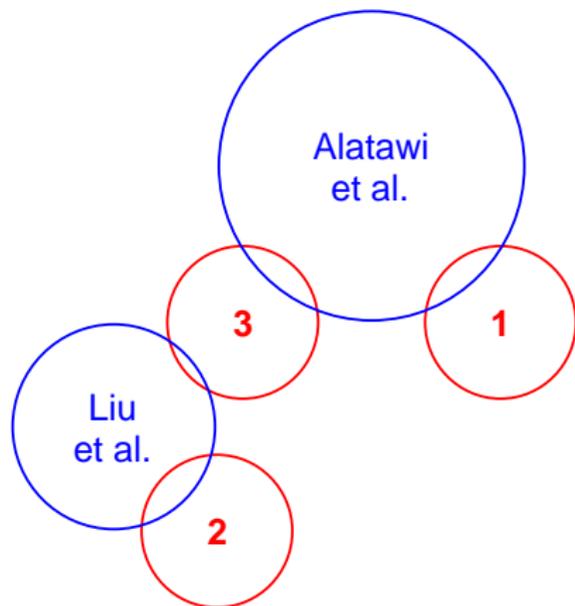
Evaluation

○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Ansätze



- ① Trigramm-basierte Abkürzungsauflösung
- ② Worteinbettungsbasierte Abkürzungsauflösung
- ③ Worteinbettungsbasierte Mehrdeutigkeitsauflösung

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○○

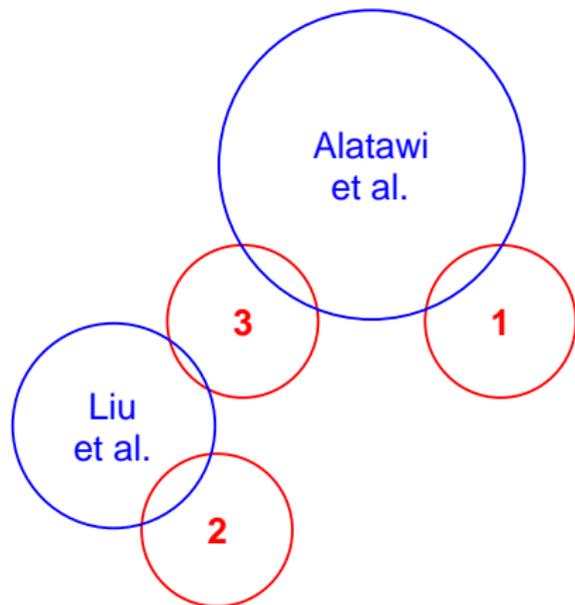
Evaluation

○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Ansätze



- ① Trigramm-basierte Abkürzungsauflösung
- ② Worteinbettungsbasierte Abkürzungsauflösung
- ③ Worteinbettungsbasierte Mehrdeutigkeitsauflösung

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○○

Evaluation

○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.

Abgekürzter
Bezeichner

```
private final String constant = "constant";  
private String blanc_whiteSpace = " "; //used for margins  
  
//how much money someone has  
static float cstmrblnc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○

Evaluation

○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.

Abgekürzter
Bezeichner

"cstmrbnc"

```
private final String constant = "constant";  
private String blanc_whiteSpace = " "; //used for margins  
  
//how much money someone has  
static float cstmrbnc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○

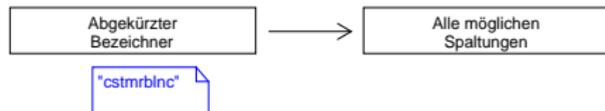
Evaluation

○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.



```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnlnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○

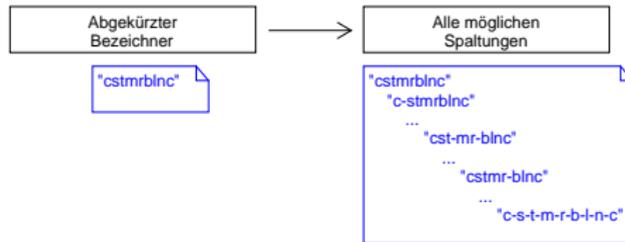
Evaluation

○○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.



```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○

Evaluation

○○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.



```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation
○○

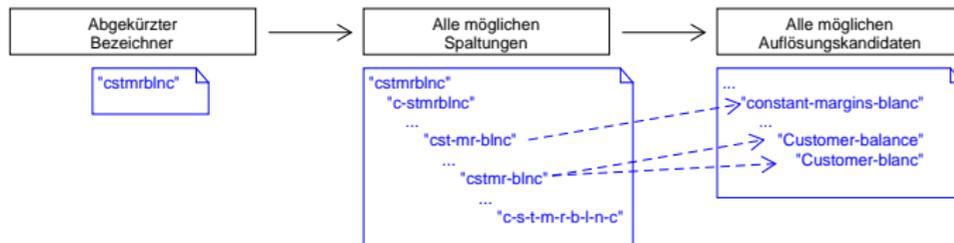
Verwandte Arbeiten
○

Ansätze
●○○○○○○○

Evaluation
○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick
○○○

Verfahren von Alatawi et al.



```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation
○○

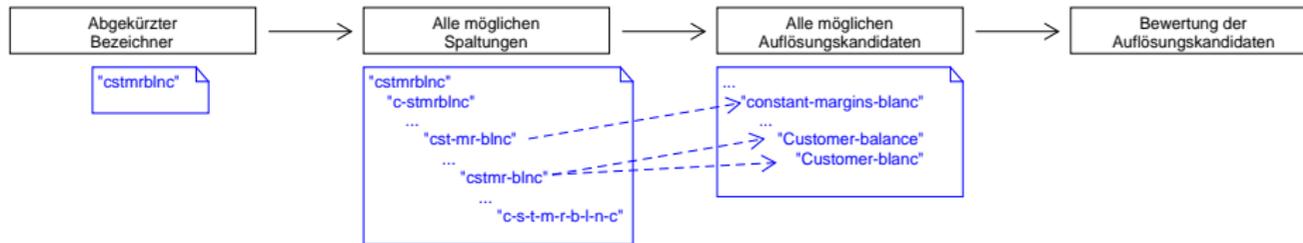
Verwandte Arbeiten
○

Ansätze
●○○○○○○○

Evaluation
○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick
○○○

Verfahren von Alatawi et al.



```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation

○○

Verwandte Arbeiten

○

Ansätze

●○○○○○○○

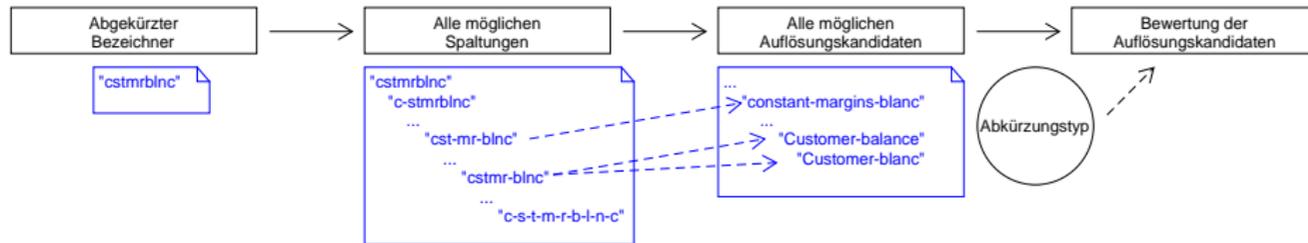
Evaluation

○○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Verfahren von Alatawi et al.

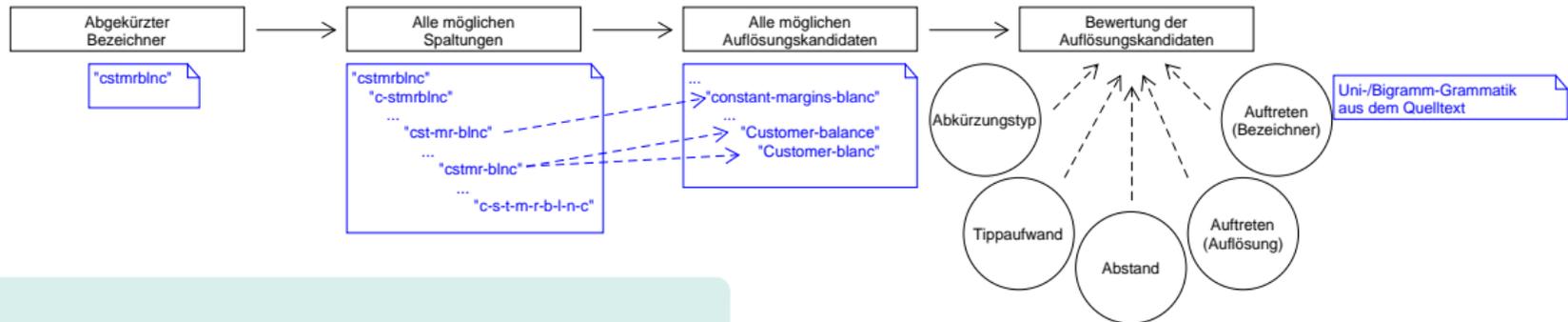


```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbinc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```


Verfahren von Alatawi et al.



```
private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
```

Motivation
○○

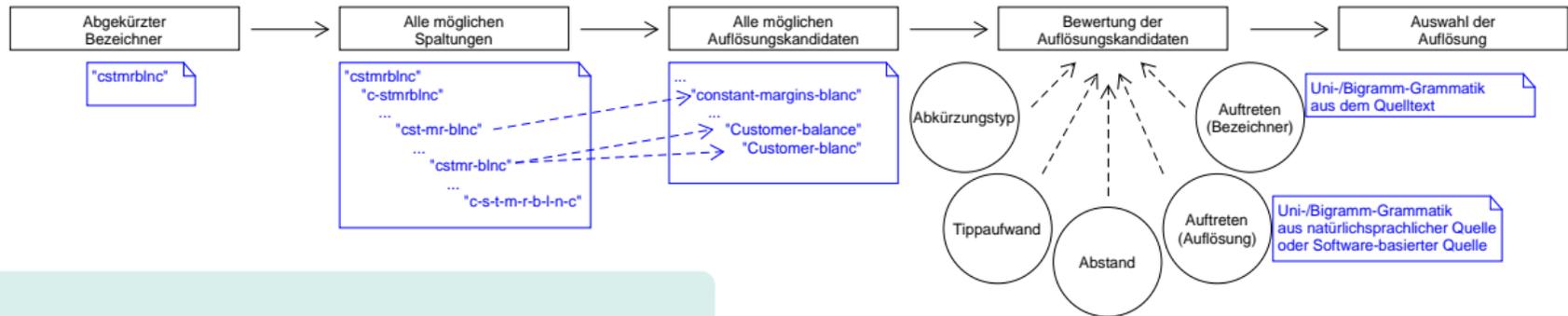
Verwandte Arbeiten
○

Ansätze
●○○○○○○○

Evaluation
○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick
○○○

Verfahren von Alatawi et al.

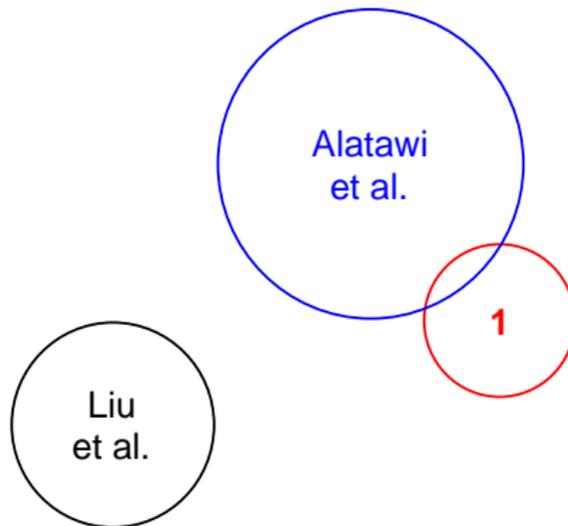


```

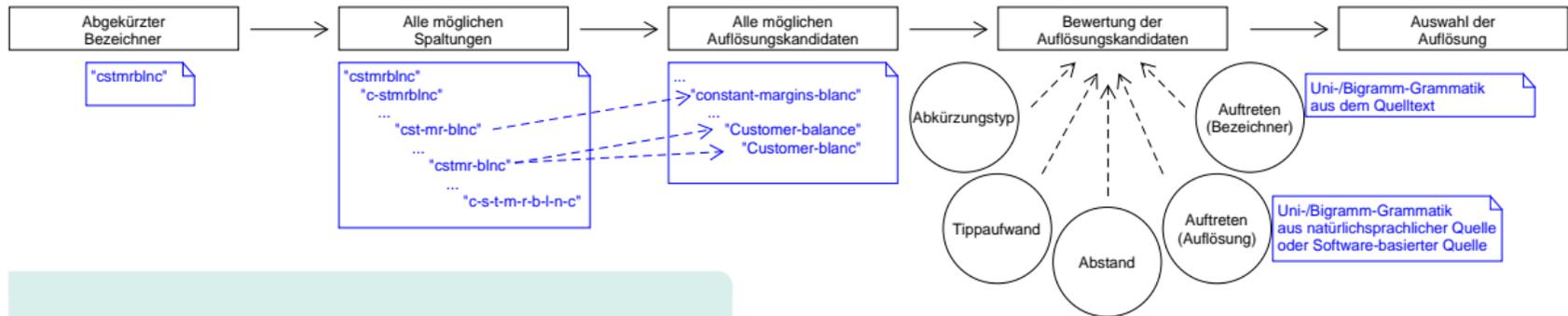
private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Ansatz 1/3: Trigramm-basiertes Verfahren



Ansatz 1/3: Trigramm-basiertes Verfahren



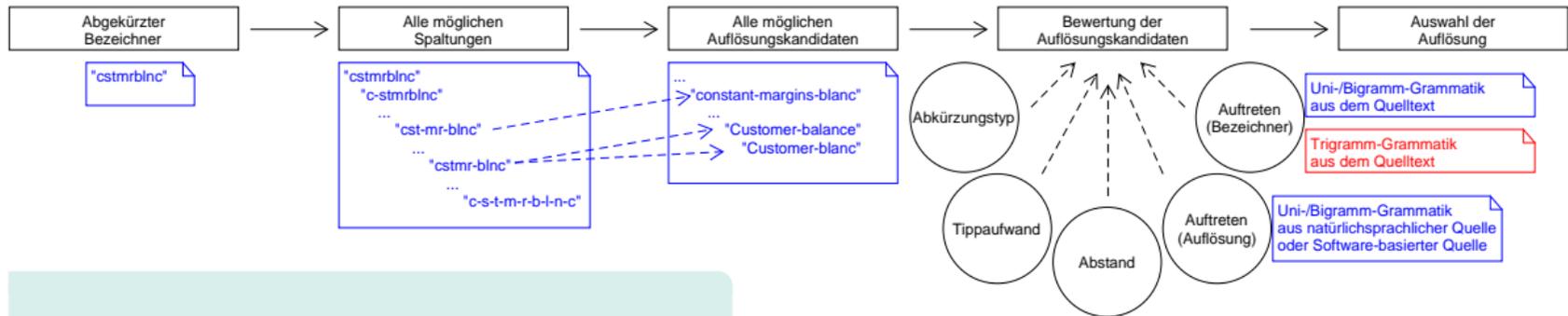
```

private final String constant = "constant";
private String blanc_whiteSpace = " "; // used for margins

// how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Motivation	Verwandte Arbeiten	Ansätze	Evaluation	Zusammenfassung und Ausblick
○○	○	○○●○○○○	○○○○○○○○○○○○○○○○○○	○○○

Ansatz 1/3: Trigramm-basiertes Verfahren

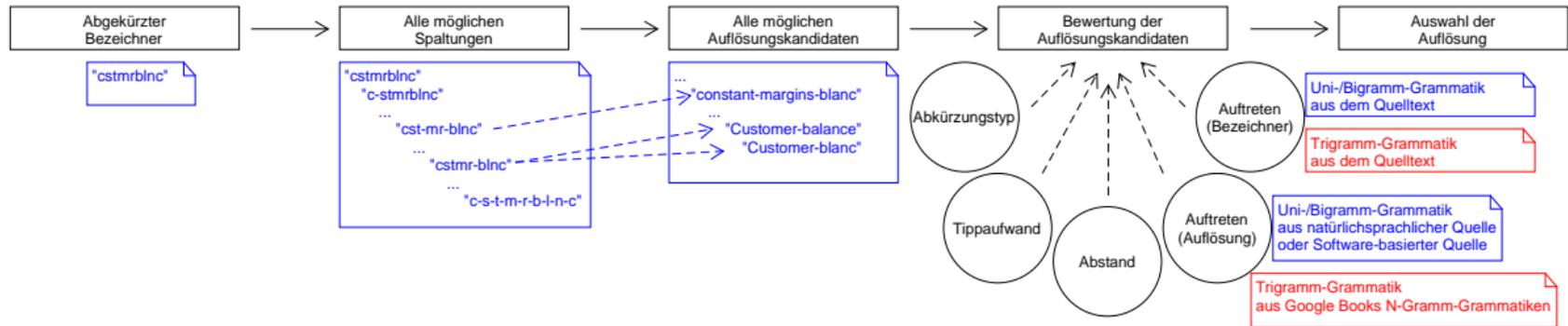


```

private final String constant = "constant";
private String blanc_whiteSpace = " "; // used for margins

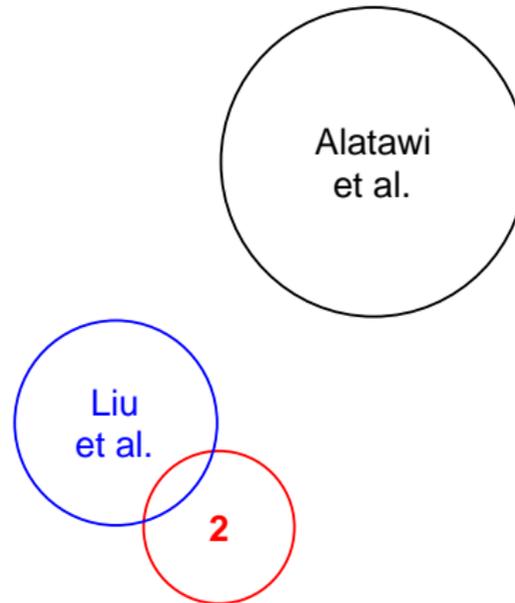
// how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Ansatz 1/3: Trigramm-basiertes Verfahren



Google Books N-Gramm-Korpus [8]

Ansatz 2/3: Warteinbettungsbasierte Auflösung



Motivation
○○

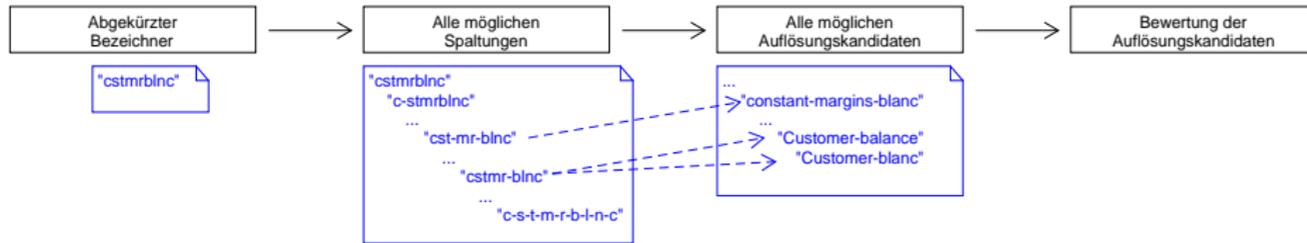
Verwandte Arbeiten
○

Ansätze
○○○●○○○

Evaluation
○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick
○○○

Ansatz 2/3: Worteinbettungsbasierte Abkürzungsauflösung

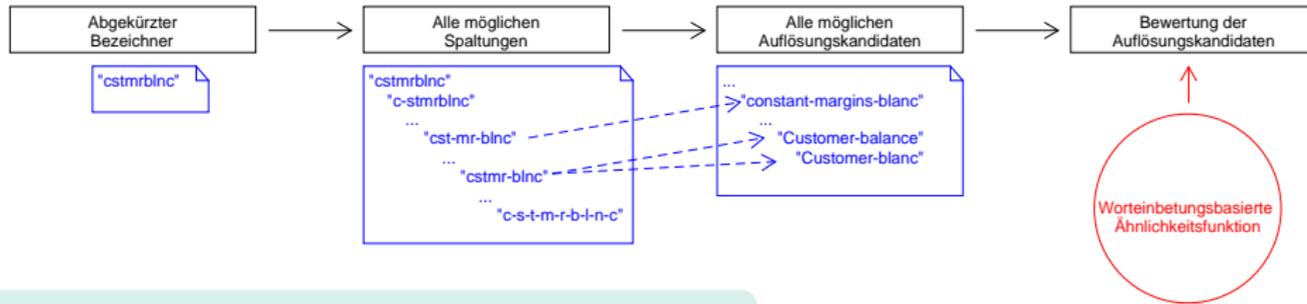


```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Ansatz 2/3: Worteinbettungsbasierte Abkürzungsauflösung

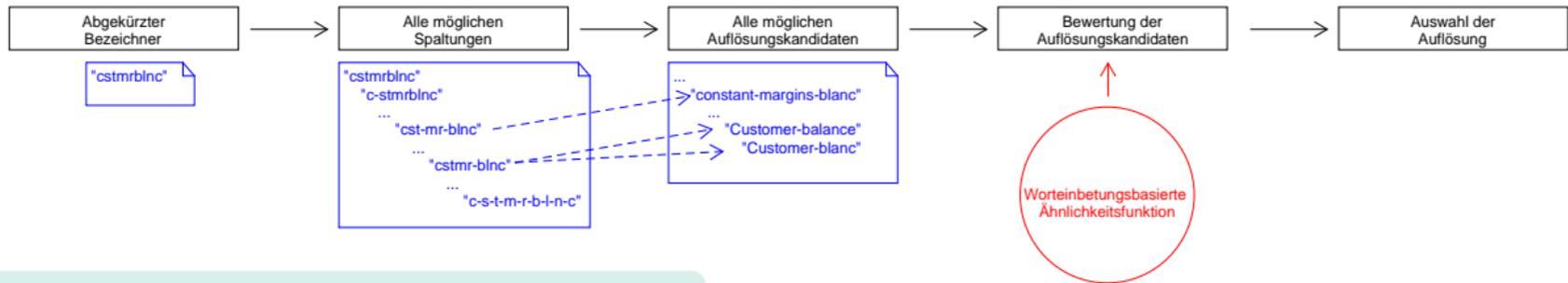


```

private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrbnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Ansatz 2/3: Worteinbettungsbasierte Abkürzungsauflösung

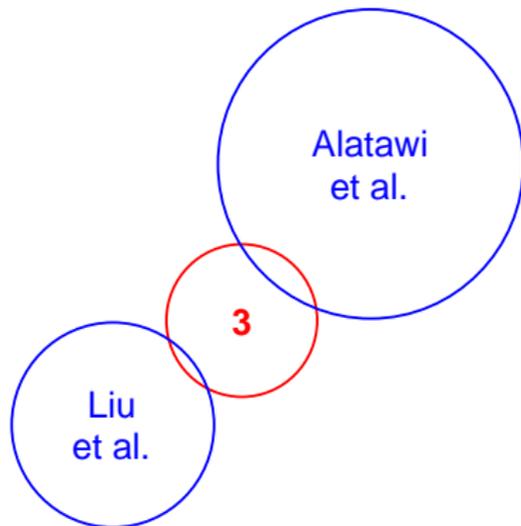


```

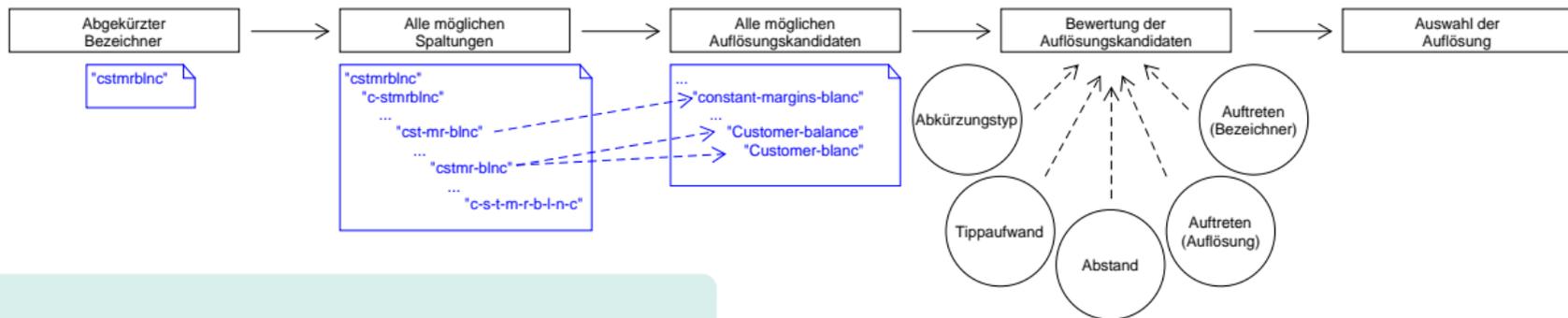
private final String constant = "constant";
private String blanc_whiteSpace = " "; //used for margins

//how much money someone has
static float cstmrblnc(String nm){
    int id = getCustomerID(name);
    return accounts.get(id).balance;
}
  
```

Ansatz 3/3: Worteinbettungsbasierte Mehrdeutigkeitsauflösung

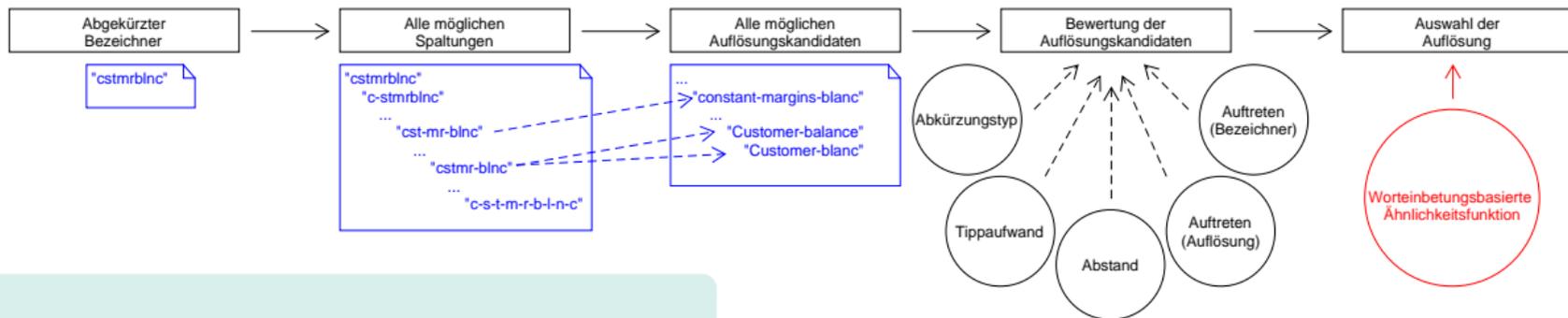


Ansatz 3/3: Worteinbettungsbasierte Mehrdeutigkeitsauflösung



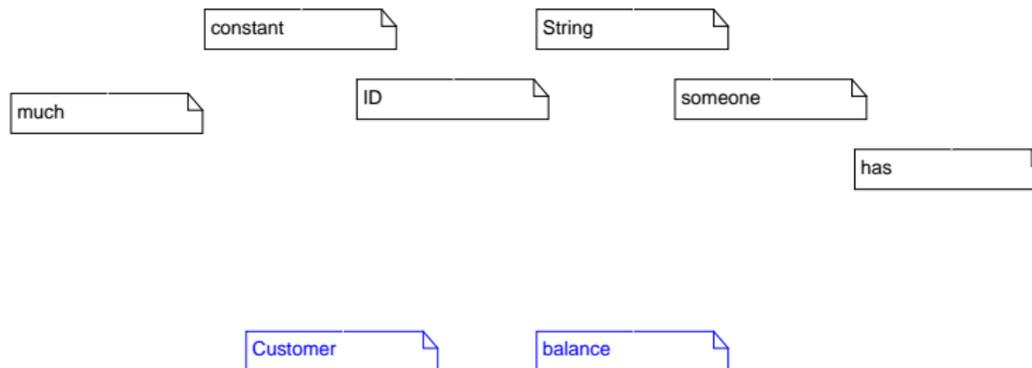
```
private final String constant = "constant";  
private String blanc_whiteSpace = " "; //used for margins  
  
//how much money someone has  
static float cstmrblnc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Ansatz 3/3: Worteinbettungs-basierte Mehrdeutigkeitsauflösung

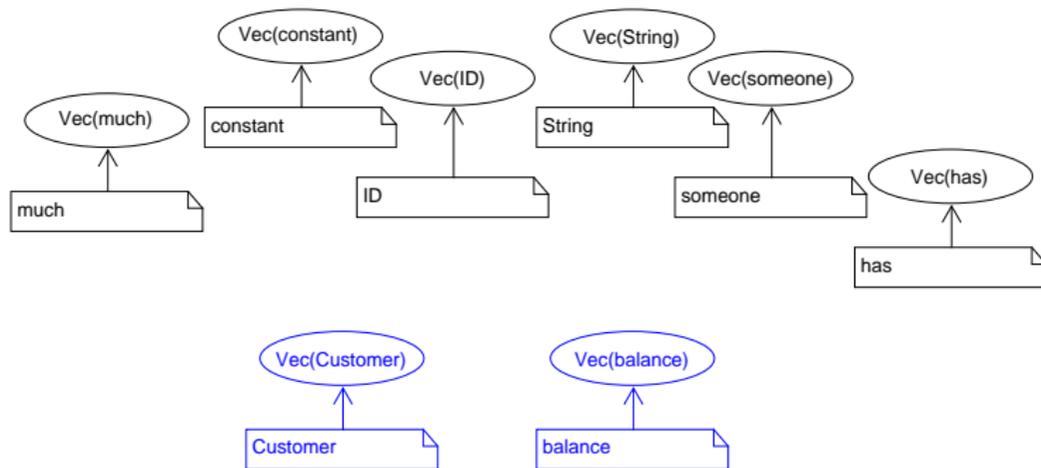


```
private final String constant = "constant";  
private String blanc_whiteSpace = " "; //used for margins  
  
//how much money someone has  
static float cstmrbnc(String nm){  
    int id = getCustomerID(name);  
    return accounts.get(id).balance;  
}
```

Ähnlichkeitsfunktionen



Ähnlichkeitsfunktionen



Motivation

○○

Verwandte Arbeiten

○

Ansätze

○○○○○○○●

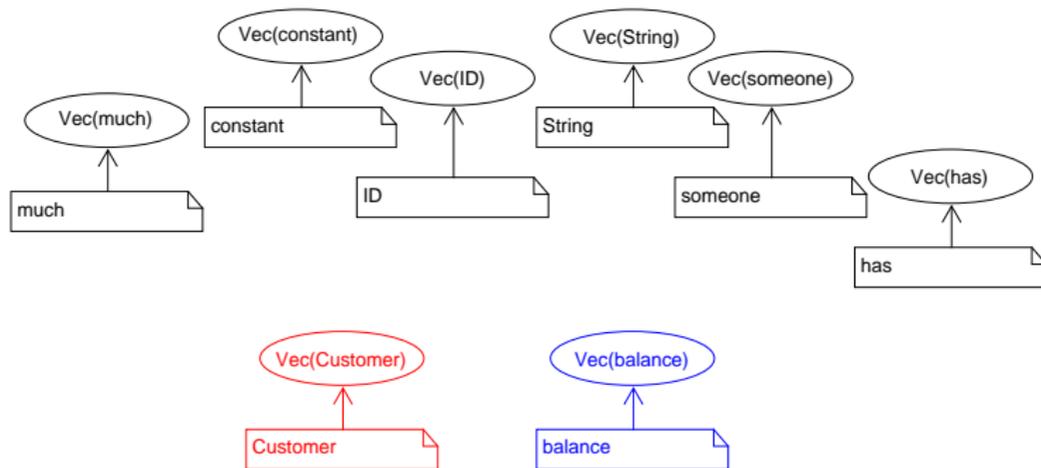
Evaluation

○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

○○○

Ähnlichkeitsfunktionen



Motivation

○○

Verwandte Arbeiten

○

Ansätze

○○○○○○○●

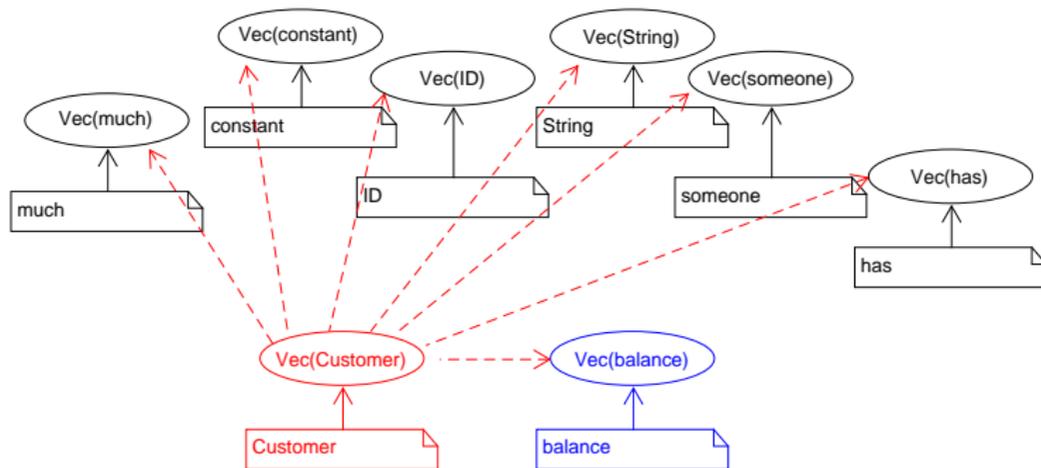
Evaluation

○○○○○○○○○○○○○○○○○

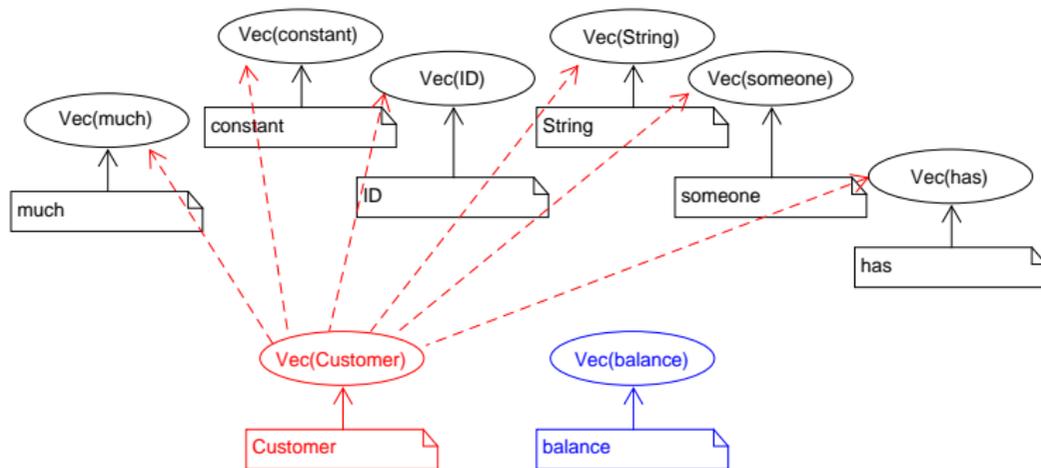
Zusammenfassung und Ausblick

○○○

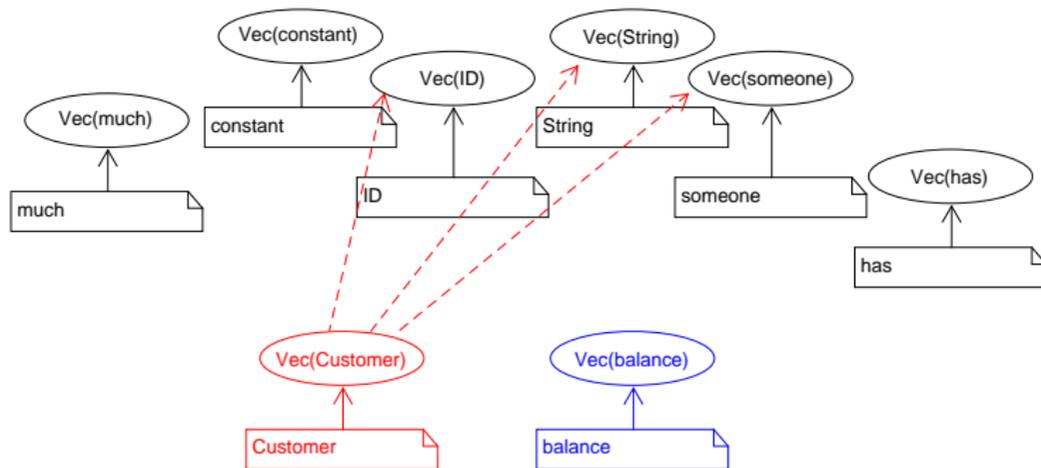
Ähnlichkeitsfunktion 1/5: Naiv



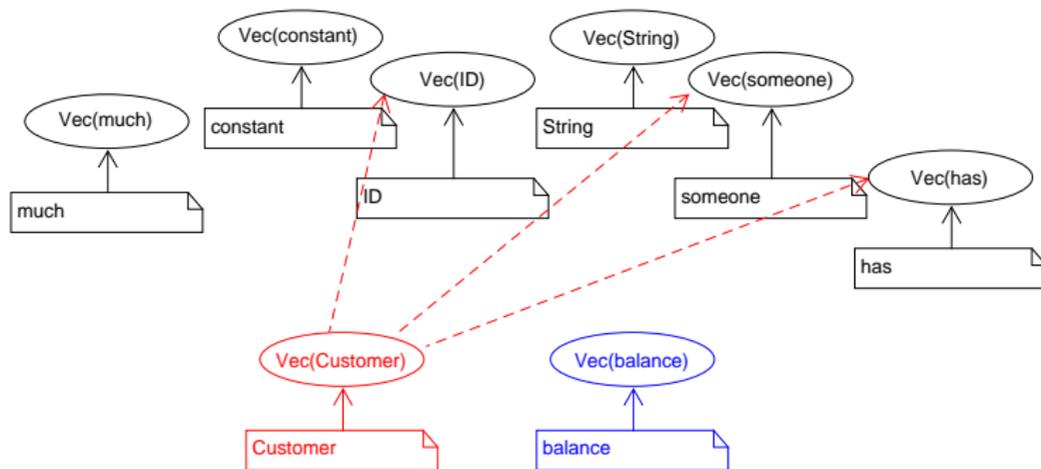
Ähnlichkeitsfunktion 2/5: Kontext



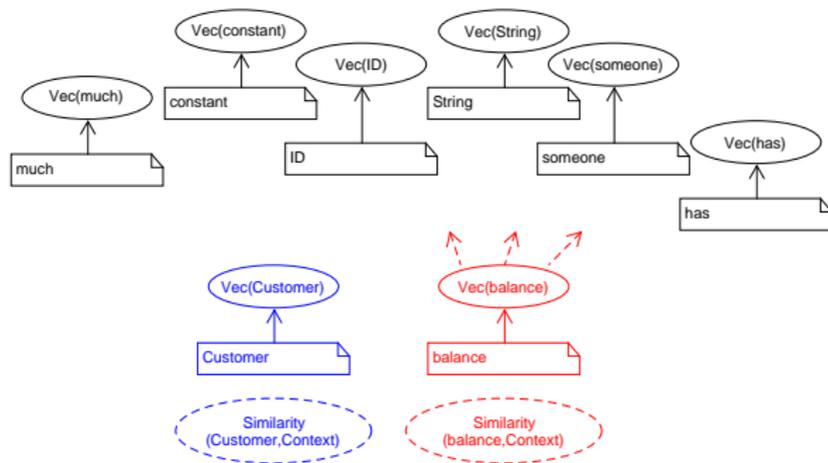
Ähnlichkeitsfunktion 3/5: Wortarten



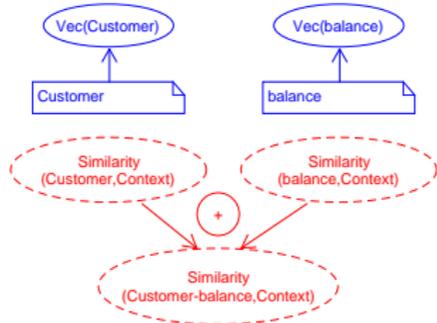
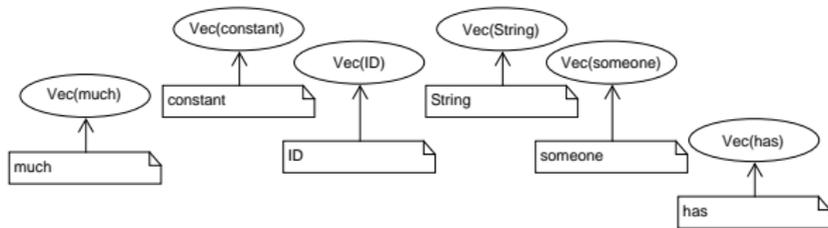
Ähnlichkeitsfunktion 4/5: Aufgeteilt



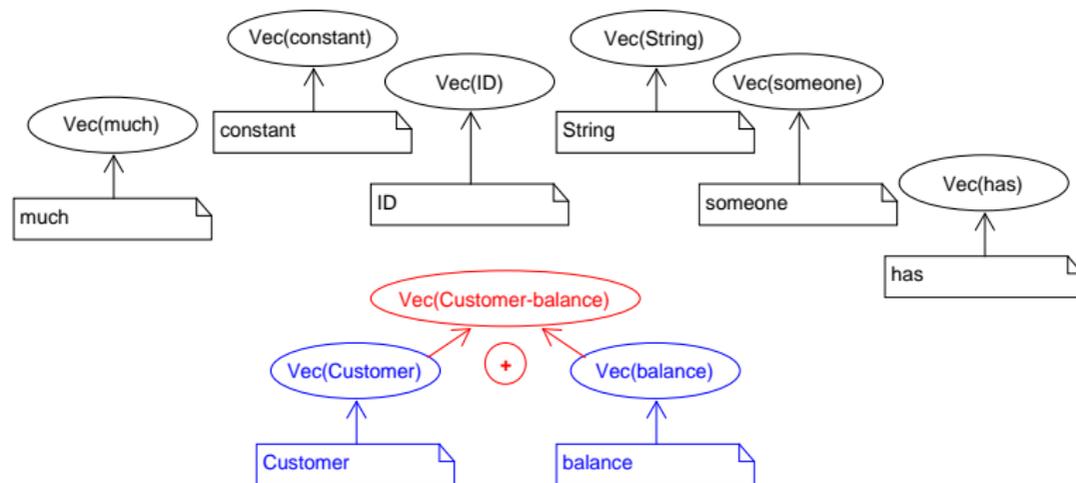
Segmentweise Ähnlichkeitsfunktionen



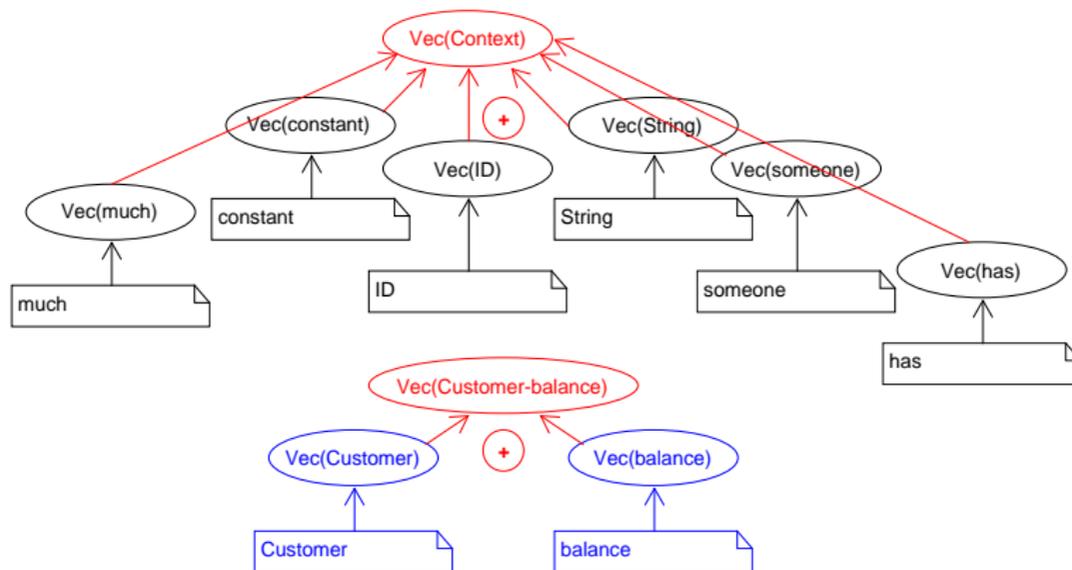
Segmentweise Ähnlichkeitsfunktionen



Ähnlichkeitsfunktion 5/5: Kombinierte Vektoren



Ähnlichkeitsfunktion 5/5: Kombinierte Vektoren



Motivation

○○

Verwandte Arbeiten

○

Ansätze

○○○○○○○●

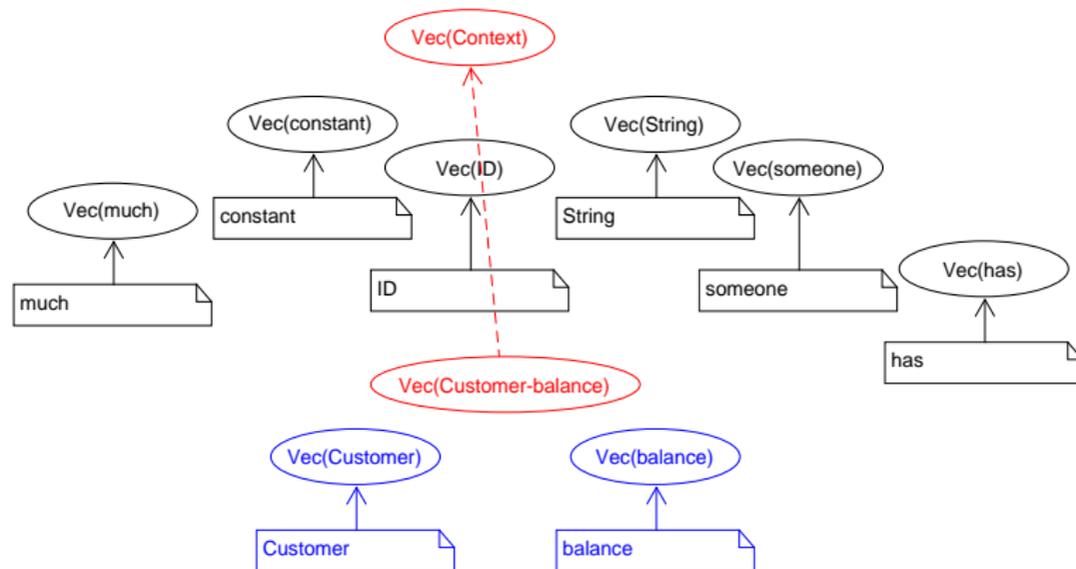
Evaluation

○○○○○○○○○○○○○○○○○

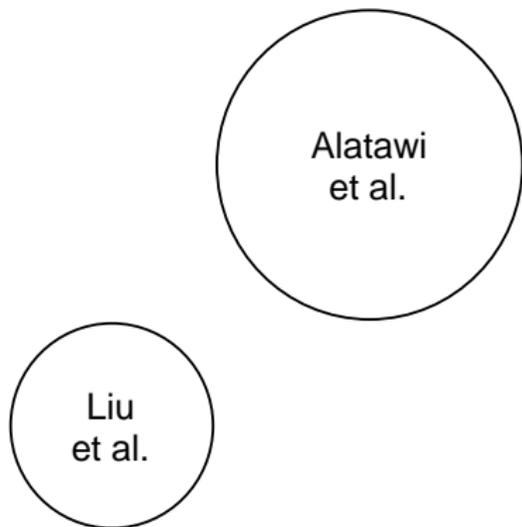
Zusammenfassung und Ausblick

○○○

Ähnlichkeitsfunktion 5/5: Kombinierte Vektoren



Forschungsfragen



Motivation

oo

Verwandte Arbeiten

o

Ansätze

oooooooo

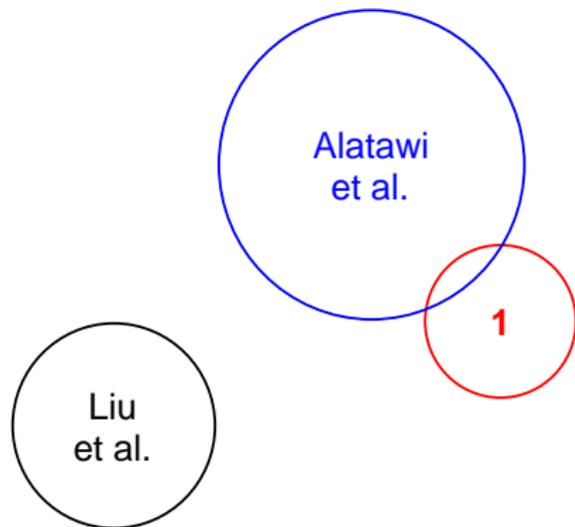
Evaluation

●oooooooooooooooooooo

Zusammenfassung und Ausblick

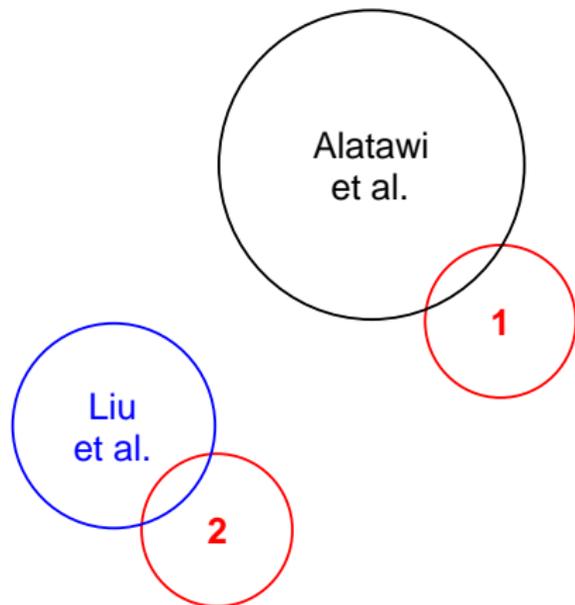
ooo

Forschungsfragen



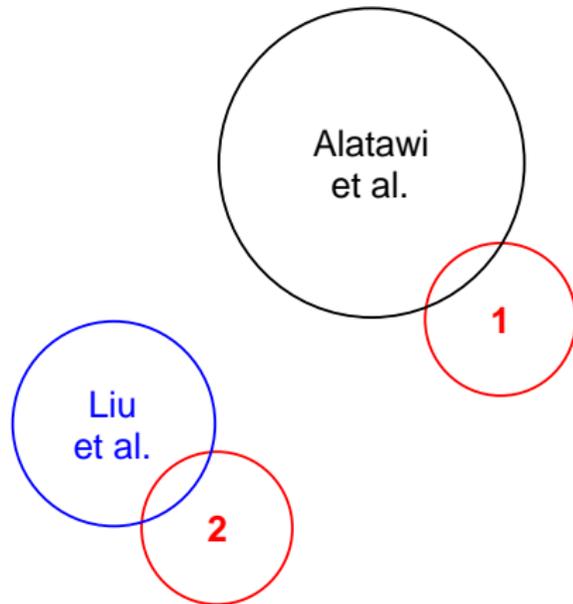
- Präzision: Trigramm-Verfahren gegen Alatawi et al.

Forschungsfragen



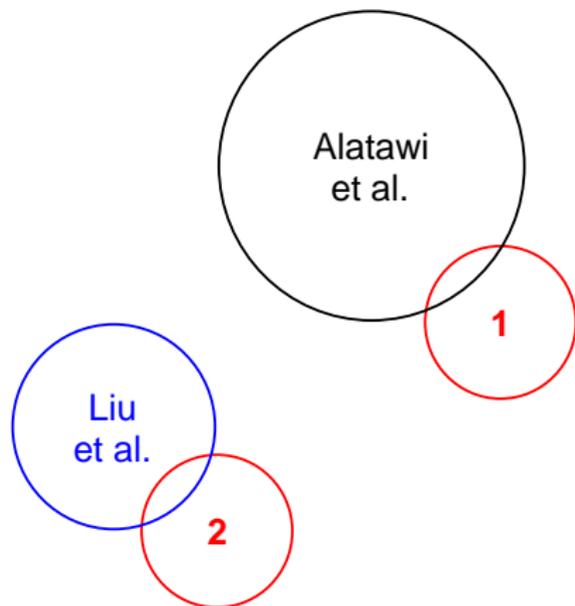
- ① Präzision: Trigramm-Verfahren gegen Alatawi et al.
- ② Am besten geeignete Ähnlichkeitsfunktion

Forschungsfragen



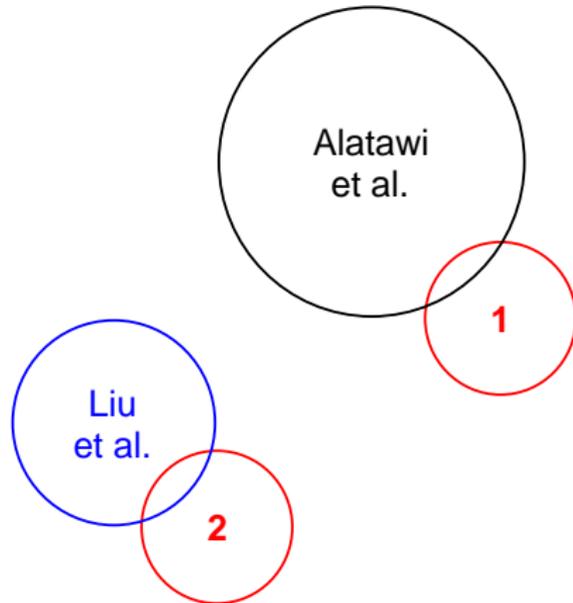
- ① Präzision: Trigramm-Verfahren gegen Alatawi et al.
- ② Am besten geeignete Ähnlichkeitsfunktion
- ③ Am besten geeignetes Worteinbettungsverfahren

Forschungsfragen



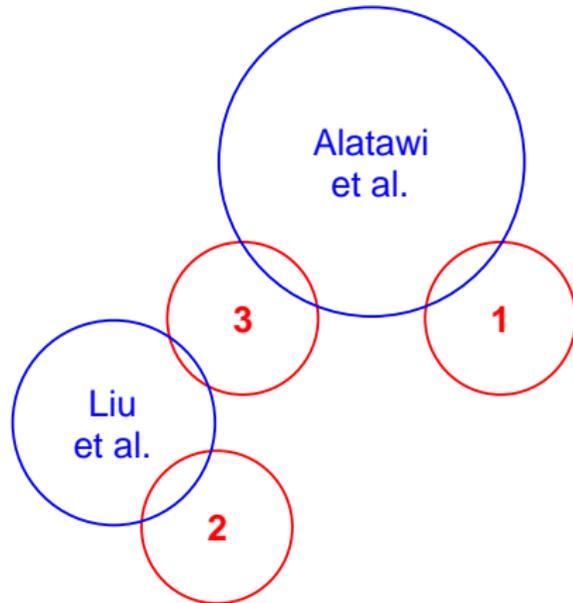
- ① Präzision: Trigramm-Verfahren gegen Alatawi et al.
- ② Am besten geeignete Ähnlichkeitsfunktion
- ③ Am besten geeignetes Worteinbettungsverfahren
- ④ Am besten geeignete Wissensquelle

Forschungsfragen



- ① Präzision: Trigramm-Verfahren gegen Alatawi et al.
- ② Am besten geeignete Ähnlichkeitsfunktion
- ③ Am besten geeignetes Worteinbettungsverfahren
- ④ Am besten geeignete Wissensquelle
- ⑤ Präzision: Worteinbettungsbasierte Abkürzungsauflösung gegen Alatawi et al.

Forschungsfragen



- ① Präzision: Trigramm-Verfahren gegen Alatawi et al.
- ② Am besten geeignete Ähnlichkeitsfunktion
- ③ Am besten geeignetes Worteinbettungsverfahren
- ④ Am besten geeignete Wissensquelle
- ⑤ Präzision: Worteinbettungsbasierte Abkürzungsauflösung gegen Alatawi et al.
- ⑥ Präzision: Worteinbettungsbasierte Mehrdeutigkeitsauflösung gegen Willkür

Beispielquelltexte und Musterlösungen

Musterlösung	echtes Programm	alle Abkürzungen	Ort im Quelltext
<i>Alatawi et al.</i> ¹	✘	✘	✔
-----	-----	-----	-----

Beispielquelltexte und Musterlösungen

Musterlösung	echtes Programm	alle Abkürzungen	Ort im Quelltext
<i>Alatawi et al.</i> ¹	✘	✘	✔
<i>Alatawi et al.</i> ²	✘	✘	✘

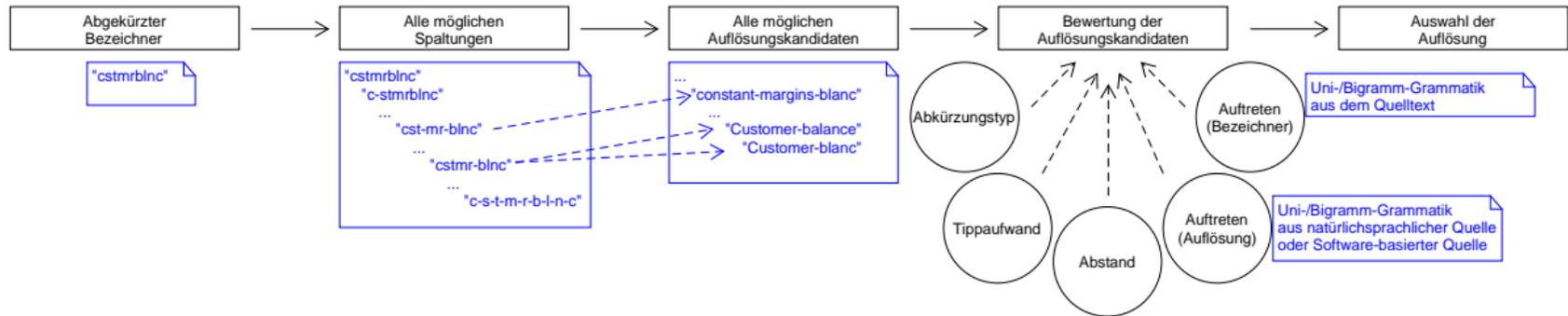
Beispielquelltexte und Musterlösungen

Musterlösung	echtes Programm	alle Abkürzungen	Ort im Quelltext
<i>Alatawi et al.</i> ¹	✗	✗	✓
<i>Alatawi et al.</i> ²	✗	✗	✗
<i>a2ps</i>	✓	✗	✗
<i>which</i>	✓	✓	✗

Beispielquelltexte und Musterlösungen

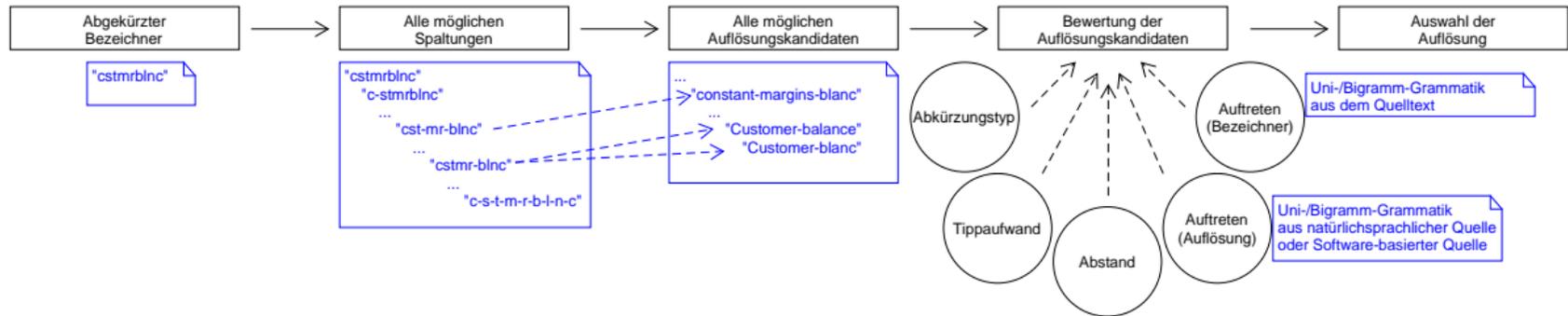
Musterlösung	echtes Programm	alle Abkürzungen	Ort im Quelltext
<i>Alatawi et al.</i> ¹	✗	✗	✓
<i>Alatawi et al.</i> ²	✗	✗	✗
<i>a2ps</i>	✓	✗	✗
<i>which</i>	✓	✓	✗
<i>Dronology</i>	✓	✓	✗

Alatawi et al. als Referenz



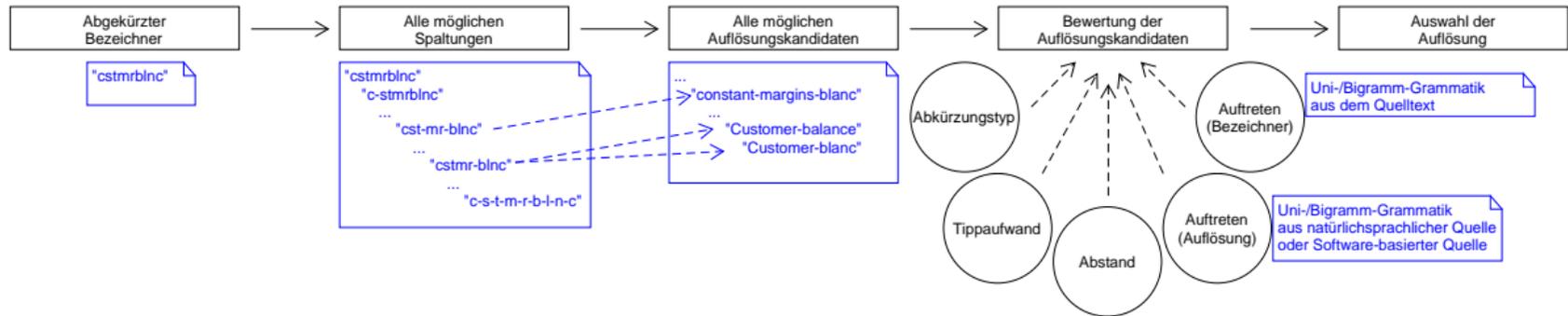
Quelltext	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	0,6703
-----	-----
-----	-----
-----	-----

Alatawi et al. als Referenz



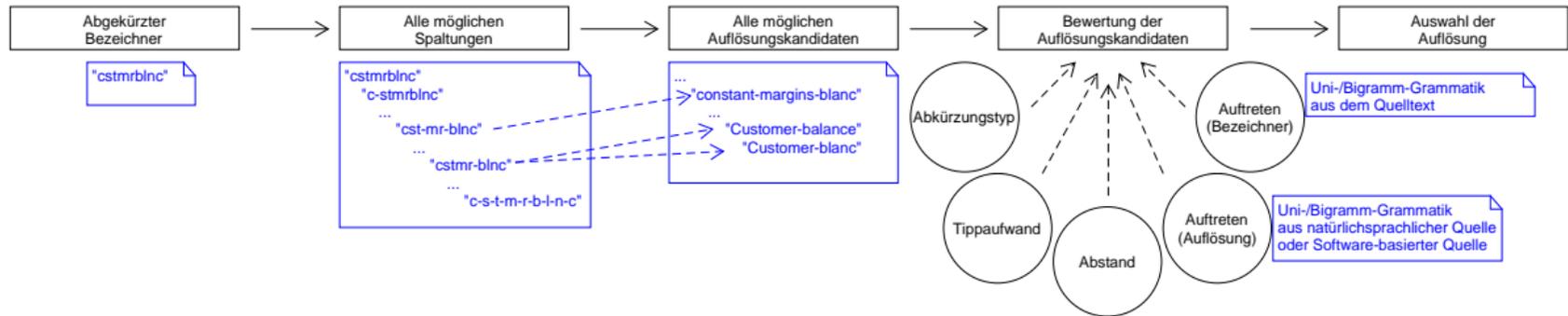
Quelltext	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	0,6703
<i>Alatawi et al.</i> ²	0,3427
-----	-----
-----	-----
-----	-----

Alatawi et al. als Referenz



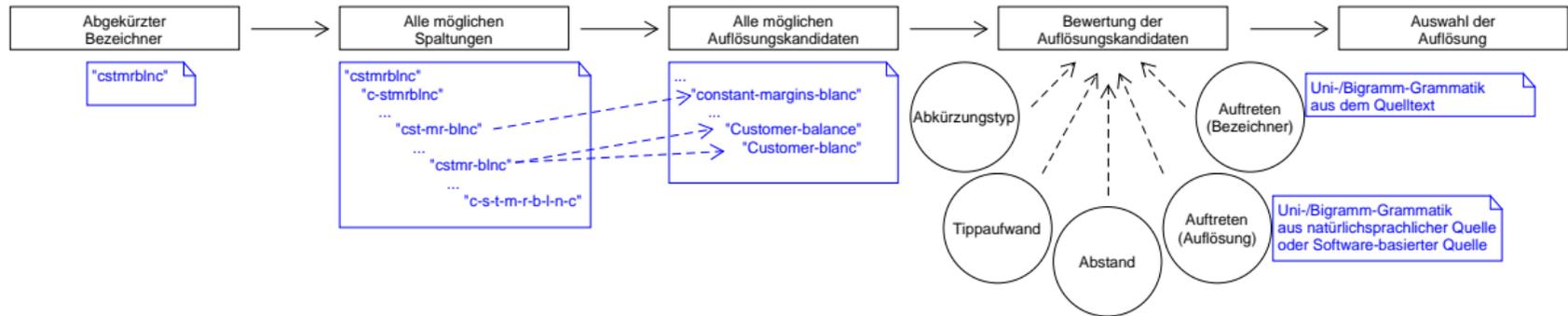
Quelltext	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	0,6703
<i>Alatawi et al.</i> ²	0,3427
<i>a2ps</i>	0,3678
-----	-----
-----	-----

Alatawi et al. als Referenz



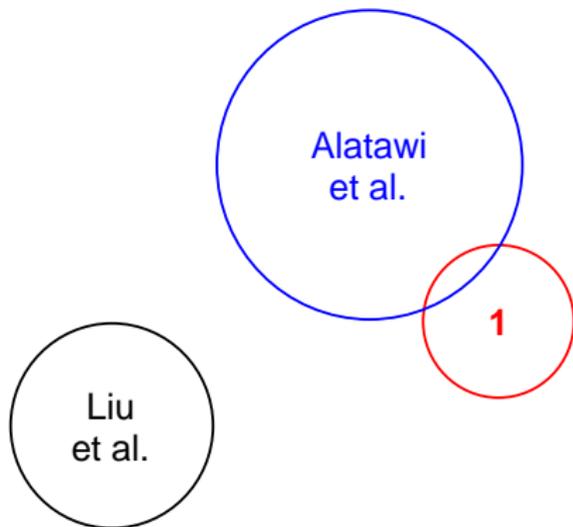
Quelltext	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	0,6703
<i>Alatawi et al.</i> ²	0,3427
<i>a2ps</i>	0,3678
<i>which</i>	0,3938

Alatawi et al. als Referenz



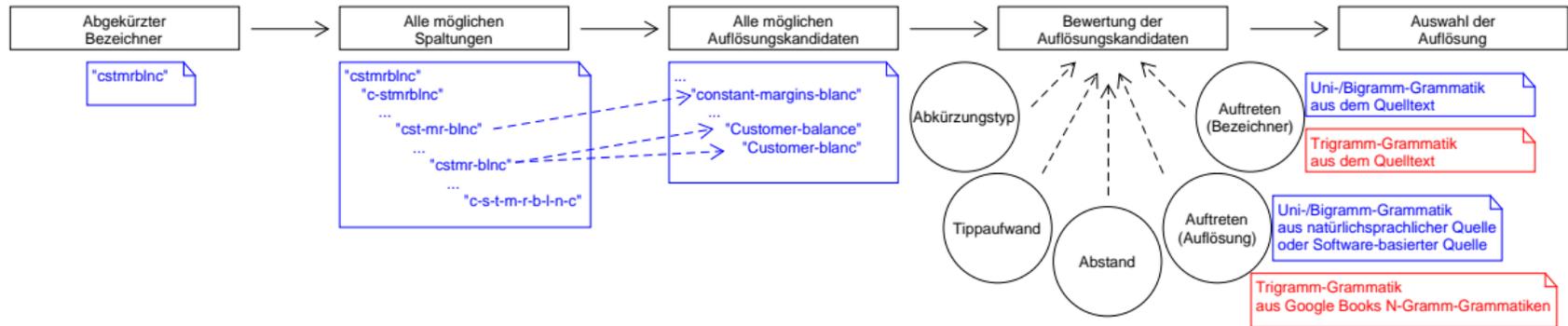
Quelltext	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	0,6703
<i>Alatawi et al.</i> ²	0,3427
<i>a2ps</i>	0,3678
<i>which</i>	0,3938
<i>Dronology</i>	0,0511

Forschungsfrage 1/6:



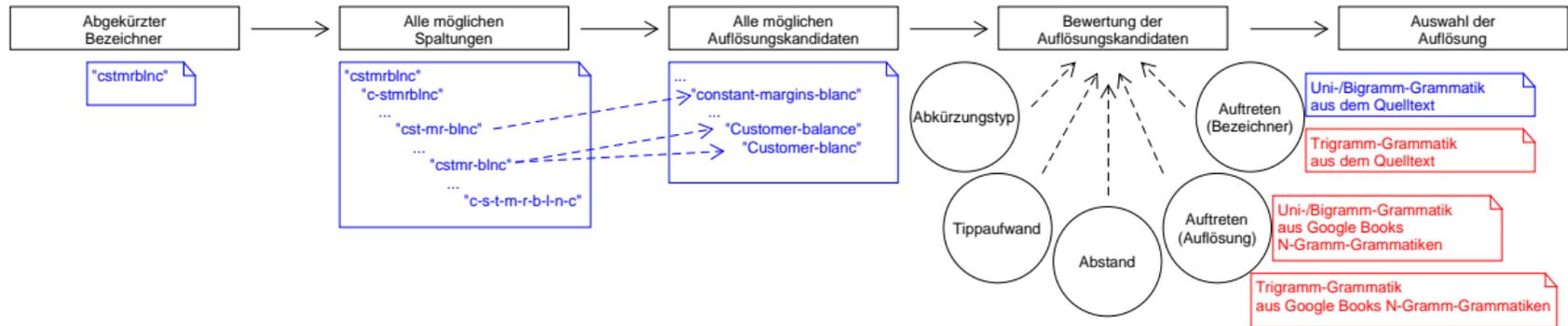
Präzision: Trigramm-Verfahren gegen Alatawi et al.

Forschungsfrage 1/6: Trigramm-basierte Auflösung



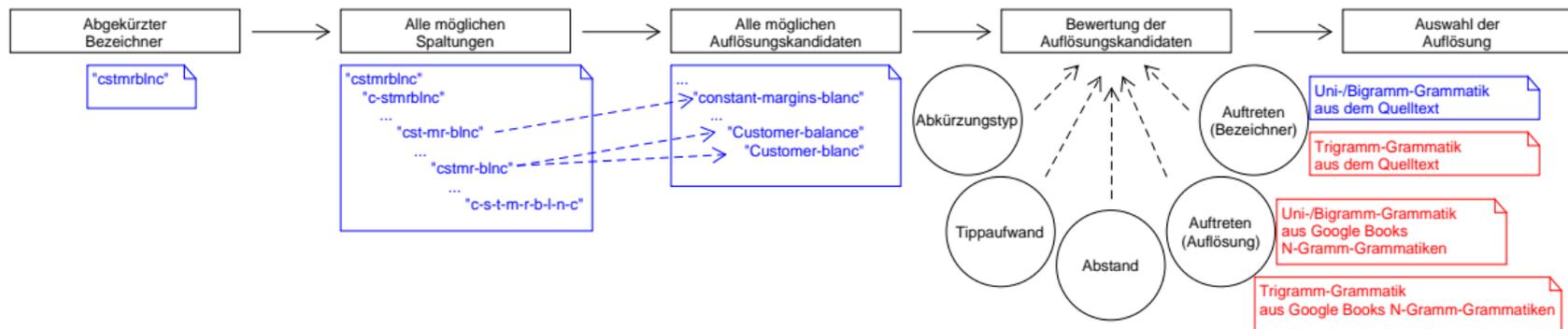
Verfahren	Präzision (Bezeichner)
Trigramm	0,7033
Bigramm	0,6703

Forschungsfrage 1/6: Trigramm-basierte Auflösung



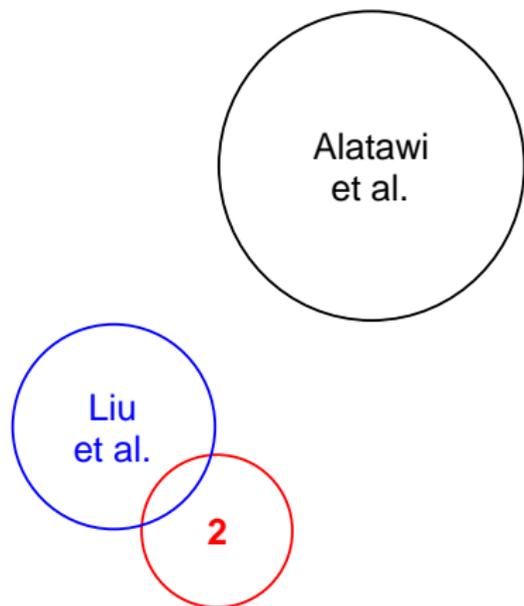
Verfahren	Präzision (Bezeichner)
Trigramm	0,5055
Bigramm	0,5055

Forschungsfrage 1/6: Trigramm-basierte Auflösung



Verfahren	Präzision (Bezeichner)	Präzision (Segment)
Trigramm	0,5055	0,7104
Bigramm	0,5055	0,6833

Forschungsfrage 2/6:



Am besten geeignete Ähnlichkeitsfunktion

Motivation
○○

Verwandte Arbeiten
○

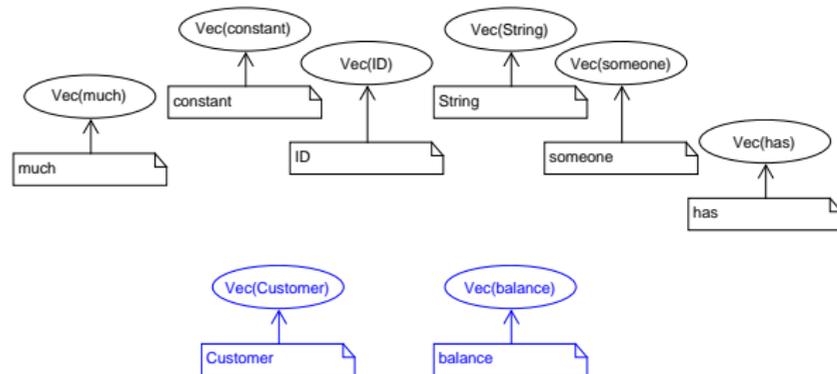
Ansätze
○○○○○○○○

Evaluation
○○○○○○●○○○○○○○○

Zusammenfassung und Ausblick
○○○

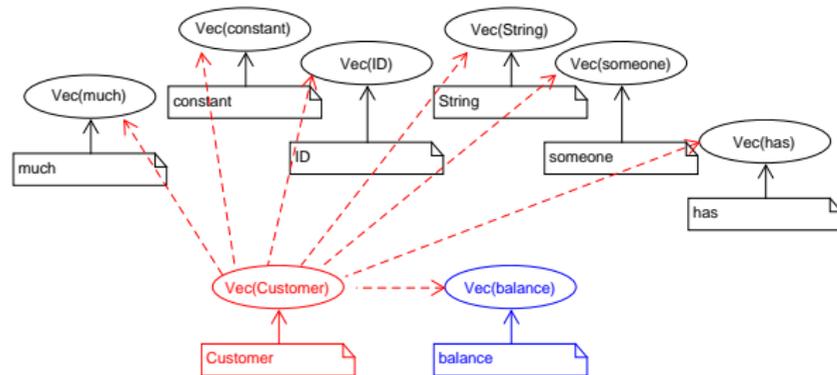
Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)



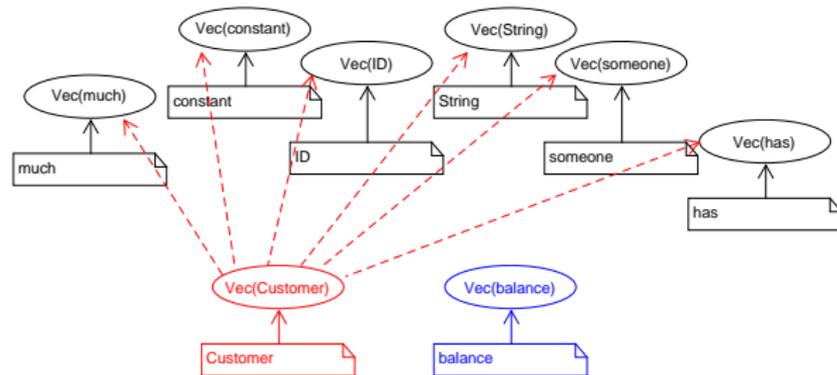
Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)
<i>Naiv</i>	0,0879



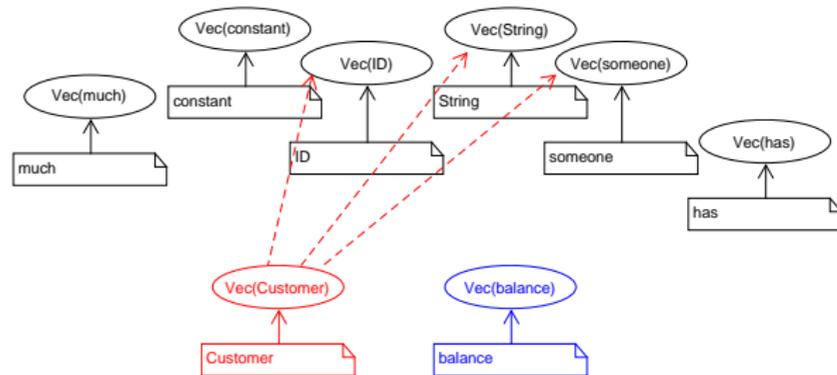
Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)
<i>Naiv</i>	0,0879
<i>Kontext</i>	0,0879



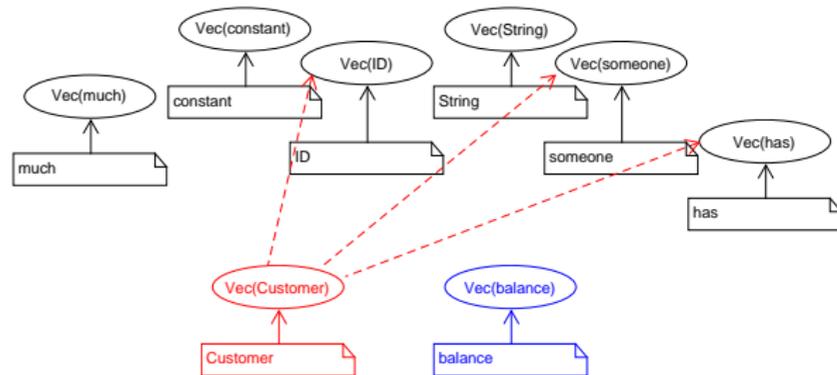
Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)
<i>Naiv</i>	0,0879
<i>Kontext</i>	0,0879
<i>Wortarten</i>	0,1099



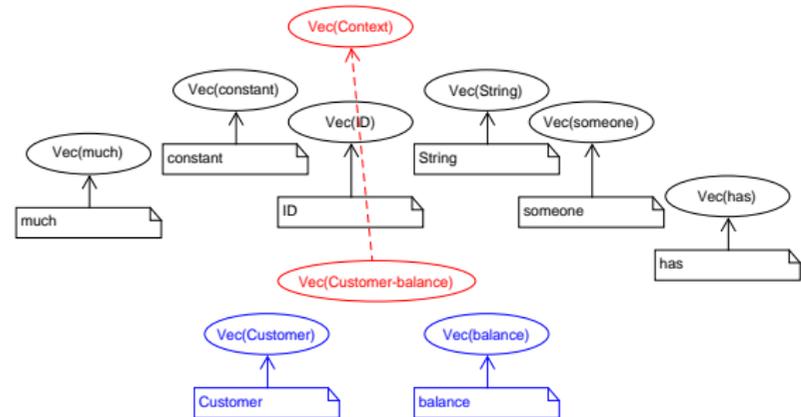
Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)
<i>Naiv</i>	0,0879
<i>Kontext</i>	0,0879
<i>Wortarten</i>	0,1099
<i>Aufgeteilt</i>	0,1429

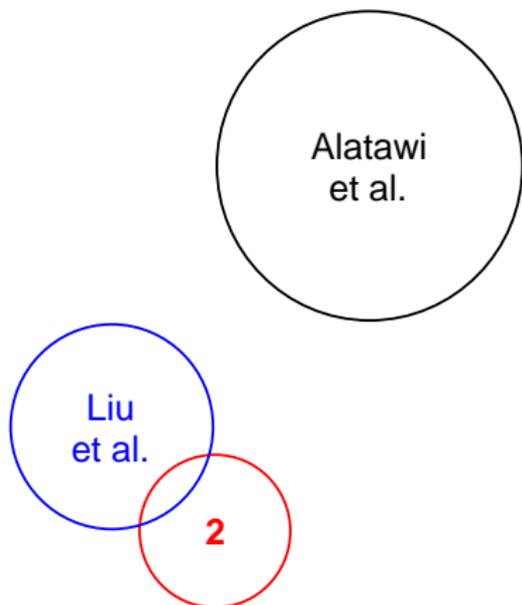


Forschungsfrage 2/6: Ähnlichkeitsfunktion

Ähnlichkeitsfunktion	Präzision (Bezeichner)
<i>Naiv</i>	0,0879
<i>Kontext</i>	0,0879
<i>Wortarten</i>	0,1099
<i>Aufgeteilt</i>	0,1429
<i>Kombinierte Vektoren</i>	0,1538

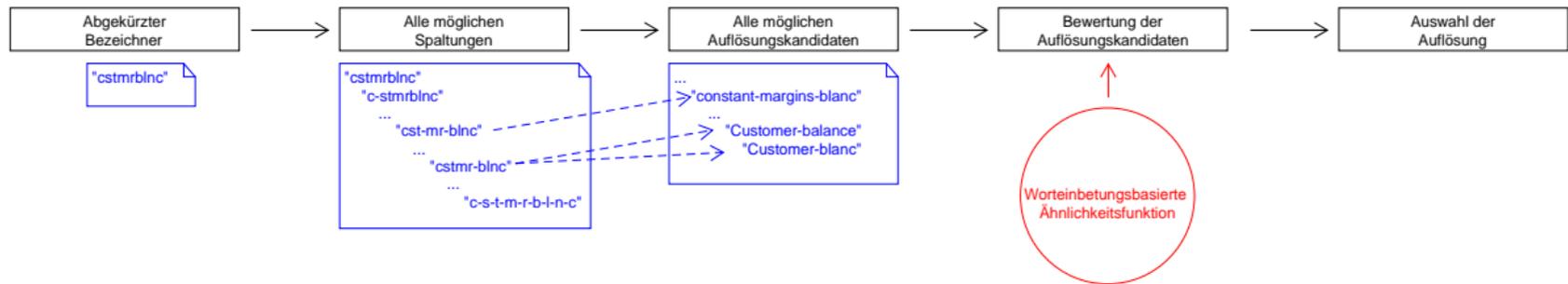


Forschungsfrage 3/6:



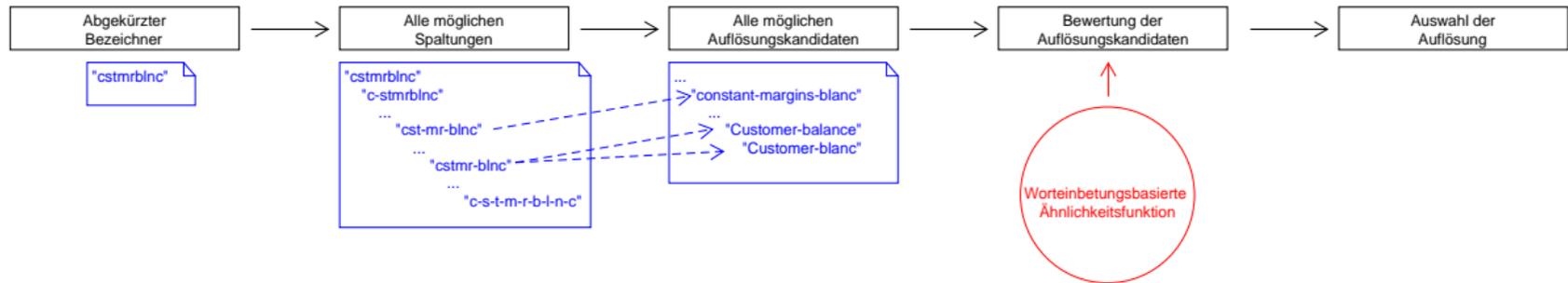
Am besten geeignetes Worteinbettungsverfahren

Forschungsfrage 3/6: Worteinbettungsverfahren



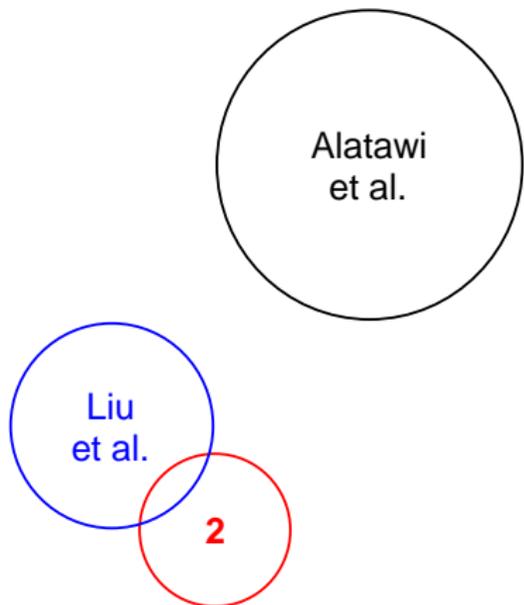
Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>SO_200</i> [5]	<i>word2vec</i>	0,1538
<i>Alatawi et al.</i> ¹	<i>Common Crawl</i> [6]	<i>fastText</i>	0,1319

Forschungsfrage 3/6: Worteinbettungsverfahren



Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>SO_200</i> [5]	<i>word2vec</i>	0,1538
<i>Alatawi et al.</i> ¹	<i>Common Crawl</i> [6]	<i>fastText</i>	0,1319
<i>Alatawi et al.</i> ¹	Quelltext	<i>word2vec</i>	0,1099
<i>Alatawi et al.</i> ¹	Quelltext	<i>fastText</i>	0,1319

Forschungsfrage 4/6:



Am besten geeignete Wissensquelle

Motivation
○○

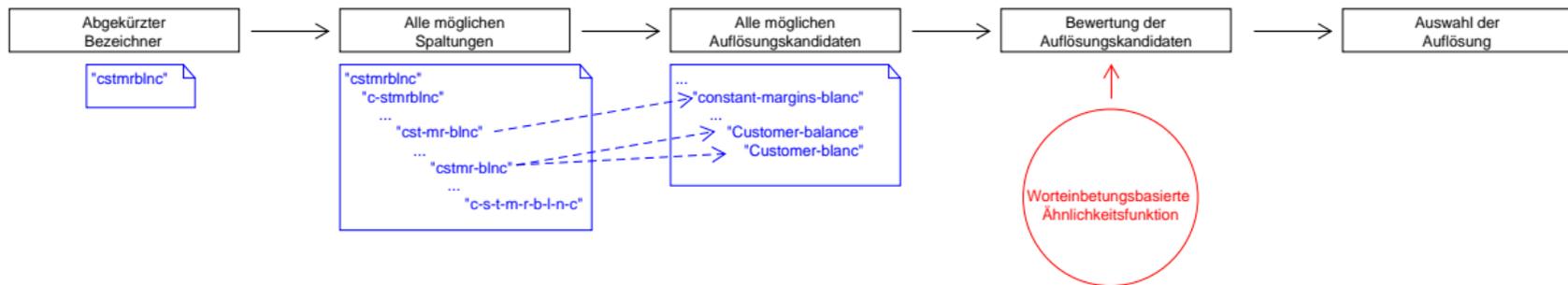
Verwandte Arbeiten
○

Ansätze
○○○○○○○○

Evaluation
○○○○○○○○○○●○○○○○○

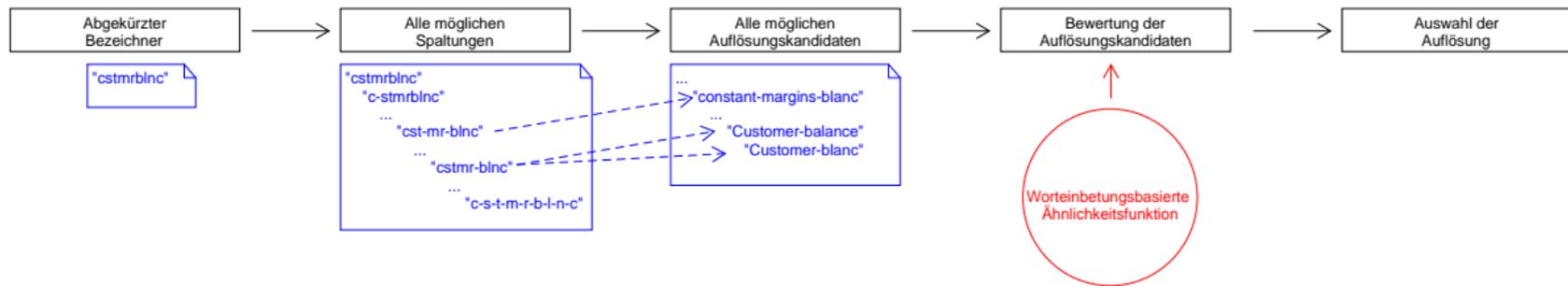
Zusammenfassung und Ausblick
○○○

Forschungsfrage 4/6: Wissensquelle



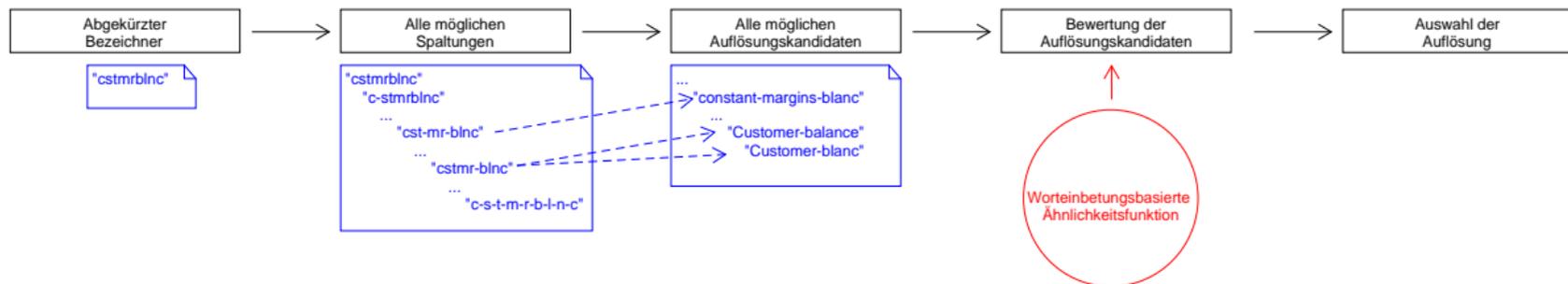
Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	SO_200 [5]	<i>word2vec</i>	0,1538
<i>Alatawi et al.</i> ¹	<i>Common Crawl</i> [6]	<i>fastText</i>	0,1319
<i>Alatawi et al.</i> ¹	Quelltext	<i>word2vec</i>	0,1099
<i>Alatawi et al.</i> ¹	Quelltext	<i>fastText</i>	0,1319

Forschungsfrage 4/6: Wissensquelle



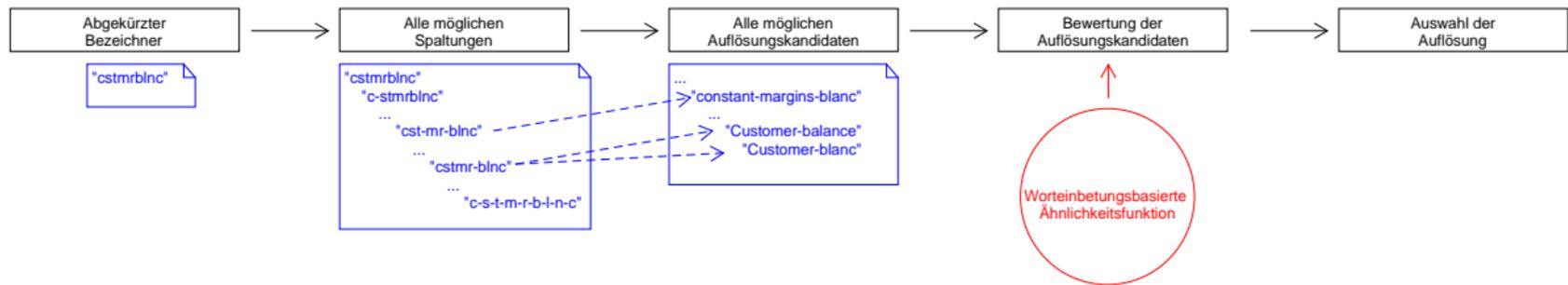
Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Dronology</i>	SO_200 [5]	<i>word2vec</i>	0,0511
<i>Dronology</i>	Quelltext	<i>word2vec</i>	0,0511

Forschungsfrage 4/6: Wissensquelle



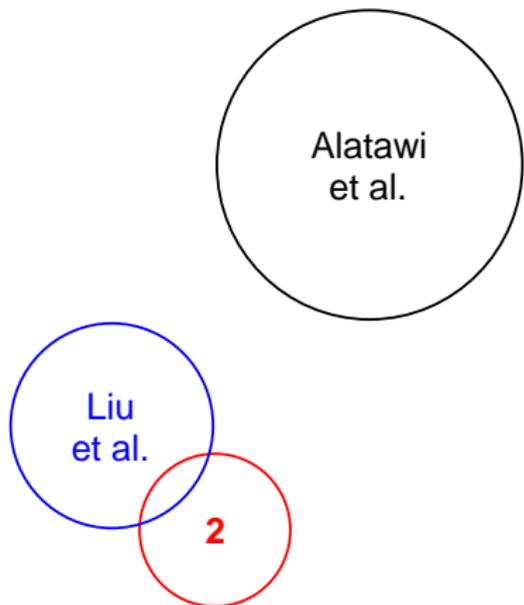
Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Dronology</i>	SO_200 [5]	<i>word2vec</i>	0,0511
<i>Dronology</i>	Quelltext	<i>word2vec</i>	0,0511
<i>Dronology</i>	acht Quelltexte	<i>word2vec</i>	0,0481

Forschungsfrage 4/6: Wissensquelle



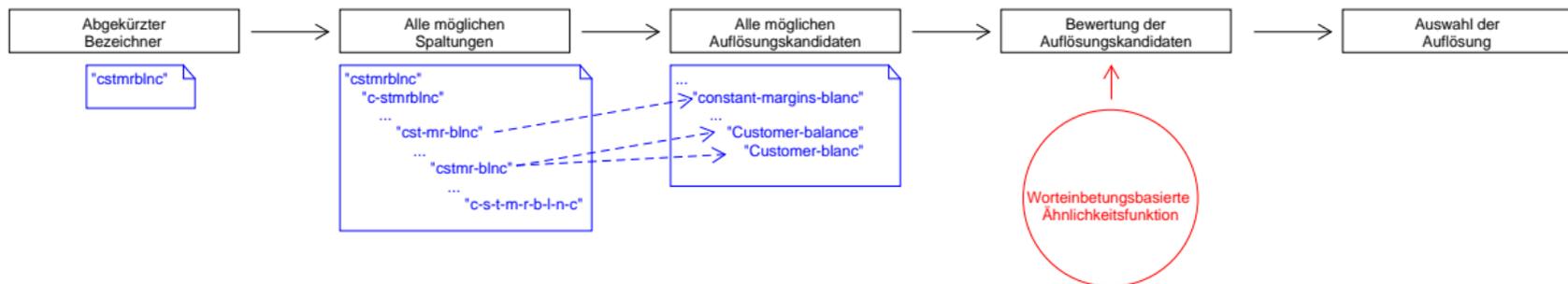
Quelltext/Musterlösung	Wissensquelle	Verfahren	Präzision (Bezeichner)
<i>Dronology</i>	SO_200 [5]	<i>word2vec</i>	0,0511
<i>Dronology</i>	Quelltext	<i>word2vec</i>	0,0511
<i>Dronology</i>	acht Quelltexte	<i>word2vec</i>	0,0481
<i>Dronology</i>	Wikipedia-Artikel	<i>word2vec</i>	0,0401

Forschungsfrage 5/6:



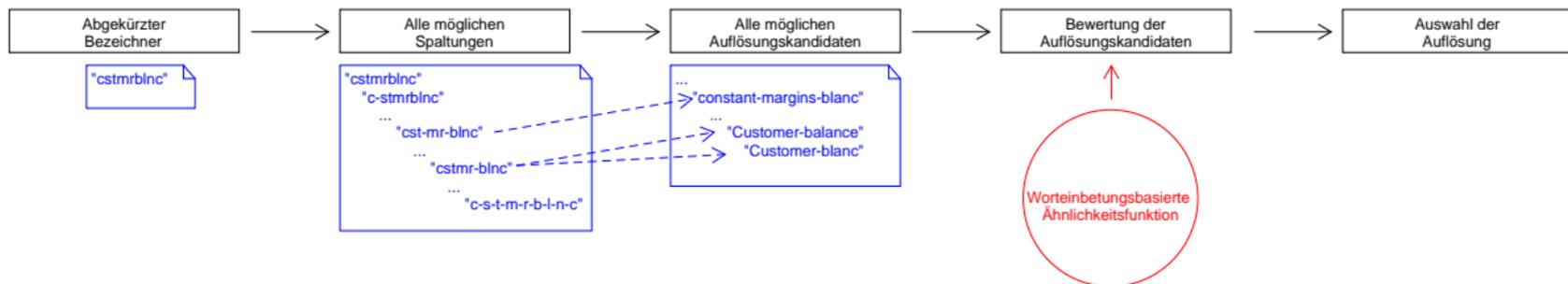
Präzision: Worteinbettungsbasierte Abkürzungsauflösung gegen Alatawi et al.

Forschungsfrage 5/6: Worteinbettungsbasierte Abkürzungsauflösung



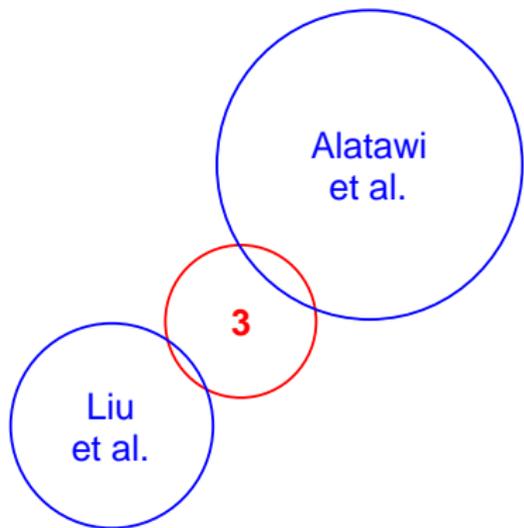
Quelltext/Musterlösung	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>word2vec</i>	0,1538
<i>Alatawi et al.</i> ¹	Bigramm	0,6703

Forschungsfrage 5/6: Worteinbettungsbasierte Abkürzungsauflösung



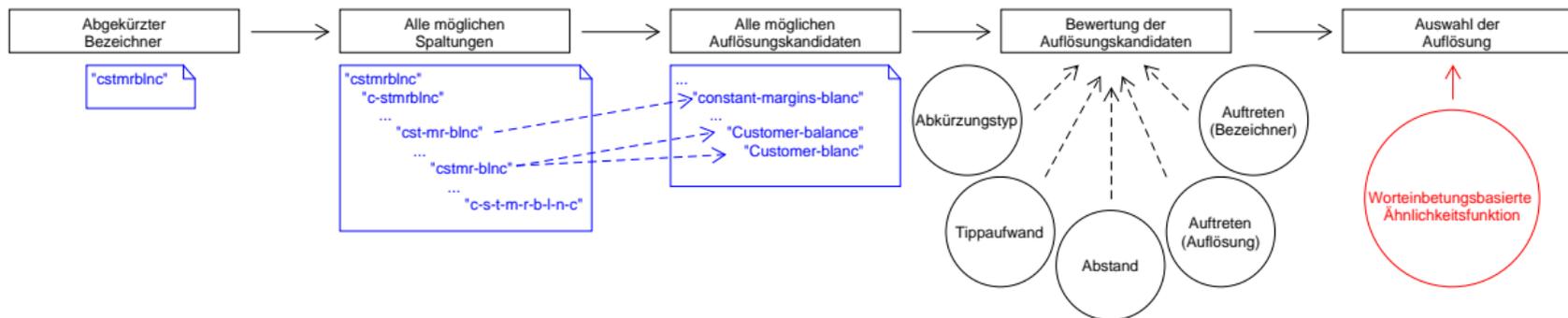
Quelltext/Musterlösung	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>word2vec</i>	0,1538
<i>Alatawi et al.</i> ¹	Bigramm	0,6703
<i>which</i>	<i>word2vec</i>	0,3952
<i>which</i>	Bigramm	0,3938

Forschungsfrage 6/6:



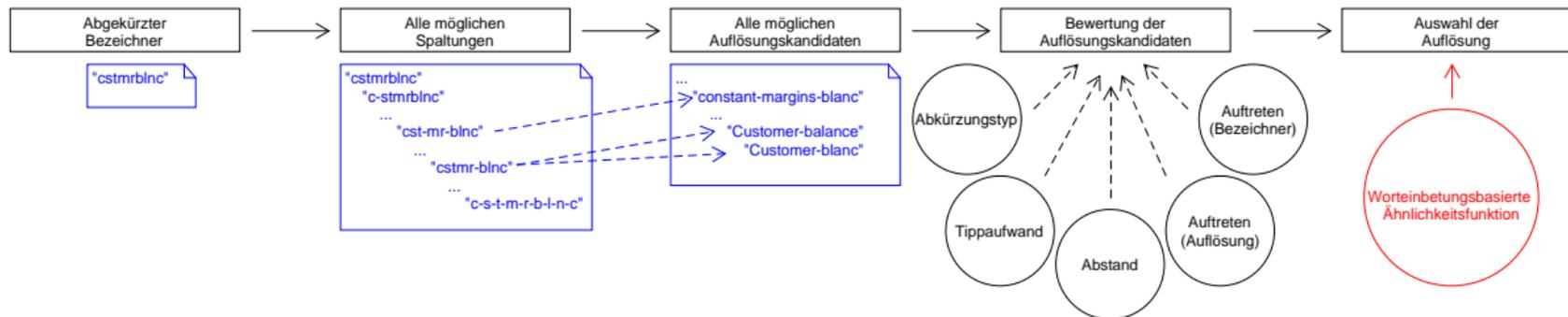
Präzision: Worteinbettungsbasierte Mehrdeutigkeitsauflösung gegen Willkür

Forschungsfrage 6/6: Worteinbettungs-basierte Mehrdeutigkeitsauflösung



Quelltext/Musterlösung	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>word2vec</i>	0,5824
<i>Alatawi et al.</i> ¹	willkürlich	0,5934

Forschungsfrage 6/6: Worteinbettungs-basierte Mehrdeutigkeitsauflösung



Quelltext/Musterlösung	Verfahren	Präzision (Bezeichner)
<i>Alatawi et al.</i> ¹	<i>word2vec</i>	0,5824
<i>Alatawi et al.</i> ¹	willkürlich	0,5934
<i>which</i>	<i>word2vec</i>	0,3943
<i>which</i>	willkürlich	0,3938

Zusammenfassung

```
static float cstmrbnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Zusammenfassung

```
static float cstmrblnc(String nm){  
    int id = getCstmrID(nm);  
    return accounts.get(id).blnc;  
}
```

Alatawi
et al.

Liu
et al.

Motivation

○○

Verwandte Arbeiten

○

Ansätze

○○○○○○○○○

Evaluation

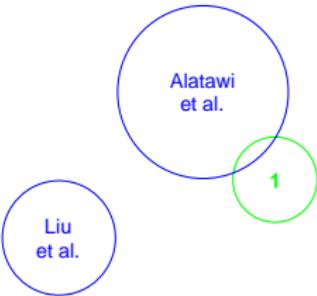
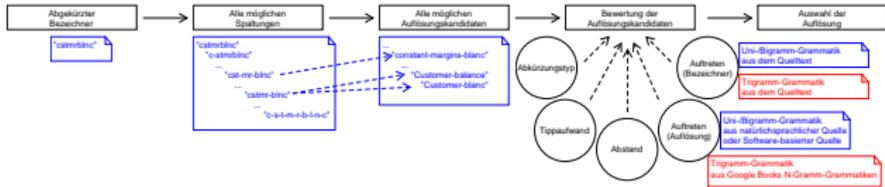
○○○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick

●○○

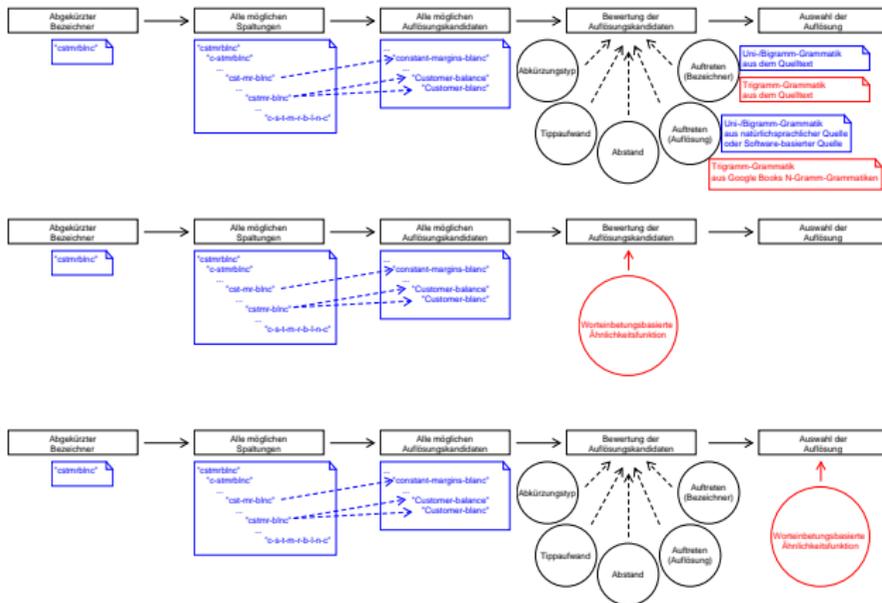
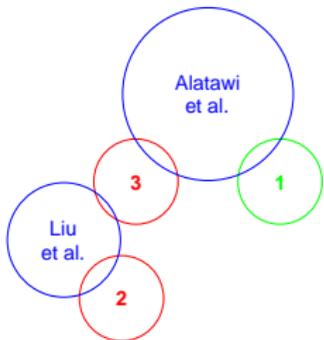
Zusammenfassung

```
static float cstmrbnlc(String nm){
    int id = getCstmrID(nm);
    return accounts.get(id).blnc;
}
```



Zusammenfassung

```
static float cstmrbnlc(String nm){
    int id = getCstmrID(nm);
    return accounts.get(id).blnc;
}
```



Motivation
○○

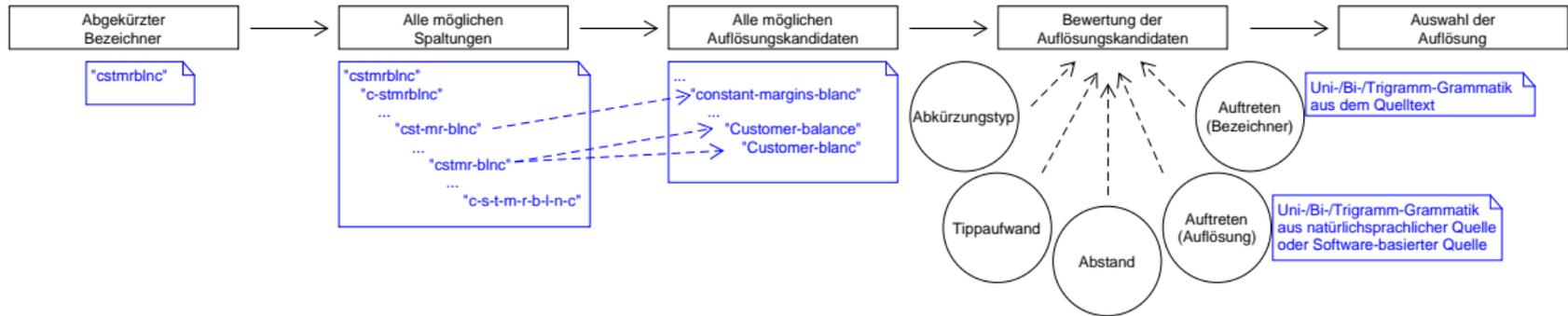
Verwandte Arbeiten
○

Ansätze
○○○○○○○○

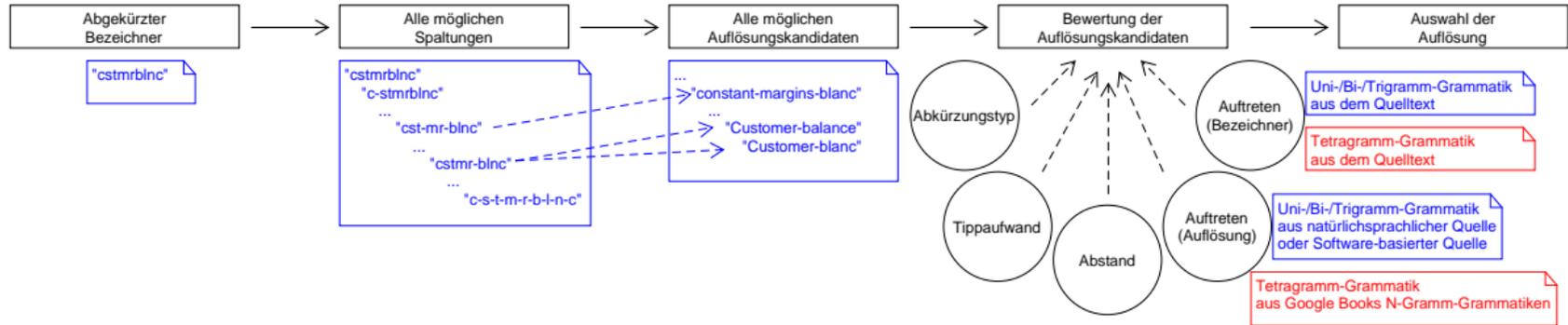
Evaluation
○○○○○○○○○○○○○○○○

Zusammenfassung und Ausblick
●○○

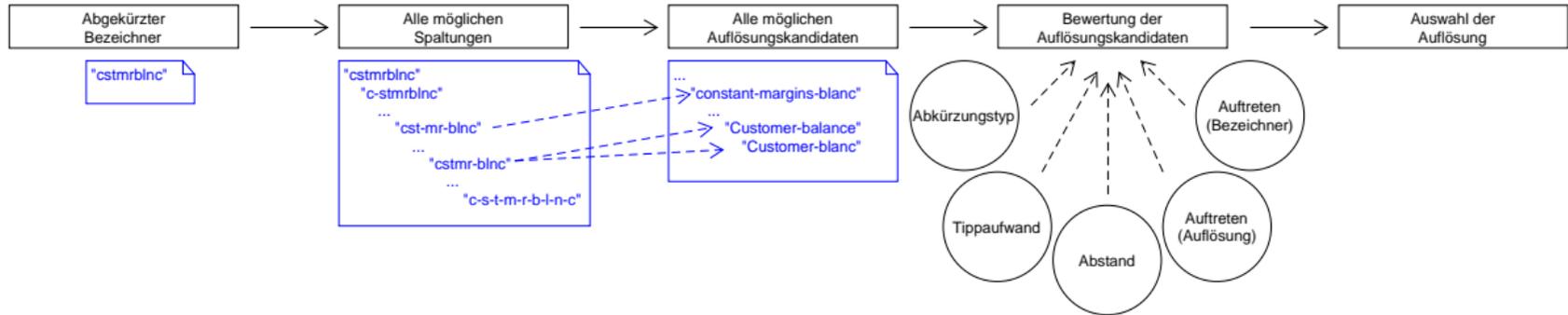
Ausblick: N-Gramme



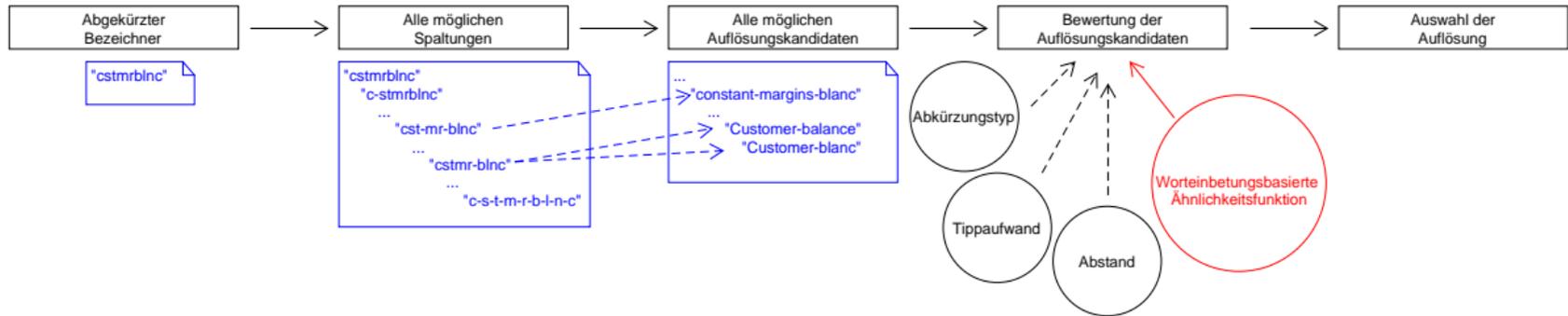
Ausblick: N-Gramme



Ausblick: Worteinbettungen



Ausblick: Worteinbettungen



Literatur I

- [1] Abdulrahman Alatawi, Weifeng Xu und Dianxiang Xu. „Bayesian Unigram-Based Inference for Expanding Abbreviations in Source Code“. In: *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). ISSN: 2375-0197. Nov. 2017, S. 543–550. DOI: 10.1109/ICTAI.2017.00088.
- [2] Abdulrahman Alatawi, Weifeng Xu und Jie Yan. „The Expansion of Source Code Abbreviations Using a Language Model“. In: *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). Bd. 02. ISSN: 0730-3157. Juli 2018, S. 370–375. DOI: 10.1109/COMPSAC.2018.10260.
- [3] Thorsten Brants und Alex Franz. *LD Consortium*. type: dataset. 2006. DOI: 10.35111/CQPA-A498. URL: <https://catalog.ldc.upenn.edu/LDC2006T13> (besucht am 27.03.2021).

Literatur II

- [4] Anna Corazza, Sergio Di Martino und Valerio Maggio. „LINSEN: An efficient approach to split identifiers and expand abbreviations“. In: *2012 28th IEEE International Conference on Software Maintenance (ICSM)*. 2012 28th IEEE International Conference on Software Maintenance (ICSM). ISSN: 1063-6773. Sep. 2012, S. 233–242. DOI: 10.1109/ICSM.2012.6405277.
- [5] V. Efstathiou, C. Chatzilenas und D. Spinellis. „Word Embeddings for the Software Engineering Domain“. In: *2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR)*. 2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR). ISSN: 2574-3864. Mai 2018, S. 38–41.
- [6] Facebook. *English word vectors · fastText*. 2021. URL: <https://fasttext.cc/index.html> (besucht am 19.03.2021).
- [7] *Github*. 2021. URL: <https://github.com> (besucht am 06.04.2021).

- [8] Google. *Google Ngram Viewer*. 2021. URL: <http://storage.googleapis.com/books/ngrams/books/datasetv2.html> (besucht am 19. 03. 2021).
- [9] Dawn Lawrie und Dave Binkley. „Expanding identifiers to normalize source code vocabulary“. In: *2011 27th IEEE International Conference on Software Maintenance (ICSM)*. 2011 27th IEEE International Conference on Software Maintenance (ICSM). ISSN: 1063-6773. Sep. 2011, S. 113–122. DOI: 10.1109/ICSM.2011.6080778.
- [10] Yue Liu u. a. „Exploiting Task-Oriented Resources to Learn Word Embeddings for Clinical Abbreviation Expansion“. In: *Proceedings of BioNLP 15 (2015)*, S. 92–97. DOI: 10.18653/v1/W15-3810. arXiv: 1804.04225. URL: <http://arxiv.org/abs/1804.04225> (besucht am 24. 10. 2020).

- [11] Christian D. Newman u. a. „An Empirical Study of Abbreviations and Expansions in Software Artifacts“. In: *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. 2019 IEEE International Conference on Software Maintenance and Evolution (ICSME). ISSN: 2576-3148. Sep. 2019, S. 269–279. DOI: 10.1109/ICSME.2019.00040.
- [12] Weifeng Xu u. a. „Statistical Unigram Analysis for Source Code Repository“. In: *International Journal of Semantic Computing* (4. Juli 2018). Publisher: World Scientific Publishing Company. DOI: 10.1142/S1793351X18400123. URL: <https://www.worldscientific.com/doi/pdf/10.1142/S1793351X18400123> (besucht am 18. 10. 2020).

Alatawi et al.: Imitation gegen Original

- Verfahren von Alatawi et al. wurde als Referenz und Grundlage der hier vorgestellten Verfahren nachimplementiert
- Nachimplementierte Variante entspricht nicht ganz dem Original

Tabelle: Präzision des Bigramm-basierten Verfahrens von Alatawi et al., in Original und Imitation, unter Verwendung von Bigramm-Grammatik aus Software-basierten Quellen, auf dem Beispielquelltext von Alatawi et al.

Implementierung	Präzision
Original	0,7917
Imitation	0,6703

Alatawi et al.: Imitation gegen Original

			Präzision
Unigramm	NLR	Original	0,7083
		Imitation	0,5714
	SCR	Original	0,7083
		Imitation	0,5934
Bigramm	NLR	Original	0,7083
		Imitation	0,5384
	SCR	Original	0,7917
		Imitation	0,6703

$$P(ZK|B) = \left(\prod_{i=1}^n P(A_i|K_i) \right)^{1/n} * P(ZK)/P(B) \quad (1)$$

$$P(A|K) = P_{Typ} * P_{TES} * P_D \quad (2)$$

$$P_{uni}(ZK) = \left(\prod_{i=1}^n P(K_i) \right)^{1/n} \quad (3)$$

$$P_{bi}(ZK) = (P(K_1) * [(P(K_1, K_2)/P(K_1)) * ... * (P(K_{n-1}, K_n)/P(K_{n-1}))])^{1/n} \quad (4)$$

Beispiel Quelltexte und Musterlösungen

Musterlösung	ganzer Quelltext	alle Abkürzungen	alle Bezeichner	Ort im Quelltext
<i>Alatawi et al.</i> ¹	✓	✗	✗	✓
<i>Alatawi et al.</i> ²	✓	✗	✗	✗
<i>a2ps</i>	✓	✗	✗	✗
<i>which</i>	✓	✓	✗	✗
<i>Dronology</i> ¹	✓	✓	✗	✗
<i>Dronology</i> ²	✗	✓	✗	✓
<i>Dronology</i> ³	✗	✓	✓	✓

Literatur

Literatur

 Alatawi et al.
 ○○

 Datensätze
 ●○

 Evaluation
 ○○○○○○○○○○○○○○○○○

N-Gramm-Grammatiken

Alatawi et al. stellen die Software-basierter Uni- und Bigramm-Grammatiken [12] und die natürlichsprachlichen Uni- und Bigramm-Grammatiken [3], welche sie für die Evaluation ihrer eigenen Verfahren verwendet haben, online zur Verfügung. Die aus Software-basierten Quellen gewonnenen N-Gramm-Grammatiken tragen den Namen „Source Code Repository“, kurz „SCR“, und die Quelle aus der sie gewonnen wurden sind 0,7 Millionen Software-Projekte, die auf Github [7] veröffentlicht wurden. Die aus natürlicher Sprache stammenden N-Gramm-Grammatiken, unter dem Namen „Natural Language Repository“ oder „NLR“, wurden aus einem Terabyte online verfügbaren Textdaten [3], wie Bücher und Nachrichtenartikel, gewonnen.

Trigramm-basierte Abkürzungsauflösung

Trigramm-basiertes Verfahren im Direktvergleich mit Bigramm-basiertem Verfahren. Es wurde jeweils die Software-basierte Uni- und Bigramm-Grammatik verwendet. Die Trigramm-Grammatik stammt aus dem Google N-Gramm-Datensatz.

Quelltext	Verfahren	Präzision (Bezeichner)	Präzision (Token)
Alatawi et al.	Trigramm	0,7033	0,7873
Alatawi et al.	Bigramm	0,6703	0,7828
<i>Dronology</i>	Trigramm	0,0511	0,0671
<i>Dronology</i>	Bigramm	0,0511	0,0671
<i>a2ps</i>	Trigramm	0,3673	0,2537
<i>a2ps</i>	Bigramm	0,3678	0,2544
<i>which</i>	Trigramm	0,3916	0,2660
<i>which</i>	Bigramm	0,3938	0,2657

Literatur	Literatur	Alatawi et al.	Datensätze	Evaluation
		ooo	oo	●○○○○○○○○○○○○○○

Trigramm-basierte Abkürzungsauflösung

Trigramm-, Bigramm- und Unigramm-basiertes Verfahren auf dem Beispiel Quelltext von Alatawi et al. Die verwendete Uni-, Bi- und Trigramm-Grammatik stammt jeweils aus dem Google N-Gramm-Datensatz.

Verfahren	Präzision (Bezeichner)	Präzision (Token)
Trigramm	0,5055	0,7104
Bigramm	0,5055	0,6833
Unigramm	0,5385	0,7195

Worteinbettungsbasierte Ähnlichkeitsfunktionen

word2vec-basiertes Verfahren, evaluiert auf dem Beispielquelltext von Alatawi et al., unter Verwendung des *SO_200*-Worteinbettungsmodells und verschiedenen Ähnlichkeitsfunktionen.

Ähnlichkeitsfunktion	Präzision (Bezeichner)	Präzision (Token)
Naiv-Brachial	0,0879	0,2353
Kontext-Brachial	0,0879	0,2398
Kontext-Wortarten	0,1099	0,2534
Kontext-Aufgeteilt	0,1429	0,3394
Kombinierte-Vektoren	0,1538	0,3394

Worteinbettungsbasierte Ähnlichkeitsfunktionen

word2vec-basiertes Verfahren, evaluiert auf verschiedenen Beispielquelltexten, unter Verwendung des *SO_200*-Worteinbettungsmodells und verschiedenen Ähnlichkeitsfunktionen.

Quelltext	Ähnlichkeitsfunktion	Präzision (Bezeichner)	Präzision (Token)
<i>Dronology</i>	Kontext-Brachial	0,0431	0,0617
<i>Dronology</i>	Kontext-Aufgeteilt	0,0461	0,0603
<i>Dronology</i>	Kombinierte-Vektoren	0,0511	0,0651
<i>a2ps</i>	Kontext-Brachial	0,3651	0,2509
<i>a2ps</i>	Kontext-Aufgeteilt	0,3646	0,2505
<i>a2ps</i>	Kombinierte-Vektoren	0,3629	0,2470
<i>which</i>	Kontext-Brachial	0,3934	0,2660
<i>which</i>	Kontext-Aufgeteilt	0,3934	0,2681
<i>which</i>	Kombinierte-Vektoren	0,3952	0,2675

Literatur

Literatur

 Alatawi et al.
 ○○○

 Datensätze
 ○○

 Evaluation
 ○○○○●○○○○○○○○

Worteinbettungsverfahren

Vergleich zwischen dem *fastText*-basierten Verfahren, unter Einsatz des [Common Crawl] Modells, und dem *word2vec*-basierten Verfahren, unter Einsatz des *SO_200*-Worteinbettungsmodells, auf dem Beispielquelltext von Alatawi et al.

Ähnlichkeitsfunktion	Verfahren	Präzision (Bezeichner)	Präzision (Token)
Kontext-Brachial	<i>word2vec</i>	0,0879	0,2398
Kontext-Brachial	<i>fastText</i>	0,1099	0,3122
Kontext-Aufgeteilt	<i>word2vec</i>	0,1429	0,3394
Kontext-Aufgeteilt	<i>fastText</i>	0,1209	0,3258
Kombinierte-Vektoren	<i>word2vec</i>	0,1538	0,3394
Kombinierte-Vektoren	<i>fastText</i>	0,1319	0,3575

Literatur

Literatur

Alatawi et al.

○○○

Datensätze

○○

Evaluation

○○○○○●○○○○○○○○

Worteinbettungsverfahren

Vergleich zwischen dem *fastText*-basierten Verfahren und dem *word2vec*-basierten Verfahren, mit Worteinbettungsmodellen, welche auf dem Beispielquelltext selbst trainiert wurden. Die Ähnlichkeitsfunktion ist jedem Fall *Kombinierte-Vektoren*.

Quelltext	Verfahren	Präzision (Bezeichner)	Präzision (Token)
Alatawi et al.	<i>word2vec</i>	0,1099	0,3077
Alatawi et al.	<i>fastText</i>	0,1319	0,3575
<i>Dronology</i>	<i>word2vec</i>	0,0511	0,0632
<i>Dronology</i>	<i>fastText</i>	0,0511	0,0767
<i>a2ps</i>	<i>word2vec</i>	0,3619	0,2446
<i>a2ps</i>	<i>fastText</i>	0,3608	0,2477
<i>which</i>	<i>word2vec</i>	0,3943	0,2663
<i>which</i>	<i>fastText</i>	0,3934	0,2669

Literatur

Literatur

Alatawi et al.
oooDatensätze
ooEvaluation
oooooooo●oooooooo

Worteinbettungsbasierte Abkürzungsauflösung

word2vec-basiertes Verfahren, mit *Kombinierte-Vektoren*-Ähnlichkeitsfunktion und *SO_200*-Worteinbettungsmodell, im Vergleich mit dem Bigramm-basierten Verfahren von Alatawi et al., unter Nutzung der Software-basierten Uni- und Bigramm-Grammatiken.

Quelltext	Verfahren	Präzision (Bezeichner)	Präzision (Token)
Alatawi et al.	<i>word2vec</i>	0,1538	0,3394
Alatawi et al.	Bigramm	0,6703	0,7828
<i>Dronology</i>	<i>word2vec</i>	0,0511	0,0651
<i>Dronology</i>	Bigramm	0,0511	0,0671
<i>a2ps</i>	<i>word2vec</i>	0,3629	0,2470
<i>a2ps</i>	Bigramm	0,3678	0,2544
<i>which</i>	<i>word2vec</i>	0,3952	0,2675
<i>which</i>	Bigramm	0,3938	0,2657
Literatur	Literatur	Alatawi et al. ooo	Evaluation oooooooo●oooooo

Worteinbettungsbasierte Mehrdeutigkeitsauflösung

word2vec-basiertes Verfahren, unter Nutzung der *Kombinierte-Vektoren*-Ähnlichkeitsfunktion und des *SO_200*-Worteinbettungsmodells, zur Mehrdeutigkeitsauflösung des Unigramm-basierten Verfahrens von Alatawi et al., unter Nutzung der Software-basierten Unigramm-Grammatik, im Vergleich zum gleichen Verfahren mit willkürlicher Mehrdeutigkeitsauflösung.

Quelltext	Mehrdeutigkeitsauflösung	Präzision (Bezeichner)	Präzision (Token)
Alatawi et al.	<i>word2vec</i>	0,5824	0,7330
Alatawi et al.	willkürlich	0,5934	0,7376
<i>Dronology</i>	<i>word2vec</i>	0,0291	0,0535
<i>Dronology</i>	willkürlich	0,0511	0,0671
<i>a2ps</i>	<i>word2vec</i>	0,3678	0,2561
<i>a2ps</i>	willkürlich	0,3678	0,2556
<i>which</i>	<i>word2vec</i>	0,3943	0,2687
Literatur <i>which</i>	Literatur willkürlich	Alatawi et al. ○○○	Evaluation ○○○○○○○○●○○○○
		0,3938	0,2684

Mehrdeutigkeit

Bei Betrachtung aller Vorkommnisse der 80 verschiedenen abgekürzter Bezeichner aus der Musterlösung für den Beispielquelltext Alatawi et al., haben von 659 Vorkommnissen 68 mehrere gleichwertig beste Auflösungskandidaten, und zwar bis zu 76 verschiedene. Im Schnitt haben die Vorkommnisse der abgekürzten Bezeichner aus der Musterlösung 2,73 verschiedene gleichwertig beste Auflösungskandidaten. Beschränkt auf die Vorkommnisse abgekürzter Bezeichner, welche tatsächlich mehrere gleichwertig beste Auflösungskandidaten haben, sind es im Schnitt 16,69. Der Bezeichner mit 76 verschiedenen gleichwertig besten ist „it“.

Bezogen auf alle Quelltextbezeichner aus dem Beispielquelltext von Alatawi et al., bei denen von 16.565 Bezeichnern, mit im Schnitt 3,92 verschiedenen gleichwertig besten Auflösungskandidaten, 2.651 Bezeichner mehrere gleichwertig beste Auflösungskandidaten haben, und zwar im Schnitt 17,37 verschiedene, ist der Anteil und die Qualität an Bezeichnern mit mehreren Auflösungskandidaten im Schnitt durchaus vergleichbar. Allerdings hat aus dieser Menge der Bezeichner mit den meisten verschiedenen Auflösungskandidaten, nämlich „set“, 692 verschiedenen Auflösungskandidaten.

Literatur

Literatur

Alatawi et al.
○○○

Datensätze
○○

Evaluation
○○○○○○○○●○○○

Segmente je abgekürzten Bezeichner

Auflösungen aus der *a2ps*-Musterlösung enthalten im Schnitt 1,33 Token. Die Auflösungen in der Musterlösung für *Dronology* enthalten im Schnitt 2,52 Token. Es kommen bei *Dronology* also im Schnitt in einer Auflösung mehr Token vor, von denen nur eines falsch aufgelöst zu werden braucht, damit der gesamte Bezeichner falsch aufgelöst wird. Darüber hinaus werden bei *Dronology* aber auch die einzelnen Token seltener richtig aufgelöst.

Literatur

Literatur

Alatawi et al.

ooo

Datensätze

oo

Evaluation

ooooooooo●ooo

Drohnen-Steuerungs-Programme

Drohnen-Steuerungs-Programme, welche für das Training eines Wordembeddingmodells verwendet wurden.

- *PX4-Autopilot*
- *Paparazzi*
- *ArduPilot*
- *LibrePilot*
- *Flone*
- *ODM*
- *DronePan*
- *Dronology*

Literatur

Literatur

Alatawi et al.
○○○

Datensätze
○○

Evaluation
○○○○○○○○○○●○○

Normalize

Normalize [9] löst 53% aller Bezeichner im Probanden-Programm *which* vollständig richtig auf, in *a2ps* nur 32%.

Literatur

Literatur

Alatawi et al.
○○○

Datensätze
○○

Evaluation
○○○○○○○○○○○○○○●○

Die Genauigkeit der Auflösung von *LINSEN* [4] betrug zwischen 56,6% auf *a2ps* und 56,7% auf *which*, bezogen auf ganze Bezeichner, beziehungsweise zwischen 71,3% auf *a2ps* und 83,5% auf *which*, bezogen auf die einzelnen weichen Wörter, beides anhand der Musterlösung von Lawrie et al. evaluiert.