



we
focus
on
students



ABP 2017

Wo steckt der Fehler in der SQL-Anfrage?
Semantische Prüfung von Lösungen

1	Problemstellung	2
2	Das SQLearn-System	4
3	Semantische Prüfung	9
4	Erste Erfahrungen	15
5	Ausblick und Zusammenfassung	18

Aufgabe

Listen Sie alle Privatkunden aus Dortmund mit Kundennummer, Vorname, Nachname sortiert nach dem Nachnamen.



SQL-Anfrage



Rückmeldung:

Anfrageergebnis(0 Tupel):		
kundennummer	vorname	nachname



Wo liegt der Fehler?

1	Problemstellung	2
2	Das SQLearn-System	4
3	Semantische Prüfung	9
4	Erste Erfahrungen	15
5	Ausblick und Zusammenfassung	18

DB1

- 5. Buchhandlungsmodell
- (V) 7. Vorlesungsbeispiele
 - Aufgabe 1 (#1)
 - Aufgabe 2 (#4)
 - Aufgabe 3 (#100)
 - Aufgabe 4 (#101)
 - Aufgabe 5 (#102)
 - Aufgabe 6 (#103)
 - Aufgabe 7 (#105)
 - Aufgabe 8 (#106)
 - Aufgabe 9 (#107)
- 6. Transfumed I
- 7. Transfumed II
- 8. FH-Modell

53 Aufgaben mit
89 Musterlösungen

Aufgabe #4

Anforderungsniveau: 2
Anforderungslevel: 2
★★★★★
Datenbank: MySQL-Buchhandlung

Aufgabenstellung

Listen Sie alle Privatkunden aus Dortmund mit Kundennummer, Nachname und Vorname sortiert nach dem Nachnamen.

Schema und DBMS-Auswahl

Eingabe:

```
select kundennummer, vorname, nachname from kunde  
where Anrede="Frau" AND Anrede="Herr" And Ort="Dortmund"
```

Ausführen
Syntax-Hilfe

Eingabebereich

Letzte Eingabe: `select kundennummer, vorname, nachname from kunde
where Anrede="Frau" AND Anrede="Herr" And Ort="Dortmund"`

Anfrageergebnis(0 Tupel):	Ergebnis der Lösung(1 Tupel):
kundennummer vorname nachname	kundennummer vorname nachname
	8524 Max Meier

Ergebnismengenvergleich

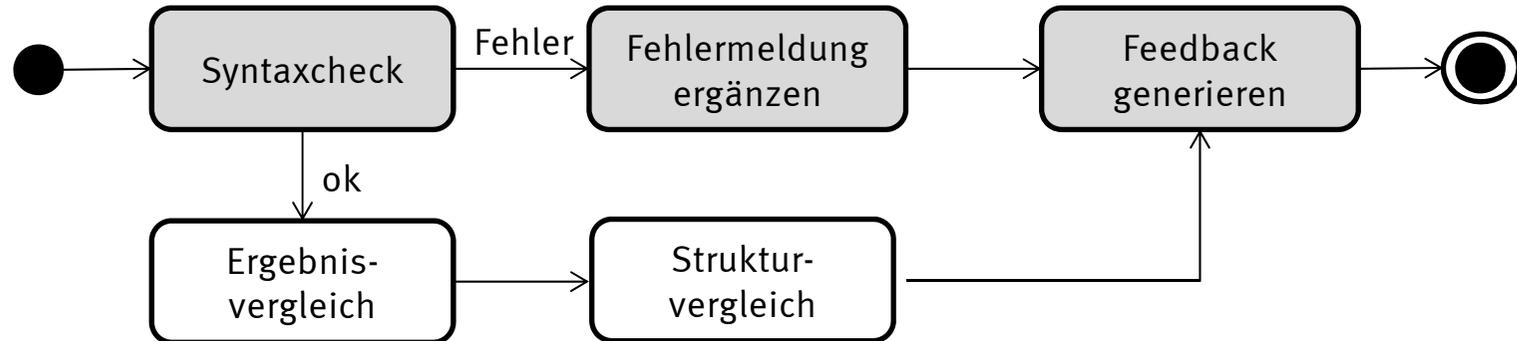
Hinweise:

- Schemaausschnitt**
KUNDE [Kundennummer, Anrede, Nachname, Vorname, Geburtsdatum, Ort]
Tipp: Nur Privatkunden besitzen die Anrede Herr oder Frau.
- Lösungsstatus**
Analyse: Das Ausführungsergebnis stimmt nicht mit der Lösung überein
Hinweis: Überprüfen Sie bitte die Where-Klausel (.=. OR .=. AND .=.)
Hinweis: Überprüfen Sie bitte die Order-By-Klausel.

Lösungshinweise

Feedback geben

Anfragebearbeitung - Syntaxprüfung



Lösung:

Select kundennummer , anrede, vorname from Kunden

Eingabe:

Select Kundennummer, anrede, vorname from Kunden

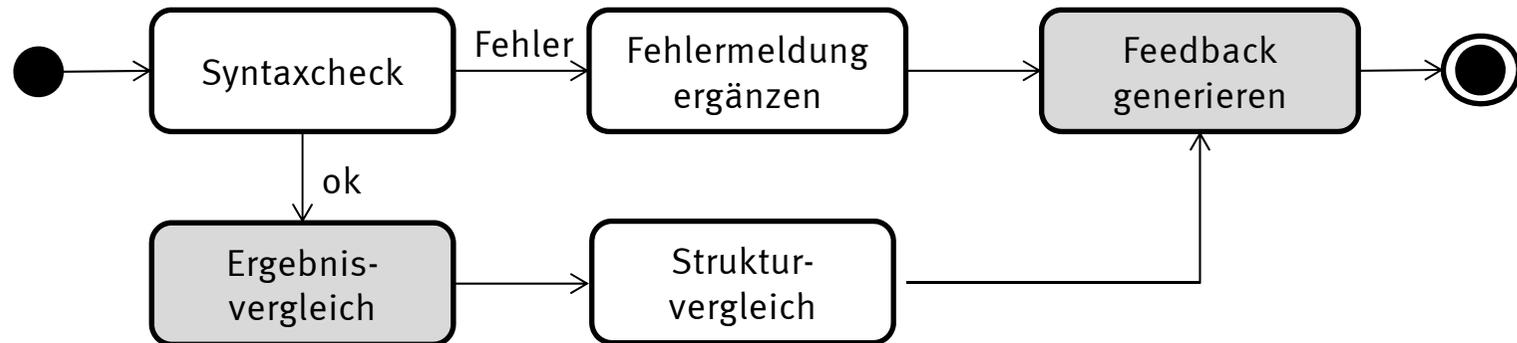
Rückmeldung:

Hinweise:

- Schemaausschnitt
KUNDEN [Kundennummer, Anrede, Name, Vorname, Geburtsdatum, Ort]
- Lösungsstatus
FEHLER: Unknown column 'Kundennnummer' in 'field list'
d.h. Ein angefragtes Spaltenattribut ist in der Tabelle nicht vorhanden. Überprüfen Sie bitte die FROM-Klausel.

ergänzte MySQL-Fehlermeldung

Anfragebearbeitung - Ergebnisvergleich



Lösung:

Eingabe:

Rückmeldung:

Select kundennummer , anrede, vorname from Kunden

Select *

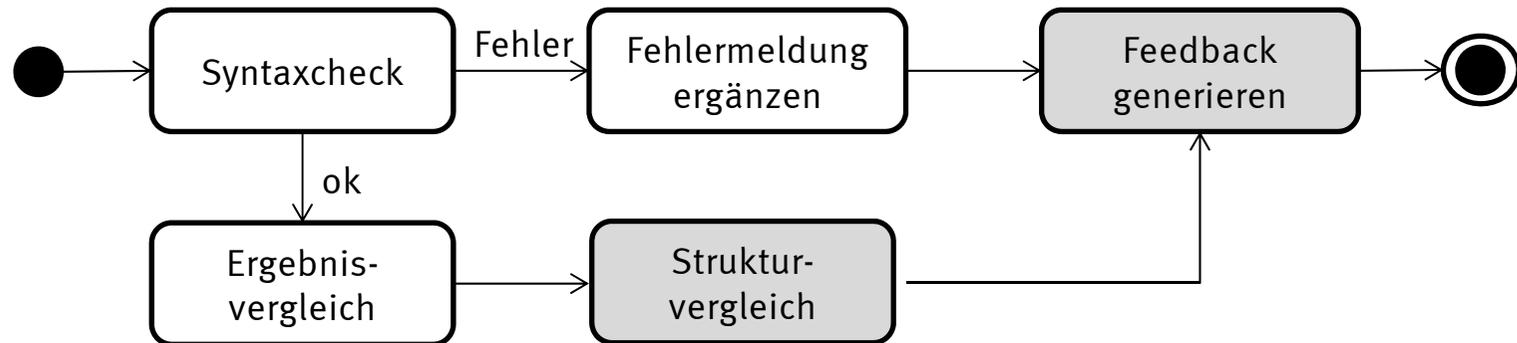
from Kunden

Anfrageergebnis(5 Tupel):						Ergebnis der Lösung(5 Tupel):		
Kundennummer	Anrede	Name	Vorname	Geburtsdatum	Ort	kundennummer	anrede	vorname
2310	Frau	Meitner	Lise	17.11.1878 00:00:00	Berlin	2310	Frau	Lise
7562	Herr	Einstein	Albert	14.03.1879 00:00:00	Princeton	7562	Herr	Albert
8365	Frau	Curie	Marie	07.11.1867 00:00:00	Paris	8365	Frau	Marie
8523		Dekanat Informatik	FH Dortmund		Dortmund	8523		FH Dortmund
8524	Herr	Meier	Max	24.12.1987 00:00:00	Dortmund	8524	Herr	Max

Hinweise:

- Schemaausschnitt**
KUNDEN [Kundennummer, Anrede, Name, Vorname, Geburtsdatum, Ort]
- Lösungsstatus**
Analyse: Das Ausführungsergebnis stimmt nicht mit der Lösung überein
Empfehlung: Überprüfen Sie bitte die Spaltenattribute.

Anfragebearbeitung – Strukturvergleich



Lösung: ... Where (Anrede = 'Herr' OR Anrede = 'Frau') AND Ort='Dortmund' ...
Eingabe: ... Where (Anrede = 'Herr' AND Anrede = 'Frau') AND Ort='Dortmund'
Rückmeldung:

Anfrageergebnis(0 Tupel):			Ergebnis der Lösung(1 Tupel):		
kundennummer	vorname	nachname	kundennummer	vorname	nachname
			8524	Max	Meier

Ergebnismengenvergleich

Hinweise:

- Schemaausschnitt**
KUNDE [Kundennummer, Anrede, Nachname, Vorname, Geburtsdatum, Ort]
Tipp: Nur Privatkunden besitzen die Anrede Herr oder Frau.
- Lösungsstatus**
Analyse: Das Ausführungsergebnis stimmt nicht mit der Lösung überein
Hinweis: Überprüfen Sie bitte die Where-Klausel (., OR ., AND .,.)
Hinweis: Überprüfen Sie bitte die Order-By-Klausel.

Lösungshinweise

1	Problemstellung	2
2	Das SQLearn-System	4
3	Semantische Prüfung	9
4	Erste Erfahrungen	15
5	Ausblick und Zusammenfassung	18

1. Syntaktische Prüfung durch Ausführung der SQL-Anfrage gegen ein DBMS
Aber: Abhängigkeit vom SQL-Dialekt

2. Semantische Prüfung der SQL-Anfrage
 - a. Eine a-priori-Erfassung aller Musterlösungen ist nicht möglich.
 - b. Heuristischer Ergebnismengenvergleich gegenüber (generierten) Datenbeständen (z.B. Chandra et. al. 2015)

Anfrageergebnis(0 Tupel):	Ergebnis der Lösung(1 Tupel):		
kundennummer vorname nachname	kundennummer	vorname	nachname
	8524	Max	Meier

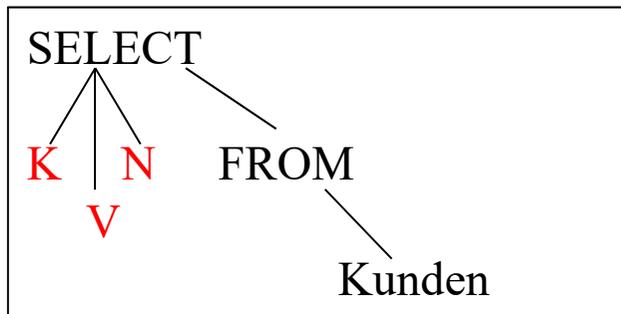
- c. Strukturvergleich durch Zerlegung der SQL-Anfrage in einen Parserbaum und dessen Vergleich mit dem Parserbaum der Musterlösung (z.B. Ifland et. al. 2014, Dollinger 2011)

Strukturvergleich von SQL-Anfragen

Lösungsansatz ÜPS (Ifland 2014):

- Knoten des Parserbaums entsprechend den SQL-Klauseln
- Nutzung der Levenshtein-Distanz zum Vergleich der Parserbäume

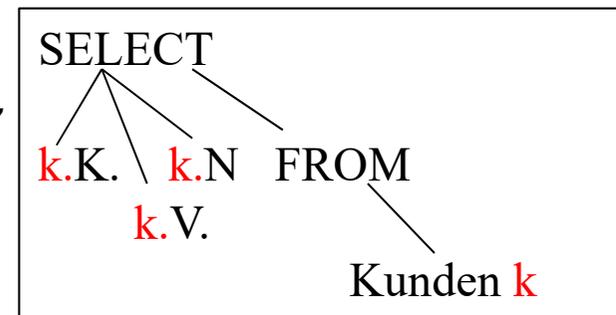
Anfrage



SELECT Kundennummer, Vorname,
Nachname
FROM Kunden

L.-Distanz 7
↔

Musterlösung



SELECT k.Kundennummer,
k.Vorname, k.Nachname
FROM Kunden k

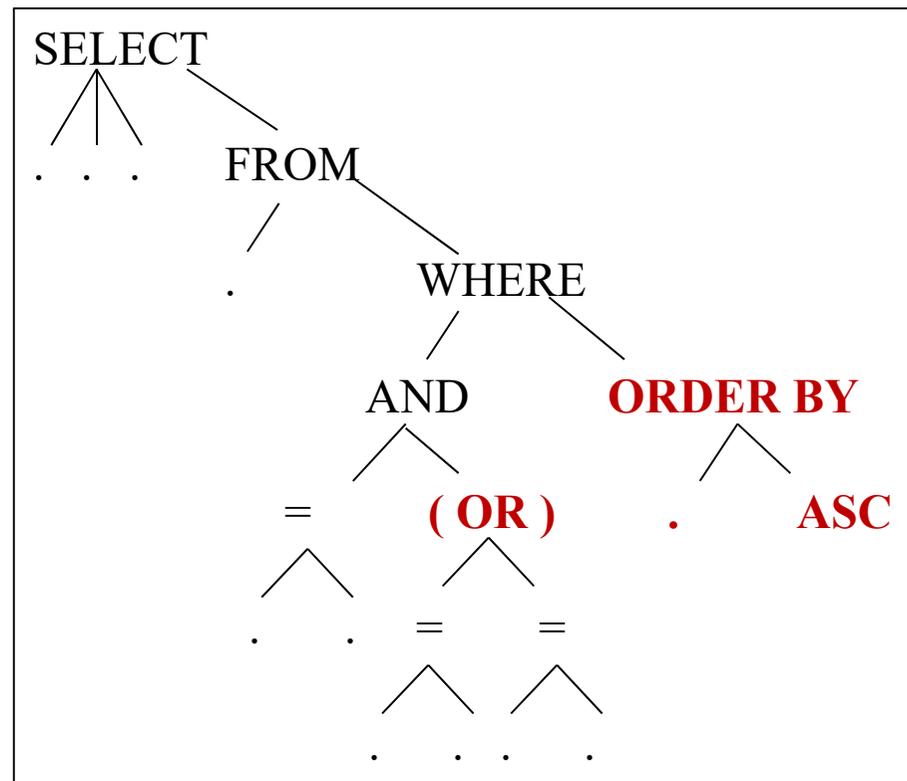
Lösung: Select Kundennummer, Vorname, Nachname FROM Kunden
WHERE (Anrede='Herr' **OR** Anrede='Frau') AND Ort='Dortmund'
ORDER BY Nachname

Anfrage: Select Kundennummer, Vorname, Nachname FROM Kunden
WHERE (Anrede='Herr' **AND** Anrede='Frau') AND Ort='Dortmund'

Vergleich der SQL-Anfragen:

- Aufbau Parserbaum
- Anwendung von log. Umformungen und Entfernen von Alias
- Verschachtelung von Anfragen möglich (pattern matching)

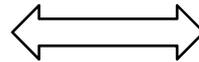
→ Heuristische Distanzfunktion zählt strukturelle und inhaltliche Unterschiede



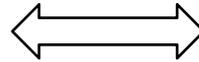
Anfrage

```
SELECT ... FROM Kunden k
WHERE (k.Anrede = 'Herr' AND
      k. Anrede='Frau')
AND k.Ort='Dortmund'
```

Distanz 1



Distanz 4



Musterlösungen

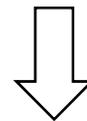
```
SELECT ... FROM Kunden WHERE
      (Anrede = 'Herr' OR Anrede = 'Frau')
AND Ort='Dortmund'
```

```
SELECT ... FROM Kunden WHERE
WHERE Anrede IS NOT NULL
AND Ort='Dortmund'
```

...

Musterlösungspool

- Berücksichtigung von SQL-Dialekten (z.B. LIMIT-Klausel)
- Äquivalente Literale (z.B. Count(*) vs. Count(Kundennummer))
- strukturell unterschiedlichen Lösungen (z.B. Gruppierung vs. Subquery)



Generierung struktureller Lösungshinweise zu der Musterlösung,
die den geringsten Abstand zur geprüften Anfrage besitzt

Hinweis: Überprüfen Sie bitte die Where-Klausel (.=, OR, .=, AND, .=)

1	Problemstellung	2
2	Das SQLearn-System	4
3	Semantische Prüfung	9
4	Erste Erfahrungen	15
5	Ausblick und Zusammenfassung	18

Beispiel 1:

Uhrzeit	Eingabe
00:59:54	select from Kunden
01:03:15	select Anrede from Kunden;
01:03:37	select Anrede, Vorname from Kunden;
01:05:03	select Anrede, Vorname from Kunden;
01:05:29	select Kundennummer , Anrede, Vorname from Kunden;

Beispiel 2:

Uhrzeit	Eingabe
16:51:06	SELECT Vorname, Nachname FROM Kunde WHERE Vorname LIKE M%
16:52:07	SELECT Vorname, Nachname FROM Kunde WHERE Vorname LIKE 'M%'
16:52:54	SELECT count(Vorname) FROM Kunde WHERE Vorname LIKE 'M%'
16:53:21	SELECT count(kundennummer) FROM Kunde WHERE Vorname LIKE 'M%'
16:53:45	SELECT count(kundennummer) FROM Kunde WHERE nachname LIKE 'M%'

1	Problemstellung	2
2	Das SQLearn-System	4
3	Semantische Prüfung	9
4	Erste Erfahrungen	15
5	Ausblick und Zusammenfassung	18

- ▶ Automatische Bewertung von SQL-Anfragen

- ▶ Vergrößerung des Musterlösungspools:
 - Filterung der Nutzereingaben nach semantisch korrekten Lösungen
 - Generierung weiterer Musterlösung
(z.B. durch Transformation vorhandener Lösungen)

- ▶ Optimierung der verwendeten heuristische Distanzfunktion

- ▶ Ausweitung der Funktionalität des SQLearn-Systems um
 - DML-Operationen
 - LINQ-Anfragen
 - SQL-Tuning

Aufgabe

Listen Sie alle Privatkunden aus Dortmund mit Kundennummer, Vorname, Nachname sortiert nach dem Nachnamen.



```
Select Kundennummer, Vorname, Nachname FROM Kunden  
WHERE (Anrede='Herr' AND Anrede='Frau') AND Ort='Dortmund'
```



Rückmeldung:

Anfrageergebnis(0 Tupel):			Ergebnis der Lösung(1 Tupel):		
kundennummer	vorname	nachname	kundennummer	vorname	nachname
			8524	Max	Meier

Hinweise:

- Schemaausschnitt**
KUNDE [Kundennummer, Anrede, Nachname, Vorname, Geburtsdatum, Ort]
Tipp: Nur Privatkunden besitzen die Anrede Herr oder Frau.
- Lösungsstatus**
Analyse: Das Ausführungsergebnis stimmt nicht mit der Lösung überein
Hinweis: Überprüfen Sie bitte die Where-Klausel (.= OR .= AND .=)
Hinweis: Überprüfen Sie bitte die Order-By-Klausel.



Da liegt der
semantische Fehler!

**Vielen Dank
für Ihre Aufmerksamkeit**