

Übung Informationsintegration

Modellierung von Datenquellen

Aufgabe 1 (GaV mit Nebenbedingungen): Gegeben seien folgendes globales Schema:

Student(MNr, Name, Semester, Studiengang)

Nebenbedingung: Semester > 5

und folgende Quellschemata:

S1(MNr, Name, Semester)

Nebenbedingung: Studiengang = 'ITMC'

S2(MNr, Name, Semester, Studiengang)

Nebenbedingung: Semester < 8

S3(MNr, Name, Semester)

Nebenbedingung: Semester > 12

S4(MNr, Name, Studiengang)

Nebenbedingung: keine

S5(MNr, Name, Semester)

Nebenbedingung: Studiengang = 'ITMC' oder Studiengang = 'MCI'

- Geben Sie ein GaV Mapping für die globale Relation Student in Form eines CREATE VIEW Statements an welches möglichst viele Quellen berücksichtigt.
- Beschreiben Sie etwaige Probleme die sich bei der Nutzung der Quellen ergeben.

Aufgabe 2 (SQL → Datalog): Gegeben sei folgendes Relationenschema:

Film(Titel, Regisseur, Genre, Jahr)

Person(PID, Name, Alter, Ort)

Spielt(Film, Person, Rolle) mit den beiden FKs Film → Film.Titel und Person → Person.PID

Gegeben seien zudem folgende zwei SQL Anfragen:

Anfrage 1:

```
SELECT p1.Name AS N, f.Titel AS T
FROM Film f, Person p1, Person p2, Spielt s1, Spielt s2
WHERE f.Titel = s1.Film
AND f.Titel = s2.Film
AND p1.PID = s1.Person
AND p2.PID = s2.Person
AND p2.Ort = 'Hamburg'
AND f.Jahr = 2015
```

Anfrage 2:

```
SELECT p2.Name AS N
FROM Film f, Person p1, Person p2, Spielt s
WHERE f.Regisseur = p1.PID
AND (p1.Name = 'S. Leone' OR p1.Name = 'S. Kubrik')
AND f.Titel = s.Film
AND p2.PID = s.Person
```

Formulieren Sie die Anfragen jeweils in einen äquivalenten Datalog Ausdruck um.

Aufgabe 3 (Mappings): Das globale Schema ist wie folgt gegeben:

Buch(BuchID, ISBN10, ISBN13, Titel, AnzSeiten, Autor, Verlag, Genre)
Preis(BuchID, Waehrung, NeuPreis) mit dem FK BuchID → Buch.BuchID
Bewertung(BID, BuchID, AnzSterne, Nutzer, Quelle) mit dem FK BuchID → Buch.BuchID

Gegeben seien zudem die beiden Quellen $S1$ (Datenbank eines Buchhändlers) und $S2$ (Datenbank des Internetdiensts 'Bookscrobler'). Das Schema von $S1$ umfasst vier Relationen und ist wie folgt definiert:

Buch(ISBN, Titel, Datum, UVP, Genre, Format)
Autor(ID, Name)
Verlag(ID, Name, Adresse)
Autorenschaft(Autor, Buch) mit den beiden FKs Autor → Autor.ID und Buch → Buch.ISBN

Für die Relation Buch ist zudem folgende Beispielinstantz gegeben:

<u>ISBN</u>	Titel	Datum	UVP	Genre	Format
3551551677	Harry Potter 1	1998	15,90	Fantasy	Gebunden
3551551685	Harry Potter 2	1999	15,90	Fantasy	Gebunden
3551551693	Harry Potter 3	1999	16,90	Fantasy	Gebunden
978-3453146976	Per Anhalter durch die Galaxis	2009	8,99	SciFi	Taschenbuch

Das Schema von $S2$ umfasst drei Relationen und ist wie folgt definiert:

Book(BID, ISBN13, Title, Year, Author)
User(ID, Name, Firstname)
Read(UID, BID, Rating) mit den beiden FKs UID → User.ID und BID → Book.BID

- Ermitteln Sie Korrespondenzen zwischen den einzelnen Quellschemata und dem globalen Schema
- Versuchen Sie ein GaV Mapping zwischen den Quellschemata und dem globalen Schema in Form von CREATE VIEW Statements zu erstellen
- Versuchen Sie ein LaV Mapping zwischen den Quellschemata und dem globalen Schema in Form von CREATE VIEW Statements zu erstellen

Aufgabe 4 (Ausführbarer Anfrageplan): Gegeben seien folgende Relationen (inkl. eines Adornments pro Relation):

Film^{bffb}(Titel, Regisseur, Genre, Jahr)
Person^{bfff}(PID, Name, Alter, Ort)
Spielt^{bff}(Film, Person, Rolle)
FilmByGenre^{fb}(Titel, Genre)

Gegeben Sei zudem folgende Datalog Anfrage:

q(N) :- Person(PID, N, A, O), Spielt(T, PID, R), Film(T, Reg, G, J), G='Krimi', J=2014

- a) Verwenden Sie den in der Vorlesung behandelten Algorithmus um festzustellen, ob diese Anfrage einen ausführbaren Plan besitzt.
- b) Ändert sich das Ergebnis aus Aufgabenteil a) wenn Sie die Relation 'FilmByGenre' hinzunehmen (d.h. folgend aufgeführte Anfrage betrachten)?

q(N) :- Person(PID, N, A, O), Spielt(T, PID, R), Film(T, Reg, G, J),
FilmByGenre(T, G), G='Krimi', J=2014