

Aufgabenblatt 3 GaV & LaV

- Letzter Abgabetermin: **Freitag, 22. Juni 2007**
- Das Aufgabenblatt gilt als bestanden, wenn beide Aufgaben bearbeitet und mindestens 20 der 40 Punkte erreicht werden.
- Die Aufgaben sollen in Zweiergruppen bearbeitet werden.
- Abgabe: Per E-Mail an Alexander Albrecht, Fachgebiet Informationssysteme.

Aufgabe 1: GaV

In dieser Aufgabe wird mit Hilfe des IBM Federation Servers ein integriertes Filminformationssystem mit folgendem Schema entwickelt:

```
MOVIE( MOVIEID, TITLE, YEAR, GENRE, DURATION )  
ACTOR(ACTORID, NAME, GENDER, INFO)  
ACTS ( ACTORID, MOVIEID, ROLE)
```

In das System sollen zum einen Flatfiles integriert werden, die Filme der letzten Jahre und Schauspieler getrennt nach Geschlecht aus der Internet Movie Database (www.imdb.com) umfassen. Zusätzlich sollen XML-Files, die ausführliche Schauspielerinformationen enthalten, sowie eine Exceldatei mit aktuellen Filmen integriert werden. In allen Datenquellen werden Filme und Schauspieler durch global eindeutige Identifier gekennzeichnet, um das Verknüpfen der Datenquellen und das Finden von Duplikaten zu vereinfachen.

- a) Installiere IBM DB2 Database Version 9 und den Websphere Federation Server lokal auf deinem Rechner. Bei Bedarf können auch Rechner unserer Fachgruppe genutzt werden. Detaillierte Informationen zur Installation und der Rechnernutzung findest du im Übungsverzeichnis. R:\InfoInt07_Naumann\Uebung\0614\.
- b) Analysiere die im Übungsverzeichnis vorhandenen Datenquellen und binde diese in DB2 als Objekte föderierter Datenbanken ein. Das genaue Vorgehen wird in der Übungsveranstaltung gezeigt. Zusätzlich gibt es im Übungsverzeichnis eine ausführliche Dokumentation. **5 P**
- c) Erstelle in DB2 das globale Schema des Filminformationssystems unter Verwendung der eingebundenen Datenquellen. Dokumentiere die verwendeten Sichten in SQL. **5 P**
- d) Beantworte die folgenden Anfragen in SQL. **6 P**
 - In welchem Film spielen die wenigsten/meisten Schauspieler mit?
 - Wieviel Star Trek Filme gibt es im Filminformationssystem?
 - In welchem Jahr hatte George Clooney am meisten zu tun?
- e) Beantworte die folgenden Anfragen in SQL. Füge deiner Lösung den jeweiligen Zugriffsplan hinzu. **9 P**
 - Zeige alle Schauspieler, die in allen James Bond Filmen der letzten 5 Jahre mitspielten.
 - Zeige alle Filmpartnerinnen von George Clooney.
 - Zeige alle Schauspieler die in mindestens zwei Filmen mit Sigourney Weaver zusammengearbeitet haben.

Aufgabe 2: LaV

Gegeben sei das folgende globale Schema:

```
FILM( TITEL, GENRE, REGISSEUR, JAHR )
SCHAUSPIELER(SCHAUSPIELER_NAME, NATIONALITÄT)
SPIELT( TITEL,SCHAUSPIELER_NAME,ROLLE,ZITATE)
```

Filmgenre seien unter anderem "ANIMATIONSFILM" und "WESTERN". Es stehen die folgenden Datenquellen zur Verfügung. Der Einfachheit halber nehmen wir an, dass jede Datenquelle nur eine Relation exportiert.

```
FILME_VOR_1970 (TITEL, GENRE)
FILME_NACH_1950 (TITEL, GENRE, SCHAUSPIELER_NAME)
WESTERN (TITEL, REGISSEUR, JAHR)
FILM_ZITATE(TITEL,JAHR,ZITATE)
ZITATE(TITEL,ROLLE,ZITATE)
SCHAUSPIELER(SCHAUSPIELER_NAME,TITEL,NATIONALITÄT)
```

- a) Gebe die LaV-Regeln an, die jede der Datenquellen als Sicht auf das globale Schema beschreiben. Verwende dafür die in der Vorlesung vorgestellte Datalog-Notation. **6 P**
- b) Aufgabe der Anfrageplanung ist es, für eine beliebige Anfrage an das globale Schema die geeigneten Datenquellen zu finden, die die Anfrage beantworten. Welche Quellen sollten zur Beantwortung der folgenden Anfragen verwendet werden? Begründe kurz deine Meinung. **3 P**
 - Suche nach Titeln von Animationsfilmen.
 - Suche nach Titeln von Filmen vor 1970.
 - Suche Schauspieler mit Hauptrollen in Westernfilmen.
- c) Gegeben sei die globale Anfrage p . Dabei muss man sich den Unterstrich in der Anfrage durch beliebige, aber unterschiedliche Variablen ersetzt denken.

$$p(T,G,R,Z) :- \text{film}(T, G, _ , _), \text{spielt}(T, _ , R, Z)$$

Begründe, unter Verwendung von Containment-Mappings, welche Datenquellen für p in Frage kommen. **6 P**