Einsatz und Realisierung von Datenbanksystemen

ERDB Übungsleitung

Maximilian {Bandle, Schüle}, Josef Schmeißer

i3erdb@in.tum.de

Folien erstellt von Maximilian Bandle & Alexander Beischl

Organisatorisches
Disclaimer

Die Folien werden von der Übungsleitung allen Tutoren zur Verfügung gestellt.

Sollte es Unstimmigkeiten zu den Vorlesungsfolien von Prof. Kemper geben, so sind die Folien aus der Vorlesung ausschlaggebend.

Falls Ihr einen Fehler oder eine Unstimmigkeit findet, schreibt an i3erdb@in.tum.de mit Angabe der Foliennummer.

Aufgabe 1

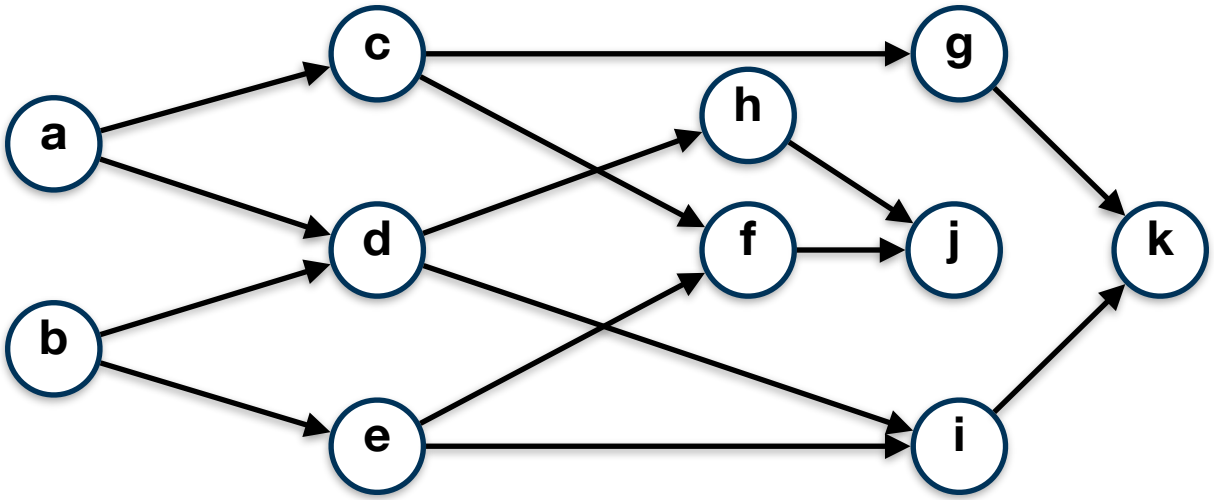
Definieren Sie das Prädikat `sg(X,Y)` das für "same generation" steht. Zwei Personen gehören zur selben Generation, wenn Sie mindestens je ein Elternteil haben, das derselben Generation angehört.

Verwenden Sie beispielsweise die folgende Ausprägung einer ElternKind Relation. Das erste Element ist hier das Kind, das Zweite ein Elternteil.

a) Definieren Sie das Prädikat in Datalog.

b) Demonstrieren Sie die naive Ausführung des Prädikats.

c) Erläutern Sie das Vorgehen bei der seminaiven Auswertung.




```
parent(c,a).
parent(d,a).
parent(d,b).
parent(e,b).
parent(f,c).
parent(g,c).
parent(h,d).
parent(i,d).
parent(i,e).
parent(f,e).
parent(j,f).
parent(j,h).
parent(k,g).
parent(k,i).
```

Aufgabe 5

Naive Auswertung

$S := \{\};$
repeat
$S' := S;$
$S := \Pi_{X,Y}\,(P\,(Z,X) \bowtie_{X=Y} P\,(Z,Y));$
$S := S\,(X,Y) \cup \Pi_{X,Y}\,(P\,(X,Z) \bowtie_{X=Y} P\,(Y,Z));$
$S := S\,(X,Y) \cup \Pi_{X,Y}\,(P\,(X,U) \bowtie (S'\,(U,V) \bowtie P\,(Y,V)));$
until $S' = S$
output $S;$

Aufgabe 5

Semi-naive Auswertung

$S := \{\}; \Delta S := \{\};$
$\Delta S := \Pi_{X,Y}(P(Z,X) \bowtie_{X=Y} P(Z,Y));$
$\Delta S := \Delta S(X,Y) \cup \Pi_{X,Y}(P(X,Z) \bowtie_{X=Y} P(Y,Z));$
$\Delta S := \Delta S(X,Y) \cup \Pi_{X,Y}(P(X,U) \bowtie (S(U,V) \bowtie P(Y,V)));$
$S := \Delta S;$
repeat
  $\Delta S' := \Delta S;$
  $\Delta S := \Pi_{X,Y}(P(Z,X) \bowtie_{X=Y} \Delta P(Z,Y))$ /*erste und*/;
    $\cup\, \Pi_{X,Y}(\Delta P(Z,X) \bowtie_{X=Y} P(Z,Y))$
    $\cup\, \Pi_{X,Y}(P(X,Z) \bowtie_{X=Y} \Delta P(Y,Z))$ /*zweite Regel*/
    $\cup\, \Pi_{X,Y}(\Delta P(X,Z) \bowtie_{X=Y} P(Y,Z))$; /*liefern $\emptyset$*/
  $\Delta S := \Delta S$
    $\cup\, \Pi_{X,Y}(\Delta P(X,U) \bowtie (S(U,V) \bowtie P(Y,V)))$ /*liefert $\emptyset$*/
    $\cup\, \Pi_{X,Y}(P(X,U) \bowtie (\Delta S'(U,V) \bowtie P(Y,V)))$ /*kann $\neq \emptyset$ sein*/
    $\cup\, \Pi_{X,Y}(P(X,U) \bowtie (S(U,V) \bowtie \Delta P(Y,V)))$; /*liefert $\emptyset$*/
  $\Delta S := \Delta S - S$; /*entferne Tupel, die schon vorhanden waren*/
  $S := S \cup \Delta S;$
until $\Delta S = \emptyset$

Deduktive Datenbanken

Naive Auswertung der Rekursion

parent <= (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)

verwandte(Vor, Nach) :- parent(Vor, Nach)
verwandte(Vor, Nach) :- verwandte(Vor, Mitte), parent(Mitte, Nach)

Schritt 0: (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)
Schritt 1: (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122), (K1, K111), (K1, K112), (K1, K121), (K1, K122)
Schritt 2: (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122), (K1, K111), (K1, K112), (K1, K121), (K1, K122)

Deduktive Datenbanken
Semi-Naive Auswertung der Rekursion

parent <= (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)

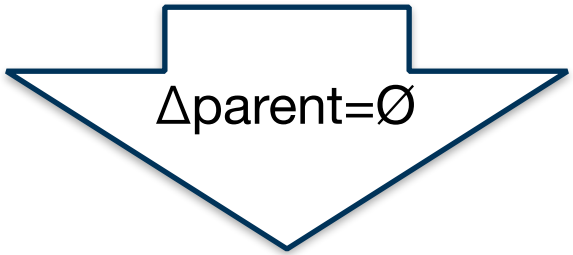
verwandte(Vor, Nach) :- parent(Vor, Nach)
verwandte(Vor, Nach) :- verwandte(Vor, Mitte), parent(Mitte, Nach)

Schritt 0: (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)

Deduktive Datenbanken
Semi-Naive Auswertung der Rekursion

Nur weiteres Auswerten der neu hinzugefügten Knoten (Δ der Relation)

verwandte(Vor, Nach) :- Δparent(Vor, Nach)
verwandte(Vor, Nach) :- Δverwandte(Vor, Mitte), parent(Mitte, Nach)
verwandte(Vor, Nach) :- verwandte(Vor, Mitte), Δparent(Mitte, Nach)



verwandte(Vor, Nach) :- Δverwandte(Vor, Mitte), parent(Mitte, Nach)

Deduktive Datenbanken
Semi-Naive Auswertung der Rekursion

parent <= (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)

verwandte(Vor, Nach) :- Δverwandte(Vor, Mitte), parent(Mitte, Nach)

Schritt 0: (K1,K11), (K1,K12), (K11,K111), (K11,K112), (K12,K121), (K12,K122)
Schritt 1: (K1, K111), (K1, K112), (K1, K121), (K1, K122)
Schritt 2: Ø

Deduktive Datenbanken
Sichere Programme

- Durch Datalog-Regeln erzeugte Relationen müssen endlich sein

- Jede Variable, die in einer Regel vorkommt, muss eingeschränkt sein:
  - die Variable im Rumpf kommt in mindestens einem normalen Prädikat vor (nicht nur in Vergleichsprädikaten) oder
  - X = c mit Konstante c existiert oder
  - ein Prädikat X = Y vorkommt, und Y bereits eingeschränkt ist

Deduktive Datenbanken
Stratifizierte Programme

$$p(...) :- q_1(...), ..., \neg q_i(...), \ldots, q_n(...).$$

- Regel p mit **negiertem** Prädikat q_i nur sinnvoll auswertbar, wenn q_i bereits materialisiert ist

- zuerst alle Regeln mit q_i auswerten => q_i materialisiert
- nur möglich, wenn q_i nicht von p abhängig ist

➡ Abhängigkeitsgraph darf keine Pfade von p nach q_i enthalten

- muss für alle Regeln gelten

Aufgabe 2

Ist folgendes Datalog-Programm stratifiziert?

$$
\begin{aligned}
p(X,Y) &:- q_1(Y,Z), \neg q_2(Z,X), q_3(X,P).\\
q_2(Z,X) &:- q_4(Z,Y), q_3(Y,X).\\
q_4(Z,Y) &:- p(Z,X), q_3(X,Y).
\end{aligned}
$$

Ist das Programm sicher – unter der Annahme, dass p, q_1, q_2, q_3, q_4 IDB- oder EDB-Prädikate sind?

Aufgabe 3

U-Bahnen sind toll! Nehmen wir als Faktenbasis die U-Bahnstationen der Linie U2 Richtung Messestadt West und der Linie U6 Richtung Klinikum Großhadern.

```
direkt(Von, Ziel, Linie).
```

Formulieren Sie ein Datalog-Prädikat, das Ihnen von Garching-Forschungszentrum aus kommend die erreichbaren Stationen inklusiver der Anzahl der Stationen angibt. Testen Sie es!

Aufgabe 4

Schreiben Sie zu dem U-Bahn-Netz-Beispiel auf der Datalog Seite (unter Examples) folgende Anfragen in Datalog:

1. Erstellen Sie den Stationsplan für den U-Bahnhof Fröttmanning, der alle Station, die ohne umsteigen erreichbar sind, auflistet.
2. Erstellen Sie für Garching-Forschungszentrum einen Plan, der alle erreichbaren Stationen, die minimale Anzahl an Umstiegen und Stops auflistet. Beschreiben Sie Ihren Ansatz ausführlich.

Fragen?