Information Integration on the Web: Homework 6

Due on February 27, 2015

Prof. Ambite & Knoblock

Tushar Tiwari

Problem 1

Write Local-As-View (LAV) rules that describe each source. [30 points]

Solution

- $S1(title, genre) \rightarrow Movie(title, director, date, genre) \bigwedge Director(director, "America", birth\_date, death\_date) \bigwedge date < 2000$
- $S2(title, date) \rightarrow Movie(title, director, date, genre) \bigwedge Director("DavidFincher", nationality, birth\_date, death\_date) \bigwedge date > 2001$
- $S3(title, director) \rightarrow Movie(title, director, date, "Sci - Fi")$
- $S4(title, URL) \rightarrow Movie(title, director, date, genre) \bigwedge Poster(title, URL)$

Problem 2

Given the query that searches for all the Drama movies by American directors released before 2000 that have poster URLs online. The returned results should be the titles of the movies and poster URLs. Write the query using the mediated schema and reformulate the query in LAV using the Bucket algorithm. Show the derivations for each step. [30 points]

Solution

query

$q(title, URL) :- Movie(title, director, date, "Drama") \bigwedge Director(director, "America", birth\_date, death\_date) \bigwedge Poster(title, URL) \bigwedge date < 2000$

filling the buckets

$Movie(title, director, date, genre)$	$Director(director, nationality, birth\_date, death\_date)$	$Poster(title, URL)$
$S1(title, genre)$ $S4(title, URL')$	$S1(title, genre')$	$S4(title, URL)$

checking containment

$S1(title, genre), S1(title, genre'), S4(title, URL), genre = "Drama" \rightarrow$
$S1(title, "Drama"), S4(title, URL) \rightarrow$
$Movie(title, director, date, "Drama") \bigwedge Director(director, "America", birth\_date, death\_date) \bigwedge date < 2000 \bigwedge Movie(title, director', date', genre') \bigwedge Poster(title, URL) \rightarrow$
$Movie(title, director, date, "Drama") \bigwedge Director(director, "America", birth\_date, death\_date) \bigwedge date < 2000 \bigwedge Poster(title, URL) \rightarrow q'(title, URL)$

$S1(title, "Drama"), S4(title, URL) \rightarrow q'(title, URL) \equiv q(title, URL)$

Problem 3

For the same query as in question 2 give the inverse rules program that answers the query, and simplify the program. [40 points]

Solution

Query

$q(title, URL) :- Movie(title, director, date, "Drama") \bigwedge Director(director, "America", birth\_date, death\_date) \bigwedge Poster(title, URL) \bigwedge date < 2000$

Inverse Rules

1. $Movie(title, f_1(title, genre), f_2(title, genre), genre) \leftarrow S1(title, genre)$
2. $Director(f_1(title, genre), "America", f_3(title, genre), f_4(title, genre)) \leftarrow S1(title, genre)$
3. $Movie(title, f_5(title, date), date, f_6(title, date)) \leftarrow S2(title, date)$
4. $Director("DavidFincher", nationality, f_7(title, date), f_8(title, date)) \leftarrow S2(title, date)$
5. $Movie(title, director, f_9(title, director), "Sci - Fi") \leftarrow S3(title, director)$
6. $Movie(title, f_{10}(title, URL), f_{11}(title, URL), f_{12}(title, URL)) \leftarrow S4(title, URL)$
7. $Poster(title, URL) \leftarrow S4(title, URL)$

Simplify

Atom 1: $Movie(title, director, date, "Drama")$ *unified with Rule* 1.
$\sigma = \{title \rightarrow t_1, genre \rightarrow "Drama", f_1(t_1, "Drama") \rightarrow dir_1, f_2(t_1, "Drama") \rightarrow date_1\}$
$\sigma(Movie(title, director, date, "Drama")) \equiv \sigma(Movie(title, f_1(title, genre), f_2(title, genre), genre)) \equiv Movie(t_1, dir_1, date_1, "Drama")$
$S1(t_1, "Drama") \leftarrow Movie(t_1, dir_1, date_1, "Drama")$

Atom 2: $Director(director, "America", birth\_date, death\_date)$ *unified with Rule* 2.
$\sigma = \{director \rightarrow dir_1, f_3(t_1, "Drama") \rightarrow bd_1, f_4(t_1, "Drama") \rightarrow dd_1\}$
$\sigma(Director(director, "America", birth\_date, death\_date)) \equiv$
$\sigma(Director(f_1(title, genre), "America", f_3(title, genre), f_4(title, genre))) \equiv$
$Director(dir_1, "America", bd_1, dd_1)$
$S1(t_1, "Drama") \leftarrow Director(dir_1, "America", bd_1, dd_1)$

Atom 3: $Poster(title, URL)$ *unified with Rule* 7.
$\sigma = \{title = t_1, URL = u_1\}$
$\sigma(Poster(title, URL)) \equiv Poster(t_1, u_1)$
$S4(t_1, u_1) \leftarrow Poster(t_1, u_1)$

$q(t_1, u_1) \leftarrow S1(t_1, "Drama") \bigwedge S4(t_1, u_1)$