CS548 - Homework 5

Name: Apurv Upasani USC ID: 4839-3102-30 Email: aupasani@usc.edu

"I, Apurv Upasani, declare that the submitted work is original and adheres to all University policies and acknowledge the consequences that may result from a violation of those rules"

Mediated schema:

- Artist(name, nationality, birth_date, death_date, biography)
- Artwork(title, name, creation_date, dimension, type), note that "name" denotes the artist name.
- Image (title, URL, tag)

Sources

- S1(title, name), which are a list of artworks authored by American artists who passed away before 1930.
- S2(title, creation_date),which are a list of artworks created before 1950 and authored by Walt Disneywho was an American. The type of all the artworks is cartoon.
- S3 (title, name),which are a list of paintings (type of artworks) created after 1990, and their authors.
- S4 (title, URL), which are a list of artworks with image URLs online and the type of which is cartoon.

1. LAV for each source

- S1(title, name) → Artwork(title, name, creation_date,dimension,type) ^ Artist(name,nationality, birth_date,death_date,biography) ^ nationality='USA' ^ death_date < 1930

- S2(title,creation_date) → Artwork(title, name, creation_date,dimension,type) ^ Artist(name,nationality, birth_date,death_date,biography) ^ creation_date < 1950 ^ name ='Walt Disney' ^ nationality='USA' ^ type='cartoon'

- S3(title,name) → Artwork(title, name, creation_date,dimension,type) ^ creation_date > 1990 ^ type='painting'

- S4(title, URL) → Artwork(title, name, creation_date,dimension,type) ^ Image (title, URL, tag) ^ type ='cartoon'

2. Query for LAV

q(title,name,URL) :- Artwork(title, name, creation_date,dimension,type) ^ Artist(name, nationality, birth_date,death_date,biography) ^ Image (title, URL, tag) ^creation_date <1940 ^type='cartoon' ^nationality = 'USA'

Query reformulation

Renaming

title = T, name= N, Creation_date = CD, Dimension = DIM, Birth_date = BD, Death_date = DD, Biography = B, TY = type, NA = nationality

Using Bucket Algorithm

q(T,N,URL) :-
Artwork(T, N, CD, DIM,TY) ^ Artist(N,NA, BD , DD ,B) ^ Image (T, URL, tag) ^ CD <1940 ^ TY = 'cartoon' ^ NA ='USA'

S1(T,N)	S1(T,N)	S4(T,URL)
S2(T,CD)	S2(T,CD)	
S3(T,N)		
S4(T,URL)		

Checking containment

1) q'(T,N,URL):- S1(T,N),S1(T,N),S4(T,URL) ^ CD < 1940 ^ TY='cartoon' ^NA = 'USA'
:- S1(T,N),S4(T,URL) ^CD <1940 ^TY='cartoon'^NA='USA'

Artwork(T,N, CD,DIM, TY) ^ Artist(N,NA,BD,DD,B) ^ NA='USA' ^ DD < 1930 ^ Artwork(T, N, CD,DIM',TY') ^ Image (T, URL, tag) ^ TY ='cartoon'^CD <1940 ^TY='cartoon'^NA='USA'

Merging like expressions

Artwork(T,N,CD,DIM,TY) ^Artist(N,NA,BD,DD,B) ^Image(T,URL,tag) ^ NA ='USA' ^DD<1930 ^TY ='cartoon' ^CD < 1940

Gives cartoons made by USA artists before 1940 who died before 1930

Hence q' $\subseteq$ q

2) q’’(T,N,URL) :- S1(T,N), S2(T,CD), S4(T,URL), CD < 1940 , TY=’cartoon’ , NA = ‘USA’

Artist(N,NA, BD,DD,B) ^ NA=’USA’ ^ DD < 1930 ^ Artwork(T, N, CD,DIM’,TY) ^ Artist(N,NA’, BD’,DD’,B’) ^ CD < 1950 ^ N =’Walt Disney’ ^ NA’=’USA’ ^ TY=’cartoon’ ^ Artwork(T, N, CD,DIM’’,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’ ^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Merging like expressions

Artwork(T,N,CD,DIM,TY) ^ Artist(N,NA,BD,DD,B)^CD<1940 ^ DD<1930 ^N = ‘Walt Disney’ ^TY=’cartoon ^ NA=’USA’

Gives cartoons created before 1940 by artist name Walt Disney who died before 1930 (may be null set)

Hence q’’ $\subseteq$ q

3) q’’’(T,N,URL) :- S2(T,N), S1(T,CD), S4(T,URL) , CD < 1940 , TY=’cartoon’ ,NA = ‘USA’

q’’’(T,N,URL) $\equiv$ q’’(T,N,URL)

Hence q’’’ $\subseteq$ q

4) q’’’’(T,N,URL) :- S2(T,CD),S2(T,CD),S4(T,URL),CD <1940, TY=’cartoon’,NA=’USA’

Artwork(T,N , CD, DIM , TY) ^ Artist(N,NA, BD, DD , B) ^ CD < 1950 ^ N =’Walt Disney’ ^ NA=’USA’ ^ TY=’cartoon’ ^ Artwork(T , N , CD ,DIM’ ,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’^ CD<1940 ^ TY=’cartoon’ ^ NA=’USA’

Merging like expressions

Artwork(T,N,CD,DIM,TY) ^ Artist(N,NA,BD,DD,B) ^Image(T,URL,tag) ^ N=’Walt Disney’ ^TY=’cartoon’ ^CD <1940 ^ NA=’USA’

Hence q’’’’ $\subseteq$ q

5) q’’’’’(T,N,URL) :- S3(T,N),S1(T,N), S4(T,URL) , CD < 1940 , TY=’cartoon’ , NA = ‘USA’

Artwork(T, N, CD,DIM, TY) ^ CD > 1990 ^ TY =’painting’ ^ Artwork(T, N, CD ,DIM’, TY) ^ Artist(N,NA, BD’,DD’ ,B’) ^ NA=’USA’ ^ DD’ < 1930 ^ Artwork(T, N, CD,DIM’’,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’ ^ CD<1940 ^ TY=’cartoon’ ^ NA=’USA’

Since sub goals do not match

Hence q’’’’’ $\not\subset$ q

6) q’’’’’’(T,N,URL) :- S3(T,N), S2(T,CD), S4(T,URL), CD < 1940 , TY=’cartoon’ , NA = ‘USA’

Artwork(T, N, CD,DIM, TY) ^ CD > 1990 ^ TY =’painting’ ^ Artwork(T,N,CD,DIM’,TY) ^ Artist(N,NA, BD,DD,B) ^ CD < 1950 ^ N =’Walt Disney’ ^ NA=’USA’ ^ TY=’cartoon’ ^ Artwork(T,NA, CD,DIM’’,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’ ^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Since sub goals do not match

Hence q’’’’’’ $\not\subset$ q

7) q’’’’’’’(T,N,URL) :- S4(T,CD), S2(T,CD),S4(T,URL) ^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’
q’’’’’’’(T,N,URL) :- S2(T,CD),S4(T,URL) ^ CD’ < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Artwork(T, N, CD, DIM ,TY) ^ Artist(N,NA, BD,DD,B) ^ CD < 1950 ^ N =’Walt Disney’ ^ NA=’USA’ ^ TY=’cartoon’^ Artwork(T, N, CD,DIM’,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Merging like expressions

Artwork(T,N,CD,DIM,TY) ^Artist(N,NA,BD,DD,B) ^ Image (T, URL,tag) ^ CD <1940 ^ N=’Walt Disney’ ^NA=’USA’ ^ TY=’cartoon’

Gives cartoons created before 1940 by Walt Disney of USA.
Hence q’’’’’’’ $\subseteq$ q

8) q’’’’’’’’(T,N,URL) :- S4(T,URL) ,S1(T,N) ,S4(T,URL) , CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’
q’’’’’’’’(T,N,URL) :- S4(T,URL) , S1(T,N) ^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Artwork(T, N, CD,DIM,TY) ^ Image (T, URL, tag) ^ TY =’cartoon’ ^ Artwork(T, N, CD,DIM’,TY) ^ Artist(N,NA,BD,DD,B) ^ NA=’USA’ ^ DD < 1930 ^ CD < 1940 ^ TY=’cartoon’ ^NA = ‘USA’

Merging like expressions

Artwork(T,N,CD,DIM,TY)^ Image (T, URL,tag)^ Artist(N,NA,BD,DD,B) ^ NA=’USA’ ^ CD < 1940^ TY=’cartoon’^DD< 1930

Gives cartoons created before 1940 by USA artists who died before 1930 (technically cartoons created before 1930)

Hence q’’’’’ $\subseteq$ q
As a result, a combination of sub queries {q’,q’’,q’’’, q’’’’,q’’’’’’,q’’’’’’’} can be used to form q(T, N, URL)

3. GAV for each mediated relation

We assume that all sources are from the same database

Artwork (T,N,_,_,_) ⟵ S1(T,N)
Artwork (T,"Walt Disney",CD,_,"cartoon") ⟵ S2(T,CD)
Artwork(T,N,_,_,"painting") ⟵ S3(T,N)
Artwork(T,_,_,_,"cartoon") ⟵ S4(T,URL)
Artist(N,"USA",_,_,_) ⟵ S1(T,N)
Artist("Walt Disney","USA",_,_,_)⟵ S2(T,CD)
Artist(N,_,_,_,_) ⟵ S3(T,N)
Image(T,URL,_) ⟵ S4(T,URL)

4. Query using GAV

q(T,N,URL) :- Artwork(T, N, CD,_,"cartoon"), Artist(N, "USA",_,_,_) ,Image (T, URL, _) ,CD <1940

Query reformulation

1. q'(T,N,URL) :- S1(T,N), S1(T,N),S4(T,URL), CD < 1940
   :- S1(T,N),S4(T,URL),CD <1940 (since S1 has the same source)

   q' is an invalid query as predicate CD not found in S1, S4
   Hence q' is discarded.

2. q''(T,N,URL) :- S2(T,CD),S1(T',N'),S4(T'',URL''),CD <1940
   :- S1(T,N),S2(T,CD),S4(T,URL),CD<1940 (T,T',T'','' are keys and T=T'=T'')

3. q'''(T,N,URL) :- S4(T,URL), S1(T',N'), S4(T'',URL),CD <1940
   :- S1(T,N),S4(T,URL),CD <1940 (since S1 has the same source)

   q''' is an invalid query as predicate CD not found in S1, S4
   Hence q' is discarded.

4. q''''(T,N,URL) :- S1(T,N), S2(T',CD), S4(T'',URL), CD < 1940
   :- S1(T,N), S2(T,CD), S4(T,URL), CD < 1940 (assume T,T',T''' are keys and T=T'=T'')

5. q'''''(T,N,URL) :- S2(T,CD),S2(T',CD), S4(T'',URL), CD<1940

   q''''' is invalid query as predicate N required in head not found in any predicate.
   Hence, q''''' is discarded

6. q^{vi}(T,N,URL) :- S4(T,URL),S2(T',CD),S4(T'',URL''), CD < 1940
   :- S4(T,URL),S2(T',CD),CD<1940

   q^{vi} is invalid query as predicate N required in head not found in any predicate.
   Hence, q''''' is discarded

7. q^{vii}(T,N,URL) :- S1(T,N),S3(T',N'),S4(T'',URL) , CD<1940

   q^{vii} is an invalid query as predicate CD not found in S1, S3, S4
   Hence q^{vii} is discarded.

8. q^{Viii}(T,N,URL) :- S2(T,CD),S3(T',N'),S4(T'',URL), CD <1940

   q^{Viii} is invalid as it returns all the paintings. Hence should be discarded

9. q^{ix}(T,N,URL) :-S4(T,URL),S4(T',URL'),S4(T'',URL''),CD <1940
   :-S4(T,URL), CD <1940 (assume T,T',T''' are keys and T=T'=T'')

   q^{ix} is an invalid query as predicate CD not found in S4
   Hence q^{ix} is discarded.

Hence q(T,N,URL) can be described by q'' and q''''.