SPARQL Query Mediation over RDF Data Sources with Disparate Contexts

Xiaoqing Zheng
School of Computer Science
Fudan University
zhengxq@fudan.edu.cn

Stuart E. Madnick
Sloan School of Management
Massachusetts Institute of Technology
smadnick@mit.edu

Xitong Li
Sloan School of Management
Massachusetts Institute of Technology
xitongli@mit.edu
Outline

- Motivational example
- Algorithm description
- System architecture
- Demonstration

LDOW2012, Lyon
Query on RDF Views

SELECT ?airline1 ?airline2 ?total
WHERE {
  GRAPH ?graph1
  {  ?airline1   depDateTime  ?depDateTime1 ;
     arrDateTime  ?arrDateTime1 ;depCity  "Boston";
     arrCity  "Tokyo";
     price  ?price1 .  }
  GRAPH ?graph2
  {  ?airline2  depDateTime  ?depDateTime2 ;
     arrDateTime  ?arrDateTime2 ;depCity  "Tokyo";
     arrCity  "Shanghai";
     price  ?price2 .  }
  FILTER ( ?depDateTime1 >= "12:30 AM 02/09/2011") .
  FILTER ( ?arrDateTime2 <= "11:30 PM 02/10/2011" ) .
  FILTER ( ?arrDateTime1 < ?depDateTime2 ) .
  LET ( ?total := ?price1 + ?price2 )
  ORDER BY ASC(?total)
  LIMIT 1
}

The answer is empty.

LDOW2012, Lyon
select ?airline1 ?airline2 ?total
where {
  graph1
  {?airline1 depDateTime ?depDateTime1 ;
    arrDateTime ?arrDateTime1 ;
    depCity "Boston" ;
    arrCity "Tokyo" ;
    price ?price1 .}
  graph2
  {?airline2 depDateTime ?depDateTime2 ;
    arrDateTime ?arrDateTime2 ;
    depCity "Tokyo" ;
    arrCity "Shanghai" ;
    price ?price2 .}

  filter (?depDateTime1 >= "9:30 AM 02/09/2011") .
  filter (?arrDateTime2 <= "11:30 PM 02/10/2011") .
  filter (?arrDateTime1 < ?depDateTime2) .

  let (?total := ?price1 + ?price2) .

  order by asc(?total)
}

# Named graph: http://japanairline.com/flights

jp241 depDateTime "2011-02-10T09:25:00Z"^^xsd:dateTime .
jp241 arrDateTime "2011-02-10T22:05:00Z"^^xsd:dateTime .
jp241 depCity "Tokyo" .
jp241 arrCity "Shanghai" .
jp241 price 25 .
SELECT ?airline1 ?airline2 ?total
WHERE {
  GRAPH ?graph1
  { ?airline1 depDateTime ?depDateTime1 ;
    arrDateTime ?arrDateTime1 ; depCity "Boston" ;
    arrCity "Tokyo" ; price ?price1 . }
  GRAPH ?graph2
  { ?airline2 depDateTime ?depDateTime2 ;
    arrDateTime ?arrDateTime2 ; depCity "Tokyo" ;
    arrCity "Shanghai" ; price ?price2 . }
  FILTER (?depDateTime1 >= "12:30 PM 02/09/2011") .
  FILTER (?arrDateTime2 <= "11:30 PM 02/10/2011") .
  FILTER (?arrDateTime1 < ?depDateTime2) .
  LET (?total := ?price1 + ?price2) .
  ORDER BY ASC(?total)
  LIMIT 1
}

# Named graph: http://usairline.com/flights
us339 depDateTime "12:30 PM 02/09/2011" .
us339 arrDateTime "7:25 AM 02/10/2011" .
us339 depCity "BOS" .
us339 arrCity "TYO" .
us339 price 950 .
us512 depDateTime "9:45 AM 02/10/2011" .
us512 arrDateTime "10:30 PM 02/10/2011" .
us512 depCity "TYO" .
us512 arrCity "SHA" .
us512 price 380 .

# Named graph: http://japanairline.com/flights
jp241 depDateTime "2011-02-10T09:25:00Z"^^xsd:dateTime .
jp241 arrDateTime "2011-02-10T22:05:00Z"^^xsd:dateTime .
jp241 depCity "Tokyo" .
jp241 arrCity "Shanghai" .
jp241 price 25 .
Semantic Heterogeneity

# Named graph: http://usairline.com/flights

us339 depDateTime "12:30 PM 02/09/2011".
us339 arrDateTime "7:25 AM 02/10/2011".
us339 depCity "BOS".
us339 arrCity "TYO".
us339 price 950.

us512 depDateTime "9:45 AM 02/10/2011".
us512 arrDateTime "10:30 PM 02/10/2011".
us512 depCity "TYO".
us512 arrCity "SHA".
us512 price 380.

# Named graph: http://japanairline.com/flights

jp241 depDateTime "2011-02-10T09:25:00Z"^^xsd:dateTime.
jp241 arrDateTime "2011-02-10T22:05:00Z"^^xsd:dateTime.
jp241 depCity "Tokyo".
jp241 arrCity "Shanghai".
jp241 price 25.

SELECT ?airline1 ?airline2 ?total
WHERE {
  GRAPH ?graph1
  {  ?airline1   depDateTime  ?depDateTime1 ;
     arrDateTime  ?arrDateTime1 ;depCity  "Boston";
     arrCity  "Tokyo";
     price  ?price1 . }
  GRAPH ?graph2
  {  ?airline2  depDateTime  ?depDateTime2 ;
     arrDateTime  ?arrDateTime2 ;depCity  "Tokyo";
     arrCity  "Shanghai";
     price  ?price2 . }
  FILTER ( ?depDateTime1 >= "9:30 AM 02/09/2011" ) .
  FILTER ( ?arrDateTime2 <= "11:30 PM 02/10/2011" ) .
  FILTER ( ?arrDateTime1 < ?depDateTime2 ) .
  LET ( ?total := ?price1 + ?price2 )
  ORDER BY ASC(?total)
  LIMIT 1
SELECT ?airline1 ?airline2 ?total
WHERE {
  GRAPH ?graph1
  { ?airline1 depDateTime ?depDateTime1 ;
    arrDateTime ?arrDateTime1 ;
    depCity "BOS";
    arrCity "TYO" ;
    price ?price1 .
  }
  GRAPH ?graph2
  { ?airline2 depDateTime ?depDateTime2 ;
    arrDateTime ?arrDateTime2 ;
    depCity "TYO" ;
    arrCity "SHA" ;
    price ?price2 .
  }
  FILTER (?depDateTime1 >= "9:30 AM 02/09/2011") .
  FILTER (?arrDateTime2 <= "11:30 PM 02/10/2011") .
  FILTER (?arrDateTime1 < ?depDateTime2) .
  LET (?total := ?price1 + ?price2) .
  ORDER BY ASC(?total)
  LIMIT 1
}

It still does not work.
A simplified flight data

<table>
<thead>
<tr>
<th>flight</th>
<th>departure</th>
<th>arrival</th>
<th>from</th>
<th>to</th>
<th>ticket price</th>
</tr>
</thead>
<tbody>
<tr>
<td>us339</td>
<td>12:30 PM 02/09/2011</td>
<td>7:25 AM 02/10/2011</td>
<td>BOS</td>
<td>TYO</td>
<td>950</td>
</tr>
<tr>
<td>us512</td>
<td>9:45 AM 02/10/2011</td>
<td>10:30 PM 02/10/2011</td>
<td>TYO</td>
<td>SHA</td>
<td>380</td>
</tr>
</tbody>
</table>
Context Definition

US_dollar hasCurrency "USD";
hasScale "1"^^xsd:integer;
hasDataType xsd:long.

Context Pool

A simplified flight data

<table>
<thead>
<tr>
<th>flight</th>
<th>departure</th>
<th>arrival</th>
<th>from</th>
<th>to</th>
<th>ticketprice</th>
</tr>
</thead>
<tbody>
<tr>
<td>us339</td>
<td>12:30 PM 02/09/2011</td>
<td>7:25 AM 02/10/2011</td>
<td>BOS</td>
<td>TYO</td>
<td>950</td>
</tr>
<tr>
<td>us512</td>
<td>9:45 AM 02/10/2011</td>
<td>10:30 PM 02/10/2011</td>
<td>TYO</td>
<td>SHA</td>
<td>380</td>
</tr>
</tbody>
</table>

LDOW2012, Lyon
US_dollar hasCurrency "USD" ;
hasScale "1"^^xsd:integer ;
hasDataType xsd:long .

flight | departure       | arrival          | from | to    | ticketprice |
-------|-----------------|------------------|------|-------|-------------|
us339  | 12:30 PM 02/09/2011 | 7:25 AM 02/10/2011 | BOS  | TYO   | 950         |
us512  | 9:45 AM 02/10/2011   | 10:30 PM 02/10/2011 | TYO  | SHA   | 380         

A simplified flight data

LDOW2012, Lyon
SELECT ?airline1 ?airline2 ?total
WHERE {
  GRAPH ?graph1
  { ?airline1 depDateTime ?depDateTime1 ;
    arrDateTime ?arrDateTime1 ;
    depCity "Boston" ;
    arrCity "Tokyo" ;
    price ?price1 . }

  GRAPH ?graph2
  { ?airline2 depDateTime ?depDateTime2 ;
    arrDateTime ?arrDateTime2 ;
    depCity "Tokyo" ;
    arrCity "Shanghai" ;
    price ?price2 . }

  FILTER ( ?depDateTime1 >= "9:30 AM 02/09/2011" ) .
  FILTER ( ?arrDateTime2 <= "11:30 PM 02/10/2011" ) .
  FILTER ( ?arrDateTime1 < ?depDateTime2 )
  LET ( ?total := ?price1 + ?price2 )
  ORDER BY ASC(?total)
  LIMIT 1
}

The answers returned should be further transformed so that they conform to the context of the receiver.

The constants should be transformed to comply with assumptions in the source contexts.

One of two arguments in expressions should be transformed so that the two arguments conform to the same context.

LDOW2012, Lyon
<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
<th>Receiver Context</th>
<th>Source Context</th>
</tr>
</thead>
</table>
| Flight    | Finding one-way cheap airfare from Boston (after 9:30 a.m., Feb 9th, 2011) to Shanghai (before 11:30 p.m., Feb 10th, 2011.) with one stop in Tokyo. | Currency is USD with a scale-factor of 1; Datetime is expressed in US style; Locations are expressed as city name. | SOURCE 1  
Currency is USD with a scale-factor of 1; Datetime is expressed in US style; Locations are expressed as IATA airport codes.  
SOURCE 2  
Currency is JPY with a scale-factor of 1000; Datetime is expressed in xsd:dateTime type; Locations are expressed as city names. |
| Factbook  | Finding the countries whose GDP is more than 50 trillion Chinese Yuan (CNY) against the CIA Factbook database | Currency is CNY with a scale-factor of trillion. | SOURCE 1  
Currency is USD with a scale-factor of 1. |
| Investment| Finding the companies where more than 50% of their profits come from their oversea branches and the profits of the branches are greater than twenty million USD. | Currency is USD with a scale-factor of 1,000. | SOURCE1  
Currency is USD with a scale-factor of 10,000.  
SOURCE2  
Currency is JPY with a scale-factor of 10,000.  
SOURCE3  
Currency is CNY with a scale-factor of 10,000. |
Thanks!

Xiaoqing Zheng
School of Computer Science
Fudan University
zhengxq@fudan.edu.cn

LDOW2012, Lyon