Extensions of SPARQL towards Heterogeneous Sources and Domain Annotations

Nuno Lopes

November 09, 2010
Scenario: Integration of Heterogeneous Sources

Additional features

Add RDB input

Represent Dynamic data as RDF
Scenario: Integration of Heterogeneous Sources

1. **XML**

2. **SOAP/WSIL**

3. **RDF**
Scenario: Integration of Heterogeneous Sources

Lowering

<XML/>

SOAP/WSDL

KML

Represent Dynamic data as RDF
Scenario: Integration of Heterogeneous Sources

Lifting

Additional features

1. Add RDB input
2. Represent Dynamic data as RDF
Scenario: Integration of Heterogeneous Sources

XSPARQL

<XML/>

SOAP/WSDL

KML

Represent Dynamic data as RDF

Add RDB input
Scenario: Integration of Heterogeneous Sources

XSPARQL + SQL

Additional features
Add RDB input
Scenario: Integration of Heterogeneous Sources

Additional features
Add RDB input
Represent Dynamic data as RDF

XSPARQL + SQL + Annotated RDF
Scenario: Integration of Heterogeneous Sources

Additional features
- Add RDB input
- Represent Dynamic data as RDF

XSPARQL + SQL + Annotated RDF
Motivation: Lifting and Lowering

Transformations between XML and RDF are not easy mainly due to the heterogeneity of RDF/XML serialisations.

Objective: language capable of integrating heterogeneous sources for the Semantic Web.

---

**Lifting**

```xml
location.kml

<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
  <Placemark>
    <name>Location of Nuno Lopes</name>
    <Point>
      <coordinates>-74.006393,40.714172,0</coordinates>
    </Point>
  </Placemark>
</kml>
```

**Lowering**

```turtle
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
_:b1 a foaf:Person;
    foaf:name "Nuno Lopes";
    foaf:based_near _:b2.
_:b2 a geo:Point;
    geo:long "+-9.041"
    geo:lat "53.272".
```

---

http://nunoilopes.org/foaf.rdf
Motivation: Lifting and Lowering

Transformations between XML and RDF are not easy mainly due to the heterogeneity of RDF/XML serialisations. The objective is to create a language capable of integrating heterogeneous sources for the Semantic Web.
Motivation: Lifting and Lowering

Transformations between XML and RDF are not easy
Mainly due to the heterogeneity of RDF/XML serialisations

Objective: language capable of integrating heterogeneous sources for the Semantic Web
XQuery + SPARQL = XSPARQL

- Syntactic extension of XQuery
- With a formally defined semantics
- Includes XML and RDF sources
XQuery + SPARQL = XSPARQL

- Syntactic extension of XQuery
- With a formally defined semantics
- Includes XML and RDF sources

Query example (Lowering)

```xml
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
<kml xmlns="http://www.opengis.net/kml/2.2">
  { for $person $name $long $lat from <http://nunolopes.org/foaf.rdf>
    where { $person a foaf:Person; foaf:name $name; foaf:based_near [ a geo:Point; geo:long $long; geo:lat $lat ] }
    return <Placemark>
      <name>{fn:concat("Location of ", $name)}</name>
      <Point><coordinates>{fn:concat($long, ",", $lat, ",0")}</coordinates></Point>
    </Placemark>
  }
</kml>
```
XQuery + SPARQL = XSPARQL

- Syntactic extension of XQuery
- With a formally defined semantics
- includes XML and RDF sources

Query example (Lowering)

```xml
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
<kml xmlns="http://www.opengis.net/kml/2.2">
  { for $person $name $long $lat from <http://nunolopes.org/foaf.rdf>
    where { $person a foaf:Person; foaf:name $name;
            foaf:based_near [ a geo:Point; geo:long $long; geo:lat $lat ] }
  return <Placemark>
    <name>{fn:concat("Location of ", $name)}</name>
    <Point><coordinates>{fn:concat($long, ",", $lat, ",0")}</coordinates></Point>
  </Placemark>
}</kml>
```
Use case: Using XSPARQL to expose location data

sensor tags are assigned to people
Use-case: Using XSPARQL to expose location data

*Sensor tags* are assigned to people.

Tag proximity is registered by *base stations*. 
Use case: Using XSPARQL to expose location data

Base stations are deployed throughout a building.
Use case: Using XSPARQL to expose location data

Sensor data readings:

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>IP Address</th>
<th>Tag</th>
<th>SSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.15</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.3.1</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.6</td>
<td>4302</td>
<td>83</td>
</tr>
</tbody>
</table>
Usecase: Using XSPARQL to expose location data

<table>
<thead>
<tr>
<th>timestamp</th>
<th>ip</th>
<th>tag</th>
<th>ssi</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.15</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.3.1</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.6</td>
<td>4302</td>
<td>83</td>
</tr>
</tbody>
</table>
Usecase: Using XSPARQL to expose location data

sensor data readings:

<table>
<thead>
<tr>
<th>timestamp</th>
<th>ip</th>
<th>tag</th>
<th>ssi</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.15</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.3.1</td>
<td>4302</td>
<td>83</td>
</tr>
<tr>
<td>2010-11-09 14:57:51</td>
<td>10.254.2.6</td>
<td>4302</td>
<td>83</td>
</tr>
</tbody>
</table>
Use case: Using XSPARQL to expose location data

Use XSPARQL to create a webpage from the annotated data.
Integrating XSPARQL with relational databases

Next steps
- Defining the syntax for querying RDBs
- Extending XSPARQL’s semantics

RDB Query syntax example (draft)

```prefix : <http://example.org/>
for locations.tag as $tag, locations.ip as $ip
from locations
construct { <$tag> :locatedIn <$ip> }
```
Integrating XSPARQL with relational databases

Next steps

- Defining the syntax for querying RDBs
- Extending XSPARQL’s semantics

RDB Query syntax example (draft)

```
prefix : <http://example.org/>

for locations.tag as $tag, locations.ip as $ip
from locations
construct { <{$tag}> :locatedIn <{$ip}> }
```
Integrating annotations in XSPARQL

- Each RDF triple is given an annotation
- Annotations refer to a specific **domain**
Integrating annotations in XSPARQL

- Each RDF triple is given an annotation
- Annotations refer to a specific domain

Temporal:

:tag4302 :locatedIn :room311 . "2010-11-09 14:57:51"
Integrating annotations in XSPARQL

- Each RDF triple is given an annotation
- Annotations refer to a specific domain

Temporal:

:tag4302 :locatedIn :room311 . "2010-11-09 14:57:51"

Fuzzy:

:tag4302 :locatedIn :room311 . "0.9"
Integrating annotations in XSPARQL

- Each RDF triple is given an annotation
- Annotations refer to a specific domain

**Temporal:**

```plaintext
:tag4302 :locatedIn :room311 . "2010-11-09 14:57:51"
```

**Fuzzy:**

```plaintext
:tag4302 :locatedIn :room311 . "0.9"
```

**Annotated SPARQL Queries:**

“*When were two people in the same room?*”

“*Who is closer to room 311?*”
Integrating annotations in XSPARQL

- Each RDF triple is given an annotation
- Annotations refer to a specific **domain**

**Temporal:**

```sparql
:tag4302 :locatedIn :room311 . "2010-11-09 14:57:51"
```

**Fuzzy:**

```sparql
:tag4302 :locatedIn :room311 . "0.9"
```

**Annotated SPARQL Queries:**

```sparql
SELECT ?l ?person WHERE {
  ?tag1 :assignedTo :nuno .
  ?tag1 :locatedIn :room311 . ?l
  ?tag2 :assignedTo ?person .
  ?tag2 :locatedIn :room311 . ?l
}
```
Current Issues / Future work

Query language integrating heterogeneous sources

- Relational Databases
- W3C RDB2RDF Working Group
Current Issues / Future work

Query language integrating heterogeneous sources

- Relational Databases
- W3C RDB2RDF Working Group

Integration of annotated data (Temporal, Fuzzy, ...) 

- Higher entailment regimes?
Current Issues / Future work

Query language integrating heterogeneous sources
- Relational Databases
- W3C RDB2RDF Working Group

Integration of annotated data (Temporal, Fuzzy, ...)
- Higher entailment regimes?

Optimisations
- Based on results from RDBs and XQuery?
**Current Issues / Future work**

<table>
<thead>
<tr>
<th>Query language integrating heterogeneous sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Relational Databases</td>
</tr>
<tr>
<td>- W3C RDB2RDF Working Group</td>
</tr>
</tbody>
</table>

- **Integration of annotated data (Temporal, Fuzzy, …)**
  - Higher entailment regimes?

- **Optimisations**
  - Based on results from RDBs and XQuery?

- **Update language**
  - How to handle the heterogeneous sources?
Above all... Avoid creating
Above all... Avoid creating

Thank you! Questions?