MATLAB-SciSPARQL integration
• Sending queries and updates directly from MATLAB interpreter
• Retrieving results row-by-row on demand
• Complete mapping between RDF terms and MATLAB types (numbers are numbers, arrays are arrays, ...)

SciSPARQL features

Array operations: slicing, projection, transposition (performed on the server)

\[
\text{SELECT} \ (\text{?A[?start:?step:, ?i]}) \ \text{AS} \ ?\text{result}) \\
\text{WHERE} \ ... \ \\
# slice and project array ?A
\]

Intra-array aggregations:
sum(), min(), etc. of array elements (performed on the server)

\[
\text{SELECT} \ (\text{array_sum(?A[?i])}) \ \text{AS} \ ?\text{result}) \\
\text{WHERE} \ ... \\
# sum up the column ?i in array ?A
\]

Automatic subscript ranges:
variables are bound to available subscripts

\[
\text{SELECT} \ \text{?i}, (\text{?A[?i]) AS} \ ?\text{result}) \\
\text{WHERE} \ ... \\
\text{FILTER} \ (\text{mod(?i, 2) = 1}) \\
# return every odd row in ?A
\]

Arrays as arithmetic operands:
extended arithmetic operators and SPARQL 1.1 aggregate functions (performed on the server)

\[
\text{SELECT} \ (\text{AVG(abs(?A-?B))}) \ \text{AS} \ ?\text{result}) \\
\text{WHERE} \ {( ?x :a ?A ; :b ?B )} \\
# get average absolute difference between :a and :b properties (element-wise if arrays)
\]

Functional RDF views:
parameterized SPARQL queries (and updates)

\[
\text{DEFINE FUNCTION} \ sse(?x) \ \\
\text{AS} \ \text{SELECT} \ \text{array_sum(sqr(?A-?B))) AS} \ ?\text{result}) \\
\text{WHERE} \ {( ?x :a ?A ; :b ?B )} \\
# get sum-of-squared error between :a and :b properties of ?x
\]

Second-order functions:
operate on functions or closures (performed on the server)

\[
\text{SELECT} \ (\text{ARGMIN}(\text{sse}) \ AS} \ ?\text{x}) \\
# get the node ?x with minimal SSE
\]

SciSPARQL update

%% Generate
...
%% Contribute
> c = newConnection(...)
> c.sparql(...)

%% Retrieve
> c.sparql(...)
%% Postprocess
...

SciSPARQL Database Manager (SSDM) features

RDF with arrays
numeric multidimensional arrays are stored as single nodes in RDF graph

Combining data and metadata in queries and updates

Client/server architecture
Scientific (e.g. experimental) data is
• contributed with complete annotation
• stored
• retrieved on demand

Opening way for data integration

No data overhead
Massive numeric arrays are projected, aggregated and filtered as part of the query answering on the server. Only the query results are sent over

Extensible server
Foreign functions can be implemented in Python, Java, C, and used in queries

Making use of existing data

Scalable data management
in-memory RDF storage for metadata.
native binary file formats for massive multidimensional numeric arrays

No performance overhead

MATLAB Client

SSDM Server

MATLAB-SciSPARQL integration

Andrej Andrejev, Xueming He, Tore Risch
Department of Information Technology, Uppsala University
http://it.uu.se/research/group/udbl/SciSPARQL

Client/server architecture

Scientific (e.g. experimental) data is
• contributed with complete annotation
• stored
• retrieved on demand

Opening way for data integration

SciSPARQL update

%% Generate
...
%% Contribute
> c = newConnection(...)
> c.sparql(...)

%% Retrieve
> c.sparql(...)
%% Postprocess
...

SciSPARQL Database Manager (SSDM) features

RDF with arrays
numeric multidimensional arrays are stored as single nodes in RDF graph

Combining data and metadata in queries and updates

Client/server architecture
Scientific (e.g. experimental) data is
• contributed with complete annotation
• stored
• retrieved on demand

Opening way for data integration

No data overhead
Massive numeric arrays are projected, aggregated and filtered as part of the query answering on the server. Only the query results are sent over

Extensible server
Foreign functions can be implemented in Python, Java, C, and used in queries

Making use of existing data

Scalable data management
in-memory RDF storage for metadata.
native binary file formats for massive multidimensional numeric arrays

No performance overhead

MATLAB Client

SSDM Server