How Semantic Technologies Can Enhance Data Access at Siemens

Evgeny Kharlamov
Nina Solomakhina

University of Oxford
Siemens

Ö.Özcep, D.Zheleznyakov, T.Hubauer
S.Lamparter, M.Roshchin, A. Soylu, S. Watson
What this Paper is About?

- In use track paper
  - based on our experience with Siemens Energy
  - part of the Optique project

- What we did
  - described and analysed monitoring of appliances by Siemens Energy, understood issues
  - set requirements to enhance Siemens monitoring routine
  - assessed why and how SW technologies can do the job
  - set a road map to improve existing SW tech to met S. reqs
  - developed the Optique platform solution and evaluation

- What we will do
  - continue the development of the solution
  - evaluations and user training
Monitoring by Siemens Energy

- Siemens
  - produces huge appliances, e.g., turbines
  - installs them in plants

- Siemens service centers
  - offers constant monitoring and diagnostics service
  - over 50 service centers world wide
  - each SC is responsible for several thousands appliances
  - their job: monitoring and diagnostic of turbine

- Monitoring and diagnostic tasks
  - reactive and preventive diagnostics
    - offline, after an issue is detected
  - predictive analyses
    - real-time, to avoid issues
Monitoring and Diagnostics Routine

- **Routine**
  - Arrival of a service ticket
  - Data access and acquisition
  - Data analyses
    - Query result visualisation
    - Data preprocessing
    - Data analysis
  - Report preparation

- **Time is vital**
  - 50 service centres
  - >1,000 requests per center per year
  - Improving access time have potential saving: €50,000,000 per year

Why does it take so long?
Data Access

- Preconfigured tools
  - different views, query catalogs
  - >4,000 queries and query patterns
  - relatively fast way to get to data

- This is often not enough
  - high variety of issues
  - existing tools are often are too specific / general

- Via an IT expert
  - per week: several new Qs, up to 35 Qs modified
  - yearly: 10% of queries are changed
  - overload with tasks
  - miscommunication, data exploration

Return the most frequent start failure and warning messages of the gas turbine T01 during the last week. Moreover, find analogous cases of failures for turbines of the same type as T01 in the last three months.
Proposed Solution

- Allow engineer to formulate queries
  - Data exploration
  - Relatively easy queries
  - Will help with relatively easy info needs
  - Enough to get a concrete understanding of the final complex query

- Does Siemens want it?
  - management say
    - we want this to happen
  - engineers say
    - we know *what* we want to ask
    - we do not have means of doing it

Is it hard to query Siemens data?
Data in a Data Centre

- Each data centre
  - Serves several thousands appliances
  - Accommodates thousands of DBs

- Sensor and event data
  - raw sensor data and pre-processed, event data

- Analytical data
  - data from previous monitoring cases

- Miscellaneous data
  - design, operating regimes of units
  - customer data: location, contact info, etc.
  - weather condition
  - some logs, e.g., dates of running the unit off/on
Data in a Data Centre

- Schemata
  - varies for different appliances: turbines, generators, compressors
  - of each turbine
    - > 150 tables
  - frequent changes
  - Up to 20 attributes per (sensor/event) table

- The actual data
  - > 100 TB of time stamped data
  - ~ 15 GB data for each turbine
  - ~ 30 GB of fresh data generated daily

---

Optique
Reqs for Direct Data Access

R1: Integrated Data Access
- unified access point to: sensor, event, analytical, and misc data

R2: Flexible Definition of Queries by Endusers
- non IT experts should be able to write queries

R3: Utilising implicit information
- to deal with missing data values

R4: Processing of Stream Data
- monitoring heavily depends on streaming data

R5: System Deployment & Maintenance Support
- deployment and maintenance should not be too costly

Return the most frequent start failure and warning messages of the gas turbine T01 during the last week. Moreover, find analogous cases of failures for turbines of the same type as T01 in the last three months.
Ontology Based Data Access

What is OBDA?
- an approach to Data integration
- virtual approach: data stays where it is
- ontology mediates users and data
- ontology and data are related by mappings

Conceptually: OBDA naturally addresses Siemens Reqs
- R1: Integrated Data Access
- R2: Flexible Definition of Qs by Endusers
- R3: Utilising implicit information
- R4: Processing of Stream Data
- R5: Deployment and Maintenance Support
OBDA System for Siemens?

- Practically: We need an end-to-end system
  - easily deployable with a semi-automatic techniques
  - accommodate relational and streaming data
  - allows for ontology and mapping import, analyses, and management
  - offers intuitive query interfaces
  - efficient and effective query rewriting
  - supports distributed query planning, optimisation, and execution.

- Existing OBDA systems address some of Siemens requirements
  - ontop, Mastro, morph – RDB, R2RQ, OntoQF, Spyder, etc.

- OBDA platform Optique integrates
  - existing solutions: ontop, ADP, Information Workbench, HermiT
  - novel components: bootstrapper, QF support, streaming language, etc.
Optique Platform

Optique demo video: www.youtube.com/user/optiqueproject/playlists
Optique Platform at Siemens

- Deployment - Fujitsu Server PRIMERGY RX600
  - 4 x 8-Core Intel 64bit processors
  - 8 x 64GB RAM
  - 7TB internal + 24TB external storage

- Demonstration data
  - 3 TB of historical data gathered in 2002-2011
  - 200 gas and steam turbines (~15 GB/turbine)
  - Data stream generator - simulates the original sensor measurements and events streaming from the turbines
Optique Platform

Optique demo video: www.youtube.com/user/optiqueproject/playlists
Siemens Ontologies and Mappings

- Based on semiformal and informal ontologies available at Siemens
- Expressed in OWL 2 QL

Stream Processing

Optique demo video: www.youtube.com/user/optiqueproject/playlists
The query framework STARQL

- Query language framework for processing of ontology level streams
- Allows non-reified handling of time/state
- Many ontology stream query languages use reification (incl. C-SPARQL, SPARQLStream, CQUELS)

<table>
<thead>
<tr>
<th>Non-reified approach</th>
<th>Reified approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time is a flow ((T, \leq_T))</td>
<td>Time is an “ordinary” attribute</td>
</tr>
</tbody>
</table>

Sensor shows value 90 at timepoint \(i\)

\[
\{s \text{ hasVal 90}\}<i> \\
\{m \text{ rdf:type Meas. m hasSens s. m hasVal 90. m hasTime i}\}
\]
STARQL - Example

Input stream:
S_Msmt = { val(s0, 90 °C)<0s>,
           val(s0,93 °C)<1s>,
           val(s0,94 °C)<2s>,
           val(s0,92 °C)<3s>,
           val(s0,95 °C)<5s>,
           ...
}

Information need for monotonicity:
Tell every 1s whether the temperature in sensor s0 increased monotonically in the last 2s.

STARQL representation:
CREATE STREAM S_out_1 AS
PULSE START = 0s, FREQUENCY = 1s
CONSTRUCT {s0 rdf:type RecMonInc}<NOW>}
FROM S_Msmt [NOW-2s, NOW]->1s
SEQUENCE BY StdSeq AS SEQ1
HAVING FORALL i < j IN SEQ1,?x,?y:
    IF {s0 val ?x}<i> AND {s0 val ?y}<j>
    THEN ?x <= ?y  }

Output stream:
S_out_1 = { RecMonInc(s0)<0s>,
           RecMonInc(s0)<1s>,
           RecMonInc(s0)<2s>,
           RecMonInc(s0)<5s>,
           ...
}

End-user interface

Optique demo video: www.youtube.com/user/optiqueproject/playlists
Diagnostics Dashboard

- Flexible, easily customised, wiki-based
- Includes several query formulation interfaces and results visualization widgets

Optique
Engineers can define and refine queries via several query formulation tools.

- They can express queries by natural language, manipulating objects from domain of interest, or expressing them in SPARQL.
Query formulation interfaces: Visual Query System (VQS)
Query Formulation Interfaces: Natural Language Querying

Query Formulation Interfaces: SPARQL

```
select distinct ?type where {
  ?x rdf:type ?type
}
```
Diagnostics Dashboard

- Widgets allow to visualise query answers, inspect query results, and export result fragment to external diagnostics tools.
- Depending on the type of data, a suitable visualisation paradigm has to be selected.
Result Visualisation Widgets
Conclusions

- What we did
  - described and analysed monitoring of appliances by Siemens Energy, understood issues
  - set requirements to enhance Siemens monitoring routine
  - assessed why and how SW technologies can do the job
  - set a road map to improve existing SW tech to met S. reqs
  - developed the Optique platform solution and evaluation
Thank you for listening

Any questions?

Optique