

Ontology-based Access to Sensor Data Streams

@jpcik Jean-Paul Calbimonte

LSIR EPFL

Ontology Summit 2015

05.03.2015

Why Streams?



Internet of Things Sensor Networks Mobile Networks Smart Devices Participatory Sensing Transportation

"It's a streaming world!"[1]

Financial Data Social Media Urban Planning Health Monitoring Marketing





Sensor Networks and the Web





Linked Data on the Web



COLE POLYTECHNIOU

4

Web of Data Linked Data

pest





Use **ontology** models to **continuously query** realtime data **streams** originated from sensors?



Continuous extensions of RDF

 As you know, "RDF -> standard model for data interchange on the Web" (http://www.w3.org/RDF/)

> <sub₁ pred₁ obj₁> <sub₂ pred₂ obj₂>

Simple model: Subject Predicate Object

- Extend RDF to model data streams
- A data stream is an (infinite) ordered sequence of data items
- A data item is a self-consumable informative unit









A simple ontology...





Data items

- With **data item** we can refer to:
 - 1. A triple

<:alice :isWith :bob>

2. A graph

<:alice :posts :p>
<:p :who :bob>
<:p :where :redRoom>

:graph1



RDF stream model



- A commonly adopted RDF stream model
 - A RDF triple is an event
 - Application time: point-based

<:alice :isWith :bob>:[1]
<:alice :isWith :carl>:[3]
<:bob :isWith :diana>:[6]



Querying RDF: SPARQL

Where is alice?

<:alice :posts :p>

<:p :who :bob>

<:p :where :redRoom>

How about streams?

SELECT ?room WHERE{
 :alice :posts ?post
 ?post :where ?room
}

low many posts about bob?
SELECT COUNT(?post) WHERE{
 ?person :posts ?post
 :post :who :bob
}



Querying data streams – The CQL model



*Stream operators



CQL-like extensions for RDF stream queries





Ontology-based access for data streams



Similar (not equals!) query languages

```
SELECT ?sensor
WHERE {
   STREAM <http://www.cwi.nl/SRBench/observations> [RANGE 10800s SLIDE 600s] {
        ?observation om-owl:procedure ?sensor ;
            om-owl:observedProperty weather:WindSpeed ;
            om-owl:result [ om-owl:floatValue ?value ] .} }
GROUP BY ?sensor HAVING ( AVG(?value) >= "74"^^xsd:float )
```

om-owl:result [om-owl:floatValue ?value] . }

GROUP BY ?sensor HAVING (AVG(?value) >= "74"^^xsd:float)



15

Classification of existing systems

	Model	Continuous execution	Union, Join, Optional, Filter	Aggregates	Time window	Triple window	R2S operator	Sequence, Co- ocurrence
TA-SPARQL	TA-RDF	Х	\checkmark	Limited	Х	Х	Х	Х
tSPARQL	tRDF	Х	\checkmark	Х	Х	Х	Х	Х
Streaming SPARQL	RDF Stream	1	\checkmark	Х	\checkmark	\checkmark	X	Х
C-SPARQL	RDF Stream	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	Rstream only	time function
CQELS	RDF Stream	~	\checkmark	√	\checkmark	\checkmark	lstream only	Х
SPARQLStrea m	(Virtual) RDF Stream	\checkmark	\checkmark	\checkmark	\checkmark	Х	\checkmark	Х
EP-SPARQL	RDF Stream	\checkmark	\checkmark	\checkmark	Х	Х	Х	\checkmark
Instans	RDF	\checkmark	\checkmark	\checkmark	Х	Х	X	Х

Disclaimer: other features may be missing

16

COLE POLYTECHNIQUE

Existing RSP systems (oversimplified!)

- C-SPARQL: RDF Store + Stream processor
 - Combined architecture



- CQELS: Implemented from scratch. Focus on performance
 - Native + adaptive joins for static-data and streaming data





Existing RSP systems (oversimplified!)

- SPARQL_{stream}: Ontology-based stream query answering
 - Virtual RDF views, using R2RML mappings
 - SPARQL stream queries over the original data streams.



- EP-SPARQL: Complex-event detection
 - SEQ, EQUALS operators





The RSP Community





W3C RSP Community Group

http://www.w3.org/community/rsp

discuss standardize Effort to formalize evangelize

our work on RDF stream processing



The RSP Community





RDF Streams Workshop! @ESWC

https://www.w3.org/community/rsp/rsp-workshop-2015/





Why Streams?

- Web standards
 Data discovery
 Data sharing
 Web queries
- Query languages Query answering Efficient processing Query Federation

Is this what we require?

- Semantics Vocabularies Data Harvesting Data linking Matching
- Ontologies
 Expressivity
 Inference
 Rule processing
 Knowledge bases



Workinprogress

Looking back 10 years ago...



"8 requirements of real-time stream processing"^[2] Keep data moving
Query with stream SQL
Handle imperfections
Predictable outcomes
Integrate stored data
Data safety & availability
Partition & scale
Respond Instantaneously

Do we address them? Do we have more requirements? Do we need to do more?



g. SIGMOD Record. 2005

Reactive Systems



Events: Event-Driven *Load:* Scalable *Failure:* Resilient *Users:* Responsive

Keep data moving
Query with stream SQL
Handle imperfections
Predictable outcomes
Integrate stored data
Data safety & availability
Partition & scale
Respond Instantaneously

Do we address them? Do we have more requirements? Do we need to do more?



Reactive RSPs

8 Requirements

Keep data moving Query with stream SQL Handle imperfections Predictable outcomes Integrate stored data Data safety & availability Partition & scale Respond Instantaneously We go beyond only these **Data Heterogeneity Data Modeling Stream Reasoning** Data discovery **Stream data linking Query** optimization ... more

Reactive Principles

Needed if we want to build relevant systems



25

Reactive RSP workflows



Minimal agreements: standards, serialization, interfaces Formal models for RSPs and reasoning Working prototypes/systems!



OSPER - Swiss Experimental Open support platform for environmental research



「managing environmental sensor data &metadata



acquisition heterogeneous sensing devices

processing summarization, filtering, compression, interpolation

querying

analysis co

continuous processing, streaming, geospatial, aggregation

sis pattern discovery, correlation, regression

discovery metadata management, semantics

provision data services, visualization, standards

Multidisciplinary research team

• Real world data + problems

Facilitating research in:

- Precipitation patterns in mountains
- Evaporation in Africa
- Return periods of Natural Hazards
- Stream flows in Alpine catchments
- Permafrost in the Alps



http://swiss-experiment.ch



OpenSense2



Crowdsourcing High-Resolution Air Quality Sensing

Air Pollution 🤶

global concern highly location-dependent time-dependent



Accurate location-dependent and real-time information on air pollution is needed



Integrated air quality measurement platform

Institutional stations

Personal mobile sensors

OpenSense infrastructure

- Heterogeneous devices and data
- Human activity assessment, lifestyle and health data



- Integration of pure statistical models and physical dispersion models
- Better coverage through crowdsensing
- Incentives for crowd data provision
- Finer temporal and spatial resolutions
- Utilitarian approach for trade-off between model complexity, privacy and accuracy
- Higher accuracy of pollution maps models



http://opensense.epfl.ch



OpenIoT FP7

Open Source Cloud solution for the Internet of Things



Established Open-source platform for Io

https://github.com/OpenIotOrg/openiot



- Integrate **sensors** & things with **cloud** computing
- Configure, deploy and use **IoT services**
- Auditing/assessing **privacy** of IoT apps in the cloud
- Semantic annotations of internet-connected objects
- Energy-efficient data harvesting
- Publish/subscribe for continuous processing and sensor data filtering
- Mobility of sensors and **QoS** aspects in IoT

e cases and validation scenarios

Smart Manufacturing

 \Box



Campus Guide



Air Monitoring





http://openiot.eu



Muchas gracias!

@jpcik

Jean-Paul Calbimonte



