Stream Reasoning For Linked Data

M. Balduini, J-P Calbimonte, O. Corcho,
D. Dell'Aglio, E. Della Valle, and J.Z. Pan

http://streamreasoning.org/sr4ld2013

ISWC 2013
Sydney, Australia

Wrap-up and conclusions

Emanuele Della Valle
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org

Share, Remix, Reuse — Legally

- This work is licensed under the Creative Commons Attribution 3.0 Unported License.
- **Your are free:**

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

- **Under the following conditions**

Attribution — You must attribute the work by inserting

  - "[source http://streamreasoning.org/sr4ld2013]" at the end of each reused slide
  - a credits slide stating
    - These slides are partially based on "Streaming Reasoning for Linked Data 2013" by M. Balduini, J-P Calbimonte, O. Corcho, D. Dell'Aglio, E. Della Valle, and J.Z. Pan http://streamreasoning.org/sr4ld2013

- To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/

Agenda

- Revisiting the research challenges
  - Relation with DSMSs and CEPs

  - Reasoning on RDF streams
  - Dealing with incomplete & noisy data
  - Engineering Stream Reasoning Applications
- What's next?
- More on Stream Reasoning at ISWC 2013

Research Challenges

- Relation with DSMSs and CEPs
  - Just as RDF relates to data-base systems?
- Data types and query languages for semantic streams
  - Just RDF and SPARQL but with continuous semantics?
- Reasoning on Streams
  - Theory: formal semantics
  - Efficiency
  - Scalability and approximation
- Dealing with incomplete & noisy data
  - Even more than on the current Web of Data
- Distributed and parallel processing
  - Streams are parallel in nature, data stream sources are distributed, …
- Engineering Stream Reasoning Applications
  - Development Environment
  - Integration with other technologies
  - Benchmarks as rigorous means for comparison

Revisiting the research challenges

Relation with DSMSs and CEPs

- Achievement
  - Somehow just as RDF, SPARQL, and OWL relate to data-base systems

DB → Semantic Web	DSMS/CEP → Semantic Web
Relational data → RDF	Data streams → RDF Streams
SQL → SPARQL	CQL/EPL/… → C-SPARQL/EP-SPARQL/…
Schema → OWL	Schema → OWL

  - But with some differences
    - Queries are registered → opportunity for query optimizations
    - Many application requires a network of queries → opportunity for inter-query optimizations
- Issues
  - It is time to bring Stream Reasoning to the Web
    - Volatile URIs
    - Serialization of RDF streams
    - Protocols: HTTP, Web sockets

Data types for semantic streams - Achievements

- **RDF streams** introduced as new data type in the Semantic Web and Linked Data research
- W3C RDF stream processor **community group started** to jointly work out a recommendation in 2014
  - http://www.w3.org/community/rsp/

Data types for semantic streams - Issues

- Multiple notions of RDF stream proposed
    - Ordered sequence (implicit timestamp)
    - One timestamp per triple (point in time semantics)
    - Two timestamps per triple (interval base semantics)
- Comparison between existing approaches

System	Data item	Time model	# of timestamps
INSTANS	triple	Implicit	0
C-SPARQL	triple	Point in time	1
$\text{SPARQL}_{\text{stream}}$	triple	Point in time	1
CQELS	triple	Point in time	1
Sparkwave	triple	Point in time	1
Streaming Linked Data	RDF graph	Point in time	1
ETALIS	triple	Interval	2

- More investigation is required to agree on an RDF stream model

Query languages for semantic streams - Achievements

- **Languages for continuous querying** of **and event processing** on RDF streams proposed
- Window base selection outperforms filter base selection
- Dynamic optimization of query plans and incremental evaluation is possible
- Multiple RDF stream processor **prototypes** implemented and deployed
- W3C RDF stream processor **community group started** to jointly work out a recommendation in 2014
  - http://www.w3.org/community/rsp/

Query languages for semantic streams - Issues

- Different syntax for S2R operator
- Semantics of query languages is similar, but not identical
- Lack of R2S operator in some cases
- Different support for time-aware operators

Query languages for semantic streams - Issues

- Comparison between existing approaches

System	S2R	R2R	Time-aware	R2S
INSTANS	Based on time events	SPARQL update	Based on time events	Ins only
C-SPARQL Engine	Logical and triple-based	SPARQL 1.1 query	timestamp function	Batch only
SPARQL$_{stream}$	Logical and triple-based	SPARQL 1.1 query	no	Ins, batch, del
CQELS	Logical and triple-based	SPARQL 1.1 query	no	Ins only
Sparkwave	Logical	SPARQL 1.0	no	Ins only
Streaming Linked Data	Logical and graph-based	SPARQL 1.1	no	Batch only
ETALIS	no	SPARQL 1.0	SEQ, PAR, AND, OR, DURING, STARTS, EQUALS, NOT, MEETS, FINISHES	Ins only

- Is it time to converge on a standard?

Query languages for semantic streams - Issues

- The **existing engines**
  - adopts **different architectural** choices and it is still unclear when each choice is best
    - C-SPARQL, ETALIS, SPARQL$_{stream}$ are wrappers for existing systems thus they are more reliable and maintainable
    - CQELS, Streaming Linked Data, INSTANS, Sparkwave are native implementations, thus they are more efficient and offer optimizations not possible in the other system
  - They have **different operational semantics**
    - for more information check out the ISWC 2013 evaluation track for "*On Correctness in RDF stream processor benchmarking*" by Daniele Dell'Aglio, Jean-Paul Calbimonte, Marco Balduini, Oscar Corcho and Emanuele Della Valle

Reasoning on Streams - Achievements

- **Stream Reasoning** research field is getting momentum
- Efficient **continuous reasoning algorithm** on RDF streams for RDFS, RDFS++, EL++, Answer Set Programming were proposed
- Multiple Stream Reasoning **proofs of concept** were implemented

Reasoning on Streams - Issues

- Issues
  - Theory still largely based on one-time semantics
    - Continuous reasoning for the following topics requires more investigations
      - Continuous conjuctive queries under OWL2QL entailment regime
      - Union of Continuous conjuctive queries under OWL2QL entailment regime
      - Continuous queries including negation (in all its possible forms)
      - Continuous recursive querying under expressive entailment regimes
      - Modelling in the ontology aggregates and functions
    - Logic based time-management
      - More expressive specification, e.g., calendar algebra
      - Windows that logically resize at runtime
  - Lack of prototypes that go beyond proof of concept
  - Explore more reasoning form beyond Q/A

Revisiting the research challenges

Dealing with incomplete & noisy data

- Data streams are incomplete and noisy!
- Achievements
  - Reasoning can help dealing with incompleteness
  - Initial works on inductive stream reasoning explored relation learning as a way to cope with those problematic aspects
- Issues
  - More research required!

Distributed and parallel processing

- Data streams are parallel and distributed in nature!
- Achievements
  - Proof of concept implemented on S4 and Storm
- Issues
  - More research required!

Engineering Stream Reasoning Applications

- Achievements
  - Deployments for
    - semantic sensor networks
    - social media analytics
    - City Data Fusion
  - Multiple benchmarks proposed
- Issues
  - It is still unclear when and where it is convenient to adopt Stream Reasoning solutions
  - Benchmarks too focused on throughput; correctness and memory allocation cost, too

Revisiting the research challenges

Wrap-up

- Data types and query languages for semantic streams
  - Notion of RDF stream :-)
  - Languages for continuous querying :-)
  - Prototypes :-)
  - Standardization :-)
- Reasoning on RDF streams
  - Theory :-|
  - Algorithms :-)
  - Prototypes :-(
- Dealing with incomplete & noisy data
  - Theory :-(
  - Algorithms :-|
  - Prototypes :-(
- Engineering Stream Reasoning Applications
  - Deployments :-)
  - Benchmarks :-|

What's next? order matters!

- Observation: order reflects recency, relevance, trustability …

Types of orders	No	Yes
Combinations	Continuous top-k Q/A	Order-aware reasoning
Relevance, Trustability, etc.	Top-k Q/A	Top-k Reasoning
Recency	DSMS/CEP	Stream reasoning
Indexes	Traditional solutions	Scalable reasoning

Semantic Technologies

Emanuele Della Valle, Stefan Schlobach, Markus Krötzsch, Alessandro Bozzon, Stefano Ceri, Ian Horrocks: **Order matters! Harnessing a world of orderings for reasoning over massive data**. Semantic Web 4(2): 219-231 (2013)

More on Stream Reasoning at ISWC 2013

- Tuesday Afternoon - **OrdRing 2013**
  - 2nd International *Workshop* on Ordering and Reasoning
  - Open Door Meeting of the *W3C RDF Stream Processing Community Group*
- Wednesday Evening - Poster session
  - M. Balduini et al. *A Restful Interface for RDF Stream Processors*
  - L. Fischer et al. *Network-Aware Workload Scheduling for Scalable Linked Data Stream Processing*
- Thursday - 11:00-12:40 **Track on Streams**
  - M. Balduini et al. *Social listening of City Scale Events using the Streaming Linked Data Framework*
  - D. Le Phuoc et al. *Elastic and scalable processing of Linked Stream Data in the Cloud*
  - *S.* Tallevi-Diotallevi et al. *Real-time Urban Monitoring in Dublin using Semantic and Stream Technologies*
  - D. Dell'Aglio et al. *In Correctness in RDF stream processor benchmarking*
  - D. Gerber et al. *Real-time RDF extraction from unstructured data streams*

Stream Reasoning For Linked Data

M. Balduini, J-P Calbimonte, O. Corcho, D. Dell'Aglio, E. Della Valle, and J.Z. Pan

http://streamreasoning.org/sr4ld2013

ISWC 2013
Sydney, Australia

POLITECNICO MILANO

UNIVERSIDAD POLITECNICA MADRID

Wrap-up and conclusions

Emanuele Della Valle
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org