Engineering of Web Stream Processing Applications

Emanuele Della Valle[(✉)], Riccardo Tommasini, and Marco Balduini

Department of Electronic, Informatics and Bioengineering, Politecnico di Milano,
Piazza Leonardo Da Vinci 32, 20191 Milan, Italy
{emanuele.dellavalle,riccardo.tommasini,marco.balduini}@polimi.it

Abstract. The goal of the tutorial is to outline how to develop and deploy a stream processing application in a Web environment in a reproducible way. To this extent, we intend to (1) survey existing research outcomes from the Stream Reasoning/RDF Stream Processing that arise in querying and reasoning on a variety of highly dynamic data, (2) introduce stream reasoning techniques as powerful tools to use when addressing a data-centric problem characterized both by variety and velocity (such as those typically found on the modern Web), (3) present a relevant Web-centric use-case that requires to address simultaneously data velocity and variety, and (4) guide the participants through the development of a Web stream processing application.

1 Introduction

More and more streams of information are becoming available on the Web. A variety of sources give origin to data streams including social networks, mobile phones, smart homes, healthcare devices and other modern infrastructures. A significant portion of this data belongs to the Web-of-Services and to the Web-of-Things ecosystems where data is published and consumed using Web standards and technologies.

From new opportunities arise new challenges. A common problem in the scenarios illustrated above is how to integrate such data and how to enable the creation of new knowledge. Reasoning techniques are a possible solution. However, while reasoners scale up in the classical, static domain of ontological knowledge, reasoning upon rapidly changing information has received attention only in the last decade [1]. The combination of reasoning techniques with data streams gives rise to Stream Reasoning. i.e., reasoning on highly dynamic flows of informations [2]. This is a high impact research area that has already started to produce results relevant to both the Semantic Web [3] and stream processing communities [4].

Learning Goals: The contents of this tutorial can be relevant for RW 2018 summer school attendees as it focuses on the engineering aspects of developing and deploying applications that use streaming data to create new knowledge.

This tutorial aims at introducing different existing approaches for querying and reasoning over data streams and providing guidelines to develop and deploy Stream Reasoning applications. In particular, the tutorial offers to the audience: (i) an overview of the use cases and the scenarios where Stream Reasoning can be used (with the advantages it brings); (ii) an overview of the current state of the art in this emerging area, with techniques and tools developed by several research groups (including but not limited to presenters' ones); (iii) a focus on a subset of the technologies to perform complex reasoning over dynamic data.

Related Events: The tutorial follows from the Stream Reasoning for Linked Data (SR4LD) tutorial series, successfully held at ESWC 2011, SemTech 2011, ISWC 2013, ISWC 2014, ISWC 2015; the RDF Stream Processing (RSP) tutorial series at ESWC 2014 and ISWC 2016; the Stream Reasoning: Managing Velocity and Variety in Big Data tutorial at DEBS 2016; and the tutorial on How to Build a Stream Reasoning Application co-located with ISWC 2017.

Technologies are now mature and reliable enough to make a step further and build a tutorial with a stronger focus on hands-on sessions. Therefore, this tutorial not only surveys existing research outcomes and introduce existing tools, but it also presents a relevant Web-centric use-case and guide the participants through application development of a Web stream processing application.

Audience: The tutorial targets researchers and practitioners interested in approaching the topic of web stream processing (both querying and reasoning) and who want to understand the current state-of-the-art as well as the future directions. The technologies and topics on this tutorial are relevant for people from IoT and sensor communities, as well as social media, pervasive health, oil industry, etc., who produce massive amounts of streaming data.

2   Presenters

Emanuele Della Valle[1] holds a PhD in Computer Science from the Vrije Universiteit Amsterdam and a Master degree in Computer Science and Engineering from Politecnico di Milano. He is assistant professor at the Department of Electronics, Information and Bioengineering of the Politecnico di Milano.

In more than 15 years of research, his research interests covered Big Data, Stream Processing, Semantic technologies, Data Science, Web Information Retrieval, and Service Oriented Architectures. He started the Stream Reasoning research field [2] by focusing the research community's interest on this theme through a number of personal initiatives (workshops, projects, journal papers) that contributed to establishing Stream Reasoning as a recognized research and industrial sector. Stream Reasoning is positioned at the intersection between Stream Processing and Artificial Intelligence. In the Big Data era, it is a method to tame simultaneously the velocity (analyzing data streams to enable real-time decisions) and variety (integrating heterogeneous data) dimensions of Big Data. Starting from 2011, he organized more than 15 tutorials and workshops on

---

[1] http://emanueledellavalle.org/.

Stream Reasoning[2]. His work on Stream Reasoning was applied in analyzing Social Media, Mobile Telecom and IoT data streams in collaboration with Telecom Italia, IBM, Siemens, Oracle, Indra, and Statoil. In 2015, he started up a company (Fluxedo) to commercialize the open source results of Stream Reasoning research.

3   Tutorial Structure

1. **Introduction to WEB Stream Processing**
   1.1 From challenges to opportunities.
   1.2 Paradigmatic shit to continuous semantics.
   1.3 The limited ability of existing Stream Processing systems to address variety.
   1.4 The Stream Reasoning research question.
   1.5 Existing Stream Reasoning systems (quick introduction and high-level comparison).
   1.6 Anatomy of a Web Stream Processing Application.
   1.7 Introduction of the tutorial running example.
2. **Publishing and Describing Streams on the Web**
   2.1 Streams on the Web: creation and publications challenges.
   2.2 VoCaLS: a vocabulary to describe RDF Streams on the WEB [5,6]
   2.3 VoCaLSing existing data streams to catalog them[3].
   2.4 TripleWave: a framework to publish RDF streams on the Web [7][4].
   2.5 Hands-on VoCaLS and TripleWave.
3. **Processing Web Streams**
   3.1 Querying and Reasoning on Web Streams: RDF Stream Processing (RSP) languages [8].
   3.2 Overview and comparison RSP query languages: C-SPARQL [9], CQELS-QL [10], $SPARQL_{stream}$ [11], RSP-QL [8].
   3.3 Overview and comparison RSP engines: C-SPARQL E.[5], CQELS[6], Yasper[7];
   3.4 Modeling an RSP Service with VoCaLS.
   3.5 Hands-on RSP engine.
4. **Building a Web Streams Application**
   4.1 Introduction of the RSP Services [12].
   4.2 Designing the Application.
   4.3 Using Jupyter notebooks and the RSP-Kernel[8] to implement the Application.
   4.4 Hands on session.
5. **Wrap-up and discussion**
   5.1 On-going research trends, real-world deployments.
   5.2 Open problems and future directions.

---

[2] http://streamreasoning.org/events.
[3] https://w3id.org/rsp/vocals.
[4] https://github.com/streamreasoning/TripleWave.
[5] https://github.com/streamreasoning/CSPARQL-engine.
[6] https://github.com/danhlephuoc/cqels.
[7] https://github.com/streamreasoning/yasper/.
[8] https://github.com/riccardotommasini/rsp-kernel.

References

1. Dell'Aglio, D., Della Valle, E., van Harmelen, F., Bernstein, A.: Stream reasoning: a survey and outlook. Data Sci. **1**(1–2), 59–83 (2017)
2. Della Valle, E., Ceri, S., van Harmelen, F., Fensel, D.: It's a streaming world! reasoning upon rapidly changing information. IEEE Intell. Syst. **24**(6), 83–89 (2009)
3. Margara, A., Urbani, J., van Harmelen, F., Bal, H.E.: Streaming the web: reasoning over dynamic data. J. Web Sem. **25**, 24–44 (2014)
4. Della Valle, E., Dell'Aglio, D., Margara, A.: Taming velocity and variety simultaneously in big data with stream reasoning: tutorial. In: DEBS, pp. 394–401. ACM (2016)
5. Sedira, Y.A., Tommasini, R., Della Valle, E.: Towards VoIS: a vocabulary of interlinked streams. In: ISWC DeSemWeb (2017)
6. Dell'Aglio, D., Le Phuoc, D., Le-Tuan, A., Ali, M.I., Calbimonte, J.-P.: On a web of data streams. In: ISWC DeSemWeb (2017)
7. Mauri, A., et al.: Triplewave: spreading RDF streams on the web. In: The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, 17–21 October 2016 (2016)
8. Dell'Aglio, D., Della Valle, E., Calbimonte, J., Corcho, Ó.: RSP-QL semantics: a unifying query model to explain heterogeneity of rdf stream processing systems. Int. J. Semant. Web Inf. Syst. **10**(4), 17–44 (2014)
9. Barbieri, D.F., Braga, D., Ceri, S., Della Valle, E., Grossniklaus, M.: C-SPARQL: a continuous query language for RDF data streams. Int. J. Semant. Comput. **4**(1), 3–25 (2010)
10. Le-Phuoc, D., Dao-Tran, M., Xavier Parreira, J., Hauswirth, M.: A native and adaptive approach for unified processing of linked streams and linked data. In: Aroyo, L. (ed.) ISWC 2011. LNCS, vol. 7031, pp. 370–388. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25073-6_24
11. Calbimonte, J., Corcho, Ó., Gray, A.J.G.: Enabling ontology-based access to streaming data sources. In: The Semantic Web - ISWC 2010 - 9th International Semantic Web Conference, ISWC, Revised Selected Papers, Part I, Shanghai, China, 7–11 November 2010, pp. 96–111 (2010)
12. Balduini, M., Della Valle, E.: A restful interface for RDF stream processors. In: Proceedings of the ISWC 2013 Posters & Demonstrations Track, Sydney, Australia, 23 October 2013, pp. 209–212 (2013)