Editorial: Special Issue on Stream Reasoning

Daniele Dell'Aglio [a], Thomas Eiter [b], Fredrik Heintz [c] and Danh Le-Phuoc [d]
[a] *Department of Informatics, University or Zurich, Zurich, Switzerland*
E-mail: dellaglio@ifi.uzh.ch
[b] *Institute of Logic and Computation, TU Vienna, Vienna, Austria*
E-mail: eiter@kr.tuwien.ac.at
[c] *Department of Computer and Information Science, Linköping University, Linköping, Sweden*
E-mail: fredrik.heintz@liu.se
[d] *Open Distributed Systems, TU Berlin, Berlin, Germany*
E-mail: danh.lephuoc@tu-berlin.de

Abstract. This editorial introduces the Special Issue on Stream Reasoning, part of the Semantic Web Journal.

Keywords: Stream reasoning, Data streams, Continuous queries

In the recent years, there has been an increasing speed and volume of data production in several domains, such as Internet of Things, Social Networks and Smart Cities. An interesting feature of such data is that its utility is often related to time: the sooner the data is processed, the higher is the value. Consequently, techniques to process huge amounts of heterogeneous streaming data in a continuous fashion are getting more and more important.

Data Stream Management Systems and Complex Event Processing [1] provide foundations for continuously querying and monitoring data streams. However, other kinds of processing are desired, such as deductive and inductive inference, where domain models provide background knowledge and context.

With Stream Reasoning we refer to a research trend that aims at studying how to introduce reasoning processes in scenarios involving streams. Up to now, we have seen two types of stream reasoning: reasoning over streams and reasoning about streams. Reasoning over streams is the incremental reasoning over information that is continuously produced and made available, typically in the context of a static domain model. Reasoning about streams is the reasoning with streams as first class entities. Stream Reasoning introduces new challenges with regard to traditional reasoning over static or slowly changing data: the data is made available in a continuous way, possibly from different sources; time is a first-class citizen; responsiveness is a key requirement; and deletion mechanisms are often required in order to satisfy performance and space constraints.

The term *Stream Reasoning* emerged a decade ago [2], and developed as a trend involving different areas of computer science, such as semantic web, databases, robotics and knowledge representation. The multidisciplinary nature of Stream Reasoning offers an ideal setting for exchanging ideas and expertise. To push this process, community members contributed to tutorials, summer schools and workshops in important and prestigious venue. Additionally, in the last five years we have organised four invitation-based events to let the community meet and discuss. We observed an increasing trend in terms of attendance and advances in the research. So far, the majority of workshop presentations focused on logic-based reasoning. However, we observe an increasing number of studies focusing on the combination of logic-based reasoning with statistics-based inference techniques to create data processing pipelines. For example, in computer vision, tasks such as object detection, semantic segmentation, and image understanding get advantage of knowledge graphs to improve the quality of the results. In the database and knowledge management communities, the process of creating and enriching knowledge graphs uses statistics-based inference techniques, e.g. deep convolutional or recurrent neural networks, and Markov logic networks, in conjunction with logic reasoners, e.g. Datalog and first-order logic reasoners. To integrate uncertainty coming from learning models with logic reasoning processes, approaches such as Softlogic [3] or extensions of Answer Set Programming (ASP) [4, 5] have been proposed to make the best of both worlds. One approach is to compute probability distributions over MTL-formulas as a result of incomplete information [6] and another is to integrate probabilistic inference more directly by extending the syntax of MTL with probabilistic statements that are computed through probabilistic inference outside the logic [7]. While most of these studies focused on processing static data, we envision that the development of new methods and techniques based on statistics-based inference techniques and logic-based reasoning will be a driving force in stream reasoning in years to come.

Stream reasoning research should also look at distribution. Most big data stream processing platforms, such as Apache Flink and Apache Spark, achieve scalability by supporting parallelization of complex iterative computing workflows and execution on cluster or cloud platforms. Edge computing is pushing the computation closer to the data sources, e.g. a resource-constrained edge device can efficiently store and process 30 million of RDF triples with 80MB of RAM [8]. Since stream data is often generated in a distributed fashion, these decentralized computing paradigms are suitable and desirable choices to develop stream reasoning solutions. This is not straightforward, since there are several theoretical and technical challenges to build scalable and robust distributed stream reasoning systems. For example, [9] proposes to use ASP to formalise distributed reasoning workflows (including data stream), by representing the federation among heterogeneous reasoners. Systems able to perform reasoning—both logic-based and statistics-based—on streams may find application in a wide are of use cases, including Internet of Things, autonomous driving and personalized medicine.

We ran this this special issue to collect the most recent and advanced research on stream reasoning. We received eight submissions, and after the review process, we selected two contributions.

In "Enhancing the Scalability of Expressive Stream Reasoning via input-driven Parallelization", the authors study how to let stream reasoning scale. The problem they focus on how to parellelise the execution of non-monotonic rules expressed in a subset of ASP. As a solution, they propose a novel input splitting technique based on the analysis of the relation between input data streams and ASP rules registered into the system. Extensive experiments analyse the performance of the proposed technique while considering different class of rules, including recursive and negation rules.

In "Ontology-mediated query answering over temporal and inconsistent data", the authors investigate how to cope with temporal and insistent data in stream reasoning. The paper considers a temporal query language that combines conjunctive queries with operators of propositional linear temporal logic (LTL) under three consistency-tolerant semantics introduced in query consistent description logic knowledge bases. The key contributions of the authors are complexity analysis of an extensive number of cases of $\mathcal{EL}_\perp$ and DL-Lite$_\mathcal{R}$. Also, another interesting contribution of the paper is its proposal for practical implementations.

We would like to thank all the people that contribute to this special issue. We thank all the authors that submitted contributions, showing another time that there is an ongoing active and passionate research on stream reasoning. We thank the Editors-in-Chief, Pascal Hitzler and Krzysztof Janowicz, which gave us the opportunity to run this special issue and supported us in the whole process. Finally, we are very grateful to the reviewers, whose effort produced valuable insights: Eva Blomqvist, David Bowden, Jean Paul Calbimonte, Daniel de Leng, Ali Intizar, Robin Keskisärkkä, Thu Le-Pham, Alessandro Margara, Alessandra Mileo, Boris Motik, Özgür Öezçep, Josiane Parreira, Konstantin Schekotihin, Patrik Schneider, Veronika Thost, Jacopo Urbani and Guohui Xiao.

References

[1] G. Cugola and A. Margara, Processing flows of information: From data stream to complex event processing, *ACM Comput. Surv.* **44**(3) (2012), 15:1–15:62. doi:10.1145/2187671.2187677.

[2] D. Dell'Aglio, E. Della Valle, F. van Harmelen and A. Bernstein, Stream reasoning: A survey and outlook, *Data Science* **1**(1–2) (2017), 59–83, Available at: https://content.iospress.com/articles/data-science/ds006. doi:10.3233/DS-170006.

[3] M.W. Chekol, G. Pirrò, J. Schoenfisch and H. Stuckenschmidt, Marrying Uncertainty and Time in Knowledge Graphs, in: *AAAI Conference on Artificial Intelligence, San Francisco, CA, USA.*, 2017, pp. 88–94.

[4] J. Suchan and M. Bhatt, Semantic Question-Answering with Video and Eye-Tracking Data: AI Foundations for Human Visual Perception Driven Cognitive Film Studies, in: *International Joint Conference on Artificial Intelligence, (IJCAI), New York, NY, USA*, 2016, pp. 2633–2639.

[5] J. Suchan, M. Bhatt, P.A. Walega and C.P.L. Schultz, Visual Explanation by High-Level Abduction: On Answer-Set Programming Driven Reasoning About Moving Objects, in: *AAAI Conference on Artificial Intelligence, New Orleans, LA, USA*, 2018, pp. 1965–1972.

[6] D. de Leng and F. Heintz, Approximate Stream Reasoning with Metric Temporal Logic under Uncertainty, in: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Honolulu, HI, USA*, 2019.

[7] M. Tiger and F. Heintz, Stream Reasoning Using Temporal Logic and Predictive Probabilistic State Models, in: *International Symposium on Temporal Representation and Reasoning, (TIME) Kongens Lyngby, Denmark*, IEEE Computer Society, 2016, pp. 196–205. doi:10.1109/TIME.2016.28.

[8] A. Le-Tuan, C. Hayes, M. Wylot and D. Le-Phuoc, RDF4Led: an RDF engine for lightweight edge devices, in: *International Conference on the Internet of Things, (IOT)*, 2018, pp. 2:1–2:8. doi:10.1145/3277593.3277600.

[9] M. Dao-Tran and T. Eiter, Streaming Multi-Context Systems, in: *International Joint Conference on Artificial Intelligence, (IJCAI), Melbourne, Australia*, 2017, pp. 1000–1007. doi:10.24963/ijcai.2017/139.