Riccardo Tommasini
Pieter Bonte
Fabiano Spiga
Emanuele Della Valle

Streaming Linked Data

From Vision to Practice

Springer

Streaming Linked Data

Riccardo Tommasini • Pieter Bonte •
Fabiano Spiga • Emanuele Della Valle

Streaming Linked Data

From Vision to Practice

Springer

Riccardo Tommasini
LIRIS Lab
Institut National des Sciences Appliquées
Villeurbanne, France

Fabiano Spiga
Tallinn, Estonia

Pieter Bonte
Department of Information Technology
Ghent University
Ghent, Belgium

Emanuele Della Valle
Department of Electronics, Informatics and
Bioengineering
Politecnico di Milano
Milano, Italy

ISBN 978-3-031-15370-9 ISBN 978-3-031-15371-6 (eBook)
https://doi.org/10.1007/978-3-031-15371-6

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my family, who always supported me in my journey, and to those who saw the potential in a rough stone and help polishing its talent. In particular, to Sherif Sakr (R.I.P.)

- Riccardo

To my loving wife and beautiful daughters, my sources of inspiration. Thank you for accompanying me along my path.

- Pieter

To my niece Sofia and to all the children of this world not suitable for children, who inspire men of good will to make it a better place.

- Fabiano

To my wife Simona and our wonderful daughter Giulia. Panta Rei, thus seize the day.

- Emanuele

Foreword

No matter what area of knowledge we may be talking about, it is always a challenging task to compile into a single book the most relevant knowledge accumulated during more than a decade of active research and application. Indeed, this books compiles and summarizes into a coherent storyline the work that was initiated more than 10 years ago by a group of researchers who had already embraced the Semantic Web principles and concepts to deal with the variety of data and identified the need to start addressing those aspects related to the velocity of data in many real-world contexts.

Having been part of this community since its early stages, including the supervision of PhD students and other early-stage researchers, and having collaborated with some of the authors of this book, as well as many others whose works are referenced in it, I must admit that I always found it difficult to collect in a single folder all the relevant reference materials that I should be providing to the next PhD student in this topic or to my students at the various research masters on Artificial Intelligence and Data Science that I participate in. Rooted on the work done for many years in tutorials around the topics of stream reasoning, RDF stream processing, and Streaming Linked Data, this book provides a clear solution for this problem and will become one of the key references in my courses around Semantic Web topics.

I want to emphasize here that this is not just a book that compiles disconnected pieces of research. By taking a quick look at the table of contents, the readers will immediately realize that the authors have gone far beyond generating an incremental compendium of the research works that have been done over the years, what would be relatively easy and would require little effort. They have created a coherent story where readers will first be able to understand the principles on which this research area has been established (around the aspects related to variety and velocity). Then they will be able to browse and learn from the historical evolution of approaches, well rooted on theory and practice. They will later understand what is the life cycle of streaming linked data projects and initiatives, based on the experience gathered by the authors in numerous projects. And finally they will understand which efforts have been done for the evaluation and continuous improvement of the theoretical frameworks and development systems around this area. All of this will provide those interested in getting up to speed into the world of streaming linked data to have a single source of reference.

And I am especially happy when seeing the latest chapter with the exercise book, something that in many situations we forget about when we are creating a book that may not only be used as a reference book for researchers but also as an initial textbook for practitioners. I will definitely use also the contents of the final chapter, the exercise book, with my Master students.

In summary, I recommend this book to all those interested in understanding how a semantic-based approach (and its corresponding technologies) can be applied in contexts that go beyond static data, either because they want to start applying these techniques in practice or because they are interested in pursuing research in the area. And I also recommend it for those like me who are teaching students at the Master level and want to provide them with an advanced textbook on a topic for which there was not such compiled material before.

Madrid, Spain
June 2022

Oscar Corcho
Ontology Engineering Group
R&D Center for Artificial Intelligence (AI.nnovation Space)
Universidad Politécnica de Madrid

Preface

A fruitful sequence of tutorials prompted a rationalization of the research journey that, across 10 years, shepherded several researchers and led to numerous contributions. In 2008, when the Stream Reasoning research question was firstly envisioned, data velocity was starting to emerge. Highly dynamic resources did not yet populate the Web, but social networks and the Internet of Things pushed toward real-time data management. With the Large Knowledge Collider (LarKC) project, the vision became a community, which gained an identity within the Semantic Web context.

The second generation of researchers developed Stream Reasoning into RDF Stream Processing. Seminal contributions strengthen the theoretical foundation of the field, which expanded beyond the Semantic Web borders to mix with inductive and deductive artificial intelligence. Stream Reasoning can now count contributions in (Temporal) logic, Robotics, Machine Learning, and data integration.

RDF Stream Processing ultimately led to Streaming Linked Data, which this book is about. While the third generation of stream reasoning researchers is approaching seniority, best practices are emerging, and stream processing is rising as the industry's de facto standard for Big Data engineering and analytics.

This book was envisioned as a resource that collects research efforts and could guide the next generation of researchers, practitioners, and industrial stakeholders. Our goal is to explain the value of data integration when it does not neglect the time-varying nature of data and the continuous essence of some information needs.

This volume focuses on theoretical and practical aspects. It provides a comprehensive perspective of algorithms, systems, and applications for streaming linked data. Finally, it introduces the readers to RSP4J, a novel open-source project that aims to gather community efforts in software engineering and empirical research.

An introductory chapter positions the work by explaining what motivates the design of specific techniques for processing data streams using Web technologies. The prerequisite chapter briefly summarizes the necessary background concepts and models needed to understand the remaining content of the book. Subsequently, Chap. 3 focuses on processing RDF streams, taming data velocity in an open environment characterized by high data variety. It introduces query answering algorithms with RSP-QL and analytics functions over streaming data. Chapter 4 focuses on publishing streams and events on the Web as a prerequisite aspect to make data findable and accessible to applications. Chapter 5 touches on the problems of benchmarking systems that analyze Web streams to foster technological progress. It surveys existing benchmarks and introduces guidelines that may support new practitioners in approaching the issue of continuous analytics. Chapter 6 presents a list of examples and exercises that will help whoever approaches the area to get used to its practices and become confident in its technological stack.

This book provides a comprehensive overview of core concepts and technological foundations of continuous analytics of Web streams. It presents real-world examples and names systems and applications. Therefore, it is of particular interest to students, lecturers, and researchers in Web data management and stream data management.

In practice, this book would not have been possible without several people: (i) the authors, who are linked by more than just respect and professional appreciation; their research groups, namely, Marco Balduini, Daniele Dell'Aglio, Femke Ongenae, Alessandro Margara, Ahmed Awad, and Radwa ElShawi; and their collaborators within the stream reasoning community, who contributed to the vision scientifically, i.e., Jean-Paul Calbimonte, Danh LePhouc, Robin Kerskisärkka,, Oscar Corcho, Alessandra Mileo, Ali Intizar, Jacopo Urbani, Boris Motik, Darko Anicic, Thomas Eiter, Patrik Schneider, and many many more.

The genesis of this book dates back to the end of 2019. A series of unfortunate events, including COVID-19 and the premature demise of Prof. Sherif Sakr,[1] slow down the writing process. Yet, we are glad to present this work to the community.

Last but not least, we thank our institutions, INSA Lyon and LIRIS Lab, Ghent University IMEC, Politecnico di Milano, and the University of Tartu. In particular, Riccardo Tommasini and Fabiano Spiga acknowledge support from the European Social Fund via IT Academy program. Pieter Bonte acknowledges his support from Research Foundation Flanders (FWO) through his postdoctoral fellowship (1266521N).

Lyon, France	Riccardo Tommasini
Gent, Belgium	Pieter Bonte
Tartu, Estonia	Fabiano Spiga
Milan, Italy	Emanuele Della Valle
April 2022	

[1] Rest in Peace.

Contents