Preface for the 4th Edition of the International Knowledge Graph Construction Workshop

David Chaves-Fraga¹,²,³, Anastasia Dimou¹,³,⁴, Ana Iglesias-Molina², Umutcan Serles⁵ and Dylan Van Assche⁶

¹KU Leuven, Department of Computer Science, Sint-Katelijne-Waver, Belgium
²Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Spain
³Flanders Make – DTAI-FET
⁴Leuven.AI – KU Leuven institute for AI, B-3000 Leuven, Belgium
⁵Semantic Technology Institute Innsbruck, Universität Innsbruck, Austria
⁶IDLab, Dept of Electronics and Information Systems, Ghent University – imec, Belgium

More and more knowledge graphs are constructed for private use, e.g., the Amazon Product Graph [1] or the Fashion Knowledge Graph by Zalando¹, or public use, e.g., DBpedia² or Wikidata³. While techniques to automatically construct KGs from existing Web objects exist (e.g., scraping Web tables), there is still room for improvement. So far, constructing knowledge graphs was considered an engineering task, however, more scientifically robust methods keep on emerging. These methods were widely questioned for their verbosity, low performance or difficulty of use, while the data sources’ variety and complexity cause further syntax and semantic interoperability issues.

Declarative methods (mapping languages) for describing rules to construct knowledge graphs and approaches to execute those rules keep on emerging. Nevertheless constructing knowledge graphs is still not a straightforward task because several existing challenges remain and yet the barriers to construct knowledge graphs are not lowered enough to be easily and broadly adopted by industry. These reasons and the vastly populated knowledge graph construction W3C Community Group⁴ show that there are still open questions that require further investigation to come up with groundbreaking solutions.

Addressing challenges related to knowledge graphs construction requires well-founded research, including the investigation of concepts and development of tools as well as methods for their evaluation. R2RML was recommended in 2012 by W3C, and since then, different extensions, alternatives and implementations were proposed [2, 3, 4]. Certain approaches followed the ETL-like paradigm, e.g., SDM-RDFizer [5], RocketRML [6], and FunMap [7], while others the query-answering paradigm, e.g., Ultrawrap [8], Morph-RDB [9] and Ontop [10].

Fourth International Workshop On Knowledge Graph Construction Co-located with the ESWC 2023, 28th May 2023, Crete, Greece

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

³https://engineering.zalando.com/posts/2018/03/semantic-web-technologies.html
²https://www.dbpedia.org/resources/knowledge-graphs/
⁴https://www.wikidata.org/wiki/Wikidata:Main_Page
⁴http://w3.org/community/kg-construct
Besides R2RML-based extensions, alternatives were proposed, e.g., SPARQL-Generate [11] and ShExML [12], as well as methods to perform data transformations while constructing knowledge graphs, e.g., FnO [13] and FunUL [14].

The fourth edition of the knowledge graph construction workshop\(^5\) has a special focus on the benchmarking of knowledge graph construction methods. We aim to put the focus of this edition to methods for efficient construction of KGs that evaluate or analyze the trade-offs of different approaches and when to choose which system. It also included:

- **Keynote.** The workshop includes the keynote from Benjamin Cogrel (Ontopic): "From Ontop to Ontopic: a virtual-first perspective on knowledge graph construction"
- **The Knowledge Graph Construction Challenge.** For the first time, the workshop includes a challenge that aims at benchmarking systems to find which RDF graph construction system optimizes for metrics e.g. execution time, CPU, memory usage, or a combination of these metrics.

The final goal of the event is to provide a venue for scientific discourse, systematic analysis and rigorous evaluation of languages, techniques and tools, as well as practical and applied experiences and lessons-learned for constructing knowledge graphs from academia and industry.

Fourteen papers were submitted. The reviews were open and public, and hosted at Open Review\(^6\). Each paper received at least three reviews from reviewers with different background and status. Each paper received a review from a senior, a junior and an industry researcher.

Six papers were accepted and four were conditionally accepted. Five of the accepted papers were long papers and the other five were short papers. The following papers were accepted for publication and presented at the workshop:

- Composable Semantic Data Transformation Pipelines with Chimera [15].
- Test-driven Knowledge Graph Construction [16].
- Towards Semantic Interpretation of Structured Data Sources in Privacy-Preserving Environments [17].
- Declarative RDF Construction from In-Memory Data Structures with RML [18].
- Preserving the Alignment of LD with Source Data [20].
- Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems [21].
- Towards a Mapping Framework for the Tenders Electronic Daily Standard Forms [22].
- Meta2KG: An Embeddings-based Approach for Transforming Metadata to Knowledge Graphs [23].
- Scaling RML and SPARQL-based Knowledge Graph Construction with Apache Spark [24].

For the first time, a Knowledge Graph Construction Challenge was organized during the workshop to evaluate the performance of different knowledge graph construction approaches in terms of execution time and resources e.g. CPU, RAM, etc. The goal of this challenge was

---

\(^5\)http://w3id.org/kg-construct/workshop/2023
\(^6\)https://openreview.net/group?id=eswc-conferences.org/ESWC/2023/Workshop/KGCW
to identify existing gaps of different approaches and not only put the focus on execution time, but also resources. The challenge consisted of 2 parts: (i) artificial data for analyzing specific parameters of the construction process e.g. joins, data size, mappings, and (ii) real-life data of the GTFS Madrid Benchmark to evaluate approaches in real use cases. We received 4 participants: CARML, SDM-RDFizer, RDFProcessingToolkit / Sansa, and RMLStreamer of which 3 submitted a final report included in the proceedings:

- Knowledge Graph Creation Challenge: Results for SDM-RDFizer [25]
- RMLStreamer with Reference Conditions in the KGCW Challenge 2023 [27]

Organizing Committee

- David Chaves-Fraga, Universidad Politécnica de Madrid & KU Leuven
- Anastasia Dimou, KU Leuven, Flanders Make, Leuven.AI
- Dylan Van Assche, Ghent University – imec – IDLab
- Ana Iglesias-Molina, Universidad Politécnica de Madrid
- Umutcan Serles, University of Innsbruck

Program Committee

- Anelia Kurteva, Delft University of Technology
- Aidan Hogan, DCC, Universidad de Chile
- Antoine Zimmermann, École des Mines de Saint-Étienne
- Beatriz Esteves, Universidad Politécnica de Madrid
- Ben De Meester, Ghent University – imec – IDLab
- Bram Steenwinckel, Ghent University – imec – IDLab
- Claus Stadler, University of Leipzig
- Davide Lanti, Free University of Bozen
- Dylan Van Assche, Ghent University – imec – IDLab
- Enrique Iglesias, Leibniz University of Hannover
- Els de Vleeschauwer, Ghent University – imec – IDLab
- Ernesto Jimenez-Ruiz, City, University of London
- Edna Ruckhaus Magnus, Universidad Politécnica de Madrid
- Giorgos Flouris, FORTH-ICS
- Franck Michel, CNRS
- Francois Scharffe, Université de Montpellier
- Hannes Voigt, TU Dresden
- Herminio García-González, Kazerne Dossin
- Josiane Xavier Parreira, Siemens Corporate Research
- Jürgen Umbrich, Vienna University of Economics and Business
References


Graph Construction with Apache Spark, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.

