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

David Chaves-Fraga[1,2,3], Anastasia Dimou[1,3,4], Ana Iglesias-Molina[2], Umutcan Serles[5] and Dylan Van Assche[6]

[1]KU Leuven, Department of Computer Science, Sint-Katelijne-Waver, Belgium
[2]Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Spain
[3]Flanders Make – DTAI-FET
[4]Leuven.AI – KU Leuven institute for AI, B-3000 Leuven, Belgium
[5]Semantic Technology Institute Innsbruck, Universität Innsbruck, Austria
[6]IDLab, 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 Zalando[1],or public use, e.g., DBpedia[2] or Wikidata[3]. 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 Group[4] 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
david.chaves@upm.es (D. Chaves-Fraga); anastasia.dimou@kuleuven.be (A. Dimou); ana.iglesiasm@upm.es (A. Iglesias-Molina); umutcan.serles@sti2.at (U. Serles); dylan.van.assche@ugent.be (D. V. Assche)



CEUR Workshop Proceedings (CEUR-WS.org)

[1]https://engineering.zalando.com/posts/2018/03/semantic-web-technologies.html
[2]https://www.dbpedia.org/resources/knowledge-graphs/
[3]https://www.wikidata.org/wiki/Wikidata:Main_Page
[4]http://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].
- Reference Conditions: Relating Mapping Rules Without Joining [19].
- 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]
- KGCW2023 Challenge Report: RDFProcessingToolkit / Sansa [26]
- 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

- Julián Arenas-Guerrero, Universidad Politécnica de Madrid
- Juliette Opdenplatz, Universität Innsbruck
- Jakub Klímek, Charles University
- Manolis Koubarakis, National and Kapodistrian University of Athens
- Lan Yang, Insight Centre for Data Analytics
- Lionel Tailhardat, Eurecom
- Maria-Esther Vidal, Leibniz University of Hannover
- Mario Scrocca, Cefriel
- Markus Schröder, German Research Center for AI
- Maxime Lefrançois, École des Mines de Saint-Étienne
- Oscar Corcho, Universidad Politécnica de Madrid
- Pano Maria, Skemu
- Samaneh Jozashoori, metaphacts GmbH
- Sergio José Rodríguez Méndez, Australian National University
- Sven Lieber, Royal Library Of Belgium
- Vladimir Alexiev, Ontotext

References

[1] X. L. Dong, X. He, A. Kan, X. Li, Y. Liang, J. Ma, Y. E. Xu, C. Zhang, T. Zhao, G. Blanco Saldana, S. Deshpande, A. Michetti Manduca, J. Ren, S. P. Singh, F. Xiao, H.-S. Chang, G. Karamanolakis, Y. Mao, Y. Wang, C. Faloutsos, A. McCallum, J. Han, Autoknow: Self-driving knowledge collection for products of thousands of types, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 2724–2734.

[2] A. Dimou, M. V. Sande, P. Colpaert, R. Verborgh, E. Mannens, R. V. de Walle, Rml: A generic language for integrated rdf mappings of heterogeneous data, in: Proceedings of the 7th Workshop on Linked Data on the Web (LDOW), 2014.

[3] D. Chaves-Fraga, F. Priyatna, I. Perez-Santana, O. Corcho, Virtual statistics knowledge graph generation from CSV files, in: Emerging Topics in Semantic Technologies: ISWC 2018 Satellite Events, Studies on the Semantic Web, IOS Press, 2018.

[4] F. Michel, L. Djimenou, C. Faron-Zucker, J. Montagnat, xR2RML: Relational and Non-Relational Databases toRDF Mapping Language, Technical Report, 2017.

[5] E. Iglesias, S. Jozashoori, D. Chaves-Fraga, D. Collarana, M.-E. Vidal, SDM-RDFizer: An RML Interpreter for the Efficient Creation of RDF Knowledge Graphs, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 3039–3046.

[6] U. Şimşek, E. Kärle, D. Fensel, RocketRML - A NodeJS implementation of a Use-Case Specific RML Mapper, in: Proceedings of the 1st Workshop on Knowledge Graph Building, 2019.

[7] S. Jozashoori, D. Chaves-Fraga, E. Iglesias, M.-E. Vidal, O. Corcho, Funmap: Efficient execution of functional mappings for knowledge graph creation, in: International Semantic Web Conference, Springer, 2020, pp. 276–293.

[8] J. F. Sequeda, D. P. Miranker, Ultrawrap: SPARQL execution on relational data, Web Semantics: Science, Services and Agents on the WWW (2013).
[9] F. Priyatna, O. Corcho, J. Sequeda, Formalisation and experiences of r2rml-based sparql to sql query translation using morph, in: Proceedings of the 23rd International Conference on World Wide Web, 2014.
[10] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, G. Xiao, Ontop: Answering SPARQL Queries over Relational Databases, Semantic Web Journal (2017).
[11] M. Lefrançois, A. Zimmermann, N. Bakerally, A SPARQL extension for generating RDF from heterogeneous formats, in: The Semantic Web: 14th International Conference, 2017.
[12] H. García-González, I. Boneva, S. Staworko, J. E. Labra-Gayo, J. M. C. Lovelle, Shexml: improving the usability of heterogeneous data mapping languages for first-time users, PeerJ Computer Science 6 (2020) e318.
[13] B. De Meester, A. Dimou, R. Verborgh, E. Mannens, An ontology to semantically declare and describe functions, in: European Semantic Web Conference, 2016, pp. 46–49.
[14] A. C. Junior, C. Debruyne, R. Brennan, D. O'Sullivan, Funul: a method to incorporate functions into uplift mapping languages, in: Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, 2016, pp. 267–275.
[15] M. Grassi, M. Scrocca, A. Carenini, M. Comerio, I. Celino, Composable Semantic Data Transformation Pipelines with Chimera, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[16] J. Mynarz, K. Haniková, V. Svátek, Test-driven knowledge graph construction, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[17] C. Karalka, G. Meditskos, N. Bassiliades, Towards Semantic Interpretation of Structured Data Sources in Privacy-Preserving Environments, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[18] I. Dasoulas, D. Chaves-Fraga, D. Garijo, A. Dimou, Declarative RDF construction from in-memory data structures with RML, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[19] E. de Vleeschauwer, S. M. Oo, B. De Meester, P. Colpaert, Reference conditions: relating mapping rules without joining, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[20] A. Randles, D. O'Sullivan, Preserving the Alignment of LD with Source Data, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[21] L. Tailhardat, Y. Chabot, R. Troncy, Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[22] E. Costetchi, A. Vassiliades, C. Nyulas, Towards a Mapping Framework for the Tenders Electronic Daily Standard Forms, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[23] N. Abdelmageed, B. König-Ries, Meta2KG: An Embeddings-based Approach for Transforming Metadata to Knowledge Graphs, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[24] C. Stadler, L. Bühmann, L.-P. Meyer, M. Martin, Scaling RML and SPARQL-based Knowledge

Graph Construction with Apache Spark, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[25] E. Iglesias, V. Maria-Esther, Knowledge Graph Creation Challenge: Results for SDM-RDFizer, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[26] S. Bin, C. Stadler, L. Bühmann, KGCW2023 Challenge Report: RDFProcessingToolkit / Sansa, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.
[27] E. de Vleeschauwer, G. Haesendonck, D. Van Assche, B. De Meester, RMLStreamer with Reference Conditions in the KGCW Challenge 2023, in: Proceedings of the 4th International Workshop on Knowledge Graph Construction, 2023.