Preface for the 3rd Edition of the International Knowledge Graph Construction Workshop

David Chaves-Fraga[1,2,3], Anastasia Dimou[1,3,4], Pieter Heyvaert[5], Freddy Priyatna[6] and Juan Sequeda[7]

[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]IDLab, Dept of Electronics and Information Systems, Ghent University – imec, Belgium
[6]Olive AI, EEUU
[7]data.world, EEUU

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

Third International Workshop On Knowledge Graph Construction Co-located with the ESWC 2022, 30th May 2022, Crete, Greece
david.chaves@upm.es (D. Chaves-Fraga); anastasia.dimou@kuleuven.be (A. Dimou); pieter.heyvaert@ugent.be (P. Heyvaert); freddy.priyatna@oliveai.com (F. Priyatna); juan@data.world (J. Sequeda)



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

others the query-answering paradigm, e.g., Ultrawrap [8], Morph-RDB [9] and Ontop [10]. 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 third edition of the knowledge graph construction workshop[5] has a special focus on the automatization of knowledge graph construction methods, analyzing their alignment with previous standard but declarative approaches using mapping rules. It also included:

- *Keynote.* The workshop includes the keynote from Javier D. Fernandez and Selena Baset (Roche): "From ETL to DIY, or how to democratize the creation of Knowledge Graphs"
- *SemTab challenge 2022.*[6]: Kick-off for the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching for 2022.

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.

Eleven papers were submitted. The reviews were open and public, and hosted at Open Review[7]. 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 one was conditionally accepted. two of the accepted papers were long papers and five were short papers. The following papers were accepted for publication and presented at the workshop:

- Transformation of Node to Knowledge Graph Embeddings for Faster Link Prediction in Social Networks [15]
- A Human-in-the-Loop Approach for Personal Knowledge Graph Construction from File Names [16]
- Continuous generation of versioned collection's members with RML and LDES [17]
- Implementation-independent Knowledge Graph Construction Workflows using FnO Composition [18]
- Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges [19]
- Devising Mapping Interoperability with Mapping Translation [20]
- Supporting Relational Database Joins for Generating Literals in R2RML [21]

Organizing Committee

- David Chaves-Fraga, Universidad Politécnica de Madrid & KU Leuven
- Anastasia Dimou, KU Leuven, Flanders Make, Leuven.AI
- Pieter Heyvaert, Ghent University - imec
- Freddy Priyatna, Olive AI
- Juan Sequeda, data.world

[5] http://w3id.org/kg-construct/workshop/2022
[6] https://www.cs.ox.ac.uk/isg/challenges/sem-tab/
[7] https://openreview.net/group?id=kg-construct.github.io/KGCW/2022/Workshop

Program Committee

- Aidan Hogan, Universidad de Chile
- Ana Iglesias-Molina, Universidad Politécnica de Madrid
- Antoine Zimmermann, École des Mines de Saint-Étienne
- Ben De Meester, Ghent University – imec
- Boris Villazón-Terrazas, Tinámica
- Christophe Debruyne, Liège University
- Dylan Van Assche, Ghent University – imec
- Edna Ruckhaus, Universidad Politécnica de Madrid
- Elvira Amador, Universidad Politécnica de Madrid
- Ernesto Jiménez-Ruiz, University of London
- Femke Ongenae, Ghent University – imec
- Franck Michel, Université Côte d'Azur
- François Scharffe, Columbia University
- Giorgos Flouris, FORTH
- Giuseppe Futia, Nexa Center
- Hannes Voigt, Neo4j
- Heiko Paulheim, University of Mannheim
- Herminio Garcia Gonzalez, Universidad de Oviedo
- Jakub Klímek, Charles University
- Julián Arenas-Guerrero, Universidad Politécnica de Madrid
- Jhon Toledo, Universidad Politécnica de Madrid
- Manolis Koubarakis, National & Kapodistrian University of Athens
- Maria-Esther Vidal, L3S & TIB
- Mario Scrocca, CEFRIEL
- Mauro Dragoni, FBKZ
- Maxime Lefrancois, École des Mines de Saint-Étienne
- Miel Vander Sande, Memoo
- Mohamed Nadjib Mami, Deutsche Post DHL Group
- Nora Abdelmageed, Friedrich-Schiller-University Jena
- Oscar Corcho, Universidad Politécnica de Madrid
- Pano Maria, Skemu
- Samaneh Jozashoori, L3S & TIB
- Semih Salihoglu, University of Waterloo
- Sergio José Rodríguez Méndez, Australian National University
- Souripriya Das, Oracle
- Sven Lieber, Royal Library of Belgium (KBR)
- Umutcan Şimşek, University of Innsbruck
- Vladimir Alexiev, Ontotext

Acknowledgments

This publication is based upon work from COST Action Distributed Knowledge Graphs, supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. http://www.cost.eu/

Funded by
the European Union

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] A. Parnami, M. Lee, Transformation of Node to Knowledge Graph Embeddings for Faster Link Prediction in Social Networks, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[16] M. Schröder, C. Jilek, A. Dengel, A Human-in-the-Loop Approach for Personal Knowledge Graph Construction from File Names, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[17] D. Van Assche, S. M. Oo, J. Rojas, P. Colpaert, Continuous generation of versioned collection's members with RML and LDES, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[18] G. De Mulder, B. De Meester, Implementation-independent Knowledge Graph Construction Workflows using FnO Composition, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[19] A. Dimou, D. Chaves-Fraga, Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[20] A. Iglesias-Molina, C. Andrea, C. Oscar, Devising Mapping Interoperability with Mapping Translation, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.
[21] C. Debruyne, Supporting Relational Database Joins for Generating Literals in R2RML, in: Proceedings of the 3rd International Workshop on Knowledge Graph Construction, 2022.