Preface for the 2nd Edition of the International Knowledge Graph Construction Workshop

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

[1] Ontology Engineering Group, Universidad Politécnica de Madrid
dchaves@fi.upm.es
[2] IDLab, Dept of Electronics and Information Systems, Ghent University – imec
{anastasia.dimou,pheyvaer.heyvaert}@ugent.be
[3] Olive AI
freddy.priyatna@oliveai.com
[4] data.world
juan@data.world

More and more knowledge graphs are constructed for private use, e.g., the Amazon Product Graph [1] or the Fashion Knowledge Graph by Zalando[5],or public use, e.g., DBpedia[6] or Wikidata[7]. 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[8] 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], FunMap [7] and CARML[9], while others the query-answering paradigm, e.g., Ultrawrap [8], Morph-RDB [9] and Ontop [10].

[5] https://engineering.zalando.com/posts/2018/03/semantic-web-technologies.html
[6] https://www.dbpedia.org/resources/knowledge-graphs/
[7] https://www.wikidata.org/wiki/Wikidata:Main_Page
[8] http://w3.org/community/kg-construct
[9] https://github.com/carml/carml

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 second edition of the knowledge graph construction workshop[10] has a special focus on knowledge graph construction methods that involve or analyze the roles of users in these processes and it also included:

- *Mapping Challenges.* As the workshop complements and aligns with the activities of the W3C Community Group on knowledge graph construction, a special track for solving a set of well-identified mapping challenges[11] was announced and different solutions were proposed.
- *Keynote.* The workshop includes the keynote from Jesús Barrasa (Neo4J): “Knowledge graphs 2021: The great convergence”
- *Discussion Panel.* A panel on machine learning techniques for knowledge graph construction was organized with distinguish researchers invited as panelist: Ernesto Jiménez-Ruiz, Franceso Osborne, Maria-Esther Vidal and Heiko Paulheim.

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.

Sixteen papers were submitted, one of which was withdrawn. The reviews were open and public, and hosted at Open Review[12]. 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.

Twelve papers were accepted and one was conditionally accepted. Eight of the accepted papers were long papers and four were short papers. The following papers were accepted for publication and presented at the workshop:

- Everything for the Users, Nothing by the Users: Lessons Learnt from an Heterogeneous Data Mapping Languages User Study [15]
- A ShExML Perspective on Mapping Challenges: Already Solved Ones, Language Modifications and Future Required Actions [16]
- Mapping Spreadsheets to RDF: Supporting Excel in RML [17]
- Demo: Knowledge Graph-Based Housing Market Analysis [18]
- JenTab: A Toolkit for Semantic Table Annotations [19]
- Stratified Data Integration [20]
- Collaborative-AI Knowledge Graph Generation: Taxonomization of IATE, the EU Terminology [21]
- Embedding-Assisted Entity Resolution for Knowledge Graphs [22]
- Integrating Nested Data Into Knowledge Graphs with RML Fields [23]
- Open Drug Knowledge Graph [24]

[10] `http://w3id.org/kg-construct/workshop/2021`
[11] `http://w3id.org/kg-construct/workshop/2021/challenges`
[12] `https://openreview.net/group?id=eswc-conferences.org/ESWC/2021/Workshop/KGCW`

– Knowledge Graph Construction with R2RML and RML: An ETL System-Based Overview [25]
– Knowledge Graph Lifecycle: Building and maintaining Knowledge Graphs [26]
– Experiences of Using WDumper to Create Topical Subsets from Wikidata [27]

Organizing Committee

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

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
– Dylan Van Assche, Ghent University – imec
– Edna Ruckhaus, Universidad Politécnica de Madrid
– Femke Ongenae, Ghent University – imec
– Francesco Osborne, The Open University
– Franck Michel, Université Côte d'Azur
– François Scharffe, Columbia University
– George Fletcher, TU Eindhoven
– Giorgos Flouris, FORTH
– Giuseppe Futia, Nexa Center
– Guohui Xiao, Ontopic
– Hannes Voigt, Neo4j
– Heiko Paulheim, University of Mannheim
– Herminio Garcia Gonzalez, Universidad de Oviedo
– Jakub Klímek, Charles University
– Josh Shinavier, Uber
– Julián Arenas-Guerrero, 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
– Oscar Corcho, Universidad Politécnica de Madrid
– Pano Maria, Skemu

- Samaneh Jozashoori, L3S & TIB
- Semih Salihoglu, University of Waterloo
- Souripriya Das, Oracle
- Sven Lieber, Ghent University – imec
- Umutcan Şimşek, University of Innsbruck
- Vladimir Alexiev, Ontotext

References

[1] Xin Luna Dong et al. "AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types". In: KDD '20. Virtual Event, CA, USA: Association for Computing Machinery, 2020, pp. 2724–2734. ISBN: 9781450379984. DOI: 10.1145/3394486.3403323. URL: https://doi.org/10.1145/3394486.3403323.

[2] Anastasia Dimou et al. "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] David Chaves-Fraga et al. "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] Franck Michel et al. *xR2RML: Relational and Non-Relational Databases toRDF Mapping Language*. Tech. rep. 2017. URL: https://hal.archives-ouvertes.fr/hal-01066663/document/.

[5] Enrique Iglesias et al. "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] Umutcan Şimşek, Elias Kärle, and Dieter Fensel. "RocketRML - A NodeJS implementation of a Use-Case Specific RML Mapper". In: *Proceedings of the 1st Workshop on Knowledge Graph Building*. 2019.

[7] Samaneh Jozashoori et al. "FunMap: Efficient Execution of Functional Mappings for Knowledge Graph Creation". In: *International Semantic Web Conference*. Springer. 2020, pp. 276–293.

[8] Juan F. Sequeda and Daniel P. Miranker. "Ultrawrap: SPARQL execution on relational data". In: *Web Semantics: Science, Services and Agents on the WWW* (2013). ISSN: 1570-8268. DOI: https://doi.org/10.1016/j.websem.2013.08.002. URL: http://www.sciencedirect.com/science/article/pii/S1570826813000383.

[9] Freddy Priyatna, Oscar Corcho, and Juan 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] Diego Calvanese et al. "Ontop: Answering SPARQL Queries over Relational Databases". In: *Semantic Web Journal* (2017).

[11] Maxime Lefrançois, Antoine Zimmermann, and Noorani Bakerally. "A SPARQL Extension for Generating RDF from Heterogeneous Formats". In: *The Semantic Web: 14th International Conference*. 2017.

[12] Herminio García-González et al. "ShExML: improving the usability of heterogeneous data mapping languages for first-time users". In: *PeerJ Computer Science* 6 (2020), e318.

[13] Ben De Meester et al. "An ontology to semantically declare and describe functions". In: *European Semantic Web Conference*. Springer. 2016, pp. 46–49.

[14] Ademar Crotti Junior et al. "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] Herminio García-González. "Everything for the users, nothing by the users: Lessons learnt from an heterogeneous data mapping languages user study". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[16] Herminio García-González. "A ShExML perspective on mapping challenges: already solved ones, language modifications and future required actions". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[17] Markus Schröder, Christian Jilek, and Andreas Dengel. "Mapping Spreadsheets to RDF: Supporting Excel in RML". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[18] Ziping Hu et al. "Demo: Knowledge Graph-Based Housing Market Analysis". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[19] Nora Abdelmageed and Sirko Schindler. "JenTab: A Toolkit for Semantic Table Annotations". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[20] Fausto Giunchiglia et al. "Stratified Data Integration". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[21] Alena Vasilevich et al. "Collaborative-AI Knowledge Graph Generation: Taxonomization of IATE, the EU Terminology". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[22] Daniel Obraczka, Jonathan Schuchart, and Erhard Rahm. "Embedding-Assisted Entity Resolution for Knowledge Graphs". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[23] Thomas Delva et al. "Integrating Nested Data into Knowledge Graphs with RML Fields". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[24] Mark Mann et al. "Open Drug Knowledge Graph". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[25] Julián Arenas-Guerrero et al. "Knowledge Graph Construction with R2RML and RML: An ETL System-Based Overview". In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[26] Umutcan Şimşek et al. “Knowledge Graph Lifecycle: Building and maintaining Knowledge Graphs”. In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.

[27] Seyed Amir Hosseini Beghaeiraveri, Alasdair Gray, and Fiona McNeill. “Experiences of Using WDumper to Create Topical Subsets from Wikidata”. In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction*. 2021.