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

David Chaves-Fraga\textsuperscript{1}, Anastasia Dimou\textsuperscript{2}, Pieter Heyvaert\textsuperscript{2}, Freddy Priyatna\textsuperscript{3}, and Juan Sequeda\textsuperscript{4}

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

More and more knowledge graphs are constructed for private use, e.g., the Amazon Product Graph \cite{amazon} or the Fashion Knowledge Graph by Zalando\textsuperscript{5}, or public use, e.g., DBpedia\textsuperscript{6} or Wikidata\textsuperscript{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\textsuperscript{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 \cite{r2rml,r2rml_extension,carml}. Certain approaches followed the ETL-like paradigm, e.g., SDM-RDFizer \cite{sdm-rdfizer}, RocketRML \cite{rocketrml}, FunMap \cite{funmap} and CARML\textsuperscript{9}, while others the query-answering paradigm, e.g., Ultrawrap \cite{ultrawrap}, Morph-RDB \cite{morph-rdb} and Ontop \cite{ontop}.

\textsuperscript{5} https://engineering.zalando.com/posts/2018/03/semantic-web-technologies.html
\textsuperscript{6} https://www.dbpedia.org/resources/knowledge-graphs/
\textsuperscript{7} https://www.wikidata.org/wiki/Wikidata:Main_Page
\textsuperscript{8} http://w3.org/community/kg-construct
\textsuperscript{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., FuO [13] and FunUL [14].

The second edition of the knowledge graph construction workshop\(^\text{10}\) 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\(^\text{11}\) 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, Francesco 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\(^\text{12}\). 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 \(^\text{15}\)
– A ShExML Perspective on Mapping Challenges: Already Solved Ones, Language Modifications and Future Required Actions \(^\text{16}\)
– Mapping Spreadsheets to RDF: Supporting Excel in RML \(^\text{17}\)
– Demo: Knowledge Graph-Based Housing Market Analysis \(^\text{18}\)
– JenTab: A Toolkit for Semantic Table Annotations \(^\text{19}\)
– Stratified Data Integration \(^\text{20}\)
– Collaborative-AI Knowledge Graph Generation: Taxonomization of IATE, the EU Terminology \(^\text{21}\)
– Embedding-Assisted Entity Resolution for Knowledge Graphs \(^\text{22}\)
– Integrating Nested Data Into Knowledge Graphs with RML Fields \(^\text{23}\)
– Open Drug Knowledge Graph \(^\text{24}\)

\(^{10}\) \url{http://w3id.org/kg-construct/workshop/2021}
\(^{11}\) \url{http://w3id.org/kg-construct/workshop/2021/challenges}
\(^{12}\) \url{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 WDump 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 Asche, 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 Klimek, 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 Lefrançois, 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, Skemn
References


