

LREC-COLING 2024

**DLnLD: Deep Learning and Linked Data
@LREC-COLING-2024**

Workshop Proceedings

Editors

Gilles Sérasset, Hugo Gonçalo Oliveira and Giedre
Valunaite Oleskeviciene

21 May, 2024
Torino, Italia

**Proceedings of the Workshop on DLnLD: Deep Learning and Linked Data
@LREC-COLING-2024**

Copyright ELRA Language Resources Association (ELRA), 2024
These proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-16-6
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Preface by the Program Chairs

Since the appearance of transformers (Vaswani et al., 2017), Deep Learning (DL) and neural approaches have brought a huge contribution to Natural Language Processing (NLP) either with highly specialized models for specific application or via Large Language Models (LLMs) (Devlin et al., 2019; Brown et al., 2020; Touvron et al., 2023) that are efficient few-shot learners for many NLP tasks. Such models usually build on huge web-scale data (raw multilingual corpora and annotated specialized, task related, corpora) that are now widely available on the Web. This approach has clearly shown many successes, but still suffers from several weaknesses, such as the cost/impact of training on raw data, biases, hallucinations, lack of explainability, among others (Nah et al., 2023).

The Linguistic Linked Open Data (LLOD) (Chiarcos et al., 2013) community aims at creating/distributing explicitly structured data (modelled as RDF graphs) and interlinking such data across languages. This collection of datasets, gathered inside the LLOD Cloud (Chiarcos et al., 2020), contains a huge amount of multilingual ontological (e.g. DBpedia (Lehmann et al., 2015)); lexical (e.g., DBnary (Sérasset, 2015), Wordnet (McCrae et al., 2014), Wikidata (Vrandečić and Krötzsch, 2014)); or linguistic (e.g., Universal Dependencies Treebank (Nivre et al., 2020; Chiarcos et al., 2021), DBpedia Abstract Corpus (Brümmer et al., 2016)) information, structured using common metadata (e.g., OntoLex (McCrae et al., 2017), NIF (Hellmann et al., 2013), etc.) and standardised data categories (e.g., lexinfo (Cimiano et al., 2011), OliA (Chiarcos and Sukhareva, 2015)).

Both communities bring striking contributions that seem to be highly complementary. However, if knowledge (ontological) graphs are now routinely used in DL, there is still very few research studying the value of Linguistic/Lexical knowledge in the context of DL. We think that, today, there is a real opportunity to bring both communities together to take the best of both worlds. Indeed, with more and more work on Graph Neural Networks (Wu et al., 2023) and Embeddings on RDF graphs (Ristoski et al., 2019), there is more and more opportunity to apply DL techniques to build, interlink or enhance Linguistic Linked Open Datasets, to borrow data from the LLOD Cloud for enhancing Neural Models on NLP tasks, or to take the best of both worlds for specific NLP use cases.

This led us to propose this workshop aims at gathering researchers that work on the interaction between DL and LLOD in order to discuss what each approach has to bring to the other. All application domains (Digital Humanities, FinTech, Education, Linguistics, Cybersecurity...) as well as approaches (NLG, NLU, Data Extraction...) were welcome, provided that the work is based on the use of BOTH Deep Learning techniques and Linguistic Linked (meta)Data.

The DLnLD workshop builds on four editions of previous workshops, namely:

- Workshop on Deep Learning and Neural Approaches for Linguistic Data, collocated with the 3rd Nexus Linguarum Plenary Meeting, in Skopje, North Macedonia and online, September 2021;
- Workshop on Linguistic Knowledge Processing with Deep Learning, hosted at the Nexus Workshop days in Jerusalem, Israel, May 2022;
- 2nd Workshop on Deep Learning and Neural Approaches for Linguistic Linked Data, collocated with the LLOD Approaches for Language Data Research and Management Conference (LLODREAM2022), in Vilnius, Lithuania, and online, September 2022;

- Workshop on Deep Learning, Relation Extraction and Linguistic Data, collocated with the Language, Data and Knowledge Conference (LDK), in Vienna, Austria, September 2023.

However, in DLnLD, the objectives were expanded to not only study how Deep Learning may be used **for** Linguistic Linked Data but also explore how Linguistic Linked Data may be leveraged by Deep Learning approaches.

The papers that are presented in this volume show that the two domains do indeed cross fertilize each other with researchers using Language Models for Linguistic Linked Data modelling or generation and others leveraging Linked Data for evaluation or post-hoc verification of LLM outputs, while others do study Graph Neural Networks as a mean to merge both worlds in specific use cases.

We do think that this is only a beginning and that research will continue towards a better entanglement of both worlds and hope this workshop only witnesses the beginning of a research trend.

This workshop is organised in the scope of COST Action CA18209 NexusLinguarum¹, supported by COST (European Cooperation in Science and Technology).

Gilles Sérasset,
Hugo Gonalo Oliveira,
Giedre Valunaite Oleskeviciene

¹<https://nexuslinguarum.eu/>

Organizing Committee

Workshop Chairs

- Gilles Sérasset, *Université Grenoble Alpes, France*
- Hugo Gonçalo Oliveira, *University of Coimbra, Portugal*
- Giedre Valunaite Oleskeviciene, *Mykolas Romeris University, Lithuania*

Program Committee

- Mehwish Alam, *Télécom Paris, Institut Polytechnique de Paris, France*
- Milana Bolatbek, *Al-Farabi Kazakh National University, Kazakhstan*
- Milan Dojchinovski, *Czech Technical University in Prague, Czech Republic*
- Basil Ell, *University of Oslo, Norway*
- Radovan Garabík, *L. Štúr Institute of Linguistics, Slovak Academy of Sciences, Slovakia*
- Katerina Gkirtzou, *Athena Research Center, Maroussi, Greece*
- Jorge Gracia del Río, *University of Zaragoza, Spain*
- Dagmar Gromann, *University of Vienna, Austria*
- Dangis Gudelis, *Mykolas Romeris University, Lithuania*
- Ilan Kernerman, *Lexicala by K Dictionaries, Israel*
- Chaya Liebeskind, *Jerusalem College of Technology, Israel*
- Marco C. Passarotti, *Università Cattolica del Sacro Cuore, Milan, Italy*
- Heiko Paulheim, *University of Mannheim, Germany*
- Alexandre Rademaker, *IBM Research Brazil and EMap/FGV, Brazil*
- Georg Rehm, *DFKI GmbH, Berlin, Germany*
- Didier Schwab, *Université Grenoble Alpes, France*
- Andon Tchechmedjiev, *IMT Mines Alès, France*
- Dimitar Trajanov, *Ss. Cyril and Methodius University – Skopje, Macedonia*
- Nicolas Turenne, *Guangdong University of Foreign Studies, China*

Table of Contents

<i>Investigating the Impact of Different Graph Representations for Relation Extraction with Graph Neural Networks</i>	
Moritz Blum, Gennaro Nolano, Basil Ell and Philipp Cimiano	1
<i>TaxoCritic: Exploring Credit Assignment in Taxonomy Induction with Multi-Critic Reinforcement Learning</i>	
Injy Sarhan, Bendegúz Toth, Pablo Mosteiro and Shihan Wang	14
<i>Combining Deep Learning Models and Lexical Linked Data: Some Insights from the Development of a Multilingual News Named Entity Recognition and Linking Dataset</i>	
Emmanuel Cartier and Emile Peetermans	31
<i>Deductive Verification of LLM Generated SPARQL Queries</i>	
Alexandre Rademaker, Guilherme Lima, Sandro Rama Fiorini and Viviane Torres da Silva	45
<i>How to Turn Card Catalogs into LLM Fodder</i>	
Mary Ann Tan, Shufan Jiang and Harald Sack	53
<i>Evaluating Large Language Models for Linguistic Linked Data Generation</i>	
Maria Pia di Buono, Blerina Spahiu and Verginica Barbu Mititelu	66
<i>Towards Automated Evaluation of Knowledge Encoded in Large Language Models</i>	
Bruno Carlos Luís Ferreira, Hugo Gonçalo Oliveira and Catarina Silva	76
<i>Self-Evaluation of Generative AI Prompts for Linguistic Linked Open Data Modelling in Diachronic Analysis</i>	
Florentina Armaselu, Chaya Liebeskind and Giedre Valunaite Oleskeviciene	86

Tutorial Program

Tuesday May 21, 2024

09:00–10:30 Session 1

Chair: Hugo Gonalo Oliveira

9:00–9:20 *Investigating the Impact of Different Graph Representations for Relation Extraction with Graph Neural Networks*

Moritz Blum, Gennaro Nolano, Basil Ell and Philipp Cimiano

9:20–9:40 *TaxoCritic: Exploring Credit Assignment in Taxonomy Induction with Multi-Critic Reinforcement Learning*

Injy Sarhan, Bendeg z Toth, Pablo Mosteiro and Shihan Wang

9:40–10:00 *Combining Deep Learning Models and Lexical Linked Data: Some Insights from the Development of a Multilingual News Named Entity Recognition and Linking Dataset*

Emmanuel Cartier and Emile Peetermans

10:00–10:20 *Deductive Verification of LLM Generated SPARQL Queries*

Alexandre Rademaker, Guilherme Lima, Sandro Rama Fiorini and Viviane Torres da Silva

11:00–13:00 Session 2

Chair: Gilles S rasset

11:00–11:20 *How to Turn Card Catalogs into LLM Fodder*

Mary Ann Tan, Shufan Jiang and Harald Sack

11:20–11:40 *Evaluating Large Language Models for Linguistic Linked Data Generation*

Maria Pia di Buono, Blerina Spahiu and Verginica Barbu Mititelu

11:40–12:00 *Towards Automated Evaluation of Knowledge Encoded in Large Language Models*

Bruno Carlos Lu s Ferreira, Catarina Silva and Hugo Gonalo Oliveira

12:00–12:20 *Self-Evaluation of Generative AI Prompts for Linguistic Linked Open Data Modelling in Diachronic Analysis*

Florentina Armaselu, Chaya Liebeskind and Giedre Valunaite Oleskeviciene

12:20–13:00 Discussions and further work on the topic

