

2nd Conference on Language, Data and Knowledge

LDK 2019, May 20–23, 2019, Leipzig, Germany

Edited by

Maria Eskevich

Gerard de Melo

Christian Fäth

John P. McCrae

Paul Buitelaar

Christian Chiarcos

Bettina Klimek

Milan Dojchinovski



Editors

Maria Eskevich 

CLARIN ERIC, Utrecht, The Netherlands
maria@clarin.eu

Gerard de Melo 

Department of Computer Science, Rutgers University–New Brunswick, NJ, USA

Christian Fäth 

Applied Computational Linguistics, Goethe University Frankfurt, Germany

John P. McCrae 

Insight Centre for Data Analytics, National University of Ireland Galway, Ireland

Paul Buitelaar 

Insight Centre for Data Analytics, National University of Ireland Galway, Ireland

Christian Chiarcos 

Applied Computational Linguistics, Goethe University Frankfurt, Germany

Bettina Klimek

Agile Knowledge Engineering and Semantic Web, University of Leipzig, Germany

Milan Dojchinovski 

Agile Knowledge Engineering and Semantic Web, University of Leipzig, Germany

ACM Classification 2012

Computing methodologies → Natural language processing; Computing methodologies → Knowledge representation and reasoning

ISBN 978-3-95977-105-4

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at <https://www.dagstuhl.de/dagpub/978-3-95977-105-4>.

Publication date

May, 2019

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <https://portal.dnb.de>.

License

This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0):
<https://creativecommons.org/licenses/by/3.0/legalcode>.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/OASlcs.LDK.2019.0

ISBN 978-3-95977-105-4

ISSN 1868-8969

<https://www.dagstuhl.de/oasics>

OASlcs – OpenAccess Series in Informatics

OASlcs aims at a suitable publication venue to publish peer-reviewed collections of papers emerging from a scientific event. OASlcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Daniel Cremers (TU München, Germany)
- Barbara Hammer (Universität Bielefeld, Germany)
- Marc Langheinrich (Università della Svizzera Italiana – Lugano, Switzerland)
- Dorothea Wagner (*Editor-in-Chief*, Karlsruher Institut für Technologie, Germany)

ISSN 1868-8969

<https://www.dagstuhl.de/oasics>

■ Contents

Preface

Maria Eskevich, Gerard de Melo, Christian Fäth and Christian Chiarcos 0:ix

Foundations: Web Technologies and Vocabularies

SPARQL Query Recommendation by Example: Assessing the Impact of Structural Analysis on Star-Shaped Queries

Alessandro Adamou, Carlo Allocca, Mathieu d'Aquin, and Enrico Motta 1:1–1:8

OWL^C: A Contextual Two-Dimensional Web Ontology Language

Sahar Aljalbout, Didier Buchs, and Gilles Falquet 2:1–2:13

Ligt: An LLOD-Native Vocabulary for Representing Interlinear Glossed Text as RDF

Christian Chiarcos and Maxim Ionov 3:1–3:15

The Shortcomings of Language Tags for Linked Data When Modeling Lesser-Known Languages

Frances Gillis-Webber and Sabine Tittel 4:1–4:15

Functional Representation of Technical Artefacts in Ontology-Terminology Models

Laura Giacomini 5:1–5:6

Language and Data: Human Language Technology

Comparison of Different Orthographies for Machine Translation of Under-Resourced Dravidian Languages

Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae 6:1–6:14

CoNLL-Merge: Efficient Harmonization of Concurrent Tokenization and Textual Variation

Christian Chiarcos and Niko Schenk 7:1–7:14

Exploiting Background Knowledge for Argumentative Relation Classification

Jonathan Kobbe, Juri Opitz, Maria Becker, Ioana Hulpuş, Heiner Stuckenschmidt, and Anette Frank 8:1–8:14

Graph-Based Annotation Engineering: Towards a Gold Corpus for Role and Reference Grammar

Christian Chiarcos and Christian Fäth 9:1–9:11

Crowd-Sourcing A High-Quality Dataset for Metaphor Identification in Tweets

Omnia Zayed, John P. McCrae, and Paul Buitelaar 10:1–10:17

2nd Conference on Language, Data and Knowledge (LDK 2019).

Editors: Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, and Milan Dojchinovski



OpenAccess Series in Informatics

OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Data and Knowledge: Entities and Relations

Inflection-Tolerant Ontology-Based Named Entity Recognition for Real-Time Applications <i>Christian Jilek, Markus Schröder, Rudolf Novik, Sven Schwarz, Heiko Maus, and Andreas Dengel</i>	11:1–11:14
Validation Methodology for Expert-Annotated Datasets: Event Annotation Case Study <i>Oana Inel and Lora Aroyo</i>	12:1–12:15
A Proposal for a Two-Way Journey on Validating Locations in Unstructured and Structured Data <i>Ilkcan Keles, Omar Qawasmeh, Tabea Tietz, Ludovica Marinucci, Roberto Reda, and Marieke van Erp</i>	13:1–13:8
Name Variants for Improving Entity Discovery and Linking <i>Albert Weichselbraun, Philipp Kuntschik, and Adrian M. P. Braşoveanu</i>	14:1–14:15
Interlinking SciGraph and DBpedia Datasets Using Link Discovery and Named Entity Recognition Techniques <i>Beyza Yaman, Michele Pasin, and Markus Freudenberg</i>	15:1–15:8

Language and Knowledge: Lexical Data

lemon-tree: Representing Topical Thesauri on the Semantic Web <i>Sander Stolk</i>	16:1–16:13
Translation-Based Dictionary Alignment for Under-Resourced Bantu Languages <i>Thomas Eckart, Sonja Bosch, Dirk Goldhahn, Uwe Quasthoff, and Bettina Klimek</i>	17:1–17:11
Cherokee Syllabary Texts: Digital Documentation and Linguistic Description <i>Jeffrey Bourns</i>	18:1–18:6
Metalexigraphy as Knowledge Graph <i>David Lindemann, Christiane Klaes, and Philipp Zumstein</i>	19:1–19:8
Cross-Dictionary Linking at Sense Level with a Double-Layer Classifier <i>Roser Saurí, Louis Mahon, Irene Russo, and Mironas Bitinis</i>	20:1–20:16
Towards the Detection and Formal Representation of Semantic Shifts in Inflectional Morphology <i>Dagmar Gromann and Thierry Declerck</i>	21:1–21:15

Applications in the Language Sciences

Opening Digitized Newspapers Corpora: Europeana’s Full-Text Data Interoperability Case <i>Nuno Freire, Antoine Isaac, Twan Goosen, Daan Broeder, Hugo Manguinhas, and Valentine Charles</i>	22:1–22:14
--	------------

Automatic Detection of Language and Annotation Model Information in CoNLL Corpora
Frank Abromeit and Christian Chiarcos 23:1–23:9

The Secret to Popular Chinese Web Novels: A Corpus-Driven Study
Yi-Ju Lin and Shu-Kai Hsieh 24:1–24:8

Predicting Math Success in an Online Tutoring System Using Language Data and Click-Stream Variables: A Longitudinal Analysis
Scott Crossley, Shamyia Karumbaiah, Jaclyn Ocumpaugh, Matthew J. Labrum, and Ryan S. Baker 25:1–25:13

Can Computational Meta-Documentary Linguistics Provide for Accountability and Offer an Alternative to “Reproducibility” in Linguistics?
Tobias Weber 26:1–26:8

■ Preface

This volume presents the proceedings of the 2nd Conference on Language, Data and Knowledge (LDK 2019) held in Leipzig, Germany, May 20–23, 2019. Language, Data and Knowledge is a bi-annual conference series on matters of human language technology, data science, and knowledge representation, initiated in 2017 by a consortium of researchers from the Insight Centre for Data Analytics at the National University of Ireland, Galway (Ireland), the Institut für Angewandte Informatik (InfAI) at the University of Leipzig (Germany), and the Applied Computational Linguistics Lab (ACoLi) at Goethe University Frankfurt am Main (Germany), and it has been supported by an international Scientific Committee of leading researchers in Natural Language Processing, Linked Data and Semantic Web, Language Resources and Digital Humanities.

The second edition of the LDK conference is hosted by the Institut für Angewandte Informatik (InfAI) in Leipzig, Germany and co-organized by the Insight Centre for Data Analytics and the Applied Computational Linguistics Lab (ACoLi). Major Sponsors were the *LiLa: Linking Latin* project, the *CID GmbH* in Germany, the *Semantic Web Company*, and *Pret-a-LLoD. Ready-to-use Multilingual Linked Language Data for Knowledge Services across Sectors* funded under the European Union’s Horizon research and innovation programme under grant agreement No. 825182. LDK 2019 has received further endorsement from the *DBpedia Association*, from the *European Lexicographic Infrastructure (ELEXIS)* project funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 731015, and from the independent research group *Linked Open Dictionaries (LiODi)* funded by the German Federal Ministry of Education and Research (BMBF).

In a biennial cycle, LDK conferences aim at bringing together researchers from across disciplines concerned with the acquisition, curation and use of language data in the context of data science and knowledge-based applications. With the advent of the Web and digital technologies, an ever increasing amount of language data is now available across application areas and industry sectors, including social media, digital archives, company records, etc. The efficient and meaningful exploitation of this data in scientific and commercial innovation is at the core of data science research, employing natural language processing and machine learning methods as well as semantic technologies and knowledge graphs.

Language data is of increasing importance to machine learning-based approaches in Human Language Technologies, Linked Data and Semantic Web research and applications that depend on linguistic and semantic annotation with lexical, terminological and ontological resources, manual alignment across language or other human-assigned labels. The acquisition, provenance, representation, maintenance, usability, quality as well as legal, organizational and infrastructure aspects of language data are therefore rapidly becoming major areas of research that are at the focus of the conference.

Knowledge graphs is an active field of research concerned with the extraction, integration, maintenance and use of semantic representations of language data in combination with semantically or otherwise structured data, numerical data and multimodal data among others. Knowledge graph research builds on the exploitation and extension of lexical, terminological and ontological resources, information and knowledge extraction, entity linking, ontology learning, ontology alignment, semantic text similarity, Linked Data and other Semantic Web technologies. The construction and use of knowledge graphs from language data, possibly and ideally in the context of other types of data, is a further specific focus of the conference.

2nd Conference on Language, Data and Knowledge (LDK 2019).

Editors: Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, and Milan Dojchinovski



Open Access Series in Informatics

OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

As in previous years, LDK 2019 features a number of collocated satellite events dedicated to the conference topics. This includes the 13th DBpedia community meeting, the 2nd Shared Task on Translation Inference Across Dictionaries (TIAD), a workshop of the W3C Ontology-Lexica Community and Business Group and a tutorial on historical text reuse (TRACER).

In addition, this edition of LDK also features an associated summer school, the 3rd Summer Datathon on Linguistic Linked Open Data (SD-LLOD-19, held in Schloss Dagstuhl – Leibniz Center for Informatics, Wadern, Germany), which complements the scientific focus of the conference with a didactic component and a hands-on experience. The SD-LLOD datathon has the main goal of giving people from industry and academia practical knowledge in the field of Linked Data and its application to natural language data and natural language annotations, from areas as diverse as knowledge engineering, lexicography, the language sciences, natural language processing and computational philology.

In total, 43 papers were submitted and reviewed by 88 reviewers. Typically, at least 3 reviews per paper resulted in 26 accepted papers. As a novel feature, LDK-2019 had a special track for short abstracts on latest development to be presented as posters during the conference. However, these are not subject to the proceedings and will be published separately.

The conference programme additionally encompasses invited talks on *Mapping the Lexicons of Signs and Words* by Christiane Fellbaum (Princeton University), and on *Schema.org Annotations and Web Tables: Underexploited Semantic Nuggets on the Web?* by Christian Bizer (Mannheim University), as well as on *The Sorbian languages* by Eduard Werner (University of Leipzig).

■ Organizing Committee

Conference Chairs

John P. McCrae (National University of Ireland Galway)
Paul Buitelaar (National University of Ireland Galway)
Christian Chiarcos (Goethe University Frankfurt)

Local Organizers

Bettina Klimek (University of Leipzig)
Milan Dojchinovski (University of Leipzig)

Program Chairs

Gerard de Melo (Rutgers University)
Maria Eskevich (CLARIN ERIC)

Proceedings Chair

Christian Fäth (Goethe University Frankfurt)



■ Scientific Advisory Committee

Francis Bond (Nanyang Technological University)
Paul Buitelaar (National University of Ireland Galway)
Christian Chiarcos (Goethe University Frankfurt)
Philipp Cimiano (Bielefeld University)
Edward Curry (National University of Ireland Galway)
Thierry Declerck (Deutsches Forschungszentrum für Künstliche Intelligenz – DFKI)
Milan Dojchinovski (University of Leipzig)
Tatjana Gornostaja (Tilde)
Jorge Gracia (University of Zaragoza)
Nancy Ide (Vassar College)
Franciska de Jong (CLARIN ERIC)
John P. McCrae (National University of Ireland Galway)
Gerard de Melo (Rutgers University)
Karin Verspoor (University of Melbourne)

2nd Conference on Language, Data and Knowledge (LDK 2019).

Editors: Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, and Milan Dojchinovski



OpenAccess Series in Informatics

OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

■ Program Committee

Eneko Agirre (University of the Basque Country)

Malihe Alikhani (Rutgers University)

Sören Auer (TIB Leibniz Information Center Science & Technology and University of Hannover)

Denilson Barbosa (University of Alberta)

Pierpaolo Basile (Dipartimento di Informatica – University of Bari)

Valerio Basile (University of Turin)

Martin Benjamin (Kamusi Project International)

Joanna Biega (Max Planck Institute for Informatics)

Michael Bloodgood (The College of New Jersey)

Francis Bond (Nanyang Technological University)

Harry Bunt (Tilburg University)

Aljoscha Burchardt (Deutsches Forschungszentrum für Künstliche Intelligenz – DFKI)

Nicoletta Calzolari (Istituto di Linguistica Computazionale – CNR)

Richard Eckart de Castilho (Ubiquitous Knowledge Processing Lab – UKP)

Philipp Cimiano (Bielefeld University)

Francesco Corcoglioniti (Fondazione Bruno Kessler)

Patrick Ernst (Max-Planck Institute for Informatics)

Besnik Fetahu (L3S Research Center)

Darja Fišer (University of Ljubljana)

Francesca Frontini (Université Paul-Valéry Montpellier 3 Praxiling UMR 5267 CNRS – UPVM3)

Luis Galárraga (Aalborg University)

Debanjan Ghosh (Rutgers University)

Jeff Good (University at Buffalo)

Gregory Grefenstette (IHMC and Biggerpan Inc)

Dagmar Gromann (TU Dresden)

Graeme Hirst (University of Toronto)

Eero Hyvönen (Aalto University and University of Helsinki – HELDIG)

Nancy Ide (Vassar College)

Sepehr Janghorbani (Rutgers University)

Richard Johansson (University of Gothenburg)

Te Taka Keegan (University of Waikato)

Roman Klinger (University of Stuttgart)

Dimitris Kontokostas (University of Leipzig)

Maria Koutraki (Leibniz Institute for Information Infrastructure – FIZ, and Karlsruher Institut für Technologie – KIT)

Udo Kruschwitz (University of Essex)

Oi Yee Kwong (The Chinese University of Hong Kong)

Piroska Lendvai (University of Göttingen)

Margot Mieskes (University of Applied Sciences Darmstadt)

Pasquale Minervini (University of Bari)

Paramita Mirza (Max Planck Institute for Informatics)

Elena Montiel-Ponsoda (Universidad Politécnica de Madrid)

Steven Moran (University of Zurich)

Andrea Moro (Sapienza University of Rome)

Hugo Gonçalo Oliveira (University of Coimbra)

Alessandro Oltramari (Bosch Research and Technology Center)

2nd Conference on Language, Data and Knowledge (LDK 2019).

Editors: Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek, and Milan Dojchinovski



OpenAccess Series in Informatics

ASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

- Petya Osenova (Sofia University, and Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences – IICT-BAS)
- Sebastian Pado (University of Stuttgart)
- Natalie Parde (University of Illinois at Chicago)
- Bolette Pedersen (University of Copenhagen)
- Pascual Pérez-Paredes (University of Cambridge)
- Maciej Piasecki (Department of Computational Intelligence Wrocław University of Science and Technology)
- Laurette Pretorius (School of Interdisciplinary Research and Graduate Studies University of South Africa)
- Gábor Prószéký (Pázmány Péter Catholic University Budapest)
- Francesca Quattri (The Hong Kong Polytechnic University)
- Alexandre Rademaker (IBM Research Brazil and EMAP/FGV)
- Dheeraj Rajagopal (Carnegie Mellon University)
- Simon Razniewski (Max Planck Institute for Informatics)
- Georg Rehm (Deutsches Forschungszentrum für Künstliche Intelligenz – DFKI)
- Nils Reiter (Institute of Natural Language Processing Stuttgart University)
- Steffen Remus (University of Hamburg)
- Laurent Romary (Institut national de recherche en informatique et en automatique – INRIA, and HUB-ISDL)
- Marco Rospocher (Università degli Studi di Verona)
- Harald Sack (Leibniz Institute for Information Infrastructure – FIZ, and Karlsruher Institut für Technologie – KIT)
- Felix Sasaki (Lambdawerk)
- Andrea Schalley (Karlstad University)
- Gilles Serasset (LIG Université Grenoble I)
- Vered Shwartz (Bar-Ilan University)
- Max Silberztein (Université de Franche-Comté)
- Aitor Soroa (Universidad del País Vasco / Euskal Herriko Unibertsitatea – UPV/EHU)
- Steffen Staab (Institut WeST University Koblenz-Landau and WAIS University of Southampton)
- Armando Stellato (University of Rome Tor Vergata)
- Stan Szpakowicz (University of Ottawa)
- Niket Tandon (Max Planck Institute for Informatics)
- Sara Tonelli (Fondazione Bruno Kessler)
- Mihaela Vela (Universität des Saarlandes)
- Marc Verhagen (Brandeis University)
- Karin Verspoor (The University of Melbourne)
- Piek Vossen (Vrije Universiteit Amsterdam)
- Sabine Schulte Im Walde (University of Stuttgart)
- Ulli Waltinger (Siemens AG – Corporate Technology)
- Eveline Wandl-Vogt (Austrian Centre for Digital Humanities @ Austrian Academy of Sciences – ACDH)
- Linlin Wang
- Aaron Steven White (University of Rochester)
- Michael Witbrock (IBM)
- René Witte (Concordia University)
- Qian Yang (Tsinghua University)
- Kalliopi Zervanou (Eindhoven University of Technology)
- Ziqi Zhang (Sheffield University)