

# 3rd Conference on Language, Data and Knowledge

LDK 2021, September 1–3, 2021, Zaragoza, Spain

Edited by

Dagmar Gromann

Gilles Sérasset

Thierry Declerck

John P. McCrae

Jorge Gracia


Julia Bosque-Gil


Fernando Bobillo


Barbara Heinisch



*Editors*

**Dagmar Gromann**   
University of Vienna, Austria  
dagmar.gromann@gmail.com

**Thierry Declerck**   
DFKI GmbH, Germany  
declerck@dfki.de


**Jorge Gracia**   
University of Zaragoza, Spain  
jogracia@unizar.es

**Fernando Bobillo**   
University of Zaragoza, Spain

**Gilles Sérasset**   
Université Grenoble Alpes, France  
gilles.serasset@imag.fr

**John P. McCrae**   
National University of Ireland Galway, Ireland  
john.mccrae@insight-centre.org

**Julia Bosque-Gil**   
University of Zaragoza, Spain

**Barbara Heinisch**   
University of Vienna, Austria  
barbara.heinisch@univie.ac.at

*ACM Classification 2012*

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

**ISBN 978-3-95977-199-3**

PRINT ISBN: 978-1-7138-3676-6

*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/978-3-95977-199-3>.

*Publication date*

August, 2021

*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 4.0 International license (CC-BY 4.0):

<https://creativecommons.org/licenses/by/4.0/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.2021.0

ISBN 978-3-95977-199-3

ISSN 1868-8969

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

## ■ Contents

Preface	
<i>Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch</i> .....	0:ix
Organizing Committee	
.....	0:xi
Scientific Advisory Committee	
.....	0:xiii
Program Committee	
.....	0:xv

### Invited Talks

The JeuxDeMots Project	
<i>Mathieu Lafourcade</i> .....	1:1–1:1
A Smell is Worth a Thousand Words: Olfactory Information Extraction and Semantic Processing in a Multilingual Perspective	
<i>Sara Tonelli</i> .....	2:1–2:1
Free/Open-Source Machine Translation for the Low-Resource Languages of Spain	
<i>Mikel L. Forcada</i> .....	3:1–3:1

### Crazy New Ideas

A Computational Simulation of Children’s Language Acquisition	
<i>Ben Ambridge</i> .....	4:1–4:3
Get! Mimetypes! Right!	
<i>Christian Chiarcos</i> .....	5:1–5:4
Mind the Gap: Language Data, Their Producers, and the Scientific Process	
<i>Tobias Weber</i> .....	6:1–6:9

### Language Data

Representing the Under-Represented: a Dataset of Post-Colonial, and Migrant Writers	
<i>Marco Antonio Stranisci, Viviana Patti, and Rossana Damiano</i> .....	7:1–7:14
Plenary Debates of the Parliament of Finland as Linked Open Data and in Parla-CLARIN Markup	
<i>Laura Sinikallio, Senka Drobac, Minna Tamper, Rafael Leal, Mikko Koho, Jouni Tuominen, Matti La Mela, and Eero Hyöönen</i> .....	8:1–8:17
Towards a Corpus of Historical German Plays with Emotion Annotations	
<i>Thomas Schmidt, Katrin Dennerlein, and Christian Wolff</i> .....	9:1–9:11

3rd Conference on Language, Data and Knowledge (LDK 2021).

Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch



OpenAccess Series in Informatics

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

Enriching a Lexical Resource for French Verbs with Aspectual Information <i>Anna Kupść, Pauline Haas, Rafael Marín, and Antonio Balvet</i> .....	10:1–10:12
Annotation of Fine-Grained Geographical Entities in German Texts <i>Julián Moreno-Schneider, Melina Plakidis, and Georg Rehm</i> .....	11:1–11:8
Supporting the Annotation Experience Through CorEx and Word Mover’s Distance <i>Stefania Pecòre</i> .....	12:1–12:15
A Twitter Corpus and Lexicon for Abusive Speech Detection in Serbian <i>Danka Jokić, Ranka Stanković, Cvetana Krstev, and Branislava Šandrih</i> .....	13:1–13:17

## Knowledge Graphs

Bias in Knowledge Graphs – An Empirical Study with Movie Recommendation and Different Language Editions of DBpedia <i>Michael Matthias Voit and Heiko Paulheim</i> .....	14:1–14:13
Enriching Word Embeddings with Food Knowledge for Ingredient Retrieval <i>Álvaro Mendes Samagaio, Henrique Lopes Cardoso, and David Ribeiro</i> .....	15:1–15:15
TatWordNet: A Linguistic Linked Open Data-Integrated WordNet Resource for Tatar <i>Alexander Kirillovich, Marat Shaekhov, Alfiya Galieva, Olga Nevzorova, Dmitry Ilvovsky, and Natalia Loukachevitch</i> .....	16:1–16:12
Explainable Zero-Shot Topic Extraction Using a Common-Sense Knowledge Graph <i>Ismail Harrando and Raphaël Troncy</i> .....	17:1–17:15
Relevance Feedback Search Based on Automatic Annotation and Classification of Texts <i>Rafael Leal, Joonas Kesäniemi, Mikko Koho, and Eero Hyvönen</i> .....	18:1–18:15
Automatic Construction of Knowledge Graphs from Text and Structured Data: A Preliminary Literature Review <i>Maraim Masoud, Bianca Pereira, John McCrae, and Paul Buitelaar</i> .....	19:1–19:9
An Ontology for CoNLL-RDF: Formal Data Structures for TSV Formats in Language Technology <i>Christian Chiarcos, Maxim Ionov, Luis Glaser, and Christian Fäth</i> .....	20:1–20:14
On the Utility of Word Embeddings for Enriching OpenWordNet-PT <i>Hugo Gonçalo Oliveira, Fredson Silva de Souza Aguiar, and Alexandre Rademaker</i> .....	21:1–21:13

## Applications for Language, Data and Knowledge

Towards Learning Terminological Concept Systems from Multilingual Natural Language Text <i>Lennart Wachowiak, Christian Lang, Barbara Heinisch, and Dagmar Gromann</i> ..	22:1–22:18
Encoder-Attention-Based Automatic Term Recognition (EA-ATR) <i>Sampritha H. Manjunath and John P. McCrae</i> .....	23:1–23:13

Universal Dependencies for Multilingual Open Information Extraction <i>Massinissa Atmani and Mathieu Lafourcade</i> .....	24:1–24:15
Inconsistency Detection in Job Postings <i>Joana Urbano, Miguel Couto, Gil Rocha, and Henrique Lopes Cardoso</i> .....	25:1–25:16
A Workbench for Corpus Linguistic Discourse Analysis <i>Julia Krasselt, Matthias Fluor, Klaus Rothenhäusler, and Philipp Dreesen</i> .....	26:1–26:9
APICS-Ligt: Towards Semantic Enrichment of Interlinear Glossed Text <i>Maxim Ionov</i> .....	27:1–27:8
Introducing the NLU Showroom: A NLU Demonstrator for the German Language <i>Dennis Wegener, Sven Giesselbach, Niclas Doll, and Heike Horstmann</i> .....	28:1–28:9
AAA4LLL – Acquisition, Annotation, Augmentation for Lively Language Learning <i>Bartholomäus Wloka and Werner Winiwarter</i> .....	29:1–29:15
Improving Intent Detection Accuracy Through Token Level Labeling <i>Michał Lew, Aleksander Obuchowski, and Monika Kutyla</i> .....	30:1–30:11
Towards Scope Detection in Textual Requirements <i>Ole Magnus Holter and Basil Ell</i> .....	31:1–31:15
Discrepancies Between Database- and Pragmatically Driven NLG: Insights from QUD-Based Annotations <i>Christoph Hesse, Maurice Langner, Anton Benz, and Ralf Klabunde</i> .....	32:1–32:9
Bridging the Gap Between Ontology and Lexicon via Class-Specific Association Rules Mined from a Loosely-Parallel Text-Data Corpus <i>Basil Ell, Mohammad Fazleh Elahi, and Philipp Cimiano</i> .....	33:1–33:21

## Use Cases in Language, Data and Knowledge

HISTORIAE, History of Socio-Cultural Transformation as Linguistic Data Science. A Humanities Use Case <i>Florentina Armaselu, Elena-Simona Apostol, Anas Fahad Khan, Chaya Liebeskind, Barbara McGillivray, Ciprian-Octavian Truică, and Giedrė Valūnaitė Oleškevičienė</i> .....	34:1–34:13
An Automatic Partitioning of Gutenberg.org Texts <i>Davide Picca and Cyrille Gay-Crosier</i> .....	35:1–35:9
A Data Augmentation Approach for Sign-Language-To-Text Translation In-The-Wild <i>Fabrizio Nunnari, Cristina España-Bonet, and Eleftherios Avramidis</i> .....	36:1–36:8
A Review and Cluster Analysis of German Polarity Resources for Sentiment Analysis <i>Bettina M. J. Kern, Andreas Baumann, Thomas E. Kolb, Katharina Sekanina, Klaus Hofmann, Tanja Wissik, and Julia Neidhardt</i> .....	37:1–37:17
Exploring Causal Relationships Among Emotional and Topical Trajectories in Political Text Data <i>Andreas Baumann, Klaus Hofmann, Bettina Kern, Anna Marakasova, Julia Neidhardt, and Tanja Wissik</i> .....	38:1–38:8

**0:viii**      **Contents**

Calculating Argument Diversity in Online Threads <i>Cedric Waterschoot, Antal van den Bosch, and Ernst van den Hemel</i> .....	39:1–39:9
Linking Discourse Marker Inventories <i>Christian Chiarcos and Maxim Ionov</i> .....	40:1–40:15
Tackling Domain-Specific Winograd Schemas with Knowledge-Based Reasoning and Machine Learning <i>Suk Joon Hong and Brandon Bennett</i> .....	41:1–41:13