2nd Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL 2023)

Torshavn, Faroe Islands
22 May 2023

Editors:

Nikolai Ilinykh
Felix Morger
Dana Dannélls
Simon Dobnik
Beáta Megyesi
Joakim Nivre

ISBN: 978-1-7138-9062-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Ableist Language Teching over Sign Language Research
Carl Börstell 1

The DA-ELEXIS Corpus - a Sense-Annotated Corpus for Danish with Parallel Annotations for Nine European Languages
Bolette S. Pedersen, Sanni Nimb, Sussi Olsen, Thomas Troelsgård, Ida Flörke, Jonas Jensen and Henrik Lorentzen 11

Sentiment Analysis Using Aligned Word Embeddings for Uralic Languages
Khalid Alnajjar, Mika Hämäläinen and Jack Rueter 19

What Causes Unemployment? Unsupervised Causality Mining from Swedish Governmental Reports
Luise Dürlich, Joakim Nivre and Sara Stymne 25

Are There Any Limits to English-Swedish Language Transfer? A Fine-grained Analysis Using Natural Language Inference
Felix Morger 30

Word Substitution with Masked Language Models as Data Augmentation for Sentiment Analysis
Larisa Kolesnichenko, Erik Velldal and Lilja Øvrelid 42

A Large Norwegian Dataset for Weak Supervision ASR
Per Erik Solberg, Pierre Beauguitte, Per Egil Kummervold and Freddy Wetjen 48

Lexical Semantics with Vector Symbolic Architectures
Adam Roussel 53

Linked Open Data compliant Representation of the Interlinking of Nordic Wordnets and Sign Language Data
Thierry Declerck and Sussi Olsen 62

Part-of-Speech tagging Spanish Sign Language data and its applications in Sign Language machine translation
Euan McGill, Luis Chiruzzo, Santiago Egea Gómez and Horacio Saggion 70

A Diagnostic Dataset for Sentiment and Negation Modeling for Norwegian
Petter Mæhlum, Erik Velldal and Lilja Øvrelid 77

Building Okinawan Lexicon Resource for Language Reclamation/Revitalization and Natural Language Processing Tasks such as Universal Dependencies Treebanking
So Miyagawa, Kanji Kato, Miho Zlazli, Salvatore Carlino and Seira Machida 86

Bridging the Resource Gap: Exploring the Efficacy of English and Multilingual LLMs for Swedish
Oskar Holmström, Jenny Kunz and Marco Kuhlmann 92

Phonotactics as an Aid in Low Resource Loan Word Detection and Morphological Analysis in Sakha
Petter Mæhlum and Sardana Ivanova 111

Vector Norms as an Approximation of Syntactic Complexity
Adam Ek and Nikolai Ilinykh 121

Low-Resource Techniques for Analysing the Rhetorical Structure of Swedish Historical Petitions
Ellinor Lindqvist, Eva Pettersson and Joakim Nivre 132