

Fourth Workshop on Resources for African Indigenous Languages (RAIL 2023)

Dubrovnik, Croatia
6 May 2023

ISBN: 978-1-7138-9061-4

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

| | |
|---|-----|
| <i>Automatic Spell Checker and Correction for Under-represented Spoken Languages: Case Study on Wolof</i> Thierno Ibrahima Cissé and Fatiha Sadat | 1 |
| <i>Unsupervised Cross-lingual Word Embedding Representation for English-isiZulu</i> Derwin Ngomane, Rooweither Mabuya, Jade Abbott and Vukosi Marivate | 11 |
| <i>Preparing the Vuk’uzenzele and ZA-gov-multilingual South African multilingual corpora</i> Richard Lastrucci, Jenalea Rajab, Matimba Shingange, Daniel Njini and Vukosi Marivate | 18 |
| <i>SpeechReporting Corpus: annotated corpora of West African traditional narratives</i> Ekaterina Aplonova, Izabela Jordanoska, Timofey Arkhangelskiy and Tatiana Nikitina | 26 |
| <i>A Corpus-Based List of Frequently Used Words in Sesotho</i> Johannes Sibeko and Orphée De Clercq | 32 |
| <i>Deep learning and low-resource languages: How much data is enough? A case study of three linguistically distinct South African languages</i> Roald Eiselen and Tanja Gaustad | 42 |
| <i>isiXhosa Intellectual Traditions Digital Archive: Digitizing isiXhosa texts from 1870-1914</i> Jonathan Schoots, Amandla Ngwendu, Jacques De Wet and Sanjin Muftic | 54 |
| <i>Analyzing political formation through historical isiXhosa text analysis: Using frequency analysis to examine emerging African Nationalism in South Africa</i> Jonathan Schoots | 65 |
| <i>Evaluating the Sesotho rule-based syllabification system on Sepedi and Setswana words</i> Johannes Sibeko and Mmasibidi Setaka | 76 |
| <i>Towards a Swahili Universal Dependency Treebank: Leveraging the Annotations of the Helsinki Corpus of Swahili</i> Kenneth Steimel, Sandra Kübler and Daniel Dakota | 86 |
| <i>Comparing methods of orthographic conversion for Bàsàá, a language of Cameroon</i> Alexandra O’neil, Daniel Swanson, Robert Pugh, Francis Tyers and Emmanuel Ngue Um | 97 |
| <i>Vowels and the Igala Language Resources</i> Mahmud Momoh | 106 |
| <i>Investigating Sentiment-Bearing Words- and Emoji-based Distant Supervision Approaches for Sentiment Analysis</i> Ronny Mabokela, Mpho Roborife and Turguy Celik | 115 |
| <i>Natural Language Processing in Ethiopian Languages: Current State, Challenges, and Opportunities</i> Atnafu Lambebo Tonja, Tadesse Destaw Belay, Israel Abebe Azime, Abinew Ali Ayele, Moges Ahmed Mehamed, Olga Kolesnikova and Seid Muhie Yimam | 126 |