

NLP for Indigenous Languages of the Americas: Transfer Learning Meets Translation

Katharina Kann¹

¹University of Colorado Boulder, USA

Abstract

Developing human language technology for truly low-resource languages, such as Indigenous languages, is challenging. Not only do we lack annotated training data, but even unlabeled data are often only available in small amounts. To make things worse, many truly low-resource languages are not represented in the pretraining data of multilingual language models. This talk will be centered around how to build NLP systems for Indigenous languages of the Americas. We will talk about the creation of AmericasNLI, a natural language inference dataset for Indigenous languages. Then, we will discuss model adaptation and translation-based approaches for the task. We will end this talk with a discussion of open questions and challenges for the development of NLP systems for Indigenous languages of the Americas.

The International Conference and Workshop on Agglutinative Language Technologies as a challenge of Natural Language Processing, ALTNLP'22, June 7-8, Koper, Slovenia

✉ katharina.kann@colorado.edu (K. Kann)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)