IWCLUL 2021

The Seventh International Workshop on Computational Linguistics of Uralic Languages

Proceedings of the Workshop

September 23–24, 2021

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-82-4

Preface

The seventh edition of ACL SIGUR's meeting IWCLUL is organised in conjunction with Electronic writing of the peoples of the Russian federation (EWPRF 2021) in Syktyvkar, Russia, the actual event was organised Online only due to the situation in 2021. This is the seventh event in the series and second hosted in Russia.

For the current proceedings of The Seventh International Workshop on Computational Linguistics for Uralic Languages, we accepted 8 high-quality submissions about topics ranging from overviews and insights into traditional language technology and resources all the way to modern neural network approaches in Uralic context and speech technology. The papers cover a wide range of Uralic languages from Finnish and North Sámi to Udmurt and Komi-Zyrian with several papers giving insight on whole range of Uralic langauges. Whereas some papers describe language-specific research, others compare different languages or work on small Uralic languages in general. These contributions are all very important for the preservation and development of Uralic languages as well as for future linguistic investigations on them.

As the conference is organised in collaboration with EWPRF, we have two full days of presentations as well as a round table, a regular business meeting of ACL SIGUR and time for discussions. The current proceedings include written papers of all of the IWCLUL oral presentations.

— The board of ACL SIGUR, October 13, 2021, Online / Syktyvkar

Organizing Committee

Association of Computational Linguistics' Special Interest Group for Uralic Languages (ACL SIGUR: https://acl-sigur.github.io)

Local organisers at Syktyvkar: http://conference.krags.ru

Programme Committee

- Flammie A Pirinen, UiT norgga árktalaš universitehta
- Timofey Arkhangelskiy Universität Hamburg
- Trond Trosterud, UiT
- Thierry Poibeau, LaTTiCe-CNRS
- Andrew Krizhanovsky, Institute of Applied Mathematical Research of the Karelian Research Centre of the Russian Academy of Sciences ordinary
- Svetlana Toldova, Higher School of Economics
- Mika Hämäläinen, University of Helsinki
- Francis M. Tyers, Indiana University Bloomington
- Csilla Horváth, Research Institute for Linguistics, Hungarian Academy of Sciences
- Jeremy Bradley, Ludwig Maximilian University of Munich
- Veronika Vincze, Hungarian Academy of Sciences, Research Group on Articial Intelligence
- Michael Rießler, Albert-Ludwigs-Universität Freiburg
- Andrey Kutuzov, University of Oslo
- Jack Rueter, University of Helsinki
- Joshua Wilbur, Univerity of Tartu
- Tommi Jauhiainen, University of Helsinki

Table of Contents

Conference Program

Thursday, September 23, 2021

13:45–14:00 *A never-published atlas of Udmurt dialects*
László Fejes

14:00–14:15 *Digitizing print dictionaries using TEI: The Abaev Dictionary Project*
Oleg Belyaev, Irina Khomchenkova, Julia Sinitsyna and Vadim Dyachkov

Friday, September 24, 2021

11:30–11:45 *Keyword spotting for audiovisual archival search in Uralic languages*
Nils Hjortnaes, Niko Partanen and Francis M. Tyers

11:45–12:00 *Evaluating Transferability of BERT Models on Uralic Languages*
Judit Ács, Dániel Lévai and Andras Kornai

12:00–12:15 *No more fumbling in the dark - Quality assurance of high-level NLP tools in a multi-lingual infrastructure*
Linda Wiechetek, Flammie A Pirinen, Børre Gaup and Thomas Omma

12:15–12:30 *Low-Resource ASR with an Augmented Language Model*
Timofey Arkhangelskiy

12:30–12:45 ***Discussion***

12:45–13:00 *The Current State of Finnish NLP*
Mika Hämäläinen and Khalid Alnajjar

13:00–13:15 *Overview of Open-Source Morphology Development for the Komi-Zyrian Language: Past and future*
Jack Rueter, Niko Partanen, Mika Hämäläinen and Trond Trosterud

14:30-15:30 ***ACL SIGUR meeting***