ACL SIGUR 2020

The Sixth International Workshop on Computational Linguistics for Uralic Languages

Proceedings of the Conference

January 10 — 11, 2020
Wien, Austria

© 2020 The Association for Computational Linguistics

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 978-1-952148-00-2

Introduction

IWCLUL 2020 was held in Wien, Austria.

Organizers:

ACL SIGUR:
Tommi A. Pirinen, University of Hamburg
Francis M. Tyers, Indiana University Bloomington

Local Organizers:

Jeremy Bradley, University of Wien
Johannes Hirvonen, University of Wien

Program Committee:

Tommi A. Pirinen, University of Hamburg
Francis M. Tyers, Indiana University Bloomington
Jeremy Bradley, University of Vienna
Eszter Simon, Research Institute for Linguistics, Hungarian Academy of Sciences
Kadri Muischnek, University of Tartu
Michael Rießler, University of Eastern Finland
Filip Ginter, University of Turku
Timofey Arkhangelskiy, University of Hamburg / Alexander von Humboldt Foundation
Veronika Vincze, Research Group on Articial Intelligence, Hungarian Academy of Sciences
Roman Yangarber, University of Helsinki
Miikka Silfverberg, University of Helsinki

Invited Speaker:

Timofey Arkhangelskiy, University of Hamburg / Alexander von Humboldt Foundation

Table of Contents

A pseudonymisation method for language documentation corpora: An experiment with spoken Komi
Rogier Blokland, Niko Partanen and Michael Rießler 1

Effort-value payoff in lemmatisation for Uralic languages
Nick Howell, Maria Bibaeva and Francis M. Tyers 9

On the questions in developing computational infrastructure for Komi-Permyak
Jack Rueter, Niko Partanen and Larisa Ponomareva 15

On Editing Dictionaries for Uralic Languages in an Online Environment
Khalid Alnajjar, Mika Hämäläinen and Jack Rueter 26

Towards a Speech Recognizer for Komi, an Endangered and Low-Resource Uralic Language
Nils Hjortnaes, Niko Partanen, Michael Rießler and Francis M. Tyers 31

Hunting for antiharmonic stems in Erzya
László Fejes 38

apPILcation: an Android-based Tool for Learning Mansi
Gábor Bobály, Csilla Horváth and Veronika Vincze 48

Conference Program

Friday, January 10th, 2020

9:00–10:00 ***Registration***

10:00–10:15 ***Opening***

10:15–11:00 *Invited Talk: Timofey Arkhangelskiy*

11:00–11:30 ***Coffee break***

11:30–13:00 **Session 1**

11:30–12:00 *A pseudonymisation method for language documentation corpora: An experiment with spoken Komi*
Rogier Blokland, Niko Partanen and Michael Rießler

12:00–12:30 *Effort-value payoff in lemmatisation for Uralic languages*
Nick Howell, Maria Bibaeva and Francis M. Tyers

12:30–13:00 *On the questions in developing computational infrastructure for Komi-Permyak*
Jack Rueter, Niko Partanen and Larisa Ponomareva

13:00–14:00 ***Lunch***

Friday, January 10th, 2020 (continued)

14:00–15:00 Session 2

14:00–14:30 *On Editing Dictionaries for Uralic Languages in an Online Environment*
Khalid Alnajjar, Mika Hämäläinen and Jack Rueter

14:30–15:00 *Towards a Speech Recognizer for Komi, an Endangered and Low-Resource Uralic Language*
Nils Hjortnaes, Niko Partanen, Michael Rießler and Francis M. Tyers

15:00–15:30 ***Coffee break***

15:30–17:00 Session 3

15:30–16:00 *Hunting for antiharmonic stems in Erzya*
László Fejes

16:00–16:30 *apPILcation: an Android-based Tool for Learning Mansi*
Gábor Bobály, Csilla Horváth and Veronika Vincze

16:30–17:00 ***Closing***

17:00–18:30 ***SIGUR business meeting***

20:00 ***Dinner***

Saturday, January 11th, 2020

9:00–16:00 ***Tutorials***