IWCLUL

First International Workshop on Computational Linguistics for Uralic Languages

Proceedings of the Workshop

January 16th, 2015
Tromsø, Norway

WWW address: http://dx.doi.org/10.7557/scs.2015.2
eISSN: 2387-3086
DOI: (whole proceedings) 10.7557/scs.2015.2, specific articles, see footers
Editors contact: iwclul-2015@googlegroups.com

Preface

The Uralic languages are an interesting group of languages from computational-linguistic perspective. They share large parts of morphological and morphophonological complexity that is not present in the Indo-European group which has traditionally dominated computational-linguistic research. This can be seen for example in number of word forms per word, which in Indo-European languages is in range of ones or tens whereas for Uralic languages it is in range of hundreds and thousands. Furthermore, Uralic languages share a lot of geo-political aspects: the national languages of the group—Finnish, Estonian and Hungarian—are small languages and only moderately resourced in terms of computational-linguistic resources while being stable and not in threat of extinction, the recognised minority languages of western-European states—such as North Sámi and Võro—are clearly in category of lesser resourced and more threatened, whereas the majority of Uralic languages in the east of Europe and Russia are close to extinction. Common to all rapid development of more advanced computational-linguistic methods is required for continued vitality of the languages in everyday life, to enable archiving and use of the languages with computers and other devices such as mobile applications.

The research of computational linguistics and Uralistics is being carried out in a handful of universities, research institutes and other sites by relatively few researchers. Our intention with organising this conference is to gather these researchers together in order to share ideas and resources, and avoid duplicating efforts in gathering and enriching these scarce resources, and hopefully to found an ongoing tradition of concentrated effort in collecting and improving language resources and technologies for the survival of the Uralic languages.

For the conference we received 14 high-quality submissions about topics including computational lexicography, language documentation, optical character recognition, web-as-corpus and automatic and rule-based morphological analysis methods. These are all very important for preservation and development of Uralic languages. We also received a broad coverage of languages in the submissions: North Sámi, Khanty, Mansi, Udmurt, Erzya, Moksha, Finnish and Estonian.

The conference was held at UiT Norgga árktalaš universitehta, Norway, on January

16th 2015, and consisted of poster sessions, three talks, two tutorials, and an invited speech, The articles related to poster sessions and the talks are included in this proceedings.
—Tommi A Pirinen, Francis M. Tyers, Trond Trosterud,
Conference organisers,
2015, Tromsø

Organisers

- Tommi A. Pirinen, Ollscoil Chathair Bhaile Átha Cliath
- Francis M. Tyers, UiT Norgga árktalaš universitehta
- Trond Trosterud, UiT Norgga árktalaš universitehta

Programme committee

- Тимофей Архангельский, Национальный исследовательский университет ”Высшая школа экономики”
- Lars Borin, Göteborgs universitet
- Марина Серафимовна Федина, Финн-йӧгра кывъяслы информатика отсӧг кузя регионкостса лаборатория
- Mark Fishel, Tartu ülikool
- Mikel L. Forcada, Universitat d’Alacant
- Mans Hulden, University of Colorado at Boulder
- Heiki-Jaan Kaalep, Tartu ülikool
- András Kornai, Budapesti Műszaki és Gazdaságtudományi Egyetem
- Krister Lindén, Helsingin yliopisto
- Tommi A. Pirinen, Ollscoil Chathair Bhaile Átha Cliath
- Gabór Prószéky, Pázmány Péter Katolikus Egyetem
- Aarne Ranta, Chalmers tekniska högskola
- Jack Rueter, Helsingin yliopisto
- Trond Trosterud, UiT Norgga árktalaš universitehta
- Francis M. Tyers, UiT Norgga árktalaš universitehta
- Sami Virpioja, Aalto-yliopisto
- Anssi Yli-Jyrä, Helsingin yliopisto

Contents