

LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

**IWLTP 2020 – 1st International Workshop on
Language Technology Platforms**

PROCEEDINGS

Edited by:
Georg Rehm, Kalina Bontcheva, Khalid Choukri,
Jan Hajič, Stelios Piperidis, Andrejs Vasiļjevs

Proceedings of the LREC 2020 Workshop IWLTP 2020 – 1st International Workshop on Language Technology Platforms

Edited by: Georg Rehm, Kalina Bontcheva, Khalid Choukri, Jan Hajič, Stelios Piperidis, Andrejs Vasiljevs

Acknowledgement: this work has received funding from the EU's Horizon 2020 research and innovation programme through the contract ELG (grant agreement no. 825627).

<https://www.european-language-grid.eu/iwltp-2020/>



ISBN: 979-10-95546-64-1

EAN: 9791095546641

For more information:

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: lrec@elda.org

© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License



Preface

With the increasing number of platforms, grids and infrastructures in the wider area of Language Technologies (LT), NLP, NLU, speech, interaction and language-centric AI, there is also a growing need for sharing experiences, approaches and best practices to learn and benefit from the work of others and also, practically, to start a collaboration towards platform interoperability.

The 1st International Workshop for Language Technology Platforms (IWLTP 2020) addresses all smaller and larger language grids, language-related infrastructures, platform initiatives as well as collaborative research projects that touch upon LT platforms, especially platform interoperability and related topics, both in Europe and world-wide. Its objective is to exchange and discuss observations, experiences, solutions, best practices as well as current and future challenges. The workshop also addresses the issue of fragmentation in the Language Technology landscape. Instead of “platform islands” that simply exist side by side, possibly even competing with each other, initiatives should discuss how their platforms can be made interoperable and how they can interact with one another to create synergies towards a productive LT platform ecosystem.

The EU project European Language Grid (ELG; 2019-2021) is creating a platform that will provide thousands of data sets and hundreds of LT services. ELG aims to promote technologies tailored to all European languages and cultures, adapted to their social and economic needs. At the same time, there are several established platforms or infrastructure-related initiatives as well as emerging new ones, both on the European but also on the national level as well as on other continents. Some of the initiatives are more language-related and have a strong industry focus, others are mainly research-oriented. Moreover, there are digital public service initiatives, and platforms, in which language is only one aspect of many. With all these established and emerging initiatives, there is a risk of even stronger fragmentation in the Language Technology field, which is already highly fragmented, at least in Europe. Our approach is to bring these initiatives together to discuss ways not only of preventing further fragmentation but, crucially, of reversing it. This will only be possible if interoperability and mutual data exchange is ensured and if metadata formats and technical requirements are compatible, among others.

A total of 30 papers were submitted to IWLTP 2020, 17 of which were accepted (acceptance rate: 56.7%). The organisers would like to thank all contributors for their valuable submissions and all members of the Programme Committee for reviewing the submitted papers. Due to the ongoing SARSCoV-2 pandemic, the workshop cannot be held as originally foreseen. Together with the organisers of LREC 2020 we will explore if we can organise the workshop at a later point in time or if we can organise it as a virtual event.

G. Rehm, K. Bontcheva, K. Choukri, J. Hajič, S. Piperidis, A. Vasiljevs

May 2020

Organising Committee and Programme Committee

Organising Committee

Georg Rehm (DFKI GmbH, Germany) – main editor and chair
Kalina Bontcheva (University of Sheffield, UK)
Khalid Choukri (ELDA, France)
Jan Hajič (Charles University, Czech Republic)
Stelios Piperidis (ILSP, R. C. “Athena”, Greece)
Andrejs Vasiljevs (Tilde, Latvia)

Programme Committee

Albina Auksoriūtė (Institute of the Lithuanian Language, Lithuania)
António Branco (University of Lisbon, Portugal)
Gerhard Budin (University of Vienna, Austria)
Walter Daelemans (University of Antwerp, Belgium)
Christian Dirschl (Wolters Kluwer, Germany)
Maria Gavriilidou (ILSP, R. C. “Athena”, Greece)
Stefan Geißler (Kairntech, France)
José Manuel Gómez Pérez (Expert System Iberia, Spain)
Manuel Herranz (Pangeanic, Spain)
Ilan Kernerman (K Dictionaries, Israel)
Svetla Koeva (Bulgarian Academy of Sciences, Bulgaria)
Simon Krek (Jožef Stefan Institute, Slovenia)
Cvetana Krstev (University of Belgrade, Serbia)
Kristen Lindén (University of Helsinki, Finland)
Dora Loizidou (University of Cyprus, Cyprus)
Jean-Pierre Lorré (Linagora, France)
Bernardo Magnini (Fondazione Bruno Kessler, Italy)
Jan Odijk (Utrecht University, Netherlands)
Maciej Ogrodniczuk (Polish Academy of Sciences, Poland)
Christoph Prinz (Sail Labs, Austria; LT Innovate, Belgium)
Sören Räuchle (3pc, Germany)
Artem Revenko (Semantic Web Company, Austria)
Eiríkur Rögnvaldsson (University of Iceland, Iceland)
Bolette Sandford Pedersen (University of Copenhagen, Denmark)
Inguna Skadina (University of Latvia, Latvia)
Marko Tadić (University of Zagreb, Croatia)
Dan Tufiş (Romanian Academy of Sciences, Romania)
Tamás Váradi (Hungarian Academy of Sciences, Hungary)
Phillippe Wacker (LT Innovate, Belgium)
Andy Way (Dublin City University, Ireland)
François Yvon (CNRS – LIMSI, France)
Matteo Zanioli (Alpenite, Italy)

Table of Contents

Infrastructure for the Science and Technology of Language PORTULAN CLARIN

António Branco, Amália Mendes, Paulo Quaresma, Luís Gomes, João Silva
and Andrea Teixeira 1

On the Linguistic Linked Open Data Infrastructure

Christian Chiarcos, Bettina Klimek, Christian Fäth, Thierry Declerck and John P. McCrae 8

Architecture of a Scalable, Secure and Resilient Translation Platform for Multilingual News Media

Susie Coleman, Andrew Secker, Rachel Bawden, Barry Haddow and Alexandra Birch 16

CoBiLiRo: A Research Platform for Bimodal Corpora

Dan Cristea, Ionuț Pistol, Șerban Boghiu, Anca-Diana Bibiri, Daniela Gîfu, Andrei Scutelnicu, Mihaela Onofrei, Diana Trandabăț and George Bugeag 22

CLARIN: Distributed Language Resources and Technology in a European Infrastructure

Maria Eskevich, Franciska de Jong, Alexander König, Darja Fišer, Dieter Van Uytvanck, Tero Aalto, Lars Borin, Olga Gerassimenko, Jan Hajic, Henk van den Heuvel, Neeme Kahusk, Krista Liin, Martin Matthiesen, Stelios Piperidis and Kadri Vider 28

ELRI: A Decentralised Network of National Relay Stations to Collect, Prepare and Share Language Resources

Thierry Etchegoyhen, Borja Anza Porras, Andoni Azpeitia, Eva Martínez Garcia, José Luis Fonseca, Patricia Fonseca, Paulo Vale, Jane Dunne, Federico Gaspari, Teresa Lynn, Helen McHugh, Andy Way, Victoria Arranz, Khalid Choukri, Hervé Pusset, Alexandre Sicard, Rui Neto, Maite Melero, David Perez, António Branco, Ruben Branco and Luís Gomes 35

Removing European Language Barriers with Innovative Machine Translation Technology

Dario Franceschini, Chiara Canton, Ivan Simonini, Armin Schweinfurth, Adelheid Glott, Sebastian Stüker, Thai-Son Nguyen, Felix Schneider, Thanh-Le Ha, Alex Waibel, Barry Haddow, Philip Williams, Rico Sennrich, Ondřej Bojar, Sangeet Sagar, Dominik Macháček and Otakar Smrž 44

Eco.pangeamt: Industrializing Neural MT

Mercedes García-Martínez, Manuel Herranz, Amando Estela, Ángela Franco and Laurent Bié 50

The Kairntech Sherpa – An ML Platform and API for the Enrichment of (not only) Scientific Content

Stefan Geißler 54

Towards Standardization of Web Service Protocols for NLPaaS

Jin-Dong Kim, Nancy Ide and Keith Suderman 59

NTeALan Dictionaries Platforms: An Example Of Collaboration-Based Model

Elvis Mboning, Daniel Baleba, Jean Marc Bassahak, Ornella Wandji and Jules Assoumou 66

A Workflow Manager for Complex NLP and Content Curation Pipelines

Julián Moreno-Schneider, Peter Bourgonje, Florian Kintzel and Georg Rehm 73

A Processing Platform Relating Data and Tools for Romanian Language

Vasile Păiș, Radu Ion and Dan Tufiş 81

LinTO Platform: A Smart Open Voice Assistant for Business Environments

Ilyes Rebai, Kate Thompson, Sami Benhamiche, Zied Sellami, Damien Laine
and Jean-Pierre Lorre 89

Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability

Georg Rehm, Dimitrios Galanis, Penny Labropoulou, Stelios Piperidis, Martin Weiß, Ricardo Usbeck, Joachim Köhler, Miltos Deligiannis, Katerina Gkirtzou, Johannes Fischer, Christian Chiarcos, Nils Feldhus, Julián Moreno-Schneider, Florian Kintzel, Elena Montiel, Víctor Rodríguez-Doncel, John P. McCrae, David Laqua, Irina Patricia Theile, Christian Dittmar, Kalina Bontcheva, Ian Roberts, Andrejs Vasiljevs and Andis Lagzdīņš 96

The COMPRISE Cloud Platform

Raivis Skadiņš and Askars Salimbajevs 108

From Linguistic Research Projects to Language Technology Platforms: A Case Study in Learner Data

Annanda Sousa, Nicolas Ballier, Thomas Gaillat, Bernardo Stearns, Manel Zarrouk, Andrew Simpkin and Manon Bouyé 112

List of Authors 121