



3rd Call for Papers

CCURL 2014

Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era

<http://www.ilc.cnr.it/ccurl2014/>

*Reykjavík, ICELAND, 26th May 2014, in conjunction with LREC 2014
Submission deadline: 20th February 2014 (all dates below)*

The LREC 2014 Workshop “CCURL 2014: Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era” will be held in Reykjavík, ICELAND, on Monday 26th May 2014.

Under-resourced languages suffer from a chronic lack of available resources (human-, financial-, time- and data-wise), and of the fragmentation of efforts in resource development. This often leads to small resources only usable for limited purposes or developed in isolation without much connection with other resources and initiatives. The benefits of reusability, accessibility and data sustainability are, more often than not, out of the reach of such languages.

Yet, these languages are those that could most profit from emergent collaborative approaches and technologies for language resource development. Given the high cost of language resource production, and given the fact that in many cases it is impossible to avoid the manual construction of resources (e.g. if accurate models are requested or if there is to be reliable evaluation) it is worth considering the power of social and collaborative media to build resources, especially for those languages where there are no or limited language resources built by experts yet.

Collaborative, Web 2.0 and Web 3.0 / Semantic Web methods and methodologies for data collection, annotation and sharing seem particularly well-suited for collecting the data

needed for the development of language technology applications for under-resourced languages. Indeed, the collaborative accumulation and creation of data appears to be the best and most practicable way to achieve better and faster language coverage and in purely economic terms could well deliver a higher return on investment than expected. Moreover, it is a good way to approach a small population of speakers who live in remote countries, or are scattered in diaspora all over the world.

The workshop aims at gathering together professionals involved with language resources for under-resourced languages. The expectation is that both academic researchers and industry practitioners will participate.

Some specific questions that the workshop will aim to answer include the following:

- How can collaborative approaches and technologies be fruitfully applied to the development and sharing of resources for under-resourced languages?
- How can small language resources be re-used efficiently and effectively, reach larger audiences and be integrated into applications?
- How can they be stored, exposed and accessed by end users and applications?
- How can research on such languages benefit from semantic and semantic Web technologies, and specifically the Linked Data framework?

We invite papers reporting on collaborative methodologies for the development of language resources for under-resourced languages, the processes involved, as well as on issues relating to their usability, e.g. design guidelines, standards for building and sharing resources, storage and exchange formats, interoperability issues, etc.

We therefore specifically encourage submissions about:

- Experiences in the creation of Linked Open Data and/or Linguistic Linked Open Data for under-resourced languages;
- Using existing Linked Open Data knowledge resources such as DBpedia, Freebase, YAGO, Lexvo, schema.org, etc. in semantics-driven approaches to resource development for under-resourced languages;
- Scaling existing language resource infrastructures to thousands of languages;
- Crowd-sourcing of linguistic data and annotations;
- Collaborative bootstrapping of language resources and language technologies (LRTs) for under-resourced languages from existing LRTs for better-resourced languages;
- Mining the Web and social media for linguistic data;
- Developing and/or using language-independent software frameworks for under-resourced languages and other collaborations across language groups;
- Ethical, sociological and practical issues in collaborative approaches and technologies;

- Usability of existing infrastructures for the development of collaboratively created resources.

SUBMISSIONS

- Papers must describe original unpublished work, either completed or in progress.
- Each submission will be reviewed by three programme committee members. The paper review will be blind, so papers should not include authors' names and affiliations.
- Accepted papers will be presented either as oral presentations or posters and will be published in the workshop proceedings.
- Papers should be formatted according to the stylesheet provided on the LREC 2014 Web site and should not exceed 8 pages for oral presentations, and 4 pages for posters, including references and appendices. Papers should be submitted in PDF unprotected format to the workshop START page: <https://www.softconf.com/lrec2014/CCURL/>.
- When submitting a paper from the START page, authors will be asked to provide essential information about resources (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a new result of your research. Moreover, ELRA encourages all LREC authors to share the described LR (data, tools, services, etc.), to enable their reuse, replicability of experiments, including evaluation ones, etc. For further information, please refer to this URL: <http://lrec2014.lrec-conf.org/en/calls-for-papers/lrec-2014-special-highlight/>.

IMPORTANT DATES

- 20th February 2014: Paper submission deadline
- 10th March 2014: Notification of acceptance
- 26th March 2014: Camera-ready papers due
- 26th May 2014: Workshop

ORGANISING COMMITTEE

- Laurette Pretorius - University of South Africa, South Africa
- Claudia Soria - CNR-ILC, Italy
- Eveline Wandl-Vogt - Austrian Academy of Sciences, ICLTT, Austria
- Thierry Declerck - DFKI GmbH, Language Technology Lab, Germany
- Kevin Scannell - St. Louis University, USA
- Joseph Mariani - LIMSI-CNRS & IMMI, France

PROGRAMME COMMITTEE

- Deborah W. Anderson - University of Berkeley, Linguistics, USA
- Sabine Bartsch - Technische Universität Darmstadt, Germany
- Delphine Bernhard - LILPA, Strasbourg University, France
- Bruce Birch - The Minjilang Endangered Languages Publications Project, Australia
- Paul Buitelaar - DERI, Galway, Ireland
- Peter Bouda - CIDLeS - Interdisciplinary Centre for Social and Language Documentation, Portugal
- Steve Cassidy - Macquarie University, Australia
- Christian Chiarcos - University of Potsdam, Germany
- Katrien Depuydt - Instituut voor Nederlandse Lexicologie, The Netherlands
- Vera Ferreira - CIDLeS - Interdisciplinary Centre for Social and Language Documentation, Portugal
- Claudia Garad - wikimedia.AT, Austria
- Dafydd Gibbon - Bielefeld University, Germany
- Oddrun Gronvik - Instituut for lingvistike og nordiske studier, University of Oslo, Norway
- Yoshihiko Hayashi - University of Osaka, Japan
- Dominic Jones - Trinity College Dublin, Ireland
- Daniel Kaufman - Endangered Language Alliance, USA
- Andras Kornai - Hungarian Academy of Sciences, Hungary
- Simon Krek - Jožef Stefan Institute, Slovenia
- Tobias Kuhn - ETH, Zurich, Switzerland
- Leonel Ruiz Miyares - Centro de Linguística Aplicada (CLA), Cuba
- Karlheinz Mörth - Austrian Academy of Sciences, ICLTT, Austria
- Steven Moran - University of Washington, USA
- Roberto Navigli - Università Degli Studi di Roma La Sapienza, Italy
- Kellen Parker - National Tsing Hua University, China
- Patrick Paroubek - LIMSI-CNRS, France
- Maria Pilar Perea i Sabater - Universitat de Barcelona, Spain
- Ulrich Schäfer - DFKI GmbH, Germany
- Caroline Sporleder - Universität Trier, Germany
- Nick Thieberger - University of Melbourne, Australia
- Piek Vossen - VU Amsterdam, The Netherlands
- Marianne Vergez-Couret - Toulouse University, France
- Michael Zock - LIF-CNRS, France