

# **CCURL 2018**

## **Collaboration and Computing for Under-Resourced Languages**

### **‘Sustaining knowledge diversity in the digital age’**

a Workshop to be held as part of the 11<sup>th</sup> edition of the Language Resources and Evaluation Conference (LREC 2018) at the Phoenix Seagaia Resort in Miyazaki (Japan)

## **1<sup>ST</sup> CALL FOR PAPERS**

Date: 12<sup>th</sup> May 2018

Web site: <http://www.ilc.cnr.it/ccurl2018>

Submission deadline: 13<sup>th</sup> January 2018

### **WORKSHOP DESCRIPTION AND OBJECTIVE**

The 3<sup>rd</sup> CCURL Workshop, entitled ‘Sustaining knowledge diversity in the digital age’, will take place on Saturday 12<sup>th</sup> May 2018 in Miyazaki (Japan), in conjunction with LREC 2018. This workshop aims at gathering together academics, industrial researchers, knowledge experts, digital language resource and technology providers, software developers, but also language activists and community representatives in order to identify the current capacity for and the difficulties in creating and sustaining the digital representation of traditional knowledge.

The diversity of cultures is a distinctive footprint of the way humans have been coping with the environment over time; unique visions of the world and knowledge are expressed by indigenous languages. Preservation and sharing of the traditional knowledge encoded by languages is being increasingly recognised as a step towards a sustainable and durable interaction of mankind with the environment. However, as language diversity is decreasing, the maintenance and transmission of such knowledge is at risk. Digital language resources can help avoid the disappearance of diverse knowledge systems, ensure their preservation and transmission, and foster their cross-fertilisation. The vast majority of this knowledge is poorly represented in digital form (only four out of the 522 indigenous languages of Latin America are represented by Wikipedia projects, for example). Moreover, as this knowledge is encoded in under-resourced (minority, endangered or minoritised) languages, specific methods and models of resource

development are required to circumvent the problems affecting low-resourced languages, such as low investments, data sparsity, fragmentation of efforts, speaker communities' lack of involvement, to cite just a few. Specific problems arise as well: low digital literacy, the issue of community ownership and control over content, or the need to include audio and video to accommodate languages that are unwritten or having no orthography standard.

## **TOPICS OF INTEREST**

We solicit papers and posters related to the following non-exclusive topics:

- models and methods for the development of language resources for representing traditional knowledge;
- experiences about forms of collaboration among research, industry and local communities;
- involvement of speakers' communities and ethical issues related to knowledge protection;
- replicability of experiences;
- use of knowledge resources for cultural heritage preservation and education;
- use of video and audio as complementary or alternative ways to writing in order to accommodate languages not spoken or with unstable orthographies;
- innovative data collection and data annotation methodologies;
- semantic and semantic web technologies for representing indigenous knowledge systems in indigenous languages.

## **SUBMISSION AND PUBLICATION**

We accept submission of long papers (up to 8 pages), short papers (up to 4 pages) and poster papers (up to 4 pages) to be presented as a long or short oral presentation at the workshop. The papers of the workshop will be published in online proceedings. Papers are expected to address the workshop main theme. They can contain an analysis and insight into existing methods and problems; a description of resources; an overview of the literature or of the current initiatives, or a combination of the above. Authors must declare if part of the paper contains material previously published elsewhere. Each submission will be reviewed by three programme committee members. In compliance with the LREC rules, papers must not be anonymized. Papers should be formatted according to the stylesheet provided by LREC 2018 (<http://lrec2018.lrec-conf.org/en/submission/authors-kit/>) and should not exceed 8 pages, including references and appendices. Papers should be submitted in PDF unprotected format to the workshop START page (URL will be provided in due time). The formatting template must be strictly adhered to and deadlines met.

## IMPORTANT DATES

- Paper submission deadline: 13<sup>th</sup> January 2018
- Notification of acceptance: 14<sup>th</sup> February 2018
- Camera-ready paper: 7<sup>th</sup> March 2018
- Workshop date: 12<sup>th</sup> May 2018

## IDENTIFY, DESCRIBE AND SHARE YOUR LRS!

- Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about “Sharing LRs” (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new “regular” feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.
- As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2018 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, [www.islrn.org](http://www.islrn.org)), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

## ORGANISING COMMITTEE

- Laurent Besacier (LIG-IMAG, France)
- Laurette Pretorius (University of South Africa, South Africa)
- Claudia Soria (CNR-ILC, Italy)

The Workshop is endorsed by SIGUL, a joint ELRA-ISCA Special Interest Group on Under-resourced Languages (<http://www.elra.info/en/sig/sigul/>).

## PROGRAMME COMMITTEE

- Tunde Adegbola (African Languages Technology Initiative, Nigeria)
- Gilles Adda (LIMSI/IMMI CNRS, France)
- Shyam Agrawal (KIIT Group of Colleges, India)
- Amir Aharoni (Wikimedia Foundation)

- Antti Arppe (University of Alberta, Canada)
- Victoria Arranz (ELRA/ELDA, France)
- Martin Benjamin (the Kamusi Project, Switzerland)
- Laurent Besacier (LIG-IMAG, France)
- Bruce Birch (The Minjilang Endangered Languages Publications Project, Australia)
- Steven Bird (Charles Darwin University, Australia)
- Luong Chi-Mai (IOIT, Vietnam)
- Khalid Choukri (ELRA/ELDA, France)
- Chris Cieri (LDC, USA)
- Thierry Declerck (DFKI, Germany)
- Sebastian Drude (The Vigdís International Centre for Multilingualism and Intercultural Understanding, Iceland)
- Vera Ferreira (CIDLeS - Interdisciplinary Centre for Social and Language Documentation, Portugal)
- Mikel Forcada (Universitat d'Alacant, Spain)
- Dafydd Gibbon (Bielefeld University, Germany)
- Tatjana Gornostaja (Tilde, Latvia)
- John Judge (ADAPT DCU, Ireland)
- Andras Kornai (Hungarian Academy of Sciences, Hungary)
- Joseph Mariani (LIMSI-CNRS, France)
- Yohei Murakami (Kyoto University, Japan)
- Satoshi Nakamura (NARA Institute of Science and Technology, Japan)
- Girish Nath Jha (JNU, India)
- Guy de Pauw (Textgain, Belgium)
- Laurette Pretorius (University of South Africa, South Africa)
- Sakriani Sakti (NAIST, Japan)
- Kevin Scannell (Saint Louis University, Missouri, USA)
- Claudia Soria (CNR-ILC, Italy)
- Oliver Stegen (SIL International, USA)
- Francis Tyers (Moscow Higher School of Economics, Russia)
- Trond Trosterud (Tromsø University, Norway)
- Kadri Vider (University of Tartu, Estonia)
- Eveline Wandl-Vogt (Austrian Academy of Sciences, Austria)