Call for papers

Collaboration: interoperability between people in the creation of language resources for less-resourced languages

LREC 2008 pre-conference workshop
Marrakech, Morocco: afternoon of Tuesday 27th May 2008
Organised by the SALTMIL Special Interest Group of ISCA

SALTMIL: http://ixa2.si.ehu.es/saltmil/
LREC 2008:   http://www.lrec-conf.org/lrec2008/
Call For Papers: http://ixa2.si.ehu.es/saltmil/en/activities/lrec2008/lrec-2008-workshop-cfp.html
Paper submission: http://www.easychair.org/conferences/?conf=saltmil2008

Papers are invited for the above half-day workshop, in the format outlined below. Most submitted papers will be presented in poster form, though some authors may be invited to present in lecture format.

Context and Focus

The minority or "less resourced" languages of the world are under increasing pressure from the major languages (especially English), and many of them lack full political recognition. Some minority languages have been well researched linguistically, but most have not, and the majority do not yet possess basic speech and language resources which would enable the commercial development of products. This lack of language products may accelerate the decline of those languages that are already struggling to survive. To break this vicious circle, it is important to encourage the development of basic language resources as a first step.

In recent years, linguists across the world have realised the need to document endangered languages immediately, and to publish the raw data. This raw data can be transformed automatically (or with the help of volunteers) into resources for basic speech and language technology. It thus seems necessary to extend the scope of recent workshops on speech and language technology beyond technological questions of interoperability between digital resources: the focus will be on the human aspect of creating and disseminating language resources for the benefit of endangered and non-endangered less-resourced languages.

Topics

The theme of "collaboration" centres on issues involved in collaborating with:

- Trained researchers.
- Non-specialist workers (paid or volunteers) from the speaker community.
- The wider speaker community.
- Officials, funding bodies, and others.

Hence there will be a corresponding need for the following:

- With trained researchers: Methods and tools for facilitating collaborative working at a distance.
- With non-specialist workers: Methods and tools for training new workers for specific tasks, and laying the foundations for continuation of these skills among native speakers.
- With the wider speaker community: Methods of gaining acceptance and wider publicity for the work, and of increasing the take-up rates after completion of the work.
- With others: Methods of presenting the work in non-specialist terms, and of facilitating its progress.

Topics may include, but are not limited to:

- Bringing together people with very different backgrounds.
- How to organize volunteer work (some endangered languages have active volunteers).
- How to train non-specialist volunteers in elicitation methods.

- Working with the speaker community: strengthening acceptance of ICT and language resources among the speaker community.
- Working collaboratively to build speech and text corpora with few existing language resources and no specialist expertise.
- Web-based creation of linguistic resources, including web 2.0.
- The development of digital tools to facilitate collaboration between people.
- Licensing issues; open source, proprietary software.
- Re-use of existing data; interoperability between tools and data.
- Language resources compatible with limited computing power environments (old machines, the $100 handheld device, etc.)
- General speech and language resources for minority languages, with particular emphasis on software tools that have been found useful.

Important dates

29 February 2008	Deadline for submission
17 March 2008	Notification
31 March 2008	Final version
27 May 2008	Workshop

Organisers

- Briony Williams: Language Technologies Unit, Bangor University, Wales, UK
- Mikel Forcada: Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, Spain
- Kepa Sarasola: Dept. of Computer Languages, University of the Basque Country

Programme Committee

- Briony Williams: Bangor University, Wales, UK
- Mikel Forcada: Universitat d'Alacant, Spain
- Kepa Sarasola: University of the Basque Country
- Atelach Alemu Argaw: Stockholm University, Sweden
- Julie Berndsen, University College Dublin, Ireland
- Shannon Bischoff, Universidad de Puerto Rico, Puerto Rico
- Lori Levin, Carnegie-Mellon University, USA
- Climent Nadeu, Universitat Politècnica de Catalunya, Spain
- Juan Antonio Pérez-Ortiz, Universitat d'Alacant, Spain
- Bojan Petek, University of Ljubljana, Slovenia
- Oliver Streiter, National University of Kaohsiung, Taiwan

Submission information

We expect short papers of max 3500 words (about 4-6 pages) describing research addressing one of the above topics, to be submitted as PDF documents by uploading to the following URL:

http://www.easychair.org/conferences/?conf=saltmil2008

The final papers should not have more than 6 pages, adhering to the stylesheet that will be adopted for the LREC Proceedings (to be announced later on the Conference web site).