Call for Papers

Post LREC-2008 Workshop:

Sustainability of Language Resources and Tools
for Natural Language Processing

Marrakech, Morocco
Saturday, 31 May 2008

*********************************************************************


Sustainability of Language Resources and Tools for Natural Language
Processing


One of the problems in Natural Language Processing and related fields is
that
the sustainability of language resources (e.g., corpora) and of language
technology
tools (e.g. annotation or query tools) are neglected on a regular basis.

This results in, for example, tools whose algorithms and data structures
are poorly
documented and whose area of application is evident only to the people
who
built the software. Similar issues arise with regard to language
resources:
often, these are tailored to the needs of an individual application or
of
a project with a very specific research question. When the project is
finished it
becomes next to impossible (especially for third parties) to gain access
to the
resource that may have taken several months or even years to create.

The very complex question of how to ensure or maybe even guarantee
sustainability is related to several key issues spanning a broad
spectrum
across several closely related fields: in the area of language
documentation,
seven dimensions of portability (content, format, discovery, access,
citation,
preservation, rights) have been suggested. Another area of research is
primarily concerned with annotation technology, especially the problem
of
building generic annotation frameworks as well as representing several
different layers of linguistic annotation referring to one specific set
of
primary data by means of standoff annotation. Closely related work deals
with
the standardisation of annotation frameworks, especially with regard to
the
level of impact a specific linguistic theory has on their vocabularies
and
markup grammars. A last area concerns the fostering of sustainability
through specific
Software Engineering processes for Computational Linguistics and Natural

Language
Processing tools, applications and resources. At the moment, we are not aware
of previous work in this latter field.


Providing sustainability for linguistic tools and language resources becomes
increasingly important for the research community. Nowadays, this is also
acknowledged by funding organisations -- they often encourage research
projects to make sure that language resources will still be accessible and
(re-)usable in ten, 15, or 20 years time.

The problem of ensuring sustainability is a multi-faceted one and depends on
several individual subtasks. At least one of these tasks should
be addressed by contributions to this workshop. The topics of interest include
but are not limited to:

- Archiving linguistic data and resources

- Annotation technology, e.g., generic corpus annotation frameworks; the
  relationship of linguistic theories to corpus annotation; metadata
  annotation schemes, and related tools and applications

- Reusability of treebanks, e.g., annotations according to one specific
  linguistic framework should be applicable to NLP tasks that are based on
  different linguistic paradigms

- Sustainability in Software Engineering for Computational Linguistics

- Copyright issues, e.g., legal restrictions, copyright of web pages (for
  example, in a web as corpus approach), software patents, intellectual
  property, national and international issues etc.

- Privacy protection, e.g., automatic anonymisation of language data

- Sustainability, maintenance, and adaptability of NLP applications and tools,
  e.g., to new domains, to new linguistic resources, or even to new
  linguistic frameworks or theories

- Querying linguistic data, e.g., the usability and adaptability of query
  interfaces or query toolboxes

- Usability and acceptance of NLP software, e.g., corpus query interfaces



SUBMISSION INSTRUCTIONS

Submissions should not exceed ten (10) pages, including references. We
strongly recommend the use of the LaTeX style files or Microsoft Word
document template that will be made available on the LREC Conference

Web site. A description of the required format will be made available to

those who are unable to make direct use of these style files.

Submission will be electronic. The only accepted format for submitted papers is Adobe PDF. The papers must be submitted no later than 15th February 2008. Papers submitted after that time will not be reviewed. For details of the submission procedure, please consult the submission webpage reachable via the workshop website.

Questions regarding the submission procedure should be directed to andreas.witt(a)uni-tuebingen.de

IMPORTANT DATES

Deadline for submission of Papers  :  15th February 2008
Notification of Acceptance         :  18th March 2008
Deadline for final paper submission:   2nd April 2008

ORGANIZING COMMITTEE

Lou Burnard, Oxford University
Khalid Choukri, ELRA/ELDA
Georg Rehm, Tübingen University
Thomas Schmidt, University of Hamburg
Andreas Witt, Tübingen University

PROGRAM COMMITTEE

o Helen Aristar-Dry, Eastern Michigan University, USA
o Jeannine Beeken, Instituut voor Nederlandse Lexicologie, The Netherlands
o Jean Carletta, University of Edinburgh, School of Informatics, UK
o Dan Cristea, University of Iasi, Romania
o Stefanie Dipper, Bochum University, Germany
o Jost Gippert, Johann-Wolfgang-Goethe-Universität Frankfurt, Germany
o Erhard Hinrichs, Tübingen University, Germany
o Marc Kupietz, Institut für Deutsche Sprache Mannheim, Germany
o Sandra Kübler, Indiana University, Computational Linguistics, USA
o D. Terence Langendoen, NSF, USA
o Joakim Nivre, Växjö University & Uppsala University, Sweden
o Massimo Poesio, University of Trento, Italy
o Kiril Ribarov, Charles University Prague, Czech Republic
o Laurent Romary, Max-Planck Digital Library, Germany
o Hinrich Schuetze, Stuttgart University, Germany
o Serge Sharoff, University of Leeds, UK
o Gary F. Simons, SIL International, USA
o Manfred Stede, Potsdam University, Germany
o Simone Teufel, University of Cambridge, Computer Laboratory, UK
o Peter Wittenburg, MPI for Psycholinguistics, Nijmegen, The Netherlands
o Martin Wynne, Oxford Text Archive, UK
o Heike Zinsmeister, Heidelberg University, Germany