

FIRST CALL FOR PAPERS

Workshop

"Quality assurance and quality measurement for language and speech resources"

on Saturday, May 27th 2006, in conjunction with

LREC 2006, The 5th INTERNATIONAL CONFERENCE ON
LANGUAGE RESOURCES AND EVALUATION

Genoa, Italy, 24-26 May 2006

Workshop description:

=====

The workshop aims at

- bringing together experience with and insights in quality assurance and measurement for language and speech resources in a broad sense (including multimodal resources, annotations, tools, etc),
- covering both qualitative and quantitative aspects,
- identifying the main tools and strategies,
- analysing the strengths and weaknesses of current practice,
- establishing what can be seen as current best practice,
- reflecting on trends and future needs.

It can be seen as a follow-up of the workshop on speech resources that took place at LREC 2004, but the scope is wider as we include both language and speech resources. We feel that there is a lot to be gained by bringing these communities together, if only because the speech community seems to have a longer tradition in resources evaluation than the written language community.

Relevance:

=====

Quality assurance is an important concern for both the provider, the distributor and the user of language and speech resources. The concept of quality is only meaningful if both the producer and the user of the resources can rely on the same set of quality criteria, and if there are effective procedures to check whether these criteria are met. The universe of possible types of language resources is huge and evolves over time, and there is no universal set of qualitative or quantitative criteria and tests that can be applied to all sorts of resources.

In this workshop we will try to investigate what sorts of criteria, tests and measures

are being used by providers, users and distribution agencies such as ELRA and LDC, and we will try to distill from this current practice general recommendations for quality assurance and measurement for language and speech resources.

The workshop will look at quality assurance and quality measures both from the provider, the distributor and the user point of view, and will explicitly address special problems in connection with very large corpora, including numerical measures, comparison of corpora, exchange formats, etc.

Format:

=====

The workshop will be a full-day event, and will include

- (1) invited presentations (25+5 minutes) from data providers, distributors, or validators who are working on the basis of an explicit QA framework
- (2) submitted papers (15+5 minutes) by others who can report on relevant QA experience in the production, validation or use of resources
- (3) a round-table discussion aiming at establishing best practice

Papers:

=====

We invite papers that

- describe or critically analyze existing quality measures used to compare or validate resources
- describe or critically analyze existing quality assurance practices in resources production
- describe or critically analyze existing approaches to quality validation or measurement of third party resources
- describe future directions aimed at improving quality assurance, validation and measurement procedures for language and speech resources

Timetable:

=====

- Paper submission deadline: Feb 17, 2006
- Notification of acceptance: March 10, 2006
- Final version of paper: April 10, 2006
- Workshop: May 27, 2006 (full day)

Submissions:

=====

Abstracts should be in English, and up to 4 pages long. Submission format is PDF.

Papers will be reviewed by at least 3 members of the scientific committee. The

reviews are NOT anonymous.

Accepted papers are up to 6 pages long, and should be submitted in the format specified for the proceedings by the LREC organizers. The URL will be published on the Workshop Site (see below).

Submissions should be sent to Steven.Krauwer@let.uu.nl

Workshop and core scientific committee:

=====

Co-chairs:

- Steven Krauwer (UU/ELSNET, steven.krauwer@let.uu.nl)
- Uwe Quasthoff (Leipzig, quasthoff@informatik.uni-leipzig.de)

Members:

- Simo Goddijn (INL, goddijn@inl.nl)
- Jan Odijk (ELRA/Scansoft/UU, jan.odijk@scansoft.com)
- Khalid Choukri (ELDA, choukri@elda.org)
- Nicoletta Calzolari (ILC-CNR/WRITE, glottolo@ilc.cnr.it)
- Bente Maegaard (CST, bente@cst.dk)
- Chris Cieri (LDC, ccieri@ldc.upenn.edu)
- Chu-ren Huang (Ac Sin, churen@gate.sinica.edu.tw)
- Takenobu Tokunaga (TIT, take@cl.cs.titech.ac.jp)
- Harald Hoege (Siemens, harald.hoege@siemens.com)
- Henk van den Heuvel (CLST/SPEX, H.vandenHeuvel@let.ru.nl)
- Dafydd Gibbon (Bielefeld, gibbon@spectrum.uni-bielefeld.de)
- Key-Sun.Choi (KORTERM, Key-Sun.Choi@kaist.ac.kr)
- Jorg Asmussen, (DSL, ja@dsl.dk)

Scientific committee:

=====

We will include other experts as needed for the review process or for the completion of the programme.

Main contact and further info:

=====

- Contact: Steven Krauwer, steven.krauwer@let.uu.nl
- Workshop URL: <http://utrecht.elsnet.org/lrec2006qa>
- Conference URL: <http://www.lrec-conf.org/lrec2006>

This workshop is supported by ELSNET and WRITE (the international coordination committee for written language resources and evaluation).