

**Language Technology Service Platforms:
Synergies, Standards, Sharing**

31 May 2014

PROGRAMME

Editor:

Nicoletta Calzolari

Workshop Draft Programme

09:00 – 10:00 – Presentations of the initiatives involved in the organisation

10:00 – 10:30 – Session

Yohei Murakami et al., *Ontology for Language Service Interoperability*

Asuncion Gomez Perez, *Linked Data for sharing, discovery and re-use of Language Resources at a Web scale*

10:30 – 11:00 Coffee break

11:00 – 13:00 – Session

Nancy Ide et al., *LAPPS/ISO Exchange Vocabulary for web service interoperability*

Steve Cassidy, *Alveo: Combining Tools and Resources in National Infrastructure to support Human Communication Research*

Steven Krauwer, *Collaboration platforms: the CLARIN vision*

Eric Nyberg and Di Wang, *Open Advancement*

Hans Uszkoreit, *Can one single European language platform meet all our demands?*

Andrejs Vasiljevs, *MLi – towards European multilingual infrastructure for public and industry services*

13:00 – 14:30 Lunch break

14:30 – 16:00 – Session

Khalid Choukri, *Language Resources identification and citation: implications for a platform*

Nicoletta Calzolari, *Enabling sharing and replicability of research results: consequences for a platform*

Ed Hovy, *The Role of Language Technology and Standards in Modernizing Scholarly Publication: The FORCE11 Initiative*

Chris Cieri and Denise DiPersio, *Licensing for globally distributed services*

16:00 – 16:30 Coffee break

16:30 – 18:30 – *Concluding Discussion – Visions and strategies for the future of Language Infrastructures: towards Common Resolutions*

Discussants: All the Speakers ... and Workshop participants

Workshop Organisers

Nicoletta Calzolari - ILC-CNR, Italy and ELRA, France

Khalid Choukri - ELRA, France

Christopher Cieri - LDC, USA

Tomaž Erjavec - Jožef Stefan Institute, Slovenia

Nancy Ide - Vassar College, USA

Toru Ishida - Kyoto University, Japan

Oleksandr Kolomiyets - KU Leuven, Belgium

Joseph Mariani - LIMSI-CNRS and IMMI, France

Yohei Murakami - Kyoto University, Japan

Satoshi Nakamura - Nara Institute of Science and Technology, Japan

Senja Pollak - Jožef Stefan Institute, Slovenia

James Pustejovsky - Brandeis University, USA

Georg Rehm - DFKI GmbH, Germany

Herman Stehouwer - Max Planck, Germany

Hans Uszkoreit - DFKI GmbH, Germany

Andrejs Vasiljevs - Tilde, Latvia

Peter Wittenburg - Max Planck Institute for Psycholinguistics, The Netherlands

Introduction

Motivation and background

Increasingly, Human Language Technology (HLT) requires sophisticated infrastructures to support research, development, innovation, collaboration and deployment as services ready for production use. To address this need, several supporting infrastructures have been established over the past few years, and others are being built or planned.

The LREC 2014 Workshop on Language Technology Service Platforms brings together major infrastructural/coordination initiatives from all over the world. The overall goal is to explore means by which these and other infrastructure projects can best collaborate and interoperate in order to avoid duplication of effort, fragmentation, and above all, to facilitate the use and reuse of technologies and resources distributed throughout the world.

Focus

Web services are an increasingly common means to provide access to language technologies and resources. These services typically work in combination with repositories of language resources and workflow managers. This development brings with it its own set of issues in relation to collaboration and interoperability, including:

- interoperability of input to and output from language technologies deployed as web services;
- means to provide services for evaluation/replicability of results and iterative development;
- means to support multi-site collaborative work;
- licensing and cataloguing of language technologies and resources;
- sharing and access mechanisms to language technologies and resources;
- quality assessment and sustainability of language technologies and resources.

Aims

This workshop aims to foster discussion on these (and related) issues in order to arrive at a set of concrete plans for future collaboration and cooperation as well as immediate next steps.

General discussions will focus on the following questions: How can the various infrastructures collaborate, in both the near and long-term future? What are the steps needed in order to share both language technologies and resources? How can the projects and initiatives (including not only those involved in the workshop, but also others) join forces in order to eventually create a global infrastructure for Human Language Technologies?

The goal is to leave the workshop with a resolution that 1. lists all active infrastructure and platform initiatives, 2. describes the consensus of all initiatives involved in the workshop, 3. outlines the requirements for collaboration and 4. proposes solutions.

Researchers and technologists interested in platforms, services, sharing of language resources etc. are encouraged to participate in the workshop in order to make sure that their voice is heard. As described above, the consensus and outcome of the workshop will be put down in writing in a short resolution document meant to be used by the whole community for public relation and dissemination purposes, especially with regard to discussions with journalists, administrators, politicians and funding agencies.

Preliminary program plan

The first session will provide short introductions to the infrastructural/coordination initiatives involved in the organisation.

In order to outline some concrete next steps for the immediate future, there will be sessions devoted to surveying two to four currently implemented solutions to crucial problems, with an eye toward assessing and comparing the various solutions in order to determine immediate action items. These sessions will address topics such as:

- interoperability and the use of standards, for example, syntax and semantics used to exchange information between web services and/or technologies that may not have been developed at the same site (i.e., that do not necessarily utilize the same formats, categories, etc.)
- implemented means to provide evaluation/replicability and means to enable multi-site collaboration
- licensing for data and tools shared over networks and services.

Contribute to an overview of the Language Resources and Technologies landscape!

In order to facilitate the discussion we ask workshop participants to answer the following questions and to send their answers to the organisers (see Contact mail below) at the beginning of May. A summary of the responses will be provided at the workshop to inform and to focus the discussion.

- (1) *Access* – How do you make information about your tools and/or resources available to the world? How and where do you find information on tools and resources you would like to use?
- (2) *Obstacles to Data and Technology Exchange* – What do you see as the major obstacle(s) to the exchange of data between technologies?
- (3) *Data or Technology Gaps* – Are there tools, technologies or resources that do not exist at this time that are required to answer your research or development questions?
- (4) *Interoperability and Standards* – What syntax and semantics do you use to exchange information between web services and/or tools that may not have been developed at the same site (i.e., do not necessarily utilize the same formats, categories, etc.)?
- (5) *Evaluation* – What have you implemented to provide evaluation/replicability?
- (6) *Licensing* – How are you handling licensing for data shared over networks and services?
- (7) *Collaboration* – How would you propose to promote collaboration among the various infrastructure projects located around the world?

We also welcome any additional comments or views that you wish to express.

We look forward to welcoming you in Reykjavik!