INVITED SPEAKER:

Max Froumentin, W3C

WORKSHOP ORGANIZERS:

Nancy Ide, Vassar College
Laurent Romary, Loria/CNRS
Graham Wilcock, University of Helsinki

PROGRAM COMMITTEE:

Kalina Bontcheva, University of Sheffield, UK
Barrett Bryant, University of Alabama at Birmingham, USA
Paul Buitelaar, DFKI, Germany
Key-Sun Choi, KAIST, Korea
Hamish Cunningham, University of Sheffield, UK
Thierry Declerck, DFKI, Germany
David Durand, Brown University, USA
Tomaž Erjavec, Jožef Stefan Institute, Slovenia
Nancy Ide, Vassar College, USA
Ewan Klein, University of Edinburgh, UK
Jimmy Lin, MIT, USA
Chieko Nakabasami, Toyo University, Japan
Naoyuki Nomura, Justsystem/Hosei University, Japan
Antonio Pareja-Lora, UCM, Spain
Laurent Romary, Loria/CNRS, France
Manfred Stede, University of Potsdam, Germany
Henry Thompson, University of Edinburgh, UK
Fabio Vitali, University of Bologna, Italy
Graham Wilcock, University of Helsinki, Finland (Chair)

WORKSHOP WEBSITE:

http://www.ling.helsinki.fi/~gwilcock/NLPXML-2004/

FURTHER INFORMATION:

Graham Wilcock
University of Helsinki
00014 Helsinki, Finland
graham.wilcock@helsinki.fi

PREFACE

This is the fourth in the NLPXML series of workshops on natural language processing and XML. The first two workshops (NLPXML-2001 in Tokyo and NLPXML-2002 in Taipei) were primarily concerned with XML-based NLP tools and the use of XML in a wide range of NLP tasks. As XML rapidly became fully accepted within the NLP community, the theme of the third workshop (NLPXML-2003 in Budapest) shifted to the new challenges and opportunities of the Semantic Web. The present workshop continues this theme by focussing on RDF/RDFS and OWL.

While "plain XML" has become the established standard for data representation, especially for purposes of interchange and software interoperability, the use of RDF (Resource Definition Framework), RDFS (RDF Schema) and OWL (Web Ontology Language) for NLP applications remains relatively limited. However, the combination of XML, RDF/RDFS and OWL provides an overall architecture for NLP resources of all kinds whose implications are still being worked out within the NLP community.

NLPXML-2004 provides a forum for presentation and discussion of practical applications of RDF, RDFS and OWL in NLP. The workshop is intended not only for those already using Semantic Web technologies, but also members of the NLP community who seek a fuller understanding of the motivations and implications of these standards for the field.

The workshop program is designed to move progressively from XML, via metadata and RDF, to ontologies and OWL, and finally to their combined use for NLP applications. In this way, the workshop aims to clarify the respective roles of XML, RDF/RDFS and OWL in language technology, in relation to the growth of the Semantic Web.

We are delighted to welcome Max Froumentin of the W3C Multimodal Interaction working group as an invited speaker. His talk on Extensible MultiModal Annotation (EMMA) markup language will include reasons both for and against using RDF as part of EMMA.

We thank the members of the NLPXML-2004 program committee for their timely and informative reviews. We would also like to thank the ACL-2004 Workshops Chair, Srinivas Bangalore, and all the other organizers of ACL-2004 for their hard work and support.

Graham Wilcock
Nancy Ide
Laurent Romary
June 2004

Table of Contents

Techniques for Text Planning with XSLT
Mary Ellen Foster and Michael White 1

Towards Metadata Interoperability
Peter Wittenburg, Daan Broeder and Paul Buitelaar 9

A Web Application using RDF/RDFS for Metadata Navigation
Xi Guo, Mark Chaudhary, Christopher Dozier, Yogi Arumainayagam
and Venkatesan Subramanian 17

The Semantics of Markup: Mapping Legacy Markup Schemas to a Common Semantics
Gary Simons, William Lewis, Scott Farrar, Terence Langendoen, Brian Fitzsimons
and Hector Gonzalez 25

Invited talk: *Extensible MultiModal Annotation markup language (EMMA)*
Max Froumentin 33

Converting Mikrokosmos Frames into Description Logics
Pablo Beltrán-Ferruz, P González-Calero and P Gervás 35

Feeding OWL: Extracting and Representing the Content of Pathology Reports
David Schlangen, Manfred Stede and Elena Paslaru Bontas 43

An Extensible Framework for Efficient Document Management using RDF and OWL
Erica Meena, Ashwani Kumar and Laurent Romary 51

Towards Ontology-based Natural Language Processing
Dominique Estival, Chris Nowak and Andrew Zschorn 59

Technical Program Schedule

Sunday, July 25

09:15-09:30 Welcome and Introduction

09:30-10:00 *Techniques for Text Planning with XSLT*
Mary Ellen Foster and Michael White

10:00-10:30 **Break**

10:30-11:00 *Towards Metadata Interoperability*
Peter Wittenburg, Daan Broeder and Paul Buitelaar

11:00-11:30 *A Web Application using RDF/RDFS for Metadata Navigation*
Xi Guo, Mark Chaudhary, Christopher Dozier, Yogi Arumainayagam and Venkatesan Subramanian

11:30-12:00 *The Semantics of Markup: Mapping Legacy Markup Schemas to a Common Semantics*
Gary Simons, William Lewis, Scott Farrar, Terence Langendoen, Brian Fitzsimons and Hector Gonzalez

12:00-13:30 **Lunch**

13:30-14:10 **Invited talk**:
Extensible MultiModal Annotation markup language (EMMA)
Max Froumentin (W3C)

14:10-14:20 **Mini-break**

14:20-14:50 *Converting Mikrokosmos Frames into Description Logics*
Pablo Beltrán-Ferruz, P González-Calero and P Gervás

14:50-15:20 *Feeding OWL: Extracting and Representing the Content of Pathology Reports*
David Schlangen, Manfred Stede and Elena Paslaru Bontas

15:20-15:50 **Break**

15:50-16:20 *An Extensible Framework for Efficient Document Management using RDF and OWL*
Erica Meena, Ashwani Kumar and Laurent Romary

16:20-16:50 *Towards Ontology-based Natural Language Processing*
Dominique Estival, Chris Nowak and Andrew Zschorn

16:50-17:30 **Panel session**:
The Roles of RDF and OWL in Language Technology

THIS IS A BLANK PAGE PLEASE IGNORE

Author Index

Arumainayagam, Yogi 1
Beltrán-Ferruz, Pablo 3
Broeder, Daan 1
Buitelaar, Paul 1
Chaudhary, Mark 1
Dozier, Christopher 1
Estival, Dominique 3
Farrar, Scott 1
Fitzsimons, Brian 1
Foster, Mary Ellen 1
Froumentin, Max 1
Gervás, P 3
Gonzalez, Hector 1
González-Calero, P 3
Guo, Xi 1
Kumar, Ashwani 3
Langendoen, Terence 1
Lewis, William 1
Meena, Erica 3
Nowak, Chris 3
Paslaru Bontas, Elena 3
Romary, Laurent 3
Schlangen, David 3
Simons, Gary 1
Stede, Manfred 3
Subramanian, Venkatesan 1
White, Michael 1
Wittenburg, Peter 1
Zschorn, Andrew 3