EACL-2006

11th Conference
of the European Chapter of the
Association for Computational Linguistics

Proceedings of the 5th Workshop
on NLP and XML (NLPXML-2006):

Multi-Dimensional Markup
in Natural Language Processing

April 4, 2006
Trento, Italy

The conference, the workshop and the tutorials are sponsored by:

Center for the Evaluation of Language and Communication Technologies

Celct
c/o BIC, Via dei Solteri, 38
38100 Trento, Italy
`http://www.celct.it`

Xerox Research Centre Europe
6 Chemin de Maupertuis
38240 Meylan, France
`http://www.xrce.xerox.com`

CELI s.r.l.
Corso Moncalieri, 21
10131 Torino, Italy
`http://www.celi.it`

Thales
45 rue de Villiers
92526 Neuilly-sur-Seine Cedex, France
`http://www.thalesgroup.com`

EACL-2006 is supported by

Trentino S.p.a. and Metalsistem Group

Order copies of ACL proceedings from:
Priscilla Rasmussen,
Association for Computational Linguistics (ACL),
3 Landmark Center,
East Stroudsburg, PA 18301 USA

Phone `+1-570-476-8006`
Fax `+1-570-476-0860`
E-mail: `acl@aclweb.org`
On-line order form: `http://www.aclweb.org/`

INTRODUCTION

We are delighted to introduce the EACL-2006 workshop on Multi-Dimensional Markup in Natural Language Processing. This is the fifth in the NLPXML series of workshops on natural language processing and XML.

The first two NLPXML workshops (at NLPRS-2001 in Tokyo and at COLING-2002 in Taipei) were 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 and fourth workshops (at EACL-2003 in Budapest and at ACL-2004 in Barcelona) shifted to the new challenges and opportunities of the Semantic Web, with the focus on RDF and OWL rather than XML. The present workshop moves the focus firmly back to XML.

The special theme of this workshop is multi-dimensional markup. Our goal is to bring together researchers from several different fields—natural language processing, corpus linguistics, markup languages, and information retrieval—to discuss theoretical and practical issues related to the integration of different layers of text annotation. One particularly interesting challenge arises from the difficulty of combining annotations resulting from disparate NLP systems in a single hierarchical structure. For downstream applications that rely on a range of linguistic annotations, problems such as crossing boundaries and overlapping elements from different sources make it difficult to query data with multiple layers of annotation. We are particularly pleased to present papers that discuss the problems associated with such integration as well as those that provide solutions within the four fields mentioned.

An unusual and, we believe, a particularly attractive feature of the workshop program is the emphasis on live demonstrations of practical working systems. We have included two separate demo sessions in the program, with a total of 12 system demos. Short descriptions of the demos are included in these proceedings, in addition to the six full workshop papers.

We would like to thank the members of the NLPXML-2006 program committee for their prompt and expert reviews. In more than one case the reviewers' detailed and well-informed comments enabled significant improvements to the papers included in this volume. We also thank the organizers of EACL-2006 for their support.

David Ahn
Erik Tjong Kim Sang
Graham Wilcock
February 2006

WORKSHOP ORGANIZERS:

David Ahn, University of Amsterdam
Erik Tjong Kim Sang, University of Amsterdam
Graham Wilcock, University of Helsinki

PROGRAM COMMITTEE:

David Ahn, University of Amsterdam (co-chair)
Wouter Alink, NFI, The Hague
Paul Buitelaar, DFKI, Saarbruecken
Jean Carletta, University of Edinburgh
Hamish Cunningham, University of Sheffield
Tomaz Erjavec, Jozef Stefan Institute, Ljubljana
Claire Grover, University of Edinburgh
Nancy Ide, Vassar, New York
Amy Isard, University of Edinburgh
Mounia Lalmas, University of London
Maarten Marx, University of Amsterdam
Guenter Neumann, DFKI, Saarbruecken
Laurent Romary, Loria, Nancy
Valentin Tablan, University of Sheffield
Henry Thompson, University of Edinburgh
Erik Tjong Kim Sang, University of Amsterdam
Arjen de Vries, CWI, Amsterdam
Graham Wilcock, University of Helsinki (co-chair)

WORKSHOP WEBSITE:

`http://ilps.science.uva.nl/nlpxml2006/`

WORKSHOP PROGRAM

Tuesday, April 4

09:00-09:05	Welcome
09:05-09:30	*Representing and Querying Multi-dimensional Markup for Question Answering* Wouter Alink, Valentin Jijkoun, David Ahn, Maarten de Rijke, Peter Boncz, and Arjen de Vries
09:30-10:00	*Annotation and Disambiguation of Semantic Types in Biomedical Text: A Cascaded Approach to Named Entity Recognition* Dietrich Rebholz-Schuhmann, Harald Kirsch, Sylvain Gaudan, Miguel Arregui, and Goran Nenadic
10:00-10:30	*Tools to Address the Interdependence between Tokenisation and Standoff Annotation* Claire Grover, Michael Matthews, and Richard Tobin
10:30-11:00	Break
11:00-11:30	*Towards an Alternative Implementation of NXT's Query Language via XQuery* Neil Mayo, Jonathan Kilgour, and Jean Carletta
11:30-11:45	Demo boosters, 1
11:45-12:30	Demo session, 1
12:30-14:30	Lunch
14:30-15:00	*Multi-dimensional Annotation and Alignment in an English-German Translation Corpus* Silvia Hansen-Schirra, Stella Neumann, and Mihaela Vela
15:00-15:15	Demo boosters, 2
15:15-16:00	Demo session, 2
16:00-16:30	Break
16:30-17:00	*Querying XML documents with multi-dimensional markup* Peter Siniakov
17:00-18:00	Panel discussion and closing

Table of Contents

Full papers