

The Semantic Web and Language Technology, Its Potential and Practicalities: EUROLAN-2003

Dan Cristea, Nancy Ide, and Dan Tufis

■ EUROLAN, which has been held biennially since 1993, is one of the most significant European summer schools in the area of natural language processing. Each of the EUROLAN sessions has focused on an area of timely interest to researchers in the field; this year's EUROLAN involved students in tutorials and hands-on sessions concerned with semantic web technologies as applied to language processing, ontology creation and use, and consideration of the semantic web's potential and limitations.

The Sixth biennial EUROLAN International Summer School was held in Bucharest, Romania, from 28 July to 8 August. This year's school was organized by the Faculty of Computer Science at the A. I. Cuza University of Iasi, the Research Institute for Artificial Intelligence at the Romanian Academy in Bucharest, and the Department of Computer Science at Vassar College. It was the most successful in its 10-year history, with 119 registered participants from 23 countries. Hosted by the Romanian Academy, the most prestigious cultural and scientific institution in the country, the event was given significant attention in the media.

All six of the EUROLAN summer schools to date have been dedicated

to some aspect of language technology that has gained recent attention within the research community. The topic of this year's school was concerned with the impact and relevance of semantic web technology for language-processing applications. Half-day seminars, followed by afternoon hands-on tutorial sessions, were led by major researchers in the field, who not only showed how semantic web technology can be used for language processing but also considered its potential in terms of what it might—and might not—bring to us in the future.

Among the lecturers, opinions differed about what the potential and limitations might be of a true semantic web as envisaged by its designers and how great its impact will really be. Although it is not yet clear whether the current vision of the semantic web will indeed reach its expectations, there are more and more opinions that it represents a major technological step that will permanently alter the thinking about data accessibility and processing strategies. The semantic web community brings together several research communities that have previously worked independently of one another, including researchers in knowledge representation, ontology development, and knowledge exploitation; web developers (including de-

velopers of standards such as XML, RDF, and OWL); with those who directly need to exploit knowledge in various applications. Linguistic knowledge at its various levels, including morphological, syntactic, lexical, semantic, discourse level, and pragmatic knowledge, is fundamental to the mere use of the semantic web by providing natural and easy-to-use access to vast stores of information.

The early sessions of the school focused on the major underlying technologies of the semantic web, and in hands-on sessions, students gained practical experience using RDF, RDFS and OWL to represent various levels of linguistic knowledge. Later in the school, the focus turned to ontologies, which is where the true power of the semantic web lies. It has long been recognized that the ability to exploit ontological descriptions of language data to enable inferencing capabilities would drastically enhance semantic processing; the semantic web offers the possibility that such ontologies might in fact be readily available to use in language-processing applications. At the same time, the difficulties of arriving at universally acceptable stores of ontological information are equally well known, especially for a domain as large as that required for language processing. EUROLAN lecturers treated the topic of ontology development from a variety of perspectives, from the purely practical to the philosophical. In the end, with the difficulties recognized, it nevertheless seemed apparent that we need to start somewhere, somehow, even if what emerges is a variety of ontological stores from which to choose.

The EUROLAN summer school also included a workshop on ontologies and information extraction, a student workshop on applied natural language processing for the semantic web, and working meetings on topics of particular interest related to the school's theme.

The following scholars and researchers served as professors and trainers at the school: Thierry Declerck, with practical session taken over by Adrian Raschip (both of DFKI Saarbruecken, Germany); James



Hendler (University of Maryland at College Park); Jerry Hobbs (University of Southern California at Marina del Rey); Nancy Ide (Vassar College); Alessandro Lenci (University of Pisa, Italy); Robert Meersman, helped in the practical session by Peter Spyns and Jan De Bo (all of Vrije Universiteit Brussel, Belgium); Srini Narayanan (SRI International, Menlo Park, SUA); Sergei Nirenburg and Marjorie McShane (both of University of Maryland); Hans Uszkoreit and Wolfgang Wahlster (DFKI and University of Saarbruecken); and Christopher Brewster and Alexei Dingli (both of University of Sheffield, United Kingdom).

The dense scientific program and correspondingly lively social events in the evenings (a Bucharest sight-seeing tour and a full-day excursion to Peles, the former Royal Palace in Sinaia, and "Dracula's castle" in Bran Transylvania) combined to create an atmosphere of camaraderie and enthusiasm among students and professors alike. We gratefully acknowledge the sponsorship of the American Association for Artificial Intelligence, which has contributed to this very successful event.

Dan Cristea is professor of the Department of Artificial Intelligence and Optimization of the Faculty of Computer Science of the A. I. Cuza University, Iasi, Romania, and the dean of this faculty. He is also the director of the master program in Computational Linguistics in the A. I. Cuza University and a principal researcher at the Institute for Theoretical Computer Science of the Romanian Academy—the Iasi branch. Cristea's main research interests are in discourse structure and the relationship between structure and referentiality, discourse parsing, anaphora resolution, and wordnet. His e-mail address is dcristea@infoiasi.ro.



Nancy Ide is professor and chair of the Department of Computer Science at Vassar College. Her research in computational linguistics includes work in the areas of word-sense disambiguation and discourse analysis, where she has published numerous papers over the past 15 years. She has been actively involved in language re-

source development and representation since 1987, founded the Text Encoding Initiative, and is currently a project leader in the International Organization for Standardization subcommittee for language resources (ISO TC37 SC4). Ide is coeditor-in-chief of the journal *Computers and the Humanities* and coeditor of the series Text, Speech, and Language Technology (Kluwer). Her e-mail address is ide@cs.vassar.edu.



Dan Tufis, a member of the Romanian Academy, is the director of the Research Institute for Artificial Intelligence of the Romanian Academy and a professor in the Department of Artificial Intelligence and Optimiza-

tion of the Faculty of Computer Science, A. I. Cuza University, Iasi. Tufis is also the director of the Computational Linguistic Centre at the Faculty of Letters of the University of Bucharest. He is the current president of the Romanian Academy Commission for Language Technology. His domains of interests are AI and statistical natural language processing. Currently, he is involved in the development of the Romanian wordnet and ontology-based natural language-processing applications. His e-mail address is tufis@racai.ro.

**Authors
Please Note:**

**Your Formatted
AAAI-04
Proceedings Papers
Are Due
April 6, 2004.**

**Extensions Aren't
Granted,
So Please Plan
Accordingly!**