

2006



COLING • ACL

COLING • ACL 2006

2nd Workshop on Ontology Learning and Population:
Bridging the Gap between Text and Knowledge

Proceedings of the Workshop

Chairs:

Paul Buitelaar, Philipp Cimiano and Berenike Loos

22 July 2006
Sydney, Australia

Production and Manufacturing by
BPA Digital
11 Evans St
Burwood VIC 3125
AUSTRALIA

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 1-932432-77-9

Table of Contents

Preface	v
Organizers	vii
Workshop Program	ix
<i>Enriching a Formal Ontology with a Thesaurus: an Application in the Cultural Heritage Domain</i> Roberto Navigli and Paola Velardi	1
<i>Multilingual Ontology Acquisition from Multiple MRDs</i> Eric Nichols, Francis Bond, Takaaki Tanaka, Sanae Fujita and Dan Flickinger	10
<i>LEILA: Learning to Extract Information by Linguistic Analysis</i> Fabian M. Suchanek, Georgiana Ifrim and Gerhard Weikum	18
<i>Ontology Population from Textual Mentions: Task Definition and Benchmark</i> Bernardo Magnini, Emanuele Pianta, Octavian Popescu and Manuela Speranza	26
<i>Efficient Hierarchical Entity Classifier Using Conditional Random Fields</i> Koen Deschacht and Marie-Francine Moens	33
<i>Taxonomy Learning using Term Specificity and Similarity</i> Pum-Mo Ryu and Key-Sun Choi	41
<i>Towards Large-scale Non-taxonomic Relation Extraction: Estimating the Precision of Rote Extractors</i> Enrique Alfonseca, Maria Ruiz-Casado, Manabu Okumura and Pablo Castells	49
<i>A hybrid approach for extracting semantic relations from texts</i> Lucia Specia and Enrico Motta	57
Author Index	65

Preface

An ontology is an explicit and formal specification of a shared conceptualization of a domain of interest. Ontologies formalize the intensional aspects of a domain, whereas the extensional part is provided by a knowledge base that contains assertions about instances of concepts and relations as defined by the ontology. The process of defining and instantiating a knowledge base is referred to as knowledge mark-up or ontology population, whereas (semi-)automatic support in ontology development is usually referred to as ontology learning.

Ontologies have been broadly used in knowledge management applications, including Semantic Web applications and research. In recent years, ontologies have regained interest also within the NLP community, specifically in such applications as information extraction, text mining and question answering. However, as ontology development is a tedious and costly process there has been an equally growing interest in the automatic learning of ontologies. Much of this work has been focused on textual data as human language is a primary mode of knowledge transfer. In this way, textual data provide both a resource for the ontology learning process as well as an application medium for developed ontologies.

Automatic methods for text-based ontology learning and population have developed over recent years, but it is difficult to compare approaches and results. In the 1st Workshop on Ontology Learning and Population (at ECAI 2004) we addressed this issue through an emphasis on the evaluation aspects of the reported work. The proceedings of this second workshop on ontology learning and evaluation (OLP2) contain 8 high-quality peer-reviewed papers presenting novel approaches that address diverse topics within ontology learning, i.e. learning taxonomic and non-taxonomic relations, populating ontologies with named entities and instances of relations as well as lexical enrichment of ontologies. Thanks to the excellent work of the program committee we have been able to compile an interesting and high quality program.

The program is divided into three thematic parts: “Lexical Ontology Enrichment”, “Ontology Population and Ontology-based IE” and “Taxonomy and Relation Extraction”. The workshop will conclude with two invited talks by Dekang Lin and Johan Bos on the usefulness of ontology learning, leading to a hopefully vivid discussion. We hope you enjoy the workshop.

Paul Buitelaar, DFKI, Saarbrücken, Germany

Philipp Cimiano, AIFB Karlsruhe, Germany

Berenike Loos, European Media Laboratory, Heidelberg, Germany

Organizers

Chairs:

Paul Buitelaar, DFKI, Germany
Philipp Cimiano, AIFB, Univ. of Karlsruhe, Germany
Berenike Loos, European Media Lab, Germany

Program Committee:

Eneko Agirre, Basque Country University, Spain
Enrique Alfonseca, Universidad Autónoma de Madrid, Spain
Nathalie Aussenac-Gilles, IRIT- CNRS Toulouse, France
Timothy Baldwin, University of Melbourne, Australia
Roberto Basili, Università di Roma “Tor Vergata”, Italy
Johan Bos, Università di Roma “La Sapienza”, Italy
Christopher Brewster, University of Sheffield, UK
Massimiliano Ciaramita, LOA-ISTC, Italy
Nigel Collier, National Institute of Informatics, Japan
Ido Dagan, Bar Ilan University, Israel
Eric Gausier, XEROX XRCE, France
Asuncion Gomez-Perez, Universidad Politécnica de Madrid, Spain
Marko Grobelnik, Jožef Stefan Institute, Slovenia
Siegfried Handschuh, DERI Galway, Ireland
Andreas Hotho, University of Kassel, Germany
Eduard Hovy, USC, Information Sciences Institute, USA
Vipul Kashyap, Partners HealthCare System, USA
Bernardo Magnini, ITC-IRST, Italy
Diana Maynard, University of Sheffield, UK
Adeline Nazarenko, LIPN - Université Paris-Nord, France
Claire Nedellec, MIG, INRA, France
George Paliouras, NCSR “Demokritos”, Greece
Patrick Pantel, USC, Information Sciences Institute, USA
Robert Porzel, European Media Lab, Germany
Marie-Laure Reinberger, Universiteit Antwerpen, Belgium
Marta Sabou, Knowledge Media Institute, UK
Michael Sintek, DFKI, Germany
Peter Spyns, Vrije Universiteit Brussel, Belgium
Steffen Staab, University of Koblenz-Landau, Germany
Vojtech Svatek, University of Economics, Prague, Czech Rep.
Paola Velardi, Università di Roma “La Sapienza”, Italy
Dominic Widdows, MAYA Design, USA

Invited Speakers:

Johan Bos, Università di Roma “La Sapienza”, Italy
Dekang Lin, University of Alberta, Canada

Workshop Program

Saturday, 22 July 2006

9:00–9:30 Introduction to OLP and Overview of the Workshop

Session 1: Lexical Ontology Enrichment

9:30–10:00 *Enriching a Formal Ontology with a Thesaurus: an Application in the Cultural Heritage Domain*

Roberto Navigli and Paola Velardi

10:00–10:30 *Multilingual Ontology Acquisition from Multiple MRDs*

Eric Nichols, Francis Bond, Takaaki Tanaka, Sanae Fujita and Dan Flickinger

10:30–11:00 Coffee Break

Session 2: Ontology Population and Ontology-Based IE

11:00–11:30 *LEILA: Learning to Extract Information by Linguistic Analysis*

Fabian M. Suchanek, Georgiana Ifrim and Gerhard Weikum

11:30–12:00 *Ontology Population from Textual Mentions: Task Definition and Benchmark*

Bernardo Magnini, Emanuele Pianta, Octavian Popescu and Manuela Speranza

12:00–12:30 *Efficient Hierarchical Entity Classifier Using Conditional Random Fields*

Koen Deschacht and Marie-Francine Moens

12:30–14:00 Lunch Break

Session 3: Taxonomy and Relation Extraction

14:00–14:30 *Taxonomy Learning using Term Specificity and Similarity*

Pum-Mo Ryu and Key-Sun Choi

14:30–15:00 *Towards Large-scale Non-taxonomic Relation Extraction: Estimating the Precision of Rote Extractors*

Enrique Alfonseca, Maria Ruiz-Casado, Manabu Okumura and Pablo Castells

15:00–15:30 *A hybrid approach for extracting semantic relations from texts*

Lucia Specia and Enrico Motta

15:30–16:00 Coffee Break

16:00–17:30 Invited Talks and Panel Discussion by Johan Bos and Dekang Lin

17:30–18:00 Concluding Remarks

