Proceedings of the

Fourth Conference on Computational Natural Language Learning

and of the

Second Learning Language in Logic Workshop

Held in cooperation with ICGI-2000

13–14 September 2000
Lisbon, Portugal

Proceedings of the

Fourth Conference on Computational Natural Language Learning

and of the

Second Learning Language in Logic Workshop

Held in cooperation with ICGI-2000

13–14 September 2000
Lisbon, Portugal

Order additional copies from:

Association for Computational Linguistics
75 Paterson Street
New Brunswick, NJ 08901 USA
+1-732-342-9100 phone
+1-732-342-9339 fax
acl@aclweb.org

Preface

The joint *Second Learning Language in Logic (LLL-2000) Workshop* and *Fourth Conference on Computational Natural Language Learning (CoNLL-2000)* took place September 13-14, 2000, at the Instituto Superior Técnico in Lisbon, Portugal and have been co-organized with the 5th International Colloquium on Grammatical Inference (ICGI-2000).

This volume contains the papers presented during this joint event. More information is available on-line from `http://www.lri.fr/~cn/LLL-2000/` and `http://lcg-www.uia.ac.be/conll2000/`.

We would like to thank all the authors for submitting their papers and thus making these proceedings possible. We address special thanks to the members of the program committees for their great work which contributed to the high quality of these proceedings. We wish to extend our gratitude to the invited speakers for presenting us with their views on innovative results in Natural Language Processing and Machine Learning.

We are also grateful to the Local Chair Arlindo Oliveira, the members of the Organizing Committee, Ana Fred and Ana T. Freitas, and all other individuals who helped in the organization of this event.

Finally, we would like to thank the sponsors of LLL-2000 and CoNLL-2000 for their generous financial and moral support: the Network of Excellence in Inductive Logic Programming (ILPNet2), the Network of Excellence in Machine Learning (MLNet3), the Computational Linguistics in Flanders research community (CLIF), and SIGNLL (ACL's SIG on Natural Language Learning).

Claire Cardie
Walter Daelemans
Claire Nédellec
Erik Tjong Kim Sang

SPONSORS:

CLIF (Computational Linguistics in Flanders)
ILPNet2 (Network of Excellence in Inductive Logic Programming)
MLNet3 (Network of Excellence in Machine Learning)
SIGNLL (ACL's SIG for Natural Language Learning)

INVITED SPEAKERS:

Jörg-Uwe Kietz
Dan Roth

ORGANIZERS:

Claire Cardie (CoNLL)
Walter Daelemans (CoNLL)
Claire Nédellec (LLL)
Erik Tjong Kim Sang (CoNLL)

LOCAL ARRANGEMENTS CHAIR:

Arlindo Oliveira

CoNLL PROGRAM COMMITTEE:

Thorsten Brants	(Universität des Saarlandes)
James Cussens	(University of York)
Raymond Mooney	(University of Texas at Austin)
John Nerbonne	(University of Groningen)
Miles Osborne	(University of Edinburgh)
David Powers	(Flinders University)
Ronan Reilly	(University College Dublin)
Antal van den Bosch	(Tilburg University)

LLL PROGRAM COMMITTEE:

Pieter Adriaans	(Syllogic and University of Amsterdam, the Netherlands)
Roberto Basili	(University of Roma, Italy)
Gilles Bisson	(INRIA, Grenoble, France)
Henrik Boström	(University of Stockholm, Sweden)
Gosse Bouma	(University of Groningen, the Netherlands)
James Cussens	(University of York, United Kingdom)
Tomaz Erjavec	(Institute Jozef Stefan, Slovenia)
Daniel Kayser	(LIPN, Universit Paris-Nord, France)
Suresh Manandhar	(University of York, United Kingdom)
Guenter Neumann	(DFKI, Saarbrcken, Germany)
Steve Pulman	(University of Cambridge, United Kingdom)
Christer Samuelsson	(Xerox Research Center Europe, Grenoble, France)
Stefan Wrobel	(University of Magdeburg, Germany)

FURTHER INFORMATION:

CoNLL and SIGNLL
Walter Daelemans
CNTS Language Technology Group
University of Antwerp (UIA)
Universiteitsplein 1 (building A)
B-2610 Antwerpen, Belgium
e-mail: daelem@uia.ua.ac.be

LLL
Claire Nédellec
Laboratoire de Recherche en informatique (LRI)
UMR 8623 CNRS
Bat 490, Université Paris-Sud
F-91405 Orsay cedex, France
e-mail: cn@lri.fr

Table of Contents

CoNLL-2000 Short Papers

CoNLL-2000 Shared Task Papers

LLL-2000 Invited Paper

LLL-2000 Papers

Fourth Conference on

Computational Natural Language Learning

(CoNLL-2000)

Preface

CoNLL-2000 is the fourth in a series of meetings organized by SIGNLL, the ACL's SIG on Natural Language Learning. Previous meetings were organized in Madrid, Sydney, and Bergen, co-located with different, but always computational linguistics-oriented, events. We are pleased that this time we could combine efforts with the grammar induction and inductive logic programming for language processing communities.

It is the explicit wish of the SIGNLL board to have the CoNLL meeting address all aspects of computational natural language learning, including issues that are not regularly discussed at computational linguistics meetings, such as computational models of human language acquisition, computational models of the origins and evolution of language, biologically-inspired learning methods, etc.

We are thrilled by the quality and quantity of the submissions, which allowed us to set up an intense but rewarding program with one invited talk, 12 long talks, and joint paper sessions with LLL-2000 and ICGI-2000. On top of that, we introduced two innovations: there are 12 *bullet presentations*, short talks accompanied by a poster presentation, and a *shared task session* in which 11 authors report on how their machine learning method performed on our shared task — the identification of syntactic constituents in text (chunking). In this part of the proceedings, you will find 37 papers providing a useful record of all presentations.

You can find out more about SIGNLL and its activities at `http://www.aclweb.org/signll/`.

Claire Cardie
Walter Daelemans
Erik Tjong Kim Sang

Ithaca and Antwerp, 2000

Second Learning Language in Logic Workshop

(LLL-2000)

Preface

LLL-2000 is the follow-up of the first LLL workshop held in 1999 in Bled (Slovenia), and co-located with the International Conference on Machine Learning and the International Conference on Logic Programming. This year LLL was integrated with the Fourth Conference on Language Learning (CoNLL) and the Fifth International Colloquium on Grammatical Inference (ICGI) with which LLL shares strong common scientific interests in language learning. The registration to ICGI, CoNLL and LLL was a joint registration so that registrants could freely move between the three events.

As in the first edition, LLL has attracted pluridisciplinary submissions from the three research fields — Natural Language Processing (NLP), Machine Learning and Computational Logic, demonstrating the growing interest in NLP methods based on ILP or non-classic logics, and hybrid methods. Relational learning more and more appears as complementary to data analysis in many NLP domains. Relational learning and logic-based learning prove here again their capacity to learn complex structured linguistic resources and knowledge such as ontology and grammar from corpora and explicit background knowledge.

The scientific program of LLL-2000 consisted of one invited talk by Jörg-Uwe Kietz on the acquisition of ontology and seven paper presentations. Six of them are reported here and the paper by Christophe Costa Florencio, accepted for presentation by both LLL and ICGI, has been published in the ICGI proceedings. The joint sessions with ICGI and CoNLL included one invited talk by Dan Roth and paper and poster presentations.

Claire Nédellec

Orsay, 2000

Author Index