

Australasian Language Technology Association Workshop 2009

Proceedings of the Workshop



Editors:
Luiz Augusto Pizzato
Rolf Schwitter

3-4 December 2009
University of New South Wales
Sydney, Australia

Australasian Language Technology Association Workshop 2009
(ALTA 2009)

URL: <http://www.alta.asn.au/events/alta2009/index.html>

Sponsors:



HCSNet

Volume 7, 2009
ISSN: 1834-7037

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Workshop (ALTA) 2009, held as part of HCSNet (ARC Network in Human Communication Science) SummerFest at the University of New South Wales, Sydney, Australia on December 3-4, 2009. This is the seventh annual installment of the ALTA workshop in its most-recent incarnation, and the continuation of an annual workshop series that has existed in various forms Down Under since the early 1990s.

The goals of the workshop are:

- to bring together the growing Language Technology (LT) community in Australia and New Zealand and encourage interactions;
- to encourage interactions and collaboration within this community and with the wider international LT community;
- to foster interaction between academic and industrial researchers, to encourage dissemination of research results;
- to provide a forum for the discussion of new and ongoing research and projects;
- to provide an opportunity for the broader artificial intelligence community to become aware of local LT research; and, finally, to increase visibility of LT research in Australia, New Zealand and overseas.

This years ALTA Workshop includes full paper presentations as well as short paper presentations. Of the 27 papers submitted to ALTA, 18 papers were selected by the program committee for publication. One of these 18 papers was withdrawn by the authors after the review process. The remaining 17 papers appear in these proceedings; 10 papers as full papers (9 pages in length) and 7 papers as short papers (5 pages in length).

Each paper was independently peer-reviewed by at least two members of an international program committee, in accordance with the DEST requirements for E1 conference publications. The ALTA Workshop is classified as a Tier B conference in the Computing Research and Education Association of Australasia (CORE) classification system which will form the basis of ranking computer science conference publications for the ARC.

We would like to thank all the authors who submitted papers to ALTA; the members of the program committee for the time and effort they put into the review process; and to our international plenary speaker, Prof. David Traum, University of Southern California, Institute for Creative Technologies.

Our thanks also go to HCSNet for its invaluable financial and organisational support, and for bringing together a large number of research communities under one multidisciplinary roof. In particular, we would like to thank Kym Buckley, Chris Cassidy, and Ben Phelan for their perfect organization of this event.

Luiz Augusto Pizzato and Rolf Schwitter
Program Co-Chairs

ALTA 2009 Committees

Workshop Co-Chairs

- Luiz Pizzato (University of Sydney)
- Rolf Schwitter (Macquarie University)

Workshop Local Organiser

- HCSNet (The ARC Network in Human Communication Science)

Program Committee

- Timothy Baldwin (University of Melbourne)
- Steven Bird (University of Melbourne)
- Lawrence Cavedon (National ICT Australia, Melbourne)
- Eric Choi (National ICT Australia, Sydney)
- Peter Clark (Boeing)
- Nathalie Colineau (CSIRO)
- Nigel Collier (NII – National Institute of Informatics, Japan)
- Robert Dale (Macquarie University)
- Jean-Yves Delort (CMCRC)
- Mark Dras (Macquarie University)
- Dominique Estival (University of Sydney)
- Caroline Gasperin (USP/ICMC, Brazil)
- Ben Hachey (CMCRC)
- Michael Hess (University of Zuerich)
- Jerry R. Hobbs (USC/ISI)
- Achim Hoffmann (University of New South Wales)
- Matthew Honnibal (University of Sydney)
- Alistair Knott (University of Otago)
- Kazunori Komatani (Kyoto University)
- Andrew Lampert (CSIRO)
- David Martinez (University of Melbourne)
- Tara E. McIntosh (University of Sydney)
- Diego Moll Aliod (Macquarie University)
- Eamonn Newman (Dublin City University)
- Cecile Paris (CSIRO)
- David M. W. Powers (Flinders University)
- Adam Saulwick (DSTO)
- Harold Somers (Dublin City University)
- Jette Viethen (Macquarie University)
- Martin Volk (University of Zuerich)
- Menno van Zaanen (Tilburg University)
- Simon Zwartz (Macquarie University)

ALTA 2009 Program

Thursday, 3rd December 2009

9:00-10:00	HCSNet Plenary talk: Eckart Altenmuller
10:00-10:30	Coffee break
10:30-10:45	ALTA Opening remarks
<hr/>	
Session 1 - 10:45 - 12:30	
Full papers	
10:45-11:15	Ben Hachey <i>Evaluation of Generic Relation Identification</i>
11:15-11:45	Will Radford, Ben Hachey, James R Curran and Maria Milosavljevic <i>Tracking Information Flow in Financial Text</i>
Short papers	
11:45-12:00	Christopher Chua, Maria Milosavljevic and James R. Curran <i>A Sentiment Detection Engine for Internet Stock Message Boards</i>
12:00-12:15	Su Nam Kim, Timothy Baldwin and Min-Yen Kan <i>Extracting Domain-Specific Words - A Statistical Approach</i>
12:15-12:30	Jon Patrick and Min Li <i>A Cascade Approach to Extracting Medication Events</i>
<hr/>	
12:30-2:00	Lunch
2:00-3:00	HCSNet Plenary talk: Jonathan Harrington
3:00-3:30	Coffee break
<hr/>	
Session 2 - 3:30-4:30	
Full papers	
3:30-4:00	Nicky Ringland, Joel Nothman, Tara Murphy and James R Curran <i>Classifying articles in English and German Wikipedia</i>
Short papers	
4:00-4:15	Sam Tardif, James R Curran and Tara Murphy <i>Improved Text Categorisation for Wikipedia Named Entities</i>
4:15-4:30	Aleksandar Igetic, Catherine Watson, Jonathan Teutenberg, Elizabeth Broadbent, Rie Tamagawa and Bruce MacDonald <i>Towards a flexible platform for voice accent and expression selection on a Healthcare Robot</i>
<hr/>	
4:30-5:30	2009 Annual General Meeting of the Australasian Language Technology Association

Friday, 4th December 2009

9:00-10:00 HCSNet Plenary talk: David Traum

10:00-10:30 Coffee break

Session 3 - 10:30-12:30

Full papers

10:30-11:00 Simon Musgrave and Michael Haugh
The AusNC Project: Plans, Progress and Implications for Language Technology

11:00-11:30 James Breen and Timothy Baldwin
Corpus-based Extraction of Japanese Compound Verbs

11:30-12:00 Meladel Mistica, Wayan Arka, Timothy Baldwin and Avery Andrews
Double Double, Morphology and Trouble: Looking into Reduplication in Indonesian

12:00-12:30 Sze-Meng Jojo Wong and Mark Dras
Contrastive Analysis and Native Language Identification

12:30-2:00 Lunch

2:00-3:00 HCSNet Plenary talk: Mark Sanderson

3:00-3:30 Coffee break

Session 4 - 3:30-5:30

Full papers

3:30-4:00 Jonathan K. Kummerfeld, Jessika Roesner and James Curran
Faster parsing and supertagging model estimation

4:00-4:30 Tim Dawborn and James R. Curran
CCG parsing with one syntactic structure per n-gram

4:30-5:00 Colin White and Rolf Schwitter
An Update on PENG Light

Short papers

5:00-5:15 James Constable and James Curran
Integrating Verb-Particle Constructions into CCG Parsing

5:15-5:30 Bahadorreza Ofoghi and John Yearwood
From Lexical Entailment to Recognizing Textual Entailment Using Linguistic Resources

5:30 Close

Contents

HCSNet Plenary Speaker	1
<i>Spoken Dialogue Models for Virtual Humans</i>	
David Traum	1
Full papers	2
<i>Evaluation of Generic Relation Identification</i>	
Ben Hachey	2
<i>Tracking Information Flow in Financial Text</i>	
Will Radford, Ben Hachey, James R Curran and Maria Milosavljevic	11
<i>Classifying articles in English and German Wikipedia</i>	
Nicky Ringland, Joel Nothman, Tara Murphy and James R Curran	20
<i>The AusNC Project: Plans, Progress and Implications for Language Technology</i>	
Simon Musgrave and Michael Haugh	29
<i>Corpus-based Extraction of Japanese Compound Verbs</i>	
James Breen and Timothy Baldwin	35
<i>Double Double, Morphology and Trouble: Looking into Reduplication in Indonesian</i>	
Meladel Mistica, Wayan Arka, Timothy Baldwin and Avery Andrews	44
<i>Contrastive Analysis and Native Language Identification</i>	
Sze-Meng Jojo Wong and Mark Dras	53
<i>Faster parsing and supertagging model estimation</i>	
Jonathan K. Kummerfeld, Jessika Roesner and James Curran	62
<i>CCG parsing with one syntactic structure per n-gram</i>	
Tim Dawborn and James R. Curran	71
<i>An Update on PENG Light</i>	
Colin White and Rolf Schwitter	80
Short papers	89
<i>A Sentiment Detection Engine for Internet Stock Message Boards</i>	
Christopher Chua, Maria Milosavljevic and James R. Curran	89
<i>Extracting Domain-Specific Words - A Statistical Approach</i>	
Su Nam Kim, Timothy Baldwin and Min-Yen Kan	94
<i>A Cascade Approach to Extracting Medication Events</i>	
Jon Patrick and Min Li	99
<i>Improved Text Categorisation for Wikipedia Named Entities</i>	
Sam Tardif, James R Curran and Tara Murphy	104
<i>Towards a flexible platform for voice accent and expression selection on a Healthcare Robot</i>	
Aleksandar Igetic, Catherine Watson, Jonathan Teutenberg, Elizabeth Broadbent, Rie Tamagawa and Bruce MacDonald	109
<i>Integrating Verb-Particle Constructions into CCG Parsing</i>	
James Constable and James Curran	114
<i>From Lexical Entailment to Recognizing Textual Entailment Using Linguistic Resources</i>	
Bahadorreza Ofoghi and John Yearwood	119