

Australasian Language Technology Association Workshop 2011

Proceedings of the Workshop



Editors:

Diego Mollá

David Martinez

1-2 December 2011

Australian National University

Canberra, Australia

Australasian Language Technology Association Workshop 2011
(ALTA 2011)

URL: <http://www.alta.asn.au/events/alta2011>

Sponsors:



Australian
National
University



**Appen
ButlerHill**

Volume 9, 2011
ISSN: 1834-7037

ALTA 2011 Workshop Committees

Workshop Co-Chairs

- Diego Mollá (Macquarie University)
- David Martinez (NICTA Victoria Research Lab and University of Melbourne)

Workshop Local Organisers

- Hanna Suominen (NICTA Canberra Research Lab and ANU)
- Wray Buntine (NICTA Canberra Research Lab and ANU)

Program Committee

- Timothy Baldwin (University of Melbourne)
- Francis Bond (Nanyang Technological University, Singapore)
- Wray Buntine (NICTA Canberra Research Lab and ANU)
- Lawrence Cavedon (NICTA Victoria Research Lab and University of Melbourne)
- Eric Choi (NICTA)
- Jean-Yves Delort (Google, Zurich)
- Mark Dras (Macquarie University)
- Rebecca Dridan (University of Melbourne)
- Dominique Estival (University of Western Sydney)
- Tanja Gaustad (Tilburg University, The Netherlands)
- Ben Hachey (Macquarie University and CMCRC)
- Nitin Indurkya (University of New South Wales and eBay Research Labs)
- Alistair Knott (University of Otago, New Zealand)
- Kazunori Komatani (Nagoya University, Japan)
- Meladel Mistica (ANU)
- Scott Nowson (Appen Butler Hill)
- Cécile Paris (CSIRO)
- Son Bao Pham (Vietnam National University, Vietnam)
- Luiz Augusto Pizzato (University of Sydney)
- David Powers (Flinders University)
- Adam Saulwick (DSTO)
- Andrea Schalley (Griffith University)
- Rolf Schwitter (Macquarie University)
- Hanna Suominen (NICTA Canberra Research Lab and ANU)
- Menno Van Zaanen (Tilburg University, The Netherlands)
- Jette Viethen (Tilburg University, The Netherlands)
- Wayne Wobcke (University of New South Wales)
- Simon Zwarts (Google, Sydney)

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Workshop (ALTA) 2011, held at the Australian National University (ANU), Canberra, Australia on December 1-2, 2011. This is the ninth annual instalment 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 collaboration within the 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 year's ALTA Workshop includes 18 peer-reviewed papers, of which 12 have been presented orally, and 6 have been presented as posters. We received a total of 29 submissions. Each paper in the 'peer-reviewed papers' and the 'peer-reviewed posters' section 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 review process was double-blind: Great care was exercised to avoid all conflicts of interest whenever an author also served as program committee/co-chair or the reviewer worked at the same institution as an author. Such conflicts of interest were resolved by transferring the reviewing task to other members of the program committee.

The proceedings also include the abstracts of the keynote presentation by Wray Buntine (National ICT Australia Ltd – NICTA) and of the special presentation by Dominique Estival (MARCS Auditory Laboratories) about the Australian Computational Linguistics Olympiad (OzCLO). There is also description of the second ALTA shared task by Diego Mollá and Abeed Sarker (Macquarie University).

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, the local organisers for their commitment and work organising this conference, and our invited speaker: Wray Buntine.

Finally, we would like to thank our sponsors, NICTA, ANU and Appen Butler Hill for supporting the workshop.

Diego Mollá and David Martinez
Program Co-Chairs

ALTA 2011 Program

Thursday, 1st December 2011

9:00-10:00 Keynote: *Discovery in Text: Visualisation, Topics and Statistics* by Wray Buntine (NICTA, Canberra)

10:00-10:30 Coffee break

10:30-10:40 ALTA Opening remarks

Session 1 - 10:40 - 12:20

Paper Presentations

10:40-11:05 Benjamin Börschinger and Mark Johnson
A Particle Filter algorithm for Bayesian Wordsegmentation

11:05-11:30 Bevan Jones, Mark Johnson and Sharon Goldwater
Formalizing Semantic Parsing with Tree Transducers

11:30-11:55 Mark Johnson
Parsing in Parallel on Multiple Cores and GPUs

11:55-12:20 Mehdi Parviz, Mark Johnson, Blake Johnson and Jon Brock
Using Language Models and Latent Semantic Analysis to Characterise the N400m Neural Response

12:45-14:00 Lunch break

Session 2 - 14:00 - 15:15

Paper Presentations

14:00-14:25 Shunichi Ishihara
A Forensic Authorship Classification in SMS Messages: A Likelihood Ratio Based Approach Using N-gram

14:25-14:50 Su Nam Kim and Lawrence Cavendon
Classifying Domain-Specific Terms Using a Dictionary

14:50-15:15 Stephen Merity and James R. Curran
Frontier Pruning for Shift-Reduce CCG Parsing

15:15-15:45 Coffee break

Session 3 - 15:45 - 16:15

Poster presentations (5 mins per poster)

John Cocks and Te Taka Keegan
A word-based approach for diacritic restoration in Māori

Nobuagi Akagi and Francesco-Alessio Ursini
The Interpretation of Complement Anaphorae: the case of The Others

Francesco-Alessio Ursini and Nobuagi Akagi
The Interpretation of Plural Pronouns in Discourse: The Case of They

Jenny McDonald, Alistair Knott, Richard Zeng and Ayelet Cohen
Learning from student responses: A domain-independent natural language tutor

Md. Waliur Rahman Miah, John Yearwood and Sid Kulkarni
Detection of child exploiting chats from a mixed chat dataset as a text classification task

Marcin Nowina-Krowicki, Andrew Zschorn, Michael Pilling and Steven Wark
ENGAGE: Automated Gestures for Animated Characters

16:15-17:25 **Presenters put up their posters**

17:25-18:45 **Poster session with ALS (drinks and nibbles provided)**

19:30 **Dinner with ADCS**

Friday, 2nd December 2011

Session 4 - 9:00 - 10:40

Joint Session with ADCS

Li Wang, Diana Mccarthy and Timothy Baldwin

Predicting Thread Linking Structure by Lexical Chaining

Diego Mollá and Maria Elena Santiago-Martinez

Development of a Corpus for Evidence Based Medicine Summarisation

Mike Symonds, Peter Bruza, Laurianne Sitbon and Ian Turner

Tensor Query Expansion: A cognitively motivated relevance model

Yan Shen, Yuefeng Li, Yue Xu, Renato Lannella, Abdulmohsen Algarni and

Xiaohui Tao

An Ontology-based Mining Approach for User Search Intent Discovery

10:40-11:10 Coffee break

Session 5 - 11:10-12:25

Paper Presentations

11:10-11:35 **(best paper award)** François Lareau, Mark Dras, Benjamin Börschinger and Robert Dale

Collocations in multilingual text generation: Lexical Functions meet Lexical Functional Grammar

11:35-12:00 Abeed Sarker, Diego Mollá and Cécile Paris

Outcome Polarity Identification of Medical Papers

12:00-12:25 Sze-Meng Jojo Wong, Mark Dras and Mark Johnson

Topic Modeling for Native Language Identification

12:25-14:00 Lunch break

14:00-15:00 **Prizes and ALTA AGM Meeting**

Special presentation - 15:00-15:30

Dominique Estival

OzCLO: The Australian Computational Linguistic Olympiad

15:30-16:00 Coffee break

Special presentation - 16:00-16:45

Diego Mollá and Abeed Sarker

Automatic Grading of Evidence: The 2011 ALTA Shared Task

16:45-17:00 **Wrap-up**

17:25-18:45 **Poster session with ALS (drinks and nibbles provided)**

Contents

Non-reviewed papers	1
<i>Discovery in Text: Visualisation, Topics and Statistics</i> Wray Buntine	2
<i>OzCLO: The Australian Computational Linguistic Olympiad</i> Dominique Estival	3
<i>Automatic Grading of Evidence: the 2011 ALTA Shared Task</i> Diego Mollá and Abeed Sarker	4
Peer-reviewed papers: Oral presentations	9
<i>A Particle Filter algorithm for Bayesian Wordsegmentation</i> Benjamin Börschinger and Mark Johnson	10
<i>Formalizing Semantic Parsing with Tree Transducers</i> Bevan Jones, Mark Johnson and Sharon Goldwater	19
<i>Parsing in Parallel on Multiple Cores and GPUs</i> Mark Johnson	29
<i>Using Language Models and Latent Semantic Analysis to Characterise the N400m Neural Response</i> Mehdi Parviz, Mark Johnson, Blake Johnson and Jon Brock	38
<i>A Forensic Authorship Classification in SMS Messages: A Likelihood Ratio Based Approach Using N-gram</i> Shunichi Ishihara	47
<i>Classifying Domain-Specific Terms Using a Dictionary</i> Su Nam Kim and Lawrence Cavedon	57
<i>Frontier Pruning for Shift-Reduce CCG Parsing</i> Stephen Merity and James R. Curran	66
<i>Predicting Thread Linking Structure by Lexical Chaining</i> Li Wang, Diana Mccarthy and Timothy Baldwin	76
<i>Development of a Corpus for Evidence Based Medicine Summarisation</i> Diego Mollá and Maria Elena Santiago-Martinez	86
<i>Collocations in Multilingual Natural Language Generation: Lexical Functions meet Lexical Functional Grammar</i> François Lareau, Mark Dras, Benjamin Börschinger and Robert Dale	95
<i>Outcome Polarity Identification of Medical Papers</i> Abeed Sarker, Diego Mollá and Cécile Paris	105
<i>Topic Modeling for Native Language Identification</i> Sze-Meng Jojo Wong, Mark Dras and Mark Johnson	115

Peer-reviewed papers: Poster presentations	125
<i>A word-based approach for diacritic restoration in Māori</i> John Cocks and Te Taka Keegan	126
<i>The Interpretation of Complement Anaphorae: the case of The Others</i> Nobuagi Akagi and Francesco-Alessio Ursini	131
<i>The Interpretation of Plural Pronouns in Discourse: The Case of They</i> Francesco-Alessio Ursini and Nobuagi Akagi	140
<i>Learning from student responses: A domain-independent natural language tutor</i> Jenny Mcdonald, Alistair Knott, Richard Zeng and Ayelet Cohen	148
<i>Detection of child exploiting chats from a mixed chat dataset as a text classification task</i> Md. Waliur Rahman Miah, John Yearwood and Sid Kulkarni	157
<i>ENGAGE: Automated Gestures for Animated Characters</i> Marcin Nowina-Krowicki, Andrew Zschorn, Michael Pilling and Steven Wark	166