Australasian Language Technology Association Workshop 2012

Proceedings of the Workshop

Editors:
Paul Cook
Scott Nowson

4–6 December 2012
Otago University
Dunedin, New Zealand

Australalasian Language Technology Association Workshop 2012
(ALTA 2012)

http://www.alta.asn.au/events/alta2012

Sponsors:

NICTA

Microsoft® Research

Appen ButlerHill

UNIVERSITY of OTAGO

SAPERE AUDE

Te Whare Wānanga o Otāgo

NEW ZEALAND

Volume 10, 2012
ISSN: 1834-7037

ALTA 2012 Workshop Committees

Workshop Co-Chairs

- Paul Cook (The University of Melbourne)
- Scott Nowson (Appen Butler Hill)

Workshop Local Organiser

- Alistair Knott (University of Otago)

Programme Committee

- Timothy Baldwin (University of Melbourne)
- Lawrence Cavedon (NICTA and RMIT University)
- Nathalie Colineau (CSIRO - ICT Centre)
- Rebecca Dridan (University of Oslo)
- Alex Chengyu Fang (The City University of Hong Kong)
- Nitin Indurkhya (UNSW)
- Jong-Bok Kim (Kyung Hee University)
- Alistair Knott (University of Otago)
- Oi Yee Kwong (City University of Hong Kong)
- Francois Lareau (Macquarie University)
- Jey Han Lau (University of Melbourne)
- Fang Li (Shanghai Jiao Tong University)
- Haizhou Li (Institute for Infocomm Research)
- Marco Lui (University of Melbourne)
- Ruli Manurung (Universitas Indonesia)
- David Martinez (NICTA VRL)
- Tara McIntosh (Wavii)
- Meladel Mistica (The Australian National University)
- Diego Mollá (Macquarie University)
- Su Nam Kim (Monash University)
- Luiz Augusto Pizzato (University of Sydney)
- David Powers (Flinders University)
- Stijn De Saeger (National Institute of Information and Communications Technology)
- Andrea Schalley (Griffith University)
- Rolf Schwitter (Macquarie University)
- Tony Smith (Waikato University)
- Virach Sornlertlamvanich (National Electronics and Computer Technology Center)
- Hanna Suominen (NICTA)
- Karin Verspoor (National ICT Australia)

Preface

The precious volume you are currently reading contains the papers accepted for presentation at the Australasian Language Technology Association Workshop (ALTA) 2012, held at the University of Otago in Dunedin, New Zealand on 4–6 December 2012. We are excited that this tenth anniversary edition of the ALTA Workshop sees ALTA leaving Australia for the first time, and becoming a truly Australasian workshop. Sadly we say goodbye to the Aussie bush hat on the conference webpage, but it is in the spirit of Mick "Crocodile" Dundee that we cross the Tasman.

The goals of the workshop are to:

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

This year's ALTA Workshop presents 14 peer-reviewed papers, including eleven full and three short papers. We received a total of 18 submissions. Each paper, full and short, was reviewed by at least three members of the program committee. With the more-international flavour of the workshop, this year's program committee consisted of more members from outside of Australia and New Zealand than in past years. The reviewing for the workshop was double blind, and done in accordance with the DIISRTE requirements for E1 conference publications. Furthermore, great care was taken to avoid all conflicts of interest; in particular, no paper was assessed by a reviewer from the same institution as any of the authors. In the case of submissions by a programme co-chair, the double-blind review process was upheld, and acceptance decisions were made by the non-author co-chair.

In addition to peer-reviewed papers, the proceedings include the abstracts of the invited talks by Jen Hay (University of Canterbury) and Chris Brockett (Microsoft Research), both of whom we are honoured to welcome to ALTA. Also within, you will find an overview of the ALTA Shared Task and three system descriptions by shared task participants. These contributions were not peer-reviewed.

We would like to thank, in no particular order: all of the authors who submitted papers to ALTA; the fellowship of the program committee for the time and effort they put into maintaining the high standards of our reviewing process; our Man In Dunedin, the local organiser Alistair Knott for taking care of all the physical logistics and lining up some great social events; our invited speakers Jen Hay and Chris Brockett for agreeing to share their wisdom with us; the team from NICTA and James Curran for agreeing to host two fascinating tutorials, and; Diego Mollá and David Martinez, the program co-chairs of ALTA 2011, for their valuable help and support. We would like to acknowledge the constant support and advice of the out-going ALTA Executive Committee and in particular President Timothy Baldwin.

Finally, we gratefully recognise our sponsors: NICTA, Microsoft Research, Appen Butler Hill, and the University of Otago. Their generous support enabled us to offer travel subsidies to six students to attend and present at ALTA.

Paul Cook and Scott Nowson
Programme Co-Chairs

ALTA 2012 Programme

The proceedings are available online at `http://www.alta.asn.au/events/alta2012/proceedings/`

Tuesday 4 December 2012 Pre-workshop tutorials

Biomedical Natural Language Processing (Owheo 106)

A Crash Course in Statistical Natural Language Processing (Lab F)

Wednesday 5 December 2012

08:50–09:00 Opening remarks (Owheo 106)

09:00–10:00 Invited talk (Owheo 106; Chair: Paul Cook)
Jennifer Hay
Using a large annotated historical corpus to study word-specific effects in sound change

---

10:00–10:30 Coffee

Session 1 (Owheo 106; Chair: Ingrid Zukerman)

10:30–11:00 Angrosh M.A., Stephen Cranefield and Nigel Stanger
A Citation Centric Annotation Scheme for Scientific Articles

11:00–11:30 Michael Symonds, Guido Zuccon, Bevan Koopman, Peter Bruza and Anthony Nguyen
Semantic Judgement of Medical Concepts: Combining Syntagmatic and Paradigmatic Information with the Tensor Encoding Model

11:30–12:00 Teresa Lynn, Jennifer Foster, Mark Dras and Elaine Uí Dhonnchadha
Active Learning and the Irish Treebank

12:00–13:30 Lunch

Session 2 (Owheo 106; Chair: Diego Mollá)

13:30–14:00 Marco Lui, Timothy Baldwin and Diana McCarthy
Unsupervised Estimation of Word Usage Similarity

14:00–14:30 Mary Gardiner and Mark Dras
Valence Shifting: Is It A Valid Task?

14:30–15:00 **ALTA 2012 best paper** Minh Duc Cao and Ingrid Zukerman
Experimental Evaluation of a Lexicon- and Corpus-based Ensemble for Multi-way Sentiment Analysis

---

15:00–15:30 Coffee

Session 3 (Owheo 106; Chair: Chris Brockett)

15:30–16:00 James Breen, Timothy Baldwin and Francis Bond
Extraction and Translation of Japanese Multi-word Loanwords

16:00–16:30 Yvette Graham, Timothy Baldwin, Aaron Harwood, Alistair Moffat and Justin Zobel
Measurement of Progress in Machine Translation

16:30–17:30 ALTA business meeting (Owheo 106)

19:30– Conference dinner (Filadelfio's, 3 North Road)

Thursday 6 December 2012

09:00–10:00 Invited talk (Owheo 206; Chair: Timothy Baldwin)
Chris Brockett
Diverse Words, Shared Meanings: Statistical Machine Translation for Paraphrase, Grounding, and Intent

10:00–10:30 Coffee

Session 4: ALTA/ADCS shared session (Owheo 106; Chair: Alistair Knott)

10:30–11:00 **ADCS paper** Lida Ghahremanloo, James Thom and Liam Magee
An Ontology Derived from Heterogeneous Sustainability Indicator Set Documents

11:00–11:30 **ADCS paper** Bevan Koopman, Peter Bruza, Guido Zuccon, Michael John Lawley and Laurianne Sitbon
Graph-based Concept Weighting for Medical Information Retrieval

11:30–12:00 Abeed Sarker, Diego Mollá-Aliod and Cecile Paris
Towards Two-step Multi-document Summarisation for Evidence Based Medicine: A Quantitative Analysis

12:00–12:30 Alex G. Smith, Christopher X. S. Zee and Alexandra L. Uitdenbogerd
In Your Eyes: Identifying Clichés in Song Lyrics

12:30–14:00 Lunch

Session 5: ALTA Shared Task and poster boasters (Owheo 206; Chair: Karin Verspoor)

14:00–14:30 Iman Amini, David Martinez and Diego Molla
ALTA 2012 Shared Task overview

14:30–14:50 ALTA poster boasters

Paul Cook and Marco Lui
langid.py for better language modelling

Robert Fromont and Jennifer Hay
LaBB-CAT: an Annotation Store

Jenny Mcdonald, Alistair Knott and Richard Zeng
Free-text input vs menu selection: exploring the difference with a tutorial dialogue system.

Jared Willett, Timothy Baldwin, David Martinez and Angus Webb
Classification of Study Region in Environmental Science Abstracts

ALTA Shared Task poster boasters

Marco Lui
Feature Stacking for Sentence Classification in Evidence-based Medicine

Abeed Sarker
Multi-class classification of medical sentences using SVMs

14:50–15:00 Awards and final remarks (Owheo 206)

15:00–15:30 Coffee

15:30–17:00 Poster session with ADCS (Owheo 106)

19:30–21:30 Boat trip: Meet at 19:00 at the wharf, 20 Fryatt St.

Contents