Australasian Language Technology Association Workshop 2013

Proceedings of the Workshop

Editors:
Sarvnaz Karimi
Karin Verspoor

4–6 December 2013
Queensland University of Technology
Brisbane, Australia

Australasian Language Technology Association Workshop 2013
(ALTA 2013)

http://www.alta.asn.au/events/alta2013

Online Proceedings:
http://www.alta.asn.au/events/alta2013/proceedings/

Gold Sponsors:

As Australia's national science agency, CSIRO shapes the future using science to solve real issues. Our research makes a difference to industry, people and the planet. We're doing cutting-edge research in collaboration technologies, social media analysis tools and trust in online communities. Our people work closely with industry and communities to leave a lasting legacy.

NICTA is Australia's Information Communications Technology (ICT) Research Centre of Excellence and the nation's largest organisation dedicated to ICT research. NICTA's primary goal is to pursue high-impact research excellence and, through application of this research, to create national benefit and wealth for Australia.

Silver Sponsor:

Research happens across all of Google, and affects everything we do. Research at Google is unique. Because so much of what we do hasn't been done before, the lines between research and development are often very blurred. This hybrid approach allows our discoveries to affect the world, both through improving Google products and services, and through the broader advancement of scientific knowledge.

Volume 11, 2013 ISSN: 1834-7037

ALTA 2013 Workshop Committees

Workshop Co-Chairs

- Sarvnaz Karimi (CSIRO)
- Karin Verspoor (National ICT Australia)

Workshop Local Organiser

- Laurianne Sitbon (Queensland University of Technology)

Programme Committee

- Timothy Baldwin (University of Melbourne)
- Steven Bird (University of Melbourne)
- Wray Lindsay Buntine (NICTA)
- Lawrence Cavedon (NICTA and RMIT University)
- Nathalie Colineau (DSTO)
- Dominique Estival (University of Western Sydney)
- Graeme Hirst (University of Toronto)
- Nitin Indurkhya (UNSW)
- Su Nam Kim (Monash University)
- Francois Lareau (Macquarie University)
- Andrew MacKinlay (NICTA)
- David Martinez (NICTA)
- Meladel Mistica (The Australian National University)
- Diego Mollá (Macquarie University)
- Scott Nowson (Xerox Research Centre Europe)
- Cecile Paris (CSIRO)
- Luiz Augusto Pizzato (University of Sydney)
- Andrea Schalley (Griffith University)
- Rolf Schwitter (Macquarie University)
- Hanna Suominen (NICTA)
- Stephen Wan (CSIRO)
- Ingrid Zukerman (Monash University)

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Association Workshop (ALTA) 2013, held at the Queensland University of Technology in Brisbane, Australia on 4–6 December 2013.

We would like to declare this the ALTA Year of the Woman, in recognition of the first-time all-female organisation of the conference, including our local organiser in Brisbane, Laurianne Sitbon. We initially thought to have a female-only line-up for the keynote speakers, but settled with a population representative 50%. Please note that no gender biases were intentionally introduced into paper reviewing or acceptance; however, the authors on the accepted papers are fully one third female.

The goals of the workshop are to:

- bring together the Language Technology (LT) community in the Australasian region and encourage interactions and collaboration;
- 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;
- increase visibility of LT research in Australasia and overseas and encourage interactions with the wider international LT community.

This year's ALTA Workshop presents 16 peer-reviewed papers, including nine full, four short papers, and three papers that will be presented as posters. We received a total of 24 submissions. Each paper, apart from two submissions that were deemed outside of the conference scope at submission time, was reviewed by three members of the program committee. 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 involving a co-chair, the double-blind review process was upheld, and acceptance decisions were made by the non-author co-chair.

The proceedings include abstracts of the invited talks by Mark Steedman and Bonnie Webber, both from the University of Edinburgh. We are delighted to take advantage of their visit to Australia to bring them to Brisbane and are honoured to welcome them to ALTA. This volume also contains an overview of the ALTA Shared Task, its use as a class project, and a description of the winning system. 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 program committee for the time and effort they put into maintaining the high standards of our reviewing process; the local organiser Laurianne Sitbon for taking care of all the physical logistics and lining up some great social events; our invited speakers Mark Steedman and Bonnie Webber for agreeing to share their extensive experience and insights with us; the team from the NeCTAR Human Communication Science virtual laboratory (HCSvLab) and David Milne for agreeing to host two fascinating tutorials, and Paul Cook and Scott Nowson, the program co-chairs of ALTA 2012, for their valuable help and support. We would like to acknowledge the constant support and advice of the ALTA Executive Committee for providing input critical to the success of the workshop.

Finally, we gratefully recognise our sponsors: CSIRO, Google, and NICTA. Their generous support enabled us to fund student paper awards, as well as offer travel subsidies to three students to attend and present at ALTA. The University of Queensland also sponsored afternoon tea on Friday afternoon. We thank them as well.

Sarvnaz Karimi and Karin Verspoor
Programme Co-Chairs

ALTA 2013 Programme

ALTA will be held in P-block, the Queensland University of Technology Gardens Point campus.

Wednesday 4 December 2013 Pre-workshop tutorials (Room P-504)

10:00–14:30 (Lunch break 12-1) *Working with the HCS vLab*
Dominique Estival (University of Western Sydney) and Steve Cassidy (Macquarie University)

15:00–17:45 (Break 16:30-16:45) *Applying Wikipedia as a machine-readable knowledge base*
David Milne (CSIRO)

Thursday 5 December 2013

08:50–09:00	Opening remarks
09:00–10:00	Invited talk (Room P-512) Bonnie Webber *Concurrent Discourse Relations*
10:00–10:30	Coffee (Level 5)
Session 1 (Room P-512)	
10:30–11:00	Marco Lui and Paul Cook *Classifying English Documents by National Dialect*
11:00–11:30	Tim O'Keefe, Kellie Webster, James R. Curran and Irena Koprinska *Examining the Impact of Coreference Resolution on Quote Attribution*
11:30–12:00	Yvette Graham, Timothy Baldwin, Alistair Moffat and Justin Zobel *Crowd-Sourcing of Human Judgments of Machine Translation Fluency*
12:00–13:30	Lunch (Terrace, Level 6)
Session 2 (Room P-512)	
13:30–14:00	Shunichi Ishihara *The Effect of the Within-speaker Sample Size on the Performance of Likelihood Ratio Based Forensic Voice Comparison: Monte Carlo Simulations*
14:00–14:30	Hanna Suominen and Gabriela Ferraro *Noise in Speech-to-Text Voice: Analysis of Errors and Feasibility of Phonetic Similarity for Their Correction*
14:30–15:00	Robert Power, Bella Robinson and David Ratcliffe *Finding Fires with Twitter*
15:00–15:30	Coffee (Level 5)
Session 3 (Room P-512)	
15:30–16:00	Asif Ekbal, Sriparna Saha, Diego Molla and K Ravikumar *Multi-Objective Optimization for Clustering of Medical Publications*
16:00–16:20	Tatyana Shmanina, Ingrid Zukerman, Antonio Jimeno Yepes, Lawrence Cavedon and Karin Verspoor *Impact of Corpus Diversity and Complexity on NER Performance*
16:20–16:50	Rolf Schwitter *Working with Defaults in a Controlled Natural Language*
16:50–17:20	ALTA business meeting
19:00–	Conference dinner (Plough Inn, South Bank, Brisbane)

Friday 6 December 2013

09:00–10:00	Joint ADCS/ALTA Invited talk (Room P-421) Mark Steedman *Robust Computational Semantics*
10:00–10:30	Coffee (Level 5)
Session 4: ALTA/ADCS joint session (Room P-421)	
10:30–11:00	**ADCS paper** Takumi Sonoda and Takao Miura *Conditional Collocation in Japanese*
11:00–11:30	**ADCS paper** Hanna Suominen and Leif Hanlen *Visual Text Summarision for Surveillance and Situational Awareness in Hospitals*
11:30–12:00	Antti Puurula *Cumulative Progress in Language Models for Information Retrieval*
12:00–12:30	Oldooz Dianat, Cecile Paris and Stephen Wan *A Study: From Electronic Laboratory Notebooks to Generated Queries for Literature Recommendation*
12:30–13:30	Lunch (Terrace, Level 6)
Session 5: ALTA Shared Task (Room P-512)	
13:30–13:45	Diego Molla *ALTA 2013 Shared Task overview*
13:45–14:00	Marco Lui and Li Wang *Recovering Casing and Punctuation using Conditional Random Fields*
Session 6 and Poster Boasters (Room P-512)	
14:00–14:30	Shunichi Ishiharal *A Comparative Study of Likelihood Ratio Based Forensic Text Comparison in Procedures: Multivariate Kernel Density vs. Gaussian Mixture Model-Universal Background Model*
14:30–14:50	Farshid Zavareh, Ingrid Zukerman, Su Nam Kim and Thomas Kleinbauer *Error Detection in Automatic Speech Recognition*
14:50–15:05	Awards and final remarks
15:05–15:20	ALTA poster boasters
	Jared Willett, David Martinez, J. Angus Webb and Timothy Baldwin *Automatic Climate Classification of Environmental Science Literature*
	Jason Brown and Sam Mandal *Rhythm, Metrics, and the Link to Phonology*
	Shunichi Ishihara *Differences in Speaker Individualising Information between Case Particles and Fillers in Spoken Japanese*
15:20–17:00	Poster session with ADCS (Level 4)

Contents

Short papers **90**

Poster papers **111**

ALTA Shared Task papers **131**