Australasian Language Technology Association Workshop 2014

Proceedings of the Workshop

Editors:
Gabriela Ferraro
Stephen Wan

26 – 28th of November, 2014
RMIT
Melbourne, Australia

Australasian Language Technology Association Workshop 2014
(ALTA 2014)

http://www.alta.asn.au/events/alta2014

Online Proceedings:
http://www.alta.asn.au/events/alta2014/proceedings/

Volume 12, 2014 ISSN: 1834-7037

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.

SEEK is a diverse group of companies, comprised of a strong portfolio of online employment, educational, commercial and volunteer businesses. SEEK is listed on the Australian Securities Exchange, where it is a top 50 company with a market capitalization close to A$6 billion. With exposure to 2.5 billion people and over 20 per cent of global GDP, SEEK makes a positive contribution to peoples lives on a global scale. SEEK Australia currently receives over 26.6 million visits per month. The SEEK experience is seamless across desktop, mobile and iPad and currently over 54 per cent of all visits to seek.com.au are via mobile devices.

Silver Sponsors:

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.

Other Sponsors:

We would like to thank IBM Research for sponsoring the prize for this year's ALTA 2014 Shared Task.

ALTA 2014 Workshop Committees

Workshop Co-Chairs

- Gabriela Ferraro (National ICT Australia)
- Stephen Wan (CSIRO)

Workshop Local Organiser

- Lawrence Cavedon (RMIT)

Programme Committee

- Timothy Baldwin (University of Melbourne)
- Wray Buntine (Monash University)
- Alicia Burga (Universitat Pompeu Fabra)
- Lawrence Cavedon (RMIT University)
- Nathalie Colineau (DSTO)
- Trevor Cohn (University of Melbourne)
- Lan Du (Macquarie University)
- Dominique Estival (University of Western Sydney)
- Ben Hachey (University of Sydney)
- Gholamreza Haffari (Monash University)
- Graeme Hirst (University of Toronto)
- Nitin Indurkhya (University of New South Wales)
- Sarvnaz Karimi (CSIRO)
- Su Nam Kim (Monash University)
- Alistair Knott (University of Otago)
- François Lareau (Université de Montréal)
- David Martinez (University of Melbourne)
- Tara McIntosh (Google)
- Meladel Mistica (Intel Corporation)
- Diego Mollá (Macquarie University)
- Anthony Nguyen (Australian e-Health Research Centre, CSIRO)
- Joel Nothman (University of Sydney)
- Scott Nowson (Xerox Research Centre Europe)
- Cécile Paris (CSIRO)
- David Powers (Flinders University)
- Lizhen Qu (NICTA)
- Will Radford (Xerox Research Centre Europe)
- Horacio Saggion (Universitat Pompeu Fabra)
- Andrea Schalley (Griffith University)
- Rolf Schwitter (Macquarie University)
- Karin Verspoor (University of Melbourne)
- Ingrid Zukerman (Monash University)

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Association Workshop (ALTA) 2014, held at the RMIT University in Melbourne, Australia on 26–28th of November, 2014.

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, we are pleased to present 20 peer-reviewed papers selected for the ALTA Workshop, including 10 full papers, 6 short papers, and 4 papers that will be presented as posters. We received a total of 30 submissions. Each paper 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.

This volume covers a diverse of set topics as represented by the selected papers. This year, a number of papers describe applications of language technology, with domains ranging from biomedicine to emergency management. As in previous years, we aim to provide to foster the career development for research students by providing opportunities to receive feedback. Our hope is that both students and staff alike will enjoy the papers presented and that the workshop will continue to be a forum for our community to build new research relationships and collaborations.

The proceedings include the abstract of the invited talk by Dr. Jennifer Lai, from IBM Research. This volume also contains an overview of the 2014 ALTA Shared task and descriptions of the systems developed by three of the participating teams. 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 Lawrence Cavedon for taking care of all the physical logistics and lining up some great social events; our invited speakers Jennifer Lai and Maarten de Rijke for agreeing to share their extensive experience and insights with us; Trevor Cohn for agreeing to host a great tutorial, and Sarvnaz Karimi and Karin Verspoor, the program co-chairs of ALTA 2013, for their valuable help and guidance in preparing this volume. We would also 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, SEEK, Google, and IBM Research. Their generous support enabled us to fund student paper awards, travel subsidies to attend and present at ALTA, catering for the event, and to fund the prize for the ALTA 2014 Shared Task.

Gabriela Ferraro and Stephen Wan
ALTA Workshop Co-Chairs

ALTA 2014 Programme

Wednesday 26th of November 2014 Pre-workshop tutorials

13:30–17:00 (Break 15:00-15:30)	*Gaussian Processes for NLP* Trevor Cohn (University of Melbourne)

Thursday 27th of November, 2014

08:50–09:00	Opening remarks
09:00–10:00	Joint ALTA-ADCS Keynote Maarten de Rijke *Diversity, Intent, and Aggregated Search*
10:00–10:30	Coffee break
Session 1	
10:30–11:00	Sunghwan Kim, John Pate and Mark Johnson *The Effect of Dependency Representation Scheme on Syntactic Language Modelling*
11:00–11:30	Haoxing Wang and Laurianne Sitbon *Multilingual lexical resources to detect cognates in non-aligned texts*
11:30–12:00	Tudor Groza and Karin Verspoor *Automated Generation of Test Suites for Error Analysis of Concept Recognition Systems*
12:00–13:30	Lunch break
Session 2: Joint ALTA-ADCS Session	
13:30–14:00	Jie Yin, Sarvnaz Karimi and John Lingad (ADCS Paper) *Pinpointing Locational Focus in Tweets*
14:00–14:30	Johannes Schanda, Mark Sanderson and Paul Clough (ADCS Paper) *Examining New Event Detection*
14:30–15:00	Kristy Hughes, Joel Nothman and James R. Curran (ALTA Paper) *Trading accuracy for faster named entity linking*
15:00–15:30	Alexander Hogue, Joel Nothman and James R. Curran (ALTA Paper) *Unsupervised Biographical Event Extraction Using Wikipedia Traffic*
15:30–16:00	Coffee break
Session 3	
16:00–16:30	Su Nam Kim, Ingrid Zukerman, Thomas Kleinbauer and Masud Moshtaghi *A Comparative Study of Weighting Schemes for the Interpretation of Spoken Referring Expressions*
16:30–16:45	Mohammad Aliannejadi, Masoud Kiaeeha, Shahram Khadivi and Saeed Shiry Ghidary *Graph-Based Semi-Supervised Conditional Random Fields For Spoken Language Understanding Using Unaligned Data*
16:45–17:00	Dominique Estival and Steve Cassidy *Alveo, a Human Communication Science Virtual Laboratory*
17:00–17:30	Awards and ALTA business meeting
19:00–	Conference dinner

Friday 28th of November, 2014

09:00–10:00	ALTA Keynote Jennifer Lai *Deep QA: Moving beyond the hype to examine the challenges in creating a cognitive assistant for humans*
10:00–10:30	Coffee break
Session 4	
10:30–10:45	Jennifer Biggs and Michael Broughton *OCR and Automated Translation for the Navigation of non-English Handsets: A Feasibility Study with Arabic*
10:45–11:00	Simon Kocbek, Karin Verspoor and Wray Buntine *Exploring Temporal Patterns in Emergency Department Triage Notes with Topic Models*
11:00–11:30	Bella Robinson, Hua Bai, Robert Power and Xunguo Lin *Developing a Sina Weibo Incident Monitor for Disasters*
11:30–12:00	Michael Niemann *Finding expertise using online community dialogue and the Duality of Expertise*
Session 5: ALTA Shared Task	
12:00–12:450	Diego Mollá *ALTA 2014 Shared Task overview*
12:00–12:450	Shared Task Winner (TBA) *Presentation Title TBA*
12:30–13:30	Lunch break
Session 6	
14:00–14:30	Diego Mollá, Christopher Jones and Abeed Sarker *Impact of Citing Papers for Summarisation of Clinical Documents*
14:30–14:45	Tatyana Shmanina, Lawrence Cavedon and Ingrid Zukerman *Challenges in Information Extraction from Tables in Biomedical Research Publications: a Dataset Analysis*
14:45–15:00	Antonio Jimeno Yepes, Andrew MacKinlay, Justin Bedo, Rahil Garvani and Qiang Chen *Deep Belief Networks and Biomedical Text Categorisation*
14:30–14:45	Fumiyo Fukumoto, Shougo Ushiyama, Yoshimi Suzuki and Suguru Matsuyoshi *The Effect of Temporal-based Term Selection for Text Classification*
Session 7	
15:00–15:15	Final remarks
15:15–17:00	Poster session with ADCS

Contents