Australasian Language Technology Association Workshop 2010

Proceedings of the Workshop

Editors:
Nitin Indurkhya
Simon Zwarts

9-10 December 2010
University of Melbourne
Melbourne, Australia

Australasian Language Technology Association Workshop 2010 (ALTA 2010)
URL: http://www.alta.asn.au/events/alata2010

Sponsors:

NICTA

THE UNIVERSITY OF
MELBOURNE

Volume 8, 2010
ISSN: 1834-7037

ALTA 2010 Workshop Committees

Workshop Co-Chairs

- Nitin Indurkhya (University of New South Wales and eBay Research Labs)
- Simon Zwarts (Macquarie University)

Workshop Local Organisers

- David Martinez (University of Melbourne)
- Steven Bird (University of Melbourne)

Program Committee

- Achim Hoffmann (University of New South Wales)
- Adam Saulwick (DSTO)
- Alistair Knott (University of Otago, New Zealand)
- Andrea Schalley (Griffith University)
- Andrew Lampert (CSIRO)
- Ben Hachey (CMCRC)
- Caroline Gasperin (Universidade de Sao Paulo, Brasil)
- Cecile Paris (CSIRO)
- David M. W. Powers (Flinders University)
- David Martinez (University of Melbourne)
- Diego Molla Aliod (Macquarie University)
- Dominique Estival (University of Western Sydney)
- Dongqiang Yang (Flinders University)
- Eric Choi (NICTA)
- Francis Bond (Nanyang Technological University, Singapore)
- Jean-Yves Delort (CMCRC)
- Jette Viethen (Macquarie University)
- Kazunori Komatani (Nagoya University, Japan)
- Luiz Augusto Sangoi Pizzato (University of Sydney)
- Mark Dras (Macquarie University)
- Matthew Honnibal (University of Sydney)
- Menno van Zaanen (Tilburg University, The Netherlands)
- Nigel Collier (National Institute of Informatics, Japan)
- Rolf Schwitter (Macquarie University)
- Scott Nowson (Appen Pty Ltd)
- Son Bao Pham (Vietnam National University, Vietnam)
- Steven Bird (University of Melbourne)
- Tara McIntosh (NICTA)
- Timothy Baldwin (University of Melbourne)
- Wayne Wobcke (University of New South Wales)

Preface

This volume contains the papers accepted for presentation at the Australasian Language Technology Workshop (ALTA) 2010, held at the University of Melbourne, Melbourne, Australia on December 9-10, 2010. This is the eighth 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 interactions and 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 full papers (9 pages in length) as well as short papers (5 pages in length). We received a total of 19 submissions including 5 short papers. 12 papers including 2 short papers were selected by the program committee for publication in these proceedings.

Each paper in the 'reviewed papers' 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.

A key feature of this year's workshop is a special session of invited speakers from industry. In this 'industry session', which aims to bridge the gap between academia and industry, members of different companies talk about how they apply language technology (research) in their work. Another exciting feature of this year's workshop is the Language Technology Programming Competition. It is formatted as a "shared task": all participants compete to solve the same problem. The problem highlights an active area of research and programming in the area of language technology. Details of the shared task are published in the proceedings and also presented in a special session (along with the winner of the competition.)

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 invited speakers: Rodolfo Delmonte (University Ca'Foscari, Italy) and Casey Whitelaw (Google Inc, Sydney). We would also like to thank all the invited speakers for our industry session for making it such a success and hope that it will feature in future ALTA workshops as well.

Finally, we would like to thank our sponsors, NICTA and the University of Melbourne, for supporting the workshop.

Nitin Indurkhya and Simon Zwarts
Program Co-Chairs

ALTA 2010 Program

Thursday, 9th December 2010

9:00-10:00	Keynote: *Opinion Mining, Subjectivity and Factuality* by Rodolfo Delmonte (University Ca'Foscari, Italy)
10:00-10:30	Coffee break
10:30-10:40	ALTA Opening remarks
Session 1 - 10:40 - 12:45	
	Paper Presentations
10:40-11:05	Shunichi Ishihara *Variability and Consistency in the Idiosyncratic Selection of Fillers in Japanese Monologues: Gender Differences*
11:05-11:30	Michael Curtotti and Eric McCreath *Corpus Based Classification of Text in Australian Contracts*
11:30-11:55	Li Wang, Su Nam Kim and Timothy Baldwin *Thread-level Analysis over Technical User Forum Data*
11:55-12:20	Susan Howlett and Mark Dras *Dual-Path Phrase-Based Statistical Machine Translation*
12:20-12:45	Yue Li and David Martinez *Information Extraction of Multiple Categories from Pathology Reports*
12:45-2:00	Lunch
2:00-3:00	Keynote: *Language Technology: A View From The Trenches* by Casey Whitelaw (Google Inc, Sydney)
3:00-3:30	Coffee break
Session 2 - 3:30 - 5:30	
	Industry Session
	1. BinaryPlex (Tim Bull)
	2. NICTA (Wray Buntine)
	3. Pacific Brands (Yuval Marom)
	4. Lexxe (Hong Liang Qiao)
	5. Digital Sonata (Vadim Berman)
	6. CSIRO (Stephen Wan)
	7. Atex (Geoff Wilson)
	8. eBay (Nitin Indurkhya)
	9. Appen (Scott Nowson)
	10. Vicnet/State Library of Victoria (Andrew Cunningham)
	11. DSTO (Adam Saulwick)
5:30-	**Informal Q&A followed by Conference Dinner (jointly with ADCS) at 6pm**

Friday, 10th December 2010

Session 3 - 9:00 - 10:40	
	Joint Session with ADCS including ALTA presentations:
	Marco Lui and Timothy Baldwin
	Classifying User Forum Participants: Separating the Gurus from the Hacks, and Other Tales of the Internet
	Alexandra Uitdenbogerd
	Fun with Filtering French
10:40-11:10	Coffee break
11:10-12:00	LT shared talk report
12:00-1:00	**ALTA AGM Meeting** Free Pizza for attendees at the end!!
1:00-2:00	Lunch
Session 4 - 2:00-3:15	
	Paper Presentations
2:00-2:25	Diego Molla
	A Corpus for Evidence Based Medicine Summarisation
2:25-2:50	Sze-Meng Jojo Wong and Mark Dras
	Parser Features for Sentence Grammatical Classification
2:50-3:15	Jette Viethen and Robert Dale
	Speaker-Dependent Variation in Content Selection for Referring Expression Generation
3:15-3:45	Coffee break
Session 5 - 3:45-4:35	
	Paper Presentations
3:45-4:10	Dominick Ng, Matthew Honnibal and James R. Curran
	Reranking a wide-coverage CCG *parser*
4:10-4:35	Simon Zwarts, Mark Johnson and Robert Dale
	Repurposing Corpora for Speech Repair Detection: Two Experiments

Contents