Australasian Language Technology Association Workshop

(ALTA 2014)

Melbourne, Australia
26-28 November 2014

Editors:

Gabriela Ferraro
Stephen Wan

ISBN: 978-1-63439-636-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2014) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Contents

Invited talk **1**

Deep QA: Moving beyond the hype to examine the challenges in creating a cognitive assistant for humans
Jennifer Lai **2**

Full papers **3**

The Effect of Dependency Representation Scheme on Syntactic Language Modelling
Sunghwan Kim, John Pate and Mark Johnson **4**

Multilingual lexical resources to detect cognates in non-aligned texts
Haoxing Wang and Laurianne Sitbon **14**

Automated Generation of Test Suites for Error Analysis of Concept Recognition Systems
Tudor Groza and Karin Verspoor **23**

Trading accuracy for faster named entity linking
Kristy Hughes, Joel Nothman and James R. Curran **32**

Unsupervised Biographical Event Extraction Using Wikipedia Traffic
Alexander Hogue, Joel Nothman and James R. Curran **41**

A Comparative Study of Weighting Schemes for the Interpretation of Spoken Referring Expressions
Su Nam Kim, Ingrid Zukerman, Thomas Kleinbauer and Masud Moshtaghi **50**

Developing a Sina Weibo Incident Monitor for Disasters
Bella Robinson, Hua Bai, Robert Power and Xunguo Lin **59**

Finding expertise using online community dialogue and the Duality of Expertise
Michael Niemann **69**

Impact of Citing Papers for Summarisation of Clinical Documents
Diego Molla, Christopher Jones and Abeed Sarker **79**

The Effect of Temporal-based Term Selection for Text Classification
Fumiyo Fukumoto, Shougo Ushiyama, Yoshimi Suzuki and Suguru Matsuyoshi **88**

Short papers **97**

Graph-Based Semi-Supervised Conditional Random Fields For Spoken Language Understanding Using Unaligned Data
Mohammad Aliannejadi, Masoud Kiaeeha, Shahram Khadivi and Saeed Shiry Ghidary **98**

Alveo, a Human Communication Science Virtual Laboratory
Dominique Estival and Steve Cassidy **104**

OCR and Automated Translation for the Navigation of non-English Handsets: A Feasibility Study with Arabic
Jennifer Biggs and Michael Broughton **108**

Exploring Temporal Patterns in Emergency Department Triage Notes with Topic Models
Simon Kocbek, Karin Verspoor and Wray Buntine **113**

Challenges in Information Extraction from Tables in Biomedical Research Publications: a Dataset Analysis
Tatyana Shmanina, Lawrence Cavedon and Ingrid Zukerman **118**

Deep Belief Networks and Biomedical Text Categorisation
Antonio Jimeno Yepes, Andrew MacKinlay, Justin Bedo, Rahil Garvani and Qiang Chen **123**

Poster papers **128**

Sinhala-Tamil Machine Translation: Towards better Translation Quality
Randil Pushpananda, Ruvan Weerasinghe and Mahesan Niranjan **129**

Analysis of Coreference Relations in the Biomedical Literature
Miji Choi, Karin Verspoor and Justin Zobel **134**

Finnish Native Language Identification
Shervin Malmasi and Mark Dras **139**

A Data-driven Approach to Studying Given Names and their Gender and Ethnicity Associations
Shervin Malmasi **145**

ALTA Shared Task papers **150**

Overview of the 2014 ALTA Shared Task: Identifying Expressions of Locations in Tweets
Diego Mollá and Sarvnaz Karimi **151**

Identifying Twitter Location Mentions
Bo Han, Antonio Jimeno Yepes, Andrew MacKinlay and Qiang Chen **157**

A Multi-Strategy Approach for Location Mining in Tweets: AUT NLP Group Entry for ALTA-2014 Shared Task
Parma Nand, Rivindu Perera, Anju Sreekumar and He Lingmin **163**

Automatic Identification of Expressions of Locations in Tweet Messages using Conditional Random Fields
Fei Liu, Afshin Rahimi, Bahar Salehi, Miji Choi, Ping Tan and Long Duong **171**