

NAACL HLT 2009

**Software Engineering,
Testing, and Quality
Assurance for
Natural Language
Processing
(SETQA-NLP 2009)**

Proceedings of the Workshop

June 5, 2009
Boulder, Colorado

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2009 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-932432-32-9

Introduction

Software engineering is a first-class research topic in computer science, but generally has not been treated as such within the natural language processing community. However, the need for well-engineered NLP components is increasing as NLP begins to show up outside our research community: bioinformatics, the search industry, education applications, etc. In addition, NLP research itself, e.g., when it involves large data sets, often requires a high level of software quality. Simply applying standard software engineering practices to NLP often fails due to the unique characteristics of natural language as an input type.

The goals of this workshop include raising awareness of the need for good software engineering practices in NLP, stimulating research on same, and providing a forum for sharing current work in this area. We are grateful to the authors for sharing their work, our invited speaker, and to the program committee for their efforts.

Kevin Bretonnel Cohen and Marc Light

Organizers:

Kevin Bretonnel Cohen, Center for Computational Pharmacology, University of Colorado School of Medicine and The MITRE Corporation

Marc Light, Thomson Reuters

Program Committee:

William A. Baumgartner Jr., University of Colorado School of Medicine

Shannon Bradshaw, Drew University

Bob Carpenter, Alias-i

Hamish Cunningham, University of Sheffield

Dan Flickinger, Stanford University

Michael Gamon, Microsoft

Tracy Holloway King, Microsoft/PowerSet

James Lyle, Microsoft

Stephan Oepen, Stanford University

Jeff Reynar, DBT Labs

Kevin Markey, Silver Creek Systems

Charles Schafer, Google

Jun'ichi Tsujii, University of Tokyo and UK National Centre for Text Mining

Martin Volk, University of Stockholm

Scott Waterman, Microsoft/PowerSet

Ken Williams, Thomson Reuters

Invited Speaker:

Ted Pedersen, University of Minnesota Duluth

Table of Contents

<i>Building Test Suites for UIMA Components</i>	
Philip Ogren and Steven Bethard	1
<i>Context-Dependent Regression Testing for Natural Language Processing</i>	
Elaine Farrow and Myroslava O. Dzikovska	5
<i>Using Paraphrases of Deep Semantic Representations to Support Regression Testing in Spoken Dialogue Systems</i>	
Beth Ann Hockey and Manny Rayner	14
<i>Integrated NLP Evaluation System for Pluggable Evaluation Metrics with Extensive Interoperable Toolkit</i>	
Yoshinobu Kano, Luke McCrohon, Sophia Ananiadou and Jun'ichi Tsujii	22
<i>Tightly Packed Tries: How to Fit Large Models into Memory, and Make them Load Fast, Too</i>	
Ulrich Germann, Eric Joanis and Samuel Larkin	31
<i>Scaling up a NLU system from text to dialogue understanding</i>	
Rodolfo Delmonte, Antonella Bristot, Gloria Voltolina and Vincenzo Pallotta	40
<i>Towards Agile and Test-Driven Development in NLP Applications</i>	
Jana Sukkarieh and Jyoti Kamal	42
<i>Grammar Engineering for CCG using Ant and XSLT</i>	
Scott Martin, Rajakrishnan Rajkumar and Michael White	45
<i>Web Service Integration for Next Generation Localisation</i>	
David Lewis, Stephen Curran, Kevin Feeney, Zohar Etzioni, John Keeney, Andy Way and Reinhard Schäler	47
<i>Distributed Parse Mining</i>	
Scott Waterman	56
<i>Modular resource development and diagnostic evaluation framework for fast NLP system improvement</i>	
Gaël de Chalendar and Damien Nouvel	65
<i>Integrating High Precision Rules with Statistical Sequence Classifiers for Accuracy and Speed</i>	
Wenhui Liao, Marc Light and Sriharsha Veeramachaneni	74

Conference Program

Friday, June 5, 2009

- 9:00–9:30 *Building Test Suites for UIMA Components*
Philip Ogren and Steven Bethard
- 9:30–10:00 *Context-Dependent Regression Testing for Natural Language Processing*
Elaine Farrow and Myroslava O. Dzikovska
- 10:00–10:30 *Using Paraphrases of Deep Semantic Representions to Support Regression Testing
in Spoken Dialogue Systems*
Beth Ann Hockey and Manny Rayner
- 10:30–11:00 Morning Break
- 11:00–11:30 *Integrated NLP Evaluation System for Pluggable Evaluation Metrics with Extensive
Interoperable Toolkit*
Yoshinobu Kano, Luke McCrohon, Sophia Ananiadou and Jun'ichi Tsujii
- 11:30–12:00 *Tightly Packed Tries: How to Fit Large Models into Memory, and Make them Load
Fast, Too*
Ulrich Germann, Eric Joanis and Samuel Larkin
- 12:00–12:30 Poster Session
- Scaling up a NLU system from text to dialogue understanding*
Rodolfo Delmonte, Antonella Bristot, Gloria Voltolina and Vincenzo Pallotta
- Towards Agile and Test-Driven Development in NLP Applications*
Jana Sukkarieh and Jyoti Kamal
- Grammar Engineering for CCG using Ant and XSLT*
Scott Martin, Rajakrishnan Rajkumar and Michael White
- 12:30–2:00 Lunch Break
- 2:00–3:00 Invited Talk by Ted Pedersen: The road from good software engineering to good
science ... is a two way street
- 3:00–3:30 *Web Service Integration for Next Generation Localisation*
David Lewis, Stephen Curran, Kevin Feeny, Zohar Etzioni, John Keeney, Andy
Way and Reinhard Schäler

Friday, June 5, 2009 (continued)

3:30–4:00 Afternoon Break

4:00–4:30 *Distributed Parse Mining*
Scott Waterman

4:30–5:00 *Modular resource development and diagnostic evaluation framework for fast NLP system improvement*
Gaël de Chalendar and Damien Nouvel

5:00–5:30 *Integrating High Precision Rules with Statistical Sequence Classifiers for Accuracy and Speed*
Wenhui Liao, Marc Light and Sriharsha Veeramachaneni