ACL-IJCNLP 2009

UCNLG+Sum 2009

2009 Workshop on Language Generation and Summarisation

Proceedings of the Workshop

6 August 2009
Suntec, Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

©2009 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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 978-1-932432-51-0 / 1-932432-51-5

Preface

The Workshop on Language Generation and Summarisation (UCNLG+MT) took place in Singapore on 6th August 2009, as part of ACL-IJCNLP'09. It was the third of the UCNLG workshops which have the general aims

1. to provide a forum for reporting and discussing corpus-oriented methods for generating language;
2. to foster cross-fertilisation between NLG and related fields by looking for common ground through corpus-oriented approaches; and
3. to promote the sharing of data and methods in all language generation research.

Each of these workshops has a special theme: at the first workshop (at Corpus Linguistics in 2005) it was the use of corpora in NLG, at the second (at MT Summit in 2007) it was Language Generation and Machine Translation. The special theme of this third UCNLG workshop was Language Generation and Text Summarisation. The core aim was to provide a forum for NLG and summarisation researchers to examine the similarities and differences between their current approaches to generating language, and to explore the potential for cross-fertilisation and the extent to which resources and techniques can be shared between the NLG and summarisation fields.

The call for papers issued at the end of January 2009 elicited a good number of high-quality submissions, each of which was peer-reviewed by three members of the programme committee. The interest in the workshop from leading researchers in both fields and the quality of submissions was high, so we aimed to be as inclusive as possible within the practical constraints of the workshop. In the end we accepted six submissions as long papers and four as short papers.

The resulting workshop programme packed a lot of exciting content into one day. We were delighted to start the workshop with a keynote presentation from Prof Kathy McKeown, one of the most eminent researchers in NLG and its application to summarisation. Our technical programme included papers on structuring abstracts (Saggion), selecting content for summaries (Cheung, Carenini and Ng), sentence compression (Xu and Grishman; Cordiero, Dias and Brazdil), sentence revision for summarisation (Tanaka et al.), evaluating summaries (Owczarzak and Dang), corpus-based generation of directions (Schuldes et al.), reducing redundancy in summarisation (Hendrickx et al.), generation of narrative content (Caropreso et al.) and summarisation of non-textual content (Kumar et al.). The programme also included a session reporting the results of the Generation of Referring Expressions in Context (GREC) shared task evaluations (part of the Generation Challenges 2009 initiative), and still had space for a general discussion on synergies between NLG and summarisation.

We would like to thank all the people who have contributed to the organisation and delivery of this workshop: the authors who submitted such high quality papers; the programme committee for their prompt and effective reviewing; our keynote speaker, Kathy McKeown; our panelists (at the time of writing), Ed Hovy, Kathy McKeown and Donia Scott; the ACL-IJCNLP 2009 Organising Committee, especially the workshop chairs, Jimmy Lin and Yuji Matsumoto, and Jing-Shin Chang; all the particpants in the workshop and future readers of these proceedings for your shared interest in this exciting area of research.

August 2009 — Anja Belz, Roger Evans and Sebastian Varges

Workshop Organisers:
Anja Belz, University of Brighton, UK
Roger Evans, University of Brighton, UK
Sebastian Varges, University of Trento, Italy

Programme Committee:
Enrique Alfonseca, Google Zurich, Switzerland
Srinivas Bangalore, AT&T, USA
Robert Dale, Macquarie University, Australia
Daniel Marcu, ISI, University of Southern California, USA
Chris Mellish, Universiy of Aberdeen, UK
Ani Nenkova, University of Pennsylvania, USA
Amanda Stent, SUNY, USA
Michael Strube, EML Research, Germany
Stephen Wan, Macquarie University, Australia
Mike White, Ohio State University, USA
Jianguo Xiao, Peking University, China

Invited Speaker:
Kathy McKeown, Columbia University, USA

Panelists:
Ed Hovy, ISI, University of Southern California, USA
Kathy McKeown, Columbia University, USA
Donia Scott, Open University, UK
(at time of writing)

Table of Contents

Invited Paper

Query-focused Summarization Using Text-to-Text Generation: When Information Comes from Multilingual Sources
Kathy McKeown 3

Long Papers

Optimization-based Content Selection for Opinion Summarization
Jackie Chi Kit Cheung, Giuseppe Carenini and Raymond T. Ng 7

Unsupervised Induction of Sentence Compression Rules
Joao Cordeiro, Gael Dias and Pavel Brazdil 15

Evaluation of Automatic Summaries: Metrics under Varying Data Conditions
Karolina Owkzarzak and Hoa Trang Dang 23

A Classification Algorithm for Predicting the Structure of Summaries
Horacio Saggion 31

Syntax-Driven Sentence Revision for Broadcast News Summarization
Hideki Tanaka, Akinori Kinoshita, Takeshi Kobayakawa, Tadashi Kumano and Naoto Katoh . . . 39

A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression
Wei Xu and Ralph Grishman 48

Short Papers

Visual Development Process for Automatic Generation of Digital Games Narrative Content
Maria Fernanda Caropreso, Diana Inkpen, Shahzad Khan and Fazel Keshtkar 59

Reducing Redundancy in Multi-document Summarization Using Lexical Semantic Similarity
Iris Hendrickx, Walter Daelemans, Erwin Marsi and Emiel Krahmer 63

Non-textual Event Summarization by Applying Machine Learning to Template-based Language Generation
Mohit Kumar, Dipanjan Das, Sachin Agarwal and Alexander Rudnicky 67

Creating an Annotated Corpus for Generating Walking Directions
Stephanie Schuldes, Michael Roth, Anette Frank and Michael Strube 72

GREC'09 Papers

The GREC Main Subject Reference Generation Challenge 2009: Overview and Evaluation Results
Anja Belz, Eric Kow, Jette Viethen and Albert Gatt 79

The GREC Named Entity Generation Challenge 2009: Overview and Evaluation Results
Anja Belz, Eric Kow and Jette Viethen 90

ICSI-CRF: The Generation of References to the Main Subject and Named Entities Using Conditional Random Fields
Benoit Favre and Bernd Bohnet 99

UDel: Generating Referring Expressions Guided by Psycholinguistc Findings
Charles Greenbacker and Kathleen McCoy 101

JUNLG-MSR: A Machine Learning Approach of Main Subject Reference Selection with Rule Based Improvement
Samir Gupta and Sivaji Bandopadhyay 103

UDel: Extending Reference Generation to Multiple Entities
Charles Greenbacker and Kathleen McCoy 105

WLV: A Confidence-based Machine Learning Method for the GREC-NEG'09 Task
Constatin Orasan and Iustin Dornescu 107

Workshop Program

6 August, 2009

Morning Session 1: Sentence Compression and Revision

08:30–09:00 *Unsupervised Induction of Sentence Compression Rules*
Joao Cordeiro, Gael Dias and Pavel Brazdil

09:00–09:30 *A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression*
Wei Xu and Ralph Grishman

09:30–10:00 *Syntax-Driven Sentence Revision for Broadcast News Summarization*
Hideki Tanaka, Akinori Kinoshita, Takeshi Kobayakawa, Tadashi Kumano and Naoto Katoh

10:00–10.30 Coffee Break

Morning Session 2: Invited Talk / Content Selection

10:30–11:30 *Query-focused Summarization Using Text-to-Text Generation: When Information Comes from Multilingual Sources*
Kathy McKeown

11:30–12:00 *A Classification Algorithm for Predicting the Structure of Summaries*
Horacio Saggion

12:00–12:30 *Optimization-based Content Selection for Opinion Summarization*
Jackie Chi Kit Cheung, Giuseppe Carenini and Raymond T. Ng

12:30–13:50 Lunch

6 August, 2009 (continued)

Afternoon Session 1: Evaluation

13:50–15:00 GREC 2009 Shared Task Evaluation results session

The GREC Main Subject Reference Generation Challenge 2009: Overview and Evaluation Results
Anja Belz, Eric Kow, Jette Viethen and Albert Gatt
The GREC Named Entity Generation Challenge 2009: Overview and Evaluation Results
Anja Belz, Eric Kow and Jette Viethen
ICSI-CRF: The Generation of References to the Main Subject and Named Entities Using Conditional Random Fields
Benoit Favre and Bernd Bohnet
UDel: Generating Referring Expressions Guided by Psycholinguistc Findings
Charles Greenbacker and Kathleen McCoy
JUNLG-MSR: A Machine Learning Approach of Main Subject Reference Selection with Rule Based Improvement
Samir Gupta and Sivaji Bandopadhyay
UDel: Extending Reference Generation to Multiple Entities
Charles Greenbacker and Kathleen McCoy
WLV: A Confidence-based Machine Learning Method for the GREC-NEG'09 Task
Constatin Orasan and Iustin Dornescu

15:00–15:30 *Evaluation of Automatic Summaries: Metrics under Varying Data Conditions*
Karolina Owkzarzak and Hoa Trang Dang

15:30–16:00 Coffee Break

Afternoon Session 2: Short Papers/Discussion

16:00–16:20 *Visual Development Process for Automatic Generation of Digital Games Narrative Content*
Maria Fernanda Caropreso, Diana Inkpen, Shahzad Khan and Fazel Keshtkar

16:20–16:40 *Reducing Redundancy in Multi-document Summarization Using Lexical Semantic Similarity*
Iris Hendrickx, Walter Daelemans, Erwin Marsi and Emiel Krahmer

16:40–17:00 *Non-textual Event Summarization by Applying Machine Learning to Template-based Language Generation*
Mohit Kumar, Dipanjan Das, Sachin Agarwal and Alexander Rudnicky

17:00–17:20 *Creating an Annotated Corpus for Generating Walking Directions*
Stephanie Schuldes, Michael Roth, Anette Frank and Michael Strube

17:20–18:00 Panel-led discussion on synergies between summarisation and NLG, including shared tasks