ACL HLT 2011

Workshop on Monolingual Text-To-Text Generation

Proceedings of the Workshop

24 June, 2011
Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

©2011 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-13 9781937284053

Introduction

The ability to perform monolingual text-to-text generation is an important step in solving many natural language processing problems. For example, when generating novel text at the sentence-level, abstractive summarization systems may need to compress sentences or fuse multiple sentences together; the evaluation of translation systems may require additional paraphrases to use as reference gold standards; and answers to questions may need to be generated automatically from extracted sentences.

The community of researchers examining monolingual text-to-text generation has grown steadily in recent years, introducing the need for a focused venue to communicate results in this area. To this end, we proposed and organised this workshop at ACL with endorsement from SIGGEN. We hope that this is the first of many text-to-text generation workshops to come.

We were excited to receive 18 submissions which were judged in accordance with the standard reviewing practices of the ACL 2011 main conference. As we intended that the workshop serve as a new forum for the community, our aim in the selection process was to choose high quality papers which would spark discussion amongst the participants.

We selected seven long papers and four short papers. Together, they tackle a diverse range of research questions: reflecting upon the scope of what might be generated in a text-to-text process, examining new generation methods, and addressing the ever challenging issue of evaluation.

We would like to thank everyone involved in the preparation of this workshop. We were very happy to receive such an enthusiastic response from the community when we proposed the workshop. We would specifically like to thank Noah Smith for his invited talk. We would also like to thank the reviewers who helped us to put together this wonderful program. Finally, we are grateful for the guidance provided by the steering committee on the direction of this workshop.

We hope you find the program challenging and the resulting discussion engaging.

Katja and Stephen

Organizers:

Katja Filippova, Google
Stephen Wan, CSIRO

Program Committee:

Anja Belz, University of Brighton
Bernd Bohnet, University of Stuttgart
Aoife Cahill, University of Stuttgart
Chris Callison-Burch, Johns Hopkins University
Robert Dale, Macquarie University
Mark Dras, Macquarie University
Michel Galley, Microsoft
Kevin Knight, University of Southern California, ISI
Emiel Krahmer, Tilburg University
Mirella Lapata, University of Edinburgh
Nitin Madnani, ETS
Erwin Marsi, NTNU
Kathleen McKeown, Columbia University
Ryan McDonald, Google
Cécile Paris, CSIRO
Michael Strube, HITS
Michael White, Ohio State University
David Zajic, University of Maryland

Invited Speaker:

Noah Smith, Carnegie Mellon University

Table of Contents

Learning to Simplify Sentences Using Wikipedia
Will Coster and David Kauchak 1

Web-based Validation for Contextual Targeted Paraphrasing
Houda Bouamor, Aurélien Max, Gabriel Illouz and Anne Vilnat 10

An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction
Stefan Bott and Horacio Saggion 20

Comparing Phrase-based and Syntax-based Paraphrase Generation
Sander Wubben, Erwin Marsi, Antal van den Bosch and Emiel Krahmer 27

Text Specificity and Impact on Quality of News Summaries
Annie Louis and Ani Nenkova 34

Towards Strict Sentence Intersection: Decoding and Evaluation Strategies
Kapil Thadani and Kathleen McKeown 43

Learning to Fuse Disparate Sentences
Micha Elsner and Deepak Santhanam 54

Framework for Abstractive Summarization using Text-to-Text Generation
Pierre-Etienne Genest and Guy Lapalme 64

Creating Disjunctive Logical Forms from Aligned Sentences for Grammar-Based Paraphrase Generation
Scott Martin and Michael White 74

Paraphrastic Sentence Compression with a Character-based Metric: Tightening without Deletion
Courtney Napoles, Chris Callison-Burch, Juri Ganitkevitch and Benjamin Van Durme 84

Evaluating Sentence Compression: Pitfalls and Suggested Remedies
Courtney Napoles, Benjamin Van Durme and Chris Callison-Burch 91

Conference Program

Friday June 24, 2011

Session 1: 9:00 - 10:30

Learning to Simplify Sentences Using Wikipedia
Will Coster and David Kauchak

Web-based Validation for Contextual Targeted Paraphrasing
Houda Bouamor, Aurélien Max, Gabriel Illouz and Anne Vilnat

An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction
Stefan Bott and Horacio Saggion

Comparing Phrase-based and Syntax-based Paraphrase Generation
Sander Wubben, Erwin Marsi, Antal van den Bosch and Emiel Krahmer

Morning break: 10:30 - 11:00

Invited Talk (Noah Smith) and Discussion: 11:00 - 12:30

Lunch break: 12:30 - 14:00

Session 3: 14:00 - 15:30

Text Specificity and Impact on Quality of News Summaries
Annie Louis and Ani Nenkova

Towards Strict Sentence Intersection: Decoding and Evaluation Strategies
Kapil Thadani and Kathleen McKeown

Learning to Fuse Disparate Sentences
Micha Elsner and Deepak Santhanam

Friday June 24, 2011 (continued)

Afternoon break: 15:30 - 16:00

Session 4: 16:00 - 17:30

Framework for Abstractive Summarization using Text-to-Text Generation
Pierre-Etienne Genest and Guy Lapalme

Creating Disjunctive Logical Forms from Aligned Sentences for Grammar-Based Paraphrase Generation
Scott Martin and Michael White

Paraphrastic Sentence Compression with a Character-based Metric: Tightening without Deletion
Courtney Napoles, Chris Callison-Burch, Juri Ganitkevitch and Benjamin Van Durme

Evaluating Sentence Compression: Pitfalls and Suggested Remedies
Courtney Napoles, Benjamin Van Durme and Chris Callison-Burch