

Text Summarization Branches Out

Proceedings of the ACL-04 Workshop

Marie-Francine Moens and Stan Szpakowicz, co-chairs

Held in cooperation with ACL-2004

25-26 July 2004
Forum Convention Centre
Barcelona, Spain

A Word from the Co-Chairs

Text summarization is still largely in a research phase, and has so far focused on news text, but it is increasingly becoming a tool for information search and selection in a variety of media. For example, summarizing is a necessity when showing content on the screen of a mobile device. Texts integrated in multimedia documents have different genres or types, but they all require the same flexibility in the presentation of summaries by allowing parameterized compression rates and integration in a mixed-media format.

Recently, text summarization technologies have advanced a lot, thanks in a large measure to the Document Understanding Conferences (DUC), sponsored by the Defense Advanced Research Projects Agency (DARPA) and organized by the National Institute of Standards and Technology in the USA. The ACL-2004 text summarization workshop in Barcelona aims both to broaden the scope of summarization beyond textual news stories and to make more people interested in this challenging field that intersects natural language processing and information retrieval. That was the thinking behind the title of our workshop: Text Summarization Branches Out.

We received 33 papers from 12 countries: Belgium, Brazil, Canada, China, France, Germany, India, Israel, Japan, Spain, UK and USA. The submissions covered a diverse selection of cutting-edge topics. We accepted 14 full papers and 3 short papers. Curiously, the most prominently represented is work on the quality of summaries. Reliable evaluation metrics and procedures are essential if we want to advance the state of the art in summarization. There are interesting studies that compare various summarization technologies and compute how their results correlate with human-made summaries. Another sizable group of papers discuss sentence compression and information fusion, novel and useful approaches to summarization in a world of small mobile devices with their miniature screens. The workshop moves beyond news stories: we have papers that deal with legal texts, figures and graphics, subtitles, technical reports, computer product reviews and email.

As far as the summarization techniques are concerned, the papers show a mix of statistical techniques and linguistically motivated natural language processing techniques, including semantic analysis and discourse analysis. Automated reasoning techniques allow fusion and understanding of content. Machine learning, supervised or unsupervised, still has a major role to play. The workshop features two panels. The first panel will look backward, attempting to summarize (yes!) progress especially in the last ten years. The second panel will look forward to the near and more distant future of summarization technologies.

We are hugely obliged to Roxana Angheluta for being the "chair" of our instance of CyberChair, Richard van de Stadt's fantastic application that helped us organizing the workshop. We thank Eduard Hovy and Dragomir Radev for their crucial advice and valuable comments. We are indebted to Inderjeet Mani for agreeing to give the invited talk. We acknowledge the assistance of the ACL-2004 workshop chairs, and of local organizers who helped schedule the event. Last, but certainly not least, we are very grateful to the members of the Program Committee for the time they generously devoted to reviewing the papers. Three referees read each paper in a blind-reviewing process.

Welcome to the ACL-2004 workshop Text Summarization Branches Out. Enjoy!

Marie-Francine Moens
Stan Szpakowicz

ORGANIZERS:

Eduard Hovy, Information Sciences Institute, University of Southern California, USA
Marie-Francine Moens (co-chair), Interdisciplinary Centre for Law & Information Technology, Katholieke Universiteit Leuven, Belgium
Dragomir Radev, School of Information and Department of Electrical Engineering and Computer Science, University of Michigan, USA
Stan Szpakowicz (co-chair), School of Information Technology and Engineering, University of Ottawa, Canada

PROGRAM COMMITTEE:

Regina Barzilay, Computer Science and Artificial Intelligence Lab, MIT, USA
Hercules Dalianis, Royal Institute of Technology, Sweden
Chiori Hori, NTT, Japan
Eduard Hovy, Information Sciences Institute, University of Southern California, USA
Hongyan Jing, IBM T.J. Watson Research Center, USA
Kathy McKeown, Computer Science Department, Columbia University, USA
Chin-Yew Lin, Information Sciences Institute, University of Southern California, USA
Inderjeet Mani, Department of Linguistics, Georgetown University, USA
Daniel Marcu, Information Sciences Institute, University of Southern California, USA
Marie-Francine Moens, Interdisciplinary Centre for Law & Information Technology, Katholieke Universiteit Leuven, Belgium
Dragomir Radev, School of Information and Department of Electrical Engineering and Computer Science, University of Michigan, USA
Horacio Rodriguez, Departamento de LSI, Universitat Politecnica de Catalunya, Spain
Horacio Saggion, Department of Computer Science, University of Sheffield, UK
Judith Schlesinger, IDA/Center for Computing Sciences, USA
Karen Sparck Jones, Computer Laboratory, Cambridge University, UK
Stan Szpakowicz, School of Information Technology and Engineering, University of Ottawa, Canada
John Tait, School of Computing and Technology, University of Sunderland, UK
Simone Teufel, Computer Laboratory, University of Cambridge, UK
Peter Turney, NRC Ottawa, Canada
Hans van Halteren, Department of Language and Speech, University of Nijmegen, The Netherlands

Table of Contents

<i>Invited Lecture: Narrative Summarization</i> Inderjeet Mani	1
<i>Extending Document Summarization to Information Graphics</i> Sandra Carberry, Stephanie Elzer, Nancy Green, Kathleen McCoy and Daniel Chester	3
<i>The Effects of Human Variation in DUC Summarization Evaluation</i> Donna Harman and Paul Over	10
<i>Paragraph-, Word- and Coherence-Based Approaches to Sentence Ranking: A Comparison of Algorithm and Human Performance</i> Florian Wolf and Edward Gibson	18
<i>Vocabulary Usage in Newswire Summaries</i> Terry Copeck and Stan Szpakowicz	19
<i>Legal Text Summarization by Exploration of the Thematic Structure and Argumentative Roles</i> Atefeh Farzindar and Guy Lapalme	27
<i>A Rhetorical Status Classifier for Legal Text Summarisation</i> Ben Hachey and Claire Grover	35
<i>Task-Focused Summarization of Email</i> Simon Corston-Oliver, Eric Ringger, Michael Gamon and Richard Campbell	43
<i>Hybrid Text Summarization: Combining External Relevance Measures with Structural Analysis</i> Gian Lorenzo Thione, Martin van den Berg, Livia Polanyi and Chris Culy	51
<i>Template-Filtered Headline Summarization</i> Liang Zhou and Eduard Hovy	56
<i>Handling Figures in Document Summarization</i> Robert P. Futrelle	61
<i>Automatic Evaluation of Summaries Using Document Graphs</i> Eugen Santos Jr., Ahmed A. Mohamed and Qunhua Zhao	66
<i>ROUGE: A Package for Automatic Evaluation of Summaries</i> Chin-Yew Lin	74
<i>Evaluation Measures Considering Sentence Concatenation for Automatic Summarization by Sentence or Word Extraction</i> Chiori Hori, Tsutomu Hirao and Hideki Isozaki	82
<i>Sentence Compression for Automated Subtitling: A Hybrid Approach</i> Vincent Vandeghinste and Yi Pan	89
<i>Generic Sentence Fusion is an Ill-Defined Summarization Task</i> Hal Daumé III and Daniel Marcu	96
<i>Event-Based Extractive Summarization</i> Elena Filatova and Vasileios Hatzivassiloglou	104
<i>Chinese Text Summarization Based on Thematic Area Detection</i> Po Hu, Tingting He and Donghong Ji	112

WORKSHOP PROGRAM

Sunday, July 25

- 8:25-8:30 Welcome
- 8:30-9:30 *Invited Lecture: Narrative Summarization*
Inderjeet Mani
- 9:30-10:00 *Extending Document Summarization to Information Graphics*
Sandra Carberry, Stephanie Elzer, Nancy Green, Kathleen McCoy and Daniel Chester
- 10:00-10:30 Coffee Break
- 10:30-11:00 *The Effects of Human Variation in DUC Summarization Evaluation*
Donna Harman and Paul Over
- 11:00-11:30 *Paragraph-, Word- and Coherence-Based Approaches to Sentence Ranking: A Comparison of Algorithm and Human Performance*
Florian Wolf and Edward Gibson
- 11:30-12:00 *Vocabulary Usage in Newswire Summaries*
Terry Copeck and Stan Szpakowicz
- 12:00-13:50 Lunch
- 13:50-15:20 *Panel 1: Text Summarization: A Look at the Last Decades*
Eduard Hovy, Donna Harman, Marie-Francine Moens, Judith Schlesinger and Hans van Halteren
- 15:20-15:40 Coffee Break
- 15:40-16:10 *Legal Text Summarization by Exploration of the Thematic Structure and Argumentative Roles*
Atefeh Farzindar and Guy Lapalme
- 16:10-16:40 *A Rhetorical Status Classifier for Legal Text Summarisation*
Ben Hachey and Claire Grover
- 16:40-17:10 *Task-Focused Summarization of Email*
Simon Corston-Oliver, Eric Ringger, Michael Gamon and Richard Campbell
- 17:10-17:30 *Hybrid Text Summarization: Combining External Relevance Measures with Structural Analysis*
Gian Lorenzo Thione, Martin van den Berg, Livia Polanyi and Chris Culy
- 17:30-17:50 *Template-Filtered Headline Summarization*
Liang Zhou and Eduard Hovy
- 17:50-18:10 *Handling Figures in Document Summarization*
Robert Futrelle

Monday, July 26

- 8:30-9:00 *Automatic Evaluation of Summaries Using Document Graphs*
Eugen Santos Jr., Ahmed A. Mohamed and Qunhua Zhao
- 9:00-9:30 *ROUGE: A Package for Automatic Evaluation of Summaries*
Chin-Yew Lin
- 9:30-10:00 *Evaluation Measures Considering Sentence Concatenation for Automatic Summarization by Sentence or Word Extraction*
Chiori Hori, Tsutomu Hirao and Hideki Isozaki
- 10:00-10:30 Coffee Break
- 10:30-12:00 *Panel 2: Text Summarization: What Lies Ahead*
Stan Szpakowicz, Eduard Hovy, Daniel Marcu, Dragomir Radev and Simone Teufel
- 12:00-13:30 Lunch
- 13:30-14:00 *Sentence Compression for Automated Subtitling: A Hybrid Approach*
Vincent Vandeghinste and Yi Pan
- 14:00-14:30 *Generic Sentence Fusion is an Ill-Defined Summarization Task*
Hal Daumé III and Daniel Marcu
- 14:30-15:00 *Event-Based Extractive Summarization*
Elena Filatova and Vasileios Hatzivassiloglou
- 15:00-15:30 *Chinese Text Summarization Based on Thematic Area Detection*
Po Hu, Tingting He and Donghong Ji
- 15:30-15:35 Closing Remarks

THIS IS A BLANK PAGE

PLEASE IGNORE

Author Index

Campbell, Richard	43
Carberry, Sandra	3
Chester, Daniel	3
Copeck, Terry	19
Corston-Oliver, Simon	43
Culy, Chris	51
Daumé III, Hal	96
Elzer, Stephanie	3
Farzindar, Atefeh	27
Filatova, Elena	104
Futrelle, Robert P.	61
Gamon, Michael	43
Gibson, Edward	18
Green, Nancy	3
Grover, Claire	35
Hachey, Ben	35
Harman, Donna	10
Hatzivassiloglou, Vasileios	104
He, Tingting	112
Hirao, Tsutomu	82
Hori, Chiori	82
Hovy, Eduard	56
Hu, Po	112
Inderjeet Mani	1
Isozaki, Hideki	82
Ji, Donghong	112
Lapalme, Guy	27
Lin, Chin-Yew	74
Marcu, Daniel	96
McCoy, Kathleen	3
Mohamed, Ahmed A.	66
Over, Paul	10
Pan, Yi	89
Polanyi, Livia	51
Ringger, Eric	43
Santos Jr., Eugen	66
Szpakowicz, Stan	19
Thione, Gian Lorenzo	51
Van den Berg, Martin	51
Vandeghinste, Vincent	89
Wolf, Florian	18
Zhao, Qunhua	66
Zhou, Liang	56