

Intelligent Scalable Text Summarization

**Proceedings of a Workshop
Sponsored by the
Association for Computational Linguistics**

**Edited by
Inderjeet Mani
and
Mark Maybury**

Supported by The MITRE Corporation

**11 July 1997
Universidad Nacional de Educación a Distancia
Madrid, Spain**

© 1997, Association for Computational Linguistics

Order additional copies from

ACL

P O Box 6090

Somerset, NJ, 08875 USA

+1-908-873-3898

acl@bellcore.com

PREFACE

This volume contains the papers presented at the (first) workshop on Intelligent Scalable Text Summarization, held on 11 July 1997 in conjunction with the joint conference of the 35th Annual Meeting of the Association for Computational Linguistics (ACL) and the 8th European Chapter of the ACL (ACL-EACL'97 Joint Conference)

With the explosion in the quantity of on-line information in recent years, demand for text summarization technology appears to be growing. Commercial companies are increasingly starting to offer text summarization capabilities, often bundled with information retrieval tools and database systems. These recent developments offer opportunities as well as substantial challenges for research in text summarization. In general, such developments create a practical need for summarization systems which scale up when applied to large volumes of unrestricted text. While there have been focused workshops in the past on text summarization, they have pre-dated the tremendous expansion of on-line information access fueled by the recent growth of the World Wide Web. This workshop is aimed at researchers interested in advancing the scientific frontiers of text summarization to meet these new practical challenges and opportunities.

The papers presented here include statistical paradigms for intelligent text summarization, methods for identifying summary topics in text, exploitation of recent advances in information extraction in summarization, evaluation methods and metrics, and theoretical foundations, including cognitive models. Of particular interest to the issue of scalability is the increased emphasis on hybrid approaches which combine statistical techniques with robustly extracted linguistic representations. The papers in this one-day workshop offer only a sample of the emerging approaches, other relevant topics, which may be touched on in the discussion sessions, include the exploitation of corpus-based resources, multi-document summarization, multilingual summarization, and the use of multi-modal information presentation strategies in summarization.

We would like to first of all thank the authors, whose research contributions have made this workshop possible. We also appreciate the efforts of other members of the Program Committee in reviewing papers and assisting in the design and running of the workshop: Udo Hahn, Julian Kupiec, Kathy McKeown, Boyan Onyshkevych, Dragomir Radev, Lisa Rau, and Kazuo Tanaka. Dragomir Radev also provided valuable services as our Webmaster. Thanks are due also to the following: Harald Trost (the ACL workshop chair) and Felisa Verdejo (local arrangements), for their logistics support at various points. In addition,

we thank Anne Roemer at MITRE, for her help in assembling the master copy of these Proceedings. Finally, we are grateful to the ACL for sponsoring the workshop, and to the MITRE Corporation for supporting the publication of these Proceedings.

Inderjeet Mani

Mark Maybury

Program Chairs

PROGRAM COMMITTEE

Udo Hahn	<i>University of Freiburg</i>
Julian Kupiec	<i>Xerox Palo Alto Research Center</i>
Inderjeet Mani	<i>The MITRE Corporation</i>
Mark Maybury	<i>The MITRE Corporation</i>
Kathy McKeown	<i>Columbia University</i>
Boyan Onyshkevych	<i>US Department of Defense</i>
Dragomir Rade	<i>Columbia University</i>
Lisa Rau	<i>SRA International</i>
Kazuo Tanaka	<i>NTT Human Interface Laboratories</i>

WORKSHOP PROGRAM

11 July 1997

*Universidad Nacional de Educación a Distancia
Madrid, Spain*

OPENING SESSION

- 9 00-9 10 Opening Remarks
Inderjeet Mani
- 9 10-9 40 Keynote Address Where are we now? Where should we go?
Karen Sparck Jones
- 9 40-9 50 Discussion

SESSION 1: TOPIC-LEVEL ABSTRACTION (chair Udo Hahn)

- 9 50-10 10 Salience-based Content Characterization of Text Documents
Brammir Boguraev and Christopher Kennedy
- 10 10-10 30 Using Lexical Chains for Text Summarization
Regina Barzilay and Michael Elhadad
- 10 30-10 50 Automated Text Summarization in SUMMARIST
Eduard Hovy and Chin Yew Lin
- 10 50-11 05 Discussion
- 11 05-11 20 BREAK

SESSION 2: EVALUATION METHODS (chair Boyan Onyshkevich)

- 11 20-11 40 How to Appreciate the Quality of Automatic Text Summarization
Jean-Luc Minel, Sylvaine Nugier, and Gérard Piat
- 11 40-12 00 A Proposal for Task-Based Evaluation of Text Summarization Systems
Thérèse F Hand
- 12 00-12 15 Discussion
- 12 15-2 00 LUNCH

SESSION 3: STATISTICAL APPROACHES (chair Eduard Hovy)

- 2 00-2 20 Automatic Text Summarization by Paragraph Extraction
Mandar Mitra, Amit Singhal, and Chris Buckley
- 2 20-2 40 Goal Directed Approach for Text Summarization
Ryo Ochitani, Yoshio Nakao, and Fumihito Nishino
- 2 40-3 00 Statistical Methods for Retrieving Most Significant Paragraphs in Newspaper Articles
José Abracos and Gabriel Pereira Lopes
- 3 00-3 20 Sentence extraction as a classification task
Simone H. Teufel and Marc Moens
- 3 20-3 35 Discussion
- 3 35-4 00 BREAK

SESSION 4: COMBINING DISCOURSE-LEVEL FEATURES

(chair Inderjeet Mani)

- 4 00-4 20 A Scalable Summarization System Using Robust NLP
Chinatsu Aone, Mary Ellen Okurowski, James Gorlinsky, Bjornar Larsen
- 4 20-4 40 COSY-MATS: An Intelligent and Scalable Summarization Shell
Maria Aretoulaki
- 4 40-5 00 From Discourse Structures to Text Summaries
Daniel Marcu
- 5 00-5 15 Discussion
- 5 15-5 30 BREAK

SESSION 5: SUMMARIZATION MODELS (chair Mark Maybury)

- 5 30-5 50 SimSum: Simulation of summarizing
Brigitte Endres-Niggemeyer
- 5 50-6 10 A Formal Model of Text Summarization Based on Condensation Operators
of a Terminological Logic
Ulrich Reimer and Udo Hahn
- 6 10-6 25 Discussion
- 6 25-6 35 Wrap-Up
Mark Maybury

TABLE OF CONTENTS

<i>Summarising Where are we now? Where should we go?</i> , Karen Sparck Jones	1
<i>Salience-based Content Characterization of Text Documents</i> , Branimir Boguraev and Christopher Kennedy	2
<i>Using Lexical Chains for Text Summarization</i> , Regina Barzilay and Michael Elhadad	10
<i>Automated Text Summarization in SUMMARIST</i> , Eduard Hovy and Chin Yew Lin	18
<i>How to Appreciate the Quality of Automatic Text Summarization</i> , Jean-Luc Minel, Sylvaine Nugier, and Gérald Piat	25
<i>A Proposal for Task-Based Evaluation of Text Summarization Systems</i> , Thérèse F Hand	31
<i>Automatic Text Summarization by Paragraph Extraction</i> , Mandar Mitra, Amit Singhal, and Chris Buckley	39
<i>Goal Directed Approach for Text Summarization</i> , Ryo Ochitani, Yoshio Nakao, and Fumihito Nishino	47
<i>Statistical Methods for Retrieving Most Significant Paragraphs in Newspaper Articles</i> , José Abracos and Gabriel Pereira Lopes	51
<i>Sentence Extraction as a Classification Task</i> , Simone H Teufel and Marc Moens	58
<i>A Scalable Summarization System Using Robust NLP</i> , Chinatsu Aone, Mary Ellen Okurowski, James Gorklinsky, and Bjornar Larsen	66
<i>COSY-MATS An Intelligent and Scalable Summarization Shell</i> , Maria Aretoulaki	74
<i>From Discourse Structures to Text Summaries</i> , Daniel Marcu	82
<i>SimSum Simulation of summarizing</i> , Brigitte Endres-Niggemeyer	89
<i>A Formal Model of Text Summarization Based on Condensation Operators of a Terminological Logic</i> , Ulrich Reimer and Udo Hahn	97

AUTHOR INDEX

José Abracos	51
Chinatsu Aone	66
Maria Aretoulaki	74
Regina Barzilay	10
Branimir Boguraev	2
Chris Buckley	39
Michael Elhadad	10
Brigitte Endres-Niggemeyer	89
James Gorlinsky	66
Udo Hahn	2
Thérèse F Hand	97
Eduard Hovy	18
Christopher Kennedy	2
Bjornar Larsen	66
Chin Yew Lin	18
Gabriel Pereira Lopes	51
Daniel Marcu	82
Jean-Luc Minel	25
Mandar Mitra	39
Marc Moens	58
Yoshio Nakao	47
Fumihito Nishino	47
Sylvaine Nugier	25
Ryo Ochitani	47
Mary Ellen Okurowski	66
Gérald Piat	25
Ulrich Reimer	97
Amit Singhal	39
Karen Sparck-Jones	1
Simone H Teufel	58