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, mult-document summarization, multilingual summarization, and the use of multimodal 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  University of Freiburg
Julian Kupiec  Xerox Palo Alto Research Center
Inderjeet Mani  The MITRE Corporation
Mark Maybury  The MITRE Corporation
Kathy McKeown  Columbia University
Boyan Onyshkevych  US Department of Defense
Dragomir Rade  Columbia University
Lisa Rau  SRA International
Kazuo Tanaka  NTT Human Interface Laboratories
WORKSHOP PROGRAM
11 July 1997
Universidad Nacional de Educación a Distancia
Madrid, Spain

OPENING SESSION
9 00-9 10 Opening Remarks
Inderjeet Man
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
Brammert 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 Gérald 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
Jose 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 Mann)

4 00-4 20  A Scalable Summarization System Using Robust NLP
Chinatsu Aone, Mary Ellen Okuroski, James Gorinsky, 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

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