FSMNLP'98

Proceedings of the International Workshop on Finite State Methods in Natural Language Processing

> Lauri Karttunen Kemal Oflazer

June 30 - July 1, 1998

Bilkent University Ankara, Turkey

| | | · | | |
|--|--|---|--|--|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

FSMNLP'98

Proceedings of the International Workshop on Finite State Methods in Natural Language Processing

Lauri Karttunen Kemal Oflazer

June 30 - July 1, 1998

Bilkent University Ankara, Turkey

Published by Bilkent University.

ISBN 975-7679-34-8

Copyright © 1998 by Bilkent University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise for commercial purposes, without the written permission of the Publisher, Bilkent University Faculty of Engineering, TR-06533 Bilkent, Ankara, Turkey.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods products, instructions or ideas contained in the material herein.

Printed and bound by Meteksan A. Ş., Ankara, Turkey. May 1998.

FSMNLP'98

INTERNATIONAL WORKSHOP ON FINITE STATE METHODS IN NATURAL LANGUAGE PROCESSING

June 30-July 1, 1998

Bilkent University Ankara, Turkey

Supported By

EACL - European Chapter of the Association for Computational Linguistics
TÜBİTAK-Turkish Scientific and Technical Research Council
NATO Science for Stability Programme TU-LANGUAGE Project

PROGRAMME AND ORGANIZING COMMITTEE

CO-CHAIRS

| Lauri Karttunen | Xerox Research Center Europe | Kemal Oflazer | Bilkent University |
|---------------------|------------------------------|---|--------------------------|
| Kenneth R. Beesley | Xerox Research Center Europe | Eric Brill Jerry Hobbs Martin Kay András Kornai Tomasz Kowaltowski Mehryar Mohri Richard Sproat Yves Schabes Atro Voutilainen | Johns Hopkins University |
| Eva Ejerhed | Umea University | | SRI International |
| Ronald M. Kaplan | Xerox PARC | | Xerox PARC |
| George Kiraz | Bell Laboratories | | Bolt Beranek Newman |
| Kimmo Koskenniemi | University of Helsinki | | University of Campinas |
| Cláudio L. Lucchesi | University of Campinas | | AT&T Labs Research |
| Mark-Jan Nederhof | DFKI | | Bell Laboratories |
| Emmanuel Roche | Teragram Corp. | | Teragram Corp. |
| Gertjan Van Noord | University of Groningen | | University of Helsinki |

PREFACE

Recent years has seen a substantial increase in the use of finite state techniques in many aspects of natural language processing as mature tools for building large scale finite-state systems from various research laboratories and universities become available. This trend was by no means foreseen as late as ten years ago given the well-known demonstration by Noam Chomsky in 1957 that finite-state methods are inherently incapable of representing the full richness of constructions in a natural language. Nevertheless, it is evident now that there are many subsets of natural language that are adequately covered by finite-state means and that there are many other areas where finite-state approximations of more powerful formalisms are of great practical benefit. The discovery that systems of phonological rewrite rules and optimality contraints are within the finite-state domain has made an important theoretical leap forward. We expect that similar discoveries are yet to be made in other areas of linguistics.

FSMNLP'98, International Workshop on Finite State Methods in Natural Language Processing was conceived, with support and motivation from EACL, as a forum to bring together recent contributions in all aspects of the theory and applications of finite state machinery in language processing. As you may have already observed by reviewing the workshop programme and by looking at the papers to be presented, the mix of the contributions is international and shows wide-spread interest in this technology. They range over a variety of topics of great interest to the NLP community.

We thank members of the Programme Committee who, with their prompt and dedicated reviews of the submissions, enabled us to select the contributions for these proceedings. We thank Bilkent University for providing the facilities and contributing to the logistics of the workshop. We also thank the NATO Science for Stability Programme (Phase III) which let us use funds from the TU-LANGUAGE Project, and TBITAK, the Turkish Scientific and Technical Research Council, for additional financial support. Last but not least, we thank Professor Bülent Özgüç, Dean of the Faculty of Art, Design and Architecture, for his help in designing the cover and in the production of the proceedings.

We hope you enjoy the workshop.

Lauri Karttunen and Kemal Oflazer

| | t · · · · | |
|---|-----------|--|
| | | |
| | | |
| | | |
| · | | |
| | | |
| | | |
| | | |
| · | | |
| | | |
| | | |
| | | |
| | | |

WORKSHOP PROGRAMME

and

TABLE OF CONTENTS

| TUESDAY, | June 30, 1998 | Page No. | | | |
|-----------------|---|---|--|--|--|
| • 9:45 - 10:00 | Opening Remarks | | | | |
| • 10:00 - 11:00 | PLENARY TALK: The Proper Treatment of Optimality in Computational Phonology(1) | | | | |
| | Lauri Karttunen | Xerox Research Centre Europe, France | | | |
| • 11:00 - 11:30 | BREAK | | | | |
| • 11:30 - 12:00 | Context-free Parsing Through Mark-Jan Nederhof | Regular Approximation(13) DFKI, Germany | | | |
| • 12:00 - 12:30 | | Case Study on Finite State Parsing(25) University of Helsinki, Finland | | | |
| • 12:30 - 14:00 | LUNCH | | | | |
| • 14:00 - 14:30 | Robust Parsing Using a Hidde Wide R. Hogenhout Yuji Matsumoto | n Markov Model(37) Canon Research Centre Europe, United Kingdom Nara Institute of Technology, Japan | | | |
| • 14:30 - 15:00 | Incremental Construction of Minimal Acyclic Finite State Automata and Transducers | | | | |
| | Jan Daciuk Bruce W. Watson | University of Gdansk, Poland Ribbit Software, Canada and University of Pretoria, South Africa | | | |
| | Richard E. Watson | Ribbit Software, Canada | | | |
| • 15:00 - 15:30 | | et Construction(57) University of Groningen, The Netherlands | | | |
| • 15:30 - 16:00 | BREAK | | | | |
| • 16:00 - 16:30 | Learning Finite State Models i David Picó Enrique Vidal | For Language Understanding(69) Polytechnic University of Valencia Spain | | | |
| • 16:30 - 17:00 | A Multilingual Natural Langua Aarne Ranta | age Interface to Regular Expressions(79) Xerox Research Centre Europe, France | | | |

WEDNESDAY, July 1, 1998

| | Implementing Voting Cons Kemal Oflazer Gökhan Tür | straints with Finite State Transducers(91) Bilkent University, Turkey |
|-----------------|---|--|
| • 10:30 - 11:00 | Feature Structures, Unific Rémi Zajac | ation and Finite State Transducers |
| • 11:00 - 11:30 | Using Genericity to Create Sandro Pedrazzini Marcus Hoffman | e Customizable Finite State Tools(110) University of Basel and IDSIA Switzerland |
| • 11:30 - 12:00 | Constraining Separated M mars Kenneth R. Beesley | orphotactic Dependencies in Finite State Gram(118) |