

Fifth Message Understanding Conference (MUC-5)

**Proceedings of a Conference
Held in
Baltimore, Maryland
August 25-27, 1993**

Sponsored by:

**Advanced Research Projects Agency
Software and Intelligent Systems Technology Office**

This document contains copies of reports and
other materials prepared for use at the
Fifth Message Understanding Conference.

**APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED**

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the United States Government.

Distributed by
Morgan Kaufmann Publishers, Inc.
340 Pine Street, 6th Floor
San Francisco, California 94104
ISBN 1-55860-336-0
Printed in the United States of America

Preface

This volume documents the proceedings of the Fifth Message Understanding Conference (MUC-5), which was held on 25-27 August, 1993, in Baltimore, Maryland. The conference was sponsored by the Advanced Research Projects Agency, Software and Intelligent Systems Technology Office (Thomas Crystal, Program Manager) and was organized by the MUC-5 program committee: Beth Sundheim, chair (NCCOSC/NRaD), Sean Boisen (BBN), Lynn Carlson (DoD), Nancy Chinchor (SAIC), Jim Cowie (CRL/NMSU), Ralph Grishman (NYU), Jerry Hobbs (SRI), Joe McCarthy (UMass), Mary Ellen Okurowski (DoD), Boyan Onyshkevych (DoD), Lisa Rau (GE), and Carl Weir (UNISYS).

The topic of the conference was performance assessment of text analysis software systems designed to extract information from free text. Prior to the conference, systems were developed and tested on information extraction performance in the domains of joint business ventures and/or microelectronic chip fabrication in English and/or Japanese. The conference was attended by representatives of organizations that participated in the evaluation, Government representatives, and other invited guests. Sessions included a tutorial on information extraction offered by Ralph Grishman (NYU), introductions to the information extraction evaluation tasks by Boyan Onyshkevych and Mary Ellen Okurowski (DoD), a paper describing a "generic" information extraction system by Jerry Hobbs (SRI), papers on systems and test results given by the participating organizations in the evaluation, system demonstrations, presentations on the evaluation design and overviews of the test results by Beth Sundheim (NCCOSC/NRaD), Nancy Chinchor (SAIC), Mary Ellen Okurowski (DoD), Boyan Onyshkevych (DoD), and Craig Will (IDA), workshops on special topics with presentations by some of the evaluation participants, and discussion meetings to critique the evaluation and make recommendations for future evaluations.

MUC-5 was closely associated with phase one of the ARPA Tipster Text Program. Some of the papers in this volume appear also in a published report on Tipster. Those papers are identified in this volume at the beginning of the section in which they appear.

TABLE OF CONTENTS

Preface

PAPERS: INFORMATION EXTRACTION TASK

Corpora and Data Preparation	1
L. Carlson, B. Onyshkevych, and M.E. Okurowski (U.S. Department of Defense)	
Tasks, Domains, and Languages	7
B. Onyshkevych, M.E. Okurowski, and L. Carlson (U.S. Department of Defense)	
Template Design for Information Extraction	19
B. Onyshkevych (U.S. Department of Defense)	

PAPERS: EVALUATION

Tipster/MUC-5 Information Extraction System Evaluation	27
B. Sundheim (Naval Command, Control and Ocean Surveillance Center)	
Domain and Language Evaluation Results	45
M.E. Okurowski (U.S. Department of Defense)	
Comparing Human and Machine Performance for Natural Language Information Extraction: Results for English Microelectronics from the MUC-5 Evaluation	53
C. Will (Institute for Defense Analyses)	
MUC-5 Evaluation Metrics	69
N. Chinchor (Science Applications International Corporation) and B. Sundheim (Naval Command, Control and Ocean Surveillance Center)	
The Statistical Significance of the MUC-5 Results	79
N. Chinchor (Science Applications International Corporation)	

PAPERS: SYSTEMS

The Generic Information Extraction System	87
J. Hobbs (SRI International)	
BBN: Description of the PLUM System as Used for MUC-5	93
R. Weischedel, D. Ayuso, S. Boisen, H. Fox, R. Ingria, T. Matsukawa, and C. Papageorgiou (BBN Systems and Technologies), D. MacLaughlin, M. Kitagawa, and T. Sakai (Boston University), J. Abe, H. Hosihi, and Y. Miyamoto (University of Connecticut), and S. Miller (Northeastern University)	
GE-CMU: Description of the SHOGUN System Used for MUC-5	109
P. Jacobs, G. Krupka, and L. Rau (GE Research and Development), M. Mauldin, T. Mitamura, and T. Kitani (Carnegie Mellon University), I. Sider and L. Childs (Martin Marietta)	
Language Systems Inc: Description of the DBG System as Used for MUC-5	121
C. Montgomery, R. Stumberger, B. Glover Stalls, N. Li, R. Belvin, and S. Hirsh Litenatsky (Language Systems, Inc.)	
MITRE-Bedford: Description of the <i>Alembic</i> System as Used for MUC-5	137
J. Aberdeen, J. Burger, D. Connolly, S. Roberts, and M. Vilain (The MITRE Corporation)	

NEC: Description of the VENIEX System as Used for MUC-5	147
K. Muraki, S. Doi, and S. Ando (NEC Corporation)	
CRL/Brandeis: Description of the <i>Diderot</i> System as Used for MUC-5	161
J. Cowie, L. Guthrie, W. Jin, R. Wang, and T. Wakao (Computing Research Laboratory, New Mexico State University), J. Pustejovsky and S. Waterman (Brandeis University)	
New York University: Description of the Proteus System as Used for MUC-5	181
R. Grishman and J. Sterling (New York University)	
PRC Inc.: Description of the PAKTUS System Used for MUC-5	195
B. Loatman and C.-H. Yang (PRC Inc.)	
SRA: Description of the SOLOMON System as Used for MUC-5	207
C. Aone, S. Flank, D. McKee, and P. Krause (Systems Research and Applications)	
SRI: Description of the JV-FASTUS System Used for MUC-5	221
D. Appelt, J. Hobbs, J. Bear, D. Israel, M. Kameyama, and M. Tyson (Artificial Intelligence Center, SRI International)	
TRW: Description of the DEFT System as Used for MUC-5	237
W. Noah and R. Weeks (TRW Systems Development Division)	
UNISYS: Description of the CBAS System Used for MUC-5	249
C. Weir and R. Fritzson (Unisys Corporation)	
University of Manitoba: Description of the NUBA System as Used for MUC-5	263
D. Lin (University of Manitoba)	
UMass/Hughes: Description of the CIRCUS System Used for MUC-5	277
W. Lehnert, J. McCarthy, S. Soderland, E. Riloff, C. Cardie, J. Peterson, and F. Feng (University of Massachusetts), C. Dolan and S. Goldman (Hughes Research Laboratories)	
Description of the LINK System Used for MUC-5	293
S. Lytinen, R. Burrige, P. Hastings, and C. Huyck (University of Michigan)	
USC: Description of the SNAP System Used for MUC-5	305
D. Moldovan, S. Cha, M. Chung, T. Gallippi, K. Hendrickson, J.-T. Kim, C. Lin, and C. Lin (University of Southern California)	
Sussex University: Description of the Sussex System Used for MUC-5	321
R. Gaizauskas, L. Cahill, and R. Evans (University of Sussex)	
 SUMMARIES OF TOPIC SESSIONS	
Summary of Workshop on Lexicons for Text Extraction (J. Pustejovsky, Brandeis University)	339
Report from the Text Analysis Techniques Topic Session (R. Stumberger, Language Systems, Inc.) ..	341
Information Extraction for the Future (P. Jacobs, GE Research and Development)	343
Topic Session on Discourse (D. Ayuso, BBN Systems and Technologies)	345
Tools and Techniques for Rapid Porting (J. McCarthy, University of Massachusetts)	347
Information Extraction and Evaluation (L. Rau, GE Research and Development)	349
 APPENDICES	
Appendix A: System Walkthrough Reference Materials	353
Appendix B: MUC-5 Test Scores	362