

TIPSTER TEXT PROGRAM

PHASE I

Proceedings of a Workshop held at

Fredricksburg, Virginia

September 19-23, 1993

Sponsored by:

Advanced Research Projects Agency

This document contains copies of reports prepared for the ARPA TIPSTER Text Program - Phase I Workshop. Included are reports from ARPA sponsored program and other materials prepared for use at the workshop.

**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 1-55860-337-9
Printed in the United States of America

Proceedings of the TIPSTER Text Program

Phase I

I. Overview

A.	TIPSTER Phase I: Program Overview	1
	<i>Roberta H. Merchant</i>	
B.	The TIPSTER Program	3
	<i>Thomas H. Crystal</i>	
C.	The Message Understanding Conferences	5
	<i>Beth M. Sundheim</i>	
D.	The Text Retrieval Conferences	7
	<i>Donna Harman</i>	

II. Document Detection

A.	Overview	9
	<i>Donna Harman</i>	
B.	Data Preparation	17
	<i>Donna Harman</i>	
C.	Summary of Results	33
	<i>Donna Harman</i>	
D.	System Overviews	
1.	INQUERY System - University of Mass/MCC	47
	<i>John Broglio, James P. Callan, W. Bruce Croft</i>	
2.	TIPSTER Phase I Final Report - HNC, Inc.	69
	<i>Bill Caid, Stephen Gallant, Joel Carleton, David Sudbeck</i>	
3.	DR-LINK System: Phase I Summary - Syracuse University	93
	<i>Elizabeth D. Liddy, Sung H. Myaeng</i>	
4.	TRW Japanese Fast Data Finder - TRW	113
	<i>Matt Mettler</i>	

III. Information Extraction

A. Overview	117
<i>Mary Ellen Okurowski</i>	
B. Data Preparation	
1. Tasks, Domains, and Languages for Information Extraction	123
<i>Boyan Onyshkevych, Mary Ellen Okurowski, Lynn Carlson</i>	
2. Corpora and Data Preparation for Information Extraction	135
<i>Lynn Carlson, Boyan Onyshkevych, Mary Ellen Okurowski</i>	
3. Template Design	141
<i>Boyan Onyshkevych</i>	
C. System Evaluation	147
<i>Beth M. Sundheim</i>	
D. System Overviews	
1. An Analysis of the Joint Venture Japanese Text Prototype and Its Effect on Systems Performance - ORD	165
<i>Steve Maiorano</i>	
2. Comparing Human and Machine Performance for Natural Language Information Extraction: Results from the TIPSTER Text Evaluation - IDA	179
<i>Craig A. Will</i>	
3. BBN's PLUM Probabilistic Language Understanding System - BBN	195
<i>Ralph Weischedel, Damaris Ayuso, Sean Boisen, Heidi Fox, Tomoyoshi Matsukawa, Constantine Papageorgiou, Dawn MacLaughlin, Tsutomu Sakai, June Abe, Hiroto Hosihi, Yoichi Miyamoto, Scott Miller</i>	
4. The TIPSTER/SHOGUN Project - GE/Martin Marietta/CMU	209
<i>Paul S. Jacobs, George Krupka, Lisa Rau, Michael L. Mauldin, Teruko Mitamura, Tsuyoshi Kitani, Ira Side, Lois Childs</i>	
5. The DIDEROT System - CRL/Brandeis	223
<i>Jim Cowie, Louise Guthrie, Wang Jin, William Ogden, James Pustejovsky, Rong Wang, Takahiro Wakao, Scott Waterman, Yorik Wilks</i>	
6. Description of the CIRCUS Systems Used for TIPSTER Text - UMASS/HUGHES	241
<i>W. Lehnert, J. McCarthy, S. Soderland, E. Riloff, C. Cardie, J. Peterson, F. Feng</i>	
Dictionary Construction by Domain Experts	257
<i>Ellen Riloff, Wendy G. Lehnert</i>	

IV. APPENDIX I

TIPSTER/MUC-5 Information Extraction Test Scores