Proceedings

TIPSTER TEXT PROGRAM

PHASE III

October 1996 - October 1998

Sponsored by:

Defense Advanced Research Projects Agency

This document contains copies of reports prepared for the TIPSTER Text Program Phase III. Included are reports from government sponsored contracts and government representatives.

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 Defense Advanced Research Projects Agency or the United States Government.

Distributed by
Morgan Kaufmann
340 Pine Street, 6th Floor
San Francisco, CA 94104
http://www.mkp.com

Senior Editor: Denise E. M. Penrose
Assistant Editor: Marilyn Uffner Alan

Production and Manufacturing by
Omnipress, Inc.
Post Office Box 7214
Madison, WI 53707-7214

Production Manager: Jodi Hauck, Omnipress

Copyright © 1999 Defense Advanced Research Projects Agency
ISBN 1-55860-610-6
ISSN 1525-5328

Table of Contents

A –THE TIPSTER PROGRAM

The TIPSTER Text Program Overview 3
F. Ruth Gee (Office of Advanced Analytic Tools)

TIPSTER Phase III Accomplishments 7
F. Ruth Gee (Office of Advanced Analytic Tools)

B – ARCHITECTURE DEVELOPMENTS

TIPSTER Lessons Learned: The SE/CM Perspective 17
Harold Corbin, Aaron Ternin (Litton/PRC)

The Common Pattern Specification Language 23
Douglas E. Appelt (SRI International), Boyan Onyshkevych (Department of Defense)

C – TIPSTER PROJECTS: A GOVERNMENT PERSPECTIVE

Project Penlight – A Government Perspective 33
Michael J. Chrzanowski (Department of Defense)

Project Underline – A Government Perspective 35
Michael J. Chrzanowski (Department of Defense)

The Cornell TIPSTER Phase III Project 37
F. Ruth Gee (Office of Advanced Analytic Tools)

The SRI TIPSTER III Project 39
Steven Maiorano (Office of Advanced Analytic Tools)

Reflections of Accomplishments in Natural Language Based Detection and Summarization 41
Susan R. Viscuso (Office of Advanced Analytic Tools)

D – INFORMATION EXTRACTION

Coreference Resolution Strategies from an Application Perspective 45
Lois C. Childs, David Dadd, Norris Heintzelman (Lockheed Martin Corporation)

Extracting and Normalizing Temporal Expressions 51
Lois C. Childs, David Cassel (Lockheed Martin Corporation)

Table of Contents

Research in Information Extraction: 1996-1998 57
Ralph Grishman (NYU)

Information Extraction Research and Applications: Current Progress and Future Directions 61
Andrew Kehler, Jerry R. Hobbs, Douglas Appelt, John Bear, Matthew Caywood, David Israel, Megumi Kameyama, David Martin, Claire Monteleoni (SRI International)

Algorithms That Learn to Extract Information – BBN: TIPSTER Phase III 75
Scott Miller, Michael Crystal, Heidi Fox, Lance Ramshaw, Richard Schwartz, Rebecca Stone, Ralph Weischedel (BBN Technologies)

Japanese IE System and Customization Tool 91
Chikashi Nabata, Satoshi Sekine, Roman Yangarber (NYU)

Transforming Examples into Patterns for Information Extraction 97
Roman Yangarber, Ralph Grishman (NYU)

E – INFORMATION RETRIEVAL

The Smart/Empire TIPSTER IR System 107
Chris Buckley, Janet Walz, Claire Cardie, Scott Mardis, Mandar Mitra, David Pierce, Kiri Wagstaff (Cornell University/SaBIR Research)

Dynamic Data Fusion 123
Ted Diamond, Elizabeth Liddy (TextWise LLC)

Improving English and Chinese Ad-Hoc Retrieval: TIPSTER Text Phase III Final Report 129
Kui-Lam Kwok (Queens College, CUNY, Computer Science Department)

Enhancing Detection Through Linguistic Indexing and Topic Expansion 139
Tomek Strzalkowski, Gees C. Stein, G. Bowden Wise (GE Corporate Research and Development)

SECTION F –TEXT SUMMARIZATION

Overview of the University of Pennsylvania's TIPSTER Project 151
Breck Baldwin, Thomas S. Morton, Amit Bagga (University of Pennsylvania, Duke University)

An NTU-Approach to Automatic Sentence Extraction for Summary Generation ... 163
Kuang-hua Chen, Sheng-Jie Huang, Wen-Cheng Lin, Hsin-Hsi Chen (National Taiwan University)

Improving Robust Domain Independent Summarization 171
Jim Cowie, Eugene Ludovik, Hugo Molina-Salgado (NMSU)

Table of Contents

Automatic Text Summarization in TIPSTER 179
Therese Firmin, Inderjeet Mani (Department of Defense, MITRE Corporation)

Summarization: (1) Using MMR for Diversity-Based Reranking and (2) Evaluating Summaries 181
Jade Godstein, Jaime Carbonell (Carnegie Mellon University, Language Technologies Institute)

Automated Text summarization and the Summarist System 197
Eduard Hovy, Chin-Yew Lin (University of Southern California, Information Sciences Institute)

Multiple & Single Document Summarization Using DR-LINK 215
Mary McKenna, Dr. Elizabeth Liddy (TextWise LLC)

A Text-Extraction Based Summarizer 223
Tomek Strzalkowski, Gees C. Stein, G. Bowden Wise (GE Corporate Research and Development)

G – EVALUATING THE TECHNOLOGIES

TIPSTER Information Extraction Evaluation: The MUC-7 Workshop 233
Elaine Marsh (Navy Center for Applied Research in Artificial Intelligence)

MUC/MET Evaluation Trends 235
Nancy Chinchor (SAIC)

The Text Retrieval Conferences (TRECs) 241
Ellen M. Voorhees, Donna Harman (National Institute of Standards and Technology)