

# THE LOCKHEED MARTIN TIPSTER II PROJECT

*Steven Maiorano*  
Office of Research & Development  
Washington, D.C. 20505  
E-mail: smaoran@ord.gov  
Telephone: 703-351-2701

## Introduction

The Lockheed Martin TIPSTER II project focused on several research areas. In addition, three demonstration projects are underway to exhibit the feasibility of transitioning the research into an operational setting.

## TIPSTER II Research Areas

Lockheed Martin's research accomplishments were threefold:

1. They defined a methodology for constructing domain-specific extraction systems in a cost-effective, reliable way. The MUC-6 evaluation system for the scenario template was constructed in three person weeks from the generic system and performed very well.

2. They provided insight into the benefit of particular techniques for integrating extraction and detection technology through a series of experimental results. The best of these experimental techniques will be integrated in the joint Lockheed Martin/GE CR&D/NYU/Rutgers submissions for TREC-5.

3. They contributed to defining a TIPSTER architecture, testing it, and implementing a multi-platform version of the document manager defined by the architecture. The document manager specifies a set of functions that should be used to get text for text processing and to store the output of text processing. Conforming to the TIPSTER architecture allows various extraction and detection modules to work together.

## The Generic Extraction System

To accomplish the goal of providing a core extraction capability that can be easily customized to a particular domain, Lockheed Martin developed a generic text processing system. The generic system incorporates their research in identifying a range of basic entities

in text together with a set of tools for easily constructing recognizers for new objects. In addition, the group has advanced the state of the art for identifying co-referential noun phrases. This work was evaluated during MUC-6 and performed extremely well. The ability to resolve co-references provides a sound basis for all forms of link analysis [1]

## The Integration of Detection and Extraction

A range of experiments were performed here to integrate extraction capabilities into the SMART retrieval system. This work was performed jointly with GE CR&D [2] and shows promising results for large-scale integration; The work will be integrated for the TREC-5 evaluation.

In addition, a range of experiments were performed to extend previous classification results [3] to routing, and to compare the resulting system with those used in TREC-4. The comparison will be described in the full paper and the results will be incorporated into a submission for the TREC-5 routing systems [4]

## The Architecture Working Group

Lockheed Martin has participated in the Architecture Working Group (AWG), and provided a testbed of the architecture functions being defined throughout the course of the contract. Lockheed Martin has implemented a document manager that accesses documents stored in a variety of databases. They have both a PC and Sun version at this time, and currently access information stored in either Sybase or Oracle. The NMSU document manager was also integrated into the Lockheed Martin text processing system (NLToolset) in a flexible way that will allow other document managers to be included easily.

*This material has been reviewed by the CIA. That review neither constitutes CIA authentication of information nor implies CIA endorsement of the author's views.*

## The Demonstration Projects

### ADEPT

The Advanced Data Extraction and Preparation via Tipster (ADEPT) Program is a demonstration project aimed at alleviating problems currently being faced by the Office of Information Resources (OIR). OIR has been chartered to implement enterprise-wide systems to collect, archive, distribute, and manage information obtained from unclassified data sources typically called "Open Sources" [5].

### CANIS

The Cable Abstracting and Indexing System Program is aimed at providing assistance to government analysts in indexing and abstracting their cable traffic. The goal of the project is to reduce the amount of manual time that an analyst will spend reviewing and processing each cable [6].

### FTM

The Free Text Management project is a demonstration system that will be located in the Federal Intelligent Document Understanding Laboratory (FIDUL); the project will include TIPSTER detection and extraction tools integrated with a common document manager.

## Summary

Overall, Lockheed Martin has demonstrated improvements in the technology to process texts and to do extraction, the ability to move the research into a realistic setting, and the ability to develop systems conforming to the architecture.

## References

[1] Childs, L.; "An Evaluation of Coreference Resolution Strategies for Acquiring Associated Information;" in this volume.

[2] Guthrie, L., Strzalkowski, T., Fang Lin, and Jin Wang; "Integration of Document Detection and Information Extraction;" in this volume.

[3] Guthrie, L., Walker, E., and Guthrie, J.; "Document Classification by Machine: Theory and Practice;" in Proceedings of the 16th International Conference on Computational Linguistics (COLING 94); Kyoto, Japan; 1059-1063; 1994.

[4] Guthrie, L. and Leistenstider, J.; "A Simple Probabilistic Approach to Classification and Routing;" in this volume.

[5] Kielty, J., and Sider, I.; "Advanced Data Extraction and Preparation via Tipster (ADEPT);" in this volume.

[6] Sider, I., Baker, J., Highbie, L., and Howard, T.; "Cable Abstracting and Indexing System (CANIS) Prototype;" in this volume