

# THE NYU TIPSTER II PROJECT

*Dr. Sarah M. Taylor*  
Office of Research and Development  
Washington, DC 20505  
E-mail: sarah@ucia.gov  
Telephone: 703-351-2565

## Introduction

NYU supported the TIPSTER Phase II effort in the development of the TIPSTER Architecture, with enhancements to both their Detection and Extraction systems, and with experiments in the combined use of Extraction and Detection for document retrieval.

## Architecture Effort

Dr. Ralph Grishman steered the Architecture Working Group, and led its subset, the Contractors' Architecture Working Group (CAWG), in the development of the Architecture Design and the implementation of that design in the two Architecture demonstrations, the first shown at the TIPSTER Phase II 6-Month Workshop in November 1994, and the second at the 12-Month Workshop in May 1995. He continues to lead these two groups through the period of revisions to the Architecture as a result of its use in TIPSTER demonstration projects and elsewhere.

## Enhancements to Extraction

NYU's Proteus extraction system was modified to use combined syntactic and semantic pattern based methods. The results were demonstrated in MUC-6, where the modified system achieved excellent results, particularly in the Coreference Task. The results of this work, which incorporates much that has been learned from other TIPSTER participants using this method, increase our understanding of the role of syntax in Information Extraction and will be used to further our progress toward the goal of an extraction system which can be configured for a task by the user.

*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.*

## Enhancements to Detection

TIPSTER Detection research at NYU has been guided by Dr. Tomek Strzalkowski, who began at NYU, but moved to GE Corporate Research and Development in January 1995. NYU uses a modified version of NIST's Prize engine in its experiments. The NYU system has been modified to be TIPSTER Architecture compliant and has participated in TREC-3 and TREC-4 under the TIPSTER contract. TREC Performance of this system, which uses a NLP module to enhance the Prize statistical core, has steadily improved, not only measured against itself, but in relation to other participating systems. The NLP module is used to identify appropriate multi-word phrases for use in indexing and in processing the user's natural language search requests.

## Experiments in Combined Detection and Extraction

In the final six months of TIPSTER Phase II, the combined NYU system, using the TIPSTER Architecture to enable integration of the Detection and Extraction systems, will be used to experiment with ways of using Extraction to improve Detection. This work will be focussed on improving the user's ability to identify news articles which contain information such as: what a certain person has said about a certain topic, who of significance has travelled to specific places, who has specified kinds of relationships with certain kinds of companies. TREC data will be used in these experiments. One goal of the experiments is to contribute to the on-going TIPSTER work in the use of combined Detection and Extraction systems. Another goal is to understand better what our users need from their retrieval systems and demonstrate an ability to provide retrieval improvements which are meaningful to them.