Project Penlight - A Government Perspective

Michael J. Chrzanowski
Department of Defense
9800 Savage Road, Fort Meade, MD 20755
mjchrza@romulus.ncsc.mil
(301) 688-9131

The purpose of the TIPSTER contract with the University of Pennsylavnia (Project PENLIGHT) was to promote and further develop Information Extraction and Text Summarization by the application of coreference resolution, an integral part of IE systems, to Text Summarization.

During the first two years of this effort, which began on 1 October 1996, the University of Pennsylvania was charged with two parallel tasks. The greater part of its effort was to go into the design and implementation of the coreference engine. The second task, essentially an independent effort, was to automatically detect discourse segments.

The University of Pennsylvania has been very successful in developing the coreference engine. The work in this area has been recognized as valuable by other researchers in the Summarization effort. As a result, the University will be making its coreference software (known as CAMP) available to the research community on the World Wide Web. The potential importance of this work as a whole to the automatic text summarization effort has caused the Government to slighly re-align its priorities over the course of the first two contract years. The Government instructed the University of Pennsylvania to spend almost all of its efforts in improving the coreference resolution engine and making it available to the research community rather than applying this engine to the problems of text summarization.

While the results of the coreference resolution development work have been impressive, the results of the application of this effort to text summarization have also achieved success, particularly in the area of query-specific summarization. The University of Pennsylvania approach is to resolve all coreferences within the full text to the query and select all sentences which are coreferent. Results from the dry-run and SUMMAC evaluations imply that coreference resolution is a necessary element of good summarization.

The University of Pennsylvania continues its efforts in resolving cross-document coreference which will be necessary to perform multi-document summarization.