

Project Underline - 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 Carnegie Group, Inc. (CGI) of Pittsburgh, PA is to promote and further develop automatic Text Summarization using a Maximal Marginal Relevance (MMR) metric to generate summaries of documents that are directly relevant to the information need of an individual user. CGI subcontracts with Carnegie Mellon University to perform most of its linguistic research.

CGI extended the MMR metric and its use in re-ranking documents and subdocuments according to the combined criteria of query relevance and maximal novel information from previously selected documents or subdocuments. MMR refinement included: 1) parametric analysis, 2) testing in the context of a functioning named IR system, and 3) modification or extension of the metric as appropriate. CGI also adapted the MMR metric to support sub-document MMR ranking, where the subdocuments are all parts of one source document. Finally, CGI provided two prototypes of the MMR-based text summarization system.: a developmental prototype and an operational prototype.

The developmental prototype (Scout DP-1) provided the ability to query using specific free text or another document, provided a list of documents meeting that query from the IR system, and summarized documents linking the selected sentences back to the original text. The delivery of the operational prototype (Scout OP-1) in late August 1998 which extended the capabilities of DP-1, improved the user interface and added a keyword in context index feature offered the first real opportunity for end users to interact with the automatic text summarization system and provide comment.

Carnegie Group Inc. participated in both the dry run and SUMMAC evaluations. CGI's performance was encouraging and relied not only on the MMR metric but on traditional sentence selection methods, such as, always select the first sentence of a document.

The MMR metric has proven to be extremely flexi-

ble and has application in many facets of retrieval and summarization. MMR is expected to prove most useful in the summarization of multiple documents.