EVALUATING TEXT UNDERSTANDING SYSTEMS

Beth M. Sundheim

Naval Ocean Systems Center
Code 444
San Diego, CA 92152-5000

PROJECT GOALS

The Naval Ocean Systems Center is extending the scope of previous efforts in the area of evaluating English text analysis systems and is seeking to refine the methodology in order to obtain performance benchmarks on an information extraction task for recall, precision, overgeneration, and fallout for a variety of systems. The methodology is also intended to enable the collection of qualitative data on the relative validity of the text analysis techniques as applied to the task of information extraction.

RECENT RESULTS

The third evaluation began in October, 1990; a dry-run phase was completed in February, 1991. Twelve sites reported results for the dry-run test at a meeting held on 13-15 February, 1991. The test required extracting information on terrorist incidents (incident type, date, location, perpetrator, target, instrument, outcome, etc.) from relevant messages in a blind test on 100 previously unseen texts in the test set. The results of this test are summarized in a paper found elsewhere in this volume.

PLANS FOR THE COMING YEAR

Official testing will be done in May, 1991, and the Third Message Understanding Conference (MUC-3) will be held May 21-23 at the Naval Ocean Systems Center. A proceedings will be published on the basis of this conference. The results of the evaluation will be analyzed to discover whether conclusions can be drawn concerning the correlation among task performance, text analysis capabilities, and theoretical approach.

In addition to the official measures, unofficial measures will be obtained of performance on particular linguistic phenomena (e.g., conjunction), as measured by the information extracted in particular sets of instances. That is, text segments exemplifying a selected phenomenon will be marked for special scoring if successful handling of the phenomenon seems to be required in order to fill one or more template slots correctly for that segment.