Data Access for Situation Handling (DASH)

J.R.Hobbs, Principal Investigator

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park CA 94025

PROJECT GOALS

The objective of this research is to design a template specification language for representing situations and creating data structures for storing information extracted from texts. The language will be used to create structured descriptions of those aspects of situations about which information is required. These structured descriptions will function to

- characterize the desired information

- guide the process of extracting information from texts

- provide a representation that can be stored in a database for later access

Experience in developing computer systems to extract data from text has shown that determining how to structure the information for storage in a database and later use can be as significant a problem as configuring a natural language system to process texts and extract the information. This should not be surprising, since defining the representation structures is a special instance of the problem of knowledge representation, which is an open problem of substantial depth.

RECENT RESULTS

The main part of our effort in the first phase of DASH has been close analysis of the templates from MUC-4, MUC-5, and the Warbreaker Message Handling Tasks. We have also attempted to provide templates for the a small set of TREC topic descriptions and narratives. This analysis has been aimed at explicating general principles of template design that might be presented in a manual for template designers. Specifically we have proposed principles involving the following aspects of template design:

1. The choice of basic ontology

2. Issues of temporal granularity

3. Criteria for choosing among alternative representations of relations

4. How to represent events

5. Entities vs. Entity Snapshots—time-slices of parts of entities

6. The nature of slot fills

7. The status of backpointers

We shall present a report on this research at the Human Language Technology Workshop.

PLANS FOR THE COMING YEAR

We plan to continue with our specifications of templates for TREC topics, with special emphasis on issues involving the representation of spatial and temporal relations and economic facts. We shall be looking for generalizations that might generate further principles of template design. We shall be compiling these principles, with examples of their application, in a preliminary version of a manual for template designers.