

SKIMMING NEWSPAPER STORIES BY COMPUTER

Gerald DeJong
Computer Science Department
Yale University
New Haven, Connecticut 06520

INTRODUCTION

FRUMP (Fast Reading Understanding and Memory Program) is a system being developed at Yale to skim newspaper stories. The United Press International news service has recently been connected to the Yale computer to provide FRUMP with real-world data. FRUMP's understanding is based on the concept of a script. A script is a data structure used to represent knowledge about very stylized events.

FRUMP uses what we call "sketchy scripts". The crucial difference between sketchy scripts and the detailed scripts used in other Yale projects [Cullingford, 1977] is that only the concepts corresponding to the most important events in a detailed script are included. Each of these important events is represented by a "request" in conceptual dependency format. The requests contain blank slots that are filled when the story is processed. There is a different sketchy script for each type of newspaper story that FRUMP can process. The blank slots are the script variables for that script. At this writing FRUMP can understand six types of news events: earthquakes, vehicle accidents, heads of states dying, floods, military fighting and meetings of heads of state. We expect FRUMP to have many more scripts by the end of the summer. For FRUMP, understanding is marking the proper requests as satisfied and binding its script variables to things found in the text.

FRUMP is made up of two parts: the conceptual fragment parser and the script applier. The parser was inspired by Becker's phrasal lexicon [Becker 1975]. It tries to satisfy the requests of the sketchy script with sentences or fragments of sentences from the input. The script applier guides the parser in selecting meanings for ambiguous phrases, binds script variables, and loads new requests when appropriate. The top down nature of FRUMP allows it to be very efficient. A news story is typically processed in ten seconds of CPU time on a DEC KA10 processor.

CLASSIFICATION 0 NEWS STORIES

A file is maintained of understood news events. Each file entry consists of the requests and bound script variables of the sketchy script used for understanding that event. Before a

The research described in this paper was supported by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under contract N00014-75-C-1111.

story can be processed, FRUMP must classify the story (to select the proper sketchy script) and decide whether it is an update of a previous story or a report of a new event. This is done by making an initial scan through the first paragraph of the story with a special set of requests active. The set of requests is made up of one key request from each of FRUMP's scripts. A key request is a request which, if satisfied, strongly indicates that the associated script is appropriate for understanding the story. When one of these requests is satisfied, FRUMP knows not only what script to use but also (from the script variables of the request) some information about this particular story. That information is matched against the corresponding information of previously understood stories to decide if the new story is an update. For a more complete description of FRUMP, see [DeJong, 1977].

AN EXAMPLE

The following consists of the lead paragraphs of an actual article. The summaries are produced by canned templates. The templates contain blanks filled by properties of the script variables. Since the internal representation is conceptual dependency, it is as easy to generate summaries in Spanish and Russian as English.

INPUT:

3 - 4 PISA, ITALY - - OFFICIALS TODAY SEARCHED FOR THE BLACK BOX FLIGHT RECORDER ABOARD AN ITALIAN AIR FORCE TRANSPORT PLANE TO DETERMINE WHY THE CRAFT CRASHED INTO A MOUNTAIN KILLING 44 PERSONS.

THEY SAID THE WEATHER WAS CALM AND CLEAR, EXCEPT FOR SOME GROUND LEVEL FOG, WHEN THE U S MADE HERCULES C130 TRANSPORT PLANE HIT MT. SERRA MOMENTS AFTER TAKEOFF THURSDAY.

THE PILOT, DESCRIBED AS ONE OF THE COUNTRY'S MOST EXPERIENCED, DID NOT REPORT ANY TROUBLE IN A BRIEF RADIO CONNVERSATION BEFORE THE CRASH.

ENGLISH SUMMARY: 44 PEOPLE WERE KILLED WHEN AN AIRPLANE CRASHED INTO A MOUNTAIN IN ITALY.

SPANISH SUMMARY: HUBO UN ACCIDENTE DE AVION EN ITALIA QUE RESULTO EN 44 MUERTOS.

RUSSIAN SUMMARY: V AVIATIONNOI KATASTROFE ITALII 44 CHELOVEK BYLO UBITO.

REFERENCES

- Becker, J.,
"The Phrasal Lexicon", Proc. Theoretical Issues In Natural Language Processing Workshop, Cambridge, Mass. June, 1975.
- Cullingford, R.,
"Organizing World Knowledge for Story Understanding by Computer", Ph.D. Thesis, Engineering and Applied Science Dept., Yale University, 1977.
- DeJong, G.,
"Skimming Newspaper Stories by Computer", Research Report 104, Computer Science Department, Yale University, May 1977.