An interactive text-editing system in support of Russian translation by machine

by DAVID A. LUTHER

USAF/RADC
Rome, New York

and

CRISTINE MONTGOMERY and RONALD M. CASE

Operating Systems, Inc.
Woodland Hills, California

ABSTRACT

An interactive, text-editing system was designed and built to support the pre- and post-processing of machine translated scientific and technical Russian literature.

Sixteen independent CRT/keyboard terminals are supported on five small processors with a distributed data base and distributed processing.

A "free-form" text editor has been provided for the creation or modification of textual files. Unusually powerful functions for technical editors have been implemented such as "restore text" and "alternate word list." Software character generation is used to display any of a mix of 256 symbols from four logical alphabets.

INTRODUCTION

The USAF Foreign Technology Division (FTD) and Rome Air Development Center (RADC) have collaborated for some time in the development and application of computer or machine translation systems. They have been specifically applied to the conversion of foreign scientific and technical literature into English. The quality of the current translation system is under constant improvement and has reached a satisfactory level. However, an analysis of the current operational environment accentuated the need for total system improvement, with focus on automating the supporting functions of input and output, which contribute most heavily to total machine translation costs.

In this paper we describe an interactive, text-editing system designed and built to support the keyboarding of cyrillic text, and the editing and composition of machine translated English. Sixteen display/keyboard terminals have access to sophisticated text manipulation functions and can display any of a repertoire of 256 symbols. A high level of interaction is maintained by a unique system of five small distributed processors.

FUNCTIONAL DESCRIPTION

The machine-translation process divides into three functional areas: input or keyboarding of foreign material, translation by computer software, and editing and recomposition. The results of a technical analysis served to document the feasibility of a machine-aided editing system oriented to the use of CRT display devices used to interact directly with the machine-translated output. A later study reported on user reaction to an experimental text-editing terminal installed within the FTD operational environment. In conjunction with the ultimate users, requirements were identified and specifications developed.

The resulting edit system was designed to provide the means for 16 independent users at different editing stations to create, view and edit different documents at CRT/keyboard terminals. Transfer to and from the translation system computer, an IBM 360/65, is via magnetic tape. The software required to support all editing and text manipulation is organized on two levels. At the bottom level, the majority of the editing functions are supported on four identical IMLAC PDS-4 processors; each processor supports four terminals. Above them is a PDP 11/05 which supports complex editing functions, file handling and peripherals.

FUNCTIONAL CAPABILITIES

The edit system is logically organized such that one of the terminals is designated as the system operator. While so designated, the user at this terminal can have access to system level commands such as assigning individuals to terminals, for instance. This function, like all others, is completely terminal oriented; that is, all commands and edits, all actions and interactions take place through the terminal.

End-user requests and commands at a terminal fall into
three modes: command string, edit and review modes. Command string mode deals with start-up and utility functions. Its syntax has a traditional form, i.e., a Command Word and a parameter(s).

The Edit mode is the most interesting and powerful of the three. It is provided for the creation and modification of textual files. The CRT display is a 1000 character "window" on a scroll of text. The text is displayed on 80 character lines in "free-form," characteristic of technical documents. All changes are made directly and immediately on the text displayed in the window using the keyboard for cursor positioning (pointing at the text), function selection and literal input.

In Review mode a specified document will be displayed at the requesting terminal but no editing will be permitted. More than one terminal may be reviewing the same document, and each terminal may be reviewing a different portion of the document.

EDIT MODE

As originally envisioned, the edit system was to be used primarily to edit and correct text material produced by the machine translation process. Its use here by the human operator or translator would be necessarily restricted to scanning and correcting functions. However, the projected use of the system evolved and grew to include human translation and cyrillic input; both applications stressing text input or creation. The combined functional requirement grew thereby to cover a gamut of editorial and composition functions.

When in Edit Mode, the user will normally find his terminal in an "insert" status. That is, any literal key that is struck will put a character on the display at the position indicated by the cursor. The cursor is positioned by way of cursor control keys on the keyboard. If the cursor is in the midst of existing text, the file will be opened up and the characters to right of the cursor will be pushed to the right. Full word wraparound (preservation of word integrity) will occur at the right margin; the word pushed off will be inserted at the beginning of the next line, and so on.

Characters and words can be deleted by pointing at the character or any part of the word, respectively, followed by pushing the appropriate function key. Another character deletion function, called rubout, will remove the character to the left of the cursor. This is extremely useful for the occasional slip of the finger while creating text. Any material deleted by any method is replaced by a special null character, displayed as a "bullet." Upon execution of the close-up function, the nulls will be removed and the text file closed-up from the bottom. This approach is much more satisfying than a text file which is changed for every occurrence of a delete.

A special capability to restore a line of text following the execution of an editing function is available. The line will be restored to what it was just prior to the last edit. This capability is intended for recovery from an inadvertent edit or for that user who suddenly changes his mind.

SPECIAL FUNCTIONS

Two powerful functions have been included that are particularly useful for editing machine translations. They are the Select and Alternate Word List functions.

By positioning the cursor under one word of a series of embedded alternate words supplied by the machine translation algorithm, and striking the select key, the user can select one word to remain and cause all the others to be deleted. This capability puts the human translator back into the loop to make difficult choices which may be beyond the capability of an automatic translation algorithm to perform.

The alternate word list allows each individual to set up a special dictionary consisting of word pairs, one of which is the string to be replaced and the other is the replacing string. During an editing session, a user need only point to an occurrence of a string to be replaced, strike the alternate word function key and a swap will be made.

HARDWARE/SOFTWARE IMPLEMENTATION

The programs which support all of the highly interactive tasks are running on four identical IMLAC PDS-4 processors. On any one IMLAC the program is shared by four keyboard/display units. Each unit has one-fourth of the display refresh capacity as well; display refresh in the IMLAC is handled by a separate processor. One significant reason the IMLAC was selected is the fact that characters are generated in software. This made possible the definition of 2 or 256 separate symbols. They are divided into four groups: roman, cyrillic, Greek and math/technical. A mode key on the keyboard switches the way in which a key strike is interpreted. Combinations of the four alphabets may be mixed on the display.

Each IMLAC is connected by a DMA interface to the PDP 11/05. The PDP is called upon to load and unload files, for execution of complex functions such as move and copy, and for mass memory storage. A scroll at the IMLAC beyond its buffer limit will cause a request to be sent to the PDP for data.

This configuration is a true example of division of labor. Considering the tasks to be supported, the processors are very small and are being used to their maximum. However, under no circumstances to date has there been anything less than immediate response, that is, no delay apparent to the user.

REFERENCES


From the collection of the Computer History Museum (www.computerhistory.org)