

MACHINE TRANSLATION WORK BY THE COMPUTER ELLIOTT 803

by M. Stein

I wish here to give an account of the solution of some problems connected with MT by the computer ELLIOTT 803.

The computer ELLIOTT 803 has the advantage over other available machines, that it has, on the one hand, a memory with larger capacity, and a machine word containing 39 bits, on the other.

Transliteration

The words are punched on tapes. Before punching the Russian text must be transliterated, i.e. the Cyrillic alphabet must be transcribed into letters used by Siemens. We use the following transliteration:

а	а	л	l	ч	4
б	б	м	m	ш	ε
в	v	н	n	щ	w
г	g	о	o	ъ	5
д	d	п	p	ы	y
е	e	р	r	ь	6
ё	e	с	2	э	3
х	x	т	t	ю	q
э	z	у	u	я	9
и	i	ф	f		
и	j	х	h		
к	k	ц	c		

Input

Six bits come up for each of the signs in the words of the machine. The sixth sign (from right to left) is one or zero, accordingly to whether the sign in question is a letter or a figure. So a bit-combination of six signs can be put in a machine word. The remaining three bits may be used to store different data.

As the longest word may consist of 18 letters at most, three cells are required (maximum) to store one word. In the cell the 37th bit is one, if the content of the following cell is a continuation of the preceding one.

The words with different length are placed in 18 different groups: B_1, B_2, \dots, B_{18} . These are the initial addresses of cells, which the words consisting of 1, 2, ..., 18 letters, respectively, are put in.

The place in the storage of the given word can be computed,

$$B_j + d I_j \quad (1 \leq j \leq 18)$$

where $d = 1, 2$ or 3 accordingly to whether the word in question needs 1, 2 or 3 cells. It shows how many words, consisting of 1, 2, ..., 18 letters, have already been put in.

Storage

The words of the different groups are arranged in growing order. This arrangement makes the search of words easier.

The Hungarian words are put in a separate place. The sequence of the ordered Russian words determine the order of the Hungarian ones.

One extra cell belongs to every Russian word. This cell contains the different data of the word in question. (For instance, the code of the form class.) One part of this cell is used for storing the address of the adequate Hungarian word.

13 bits are required to determine an address. So 26 bits remain for other data. From this "informational cell" any needed information can be separated by logical multiplication by different constants.

Dictionary look up

First, it shall be determined, how many letters the unknown word consists of. After, the unknown word shall be looked up only in the group of words with given length.

The searching of words is executed by a bisecting method, i.e., if we denote the unknown word by S , and the words of the same length as S , by S_1, S_2, \dots, S_n , and if $S \ll S_n$, so it must be decided, which of the following three formulae holds true:

- $$\begin{aligned} \text{i/ } S &= S_j & /j &= \frac{n}{2} \\ \text{ii/ } S &< S_j \\ \text{iii/ } S &> S_j \end{aligned}$$

If ii/ holds true, the same examination must be carried out for $j = j - j/2$, and if iii/ holds true, for $j = j + j/2$.

This procedure must be continued until i/ holds true for any j .

If there does not exist a j which satisfies the above requirement then the next letter from the end of the word is to be cut. The procedure begins all over again.

Morphological analysis

Dictionary look up is at the same time the starting point of morphological analysis. On the basis of the endings which have been cut, the morphological class of the word under consideration can be determined. Some ambiguities can only be solved after a syntactic analysis have been carried out. The checking of programs concerning syntactical analysis has just begun.