Attempt at Total Input of Records

Suitō Makoto

This “attempt at a total input of records” aims to utilize the full text of diaries and compilations for retrieval and statistics by inputting them into a computer. This is an attempt to be able to know immediately, for example, in what article of what date the word “festival” appears in a diary, by using a computer.

It is natural that diaries so far written have a variety of characteristics. Some are suitable for input, but others are not. For example, the length of sentences written in diaries differs day by day. Some were rewritten or inserted after the original was written, and some include Chinese characters which are not used today. Therefore, it is not simply a matter of inputting sentences. The first barrier was the arrangement of conditions for input.

The second one was the difficulty of computer operation.

Anyway, the following two goals were set, and some barriers overcome.

The system must be easy to operate so that anybody can use it.

Data must be retrievable by any optional character string, without preparing keywords in advance. At last, the system started operating: But once it started, new problems occurred, for example, the time required for retrieval. At first, the ultimate goal was to be able to retrieve data by any optional character string for any length of time. As a result, the system was found to take 12 or 13 minutes for a retrieval from 400,000 characters. I was excited when I succeeded in retrieving data for the first time, but this length of time was a considerable barrier to the repeated retrieval of data by more than one word. Then, the curtailment of retrieval time became an important task. Now retrieval time has been reduced to only one or two minutes. This paper is the record of this hard struggle.