



ABSTRACT

RECORDING MEDIUM OF CHARACTER DATA OF FULL TEXT AND
CHARACTER STRING COLLATING APPARATUS

All two-character chains including two general characters and all three-character chains including one special character between two general characters are detected from a registration character string in which a large number of special characters not having any meaning are frequently arranged, or all two-character chains including two general or symbolic characters are detected from a converted registration character string produced by changing each special character of the registration character string to one type of symbolic character determined in correspondence to one general character adjacent to the special character. Also, occurrence frequencies of the general or symbolic characters of each chain are counted and stored in a recording medium with the registration character chains. When a retrieval character chain is input, occurrence frequencies of particular character chains corresponding to all retrieval character chains detected from the retrieval character string in the same manner are read out from the recording medium and are collated with each other, and a particular character string agreeing with the retrieval character chain is retrieved from the registration character string. Because an occurrence frequency of any special character is not counted or the special characters are changed to various types of symbolic characters, a recording area required for the occurrence



frequencies of the registration character chains can be reduced.

Figure 4.