ABSTRACT

METHODS AND SYSTEMS FOR IMPLEMENTING APPROXIMATE STRING MATCHING WITHIN A DATABASE

A computer-based method for character string matching of a candidate character string with a plurality of character string records stored in a database is described. The method includes a) identifying a set of reference character strings in the database, the reference character strings identified utilizing an optimization search for a set of dissimilar character strings, b) generating an n-gram representation for one of the reference character strings in the set of reference character strings, c) generating an n-gram representation for the candidate character string, d) determining a similarity between the n-gram representations, e) repeating steps b) and d) for the remaining reference character strings in the set of identified reference character strings, and f) indexing the candidate character string within the database based on the determined similarities between the n-gram representation of the candidate character string and the reference character strings in the identified set.

Fig. 4