

DATA IDENTIFICATION METHOD, APPARATUS, DEVICE, AND READABLE MEDIUM

ABSTRACT

Implementations of the present specification disclose a data identification method, apparatus, device, and a computer-readable medium. A solution includes: obtaining a first data set, data samples in the first data set being at least a part of data of a to-be-identified field; obtaining a state transition matrix set generated based on statistics of data samples in a second data set, a data type of the data samples in the second data set being known; determining sample state transition probabilities corresponding to the data samples in the first data set based on the state transition matrix set; determining a ratio between a number of data samples in the first data set whose sample state transition probabilities are greater than a first threshold and a total number of the data samples in the first data set; and determining data corresponding to the to-be-identified field as being of a same data type as the data samples in the second data set in response to that the ratio is greater than a second threshold.

Fig. 1