

METHOD, APPARATUS, DEVICE, AND READABLE MEDIUM FOR IDENTIFYING PRIVATE DATA

ABSTRACT

Implementations of the present specification disclose a method, apparatus, device, and a readable medium for identifying private data. A solution includes: obtaining a first length distribution vector and first character distribution statistical information of a first data set, the first data set being a set of private data samples; obtaining a second length distribution vector and second character distribution statistical information of a second data set, the second data set being a set of at least a part of data samples corresponding to a to-be-identified field; calculating a length distribution similarity and a character distribution similarity between the first data set and the second data set, and determining whether data corresponding to the to-be-identified field is private data of a same type as the first data set based on the length distribution similarity and the character distribution similarity.

FIG. 1