

METHODS AND APPARATUSES FOR ESTIMATING WORD SEGMENT FREQUENCY IN DIFFERENTIAL PRIVACY PROTECTION DATA

ABSTRACT

5 This specification provides a method and an apparatus for estimating a word
segment frequency in differential privacy protection data, and an electronic device.
According to the method, each piece of word segment information that is reported by a
terminal device and that is subject to local differential privacy processing is obtained;
N groups of word segment information are obtained through division, so each piece of
10 word segment information of the same group corresponds to the same target quantity;
each group of estimated data that is corresponding to each group of word segment
information and that represents unbiased word segment frequency estimation is
determined; and each layer of nodes of a prefix tree used to record a word segment
frequency is generated layer by layer based on each group of estimated data. In a
15 process of generating an nth layer of nodes of the prefix tree, some candidate n-tuple
word segments are selected as n-tuple word segments represented by the nth layer of
nodes, and it is not necessary to traverse all n-tuple word segments formed by
predetermined word units. This greatly reduces a calculation amount and improves
calculation efficiency, and the n-tuple word segments represented by the nth layer of
20 nodes and selected based on frequency salient distribution information of word
segments are more reasonable.

[Fig. 2]