

CLAIMS

1. A method for combating malicious information based on order clustering, the method comprising:
 - generating a unique order ID and an order content based on an information about an order (101, 301);
 - converting the order content to an order vector (102, 302);
 - performing a clustering operation based on the unique order ID and the order vector to obtain an order clustering result (103, 303); and
 - determining whether the content of the respective order in the cluster comprises malicious information by analyzing a content of a respective order in each cluster to (304), and
 - in response to determining that the content of the respective order comprises malicious information, performing automatic malicious information processing for the respective order in the cluster that comprises malicious information (304).
2. The method of claim 1, wherein the order content is an order randomly selected from the cluster.
3. The method of any one of claims 1 and 2, wherein the order comprises information transferred to a user from an external source.
4. The method of any one of claims 1 to 3, wherein generating the unique order ID and the order content comprises: generating the unique order ID based on a user ID, an order ID, and a service type in the information about the order; and generating the order content based on an order title or an order remark in the information about the order.
5. The method of claim 4, wherein the order title comprises a product description and an order description.
6. The method of claim 4, wherein the order remark comprises a promotional information of an external source initiating the order.

7. The method of any one of claims 1 to 6, wherein content of an order in the cluster is mapped to a characteristic string,
determining whether a mapped string of content of an order to be analyzed is the same as the characteristic string, and
in response to determining that the mapped string of content of an order to be analyzed is the same as the characteristic string, adding the order to be analyzed to the cluster.
8. The method of claim 7, in response to determining that the mapped string of content of an order to be analyzed is the same as the characteristic string, marking the cluster as a junk class.
9. The method of any one of claims 1 to 8, wherein determining whether the content of the respective order in the cluster comprises malicious information is based on a keyword.
10. The method of any one of claims 1 to 9, wherein performing automatic malicious information processing for the orders in the cluster that comprises malicious information comprises one or more of recording a user ID in the cluster, sending the user ID to an administrator for banning the user ID, sending alarm information to the user ID freezing the transaction or blocking the user account associated with the order.
11. The method of any one of claims 1 to 10, wherein the clustering operation comprises a mean-shift clustering algorithm that is a non-parametric multi-model segmentation method.
12. The method of any one of claims 1 to 11, further comprising importing the order to a Hadoop distributed file system (HDFS).
13. The method of any one of claims 1 to 12, wherein converting the order content to the order vector comprises: converting the order content to a term frequency–inverse document frequency vector.
14. An apparatus for order clustering in an electronic transaction, the apparatus comprising a plurality of modules configured to perform the method of any one of claims 1 to 13.