ABSTRACT

Thesis work: 140 pp., 39 fig., 16 tabl., 4 applications, 23 sources.

The object of study is the clustering methods. The subject of study is the event logs of web users.

The aim of the study:

1) research methods and algorithms for segmentation and its validation;

2) develop software that implements clustering algorithms;

3) obtaining segmentation recommended for zastuvannya.

Theoretical and methodological basis of the study is the work of foreign foreign scientists in the field of data mining, mathematical modeling, data segmentation and marketing.

During the thesis created software to determine the segments of users creating their descriptions, and present the results of the program on real data.

The methodology is implemented on the basis of already known algorithms and Kmeans Bisect Kmeans using their own development, which include non-linearities play session and use features of graph for weight characteristics.

The software is implemented using the Python programming language and framework for working with large data Apache Spark. Recommendations for further research.

DATA MINING, CLUSTERIZATION, SPARK, BIG DATA, SEGMENTATION, GRAPH THEORY, WEB-USERS.