

РЕФЕРАТ

Дипломна робота: 140 с., 39 рис., 16 табл., 4 додатки, 23 джерела.

Об'єктом дослідження є методи кластеризації. Предметом дослідження є журнал подій веб-користувачів.

Мета дослідження:

- 1) розробити необхідний інструментарій для перетворення журналу в дата сет для сегментації
- 2) дослідження методів та алгоритмів для сегментації та її перевірки;
- 2) розробка ПЗ, який реалізує алгоритми кластеризації;
- 3) опис сегментації, що рекомендується для застосування.

Теоретичною та методологічною основою дослідження є праці закордонних вчених в галузі інтелектуальної обробки даних, математичного моделювання, сегментації даних та маркетингу.

В ході дипломної роботи створено програмний продукт для визначення сегментів користувачів, створення їх описів, а також представлено результати роботи програми на реальних даних.

Методологія реалізована на основі уже відомих алгоритмів Kmeans та Bisect Kmeans з використанням власних розробок, що містять відтворення нелінійності сесії та використання особливостей графів для отримання вагових характеристик.

Програмний продукт реалізовано за допомогою мови програмування Python та фреймворку для роботи з великими даними Apache Spark. Надано рекомендації до подальших досліджень.

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, КЛАСТЕРИЗАЦІЯ, СПАРК, ВЕЛИКІ ДАНІ, СЕГМЕНТАЦІЯ, ТЕОРІЯ ГРАФІВ, ВЕБ-КОРИСТУВАЧІ.