

Аннотация дипломной работы

Тема: Методы кластеризации и визуализации данных

ФИО студента: Андреюк Илья Геннадьевич

Научный руководитель: Соболева Татьяна Валентиновна

Кафедра (специальность, специализация): многопроцессорных систем и сетей (информатика, программное обеспечение вычислительных систем)

45 страниц, 19 рисунков, 8 источников.

Ключевые слова: КЛАСТЕРИЗАЦИЯ ДАННЫХ, МАШИННОЕ ОБУЧЕНИЕ, ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ, КЛАСТЕРЫ, КЛАСТЕРНЫЙ АНАЛИЗ, РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ, ВИЗУАЛИЗАЦИЯ ГРАФОВ, GOOGLE MAPS API, ПОИСК РЕЗКО ВЫДЕЛЯЮЩИХСЯ ЗНАЧЕНИЙ

Цель работы: провести обзор существующих методов кластеризации и визуализации данных, разработать приложение, визуализирующее отношения между объектами.

Описание работы студента: в работе были рассмотрены понятия, цели и задачи рекомендательной системы, разработан рекомендательный алгоритм на основе оценок пользователей веб-сайта MovieLens.

На основе данных, полученных в результате работы алгоритма рекомендации, была получена визуализация данных в двумерном пространстве. В ходе работы был проведён сравнительный анализ двух силовых алгоритмов: Фрюхтермана-Рейнголда и Камада-Кавайи. В результате получено, что алгоритм Камада-Кавайи даёт более качественную визуализацию, чем алгоритм Фрюхтермана-Рейнголда.

Изучены часто используемые методы кластеризации данных с целью разработки приложения на основе работы рекомендательного алгоритма. В работе подробно описаны и проанализированы алгоритмы кластеризации, а алгоритм k -средних был успешно применён к визуализации на разных уровнях приближения.

Разработано веб-приложение и кластеризованные данные были отображены с применением движка *GoogleMapsAPI*.

Полученные результаты могут быть использованы для разработки рекомендательной системы. Преимуществом такой рекомендательной системы является то, что все данные находятся в одном централизованном месте, и пользователю такой системы нет необходимости пользоваться какими-либо другими источниками для получения рекомендаций.