

РЕЦЕНЗИЯ

на выпускную квалификационную работу бакалавра

Васильева Игоря Сергеевича

«Анализ социальных сетей с помощью технологий больших данных»

В выпускной квалификационной работе бакалавра Васильева Игоря Сергеевича описан процесс работы с графами, построенными на основе неструктурированных данных, взятых с сайтов интернет-рекрутмента. Данный класс задач представляет собой совокупность методов работы с данными: архитектура, сбор, хранение и анализ. В своей работе автор рассматривает каждый аспект работы, за исключением архитектуры. Каждый из этих этапов, вообще говоря, представляет собой отдельную задачу исследования, однако, учитывая количество коммерческих и бесплатных решений, на данный момент, более интересной является задача разработки системы, включающей в себя каждый из этих этапов. Представленную работу можно рассматривать как прототип такой системы. Помимо навыков освоения готовых технологий, таких как платформа Apache Spark, утилиты word2vec и doc2vec, автор продемонстрировал метод улучшения существующего алгоритма Гейла-Шепли для нахождения соответствия между элементами графа, правда с увеличением вычислительной сложности, что говорит о выполнении исследовательской деятельности. Работа состоит из введения, постановки задачи, обзора литературы, четырех глав, вывода, заключения, перспективах дальнейшей работы, а также перечня использованной литературы. В первой главе, состоящей из четырех разделов, приводятся общие сведения об анализе графов. Во второй главе, состоящей из трех разделов, описываются особенности анализа графа, построенного на данных, полученных из открытых источников, включая методы и утилиты работы с естественными языками. В третьей главе, состоящей из двух разделов, описывается процесс обработки данных для решения поставленной задачи, а также содержится описание особенностей работы с данными в Apache Spark. В четвертой главе, состоящей из 5 разделов, автор описывает структуру написанной программы для анализа и визуализации результатов.

Замечания по работе:

- В работе не хватает сравнения эффективности использования различных готовых инструментов, используемых автором, для решения отдельных этапов задачи.
- Одним из готовых решений, используемых автором, для визуализации полученного графа, является GraphX – Apache Spark API для визуализации графа, однако, скриншотов, демонстрирующих построенных граф не приводится.
- Автор приводит примеры задач, которые можно решить с помощью анализа построенного графа – самые востребованные профессии, выделение кластеров профессий и соискателей и так далее, однако, ни одна из этих задач не была решена хотя бы показательно.
- В целом, проделанной работой автор продемонстрировал владение современными методами и инструментами для работы с данными. Однако отсутствие сравнения готовых инструментов, а также решения конкретной аналитической задачи не позволяет по критериям для оценивания выпускной квалификационной работы на факультете ПМ-ПУ поставить отличную оценку. Считаю, что выпускная квалификационная работа Васильева Игоря Сергеевича заслуживает оценки **ХОРОШО**.

Рецензент,

Инженер-программист

Лаборатории Информационных Технологий

Объединенного Института Ядерных Исследований

27.05.2016



Румянцев Б.Д.