

Отзыв на выпускную квалификационную работу бакалавра Васильева Игоря Сергеевича

“Анализ социальных сетей с помощью технологий больших данных”

Выпускная квалификационная работа Васильева И.С. посвящена исследованию связей в профессиональных социальных сетях, объединяющих работодателей и соискателей. Целью работы является анализ данных из баз резюме и вакансий и разработка эффективного метода их сопоставления. В силу объемов данных, потенциально возникающих в ходе решения этой задачи, в работе рассматриваются технологии больших данных, которые могут позволить создать масштабируемую реализацию системы для работы с рассматриваемыми социальными сетями.

Автор приводит подробный анализ социальных графов в целом, их характеристик и метрик. Кратко рассмотрены основные алгоритмы анализа графов. Особое внимание уделено построению и анализу графа соискателя-работодателя. Анализ такого графа может дать ответы на многие актуальные вопросы, например каковы самые востребованные профессии, каковы профессии, требующие определенную совокупность навыков, каковы наиболее подходящие вакансии для соискателя и т.п. В качестве варианта решения поставленной задачи предлагается использовать алгоритмы нахождения максимального паросочетания в двудольном графе, в частности алгоритм Гейла-Шепли.

Для подготовки практической реализации автор рассматривает ряд подходов и существующих инструментов, уделяя особое внимание механизмам векторного представления слов и предложений Word2Vec и Doc2Vec для разбора резюме и вакансий, а также описывает применение Apache Spark для организации распределенной обработки данных. В практической части работы детально описаны подготовка данных и сбор информации о вакансиях и резюме из открытых источников, реализация классификации документов и ключевых навыков, анализ резюме и вакансий, определение связей между резюме и вакансиями с помощью инструментов Apache Spark. Результатом работы стало построение графа соискателя-работодателя на основе подхода, описанного в теоретической части, и применения современных инструментов обработки и анализа данных. Недостатком работы следует считать отсутствие детального графического представления полученного результата, отсутствие анализа полученного графа соискателя-работодателя, а также оценки эффективности разработанной системы. Тем не менее, основная цель работы, состоящая в разработке методов по сопоставлению данных из баз резюме и вакансий с использованием алгоритмов на графах и технологий больших данных, достигнута.

Несмотря на отмеченные недостатки, считаю, что работа Васильева И.С. соответствует требованиям для присвоения степени бакалавра с оценкой «отлично», а самого Васильева И.С. рекомендую к поступлению в магистратуру.

Научный руководитель,
доцент кафедры КМиМС, PhD



В.В. Корхов