

## ОТЗЫВ

Научного руководителя

На выпускную квалификационную работу бакалавра на тему «Оценка качества слабоструктурированных данных при сопоставлении независимых источников», представленную Калиной Алексеем Игоревичем в 2018 году.

Задачи оценки качества данных относятся к числу наиболее сложных и одновременно весьма важных для практических задач, при решении которых используются внешние источники данных, а также любых задач интеграции. Исследования методов оценки качества данных ведутся уже несколько десятилетий, однако получить убедительные результаты пока не удастся, несмотря на большой объем опубликованных материалов. Для иллюстрации достаточно указать, что только количество размерностей, по которым может производиться оценка качества. Составляет несколько десятков.

В отличие от большинства работ, связанных с оценкой качества, в работе Алексея Игоревича рассматривается задача оценки и улучшения качества не реляционных, а слабоструктурированных данных. В работе предложен метод оценки и одновременного улучшения качества посредством сравнения данных из нескольких различных источников. Для реализации этого метода пришлось проработать методы идентификации и отождествления объектов из разных источников и сопоставления значений их атрибутов.

Эксперименты проводились на двух типах коллекций данных, в том и другом случае удалось существенно повысить качество и устранить ошибки в данных автоматическими методами, что и составляет основную ценность работы.

По результатам работы подготовлена публикация, представленная на конференции SEIM218 в апреле текущего года, полный текст работы публикуется в сборнике, индексируемом в РИНЦ.

В связи с вышеизложенным считаю, что работа А.И. Калины заслуживает оценку «отлично».

Научный руководитель



Б.А. Новиков

Д.ф.-м.н., проф.

Проф. Кафедры инфомационно-аналитических систем СПбГУ