

ОТЗЫВ НАУЧНОГО РУКОВОДИТЕЛЯ

на диссертацию Малых Валентина Андреевича

«Методы сравнения и построения устойчивых к шуму программных систем в задачах обработки текстов»,

представленную на соискание ученой степени кандидата технических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

В современном мире текстов становится все больше, наблюдается так называемый информационный взрыв. При таком объеме производимого текста часть его неизбежно будет содержать различного рода опечатки и ошибки. Эти ошибки (шумы) лишь немного мешают при чтении текстов, но существенно осложняют их автоматическую обработку, которая занимает существенное место в широком классе задач информационного поиска, аналитики и управления.

Целью диссертационной работы Малых В.А. является исследование и разработка методов сравнения различных программных систем обработки текста по их устойчивости к шуму. Существующие методы сравнения не рассматривали те классы задач, которые находятся в центре внимания диссертанта: классификацию текстов, распознавание именованных сущностей и извлечение точек зрения. Помимо общих методов сравнения, дающих основу для выбора любых устойчивых методов, диссертантом были разработаны специальные методы для рассмотренных классов задач.

Актуальность выбранной темы проистекает из того, что существующие решения по исправлению орфографии обладают отнюдь не идеальным качеством. Как показал диссертант в своей работе, для рассматриваемых задач подавление шума с помощью систем проверки орфографии может в некоторых случаях даже ухудшать качество программной системы относительно системы, не использующей проверку орфографии. В большинстве же случаев, как показано в диссертации, программные системы, не устойчивые к шуму, пусть даже с проверкой орфографии, проигрывают в качестве системам, устойчивым к шуму изначально.

Выбранные диссертантом классы задач относятся к самым распространенным в анализе текстов. К задаче классификации текстов сводится большое число практических задач обработки потока сообщений. Распознавание именованных сущностей широко применяется в извлечении из текстов информационных единиц, например, адресов и названий. Извлечение точек зрения применяется в анализе мнений пользователей о продуктах, а также при классификации источников сообщений. В целом, рассмотренные классы задач охватывают широкий спектр практически востребованных применений анализа текстов и выбраны диссертантом с целью улучшить качество работы реально работающих программных систем.

При разработке и реализации методов, описанных в диссертации, Малых В.А. проделал большой объем работы, в том числе проанализировал и протестировал множество программных систем, провел многочисленные эксперименты.

В процессе подготовки диссертации Малых В.А. продемонстрировал способность самостоятельно выполнять научные исследования. Диссертант обладает фундаментальной осведомленностью о том, что уже сделано в научной области, которой он занимается.

Его собственный вклад в эту область так же существенен.

Предложенные им векторные представления слов и методы анализа, основанные на современных подходах, связанных с машинным обучением и нейронными сетями специального вида, несомненно, станут одной из ветвей в развитии математического и программного обеспечения компьютеров и компьютерных систем.

Считаю, что диссертационная работа соответствует всем требованиям, предъявляемым ВАК к работам на соискание ученой степени кандидата технических наук по специальности 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, а ее автор, Малых Валентин Андреевич, заслуживает присуждения ему ученой степени кандидата технических наук.

Научный руководитель:
зав. отделением ФИЦ ИУ РАН,
чл.-корр. РАН, д.т.н., профессор

В.Л. Арлазаров

28 февраля 2019 года