

**ОТЗЫВ**  
научного руководителя  
проф. Мухамедиева Р.И.

на диссертационную работу Якунина Кирилла Олеговича «Разработка моделей и методов сбора, анализа и классификации медиа-публикаций на базе методов обработки естественных языков», представленной на соискание степени «доктор философии» (PhD) по специальности «6D070400 - Вычислительная техника и программное обеспечение»

Задачи анализа естественно-языковых текстов пользуются большой популярностью в связи с большим количеством практических приложений, начиная от поисковых систем и заканчивая системами выработки персональных рекомендаций.

Автоматический анализ текстов часто является сложной задачей несмотря на значительной количество методов и средств. В частности, в настоящее время хорошо отработаны технологии обработки на основе статистической модели языка. Глубокие нейросетевые модели, так называемые трансформеры, показывают высокие результаты в задачах классификации.

Однако соединение достоинств различных методов становится сложным при обработке больших текстовых корпусов. Имеются также значительные технические и организационные трудности в подготовке размеченных корпусов для обучения современных моделей обработки данных. Глубокие модели являются трудно интерпретируемыми. Сложности могут возникнуть и при создании технических комплексов, способных обрабатывать миллионы публикаций. Преодоление указанных сложностей составило основу диссертационной работы.

Кирилл Якунин, на основе алгоритмов тематического моделирования, предложил модели и методы, позволяющие обучать классификационные модели, требующие минимального объема ручной разметки. Полученная модель обладает относительно низкой сложностью и легко интерпретируется. В результате, предложенная в работе тематическая векторизация текстов, а также предложенный метод расчета тематического межкорпусного дисбаланса позволяют снизить объем ручной разметки с более чем десятков тысяч текстов до сотен, а в ряде случаев до нескольких десятков текстов.

В ходе исследования, соискатель выполнил аналитический обзор и апробирование известных методов решения задач классификации текстов, имеющихся средств медиа мониторинга, предложил собственный метод многокритериальной оценки масс-медиа на базе байесовской системы агрегации, метода анализа иерархий (АНР) и тематического моделирования.

Практическая ценность работы заключается в том, что на основе предложенных методов и алгоритмов под руководством соискателя и при его непосредственном участии была разработана распределенная информационная система, позволяющая проводить автоматический мультикритериальный анализ медиа-пространства. Система внедрена в МОН РК в 2020г.

В ходе исследования Кирилл Якунин опубликовал более двух десятков статей, в том числе пять входящих в первый и второй квартили.

Все результаты получены автором лично. Научные руководители принимали участие в формулировке задач, согласовании плана проведения исследования, а также, в оценке результатов.

Представленная работа написана ясно, на хорошем научном языке, является завершённой, удовлетворяющей требованиям, предъявляемым к диссертациям на соискание степени доктора философии (PhD) по специальности 6D070400 - Вычислительная техника и программное обеспечение.

Кирилл Якунин является сложившимся молодым ученым, сочетающим практические навыки, высокий уровень теоретической подготовки и широкий кругозор в области новых технологий обработки данных. Он продолжает научную работу. По завершении работы над диссертацией им в соавторстве и лично подготовлены к публикации две работы, посвященные применению разработанных методов для решения практических задач обработки крупных массивов текстовой информации.

Уровень научной подготовки соискателя и представленная к защите диссертационная работа позволяют считать, что Якунин Кирилл Олегович заслуживает присуждения ему ученой степени PhD по специальности «6D070400 - Вычислительная техника и программное обеспечение».

Научный консультант:  
Д.инж.з., профессор



Мухамедиев Р.И.  
01.06.2021г.