

**РЕЦЕНЗИЯ**  
на выпускную квалификационную работу  
обучающегося 4 курса факультета ПМ-ПУ СПбГУ  
Тимофеева Александра Ивановича  
«Автоматическое выделение информативных тем документов с  
использованием латентного размещения Дирихле»

Выпускная работа А. И. Тимофеева посвящена задаче анализа текстовых документов с целью выделения из полного текста некоторого сравнительно небольшого количества тем. Работа с обобщенными описаниями коллекции документов, само собой, значительно проще, чем с полными содержаниями. Извлечение таких скрытых тематических характеристик текстов является частью таких важных прикладных задач, как информационный поиск (поиск документов не только по точному совпадению запроса, но и по теме), анализ социальных сетей, создание рекомендательных систем и пр.

Во введении объясняется интерес к задаче выделения информативных тем. Говорится о необходимости развивать имеющиеся методы для расширения круга приложений.

В первой главе описываются существующие вероятностные тематические модели. Они представлены в порядке, позволяющем подвести читателя к обоснованию выбора модели латентного размещения Дирихле и её модификации.

Во второй главе описываются методы сравнения вероятностных распределений, используемые в работе, и способ прогнозирования ARIMA.

В третьей главе приводятся результаты эксперимента, проведённого на примере открытой коллекции Book-Crossing Dataset. Прогноз построенный по обучающей выборке даёт хорошее приближение к реальным данным.

ВКР завершается заключением, в котором суммируются полученные результаты и делаются выводы, и списком использованной литературы.

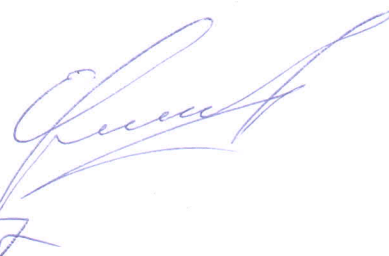
Работа полностью соответствует содержанию. Поставленная задача решена в полном объёме, результаты представленного эксперимента на тестовой коллекции аннотаций к книгам на английском языке показывают хорошие перспективы использования предложенного в работе метода.

Работа написана понятным языком, все используемые понятия и технологии подкреплены со ссылками на источники.

В ходе чтения было обнаружено несколько опечаток, вероятнее всего возникших в ходе изменения текста и добавления в него новых разделов. Так на стр. 7 после формулы (1), автор ссылается на (2). Такой формулы, однако, в работе нет вовсе. Видимо, имеется в виду ссылка на (1). Кроме того, рисунки, которые должны быть 12-м и 13-м, подписаны как 6-й и 7-й, а ссылка рядом с ними дана на рис. 9, хотя должна на тот же 12-й.

Отмеченные недостатки не влияют на высокое качество проделанной работы. По моему мнению, проведённое исследование свидетельствует о способностях студента к самостоятельному выполнению научных исследований и практической реализации теоретических результатов. Считаю, что выпускная квалификационная работа А. И. Тимофеева заслуживает оценки «отлично», а сам он — присвоения квалификации бакалавра.

Рецензент, к. ф.-м. н.,  
доцент каф. информационных систем



Еремин А. С.

01.06.2017