

ОТЗЫВ

На автореферат диссертации Карповича Сергея Николаевича «Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов», представленной к защите на соискание ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Одним из перспективных направлений автоматической обработки текстовых документов является вероятностное тематическое моделирование. В этом направлении чрезвычайно актуальной и важной является задача исследования новых возможностей построения вероятностных тематических моделей на коллекциях и потоках текстовых документов. Работа Карповича С.Н. посвящена решению этой задачи и направлена на создание общедоступного инструмента для анализа коллекций и потоков текстовых документов с помощью вероятностного тематического моделирования.

Автореферат дает ясное представление о структуре и содержании диссертационной работы, а также об основных полученных результатах. Научная новизна основных полученных результатов заключается в следующем:

1. Создан специальный русскоязычный корпус текстов SCTM-ru.
2. Разработан метод построения вероятностной тематической модели через обучение на размеченных данных.
3. Разработан алгоритм многозначной классификации текстовых документов, основанный на вероятностном тематическом моделировании.
4. Разработан метод определения тем для нового слова в динамических вероятностно тематических моделях.

На основании автореферата может быть сделан обоснованный вывод о том, что научные результаты, полученные С.Н. Карповичем, открывают новые свойства вероятностных тематических моделей и дополняют функционал вероятностного тематического моделирования новыми методами.

Практическая значимость работы заключается в возможности использования полученных результатов для эффективного решения актуальных прикладных задач в таких

направлениях, как информационный поиск, рекомендательные системы, системы принятия решений.

По содержанию автореферата нужно отметить следующие незначительные замечания. Во-первых, на рисунке 1 отображена концептуальная схема вероятностного тематического моделирования. В каждой из трех матриц вероятностной тематической модели в качестве значений ячеек указаны одни и те же выражения: Value 1, Value 2.... В этом случае остается непонятно - это одни и те же значения в каждой матрице или значения этих матриц различаются (то же относится к рис.4). Во-вторых, на схеме алгоритма определения тем «нового слова» (рис.5) допущена грамматическая ошибка: необходимо заметить слово «косинусного» на «косинусного».

Указанные замечания не являются принципиальными. Выполненное диссертационное исследование является законченной научно-исследовательской работой, обладающей научной новизной и практической значимостью, и соответствует требованиям п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 № 842, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Бабкин Эдуард Александрович,

К.т.н., доцент, PhD

Адрес: 603155, Россия, Нижний Новгород, ул. Б. Печерская, д.25/12

Телефон: +7-9107969761

Эл.почта: eababkin@hse.ru

Место работы: Национальный исследовательский университет Высшая школа экономики (Нижний Новгород),

Факультет информатики, математики и компьютерных наук.

Должность: профессор кафедры информационных систем и технологий

12.10.2017 V V