

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».

Диссертационная работа посвящена актуальной проблеме обработки потоков текстовых документов, связанной с автоматическим анализом потоков и коллекций на основе разработки новых алгоритмов и методов вероятностного тематического моделирования. Новизна работы состоит в следующем:

- создан русскоязычный корпус текстов, пригодный для проведения экспериментов с различными алгоритмами обработки текстов на естественном языке;
- предложен новый алгоритм многозначной классификации, использующий вероятностные тематические модели;
- разработан алгоритм построения динамической тематической модели, в котором темы «новых слов» определяются с помощью произведения Адамара векторов тем документов, в которых встретилось «новое слово».

Практическая значимость полученных автором результатов состоит в применении вероятностного тематического моделирования для решения задач автоматического анализа потока текстовых документов. Достоверность полученных результатов подтверждается применением апробированного математического аппарата, отсутствием противоречий с другими работами в данной предметной области. Основные научные результаты опубликованы в 3 статьях, входящих в Перечень ВАК, и одной работе в международном издании, индексируемом в реферативных базах Web of Science и Scopus.

Замечания по автореферату диссертации:

- автором разработан новый русскоязычный корпус текстовых документов SCTM-ru, но в тексте автореферата нет данных о сравнении этого корпуса с существующими корпусами, что не дает возможность в полной мере оценить значимость этой разработки.
- применение произведения Адамара очевидно сокращает трудоемкость вычислений, что важно при потоковой обработке, однако автор не указывает на возможные ошибки тематической классификации, порождаемые применением этого аппарата;
- на стр. 3 автореферата при описании текстового потока автор пишет «... время происшествия», автор отзыва считает, что более корректно — время регистрации документа.

Указанные замечания не являются принципиальными для выполненного исследования и представленных в научной новизне результатов. Судя по автореферату, выполненное исследование является завершенным научным трудом. Полученные результаты дают основание для признания диссертационной работы соответствующей как заявленной специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и

компьютерных сетей», так и требованиям, установленным в п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 № 842, предъявляемым к кандидатским диссертациям, а ее автор — Карпович Сергей Николаевич заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Сведения о составителе отзыва:

ФИО: Ульянов Михаил Васильевич

Ученая степень: доктор технических наук

Ученое звание: профессор

Почтовый адрес: Россия, 117997, Москва, ул. Профсоюзная, д. 65

Телефон: +7 (495) 334-87-51, 8-916-589-94-04

Адрес электронной почты: muljanov@mail.ru

Наименование организации:

ФГБУН Институт проблем управления им. В.А. Трапезникова РАН

Должность: ведущий научный сотрудник лаборатории №68.

12.10.2017