

федеральное государственное автономное  
образовательное учреждение высшего образования  
«Санкт-Петербургский национальный  
исследовательский университет  
информационных технологий,  
механики и оптики» (Университет ИТМО)

Кронверкский проспект, д. 49, г. Санкт-Петербург,  
Российская Федерация, 197101  
тел.: (812) 232-97-04 | факс: (812) 232-23-07  
od@mail.ifmo.ru | www.ifmo.ru

19.07.2017 № 22-1/924

3

## ЗАКЛЮЧЕНИЕ

Санкт-Петербургского национального исследовательского университета  
информационных технологий, механики и оптики (Университет ИТМО)  
Министерства образования и науки Российской Федерации

Диссертация «Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов» выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики».

В период подготовки диссертации соискатель Карпович Сергей Николаевич работал в «Санкт-Петербургском национальном исследовательском университете информационных технологий, механики и оптики» (Университет ИТМО) Министерства образования и науки Российской Федерации, международная научная лаборатория «Интеллектуальные технологии для социо-киберфизических систем», программист.

В 2006 г. окончил Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» имени В.И. Ульянова (Ленина) по специальности «Автоматизированные системы обработки информации и управления» диплом ВСГ №0242587.

В настоящее время является аспирантом «Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики» (Университет ИТМО) Министерства образования и науки Российской Федерации, по направлению/специальности «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Справка об обучении № 32/2017, выдана в 2017 г. федеральным государственным автономным образовательным учреждением высшего

образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Министерства образования и науки Российской Федерации.

Научный руководитель – доктор технических наук, профессор Александр Викторович Смирнов, основное место работы – Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), заведующий лабораторией интегрированных систем автоматизации.

По итогам рассмотрения принято следующее заключение:

1. Личное участие соискателя ученой степени в получении результатов, изложенных в диссертации.

Содержание диссертации и основные положения, выносимые на защиту, отражают персональный вклад автора в опубликованных работах. Подготовка к публикации полученных результатов проводилась автором самостоятельно с консультациями с соавторами. Представленные к защите результаты получены лично автором.

2. Степень достоверности результатов, проведенных соискателем ученой степени исследований.

Достоверность подтверждена аналитическим образом исследований и разработок в области обработки текстов на естественном языке, положительными итогами практического применения результатов диссертационной работы для анализа коллекций и потоков текстовых документов, а также апробацией основных научно-практических положений в печатных трудах и докладах на всероссийских и международных конференциях.

3. Новизна и практическая значимость результатов исследования заключаются в том, что:

- Создан русскоязычный корпус текстов SCTM-ru, позволяющий исследовать алгоритмы вероятностного тематического моделирования и отличающийся от других корпусов наличием оригинального текста документа и метатекстовой разметки: автор, время описанных событий, тема. Текст и метатекстовая разметка необходимы для построения различных видов ВТМ. Источником данных корпуса является сайт «Русские Викиновости».
- Предложен метод расчета матриц ВТМ на основе обучения с учителем (авторами документов) с учетом заданных связей документов и тем, позволяющий упростить построение ВТМ за счет отсутствия итераций.
- Предложен алгоритм многозначной классификации текстовых документов ml-PLSI, основанный на вероятностном тематическом моделировании и заключающийся в использовании матрицы «слово-

тема» ВТМ для классификации документов, позволяющий определять темы «новых-документов» при анализе потока текстовых документов в динамической тематической модели.

- Предложен метод определения тем «нового слова», отличающийся вычислительной эффективностью и использующий произведение Адамара тематических векторов документов, где это слово встретилось, вместо альтернативных подходов с использованием распределения Дирихле или процесса Дирихле, позволяющий определять вектора тем для «новых-слов» в потоке текстовых документов при построении динамической тематической модели.
- На основе микросервисной архитектуры разработан прототип комплекса программных средств для анализа потока текстовых документов с использованием вероятностного тематического моделирования, отличающийся от систем, реализованных на базе других архитектурных решений, большей свободой в выборе подходящих способов для решения конкретных практических задач, позволяющий визуализировать промежуточные и конечные результаты вероятностного тематического моделирования.

#### 4. Ценность научных работ соискателя ученой степени.

Ценность научных работ аспиранта состоит в создании корпуса текстов, методов, алгоритма и комплекса программных средств вероятностного тематического моделирования потока текстовых документов. Применение полученных результатов позволяет анализировать коллекции и потоки текстовых документов. Научные результаты были представлены на 3 всероссийских и 6 международных конференциях. Результаты диссертационной работы имеют высокую значимость для решения задач анализа текстов на естественном языке, информационном поиске и в сервисах рекомендаций.

5. Диссертация соответствует научной специальности: (05.13.11) «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей», а также требованиям, установленным п. 14 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ № 842 от 24.09.2013 г. (ред. от 10.06.2017).

6. Полнота изложения материалов диссертации в работах, опубликованных соискателем.

Основное содержание диссертации опубликовано в 12 статьях, из них 2 публикации в изданиях, рецензируемых Web of Science или Scopus, 3 публикации в журналах из перечня ВАК и 1 свидетельство о регистрации программы для ЭВМ.

6.1. Научные издания, входящие в международные реферативные базы данных и системы цитирования:

6.1.1. Smirnov A. Topic Model Visualization with IPython // A. Smirnov, N. Teslya, **S. Karpovich**, A. Grigorev // Open Innovations Assoc. FRUCT, Proc. of 20th Conf. – 2017. – Pp. 131-137 (**Web of Science, Scopus**) – 0.44 п.л. – 0.22 п.л. (авторский вклад 50%)

6.1.2. **Карпович С.Н.** Многозначная классификация текстовых документов с использованием вероятностного тематического моделирования ml-PLSI // Труды СПИИРАН. – СПб., 2016. – Т. 4. – №. 47. – С. 92-104 (**ВАК, Scopus**) – 0.76 п.л. (авторский вклад 100%)

6.2. Научные издания, входящие в перечень российских рецензируемых журналов:

6.2.1. **Карпович С.Н.** Русскоязычный корпус текстов SCTM-ru для построения тематических моделей // Труды СПИИРАН. – СПб., 2015. – № 39. С. 123-142. УДК 004.912 (**ВАК**) – 1.16 п.л. (авторский вклад 100%)

6.2.2. **Карпович С.Н.** Многозначная классификация текстовых документов с использованием вероятностного тематического моделирования ml-PLSI // Труды СПИИРАН. – СПб., 2016. – Т. 4. – №. 47. – С. 92-104 (**ВАК, Scopus**) – 0.76 п.л. (авторский вклад 100%)

6.2.3. **Карпович С.Н.** Тематическая модель с бесконечным словарем // Информационно-управляющие системы. 2016. №6 С. 43-49. doi:10.15217/issn1684-8853.2016.6.43 (**ВАК**) – 1.12 п.л. (авторский вклад 100%)

6.3. Публикации, которые приравниваются к рецензируемым научным изданиям:

6.3.1. Свидетельство о регистрации программы для ЭВМ «Система для анализа текстовых документов с использованием вероятностного тематического моделирования» // **Карпович С.Н.** №2017615118 от 3 мая 2017.

6.4. Публикации в иных изданиях:

6.4.1. **Карпович С.Н.** Проблемы базового образования и профессиональной подготовки в сфере информационных технологий // Современное образование: содержание, технологии, качество: XX междунар. научно-метод. конф. СПб., 23 апр.2013. Материалы конф. Т.1. – С.66-68. – 0.12 п.л. (авторский вклад 100%)

6.4.2. **Карпович С.Н.** Региональная информатика (РИ-2014). XIУ Санкт-Петербургская международная конференция «Региональная информатика (РИ-2014)». СПб, 29-31 окт.2014.: Материалы конференции. / СПОИСУ. – СПб, 2014. – 637 с. – 0.05 п.л. (авторский вклад 100%)

6.4.3. **Карпович С.Н.** Создание алгоритмов обработки неструктурированных данных большого объема // Материалы конф. Региональная информатика (РИ-2014). XIУ СПб Междунар. конф.-СПб, 29-31 окт. 2014.- С.39. – 0.06 п.л. (авторский вклад 100%)

6.4.4. **Карпович С.Н.** Универсальный потоковый анализ и комбинирование тематических моделей в обработке неструктурированных данных большого объёма // 68-я Научно-техническая конференция профессорско-преподавательского состава СПбГЭТУ "ЛЭТИ" Санкт-Петербург, 28 января-5 февраля 2015 г. Сборник докладов студентов, аспирантов и молодых ученых. – 0.06 п.л. (авторский вклад 100%)

6.4.5. **Карпович С.Н.** Автоматическая обработка текстов в тематическом моделировании // Современное образование: содержание, технологии, качество: XX междунар. науч.-метод. конф.: Изд-во СПбГЭТУ «ЛЭТИ», СПб, 22 апр. 2015 г., Т.2 – С. 167-168. – 0.09 п.л. (авторский вклад 100%)

6.4.6. **Карпович С.Н.** Русскоязычный корпус текстов СКТМ-ру для построения тематических моделей // Международная конференция «Корпусная лингвистика». 22-26 июня 2015. Сборник докладов конференции (РИНЦ) – 0.5 п.л. (авторский вклад 100%)

6.4.7. **Карпович С.Н.** Информационно-психологическая безопасность тематической модели с непрерывным временем для контроля публикаций в сети интернет // Информационная безопасность регионов России (ИБРР-2015). IX Санкт-Петербургская межрегиональная конференция. Санкт-Петербург, 28-30 октября 2015 г.: Материалы конференции / СПОИСУ. - СПб., 2015. - С. 309-310. 0.07 п.л. (авторский вклад 100%)

6.4.8. **Карпович С.Н.** Визуализация тематических моделей с помощью iPython // Международная конференция «Корпусная лингвистика». 2017. Сборник докладов конференции (РИНЦ) – 0.5 п.л. (авторский вклад 100%)

Диссертация «Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов» Карпович Сергей Николаевич соответствует требованиям п. 9 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ № 842 от 24.09.2013 г. (ред. от 10.06.2017) и пунктам 3 — «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем», и 4 — «Системы управления базами данных и знаний» паспорта специальности ВАК (05.13.11) по специальности «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Диссертация рекомендуется к защите на соискание ученой степени кандидата технических наук по специальности (05.13.11) «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Заключение принято на совместном заседании кафедры информационных систем и кафедры информационных технологий Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (Университет ИТМО).

Присутствовало на заседании 8 чел. в числе которых три доктора наук и

четыре кандидата наук.

Результаты голосования: «за» - 8 чел., «против» - 0 чел., «воздержалось» - 0 чел., протокол № 6 от « 27 » июня 2017 г.

Председательству

д.т.н., проф., за  
кафедрой информ.  
систем

Парфенов В.Г.

Секретарь заседани  
к.т.н., доцент  
компьютерных техн

Буздалов М.В.

Сведения о составителя

ФИО: Парфенов Влади

Ученая степень: доктор

Ученое звание: профессор

Место работы: Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Должность: заведующий кафедрой информационных систем

Почтовый адрес: 197101, г. Санкт-Петербург, Кронверкский проспект, д.49.

Телефон: +7 (812) 233-42-98

Адрес электронной почты: parfenov@mail.ifmo.ru

ФИО: Буздалов Максим Викторович

Ученая степень: кандидат технических наук

Ученое звание: профессор

Место работы: Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Должность: доцент кафедры компьютерных технологий

Почтовый адрес: 197101, г. Санкт-Петербург, Кронверкский проспект, д.49.

Телефон: +7 (812) 233-42-98

Адрес электронной почты: mbuzdalov@corp.ifmo.ru