



**ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ
СЛОЖНЫМИ СИСТЕМАМИ
РОССИЙСКОЙ АКАДЕМИИ НАУК**

443020, г. Самара, ул. Садовая, 61 тел. (846) 332-39-27 факс (846) 333-27-70, e-mail: iccs@iccs.ru, http://www.iccs.ru
ОКПО 04874434, ОГРН 1026301420881, ИНН/КПП 6317028447/631701001

Исх. №56-99
от 12.10.2017 г.

В диссертационный совет
Д-002.199.01 при СПИИРАН

ОТЗЫВ

о диссертации Карповича Сергея Николаевича
«Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов»,
на соискание учёной степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

(по автореферату)

В последние годы текстовая информация в электронном виде генерируется в огромных количествах. Преобразование таких неструктурированных данных на естественном языке – слов, предложений, документов и их взаимосвязей – в компактную форму, позволяющую человеку лучше понимать текстовые документы и уменьшать нагрузку при работе с большими коллекциями текстовых документов – одна из актуальнейших задач современности. Поэтому диссертация С.Н. Карповича, посвященная разработке методов вероятностного тематического моделирования текстов, является актуальным исследованием.

Научные положения, выносимые С.Н. Карповичем на защиту, являются обоснованными. Применение методов, выдвигаемых соискателем, вызвано требованиями, предъявляемыми со стороны пользователей и разработчиков систем вероятностного моделирования тем. Эти требования выявлены и сформулированы автором диссертационной работы на основе обширного анализа состояния исследований в избранной предметной области.

Полученные результаты, в частности, метод расчета матриц вероятностной тематической модели (ВТМ) на основе обучения с учителем, алгоритм многозначной классификации документов и метод определения нового слова обладают научной новизной. Совместное использование предложенных методов представляет интерес в области построения ВТМ, основанный на возможности анализа потока текстовых документов, тогда как подавляющее большинство существующих методов

применимо лишь к статическим корпусам документов. Достоверность теоретических предположений подтверждается результатами испытаний методов на реальных данных и сравнением с результатами.

Практическая значимость полученных результатов состоит в значительном ускорении построения ВТМ документа и возможности анализа потока текстовых документов. Автором были также разработаны средства для визуализации результатов работы предложенных методов, которые позволяют оценить динамику изменения интереса к определенным темам.

Среди замеченных в автореферате недостатков обозначу следующее.

1. Автор выбирает для исследования вероятностное моделирование тем, не позиционируя эту модель среди других конкурентных подходов к малоразмерному представлению текстов – неотрицательному матричному разложению (NMF), моделям word2vec, t-sne и т.п.
2. В тексте автореферата приведены количественные оценки качества работы разработанного метода, однако важна и скорость работы соответствующих алгоритмов.
3. На рисунках 2 и 6 связи блоков имеют цифровые обозначения, которые не комментируются в тексте автореферата.

В целом считаю, что представленная к защите диссертация Карповича С.Н. *«Математическое и программное обеспечение вероятностного тематического моделирования потока текстовых документов»* соответствует требованиям п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24.09.13 №842, предъявляемым к кандидатским диссертациям, и автор диссертации заслуживает присуждения учёной степени кандидата технических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

рнов

Сведения о составителе отзыва:

Смирнов Сергей Викторович

д.т.н.

Федеральное государственное бюджетное учреждение науки Институт проблем управления сложными системами Российской академии наук (ИПУСС РАН)

Заместитель директора по научной работе

443020, г. Самара, ул. Садовая, 61

Тел.: +7(846)332-27-70, smirnov@iccs.ru

Подпись Смирнова С.В. удостоверяю

Старший инспектор отдел

ИПУСС РАН

Боровик Н.В.