

ОТЗЫВ

на автореферат диссертации Милковой Марии Александровны «**Разработка и экспериментальная апробация методики построения тематической модели при работе с научной информацией**», представленной на соискание учёной степени кандидата экономических наук по специальности 08.00.13 – «Математические и инструментальные методы экономики».

Актуальность

Мировой поток научной информации не только растёт количественно во времени, но также подвержен качественным изменениям. Появляются новые темы исследований, язык науки обогащается новыми терминами или старые термины начинают употребляться в другом контексте, новые авторы публикаций набирают вес в научном мире. Поисковые запросы для информационного обеспечения научной или управленческой деятельности, со временем устаревают и нуждаются в обновлении. Отслеживание изменений и обновление запросов экспертными методами – трудоёмкая работа, результаты которой зависят от квалификации и субъективных предпочтений экспертов. В этих условиях тема диссертации представляется актуальной, так как тематическая кластеризация коллекций документов и выявление скрытых тем, выявление наиболее значимых терминов и авторов в каждой теме – прямой путь к автоматизации работы, которая пока ещё в большинстве случаев выполняется вручную.

Новизна и теоретическая значимость

Предложена и экспериментально исследована формализация алгоритма АРТМ, позволяющая снизить трудоёмкость подбора параметров при построении тематической модели на коллекции текстов. Проведён анализ чувствительности результатов тематического моделирования к изменению начальных инициализаций матриц. Показано, что в случае, если темы являются хорошо интерпретируемыми, их состав существенно не меняется в зависимости от начальных приближений.

Практическая значимость

Автором предложена и апробирована методика автоматического составления рейтинга отраслей импортозамещения на основе применения инструментария тематического моделирования к коллекции патентных документов, релевантных государственной программе и отраслевым планам импортозамещения. Согласно сайтам <https://programs.gov.ru/> и <https://spending.gov.ru/gp/> в России в настоящее время действуют более 40 государственных программ. Методика автора может оказаться востребованной в процедурах мониторинга реализации программ с использованием документов различного вида (патенты, научные публикации, нормативные документы).

Достоверность результатов обоснована строгим применением математического аппарата теории матриц, теории вероятностей и математической статистики. Эксперименты проводились на выборках текстов достаточно большого объёма – более 150 тыс. патентных документов (название, реферат) и более 37 тыс. научных статей (название, аннотация).

Недостатки и дискуссионные вопросы

К языку, стилю изложения и оформлению автореферата замечаний нет, однако есть некоторые вопросы по содержанию и смыслу работы.

Стр. 4-6. В разделе *«Степень научной разработанности проблемы»* непонятно, какое непосредственное отношение к теме диссертации имеют нейроэкономика и поведенческая экономика, а также психологические и поведенческие аспекты принятия решений.

Стр. 14. Не указано, какому множеству принадлежит s – индекс суммирования в формуле (3).

Стр. 19. Не раскрыто, на основании чего *«веса модальностей предлагается брать равными 0.5 для слов, 1.0 для биграмм, авторов и ссылок»*. Впрочем, далее на стр. 21 читаем, что *«оптимальный вес модальности биграмм определялся экспериментально»*. Желательно было раскрыть суть экспериментального определения веса модальностей.

Поскольку эксперименты автора носили не умоглядный характер, а выполнялись с использованием разработанного им программного обеспечения на языке Python и свободно распространяемой библиотеки BigARTM, в автореферате хотелось бы увидеть сведения о технических характеристиках компьютера и машинном времени, затраченном на построение тематических моделей двух коллекций (статей и патентов) с различным количеством документов. Это повысило бы практический интерес читателей к автореферату.

Вывод

Отмеченные выше недостатки не являются критически значимыми, поэтому не снижают высокий уровень диссертационной работы. Считаю, что область исследований диссертации соответствует пунктам 2.6 и 2.8. паспорта специальности 08.00.13, а сама диссертация отвечает критериям научно-квалификационной работы согласно разделу II «Положения о присуждении ученых степеней». Автор диссертации, Милкова Мария Александровна, заслуживает присуждения учёной степени кандидата экономических наук по специальности 08.00.13 – «Математические и инструментальные методы экономики».

Начальник Отдела программных систем Управления информационных систем ВИНТИ РАН, кандидат технических наук по специальности 05.25.05 – «Информационные системы и процессы»

Федорен

Федорен Олег Владимирович

*Личную копию автореферата
Милковой Марии Александровны
получил и заверил под
2 л. специального назначения
отдела кадров
Дир. Луцкинова
29.11.2021*



29 ноября 2021 г.

e-mail: ovf@viniti.ru

тел.: (499) 155-42-06

Федеральное государственное бюджетное учреждение науки
Всероссийский институт научной и технической информации Российской академии наук
(ВИНИТИ РАН)

<http://www.viniti.ru/>

Адрес: Россия, 125190, Москва, А-190, ул. Усиевича, д. 20