



УТВЕРЖДАЮ:

Проректор по научной работе и
международным связям
ФГБОУ ВО «Уральский
государственный университет
путей сообщения»

С.В. Бушуев

21 октября 2016 г.

ЗАКЛЮЧЕНИЕ

Федерального государственного бюджетного образовательного
учреждения высшего образования
«Уральский государственный университет путей сообщения»

Диссертация «Алгоритмы интеллектуального поиска на основе метода категориальных векторов» выполнена на кафедре естественнонаучных дисциплин в Федеральном государственном бюджетном образовательном учреждении высшего образования «Уральский государственный университет путей сообщения» Федерального агентства железнодорожного транспорта.

В период подготовки диссертации Бондарчук Дмитрий Вадимович обучался в очной аспирантуре в Федеральном государственном бюджетном образовательном учреждении высшего образования «Уральский государственный университет путей сообщения» с 01 июня 2012 по 30 июня 2015 г., с 2012 г. по настоящее время работает ассистентом кафедры естественнонаучных дисциплин (ранее, до 01.10.2016 г. – кафедра «Высшая и прикладная математика», Приказ ректора УрГУПС о реорганизации кафедр № 545 от 12.09.2016 г.).

В 2012 г. Бондарчук Д.В. окончил Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Национальный исследовательский ядерный университет МИФИ» с присуждением квалификации инженера-программиста по специальности «Программное обеспечение вычислительной техники и автоматизированных систем».

Удостоверение о сдаче кандидатских экзаменов выдано 10 октября 2016 г. Федеральным государственным автономным образовательным учреждением высшего образования «Южно-Уральский государственный университет» (национальный исследовательский университет).

Научный руководитель – Тимофеева Галина Адольфовна, профессор, заведующая кафедрой естественнонаучных дисциплин Федерального государственного бюджетного образовательного учреждения высшего образования «Уральский государственный университет путей сообщения», доктор физико-математических наук, профессор.

По результатам рассмотрения диссертации «Алгоритмы интеллектуального поиска на основе метода категориальных векторов» принято следующее **заключение**:

Актуальность темы и направленность исследования

Диссертационная работа Д.В. Бондарчука посвящена совершенствованию методов и разработке алгоритмов интеллектуального анализа текстовых данных. Актуальность темы диссертационной работы Д.В. Бондарчука определяется тем, что в настоящее время работа значительного числа информационных систем, систем поддержки принятия решений и поисковых систем в различных отраслях экономики и науки основана на обработке больших объемов текстовой информации. Наиболее перспективным подходом к решению задач классификации, кластеризации и упорядочения текстовых данных является применение технологий, основанных на машинном обучении. Большинство современных методов и алгоритмов интеллектуальной обработки текстовых данных опираются на знания о структуре языка, статистику встречаемости слов в тексте и т.п. В качестве недостатка существующих на сегодняшний день методов и алгоритмов можно выявить недостаточный учет взаимодействия элементов информации меж-

ду собой и отношения пользователя к знанию, вследствие чего снижается релевантность поиска. Таким образом, актуальной является задача улучшения качества интеллектуального анализа текстовых данных за счет учета семантической и лексикографической взаимосвязи слов, решения проблемы лексической многозначности и разработки методов, обеспечивающих выдачу непустого упорядоченного по степени полезности результата для любого пользовательского запроса.

Личное участие автора в получении научных результатов

В ходе диссертационного исследования Д.В. Бондарчуком были получены следующие основные результаты:

1. Разработаны модель образа текстового документа и соответствующий метод отображения текста в семантическое пространство, обеспечивающие компактное представление документа в оперативной памяти на основе матрицы корреспонденций термов, которая подвергается ортогональному разложению.
2. Разработан алгоритм интеллектуального анализа текстов, гарантирующий непустой результат независимо от распределения обучающей выборки по категориям на основе использования вычисления категориальных векторов для упорядочения результирующей выборки по степени релевантности запросу пользователя.
3. Предложен метод перевзвешивания термов векторной модели с помощью вычисления их семантической взаимосвязи друг с другом на основе авторской адаптации алгоритма Леска.
4. На основе разработанных методов и подходов реализован алгоритм подбора рекомендаций. Проведены вычислительные эксперименты, подтверждающие более высокую эффективность разработанного алгоритма по сравнению с существующими.

Все исследования, связанные с пунктами 1–4, были выполнены

Д.В. Бондарчуком единолично. Г.А. Тимофеевой принадлежит постановка задачи и научное консультирование при выполнении исследований.

Степень достоверности результатов исследования

Все аналитические результаты, приведенные в диссертационной работе, сформулированы в виде утверждений, снабженных подробными доказательствами. Достоверность полученных результатов подтверждена большим количеством вычислительных экспериментов, проведенных автором на модельных и реальных текстовых данных.

Научная новизна результатов

Научная новизна проведенного исследования заключается в том, что автором предложены и обоснованы новые методы и подходы к интеллектуальному анализу текстовых данных, в том числе: выделение семантического ядра с помощью ортогонального разложения матрицы корреспонденций термов, использование в векторной модели переопределения веса термина с учетом семантической близости, статистический метод вычисления семантической близости термов, основанный на сборе контекстных множеств термов, впервые дано математическое обоснование применения нормированной терм-документной матрицы.

Практическая значимость полученных результатов

Предложенные в работе методы и подходы могут найти применение в системах автоматической рубрикации, фильтрации спама, системах интеллектуального и научного поиска, системах поддержки принятия решений, системах формирования персональных рекомендаций, а так же в системах поиска по глобальной сети Интернет. Разработанные алгоритмы позволяют производить поиск, классификацию и формировать персональные рекомендации пользователю, а так же выдавать ему результат в виде списка, отсортированного по степени релевантности запросу. Эффективность разработанных мето-

дов и алгоритмов подтверждается их внедрением в коммерческих организациях для интеллектуального анализа текстовых данных.

Материалы диссертации полно представлены в работах, опубликованных соискателем.

**Статьи, опубликованные в рецензируемых научных журналах,
определенных ВАК:**

1. Бондарчук Д.В. Статистический способ определения семантической близости термов // Системы управления и информационные технологии. – 2015. – Т. 61, № 3. – С. 55–57.
2. Бондарчук Д.В. Алгоритм построения семантического ядра для текстового классификатора // В мире научных открытий. – 2015. – Т. 68, № 8.2. – С. 713–724.
3. Бондарчук Д.В., Тимофеева Г.А. Выделение семантического ядра на основе матрицы корреспонденций термов // Системы управления и информационные технологии. – 2015. – Т. 61, № 3.1. – С. 134–139.
4. Бондарчук Д.В., Тимофеева Г.А. Применение машинного обучения для формирования персональных рекомендаций в сфере трудоустройства // Экономика и менеджмент систем управления. – 2015. – Т. 18, № 4.2. – С. 215–221.
5. Бондарчук Д.В., Тимофеева Г.А. Математические основы метода категориальных векторов в интеллектуальном анализе данных // Вестник Уральского государственного университета путей сообщения. – 2015. – 4(28). – С. 4–8.

Статьи, опубликованные в научных журналах и изданиях, индексируемых в WoS и Scopus:

6. Bondarchuk D. V., Timofeeva G. A. Vector space model based on semantic relatedness // AIP Conference Proceedings, Vol. 1690, Proceedings of 41st International Conference "Applications of Mathematics in Engineering and Economics"(AMEE'15). – 2015. – Pp.1–5.

7. Bondarchuk D.V., Martynenko A.V. Spectral properties of a matrix of correspondences between terms // CEUR Workshop Proceedings, Vol. 1662, Proceedings of 47th International Youth School-Conference "Modern Problems in Mathematics and its Applications"(MPMA 2016). – 2016. – Pp. 186–190.

Статьи, опубликованные в рецензируемых научных журналах и изданиях, индексируемых в РИНЦ:

8. Бондарчук Д.В. Использование латентно-семантического анализа в задачах классификации текстов по эмоциональной окраске // Бюллетень результатов научных исследований. – 2012. – 2(3). – С. 146–151.
9. Бондарчук Д.В. Выбор оптимального метода интеллектуального анализа данных для подбора вакансий // Информационные технологии моделирования и управления. – 2013. – 6(84). – С. 504–513.
10. Бондарчук Д.В. Интеллектуальный метод подбора персональных рекомендаций, гарантирующий получение непустого результата // Информационные технологии моделирования и управления. – 2015. – Т. 2(92). – С. 130–138.

Диссертация «Алгоритмы интеллектуального поиска на основе метода категориальных векторов» Бондарчука Дмитрия Вадимовича рекомендуется к защите на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – теоретические основы информатики.

Заключение принято на заседании кафедры естественнонаучных дисциплин Федерального государственного бюджетного образовательного учреждения высшего образования «Уральский государственный университет путей сообщения».

Диссертация заслушана и обсуждена на открытом заседании кафедры 21 октября 2016 г., протокол № 4.

ПРИСУТСТВОВАЛИ: зав. кафедрой естественнонаучных дисциплин Тимофеева Галина Адольфовна (доктор физико-математических наук, профессор), профессор Баутин Сергей Петрович (доктор физико-математических наук, профессор), профессор Дерябин Сергей Львович (доктор физико-математических наук, профессор), профессор Стружанов Валерий Владимирович (доктор физико-математических наук, профессор), профессор Титов Сергей Сергеевич (доктор физико-математических наук, профессор), профессор Румянцев Сергей Алексеевич (доктор физико-математических наук), профессор Фишбейн Лев Абрамович (кандидат физико-математических наук, доцент), декан электротехнического факультета Башуров Вячеслав Владимирович (кандидат физико-математических наук, доцент), доцент Волкова Татьяна Александровна (кандидат физико-математических наук, доцент), доцент Завьялова Татьяна Викторовна (кандидат физико-математических наук, доцент), доцент Завалицин Дмитрий Станиславович (кандидат физико-математических наук, доцент), доцент Мартыненко Александр Валериевич (кандидат физико-математических наук, доцент), доцент Медведева Наталья Валерьевна (кандидат физико-математических наук, доцент), доцент Скачков Павел Павлович (кандидат физико-математических наук, доцент), доцент Спевак Лев Фридрихович (кандидат физико-математических наук, доцент), доцент Розенберг Валерий Львович (кандидат физико-математических наук, доцент), доцент Филиппов Сергей Дмитриевич (кандидат физико-математических наук, доцент), доцент Чуев Николай Павлович (кандидат физико-математических наук, доцент), доцент Мезенцев Алексей Владимирович (кандидат физико-математических наук), доцент Садов Алексей Павлович (кандидат физико-математических наук), доцент Куликова Ольга Валентиновна (кандидат педагогических наук, доцент), доцент Гончарь Петр Сергеевич (кандидат педагогических наук, доцент), доцент Гниломедов Павел Ива-

нович (кандидат педагогических наук, доцент), доцент Пирогова Ирина Николаевна, старший преподаватель Борисова Надежда Олеговна, старший преподаватель Филиппова Елена Геннадьевна, старший преподаватель Башурова Оксана Анатольевна, ассистент Куликова Ирина Валерьевна, ассистент Габдулхаев Вадим Фатикович, ассистент Кирьянова Анна Сергеевна, ассистент Козлов Петр Алексеевич.

ПРИГЛАШЕНЫ: зав. кафедрой «Информационные технологии и защита информации» Зырянова Татьяна Юрьевна (кандидат технических наук, доцент), доцент кафедры «Информационные технологии и защита информации» Паршин Константин Анатольевич (кандидат технических наук, доцент), старший преподаватель кафедры «Информационные технологии и защита информации» Медведев Никита Владимирович (кандидат технических наук).

Присутствовало на заседании 34 человека, из них сотрудников кафедры – 31. Результаты голосования: «за» – 34 чел., «против» – нет, «воздержалось» – нет, протокол № 4 от 21.10.2016 г.

Мартыненко Александр Валериевич
кандидат физико-математических наук, доцент,
зам. зав. кафедрой
естественнонаучных дисциплин


