

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.224.03,
СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО
УЧРЕЖДЕНИЯ

«ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
«ИНФОРМАТИКА И УПРАВЛЕНИЕ»
РОССИЙСКОЙ АКАДЕМИИ НАУК»

ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ ДОКТОРА НАУК

Аттестационное дело № _____

Решение диссертационного совета 24.1.224.03 от 19 февраля 2026 г., № 2

О присуждении Бутенко Юлии Ивановне, гражданке Российской Федерации, учёной степени доктора технических наук.

Диссертация «Модели и методы автоматической обработки научно-технических текстов в параллельном корпусе» по специальности 2.3.8 – «Информатика и информационные процессы» принята к защите 18 сентября 2025 г., протокол № 8, диссертационным советом 24.1.224.03 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН), 119333, г. Москва, ул. Вавилова, д. 44/2, созданным на основании приказа Министерства науки и высшего образования Российской Федерации № 1732/нк от 13 декабря 2022 г.

Соискатель Бутенко Юлия Ивановна, дата рождения 01.03.1987, диссертацию на соискание учёной степени кандидата технических наук «Онтологические модели и методы формирования нормативного профиля при сертификации программного обеспечения» защитила в 2015 году в диссертационном совете в Национальном техническом университете «Харьковский политехнический институт» Министерства образования и науки Украины. В настоящее время работает в Федеральном государственном автономном образовательном учреждении высшего образования «Московский

государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана) в должности доцента.

Диссертация выполнена на кафедре теоретической информатики и компьютерных технологий МГТУ им. Н.Э. Баумана.

Официальные оппоненты:

1. Барахнин Владимир Борисович, доктор технических наук, доцент, ведущий научный сотрудник лаборатории информационных ресурсов Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр информационных и вычислительных технологий»;

2. Елизаров Александр Михайлович, доктор физико-математических наук, профессор, профессор Института информационных технологий и интеллектуальных систем Федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет»;

3. Котельников Евгений Вячеславович, доктор технических наук, доцент, профессор Школы вычислительных социальных наук Автономной некоммерческой образовательной организации высшего образования «Европейский университет в Санкт-Петербурге»

дали положительные отзывы на диссертацию.

Ведущая организация – Федеральное государственное бюджетное учреждение науки Институт проблем управления имени В. А. Трапезникова Российской академии наук – в своем положительном отзыве, подписанном Губановым Дмитрием Алексеевичем, доктором технических наук, ведущим научным сотрудником лаборатории № 25, указала:

В диссертационной работе предложен комплекс моделей, методов и программных средств, разработанных на основе фундаментальных лингвистических теорий и положений, что обеспечивает возможность эффективного решения широкого круга задач автоматической обработки научно-технических текстов. Практическая ценность диссертационной работы заключается в разработке программных средств для обработки научно-

технических текстов в основе параллельного корпуса, применение которых позволяет автоматизировать трудоемкие процедуры по разметке и выравниванию текстов при формировании корпусов. Помимо непосредственного использования в корпусной лингвистике, предложенные решения могут быть востребованы при решении ряда прикладных задач в области терминографии, перевода, наукометрии, машинного обучения и инженерии знаний. Тема и содержание диссертации полностью соответствуют выбранной специальности 2.3.8 – «Информатика и информационные процессы». В отзыве ведущей организации отмечается, что диссертация Бутенко Ю.И. является завершённой научно-квалификационной работой, в которой решается научная задача, имеющая важное практическое значение, а ее автор, Бутенко Ю.И., заслуживает присуждения учёной степени доктора технических наук по специальности 2.3.8 – «Информатика и информационные процессы».

Соискатель имеет 155 опубликованных научных работ. По теме диссертации опубликовано 60 научных работ, из которых 27 статей в научно-технических журналах, входящих в перечень ВАК, 20 – в изданиях, входящих в международные наукометрические базы Scopus и Web of Science. В трудах российских и международных конференций опубликовано 29 работ.

Недостоверные сведения об опубликованных соискателем работах в диссертации отсутствуют. Диссертация соответствует п. 14 Положения о присуждении учёных степеней. Автор подробно указала личный вклад в опубликованные с соавторами работы.

Наиболее значимые работы Бутенко Ю.И. по теме диссертации:

1. Бутенко Ю.И. Метод выявления русскоязычных машинно-сгенерированных текстов по особенностям актуального членения предложения // Научно-техническая информация: Серия 1. Организация и методика информационной работы. 2025. №6. С. 19-26. DOI: 10.36535/0548-0019-2025-06-3. (ВАК, переводная версия Scopus, WoS)

2. Бутенко Ю.И. Метод извлечения многокомпонентных терминологических единиц с правыми определениями из научно-технических текстов // Вестник Новосибирского государственного университета. Серия:

Информационные технологии. 2024. Т. 22. № 3. С. 5-14. DOI: 10.25205/1818-7900-2024-22-3-5-14. (ВАК К-2)

3. Бутенко Ю.И. Извлечение номенклатурных наименований из англо- и русскоязычных научно-технических текстов // Искусственный интеллект и принятие решений. 2024. №3. С. 95-103. DOI:10.14357/20718594240309. (RSCI)

4. Бутенко Ю.И. Метод выравнивания многокомпонентных англо- и русскоязычных терминов в параллельном корпусе на основе пакета SimAlign // Научно-техническая информация: Серия 2. Информационные процессы и системы. 2024. №8. С. 29-38. DOI: 10.36535/0548-0027-2024-08-4. (RSCI, переводная версия WoS)

5. Бутенко Ю.И., Сидняев Н.И., Синева Е.Е. Стратегии поиска в пространстве состояний баз знаний // Научно-техническая информация: Серия 2. Информационные процессы и системы. 2024. №6. С. 25-39. DOI:10.36535/0548-0027-2024-06-4. (RSCI, переводная версия WoS)

6. Бутенко Ю.И., Солошенко К.А. Лексический тренажер по иностранному языку для студентов технических специальностей МГТУ им. Н.Э. Баумана // Экономика. Информатика. 2024. Т. 51. № 1. С. 189–200. DOI: 10.52575/2712-746X-2024-51-1-189-200. (ВАК К-2)

7. Бутенко Ю.И. Использование базы данных структурных трансформаций для извлечения многокомпонентных терминологических единиц // Системы и средства информатики. 2023.Т. 33, №1. С. 35-44. DOI: 10.14357/08696527230104. (RSCI)

8. Бутенко Ю.И., Галетка М.Л. Синева Е.Е. Создание системы разметки семантических ролей в научно-технических текстах по авиации и космонавтике // Научно-техническая информация: Серия 2. Информационные процессы и системы. 2022. №10. С. 23-32. DOI:10.36535/0548-0027-2022-10-4. (RSCI, переводная версия WoS)

9. Бутенко Ю.И. Модель текста научно-технической статьи для разметки в корпусе научно-технических текстов // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2022. Т. 20. № 3. С. 5-13. DOI: 10.25205/1818-7900-2022-20-3-5-13. (ВАК К-2)

10. Бутенко Ю.И. Строганов Ю.В. Сапожков А.М. Система извлечения многокомпонентных терминов и их переводных эквивалентов из параллельных научно-технических текстов // Научно-техническая информация: Серия 2. Информационные процессы и системы. 2022. №9. С. 12-21. DOI: 10.36535/0548-0027-2022-09-3. (RSCI)

11. Бутенко Ю.И. Модель учебно-научного текста для разметки корпуса научно-технических текстов // Экономика. Информатика. 2021. Т. 48. № 1. С. 123–129. DOI:10.52575/2687-0932-2021-48-1-123-129. (ВАК К-2)

12. Бутенко Ю.И., Строганов Ю.В., Сапожков А.М. Метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов // Прикладная информатика. 2021. №6. С. 21-27. DOI: 10.37791/2687-0649-2021-16-6-21-27. (RSCI)

13. Бутенко Ю.И., Сидняев Н.И., Болотова Е.Е. Использование падежной грамматики при информационном поиске в базе знаний экспертной системы о конструкциях летательных аппаратов // Системы и средства информатики. 2021. №3. С. 80-87. DOI: 10.14357/08696527210307. (RSCI)

14. Бутенко Ю.И. Метод разрешения лексической многозначности поискового запроса на основе онтологий // Прикладная информатика. 2020. Т. 15, №5. С. 103-110. DOI: 10.37791/2687-0649-2020-15-5-103-110. (RSCI)

15. Butenko Yu.I., Stroganov Yu.V., Sapozhkov A.V. System for extracting multicomponent terms and their translated equivalents from parallel scientific and technical texts // AIP Conference Proceedings: International conference on modeling in engineering 2021, Moscow, Russia, October, 26-27. 2023. Vol. 2833(1), P. 030015. DOI: 10.1063/5.0151707. (Scopus)

На автореферат поступили отзывы, которые подписали:

1. Сафонов Константин Владимирович, доктор физико-математических наук, профессор, директор Института информатики и телекоммуникаций ФГБОУ ВО «Сибирский государственный университет имени академика М.Ф. Решетнёва». Отзыв положительный. Замечания носят рекомендательный характер.

2. Ильичев Андрей Теймуразович, доктор физико-математических наук, профессор, ведущий научный сотрудник Математического института им. В.А. Стеклова Российской академии наук. Отзыв положительный. Замечаний нет.

3. Катышев Павел Алексеевич, доктор филологических наук, профессор, заведующий кафедрой общего и русского языкознания Института Пушкина. Отзыв положительный. Замечаний нет.

4. Безуглов Дмитрий Анатольевич, доктор технических наук, профессор, заместитель директора по научной работе Ростовского филиала Российской таможенной академии. Отзыв положительный. Замечания носят рекомендательный характер.

5. Хорошилов Александр Алексеевич, доктор технических наук, ведущий научный сотрудник отдела информационных технологий ФГБУН Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН). Отзыв положительный. Замечания носят рекомендательный характер.

6. Михеева Наталья Федоровна, доктор филологических наук, профессор, профессор кафедры теории и практики иностранных языков ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы». Отзыв положительный. Замечаний нет.

7. Смирнов Борис Петрович, доктор технических наук, научный руководитель Закрытого акционерного общества «Научно-производственный центр информационных региональных систем». Отзыв положительный. Замечания носят рекомендательный характер.

Также имеется 2 акта о внедрении результатов диссертации, полученных от следующих организаций: ООО «Агентство переводов «Эдельвейс», ООО «Инженерное бюро ВССТ».

Выбор официальных оппонентов и ведущей организации обосновывается их достижениями в области обработки естественного языка, что подтверждается их исследованиями и публикациями в высокорейтинговых научных журналах.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

1. **Разработаны** модели структуры научно-технических текстов, использующие межуровневые элементы и оценки значимости каждого структурного элемента при создании параллельного корпуса, что позволяет более эффективно обрабатывать научно-технические тексты на разных уровнях языковой системы.

2. **Предложены** модели выравнивания и методы разметки соответствия англо- и русскоязычных терминологических единиц из научно-технических текстов, отличающиеся от существующих возможностью извлечения терминов с правыми определениями, что позволяет использовать эти модели и методы при обработке текстов при создании параллельного корпуса.

3. **Предложены** модели номенклатурных наименований в научно-технических текстах на русском и английском языках, **разработан** метод разметки и выравнивания номенклатурных наименований, что позволяет повысить эффективность разметки научно-технических текстов за счет учета лексических единиц, в состав которых входят произвольные буквенно-числовые последовательности в том числе символы разных алфавитов.

4. **Предложены** методы выявления машинно-сгенерированных и машинно-переведенных текстов, а также их фрагментов на основе семантико-синтаксических особенностей русского языка. **Разработана** процедура определения особенностей синтаксической структур текстов, как основа для выявления машинных и антропогенных научно-технических текстов.

5. **Разработан** прототип системы управления корпусными данными, который в отличие от существующих корпусных менеджеров позволяет управлять корпусными данными на разных этапах их обработки, а также формировать различные наборы данных для машинного обучения. **Разработаны** система разметки и выравнивания многокомпонентных терминов, система семантико-синтаксической разметки научно-технических текстов.

Теоретическая значимость исследования заключается в разработке проблем описания, представления и анализа структуры текста; на этом основании предложены новые подходы к комплексному описанию структуры научно-технических текстов, включающему межуровневые элементы, разработаны

методы извлечения и выравнивания многокомпонентных терминологических единиц и номенклатурных наименований из параллельных научно-технических текстов, разработаны принципы и методы выявления машинно-сгенерированных и машинно-переведенных научно-технических текстов и их фрагментов. Результаты диссертации могут быть использованы при проведении научно-исследовательских работ в области информатики, машинного обучения, инженерии знаний и прикладной лингвистики.

Практическая значимость исследования состоит в применимости предложенных в диссертации подходов и методов к построению эффективных доверенных систем автоматической обработки научно-технических текстов. Значимость подтверждается внедрением результатов диссертационной работы в ряд прикладных систем текстовой аналитики, используемых в коммерческих организациях. Методические результаты работы могут быть использованы в системах автоматической обработки естественных языков для специальных целей, при разработке различных информационно-поисковых систем широкого назначения.

Достоверность и обоснованность результатов подтверждается непротиворечивостью и согласованностью с известными фактами и исследованиями в рассматриваемой области, высокой степенью сходимости теоретических результатов с данными экспериментов и определяется применением теоретических и методологических основ разработок ведущих ученых в области обработки естественного языка, корректным и обоснованным использованием математического аппарата, экспериментальными исследованиями разработанных моделей и методов.

Личный вклад соискателя в основные результаты работы является определяющим. Соискателю принадлежит определяющая роль при решении задач развития теоретических основ создания информационных моделей и методов обработки научно-технических текстов. Соискатель предложил общий подход к извлечению англо- и русскоязычных многокомпонентных терминов на основе синтаксических шаблонов, подкрепленных морфологической информацией о каждой словоформе, а также подход к созданию учебных и специальных словарей

на основе параллельного корпуса научно-технических текстов. Соискатель разработал общий подход к установлению семантических ролей в научно-технических текстах, принципы семантико-синтаксического анализа научно-технических текстов. Соискателю лично принадлежит постановка задачи анализа композиционной структуры научно-технических текстов и проработка основных подходов к их анализу. Соискателем проработаны общие принципы информационного поиска в сложно-структурированных научно-технических текстах на английском и русском языках. Соискателем предложены различные подходы к использованию параллельного корпуса в лингвистике и лингводидактике.

В ходе защиты диссертации были заданы вопросы, уточняющие главные результаты диссертации и используемые обозначения и термины. Был высказан ряд замечаний, в частности следующие:

1. На стр. 26 диссертации указано, что одним из ключевых аспектов проектирования параллельного корпуса научно-технических текстов являются их метаописания. Однако в диссертации не говорится, какие именно стандарты метаописаний использованы.

2. Предложен метод извлечения многокомпонентных терминов, основанный на выделении левого, правого определений и ядерного элемента. Следовало бы привести развернутые примеры типовых корректных или ошибочных извлечений и описать ограничения метода на специфических для научно-технического текста конструкция, что позволило бы точнее очертить его область применимости.

3. В главе 3 не описан отдельный механизм автоматического распознавания важной категории научно-технической лексики, как аббревиатуры и сокращенные обозначения, в следствии чего их корректная обработка фактически выносится на этап экспертной правки, что снижает общий уровень автоматизации решения.

4. В докладе и автореферате используется словосочетание «модель научно-технического текста». Это не точно, поскольку модель — это упрощённое представление объекта или явления, которое сохраняет его существенный смысл, соответственно, невозможно представить модель, например, научной статьи. Правильно было бы говорить «модель структуры научно-технического текста».

5. Предложены методы выявления машинно-сгенерированных текстов, но используемые при этом критерии не описаны.

Соискатель Бутенко Ю.И. ответила на задаваемые ей вопросы и привела собственную аргументацию, удовлетворившую авторов вопросов.

На заседании 19 февраля 2026 года диссертационный совет принял решение присудить Бутенко Ю.И. учёную степень доктора технических наук за разработанные модели, методы и программные средства, которые вносят значительный вклад в развитие в стране эффективных систем автоматической обработки научно-технических текстов.

При проведении тайного голосования диссертационный совет в количестве 22 человек, из них 5 докторов наук по специальности 2.3.8 – «Информатика и информационные процессы», участвовавших в заседании, из 29 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за – 17, против – 1, недействительных бюллетеней 4.

Председатель
диссертационного совета 24.1.224.03
д.ф.-м.н.



К.В. Воронцов

Ученый секретарь
диссертационного совета 24.1.224.03
к.т.н.

И.А. Рейер

19 февраля 2026 г.