

**Заключение диссертационного совета 24.1.172.01, созданного на базе  
Федерального государственного бюджетного научного учреждения  
«Научно-исследовательский институт биомедицинской химии имени  
В.Н. Ореховича», по диссертации на соискание ученой степени  
кандидата наук**

аттестационное дело № \_\_\_\_\_

решение диссертационного совета от 23 апреля 2026 г. № 4

О присуждении Бизюковой Надежде Юрьевне, гражданке РФ, учёной степени кандидата биологических наук.

Диссертация «Формирование знаний о биологической активности низкомолекулярных органических соединений на основе автоматизированного анализа текстов» по специальности 1.5.8. Математическая биология, биоинформатика, принята к защите 12 февраля 2026, протокол № 1, диссертационным советом 24.1.172.01 на базе Федерального государственного бюджетного научного учреждения (ФГБНУ) «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», 119121, Россия, Москва, ул. Погодинская, д. 10, стр. 8, созданного Приказом Минобрнауки России № 105/нк от 11.04.2012 г. с внесением изменений Приказом Минобрнауки России № № 561/нк от 03 июня 2021 г. и Приказом Минобрнауки России № 1959/нк от 12 октября 2023 г.

Соискатель Бизюкова Надежда Юрьевна, 1998 года рождения.

В 2021 году соискатель окончила Федеральное государственное автономное образовательное учреждение высшего образования (ФГАОУ ВО) «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Министерства здравоохранения Российской Федерации. В 2025 году окончила очную аспирантуру по научной специальности 1.5.8. Математическая биология, биоинформатика; отрасль

науки 1.5. – Биологические науки, при ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича». Работает младшим научным сотрудником в лаборатории анализа больших данных для цифровой фармакологии ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича».

Диссертация выполнена в лаборатории анализа больших данных для цифровой фармакологии ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича».

Научный руководитель – кандидат биологических наук Тарасова Ольга Александровна, старший научный сотрудник, заведующая лабораторией анализа больших данных для цифровой фармакологии ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича». Научный консультант – доктор биологических наук, академик РАН Поройков Владимир Васильевич, главный научный сотрудник, заведующий отделом биоинформатики ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича».

Официальные оппоненты:

Орлов Юрий Львович, доктор биологических наук, профессор РАН, ФГАОУ ВО Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет), профессор кафедры информационных технологий и обработки медицинских данных;

Девяткин Дмитрий Алексеевич, кандидат физико-математических наук, Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук", руководитель отдела 73;

ведущая организация – ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», гор. Новосибирск, в своём положительном заключении, подписанном кандидатом биологических наук, заведующим лабораторией компьютерной

протеомики, ведущим научным сотрудником Иванисенко Владимиром Александровичем, высказала ряд замечаний и вопросов, носящих в основном уточняющий и дискуссионный характер, которые касаются устойчивости результатов предложенного интегрального подхода; уточнения границ применимости предложенного подхода; степени экстраполяции полученных оценок полноты извлеченных сведений, представленных в базе данных ChEMBL, на другие классы соединений и иные биомедицинские направления; обработки случаев неполной или неоднозначной нормализации, в особенности для редко встречающихся соединений, новых обозначений и вариантов написания; систематической экспертной валидации созданной базы данных и веб-ресурса. Отмечаются также отдельные опечатки, редакционные и технические неточности по тексту работы. Вместе с тем указанные замечания и вопросы не снижают общей высокой оценки диссертационной работы.

Соискатель имеет 36 опубликованных работ, в том числе по теме диссертации 17 работ, опубликованных в рецензируемых научных изданиях – 5.

Наиболее значимые научные работы по теме диссертации:

1. Biziukova N.Yu., Ivanov S.M., Tarasova O.A. Identification of Proteins and Genes Associated with Hedgehog Signaling Pathway Involved in Neoplasm Formation Using Text-Mining Approach // Big Data Mining and Analytics. – 2024. – Т. 7. – № 1. – С. 107–130. [WoS, Scopus, РИНЦ, Q1, ИФ 6,2] К1, 2,79 усл. печ. л.

Исследование направлено на извлечение и систематизацию информации о белках и генах, связанных с сигнальным путём Hedgehog и развитием опухолевых заболеваний. Автоматизированный анализ релевантных публикаций позволил выявить молекулярные механизмы, вовлечённые в неопластические процессы, и расширить понимание аномальной активации пути Hedgehog, значимой при различных видах рака. Вклад соискателя заключается в разработке и реализации алгоритмов извлечения наименований объектов (белков и генов), включая обучение модели и оценку качества

распознавания. Соискателем сформулирован список правил для поиска функциональной взаимосвязи между белками/генами и различными неопластическими процессами.

2. Тарасова О.А., Бизюкова Н.Ю., Столбова Е.А., Столбов Л.А., Такташов Р.Р., Карасев Д.А., Ионов Н.С., Иванов С.М., Дмитриев А.В., Рудик А.В., Дружиловский Д.С., Соболев Б.Н., Филимонов Д.А., Поройков В.В. Извлечение информации о взаимодействии вирусов с организмом человека и о противовирусных соединениях на основе интеллектуального анализа больших коллекций текстов // Биомедицинская химия. – 2024. – Т. 70. – № 6. – С. 469–474. [Scopus, РИНЦ, ВАК, Q4, ИФ 1,05] К3, 0,7 усл. печ. л.

В работе развивается методология автоматического извлечения данных о взаимодействиях вирусов с организмом человека и соединений с потенциальной противовирусной активностью из больших массивов литературных источников. Полученная информация интегрируется с биомедицинскими базами данных и формирует основу базы знаний для поиска новых терапевтических стратегий и приоритизации соединений-кандидатов. Задачи соискателя включали построение архитектуры алгоритма извлечения знаний, настройку и применение алгоритмов распознавания наименований химических и биологических объектов, обработку больших массивов текста и представление полученных структурированных результатов в формате, удобном для интеграции в базу знаний.

3. Biziukova N.Y., Rudik A.V., Dmitriev A.V., Tarasova O.A., Filimonov D.A., Poroikov V.V. XenoMet: A Corpus of Texts to Extract Data on Metabolites of Xenobiotics // ACS Omega. – 2025. – Т. 10. – № 3. – С. 2459–2471. [WoS, Scopus, РИНЦ, Q1, ИФ 4,3] К1, 1,51 усл. печ. л.

Статья посвящена созданию XenoMet – специализированного корпуса текстов для автоматического извлечения информации о биотрансформации ксенобиотиков: названия субстратов, метаболитов и типов реакций. Корпус формировался на основе автоматического поиска релевантных публикаций PubMed, полуавтоматической аннотации CNER-методами и последующей

ручной верификации. Показаны примеры его применения для построения метаболических деревьев и обучения CRF-моделей извлечения фактов о метаболизме. Вклад соискателя включал формирование корпуса: разработку и выполнение стратегии сбора текстов, участие в аннотации сущностей, создании и корректировке словарей реакций, программной реализации алгоритма распознавания и извлечения ассоциаций, а также анализ качества работы модели.

На диссертацию и автореферат поступили отзывы от:

официального оппонента Орлова Юрия Львовича, доктора биологических наук, профессора РАН, ФГАОУ ВО Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет), профессора кафедры информационных технологий и обработки медицинских данных, в котором имеются некоторые замечания, вопросы и рекомендации. В частности, отмечается, что в работе недостаточно подробно раскрыт вопрос обработки омонимии и неоднозначности терминов; при оценке значимости ассоциаций, в значительной степени опирающейся на частоту их упоминаний в текстах, было бы полезно дополнительно учитывать значимость источников и характер описываемых взаимосвязей; при описании веб-сервиса SigNER, предназначенного для распознавания сущностей; нет пояснения, каким образом пользователь может воспользоваться функционалом извлечения ассоциаций. Представленные в обзоре литературы примеры практического применения методов интеллектуального анализа текстов в биомедицинских исследованиях стоило бы дополнить российскими разработками такими, как ANDSystem, ANDVisio и ANDDigest. Указывается также на некоторые технические и редакционные неточности. Однако высказанные замечания и вопросы носят уточняющий характер, не снижают общей положительной оценки работы и не умаляют значимости полученных результатов;

официального оппонента Девяткина Дмитрия Алексеевича, кандидата физико-математических наук, Федерального государственного учреждения "Федеральный исследовательский центр "Информатика и управление" Российской академии наук", руководителя отдела 73, в котором имеются вопросы и замечания. Отмечается, что в диссертации недостаточно подробно раскрыт вопрос распространения ошибок между последовательными этапами обработки текстов; недостаточно подробно обсуждён риск появления ложноположительных ассоциаций при их извлечении, особенно в случае текстов с высокой плотностью терминов; полезно было бы более чётко развести ассоциации, извлекаемые по семантическим шаблонам, и ассоциации, основанные лишь на совместном упоминании; на стр. 40 отсутствует описание параметров оценок качества классификации текстов и их фрагментов (TP, FN, FP). Перечисленные вопросы и замечания носят в основном уточняющий характер, не снижают общей положительной оценки работы и не умаляют значимости полученных результатов.

В положительном отзыве на автореферат кандидата физико-математических наук, старшего научного сотрудника кафедры биофизики физического факультета ФГАОУ ВО «Московский государственный университет имени М.В. Ломоносова» Генералова Евгения Александровича отмечается, что отдельные положения автореферата могут быть предметом дальнейшего обсуждения. В частности, это касается ограничений предложенного метода при извлечении ассоциаций из текстов с высокой лингвистической сложностью; преимуществ подхода на основе MeSH-терминов при формировании коллекций публикаций; анализа достоверности результатов проведённого сопоставления с базой ChEMBL; описания методов машинного обучения для отбора без детальной информации о параметрах моделей и процедуре их настройки. Указанные замечания носят частный характер и не снижают общей высокой оценки выполненного исследования.

В положительном отзыве на автореферат кандидата биологических наук, старшего научного сотрудника лаборатории радионуклидных и лучевых технологий в экспериментальной онкологии Научно-исследовательского института клинической и экспериментальной радиологии ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Минздрава России Поздняковой Натальи Владимировны отмечено, что в автореферате не обсуждается достоверность новых ассоциаций и возможность ложноположительных результатов при сопоставлении с базой ChEMBL; не рассматривается влияние качества исходных публикаций на результаты анализа; ограниченно представлено обсуждение того, насколько полно извлекаются отрицательные результаты. Указанные замечания носят частный характер и не снижают общей положительной оценки работы.

В положительном отзыве на автореферат доктора химических наук, ведущего научного сотрудника лаборатории физиологически активных веществ ФГБУН Новосибирский институт органической химии им. Н.Н. Ворожцова Сибирского отделения Российской академии наук Яровой Ольги Ивановны есть некоторые замечания. Отмечается, что было бы интересно подробнее рассмотреть функциональные возможности разработанного веб-интерфейса как практического аспекта результатов работы; указывается на некорректность терминологии в таблице 1; спрашивается о временных и вычислительных затратах на полный цикл анализа при обработке коллекции из более чем 400 тысяч публикаций по противовирусной активности.

В положительном отзыве на автореферат доктора медицинских наук, доцента, заместителя руководителя Центра предиктивной генетики, фармакогенетики и персонализированной терапии ФГБНУ «Российский научный центр хирургии им. акад. Б.В. Петровского» Мирзаева Карина Бадавиевича есть отдельные замечания. Указано, что целесообразно было бы более подробно обсудить ограничения предложенного метода и условия его применимости, а также расширить сравнительный анализ с существующими

подходами в области извлечения биомедицинских знаний из текстов. Отмеченные замечания носят рекомендательный характер и не влияют на общую положительную оценку работы.

Выбор официальных оппонентов и ведущей организации обоснован тем, что Орлов Юрий Львович – один из ведущих специалистов в области биоинформатики, компьютерной геномики, в области анализа данных высокопроизводительного геномного секвенирования, анализа интерактома, цифровой медицины; Девяткин Дмитрий Алексеевич – ведущий специалист в области методов машинного обучения и анализа текстов, распределённых систем обработки информации, систем искусственного интеллекта;

ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук» - один из ведущих научных центров, занимающийся фундаментальными и прикладными разработками в области интеграция молекулярных, клеточных, онтогенетических и популяционных исследований для понимания генетических механизмов изменчивости и эволюции. Одним из основных направлений института являются медико-биологические исследования по разработке на основе современных методов биоинформатики, молекулярной и клеточной биологии подходов к выявлению новых фармакологических мишеней и созданию лечебных препаратов нового поколения.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- впервые разработан целостный подход к автоматизированному извлечению сведений о биологической активности низкомолекулярных органических соединений, ориентированный на последовательное решение взаимосвязанных задач обработки медицинских текстов;
- создана база знаний, позволяющая аккумулировать и структурировать информацию о химических соединениях и их биологических эффектах, что делает возможным её использование при решении задач медицинской химии и фармакологии.

Теоретическая значимость исследования обоснована тем, что работа связана с развитием подходов к системному представлению знаний, извлекаемых из биомедицинских текстов; показано, каким образом разнородные сведения, содержащиеся в научных публикациях могут быть приведены к единой структуре, обеспечивающей их сопоставимость и дальнейший анализ. Получены новые научные результаты, имеющие существенное значение для развития методов математической биологии, биоинформатики и автоматизированного анализа биомедицинских текстов.

Значение полученных соискателем результатов исследования для практики определяется возможностью использования разработанных решений для автоматизированного извлечения информации из больших массивов научных публикаций, выявлять закономерности в действии соединений, развитии заболеваний и др. Данный подход может быть использован, в частности, для разработки диагностических систем, новых лекарств или их репозиционирования.

Оценка достоверности результатов исследования подтверждается использованием современных разнообразных методов статистической обработки, использованием данных из публикаций реферируемых журналов, валидированием созданного подхода на нескольких репрезентативных выборках из разных областей биомедицинских данных. Результаты диссертации опубликованы в реферируемых журналах и докладывались на российских и международных конференциях, что свидетельствует об их значимости и востребованности.

Личный вклад соискателя состоит в проведении анализа литературы по теме работы, создании коллекций публикаций по разным темам, разработке метода по автоматическому извлечению данных из научных публикаций и его программной реализации. Автором проведена валидация предлагаемого подхода, оценена его точность и воспроизводимость.

В ходе защиты диссертации не было высказано критических замечаний.

На заседании 23 апреля 2026 г. диссертационный совет принял решение присудить Бизюковой Надежде Юрьевне учёную степень кандидата биологических наук за решение актуальной задачи по созданию автоматизированной системы интеллектуального анализа текстов научных публикаций для выявления и структурирования данных о биологической активности химических соединений и поиска новых направлений фармакологического применения известных лекарств.

При проведении тайного голосования диссертационный совет в количестве 20 человек, из них 9 докторов наук по специальности 1.5.8. Математическая биология, биоинформатика, участвовавших в заседании, из 23 человек, входящих в состав совета, проголосовали: за – 20, против - 0, недействительных бюллетеней – 0.

Зам. председателя Диссертационного совета  
доктор биологических наук, профессор

Ученый секретарь Диссертационного совета  
кандидат химических наук

23 апреля 2026 года



Медведев А.Е.

Карпова Е.А.