

## **ОТЗЫВ**

официального оппонента д.б.н., проф. РАН, Орлова Юрия Львовича  
на диссертационную работу  
Бизюковой Надежды Юрьевны

«Формирование знаний о биологической активности низкомолекулярных органических соединений на основе автоматизированного анализа текстов», представленную на соискание  
ученой степени кандидата биологических наук по специальности  
1.5.8. – Математическая биология, биоинформатика

Диссертационная работа Бизюковой Надежды Юрьевны посвящена разработке методов автоматизированного анализа биомедицинских текстов с целью извлечения и систематизации сведений о биологической активности низкомолекулярных органических соединений. Представленная работа соответствует специальности “Математическая биология, биоинформатика”. Особый интерес представляют предложенные в работе подходы к интеграции методов обработки естественного языка и машинного обучения, позволяющие получать целостную картину взаимосвязей между химическими соединениями и их биологическими эффектами на основе анализа больших массивов научных публикаций.

**Актуальность** диссертационного исследования Н.Ю. Бизюковой обусловлена стремительным ростом объема биомедицинской информации, объема экспериментальных данных и публикаций, необходимостью разработки эффективных методов её автоматизированного анализа. Создание инструментов интеллектуального анализа текстов, позволяющих систематизировать знания о биологической активности химических соединений, значимо для решения биологических задач. Особенно важна разработка независимых отечественных программных инструментов, инструментов интеллектуального анализа текстов, позволяющих систематизировать знания о биологической активности химических соединений.

### **Структура диссертационной работы**

Текст диссертации Н.Ю. Бизюковой изложен на 150 страницах и включает в себя 17 рисунков и 12 таблиц. Работа содержит три главы, включает введение, обзор литературы, описание материалов и методов, результатов работы и их обсуждения, заключение и выводы. Кроме того, текст содержит Список сокращений, Список литературы (188 источников) и 2 Приложения.

### **Характеристика диссертации**

Обзор литературы содержит полное и систематизированное изложение современных подходов к интеллектуальному анализу биомедицинских текстов, включая методы распознавания сущностей и извлечения ассоциаций. Автором подробно рассмотрены

существующие алгоритмы и инструменты, что позволяет обоснованно подвести к постановке целей и задач исследования.

Раздел «Материалы и методы» изложен детально, описаны используемые подходы к формированию корпуса текстов, методы обработки естественного языка и алгоритмы извлечения информации, а также принципы построения базы данных. Результаты работы Н.Ю. Бизюковой изложены последовательно, с логичным переходом от описания отдельных этапов метода к его практическому применению. Приведены примеры использования разработанного подхода для анализа биомедицинских данных.

Текст диссертации в целом написан понятным научным языком, материал структурирован и легко воспринимается. Иллюстрации и таблицы выполнены аккуратно и наглядно, что способствует лучшему восприятию представленных результатов. В работе встречаются отдельные опечатки и незначительные стилистические неточности, не влияющие на общее качество изложения.

**Научная новизна** диссертационной работы Н.Ю. Бизюковой заключается в разработке интегрального подхода к автоматизированному анализу биомедицинских текстов, обеспечивающего извлечение и структурирование сведений о биологической активности низкомолекулярных органических соединений. В отличие от существующих решений, ориентированных, как правило, на отдельные типы сущностей или ограниченные виды взаимосвязей, в представленной работе реализован подход, позволяющий получать целостное представление о биологической активности соединений, включая не только конечные эффекты, но и механизмы их действия, представленные в виде цепочек и каскадов биологических процессов.

Дополнительным аспектом новизны является возможность масштабирования разработанного метода за счёт расширения набора правил и паттернов для извлечения ассоциаций без существенного изменения архитектуры подхода или переобучения моделей, что делает его перспективным для дальнейшего развития и адаптации к смежным задачам.

### **Теоретическая и практическая значимость**

Теоретическая значимость работы Н.Ю. Бизюковой заключается в формализации подхода к извлечению взаимосвязей между биологическими объектами из неструктурированных текстовых данных и развитию методов их системного анализа в рамках задач биоинформатики. Практическая значимость определяется созданием российского инструмента (доступного онлайн на сайте Института <https://way2drug.com/>), позволяющего автоматизировать получение и структурирование сведений о биологической активности химических соединений. Такие автоматически структурированные сведения могут быть использованы при анализе научной литературы и поиске новых направлений фармакологического применения известных соединений.

## **Обоснованность научных выводов и положений**

Объем проведенных исследований и вычислительных экспериментов является достаточным для получения достоверных результатов, а сформулированные в работе выводы согласуются с полученными данными. Обоснованность научных положений подтверждается использованием репрезентативных коллекций биомедицинских текстов, применением современных методов обработки естественного языка и машинного обучения, а также сопоставлением извлеченной информации с известными экспериментальными и клиническими данными. Достоверность результатов обеспечивается корректностью примененных методических подходов и их проверкой на независимых выборках. Основные положения диссертационной работы были опубликованы (6 статей в рецензируемых журналах, всего 17 публикаций), представлены и обсуждены на российских и международных научных конференциях и симпозиумах, что свидетельствует о научной значимости и признании полученных результатов научным сообществом.

## **Вопросы и замечания**

1. В работе используется комбинированный подход к распознаванию и нормализации наименований объектов, однако представляется недостаточно подробно раскрытым вопрос обработки омонимии и неоднозначности терминов. В частности, не вполне ясно, каким образом осуществляется нормализация аббревиатур, которые могут иметь несколько расшифровок в зависимости от контекста.
2. Интересным направлением развития работы могло бы стать введение механизма ранжирования извлеченных ассоциаций. В текущем виде оценка значимости ассоциаций в значительной степени опирается на частоту их упоминания в текстах. Вместе с тем, учитывая различную научную ценность источников, а также различия в модальности утверждений (например, подтвержденные эффекты, предположительные связи, отрицательные результаты), было бы полезно дополнительно учитывать значимость источников и характер описываемых взаимосвязей.
3. В диссертации описан веб-сервис SigNER, предназначенный для распознавания наименований сущностей, однако нет пояснения, каким образом пользователь может воспользоваться функционалом извлечения ассоциаций.
4. При описании биологической активности химических соединений основной акцент сделан на качественных характеристиках. Включение количественных параметров (таких как IC<sub>50</sub>, K<sub>i</sub>, EC<sub>50</sub> и др.) могло бы существенно расширить практическую ценность разработанной системы.
5. В работе основное внимание уделено анализу заголовков и аннотаций. В качестве перспективного направления развития стоит отметить возможность более широкого

использования полных текстов научных публикаций. Включение анализа полных текстов позволило бы извлекать более детализированную и полную информацию.

6. Следует отметить, что в обзоре литературы сравнительно ограниченно представлены примеры практического применения методов интеллектуального анализа текстов в биомедицинских исследованиях. Стоило бы отметить российские разработки ANDSystem, ANDVisio, ANDDigest (группа В.А.Иванисенко, ИЦиГ СО РАН, Новосибирск).
7. В обзоре литературы не рассмотрены современные подходы к анализу текстов с использованием больших языковых моделей (LLM). Сравнение предложенного метода с подобными подходами представляло бы интерес с точки зрения оценки эффективности и перспектив дальнейшего развития.
8. В тексте диссертации выявлены отдельные редакционные и технические неточности. В частности, на стр. 9, по-видимому, допущена ошибка в числовых данных, характеризующих динамику количества научных журналов “ежегодно количество журналов в период с 2015 по 2024 год увеличивалось: от 395 (в 2015 году) до 56 (в 2024 году) (рисунок 1)” – речь идет о динамике, но не числе журналов вообще. На стр. 25 присутствует редакционная вставка (уточнение в скобках), вероятно не удалённая при окончательном редактировании текста. На стр. 26–27 неточность в приведённом примере биомедицинских терминов («трансформирующий фактор роста бета (Epidermal Growth Factor Receptor (EGFR))»). Кроме того, в тексте диссертации присутствует пустая страница (стр. 141), не несущая смысловой нагрузки. В целом по работе встречается незначительная неоднородность в использовании русскоязычной и англоязычной терминологии. В тексте автореферата приведены ссылки на научную литературу, не указанную в автореферате (например (Poroikov V. et al., 2019; Аладышева Ж. И. и др., 2019)), но обсуждаемую в основном тексте диссертации. Следовало бы избегать ссылок на литературу в автореферате или привести их кратко в конце самого автореферата.

Высказанные замечания и вопросы носят уточняющий характер, не снижают общей положительной оценки работы и не умаляют значимости полученных результатов.

## **Заключение**

Содержание диссертации Н.Ю. Бизюковой полностью отражено в публикациях автора в журналах, рекомендованных ВАК, а также в других рецензируемых изданиях, и представлено на международных и российских научных конференциях. Основные результаты представлены в автореферате, который адекватно отражает содержание диссертации. Поставленные в диссертации задачи решены полностью. Таким образом, диссертационная работа Бизюковой Надежды Юрьевны «Формирование знаний о биологической активности низкомолекулярных органических соединений на основе автоматизированного анализа текстов» является завершённой научно-квалификационной работой, в которой решена научная задача, имеющая

значение для развития биоинформатики и математической биологии. Работа соответствует пункту 15 паспорта специальности 1.5.8 – «Математическая биология, биоинформатика» (математический и компьютерный анализ биомедицинских текстов, извлечение информации о биологических объектах и их взаимосвязях). По своей научной новизне, теоретической и практической значимости диссертация соответствует требованиям пункта 9 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842 (в действующей редакции), предъявляемым к диссертациям на соискание ученой степени кандидата наук. Автор диссертации, Бизюкова Надежда Юрьевна, заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8 – «Математическая биология, биоинформатика».

**Официальный оппонент:**

Орлов Юрий Львович, д.б.н., профессор РАН,  
профессор кафедры информационных технологий и обработки медицинских данных ФГАОУ  
ВО Первый МГМУ имени И.М. Сеченова Минздрава России (Сеченовский Университет).  
ул. Трубецкая, д. 8, стр. 2, г. Москва, 119048, Россия.  
Email y.orlov@sechenov.ru Тел. +7 (495) 609-14-00, +7(499) 248-01-81

Доктор биологических наук, профессор РАН

Орлов Юрий Львович



A handwritten signature in blue ink, corresponding to the name "Орлов Юрий Львович" mentioned in the text above.

27/03/2026 г.

(Специальность 1.5.8 – Математическая биология, биоинформатика)