

ОТЗЫВ

на автореферат диссертации Бизюковой Надежды Юрьевны
«Формирование знаний о биологической активности низкомолекулярных органических соединений», представленной на соискание ученой степени кандидата биологических наук по специальности 1.5.8 – Математическая биология, биоинформатика.

Диссертационная работа Бизюковой Надежды Юрьевны посвящена разработке методов получения структурированных знаний о биологической активности низкомолекулярных органических соединений на основе анализа биомедицинских текстов. В современных условиях ключевой проблемой является не столько дефицит данных, сколько их фрагментированность и распределённость по большому числу публикаций, что существенно затрудняет их комплексное использование в научных и прикладных задачах. В этой связи особую значимость приобретают подходы, позволяющие не только извлекать информацию из текстов, но и формировать целостные и согласованные представления о механизмах действия химических соединений.

В работе предложен комплексный подход, охватывающий полный цикл преобразования текстовой информации в структурированные данные: от формирования выборок публикаций до извлечения и последующей нормализации сведений. Значимым вкладом является обоснованный выбор стратегии формирования текстовых коллекций, сочетающей использование экспертно заданных критериев (в частности, MeSH-аннотаций) и методов машинного обучения. Такой подход позволяет учитывать как содержательную релевантность текстов, так и масштабируемость решения при переходе к различным предметным областям.

Существенным элементом исследования является разработка подхода к распознаванию биомедицинских объектов, ориентированного на одновременную работу с различными типами сущностей. В отличие от узкоспециализированных решений, предложенная комбинация статистических моделей и правил, учитывающих формальные особенности записи отдельных категорий объектов, обеспечивает устойчивость метода при анализе разнородных текстовых данных и снижает зависимость от специфики корпуса.

Отдельно следует отметить разработанный метод выявления взаимосвязей между объектами, позволяющий переходить от фиксации отдельных фактов к реконструкции функциональных взаимодействий. Использование лингвистических шаблонов в сочетании с количественной оценкой контекстной близости сущностей способствует повышению достоверности извлекаемых ассоциаций. Включение этапа нормализации, основанного на сопоставлении с внешними онтологиями и базами данных, обеспечивает интерпретируемость результатов и возможность их интеграции в существующие информационные ресурсы.

Практическая значимость диссертационного исследования определяется не только разработкой алгоритмического подхода, но и его доведением до уровня прикладного инструмента. В работе реализована база данных, обеспечивающая хранение нормализованных сведений о химических соединениях, биологических объектах и выявленных между ними связях, с сохранением контекста их извлечения.

Преимуществом является возможность автоматического пополнения ресурса по мере обработки новых публикаций. Создание веб-интерфейса расширяет доступ к результатам исследования и делает их пригодными для использования в широком спектре задач – от анализа молекулярных механизмов до поиска перспективных направлений фармакологических исследований.

Важным результатом является проведённый анализ согласованности извлечённых данных с существующими фактографическими ресурсами. Показано, что предложенный подход способен воспроизводить значительную часть уже известных сведений, при этом выявляя дополнительные ассоциации, не отражённые в базах данных. Это указывает на потенциал разработанного метода как инструмента не только для агрегации информации, но и для выявления новых знаний, расширяющих текущие представления о биологической активности соединений и их взаимодействии с биологическими системами.

Следует отметить высокий уровень методической проработки исследования. Автором продемонстрировано уверенное владение современными подходами интеллектуального анализа текстов, а также умение комбинировать методы различной природы – от статистических моделей до правил, основанных на формальных признаках. Выбранные методы адекватны поставленным задачам, а достоверность полученных результатов подтверждена использованием стандартных метрик оценки качества, анализом ошибок и сопоставлением с независимыми источниками данных. Логика изложения материала и формулировка выводов свидетельствуют о целостности и внутренней согласованности проведённого исследования.

Результаты работы прошли апробацию в научном сообществе и отражены в публикациях в рецензируемых изданиях, а также в материалах российских и международных конференций. Это подтверждает их востребованность и соответствие современному уровню исследований в области биоинформатики и медицинской химии.

Автореферат оформлен на высоком научном уровне, последовательно отражает цели, задачи, методы и основные результаты диссертационной работы и позволяет сформировать полное представление о её содержании и значимости.

Отдельные положения автореферата могут быть предметом дальнейшего обсуждения.

1. В автореферате недостаточно подробно раскрыты ограничения предложенного метода при извлечении ассоциаций из текстов с высокой лингвистической сложностью (например, при наличии отрицаний, модальных конструкций или противоречивых утверждений).
2. В разделе, посвящённом формированию коллекций публикаций, показано преимущество подхода на основе MeSH-терминов, однако остаётся не до конца раскрытым вопрос чувствительности метода к качеству индексирования публикаций и потенциального смещения выборки, связанного с неполной или запаздывающей MeSH-аннотацией.
3. Несмотря на использование нормализации сущностей с привлечением внешних баз данных (например, ChEMBL, Disease Ontology), в автореферате не обсуждаются случаи неоднозначного сопоставления и их влияние на

- согласованность базы данных, что является важным аспектом при интеграции разнородных источников.
4. Проведённое сопоставление с базой ChEMBL демонстрирует как совпадения, так и значительное число новых ассоциаций, однако в автореферате ограниченно представлен анализ достоверности этих новых связей (например, через экспертную валидацию или независимые источники)
 5. При описании методов машинного обучения для отбора не приводится детальной информации о параметрах моделей и процедуре их настройки, что ограничивает воспроизводимость экспериментов.

Указанные замечания носят частный характер и не снижают общей высокой оценки выполненного исследования.

Представленные в автореферате материалы позволяют заключить, что диссертационная работа Бизюковой Надежды Юрьевны «Формирование знаний о биологической активности низкомолекулярных органических соединений» является завершённым и методически обоснованным научным исследованием, выполненным на высоком научном уровне. Работа соответствует требованиям пункта 9 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842. Автор заслуживает присуждения ученой степени кандидата биологических наук по специальности 1.5.8 – Математическая биология, биоинформатика.

Отзыв подготовил:

К.ф.-м.н., с.н.с. кафедры биофизики физического факультета
МГУ имени М.В.Ломоносова
Генералов Евгений Александрович

«02» апреля 2026г.

Контактные данные:

ФГБОУ ВО «Московский государственный университет имени М.В.Ломоносова»,
физический факультет, кафедра биофизики
119991, ГСП-1, Москва, Ленинские горы, МГУ имени М.В. Ломоносова, д. 1, стр. 2
Телефон: 8 (966) 058-20-28
E-mail: generalovea@my.msu.ru



Подпись *Генералова Е.А.*
УДОСТОВЕРЯЮ *Архангельская*
ЗАВ.КАНЦЕЛЯРИЕЙ
В.В. АРХАНГЕЛЬСКАЯ