

Создание информационной системы поддержки принятия врачебных решений на основе методов доказательной медицины

Г.С. Лебедев^{1,2}, Э.Н. Фартушный¹, И.А. Шадеркин^{1,2}, Г.С. Клименко^{1,3},
И.В. Рябков^{1,2}, П.Б. Кожин¹, К.А. Кошечкин¹, Г.П. Радзиевский¹, И.В. Фомина²

¹ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия

²ФГБУ ЦНИИОИЗ Минздрава России, Москва, Россия

³Фонд развития цифровой экономики, Москва, Россия

Для корреспонденции:

Лебедев Георгий Станиславович, geramail@rambler.ru

Bilding of the medical decision support system on the basis of providing medicine based on evidence-based medicine

G.S. Lebedev^{1,2}, E.N. Fartushniy¹, I.A. Shaderkin^{1,2}, H.S. Klimenko^{1,3}, I.V. Ryabkov^{1,2}, ,
P.B. Kozhin¹, K.A. Koshechkin¹, G.P. Radzievskiy¹, I.V. Fomina²

¹I.M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

²Federal Research Institute for Health Organization and Informatics, Moscow, Russia

³Digital Economy Development Fund, Moscow, Russia

This article presents a new study related to the creation of a medical decision support system with an intellectual analysis of scientific data (texts of medical care standards, clinical guidelines, instructions for the use of medicines, scientific publications of evidence-based medicine). Such a system is designed to provide the possibility of making medical decisions in pharmacotherapy, taking into account personalized medical data due to the optimal prescription of medicines and the use of medical technologies, reducing the frequency of undesirable reactions while using two or more drugs for different indications. The technical goal of the study is to create an intelligent automated information system to support the adoption of medical decisions and its implementation in clinical practice.

Key words: Medical decision support systems, evidence-based medicine, intellectual systems, text mining

Важнейшей проблемой системы здравоохранения в рамках реализации Стратегии развития здравоохранения до 2025 является обеспечение доступности, эффективности и безопасности медицинской помощи, оказываемой населению Российской Федерации. Важной проблемой является тот факт, что, организации и учреждения, участвующие в обращении лекарствен-

ных средств и использовании медицинских технологий, часто не используют методологию оценки эффективности их применения. В стратегии научно-технологического развития России до 2035 г, утвержденной в декабре 2016 г., президентом РФ В.В. Путиным, сказано: «В ближайшие 10-15 лет приоритетами научно-технологического развития Российской Федерации следует считать те направления, которые поз-

волят получить научные и научно-технические результаты и создать технологии, являющиеся основой инновационного развития внутреннего рынка продуктов и услуг, устойчивого положения России на внешнем рынке, и обеспечат:

(а) переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта;

(в) переход к персонализированной медицине, высокотехнологичному здравоохранению и технологиям здоровье-сбережения, в том числе за счёт рационального применения лекарственных препаратов [1].

К приоритетным проектам, согласно Государственной программе Российской Федерации «Развитие здравоохранения», относится развитие и внедрение инновационных методов диагностики, профилактики и лечения, а также основ персонализированной медицины; совершенствование процессов организации медицинской помощи на основе внедрения информационных технологий [2]. В рамках этой программы начата реализация Федерального проекта «Создание цифрового контура здравоохранения на основе государственной информационной системы в сфере здравоохранения Российской Федерации». Создание цифрового контура подразумевает, в том числе, и внедрение интеллектуальных систем поддержки принятия врачебных решений.

Стремительное развитие информационных технологий, цифровизация больших объемов баз данных требуют принципиально новых подходов и постоянного совершенствования методов их анализа. Постоянно обновляющиеся объемы биомедицинской информации и растущее в геометрической прогрессии количество новых публикации требуют разработки эффективных и качественных методов тематического рубрицирования документов и категоризации, извлечение фактов и знаний. Приоритетным фундаментальным направлением работы методов интеллектуального анализа научных публикаций является их использование в системах принятия информированных врачебных решений, включающих превентивную и персонализированную медицину. Знание результатов последних исследований клинического применения лекарственных препаратов и оценка диагностической точности высокотехнологичных медицинских методов, позволяет существенно повысить эффективность оказания медицинской помощи и экономить значительные материальные ресурсы.

Интеллектуальные методы поиска и извлечения знаний из больших баз данных позволят врачам-специалистам проводить высокоэффективное лечение и мониторинг его результатов с позиций персонализированной медицины.

По состоянию на 2019 год, существует большое количество баз медицинских данных, разработанных на принципах доказательной медицины – более 30 миллионов ссылок, проиндексированных через PubMed – крупнейшую базу биомедицинской литературы, разработанную и поддерживаемую Национальным центром биотехнологической информации (NCBI). Поисковая система Entrez NCBI, интегрированная с PubMed, предоставляет доступ к разнообразному набору из 38 баз данных [3]. В настоящее время PubMed индексирует публикации из 5254 журналов по биологии и медицине, начиная с 1948 года. На современном этапе PubMed служит основным инструментом для поиска биомедицинской литературы. Ежедневно системой обрабатывается несколько миллионов запросов, формируемых пользователями, с целью быть в курсе последних достижений и выделять приоритетные исследования в своих областях.

Несмотря на предоставление PubMed-ом эффективных интерфейсов поиска, его пользователям становится все сложнее находить информацию, соответствующую их индивидуальным потребностям. Подробные пользовательские запросы порождают поисковую выдачу, содержащую тысячи релевантных документов.

Популярной базой достоверной медицинской информации является Cochran Library – база знаний, формируемая независимой сетью исследователей, состоит из более чем 37 тысяч ученых в 130 странах мира. Результаты исследований в виде систематических обзоров и мета-анализов рандомизированных исследований публикуются в базе данных Кокрейновской библиотеки [4].

Большое количество медицинских русскоязычных статей, результатов клинических исследований находится в Центральной научной медицинской библиотеке (ЦНМБ) Первого МГМУ им. И.М. Сеченова, которая является русскоязычным аналогом PubMed. ЦНМБ является головной медицинской библиотекой России [5].

По данным Clarivate Analytics, другая база цитирования научных публикаций Web of Science Core Collections содержит более 1,4 миллиарда ссылок из более 20 тысяч источников публикаций (<https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/>; февраль 2019 г.).

База данных Microsoft Academic Search (<https://academic.microsoft.com>; февраль 2019 ►►

года) по разделу Медицина составляет более 26 миллионов публикаций и более 17 миллионов по разделу Биология. При этом наиболее стремительный рост объемов наблюдается по разделу Биохимия – более 5 млн публикаций, из которых более миллиона относятся к категории Генетика.

Google Scholar не сообщает об объеме ссылок, которые можно идентифицировать с помощью их поисковой системы, однако по запросу “персонализированная медицина” (<https://scholar.google.ru/scholar?hl=ru&asdt=0%2C5&q=personalized+medicine&btnG>; февраль 2019 года) объем поисковой выдачи составил 1 млн. 210 тыс. документов.

Примерное количество ежедневных публикаций только в указанных индексируемых базах данных по медицинской тематике уже сегодня превысило 15 тыс. публикаций/день.

Указанные ресурсы служат информационной базой для создания современных систем поддержки клинических решений, основанных на доказательных резюме информации – DumaMed (США), UptoDate (США), EBMG- Evidence Based Medicine Guidelines (Финляндия) и подобных.

Очевидно, что обработка больших объемов информации без применения специальных методов анализа биомедицинского контента просто невозможна. В последнее десятилетие методы интеллектуального анализа текста (в англоязычной транскрипции обозначаемый TextMining) широко используются для решения задач извлечения информации из массивов и коллекций текстовых документов [6,7].

В настоящее время не существует базового сборника описания принципов анализа информации, который можно считать методическим руководством по интеллектуальному анализу биомедицинского текста. В ряде работ описываются фундаментальные методы обработки естественного языка подачи информации [8,9].

В рамках Дорожной карты Национальной Технологической Инициативы «Хелснет» указана необходимость создания систем поддержки принятия решений (СПИР) в сфере превентивной медицины с использованием технологий эволюционного моделирования, цифровой модели знаний о здоровье человека и свойствах средств коррекции, обработки больших объемов данных и индивидуального мониторинга функционального состояния, а также телемедицинских консультаций населению. Системы поддержки принятия решений (СПИР) – класс специализированных медицинских информационных систем, зарегистрированных в установленном порядке для медицинского приме-

нения и являющихся частью медицинских информационных систем для локального использования в медицинских организациях [10].

Современные методы интеллектуального анализа биомедицинских текстов направлены на решение отдельных задач из перечисленного ниже списка:

- Информационный поиск (information retrieval);
- Распознавание (named entity recognition) и идентификация именованных сущностей (named entity identification);
- Выявление связей между объектами (association extraction) [11];
- Выявление связей между событиями (event extraction) [12];
- Идентификация и выявление процессов (pathway extraction).

В настоящее время информационный поиск ограничивается представлением коллекций документов из наборов неструктурированных данных, удовлетворяющих информационному запросу. Он не решает задачу анализа информации и выявления скрытых семантических шаблонов, что является основной целью интеллектуального анализа текста. Информационный поиск связан с известными проблемами поиска соответствующих документов в ответ на конкретную потребность в информации. Традиционный подход контекстного поиска по предварительно индексированным документам не обеспечивает поисковую выдачу нужного качества и требует от пользователя дополнительных временных затрат на изучение предоставленных документов.

Исследование подходов к динамическому тематическому моделированию является активной областью исследований [13-18]. Однако опыт практического применения метода для выявления тематики медицинских документов нам неизвестен. По всей видимости, это связано со специфической лингвистического обеспечения задач анализа биомедицинского контента. Несмотря на наличие наднациональных тезаурусов и онтологий (ICD-10, MeSH, SNOMED-CT, LIONC) выделение именованных сущностей (болезни, гены, белки, лекарственные препараты, клинические анализы), токенизация биомедицинской литературы остается серьезной проблемой, обусловленной в том числе несогласованностью названий известных сущностей, например симптомы, лекарства, и др., нестандартных сокращений, специфики разных языков. Идентификация именованных объектов позволит связать объекты, представляющие интерес, с информацией, которая не детализирована в публи-

кации. Нормализация именованных сущностей на различных языках представления информации является важнейшим условием для построения эффективных предсказательных тематических моделей по направлению персонализированной медицины.

Кроме этого, мы полагаем, что для выявления приоритетных направлений развития превентивной и персонализированной медицины репозиторий обрабатываемых документов не должен ограничиваться только научными публикациями.

В практической библиометрике для описания документов, не опубликованных научными издательствами, используется термин «серая» литература и он может образовывать жизненно важный компонент обзоров фактических данных, таких как систематические обзоры и систематические карты [19], экспресс-оценки фактов [20] и краткие обзоры [21]. «Серая» литература в самом общем случае включает [19]:

- отчеты о научно-исследовательских работах и опытно-конструкторских разработках (НИОКР);
- кандидатские и докторские диссертации;
- труды конференций;
- патенты;
- выпускные квалификационные работы учащихся зарубежных престижных учебных заведений, специализирующихся в данной тематике;
- рецензии и отзывы, материалы экспертиз;
- документы, материалы и публикации специализированных веб-сайтов;
- доклады и выступления, рабочие материалы семинаров, круглых столов, препринты;
- документы тендеров и конкурсов, контрактов, договоров и соглашений;
- аналитические прогнозы, учебно-методические материалы, плакаты, слайды, иллюстрации презентаций обзоры и другие материалы.

По нашему мнению, «серая» литература может оказаться весьма полезной в задачах поиска или синтеза новых направлений и тем исследований, несмотря на то, что она не публикуется официально так же, как традиционная академическая литература, например [19]. Как правило, требуются значительные усилия для поиска «серой» литературы в попытке включить данные, хранимые практиками, а также для учета возможного смещения публикаций. Смещение публикаций – это тенденция к тому, что значимые позитивные исследования будут с большей вероятностью публиковаться, чем несущественные или негативные исследования, что приводит к повышенной вероятности завышения величины эффекта в мета-анализе и других синтезах [22].

Включение «серой» литературы в коллекцию анализируемых документов является дополнительным условием при изучении и выявлении трендов и направлений развития персонализированной медицины. Благодаря включению в изучаемую выборку всех доступных документированных данных обеспечивается более точное прогнозирование и снижается подверженность предвзятости.

В инициализированном нами исследовании мы предлагаем разработать метод выявления перспективных направлений развития персонализированной медицины посредством построения мультязычной динамической вероятностной тематической модели (Dynamic probabilistic topic model) на коллекции изучаемых документов. Модель будет носить мультязычный характер, учитывать n-язычный словарь (по количеству видов языков представленных в отобранной коллекции документов) и онтологических связей между документами сравнимых коллекции.

■ АКТУАЛЬНОСТЬ ИССЛЕДОВАНИЯ

Системный анализ отечественной и зарубежной практики диагностики и фармакотерапии позволил выделить следующие основные факторы, определяющие эффективность и безопасность медикаментозного лечения.

Первый фактор характеризуется практикой назначения лекарственных средств с низким уровнем доказанной клинической эффективности. По экспертным оценкам эффективность применения фармакотерапии в России составляет не более 60%. Согласно стратегии развития здравоохранения до 2025 ставится задача по снижению смертности в трудоспособном возрасте до 380 на 100 тысяч человек. В настоящее время показатель смертности по причине врачебных ошибок достигает 100 тыс. смертей в год в РФ [<http://svpressa.ru/society/article/63391/>]. Даже применение лекарственной терапии на основании стандартов медицинской помощи может приводить к неэффективности в 40% случаев за счет полипрагмазии и индивидуальных особенностей больных. В том числе и применение стандартов лечения, рекомендованных Британским формуляром, являющимся наиболее востребованным в мире, неэффективно у 20% пациентов. В этой связи возможность использования наряду со стандартами лечения клинических рекомендаций, требует создания и внедрения в практику здравоохранения экспертных систем поддержки ►►

принятия решения на основе доказательной и персонализированной медицины.

Второй фактор связан со значительным количеством побочных реакций и смертей, вызванных полипрагмазией. Потенциально опасные сочетания лекарственных средств являются серьезной клинической, социальной и экономической проблемой системы здравоохранения и государства в целом. На текущий момент 17-23% назначаемых врачами сочетаний лекарственных средств – потенциально опасные. Хотя только у 6-8% пациентов, получающих такие комбинации лекарственных средств, развиваются нежелательная реакция, по экспертным оценкам в мире от нежелательных реакций ежегодно умирают до полумиллиона больных. Причина смерти трети из них – взаимодействия лекарственных средств, связанные с применением потенциально опасных сочетаний. Кроме того, расходы на лечение нежелательных реакций, возникающих при применении потенциально опасных комбинаций, составляют половину от затрат на лечение всех лекарственных осложнений. Существующие отечественные и зарубежные системы оценки позволяют предсказать взаимодействие только 2-х лекарственных препаратов. В практической медицине, особенно в гериатрии, число препаратов, применяемых одновременно, в среднем, приближается к десяти. При этом опасность развития побочных эффектов при нежелательном взаимодействии препаратов увеличивается пропорционально числу применяемых одновременно препаратов.

Вышеперечисленные два фактора определяют третий фактор – неоптимальное обращение лекарственных средств в медицинских организациях. Он характеризуется тем, что закупаются препараты с низкой эффективностью, без учета их взаимодействия при одновременном применении, что с одной стороны, ведет к увеличению средств, затраченных на закупку лекарственных средств, увеличивает количество не востребуемых препаратов на складе, срок действия которых истекает до их возможного применения, а с другой стороны ухудшает качество лечения пациентов и ведет к необоснованным потерям и уменьшению продолжительности жизни.

Эти три фактора ведут к увеличению расходов на лекарственное обеспечение и ухудшают восполнение человеческого капитала. Разрабатываемые в рамках проекта научно-технологические решения позволят уменьшить или даже исключить влияние этих факторов – создать технологическое решение оценки медицинских тех-

нологий на основе доказательной медицины, сформировать подходы к научно-обоснованному безопасному назначению лекарственных средств, особенно в условиях полипрагмазии и обеспечить рациональное обращение лекарственных средств в медицинских организациях.

Таким образом, повышение эффективности и безопасности фармакотерапии может быть достигнуто наряду с административными нормативными и другими мероприятиями внедрением предлагаемых научно-технологических решений на основе подходов доказательной и персонализированной медицины с применением методов искусственного интеллекта, что является актуальной задачей. Разработка решений для оценки медицинских технологий, экспертной системы, базирующейся на последних достижениях персонализированной медицины и предсказывающей опасные взаимодействия 3-х и более лекарств, автоматизация рационального обращения лекарственных средств, является важной социальной, клинической и экономической задачей здравоохранения.

■ **НОВИЗНА ИССЛЕДОВАНИЯ**

В настоящее время интеллектуальный анализ биомедицинских текстов и документов является важнейшим направлением исследований в персонализированной и доказательной медицине.

Эффективные и робастные методы поиска прецедентов и доказательств по узкоспециализированным направлениям медицинской практики способны привести к формированию нового научного направления, принципиальным образом изменить качество оказания медицинской помощи, изменить существующую научно-методическую практику построения клинических систем поддержки принятия врачебных решений.

По состоянию на 2019 год нам неизвестны решения, работающие на многоязычном контенте.

Создание национальной системы интеллектуального анализа научных публикаций доказательной медицины, с целью принятия информированных врачебных решений и мониторинга приоритетных направлений развития превентивной и персонализированной медицины, которая будет по ряду показателей, в том числе по использованию последних данных по доказательной медицине, по своему функционалу, по покрытию разных врачебных специальностей

иметь существенные преимущества перед иностранными системами UpToDate, DynaMed Plus. Так же существенным преимуществом отечественной системы будет являться и тот факт, что она будет полностью адаптирована к российским условиям и будет отвечать всем потребностям отечественной клинической практики.

С этой целью в систему будут включены отечественные клинические рекомендации, законные и подзаконные акты МЗ РФ, которые по необходимости будут доступны пользователям.

На первом этапе планируется создание функциональной системы анализа научных публикаций для разработки и написания оригинального контента. Контент создаваемой системы будет создан с учетом следующих особенностей:

- Анализировать генетические особенности и результаты биомаркеров с целью выявления предрасположенностей к развитию заболеваний.
- Применять персонализированные методы лечения заболеваний и коррекцию состояний, в том числе персонализированное применение лекарственных препаратов и БКП, включая таргетное (мишень-специфическое), основанное на анализе генетических особенностей и иных биомаркеров;
- Поиск в информационных базах данных контента по пациентам входящим в субпопуляции дифференцированного диагноза и применения показавших свою эффективность схем лечения для них, что позволит существенно сократить время на разработку схем лечения.
- Использовать биомаркеры для мониторинга эффективности лечения.
- Осуществлять машинный мониторинг новейшей информации по медицинским базам данных по поисковым запросам пользователей системы для проведения дальнейших клинических исследований.
- Проводить фармакокинетические исследования для предсказания реакции пациента на направленную терапию и уменьшение побочных эффектов, особенно в ситуациях онкологических и социально значимых инвалидизирующих заболеваний.
- Использовать встроенные системы комплексной обработки медицинских изображений для диагностического поиска по базам данных, содержащих результаты исследований, проведенных при помощи визуальных методов диагностики.

Разработка специальных компьютерных программ, способных оперировать многочисленными электронными данными, учитывать сложные взаимосвязи между различными факторами и подска-

зывать врачу оптимальные индивидуальные решения.

■ ЗАДАЧИ ИССЛЕДОВАНИЯ

В рамках заявленной проблематики исследований предполагается решение следующей группы задач:

- Разработка методики и формирование мультиязычного реестра источников информации для прогнозирования трендов и перспектив развития персонализированной медицины;
- Разработка технологии парсинга информационных источников и формирования датасетов анализируемых документов;
- Разработка методов предобработки коллекций документов, нормализации данных с использованием биомедицинских тезаурусов и онтологий с учетом специфики персонализированной медицины;
- Разработка методов динамического мультиязычного тематического моделирования коллекций документов персонализированной медицины;
- Проведение машинного обучения (machine learning) разработанной мультиязычной динамической тематической модели на коллекции документов персонализированной медицины и оценка качества машинного обучения на известных метриках;
- Разработка методов визуального представления результатов тематического моделирования с учетом специфики персонализированной медицины.
- разработка соответствующего программного обеспечения и проведение его апробации в Университетских клинических больницах Сеченовского университета (Национальная система поддержки клинических решений, основанная на принципах доказательной медицины Sechenov-DataMed – www.datamed.pro)
- Популяризация результатов исследования, доведение их до широкого круга потенциальных пользователей, включая организаторов здравоохранения, руководителей медицинских организаций и практикующих врачей.

■ ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На первом этапе с точки зрения программных решений будет разработан комплекс методов, их модельная алгоритмическая реализация, проведены машинные/вычислительные эксперименты, проведена оценка качества, подтверждающая ►►

их эффективность и работоспособность для решения поставленных задач (структура разрабатываемой системы представлена на рис. 1).

По истечении первого этапа исследований нами будет доказана эффективность применения разрабатываемых методов при обработке биомедицинского контента с учетом всей лингвистической специфики.

В основе методики определения источников будут использоваться многокритериальные методы библиографического поиска с учитывающие значимость и ранжирование источников. При этом предполагается использовать не только научные индексируемые источники публикаций, но и так называемую «серую», но высоко-достоверную область знаний, например патентные базы данных по соответствующим разделам:

- Разработка методики и формирование мультязычного реестра источников информации для прогнозирования трендов и перспектив развития персонализированной медицины.

- Разработка технологии парсинга информационных источников и формирования датасетов анализируемых документов.

- Разработка методов предобработки коллекций документов, нормализации данных с использованием биомедицинских тезаурусов и онтологий с учетом специфики персонализированной медицины;

- Разработка методов динамического мультязычного тематического моделирования коллекций документов персонализированной медицины;

- Проведение машинного обучения (machine learning) разработанной мультязычной динамиче-

ской тематической модели на коллекции документов персонализированной медицины и оценка качества машинного обучения на известных метриках;

- Разработка методов визуального представления результатов тематического моделирования с учетом специфики персонализированной медицины.

В результате программной реализации указанных методов планируется получить следующие результаты по группам функций:

Организация поиска:

- управление поисковыми сессиями;
- работа с историей поисковых запросов;
- работа с персональной картотекой документов и с специализированной рабочей тетрадью запросов пользователя.

Составление поискового запроса:

- выбор поисковых массивов и поисковых полей;

- редактирование запроса в двух режимах: экспертный (индивидуально составленный) и запрос по образцу;

- вставка значений из поискового индекса;
- использование мастера создания поисковых выражений;

- уточнение расширений терминов запроса;
- перевод терминов запроса на иностранные языки.

Выполнение запроса:

- контроль орфографических и синтаксических ошибок;

- контроль целостности запроса на основе классификационных биомедицинских индексов документов;



Рис. 1 Структура справочной системы

- возможность поиска документов в персональной картотеке;
- многоязычный поиск на шести языках: русский, английский, испанский, немецкий, французский, итальянский;
- поиск лексически и семантически похожих документов;
- поиск документов с использованием запроса на естественном языке (ЕЯ).

Работа с результатами поиска в режиме реального времени:

- анализ результатов поиска с использованием частотных распределений значений полей в найденных документах;
- исследование тематической структуры поисковой выдачи с использованием кластерного анализа результатов поиска. Тематическое объединение документов на основе схожести текста документов и других значимых полей (реферат, формула и т.д.);
- построение семантической карты связей ключевых тем поисковой выдачи документов для семантической навигации по связям тем с целью поиска наиболее релевантных заявке документа, а также для погружения в документы, содержащие эти темы и связи. Использование биомедицинских тезаурусов позволит содержательно анализировать связи и темы, обобщенные до доминант синсетов;
- работа с итоговой подборкой документов;
- поиск документов, похожих по содержанию на выбранный документ;
- группировка, экспорт и печать результатов поиска;
- управление настройками отображения результатов поиска.

Просмотр документов:

- навигация по документам;
- навигация по подсвеченным терминам, релевантным запросу;

- работа с графическими объектами документов;
- перевод документа;
- поиск на основе значения библиографического поля;
- просмотр кодов генетических последовательностей.

Планируемые функциональные результаты сведены в таблицу.

■ ЗАКЛЮЧЕНИЕ

Таким образом, при выполнении проекта будут созданы следующие ресурсы, составляющие базу знаний (семантическую сеть применения лекарственных средств):

- База данных содержащая формализованные нормативные данные, определяющие порядок применения фармакотерапевтических методов, таких как стандарты медицинской помощи, клинические рекомендации;
- База данных, содержащая формализованные инструкции по применению лекарственных средств, включая соответствие лекарственных средств диагнозам, полу и возрасту пациентов, противопоказания к применению, сведения о несовместимости лекарственных средств, сведения о комплексности применения нескольких препаратов;
- База данных, содержащая формализованные данные научных публикаций доказательной медицины со сведениями об оценке лекарственных средств и формализованные алгоритмы расчета показателей их эффективности;
- Программа для ЭВМ, представляющая из себя информационную систему поддержки принятия врачебного решения в фармакотерапии, интегрированную с медицинской информационной системой. ►

Таблица. Функциональные результаты исследования

Функция	Определение	Описание
Поиск на ЕЯ	Поиск документов по запросу на естественном языке	<ul style="list-style-type: none"> • извлечение лексических терминов (ЛТ) из запроса с использованием морфо- и синтаксических анализаторов системы; • вычисление весов лексических терминов (ключевых тем в канонической форме); • выполняется нечеткий поиск в поисковой системе с учетом весов извлеченных ЛТ.
Поиск похожих биомедицинских документов	Поиск документов, терминологические векторы (ТВД) которых близки к ТВД указанного документа	<ul style="list-style-type: none"> • для каждого документа на этапе индексации строится его семантическое представление – семантическая сеть, на основе которой также формируются список ключевых тем и др. аналитическая информация с учетом особенностей документов; • выполняется поиск похожих документов на основе меры сходства.
Кластерный анализ поисковой выдачи	Объединение документов в тематические группы (кластеры)	<ul style="list-style-type: none"> • кластеризация подборки документов осуществляется на основе меры сходства.
Семантическая сеть	Выявление ключевых тем и их взаимосвязей внутри документа, формирование бинарной семантической сети	<ul style="list-style-type: none"> • компоненты системы для каждого документа на этапе индексации извлекают лексические термины и их связи; • лексические термины и связи между ними сохраняются в базе данных.
Семантическая карта	Динамическое объединение семантических сетей документов поисковой выдачи, анализ причинно-следственных связей между темами документов	<ul style="list-style-type: none"> • визуализация в виде семантической карты используется для семантической навигации в подборке документов.

Социальная значимость проекта заключается во внедрении разрабатываемых решений в информационные системы медицинских организаций, позволяющих при обращении лекарственных средств добиться уменьшения трудопотерь пациентов, увеличить продолжительность жизни, уменьшить количество случаев непреднамеренной и преждевременной смерти, уменьшить количество случаев хронизации своевременно выявленных заболеваний. Коммерциализация проекта будет достигнута:

- за счет продажи информационно-аналитической системы поддержки принятия решений в медицинские организации и аптечные учреждения (1000-1500 медицинских организаций, 6000 аптечных учреждений);
- за счет продажи мобильных приложений ин-

формационно-аналитическая системы поддержки принятия решений отдельным врачам, гражданам Российской Федерации, следящих за своим здоровьем (30000-50000 приложений),

- за счет абонентского доступа к базе знаний из других приложений для разработчиков медицинских информационных систем (30000-50000 лицензий). ▀

Разрабатывается при поддержке гранта Программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014-2020 годы», соглашение RFMEFI60819X0278..

Авторы заявляют об отсутствии конфликта интересов.

РЕЗЮМЕ

В настоящей статье представлено новое исследование, связанное с созданием системы поддержки принятия врачебных решений с интеллектуальным анализом научных данных (текстов стандартов медицинской помощи, клинических рекомендаций, инструкций по применению лекарственных средств, научных публикаций доказательной медицины). Такая система предназначена для обеспечения возможности принятия врачебных решений в фармакотерапии, учитывающих персонализированные медицинские данные за счет оптимального назначения лекарственных средств и применения медицинских технологий, снижения частоты возникновения нежелательных реакций при одновременном применении двух и более лекарственных средств по разным показаниям. Технической целью выполнения исследования является создание интеллектуальной автоматизированной информационной системы поддержки принятия врачебных решений и внедрение ее в клиническую практику.

Ключевые слова: системы поддержки принятия врачебных решений, доказательная медицина, интеллектуальные системы, интеллектуальный анализ текстов.

ЛИТЕРАТУРА

1. Указ Президента Российской Федерации от 1 декабря 2016 г. № 642 "О Стратегии научно-технологического развития Российской Федерации".
2. Постановление Правительства Российской Федерации от 26 декабря 2017 г. № 1640 «Об утверждении государственной программы Российской Федерации «Развитие здравоохранения».
3. Jurafsky D, Martin JH. *Speech and Language Processing*, 2nd edn. Upper Saddle River: Pearson, 2009.
4. Кокрейновская библиотека <https://www.cochrane.org>.
5. Федеральная электронная медицинская библиотека, Министерства здравоохранения Российской Федерации <http://www.femb.ru/>.
6. Text Mining for Precision Medicine: Bringing Structure to EHRs and Biomedical Literature to Understand Genes and Health.
7. Text Mining to Support Gene Ontology Curation and Vice Versa. Ruch P. *Methods Mol Biol*. 2017;1446:69–84.
8. Jurafsky D, Martin JH. *Speech and Language Processing*, 2nd edn. Upper Saddle River: Pearson, 2009.
9. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2010;38:D5 D16.
10. Постановление Правительства РФ от 29.09.2017 N 1184 (ред. от 10.09.2018) "О порядке разработки и реализации планов мероприятий ("дорожных карт") по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации Национальной технологической инициативы и внесении изменений в некоторые акты Правительства Российской Федерации" (вместе с "Положением о разработке и реализации планов мероприятий ("дорожных карт") по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации Национальной технологической инициативы") http://www.consultant.ru/document/Cons_doc_LAW_279244/
11. Rebholz-Schuhmann D, Oellrich A, Hoehndorf R. Text-mining solutions for biomedical research: enabling integrative biology. *Nat Rev Genet* 2012; 13:829–39.
12. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. *An Introduction to Information Retrieval*. Cambridge University Press New York, NY, USA ©2008. 496 p.
13. Fayyad U., Piatetsky-Shapiro G., Smyth P. From Data Mining to Knowledge Discovery: an Overview // *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996. pp. 1–34.
14. Ingo Feinerer, Kurt Hornik, David Meyer. Text Mining Infrastructure in R. *Journal of Statistical Software*. V. 25, Issue 5, March 2008. 54 p.
15. Jiawei Han, Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann; 2 edition, January 13, 2006. 800 p.
16. <https://www.cs.cmu.edu/~mccallum/bow/>.
17. DiMeX: A Text Mining System for Mutation–Disease Association Extraction April 2016 *PLoS ONE* 11(4).
18. Joint Event Extraction via Recurrent Neural Networks Thien Huu Nguyen, Kyunghyun Cho and Ralph Grishman Computer Science Department, New York University, New York, NY 10003, USA.
19. Павлов Л. П. Серая литература как источник научной и технической информации Федеральное государственное автономное научное учреждение «Центр информационных технологий и систем органов исполнительной власти», Москва, Россия.
20. Бурнер К, Чен С, Бойакс КВ. Визуализация областей знаний. *Annu Rev Inform Sci*. 2003; 37 (1): 179–255. doi: 10.1002 / aris.1440370106.
21. Апора С.К., Портер А.Л., Юсти Дж., Шапира П. Захват новых разработок в появляющейся технологии: обновленная стратегия поиска для определения результатов исследований в области нанотехнологий. *Наукометрия*. 2013; 95(1): 351–370. doi: 10.1007 / s11192–012–0903–6.
22. Exploring the Relationship between the Engineering and Physical Sciences and the Health and Life Sciences by Advanced Bibliometric Methods Ludo Waltman1 *, Anthony F. J. van Raan1, Sue Smart2 *PLoS One*. 2014; 9(10): e111530. Published online 2014 Oct 31. doi: 10.1371/journal.pone.0111530.
23. Chan I.S., Ginsburg G.S. Personalized medicine: progress and promise. *Annu Rev Genomics Hum Genet* 2011; 12:217–44.
24. Johnson A.D., O'Donnell C.J. An open access database of genomewide association results. *BMC Med Genet* 2009; 10:6.