

Научная статья

УДК 004.891

DOI 10.25205/1818-7900-2025-23-3-5-22

Подход к разработке интеллектуальных персонализированных медицинских информационных систем

Галина Борисовна Загорулько^{1,2}
Владислав Алексеевич Устюгов¹

¹Новосибирский государственный университет
Новосибирск, Россия

²Институт систем информатики им. А. П. Ершова СО РАН
Новосибирск, Россия

gal@iis.nsk.su
v.ustyugov@g.nsu.ru

Аннотация

Статья посвящена проблемам создания интеллектуальных информационных систем для решения различных задач медицины с учетом персонализированного подхода. Рассмотрено текущее состояние персонализированной медицины, сформулированы основные требования к персонализированной медицинской системе, предложена архитектура типовой системы. Подчеркнута актуальность создания таких систем, научная новизна предлагаемого подхода и его практическая ценность. Описаны программные компоненты системы, удовлетворяющей перечисленным требованиям. Обоснован выбор методов и средств, обеспечивающих их разработку и функционирование. Обладающая описанными структурными компонентами и функциональными возможностями система будет способствовать более эффективному взаимодействию врачей и пациентов, повышению качества медицинского обслуживания и улучшению результатов лечения.

Ключевые слова

персонализированная медицина, искусственный интеллект, машинное обучение, обработка медицинской информации, обработка медицинских текстов на естественном языке, поддержка принятия врачебных решений

Для цитирования

Загорулько Г. Б., Устюгов В. А. Подход к разработке интеллектуальных персонализированных медицинских информационных систем // Вестник НГУ. Серия: Информационные технологии. 2025. Т. 23, № 3. С. 5–22.
DOI 10.25205/1818-7900-2025-23-3-5-22

An Approach to Developing Intelligent Personalized Medical Information Systems

Galina B. Zagorulko^{1,2}, Vladislav A. Ustyugov¹

¹Novosibirsk State University,
Novosibirsk, Russian Federation

²A. P. Ershov Institute of Informatics Systems of SB RAS,
Novosibirsk, Russian Federation

gal@iis.nsk.su
v.ustyugov@g.nsu.ru

Abstract

The paper is devoted to the problems of creating intelligent information systems for solving various medical problems taking into account the personalized approach. The current state of personalized medicine is considered, the main requirements for a personalized medical system are formulated, and the architecture of a typical system is proposed. The actuality of creating such systems, the scientific novelty of the proposed approach and its practical value are emphasized. The software components of the system that meet the listed requirements are described. The choice of methods and tools to ensure their development and operation is substantiated. The system with the described structural components and functional capabilities will contribute to more effective interaction between doctors and patients, raising the medical care quality and improving the treatment results.

Keywords

personalized medicine; artificial intelligence; machine learning; medical information processing; natural language processing of medical texts; support for medical decision-making

For citation

Zagorulko G. B., Ustyugov V. A. An approach to developing intelligent personalized medical information systems. *Vestnik NSU. Series: Information Technologies*, 2025, vol. 23, no. 2, pp. 5–22 (in Russ.) DOI 10.25205/1818-7900-2025-23-3-5-22

Введение

Персонализированная медицина представляет собой совокупность научных знаний, методов и практических действий по оказанию медицинской помощи пациенту, учитывающих его индивидуальные особенности. К таким особенностям относятся не только традиционные факторы, рассматриваемые в медицине на протяжении длительного времени, такие как возраст, пол, телосложение, индекс массы тела, наличие сопутствующих заболеваний, биохимические показатели и семейный анамнез, позволяющий оценить предрасположенность к наследственным патологиям, но и, прежде всего, молекулярные характеристики организма. Важнейшую роль в этом аспекте также играют данные, полученные в ходе геномных и постгеномных исследований, реализуемых с использованием передовых методов молекулярной биологии и медицины.

Существенный прорыв в развитии персонализированной медицины стал возможен благодаря расшифровке генома человека, представляющего собой своего рода генетическую карту организма. В результате ученые смогли идентифицировать около 25 тысяч человеческих генов, декодировать последовательности нуклеотидов в структуре ДНК, что стало основой для построения индивидуальных «генетических паспортов» [1; 2].

Для того чтобы результаты научных исследований стали достоянием практикующих врачей и могли успешно использоваться для лечения пациентов, необходима разработка инструментов, обеспечивающих информационно-аналитическую поддержку процессов здравоохранения.

Со временем у каждого пациента накапливается значительный объем данных, связанных с его здоровьем, которые требуют надежного хранения, структуризации и обработки. С увеличением объема этой информации также существенно растет сложность ее анализа врачами,

что сильно сказывается на точности диагностики и эффективности лечения. В связи с этим ключевую роль в современной системе здравоохранения, и особенно в персонализированной медицине, начинают играть информационные технологии, обеспечивающие эффективное управление медицинскими данными.

Современные цифровые технологии позволяют проводить онлайн-консультации как между врачом и пациентом, так и между самими специалистами. Пациенты все чаще используют различные гаджеты и мобильные приложения для мониторинга своего здоровья в режиме реального времени, что дает им возможность быстро реагировать на любые изменения в состоянии. Это способствует более глубокому пониманию того, каким образом следует разрабатывать стратегии лечения для каждого человека, создавая условия для перехода от универсальных медицинских подходов к персонализированному оказанию помощи с учетом индивидуальных потребностей пациентов.

Большие объемы информации, которые необходимо учитывать врачу при постановке диагноза и выбора стратегии лечения с одной стороны, а также наличие большого объема данных, накопленных в результате цифровой трансформации процессов здравоохранения, с другой, диктуют необходимость и создают предпосылки использования результатов последних исследований в области искусственного интеллекта.

Таким образом, представляется актуальной задача построения персонализированных медицинских информационных систем, которые бы позволили сочетать в себе достижения как в области информационных технологий и искусственного интеллекта, так и передовой опыт в области персонализированной медицины.

В статье описывается подход к разработке интеллектуальных персонализированных медицинских информационных систем (ИПМИС), требования к их функциональным возможностям, программные компоненты таких систем, способы и средства их программной реализации. Обосновываются принципы, на которых должны строиться системы такого класса.

1. Существующие решения

Исследования показали, что полноценных интеллектуальных решений в области персонализированной медицины, как на зарубежном, так и на российском рынке практически нет. На российском рынке существуют лишь три платформы, демонстрирующие движение в сторону индивидуального подхода к оказанию медицинской помощи:

1. *REMSMED* [3]. Телемедицинская платформа, решающая комплекс задач по лечению пациентов и позволяющая проводить предварительный скрининг здоровья и прогнозировать риск развития заболеваний. REMSMED актуальна для различных специализаций и позволяет составлять персонализированные планы лечения с отслеживанием динамики состояния пациента и оперативно корректировать лечение, проводить видеоконсультации. Пациенты получают доступность квалифицированной медицинской помощи и возможность самостоятельного мониторинга здоровья.

2. *WebioMed* [4]. Платформа прогнозной аналитики и управления рисками в здравоохранении на основе машинного обучения. Предназначена для автоматического анализа обезличенных медицинских данных с целью прогнозирования возможного развития заболеваний и их осложнений на персональном и популяционном уровне. Включает несколько подсистем:

- система поддержки принятия врачебных решений (СППВР);
- система поддержки принятия управленческих решений;
- система извлечения данных из неструктурированных медицинских записей;
- множество моделей машинного обучения, к которым можно получать доступ по API.

3. *MediCaSe* [5]. Система не является полноценной платформой, удовлетворяющей аспектам персонализированной медицины. Однако используется как вспомогательный инструмент

диагностики в некоторых клиниках. Система проводит опрос по алгоритму «дерева решений». В основном это вопросы «закрытого типа» – на них можно ответить либо «да», либо «нет». Заподозрив какую-либо патологию, искусственный интеллект, как и врач, подробнее расспрашивает об ее симптомах, активировав дополнительные ветви опроса. Таким образом, симптом-чекер MeDiCase позволяет быстро провести первичную оценку симптомов и сориентироваться в дальнейших действиях.

В табл. 1 представлены преимущества и недостатки каждой платформы в контексте персонализированного подхода и применения инструментов искусственного интеллекта.

Таблица 1

Сравнительный анализ существующих медицинских интеллектуальных платформ

Table 1

Comparative analysis of existing medical intelligent platforms

Платформа	Преимущества	Недостатки
REMSMED	Платформа демонстрирует явное движение в сторону персонализации медицинского обслуживания	Практически не используются инструменты искусственного интеллекта. Все процессы (анализ, предоставление рекомендаций и т. д.) осуществляются врачами
WebioMed	Платформа обладает развитыми инструментами искусственного интеллекта и использует множество моделей машинного обучения для анализа медицинских рисков	Практически нет персонализации медицинского обслуживания, так как все рекомендации являются общими – платформа используется, скорее, для общего анализа и предоставления статистики
MediCaSe	Обладает очень мощными моделями диагностики, позволяющими в удобном формате опросника получить качественные рекомендации	Является лишь вспомогательным инструментом для первичной диагностики и не является полноценным инструментом создания интеллектуальных информационных систем

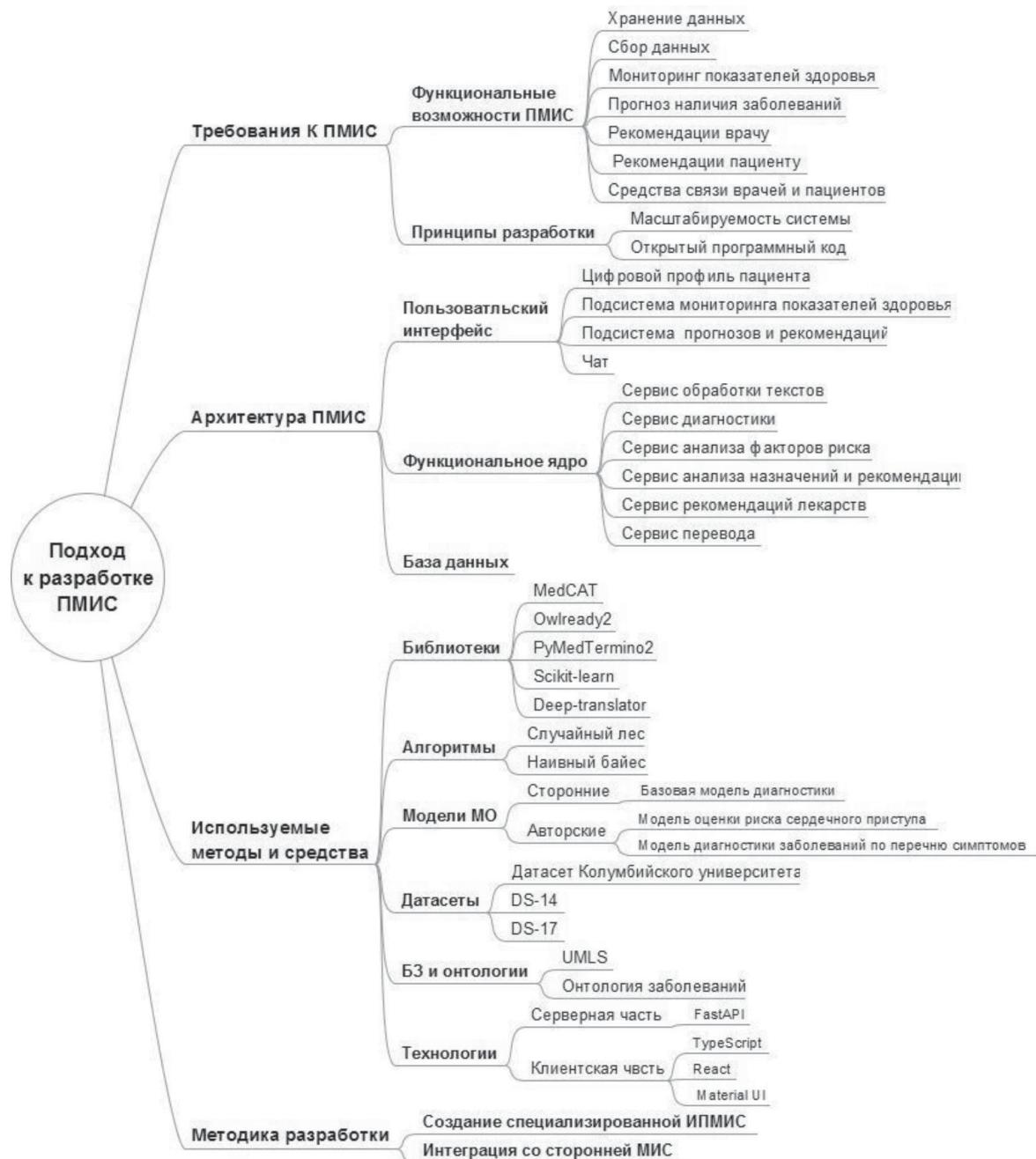
2. Подход к разработке ИПМИС

На рис. 1 представлены основные положения предлагаемого подхода к разработке ИПМИС. Перечислены требования к персонализированным системам, их структурные компоненты. Определена архитектура типовой ИПМИС.

Большое внимание уделено методам и средствам, используемым как для разработки ИПМИС, так и для обеспечения их дальнейшего функционирования. Таким образом, предлагается типовая персонализированная медицинская система с базовым набором функциональных возможностей. Также предлагается методика, позволяющая специализировать типовую систему на определенную область медицины или на решение определенного класса задач. Кроме того, разработанные компоненты могут быть использованы для интеграции с существующими медицинскими информационными системами (МИС).

Рассмотрим, каким **требованиям** должна удовлетворять ИПМИС. Существующие исследования, посвященные разработке персонализированных медицинских систем, формулируют ряд требований к ним [6]. Обобщая информацию, можно сделать вывод, что система такого класса должна обладать следующими фундаментальными **функциональными возможностями**:

- 1) хранение данных о пациентах в едином цифровом профиле;



Rис 1. Подход к разработке ИПМИС
Fig. 1. The approach to IPMIS development

- 2) сбор данных для формирования цифрового профиля как ручными, так и автоматическими методами;
- 3) отслеживание динамики. Удобный мониторинг показателей здоровья для оценки динамики и отклонений от норм;
- 4) предоставление индивидуальных прогнозов о наличии заболеваний;
- 5) предоставление индивидуальных рекомендаций врачу и пациенту;
- 6) предоставление средств удобной связи для врачей и пациентов.

Для ИПМИС крайне важно иметь возможность оперативно пополнять свою базу знаний, быстро настраиваться на решение новых задач, использовать передовые методы искусственно-

го интеллекта по мере их появления. Поэтому такие системы должны строиться на **принципах** открытого кода и масштабируемости, который подразумевает использование сервис-ориентированного подхода.

3. Архитектура ИПМИС

Типовая ИПМИС имеет клиент-серверную архитектуру, где клиентская часть является пользовательским интерфейсом, а серверная часть – функциональным ядром системы.

3.1. Пользовательский интерфейс

Пользовательский интерфейс ИПМИС реализован как веб-приложение. Он предоставляет пользователям удобный доступ ко всем функциональным возможностям системы и позволяет взаимодействовать со всем API серверной части, составляя вместе с ней полноценную медицинскую систему с базовой функциональностью. Разработанный интерфейс также можно использовать в качестве микросервисной клиентской части в случае интеграции с существующими МИС. В таком сценарии данный интерфейс вместе со всеми страницами и панелями может открываться при выборе соответствующего пациента в МИС, а все отображаемые данные и прогнозы будут относиться к выбранному конкретному пациенту.

3.1.1. Цифровой профиль

Цифровой профиль пациента представляет собой централизованное хранилище медицинской информации, которое обеспечивает удобный доступ к данным и возможность их управления. Этот инструмент играет важную роль, предоставляя врачам, пациентам и другим участникам здравоохранения эффективный способ обмена, анализа и хранения данных. Формирование цифрового профиля осуществляется путем интеграции данных из различных электронных медицинских карт (ЭМК) и записей о медицинских обращениях пациента.

Информация, содержащаяся в цифровом профиле, охватывает как структурированные персональные данные о здоровье пациента, так и медицинскую документацию. К основным данным о здоровье относятся:

- 1) общие сведения;
- 2) результаты лабораторных исследований;
- 3) вакцинации и профилактические мероприятия;
- 4) аллергии и непереносимости;
- 5) семейный анамнез;
- 6) информация об образе жизни.

Кроме того, цифровой профиль включает различные медицинские документы:

- 1) историю болезней с анамнезами и окончательными диагнозами;
- 2) результаты лабораторных и диагностических исследований;
- 3) назначения и рекомендации врачей;
- 4) прочую медицинскую документацию.

Главная задача цифрового профиля заключается в консолидации всех ключевых данных о здоровье пациента в одном месте. Это упрощает доступ к информации, повышает точность диагностики, улучшает качество лечения и стимулирует пациентов активнее участвовать в процессе управления своим здоровьем.

На странице цифрового профиля данные о здоровье, разделенные на категории, содержатся в специальных выпадающих списках. Напротив каждого показателя есть дата последнего

изменения. С помощью кнопок «Редактировать», «Удалить» и «Добавить» врач может вручную ввести нужную информацию и обновить данные. На рис. 2 показаны структурированные данные цифрового профиля пациента.

Общая информация

Возраст	30	Пол	Мужской	Дата рождения	27.07.1994	Рост, см	170
Вес, кг	60	ИМТ, кг/м ²	24.2	Температура, °C	36	Систолическое АД, мм рт.ст.	160
Диастолическое АД, мм рт.ст.	100	ЧСС, уд/мин	90	Частота дыхания, число/мин	30		

Лабораторные исследования

Вакцинации и профилактические мероприятия

Аллергии и непереносимости

Семейный анамнез

Образ жизни

Добавить

Рис. 2. Данные о здоровье на странице цифрового профиля

Fig. 2. Health data on the digital profile page

Секция с документами содержит интерактивные таблицы, с помощью которых можно искать, фильтровать и сортировать документы, а также открыть соответствующий документ. При нажатии на кнопку «Анализ текста» выделяются распознанные медицинские термины (рис. 3). При нажатии на любой из них отображается информация из базы знаний UMLS [7]. Эта возможность обеспечивается сервисом анализа текста (см. п. 3.2.1).

Артериальная гипертензия

Код МКБ: I10
Дата постановки: 2024-11-20
Врач: Иванов Петр Сергеевич
Специальность: Терапевт
Нозология: Сердечно-сосудистые заболевания
Тип заболевания: Хроническое
Учреждение: Городская поликлиника №1

Анамнез

Пациент обратился с жалобами на периодические головные боли, учащенное сердцебиение, повышенное давление мм рт. ст., ощущение усталости.
История болезни: гипертония у пациента диагностирована впервые 5 лет назад.
Семейный анамнез отягощен: у отца и матери артериальная гипертензия.
Обострение состояния отмечается на фоне стрессов и нерегулярного образа жизни.

Клинические данные

При осмотре: артериальное давление - 150/95 мм рт. ст., частота сердечных сокращений - 85 уд./мин.
Данные ЭКГ: признаки гипертрофии левого желудочка.
Биохимический анализ крови: повышение уровня холестерина до 6,5 ммоль/л.
Анализы мочи: без особенностей.

Артериальная гипертензия

Термин: "артериальная гипертензия"
Каноническое название: Arterial hypertension
CUI: C0020538
Тип: Заболевание
Код МКБ-10: I10

Рис. 3. Распознавание терминов с помощью сервиса обработки текстов

Fig. 3. Recognition of terms using a text processing service

3.1.2. Подсистема мониторинга показателей здоровья для оценки динамики и отклонений от норм

Основная цель данного компонента – отобразить конкретные биометрические показатели пациента вместе с их допустимыми диапазонами, датами измерений и динамикой изменений. Кроме того, врач должен иметь возможность редактировать те или иные показатели и добавлять новые.

Подсистема показателей здоровья также играет важную роль в персонализации медицинского обслуживания. Возможность устанавливать индивидуальные целевые уровни позволяет учитывать различия между пациентами, включая их физиологические параметры, генетические особенности и образ жизни, а также оперативно выявлять отклонения от нормы, оценивать эффективность проводимого лечения и принимать обоснованные решения на основе объективных данных. Например, резкий скачок артериального давления или отклонения уровня холестерина от целевого значения сигнализируют о необходимости скорректировать терапевтический план. Это особенно важно для управления хроническими заболеваниями, которые требуют постоянного контроля и своевременной корректировки.

На странице показателей здоровья все показатели синхронизируются с показателями лабораторных измерений и некоторыми общими показателями, содержащимися на странице цифрового профиля (рис. 4).

Холестерин		Индекс массы тела		Артериальное давление	
Нет данных		24.2 кг/м³		160/100 мм рт.ст.	
Дата обновления Нет данных	Целевой уровень 3.6-5.2	Дата обновления 10-20-2024	Целевой уровень 18.5-24.9	Дата обновления 10-20-2024	Целевой уровень 120-80
Частота сердечных сокращений		Глюкоза		Калий	
90 уд/мин		7.5 ммоль/л		3.0 ммоль/л	
Дата обновления 10-20-2024	Целевой уровень 60-80	Дата обновления 10-20-2024	Целевой уровень 3.9-5.6	Дата обновления 10-20-2024	Целевой уровень 3.5-5.1
Витамин D		SpO2			
38 нг/мл		98 %			
Дата обновления 10-20-2024	Целевой уровень 20-50	Дата обновления 10-20-2024	Целевой уровень 95-100		
					Добавить новый показатель...

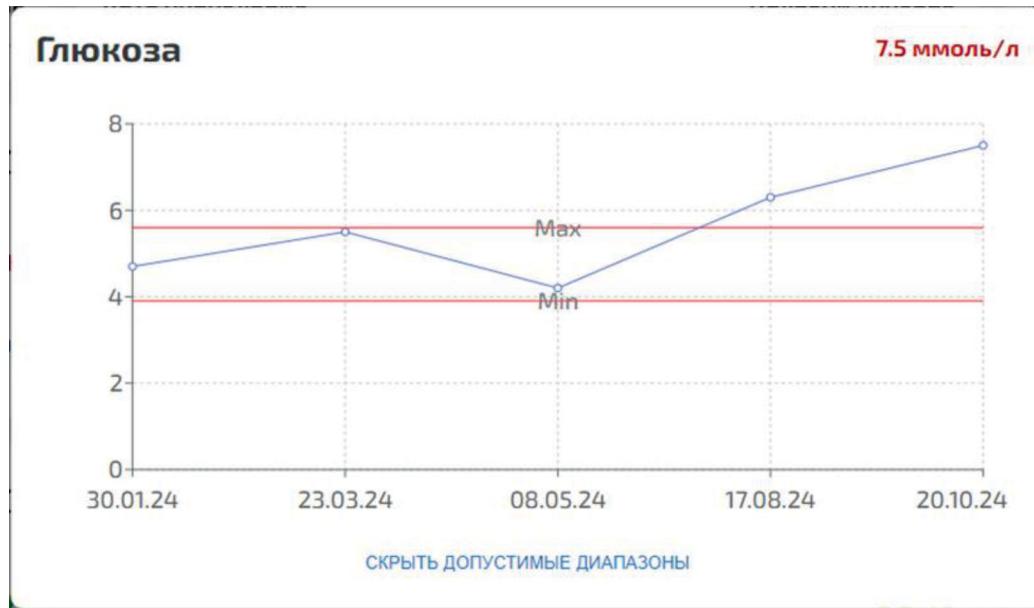
Rис. 4. Страница показателей здоровья
Fig. 4. The health indicators page

При нажатии на кнопку редактирования или добавлении нового показателя откроется специальная форма для внесения данных: название показателя, значение, единица измерения, дата измерения, целевой уровень.

Также для каждого показателя можно открыть окно с графиком, иллюстрирующим динамику его изменения (рис. 5).

3.1.3. Подсистема предоставления индивидуальных прогнозов и рекомендаций врачу и пациенту

Главную роль в персонализированном подходе играет возможность предоставления прогнозов и других персональных рекомендаций на основе текущего цифрового профиля. Для этого может быть использована специальная панель проблем пациента. Основная цель – отобразить текущее положение дел пациента в доступной форме. Данная панель должна быть



Rис. 5. График с динамикой изменения показателя «Глюкоза»
Fig. 5. Graph with the dynamics of changes in the Glucose indicator

доступна как врачу, собирающемуся ставить диагноз или прописывать/корректировать план лечения, так и пациенту, интересующемуся своим состоянием здоровья. Информация должна иметь следующее содержание:

1. **Симптомы.** Содержит список симптомов, введенных врачом вручную или извлеченных из текстов последних анамнезов;
 2. **Заболевания.** Содержит список последних заболеваний, извлеченных из истории болезней, а также хронические заболевания;
 3. **Оценка рисков.** Содержит оценку проблем пациента по различным нозологиям (классификациям заболеваний). Оценка осуществляется врачами с некоторой периодичностью на основе комплексного анализа данных пациента – динамики показателей здоровья и историй болезней. Оценка имеет несколько уровней: риск не определен, низкий риск, средний и высокий;
 4. **Подозрения.** Содержит список возможных заболеваний, прогнозируемых сервисом диагностики;
 5. **Факторы риска.** Содержит возможные риски с учетом индивидуальных особенностей, сформированные сервисом анализа факторов риска;
 6. **Лекарства.** Содержит список прописанных и принимаемых пациентом в данный момент лекарств вместе дозировками;
 7. **Рекомендации.** Содержит изложенные в краткой форме различные рекомендации как от врачей, так и от специальных сервисов (в частности, сервиса рекомендаций лекарств).
- Эффективным способом отображения подобной информации в интерфейсе пользователя являются карточки с соответствующими пунктами. На рис. 6 показана страница панели состояния пациента
- Помимо краткой информации, у каждой карточки также есть специальная кнопка «Подробнее», открывающая окно с описанием источников той или иной информации (рис. 7). Источниками могут быть как сами врачи, так и различные сервисы оценок, анализов и соответствующие модели машинного обучения.

Симптомы	Подробнее	Заболевания	Подробнее	Оценка рисков	Подробнее
Повышенная глюкоза крови Высокие уровни кетонов		Артериальная гипертензия Сердечная недостаточность ОРВИ	I11 I50.0 J00-J02	Сердечно-сосудистые заболевания Заболевания системы пищеварения Заболевания органов дыхания Заболевания эндокринной системы Заболевания почек и мочевыводительной системы Заболевания опорно-двигательного аппарата Заболевания кожи и волосистых покровов	Высокие Средние Низкие Низкие Низкие Низкие Неизвестно Неизвестно
Подозрения	Подробнее	Факторы риска	Подробнее		
Сахарный диабет	E10	Высокий шанс сердечного приступа Целевое АД не достигнуто Целевое ЧСС не достигнуто Артериальная гипертензия Сердечная недостаточность			
Лекарства	Подробнее	Рекомендации	Подробнее		
Лозартан Эналаприл Амлодипин	50 мг./ден. 5 мг./ден. 5 мг./ден.	Принимать Омега 3 Альфа-липоевая кислота Магний Ограничить употребление сахара Регулярные аэробные нагрузки		1000 мг./ден. 600 мг./ден. 400 мг./ден.	

Рис. 6. Страница панели состояния пациента
Fig. 6. Patient Status panel page

Факторы риска	
Высокий шанс сердечного приступа	
Источник: Модель оценки риска сердечного принципа. Точность работы - 95%	
Дата фиксации: 2024-11-15	
Целевое АД не достигнуто	
Источник: Оценка динамики показателей здоровья	
Дата фиксации: 2024-11-18	
Целевое ЧСС не достигнуто	
Источник: Оценка динамики показателей здоровья	
Дата фиксации: 2024-11-20	
Артериальная гипертензия	
Источник: Хроническое заболевание из истории болезней	
Дата фиксации: 2024-11-12	
Сердечная недостаточность	
Источник: Хроническое заболевание из истории болезней	
Дата фиксации: 2024-11-19	

Рис. 7. Пример отображения информации об источниках для факторов риска
Fig. 7. Example of displaying information about sources for risk factors

3.1.4. Средства удобной связи врачей и пациентов

Персонализированный подход невозможен без удобной связи пациента с лечащим врачом. Поэтому был предусмотрен базовый элемент телемедицины – чат, с помощью которого пациент может обратиться к соответствующему лечащему врачу с проблемой или с целью уточнения какого-либо вопроса в режиме онлайн. На странице чата в пользовательском интерфейсе присутствует панель инструментов вместе со специальными формами для ввода, позволяющими врачу:

1) сделать заметки, например, зафиксировав анамнез, состояние пациента и какие-либо базовые измерения (температура тела, давление и т. д.). Соответствующая форма представлена на рис. 8. После фиксации данная заметка автоматически помещается в лист текущих анамнезов на странице цифрового профиля пациента;

2) составить индивидуальный план лечения вместе со списком препаратов и их дозировками.

Таким образом, пациент и врач всегда могут поддерживать удобный дистанционный контакт, а информация о процессах взаимодействия – анамнезы, планы лечения – будет пополнять цифровой профиль полезными данными, необходимыми для работы интеллектуальных сервисов и отображения текущего состояния пациента в цифровом профиле.

3.1.5. Средства реализации пользовательского интерфейса

Для реализации пользовательского интерфейса был выбран современный технологический стек, включающий TypeScript, React и Material UI (MUI). Такой выбор обеспечивает высокую надежность, скорость разработки и соответствие стандартам удобства и доступности пользовательского интерфейса.

1. TypeScript. Язык обеспечивает строгую типизацию, снижает количество ошибок и упрощает работу с данными.

2. React. Фреймворк предоставляет удобные средства для создания веб-интерфейсов с компонентной архитектурой и эффективным управлением состоянием [8].

3. MUI (Material UI). Данная библиотека ускоряет разработку за счет готовых адаптивных компонентов, а также поддерживает современные стандарты доступности веб-интерфейсов [9].

The screenshot shows a user interface for recording anamnesis data. At the top, the title 'Анамнез' is displayed. Below it, there are input fields for 'ФИО пациента' (Ivan Ivanov) and 'Класс заболеваний' (Respiratory system). A large text area labeled 'Общее состояние' is currently empty. Below this, a section for 'Симптомы' lists three items: 'Кашель', 'Боль в груди при дыхании', and 'Хрипсы', each with a delete icon. A button labeled 'Добавить симптом' is also present. Further down, there are fields for 'Температура тела (°C)', 'Систолическое давление (мм рт.ст.)', 'Диастолическое давление (мм рт.ст.)', and 'Пульс (ЧСС, уд/мин)'. At the bottom right is a blue 'СОХРАНИТЬ' (Save) button.

Рис. 8. Форма для фиксации анамнеза
Fig. 8. Anamnesis record form

3.2. Функциональное ядро ИПМИС

Функциональные возможности разработанной системы обеспечиваются набором соответствующих программных сервисов.

1. **Сервис обработки текстов.** Позволяет анализировать все медицинские документы с целью получения информации о терминах из базы знаний UMLS, а также автоматически извлекать некоторые необходимые сущности для дальнейшего использования в моделях машинного обучения, в частности, симптомы из анамнезов.

2. **Сервис диагностики.** Позволяет автоматически формировать прогнозы о наличии возможных заболеваний. Диагностика осуществляется как с помощью обученных на различных медицинских данных моделей машинного обучения, так и с помощью онтологии.

3. **Сервис анализа факторов риска.** Позволяет отображать факторы, на которые стоит обратить внимание при работе с пациентом. К базовым факторам риска относятся:

- различные отклонения показателей здоровья от допустимых норм, прописанных в эталонном цифровом профиле;
- хронические заболевания из списка окончательных диагнозов (истории болезней);
- различные индивидуальные непереносимости и аллергии, в том числе побочные эффекты от лекарств;
- высокие оценки рисков по классам заболеваний, полученные соответствующими моделями, например, моделью оценки риска сердечного приступа.

4. **Сервис анализа назначений и рекомендаций.** Используется для анализа назначений и рекомендаций врачей из записей в соответствующей секции цифрового профиля.

5. **Сервис рекомендаций лекарств.** Данный сервис предоставляет индивидуальные рекомендации лекарств для лечения тех или иных текущих заболеваний на основе предыдущих клинических случаев из цифрового профиля и онтологии заболеваний.

Для реализации перечисленных сервисов были использованы современные методы, библиотеки, модели, описанные ниже.

3.2.1. Инструменты обработки медицинских текстов

При работе с медицинскими данными основная часть наиболее ценной информации содержится в медицинских текстах. Заметки, сделанные врачами, документируют множество деталей, включая имеющиеся заболевания, методы лечения, медикаменты или процедуры, позволяют формировать представление о состоянии пациента. Обработка слабоструктурированных медицинских записей из электронных медицинских карт представляет собой сложную задачу. Для эффективной обработки таких данных необходимы инструменты, способные выполнять извлечение терминов и их связывание с терминологией предметной области.

MedCAT (Medical Concept Annotation Toolkit) [10] – это библиотека Python с открытым исходным кодом, предназначенная для извлечения и связывания терминов (Named Entity Recognition + Linking, NER+L), разработанная с учетом специфики медицинской терминологии. Ее ключевой особенностью является использование самообучающегося алгоритма машинного обучения, который позволяет извлекать концепты, используя любую концептуальную словарную базу, в частности, базу знаний UMLS.

Основные компоненты MedCAT включают:

MedCAT. Ядро библиотеки, реализующее алгоритмы извлечения и связывания сущностей;

MedCATtrainer. Веб-интерфейс для аннотации и обучения моделей, позволяющий специалистам в области здравоохранения адаптировать модели под конкретные задачи;

MedCATservice. REST API для интеграции с другими системами и приложениями.

MedCAT использует векторные представления слов и концептов, обученные на больших корпусах медицинских текстов, таких как MIMIC-III, что позволяет эффективно обрабатывать

вать сложные и многозначные термины. Благодаря механизму самообучения, библиотека способна адаптироваться к новым данным без необходимости ручной аннотации больших объемов текста.

Одним из ключевых преимуществ MedCAT является тесная интеграция с базой знаний UMLS [7; 11]. UMLS (Unified Medical Language System) – это обширная метаонтология, разработанная Национальной медицинской библиотекой США, объединяющая более 200 специализированных медицинских словарей, включая SNOMED-CT, ICD-10, MeSH и др. Она предназначена для стандартизации терминов, используемых в клинической практике, научных исследованиях и медицинских информационных системах. В UMLS используется концептуально-ориентированный подход: каждому понятию присваивается уникальный идентификатор (CUI), который связывает различные синонимы, аббревиатуры и переводы.

Интеграция MedCAT с UMLS обеспечивает стандартизированное представление медицинских концептов и облегчает их использование в различных практических и исследовательских задачах. MedCAT поддерживает связывание извлеченных сущностей с уникальными идентификаторами концептов из данных онтологий вместе с сопутствующей информацией о термине: каноническое название, определение, семантический тип, классификация по МКБ-10.

В ИПМИС MedCAT используется для извлечения из текстов анамнезов информации, полезной для медицинской диагностики:

1. Получение кодов семантических типов извлеченных терминов (ТУ) позволяет четко их классифицировать: выделить симптомы и заболевания;
2. Получение кодов концептов позволяет получить больше информации о той или иной сущности путем соответствующего запроса к базе UMLS. Например, получить подробную информацию об извлеченном концепте: определение, код, каноническое название и т. д. (рис. 9).

<p>Description: Intracerebral hemorrhage (very acute clinical changes occurred immediately). CC: Left hand numbness, on presentation; then developed lethargy later that day.</p> <p>HX: On the day of presentation, this 72 y/o RHM suddenly developed generalized weakness and lightheadedness, and could not rise from a chair. Four hours later he experienced sudden left hand numbness lasting two hours. There were no other associated symptoms except for the generalized weakness and lightheadedness. He denied vertigo.</p> <p>He had been experiencing falling, spells without associated LOC up to several times a month for the past year.</p> <p>MEDS : procardia SR, Lasix, Ectoin, KCL, Digoxin, Colace, Coumadin.</p> <p>PMH : 1/8/92 evaluation for presyncope (Echocardiogram showed: AV fibrosis / calcification , AV stenosis /insufficiency, MV stenosis with annular calcification and regurgitation , moderate TR, Decreased LV systolic function, severe LAE , MRI brain : focal areas of increased T2 signal in the left cerebellum and in the brainstem [probably representing microvascular ischemic disease] , IVG (MUGA) scan revealed: global hypokinesis of the LV and biventricular dysfunction , RV ejection Fx 45% and LV ejection Fx 39%. He was subsequently placed on coumadin severe valvular heart disease), 2) HTN , 3) Rheumatic fever and heart disease , 4) COPD , 5) ETOH abuse , 6) colon polyps , 7) CAD , 8) CHF , 9)Appendectomy, 10)junctional tachycardia .</p>	<table border="1"> <tbody> <tr> <td>Pretty Name</td> <td>Procardia</td> </tr> <tr> <td>Identifier</td> <td>C0700861</td> </tr> <tr> <td>Type</td> <td>Pharmacologic Substance</td> </tr> <tr> <td>Confidence Score</td> <td>1</td> </tr> <tr> <td>Start Index</td> <td>631</td> </tr> <tr> <td>End Index</td> <td>640</td> </tr> <tr> <td>ICD-10 Code</td> <td>-</td> </tr> <tr> <td>id</td> <td>24</td> </tr> <tr> <td>Status</td> <td>Affirmed</td> </tr> </tbody> </table>	Pretty Name	Procardia	Identifier	C0700861	Type	Pharmacologic Substance	Confidence Score	1	Start Index	631	End Index	640	ICD-10 Code	-	id	24	Status	Affirmed
Pretty Name	Procardia																		
Identifier	C0700861																		
Type	Pharmacologic Substance																		
Confidence Score	1																		
Start Index	631																		
End Index	640																		
ICD-10 Code	-																		
id	24																		
Status	Affirmed																		

Рис. 9. Пример извлечения медицинских терминов из текста с использованием библиотеки MedCAT

Fig. 9. An example of extracting medical terms from a text using the MedCAT library

3.2.2. Модели машинного обучения

Средства анализа данных и алгоритмы Python, предоставляемые программным пакетом scikit-learn [12], обладают мощными возможностями, позволяющими решать различные задачи классификации данных. Алгоритм случайного леса с построением деревьев решений способен с высокой точностью прогнозировать заболевание по данным вида «болезнь – симптомы». На базе данного алгоритма построена модель диагностики заболеваний по симптомам.

В качестве исходного набора данных используется таблица, соотносящая симптомы с болезнями, сгенерированная автоматически исследователями Колумбийского университета на основе информации из текстовых справок о выписке пациентов Нью-Йоркской больницы,

поступивших в течение 2004 года. В первом столбце указано заболевание, во втором – соответствующие симптомы. На основе этих заметок были рассчитаны связи вида «болезнь – симптомы» для 150 наиболее частых заболеваний [13].

Помимо базовой модели диагностики, в процессе исследований были реализованы другие вспомогательные модели для реализации соответствующих сервисов диагностики и оценки рисков:

1. **Модель оценки риска сердечного приступа** [14]. Модель построена на основе датасета, содержащего 14 признаков: пол, возраст, уровень холестерина, кровяное давление, частота сердечных сокращений и т. д. Целевая переменная – попадание пациента в группу риска сердечного приступа (1 – да, 0 – нет).

2. **Модель диагностики заболеваний по перечню симптомов** [15]. Модель построена на основе датасета, имеющего 17 признаков, которые могут быть заполнены соответствующими симптомами (всего 132 вида симптомов), и целевое заболевание. На данный момент для сопоставления доступно 42 различных заболевания. Кроме того, к датасету прилагаются наборы данных с весами тех или иных симптомов (их влиянием на конечный диагноз), описанием симптомов и рекомендациями действий при тех или иных заболеваниях.

3.2.3. Онтологии

Помимо указанных выше моделей, в качестве инструментов для осуществления прогнозов заболеваний, оценки рисков и предоставления рекомендаций предлагается использовать стороннюю онтологию заболеваний (рис. 10) [16], которая при необходимости может быть дополнена новыми сущностями.

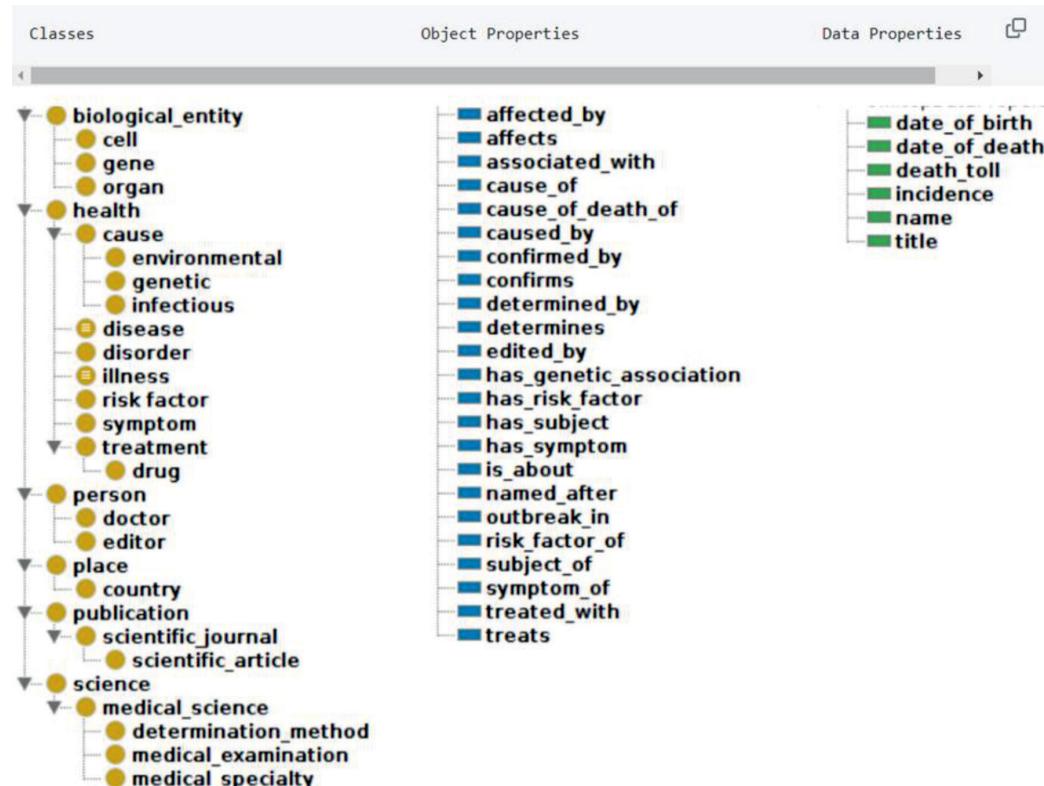


Рис. 10. Основные элементы онтологии заболеваний

Fig. 10. The main elements of the disease ontology

Данная онтология применима для следующих целей:

1. Диагностики заболеваний по перечню симптомов. Онтология семантически связывает 90 симптомов и 135 заболеваний отношениями *symptom_of* и *has_symptom*;
2. Поиска подходящего лечения, в частности лекарств. Для этого в онтологии присутствует класс *treatment* с разделом *drug* и перечнем лекарств, используемых для лечения заболеваний (отношение *treated_with*);
3. Поиска факторов риска. В онтологии присутствует раздел *risk_factors* и отношение *has_risk_factor*, что позволяет соотносить заболевания с факторами риска;
4. Поиска медицинской специальности, с которой связано данное заболевание (класс *medical_speciality*, отношение *associated_with*).

Для работы с онтологиями предлагается использовать пакет Owlready2 [17]. Пакет позволяет загружать, изменять, сохранять онтологии. Поддерживает правила вывода, используя входящие в комплект решатели HermiT и Pellet.

Кроме того, пакет предоставляет некоторый функционал ORM (объектно-реляционного отображения), позволяя использовать встроенные методы поиска, генерируя нужные SPARQL-запросы. Все это значительно упрощает процесс навигации по онтологиям. Особый интерес также представляет встроенный модуль PyMedTermino2. Этот модуль предоставляет удобный доступ к основным медицинским терминам и поддерживает все термины из онтологий UMLS, а также полнотекстовый поиск по терминам.

Заключение

В статье был описан подход к разработке интеллектуальных персонализированных медицинских информационных систем. В ходе исследования был проведен сравнительный анализ существующих на российском рынке доступных медицинских систем, ориентированных на персонализированный подход. На основе проведенного анализа были сформулированы основные функциональные возможности, которыми должна обладать ИПМИС. К ним относятся: хранение данных о пациентах в цифровом профиле, сбор данных для формирования цифрового профиля, отслеживание динамики и выявление факторов риска, предоставление индивидуальных диагнозов и прогнозов о наличии заболеваний и рекомендаций врачу и пациенту, а также предоставление удобных средств их связи в режиме онлайн. Также были выделены подходы и принципы, на основе которых должны строиться и функционировать ИПМИС.

В рамках предлагаемого подхода была определена архитектура типовой ИПМИС. Пользовательский интерфейс системы, реализованный как веб-приложение, предоставляет доступ к функциональным возможностям, поддерживаемым набором специальных сервисов. Все программные компоненты ИПМИС были реализованы с помощью современных методов, средств, фреймворков, библиотек, моделей.

Разработанные компоненты обладают открытым исходным кодом. Для работы с ними предоставляют API, что позволяет строить ИПМИС, специализированные на разные области медицины, на решение определенных задач, наращивать их функциональность, а также интегрировать их с существующими медицинскими информационными системами, добавляя возможности ведения учета индивидуальных параметров пациентов, осуществления мониторинга показателей здоровья, а также использовать методы машинного обучения как для автоматизации рутинной работы врача, так и для получения индивидуальных рекомендаций пациенту.

Научная новизна работы заключается в определении ключевых компонентов ИПМИС и разработке архитектурного и технологического решения, позволяющего интегрировать их в рамках единой системы, которая представляет собой веб-приложение, удовлетворяющее всем необходимым требованиям персонализированной системы с применением методов и технологий искусственного интеллекта, а именно: возможности учета биометрических параметров

пациентов и предоставления индивидуальных прогнозов и рекомендаций с помощью онтологий и моделей машинного обучения – моделей-классификаторов для диагностики заболеваний и оценки рисков, а также моделей обработки медицинских текстов с опорой на крупнейший медицинский тезаурус UMLS.

Практическая ценность предлагаемого подхода и разработанных на его основе систем состоит в снижении трудозатрат как разработчиков ИПМИС, так и практикующих врачей в процессе их функционирования. Кроме того, такие системы будут полезны пациентам, следящим за своим здоровьем. Они получат возможность наглядного мониторинга своих показателей здоровья, удаленного взаимодействия с лечащим врачом, что будет способствовать более осознанному и ответственному следованию его рекомендациям.

Дальнейшее развитие подхода будет проходить за счет использования новых моделей машинного обучения, развития онтологии заболеваний, добавления в цифровой профиль данных генетического паспорта пациента и разработки сервисов для работы с ними.

Список литературы

1. Дедов И. И., Тюльпаков А. Н., Чехонин В. П., Баклаушев В. П., Арчаков А. И., Мошковский С. А. Персонализированная медицина: современное состояние и перспективы // Вестник РАМН. 2012. № 12. С. 4–6.
2. Кайшева А. Л., Гришин Д. В., Каменский П. А., Федорончук Т. В., Мальсагова К. А., Чойнзонов Е. Л., Лисица А. В. Современные направления развития постгеномных медицинских технологий // Патологическая физиология и экспериментальная терапия, 2018. № 62 (3). С. 95–105.
3. Платформа дистанционного контроля состояния здоровья RemsMed. Официальный сайт: [Электронный ресурс]. URL: <https://remsmed.ru/>
4. Платформа прогнозной аналитики WebioMed. Официальный сайт: [Электронный ресурс]. URL: <https://webiomed.ru/>
5. Симптом-чекер MeDiCase. Официальный сайт: [Электронный ресурс]. URL: <https://medicase.pro/?ysclid=1qjle05znm907664215>
6. Орлов С. П., Золотов В. П., Якимов А. В. Архитектура интернет-платформы для персонализированной медицины. // Вестник Самарского гос. техн. ун-та. Серия: Технические науки. 2015. № 4.
7. Официальный сайт UMLS: [Электронный ресурс]. URL: <https://www.nlm.nih.gov/research/umls/index.html>
8. Документация фреймворка React: [Электронный ресурс]. URL: <https://react.dev/reference/>
9. Документация библиотеки MUI: [Электронный ресурс]. URL: <https://mui.com/material-ui/>
10. Kraljevic Z., Searle T., Shek A., Roguski L., Noor K., Bean D., Teo J., Windsor C., Gorrell G., Roberts A., Stewart R., Dobson R., Shah A. Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit // Artificial Intelligence in Medicine. 2021. Vol. 117.
11. Chanda A. K., Bai T., Yang Z., Vucetic S. Improving medical term embeddings using UMLS Metathesaurus // BMC Medical Informatics and Decision Making. 2022. Vol. 22. No. 1. P. 114.
12. Документация пакета Scikit-learn: [Электронный ресурс]. URL: <https://scikit-learn.org/stable/index.html>
13. Disease-Symptom Knowledge Database: [Электронный ресурс]. URL: <https://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html>
14. Модель оценки рисков сердечного приступа: [Электронный ресурс]. URL: https://colab.research.google.com/drive/1wOUD_gBQgjqcOi84h0Gzj3XHlp07vju?usp=sharing

15. Модель диагностики заболеваний по перечню симптомов: [Электронный ресурс]. URL: <https://colab.research.google.com/drive/1Kd3mdqXMDQy8bgTfEz4Mu6ZUXVILtVM?usp=sharing>
16. Dorsa-Arezoiji. Diseases-Ontology: [Электронный ресурс]. URL: <https://github.com/Dorsa-Arezoiji/Diseases-Ontology>
17. Документация пакета Owlready2: [Электронный ресурс]. URL: <https://owlready2.readthedocs.io/en/latest/>

References

1. **Dedov I. I., Tyul'pakov A. N., Chekhonin V. P., Baklaushev V. P., Archakov A. I., Moshkovskii S. A.** Personalized medicine: State-of-the-art and prospects. *Vestnik RAMN*, 2012, no. 12, pp. 4–6.
2. **Kaisheva A. L., Grishin D. V., Kamensky P. A., Fedoronchuk T. V., Malsagova K. A., Choynzonov E. L., Lisitsa A. V.** Modern directions of development of postgenomic medical technologies. *Pathological physiology and experimental therapy*, 2018, no. 62 (3), pp. 95–105.
3. Remote health monitoring platform “RemsMed”. Official website: [Electronic resource]. URL: <https://remsmed.ru/>
4. Predictive analytics platform “WebioMed”. Official website: [Electronic resource]. URL: <https://webiomed.ru/>
5. Symptom-checker MeDiCase. Official website: [Electronic resource]. URL: <https://medicase.pro/?ysclid=lqjle05znm907664215>
6. **Orlov S. P., Zolotov V. P., Yakimov A. V.** The architecture of the web-based platform for personalized medicine. *Vestnik of Samara State Technical University. The series “Technical Sciences”*, 2015, no. 4.
7. Documentation of the ScispaCy library: [Electronic resource]. URL: <https://allenai.github.io/scispacy/>
8. UMLS official Website: [Electronic resource]. URL: <https://www.nlm.nih.gov/research/umls/index.html>
9. Documentation of the Scikit-learn package: [Electronic resource]. URL: <https://scikit-learn.org/stable/index.html>
10. Disease-Symptom Knowledge Database: [Electronic resource]. URL: <https://people.dbmi.columbia.edu/~friedma/Projects/DiseaseSymptomKB/index.html>
11. Heart attack risk assessment model: [Electronic resource]. URL: https://colab.research.google.com/drive/1wOUD_gBQgjqcOi84h0Gzj3XHlip07vju?usp=sharing
12. A model for diagnosing diseases based on a list of symptoms: [Electronic resource]. URL: <https://colab.research.google.com/drive/1Kd3mdqXMDQy8bgTfEz4Mu6ZUXVILtVM?usp=sharing>
13. Dorsa-Arezoiji. Diseases-Ontology: [Electronic resource]. URL: <https://github.com/Dorsa-Arezoiji/Diseases-Ontology>
14. Documentation of the Owlready2 package: [Electronic resource]. URL: <https://owlready2.readthedocs.io/en/latest/>

Информация об авторах

Загорулько Галина Борисовна, доцент, старший научный сотрудник

Устюгов Владислав Алексеевич, студент магистратуры

Information about the Authors

Galina B. Zagorulko, Associate Professor, Senior Researcher

Vladislav A. Ustyugov, Graduate Student

*Статья поступила в редакцию 16.05.2025;
одобрена после рецензирования 10.07.2025; принятa к публикации 10.07.2025*

*The article was submitted 16.05.2025;
approved after reviewing 10.07.2025; accepted for publication 10.07.2025*