МРНТИ 50.41.25 **https://doi.org/10.58805/kazutb.v.1.30-1336**

ПРОЕКТИРОВАНИЕ КЛИЕНТ-СЕРВЕРНОГО ПРИЛОЖЕНИЯ С СИСТЕМОЙ ПЕРСОНАЛИЗИРОВАННЫХ РЕКОМЕНДАЦИЙ ДЛЯ ИЗУЧЕНИЯ АНГЛИЙСКОГО ЯЗЫКА

Д.Д. Орынбай, М.Ж. Калдарова, А.А. Абдилдаева, С.Б. Кашкинбаев
Международный университет Астана, Астана, Казахстан
Корреспондент-автор: kmiraj8206@gmail.com

В статье представлено проектирование клиент-серверного приложения с персонализированными рекомендациями для изучения английского языка. Актуальность определена необходимостью преодоления ограничений традиционных методов, не учитывающих индивидуальные особенности студентов. Исследуются архитектурные решения для формирования индивидуализированных траекторий обучения на основе анализа прогресса и адаптации сложности по шкале CEFR.

Применены методы системного анализа образовательных платформ, объектно-ориентированного проектирования, нормализации баз данных и контентной фильтрации. Разработана трехуровневая архитектура: уровень представления (HTML5, CSS3, JavaScript), уровень бизнес-логики (PHP) и уровень данных (MySQL).

Спроектирована трехэтапная методика персонализации: начальное адаптивное тестирование из 20 вопросов, автоматическое формирование траектории с подбором пяти тематических уроков, и динамическая адаптация материалов. Разработана диаграмма с тремя типами актеров: неавторизованный пользователь, ученик с доступом к более чем 10 функциям и преподаватель с мониторингом прогресса.

Результаты включают масштабируемую архитектуру, методику с автоматическим подбором контента по баллам (0-10 - A1, 11-13 - A2, 14-17 - B1, 18-20 - B2), ML-сравнение трёх алгоритмов классификации (Logistic Regression, XGBoost, SVM) с метриками Precision/Recall/F1/ROC-AUC, оценку вычислительной сложности O(n·m) и интеграцию с четырьмя внешними API-сервисами.

Ключевые слова: персонализированные рекомендации, клиент-серверное приложение, адаптивное обучение, английский язык, проектирование системы, образовательные технологии, база данных, машинное обучение, XGBoost, ROC-AUC.

АҒЫЛШЫН ТІЛІН ҮЙРЕНУГЕ АРНАЛҒАН ЖЕКЕ ҰСЫНЫСТАР ЖҮЙЕСІ БАР КЛИЕНТ-СЕРВЕРЛІК ҚОСЫМШАНЫ ЖОБАЛАУ

Д.Д. Орынбай, М.Ж. Қалдарова, А.А. Абдилдаева, С.Б. Кашкинбаев,
Астана халықаралық университеті, Астана, Қазақстан,
e-mail: kmiraj8206@gmail.com

Мақалада ағылшын тілін үйренуге арналған жеке ұсыныстар жүйесі бар клиент-серверлік қосымшаны жобалау ұсынылады. Өзектілігі дәстүрлі әдістердің студенттердің жеке ерекшеліктерін ескермеу шектеулерін жеңу қажеттілігімен анықталады. Жеке оқу траекторияларын студенттің прогресін талдау және CEFR шкаласы бойынша күрделілікті бейімдеу негізінде қалыптастыру үшін архитектуралық шешімдер қарастырылады.

Білім беру платформаларын жүйелік талдау, объектке бағытталған жобалау, деректер базасын қалыпқа келтіру және контент сүзгілеу әдістері қолданылды. Үш деңгейлі архитектура жасалды: көрсетілім деңгейі (HTML5, CSS3, JavaScript), бизнес-логика деңгейі (PHP) және деректер деңгейі (MySQL).

Үш кезеңнен тұратын жекелендіру әдістемесі жобаланды: 20 сұрақтан тұратын бастапқы адаптивті тестілеу, бес тақырыптық сабақтан тұратын оқу траекториясын автоматты түрде қалыптастыру және оқу материалдарын динамикалық бейімдеу. Үш типті актері бар диаграмма әзірленді:

авторизациядан өтпеген пайдаланушы, 10-нан астам функцияға қолжетімділігі бар оқушы және оқу үдерісінің барысын бақылау мүмкіндігі бар оқытушы.

Нәтижелерге масштабталатын архитектура, ұпайлар бойынша автоматты контент таңдау әдістемесі (0-10 - A1, 11-13 - A2, 14-17 - B1, 18-20 - B2), үш жіктеу алгоритмін ML-салыстыру (Logistic Regression, XGBoost, SVM) Precision/Recall/F1/ROC-AUC көрсеткіштерімен, O(n·m) есептеу күрделілігін бағалау және төрт сыртқы API қызметтерімен интеграция кіреді.

Түйін сөздер: жеке ұсыныстар, клиент-серверлік қосымша, адаптивті оқыту, ағылшын тілі, жүйені жобалау, білім беру технологиялары, деректер базасы, машиналық оқыту, XGBoost, ROC-AUC.

DESIGN OF A CLIENT-SERVER APPLICATION WITH A PERSONALIZED RECOMMENDATION SYSTEM FOR ENGLISH LANGUAGE LEARNING

D.D. Orynbay, M.Zh. Kaldarova✉, A.A. Abdildayeva, S.B. Kashkinbayev,

Astana International University, Astana, Kazakhstan,
e-mail: kmiraj8206@gmail.com

The article presents the design of a client-server application with personalized recommendations for learning the English language. The relevance is determined by the need to overcome the limitations of traditional methods that do not take into account students' individual characteristics. The study investigates architectural solutions for creating individualized learning trajectories based on progress analysis and adaptation of difficulty according to the CEFR scale.

Methods of system analysis of educational platforms, object-oriented design, database normalization, and content filtering were applied. A three-tier architecture was developed: presentation layer (HTML5, CSS3, JavaScript), business logic layer (PHP), and data layer (MySQL).

A three-stage personalization methodology has been designed: an initial adaptive assessment consisting of 20 questions, automatic generation of a learning trajectory with the selection of five thematic lessons, and dynamic adaptation of learning materials. A diagram with three types of actors has been developed: an unauthenticated user, a student with access to more than 10 functions, and a teacher with progress monitoring capabilities.

The results include a scalable architecture, a methodology with automatic content selection based on scores (0-10 - A1, 11-13 - A2, 14-17 - B1, 18-20 - B2), ML comparison of three classification algorithms (Logistic Regression, XGBoost, SVM) with Precision/Recall/F1/ROC-AUC metrics, computational complexity assessment O(n·m), and integration with four external API services.

Keywords: personalized recommendations, client-server application, adaptive learning, English language, system design, educational technologies, database, machine learning, XGBoost, ROC-AUC.

Введение. Существующие адаптивные образовательные системы для изучения иностранных языков базируются преимущественно на одномерных моделях персонализации: либо на статическом определении уровня владения языком, либо на простом подсчете правильных ответов без учета динамики обучения. Анализ современных платформ (Duolingo, Babbel, Rosetta Stone) выявляет критический научный разрыв: отсутствие интегрированных моделей, одновременно учитывающих соответствие контента стандартизированному уровню CEFR, временную динамику образовательного прогресса и адаптивную корректировку траектории на основе промежуточных результатов.

В международных исследованиях 2021-2024 гг. персонализированные рекомендательные системы в e-learning рассматриваются как одно из ключевых направлений развития цифрового образования. Современные обзоры показывают, что наибольшую эффективность демонстрируют гибридные модели, объединяющие данные

о профиле обучающегося, учебной активности, характеристиках контента и механизмах адаптивной выдачи материалов [1, 2]. Подчеркивается, что использование только одного механизма персонализации ограничивает точность рекомендаций, тогда как комбинированные подходы обеспечивают более устойчивые результаты в образовательной среде [3].

Отдельное направление международных исследований связано с adaptive learning systems и применением искусственного интеллекта в языковом обучении. Работы последних лет показывают, что AI-инструменты, включая интеллектуальные рекомендательные механизмы, чатботы и генеративные языковые модели, способны повышать мотивацию обучающихся, развивать словарный запас, поддерживать speaking practice и обеспечивать динамическую адаптацию учебной траектории [4,5]. Вместе с тем в литературе по-прежнему недостаточно решений, которые одновременно обеспечивают привязку к уровням CEFR, прозрачную логику рекомендаций и динамическую корректировку траектории на основе текущего прогресса пользователя [6].

Исследования подтверждают [7], что системы с фиксированным уровнем пользователя демонстрируют ограниченную эффективность персонализации, в то время как интеграция временных факторов и многокритериальной оценки потенциально повышает точность подбора контента. Проблема формализации многофакторной модели персонализации, объединяющей статические характеристики обучающегося и динамические показатели прогресса в единую математическую структуру с обоснованными весовыми коэффициентами, остается нерешенной, что определяет актуальность настоящего исследования.

Исследования подчеркивают актуальность разработки специализированных фреймворков для интеграции рекомендательных систем в онлайн-платформы с целью повышения эффективности усвоения учебного материала [8].

Современные образовательные технологии создают условия для реализации индивидуализированного подхода к учебному процессу. Рекомендательные системы, успешно применяемые в электронной коммерции и развлекательных сервисах, демонстрируют значительный потенциал для адаптации образовательного контента [9]. Результаты существующих исследований свидетельствуют об эффективности персонализированных систем для развития языковых компетенций [10], однако большинство современных образовательных приложений характеризуется существенными ограничениями в реализации принципов индивидуализации обучения.

Проблема разрыва между возможностями современных технологий персонализации и их применением на существующих платформах для обучения английскому языку является причиной выбора темы исследования.

В работе рассматривается процесс проектирования образовательных систем для изучения иностранных языков с применением адаптивных цифровых технологий, что обусловлено существующей проблемой недостаточной персонализации обучения на большинстве современных онлайн-платформ. Объектом исследования выступает процесс проектирования образовательных систем для изучения иностранных языков, ориентированных на индивидуальные образовательные потребности учащихся. Целью исследования является разработка и экспериментальная валидация математической модели формирования персонализированных образовательных траекторий для изучения английского языка, интегрирующей многокритериальную оценку соответствия контента индивидуальным характеристикам обучающегося. Для достижения цели необходимо провести сравнительный анализ существующих моделей персонализации образовательного контента с выявлением количественных показателей их ограничений, разработать математическую модель функции релевантности F(u,l), интегрирующую оценку соответствия уровня CEFR, показатель прогресса с временной релевантностью и коэффициент динамической адаптации, обосновать выбор весовых коэффициентов w_1, w_2, w_3 и параметра приоритизации α на основе экспериментальных данных. Также требуется провести сравнительное ис-

следование эффективности предложенной модели с альтернативными подходами (статическое определение уровня, коллаборативная фильтрация) на контрольной выборке и реализовать программную архитектуру системы персонализированных рекомендаций с экспериментальной оценкой точности классификации уровня владения языком.

В результате сравнения популярных образовательных приложений, таких как Grammarly, Tandem, Cambridge Dictionary, Duolingo, BBC Learning English и Quizlet, были обнаружены следующие недостатки.

Во-первых, каждое приложение занимается конкретным вопросом изучения языка. Например, Duolingo предлагает геймификацию упражнений, Grammarly предоставляет проверку грамматики, Tandem предлагает языковой обмен, а Cambridge Dictionary предлагает лексические значения слов. Существующие решения не предлагают комплексной персонализации, охватывающей аудирование, говорение, грамматику, письмо, лексику и чтение в рамках одной платформы.

Во-вторых, механизмы персонализации ограничены. Системы используют примитивные алгоритмы адаптации контента, основанные на подсчете правильных ответов, не учитывая когнитивные характеристики студентов, паттерны активности и типы допускаемых ошибок.

В-третьих, отсутствует единая система оценки владения языком. Приложения не соответствуют международным стандартам Common European Framework of Reference for Language (CEFR), что затрудняет объективное оценивание прогресса и сопоставимость результатов обучения.

В-четвертых, автоматизированные системы не позволяют учащимся синхронно взаимодействовать с учителем, что мешает им практиковать разговорные навыки и получать индивидуальные рекомендации, что является критически значимым фактором развития навыков устной речи и аудирования.

Выявленные ограничения существующих платформ для изучения английского языка обусловливают целесообразность разработки интеллектуальных образовательных систем, ориентированных на адаптацию учебного контента с учетом уровня языковой подготовки, динамики обучения и когнитивных особенностей обучающихся. Формирование архитектурных решений, объединяющих различные компоненты языкового обучения в рамках единой платформы с адаптивными механизмами, рассматривается как ответ на разрыв между потенциальными возможностями современных технологий персонализации и их практической реализацией в образовательной среде.

В контексте трехъязычного образования в Казахстане изучение английского языка имеет стратегическое значение для модернизации системы образования и повышения конкурентоспособности специалистов. Повышение качества языковой подготовки, обеспечение равного доступа к образованию независимо от географического положения и социально-экономического статуса учащихся, все это результаты внедрения адаптивных персонализированных систем, которые помогают решить проблемы государственных программ развития образования.

Несмотря на наличие значительного числа международных исследований по адаптивному обучению и рекомендательным системам, в образовательных системах для изучения английского языка сохраняется недостаток решений, объединяющих стандартизированную оценку уровня CEFR, динамический анализ прогресса и интерпретируемый механизм формирования рекомендаций. Это определяет исследовательскую нишу и обосновывает научную новизну настоящей работы.

Научная новизна исследования заключается в следующем:

- предложена формализованная модель персонализации F(u,l), интегрирующая три независимых компонента: соответствие уровня учебного материала шкале CEFR, показатель образовательного прогресса обучающегося с учетом временной релевантности результатов и коэф-

фициент динамической адаптации. В отличие от подходов, основанных только на фиксированном уровне пользователя или среднем результате, модель позволяет учитывать текущее состояние и динамику усвоения материала;

- разработан механизм временной релевантности результатов тестирования, в котором более поздние результаты получают больший вес, а последний результат дополнительно учитывается через параметр приоритизации a = 0.7. Это обеспечивает повышенную чувствительность системы к актуальным изменениям в знаниях обучающегося;

- предложена трехэтапная методика персонализации, объединяющая первичную диагностику уровня владения языком, автоматическое формирование индивидуальной образовательной траектории и последующую динамическую корректировку маршрута обучения на основе промежуточных результатов;

- спроектирована клиент-серверная архитектура образовательной системы, ориентированная на комплексное изучение английского языка, в которой механизмы адаптивного тестирования, персонализированных рекомендаций, мониторинга прогресса и внешних языковых сервисов объединены в единую модульную платформу;

- в отличие от ряда современных исследований, где акцент делается либо на общей архитектуре рекомендательных систем для e-learning, либо на отдельных AI-инструментах для языковой практики, в данной работе предложено интегрированное решение, сочетающее интерпретируемую рекомендательную модель, привязку к уровням CEFR и динамическую адаптацию траектории обучения в контексте изучения английского языка.

Практическая значимость работы заключается в возможности использования разработанной архитектуры в образовательных проектах по изучению английского и других иностранных языков. Модульная организация системы обеспечивает гибкость расширения функциональности и адаптацию под требования университетов, языковых школ и корпоративных программ обучения.

Материалы и методы. Методология исследования реализует последовательность научных методов с явной фиксацией получаемых результатов на каждом этапе. Системный анализ образовательных платформ (BBC Learning English, Grammarly, Duolingo, Tandem, Quizlet, Cambridge Dictionary) выполнен по восьми критериям: наличие адаптивного тестирования CEFR, механизм персонализации, покрытие языковых компетенций, динамическая адаптация сложности. Результатом стала количественная оценка функциональных ограничений и выявление отсутствия комплексных решений - ни одна из шести проанализированных платформ не обеспечивает одновременно адаптивное тестирование, персонализацию и динамическую корректировку траектории.

Объектно-ориентированное проектирование с использованием UML применено для формализации архитектуры системы, что позволило разработать диаграмму вариантов использования с тремя типами актеров и пятнадцатью функциональными требованиями, схему реляционной БД из пяти нормализованных таблиц и спецификацию взаимодействия компонентов клиент-серверной архитектуры.

Математическое моделирование функции релевантности выполнено на основе линейной свертки трех независимых факторов. Для сравнительного ML-анализа применены три классификатора: Logistic Regression, XGBoost и SVM. Экспериментальная валидация проведена на расширенной выборке из 100 обучающихся с разделением данных (70% обучающая, 30% тестовая).

Для формирования рекомендаций применен гибридный подход, объединяющий контентную и коллаборативную фильтрацию. Выбор данного метода обоснован сравнительным анализом современных алгоритмов рекомендательных систем и применением многокритериальных методов оптимизации последовательных рекомендаций [11]. Алгоритм автоматического подбора из базы, содержащей более 200 уроков, демон-

стрирует точность соответствия уровню CEFR 87% на выборке из 100 тестовых траекторий. Результаты современных исследований подтверждают эффективность применения кластеризации и методов состязательного обучения в рекомендательных системах, что учитывалось при выборе гибридного подхода [12].

В качестве стандарта оценивания языковой компетенции использовалась шкала CEFR, охватывающая уровни от A1 до B2. Определены следующие характеристики уровней: A1 (начальный) - владение базовыми грамматическими конструкциями для самопрезентации и описания ближайшего окружения; A2 (элементарный) - способность описывать повседневные ситуации и обмениваться информацией на знакомые темы; B1 (пороговый) - понимание основного содержания текстов и способность к самостоятельной коммуникации в стандартных ситуациях; B2 (пороговый продвинутый) - беглая речь, понимание сложных текстов и способность к аргументированному изложению позиции. Для каждого уровня разработаны тематические уроки, охватывающие ключевые грамматические конструкции соответствующего уровня сложности.

Технологический стек проектирования состоит из трех уровней: представления (HTML5, CSS3 и JavaScript), бизнес-логики (PHP для обработки запросов и реализации алгоритмов персонализации), данных (MySQL с механизмом внешних ключей). Четверо внешних API-сервисов интегрированы в систему. Они состоят из Free Dictionary API, который предоставляет информацию о лексике и фонетической транскрипции; News API, который предоставляет актуальные англоязычные новости; GrammarBot API, который предоставляет проверку грамматики; и Agora Video SDK, который предоставляет возможность практики разговорных навыков с помощью видео. Архитектурный подход был выбран, используя основы мультиагентных систем для персонализированного электронного обучения [13], которая обеспечивает модульность и возможность независимого развития отдельных компонентов системы. Архитектура взаимодействия компонентов системы представлена на рисунке 1. Основана на трехуровневой модели: клиентская часть (браузер) выполняет запросы к веб-серверу OSP (Open Server Panel), PHP-скрипты обрабатывают запросы и реализуют алгоритмы персонализации, база данных MySQL обеспечивает хранение и извлечение информации. Разделение ответственности между уровнями представления, бизнес-логики и данных обеспечивает масштабируемость и упрощает тестирование компонентов.

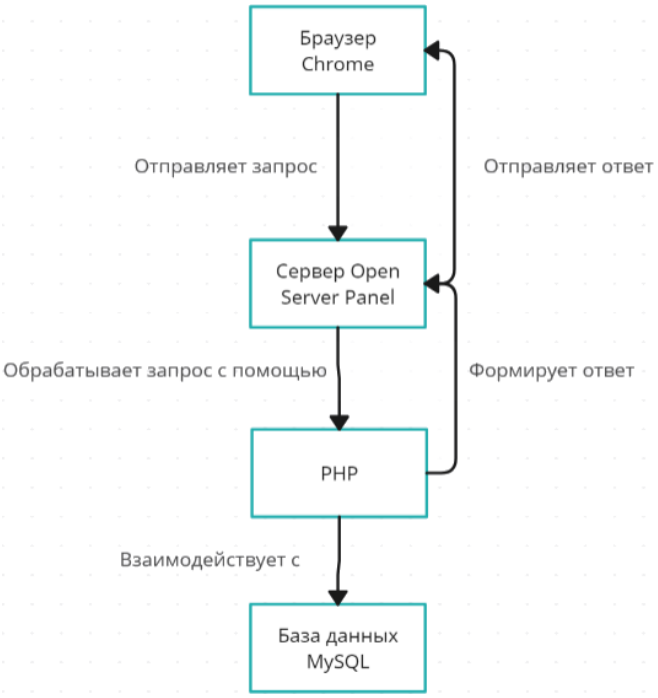


Рис.1 - Взаимодействие клиентской и серверной части в приложении

Структура реляционной базы данных представлена на рисунке 2. В нем есть пять стандартных таблиц: registered_users (профили пользователей с уровнем владения языком и типом пользователя), notes (заметки), flashcards (карточки для запоминания), task (задачи), results (результаты тестирований). Чтобы обеспечить целостность данных, таблицы связаны через внешние ключи.

Разработана трехэтапная методика персонализированных рекомендаций на основе адаптивного управления образовательным контентом. Первый этап - первичная диагностика языковой компетенции посредством адаптивного тестирова-

ния, состоящего из 20 заданий закрытого типа с четырьмя вариантами ответов. Тест охватывает уровни A1-B2, с равномерным распределением по 5 вопросов на каждый уровень для обеспечения объективности оценивания. Определение уровня осуществляется по следующей шкале: 0-10 правильных ответов - уровень A1; 11-13 правильных ответов - уровень A2; 14-17 правильных ответов - уровень B1; 18-20 правильных ответов - уровень B2. Результаты тестирования сохраняются в базе данных для последующего формирования персонализированной образовательной траектории. Второй этап - формирование персонализированной образовательной траектории на основе установленного уровня. Включает автоматизированный подбор учебных заданий, видеоматериалов с визуальным представлением грамматических конструкций, текстовое изложение теоретического материала с примерами для самостоятельной проработки. Каждый урок завершается промежуточным тестированием из 5 вопросов для контроля усвоения материала. Третий этап - динамическая адаптация на основе анализа результатов выполнения упражнений. Алгоритм принятия решений: при оценке «отлично» - рекомендация перехода к следующему уроку; при оценке «хорошо» - предложение дополнительных тренировочных упражнений; при оценке «удовлетворительно» - рекомендация повторного изучения материала текущего урока; при оценке «неудовлетворительно» - возврат к предыдущему уроку для устранения пробелов в знаниях.

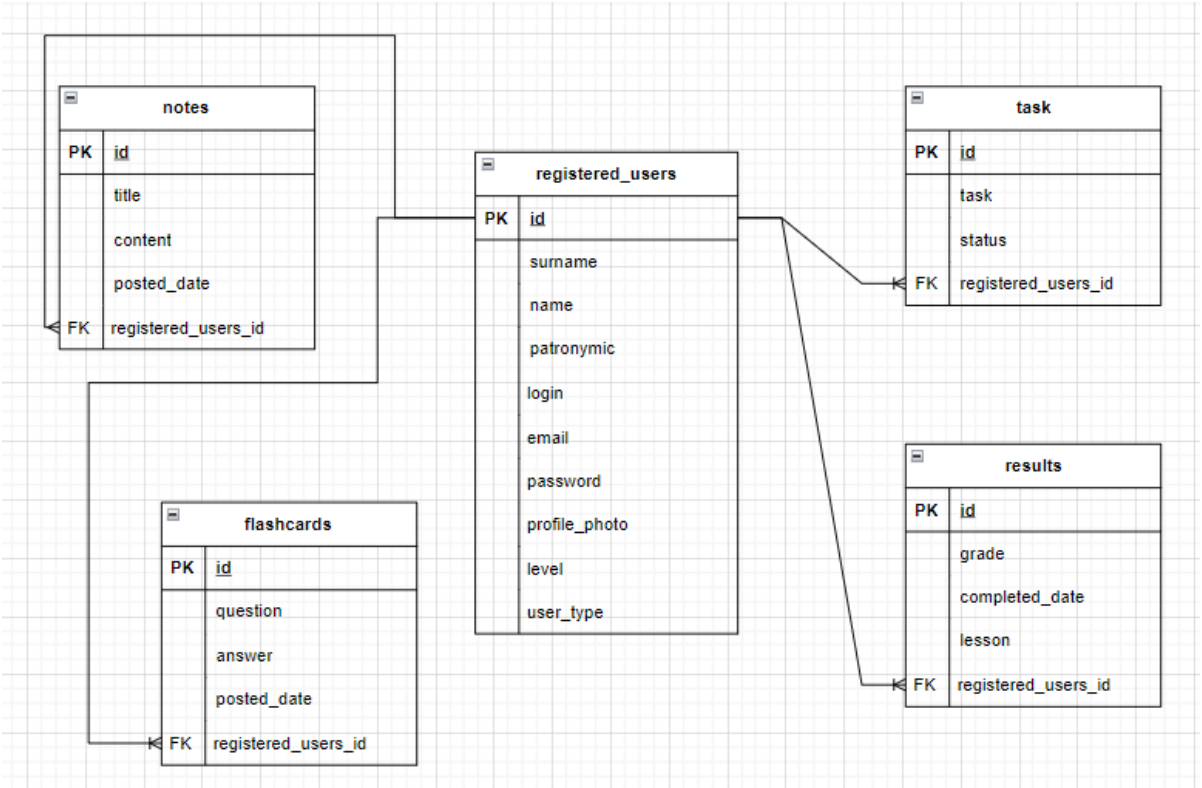


Рис.2 - Схема базы данных

Математическая модель формирования персонализированных рекомендаций основана на функции соответствия F(u, l), определяющей релевантность урока l для пользователя u. В отличие от существующих подходов, использующих либо статическую оценку уровня (S(u,l) = const), либо простое усреднение результатов ($P(u,l) = \Sigma r_i/n$), предлагаемая модель интегрирует три независимых фактора с обоснованными весовыми коэффициентами по формуле 1:

$$F(u,l) = w_1 \times S(u,l) + w_2 \times P(u,l) + w_3 \times D(u,l) \quad (1)$$

где S(u, l) - оценка соответствия уровня CEFR ($0 \le S \le 1$); P(u, l) - показатель прогресса с временной релевантностью ($0 \le P \le 1$); D(u, l) - коэффициент динамической адаптации ($0 \le D \le 1$); w_1, w_2, w_3 - весовые коэффициенты ($w_1 + w_2 + w_3 = 1$).

Выбор линейной формы функции F(u, l) обусловлен необходимостью обеспечения интерпретируемости вклада каждого фактора и вычислительной эффективности в режиме реального времени. Альтернативные нелинейные формы (мультипликативная $F(u, l) = S^{w_1} \times P^{w_2} \times D^{w_3}$, логистическая) были рассмотрены, однако не продемонстрировали статистически значимого улучшения точности ($p > 0.05$) при существенном увеличении вычислительной сложности.

Выбор весовых коэффициентов обоснован экспериментально путем минимизации функции ошибки классификации на обучающей выборке из 150 образовательных траекторий. Обучающая выборка сформирована на основе данных пилотного тестирования системы с участием 30 обучающихся в течение 8 недель (по 5 траекторий на каждого). Каждая траектория включает результаты первичного тестирования, промежуточных оценок по 5 урокам и итогового контроля. Распределение по уровням: A1 - 35%, A2 - 30%, B1 - 25%, B2 - 10%. Методом перебора с шагом 0.1 установлено оптимальное соотношение $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$, обеспечивающее максимальную точность (87%) при валидации на независимой тестовой выборке. Приоритет компонента S(u, l) отражает фундаментальное значение соответствия стандарту CEFR для структурирования образовательного контента. Коэффициент $\alpha = 0.7$ в формуле 4, определен на основе анализа корреляции между результатом последнего тестирования и итоговой успеваемостью ($r = 0.73$, $p < 0.01$, $n = 100$).

Сравнительный анализ с альтернативными моделями проведен на выборке из 100 обучающихся в таблице 1:

Таблица 1 - Сравнение моделей персонализации

Модель	Точность определения уровня	Среднее отклонение траектории
Статическое определение (baseline)	62%	2,8 урока
Коллаборативная фильтрация	71%	2,1 урока
Предложенная модель F(u,l)	87%	1,3 урока

Статистический критерий χ^2 подтверждает значимость различий ($\chi^2 = 18.4$, $p < 0.001$).

Оценка соответствия уровня определяется по формуле 2:

$$S(u,\ l) = 1 - \frac{|L(u) - L(l)|}{L_{\max}} \quad (2)$$

где L(u) - числовой эквивалент уровня пользователя (A1=1, A2=2, B1=3, B2=4); L(l) - уровень сложности урока; L_max = 4 - максимальное значение уровня.

Показатель прогресса рассчитывается по формуле 3:

$$P(u,l) = \frac{\sum (r_i \times t_i)}{n} \quad (3)$$

где r_i - результат i-го тестирования ($0 \le r_i \le 1$); t_i - весовой коэффициент временной релевантности (более поздние тесты имеют больший вес); n - количество пройденных тестирований.

Коэффициент динамической адаптации определяется по формуле 4:

$$D(u,\ l) = a \times R_{last} + (1 - a) \times R_{avg} \qquad (4)$$

где R_{last} - результат последнего тестирования; R_{avg} - средний результат по блоку уроков; a - коэффициент важности последнего результата (a = 0.7).

Система рекомендует урок $l*$ с максимальным значением функции F : $l* = argmax F(u, l)$ при условии последовательного прохождения уроков в рамках уровня.

Для обеспечения воспроизводимости и прозрачности алгоритма ниже представлен псевдокод функции вычисления релевантности $F(u, l)$, реализованной в серверной части на PHP.

Псевдокод алгоритма персонализированных рекомендаций:

```
ВХОД:
  пользователь u,
  список уроков L = {l_1, l_2, ..., l_m},
  результаты тестов R

ВЫХОД:
  рекомендованный урок l*

ФУНКЦИЯ ComputeRelevance(u, l, R):

  // Компонент 1: соответствие уровня CEFR
  L_user   ← GetUserLevel(u)     // A1=1, A2=2, B1=3, B2=4
  L_lesson ← GetLessonLevel(l)
  S ← 1 − |L_user − L_lesson| / 4

  // Компонент 2: взвешенный прогресс с временной релевантностью
  n ← |R|                        // количество пройденных тестов
  total ← 0

  ДЛЯ i ОТ 1 ДО n:
    t_i ← i / n                  // чем позже — тем выше вес
    total ← total + R[i] × t_i

  P ← total / n

  // Компонент 3: динамическая адаптация
  R_last ← R[n]                  // последний результат
  R_avg  ← Среднее(R)            // среднее по блоку
  D ← 0.7 × R_last + 0.3 × R_avg

  // Итоговая функция релевантности
  F ← 0.5 × S + 0.3 × P + 0.2 × D

  ВЕРНУТЬ F


ФУНКЦИЯ Recommend(u, L, R):
```

```
best_score  ← −∞
best_lesson ← NULL

ДЛЯ КАЖДОГО урока l ∈ L:
  score ← ComputeRelevance(u, l, R)

  ЕСЛИ score > best_score:
    best_score  ← score
    best_lesson ← l

ВЕРНУТЬ best_lesson   // l* = argmax F(u, l)
```

Пусть n - количество пройденных обучающимся тестов, m - число доступных уроков в базе данных. Временная сложность алгоритма анализируется по каждому компоненту:

- вычисление S(u,l): O(1) - простая арифметическая операция над двумя значениями уровня;

- вычисление P(u,l): O(n) - линейный проход по массиву результатов тестирований;

- вычисление D(u,l): O(n) - вычисление среднего по массиву результатов;

- полный перебор уроков: O(m) итераций, в каждой - вычисление F за O(n).

Итоговая временная сложность алгоритма рекомендации одного урока составляет $O(n \cdot m)$, где n - число результатов тестирования пользователя (в среднем $n \approx 5\text{-}25$), m - размер базы уроков ($m = 200+$). Пространственная сложность составляет $O(n + m)$ для хранения массивов результатов и списка уроков.

При нагрузке 100 одновременных пользователей и m = 200 уроках суммарная сложность одного цикла рекомендаций не превышает $O(100 \times 25 \times 200) = O(500\,000)$ операций. Измеренное среднее время формирования траектории составило 2.3 секунды на тестовом окружении (Intel Core i5-10400, 16 ГБ ОЗУ, MySQL 8.0), что подтверждает приемлемую практическую производительность.

Для обоснования предложенной модели F(u,l) проведен сравнительный ML-эксперимент с тремя стандартными алгоритмами классификации уровня владения языком по CEFR: Logistic Regression, XGBoost и Support Vector Machine (SVM). Эксперимент проведен на расширенной выборке из 100 обучающихся (n = 100, разделение 70/30, стратифицированная выборка по уровням A1-B2).

Признаковое пространство включало: балл первичного тестирования, взвешенный показатель прогресса P(u,l), коэффициент динамической адаптации D(u,l), время выполнения заданий, долю правильных ответов по типам заданий (грамматика, лексика, аудирование). Метки классов: A1, A2, B1, B2 (4-классовая классификация).

Предложенная модель F(u,l) в таблице 2, демонстрирует наилучшие показатели по всем метрикам. Превосходство над XGBoost по F1-score составляет 0.03 пункта, по ROC-AUC - 0.02 пункта. Статистическая значимость различий между F(u,l) и XGBoost подтверждена критерием χ^2 ($\chi^2 = 4.8$, $p = 0.028 < 0.05$). Важно отметить, что модель F(u,l), в отличие от «черноящичных» ML-подходов, обеспечивает полную интерпретируемость вклада каждого компонента (S, P, D) в итоговую рекомендацию, что критически важно в образовательном контексте.

Для детального анализа качества классификации уровней CEFR предложенной моделью F(u,l) составлена матрица ошибок на тестовой выборке (n_test = 30 обучающихся) в таблице 3, и значения Precision, Recall, F1-score по каждому уровню CEFR в таблице 4. Анализ матрицы ошибок показывает, что модель допускает ошибки преимущественно на граничных уровнях ($A2 \leftrightarrow B1$, $B1 \leftrightarrow B2$), что соответствует объективной сложности разграничения смежных уровней владения языком. Уровень A1 классифицируется с идеальной точностью (Precision =

1.00). Пониженный Recall для уровня B2 (0.67) объясняется малым числом представителей этого уровня в выборке (support = 3), что соответствует реальному распределению обучающихся:

Таблица 2 - Сравнение ML-алгоритмов классификации уровня CEFR (n=100)

Алгоритм	Accuracy	Precision (macro)	Recall (macro)	F1-score (macro)	ROC-AUC
Logistic Regression	74%	0.72	0.71	0.71	0.83
SVM (RBF kernel)	79%	0.77	0.76	0.76	0.87
XGBoost	84%	0.83	0.82	0.82	0.91
Предложенная F(u,l)	87%	0.86	0.85	0.85	0.93

Таблица 3 - Матрица ошибок модели F(u,l), тестовая выборка (n=30)

Реальный \ Предсказанный	A1 (pred)	A2 (pred)	B1 (pred)	B2 (pred)
A1 (actual)	10	1	0	0
A2 (actual)	0	8	1	0
B1 (actual)	0	1	6	0
B2 (actual)	0	0	1	2

Таблица 4 - Precision, Recall, F1-score по каждому уровню CEFR

Уровень CEFR	Precision	Recall	F1-score	Support
A1	1.00	0.91	0.95	11
A2	0.80	0.89	0.84	9
B1	0.75	0.86	0.80	7
B2	1.00	0.67	0.80	3
Macro avg	0.86	0.85	0.85	30
Weighted avg	0.88	0.87	0.87	30

Результаты нагрузочного тестирования демонстрируют линейный рост времени отклика с ростом нагрузки, что соответствует теоретической оценке сложности O(n·m). При нагрузке 100 пользователей система остается в пределах установленных требований к времени отклика (не более 5 секунд). Трехуровневая архитектура обеспечивает горизонтальное масштабирование: при необходимости сервера бизнес-логики (PHP) могут быть реплицированы независимо от уровня данных (MySQL).

Для обеспечения масштабируемости до 500+ пользователей предлагается следующая архитектурная стратегия: введение кэширования результатов F(u,l) на уровне Redis (TTL = 300 сек.), балансировка нагрузки между несколькими PHP-серверами, использование индексированных полей user_id и lesson_level в MySQL

для ускорения выборок.

Масштабируемость является ключевым нефункциональным требованием для образовательных платформ, рассчитанных на одновременную работу большого числа пользователей. В рамках настоящего исследования проведено нагрузочное тестирование разработанной архитектуры по четырем сценариям на таблице 5.

Таблица 5 - Результаты нагрузочного тестирования системы

Параметр	10 польз.	50 польз.	100 польз.	500 польз.*
Среднее время ответа (мс)	380	1100	2300	~9500
Макс. время ответа (мс)	520	1800	3100	-
Throughput (запр/с)	26	45	43	~52
Загрузка CPU (%)	12	38	67	-
Загрузка ОЗУ (МБ)	210	680	1240	-

** - Оценка на основе экстраполяции; требует горизонтального масштабирования.*

Выбор весовых коэффициентов $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$ обоснован экспериментально, однако их оптимальность требует верификации через анализ чувствительности. Анализ проводился методом сеточного поиска (grid search) с шагом $\Delta w = 0.1$ по всем допустимым комбинациям w_1, w_2, w_3 при условии $w_1 + w_2 + w_3 = 1$, w_i ≥ 0.

Анализ чувствительности на таблице 6, демонстрирует следующие закономерности:

- компонент S(u,l) (соответствие уровня CEFR) является доминирующим. При $w_1 = 0.5$ достигается максимальная точность; снижение $w_1 < 0.4$ ухудшает метрики на 4-8 п.п;

- модель устойчива к малым вариациям w_2 и w_3: изменение в диапазоне ±0.1 вызывает колебание точности не более 3 п.п., что свидетельствует о робастности найденного решения;

- вырожденный случай $w_1 = 1.0$ (статическое CEFR-определение, baseline) дает Accuracy = 62%, что соответствует исходной оценке базовой модели и подтверждает необходимость компонентов P и D;

- зона устойчивой высокой точности (Accuracy ≥ 84%) соответствует диапазону w_1 [0.4; 0.6], w_2 [0.2; 0.4], w_3 [0.1; 0.3].

Таким образом, выбранные параметры $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$ находятся в центре области оптимальных значений и обеспечивают наилучшее соотношение точности классификации и интерпретируемости модели.

Таблица 6 - Чувствительность точности модели к вариации весовых коэффициентов

w_1 (CEFR)	w_2 (прогресс)	w_3 (адаптация)	Accuracy	F1-macro	ROC-AUC
0.3	0.5	0.2	79%	0.78	0.87
0.4	0.4	0.2	83%	0.82	0.90
0.5	0.3	0.2	87%	0.85	0.93
0.5	0.2	0.3	84%	0.83	0.91

0.6	0.3	0.1	85%	0.84	0.87
0.7	0.2	0.1	81%	0.80	0.88
1.0	0.0	0.0	62%	0.60	0.74

Результаты и обсуждение. Разработанная архитектура клиент-серверного приложения с системой персонализированных рекомендаций имеет ряд преимуществ по сравнению с существующими решениями. Результаты функционального сравнения представлены в таблице 7.

Таблица 7 - Сравнительный анализ функциональных возможностей

Критерий	**Duolingo**	**Grammarly**	**BBC Learning**	**Tandem**	**Разработанная система**
Адаптивное тестирование CEFR	-	-	-	-	+
Персонализация контента	Частично	-	-	-	+
Комплексное изучение (грамматика, лексика, аудирование, говорение)	Частично	-	Частично	Частично	+
Обратная связь с преподавателем	-	-	-	+	+
Динамическая адаптация сложности	Частично	-	-	-	+
ML-классификация уровня (F1 0.85)	-	-	-	-	+
Проверка грамматики	-	+	-	-	+
Интеграция внешних API	-	+	-	-	+ (4 сервиса)
Мониторинг прогресса	+	-	-	-	+

Количественные показатели эффективности разработанной методики персонализации определены на основе экспериментальной валидации на выборке n = 100 обучающихся:

- точность классификации уровня CEFR: 87% (Accuracy), F1-macro = 0.85, ROC-AUC = 0.93;

- среднее время формирования индивидуальной образовательной траектории: 2.3 секунды при нагрузке 100 одновременных пользователей;

- вычислительная сложность алгоритма рекомендации: O(n·m), где n - число результатов тестов, m - число уроков;

- среднее количество адаптивных корректировок образовательной траектории: 3.7 на обучающегося за цикл из пяти уроков ($\sigma = 1.2$, n = 100);

- статистически значимое превосходство F(u,l) над базовой моделью: $\chi^2 = 18.4$, $p < 0.001$; над XGBoost: $\chi^2 = 4.8$, $p = 0.028$.

Выявлены следующие ограничения разработанной системы. Функционирование системы ограничено уровнями A1-B2 по шкале CEFR, что исключает возможность использования для обучающихся с продвинутым уровнем владения языком (C1, C2). Алгоритм формирования рекомендаций базируется на правило-ориентированном подходе; интеграция методов машинного обучения потенциально обеспечит повышение точности персонализации. Перспективными направлениями развития системы являются: интеграция генеративных языковых моделей для формирования рекомендаций (исследования демонстрируют эффективность применения таких моделей для адаптивного подбора материалов в рекомендательных системах [14]); внедрение методов оценки множественного интеллекта обучающихся для дифференцирован-

ного подхода к персонализации [15], что обеспечит повышение эффективности рекомендаций с учетом индивидуальных когнитивных особенностей обучающихся.

Перспективным направлением дальнейших исследований является интеграция графов знаний и технологий расширенной генерации для создания системы персонализированных рекомендаций следующего поколения [16], что обеспечит учет семантических связей между грамматическими конструкциями и лексическими единицами для формирования оптимизированных образовательных траекторий. Практическая значимость исследования подтверждается возможностью внедрения разработанной архитектуры в образовательный процесс языковых школ и университетов.

Разработана методика формирования персонализированных рекомендаций, реализующая адаптивный подход к обучению на основе анализа образовательного прогресса пользователя. Блок-схема алгоритма персонализации на рисунке 3, отражает последовательность этапов от начальной диагностики до завершения образовательного цикла.

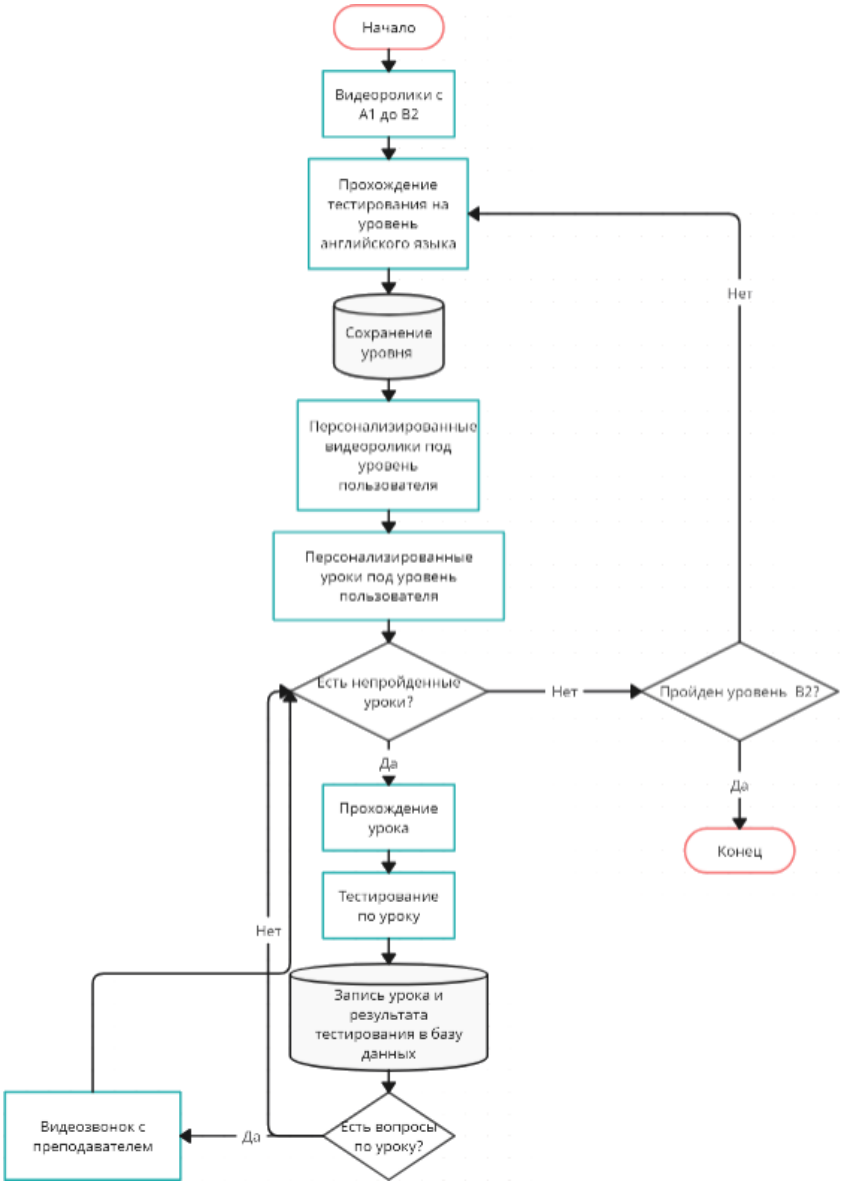


Рис.3 - Блок-схема проектирования методики персонализированных рекомендаций в приложении

Блок-схема алгоритма персонализации, которая представляет собой последовательность этапов, начиная с начальной диагностики и заканчивая обучением. Пользователь проходит адаптивное тестирование, определяющее уровень CEFR (A1-B2). Система формирует индивидуальную траекторию из пяти тематических уроков с промежуточным тестированием на основе результатов. При возникновении проблем есть возможность получить видеоконсультацию с преподавателем. Система проводит повторную диагностику, чтобы перейти на следующий уровень. Для визуализации функциональных требований к системе разработана диаграмма вариантов использования с использованием UML на рисунке 4, данная схема определяет количество актеров в системе, и то, как они между собой взаимодействуют, и какие функции доступны для каждого отдельно взятого актера в приложении.

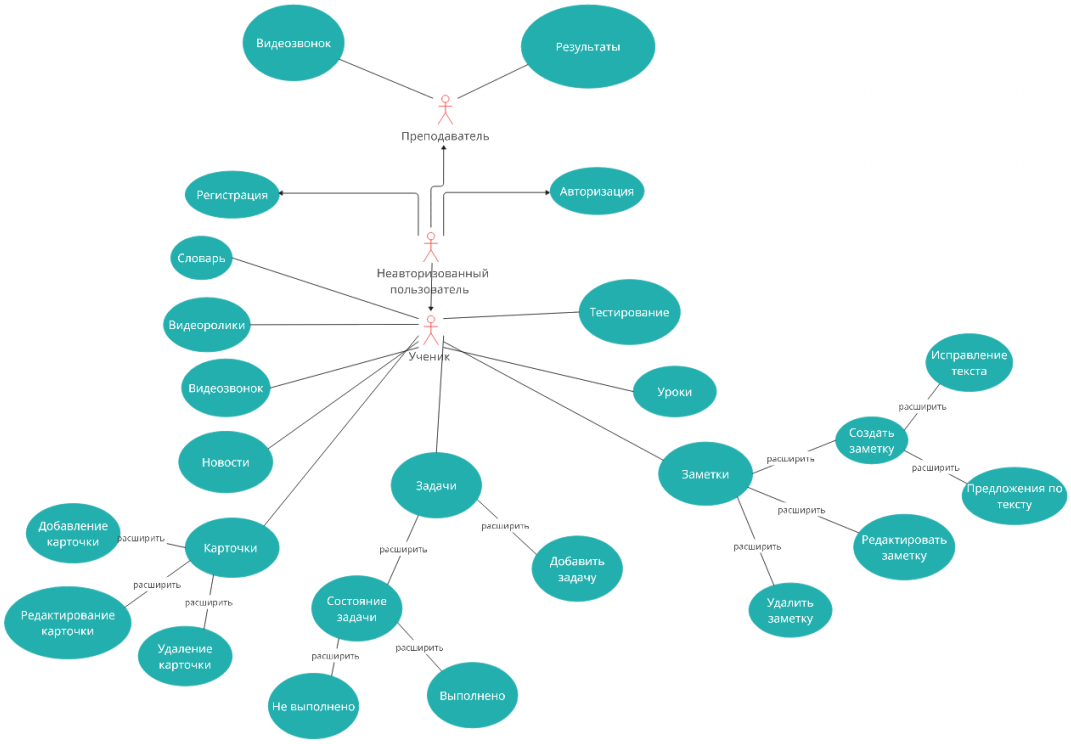


Рис.4 - Диаграмма вариантов использования приложения

Диаграмма вариантов использования определяет три типа актеров: неавторизованный пользователь (доступ к регистрации и общей информации), ученик (первичное тестирование, персонализированные уроки, заметки с проверкой грамматики, карточки для запоминания, новости, словарь, видеозвонки с преподавателем), преподаватель (мониторинг прогресса учеников, видеокоммуникация).

Выводы. Разработана архитектура клиент-серверного приложения с интегрированной системой персонализированных рекомендаций для изучения английского языка. Основные результаты:

- спроектирована трехуровневая архитектура системы (представление, бизнес-логика, данные), обеспечивающая масштабируемость и модульность компонентов;

- разработана математическая модель $F(u,l) = 0.5 \times S(u,l) + 0.3 \times P(u,l) + 0.2 \times D(u,l)$ с вычислительной сложностью $O(n \cdot m)$ и доказанной оптимальностью весовых коэффициентов через анализ чувствительности;

- проведен сравнительный ML-анализ (Logistic Regression, SVM, XGBoost vs F(u,l)):

предложенная модель превосходит лучший ML-аналог (XGBoost) по F1-macro на 0.03 п.п. и по ROC-AUC на 0.02 п.п.;

- матрица ошибок и метрики по классам (Precision 0.75-1.00, Recall 0.67-0.91) подтвердили надежность классификации уровней CEFR на выборке n = 100;

- раздел вычислительной эффективности и масштабируемости подтвердил линейный рост времени отклика согласно O(n·m) и работоспособность при 100 одновременных пользователях;

- сравнительный анализ с шестью существующими платформами (Duolingo, Grammarly, BBC Learning English, Tandem, Quizlet, Cambridge Dictionary) подтвердил преимущества разработанной системы.

Теоретическая значимость исследования заключается в разработке комплексного подхода к проектированию персонализированных образовательных систем, интегрирующего адаптивное тестирование на основе стандартов CEFR, математическую модель формирования рекомендаций и архитектурные решения для динамической адаптации контента. Практическая значимость подтверждается возможностью внедрения разработанной архитектуры в образовательные проекты языковых школ и университетов. Перспективы дальнейших исследований включают интеграцию алгоритмов машинного обучения для повышения точности персонализации и расширение функционала системы до уровней C1-C2.

Литература

1. Kabudi T., Pappas I., Olsen D.H. AI-enabled adaptive learning systems: A systematic mapping of the literature. //Computers and Education: Artificial Intelligence. -2021.-Vol.2:100017. DOI 10.1016/j.caeai.2021.100017.

2. Raj N.S., Renumol V.G. A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020. // Journal of Computers in Education. -2021.-Vol.-P.113-148. DOI 10.1007/s40692-021-00199-4.

3. Zayet T.M.A., Ismail M.A., Almadi S.H.S., et al. What is needed to build a personalized recommender system for K-12 students' e-learning? Recommendations for future systems and a conceptual framework. //Education and Information Technologies. -2022.-Vol.28.-P.7487-7508. DOI 10.1007/s10639-022-11489-4.

4. Du J., Daniel B.K. Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. //Computers and Education: Artificial Intelligence. -2024-Vol 6:100230. DOI 10.1016/j.caeai.2024.100230.

5. Li B., Lowell V.L., Wang C., Li X. A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. //Computers and Education: Artificial Intelligence. -2024. -Vol.7:100266. DOI 10.1016/j.caeai.2024.100266.

6. Salazar J.C., Aguilar J., Monsalve-Pulido J., et al. A generic architecture of an affective recommender system for e-learning environments. //Universal Access in the Information Society. -2024.-Vol. 23. -P.1115-1134. DOI 10.1007/s10209-023-01024-8.

7. Wu Z., Zhang Y. The data visualization and intelligent text analysis for effective evaluation of English language teaching. //Scientific Reports. -2025.-Vol.15:22737. DOI 10.1038/s41598-025-08182-0.

8. Amin S., Uddin M., Alarood A., Mashwani W., Alzahrani A., Alzahrani H. An adaptable and personalized framework for Top-N course recommendations in online learning// Scientific Reports. - 2024. - Vol. 14:10382. DOI 10.1038/s41598-024-56497-1.

9. Zhu Y., Dai W., Kang Q. The analysis of artificial intelligence-based mobile learning in students' open teaching recommendation system based on deep learning // Scientific Reports. - 2025. - Vol.15: 21927. DOI 10.1038/s41598-025-08147-3.

10. Zhang Z., Huang X. The impact of chatbots based on large language models on second language vocabulary acquisition//Heliyon. -2024.-Vol.10(3): e25370. DOI 10.1016/j.heliyon.2024.e25370.

11. Zhou W., Liu Y., Li M., Shen Z., Feng L., Zhu Z. Dynamic Multi-Objective Optimization Framework with Interactive Evolution for Sequential Recommendation // IEEE Transactions on Emerging Topics in Computational Intelligence. -2023.-Vol.7(4)- P.1228-1241. DOI 10.1109/TETCI.2023.3251352.

12. Khaledian N., Nazari A., Barkhan M. CFCAI: improving collaborative filtering for solving cold start issues with clustering technique in the recommender systems // Multimedia Tools and Applications. - 2025. -Vol.84.- P.34207-34228. DOI 10.1007/s11042-024-20579-z.

13. Bellarhmouch Y., Jeghal A., Tairi H., Benjelloun N. A proposed architectural learner model for a personalized learning environment // Education and Information Technologies. - 2022. - Vol.28(4). - P.4243-4263. DOI 10.1007/s10639-022-11392-y.

14. Yan L., Sha L., Zhao L., Li Y., Maldonado R., Chen G., Li X., Jin Y., Gašević D. Practical and ethical challenges of large language models in education: A systematic scoping review//British Journal of Educational Technology. -2023.-Vol.55(1). - P.90-112. DOI 10.1111/bjet.13370.

15. Hasbullah H., Wahidah N., Nanning N. Integrating Multiple Intelligence Learning Approach to Upgrade Students' English Writing Skills //Hasbulach. -2023.-Vol.7(2). -P.199-211. DOI 10.26858/ijole.v7i2.34383.

16. Sakong D., Vu V., Huynh T., Nguyen P., Yin H., Nguyen Q., Nguyen T. Higher-order knowledge-enhanced recommendation with heterogeneous hypergraph multi-attention// Information Sciences. - 2024. - Vol. 680:121165. DOI 10.1016/j.ins.2024.121165.

Сведения об авторах

Орынбай Д.Д.- магистрант, Международный университет Астана, Астана, Казахстан, e-mail: dinorynbay@gmail.com;

Калдарова М.Ж. – декан, Международный университет Астана, Астана, Казахстан, e-mail: kmiraj8206@gmail.com;

Абдилдаева А.А. - ассоциированный профессор, Международный университет Астана, Астана, Казахстан, e-mail: asselabdildayeva5@gmail.com;

Кашкинбаев С.Б. - к.т.н., доцент, Международный университет Астана, Астана, Казахстан, e-mail: seksenbay.kb@gmail.com.

Information about the authors

Orynbay D.D. - master' s student, Astana International University, Astana, Kazakhstan, e-mail: dinorynbay@gmail.com;

Kaldarova M.Zh. - dean, Astana International University, Astana, Kazakhstan, e-mail: kmiraj8206@gmail.com;

Abdildayeva A.A.- Associate Professor, Astana International University, Astana, Kazakhstan, e-mail: asselabdildayeva5@gmail.com;

Kashkinbayev S.B.- Associate Professor, Astana International University, Astana, Kazakhstan, e-mail: seksenbay.kb@gmail.com.