

УДК 004.42, 004.75, 004.421

Научная статья

DOI: 10.35330/1991-6639-2024-26-3-21-31

EDN: HHGDFE

Разработка сервиса коротких мультиязычных переводов на основе распределенной системы обработки аннотированной информации

К. Ф. Край, А. З. Энес

Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. И. Арманд, 37-а

Аннотация. Разработка сервиса коротких двуязычных переводов представляет собой актуальную задачу создания систем автоматического перевода текстов для средних и малых языков, для которых не существуют аннотированные базы данных, необходимые для обучения. В рамках данной работы реализуется система для автоматического перевода коротких текстов с одного языка на другой, особенностью которой является использование распределенной архитектуры для сбора, обработки, хранения и использования информации, предоставляемой пользователями системы на основе оригинальной модели мотивации, что позволяет значительно увеличить скорость создания обучающей выборки и качество обработки переводов.

Ключевые слова: машинное обучение, распределенная система, обработка естественного языка, сбор данных, веб-сервис, краудсорсинг, машинный перевод, искусственный интеллект, аннотирование информации, анализ текстов, перевод, текст, обработка информации

Поступила 22.04.2024, одобрена после рецензирования 27.05.2024, принята к публикации 31.05.2024

Для цитирования. Край К. Ф., Энес А. З. Разработка сервиса коротких мультиязычных переводов на основе распределенной системы обработки аннотированной информации // Известия Кабардино-Балкарского научного центра РАН. 2024. Т. 26. № 3. С. 21–31. DOI: 10.35330/1991-6639-2024-26-3-21-31

MSC: 68M14, 68T50

Original article

Development of a short bilingual translation service based on a distributed annotated information processing system

K.F. Krai, A.Z. Enes

Institute of Computer Science and Problems of Regional Management –
branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
360000, Russia, Nalchik, 37-a I. Armand street

Abstract. The development of a service for short bilingual translations is an urgent task of the creation of automatic text translation systems for medium and small languages, for which there are no annotated databases necessary for training. As part of this work, a system is being implemented for automatically translating short texts from one language to another, the feature of which is the use of a distributed architecture for collecting, processing, storing and using information provided by system users based on the original motivation model, which allows you to significantly increase the speed of creating a training sample and the quality of translation processing.

Keywords: machine learning, distributed system, natural language processing, data collection, web service, crowdsourcing, machine translation, artificial intelligence, information annotation, text analysis, translation, text, information processing

Submitted 22.04.2024,

approved after reviewing 27.05.2024,

accepted for publication 31.05.2024

For citation. Krai K.F., Enes A.Z. Development of a short bilingual translation service based on a distributed annotated information processing system. *News of the Kabardino-Balkarian Scientific Center of RAS*. 2024. Vol. 26. No. 3. Pp. 21–31. DOI: 10.35330/1991-6639-2024-26-3-21-31

ВВЕДЕНИЕ

В эпоху активного развития информационных технологий и ускоренного процесса глобализации вопросы разработки и реализации инновационных цифровых инструментов, направленных на сохранение и поддержку языкового и культурного многообразия, приобретают особую актуальность. В контексте решения этой задачи на передний план выходят исследования, связанные с созданием сервисов перевода для языков малых народов, что имеет значение не только для сохранения уникального лингвистического наследия, но и для обеспечения равных условий интеграции носителей данных языков в мировой информационный дискурс.

В сегодняшней динамичной сфере технологий с искусственным интеллектом (ИИ) сбор и анализ данных занимают ключевую позицию. Эти процессы не только являются основой для обучения и развития систем ИИ, но и необходимы для точной и эффективной работы таких систем, в том числе в области автоматизированного перевода. Тщательно собранные и проанализированные данные позволяют ИИ-системам адаптироваться к различным культурным и контекстуальным особенностям, что крайне важно для удовлетворения специфичных требований пользователей и улучшения взаимодействия человека и машины. Кроме того, сбор данных и их аналитика имеют значительное влияние на общественные аспекты использования ИИ, как подчеркивают эксперты, обширные и точные данные способствуют повышению производительности, качества обслуживания и эффективности автоматизации [1].

Таким образом, разработка методологии создания высокоэффективных алгоритмов машинного обучения, адекватно воспроизводящих лингвистические структуры и идиоматику языков малых народов, является **актуальной**, предполагает выполнение ряда работ, включающих в себя: сбор и аналитическую обработку необходимых лингвистических данных, разработку специализированных моделей искусственного интеллекта, а также тестирование и адаптацию этих моделей в соответствии с лингвокультурными и контекстуальными характеристиками данных языков.

1. ПРОБЛЕМА РАЗРАБОТКИ АННОТИРОВАННЫХ БАЗ ДАННЫХ ДЛЯ ОБУЧЕНИЯ СИСТЕМ МАШИННОГО ПЕРЕВОДА МАЛЫХ И СРЕДНИХ ЯЗЫКОВ

Термин «языки с ограниченными ресурсами», введенный в [2], описывает языки, которые могут обладать некоторыми или всеми из следующих характеристик: отсутствие уникальной системы письма или стабильной орфографии, ограниченное присутствие в интернете, нехватка лингвистической экспертизы, дефицит электронных ресурсов для обработки естественного языка, двуязычные электронные словари, словари произношения, транскрибированные данные речи и т.д. Нехватка ресурсов требует разработки новаторских методологий сбора данных или моделей, позволяющих делиться информацией

между языками [2]. Малые и средние языки относятся к «языкам с ограниченными ресурсами», которые сталкиваются с ограничениями в доступе к ресурсам для современных технологий обработки языка.

Данная статья посвящена разработке сервиса коротких двуязычных переводов на основе распределенной системы обработки аннотированной информации для языков малых народов. Проблема недостатка данных для обучения алгоритмов машинного перевода является значительным препятствием, поскольку большинство существующих технологий, включая нейронные сети и методы машинного обучения, требуют значительного объема текстовых данных, разметка которых, как правило, выполняется вручную большими коллективами профессионально подготовленных носителей обоих языков. Создание обучающих выборок для таких языков представляет собой длительный и дорогостоящий процесс, организация которого требует существенных трудовых и материальных затрат. Это приводит к тому, что носители языков с ограниченным количеством доступных материалов и недостаточной представленностью в Интернете сталкиваются с трудностями в создании точных и надежных систем перевода. В связи с этим можно утверждать, что важную роль играют оптимизация и упрощение процедур сборки лингвистических данных для этих языков, что необходимо для эффективного обучения систем искусственного интеллекта.

Объектом данного исследования является процесс распределенного создания и аннотирования базы данных для обучения системы машинного перевода для малых и средних языков.

Предмет исследования – возможность его организации на основе реализации алгоритмов мотивации носителей малых и средних языков к удаленному созданию записей в таких базах данных и голосованию в интересах ранжирования вариантов перевода по релевантности и качеству.

Цель данного исследования – разработка сервиса коротких мультязычных переводов, позволяющего переводить текст с любого языка на любой другой язык за счет взаимодействия пользователей.

В данной работе рассматриваются следующие **задачи**, которые необходимо решить для ее достижения:

- разработка и реализация алгоритма обработки запросов сервиса коротких мультязычных переводов;
- разработка и реализация алгоритма мотивации процесса сбора данных и алгоритма голосования пользователей сервиса коротких мультязычных переводов

2. ОБЩАЯ АРХИТЕКТУРА СЕРВИСА КОРОТКИХ МУЛЬТЯЗЫЧНЫХ ПЕРЕВОДОВ

Программа представляет собой веб-сервер, пользователи которого могут оставлять запрос на перевод отдельных слов, фраз и предложений в аудио- и/или текстовой форме. Программа пересылает запрос всем пользователям, владеющим указанным языком, и предлагает ответить на запрос, проверяет наличие готового перевода и выводит список всех вариантов с ранжированием по популярности. Кроме того, в программе реализованы функционал оценки перевода и система мотивации пользователей.

Несмотря на то, что инструменты машинного перевода достигли высокого уровня развития и постоянно улучшаются, они по-прежнему не способны обеспечить абсолютно точный перевод произвольных текстов, что становится очевидным при их практическом применении [3]. Тем не менее вовлечение большого числа людей в процесс перевода не просто осуществимо, но и в некоторых случаях может обеспечить уровень качества, который не уступает более дорогостоящим и требующим значительного времени подходам [4].

В последние десятилетия в дисциплине переводоведения наблюдается заметное усиление интереса к принципам коллективного перевода с привлечением широкой общественности (краудсорсинг) и онлайн-сотрудничеству в процессе перевода текстов. Данное направление, отличающееся высокой динамикой и изменчивостью, привлекает все большее число научных сотрудников. Широкое внедрение этих различных методов перевода обещает реализовать новые перспективные алгоритмы перевода, изменить фундаментальные принципы дисциплины, способствует прогрессу в сфере разработки технологий и оказывает влияние на общественное восприятие процесса перевода [5].

Разрабатываемая нами распределенная система обработки информации [6] предназначена для применения в качестве веб-сервиса для сбора данных, нацеленных на обучение систем понимания естественного языка, а также данная система может использоваться для обеспечения обработки аудиозапросов в научных и образовательных организациях [7].

Реализация распределенного подхода к сбору данных для обучения систем пониманию естественного языка влечет за собой ряд ключевых преимуществ: расширение и диверсификацию набора данных, ускорение процесса их сбора и обработки, а также снижение затрат по сравнению с традиционными методами сбора данных, требующими значительного участия профессионалов высокой квалификации.

В рамках реализации поставленной задачи была разработана алгоритмическая модель, предназначенная для распределения задач по переводу текстовых данных с любого языка на любой другой язык.

В разрабатываемом нами распределенном сервисе коротких двуязычных переводов [6, 7] подчеркивается важный аспект использования краудсорсинга в переводе. Краудсорсинговые переводы открывают перед участниками уникальные возможности для обмена информацией, саморазвития и поддержки многоязычного контента в интернете. Это привлекает не только профессиональных переводчиков, но и энтузиастов, стремящихся внести свой вклад в распространение знаний [8].

Сервис представляет собой платформу агрегирования больших объемов лингвистической информации путем привлечения широкого круга участников (волонтеров или оплачиваемых исполнителей) из различных регионов и культурных контекстов. Это позволяет создавать многоязычные, культурно и контекстуально разнообразные датасеты, необходимые для обучения систем искусственного интеллекта и машинного обучения, способных эффективно воспринимать и анализировать человеческий язык в его естественной форме.

Семантические и контекстные особенности, культурно-специфические элементы, тональность текста и его стилистические характеристики значительно увеличивают требования к объему и качеству данных для обучения, а также к степени сложности алгоритмов, обеспечивающих процесс перевода. Семантический и контекстный анализ требует от системы не только распознавания и перевода отдельных слов или фраз, но и глубокого понимания предложений, абзацев и всего текста в целом [9].

Решение этих задач возможно с применением сложных алгоритмов машинного обучения, включая глубокое обучение, которые могут обрабатывать большие объемы аннотированных данных и учиться на примерах. Включение механизмов искусственного интеллекта, способных к самообучению и адаптации, позволяет системе на основе анализа предыдущих ошибок и успехов постоянно совершенствовать процесс перевода.

Разрабатываемая нами распределенная система рассчитана на взаимодействие между собой двух категорий пользователей – клиентов и экспертов. На рисунке 1 представлена схема краудсорсинга с обратной связью, где:

- Клиенты (A_1, A_2, A_n) отправляют Запросы для получения (n) вариантов перевода;
- Сервис отправляет Запрос Экспертам ($W_1, W_2, W_3, \dots, W_n$) для дальнейшей обработки;

- Сервис регистрирует полученный от Экспертов перевод, направляет эти сведения для дальнейшей обработки и голосования ($V_1, V_2, V_3 \dots V_n$);
- Сервис определяет и сохраняет итоги выполнения исходных задач, а также генерирует обратную связь в виде оценки качества выполнения задачи и отправляет вознаграждения в соответствии с количеством голосов за выбранный вариант ответа ($W_n V_n X \%$).

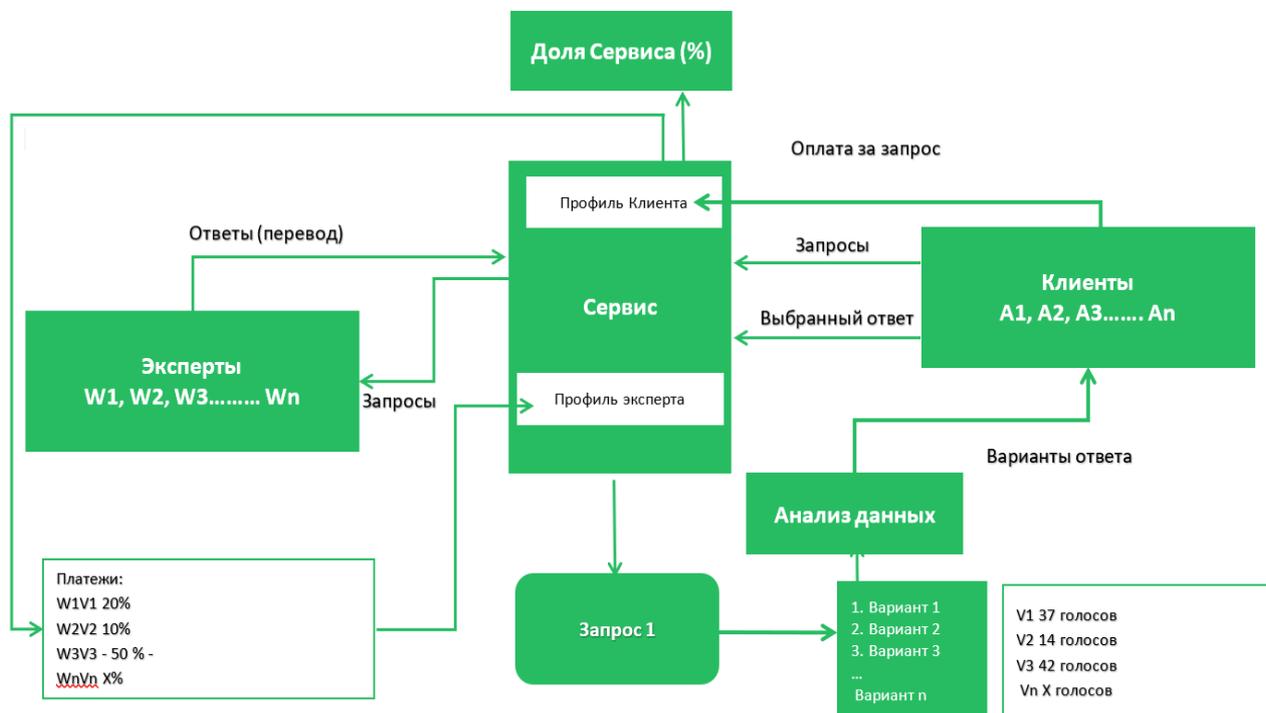


Рис. 1. Схема распределенной системы обработки аннотированной информации

Fig. 1. Scheme of a distributed system for processing annotated information

3. АЛГОРИТМ ОБРАБОТКИ ЗАПРОСОВ СЕРВИСА КОРОТКИХ МУЛЬТИЯЗЫЧНЫХ ПЕРЕВОДОВ

Алгоритм обработки запросов в системе представляет собой комплексный, многофазовый процесс, который включает ряд операций, связанных с созданием, обработкой и управлением данными, выполнением операций с базой данных, а также реализацией механизма отправки электронных сообщений. Этот процесс опирается на использование различных программных классов и методов.

В начальной фазе процесса происходит создание объекта «Запрос». Этому объекту присваиваются атрибуты, значения которых формируются на основе параметров запрашиваемых операций и дополнительной информации. Это подготавливает информацию к более глубокой обработке на последующих этапах.

Одним из ключевых аспектов алгоритма является сохранение созданных объектов «Запрос» в реляционной базе данных. Этот шаг позволяет систематизировать и агрегировать информацию, связанную с запросами пользователей, что способствует более эффективному управлению данными.

Процесс также включает механизмы для контроля передвижения средств, а также систему создания и отправки электронных сообщений. В рамках последней происходят формирование содержимого письма, идентификация получателей и сам процесс отправки сообщений, что обеспечивает коммуникацию между пользователями и системой.

Графическое представление данного алгоритма в виде блок-схемы, иллюстрирующей последовательность операций в системе аннотированной обработки информации, демонстрируется на рисунке 2.

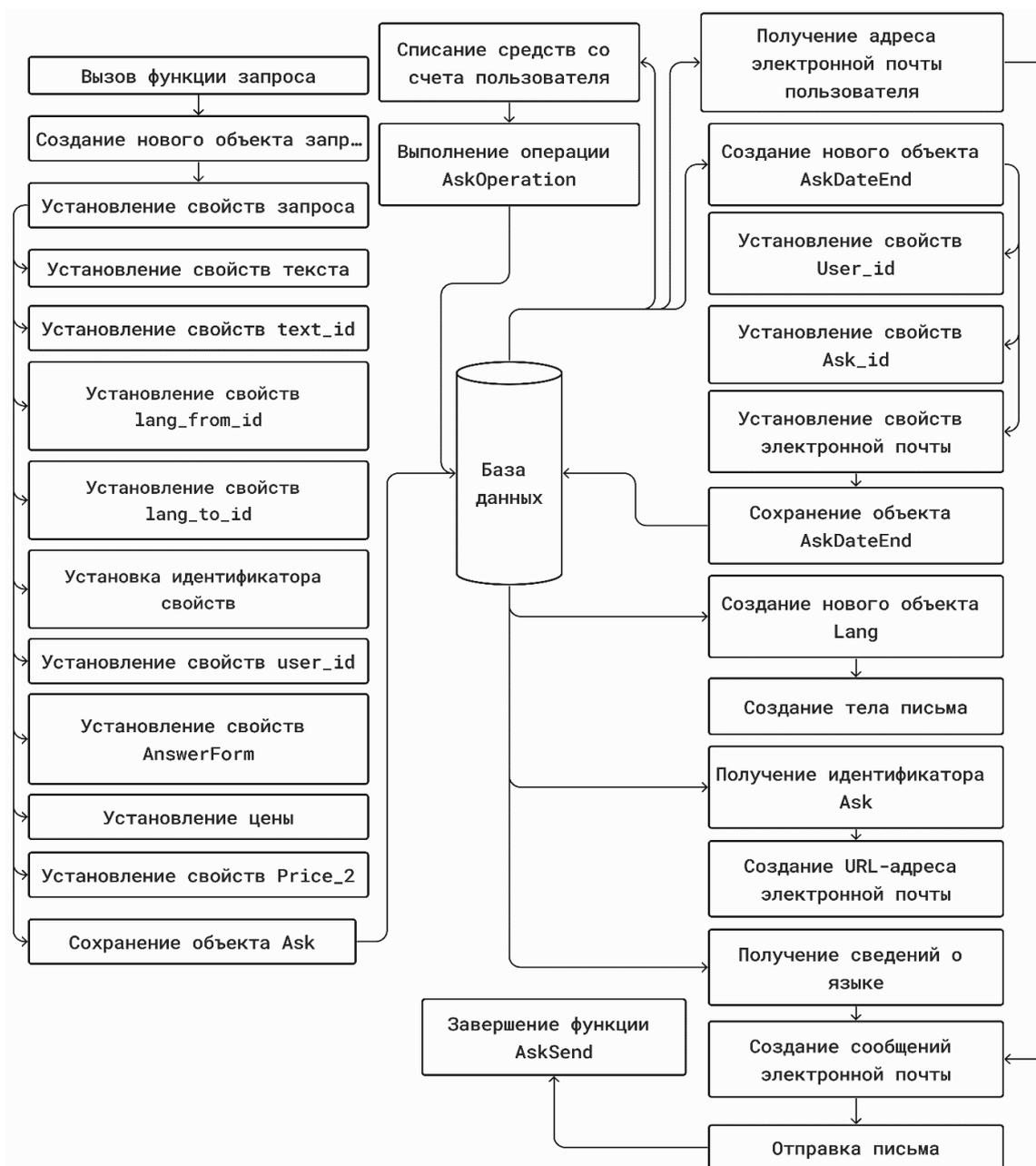


Рис. 2. Алгоритм обработки запросов распределенной системы

Fig. 2. Algorithm for processing requests of a distributed system

4. РАЗРАБОТКА И РЕАЛИЗАЦИЯ АЛГОРИТМА МОТИВАЦИИ ПРОЦЕССА СБОРА ДАННЫХ И АЛГОРИТМА ГОЛОСОВАНИЯ ПОЛЬЗОВАТЕЛЕЙ СЕРВИСА КОРОТКИХ МУЛЬТИЯЗЫЧНЫХ ПЕРЕВОДОВ

Основной принцип мотивационной модели заключается в том, что каждый перевод, предоставленный экспертом на запрос клиента, сохраняется в системе и становится активом этого эксперта. Таким образом, каждый раз, когда другие клиенты будут обращаться за тем же самым переводом и выбирать данный вариант ответа для оплаты, эксперт, владеющий этим активом, будет получать соответствующее вознаграждение.

Таким образом, система мотивирует экспертов предоставлять качественные переводы, так как они будут получать вознаграждение каждый раз, когда их работа используется другими клиентами. Это способствует повышению качества сбора данных и обработки информации в распределенной системе.

Алгоритм обрабатывает пользовательские запросы с применением механизмов проверки уникальности вопросов и ответов, вычислением наград и обновлением базы данных. В начальной фазе исполнения кода осуществляется верификация языка взаимодействия, а также проверка на предмет прежнего поступления идентичных запросов или ответов от пользователя. Далее следует этап присуждения вознаграждений, зависящий от множества критериев. В случае выявления новизны ответа он инкорпорируется в базу данных. Процедуры обновления данных в базе включают корректировки записей о собственниках и значениях цен, а также инициацию рассылки уведомлений пользователям по электронной почте. В коде присутствуют элементы повторения операций, которые предоставляют потенциальные возможности для повышения эффективности обработки ошибок и усиления мер по защите конфиденциальности пользовательских данных. На рисунке 3 демонстрируется блок-схема алгоритма, которая иллюстрирует процесс обработки ответов в рамках распределенной системы, показывая последовательные этапы аннотированной обработки информации.

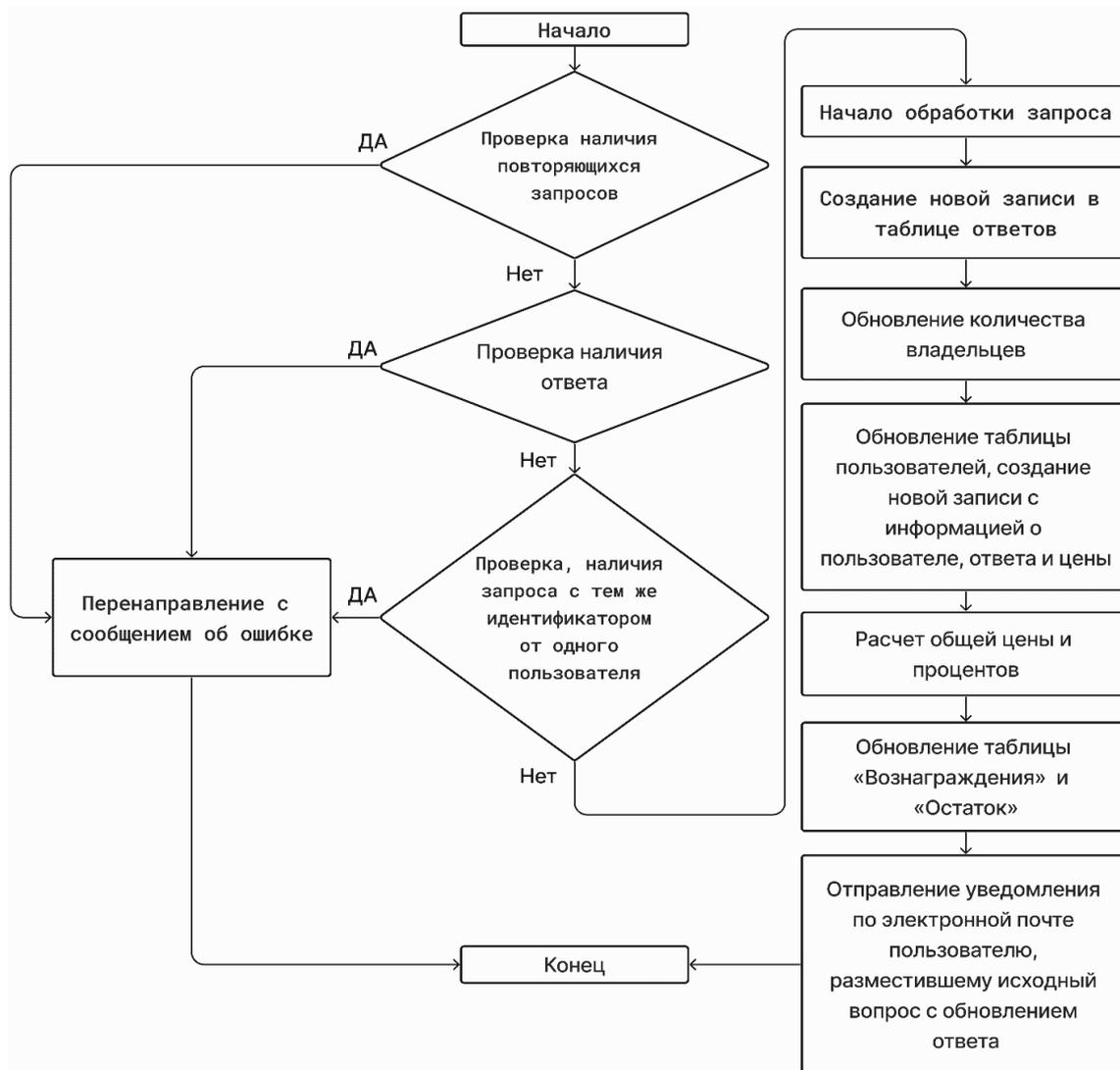


Рис. 3. Алгоритм обработки ответов распределенной системы

Fig. 3. Algorithm for processing responses from a distributed system

На начальном этапе реализуются процедуры идентификации и блокировки повторных запросов, что позволяет исключить избыточную нагрузку на инфраструктуру системы и предотвратить дублирование данных в базе. Механизмы проверки включают:

- проверку на повторение вопроса – используется алгоритм сравнения строк для идентификации одинаковых запросов от одного пользователя, где эффективность данного процесса может быть увеличена за счет применения методов оптимизации текстовых данных, таких как стемминг и лемматизация;

- проверку ответа на тот же вопрос – обеспечивает исключение множественных ответов от одного пользователя по одному запросу, что предотвращает дублирование информации и улучшает качество данных;

- проверку запроса на тот же вопрос: предотвращает создание идентичных запросов, что поддерживает структурированность и чистоту базы данных.

На втором этапе проводятся обработка и хранение данных, где последующая обработка запроса включает в себя несколько ключевых операций, направленных на обработку и анализ представленной информации:

- расчет вознаграждения – применение математической модели для расчета вознаграждения, основываясь на процентных значениях;

- создание записей ответа – размещение данных ответа в специализированной таблице `ans_repeats` способствует упорядочиванию информации и облегчает последующий доступ к ней;

- обновление количества владельцев – увеличение числа владельцев позволяет отслеживать популярность и актуальность конкретных запросов;

- расчет общей стоимости и процента – представляет собой аналитический этап, на котором осуществляется совокупный анализ данных, направленный на определение статистически значимых параметров;

- обновление вознаграждений – на этом этапе происходит корректировка размеров вознаграждений в соответствии с полученными данными, что является ключевым моментом в мотивационной части системы.

Рисунки 4 и 5 демонстрируют результаты работы алгоритмов обработки информации разработанного сервиса коротких мультязычных переводов.

Ответы

№	Совладельцы	Варианты	Дата получения	Ориентировочный доход (Rub)	Текущая доступная доля, %	Текущий доступный бонус, %	Total Income		
	Карина	НОбЭ УЭФ1КЪЫМ	2024-05-14 00:23:21	1	3	15	35	<div style="background-color: #28a745; color: white; padding: 5px; text-align: center;">СОГЛАСИТЬСЯ & СТАТЬ СОВЛАДЕЛЬЦЕМ</div> <div style="background-color: #dc3545; color: white; padding: 5px; text-align: center; margin-top: 5px;">Пожаловаться</div>	
	Sati	УЭФ1КЪЫМ НОбЭ	2024-03-16 13:13:49	3	3	15	5	12	<div style="background-color: #28a745; color: white; padding: 5px; text-align: center;">СОГЛАСИТЬСЯ & СТАТЬ СОВЛАДЕЛЬЦЕМ</div> <div style="background-color: #dc3545; color: white; padding: 5px; text-align: center; margin-top: 5px;">Пожаловаться</div>

ОТВЕТИТЬ

Рис. 4. Интерфейс разработанного сервиса с демонстрацией возможности наличия нескольких вариантов ответа

Fig. 4. Interface of the developed service with demonstration of capabilities to have multiple answers

Вопросы		Ответы			
МахуэфI тхъэм къывит(Русский → Адыгэбзэ - Кабардинский) Answers					
	Дата получения	Доля, %	Доля (руб)	Бонус, %	Бонус (Rub)
Arsen	2023-11-28 06:36:05	45.5	13.65	0	0
Olga	2023-12-01 08:24:27	19.5	5.85	0	0

Рис. 5. Интерфейс разработанного сервиса с демонстрацией результатов расчета долей владения конкретным ответом

Fig. 5. Interface of the developed service with demonstration of results calculating ownership shares of a specific answer

Заключительный этап алгоритма включает в себя процесс уведомления пользователя о завершении обработки его запроса, что является финальной стадией в цикле обработки данных и обеспечивает необходимую обратную связь.

ЗАКЛЮЧЕНИЕ

Данный информационно-технологический сервис представляет собой результат успешного применения современных методологий и технологий в области обработки данных и интерактивного взаимодействия с пользователями. Реализация такого комплексного подхода способствует значительному улучшению качества и скорости обработки пользовательских запросов, а также поддерживает динамическое совершенствование функционала сервиса. В процессе исследования были разработаны и внедрены алгоритмы обработки запросов, мотивации пользователей и голосования, что сыграло ключевую роль в обеспечении эффективной работы сервиса коротких мультязычных переводов и повышении качества переводов. Принципы мотивации пользователей, изложенные в работе, играют важную роль в повышении стандартов предоставления информации. Стимулирование экспертов для высококачественной деятельности и обеспечение их интереса к непрерывному развитию и совершенствованию своих профессиональных навыков приводят к положительной динамике во всех аспектах информационного взаимодействия. Это дополнительно стимулирует повышение стоимости и актуальности информационного сервиса, что является основой для дальнейших технологических инноваций и развития систем, использующих искусственный интеллект. Интеграция механизмов предотвращения дублирования информации и оптимизация процессов работы с данными способствуют сокращению используемых ресурсов. Это особенно важно в условиях увеличения объемов данных и возрастающих запросов к пользовательской эффективности и удобству сервисов.

REFERENCES

1. Schultz T., Kirchhoff K. Multilingual speech processing. 2006. ISBN: 9780080457628
2. Bisani M., Ney H. Joint-sequence models for grapheme-to-phoneme conversion. *Speech Communication*. 2008. Vol. 50. No. 5. Pp. 434–451. DOI: 10.1016/j.specom.2008.01.002.hal-00499203

3. Kartseva E.Yu., Margaryan T.D., Gurova G.G. *Razvitie mashinnogo perevoda i ego mesto v professional'noy mezhkul'turnoy kommunikacii* [Development of machine translation and its place in professional intercultural communication]. *RUDN journal of language studies, semiotics and semantics*. 2016. No. 3. Pp. 155–164. EDN: WHBQCV. (In Russian)

Карцева Е. Ю., Маргарян Т. Д., Гурова Г. Г. Развитие машинного перевода и его место в профессиональной межкультурной коммуникации // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2016. № 3. С. 155–164. EDN: WHBQCV

4. Corney J., Lynn A., Torres-Sanchez C. et al. Towards crowdsourcing translation tasks in library cataloguing, a pilot study. *4th IEEE International conference on digital ecosystems and technologies*. Dubai. United Arab Emirates. 2010. Pp. 572–577. DOI: 10.1109/DEST.2010.5610593

5. Jiménez-Crespo M.A. Crowdsourcing and online collaborative translations. 2017. Amsterdam/Philadelphia: John Benjamins Publishing Company. 320 p. ISBN: 1932–8036/2017BKR0009

6. Certificate of state registration of a computer program No. 2024610266 Russian Federation. *Programma dlya sozdaniya i podderzhki infrastruktury mashinnogo obucheniya mul'tiagentnoy nejrokognitivnoy sistemy prinjatiya resheniy ponimaniyu vyskazyvaniy* [Program for creating and supporting machine learning infrastructure for a multi-agent neurocognitive decision-making system for understanding statements]. No. 2023688479: application. 12/19/2023: publ. 01/09/2024 / A.Z. Enes, Z.V. Nagoev, K.F. Krai. Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. EDN: IRFXLN. (In Russian)

Свидетельство о государственной регистрации программы для ЭВМ № 2024610266 РФ. Программа для создания и поддержки инфраструктуры машинного обучения мультиагентной нейрокогнитивной системы принятия решений пониманию высказываний: № 2023688479, заявл. 19.12.2023: опублик. 09.01.2024 / А. З. Энес, З. В. Нагоев, К. Ф. Край. Кабардино-Балкарский научный центр РАН. EDN: IRFXLN

7. Certificate of state registration of a computer program No. 2024610267 Russian Federation. *Programma dlya raspredelennoy platformy audio perevodov* [Program for a distributed audio translation platform]: No. 2023688483: application. 12/19/2023: publ. 01/09/2024 / A.Z. Enes, Z.V. Nagoev, K.F. Krai. Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences. EDN: GGOBJH. (In Russian)

Свидетельство о государственной регистрации программы для ЭВМ № 2024610267 РФ. Программа для распределенной платформы аудио переводов: № 2023688483 заявл. 19.12.2023, опублик. 09.01.2024 / А. З. Энес, З. В. Нагоев, К. Ф. Край; Кабардино-Балкарский научный центр РАН. EDN: GGOBJH

8. McDonough D.J. Analyzing the crowdsourcing model and its impact on public perceptions of translation. *The Translator*. 2012. No. 18(2). Pp. 167–191. DOI: 10.1080/13556509.2012.10799507

9. Yasir M., Akbar Kh., Bashir M. The Impact of artificial intelligence on language translation: A Review. *IEEE Access*. 2024. No. 12. Pp. 25553–25579. DOI: 10.1109/ACCESS.2024.3366802

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Финансирование. Исследование проведено без спонсорской поддержки.

Funding. The study was performed without external funding.

Информация об авторах

Край Карина Фаезовна, мл. науч. сотр. отдела «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН; 360004, Россия, г. Нальчик, ул. И. Арманд, 37-а;

kraykarina@mail.ru, ORCID: <https://orcid.org/0000-0002-6927-7361>, SPIN-код: 5967-0267

Энес Ахмед Зюльфикар, мл. науч. сотр. отдела «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН; 360004, Россия, г. Нальчик, ул. И. Арманд, 37-а;

ahmedenes@mail.ru, ORCID: <https://orcid.org/0000-0003-3633-4910>, SPIN-код: 3643-1808

Information about the authors

Karina F. Krai, Junior Researcher of the Laboratory of Computational Linguistics, Institute of Computer Science and Problems of Regional Management – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

360000, Russia, Nalchik, 37-a I. Armand street;

kraykarina@mail.ru, ORCID: <https://orcid.org/0000-0002-6927-7361>, SPIN-code: 5967-0267

Akhmed Z. Enes, Junior Researcher of the Laboratory of Computational Linguistics, Institute of Computer Science and Problems of Regional Management – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

360000, Russia, Nalchik, 37-a I. Armand street;

ahmedenes@mail.ru, ORCID: <https://orcid.org/0000-0003-3633-4910>, SPIN-code: 3643-1808