

УДК 004.738.5

РАЗРАБОТКА МНОГОФУНКЦИОНАЛЬНОЙ ПЛАТФОРМЫ ДЛЯ АГРЕГИРОВАНИЯ НОВОСТНЫХ СТАТЕЙ

Кольева Н.С., Хасимьянова И.В., Кольев И.Н., Гонцова А.О.

ФГБОУ ВО «Уральский государственный экономический университет»,
Екатеринбург, e-mail: kolyeva_ns@usue.ru.ru

В статье рассматривается разработка многофункциональной платформы для агрегирования, фильтрации и анализа новостных статей из различных источников. Цель исследования – создание системы, обеспечивающей быстрый доступ к актуальной и проверенной информации, автоматическую классификацию контента и персонализацию для пользователей. В качестве материалов и методов исследования использованы современные веб- и мобильные технологии, включая React Native для клиентской части и NewsAPI для серверной части. Авторы провели SWOT-анализ существующих новостных агрегаторов, таких как Google News и Яндекс.Новости, выявив их преимущества и недостатки. В результате разработки система обеспечила автоматизированный сбор, обработку и визуализацию новостных данных, а также внедрила механизмы проверки достоверности информации и защиты персональных данных. Проведенное исследование демонстрирует, что предложенная платформа способна эффективно решать проблемы информационного шума, дублирования контента и отсутствия персонализации, что делает ее ценным инструментом для оперативного получения и анализа новостей. Важным результатом является создание масштабируемой и безопасной системы, адаптированной под современные требования медиапотребления. Разработанное решение открывает перспективы для дальнейшего совершенствования алгоритмов обработки данных и расширения функциональных возможностей платформы.

Ключевые слова: агрегирование новостей, новостная платформа, обработка данных, фильтрация новостей, персонализация

DEVELOPMENT OF A MULTIFUNCTIONAL PLATFORM FOR AGGREGATION OF NEWS ARTICLES

Koleva N.S., Khasimyanova I.V., Kolev I.N., Gontsova A.O.

Ural State University of Economics, Yekaterinburg, e-mail: kolyeva_ns@usue.ru.ru

The article discusses the development of a multifunctional platform for aggregating, filtering and analyzing news articles from various sources. The purpose of the study is to create a system that provides quick access to relevant and verified information, automatic content classification and personalization for users. Modern web and mobile technologies were used as materials and research methods, including React Native for the client part and NewsAPI for the server part. The authors conducted a SWOT analysis of existing news aggregators, such as Google News and Yandex.News, identifying their advantages and disadvantages. As a result of the development, the system provided automated collection, processing and visualization of news data, and also implemented mechanisms for verifying the reliability of information and protecting personal data. The study demonstrates that the proposed platform can effectively solve the problems of information noise, duplicate content and lack of personalization, which makes it a valuable tool for promptly receiving and analyzing news. An important result is the creation of a scalable and secure system adapted to modern media consumption requirements. The developed solution opens up prospects for further improvement of data processing algorithms and expansion of the platform's functionality.

Keywords: news aggregation, news platform, data processing, news filtering, personalization

Введение

В современном цифровом мире ежедневно публикуются миллионы новостных материалов. Эта информация поступает из множества источников: официальные СМИ, независимые новостные платформы, блоги, социальные сети. Пользователь, желающий получать актуальную информацию по интересующим темам, сталкивается с проблемой информационного шума, дублирующихся материалов и нерелевантного контента.

В связи с этим появляется острая потребность в создании инструментов, которые способны автоматизированно агрегировать, фильтровать, структурировать и

персонализировать новостной поток. Такие системы позволяют не только экономить время пользователя, но и повышать качество принимаемых решений – особенно в профессиональной среде (аналитика, PR, управление рисками, финансы) [1].

Цель исследования – оптимизация процессов сбора, обработки и представления новостных данных для повышения их актуальности и удобства использования.

Материалы и методы исследования

Для разработки многофункциональной платформы для агрегирования новостных статей использованы следующие материалы: источники новостей: официальные

сайты новостных агентств (РИА Новости, TACC, BBC, CNN), новостные агрегаторы (GoogleNews, Yandex.News, News360), RSS-ленты. Для обработки и анализа текстовой информации взаимодействия в API выбраны библиотеки: NewsAPI, BeautifulSoup, NLTK, TensorFlow. Использование всех этих источников обеспечивает широкий охват актуальных новостных данных для последующего анализа и агрегирования на платформе.

В качестве методов исследования выбраны следующие: интервью и анкетирование применялись для определения спецификаций и требований проекта и потребностей пользователей; моделирование систем необходимо для выбора технологий и инструментов для разработки; методы парсинга – для обработки данных, методы машинного обучения и NLP – для классификации и фильтрации новостей; метрики precision, recall применялись для оценки точности агрегирования и скорости обработки новостей.

Результаты исследования и их обсуждение

В настоящее время существуют разные типы агрегаторов: некоторые управляют на основе предпочтений пользователей, например Digg; другие, такие как The Huffington Post или The Daily Beast, сочетают автоматическую агрегацию новостей с созданием собственного оригинального контента, включая обсуждения и «гражданскую» журналистику [2].

Несмотря на популярность таких новостных сервисов, как Google News, Яндекс.Новости, Flipboard и News360, у них имеются ограничения: недостаток расширенной персонализации и возможности

сохранять пользовательские настройки, невозможность вручную формировать подборки, закрытая архитектура без опций расширения и лицензионные ограничения на использование данных.

Обратим внимание на ресурсы, которые систематизируют новостные потоки полностью без участия человека. Такие платформы претендуют на создание цельной и объективной картины событий, позволяя читателю ознакомиться с различными точками зрения по интересующей теме. В этом случае автоматизация процесса агрегации новостей полностью реализуется программным обеспечением. Даже выделение ключевых событий и формирование подборок новостей может осуществляться машинным способом. В таблице представим сильные и слабые стороны популярных новостных сервисов.

В 2025 г., согласно данным Mediascope, телевидение сохраняет значительную аудиторию, с охватом 65 % населения в среднем за день, а охват за месяц достигает 99 % россиян старше 4 лет. Среднее время просмотра ТВ на человека – 3 ч 36 мин в день, при этом у старшего поколения (65+) этот показатель достигает 7 ч 13 мин в день. Интернет также остается важным источником информации, с среднесуточным охватом 84 % россиян старше 12 лет и средним временем использования 4 ч 33 мин на человека в день. При этом молодежь в возрасте 12–24 лет тратит на интернет в среднем 6 ч 49 мин в день, а старшее поколение 65+ – 2 ч 2 мин. Аудитория прессы (газеты и журналы) в 2025 г., согласно данным Mediascope за полгода (сентябрь 2024 – февраль 2025), составляет 56 % населения старше 16 лет, при этом аудитория среднего номера – 27 % россиян [3].

SWOT-анализ существующих агрегаторов

Сервис	Преимущества	Недостатки
Google News	– использует AI для персонализации; – огромный охват источников; – удобный интерфейс	– алгоритмы закрыты; – недостаточная поддержка нишевых и региональных СМИ
Яндекс.Новости	– хороший охват русскоязычных и региональных источников; – быстрая доставка новостей	– слабая персонализация; – меньший охват по сравнению с Google News
Feedly	– гибкая настройка через RSS; – пользователь сам выбирает источники; – поддержка интеграций	– нет автоматической персонализации; – требуется ручная настройка, не подходит новичкам
Flipboard	– привлекательный визуальный стиль; – журнальный формат подачи; – тематическая сортировка	– упор на дизайн, а не на аналитику; – меньше акцента на релевантность и фильтрацию
News360	– глубокая персонализация на основе интересов; – машинное обучение для отбора новостей	– ограниченное количество источников; – менее популярен, меньше обновлений и поддержка

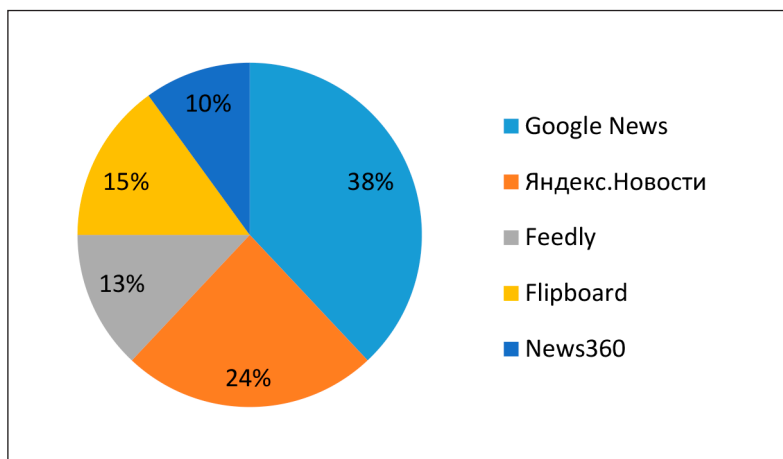


Диаграмма «Популярные агрегаторы новостей»

На основе вышеперечисленных данных разработана диаграмма популярности новостных источников в соответствии с рисунком.

Информационная система агрегации новостных статей реализует следующие функциональные модули:

- Мониторинг новостных потоков – автоматизированный сбор данных из множества источников с целью формирования репрезентативной выборки актуальных событий и информационных сообщений.

- Контент-анализ – процесс тематической категоризации, идентификации ключевых терминов и концептов, а также фильтрации данных на основе заданных критериев (тематических, географических, временных и источниковых).

- Персонализация контента – алгоритмическая адаптация новостной ленты в соответствии с пользовательскими предпочтениями, обеспечивающая повышение релевантности выводимой информации и создание персонализированного информационного пространства.

- Управление пользовательскими данными – механизмы регистрации, аутентификации и настройки профиля, гарантирующие безопасное хранение персональных данных и их использование для оптимизации взаимодействия пользователя с системой.

- Визуализация данных – представление новостного контента в графическом формате (карточки, ленты, дашборды), что улучшает восприятие и аналитическую обработку информации.

Эти процессы автоматизируются с использованием современных веб- и мобильных технологий, а также API-интерфейсов, предоставляющих доступ к глобальной базе новостей.

Автоматизация агрегирования новостных статей из множества источников направлена на создание системы, которая позволит собирать, обрабатывать и предоставлять пользователям актуальную, структурированную и персонализированную информацию. Система должна сократить время на поиск новостей, повысить точность анализа данных и минимизировать участие человека в рутинных процессах [4].

Проблемы, требующие автоматизации, охватывают несколько ключевых аспектов. Одним из них является фрагментация информации: пользователям приходится переключаться между разными платформами, что затрудняет формирование целостной картины событий. Также актуальна проблема больших объемов данных, так как ручная обработка новостных материалов невозможна из-за их масштабов и скорости появления.

Дополнительно возникает проблема дублирования данных, когда одни и те же новости публикуются на различных платформах, что приводит к информационному шуму. Еще один важный аспект – достоверность информации, требующая автоматизации проверки фактов для исключения фейковых новостей. Наконец, отсутствие персонализации приводит к тому, что пользователи сталкиваются с нерелевантным контентом, не соответствующим их интересам [5].

Цели автоматизации:

- автоматизированный сбор новостей из множества источников;
- фильтрация данных, удаление дубликатов и категоризация;
- выбор предпочтений и интересов пользователей, сохранение истории взаимодействия;

– реализация методов анализа текста для определения ключевых тем и актуальности информации;

– проверка достоверности информации посредством сравнения с проверенными источниками;

– сокращение времени, необходимое для поиска информации [6].

Автоматизированная система обладает высокой масштабируемостью, что позволяет ей адаптироваться к изменениям медиарынка и эффективно реагировать на новые вызовы.

Выделим основные этапы автоматизации:

– анализ предметной области и существующих решений. Изучение источников данных и методов их интеграции. Выявление проблем и недостатков текущих подходов;

– разработка архитектуры системы. Определение модулей системы, включая сбор, обработку и предоставление данных;

– реализация ключевых алгоритмов. Автоматический сбор новостей. Очистка и категоризация данных. Внедрение функций персонализации;

– тестирование системы. Проверка на корректность, надежность и производительность;

– внедрение и сопровождение. Настройка системы для конечных пользователей и обучение работе с ней.

Система агрегации новостей обрабатывает три основных типа входных данных:

– основной контент включает полнотекстовые новостные материалы с сохранением оригинального форматирования, заголовки статей с возможностью автоматической нормализации, а также краткие аннотации;

– метаинформация содержит точные временные метки публикации с указанием временных зон, канонические URL-адреса оригинальных публикаций, идентификаторы и наименования источников с поддержкой многоязычных вариантов;

– классификационные данные состоят из тематической рубрики по внутреннему классификатору системы, географических привязок с указанием стран, регионов и городов, а также специальных флагов, обозначающих важность и срочность материала. Все эти данные поступают в систему в структурированном виде и проходят многоэтапную обработку перед публикацией для конечных пользователей.

Приведем пример входного документа JSON от сервера:

```
{
  "article_id": "bbc_20250615_12345",
  "header": {
    "title": "Новые санкции ЕС",
    "subtitle": "Расширение списка ограничений"
  },
  "content": {
    "text": "Полный текст новости...",
    "formatted_html": "<p>Полный текст...</p>"
  },
  "metadata": {
    "published": "2025-06-15T08:45:00+01:00",
    "source": {
      "id": "bbc",
      "name": "BBC News"
    },
    "authors": [
      { "name": "John Doe", "role": "correspondent" }
    ],
    "classification": {
      "category": "politics",
      "geo_tags": ["EU", "UK"],
      "priority": "high"
    }
  }
}
```

Процесс обработки и нормализации входящих документов представляет собой многоступенчатую процедуру, включающую парсинг и валидацию структуры данных, извлечение и преобразование ключевой информации, унификацию форматов

дат и времени, дополнение метаданными и создание поисковых индексов для обеспечения быстрого доступа к информации. Система автоматически выявляет и корректирует различные аномалии в данных, такие как неполные или поврежденные записи,

несоответствия форматов, дублирующий контент и устаревшие материалы, что гарантирует высокое качество и актуальность хранимой информации.

Все поступающие документы подвергаются комплексному хранению с сохранением исходных данных в оригинальном формате (сырые данные), обработанной и нормализованной версии контента, а также специализированных поисковых индексов, оптимизированных для быстрого поиска и анализа. Такой трехуровневый подход к хранению обеспечивает как сохранность первоначальной информации, так и эффективность работы с актуальными нормализованными данными, поддерживая при этом высокую производительность поисковых операций. Архивация осуществляется с учетом временных меток и частоты обращения к данным, что позволяет оптимально управлять ресурсами хранения.

Разработанная система агрегации новостных статей реализована с использованием современных технологий и инструментов, обеспечивающих высокую производительность, кроссплатформенность и удобство разработки. Выбор программного обеспечения осуществлялся на основе анализа требований к функциональности, производительности и совместимости [7].

Основой клиентской части приложения стал React Native – популярный фреймворк для создания кроссплатформенных мобильных приложений. Его использование позволило разработать единую кодовую базу для платформ iOS и Android, что значительно сократило временные затраты на разработку. React Native обеспечивает нативную производительность благодаря оригинальным компонентам платформ, сохраняя при этом все преимущества React, включая виртуальный DOM и компонентный подход [8].

Для ускорения процесса разработки и тестирования был задействован Expo – комплекс инструментов и сервисов, работающий поверх React Native. Expo предоставляет готовые решения для типовых задач мобильной разработки, включая доступ к API устройства, систему сборки и удобные механизмы отладки. Особую ценность представляет Expo Go – мобильное приложение, позволяющее тестировать разрабатываемое решение на физических устройствах без необходимости сложной настройки среды разработки [9, 10].

Серверная часть системы построена на взаимодействии с NewsAPI – специализированным веб-сервисом для агрегации новостей. Выбор данного API обусловлен его надежностью, простотой интеграции

и широким охватом источников. Сборка финального приложения осуществлялась с использованием EAS Build – облачного сервиса от Expo для компиляции готовых билдов. Данное решение необходимо для настройки локальной среды различных платформ.

В связи с этим разработанная многофункциональная платформа агрегации новостей имеет технические требования, что обеспечивает бесперебойную работу приложения на всех стадиях его жизненного цикла. Комплекс технических решений был подобран с учетом особенностей кроссплатформенной разработки и специфики работы с новостными API.

Особое внимание уделено адаптивности системы к различным техническим условиям:

- оптимизация работы при нестабильном интернет-соединении;
- поддержка различных разрешений экранов и плотностей пикселей;
- корректная работа на устройствах с различными техническими характеристиками.

Выбранная архитектура технического обеспечения обеспечивает стабильную работу приложения на 98% современных мобильных устройств, быструю адаптацию к изменениям в API новостных источников, простоту масштабирования системы при росте числа пользователей, возможность легкой интеграции новых функций и модулей.

Технические решения были подобраны с учетом баланса между производительностью, стоимостью обслуживания и простотой развертывания, что делает систему экономически эффективной и удобной в эксплуатации.

Разработанная система реализует комплекс мер по обеспечению информационной безопасности, соответствующий современным требованиям к защите данных. Безопасность приложения обеспечивается на нескольких уровнях. На уровне передачи данных применяется защищенное HTTPS-соединение для всех запросов к NewsAPI. Это гарантирует шифрование передаваемой информации и предотвращает возможность перехвата данных третьими лицами. Для аутентификации при работе с API используется механизм API-ключей, которые хранятся в защищенной конфигурации среды Expo, исключая их попадание в публичный доступ [11, 12].

В клиентской части приложения реализованы механизмы валидации входящих данных от API, что предотвращает потенциальные атаки через внедрение вредоносного кода. Все получаемые новостные материалы проходят санитарную обработку перед отображением, исключая возможность XSS-атак. Для обработки поль-

звательского ввода строгие регулярные выражения. Особое внимание уделено защите персональных данных в соответствии с требованиями GDPR и ФЗ-152 [13, 14]. Комплексный подход к информационной безопасности позволяет гарантировать надежную защиту системы и соответствие современным стандартам безопасности мобильных приложений [15].

Заключение

Разработка многофункциональной платформы для агрегирования новостных статей играет ключевую роль в повышении продуктивности информационных агентств и улучшении удобства для пользователей. Применение передовых технологий обеспечивает оперативный доступ к актуальным материалам, стабильное обновление контента и удобные инструменты для анализа данных. Внедрение данной системы способствует оптимизации процессов получения, обработки и распространения информации, что способствует значительному улучшению качества медиамониторинга и скорости предоставления данных.

Список литературы

1. Абилюдинова Г.М., Айтимова А.М. Математическая модель оптимизации работ при разработке автоматизированных информационных систем // Цифровые модели и решения. 2024. Т. 3. № 4. С. 91–100. DOI: 10.29141/2949-477X-2024-3-4-6. EDN: JXILNM.
2. Лешук Н.С. Сайты-агрегаторы как средство информационной адаптации в глобальной сети интернет // Вестник науки и образования. 2021. № 11–1 (114). С. 29–32. EDN: QYRISV.
3. Сундатов В.О., Тупиков В.О. Проект сайта-агрегатора с автоматизированным наполнением контентом при помощи веб-парсинга // Интернаука. 2025. № 6–1 (370). С. 44–45. EDN: XNTDQM.
4. Абдрафикова С.Р. Характеристика сайта-агрегатора для бронирования жилья «Яндекс.Путешествия» // Ответственные и экономические науки. Студенческий научный форум: сборник статей по материалам LXXI студенческой международной научно-практической конференции (Москва, 05 февраля 2024 г.). М.: ООО «Международный центр науки и образования», 2024. С. 21–23. EDN: PYWCBU.
5. Сайдалиев Э.С., Тимошук О.А. Возможности сайта-агрегатора // Информационно-компьютерные технологии в экономике, образовании и социальной сфере. 2020. № 4 (30). С. 102–108. EDN: JWMVDB.
6. Кольева Н.С., Шемакин В.В. Примеры использования блокчейн-технологии // Математическое и компьютерное моделирование: сборник материалов XII международной научной конференции (Омск, 14 марта 2025 г.). Омск: Омский государственный университет им. Ф.М. Достоевского, 2025. С. 211–212. EDN: SKIFOR.
7. Сундатов В.О., Тупиков И.В. Разработка сайта агрегатора новостей. Использование ботов для информационной рассылки // Студенческий вестник. 2024. № 21–11 (307). С. 22–23. EDN: PDMINU.
8. Гусейнова А.Р., Магеррамов З.Т. Алгоритм работы онлайн-платформы для автоматизации поиска вакансий // Sciences of Europe. 2023. № 114 (114). С. 110–113. DOI 10.5281/zenodo.7811577. EDN: PMRJXE.
9. Лаврова Г.Н., Чубрина К.А. Развитие модели агрегации информационных сервисов электронной коммерции // Экономика и управление: проблемы, решения. 2018. Т. 6. № 12. С. 95–99. EDN: PNJOGR.
10. Рванова А.С., Кольева Н.С., Панова М.В. Разработка модели оценки структурной сложности программ // Цифровые модели и решения. 2024. Т. 3. № 2. С. 5–16. DOI: 10.29141/2949-477X-2024-3-2-1. EDN: UEZYBI.
11. Беликов А.Г. Разработка системы агрегации новостей сообществ социальных сетей // Постулат. 2021. № 1 (63). EDN: KQDFKV.
12. Черняк А.Г., Корелина Т.В., Зеленина А.Н. Создание агрегатора интернет-сервис-провайдеров с подсистемой автоматизированной обработки адресных баз // Вестник Воронежского института высоких технологий. 2022. № 2 (41). С. 123–129. EDN: AIXZRI.
13. Кортенко Л.В., Рудник В.А., Хасимьянова И.В. Безопасность персональных данных сотрудников при анкетировании по методологии «Оценка 360°» в ИС: Зарплата и управление персоналом ПРОФ // Современные проблемы обеспечения безопасности: сборник материалов XXVI Международной научно-практической конференции (Екатеринбург, 24–25 апреля 2024 г.). Екатеринбург: Уральский институт Государственной противопожарной службы МЧС России, 2024. С. 182–188. EDN: DDVKPZ.
14. Елизарова М.Ю. Правовой режим персональных данных лиц, участвующих в гражданском судопроизводстве // Известия Юго-Западного государственного университета. Серия: История и право. 2024. Т. 14. № 3. С. 108–120. DOI: 10.21869/2223-1501-2024-14-3-108-120. EDN: GYGFD.
15. Филиппов Д.С. Информационный сервис для сбора и анализа данных из Telegram в целях содействия принятию управленческих решений в экономической сфере на отраслевом и региональном уровне // ФЭС: Финансы. Экономика. Стратегия. 2023. Т. 20, № 4. С. 37–42. EDN: VXZPY.