

УДК 004.9

Смирнов Максим Алексеевич, магистрант, Уфимский университет науки и технологий, г. Уфа

СИСТЕМА СБОРА И ОБРАБОТКИ НОВОСТНЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ TELEGRAM И ВНЕШНИХ СЕРВИСОВ

Аннотация

В статье рассматривается проектирование программной системы для автоматизированного сбора, обработки и предоставления пользователям актуальной информации через Telegram-бота. Актуальность работы обусловлена ростом объема новостных данных и необходимостью оперативного получения релевантной информации из различных источников. Предлагаемое решение объединяет данные новостных ресурсов, погодных сервисов, валютных API и афиш мероприятий. Система выполняет фильтрацию, удаление дубликатов, структурирование и персонализацию информации с учетом предпочтений пользователя. Особое внимание уделено архитектуре серверной части, взаимодействию с внешними сервисами и организации хранения данных.

Annotation

The article discusses the design of a software system for automated collection, processing, and delivery of relevant information to users through a Telegram bot. The relevance of the study is обусловлена the growing volume of news data and the need for prompt access to useful information from various sources. The proposed solution integrates data from news resources, weather services, currency APIs, and event listings. The system performs filtering, duplicate removal, structuring, and personalization of information based on user preferences. Special attention is paid to the architecture of the server side, interaction with external services, and data storage organization.

Ключевые слова: Telegram-бот, агрегирование данных, новостные данные, внешние API, персонализация, информационная система.

Keywords: Telegram bot, data aggregation, news data, external APIs, personalization, information system

С развитием цифровых платформ и мессенджеров пользователи ежедневно сталкиваются с большим объемом новостной информации. Значительная часть этих данных не относится к интересам пользователя, быстро теряет актуальность или поступает из источников с низкой степенью доверия. Это усложняет поиск достоверных информационных источников и увеличивает время на их обработку.

В таких условиях актуальной становится задача создания программной системы, которая автоматически собирает данные из нескольких источников, выполняет их обработку и предоставляет пользователю только релевантную информацию через удобный канал взаимодействия — Telegram-бота.

Цель: спроектировать и разработать системы сбора и обработки новостных данных и Telegram-бота с интеграцией внешних сервисов, которые обеспечивают пользователю оперативный доступ к актуальной информации.

Задачи:

1. Провести анализ предметной области и существующих решений;
2. Выявить требования к системе сбора, обработки и выдачи новостной информации;
3. Разработать функциональную модель системы и декомпозировать задачу на подзадачи;
4. Спроектировать архитектуру программного решения и структуру базы данных;
5. Проанализировать технологии реализации ПО;
6. Реализовать Telegram-бота и серверную часть с интеграцией внешних сервисов;
7. Провести проверку работоспособности и оценку качества разработанного решения.

Объект исследования — процесс автоматизированного информационного обслуживания пользователей на основе новостных данных и внешних сервисов.

Предмет исследования — функциональность программного обеспечения доставки разнородных информационных данных с помощью Telegram-бота и внешних API.

Прикладная область работы относится к информационным системам, обеспечивающим агрегирование контента и его доставку конечному пользователю и является важным направлением в сервисах повседневного спроса. Основная специфика данной области заключается в следующем:

1. Новостные потоки, метеорологические данные, курсы валют и локальные афиши мероприятий размещаются на десятках независимых веб-ресурсов и мобильных приложений, каждый из которых обладает собственной архитектурой, частотой обновлений, форматом представления данных и политикой доступа;

2. Данные требуют регулярного обновления и контроля актуальности;

3. Для повышения полезности информации необходимо учитывать предпочтения пользователя (регион, категория);

4. Пользователю требуется интуитивно понятный интерфейс взаимодействия без сложной настройки.

Telegram-боты в этой области получили широкое распространение благодаря доступности, простоте использования и низкому порогу входа для пользователей. Однако для практической ценности одного канала доставки недостаточно: требуется полноценная серверная логика, обеспечивающая сбор данных, их фильтрацию, удаление дубликатов, хранение и формирование ответов по запросу.

Следовательно, оптимальное решение должно представлять собой не только бот-интерфейс, но и комплексную систему обработки данных с интеграцией внешних сервисов.

Для удобства постановки задачи была разработана контекстная диаграмма ИС «Система сбора и обработки новостных данных с использованием Telegram-бота и внешних сервисов» приведенная на Рисунок 1.

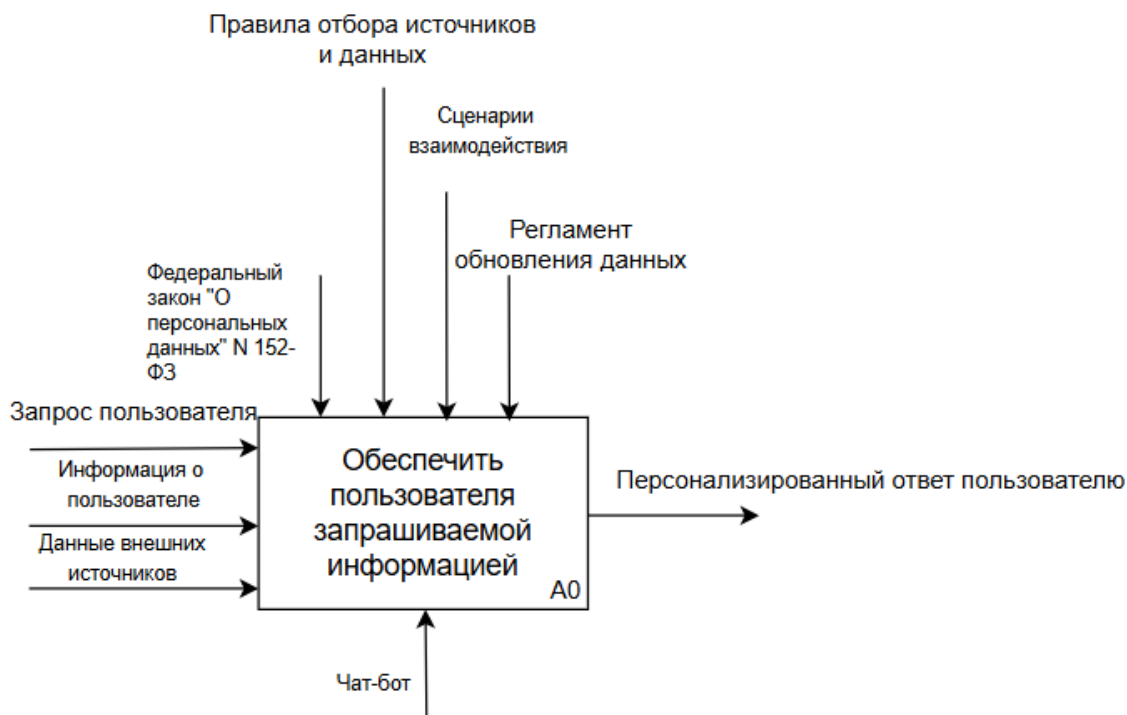


Рисунок 1 – Контекстная диаграмма

- Информация о пользователе: параметры, которые задает пользователь для персонализации;
- Данные внешних источников, т.е. новости, погода, курс валют, мероприятия.
- Правила отбора источников и данных: определяют принципы фильтрации и преобразования;
- Сценарии взаимодействия (описывает логику взаимодействия между пользователем и ботом);
- Регламент обновления данных (задает периодичность и условия фоновой актуализации).

Литература

1. Воробьев, А. В. Оценка влияния геомагнитной активности на метрологические характеристики инклинометрических информационно-измерительных систем / А. В. Воробьев, Г. Р. Воробьева // Измерительная техника. – 2017. – № 6. – С. 21-24.
2. Воробьев, А. В. Подход к восстановлению геомагнитных данных на базе концепции цифровых двойников / А. В. Воробьев, В. А. Пилипенко // Солнечно-земная физика. – 2021. – Т. 7, № 2. – С. 53-62. – DOI 10.12737/szf-72202105.

3. Воробьев, А. В. Корреляционный анализ геомагнитных данных, синхронно регистрируемых магнитными обсерваториями INTERMAGNET / А. В. Воробьев, Г. Р. Воробьева // Геомагнетизм и аэрномия. – 2018. – Т. 58, № 2. – С. 187-193. – DOI 10.7868/S0016794018020049.

4. Воробьев, А. В. Подход к обнаружению и устранению артефактов пространственных изолиний в приложениях Веб-ГИС / А. В. Воробьев, Г. Р. Воробьева // Компьютерная оптика. – 2023. – Т. 47, № 1. – С. 126-136. – DOI 10.18287/2412-6179-CO-1127.

5. Воробьев, А. В. Индуктивный метод восстановления временных рядов геомагнитных данных / А. В. Воробьев, Г. Р. Воробьева // Труды СПИИРАН. – 2018. – № 2(57). – С. 104-133. – DOI 10.15622/sp.57.5.

6. Воробьев, А. В. Концепция единого пространства геомагнитных данных / А. В. Воробьев, Г. Р. Воробьева, Н. И. Юсупова // Труды СПИИРАН. – 2019. – Т. 18, № 2. – С. 390-415. – DOI 10.15622/sp.18.2.390-415.

Literature

1. Vorobyev, A. V. Assessment of the influence of geomagnetic activity on the metrological characteristics of inclinometric information-measuring systems / A. V. Vorobyev, G. R. Vorobyeva // Measurement Techniques. – 2017. – No. 6. – Pp. 21–24.

2. Vorobyev, A. V. An approach to restoration of geomagnetic data based on the digital twin concept / A. V. Vorobyev, V. A. Pilipenko // Solar-Terrestrial Physics. – 2021. – Vol. 7, No. 2. – Pp. 53–62. – DOI 10.12737/szf-72202105.

3. Vorobyev, A. V. Correlation analysis of geomagnetic data synchronously recorded by INTERMAGNET magnetic observatories / A. V. Vorobyev, G. R. Vorobyeva // Geomagnetism and Aeronomy. – 2018. – Vol. 58, No. 2. – Pp. 187–193. – DOI 10.7868/S0016794018020049.

4. Vorobyev, A. V. An approach to detection and elimination of spatial isoline artifacts in Web-GIS applications / A. V. Vorobyev, G. R. Vorobyeva // Computer Optics. – 2023. – Vol. 47, No. 1. – Pp. 126–136. – DOI 10.18287/2412-6179-CO-1127.

5. Vorobyev, A. V. Inductive method for restoration of geomagnetic data time series / A. V. Vorobyev, G. R. Vorobyeva // SPIIRAS Proceedings. – 2018. – No. 2(57). – Pp. 104–133. – DOI 10.15622/sp.57.5.

6. Vorobyev, A. V. Concept of a unified geomagnetic data space / A. V. Vorobyev, G. R. Vorobyeva, N. I. Yusupova // SPIIRAS Proceedings. – 2019. – Vol. 18, No. 2. – Pp. 390–415. – DOI 10.15622/sp.18.2.390-415.