

СИСТЕМА АВТОМАТИЗИРОВАННОГО СБОРА И АНАЛИЗА ДАННЫХ О ПРОДАЖАХ ТОВАРОВ В ИНТЕРНЕТ-МАГАЗИНАХ

А. Р. ГАНИЕВ¹

¹sofaebola@mail.ru

¹ ФГБОУ ВО «Уфимский университет науки и технологий» (УУНИТ)

Аннотация. В данной статье рассматривается система автоматизированного сбора и анализа данных о продажах товаров в интернет-магазинах. Основное внимание уделено использованию парсера в сочетании с прокси-сервером для эффективного и анонимного сбора данных. Полученные данные редактируются и хранятся в базе данных. Для анализа данных и создания визуализации используются динамические дашборды, а интеллектуальный анализ проводится с помощью карт Кохонена. Результаты исследования показывают, что автоматизация сбора и анализа данных способствует улучшению бизнес-процессов, повышению конверсии и оптимизации маркетинговых стратегий интернет-магазинов.

Ключевые слова: автоматизированный сбор данных; анализ данных; веб-скрапинг; прокси-сервер; дашборды; карты Кохонена; интернет-магазины; электронная коммерция; визуализация данных; интеллектуальный анализ; оптимизация продаж.

Для повышения эффективности процесса реализации товаров предприятиям необходимо использовать современные информационные технологии сбора информации о потребительском спросе, продажах конкурентов, применять инструменты анализа собранной информации и оперативно принять управленческие решения. Решение этих задач проводится в несколько этапов.

На *первом этапе* проводится автоматизированный сбор данных о продажах товаров в интернет-магазинах – это процесс использования специализированных программных решений и технологий для автоматического получения, обработки и анализа информации о транзакциях, запасах, ценах и других параметрах, связанных с продажами товаров через интернет.

Этот процесс включает в себя использование метода веб-скрейпинга (*web scraping*) – это метод автоматизированного извлечения данных, или парсинг, предназначен для извлечения данных с веб-сайтов, их обработку и анализ с целью получения полезной информации для принятия бизнес-решений. Парсинг использует программное обеспечение или скрипты для сканирования HTML-кода веб-страницы и извлечения необходимой информации, которая может включать в себя цены на товары, наличие товаров на складе, отзывы клиентов и другие данные, которые могут быть полезны для анализа продаж.

Разработка парсера позволит компаниям следить за ценами на товары, оценивать их конкурентоспособность, а также оперативно реагировать на изменения в ценах и наличие товаров у конкурентов.

В качестве системы для автоматизированного сбора данных в рамках данной работы был выбран Web-scraper – инструмент, используемый для автоматизированного сбора данных, который имеет множество настроек решения разных типов задач, относящихся к автоматизации сбора данных о товарах в интернет-магазинах.

Для обеспечения стабильности и анонимности процесса парсинга необходимо использовать прокси-серверы. Это позволяет избежать блокировок со стороны сайтов, которые могут ограничивать частоту запросов с одного IP-адреса. Прокси-серверы помогут распределить запросы и уменьшить нагрузку на отдельные IP-адреса. Основным преимуществом прокси-сервера является ротация прокси-серверов, которая используется для предотвращения блокировок и обхода ограничений, используются пулы прокси-серверов, которые периодически меняются (ротироваются). Это позволяет распределять запросы между разными IP-адресами и снижать риск блокировки.

Этап парсинга данных является фундаментальным в системе автоматизированного сбора и анализа данных о продажах товаров в интернет-магазинах. Использование Web-scraper и прокси-серверов позволяет эффективно и безопасно извлекать нужную информацию, обеспечивая актуальность и полноту данных для последующего анализа и принятия решений.

Собранные с помощью парсинга данные могут быть импортированы в Microsoft Excel для предварительной обработки: очистки, фильтрации, сортировки данных, удаления дубликатов, исправления ошибок в данных

Второй этап – хранение данных – является критически важным этапом в системе автоматизированного сбора и анализа данных о продажах товаров в интернет-магазинах. Этот этап включает в себя выбор подходящей системы управления базами данных (СУБД), создание структуры базы данных, а также методы загрузки и управления данными. В данной работе рассматривается использование phpMyAdmin для управления базой данных MySQL.

MySQL – это популярная реляционная система управления базами данных, которая широко используется благодаря своей производительности, надежности и простоте в использовании. Для удобного управления базой данных MySQL через веб-интерфейс применяется phpMyAdmin, который является инструментом с открытым исходным кодом и позволяет выполнять различные операции с базами данных MySQL через веб-интерфейс. Он предоставляет пользователям удобные возможности для создания и управления базами данных, выполнения запросов и импорта/экспорта данных.

Для эффективного хранения данных о продажах товаров разрабатывается структура базы данных, представленная на рис. 1.

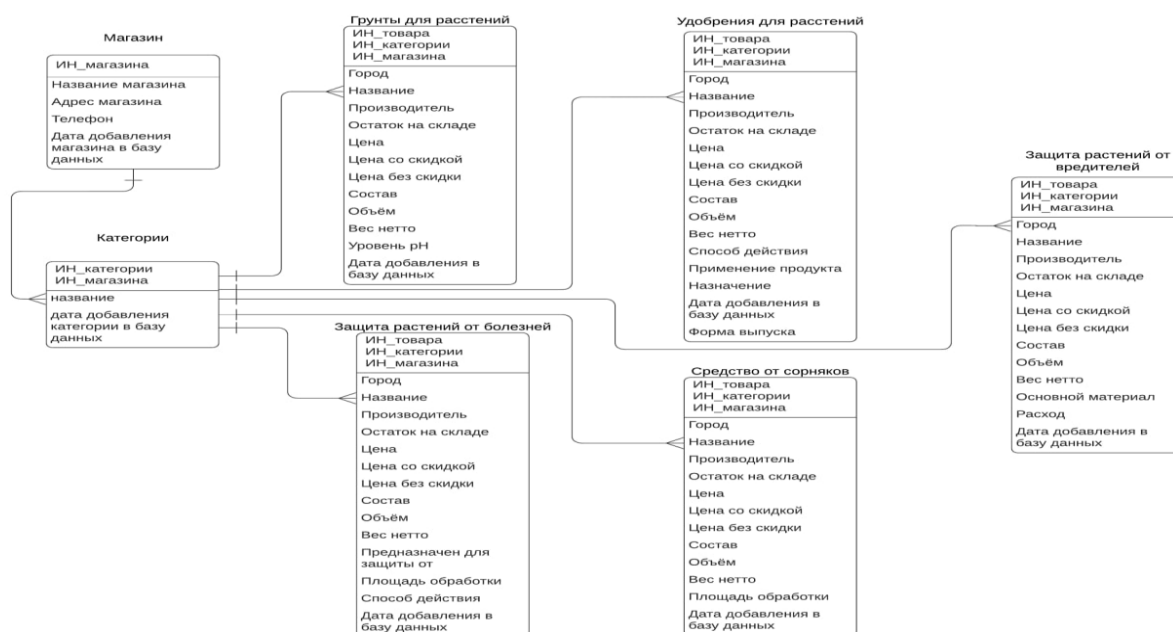


Рис. 1. Структура базы данных

Для хранения данных о продажах была создана структура базы данных, включающая таблицы для магазинов, категории и названия категорий, имеющие разные компоненты.

Этап хранения данных играет ключевую роль в системе автоматизированного сбора и анализа данных о продажах товаров в интернет-магазинах. Выбор надежной СУБД, создание продуманной структуры базы данных, правильный импорт данных и регулярное управление базой данных обеспечивают эффективное хранение и доступ к данным для последующего анализа и визуализации. Использование phpMyAdmin для управления базой данных MySQL позволяет упростить многие задачи администрирования и сделать процесс хранения данных более удобным и надежным.

На *третьем этапе* используются программные средства для создания визуальных представлений данных и применения методов анализа для выявления скрытых закономерностей и трендов. Визуализация данных помогает наглядно представить информацию, сделать её более понятной и доступной для анализа. Основным инструментом для визуализации данных, который рассматриваются в рамках данной работы, Power BI Desktop (рис. 2).

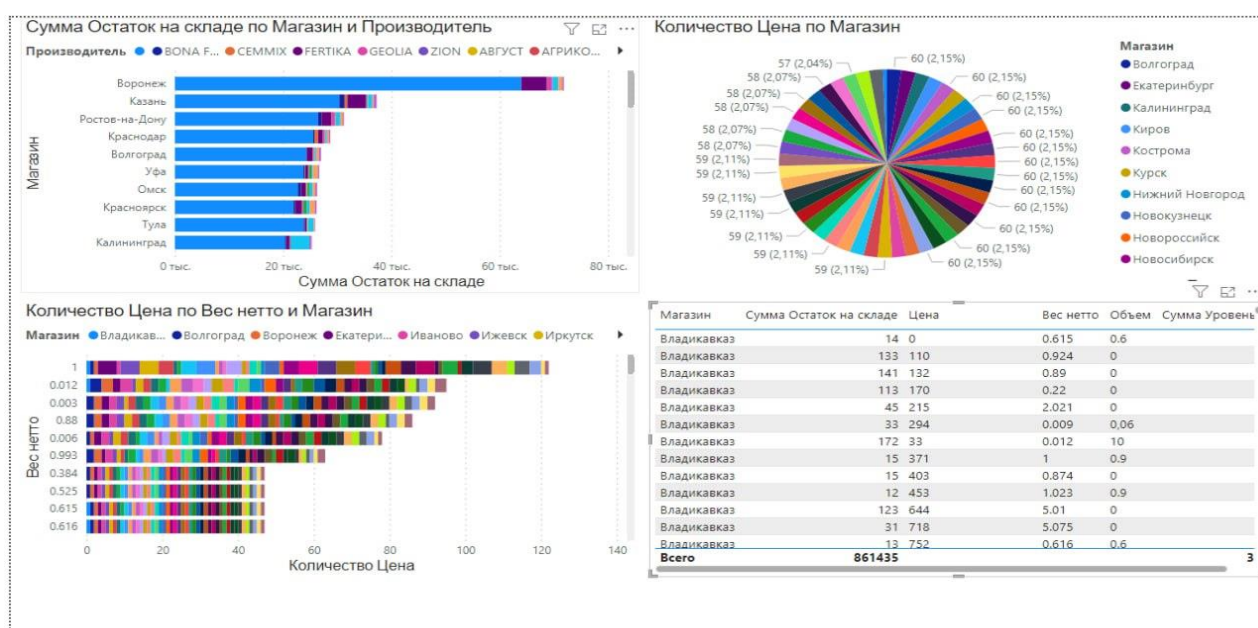


Рис. 2. Динамические дашборды

На *следующем этапе* проводится интеллектуальный анализ данных о продажах с помощью метода самоорганизующихся карт Кохонена. Этот метод помогает выявить структуры, кластеры и аномалии в данных, которые сложно обнаружить в исходном многомерном пространстве признаков. Так, с помощью данного метода были выявлены такие закономерности: наибольшее количество остатков товаров находятся на складах в городах Воронеж, Казань, Уфа и Тольятти; наивысшая цена – в Ярославле, Хабаровске, Ульяновске, Тюмени, Ростов-на-Дону и др.

Последовательное применение информационных технологий сбора данных и интеллектуальных технологий анализа данных помогло выявить важные закономерности о продажах, что позволяет лучше понять поведение потребителей и маркетинговые стратегии конкурентов, а следовательно, принять более рациональное оперативное решение. Полученные результаты демонстрируют значимость интеграции различных инструментов для достижения комплексного анализа данных. Внедрение такой системы способствует улучшению стратегических решений и повышению эффективности интернет-магазина.

СПИСОК ЛИТЕРАТУРЫ

1. **Смирнов В. А.** PHP и MySQL. Разработка динамических веб-приложений. [Электронный ресурс]. URL: https://books.google.com/books/about/PHP_MySQL_HTML_5_%D0%B8_CSS_3_%D0%A0%D0%B0%D0%B7%D1%80%D0%B0%D0%B1%D0%BE.html?id=Ol-tCwAAQBAJ (дата обращения 24.10.2024).
2. **Сохейл Д. Б.** Power BI: моделирование на экспертном уровне. М.: БХВ, 2022. 490 с.
3. **Дронов В. А.** PHP и MySQL. 25 уроков для начинающих. М.: БХВ, 2020. 426 с.
4. **Гольчевский Ю. В.** Автоматизация механизмов поиска информации на основе открытых источников в сети Интернет. М.: Б/и, 2017. 414 с.
5. **Станислав С. С.** Power BI: Бизнес-анализ и визуализация данных. [Электронный ресурс]. URL: <http://finalytics.pro/inform/power-bi-analiz-vizualizatsiya/> (дата обращения 20.10.2024).
6. **Тобиас Р. А.** Изучение SQL, с помощью phpMyAdmin. [Электронный ресурс]. URL: <https://php-myadmin.ru/learning/> (дата обращения 20.10.2024).

ОБ АВТОРАХ

ГАНИЕВ Айдар Радикович, маг. каф. ТК. Дипл. системный анализ и управление (УУНИТ, 2024).

METADATA

Title: Automated System for Collecting and Analyzing Sales Data in Online Stores.

Authors: A. R. Ganiev

Affiliation: Ufa University of Science and Technology (UUST), Russia.

Email: sofaebola@mail.ru

Language: Russian.

Source: Molodezhnyj Vestnik UGATU (scientific journal of Ufa University of Science and Technology), no. 2 (33), pp. 28-31, 2025. ISSN 2225-9309 (Print).

Abstract: This article discusses an automated system for collecting and analyzing sales data in online stores. The focus is on using a parser in combination with a proxy server for efficient and anonymous data collection. The collected data is edited and stored in a database, analyzed using dashboards, and evaluated with Kohonen maps for business optimization and decision-making.

Key words: Automated data collection, data analysis, web scraping, proxy server, dashboards, Kohonen maps, online stores, e-commerce, data visualization, data mining, sales optimization.

About authors:

GANIEV Aidar Radikovich, Master of Technics & Technology (UUST, 2024).