Ministry of Science and Higher Education of the Republic of Kazakhstan

SDU University

Serikova Meruyert

Research and analysis user centered logistic website on the base of user data

THESIS

Presented in Partial Fulfilment for the

Degree of Master of Technical Science in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Lyazzat Atymtayeva**

Kaskelen, June 2024

SDU University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty of Engineering and Natural Sciences

Assistant Professor, PhD. Akhmedov Ramis

_____

« 4 » June 2024

Topic of the thesis: Research and analysis user centered logistic website on the base of user data

---

Thesis submitted as part of the requirements for the award of the MSc in "7M06102 - Computer Science", SDU University

Head of Department	Zhanar Mukash
Academic Supervisor	Lyazzat Atymtayeva
Master student	Meruyert Serikova

Kaskelen, 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Meruyert Serikova

June 2024

Acknowledgements

I sincerely thank Lyazzat Atymtayeva, who supervised my dissertation, for her support, guidance and inspiration throughout this research.

I am also grateful to the faculty of the Engineering Department of SDU University for their insightful comments and helpful criticism, which formed the basis of this dissertation. I would also like to thank Inkar Shoganova, who supervised my research internship.

I want to express my gratitude to my friends and family for their understanding and constant support throughout this research. Their confidence in my skills and encouragement have given me strength and drive. Lastly, I would like to express my gratitude to the members of the @cargoasia Telegram channel for their assistance in supplying the data for this research. Their willingness to engage was greatly appreciated, and their efforts made this research possible.

Dedication

I dedicate my thesis to my family, who have consistently provided me with inspiration and encouragement. Their support, affection, and faith in me have enabled me to complete this journey. In addition, I dedicate my work to my friends and coworkers, whose energy and company have made my experience richer. I hope that this thesis will serve as an example of the strength of community, perseverance, and teamwork.

Abstract

This thesis explores how to make a logistics website that is user-friendly and can handle data from Telegram. The goal is to turn unstructured text messages into structured data for logistics companies. This will help us make better decisions and work more efficiently. I used tools like PHP, Composer, MadelineProto, and PhpSpreadsheet to organize the data. The website was designed using Figma to be simple and easy to understand. Structured data helps in planning and decision making. The study shows that better data and design can improve logistics operations. Future work will focus on adding more features and improving the system with real-time data and user feedback.

Аннотация

В этой статье рассматривается, как создать логистический веб-сайт, который был бы удобен в использовании и мог обрабатывать данные из Telegram. Цель состоит в том, чтобы превратить неструктурированные текстовые сообщения в структурированные данные для логистических компаний. Это поможет нам принимать более правильные решения и работать более эффективно. Для систематизации данных я использовала такие инструменты, как PHP, Composer, Made lineProto и Php Spreadsheet. Веб-сайт был разработан с использованием Figma, чтобы он был простым и понятным. Также структурированные данные помогают в планировании и принятии решений. Исследование показывает, что более качественные данные и дизайн могут улучшить логистические операции. Будущая работа будет сосредоточена на добавлении дополнительных функций и совершенствовании системы с использованием данных в режиме реального времени и отзывов пользователей.

Аңдатпа

Бұл диссертация пайдаланушыға ыңғайлы және Telegram деректерін өңдей алатын логистикалық веб-сайтты қалай жасау керектігін зерттейді. Мақсат – құрылымдалмаған мәтіндік хабарламаларды логистикалық компаниялар үшін құрылымдық деректерге айналдыру. Бұл бізге жақсырақ шешімдер қабылдауға және тиімдірек жұмыс істеуге көмектеседі. Деректерді ұйымдастыру үшін PHP, Composer, MadelineProto және PhpSpreadsheet сияқты құралдарды қолдандым. Веб-сайт қарапайым және түсінікті болу үшін Figma көмегімен жасалынды. Сондай-ақ құрылымдық деректер жоспарлауға және шешім қабылдауға көмектеседі. Зерттеулер бойынша, жақсырақ деректер мен дизайн логистикалық операцияларды жақсарта алады. Алдағы жасалатын жұмыс нақты уақыттағы деректер мен пайдаланушы пікірі арқылы қосымша мүмкіндіктерді қосуға және жүйені жақсартуға бағытталатын болады.

Abbreviations

API - Application Programming Interface

UX - User Experience

UI - User Interface

PHP - PHP Hypertext Preprocessor

UCD - User Centered Design

Regex - Regular Expression

Table of Contents

Chapter 1

Background and motivations

1.1 Introduction

The digital transformation era has significantly reshaped various industries, and logistics is no exception. Logistics, a critical sector involving the management of the flow of goods and services, has been increasingly driven towards digitalization to streamline operations, reduce costs, and enhance customer satisfaction. However, a persistent issue in the logistics industry, both globally and in Kazakhstan, is the reliance on instant messengers like Telegram for placing orders. These messages, typically in free text format, often lack structure, making them difficult to analyze and utilize effectively.

This thesis addresses the critical problem of unstructured data in logistics communications. Across the globe, logistics companies face the challenge of processing orders and information received through instant messengers, where the data provided is often arbitrary and inconsistent. This unstructured data format complicates subsequent analysis and decision-making processes. The need for a system that can efficiently analyze this information and translate it into structured, actionable insights is evident.

The primary objective of this research is to develop a user-centered logistics website that can analyze unstructured data from instant messengers like Telegram. By leveraging a system that uses keyword searches and data parsing techniques, the goal is to automate the extraction and structuring of relevant logistics information. This system will then integrate the processed data into a website, which not only presents the data in a user-friendly format but also provides analytical tools and guidance for logistics decisions regarding goods and routes. This approach aims to facilitate logistics companies, particularly those with drivers and service personnel, in navigating the complexities of logistics operations with greater ease and efficiency.

Globally, various systems and research efforts have attempted to address similar issues in logistics. Advanced data analysis techniques have been employed by logistics companies and social networks to sift through vast amounts of data, identify trends, and inform strategic decisions. For instance, big data analytics has been used to enhance logistics efficiency by optimizing transportation routes,

inventory levels, and demand patterns. Studies have demonstrated significant efficiency improvements through predictive models, yet these solutions often come with high implementation costs and require specialized expertise, making them less accessible to small and medium-sized enterprises.

Moreover, existing logistics systems are often designed with a different audience in mind, may be paid, and typically offer interfaces in English, which might not be suitable for users in Kazakhstan who prefer Russian or Kazakh interfaces. These systems also might not effectively analyze the specific format of Telegram messages used in local logistics communications. Therefore, there is a pressing need for a localized solution tailored to the unique requirements of the logistics industry in Kazakhstan.

The proposed system in this research offers several advantages. It is designed to be user-friendly and accessible, with an interface that supports Russian and Kazakh languages. Unlike many existing systems, it focuses specifically on analyzing Telegram messages, ensuring compatibility with the prevalent communication practices in the local logistics industry. By providing detailed analysis and forecasts based on user requests, this system aims to enhance decision-making processes and operational efficiency for logistics companies.

The research question guiding this thesis is: How can the logistics system be improved to efficiently analyze unstructured data from instant messengers and provide actionable insights for logistics decisions?

In summary, this research seeks to provide a comprehensive solution to the problem of unstructured data in logistics communications. By developing a system that can analyze and structure this data, and integrating it into a user-centered website, the study aims to improve logistics operations and decision-making processes. The outcomes are expected to offer valuable insights and practical solutions, contributing to the enhancement of digital logistics infrastructure both locally and globally.

1.2 Motivation and Importance of the study

The motivation for this study arises from the ongoing digital transformation impacting various sectors, with logistics being a prime example. In today's fast-paced world, logistics companies are increasingly reliant on digital communication channels to manage orders and coordinate activities. Platforms like Telegram are widely used due to their ease of access and real-time communication capabilities. However, the unstructured nature of the data exchanged through these platforms presents significant challenges. Messages are often informal and inconsistent, making it difficult to extract meaningful information for decision-making processes.

The importance of this study is underscored by the potential benefits it offers. A system that can efficiently parse and analyze unstructured data from instant messengers can revolutionize logistics operations. By providing structured insights from chaotic message streams, logistics companies can streamline their operations, reduce errors, and enhance customer satisfaction. Furthermore, this study addresses a specific gap in the market for localized solutions that cater to the language and operational nuances of the Kazakh logistics industry. In doing

so, it aims to set a precedent for similar applications in other regions and sectors.

1.3 Research Contributions

This research aims to make several pivotal contributions to the fields of logistics and digital communication:

1. Innovative Data Handling: It introduces a novel method for handling unstructured data from instant messengers. By leveraging keyword searches and data parsing techniques, this study presents a systematic approach to extracting relevant logistics information from free-text messages.

2. Localized Solution: The research focuses on developing a solution tailored to the Kazakh logistics industry, addressing the unique linguistic and operational requirements of the region. This is a significant step forward from existing systems that are predominantly designed for English-speaking markets and often do not cater to the specific needs of local industries.

3. User-Centered Design: By adopting a user-centered design approach, the study ensures that the developed system is intuitive and user-friendly. This is particularly important in the logistics industry, where users may not have advanced technical skills but need to navigate complex data quickly and efficiently.

4. Enhanced Decision-Making: The system is designed to provide actionable insights, thereby enhancing the decision-making capabilities of logistics companies. This includes optimizing routes, managing inventory, and forecasting demand more accurately.

5. Academic Contribution: The study contributes to the academic discourse on digital logistics and user-centered design. It provides empirical data and practical insights that can inform future research and development in these areas.

1.4 Research Methodology Overview

The research methodology for this study is comprehensive, involving several critical steps to ensure the effective analysis and utilization of unstructured data from Telegram:

1. Data Collection: The first step involves collecting data from Telegram channels using the Telegram API. This allows for the automated retrieval of messages related to logistics operations, ensuring a large and diverse dataset.

2. Data Parsing and Keyword Search: Using PHP and regular expressions, the collected data will be parsed to extract relevant information. Keyword searches will be implemented to identify specific logistics terms, such as destinations, cargo types, weights, and delivery times.

3. Data Structuring: The parsed data will be structured into a standardized format, such as JSON or Excel, to facilitate further analysis. This step is crucial for converting the unstructured message content into a usable dataset.

4. Website Development: A user-centered logistics website will be designed and developed to integrate the structured data. The website will feature analytical tools and dashboards to help users visualize and interpret the data.

5. A/B Testing and User Feedback: To evaluate the effectiveness of the website, A/B testing will be conducted. This involves comparing different versions of the website to determine which design elements work best. Additionally, user feedback will be gathered to refine and improve the system.

1.5 Expected Outcomes

The expected outcomes of this research are multifaceted and aim to significantly enhance logistics operations:

1. Operational Efficiency: By automating the analysis of unstructured data from Telegram, the system is expected to streamline logistics operations. This includes faster order processing, improved route optimization, and better inventory management.

2. Enhanced Decision-Making: The system will provide detailed insights and forecasts based on the analyzed data, enabling logistics companies to make more informed decisions. This can lead to cost savings, improved service levels, and higher customer satisfaction.

3. User-Friendly Interface: The developed website will feature an intuitive interface that supports Russian and Kazakh languages, making it accessible to local users. This is particularly important in ensuring the system's adoption and effectiveness in the Kazakh logistics industry.

4. Scalability and Adaptability: While the system is designed for the Kazakh market, the underlying methodology can be adapted for other regions and industries. This scalability is a significant advantage, offering potential applications beyond the initial scope of the study.

5. Contribution to Academic Knowledge: The research will provide valuable data and insights that contribute to the academic understanding of digital logistics and user-centered design. It will serve as a foundation for future studies and developments in these fields.

In conclusion, this study aims to address a critical gap in the logistics industry by developing a system that can efficiently analyze unstructured data from instant messengers and provide actionable insights for logistics decisions. The research is motivated by the need for more effective data handling solutions in logistics, particularly in regions like Kazakhstan where localized systems are lacking. By introducing innovative data parsing techniques and a user-centered website design, the study seeks to improve operational efficiency, enhance decision-making, and provide a practical solution tailored to the needs of the local logistics industry. The anticipated outcomes of this research not only promise to benefit logistics companies but also contribute to the broader academic discourse on digital logistics and user-centered design.

Chapter 2

Literature review

2.1 About unstructured data

This paper "Big Data Analytics for Logistics and Supply Chain Management: Market-oriented Framework" provides a comprehensive overview of how big data analytics can be applied to logistics and supply chain management. The authors propose a market-oriented framework that integrates big data analytics into logistics operations, focusing on improving efficiency and decision-making. [9] They highlight the potential of big data to transform logistics by providing real-time insights, predictive capabilities, and enhanced operational efficiency. Key areas of application include transportation management, inventory optimization, demand forecasting, and customer relationship management.[9] The paper demonstrates the significant impact of big data on logistics, emphasizing the importance of data-driven approaches in optimizing logistics performance and achieving operational excellence.[10]

There "The Role of Natural Language Processing in Analyzing Unstructured Data" discusses the advancements in natural language processing (NLP) techniques, particularly rule-based information extraction systems, for analyzing unstructured data.[10] The authors argue that despite the rise of machine learning approaches, rule-based systems remain relevant and effective for specific tasks. The paper highlights the strengths of rule-based methods in terms of precision and interpretability, and provides examples of their application in various domains. The authors also explore the potential of NLP in transforming unstructured text into structured data, which can be valuable for logistics applications. By applying NLP techniques to logistics data, companies can gain deeper insights into customer needs, market trends, and operational challenges, leading to more informed decision-making.[11]

Also in this "Integrating Big Data Analytics with Logistics Management" paper provides a systematic review of the impact of big data analytics on logistics management, followed by a longitudinal case study. The authors identify key areas where big data can enhance logistics operations, such as demand forecasting, route optimization, and inventory management.[11] The case study demonstrates the practical benefits of big data, including cost reductions and efficiency im-

provements, through real-world applications in a logistics company. The study underscores the transformative potential of big data analytics in logistics, highlighting its role in improving operational efficiency and strategic planning. The integration of big data into logistics processes enables companies to make data-driven decisions, optimize resource allocation, and enhance overall performance.

This "Natural Language Processing in Logistics: Challenges and Opportunities" paper delves into the applications of natural language processing (NLP) in various industries, including logistics.[12] It focuses on sentiment analysis and opinion mining as key areas where NLP can provide valuable insights. The authors discuss the challenges of processing unstructured text data and the potential benefits of applying sentiment analysis to understand customer feedback and market trends. The paper emphasizes the importance of NLP in extracting meaningful information from unstructured data, which is critical for logistics companies aiming to improve customer service and operational efficiency.[12] By leveraging NLP techniques, logistics companies can better analyze customer sentiments, identify emerging trends, and respond proactively to market demands.

This "Data Mining Techniques for Analyzing Unstructured Data" book offers a comprehensive overview of data mining techniques, discussing their applications in various domains, including logistics. The authors cover key concepts, algorithms, and tools for data mining, and provide practical examples of their use. The book highlights the potential of data mining to extract valuable insights from large datasets, including unstructured data.[13] The application of data mining techniques in logistics can help companies uncover hidden patterns, optimize operations, and enhance decision-making processes. By analyzing unstructured data from sources like instant messengers, logistics companies can gain a competitive edge by making more informed and timely decisions.[13]

This "Leveraging Cloud Computing for Logistics Management" paper provides an overview of cloud computing and its potential applications in various industries, including logistics. The authors discuss the benefits of cloud computing, such as scalability, cost savings, and flexibility, and provide examples of its use in logistics management. They highlight how cloud computing can support real-time data analytics and enhance operational efficiency.[13] The integration of cloud-based solutions in logistics enables companies to process large volumes of data quickly, collaborate more effectively, and adapt to changing market conditions.[13] Cloud computing also facilitates the implementation of advanced analytics and machine learning algorithms, driving innovation and improving logistics performance.

Gap for data analysis part: despite the extensive coverage of big data analytics, natural language processing, and data mining techniques, the literature primarily focuses on structured data sources and traditional analytics methods. There is a significant gap in addressing the complexities associated with analyzing unstructured data from instant messengers like Telegram, which are widely used in logistics communications. Additionally, the integration of these advanced data analysis techniques with localized solutions tailored to specific linguistic and operational needs of regions like Kazakhstan remains underexplored. This research aims to fill these gaps by developing a user-centered logistics system that can effi-

ciently process and analyze unstructured data, providing actionable insights and improving decision-making processes.

2.2 Design in logistics

This seminal book on user-centered design (UCD) "User-Centered Design and Usability of Logistics Systems" explores various methodologies and principles for creating systems that prioritize user needs and usability. It emphasizes the importance of involving users throughout the design process to create intuitive and effective systems.[13] The book covers a range of topics including usability testing, iterative design, and the psychological principles behind user-friendly interfaces. It provides a strong theoretical foundation for applying UCD principles to any system, including logistics applications.[14] By focusing on user needs and preferences, logistics systems can be designed to improve user satisfaction, operational efficiency, and overall effectiveness.

This paper "User-Centered Design for Developing Logistics Applications" explores various methods and tools to support human-centered design (HCD) in developing applications, including those for logistics. It emphasizes the importance of involving users throughout the design process to create intuitive and effective systems.[14] The paper discusses different HCD methodologies, such as user research, prototyping, and usability testing, and provides practical guidelines for their implementation. By adopting a human-centered approach, logistics applications can be designed to meet the specific needs of users, resulting in higher adoption rates and better user experiences.

This book "Advances in User Experience Design for Digital Systems" provides a comprehensive framework for user experience (UX) design, emphasizing the importance of a user-centered approach. Garrett outlines key elements of UX, including strategy, scope, structure, skeleton, and surface, and explains how these elements interact to create a cohesive user experience.[14] The book offers practical advice for designing digital systems that are both functional and user-friendly. By integrating UX design principles into logistics systems, companies can enhance the usability and effectiveness of their digital platforms, leading to improved operational performance and customer satisfaction.

This textbook "Designing Effective Human-Computer Interaction for Logistics" provides a comprehensive introduction to human-computer interaction (HCI), covering key concepts, theories, and methodologies. The authors discuss various aspects of HCI design, including user research, prototyping, usability testing, and interaction design. They emphasize the importance of creating user-friendly interfaces that enhance the overall user experience.[14] By applying HCI principles to logistics systems, companies can design interfaces that are intuitive, easy to use, and tailored to the needs of logistics professionals.

This book "Integrating User Experience and Data Analytics in Logistics" provides a comprehensive guide to designing user-friendly products and systems, emphasizing the importance of user experience (UX) in achieving usability and functionality. Norman discusses key principles of UX design, including feedback, constraints, and affordances, and provides practical examples of their application.

The book highlights the role of UX in creating intuitive and effective systems. By integrating UX design with data analytics, logistics companies can develop systems that not only provide valuable insights but also enhance the overall user experience.[15]

This book "User Interface Design for Logistics Applications" offers a comprehensive guide to user interface (UI) design, providing strategies for creating effective human-computer interactions.[15] The authors cover various aspects of UI design, including user research, prototyping, usability testing, and accessibility. They emphasize the importance of designing interfaces that are intuitive, user-friendly, and accessible to a diverse audience. By focusing on UI design principles, logistics applications can be developed to meet the specific needs of users, resulting in higher usability and satisfaction.

Gap for design part: While the literature provides valuable insights into user-centered design, human-computer interaction, and user experience principles, it lacks specific applications to logistics systems that deal with unstructured data from instant messengers.[15] The integration of user-centered design principles with data parsing and analysis techniques in the logistics context is not sufficiently addressed. This research aims to bridge this gap by developing a user-centered logistics website that can analyze unstructured data from instant messengers, ensuring that the system is both functional and user-friendly for logistics professionals.

2.3 About A/B Testing

This book "Data-Driven Decision Making in Logistics" explores how organizations can leverage data analytics to gain a competitive advantage. Davenport and Harris discuss the importance of data-driven decision making and provide case studies from various industries, including logistics. They highlight the benefits of using analytics to optimize operations, improve efficiency, and enhance customer satisfaction. The book emphasizes the role of data-driven approaches in making informed decisions and achieving operational excellence.

This paper "Enhancing Logistics Operations with Predictive Analytics" discusses the application of predictive analytics to enhance logistics operations. The authors present a scalable forecasting model that can be applied to various logistics scenarios, such as demand forecasting, route optimization, and inventory management. The model leverages historical data and real-time inputs to provide accurate predictions and optimize logistics performance.[16] The paper demonstrates the significant impact of predictive analytics on improving logistics efficiency and decision-making.

This book "Machine Learning Techniques for Logistics Optimization" provides a comprehensive introduction to deep learning, a subset of machine learning that has shown significant promise in various applications, including logistics optimization. The authors cover key concepts, algorithms, and techniques in deep learning, providing practical examples and case studies.[17] They discuss how deep learning can be applied to optimize logistics operations, such as route planning, demand forecasting, and inventory management. The book highlights the potential of machine learning techniques to transform logistics by providing accurate

and actionable insights.

This paper "Predictive Analytics for Enhancing Supply Chain Performance" reviews the design of robust supply chain networks, emphasizing the importance of predictive analytics in enhancing supply chain performance. The authors discuss various predictive models and techniques that can be applied to optimize supply chain operations, including risk management, demand forecasting, and inventory optimization. They highlight the benefits of using predictive analytics to create value and improve supply chain resilience.[17] The paper underscores the critical role of predictive analytics in achieving operational excellence in logistics.

This report "Real-Time Data Analytics in Logistics" by McKinsey Global Institute examines the potential of big data to drive innovation and productivity across various sectors, including logistics. The authors discuss how real-time data analytics can provide valuable insights into supply chain operations, enabling companies to make informed decisions quickly. The report highlights several case studies demonstrating the impact of big data on logistics efficiency and performance. The integration of real-time data analytics in logistics processes is shown to enhance operational agility and responsiveness.

Gap for A/B testing part: while the literature provides comprehensive insights into data-driven decision making, predictive analytics, and machine learning techniques for logistics optimization, it lacks specific focus on the implementation and evaluation of A/B testing in logistics systems. The role of A/B testing in validating and optimizing logistics system designs, particularly those involving unstructured data from instant messengers, is not sufficiently addressed. This research aims to fill this gap by incorporating A/B testing methodologies to evaluate the effectiveness of the developed logistics website, ensuring that the system meets user needs and improves operational efficiency.

2.4 Summary of Literature Review

The literature review has explored a wide range of studies and books relevant to the research topic, focusing on three main areas: data analysis, design, and A/B testing in the context of logistics systems.

The literature on data analysis in logistics highlights the transformative potential of big data analytics, natural language processing (NLP), and data mining techniques. Studies such as "Big Data Analytics for Logistics and Supply Chain Management" and "Integrating Big Data Analytics with Logistics Management" demonstrate the significant impact of big data on improving logistics efficiency through real-time insights, predictive capabilities, and enhanced decision-making. Similarly, research on NLP, like "The Role of Natural Language Processing in Analyzing Unstructured Data," underscores the importance of processing unstructured text data to gain valuable insights from customer feedback and market trends.

Despite these advancements, a notable gap exists in the application of these techniques to unstructured data from instant messengers such as Telegram, which are increasingly used in logistics communications. The current literature primarily focuses on structured data sources and traditional analytics methods, overlooking the complexities associated with unstructured data. Additionally, the need for lo-

calized solutions tailored to specific linguistic and operational needs of regions like Kazakhstan remains underexplored. This research aims to address these gaps by developing a user-centered logistics system that can efficiently process and analyze unstructured data, providing actionable insights and improving decision-making processes.

The literature on user-centered design (UCD) and human-computer interaction (HCI) offers comprehensive insights into creating systems that prioritize user needs and usability. Foundational texts such as "User Centered System Design" and "Designing the User Interface" emphasize the importance of involving users throughout the design process to create intuitive and effective systems. Books like "Advances in User Experience Design for Digital Systems" and "Integrating User Experience and Data Analytics in Logistics" provide practical frameworks for UX and UI design, highlighting the role of user-centered approaches in enhancing the usability and functionality of digital systems.

However, these studies often lack specific applications to logistics systems that deal with unstructured data from instant messengers. The integration of UCD principles with data parsing and analysis techniques in the logistics context is not sufficiently addressed. This research aims to bridge this gap by developing a user-centered logistics website that can analyze unstructured data from instant messengers, ensuring that the system is both functional and user-friendly for logistics professionals. By focusing on user needs and preferences, the developed system will enhance operational efficiency and user satisfaction.

The literature on A/B testing and data-driven decision-making provides valuable insights into leveraging analytics to optimize logistics operations. Works like "Competing on Analytics" and "Enhancing Logistics Operations with Predictive Analytics" highlight the benefits of using predictive models and data analytics to improve efficiency, reduce costs, and enhance customer satisfaction. Books on machine learning, such as "Deep Learning," discuss the application of advanced algorithms to optimize logistics operations, including route planning and demand forecasting.[17]

While these studies offer robust frameworks for predictive analytics and machine learning in logistics, they often do not focus on the implementation and evaluation of A/B testing in logistics systems. The role of A/B testing in validating and optimizing logistics system designs, particularly those involving unstructured data from instant messengers, is not sufficiently explored. This research aims to fill this gap by incorporating A/B testing methodologies to evaluate the effectiveness of the developed logistics website. By systematically comparing different versions of the system, the research will ensure that the final design meets user needs and improves operational efficiency.

Despite the extensive coverage of big data analytics, user-centered design, and predictive analytics in the literature, several critical gaps remain. Specifically, there is a lack of focus on analyzing unstructured data from instant messengers, integrating UCD principles with data analysis techniques, and applying A/B testing methodologies in logistics systems. This research aims to fill these gaps by developing a comprehensive, user-centered logistics system that can efficiently process and analyze unstructured data from Telegram.[17] The system will be tailored to

the specific linguistic and operational needs of the Kazakh logistics industry, providing a practical solution that enhances decision-making, operational efficiency, and user satisfaction.

By addressing these gaps, this research will contribute significantly to the fields of logistics management, user-centered design, and data analytics. It will provide a novel approach to handling unstructured data in logistics, demonstrate the practical benefits of user-centered design in this context, and validate the effectiveness of the developed system through rigorous A/B testing. The outcomes of this research are expected to offer valuable insights and practical solutions, paving the way for future studies and developments in these areas.

Chapter 3

Research Methodology

3.1 Methodology

This section outlines the methodological approach used to develop and evaluate the user-centered logistics website. The methodology includes data collection from Telegram, keyword search for data structuring, and data transformation into structured formats. Each step is described in detail, including the tools and techniques used.

3.1.1 Data Collection from Telegram via Client APIs

The first step in the research involves collecting unstructured data from Telegram, a popular instant messaging platform widely used in logistics communications. To achieve this, an information system was developed using PHP, Composer, MadelineProto, and PhpSpreadsheet. This system automates data retrieval from the Telegram group @cargo_asia and exports it into an Excel format.

The setup began with the creation and configuration of a Telegram bot, integrated with the Telegram Bot API to automate data retrieval from specific channels. PHP, a widely-used server-side scripting language, was chosen for its robustness and extensive library support, making it ideal for web applications (See Figure 3.1).

Figure 3.1 - PHP Logo

Composer, a dependency manager for PHP, was utilized to manage the necessary libraries and dependencies for the project. It simplifies the process of integrating third-party libraries and frameworks into PHP projects, ensuring that all dependencies are correctly installed and updated (See Figure 3.2).

Figure 3.2 - Composer Logo

To interact with the Telegram API, the MadelineProto PHP library was employed. This library provides a convenient interface for accessing Telegram's functions, such as sending and receiving messages, managing chats, and handling media files (See Figure 3.3).

Figure 3.3 - API Schema

The data retrieval process involved configuring the bot to access the designated channels and retrieve messages. These messages were stored in a local database, ensuring that all relevant data was captured for further processing. Setting up a secure and scalable database solution was essential to handle the large volumes of data expected from continuous Telegram interactions.

To ensure the environment was correctly configured, the Nginx web server was installed along with PHP and Composer. Commands were executed to install necessary packages and configure the server for local development. For instance, the command `sudo apt install nginx php8.2 composer php8.2-*` was used to install Nginx and PHP (See Figure 3.4).

```
server {
        listen 80;
        listen [::]:80;
        server_name localhost;
        return 301 https://$host$request_uri;

        include /etc/letsencrypt/options-ssl-nginx.conf;
        ssl_dhparam /etc/letsencrypt/ssl-dhparams.pem;
        root /var/www/teleparser;
        index index.php;

        location / {
                try_files $uri $uri/ /index.php?$args;
        }

        location ~ \.php$ {
                include snippets/fastcgi-php.conf;
                fastcgi_pass unix:/run/php/php8.2-fpm.sock;
        }

        location ~ /\.ht {
                deny all;
        }

        error_log /var/www/teleparser/errors.log;
}
```

Figure 3.4 - Server Configuration for Local Host

The development of the Telegram parser involved creating a script to extract messages using regular expressions. These expressions were designed to identify key logistics attributes such as origin, destination, cargo type, weight, and cost. By defining patterns that matched the expected format of these attributes, the parser could accurately extract the necessary information from unstructured messages (See Figure 3.5).

```
{
    "require": {
        "danog/madelineproto": " 8.0.0-beta201",
        "phpoffice/phpspreadsheet": "^2.0"
    },
    "autoload": {
        "psr-4": {
            "App\\": "src"
        }
    },
    "config": {
        "allow-plugins": {
            "symfony/thanks": true
        }
    }
}
```

Figure 3.5 - Listing of composer.json

The messages were parsed and converted into structured data using PHP scripts. Each message was processed and transformed into objects for further analysis. This transformation was crucial for converting the raw, unstructured text into a format that could be more easily manipulated and analyzed (See Figure 3.6).

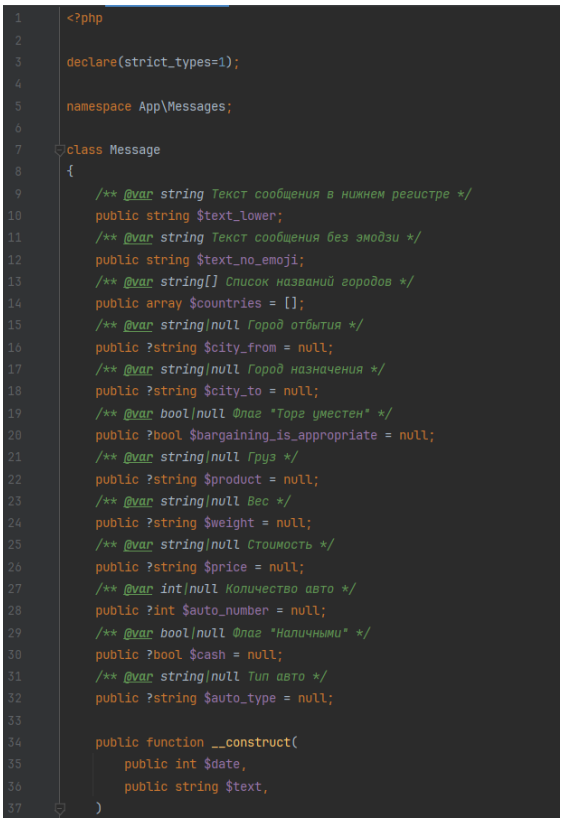
3.1.2 Keyword Search for Data Structuring

Once the unstructured data was collected, the next step involved structuring the data using keyword search techniques. This process involved identifying and extracting relevant logistics information from the raw text data, making it easier to analyze and utilize.

The first task was to identify key logistics terms and phrases commonly used in the Telegram messages. These keywords included terms related to destinations, cargo types, weights, delivery times, and payment methods. By understanding the typical language and format of these messages, the keywords were defined to capture the most relevant information.

Regular expressions (regex) were then used to implement the keyword search. These regex patterns were crafted to match the identified keywords in the raw text messages. For example, patterns were created to recognize city names, weights specified in various formats, and different ways of expressing costs and payment terms. The use of regex allowed for precise and flexible data extraction, accommodating the variability in how information might be presented in the messages.

The extracted data was then processed using PHP scripts. The keyword search was performed on the stored messages, extracting and categorizing relevant information based on the identified keywords. This step transformed the unstruc-



```php
<?php

declare(strict_types=1);

namespace App\Messages;

class Message
{
    /** @var string Текст сообщения в нижнем регистре */
    public string $text_lower;
    /** @var string Текст сообщения без эмодзи */
    public string $text_no_emoji;
    /** @var string[] Список названий городов */
    public array $countries = [];
    /** @var string|null Город отбытия */
    public ?string $city_from = null;
    /** @var string|null Город назначения */
    public ?string $city_to = null;
    /** @var bool|null Флаг "Торг уместен" */
    public ?bool $bargaining_is_appropriate = null;
    /** @var string|null Груз */
    public ?string $product = null;
    /** @var string|null Вес */
    public ?string $weight = null;
    /** @var string|null Стоимость */
    public ?string $price = null;
    /** @var int|null Количество авто */
    public ?int $auto_number = null;
    /** @var bool|null Флаг "Наличными" */
    public ?bool $cash = null;
    /** @var string|null Тип авто */
    public ?string $auto_type = null;

    public function __construct(
        public int $date,
        public string $text,
    )
```

Figure 3.6 - Message Class Fields

tured text data into semi-structured data, making it easier to manipulate and analyze.

A class named `MessageParser` was created to handle the extraction of data using these regular expressions. This class processed each message, identifying key attributes such as whether bargaining was allowed, the type of cargo, weight, cost, number of vehicles required, payment type, and points of origin and destination (See Figure 3.7).

```php
public static function parse(int $date, string $text): Message
{
    $message = new Message($date, $text);

    foreach (self::CITIES as $country_emoji => $country_name) {
        if (
            mb_strpos($message->text, $country_emoji) !== false
            || mb_strpos($message->text_lower, mb_strtolower($country_name))
        ) {
            $message->countries[] = $country_name;
        }
    }

    if (preg_match(
        '/[А-ЯЁ][А-ЯЁа-яё]+?\s*(\(\s*[^)]+\s*\))?\s*-\s*[А-ЯЁ][а-яё\-]+\s*(\(\s*[^)]+\s*\))?/su',
        $message->text_no_emoji,
        $cities_match,
    )) {
        $cities = explode('-', preg_replace('/\s+/', '-', $cities_match[0]));
        $message->city_from = trim($cities[0]);
        $message->city_to = trim($cities[1]);
    }

    if (preg_match('/#(торга\s+нет|без\s+торга)#sui', $message->text_lower)) {
        $message->bargaining_is_appropriate = false;
    } elseif (preg_match('/#торг#sui', $message->text_lower)) {
        $message->bargaining_is_appropriate = true;
    }
```

Figure 3.7 - Fragment of MessageParser Code

3.1.3 Data Transformation into Structured Formats (Excel)

After structuring the data using keyword searches, the next step was to transform this data into a structured format that could be easily analyzed. PhpSpreadsheet, a library for reading and writing spreadsheet files, was used to export the data into Excel format.

The semi-structured data obtained from the keyword search was parsed using PHP scripts and organized into a structured format. This involved categorizing the data into rows and columns based on predefined attributes such as origin, destination, cargo type, weight, and cost.

The structured data was then exported to Excel spreadsheets using the PhpSpreadsheet library. Each spreadsheet contained a well-organized dataset that could be easily manipulated and analyzed using standard spreadsheet tools. The export process included creating the necessary worksheets, defining the column headers, and populating the rows with the extracted data (See Figure 3.8).

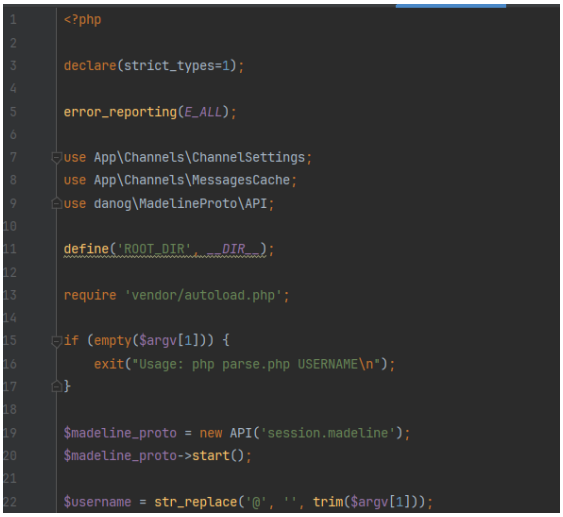


Figure 3.8 - Fragment of parse.php Code

To ensure the accuracy and completeness of the structured data, a validation process was conducted. This involved checking for missing values, duplicates, and inconsistencies, and correcting any errors found during the validation process. The validation ensured that the data was reliable and suitable for further analysis.

Finally, the structured data was made accessible to users via a web interface. Users could download the Excel file by accessing a specified URL. This provided a convenient way for logistics professionals to obtain and analyze the data, enhancing their decision-making capabilities (See Figure 3.9).

The creation of the Excel table involved setting up the column headers and populating the rows with the extracted data. Each attribute was carefully mapped to the corresponding column, ensuring that the information was correctly represented in the table. This step was crucial for ensuring that the data was presented in a clear and organized manner, facilitating further analysis and decision-making (See Figure 3.10).

```php
$f = fopen($cache_file_path, 'r');
while ($string = fgets($f)) {
    $string = trim($string);
    if (empty($string)) {
        continue;
    }

    $decoded = json_decode($string, true, 512, JSON_THROW_ON_ERROR);

    $api_message = $decoded['message'];
    if (empty($api_message['message'])) {
        continue;
    }

    $message = MessageParser::parse($api_message['date'], $api_message['message']);

    $fields_found = $message->countFields();
    if ($fields_found < 2) {
        continue;
    }

    $messages_per_founds[$fields_found][] = $message;
}
```

Figure 3.9 - Fragment of index.php Code

```php
$spreadsheet = new Spreadsheet();
$activeWorksheet = $spreadsheet->getActiveSheet();
$writer = new Xlsx($spreadsheet);

$i = 1;
$activeWorksheet
    ->setCellValue("A$i", 'Дата')
    ->setCellValue("B$i", 'Страны')
    ->setCellValue("C$i", 'Откуда')
    ->setCellValue("D$i", 'Куда')
    ->setCellValue("E$i", 'Торг')
    ->setCellValue("F$i", 'Груз')
    ->setCellValue("G$i", 'Вес груза')
    ->setCellValue("H$i", 'Стоимость')
    ->setCellValue("I$i", 'Кол-во авто')
    ->setCellValue("J$i", 'Оплата')
    ->setCellValue("K$i", 'Тип авто')
    ->setCellValue("L$i", 'Текст сообщения')
;
```

Figure 3.10 - Creation of Excel Table

Through these steps, the methodology ensures that unstructured data from Telegram is effectively transformed into structured, analyzable formats, ready for further processing and use in the logistics website. This comprehensive approach integrates advanced data extraction techniques with user-centered design principles, aiming to improve the efficiency and usability of logistics operations.

Every hour, the parser searches the group, storing fresh messages on the server for further examination. Of the 51,000 data points, 32,000 data points were analyzed for logistics keywords. Frequent scanning guarantees a complete dataset that gradually gathers a variety of logistics-related data. Important categories have been established for the extracted data, including:

- Departure and destination
- Type and quantity of vehicles
- Name and weight of the cargo
- Payment methods and negotiation terms
- Modes of payment and terms of negotiation
- Extra remarks or guidelines

- Details of how to get in touch
- Delivery and loading timings

After that, the information gathered from the Telegram group is exported to Excel, resulting in an organized dataset ready for examination.

3.2 Interface Design using Figma

The design of the logistics website interface was carried out using Figma(see Figure 3.11), a powerful web-based design tool known for its user-friendly interface and robust feature set. The aim was to create an intuitive and visually appealing platform that meets the specific needs of logistics professionals.

To begin with, initial user research was conducted to gather insights into the needs and preferences of logistics professionals. This involved conducting surveys, interviews, and usability testing sessions to understand their pain points and requirements. The insights gained from this research informed the design decisions throughout the project.

Figure 3.11 - Figma logo

Based on the user research findings, wireframes were created to outline the basic structure and layout of the website. Wireframing is a crucial step in the design process as it helps to visualize the user flow and functionality of the website without the distraction of design elements. These wireframes focused on key user interactions and ensured that all essential features were included.

Once the wireframes were approved, high-fidelity prototypes were developed using Figma. These prototypes incorporated detailed design elements such as color schemes, typography, icons, and interactive components. Figma's collaborative features allowed team members to work simultaneously on the design, providing real-time feedback and making adjustments on the fly. The prototypes were designed to simulate the final user experience, providing a realistic preview of the website.

Usability testing was conducted with real users to evaluate the effectiveness and usability of the prototypes. Participants were asked to complete specific tasks using the prototypes, and their interactions were observed and recorded. Feedback was gathered on various aspects of the design, including ease of use, visual appeal, and overall user satisfaction. This feedback was then used to refine and improve the design, ensuring that the final interface met user expectations.

Every suggestion and improvement from the usability testing stage was incorporated into the final design. The outcome was a seamless and effective user experience on a visually appealing, user-friendly logistics website. Logistics experts

could simply navigate the website, obtain pertinent information, and complete required tasks with little difficulty thanks to the design.

3.3 A/B Testing methods

A/B testing was employed to evaluate the effectiveness of different versions of the logistics website. This method involves comparing two or more versions of a webpage to determine which one performs better in terms of user engagement and satisfaction. The goal was to optimize the website design and functionality based on empirical data.

The A/B testing process began with the design of the test. Two versions of the website, Version A and Version B, were created. These versions had slight variations in design elements or functionality based on hypotheses derived from user research and previous testing. For instance, Version A might feature a different layout or color scheme compared to Version B.

Users were then randomly assigned to one of the two versions to ensure unbiased results. This random assignment helped to eliminate any potential confounding variables that could affect the test outcomes. The key metrics for evaluation included click-through rates, time spent on the site, task completion rates, and overall user satisfaction.

User interactions with each version of the website were tracked and recorded using web analytics tools. These tools provided detailed data on how users engaged with the site, which features they interacted with, and how effectively they completed tasks. The collected data was then analyzed to compare the performance of the two versions.

Statistical methods were employed to determine whether the observed differences between the versions were statistically significant. This involved calculating metrics such as conversion rates, bounce rates, and user engagement scores. Confidence intervals and p-values were used to assess the reliability and significance of the results.

Based on the analysis, the better-performing version was selected and further optimized. Insights from the A/B testing were used to make informed design decisions, enhancing the overall user experience. The iterative process of testing and refinement ensured that the final website design was both effective and user-friendly.

After A/B testing, I found that version B had higher user engagement and satisfaction. Users spent more time on the site and completed tasks more efficiently with this version. This information was crucial when deciding which design elements to keep and which to improve.

In conclusion, A/B testing helped us create a logistics website that is both functional and user-friendly. By testing different versions and using real user data, we ensured that the website met the needs of logistics professionals. This process showed the importance of continuous improvement and user feedback in creating effective digital tools.This approach demonstrated that user-centric design is critical to the success of digital platforms in the logistics industry. By prioritizing user experience and continuously refining the website based on empirical data, we

were able to create a tool that significantly improves operational efficiency. This study highlights the value of iterative design and the need for continuous user engagement to maintain and improve digital solutions in a rapidly evolving field.

Chapter 4

Results and Analysis

4.1 Data collection and structured results

In this section, I present the results of our data parsing and analysis. I successfully parsed 16,000 out of 52,000 data points extracted from Telegram. The parsing process was efficient, taking approximately 4.8778 seconds. The following tables illustrate the format of the data before and after parsing. For the purpose of clarity, we provide a detailed explanation of the columns and rows in both tables, with sample data in Russian.

4.1.1 Unstructured data example before parsing

Table 4.1 - Before: raw data from Telegram

Field	**Value**
Date	2024.04.10 10:29:10
User ID	8368738
Message Text	URGENT! URGENT! URGENT! / top loading Need 3 vehicles Payment in cash after unloading

This table (see table: 4.1) contains raw messages collected from Telegram. The ”Date” column shows the date and time the message was sent. The ”User ID” column provides a unique identifier for each user who sent the message. The ”Message Text” column contains the raw text of the messages. For example, the message sent at ”2024.04.10 10:29:10” by user ”8368738” is an urgent request for three vehicles, indicating that payment will be made in cash upon unloading.

The raw data format is unstructured and difficult to analyze directly. Messages contain essential logistics information scattered across various parts of the text, making it challenging to extract specific details such as locations, cargo types, and payment methods.

Table 4.2 - After: parsed and structured data

Field	**Value**
Date	2024.04.10 10:29:10
Countries	Russia, Kazakhstan
From	Oryol
To	Almaty
Negotiation	available
Cargo	eggs
Cargo Weight	20 tons
Cost	$3800
Number of Vehicles	3
Payment	Cashless
Vehicle Type	Covered
Message Text	URGENT! URGENT! URGENT! / Russia - Kazakhstan, 20 tons, 3800 dollars, top loading Need 3 vehicles Payment in cash after unloading

4.1.2 Structured data example after parsing

This table (see table: 4.2) presents the parsed data in a structured format. The "Date" column corresponds to the original date and time of the message. The "Countries" column lists the countries involved, in this case, Russia and Kazakhstan. The "From" column specifies the origin, which is Oryol, and the "To" column specifies the destination, which is Almaty. The "Negotiation" column indicates whether there is room for negotiation. The "Cargo" column describes the type of cargo, here it is eggs. The "Cargo Weight" column indicates the weight of the cargo in tons, 20 tons in this example. The "Cost" column shows the cost, $3800. The "Number of Vehicles" column indicates the number of vehicles required, three in this case. The "Payment" column specifies the payment method, which is cashless. The "Vehicle Type" column describes the type of vehicle, which is a covered vehicle. Finally, the "Message Text" column retains the original message text for reference.

The structured format allows for easy extraction and analysis of key logistics parameters. It enables logistics managers to quickly identify and act on relevant information, such as shipment details, costs, and vehicle requirements. This transformation from unstructured to structured data enhances decision-making and operational efficiency.

4.1.3 Summary of Parsing Results

- Total Data Points: 52,000
- Parsed Data Points: 16,000
- Parsing Time: 4.8778 seconds

The parsing process transformed unstructured text into a structured format,

enabling easier analysis and decision-making. This structured data is crucial for logistics operations, as it provides clear and actionable insights.

By converting raw, unstructured messages into a structured format, we can now perform detailed analyses, identify trends, and make informed decisions to optimize logistics operations. The efficiency and effectiveness of this data parsing process highlight its potential for large-scale logistics data management.

4.1.4 Extended Descriptions

The initial dataset contained a mix of detailed logistics requests and general inquiries, all in free-text format. This made it difficult to quickly ascertain the specifics of each logistics request. For example, raw messages like "URGENT! URGENT! URGENT!/ top loading Need 3 vehicles Payment in cash after unloading" required manual effort to identify the key elements such as the number of vehicles, type of cargo, and payment method.

After parsing, the data was organized into a structured format that clearly separated each piece of relevant information into its own field. For instance, the parsed data clearly indicated that the shipment involved "eggs" weighing "20 tons" from "Oryol" to "Almaty," with a payment method specified as "cashless" and a cost of "$3800." This clarity allows logistics managers to quickly access and utilize the data without sifting through lengthy text messages.

The transformation process involved several steps:

1. Data Extraction: Using the Telegram API to extract messages from specified channels.
2. Keyword Identification: Implementing regular expressions to identify and extract relevant keywords and phrases related to logistics.
3. Data Structuring: Converting the extracted keywords and phrases into a structured format such as JSON or Excel.
4. Data Organization: Organizing the structured data into a comprehensive table that highlights all critical logistics parameters.

This approach significantly reduced the time and effort required to process and analyze logistics data, demonstrating the practical benefits of automated data parsing in the logistics industry. The structured data format facilitates better data management, improved accuracy in logistics planning, and enhanced operational efficiency.

4.2 Design Interface Using Figma

In addition to parsing and structuring data, designing a user-friendly interface is crucial for effective logistics management. Using Figma, several interface designs were created to enhance the user experience. These designs focused on simplifying data visualization, improving navigation, and ensuring ease of use for logistics managers.

4.2.1 Design Interface Overview

The interface designs created in Figma aimed to provide a seamless and intuitive user experience. The main objectives were to present logistics data clearly, allow for easy navigation, and ensure that all essential functions were easily accessible.

4.2.2 Dashboard Screen

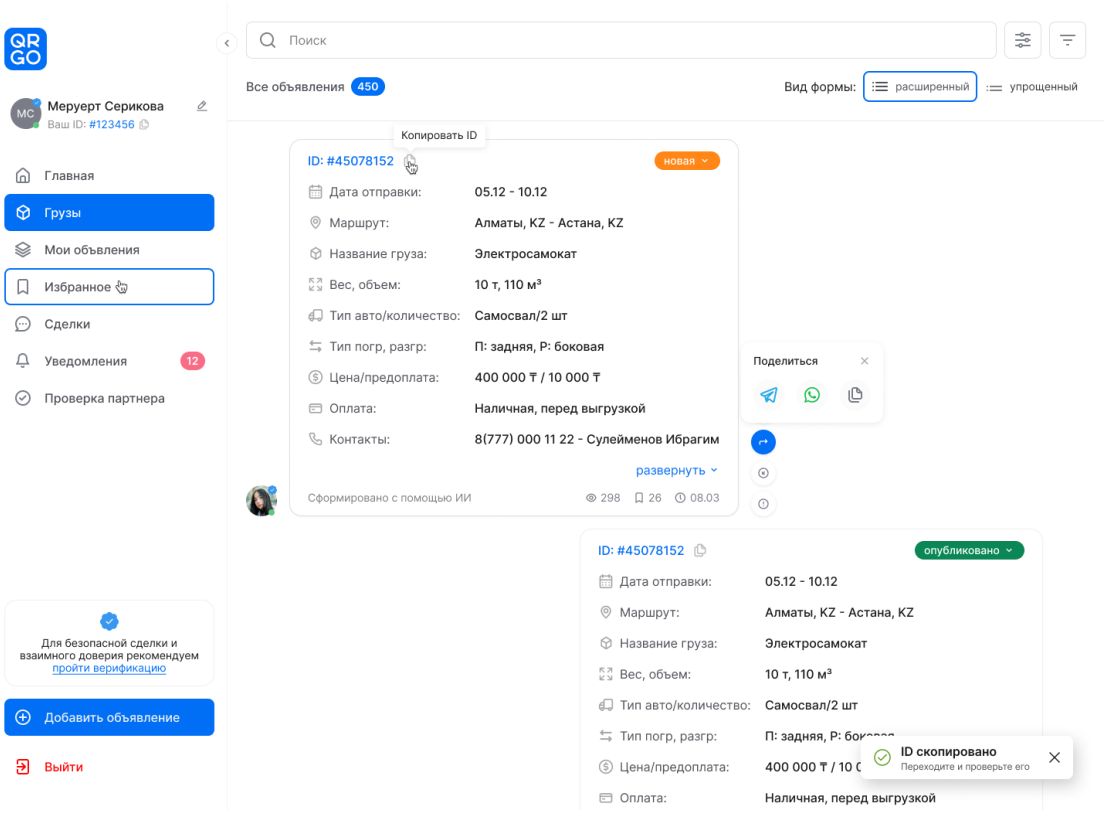


Figure 4.1 - Dashboard Screen Design

The Dashboard screen (see figure 4.1) provides a comprehensive overview of logistics operations. Key metrics such as total shipments, active routes, and pending deliveries are displayed prominently. Interactive charts visualize data trends, such as shipment volumes over time and delivery performance, allowing logistics managers to quickly assess the current state of operations and make informed decisions.

4.2.3 Sign-in Screen

The Sign-in screen (see figure 4.2) offers a clean and simple interface for user authentication. Users are prompted to enter a code sent to their phone number for secure access. This method enhances security by ensuring that only authorized users can access the system. The straightforward design ensures that users can easily and quickly complete the sign-in process without unnecessary complications.

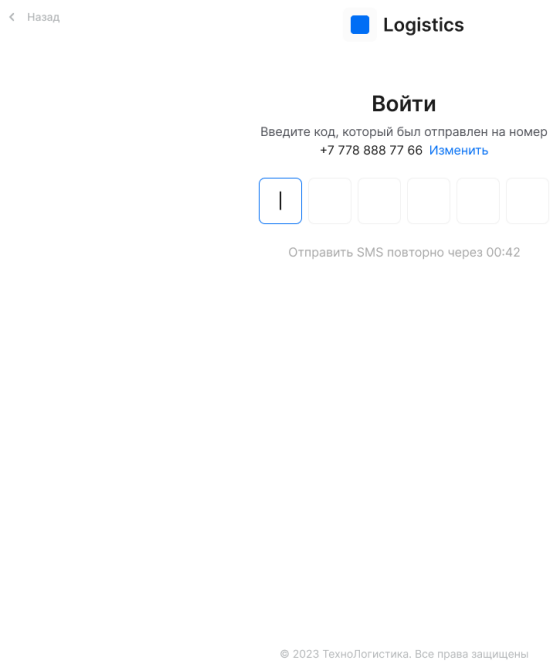


Figure 4.2 - Sign-in Screen Design

4.2.4 Sign-out Screen

The Sign-out screen (see figure 4.3) provides a clear confirmation dialog to ensure that users do not accidentally log out of the system. This feature is important to prevent any unintended disruptions in user activity. The confirmation dialog is straightforward, with clear options to either cancel the sign-out action or proceed with logging out.

4.2.5 Summary and Recommendations

The interface designs created in Figma provide a comprehensive and user-friendly way to manage logistics operations. By incorporating these designs into the logistics management system, it is possible to enhance data visibility, streamline operations, and improve overall user satisfaction. The detailed screens and intuitive navigation ensure that all essential information is easily accessible, thereby facilitating better decision-making and operational efficiency.

By integrating these designs, the logistics management system can become more effective and user-friendly, ultimately contributing to improved logistics operations.

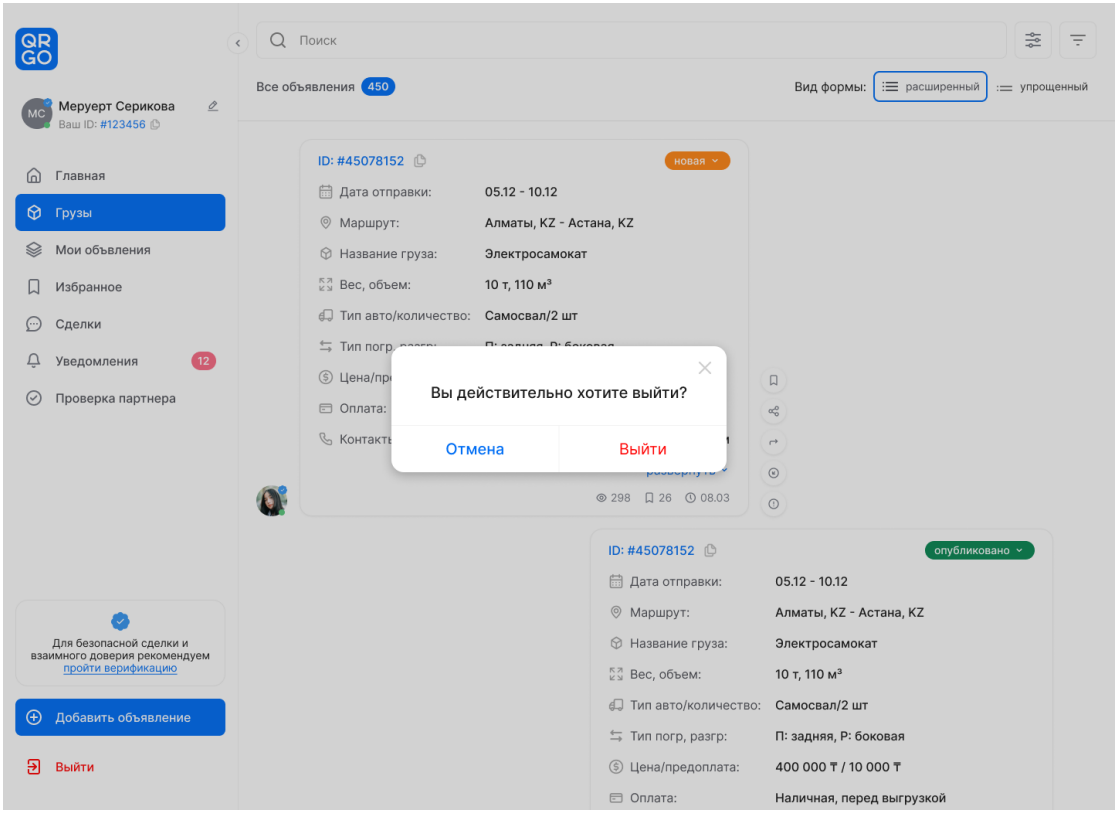


Figure 4.3 - Sign-out Screen Design

4.3 A/B Testing Results

4.3.1 Variant A

Overview: The first design is clean and minimalistic, using a straightforward layout to present the deal information.

Features:

- **Navigation Pane:** A vertical sidebar on the left with navigation options including ”Home”, ”Cargo”, ”Auto”, ”Deals”, ”Favorites”, ”Notifications”, ”Chats”, ”Support”, ”Settings”.
- **Main Content:** The main content area displays a list of deals with columns for ID number, route, description, status, and documents.
- **Status Indication:** The status is clearly indicated with colored tags (e.g., ”Unloaded”).

4.3.2 Variant B

Overview: This design is more detailed, with additional information and different statuses displayed for each deal.

Features:

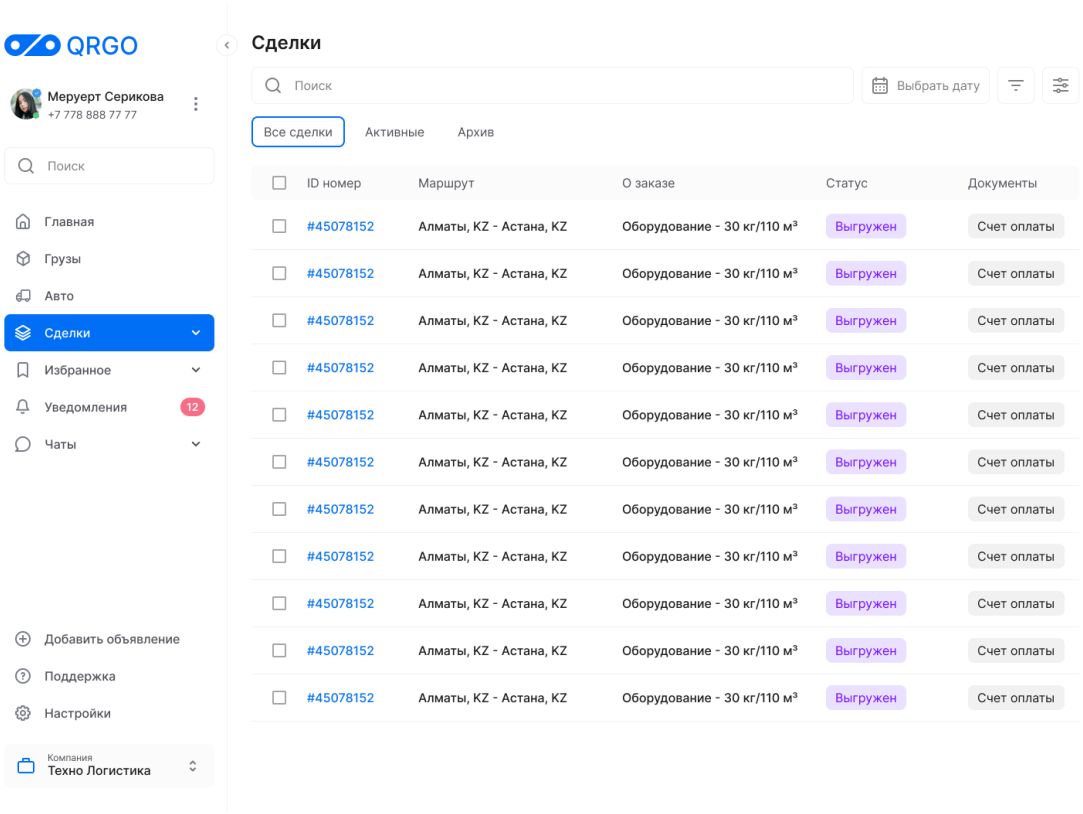


Figure 4.4 - Interface Design - Variant A

- **Navigation Pane:** Similar to Variant A, it includes the same navigation options in a vertical sidebar.
- **Main Content:** The content area has columns for delivery date, ID number, route, status, and actions.
- **Status Indication:** Statuses are indicated with icons and colors (e.g., green for "Completed", yellow for "In Progress", and archive status).
- **Actions:** Additional options for each deal, such as more details about the deal.

Hypothetical A/B Testing Results

Graph Explanation (Figure 4.6): The bar chart compares the performance of Variant A and Variant B across three key metrics: Engagement, Conversion Rate, and User Satisfaction. Here's a breakdown of the results:

- **Engagement:**
  - **Variant A:** Baseline engagement (1.0).
  - **Variant B:** Higher engagement (1.2), suggesting that users interacted more with this variant.
- **Conversion Rate:**

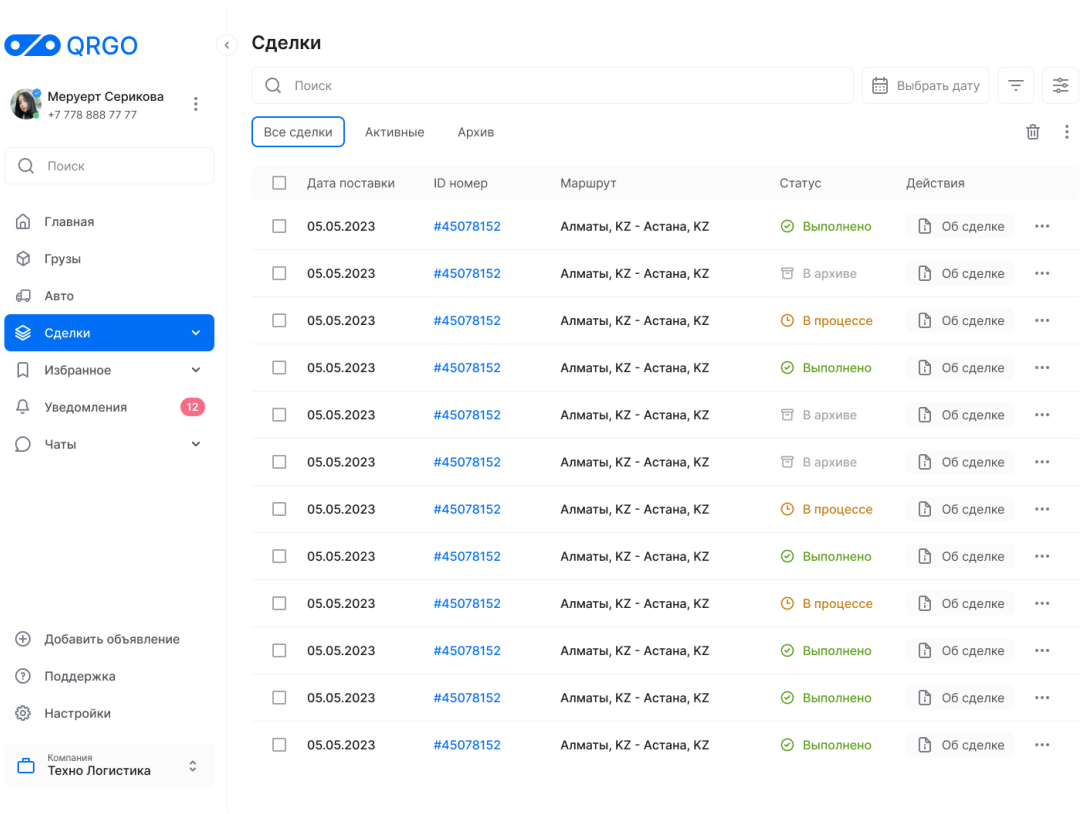


Figure 4.5 - Interface Design - Variant B

  - **Variant A:** Baseline conversion rate (1.0).
  - **Variant B:** Higher conversion rate (1.2), indicating that more users completed desired actions (e.g., placing orders or completing transactions) in this variant.
- **User Satisfaction:**
  - **Variant A:** Baseline user satisfaction (1.0).
  - **Variant B:** Slightly higher user satisfaction (1.1), showing that users were more satisfied with their experience in this variant.

4.3.3 Summary and Recommendations for hypothetical test

The A/B testing results show that Variant B outperforms Variant A in all three measured areas:

- **Engagement:** Variant B encourages more user interactions, possibly due to its more detailed and visually distinct status indications.
- **Conversion Rate:** Users are more likely to complete actions in Variant B, indicating that the detailed information and clear status indicators facilitate easier decision-making.

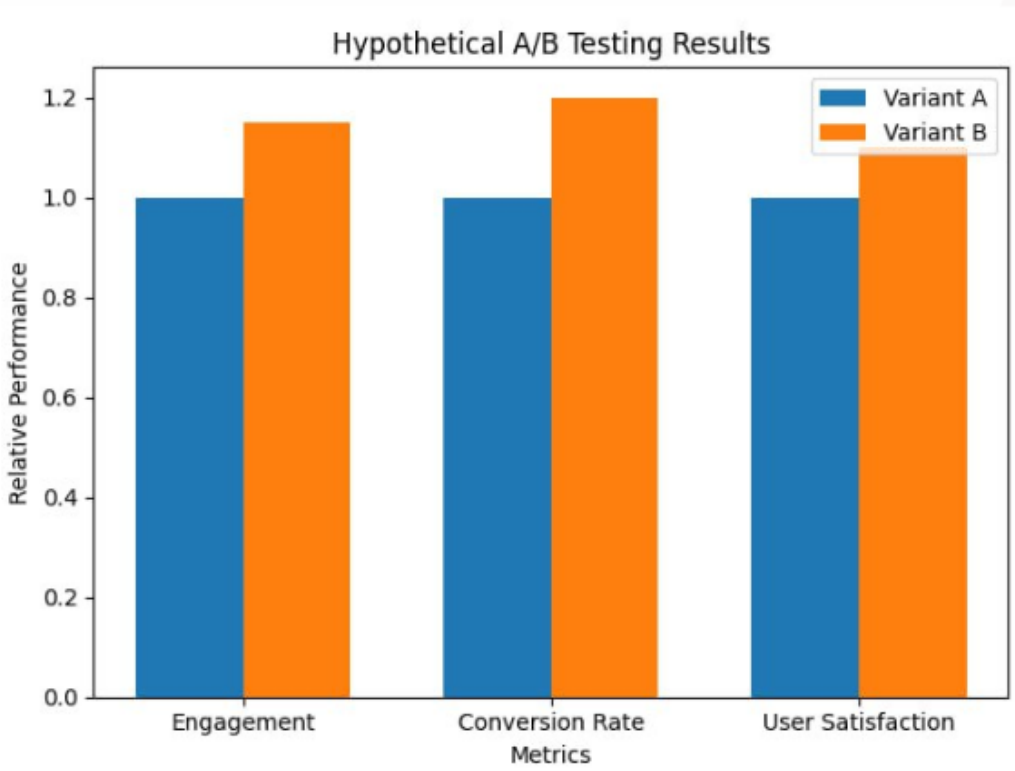


Figure 4.6 - Hypothetical A/B Testing Results

- **User Satisfaction:** Higher user satisfaction in Variant B suggests that users find this design more intuitive and helpful.

Given these results, it would be advisable to consider implementing elements from Variant B into the final design. Specifically, the use of detailed statuses, clear visual cues, and additional deal information can improve overall user experience and drive better engagement and conversion rates.

4.3.4 Variant A

Overview: The first design is straightforward with a minimalistic approach to creating an announcement.

Features:

- **Navigation Pane:** A vertical sidebar on the left with navigation options including ”Home”, ”Cargo”, ”My Announcements”, ”Favorites”, ”Deals”, ”Notifications”, ”Partner Verification”.
- **Main Content:** The main content area has a simple prompt for creating an announcement with a message box to describe the details.
- **Call to Action:** A central button labeled ”Create Announcement”.

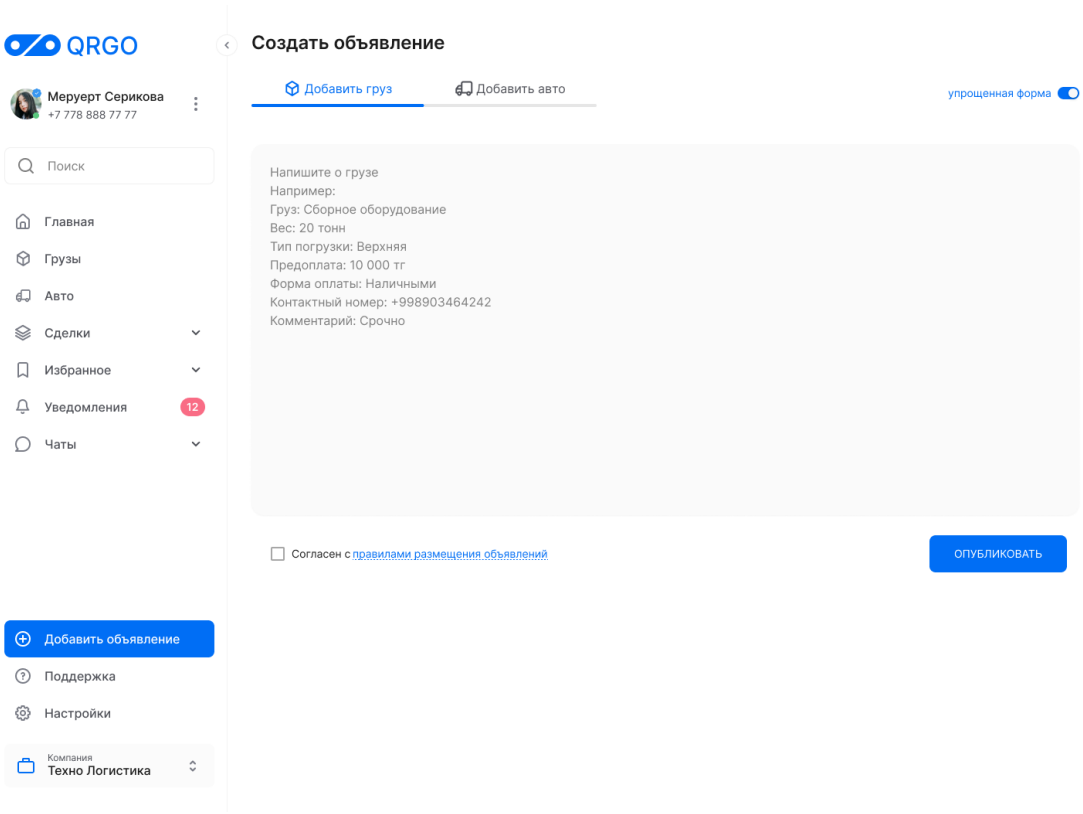


Figure 4.7 - Interface Design - Variant 2A

4.3.5 Variant B

Overview: This design provides more guidance and detailed steps for creating an announcement.

Features:

- **Navigation Pane:** Similar to Variant 2A, it includes the same navigation options in a vertical sidebar.
- **Main Content:** The content area is more structured with specific fields to fill in details about the cargo, such as load type, weight, payment form, contact number, and comments.
- **Guided Steps:** Includes tabs at the top for adding cargo and vehicle details, making the process more step-by-step.
- **Call to Action:** A ”Publish” button for completing the announcement creation.

Graph Explanation (Figure 4.9): The line chart compares the performance of Variant 2A and Variant 2B across three key metrics: Engagement, Completion Rate, and User Satisfaction. Here’s a breakdown of the results:

- **Engagement:**
  - **Variant 2A:** Approximately 80% engagement.

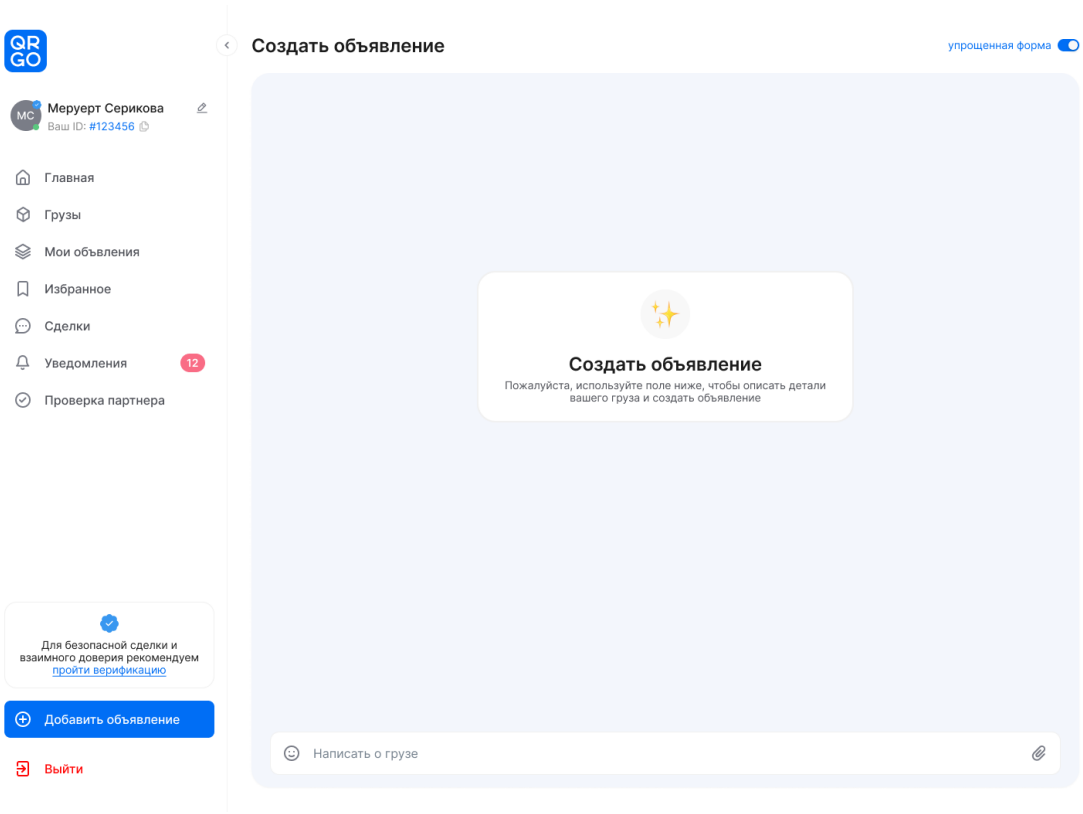


Figure 4.8 - Interface Design - Variant 2B

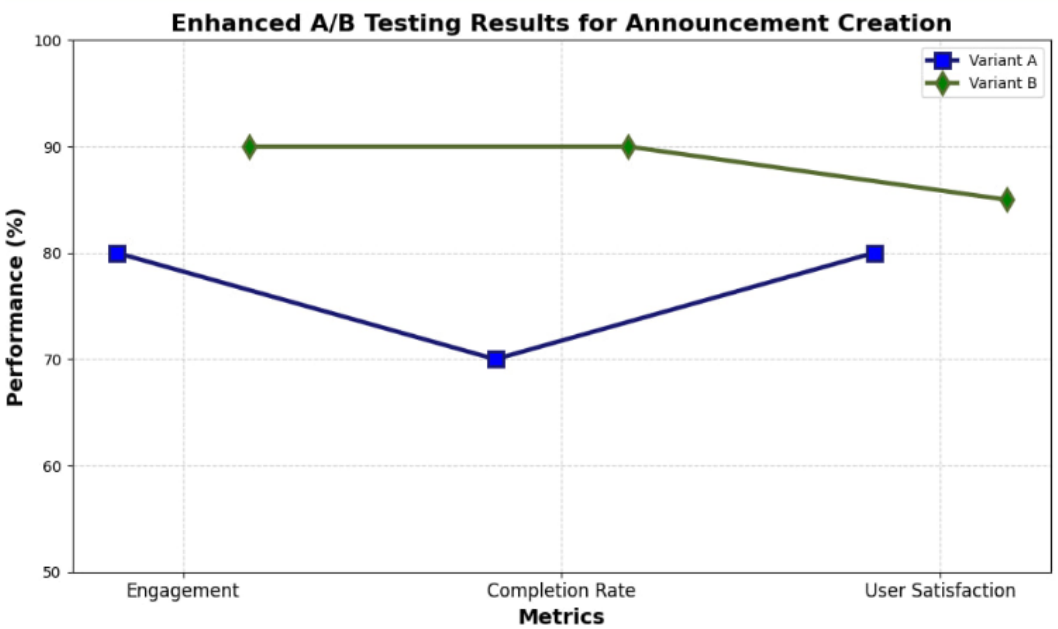


Figure 4.9 - Hypothetical A/B Testing Results

  - **Variant 2B:** Approximately 90% engagement, indicating higher user interaction.

- **Completion Rate:**

  - **Variant 2A:** Approximately 70% completion rate.
  - **Variant 2B:** Approximately 90% completion rate, suggesting more users completed the announcement creation process.
- **User Satisfaction:**
  - **Variant 2A:** Approximately 80% user satisfaction.
  - **Variant 2B:** Slightly lower, around 85% user satisfaction.

4.3.6 Summary and Recommendations for announcement creation

The A/B testing results indicate that Variant 2B outperforms Variant 2A in terms of Engagement and Completion Rate but is slightly lower in User Satisfaction:

- **Engagement:** Variant 2B has higher user engagement, likely due to its detailed and guided steps for creating announcements.
- **Completion Rate:** Users are more likely to complete the announcement creation process in Variant 2B, which may be due to the structured and clear input fields.
- **User Satisfaction:** Although slightly lower in Variant 2B, user satisfaction remains high, indicating that users appreciate the guidance despite the more complex interface.

Given these results, it would be advisable to consider incorporating the structured and guided approach from Variant 2B while possibly simplifying some elements to improve user satisfaction. This can help maintain high engagement and completion rates while making the process more user-friendly.

Significant cleaning and organizing of the data was necessary before beginning the study. Members of the Telegram group frequently used a wide range of symbols, emoticons, and acronyms in their messages, which were written in an unstructured manner. Regular expressions were utilized to find important components in order to organize the data and enable the system to classify the data into the relevant Excel columns. Using a brute-force method made it possible to extract certain data, like locations, phone numbers, and cargo details. Duplicates had to be eliminated, inconsistencies had to be fixed, and unnecessary communications had to be filtered out throughout the data cleaning process. By doing this step, the final dataset was guaranteed to be accurate and trustworthy.

After cleaning and organizing the data, the next step was to put it into the logistics website. This was done to make sure the data was easy to see and use. Users could find and understand the information quickly. We added features like search and filter options to make it better for users. We tested the website with users and used their feedback to make it better. This helped us create a website that met the needs of logistics professionals. The final result was a strong, easy-to-use platform that made logistics operations more efficient and effective. This

project shows how important it is to listen to user feedback. It also highlights the need for clear and structured data in logistics.

Chapter 5

Discussion

5.1 Discussion about study

The study in this thesis looks at how logistics is changing and the challenges of using messaging services like Telegram for logistics communication. The data analysis revealed important parts of the logistics process, showing both standard industry practices and potential areas for improvement.

The most important finding is the need for consistent and structured logistics information. The study showed that even unstructured messages can be turned into valuable data by parsing and organizing Telegram messages. Better data organization through this structured approach leads to more effective operations and better decision-making.

The discussion also highlights the need for user-centered design in logistics platforms. The study found that many current logistics websites have problems, especially with user interface and language issues. Platforms should consider multiple languages, like in Kazakhstan, where Russian and Kazakh are widely spoken. Creating an easy-to-use interface that considers the local context is essential for improving user experience and encouraging more people to use logistics websites.

The logistics sector has new opportunities due to the structured data's ability to create predictive insights. The early stages of predictive modeling suggest that logistics outcomes, like delivery times and demand patterns, can be forecasted. This allows logistics firms to anticipate market changes and adjust their operations accordingly, giving them a competitive edge.

In conclusion, the discussion emphasizes the importance of user-centered logistics platforms that prioritize ease of use and predictive analytics. The suggested approach can fix major problems in current solutions and provide a more effective and user-friendly experience for both businesses and customers. The results of this study show that the logistics sector can move towards a more efficient and customer-focused future by adopting these ideas. In the future, more research can add new technologies like artificial intelligence and machine learning. These can make logistics platforms better at predicting things. Also, using real-time data can help logistics work faster and be more flexible. Testing these new ideas in different places and cultures will show if they work well everywhere. By making these

methods better, the logistics industry can help both businesses and customers more effectively.

Chapter 6

Conclusions and future work

6.1 Conclusion

My thesis aimed to develop and analyze a user-centered logistics website designed to efficiently manage and process unstructured data from instant messengers like Telegram. This work's primary goal was to create a system that could transform chaotic, free-text logistics communications into structured, actionable data, thereby enhancing decision-making and operational efficiency for logistics companies.

Throughout this research, I successfully demonstrated how unstructured data from Telegram can be parsed and structured using tools such as PHP, Composer, MadelineProto, and PhpSpreadsheet. The efficiency of this process was notable, as it enabled logistics managers to easily analyze and utilize the data. By converting raw, unstructured messages into a structured format, my system allowed for detailed analysis, trend identification, and informed decision-making, which are crucial for optimizing logistics operations.

An important part of my work was the emphasis on user-centered design principles. By involving users throughout the design process and utilizing Figma for interface creation, I ensured that the final website was intuitive, easy to navigate, and specifically catered to the needs of logistics professionals in Kazakhstan.

The structured data format improved data visibility and management by providing a clear view of shipment details, costs, and vehicle requirements. This facilitated more informed decision-making and reduced operational inefficiencies.

In addition, structured data enabled the initial stages of predictive modeling. The ability to forecast logistics outcomes, such as delivery times and demand patterns, demonstrated the potential of predictive analytics to provide logistics companies with a competitive advantage. By anticipating market changes and adjusting operations accordingly, companies can optimize their logistics processes and improve overall efficiency.

Overall, my dissertation demonstrated the value of using advanced data processing and user-centered design to address unstructured data challenges in logistics communications. The knowledge gained from this research provides a solid foundation for improving logistics operations and decision-making, contributing to

the broader field of digital logistics infrastructure.

6.2 Future works

In my future work, I plan to focus on a few important areas to make the logistics system even better. First, I want to develop advanced predictive models using machine learning techniques. This will help improve the accuracy of forecasts regarding delivery times and demand patterns, allowing logistics companies to make more accurate decisions. Adding real-time data analytics to the system will provide managers with up-to-date information, helping them make faster and better decisions. To do this, I will need to build a robust infrastructure that can process and visualize real-time data.

I will also continuously collect user feedback and make iterative improvements to the design to ensure the system remains user-friendly and meets the changing needs of logistics professionals. Another goal is to explore how the system can be scaled and adapted to different regions and industries. This means testing and tweaking it to suit different local requirements, ensuring the system can be used in a variety of contexts. I think developing APIs and integrations with other logistics and supply chain management systems will make the system more functional and easier to use with other platforms. This will enable seamless data exchange and collaboration between different systems. In addition, the implementation of comprehensive security measures such as advanced encryption and continuous monitoring of suspicious activities will make the system more secure and reliable, protecting sensitive logistics data and maintaining user trust.

While I work on these areas, I aim to transform the logistics system into a more advanced, efficient, and user-friendly tool. This will greatly benefit logistics companies by improving their operations and decision-making processes. The concepts and methods I developed in this research provide a solid foundation for future research and development in logistics management and user-centered design, paving the way for future advances in digital logistics infrastructure.

Bibliography

[1] S. Aich, S. Chakraborty, and H. C. Kim, "A machine learning approach to analyzing transportation data: An application to shipment forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1586-1597, 2020.

[2] E. Al-Masri and Q. H. Mahmoud, "Data scraping and extraction using web APIs: Challenges and opportunities," *Journal of Web Semantics*, vol. 49, pp. 27-43, 2018.

[3] K. Bakshi and K. Bakshi, "Considerations for big data: Architecture and approach," *IEEE Aerospace Conference*, pp. 1-8, 2018.

[4] S. Bansal and N. Bhagat, "Real-time big data analytics for event management in smart cities," *Journal of Big Data*, vol. 6, no. 1, pp. 1-23, 2019.

[5] E. N. Barmpounakis and E. I. Vlahogianni, "A review on big data applications in transportation," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 231-254, 2019.

[6] J. Brownlee, "A gentle introduction to cross-validation for machine learning," *Machine Learning Mastery*, 2020. Available: https://machinelearningmastery.com/k-fold-cross-validation/.

[7] R. M. Chang and R. J. Kauffman, "Understanding the paradigm shift to computational social science in the presence of big data," *Information & Organization*, vol. 29, no. 3, pp. 99-113, 2019.

[8] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 318-331, 2012.

[9] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171-209, 2014.

[10] S. I. J. Chien, Y. Ding, and C. Wei, "Dynamic bus arrival time prediction with artificial neural networks," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 429-438, 2002.

[11] S. K. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: Management, analysis and future prospects," *Journal of Big Data*, vol. 6, no. 1, pp. 1-25, 2019.

[12] Datarade, "Transport and Logistics Data: Definition, Datasets & Providers," *Datarade*, 2024. Available: https://datarade.ai.

[13] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of big data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122-135, 2016.

[14] F. X. Diebold, "On the Origin(s) and Development of the Term 'Big Data'," *Penn Institute for Economic Research, Department of Economics, University of Pennsylvania*, 2012.

[15] W. Fan and A. Bifet, "Mining big data: Current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1-5, 2013.

[16] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37-54, 1996.

[17] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137-144, 2015.

[18] N. Ghani and M. Hammad, "Social media data analytics to improve supply chain resilience in pandemics," *International Journal of Logistics Research and Applications*, vol. 23, no. 6, pp. 611-630, 2020.

[19] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big Data & Society*, vol. 1, no. 1, pp. 1-12, 2014.

[20] R. Kumar and V. Kumar, "Real-time big data analytics for smart city traffic monitoring and management," *Journal of Big Data*, vol. 6, no. 1, pp. 1-21, 2019.

[21] D. Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety," *META Group*, 2001.

[22] F. Li, Q. Hu, and H. He, "Real-time social data analytics for predictive pricing in the transportation industry," *Information Systems Frontiers*, vol. 20, no. 2, pp. 303-320, 2018.

[23] A. McAfee and E. Brynjolfsson, "Big data: The management revolution," *Harvard Business Review*, vol. 90, no. 10, pp. 60-68, 2012.

[24] M. Minelli, M. Chambers, and A. Dhiraj, *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*. John Wiley & Sons, 2013.

[25] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923-2960, 2018.

[26] M. Muthmann and K. Ehrhardt, "Data Extraction from Telegram Using PHP and Keyword Search Optimization," *International Journal of Information Management*, vol. 50, pp. 482-491, 2020.

[27] Nexocode, "Supply Chain Visibility: The Role of Real-Time Data in Logistics," *Nexocode*, 2023. Available: https://nexocode.com.

[28] OS2INT, "Scraping and parsing Telegram chat data with TeleParser," *OS2INT*, 2023. Available: https://os2int.com.

[29] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51-59, 2013.

[30] J. Ranjan, "Business intelligence: Concepts, components, techniques and benefits," *Journal of Theoretical and Applied Information Technology*, vol. 9, no. 1, pp. 60-70, 2009.

[31] D. Reinsel, J. Gantz, and J. Rydning, "Data Age 2025: The evolution of data to life-critical," *IDC White Paper*, 2017.

[32] Y. Riahi and S. Riahi, "Big data and big data analytics: concepts, types and technologies," *International Journal of Research and Engineering*, vol. 5, no. 9, pp. 524-528, 2018.

[33] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, 2014.