

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Уфимский государственный нефтяной технический университет»**

при поддержке:
Российской академии естественных наук
Академии наук Республики Башкортостан
Общественной организации
«Профессионалы дистанционного обучения»
Ассоциации образовательных программ
«Электронное образование Республики Башкортостан»
Российского союза научных и инженерных
общественных объединений
Партнерского центра
международного сертификационного холдинга IMQ

Информационные технологии Проблемы и решения

У ф а
УНПЦ "Издательство УГНТУ"
2 0 2 2

Информационные технологии. Проблемы и решения. – Уфа: УНПЦ "Издательство УГНТУ", 2022. 2(19). 112 с.

Information technology. – Ufa: UNPC "USPTU Publishers", 2022. 2(19). 112 p.

Учредитель:

**ФГБОУ ВО Уфимский государственный
нефтяной технический университет**

2022, 2(19)

Издается с 2014 г.

РЕДКОЛЛЕГИЯ

Главный редактор

Р.Н. Бахтизин, первый проректор Уфимского государственного нефтяного технического университета, д-р физ.-мат. наук, профессор

Члены редколлегии

Ю.Н. Белоножкин, канд. экон. наук, доцент кафедры финансы и кредит Сочинского государственного университета

Й. Дарадке, доцент, заместитель декана факультета вычислительной техники и сетей Университета принца Саттама бин Абдулазиза (PSAU) - Королевство Саудовская Аравия (KSA)

Ф.У. Еникеев, д-р техн. наук, профессор кафедры вычислительной техники и инженерной кибернетики Уфимского государственного нефтяного технического университета

В.В. Ерофеев, д-р техн. наук, профессор, руководитель Челябинского регионального отделения РАЕН

Н.В. Корнеев, д-р техн. наук, профессор кафедры управления безопасностью сложных систем Губкинского университета, член-корр. РАЕН

И.М. Михайловская, ст. преподаватель кафедры вычислительной техники и инженерной кибернетики Уфимского государственного нефтяного технического университета

Е.А. Султанова, канд. техн. наук, доцент кафедры вычислительной техники и инженерной кибернетики Уфимского государственного нефтяного технического университета, член-корр. РАЕН

В.Н. Филиппов, канд. техн. наук, доцент кафедры вычислительной техники и инженерной кибернетики Уфимского государственного нефтяного технического университета, действительный член РАЕН

© ФГБОУ ВО «Уфимский государственный нефтяной технический университет», 2022

© Коллектив авторов, 2022

Полнотекстовая версия выпуска размещена в Научной электронной библиотеке elibrary.ru по ссылке:

https://elibrary.ru/title_about.asp?id=61250

Подробности на сайте: <http://vtik.net>

Отпечатано с готового электронного файла.

Подписано в печать 15.05.2022. Формат 60x80/16. Гарнитура «Таймс». Усл. печ. л. 6,51. Тираж 800 экз. Заказ 149.

Учебный научно-производственный центр "Издательство Уфимского государственного нефтяного технического университета"

Адрес учебного научно-производственного центра "Издательство Уфимского государственного нефтяного технического университета": 450064, Российская Федерация, Республика Башкортостан, г. Уфа, ул. Космонавтов, 1.

Founder:

**FSBEU NE Ufa State Petroleum
Technological University**

2022, 2(19)

Published since 2014

EDITORIAL BOARD

Editor-in-Chief

R.N. Bakhtizin, Dr. of Physical and Mathematical Sci., Professor, First Vice-Rector of Ufa State Petroleum Technological University

Editorial Board Members:

Yu. N. Belonozhkin, PhD Economic Sci. Department of Finance and Credit Sochi State university

Dr. Yousef Daradkeh, Associate Professor and Assistant Dean for Administrative Affairs, Department of Computer Engineering and Networks, Prince Sattam bin Abdulaziz University (PSAU) - Kingdom of Saudi Arabia (KSA)

F.U. Enikeev, Dr. of Technical Sci., Professor of Department of Computer Science and Engineering Cybernetics Ufa State Petroleum Technological University

V.V. Yerofeyev, Dr. Sci. Professor, Head of the Chelyabinsk regional branch of RANS

N.V. Korneev, Dr. Tech. Sci., Professor of the Department of Safety Management of Complex Systems, Gubkin University, Corresponding Member of the Russian Academy of Natural Sciences.

I.M. Mikhaylovskaya, Senior Lecturer of Department of Computer Engineering and Engineering Cybernetics Ufa State Petroleum Technological University

E.A. Sultanova, PhD, Deputy Head of Department of Computer science and Engineering cybernetics Ufa State Petroleum Technological University, corresponding member RANS

V.N. Filippov, PhD, Deputy Head of Department of Computer science and Engineering cybernetics of Ufa State Petroleum Technological University, Full member of the RANS

ОГЛАВЛЕНИЕ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В НАУКЕ, ОБРАЗОВАНИИ И ПРОИЗВОДСТВЕ

| | |
|--|----|
| Хаматянов М.И., Медведева О.А. АВТОМАТИЗИРОВАННАЯ СИСТЕМА ДЛЯ ЭЛЕКТРОННОЙ ПОДАЧИ ДОКУМЕНТОВ ПО ЗАДАННОМУ МАКЕТУ..... | 5 |
| Астаева Т.О., Зайцева Е.В. РАЗРАБОТКА ТЕХНОЛОГИИ АНИМАЦИИ ТРЕХМЕРНОЙ МОДЕЛИ В ЦЕЛЯХ ПОВЫШЕНИЯ КАЧЕСТВА ОБУЧЕНИЯ ХОРЕОГРАФИИ..... | 9 |
| Бердюгин А.В. ОПИСАНИЕ СХЕМ ИМЕНОВАННЫХ СТРУКТУРНЫХ ЕДИНИЦ ДЛЯ ИНФОЛОГИЧЕСКОЙ МОДЕЛИ..... | 15 |
| Агапов Ю.Г., Подвесовская М.А. СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК ЯЗЫКА RUTRON ДЛЯ ИЗУЧЕНИЯ МЕТОДОВ ВИЗУАЛИЗАЦИИ ДАННЫХ.... | 20 |
| Попова В.А. ПРИМЕНЕНИЕ МЕТОДИКИ СТАТИЧЕСКОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ОШИБОК В ПРОГРАММАХ НА ДИНАМИЧЕСКИХ ЯЗЫКАХ ПРОГРАММИРОВАНИЯ..... | 26 |
| Снежко Ю.А., Огилько В.М. ПОДХОД К ВОССТАНОВЛЕНИЮ ГЕОФИЗИЧЕСКИХ ДАННЫХ МЕТОДОМ РЕЗЕРВИРОВАНИЯ..... | 32 |
| Хлыбов А.В., Гиниятуллин В.М., Ермолаев Е.В. СОКРАЩЕНИЕ СВЕРТОЧНЫХ СЛОЕВ..... | 37 |
| Шайхутдинов Р.Э., Масгутова И.С. РАЗРАБОТКА ЧАТ-БОТА ДЛЯ ОБЕСПЕЧЕНИЯ НЕПРЕРЫВНОЙ РАБОТЫ ЕДИНОЙ ЦИФРОВОЙ ПЛАТФОРМЫ..... | 42 |
| Фахретдинов А.Р., Султанова Е.А., Михайловская И.М. РАЗРАБОТКА ИНФОРМАЦИОННО-ПРОГРАММНОЙ МОДЕЛИ СПЕЦИАЛИЗИРОВАННОГО ПОТОКА БЕТОННЫХ РАБОТ ДЛЯ ОБЕСПЕЧЕНИЯ НЕПРЕРЫВНОСТИ ПРОИЗВОДСТВЕННОГО ЦИКЛА..... | 46 |

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ, УПРАВЛЕНИИ И БИЗНЕСЕ

| | |
|---|----|
| Петлина Е.М. ОСНОВНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ УПРАВЛЕНИЯ..... | 57 |
| Головина Е.Ю., Фархутдинова А.И. РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ПРОВЕРКИ ЭЛЕКТРОННЫХ ПИСЕМ ДЛЯ ПОЧТОВОГО СЕРВЕРА ПРЕДПРИЯТИЯ..... | 62 |

МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ

| | |
|---|----|
| Журавлев Н.А., Фомичева О.Е. БИОМЕТРИЧЕСКАЯ СИСТЕМА КОНТРОЛЯ И УПРАВЛЕНИЯ ДОСТУПОМ НА ОСНОВЕ НЕЙРОННОЙ СЕТИ | 68 |
| Хлыбов А.В., Гиниятуллин В.М., Ермолаев Е.В. СПОСОБ ПОДБОРА СВЕРТОЧНЫХ ЯДЕР НА ПРИМЕРЕ MNIST..... | 73 |

СЕТИ И ТЕЛЕКОММУНИКАЦИИ

| | |
|--|----|
| Фокин А.Б., Шмойлов А.В., Коркин А.Г. ВЫБОР ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБСЛУЖИВАНИЯ В СОГЛАШЕНИИ ОБ УРОВНЕ ОБСЛУЖИВАНИЯ..... | 78 |
|--|----|

**СИСТЕМЫ АВТОМАТИЗИРОВАННОГО
ПРОЕКТИРОВАНИЯ И ГИС-ТЕХНОЛОГИИ**

| | |
|--|----|
| Набиева К.И., Баймурзин Д.И. ГЕОСОЦИАЛЬНЫЕ СЕТИ И КАРТОГРАФИРОВАНИЕ..... | 82 |
| Галпакова А.Б., Аннаева С.Д. РАЗРАБОТКА ГЕОИНФОРМАЦИОННЫХ СИСТЕМ УПРАВЛЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОСТРАНСТВ..... | 88 |
| Аннаев М.С., Ымамгулыев Ы.Г. ВОЗМОЖНОСТИ ВНЕДРЕНИЯ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЦИФРОВОЕ ОБРАЗОВАТЕЛЬНОЕ ПРОСТРАНСТВО..... | 91 |

**СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ
И ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА**

| | |
|---|----|
| Сильнова С.В., Мустафина Д.Б. ЗАДАЧА ПРОГНОЗИРОВАНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА НЕФТЕПРОДУКТОВ АТМОСФЕРНО-ВАКУУМНОЙ ПЕРЕГОННОЙ УСТАНОВКИ..... | 95 |
|---|----|

СИСТЕМЫ И МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ

| | |
|---|-----|
| Очилов Н.Н. АЛГОРИТМЫ ШИФРОВАНИЯ НА ОСНОВЕ МЕСТНЫХ СТАНДАРТОВ..... | 100 |
| Симонян Р.А., Корнеев Н.В. ВОЗМОЖНОСТИ REGEX ДЛЯ РЕШЕНИЯ ПРОБЛЕМ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ..... | 107 |

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В НАУКЕ, ОБРАЗОВАНИИ И ПРОИЗВОДСТВЕ

УДК 004.912

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ДЛЯ ЭЛЕКТРОННОЙ ПОДАЧИ ДОКУМЕНТОВ ПО ЗАДАННОМУ МАКЕТУ

AUTOMATED SYSTEM FOR ELECTRONIC SUBMISSION OF DOCUMENTS ACCORDING TO A GIVEN LAYOUT

Хаматянов М.И., Медведева О.А.,
ФГАОУ ВО «Казанский (Приволжский) федеральный университет»,
г. Казань, Российская Федерация

M.I. Khamatyanov, O.A. Medvedeva,
FSAEI HE “Kazan (Volga Region) Federal University”,
Kazan, Russian Federation

e-mail: medvedeva2019@gmail.com

Аннотация. В данной работе рассматривается задача разработки web-приложения для электронной подачи документов. Для выполнения задачи проанализирован процесс подачи документов и рассмотрены существующие решения. Для выполнения работы были получены необходимые шаблоны документов. Разработанное web-приложение для электронной подачи документов предоставляет следующий функционал: авторизация пользователя в системе, редактирование профиля авторизованного пользователя, создание и редактирование пользователей системы, создание и редактирование типов пользователей и отделов, создание и редактирование доступных для подачи документов и их макеты, подача документов для пользователей, просмотр и редактирование поданных документов для пользователей, просмотр и обработка поданных документов для менеджеров. Web-приложение имеет интуитивно понятный и удобный интерфейс, адаптированный под устройства с различным разрешением. Приложение предоставляет все необходимые возможности для электронной подачи документов и позволяет пользователям легко их отслеживать. С помощью функционального и юнит-тестирования были протестированы как клиентская, так и серверная части приложения. Web-приложение разработано с использованием языков программирования Java, JavaScript и фреймворков Spring Boot, Vue.js, базы данных PostgreSQL в среде программирования IntelliJ IDEA.

Abstract. In this paper, the problem of developing a web application for electronic filing of documents is considered. The necessary document templates were obtained to complete the work. The developed web-application for electronic filing of documents provides the following functionality: user authorization in the system, editing the profile of an authorized user, creating and editing system users, creating and editing types of users and departments, creating and editing documents available for filing and their layouts, filing documents for users, viewing and editing submitted documents for users, viewing and processing submitted documents for

managers. The web application has an intuitive and user-friendly interface adapted to devices with different resolutions. The application provides all the necessary features for electronic filing of documents and allows users to easily track them. Using functional and unit testing, both the client and server parts of the application were tested. The web application was developed using the Java and JavaScript programming languages and the Spring Boot and Vue.js frameworks, PostgreSQL database in the IntelliJ IDEA programming environment.

Ключевые слова: web-приложение, Java, Spring Boot, JavaScript, Vue.js, Html, Css, PostgreSQL.

Keywords: web application, Java, Spring Boot, JavaScript, Vue.js, Html, Css, PostgreSQL.

В настоящий момент существует множество решений для электронного документооборота, но потребность в создании новых приложений все еще есть. В областях, в которых имеет место частая работа с документами и их обработкой, актуальными являются системы, направленные на автоматизацию процесса обработки документов, что позволяет значительно сэкономить время и упростить обработку заявлений в электронном формате.

Для подачи документов необходимо физически принести распечатанные документы в учреждение, что не всегда удобно и не всегда есть возможность прийти и лично отдать документы, особенно в период дистанционного обучения и удаленной работы. Также не всегда предусмотрена возможность отслеживания статуса поданного документа. Решение в виде web-приложения для электронной подачи документов и их обработки поможет значительно сэкономить время, теряемое на распечатку и физическую подачу документов, а также позволит оперативно отслеживать статус поданных документов. Актуальность разработки состоит в следующем: приложение автоматизирует процесс подачи и обработки документов; приложение позволяет не привязываться к временным рамкам и подавать документы в любое удобное время; приложение облегчает процесс взаимодействия между пользователями и людьми, подписывающими документ.

Целью разработки web-приложения является автоматизация процесса подачи документов с помощью предоставления возможности их электронной подачи и обработки, используя интуитивно понятный пользовательский интерфейс.

Предметом разработки является web-приложение для электронной подачи документов. Назначение web-приложения: предоставление пользователям возможности просмотра и подачи доступных документов; предоставление пользователям возможности просматривать статус поданных документов; предоставление администраторам возможности просмотра, создания, редактирования и удаления пользователей системы; предоставление администраторам возможности просмотра, создания, редактирования и удаления доступных для подачи документов; предоставление менеджерам возможности просмотра и обработки поданных пользователями документов.

Целевой аудиторией разрабатываемого web-приложения являются работники организаций, в которых имеет место подача документов в различные ведомства. К примеру, студенты и работники университетов.

Приведем следующие классы пользователей.

Администратор – авторизованный пользователь. Имеет доступ к своему профилю с возможностью его редактировать. Обладает правом на просмотр и редактирование всех пользователей системы, правом на просмотр и редактирование доступных для подачи

документов, правом на просмотр и редактирование типов отделов менеджеров и типов пользователей.

Пользователь – авторизованный пользователь. Имеет доступ к своему профилю с возможностью его редактировать. Обладает правом на просмотр доступных ему документов для подачи. Обладает правом подать выбранный документ и следить за его статусом.

Менеджер – авторизованный пользователь. Имеет доступ к своему профилю с возможностью его редактировать. Обладает правом на просмотр и обработку доступных ему документов, поданных пользователями.

Данные приложения будут храниться в базе данных. На основе технического задания была построена ER-модель базы данных для разрабатываемого web-приложения.

Системой управления базами данных (СУБД) в разрабатываемом web-приложении была выбрана PostgreSQL. Это объектно-реляционная СУБД. Она является достаточно мощным и простым инструментом. С данной базой данных можно работать из выбранной среды разработки IntelliJ IDEA, что значительно упрощает процесс разработки.

Используя возможности выбранной СУБД построена база данных web-приложения для электронной подачи документов (рисунок 1).

Web-приложение состоит из серверной и клиентской части, общающихся посредством REST запросов.

Серверная часть написана на языке программирования Java с использованием фреймворка Spring boot, который позволяет разрабатывать современные промышленные приложения.

Клиентская часть написана с использованием фреймворка Vue.js. Он является современным и прогрессивным, позволяет решать большое количество задач и с его помощью можно создать удобные в использовании web-приложения.

Для настройки CORS и определения путей на сервере доступных только авторизованным пользователям или пользователям с определенной ролью используется Spring Security.

Выводы

Разработанное web-приложение для электронной подачи документов предоставляет следующий функционал: авторизация пользователя в системе, редактирование профиля авторизованного пользователя, создание и редактирование пользователей системы, создание и редактирование типов пользователей и отделов, создание и редактирование доступных для подачи документов и их макеты, подача документов для пользователей, просмотр и редактирование поданных документов для пользователей, просмотр и обработка поданных документов для менеджеров и др.

Web-приложение имеет интуитивно понятный и удобный интерфейс, адаптированный под устройства с различным разрешением.

Приложение предоставляет все необходимые возможности для электронной подачи документов и позволяет пользователям легко их отслеживать.

С помощью функционального и юнит-тестирования были протестированы как клиентская, так и серверная части приложения.

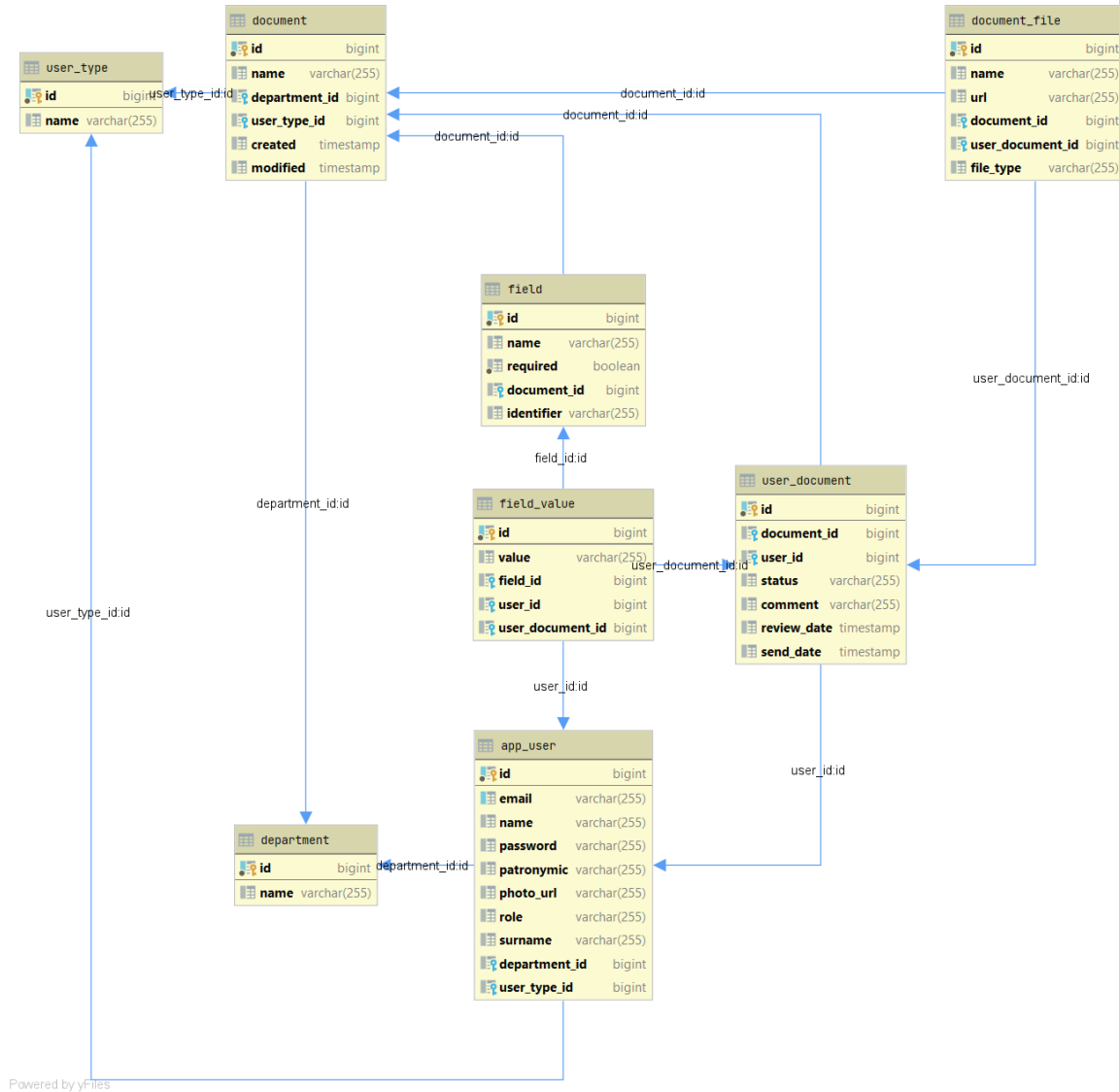


Рисунок 1. ER-модель базы данных

Литература

1. Руководство Spring [Электронный ресурс] – Режим доступа: URL: <https://spring.io/learn> (Дата обращения: 01.03.2022).
2. Руководство Vue.js [Электронный ресурс] – Режим доступа: URL: <https://ru.vuejs.org/v2/guide> (Дата обращения: 10.03.2022).
3. Возможности IntelliJ IDEA [Электронный ресурс] – Режим доступа: URL: <https://www.jetbrains.com/ru-ru/idea/features/> (Дата обращения: 04.03.2022).
4. Руководство HTML/CSS/JavaScript [Электронный ресурс] – Режим доступа: URL: <https://habr.com/ru/post/275729/> (Дата обращения: 10.03.2022).
5. Руководство Vue.js [Электронный ресурс] – Режим доступа: URL: <https://ru.vuejs.org/v2/guide> (Дата обращения: 10.03.2022).
6. Документация PostgreSQL [Электронный ресурс] – Режим доступа: URL: <https://postgrespro.ru/docs/postgresql/9.4/reference> (Дата обращения: 13.02.2022).

УДК 004.928

**РАЗРАБОТКА ТЕХНОЛОГИИ АНИМАЦИИ
ТРЕХМЕРНОЙ МОДЕЛИ
В ЦЕЛЯХ ПОВЫШЕНИЯ КАЧЕСТВА ОБУЧЕНИЯ ХОРЕОГРАФИИ**

**DEVELOPMENT OF THREE-DIMENSIONAL MODEL
ANIMATION TECHNOLOGY
IN ORDER TO IMPROVE THE QUALITY OF CHOREOGRAPHY TRAINING**

Астаева Т.О., Зайцева Е.В.,
Национальный исследовательский технологический университет «МИСиС»
г. Москва, Российская Федерация

T.O. Astaeva, E.V. Zaytseva,
National University of Science and Technology "MISIS"
Moscow, Russian Federation

email: m1606200@edu.misis.ru

Аннотация. Статья посвящена применению информационных технологий в сфере хореографического искусства путем описания разработки технологии анимации трехмерной модели для обучения хореографии, включающей 3D-модель танцора, анимированную с помощью технологий компьютерного зрения и модуля трехмерной реконструкции. Сформулирована проблема обучения хореографии, ведь развитие в сфере танцевального искусства, кроме всего прочего, связано с развитием интернета, поскольку в настоящее время для изучения хореографии, заинтересованные ученики проходят обучение не только в рамках занятий хореографией непосредственно с инструктором, но и применяют специализированные учебники и интернет-курсы, которые приобрели широкое распространение. Вышеупомянутые информационные технологии для дистанционного обучения хореографии не дают возможности полного понимания телодвижения, что актуально не только для эстетически красивого исполнения движения, но и для здоровья и безопасности учеников. Далее обозначена целесообразность применения 3D-моделей в указанной области, обоснована разработка обучающего приложения в области хореографии, основанного на представлении 3D-модели выполнения движений танцора. В результате приведена модель разработанной технологии, спроектированы информационное обеспечение обучения хореографии и логика работы обучающего приложения, а также описана разработка 3D-модели танцора.

Abstract. The article is devoted to the application of information technologies in the field of choreographic art by describing the development of a three-dimensional model animation technology for teaching choreography, including a 3D model of a dancer animated using computer vision technologies and a three-dimensional reconstruction module. The problem of teaching choreography is formulated. The development in the field of dance art, among other reasons, is connected with the development of the Internet, because at the present, interested students are trained not only as part of choreography classes directly with an instructor, but also use specialized textbooks, online classes and Internet courses that have become widely distributed. The aforementioned information technologies for distance learning

of choreography do not allow a complete understanding of body movement, which is relevant not only for the aesthetically beautiful performance of the movement but also for the health and safety of students. Further, the expediency of using 3D models in this area is indicated, the development of a training application in the field of choreography based on the representation of a 3D model of the dancer's movements is justified. As a result, a model of the developed technology is presented, information support for choreography training and the logic of the training application are designed, and the development of a 3D model of a dancer is described.

Ключевые слова: моделирование, нейронная сеть, обучение, анимация, компьютерное зрение, информационные технологии.

Keywords: modeling, neural network, training, animation, computer vision, information technology.

Современные информационные технологии применяются и в области науки и техники, и в области искусства. Танцевальное направление в искусстве как в мире, так и в Российской Федерации, ежегодно становится более популярным. Кроме того, занятия хореографией помогают избавиться или предотвратить многие физические болезни [1], а также у людей, профессионально занимающихся танцами, снижен риск неврологических болезней. Следует также отметить положительный эффект занятий хореографией для эндокринной системы.

Таким образом, хореография является не только актуальным и активно развивающимся направлением искусства в настоящее время, но и оказывает положительный эффект на как физическое, так и на психологическое или духовное здоровье человека независимо от его пола и возраста.

Несмотря на популярность, следует отметить некоторую сложность и травмоопасность занятий хореографией. Среди распространенных травм отмечают проблемы с менисками коленных суставов, которые возникают при неправильном распределении центра тяжести тела, т.к. имеют место выворотные позиции ног, которые совершенно непривычны в повседневной жизни человека и ошибочно исполняемые на танцевальных уроках.

Неправильная интерпретация танцевальных движений влечет за собой множественные травмы [2].

Таким образом, травматичность данного вида искусства достаточно высока. Учитывая количество людей, заинтересованных в занятиях хореографией, следует уделить внимание профилактике данной проблемы. Ведь достаточно распространённой является проблема неверного распознавания учеником движений учителя [3].

Наряду с занятиями непосредственно с инструктором, в последнее время стали широко применяться специализированные учебники по хореографии, а также танцевальные интернет-курсы.

С развитием интернета стало возможным изучение танцевальных элементов, где посредником между учеником и учителем выступает видео, снятое на камеру.

Обычная камера, которая снимает видео в 2D не может адекватно передать все особенности выполнения того или иного движения.

Пособия также являются неэффективными, так как имеют в учебнике как правило иллюстрации, представленные с разных сторон: спереди, сзади и сбоку.

Один или несколько ракурсов не дают возможности полного понимания представляемого телодвижения. В частности, нет возможности выполнения элемента с разных сторон. В подобных аналогичных ситуациях в других видах спорта

целесообразно применение трехмерных моделей, на которых представлены различные движения спортсменов.

Например, работа по применению имитационного и геометрического моделирования в обучении горнолыжному спорту. В ней используется анимированная 3D-модель представителя горнолыжного спорта, которая повторяет движения реального горнолыжника [4].

Также существует множество 3D-анимаций в других направлениях, таких как медицина, кинематограф, биология, компьютерные игры, реклама и маркетинг, однако в сфере обучения танцам применение данной технологии рассматривается как достаточно новый подход.

Для того чтобы разработать модель персонажа необходимо знание анатомии человека. Чтобы корректно анимировать прежде упомянутого персонажа необходимо воссоздать движения живого человека. Телодвижения должны быть неотличимы от реальных, поскольку в обучении танцевальным движениям каждая деталь выполнения действия должна быть принята во внимание.

Использование стандартной анимации не дает возможностей учета всех деталей при «оживлении» 3D-персонажа.

Разработка сложной анимации, которая рассматривается в случае с хореографией, практически невозможна без использования более совершенной технологии, ведь в первом случае предстоит воспроизведение анатомических движений человека математически. Чтобы улучшить возможность обучения off-line, целесообразно использование новой технологии на основе компьютерного зрения. Применение технологии YOLO [5] и модуля трехмерной реконструкции позволит перенести движение хореографа-тренера на фигуру персонажа в графическом редакторе, которая отобразит движение первого и, кроме того, позволит рассмотреть то или иное движение со всех сторон, имитируя работу с реальным инструктором. Другими словами, эта технология дает возможность оцифровать движения актёра и в дальнейшем использовать их для управления 3D-моделью персонажа.

Требуется разработать 3D-модель персонажа, который представляет собой фигуру человека. Далее, следует передать танцевальные движения хореографа-преподавателя разработанному 3D-персонажу. 3D-модель должна передавать особенности выполнения движений хореографа-преподавателя. Для этого следует подобрать подходящий уровень детализации.

Нижние конечности 3D-манекена необходимо разбить на пять частей. Следует выделить следующие: пальцы ног, которые отобразят движение на полупальцах; голеностопный сустав, который позволит отобразить движения стопы; коленная чашечка, которая отобразит естественное сгибание и разгибание вышеназванной части нижней конечности; бедро, которое отобразит движение верхней части нижней конечности; крестец, который отобразит связь нижней конечности с телом.

Верхние конечности 3D-персонажа следует разбить на четыре части. Необходимо выделить следующие: лучевая кость, которая отобразит движения кисти (в данном случае не является необходимым отдельно выделять пальцы кисти, т.к. стандартное положение кисти в народном танце - кулак); локтевая ямка, которая свяжет движения лучевой и локтевой кости с запястьем; головка плечевой кости, которая отображает движение руки начиная с плечевой кости; ключица/шейный отдел позвоночника, который отобразит связь верхней конечности с телом.

Тело 3D-персонажа требуется разделить на пять частей. Следует выделить следующие: копчик, который связывает движения тела с нижними конечностями; поясничный отдел, который отобразит гибкость нижней части тела; грудной отдел, которое отобразит естественную гибкость тела; нижняя часть шейного отдела, которая

связывает движения тела с верхними конечностями; атлант и осевой позвонок, которые обеспечивают движение шейного отдела; теменное отверстие, которое свяжет движение головы с шейным отделом. Таким образом, данное разделение 3D-персонажа позволит корректно передать движения народного танца хореографа-преподавателя.

Технология анимации данной модели состоит из семи этапов (рисунок 1).

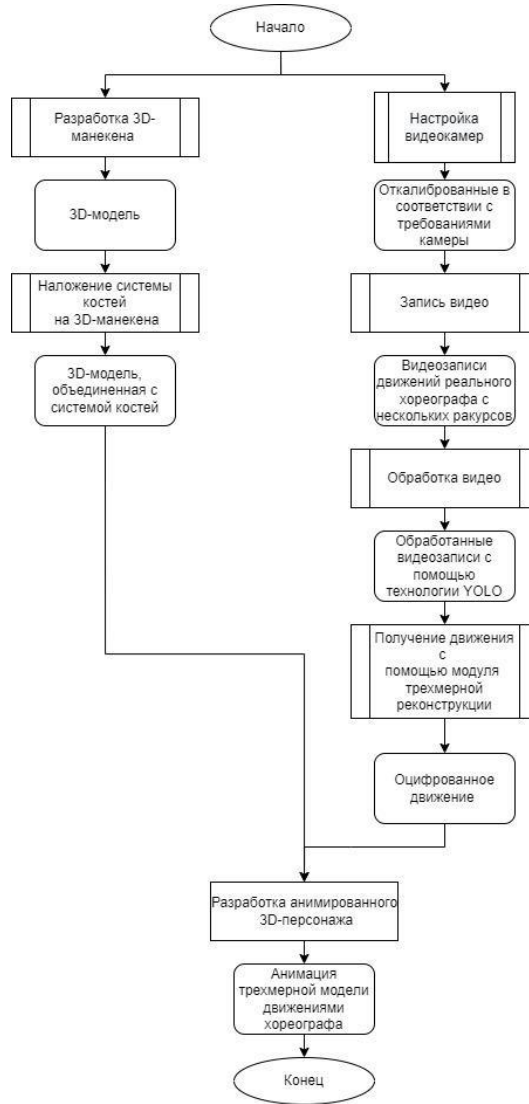


Рисунок 1. Схема процесса анимации трехмерной модели

Разработка 3D-манекена – разработка модели персонажа в программе 3Ds Max 2019.

Персонаж представляет фигуру человека в танцевальном костюме. Наложение системы костей на 3D-манекена – разработка системы костей необходимой детализации и наложение модель.

Настройка видеокамер – калибровка камер в соответствии с инструкцией, изложенной в документации к модулю трехмерной реконструкции.

Запись видео – запись танцевальных движений от реального человека-хореографа с нескольких камер.

Обработка видео – обработка полученных на предыдущем этапе видео с помощью обученной нейронной сети YOLO.

Получение движения с помощью модуля трехмерной реконструкции – загрузка обработанных видефрагментов выполнения движения как входных данных для модуля трехмерной реконструкции и получение файла с движением как выходных данных вышеуказанного модуля [6].

Разработка анимированного 3D-персонажа – объединение персонажа с системой костей, созданной на втором этапе, и движения, полученного на предыдущем этапе.

Итогом будет полноценный 3D-персонаж, выполняющий движения хореографа-преподавателя.

Для работы с анимированным персонажем использована программа 3Ds Max. Данная программа с интуитивно понятным интерфейсом позволит преподавателю и ученику легко разобраться в алгоритме работы приложения. Кроме того, вышеупомянутая программа подходит для операционной системы Windows, которая является наиболее популярной среди пользователей.

Среди используемых приложением баз данных [7], выделим «Модель персонажа» и «Движение».

Первая предусматривает хранение информации о 3D модели человека в среде 3Ds Max. Такой информацией являются данные о координатах вершин, ребер и граней персонажа.

База данных «Движение» включает в себя хранение данных о танцевальных движениях и их сочетаниях, т.е. включает в себя информацию о танцах, которые формируются из большинства движений.

Тренер-хореограф, который ведет обучение, обладает возможностью формировать новые движения, передавая информацию о движении и отправляя запрос на формирование движения.

Следующим шагом после формирования движения будет добавление этого движения в базу данных «Движение», о которой упоминалось ранее. Далее, движением может воспользоваться ученик хореографических классов. Выполнение процесса воспроизведения движения влияет на положение 3D персонажа в 3D пространстве.

Кроме возможности анимировать персонажа, ученику предоставляется право менять ракурс, что позволит рассмотреть движения под разными углами.

При запуске приложения, ученик танцевальной группы видит модель 3D-персонажа в начальном положении.

Модель 3D-персонажа визуализирует собой тело танцовщицы в танцевальном купальнике, танцевальной балетной обуви и с волосами, убранными в балетный пучок.

Следующим этапом является выбор движения в программе. Выбор любого элемента позволяет рассмотреть выполнение движения неограниченное количество раз, при этом изменяя или не изменяя положение глаза наблюдателя.

Для изменения ракурса существуют специальные инструменты, т.е. инструменты для поворота, перемещения и масштабирования сцены. Такие инструменты находятся в верхней панели приложения и располагаются по порядку: «Перемещение», «Поворот», «Масштабирование». Также, осуществляется возможность приостановить выполнение движения и рассмотреть все тонкости исполнения с разных ракурсов.

Таким образом, результатом разработки является приложение, в котором ученик имеет возможность выбрать любое движение из предложенных. Под предложенными понимаются движения, которые были добавлены учителем. После выбора определенного движения, проходящий обучение танцовщик либо выбирает другое движение, либо анимирует выбранное.

В течение анимации, у ученика существует возможность остановить анимацию на интересующем его моменте и изменить ракурс, перемещая сцену, поворачивая сцену, либо используя инструменты масштабирования, т.е. приближая и отдаляя сцену.

Следует отметить, что маневры изменения ракурса доступны и осуществляются не только в момент остановки анимации персонажа в выбранном положении, но и во время проигрывания анимации. Кроме того, предусмотрена возможность менять ракурс неограниченное количество раз.

Выводы

В работе был проведен анализ процессов в обучения хореографии и доказана актуальность разработки технологии анимации трехмерной модели и обучающего приложения. Были описаны поведение и функционирование проектируемого приложения.

Таким образом, результатом является технология анимации трехмерной модели, включающая приложение для поддержки самостоятельного обучения хореографии, разработку 3D-модели и её последующую анимацию.

3D-модель персонажа в обучающем приложении состоит из модели скелета и оболочки. Скелет разбит на 21 часть в соответствии с поставленной задачей. Части, из которых состоит скелет, связаны между собой. Такое разбиение скелета позволяет полностью передать движения танцора хореографа-преподавателю, а оболочка, в свою очередь позволяет корректно их отобразить.

Оболочка 3D-модели состоит из 256 443 полигонов и 128 388 вершин. Такое разбиение позволит достаточно детализировано, информативно и реалистично отобразить танцевальные движения живого человека. Результат работы с приложением окажет влияние на самостоятельное обучение хореографии, делая такой вариант обучения более эффективным для учеников танцевальных хореографических классов.

Литература

1. Ахмедова З. Развитие личности средствами хореографии // *Oriental Art and Culture*. 2021. №4.
2. Ботова Л.Н., Дервянко О.И. Связь кинематических характеристик приземления и эффективности удержания вертикальной позы у гимнастов 8-9 лет//Ученые записки университета Лесгафта. 2019. №10 (176). URL: <https://cyberleninka.ru/article/n/svyaz-kinematicheskikh-harakteristik-prizemleniya-i-effektivnosti-uderzhaniya-vertikalnoy-pozy-u-gimnastov-8-9-let> (дата обращения: 30.03.2022).
3. Марченкова, А.И., Марченков А.Л. Проблемы травматизма в процессе обучения хореографическому искусству/А.И. Марченкова, А.Л. Марченков. – Текст: непосредственный//Педагогика: традиции и инновации: материалы II Междунар. науч. конф. (г. Челябинск, октябрь 2012 г.). – Челябинск: Два комсомольца, 2012. – С. 172-174. – URL: <https://moluch.ru/conf/ped/archive/63/2760/> (дата обращения: 23.05.2020).
4. Ворончагина К.В. Имитационное и геометрическое моделирование в обучении горнолыжному спорту//ГИАБ. 2013. №11. URL: <https://cyberleninka.ru/article/n/imitatsionnoe-i-geometricheskoe-modelirovanie-v-obuchenii-gornolyzhnomu-sportu> (дата обращения: 23.05.2020).
5. Боков П.А., Кравченя П.Д. Экспериментальный анализ точности и производительности разновидностей архитектур YOLO для задач компьютерного зрения//Программные продукты и системы. 2020. №4.
6. Яценко А.В., Беликов А.В., Петерсон М.В., Потапов А.С. Дистилляция нейросетевых моделей для детектирования и описания ключевых точек изображений//Научно-технический вестник информационных технологий, механики и оптики. 2020. №3. URL: <https://cyberleninka.ru/article/n/distillyatsiya-neyrosetevyh->

modeley-dlya-detektirovaniya-i-opisaniya-klyuchevykh-tochek-izobrazheniy (дата обращения: 30.03.2022).

7. Фомичева О.Е., Чуков А.М. Интеллектуальный анализ данных финансово-аналитической системы университета. Задача кластеризации//Сборник научных трудов 18-й международной НПК «Новые информационные технологии в образовании (применение технологий «1С» для развития компетенций цифровой экономики) Москва, 2018. С. 88-91.

УДК 004.62

ОПИСАНИЕ СХЕМ ИМЕНОВАННЫХ СТРУКТУРНЫХ ЕДИНИЦ ДЛЯ ИНФОЛОГИЧЕСКОЙ МОДЕЛИ

DESCRIPTION OF SCHEMES OF NAMED STRUCTURAL UNITS FOR THE INFOLOGICAL MODEL

Бердюгин А.В.,
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технологический университет «СТАНКИН»,
г. Москва, Российская Федерация

A.V. Berdyugin,
Moscow State University of Technology “STANKIN”,
Moscow, Russian Federation

e-mail: a.berdyugin@stankin.ru

Аннотация. В статье приведено описание схем именованных структурных единиц (ИСЕ) для инфологической модели в рамках методологии автоматизации интеллектуального труда (МАИТ). Дано краткое описание МАИТ и инфологического моделирования. Схема ИСЕ – это производная конструкция на инфологической структуре. Схема является совокупностью элементарных схем. В свою очередь, элементарные схемы определяются тернарными связями между ИСЕ типа «Информационный Модуль», ИСЕ типа «Информационная Сущность» и ИСЕ типа «Информационный атрибут». Исходными данными для описания схем являются форма P2 (описание бинарных связей именованных структурных единиц) и форма P9 (описание тернарных связей именованных структурных единиц). Для описания схем именованных структурных единиц разработана спецификация P10, которая формируется в соответствии с определенными правилами. Данные правила описывают порядок формирования и заполнения каждого из полей указанной спецификации. В статье приведен результат апробации – пример описания схем ИСЕ для фрагмента задачи «Проектировочный расчёт плоскоремленной передачи». Описание схем именованных структурных единиц, отображаемых в спецификации P10, позволит в дальнейшем разработать описание бинарных связей между схемами ИСЕ, а также будет включено в методику описания производных конструкций инфологической структуры.

Abstract. The article describes the schemes of named structural units (NSU) for the infological model within the framework of the methodology of automation of intellectual labor

(MAIL). A brief description of MAIL and infological modeling is given. The NSU scheme is a derived construction on an infological structure. A circuit is a collection of elementary circuits. In turn, elementary schemes are determined by ternary connections between the NSU of the “Information Module” type, the NSU of the “Information Entity” type and the NSU of the “Information Attribute” type. The initial data for the description of schemes are form P2 (description of binary relations of named structural units) and form P9 (description of ternary relations of named structural units). To describe the schemes of named structural units, the P10 specification has been developed, which is formed in accordance with certain rules. These rules describe the procedure for forming and filling in each of the fields of the specified specification. The article presents the result of the approbation - an example of the description of the NSU schemes for a fragment of the problem "Design calculation of a plane-time transmission". The description of the schemes of named structural units displayed in the P10 specification will make it possible to further develop a description of binary relationships between NSU schemes, and will also be included in the methodology for describing derived constructions of the infological structure.

Ключевые слова: методология автоматизации интеллектуального труда, инфологическая модель, инфологическая структура, структурная единица, схема структурных единиц.

Keywords: the methodology for automating intellectual labor, infological modeling, infological structure, structural unit, schemes of structural units.

Методология автоматизации интеллектуального труда (МАИТ) является теоретическим базисом и методическим обеспечением при создании прикладных автоматизированных систем (ПАС). Процесс проектирования ПАС на основе МАИТ обеспечивает формирование модельных представлений предметных задач: начального, концептуального, инфологического [1, 2].

Инфологическое моделирование является синтаксическим представлением будущей автоматизированной системы, решающей класс предметных задач в вычислительной среде [3].

В статье приведено описание схем именованных структурных единиц (ИСЕ), которые являются производными конструкциями инфологической структуры. Целью данного описания является фиксация схем ИСЕ в разработанной спецификации (P10).

Схемой в МАИТ является совокупность элементарных схем, которые в свою очередь формируются на основании тернарных связей.

Тернарная связь (и элементарная схема) в инфологической модели является совокупностью ИСЕ типа «ИМ» (информационный модуль), ИСЕ типа «ИС» (информационная сущность), ИСЕ типа «ИА» (информационный атрибут). Схема определяется общей парой: информационный модуль и информационная сущность, а также неповторяющимися атрибутами.

В процессе работы было определено, что исходной информацией для формирования спецификации P10 являются формы P2 и P9, приведенные в таблицах 1 и 2.

Таблица 1 – Форма P2: Описание бинарных связей именованных структурных единиц

| Класс связи ИСЕ | Код исходной ИСЕ | Код ИСЕ | Наименование связи | Тип связи | Оценка связи | Код связи |
|-----------------|------------------|---------|--------------------|-----------|--------------|-----------|
| | | | | | | |

Таблица 2 – Форма Р9: описание тернарных связей именованных структурных единиц

| Код уровня сложности | Код «ИМ» | Ранг «ИС» в «ИМ» | Код «ИС» | Ранг «ИА» в «ИС» | Код «ИА» | Код тернарной связи |
|----------------------|----------|------------------|----------|------------------|----------|---------------------|
|----------------------|----------|------------------|----------|------------------|----------|---------------------|

Для заполнения спецификации Р2 необходима сформированная инфологическая структура, которая должна быть зафиксирована в форме Р1, где отражается состав инфологической структуры – перечень именованных структурных единиц. Пример инфологической структуры приведен на рисунке 1. На рисунке красным цветом выделена схема именованных структурных единиц третьего уровня семантической сложности (МСА3).

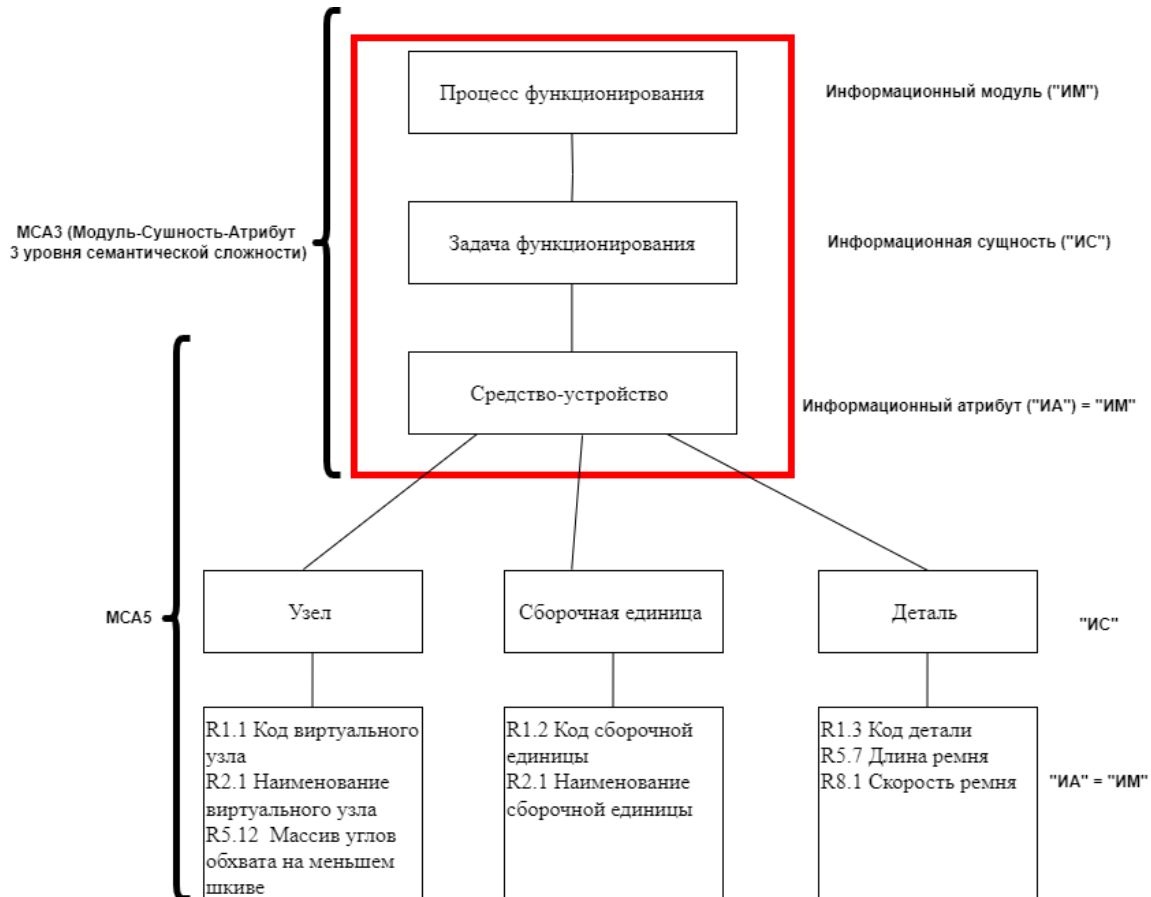


Рисунок 1. Пример инфологической структуры

В ходе разработки описания были определены правила, которые необходимо соблюдать при формировании спецификации Р10.

В форме фиксируется перечень схем именованных структурных единиц, которые определяются кодами общей пары: информационного модуля и информационной сущности.

Форма Р10 (описание схем именованных структурных единиц) приведена в таблице 3.

Таблица 3 – Форма Р10: описание схем именованных структурных единиц

| Код уровня сложности | Код схемы |
|----------------------|-----------|
| | |

Работа апробирована на различных проектных расчетных задачах.

Пример апробации для фрагмента задачи «Проектировочный расчёт плоскоременной передачи» приведен далее.

Исходной информацией для формирования спецификации P10 «описание схем именованных структурных единиц» являются фрагменты форм P2 и P9, которые представлены в таблицах 4 и 5 соответственно.

Спецификация P2 создаётся автоматически на основании алгоритма действий (системы предметных манипуляций), которая формируется в программном комплексе «ИС-2», поддерживающем проектирование прикладных автоматизированных систем по МАИТ [4-6].

Таблица 4 – Фрагмент формы P2: описание бинарных связей именованных структурных единиц

| Класс связи ИСЕ | Код исходной ИСЕ | Код ИСЕ | Наименование связи | Тип связи | Оценка связи | Код связи |
|-----------------|------------------|---------|---|-----------|--------------|-----------------|
| M30 | M3.P1.1 | M3.P1.1 | Упорядочивание ИСЕ-ИМ-3 «Процесс функционирования КП» | У | 99 | M3.P1.1M3.P1.1 |
| M3C | M3.P1.1 | C3.Z1.1 | Состав ИСЕ-ИМ-3 «Процесс функционирования КП» | С | 1 | M3.P1.1C3.Z1.1 |
| C30 | C3.Z1.1 | C3.Z1.1 | Упорядочивание ИСЕ-ИС-3 «Задача функционирования КП» | У | 99 | M3.P1.1M3.P1 |
| C3A | C3.Z1.1 | A3.K1.7 | Состав ИСЕ-ИС-3 «Задача функционирования КП» | С | 1 | C3.Z1.1A3.K1.7 |
| A3C | C3.Z1.1 | A3.K1.7 | Компоновка ИСЕ-ИС-3 «Задача функционирования КП» | К | 1 | C3.Z1.1A3.K1.7- |
| A30 | A3.K1.7 | A3.K1.7 | Упорядочивание ИСЕ-ИА-3 «Средство-устройство» | У | 99 | A3.K1.7A3.K1.7 |
| M5C | M5.K1.7 | C5.O7.1 | Состав ИСЕ-ИМ-5 «Средство-устройство» | С | 1 | M5.K1.7C5.O7.1 |
| M5C | M5.K1.7 | C5.O7.2 | Состав ИСЕ-ИМ-5 «Средство-устройство» | С | 1 | M5.K1.7C5.O7.2 |
| M5C | M5.K1.7 | C5.O7.4 | Состав ИСЕ-ИМ-5 «Средство-устройство» | С | 1 | M5.K1.7C5.O7.4 |
| C5M | M5.K1.7 | C5.O7.1 | Компоновка ИСЕ-ИМ-5 «Средство-устройство» | К | 1 | M5.K1.7C5.O7.1- |
| C5M | M5.K1.7 | C5.O7.2 | Компоновка ИСЕ-ИМ-5 «Средство-устройство» | К | 2 | M5.K1.7C5.O7.2- |
| C5M | M5.K1.7 | C5.O7.4 | Компоновка ИСЕ-ИМ-5 «Средство-устройство» | К | 3 | M5.K1.7C5.O7.4- |

Таблица 5 – Фрагмент формы P9: описание тернарных связей именованных структурных единиц

| Код уровня сложности | Код «ИМ» | Ранг «ИС» в «ИМ» | Код «ИС» | Ранг «ИА» в «ИС» | Код «ИА» | Код тернарной связи |
|----------------------|----------|------------------|----------|------------------|----------|-----------------------|
| MCA3 | M3.P1.1 | 1 | C3.Z1.1 | 1 | A3.K1.7 | M3.P1.1C3.Z1.1A3.K1.7 |
| MCA5 | M5.K1.7 | 1 | C5.O7.1 | 1 | A5.R1.3 | M5.K1.7C5.O7.1A5.R1.3 |

Продолжение таблицы 5

| Код уровня сложности | Код «ИМ» | Ранг «ИС» в «ИМ» | Код «ИС» | Ранг «ИА» в «ИС» | Код «ИА» | Код тернарной связи |
|----------------------|----------|------------------|----------|------------------|----------|-----------------------|
| MCA5 | M5.K1.7 | 1 | C5.O7.1 | 2 | A5.R2.1 | M5.K1.7C5.O7.1A5.R2.1 |
| MCA5 | M5.K1.7 | 1 | C5.O7.1 | 3 | A5.R3.1 | M5.K1.7C5.O7.1A5.R3.1 |
| MCA5 | M5.K1.7 | 2 | C5.O7.2 | 1 | A5.R1.1 | M5.K1.7C5.O7.2A5.R1.1 |
| MCA5 | M5.K1.7 | 3 | C5.O7.4 | 1 | A5.R1.2 | M5.K1.7C5.O7.4A5.R1.2 |

В процессе формирования схем именованных структурных единиц была сформирована форма P10, приведённая в таблице 6.

Таблица 6 – Форма P10: описание схем именованных структурных единиц

| Код уровня сложности | Код схемы |
|----------------------|----------------|
| MCA3 | M3.P1.1C3.Z1.1 |
| MCA5 | M5.K1.7C5.O7.1 |
| MCA5 | M5.K1.7C5.O7.2 |
| MCA5 | M5.K1.7C5.O7.4 |

Выводы

Таким образом, на основе установленных правил, а также исходных данных в виде спецификаций P2 и P9 была получена форма P10, отображающая информацию обо всех схемах именованных структурных единиц инфологической модели. Данная спецификация позволит в дальнейшем описать методику формирования схем ИСЕ и разработать форму, которая описывает бинарные связи схем именованных структурных единиц инфологической модели.

Литература

1. Новоселова, О. В. Моделирование интегрированной среды поддержки создания прикладных автоматизированных систем/О.В. Новоселова, Г.Д. Волкова, А.Г. Гаврилов //Известия высших учебных заведений. Поволжский регион. Технические науки. – 2014. – № 1(29). – С. 81-91.
2. Когнитивные технологии создания прикладных автоматизированных систем /Г.Д. Волкова, О.В. Новоселова, Е.Г. Семячкова [и др.]//Вестник МГТУ Станкин. – 2019. – № 2(49). – С. 17-22.
3. Волкова, Г.Д. Проектирование прикладных автоматизированных систем в машиностроении: учебное пособие/Г.Д. Волкова, О.В. Новоселова, Е.Г. Семячкова. – М.: МГТУ «Станкин», 2002. – 162 с.
4. Гаврилов, А.Г. Инструмент проектирования информационно-активных систем интегрированная среда «ИС-2»/А.Г. Гаврилов, Г.Д. Волкова, О.В. Новоселова//Моделирование нелинейных процессов и систем: Материалы пятой международной конференции, Москва, 16-20 ноября 2020 года. – Москва: Общество с ограниченной ответственностью Издательство «Янус-К», 2021. – С. 129-130.
5. Гаврилов, А.Г. Программный комплекс «ИС-2» и его особенности/А.Г. Гаврилов, Г.Д. Волкова, О.В. Новоселова//Цифровая экономика: технологии, управление, человеческий капитал: Материалы III всероссийской научно-практической конференции, Вологда, 25 сентября 2020 года. – Вологда: Общество с ограниченной ответственностью «Маркер», 2020. – С. 10-13.

6. Свидетельство о государственной регистрации программы для ЭВМ № 2017618975 Российская Федерация. Программный комплекс ИС-2-У: № 2017616041: заявл. 23.06.2017; опубл. 14.08.2017/А.Г. Гаврилов; заявитель федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технологический университет «СТАНКИН»

УДК 004.6

СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК ЯЗЫКА PYTHON ДЛЯ ИЗУЧЕНИЯ МЕТОДОВ ВИЗУАЛИЗАЦИИ ДАННЫХ

COMPARATIVE ANALYSIS OF PYTHON LIBRARIES FOR STUDYING DATA VISUALIZATION TECHNIQUES

Агапов Ю.Г., Подвесовская М.А.,
ФГБОУ ВО «Брянский государственный технический университет»,
г. Брянск, Российская Федерация

Yu.G. Agapov, M.A. Podvesovskaia,
FSBEI HE “Bryansk State Technical University”,
Bryansk, Russian Federation

e-mail: agapov.yuri@mail.ru

Аннотация. Визуализация данных является одним из важных инструментов, применяемых на большинстве этапов анализа данных. Целью данной работы является сравнительный анализ библиотек языка программирования Python, предназначенных для визуализации данных, с точки зрения простоты их освоения и возможности использования при изучении методов визуализации и получении практических навыков их применения. Для сравнительного анализа выбраны следующие библиотеки: Matplotlib с надстройкой Seaborn, Plotly и Bokeh. Для визуализации использованы три различных, не связанных между собой набора данных, находящихся в свободном доступе. Приводится описание результатов экспериментов, целью которых являлось применение указанных библиотек для построения двумерных диаграмм следующих типов: линейная диаграмма, диаграмма области, столбиковая диаграмма (гистограмма), диаграмма радар, круговая диаграмма, сложенная столбиковая диаграмма. Применительно к перечисленным типам диаграмм обсуждаются такие характеристики библиотек, как полнота данных для построения диаграмм, сложность предварительной обработки данных, дополнительные возможности отображения и др. По результатам обсуждения делается вывод о целесообразности использования библиотеки Plotly как наиболее подходящего решения для освоения методов визуализации данных с использованием Python.

Abstract. Data visualization is one of the important tools used in most phases of data analysis. The aim of this work is a comparative analysis of Python programming language libraries, designed for data visualization, in terms of their ease of mastering and the possibility of using them to study visualization methods and gain practical skills in their application. For the comparative analysis the following libraries were chosen: Matplotlib with Seaborn add-on, Plotly and Bokeh. For visualization, three different, unrelated, freely available data sets were

used. We describe the results of experiments whose purpose was to use these libraries to build two-dimensional diagrams of the following types: line chart, area chart, bar graph (histogram), radar chart, pie chart, stacked bar chart. In relation to the listed types of diagrams, such characteristics of libraries as completeness of data for plotting diagrams, complexity of data preprocessing, additional display possibilities, etc., are discussed. Based on the results of the discussion, a conclusion is made about the advisability of using the Plotly library as the most suitable solution for studying data visualization methods using Python.

Ключевые слова: визуализация данных, Python, Matplotlib, Seaborn, Plotly, Bokeh.

Keywords: data visualization, Python, Matplotlib, Seaborn, Plotly, Bokeh.

Введение

Важность визуализации данных на всех этапах анализа данных хорошо известна и не вызывает сомнений. Как отмечается в [1], ключевой функцией визуализации является упрощение восприятия данных для последующего анализа. Целью данной работы является сравнительный анализ библиотек языка Python, предназначенных для визуализации данных: Matplotlib [2] с надстройкой Seaborn [3] (будем использовать для данной связки запись Matplotlib/Seaborn), Plotly [4] и Bokeh [5], с точки зрения простоты их освоения и возможности использования в учебных целях, для получения начальных практических навыков применения методов визуализации данных. Выбор этих библиотек объясняется их распространенностью, а также широким диапазоном предоставляемого функционала, что, помимо прочего, обуславливает возможность их использования обучающимися в своих последующих рабочих проектах.

Постановка задачи и описание исходных данных

Поскольку речь идет о получении начальных навыков решения задач, связанных с визуализацией, то ограничимся рассмотрением возможностей указанных библиотек применительно к визуализации табличных данных с использованием двумерных визуальных моделей.

В качестве визуальных моделей были выбраны: линейная диаграмма, диаграмма области, столбиковая диаграмма (гистограмма), диаграмма радар, круговая диаграмма и сложенная столбиковая диаграмма. Выбор именно такого набора обусловлен наглядностью перечисленных моделей относительно отображаемой информации, а также охватом различных видов систем координат [1]. Подробное описание данных визуальных моделей можно найти, например, в [6].

Для визуализации были выбраны три различных, не связанных между собой набора данных, находящихся в свободном доступе:

- данные о тестовых баллах, полученных абитуриентами на вступительных экзаменах в колледжи во Вьетнаме в 2019-2020 гг. [7];
- данные о населении стран по годам [8];
- данные о пассажирах парохода «Титаник» [9].

Описание результатов

Линейная диаграмма. Для построения линейной диаграммы использовались данные о населении Франции и Японии – соответственно, график отражал сравнительную динамику роста населения этих двух стран по годам (рисунок 1).

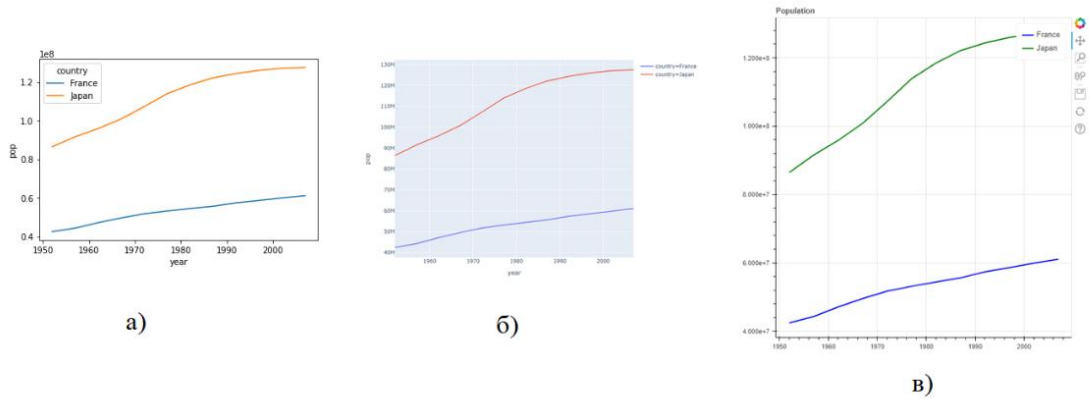


Рисунок 1. Линейная диаграмма, построенная средствами библиотек Mathplotlib/Seaborn (а), Plotly (б), Vokeh (в)

Построение данной диаграммы средствами библиотек Mathplotlib/Seaborn и Plotly трудностей не вызвало. В обоих случаях для построения используется метод, в который передаются данные и названия столбцов таблицы. Имеется возможность установки параметров отображения диаграммы, таких как цвет, тип линии, доверительный интервал и др.

В библиотеке Vokeh, в отличие от предыдущих библиотек, отсутствует возможность передачи данных одновременно по двум странам, в связи с чем потребовалось построение каждого графика по отдельности, с указанием цвета. В метод построения передаются данные по стране, название столбцов таблицы, цвет графика и подпись. Имеется возможность выбора цвета, типа линии и т.д.

Без дополнительных манипуляций в Mathplotlib/Seaborn были построены два графика с разными цветами и подписями легенды, взятыми из таблицы. В Plotly, помимо этого, строится сетка. В Vokeh также по умолчанию строится сетка, но легенда и цвета настраиваются отдельно.

Диаграмма области. Для построения диаграммы области использовались те же данные, что и для линейной диаграммы (рисунок 2).

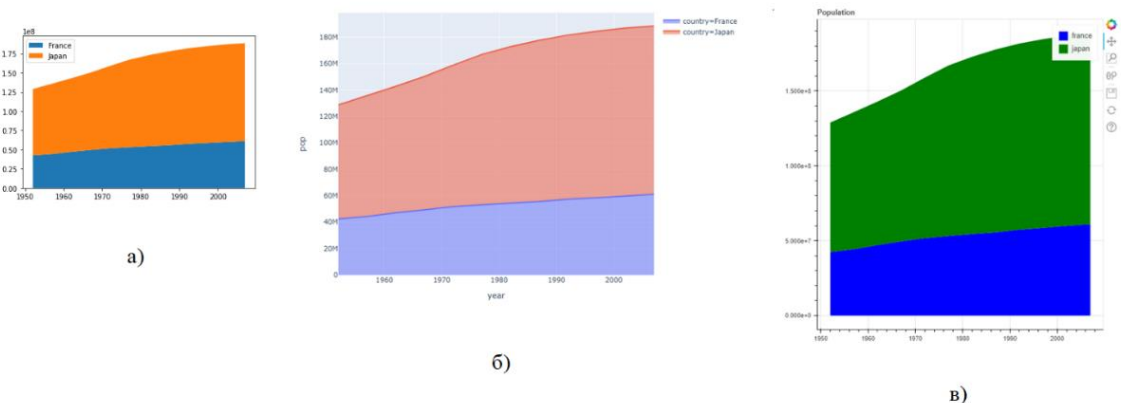


Рисунок 2. Диаграмма области, построенная средствами библиотек Mathplotlib/Seaborn (а), Plotly (б), Vokeh (в)

В Mathplotlib/Seaborn и Plotly построение диаграммы трудностей не вызвало. В метод построения передаются данные и подписи к легенде. Plotly предоставляет больше дополнительных возможностей, например, построение сложных диаграмм и сложных диаграмм с нормализованными значениями.

При использовании библиотеки Vokeh, как и в случае с линейной диаграммой, цвета необходимо задавать самостоятельно, однако для диаграммы области данные передаются одним набором, а не двумя. Легенда добавляется отдельно.

Столбиковая диаграмма (гистограмма). Для данного типа диаграммы использовались данные о пассажирах «Титаника» (были выбраны столбцы, задающие пол, класс, а также информацию о том, выжил ли пассажир). Результаты построения показаны на рисунке 3.

В Mathplotlib/Seaborn диаграмма была построена с дополнительной группировкой по классам пассажиров путем указания столбца с классами при вызове метода. Имеется возможность задания дополнительных параметров диаграммы – цвета, группировка, доверительный интервал.

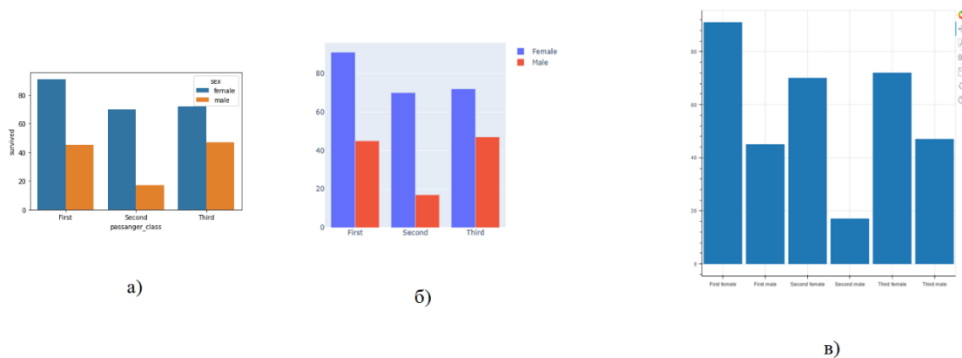


Рисунок 3. Столбиковая диаграмма, построенная средствами библиотек Mathplotlib/Seaborn (а), Plotly (б), Vokeh (в)

При использовании Plotly дополнительную группировку потребовалось выполнить вручную путем построения двух столбиковых диаграмм и объединения их в одной фигуре. Дополнительные параметры диаграмм в большей степени относятся к эстетическим, чем к функциональным (например, не задается доверительный интервал, но имеется возможность текстурирования или построения относительно нуля).

В Vokeh пришлось дополнительно объединять данные по двум столбцам исходной таблицы, поскольку группировка в явном виде не поддерживается (применить ее к имеющимся данным не удалось).

В столбиковой диаграмме, построенной средствами Mathplotlib/Seaborn, цвета устанавливаются библиотекой по умолчанию, а также по умолчанию присутствует легенда. В Plotly, дополнительно к этому, отображаются вертикальные линии сетки. В Vokeh линии также отображаются по умолчанию, но легенду необходимо добавлять отдельно.

Диаграмма радар. Для построения данной диаграммы использовались данные о результатах экзаменов, а именно средний балл по каждому из них. Среди рассматриваемых библиотек возможность построения данного типа диаграммы в явном виде реализована только в библиотеке Plotly (рисунок 4). При использовании двух других библиотек требуется написание дополнительного кода, что усложняет их использование для обучения.

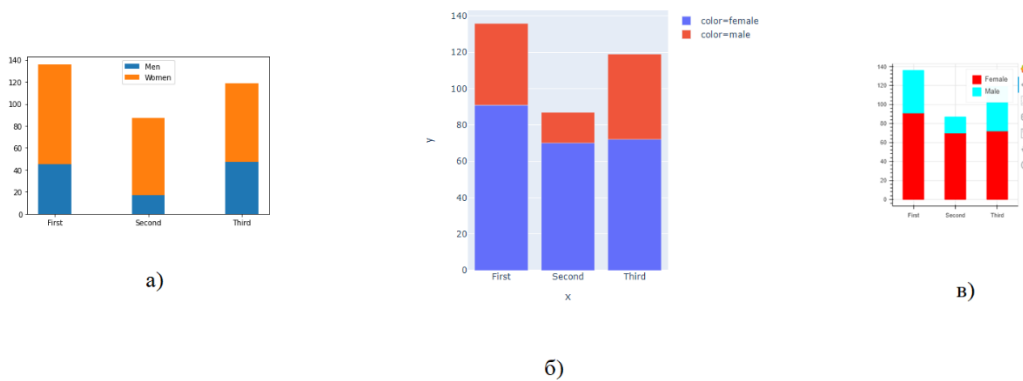


Рисунок 6. Сложенная столбиковая диаграмма, построенная средствами библиотек Matplotlib/Seaborn (а), Plotly (б), Bokeh (в)

В Matplotlib/Seaborn данная диаграмма является частным случаем обычной столбиковой диаграммы и строится путем установки начала отсчета с разных точек для разных групп (так, построение столбцов «Women» начинается со столбцов «Men»), что является ручным дополнением и может вызвать затруднения при построении впервые. Дополнительные возможности такие же, как у обычной гистограммы.

В Plotly построение сложенной гистограммы не вызвало трудностей и не потребовало дополнительных действий. Для построения указываются данные, и указываются столбцы, по которым выполняется группировка.

В Bokeh пришлось менять представление данных вручную для получения результата, что не вполне подходит для начинающих.

Легенда по умолчанию отображается только в Plotly.

Выводы

Анализ результатов проведенных экспериментов по построению диаграмм позволяет сделать следующие выводы.

Библиотека Bokeh в большинстве случаев требует дополнительных преобразований как данных, так и диаграмм, что осложняет ее применение для получения начальных навыков визуализации данных. Библиотека Matplotlib/Seaborn имеет меньше недостатков и не всегда требует непосредственной работы с данными, хотя в отдельных случаях необходимы дополнительные манипуляции с диаграммами (например, для столбиковой диаграммы). Из всех рассмотренных библиотек, с точки зрения простоты освоения и использования, лучше всех показала себя библиотека Plotly, в которой большинство вызовов функций достаточно просты и не требуют дополнительных преобразований данных.

В целом использование библиотеки Plotly можно признать наиболее подходящим решением для освоения методов визуализации данных с использованием Python. Определенным недостатком данной библиотеки можно признать используемый svg-формат представления диаграмм, что может приводить к ухудшению ее производительности на больших объемах данных. Библиотека Matplotlib/Seaborn больше подходит для разовых задач визуализации, и ее смогут легко освоить пользователи, имеющие опыт работы с системой MATLAB и подобными ей. Библиотека Bokeh является наиболее низкоуровневой, с большим количеством настроек и вызовов методов и функций, она хуже всех подходит для первых шагов в визуализации на Python.

Литература

1. Романова И.К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах//Наука и Образование. МГТУ им. Н.Э. Баумана. Электрон. журн. 2016. № 3. С. 133-167.
2. Matplotlib documentation [Электронный ресурс]. Режим доступа: <https://matplotlib.org/stable/index.html>
3. Seaborn 0.11.2 documentation [Электронный ресурс]. Режим доступа: <https://seaborn.pydata.org>
4. Plotly Python Graphing Library [Электронный ресурс]. Режим доступа: <https://plotly.com/python/>
5. Bokeh documentation [Электронный ресурс]. Режим доступа: <https://docs.bokeh.org/en/latest/>
6. Клышинский Э.С., Рысаков С.В., Шихов А.И. Обзор методов визуализации многомерных данных//Новые информационные технологии в автоматизированных системах. 2014. С. 519-530.
7. College Entrance Exam Scores in Vietnam 2019 [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/jonnyjackk/college-entrance-exam-scores-in-vietnam-2019>.
8. Garminder documentation [Электронный ресурс]. – Режим доступа: <https://www.garminder.org/data/documentation/>
9. Waskom M. Data repository for seaborn examples [Электронный ресурс]: Режим доступа: <https://github.com/mwaskom/seaborn-data>

УДК 004.043

ПРИМЕНЕНИЕ МЕТОДИКИ СТАТИЧЕСКОГО АНАЛИЗА ДЛЯ ВЫЯВЛЕНИЯ ОШИБОК В ПРОГРАММАХ НА ДИНАМИЧЕСКИХ ЯЗЫКАХ ПРОГРАММИРОВАНИЯ

APPLICATION OF STATIC ANALYSIS TECHNIQUES TO IDENTIFY ERRORS IN PROGRAMS IN DYNAMIC PROGRAMMING LANGUAGES

Попова В.А.,
ФГБОУ ВО «Иркутский государственный университет»,
г. Иркутск, Российская Федерация

V.A. Popova,
FSBEI HE “Irkutsk State University”,
Irkutsk, Russian Federation

e-mail: victorypopova1@gmail.com

Аннотация. Современные среды разработки содержат инструменты, которые позволяют без запуска программ выявлять ошибки и недостатки в исходных текстах на языках программирования, а также выполнять рефакторинг программного кода различной степени сложности. Так, повышается качество программных систем, поскольку значительное количество ошибок может быть выявлено ещё на этапе

разработки. Одной из важных проблем, которая существенно влияет на качество разрабатываемых программных систем, является возникновение ошибок несоответствия типов. Такая проблема может возникнуть в любом типизированном языке программирования. Но для языков программирования со статической типизацией ошибки несоответствия типов выявляются на этапе компиляции программы, тогда как для языков с динамической типизацией ошибки такого рода могут быть выявлены только во время выполнения программного кода. Для некоторых языков программирования с динамической типизацией разработаны инструменты, которые применяют методику статического анализа для выявления ошибок несоответствия типов на основании анализа программного кода. В данной работе описываются полученные результаты применения статического анализа к языку программирования с динамической типизацией «1С: Предприятие». Приводится процесс работы механизма статического анализа, определяется состав системы типов «1С: Предприятия», перечисляются ошибки, выявленные в процессе тестирования программного комплекса. Определяются дальнейшие этапы работы по созданию программного комплекса для статического анализа.

Abstract. Modern development environments contain tools to identify errors and flaws in source texts in programming languages without the need to run a program, as well as to perform refactoring of program code of varying degrees of complexity. Thus, the quality of software systems is improved, since a significant number of errors can be detected at the development stage. An important problem that significantly affects the quality of software systems being developed is the occurrence of type mismatch errors. This problem can occur in all programming languages. But for programming languages with static type-checking errors of type inconsistencies are detected even at the stage of compiling the program, whereas for languages with dynamic type-checking, errors of this kind can be detected only during the execution of the program code. For some programming languages with dynamic type-checking, tools have been developed that apply the static analysis methodology to identify type mismatch errors based on program code analysis. The article describes the obtained results of applying static analysis to the programming language with dynamic type-checking 1C: Enterprise. The process of operation of the static analysis mechanism is given, the composition of the type system of 1C: Enterprise is described, the errors identified during the testing of the software package are listed. The further stages of work on the creation of a software package for static analysis are determined.

Ключевые слова: качество программных систем, статический анализ кода, динамическая типизация, система типов, абстрактное синтаксическое дерево.

Keywords: quality of software systems, static code analysis, dynamic type-checking, type system, abstract syntax tree.

Введение

Для всех языков программирования существуют правила [1], которые требуется соблюдать при написании программного кода, чтобы разрабатываемая система была работоспособной, обладала необходимой функциональностью и соответствовала стандартам качества. Так, правила, ставящие в соответствие конструкциям программы типы данных, образуют логическую систему, которая называется системой типов [2, с. 17-18].

Языки программирования по характеристике типизации делятся на типизированные и бестиповые (нетипизированные).

Нетипизированными называются языки программирования, в которых доступно производить любые операции над данными, которые рассматриваются только в качестве последовательности бит различной длины [3]. К группе нетипизированных языков относятся как низкоуровневый язык ассемблер, так и высокоуровневые, например, BCPL, Forth.

В типизированных языках программирования для каждой операции определяется набор типов, к которому она может быть применима. Типизированные языки программирования по характеристике проверки соответствия типов конструкциям подразделяются на статические и динамические.

В программах, написанных на языках со статической типизацией, для устранения ошибок несоответствия типов требуется меньше затрат, поскольку ошибки выявляются в процессе разработки или тестирования, а не при использовании программы пользователями. В динамически типизированных языках программирования типы конструкций не вычисляются при компиляции, поскольку могут быть определены только в процессе выполнения программного кода.

В программах, написанных на языках программирования с динамической типизацией, существует риск выполнить недопустимое преобразование типов, из-за чего программа может аварийно завершать свою работу или не выполнять до конца ту или иную операцию, если разработчиком не была предусмотрена обработка критической ситуации. Ввиду подобных проблем пользователям становится труднее использовать программное обеспечение и требуется ожидать исправления ошибок разработчиками, на что может потребоваться достаточно большое количество времени.

Для отслеживания ошибок, возникающих в результате динамической проверки типов, следует использовать механизмы, которые позволяют анализировать код программы без её запуска. На сегодняшний день для подобных задач используется методика статического анализа программного кода.

Методика статического анализа

Для некоторых языков программирования с динамической типизацией существуют механизмы, которые позволяют выполнять статический анализ по исходному коду программ. Например, разработан алгоритм Хиндли-Милнера [4], который позволяет провести преобразования на уровне синтаксиса для автоматического определения типов выражений, используемых в программе. На системе типов Хиндли-Милнера основаны такие языки, как Standard ML, OCaml, Haskell.

Для языка программирования JavaScript компанией Google был создан механизм статической проверки типов на языке программирования Java в рамках проекта Closure Compiler.

Также существует набор технологий компилятора LLVM (Low Level Virtual Machine), которые позволяют проверять соответствия типов в поддерживаемых проектом LLVM языках программирования как со статической типизацией (C, C++), так и с динамической (Julia, PHP, Python и т.д.).

Существующие статические анализаторы программного кода осуществляют проверку корректности использования конструкций языка, исследуя абстрактное синтаксическое дерево, а также определяют вызовы методов, чтобы смоделировать поведение объектов в определённой точке выполнения программного кода [5].

Абстрактное синтаксическое дерево (Abstract Syntax Tree – AST) – это представление программного кода, которое отражает связи между элементами

выражений [6]. В AST вершинами являются операторы выражений, а листьями – аргументы (операнды).

Для того чтобы построить AST, необходимо выполнить лексический и синтаксический анализ программы [7].

Целью лексического анализа является выявление в программном коде лексем, к которым относятся ключевые слова, знаки операций, некоторые фиксированные значения (литералы) и т.д.

После определения лексем выполняется синтаксический анализ, суть которого заключается в сопоставлении лексем в соответствии с грамматикой, определённой для используемого языка программирования. На основании результатов синтаксического анализа можно выполнить построение AST.

Механизм статического анализа для языка программирования «1С: Предприятие»

К настоящему моменту для встроенного языка программирования в платформу «1С: Предприятие» не существует механизмов статического анализа, позволяющих выявлять ошибки несоответствия типов, которые могут возникать в разрабатываемых конфигурациях. В связи с этим возникла идея создания механизма статической проверки типов для программ, разработанных на платформе «1С: Предприятие».

Помимо применения правил вычисления типов по исходному тексту программы, при разработке механизма статической проверки необходимо использовать информацию о типах языка, встроенного в платформу «1С: Предприятие». Результаты работы по анализу системы типов и процессу извлечения информации о типах конфигурации «1С: Предприятия» описаны в [8] и [9] соответственно.

В рамках работы по разработке механизма статического анализа был создан формат для описания объектов конфигурации и типов платформы «1С: Предприятие». Такой формат был назван «Дерево типов конфигурации» (ДТК). Название обусловлено тем, что каждый объект конфигурации наследует функциональность одного или нескольких типов, определённых в платформе «1С: Предприятие».

Файл формата ДТК состоит из:

- типов платформы («ветки дерева») – служат основанием для объектов конфигурации;
- объектов конфигурации («листья дерева») – наследуют функциональность типов платформы.

Для того чтобы разработать средство статической проверки типов, требовалось получить представление программного кода конфигурации в формате AST, а уже далее выполнять анализ соответствия типов конструкциям языка, применяя информацию из документа формата ДТК.

Программный комплекс, который применяет методы статического анализа типов к свойствам языка программирования «1С: Предприятие», разрабатывается на языке программирования C++. К настоящему моменту реализовано следующее:

1. выполнение лексического и синтаксического анализа программного кода конфигураций «1С: Предприятия»;
2. преобразование структуры методов в представление AST;
3. выполнение процедуры статического анализа с применением информации о типах из документа формата ДТК.

Схема работы механизма статического анализа представлена на рисунке 1.

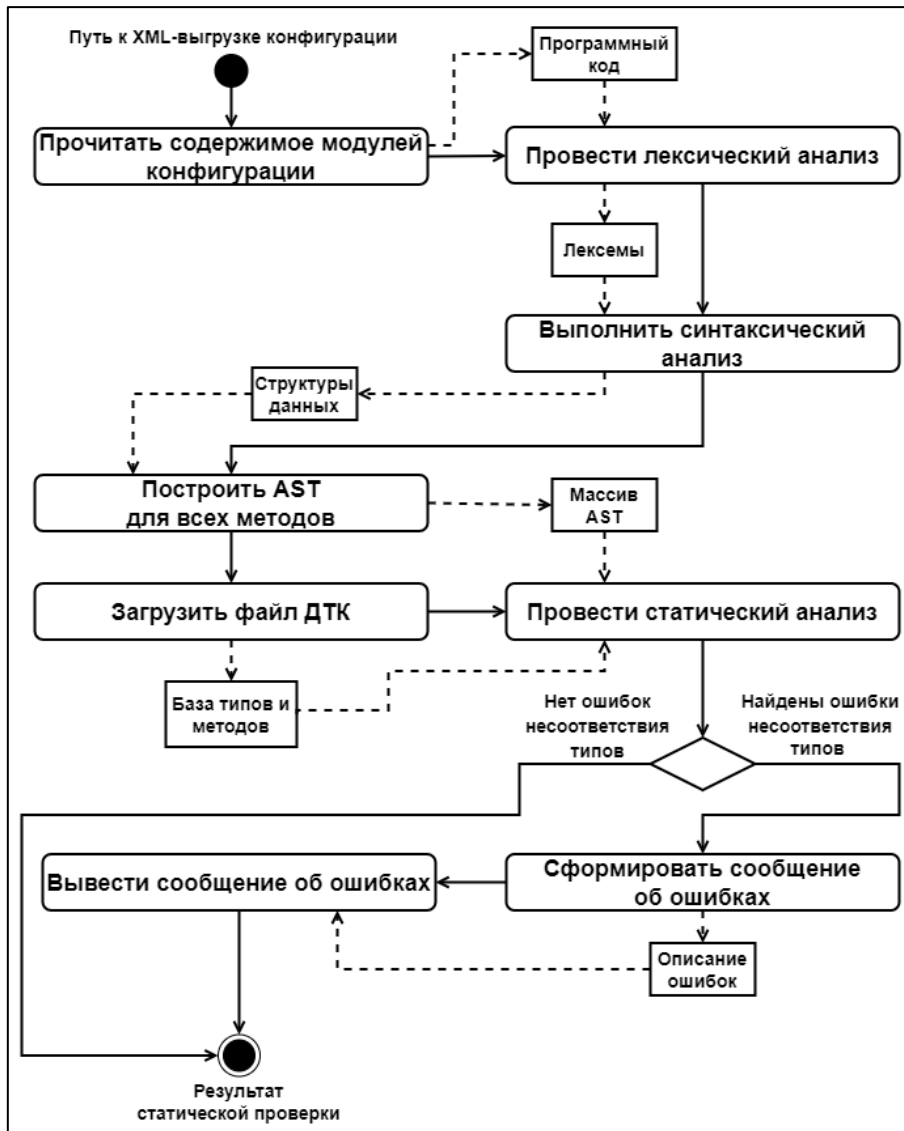


Рисунок 1. Процесс работы программы для статического анализа

Процесс статической проверки типов начинается из точек входа, которыми являются методы, определяющие такие события приложения, как, например, *ПриНачалеРаботыСистемы()* и *ПриЗавершенииРаботыСистемы()*.

В процессе тестирования программного комплекса были выявлены следующие ошибки:

- некорректной передачи параметров в метод;
- обращения к отсутствующему свойству объекта;
- попытке выполнить перебор, предполагая, что переменная имеет тип коллекции значений.

Например, если в метод платформы *СтрЗаменить(...)*, который содержит 3 параметра типа Строка, передать параметры другого типа, то будет сформировано сообщение об ошибке с указанием, какие параметры передаются некорректно. Информация об ошибках несоответствия типов записывается в файл. Для каждой ошибки указывается, в каком модуле и в какой строке она находится.

Затем осуществлялась доработка формата ДТК для того, чтобы при выполнении статического анализа также иметь возможность анализировать не только методы, но и события, которые инициируются пользователем при взаимодействии с формами,

располагающимися в пользовательском интерфейсе. Таким образом, из документации требовалось извлечь данные о событиях пользовательского интерфейса и, соответственно, дополнить файл формата ДТК.

Извлечённая информация о событиях позволила определить новые точки входа в программу: если ранее такими точками были события приложения, то теперь являются и реализованные события для элементов и команд, которые расположены в пользовательских формах. Это обусловлено тем, что выполнение программного кода осуществляется не только при запуске режима «1С: Предприятия». Некоторые фрагменты кода выполняются при работе пользователя с элементами и командами в случае, если для них переопределена функциональность событий встроенного языка.

На текущий момент ведётся работа над поиском ошибок в использовании элементов пользовательского интерфейса. Например, элементы могут быть вложены в группы, свойства которых можно настраивать в программном коде.

Выводы

В результате работы был усовершенствован на основании предыдущих наработок авторов механизм статического анализа для поиска ошибок несоответствия типов в программах, написанных на языке программирования «1С: Предприятие».

Разработанный программный комплекс может стать полезным инструментом в процессе создания конфигураций на платформе «1С: Предприятие», поскольку его функциональность позволяет находить значительное количество ошибок в программах ещё на этапе разработки.

Литература

1. Новиков К. Д., Раскатова М. В. Оптимизация программного обеспечения//Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление. – 2021. – № 1. – С. 159-165.
2. Benjamin C. Pierce. Types and Programming Languages. – Cambridge, Massachusetts London, England: The MIT Press, 2002. – 646 p.
3. Typing discipline: Untyped. – URL: <http://progopedia.com/typing/untyped/> (online, accessed 28.03.2022).
4. Milner R. A theory of type polymorphism in programming//Journal of Computer and System Sciences. – 1978. – Vol. 17. – P. 348-375.
5. Ермаков М.К., Вартанов, С.П. Применение статической бинарной инструментации с целью проведения динамического анализа программ для платформы ARM//Труды Института системного программирования РАН. – 2015. – № 1. – С. 5-24.
6. Jones J. Abstract Syntax Tree Implementation Idioms//Pattern Languages of Program Design. – 2003. – URL: <https://www.bibsonomy.org/bibtex/2a86095c92a9801ab24d2196d334abe46/gwpl> (online, accessed 28.03.2022).
7. Кузнецов М.А., Хорольский А.В. Теоретические аспекты разработки компилятора для учебных целей//Современные проблемы науки и образования. – 2013. – № 3. – С. 52-59.
8. Балюк А.С., Попова В.А. Разработка программного комплекса для конвертации конфигурации платформы «1С: Предприятие» в UML-модель//Сложные системы модели, анализ и управление. – 2021. – № 4. – С. 137-145.
9. Balyuk A.S., Popova V.A. Static type-checking for programs developed on the platform 1С: Enterprise//CEUR Workshop Proceedings: Irkutsk, 14 September 2021. – Irkutsk, 2021. – P. 101-111.

УДК 004

**ПОДХОД К ВОССТАНОВЛЕНИЮ ГЕОФИЗИЧЕСКИХ ДАННЫХ
МЕТОДОМ РЕЗЕРВИРОВАНИЯ****METHODS FOR RECOVERY OF GEOMAGNETIC DATA
GEOPHYSICAL DATA BY THE REDUNDANCY METHOD**

Снежко Ю.А., Огилько В.М.,
ФГБОУ ВПО «Уфимский государственный авиационный
технический университет»,
г. Уфа, Российская Федерация

Yu.A. Snezhko, V.M. Ogilko,
FSBEI NPE "Ufa State Aviation Technical University",
Ufa, Russian Federation

e-mail: julia.sneg2@gmail.com

Аннотация. Разработка ГИС имеет множество проблемных моментов, одним из которых является огромные массивы сырых данных с геомагнитных станций. Столь большой объем данных не может быть обработан в изначальном виде за приемлемое время, потому требуется предобработка таких данных. Помимо того, что нужно привести данные со станций в тот вид, при котором вычисления на их основе будут проводиться за адекватное время, требуются и другие виды обработки. Одним из них является восстановление пропусков в таких временных рядах. В работе рассмотрены различные методы восстановления временных рядов, в частности, геофизических данных. Эта задача, не смотря на долгие годы исследований, все еще остается актуальной. В статье описываются такие методы как аппаратное резервирование, сглаживание временного ряда, модель авторегрессии первого порядка, интегрированная модель авторегрессии, метод резервной станции, индуктивный метод, линейная интерполяция, сплайн интерполяция, кригинг и кокригинг. Приводится краткий анализ эффективности вышеперечисленных методов, их формальный вид. Проведенный анализ показал, что большая часть методов восстановления временных рядов на примере геомагнитных данных эффективны лишь на небольших фрагментах.

Abstract. The development of GIS has many problems, one of which is the huge arrays of raw data from geomagnetic stations. Such a large amount of data cannot be processed in its original form in an acceptable time, therefore, preprocessing of such data is required. In addition to the fact that it is necessary to bring the data from the stations into the form in which calculations based on them will be carried out in an adequate time, other types of processing are also required. One of them is the recovery of gaps in such time series. The paper considers various methods for restoring time series, in particular, geophysical data. This task is in spite of many years of research is still relevant. The article describes such methods as hardware redundancy, time series smoothing, first order autoregressive model, integrated autoregressive model, backup station method, inductive method, linear interpolation, spline interpolation, kriging and cokriging. A brief analysis of the effectiveness of the above methods, their formal form is given. The analysis performed showed that most of the methods for restoring time series using geomagnetic data as an example are effective only on small fragments.

Ключевые слова: интерполяция, геоданные, восстановление временных рядов, обработка временных рядов, информационное резервирование.

Keywords: interpolation, geodata, time series recovery, time series processing, information redundancy.

Введение

При разработке ГИС актуальной проблемой является оптимизация обработки большого количества данных. Обработка данных в исходном состоянии требует высокой производительности от устройства и может занять несоизмеримо большое количество времени. Кроме того, в этих данных часто наблюдаются пропуски, которые необходимо заполнить до дальнейшей обработки.

На данный момент составление достоверной карты геомагнитного поля Земли – задача, требующая точности в расположении точек. Чтобы составить такую карту нужно получить огромное количество данных с геомагнитных станций. Они расположены по всему земному шару и, в настоящее время, непрерывно регистрируют данные о геомагнитном поле.

Таким образом, этап получения данных за определенный промежуток времени сильно облегчен, но, несмотря на это, при размещении таких данных на карте остаются «белые пятна», где параметры по той или иной причине записаны не были и, в частности в сети ИНТЕРМАГНЕТ, заменены на 88888 и 99999 [1]. Для ликвидации таких пробелов требуется произвести дополнительные вычисления.

Известные решения

Наиболее очевидным методом решения проблемы пропусков геофизических данных является метод аппаратного резервирования, который подразумевает регистрацию данных параллельно с помощью нескольких устройств, но он влечет расходы на дополнительное оборудование и далеко не все организации в силу своих финансовых возможностей могут обеспечить достаточное количество резервных устройств [10]. Именно поэтому многие исследователи решают прибегнуть к аналитическим методам.

Например, с помощью линейной интерполяции, которая заключается в подборе коэффициентов полинома первой степени, который задан уравнением прямой, соединяющей известные значения уровней ряда [2]. Формально метод можно представить в следующем виде (рисунок 1):

$$\frac{(y_2 - y_1)}{(x_2 - x_1)} = \frac{(y - y_1)}{(x - x_1)};$$

$$x_1 \leq x \leq x_2; \quad y = ax + b;$$

$$a = \frac{(y_2 - y_1)}{(x_2 - x_1)}; \quad b = y_1 - ax_1,$$

Рисунок 1. Формальный вид метода линейной интерполяции

где x_1 и y_1 – первая крайняя точка пропуска и значение уровня в ней, x_2 и y_2 – вторая крайняя точка пропуска и значение уровня в ней,

x и y – пропущенная точка и значение уровня в ней,
 a, b – коэффициенты построенной прямой.

Но данный метод обеспечивает достаточно точные результаты лишь на единичных пропусках и небольших фрагментах данных, для больших серий данный способ не подходит в силу слишком большой величины среднеквадратической ошибки [3].

Так же можно, используя сплайн-интерполяцию, разбить область интерполяции на участки с небольшим количеством узлов, на которых строится свой интерполирующий полином, но при дополнительных условиях согласованности с интерполяцией на соседних участках [4].

Если в узлах x_i, x_{i+1} , заданы значения f_i, f_{i+1} , которые принимает кубический сплайн, то на частичном отрезке $[x_i, x_{i+1}]$, он принимает вид (рисунок 2):

$$S_3(x) = \frac{(x_{i+1}-x)^2(2(x-x_i)+h)}{h^3} f_i + \frac{(x-x_i)^2(2(x_{i+1}-x)+h)}{h^3} f_{i+1} + \frac{(x_{i+1}-x)^2(x-x_i)}{h^2} m_i + \frac{(x-x_i)^2(x-x_{i+1})}{h^2} m_{i+1}$$

Рисунок 2. Формальный вид кубического сплайна

Метод кластеризации состоит из двух частей: анализа временного ряда и прогнозирования на основе выделенных последовательностей. Он основывается на корреляции между данными со станций и определяет поведение рассматриваемого временного ряда и выделяет характерные последовательности. Данный метод позволяет достичь более точных результатов на фрагментах ряда большой длины, но вычисления требуют куда больших затрат в сравнении с другими.

Кригинг является еще одним методом интерполяции, в частности, он относится к статистическим методам. Он позволяет исследовать пространственные автокорреляции между геоданными, в частности, решает два типа задач: определение связанной пространственной структуры данных и прогнозирование. В основе метода лежит принцип несмещённости среднего. Использование данного метода подразумевает, что расстояние или направление между опорными точками отражает пространственную корреляцию, которая может использоваться для объяснения изменения на поверхности. Данный метод лучше всего использовать, если известно, что есть пространственно коррелированное расстояние или направленное смещение в данных [4]. Все модели семейства кригинга так или иначе сводятся к линейной регрессионной оценке. Оценка данного метода можно представить следующим образом (рисунок 3):

$$Z^*(x) = \sum_{i=1}^{n(x)} \lambda_i(x) Z(x_i)$$

Рисунок 3. Оценка кригинга

Кокригинг является еще одним статистическим методом и обобщением кригинга на случай многопеременных данных, при наличии пространственной корреляции. Основная переменная оценивается на основе ее собственных измерений и данных по другим переменным. Знание всех переменных во всех точках не требуется. Оценка обычного кокригинга функции $Z_{a_0}(x_0)$ – линейная комбинация значений различных переменных из окрестности точки x_0 (рисунок 4) [4]. Количество участников оценивания среди различных переменных n_a может быть различно:

$$Z_{\alpha_0}^*(\mathbf{x}_0) = \sum_{\alpha=1}^K \sum_{i=1}^{n_\alpha} \lambda_i^\alpha Z_\alpha(\mathbf{x}_i)$$

Рисунок 4. Оценка кокригинга

Высокая степень автокорреляции геомагнитных данных позволяет предполагать тесную линейную осциллирующую связь соседних уровней временного ряда, и, следовательно, применять прогностические модели и методы. В некотором роде можно сказать, что прогноз отсутствующего фрагмента данных будет основываться на известных уровнях ряда, предшествующих ему, так как они выступят в роли обучающей выборки. Таким образом задачу восстановления части данных можно представить как прогнозирование недостающего фрагмента ряда на основании предшествующих ему. Например, можно использовать модель авторегрессии (AR) первого порядка, которая позволяет оценивать изменение целевой переменной в зависимости от лишь одного фактора – её собственного значения в прошлом периоде авторегрессии [5]. Этот метод тоже обеспечивает хорошую точность при небольших пропусках. В случае восстановления фрагментов более 5 значений среднеквадратическая ошибка экспоненциально возрастает [6].

Интегрированная модель авторегрессии – скользящего среднего (ARIMA) показывает более качественные результаты прогнозирования пропущенных данных в сравнении с AR за счёт гибкой параметризации обработки данных [6]. Наилучший результат достигается при первом порядке авторегрессии с нулевым порядком интегрирования и единичным порядком скользящего среднего. В случае восстановления фрагментов порядка 5 – 7 значений ошибка достаточно мала.

В отличие от метода авторегрессии, при применении метода скользящего среднего каждое значение временного ряда подвержено суммарному воздействию предыдущих ошибок

Под сглаживанием временного ряда методом скользящей средней [7] подразумевается замена пропущенного фрагмента данных средним значением соседних элементов. Данный метод является одним из эмпирических методов для сглаживания и прогнозирования временных рядов. Он показывает высокую эффективность в случае единичных пропусков, но при увеличении числа пропущенных значений наблюдается экспоненциальный рост среднеквадратической ошибки. К примеру, при восстановлении 10-минутного фрагмента временного ряда геофизических данных этим методом среднеквадратическая ошибка равняется 1,3 нТл, что не является допустимым значением погрешности измерений [8].

При использовании индуктивного метода предполагается, что если пара непоследовательных фрагментов временного ряда, разделённых пропущенным фрагментом, оказывается достаточно близкой к паре фрагментов, разделённых известным фрагментом, то значения между ними, в по теореме Такенса, будут отличаться статистически незначительно. Чтобы ускорить обработку данных, берутся значения за год, так как исследования [6] показали, что увеличение временного отрезка приводит только к значительным вычислительным затратам, но не влияет на эффективность метода восстановления.

В результате исследования выяснилось, что наименьшая ошибка восстановления обеспечивает частота среза фильтра нижних частот не более 10 минут. При фрагментах до 30 минут ошибка не превышает допустимые значения. Таким образом для получения достаточно точных результатов необходимо чтобы индекс геомагнитной активности не превышал 1, данные должны быть взяты с геомагнитных станций в средних географических широтах и длительность восстанавливаемого сигнала не должна

превышать 10 значений уровня для оригинальных временных рядов и 1440 – для фильтрованных геофизических данных.

Кроме вышеперечисленных так же предлагался метод резервной станции, который представляет собой способ пространственно-информационного резервирования, суть которого заключается в том, что что сильная корреляционная связь наборов данных является основанием для взаимозаменяемости соответствующих временных рядов. То есть, пропущенные значения можно заменить на фрагмент данных другой обсерватории, который был зарегистрирован в то же время [9]. Формально метод может быть выражен следующим образом (рисунок 5):

$$\sigma_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}},$$

Рисунок 5. Формальный вид метода резервной станции

где $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ и $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$ – средние значения выборок X и Y соответственно.

Кроме того, для корректного расчета коэффициента корреляции необходимо, чтобы количество значений в исследуемых переменных X и Y было одинаковым, исследуемые переменные X и Y были распределены нормально и измерены в интервальной шкале или шкале отношений.

Выводы

Значительное число моделей и методов применяется для решения задачи пропусков во временных рядах, но большинство из них эффективны лишь на малых фрагментах, потому проблема все еще требует дальнейшего изучения.

Работа поддержана грантом РФФИ № 20-07-00011-а.

Литература

1. INTERMAGNET technical reference manual. Version 4.6 /ed. by S.-L. Benoît. – Edinburgh: INTERMAGNET, BGS, 2012.
2. Волков, Е.А. Численные методы/Е.А. Волков. – М.: Наука, 1987.
3. Soloviev, A. Mathematical tools for geomagnetic data monitoring and the INTERMAGNET Russian segment/A. Soloviev [et al.]/Data Science Journal. – 2013. – Vol. 12. – P. WDS114-WDS119
4. Демьянов В.В., Савельева Е.А. Геоestatистика: теория и практика. – 2010.
5. Бокс, Дж. Анализ временных рядов (прогноз и управление)/Дж. Бокс, Г. Дженкинс. – Вып. 1. – М.: Мир, 1974. – 408 с.
6. Воробьева Г.Р. Подход к восстановлению геомагнитных данных путём сопоставления суточных фрагментов временного ряда с равной геомагнитной активностью//Компьютерная оптика. – 2019. – Т. 43. – №. 6. – С. 1053-1063.
7. Кулаичев, А.П. Методы и средства комплексного анализа данных/А.П. Кулаичев. – М.: «ФОРУМ: ИНФРА», 2006. – 512 с.
8. Aue A., Dubart D., Hörmann S. On the prediction of stationary functional time series. Journal of the American Statistical Association, 2015, No. 110(509), pp. 378-392.
9. Воробьев А.В., Воробьева Г.Р. Методы резервирования в задачах

восстановления временных рядов геомагнитных данных – Информационные процессы, Том 19. – 2018. – №. 1

10. Воробьев, А.В. Концепция единого пространства геомагнитных данных/А.В. Воробьев, Г.Р. Воробьева, Н.И. Юсупова//Труды СПИИРАН. – 2018. – № 18(2).

УДК 004.932

СОКРАЩЕНИЕ СВЕРТОЧНЫХ СЛОЕВ

REDUCTION OF CONVOLUTIONAL LAYERS

Хлыбов А.В., Гиниятуллин В.М., Ермолаев Е.В.,
Уфимский государственный нефтяной технический университет,
ул. Космонавтов, 1, г. Уфа, Республика Башкортостан, 450064, Россия

A.V. Khlybov, V.M. Giniyatullin, E.V. Ermolaev,
Ufa State Petroleum Technological University,
Kosmonavtov Str., 1, Ufa, Republic of Bashkortostan, 450064, Russia

e-mail: brinkinvision@gmail.com

Аннотация. В статье описано исследование по упрощению существующей архитектуры нейронной сети LeNet-5 для обработки набора данных MNIST. Для достижения необходимой точности количество элементов модели является избыточным, поэтому предлагается два способа для уменьшения структуры: преобразование коэффициентов сверточных матриц и удаление слоев нейронной сети. Оказалось, что усечение десятичных знаков после запятой у коэффициентов ядер сверток первого сверточного слоя не оказывает практически никакого влияния на итоговый результат модели. После этого было проведено усечение диапазона до меньшей значности. Сведение к трехзначности диапазона изменения коэффициентов дает результаты с приемлемой точностью. Также представлены результаты работы модели при уменьшении элементов, входящих в ее состав. Удалось свести структуру сети к двум слоям: сверточному и полносвязному. При этом сверточный слой состоит только из одного ядра. Представленные результаты показывают возможность и необходимость оптимизации нейронных сетей, что позволит снизить требования к вычислительным ресурсам. Требуется разработка такой формализованной методики упрощения, с помощью которой можно четко понимать, какие элементы нейронной сети требуют изменения и к чему приведут эти изменения.

Abstract. The article describes a study to simplify the existing LeNet-5 neural network architecture for processing the MNIST dataset. To achieve the required accuracy, the number of model elements is redundant, so two methods are proposed to reduce the structure: transforming the coefficients of convolution matrices and removing layers of the neural network. It turned out that the reduction of decimal places after the decimal point for the coefficients of the convolution kernels of the first convolutional layer has practically no effect on the final result of the model. After that, the range was truncated to a lower value. Reduction to the three-valued range of coefficients gives results with acceptable accuracy. Also presented are the results of the model with a decrease in the elements included in its composition. It was

possible to reduce the network structure to two layers: convolutional and fully connected. In this case, the convolutional layer consists of only one kernel. The presented results show the possibility and necessity of optimizing neural networks, which will decrease the requirements for computing resources. It is required to develop such a formalized simplification technique, with which it is possible to clearly understand which elements of the neural network require changes and what these changes will lead to.

Ключевые слова: нейронная сеть, черный ящик, сверточный слой, весовые коэффициенты, LeNet-5.

Keywords: neural network, black box, convolutional layer, weight coefficients, LeNet-5.

Сложность существующих нейронных сетей растет вместе с объемами данных, которые необходимо обработать, и вычислительными мощностями, предназначенными для их обработки [1]. Для таких сетей уже недостаточно обычного стационарного компьютера [2].

Из-за наличия огромного количества параметров в нейронных моделях, становится сложно определить причины возникающих ошибок. Весовые коэффициенты, полученные в процессе обучения сети, невозможно никак интерпретировать («черный ящик») [3].

Отсюда возникает необходимость нахождения путей избавления от неопределенностей современных искусственных нейронных сетей. Наличие каждого слоя и нейрона нужно строго обосновывать. Необходимо четко понимать, когда целесообразно добавить элемент в нейронную сеть, а когда это добавление в результате не сыграет существенной роли. Затраченные ресурсы на получение требуемой точности не должны многократно превышать прирост этой точности.

Целью данной работы является экспериментальное исследование возможности упрощения различных параметров сверточной нейронной сети.

В качестве базовой архитектуры нейронной сети выбрана модель LeNet-5 для обработки изображений рукописных цифр MNIST [4]. В таблице 1 представлена исходная структура сети.

Таблица 1 – Структура модели LeNet-5

| Элемент | Основные параметры |
|-------------------|------------------------|
| Сверточный слой | 6 ядер 5*5, вход=28*28 |
| Maxpooling | Размер 2*2, шаг=2 |
| Сверточный слой | 16 ядер 5*5 |
| Maxpooling | Размер 2*2, шаг=2 |
| Полносвязный слой | 120 нейронов |
| Полносвязный слой | 84 нейрона |
| Полносвязный слой | 10 нейронов |

Модель была реализована с помощью библиотеки Tensorflow на языке программирования Python [5]. Первоначальная точность модели без изменений: по обучающей выборке = 99.28%, по тестовой = 98.58%.

Очевидно, что первый сверточный слой влияет на все последующие элементы модели, поэтому упрощение было начато с него. Весовые коэффициенты ядер первого

сверточного слоя преобразовывались. Для этого сокращалось количество десятичных знаков после запятой у коэффициентов.

В данном случае для значений используется тип float, поэтому количество знаков после запятой 7-8. Последовательно усекались значащие цифры после запятой, и замерялась точность сети как по обучающей выборке, так и по тестовой (таблица 2).

Из таблицы 2 следует, что усечение коэффициентов ядер первого сверточного слоя уже обученной сети практически не влияет на точность распознавания.

Таблица 2 – Усечение десятичных знаков

| Количество десятичных знаков после запятой | Точность по обучающей выборке, % | Точность по тестовой выборке, % |
|--|----------------------------------|---------------------------------|
| 8 | 99.28 | 98.58 |
| 7 | 99.57 | 98.86 |
| 6 | 99.62 | 99.14 |
| 5 | 99.54 | 98.91 |
| 4 | 99.65 | 99.06 |
| 3 | 99.77 | 98.97 |
| 2 | 99.75 | 99.02 |
| 1 | 99.65 | 98.95 |

Далее сокращалось количество слоев в сети, начиная с конца (не затрагивая выходной слой), вплоть до первого сверточного слоя. Результаты сведены в таблицу 3.

Таблица 3 – Удаление элементов нейронной сети

| Удаленный элемент сети | Точность по обучающей выборке, % | Точность по тестовой выборке, % |
|---|----------------------------------|---------------------------------|
| Полносвязный слой с 84 нейронами | 99,63 | 98,79 |
| Полносвязный слой с 120 нейронами | 99,14 | 98,69 |
| Сверточный слой с 16 ядрами + 2 maxpooling-слоя | 99,38 | 97,96 |

Из таблиц 3 видно, что количество элементов исходной сети было избыточно для набора данных MNIST. Для решения конкретной задачи (распознавание по классам рукописных цифр) вполне хватает одного сверточного слоя с 6 ядрами 5*5 и одного полносвязного слоя с 10 нейронами. Коэффициенты ядер сверточного слоя представлены в таблице 4.

Таблица 4 – Ядра первого сверточного слоя

| Ядро 1 | | | | |
|--------------|---------------|---------------|----------------|---------------|
| -0.174333170 | -0.0665594637 | -0.0831459686 | -0.658914685 | -0.939356208 |
| 0.109362051 | 0.0769770667 | 0.0623391792 | 0.360517472 | -0.129954070 |
| 0.216657057 | 0.211373553 | 0.390184462 | 0.492284626 | 0.418038458 |
| 0.0564525202 | 0.125601754 | 0.249569997 | 0.0341770165 | 0.336711198 |
| -0.412444502 | -0.505079687 | -0.217631117 | -0.382181495 | -0.0697908401 |
| Ядро 2 | | | | |
| -0.234307528 | -0.0480053611 | -0.228880003 | 0.338677675 | 0.187788695 |
| -0.157286912 | -0.764863729 | -0.510300577 | -0.00276786252 | 0.200396776 |
| -0.351606339 | -0.472955257 | -0.0965034515 | 0.159918457 | 0.175715968 |
| -0.413525492 | -0.398823708 | 0.0963655040 | 0.400702506 | 0.126797870 |
| -0.825033188 | -0.133665442 | 0.345471352 | 0.444564223 | 0.118785664 |

Продолжение таблицы 4

| Ядро 3 | | | | |
|----------------|-----------------|----------------|--------------|---------------|
| -0.294651508 | 0.111423522 | 0.199960917 | 0.445944935 | 0.0856154636 |
| 0.247437075 | 0.0515066683 | 0.140456662 | 0.111505836 | -0.385869354 |
| 0.375180751 | 0.120566905 | 0.234505773 | -0.101796642 | -0.622854829 |
| 0.425762922 | 0.0194274969 | -0.710630715 | -0.792543352 | -0.0122394534 |
| -0.299817294 | -0.652545631 | -0.0727587566 | 0.146831110 | 0.811151028 |
| Ядро 4 | | | | |
| 0.273966879 | -0.000763709191 | 0.0789451748 | 0.0556988902 | 0.456944764 |
| -0.00120917079 | 0.292878896 | 0.210437775 | 0.344226152 | 0.511845589 |
| -0.350831926 | 0.360385835 | 0.422464669 | 0.0351020023 | 0.0449313037 |
| -0.918258429 | -0.288373977 | -0.179938003 | -0.536574423 | -0.758768260 |
| -0.951341569 | -0.719964445 | -0.982500136 | -0.398845404 | -0.293000162 |
| Ядро 5 | | | | |
| 0.157561928 | 0.273087591 | 0.201148212 | -0.207157582 | -0.482943743 |
| 0.0314159133 | 0.125312895 | 0.214353770 | -0.443870068 | -0.365781724 |
| 0.0955079645 | 0.283867419 | -0.0265890863 | -0.297105253 | -0.543140829 |
| 0.0497687869 | 0.136038646 | 0.183023721 | -0.709116220 | -0.394453377 |
| 0.300111771 | 0.384916514 | -0.0416037329 | -0.611609578 | -0.370249689 |
| Ядро 6 | | | | |
| -0.130261615 | -0.0505145155 | 0.315594763 | 0.0477402397 | -0.194517374 |
| 0.199525759 | 0.174246103 | 0.0111735202 | 0.0996073112 | -0.232725248 |
| -0.122273356 | -0.202995360 | -0.00907654036 | 0.318423659 | 0.0100705335 |
| -0.294115812 | -0.133854032 | 0.0498351976 | 0.250717670 | 0.00427915901 |
| 0.292412817 | -0.247929960 | -0.379034817 | 0.281592131 | -0.0602287762 |

Коэффициенты таблицы 4 также были подвергнуты усечению количества знаков после запятой. На такой сокращенной структуре коэффициенты сверток должны оказывать большее влияние на конечный результат.

В итоге после сокращения до одного десятичного знака после запятой точность по обучающей выборке=99.70 %, по тестовой=98.34 %.

В таблице 5 представлены результаты удаления ядер первого сверточного слоя.

Таблица 5 – Сокращение количества ядер первого сверточного слоя

| Количество ядер в слое | Точность по обучающей выборке, % | Точность по тестовой выборке, % |
|------------------------|----------------------------------|---------------------------------|
| 6 | 99.83 | 98.12 |
| 5 | 99.26 | 98.02 |
| 4 | 98.96 | 97.88 |
| 3 | 98.57 | 97.65 |
| 2 | 97.70 | 96.95 |
| 1 | 93.53 | 93.21 |

Коэффициенты оставшегося ядра были усечены до одного значащего знака после запятой.

В таблице 6 представлено полученное ядро. Точность сети составила: по обучающей выборке = 93.30%, по тестовой = 92.98%.

Таблица 6 – Усеченные коэффициенты ядра сверточного слоя

| | | | | |
|------|-----|------|------|------|
| 0.4 | 0.3 | 0.5 | -0.2 | -0.3 |
| 0.3 | 0.3 | -0.3 | -0.4 | -0.4 |
| 0.5 | 0.1 | 0.1 | -0.2 | -0.9 |
| -0.1 | 0.5 | 0.1 | 0.6 | 0.1 |
| 0 | 0.2 | -0.1 | 0.4 | 0.6 |

Максимальное значение равно 0.6, минимальное -0.9. Значность ядра была сведена к трехзначности: $[-0.9; -0.4]=-1$, $[-0.3; 0.3]=0$, $[0.4; 0.9]=1$ (таблица 7). Точность по обучающей выборке составила 93.75 %, по тестовой 93.24 %.

Таблица 7. Трехзначное сверточное ядро

| | | | | |
|---|---|---|----|----|
| 1 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | -1 | -1 |
| 1 | 0 | 0 | 0 | -1 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 0 | 0 | 1 | 1 |

Выводы

Таким образом, полученные результаты говорят о возможности и необходимости упрощения существующих нейронных сетей. Упрощение возможно не только по пути сокращения структуры модели, но и при помощи замены весовых коэффициентов, представляющих собой числа с плавающей запятой, на дискретные значения. Это позволит ускорить процесс обучения сети и снизить необходимость в высокой вычислительной мощности.

Литература

1. Виноградов О.В., Морозова О.А. Аспекты применения нейронных сетей для прогнозирования чрезвычайных ситуаций//Технологии гражданской безопасности, том 18. 2021. – С. 23-26.
2. Filippova A.G., Filippov V.N., Sultanova E.A. Author's Method of Designing Unified User Interface Web-Sites World Applied Sciences Journal. 2014. Т. 29. № 6. С. 758-763.
3. Андреев О.А., Трофимов А.Т. Интерпретация весовых коэффициентов и функций активации искусственных нейронных сетей//Цифровая обработка сигналов и ее применение – DSPA-2019, Москва. 2019. – С. 565-569.
4. Module05_mnist_conv. ipynb [Электронный ресурс]. URL: https://github.com/SlinkoIgor/Neural_Networks_and_CV/blob/master/module05_mnist_conv.ipynb (дата обращения: 22.03.2022).
5. Nurgaliev A.R., Filippov V.N., Loginova M.E. Application of Information Technologies at The Design Stage of Oil and Gas Wells Information Technology. Problems and Solutions. 2020. № 1 (10). С. 21-26.

УДК 004

**РАЗРАБОТКА ЧАТ-БОТА ДЛЯ ОБЕСПЕЧЕНИЯ НЕПРЕРЫВНОЙ РАБОТЫ
ЕДИНОЙ ЦИФРОВОЙ ПЛАТФОРМЫ**

**DEVELOPMENT OF A CHAT-BOT TO ENSURE THE CONTINUOUS OPERATION
OF A SINGLE DIGITAL PLATFORM**

Шайхутдинов Р.Э., Масгутова И.С.,
Институт нефтепереработки и нефтехимии ФГБОУ ВО УГНТУ в г. Салавате,
ул. Губкина, 22 б, г. Салават, Республика Башкортостан,
Россия, 453250

Shaukhytdinov R.E., Masgutova I.S.,
Institute of Oil Refining and Petrochemistry FSBEI HE USPTU in Salavat,
Gubkin Str., 22b, Salavat, Republic of Bashkortostan,
453250, Russia

e-mail: shaykhutdinov.roma@mail.ru

Аннотация. В современных предприятиях получила широкий разворот цифровизация процессов, которая в свою очередь понесла за собой ряд проблем, связанных в большинстве случаев с компьютерной безграмотностью сотрудников, а также и со сложностью в обращении с новыми высокотехнологическими приборами. В настоящее время существует проблема бесперебойной службы обработки ИТ заявок для поддержки персонала медицинских учреждений, вызванная разными факторами, начиная от отсутствия call-центра, заканчивая плотным графиком специалистов технической поддержки. Из-за данной проблемы работоспособность и продуктивность работников сильно снижается. Одно из решений данной проблемы является создание чат-бота для обеспечения онлайн консультации и автоматизации обработки ИТ заявок. Поступающие обращения от медиков в ИТ отдел требуют немедленного реагирования, зачастую выявленные проблемы существенно влияют на работоспособность медицинских работников и требуют большего количества времени на их решение, в отличие от обращений по поводу консультации, которые в свою очередь времени почти не занимают, но преобладают количеством. Чат-боты, или виртуальные собеседники, используются в диалоговых системах для различных практических целей, включая обслуживание клиентов или сбор информации. Некоторые чат-боты используют сложные системы обработки естественного языка. По исследованиям российского рынка, 64% пользователей ответили, что преимуществом чат-ботов для бизнеса является доступность и активность 24/7. Кроме того, пользователи находят удобным получение ответов на простые вопросы, а также легкость в коммуникации.

Abstract. In modern enterprises, digitalization of processes has received a wide turn, which in turn has entailed a number of problems associated in most cases with computer illiteracy of employees, as well as with the difficulty in handling new high-tech devices. Currently, there is a problem with the uninterrupted service of processing IT requests to support the personnel of medical institutions, caused by various factors, ranging from the lack of a call center to the tight schedule of technical support specialists. Because of this problem, the efficiency and productivity of workers is greatly reduced. One of the solutions to this problem is the creation of a chatbot to provide online consultations and automate the processing of IT

requests. In most cases, doctors contact the IT department for consultations at the EDP, which puts a lot of pressure on the schedule, and other important applications are postponed to a later date. Because of this problem, the efficiency and productivity of workers is greatly reduced. One of the solutions to this problem is the creation of a chatbot to provide online consultations and automate the processing of IT requests. Incoming requests from doctors to the IT department require an immediate response, often identified problems significantly affect the performance of medical workers and require more time to resolve them, in contrast to requests for a consultation, which, in turn, take almost no time, but prevail in number. Chatbots, or virtual interlocutors, are used in conversational systems for a variety of practical purposes, including customer service or information gathering. Some chatbots use sophisticated natural language processing systems. According to Russian market research, 64% of users answered that the advantage of chatbots for business is availability and activity 24/7. In addition, users find it convenient to get answers to simple questions, as well as ease of communication.

Ключевые слова: инструктаж, консультация, заявка, ЕЦП, IT специалист, обращение, чат-бот.

Keywords: briefing, consultation, application, EDP, IT specialist, appeal, chatbot.

Были проанализированы все поступающие обращения к сотрудникам технической поддержки в течении месяца, и похожие были объединены в группы, указанные ниже:

- обращения, касающиеся неполадок с сервером. Зачастую такие проблемы выявляются сразу на нескольких ПК, случаются из-за различных факторов;
- обращение по поводу консультации, зачастую появляется из-за компьютерной безграмотности врачей;
- обращения, связанные с интернет-подключением. Неполадки зачастую появляются из-за проблем на стороне провайдера;
- обращения из-за проблем с оборудованием (сломанные мышки, клавиатуры, неисправные системные блоки или мониторы);
- обращение из-за серьезных проблем, зачастую занимающие большую часть времени сотрудника технической поддержки проявляются из-за различных факторов.

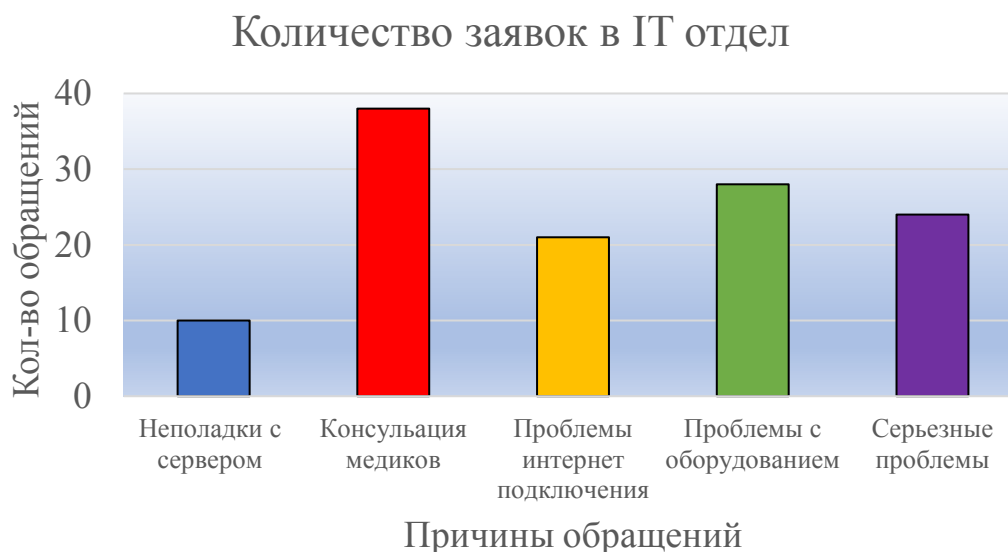


Рисунок 1. График количества обращений за месяц

Так же было определено количество времени, отводящегося на устранение неполадок.

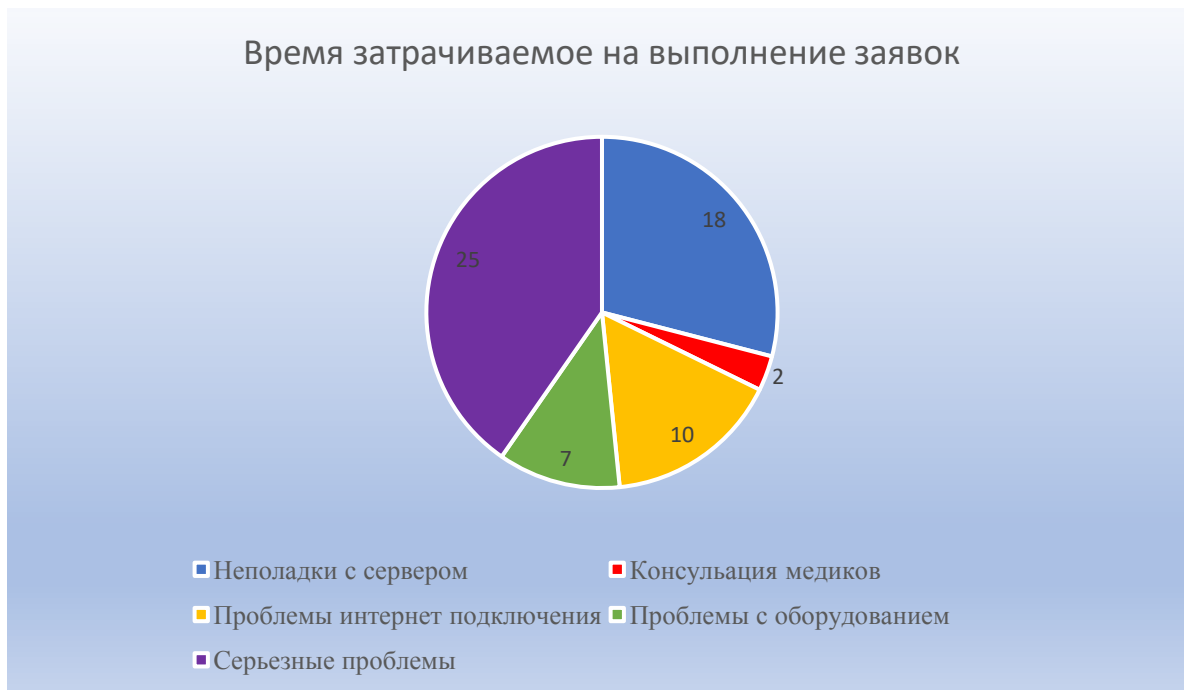


Рисунок 2. Время, затрачиваемое на выполнение заявок

На графике мы можем увидеть распределение рабочего времени ИТ специалиста, на решение тех или иных проблем. Сильно выделяется на фоне остальных время выполнения заявок красной зоны, она составляет около 4% от всего времени сотрудника технической поддержки, но судя по предыдущему графику мы можем сказать, что эта зона преобладает над остальными и занимает лидирующие позиции. Из всего этого мы можем сделать вывод, что большой промежуток времени уходит на решение проблем, не требующих присутствия специалиста. В следствии чего многие медицинские сотрудники остаются без помощи – работоспособность падает.

Одним из возможных методов решения данной проблемы, является разработка Чат-бота для своевременного реагирования на обращения медиков о консультации.

Чат-боты имеют свои особенности, в большинстве своем направленные на работу с клиентами, многие решения размещены в мессенджере «Telegram» что не позволяет достигнуть какой-либо организованности в предоставлении информации. Это одна из основных проблем данных программ. Так же можно отметить, что преобладающее большинство ботов не является отечественным решением.

Существуют разные виды чат-ботов. Одни нацелены на информирование клиентов, другие ориентированы на продажи, третьи используются исключительно в качестве личных помощников. Все зависит от функционала, заложенного в программу.

Разрабатываемый чат бот возьмет на себя задачу консультации и позволит сотрудникам тех поддержки не отвлекаться на обращения о консультации врачей, а заниматься более серьезными неполадками.

При нажатии на значок чат-бота открывается диалоговое окно, в котором врач может задать свой вопрос и получить полноценную консультацию по заданному вопросу, как указано на рисунке 3. В случае, если чат-бот не смог осуществить поддержку он предлагает обратиться к специалисту.

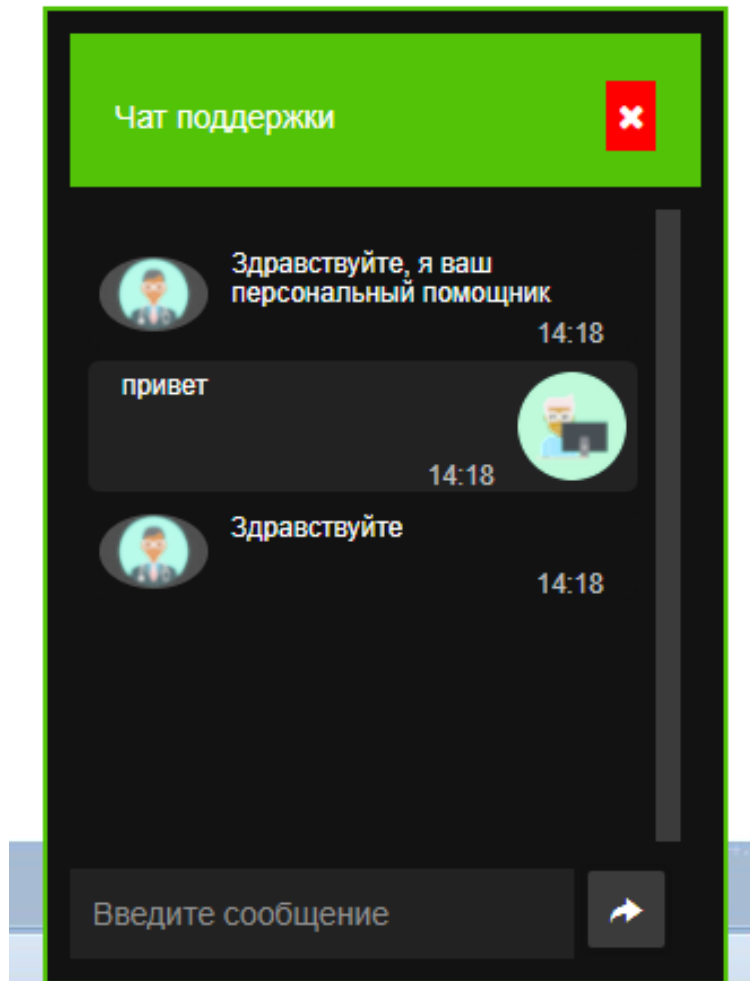


Рисунок 3. Внешний вид чат-бота

Выводы

Разрабатываемый чат-бот позволит специалистам IT отдела заниматься исключительно заявками о серьезных неисправностях.

Литература

1. <https://supermed.pro/med-chat-bots.html>
2. https://www.livemd.ru/tags/chat_boty/
3. <https://www.oracle.com/ru/chatbots/what-is-a-chatbot/>
4. <https://emailmatrix.ru/blog/chat-bots-for-business/>
5. <https://cyberleninka.ru/article/n/tehnologii-sozdaniy>
6. <https://en.wikipedia.org/wiki/Chatterbot>
7. <http://www.publishing-vak.ru/file/archive-economy-2019-10/76-kozoriz.pdf>
8. <https://plusworld.ru/daily/tehnologii/403076-2/>

УДК 004

**РАЗРАБОТКА ИНФОРМАЦИОННО-ПРОГРАММНОЙ МОДЕЛИ
СПЕЦИАЛИЗИРОВАННОГО ПОТОКА БЕТОННЫХ РАБОТ
ДЛЯ ОБЕСПЕЧЕНИЯ НЕПРЕРЫВНОСТИ ПРОИЗВОДСТВЕННОГО ЦИКЛА**

**DEVELOPMENT OF AN INFORMATION AND SOFTWARE MODEL
OF A SPECIALIZED CONCRETE WORK FLOW
TO ENSURE THE CONTINUITY OF THE PRODUCTION CYCLE**

Фахретдинов А.Р., Султанова Е.А., Михайловская И.М.,
Уфимский государственный нефтяной технический университет,
ул. Космонавтов, 1, г. Уфа, Республика Башкортостан, 450064, Россия

Fakhretdinov A.R., E.A. Sultanova, I.M. Mikhaylovskaya,
Ufa State Petroleum Technological University,
Kosmonavtov Str., 1, Ufa, Republic of Bashkortostan, 450064, Russia

e-mail: fakhretdinov.aidar@yandex.ru

Аннотация. Предметом исследования – является процесс автоматизации расчетов параметров проектирования технологии бетонных работ и обоснование «нормокомплекта» оборудования для организации специализированного потока бетонирования типового этажа. Транспортно-бетонный цикл представляет собой синхронизированное расписание работ производственного цикла по укладке бетонной смеси и динамику движения комплекта АБС, обеспечивающий непрерывность бетонирования, за счет использования выбранной технологической схемы бетонирования. Он обуславливает непрерывность технологических процессов укладки бетонной смеси, исходя из требований обязательного соблюдения времени бетонирования и уменьшения количества технологических и организационных перерывов параметров комплексного процесса. Таким образом, автоматизация представленных методик позволит существенно облегчить расчет общей продолжительности возведения монолитной части здания и обеспечить бесперебойность работ на строительной площадке.

Abstract. The subject of the study is the process of automating the calculation of the design parameters of the technology of concrete work and the justification of the "standard set" of equipment for organizing a specialized flow of concreting a typical floor. The transport-concrete cycle is a synchronized schedule of work of the production cycle for laying the concrete mixture and the dynamics of the movement of the ABS set, ensuring the continuity of concreting, through the use of the selected technological scheme of concreting. It determines the continuity of the technological processes of laying the concrete mixture, based on the requirements of the obligatory observance of the time of concreting and the reduction of the number of technological and organizational interruptions of the parameters of the complex process. Thus, the automation of the presented methods will significantly facilitate the calculation of the total duration of the construction of the monolithic part of the building and ensure the uninterrupted work at the construction site.

Ключевые слова: монолит, организационно-технологические параметры, бетонные работы, календарный график, матрицы.

Keywords: monolith, organizational and technological parameters, concrete work, calendar chart, matrices, sequence diagram, informative, autonomy, network charts, cyclogram.

Научная новизна заключается в разработке новой информационной системы, которая позволит рассчитывать организационно-технологические параметры. Специализированный поток монолитных бетонных работ состоит из арматурных, бетонных и опалубочных работ с учетом архитектурно-планировочных, конструктивных и технологических факторов проектируемых зданий и сооружений. Для успешной его организации необходимо наличие «нормокомплекта» оборудования и оснастки для изготовления арматурных изделий, опалубочных систем и приготовления товарных бетонных смесей. Для более простого расчета общей продолжительности возведения монолитной части здания и бесперебойной работы на строительной площадке было предложено автоматизировать методики организации, планирования и управления строительством.

Объемно-планировочные решения высотных зданий развиваются в трех направлениях. Основное – однофункциональное, например, офис или гостиница, второе – двухфункциональное, например, с размещением офиса на нижних этажах и гостиницы на верхних этажах, третье – многофункциональное. Примеры последнего встречаются относительно редко и в самых высоких зданиях, пространство которых трудно заполнить одной функцией. Как правило, в высотных зданиях любой объемно-планировочной структуры содержатся в нижних уровнях элементы открытой городской инфраструктуры (торговой и развлекательной), а в подземных – гаражи [1, 2].

На этапе проектирования зданий и сооружений возникает немало трудностей. Связано это, прежде всего, с большим количеством параметров, на расчет которых, без специализированного ПО, уходит много времени. Также существует повышенный риск возникновения ошибки, которая впоследствии может привести даже к обрушению. Было принято решение разработать информационную систему для автоматизации процесса проектирования зданий и сооружений.

Многие предприятия используют САПР-системы, но и это не решает проблему в полной мере. Такие системы универсальны, а каждая технология в строительстве индивидуальна, поэтому в таких системах отсутствуют многие второстепенные, но необходимые параметры, и пользователь все равно вынужден рассчитывать их вручную [3, 4, 5].

При организации, планировании и управлении строительством используются все виды моделей, но наиболее широко изобразительные (графические): линейный календарный график, циклограмма, сетевой график в форме графа, а также табличные, например, матрицы [6].

Повсеместно применяется линейный календарный график (рисунок 1), предложенный в конце прошлого столетия Г.Л. Гантом. На оси ординат этого графика выписываются в технологической последовательности виды работ с их характеристиками (объемами, стоимостями, трудоемкостями, машиноёмкостями, составами исполнителей и др.), а на оси абсцисс (после зоны, содержащей наименование и характеристики видов работ) – принятые порядковые или календарные единицы времени в количестве, достаточном для отображения всего периода производства работ. Непосредственно на сетку календарного графика наносятся горизонтальные линии, отображающие ход и сроки выполнения каждого вида работ. Под сеткой календарного графика выписывается потребность в исполнителях и их механовооруженность на каждую единицу времени.

Широкое распространение получил график потока – циклограмма (рисунок 2).

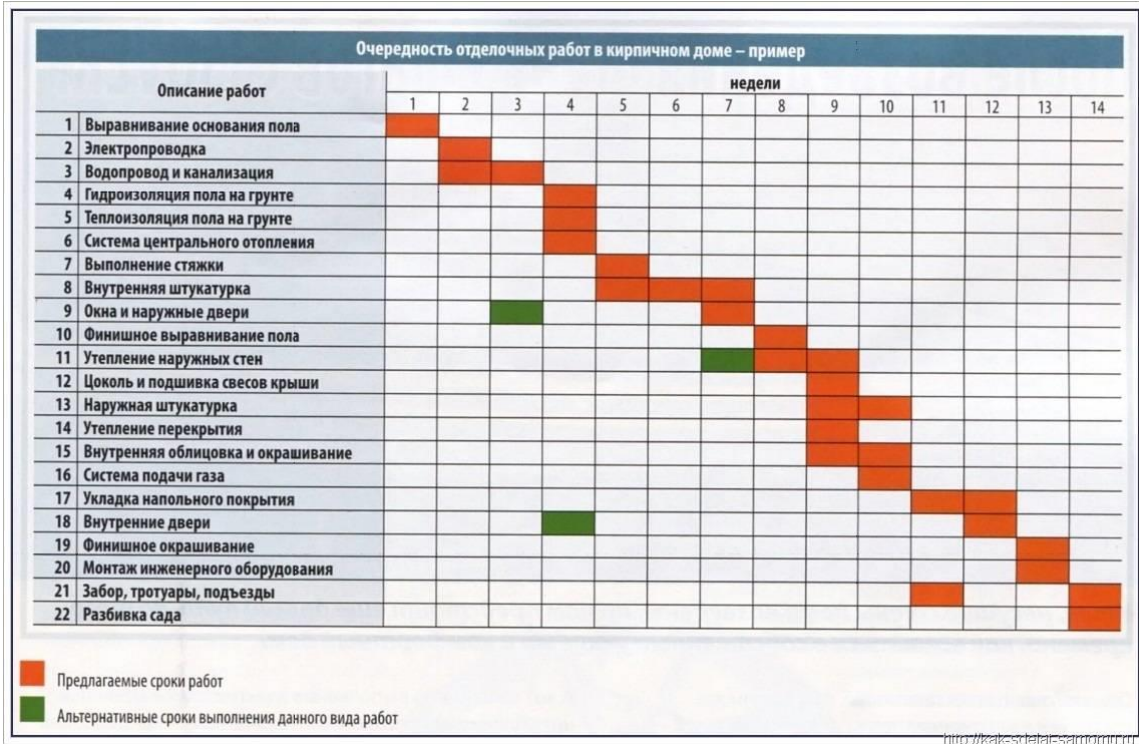


Рисунок 1. Пример календарного графика в линейной форме¹

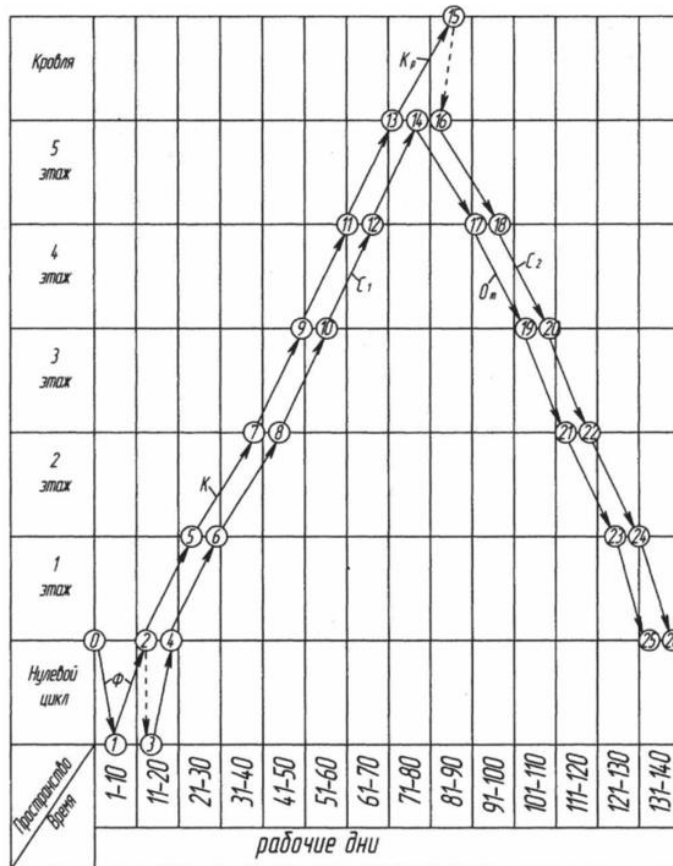


Рисунок 2. Календарный график в форме циклограммы¹

¹ Изображение из открытых источников сети Internet

При построении циклограммы учитывается, что при правильной организации работ в каждый определенный момент времени на одном частном фронте может выполняться только одна основная работа.

На рисунке 2 приведена объектная сетевая циклограмма строительства пятиэтажного четырехсекционного крупнопанельного жилого дома, где:

- 0-1 – цикл земляных работ;
- 1-2 – подготовка фундамента;
- 2-5, 5-7, 7-9, 9-11, 11-13 – возведение коробки 1-5 этажей;
- 3-4-6-8-10-12-14 – специализированные (сантехнические, электромонтажные) работы;
- 13-15 – кровельные работы;
- 14-17-19-21-23-25 – отделочные работы на 1-5 этажах;
- 16-18-20-22-24-26 – установка сантехнических приборов, электроарматуры;
- 2-3, 15-16 – организационно-технические зависимости.

Сетевой график (рисунок 3) представляет собой ориентированный граф, т.е. сеть, образуемую стрелками (работами и связями или только связями) и кружками или прямоугольниками, которые обозначают либо факт начала или окончания работ, т.е. являются событиями (кружки), либо отражают не только начало и окончание, но и сами работы (прямоугольники) [7, 8].

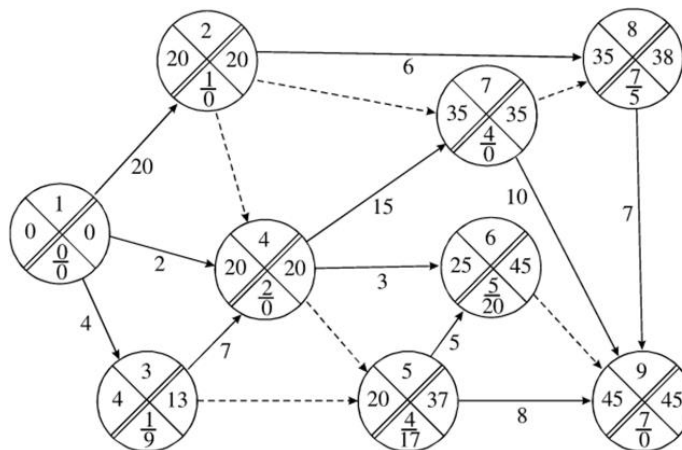


Рисунок 3. Календарный график в форме графа (сетевой график)²

Сетевые графики достаточно просты и наглядны при организации строительных работ с малым числом их видов и частных фронтов, но они весьма быстро усложняются и теряют наглядность при возрастании числа видов и фронтов работ, уступая по этим показателям линейному календарному графику и календарному графику в форме циклограммы.

Таким образом, рассмотрение наиболее характерных и распространенных форм календарных графиков показывает, что каждая из них имеет как достоинства, так и недостатки. Наличие достоинств определяет факт их использования, а наличие недостатков – постоянный поиск новых, более совершенных форм [9, 10, 11, 12].

Осуществляя поиск новой формы календарного графика, свободной от недостатков, присущих рассматриваемым формам, и включающей в себя их достоинства, была предложена комбинированная форма (рисунок 1.6). В основу комбинированной формы календарного графика положен линейный календарный график. Однако отображение основных видов работ, выполняемых непосредственно на

² Изображение из открытых источников сети Internet

частных фронтах, осуществлено в виде циклограммы, т.е. с показом самих частных фронтов. При этом предусматривается наложение связей между работами как основными, так и прочими. За сеткой календарного графика выписывается потребность в ресурсах на каждую единицу времени и при необходимости на каждый частный фронт работ. Потребность в ресурсах отображается как суммарная, так и по профессиям (трудовых ресурсов), по типоразмерам машин и механизмов, а также по видам необходимых конструкций, деталей, полуфабрикатов и материалов (материально-технических ресурсов).

Приведенные разновидности календарных графиков позволяют составить их классификацию (рисунок 4), охватывающую известные формы и позволяющую сконструировать новые путем варьирования классификационными признаками.

| | |
|--|--|
| По числу осей координат, использованных для отображения календарных графиков | <ul style="list-style-type: none"> двумерные трехмерные (с проекциями или без проекций на плоскости) |
| По однообразию структуры | <ul style="list-style-type: none"> одноструктурные комбинированные |
| По степени формализации | <ul style="list-style-type: none"> формализованные малоформализованные |
| По направлению развития | <ul style="list-style-type: none"> от пересечения координат в направлении возрастания шкал слева направо и/или сверху вниз |
| По составу показателей, выносимых на ось координат | <ul style="list-style-type: none"> в системе ОБР (ордината-виды работ) в системе ОФР (ордината-фронты работ) без четкой фиксации осей координат |
| По степени соответствия масштабов пространства реальным фронтам работ | <ul style="list-style-type: none"> с полным соответствием (нормальные) с частичным соответствием (нормализованные) |
| По полноте отображения перечня работ | <ul style="list-style-type: none"> с отображением всего перечня работ с отображением основных работ |
| По учету масштабов времени и пространства | <ul style="list-style-type: none"> в масштабе времени и пространства в масштабе времени в масштабе пространства безмасштабные |
| По характеру наклона линий, отображающих работы | <ul style="list-style-type: none"> с горизонтальными линиями с наклонными линиями |
| По степени учета отображения связей | <ul style="list-style-type: none"> с учетом и отображением связей с учетом, но без отображения без отображения и с недостаточным учетом |
| По степени отображения потребности в ресурсах на каждую единицу времени | <ul style="list-style-type: none"> потребность в ресурсах отображается потребность в ресурсах не отображается |

Рисунок 4. Классификация календарных графиков, применяемых в строительстве

В качестве табличных форм фиксации моделей организации работ следует прежде всего назвать матрицы, которые могут использоваться как самостоятельные формы фиксации моделей и как формы для подготовки исходных данных при любых других способах расчета и фиксации моделей, например, при разработке календарных графиков [13, 14, 15].

В 1981 году в связи с необходимостью расчета новых разновидностей поточных методов организации работ, А.В. Афанасьев предложил так называемые ранговые матрицы в системах ОФРР и ОВРР (рисунок 5).

Спецификой этих матриц является вынос на ось абсцисс рангов работ и фиксация в строкографах одноранговых работ либо последовательно по видам (система ОВРР), либо последовательно по фронтам (система ОФРР).

| ОФР | | Виды работ | | | |
|--------------|-----|------------|-------|-------|-------|
| | | A | Б | В | Г |
| Фронты работ | I | 'AI | 'BI | 'VI | 'GI |
| | II | 'AII | 'BII | 'BII | 'GII |
| | III | 'AIII | 'BIII | 'BIII | 'GIII |
| | IV | 'AIV | 'BIV | 'BIV | 'GIV |

| ОФРР | | Ранги работ | | | | | | |
|--------------|-----|-------------|------|-------|-------|-------|-------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Фронты работ | I | 'AI | 'BI | 'VI | 'GI | | | |
| | II | | 'AII | 'BII | 'BII | 'GII | | |
| | III | | | 'AIII | 'BIII | 'BIII | 'GIII | |
| | IV | | | | 'AIV | 'BIV | 'BIV | 'GIV |

| ОВР | | Фронты работ | | | |
|------------|---|--------------|------|-------|------|
| | | I | II | III | IV |
| Виды работ | A | 'AI | 'AII | 'AIII | 'AIV |
| | Б | 'BI | 'BII | 'BIII | 'BIV |
| | В | 'VI | 'VII | 'BIII | 'BIV |
| | Г | 'GI | 'GII | 'GIII | 'GIV |

| ОВРР | | Ранги работ | | | | | | |
|------------|---|-------------|------|-------|-------|-------|-------|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Виды работ | A | 'AI | 'AII | 'AIII | 'AIV | | | |
| | Б | | 'BI | 'BII | 'BIII | 'BIV | | |
| | В | | | 'VI | 'VII | 'BIII | 'BIV | |
| | Г | | | | 'GI | 'GII | 'GIII | 'GIV |

Рисунок 5. Форма матриц в системах

Функциональные требования информационной системы отображаются с помощью диаграммы вариантов использования (use case diagram).

Диаграмма обеспечивает описание того, что система в состоянии сделать и с кем (или чем) она будет взаимодействовать [16, 17, 18, 19, 20].

В моделируемой системе актером является пользователь программного продукта. Тип отношения – направленная ассоциация (рисунок 6).



Рисунок 6. Направленная ассоциация

Исходя из потребностей пользователя, выделяются следующие варианты использования программы:

- формирование структуры предприятия,

- учет отходов,
- формирование отчетов,
- определение перечня веществ предприятия.

Диаграмма вариантов использования представлена на рисунке 7.

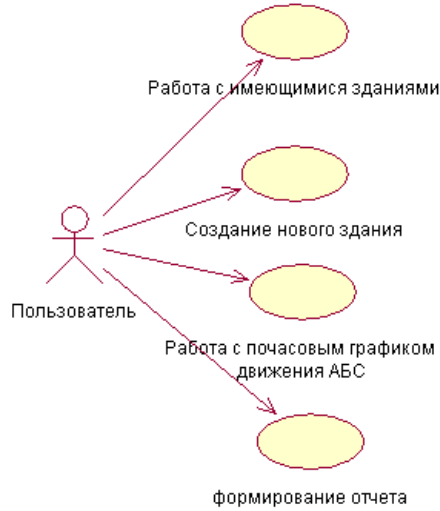


Рисунок 7. Диаграмма вариантов использования

Диаграмма последовательности действий (sequence diagram) отображает взаимодействие объектов, упорядоченное по времени. На ней показаны объекты и классы, используемые в сценарии, и последовательность сообщений, которыми обмениваются объекты, для выполнения сценария.

Диаграмма последовательности автоматизированной системы представлена на рисунках 8-10.

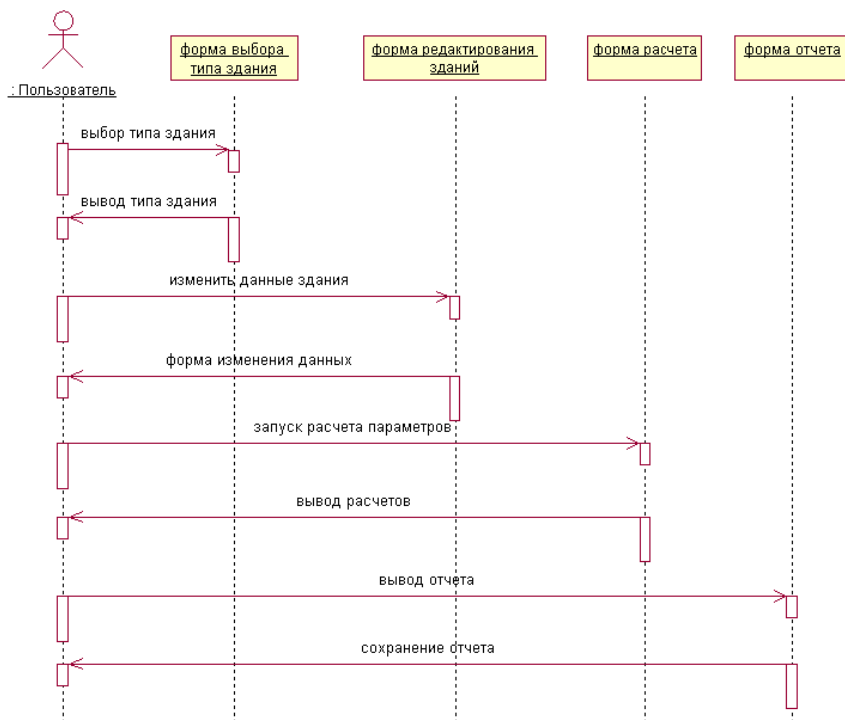


Рисунок 8. Диаграмма последовательности существующих зданий

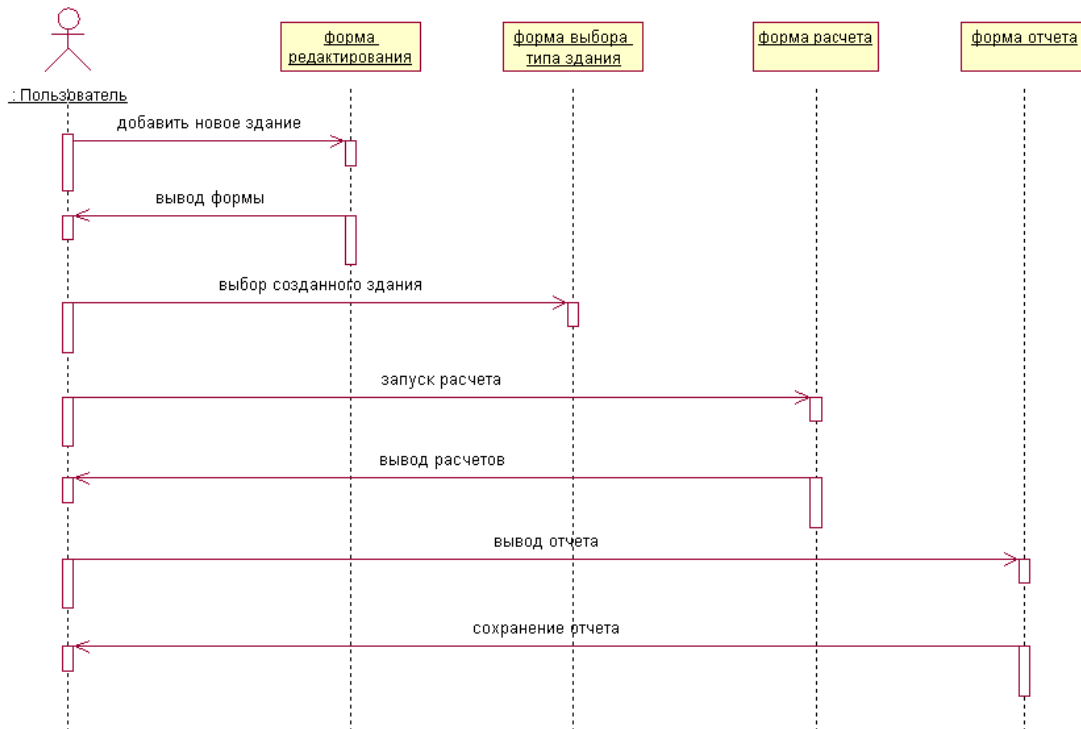


Рисунок 9. Диаграмма последовательности добавления нового здания

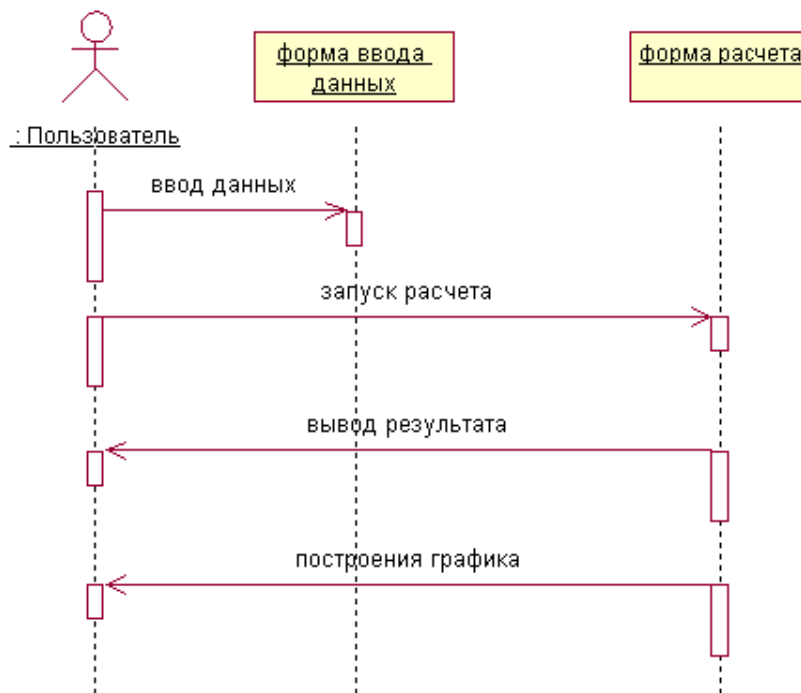


Рисунок 10. Диаграмма последовательности графика движения АВС

Информационная система должна обеспечивать возможность выполнения перечисленных ниже функций:

- вывод готовых зданий,
- добавление нового здания,
- расчет продолжительности специализированного потока,
- возможность выполнения редактирования,

- возможность удаления здания из списка,
- расчет почасового графика движения АБС,
- возможность сохранения отчета.

Потенциальными потребителями являются строительные организации, которым нужно специализированное ПО для расчетов.

Основные потребности пользователей:

- расчет транспортно-бетонного цикла;
- расчет продолжительности специализированного потока;
- формирование отчетов;
- удобство использования, интуитивно-понятный интерфейс;
- надежность и возможность сохранения результатов работы программы;
- высокая скорость работы.

Каждое программное средство должно выполнять определенные функции, а также обладать целым рядом свойств, позволяющим успешно его использовать в течение длительного периода, т.е. обладать определенными качествами.

Качество ПО – это совокупность черт и характеристик ПО, влияющих на способность удовлетворять заданные потребности пользователей.

Основными критериями качества ПС принято считать:

- функциональность,
- надежность,
- легкость применения,
- эффективность,
- сопровождаемость,
- мобильность.

Функциональность – это способность ПО выполнять набор функций, удовлетворяющих заданным или подразумеваемым потребностям пользователей. Набор указанных функций определяется во внешнем описании – в функциональной спецификации ПО. Разрабатываемое программное обеспечение обладает всеми необходимыми функциями, которое позволит автоматизировать процесс проектирования зданий и сооружений удовлетворяющими потребностям пользователей [21].

Автономность. Программное обеспечение работает независимо от других программ. Дополнительная установка стороннего программного обеспечения не требуется.

Устойчивость. Программное обеспечение полностью работоспособно в случае ввода некорректных входных данных. Это достигнуто использованием проверки данных на корректность, а также использованием обработчиков исключительных ситуаций, отвечающих за обработку ситуаций обнаружения некорректных данных.

Защищенность. Предусмотрено сохранение всей информации в случае возникновения сбоя в работе компьютера или ПО и загрузка последнего сохранения при запуске программного обеспечения.

Легкость применения – это характеристики ПО, которые позволяют минимизировать усилия пользователя по подготовке исходных данных, применению ПС и оценке полученных результатов, а также вызывать положительные эмоции определенного или подразумеваемого пользователя.

Информативность. Названия полей для ввода и объектов отображены в полном виде и соответствуют названиям, приведенным в нормативных документах. Программа формирует подробный отчет, содержащий исходные данные и результаты расчетов.

Удобство использования. Программный продукт прост в использовании, обладает интуитивно понятным интерфейсом и удобным набором функций.

Эффективность – это отношение уровня услуг, предоставляемых ПО пользователю при заданных условиях, к объему используемых ресурсов.

Временная эффективность. Вычисления и другие операции выполняются быстро, в доли секунды.

Эффективность по ресурсам. Минимальный объем требуемой оперативной памяти 256 Мб, минимальный объем доступного свободного места на жестком диске 100 Мб.

Эффективность по устройствам. Для непосредственной работы с программой требуется минимум периферийных устройств, помимо самой ЭВМ. В качестве средств ввода требуются клавиатура и мышь. В качестве средств вывода – монитор, минимальное разрешение экрана 1024x780 [22, 23].

Мобильность – это возможность переноса ПО из одной среды (окружения) в другую, в частности, с одной ЭВМ на другую.

Структурированность. Программное средство представляет собой модульную программу.

Сопровождаемость – это характеристики ПО, позволяющие минимизировать усилия по внесению в него изменений для устранения ошибок и модификации в соответствии новыми потребностями пользователей.

Обеспечена простота сопровождения наличием руководства программиста, структурированностью текста программы.

Выводы

Автоматизация представленных методик позволит существенно облегчить расчет общей продолжительности возведения монолитной части здания и обеспечить бесперебойность работ на строительной площадке.

Литература

1. Коровяков В.Ф. Роль научно-технического сопровождения строительства в повышении качества монолитного строительства//Промышленное и гражданское строительство. 2014. №5. С. 34-36.
2. Духанин П.В., Бондарь Н.Е. Сравнение новых технологий монтажа ограждающих конструкций в каркасно-монолитном строительстве//Академическая публицистика. 2020. №11. С. 330-333.
3. Ключникова О.В., Попов А.В. Технология и комплексная механизация строительства большепролетных монолитно-каркасных зданий//Перспективы науки. 2020. №12 (135). С. 114-116.
4. Гаврилюк Е.А., Манохин П.Е. Проблемы монолитного строительства // Энергосбережение, информационные технологии и устойчивое развитие. 2014. С. 12-18.
5. Манохин П.Е., Шамшурина Е.А. Особенности применения несъемной опалубки из твердого самозатухающего пенополистирола в технологии монолитного строительства//Фотинские чтения. 2015. № 2 (4). С. 177-179.
6. Анисимов С.А. Систематизация и структурный анализ технологии и организации строительства монолитного и крупнопанельного железобетонного многоэтажного строительства жилых зданий//Молодежь и научно-технический прогресс XI международной научно-практической конференции студентов, аспирантов и молодых ученых. 2018. С. 12-15.

7. Харун М.И., Коротеев Д.Д., Левицкая А.Ю., Макав Д.В. Термическая обработка бетона в монолитном строительстве//Иновации в науке и практике XV международной научно-практической конференции. 2019. С. 36-44.
8. ГОСТ 19.701–90 Схемы алгоритмов, программ, данных и систем. Условные обозначения и правила выполнения. – М.: Изд-во стандартов, 1991. – 26 с.
9. Карибов С.В., Зиганшина М.Ф. Оценка стоимости программного обеспечения методами сравнительного подхода научные основы современного прогресса//Сборник статей международной научно-практической конференции. 2016. С. 109-118.
10. Черняховская Л.Р., Никулина Н.О., Бармина О.В. Оценка эффективности поддержки принятия решений при реализации проекта по разработке программного обеспечения//Информационные технологии интеллектуальной поддержки принятия решений (ITIDS'2018). Труды VI Всероссийской конференции. 2018. С. 16-22.
11. Бринкгрив Р.Б. Дж., Энгин Э., Энгин Х.К. Валидация эмпирических формул для получения параметров модели для песков//Численные методы в геотехнической инженерии. – 2010.
12. С. Киннунен Х. Нюландер, Штамповка бетонных плит без поперечной арматуры (Королевский технологический институт, Стокгольм, 1960 г.).
13. Л. Нгуен-Мин, М. Ровнак, Т. Тран-Куок и К. Нгуен-Ким. Сопротивление продавливанию железобетонных плоских плит, армированных стальным волокном. Процедура Инжиниринг 14 (2011).
14. Л.Ф. Майя, М. Фернандес Руис, А. Муттони, С.Дж. Фостер. Прочность на продавливание сталефибробетонных плит. Инженерные сооружения 40 (2012).
15. Ли Цзиньюй, Сюй Вэньюй, Цао Цзянго, Линь Ли, Гуань Юши. Исследование механизма разрушения бетона под действием мороза. Шуйли Сюэбао. 1999(1).
16. Дженгиз Д., Серкан Т. Прочность двухосно нагруженных высокопрочных железобетонных колонн/Строительная техника и механика. Том. 44, № 5, 2012 г.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ, УПРАВЛЕНИИ И БИЗНЕСЕ

УДК 004.4

ОСНОВНЫЕ НАПРАВЛЕНИЯ РАЗВИТИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ УПРАВЛЕНИЯ

MAIN DIRECTIONS OF DEVELOPMENT OF INFORMATION CONTROL TECHNOLOGIES

Петлина Е.М.,
Ставропольский государственный педагогический институт,
г. Ставрополь, Российская Федерация

E.M. Petlina,
Stavropol State Pedagogical Institute,
Stavropol, Russian Federation

e-mail: 356620@gmail.com

Аннотация. В статье рассмотрены возможности использования информационных технологий. Изучены подходы классификации информационных технологий. Сегодня наблюдается постоянное увеличение сфер применения IT-технологий. Наблюдается упрощение интерфейса в интересах пользователя, частичной открытостью своей архитектуры, стремлением к программной совместимости друг с другом, ориентацией на определенную компьютерную платформу. Классификации программных продуктов всегда относительны. Распространена классификация информационных технологий по уровню управленческой области применения. За последнее десятилетие сформировалась новое направление в программной технике CASE. Это инструментарий для системных аналитиков, разработчиков и программистов, заменяющим им бумагу и карандаш на компьютер для автоматизации процесса проектирования и разработки программного обеспечения. CASE-технология для автоматизированных систем принятия решений является такой же составной, как и все другие, включенные в общий программный пакет системы поддержки принятия решений. Поддержка совокупностью информационных технологий позволяет подойти к вопросам применения разнообразных информационных технологий дифференцированно, с учетом особенностей функционирования различных уровней органов управления. Структурно информационные системы представлены совокупностью отдельных составляющих частей. Они делятся на функциональные, обеспечивающие и организационные. Выявлены основные направления развития ИТ.

Abstract. The article considers the possibilities of using information technologies. Approaches to the classification of information technologies have been studied. Today, there is a constant increase in the scope of IT-technologies. There is a simplification of the interface in the interests of the user, the partial openness of its architecture, the desire for software compatibility with each other, and orientation to a specific computer platform. Classifications of software products are always relative. The classification of information technologies according to the level of the managerial scope is widespread. Over the past decade, a new direction in CASE software engineering has been formed. This is a toolkit for system analysts, developers and programmers who replace paper and pencil with a computer to automate the

process of designing and developing software. CASE-technology for automated decision-making systems is as complex as all others included in the general decision support system software package. Support by a set of information technologies allows us to approach the issues of using various information technologies in a differentiated way, taking into account the peculiarities of the functioning of various levels of government. Structurally, information systems are represented by a set of individual constituent parts. They are divided into functional, providing and organizational. The main directions of IT development are identified.

Ключевые слова: информационные технологии, классификация программных продуктов, CASE-технология, информационная система, автоматизированная система управления (АСУ).

Keywords: information technologies, classification of software products, CASE-technology, information system, automated control system (ACS).

Для оценки возможности использования информационных технологий целесообразно иметь определенные рекомендации или алгоритмы ее выбора [2, 6]. Обычно, в таком случае пользуются существующими классификациями предметов исследования. Четко определенной классификации информационных технологий не существует. Это обусловлено, с одной стороны, «привязанностью» информационных технологий к аппаратным средствам, с другой – спецификой задач, решаемых конкретными информационными технологиями, с третьей – постоянной динамикой процесса создания, развития и реализации совокупности информационных технологий [5].

Современные информационные технологии характеризуются постоянным увеличением сфер применения, облегчением (упрощением) интерфейса в интересах пользователя, частичной открытостью своей архитектуры, стремлением к программной совместимости друг с другом (порой за счет существенного усложнения), ориентацией на определенную компьютерную платформу. С позиций сегодняшнего дня, можно утверждать, что все ИТ выполняют определенные функциональные задачи (комплекс задач), которые входят или могут входить в автоматизированную систему на различных уровнях органов управления. Поэтому, очень важно осуществить классификацию современных информационных технологий с позиций возможного применения ИТ, т.к. они могут внести коррективы в предъявляемые требования к автоматизированным системам для различных органов управления [7].

Классификации всегда относительны. Например, выделение одной из ИТ в группу применения в штабе Регионального управления предусматривает, что данная информационная технология должна обладать свойствами, отнесенными к группе сетевых ИТ, ИТ на основе СУБД, элементов ГИС-технологий, в то же время она должна обладать высокой степенью интеграции, а также обеспечивать возможность планирования и подготовки предложений для принятия решения на среднем (или верхнем) уровне управленческой области применения.

За последнее десятилетие сформировалась новое направление в программной технике – CASE (Computer-Aided Software/System Engineering). В настоящее время не существует общепринятого определения CASE. Содержание этого понятия обычно определяется перечнем задач, решаемых с помощью CASE, а также совокупностью применяемых методов и средств. Очень упрощенно, CASE-технология представляет собой совокупность методологий анализа, проектирования, разработки и сопровождения сложных систем программного обеспечения (ПО), поддерживаемую комплексом взаимосвязанных средств автоматизации. CASE – это инструментальный для

системных аналитиков, разработчиков и программистов, заменяющим им бумагу и карандаш на компьютер для автоматизации процесса проектирования и разработки ПО [10].

Иногда CASE-технология провозглашается как панацея от всех бед. Однако авторы этих утверждений не учитывают специфику описания предметной области применения CASE-технологии и подменяют ею порой суть системы поддержки принятия решения. В итоге анализа ИТ и задач, решаемых ими, можно утверждать, что CASE-технология для автоматизированных систем принятия решений является такой же составной, как и все другие, включенные в общий программный пакет системы поддержки принятия решений.

Поддержка совокупностью информационных технологий, представленных в данной классификации, методологий структурного системного анализа и рассмотрения организацией, как многоуровневой иерархической системы позволяет подойти к вопросам применения разнообразных ИТ дифференцированно, с учетом особенностей функционирования различных уровней органов управления [2].

Информационная система – организационно упорядоченная совокупность документов (массивов документов) и информационных технологий (библиотеки, архивы, фонды, банки данных, АРМы и т.д.), используемых для хранения, обработки и выдачи информации для достижения цели управления [10].

К главным принципам построения эффективных систем обработки данных относятся:

- принцип интеграции, который заключается в том, что введенные один раз в систему данные для обработки многократно используются для решения значительного количества задач;

- принцип системности, который заключается в обработке различных данных в разных аспектах с целью получения информации, достаточной для принятия решений на всех ступенях управления;

- принцип комплексности, который заключается в том, что процедура преобразования данных механизмуется и автоматизируется на всех фазах функционирования информационной системы.

Структурно информационные системы представлены совокупностью отдельных составляющих частей, которые называют подсистемами. Они делятся на функциональные, обеспечивающие и организационные.

Функциональные подсистемы исполняют роль поддержки моделей, методов и алгоритмов с целью получения информации для управляющего органа. Состав функциональных подсистем очень многообразен и зависит от предметной сферы применения системы обработки данных, особенностей экономической деятельности объекта подчинения [9].

Анализ, представленных информационных технологий позволяет выявить следующие основные направления развития ИТ:

- создание, развитие и модернизация ИТ идет по направлению комплексной интеграции отдельных составляющих информационных процедур;

- ИТ «привязаны» к аппаратно-программной части, дальнейшее создание программно-технических комплексов автоматизации деятельности должно учитывать набор требуемых информационных процедур, уровень программной составляющей и ограничения по аппаратной части;

- диапазон содержания и применения ИТ постоянно увеличивается, ориентируясь на растущие потребности потребителей и расширение сфер применения;

– интерфейс программных составляющих ИТ постоянно унифицируется по основным функциям программного средства, причем в большинстве в основу положен стандарт интерфейса Windows-продуктов;

– одним из направлений развития является углубленная специализация ИТ на функциональную деятельность потребителя (заказчика);

– в любой современной ИТ предусмотрена реализация связи и услуг связи.

Известно, что унификация является одним из эффективных средств экономии материальных, трудовых, финансовых, интеллектуальных и иных ресурсов, которая может использоваться для концентрации усилий на наиболее важных направлениях развития. Создание многоуровневых иерархических систем управления в государственном или ведомственном масштабе (особенно систем централизованных или смешанных) вообще невозможно без определенного уровня унификации. Понимание данной проблемы заставило правительственные и государственные ведомства искать пути упорядочения выхода на рынок программно-технических комплексов и компьютерных технологий. Так, начиная с 1995 года, производился комплекс работ по созданию государственной информационно-телекоммуникационной системе (ИТКС) в целях поддержки принятия решений высшими должностными лицами Российской Федерации, в первую очередь в интересах Администрации Президента, Правительства, Совета безопасности, Государственной Думы, отдельных ведомств и глав субъектов Федерации. Кроме «чистых» услуг связи (телефонная связь, передачи данных, телефакс, пейджинговая связь и т.п.) образованы иные услуги доставки и представления информации [3]. Они реализованы в ряде ситуационных центров в виде информационной составляющей, представляющих собой совокупность информационных фондов, информационно-технических средств и средств моделирования ситуаций, информационно-аналитической системы и информационных ресурсов. В основе функционирования ситуационных центров лежат так называемые базовые информационные технологии: сбора и накопления информации, обработки текстовой информации, ГИС-технологии, аналитической обработки, представления информации и электронных документов [8]. При представлении результатов решения комплекса функциональных задач применяются программные средства оперативного анализа информации, интеллектуального анализа данных, моделирования, видеопредставления информации и др. В настоящее время по такой технологической цепочке развернуто несколько ситуационных центров.

Основой всего функционирования АСУ является информация. Она же лежит в основе всех обеспечивающих частей системы [4]. Поэтому, в процессе предъявления требований к автоматизированной системе целесообразно предварительно определить:

- содержание информации;
- лингвистический состав информации;
- язык представления информации;
- уровень и способ кодирования информации;
- формализованное описание информации;
- принадлежность информации;
- конфиденциальность информации;
- форма представления информации;
- объем информации;
- степень связанности информации с другими информационными массивами;
- безопасность информации [1];
- защищенность информации, подверженность информации к искажению (уничтожению);
- источник получения информации;

- место расположения (хранения) информации;
- время хранения и периодичность обновления информации;
- адресность информации;
- способ получения информации и другие признаки (атрибуты, агрегатные состояния) информации.

Выводы

Современный уровень развития персональных компьютеров и программных средств характеризуется «сращиванием» техники и программного обеспечения (операционных систем, программ обработки и формирования информационных сообщений и образов, отображения информации и т.д.). Сегодня уже нельзя провести четкой границы между «техникой» и «математическим обеспечением». Это наглядно видно из приведенного описания информационных технологий и обуславливает необходимость знания особенностей применения современных программно-технических комплексов поддержки принятия решений на различных уровнях управления организацией.

Литература

1. Павлов А.С., Петлина Е.М. Основные принципы обеспечения информационной безопасности//Гуманитарно-правовые аспекты развития российского общества: сборник научных трудов по итогам региональной научно-практической конференции. Ставрополь: СФ КУ МВД РФ, 2017. С. 380-386.
2. Петлина Е.М. Автоматизированные системы управления как средство повышения эффективности работы предприятия//Общество и личность: ситуация постгуманизма. Памяти профессора Е.Н. Шиянова. Ставрополь, 2021. С. 234-237.
3. Петлина Е.М. Использование информационных технологий для формирования общих компетенций специалиста//Актуальные направления научных исследований XXI века: теория и практика. 2017. Т. 5. № 7-1 (33-1). С. 275-279.
4. Петлина Е.М. Компьютерное моделирование. Саратов, 2019.
5. Петлина Е.М. Об использовании математических методов для анализа работоспособности технических устройств//Социально-гуманитарные и естественно-технические науки и вызовы современности. Материалы международной научно-практической конференции. Ставрополь, 2017. С. 806-810.
6. Петлина Е.М., Горбачев А.В. Информационные технологии в профессиональной деятельности: учебное пособие для СПО. Саратов, 2021.
7. Петлина Е.М., Нестеров Д.С. О развитии информационных технологий в Российской Федерации//Современное гуманитарное знание о проблемах социального развития: материалы Всероссийской конференции. Ставрополь, 2021. С. 50-52.
8. Петлина Е.М., Хатагова С.В. Информатизация образования как основной принцип формирования компетенций специалиста//Инновации в образовании. 2017. № 3. С. 124-133.
9. Федеральный закон «Об информации, информационных технологиях и о защите информации» от 27.07.2006 N 149-ФЗ
10. Юрданов Д.В., Горденко Д.В., Горденко Н.В., Петлина Е.М., Павлюк Д.Н. К вопросу применения системы остаточных классов в современных устройствах цифровой обработки сигналов//Фундаментальные исследования. 2016. № 2-2. С. 318-322.

УДК 004.4

**РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА
ПРОВЕРКИ ЭЛЕКТРОННЫХ ПИСЕМ
ДЛЯ ПОЧТОВОГО СЕРВЕРА ПРЕДПРИЯТИЯ**

**SOFTWARE DEVELOPMENT
CHECKING EMAILS
FOR THE COMPANY'S MAIL SERVER**

Головина Е.Ю., Фархутдинова А.И.,
Институт нефтепереработки и нефтехимии ФГБОУ ВО УГНТУ в г. Салавате,
ул. Губкина, 22 б, г. Салават, Республика Башкортостан,
Россия, 453250

E.Yu. Golovina, A.I. Farkhutdinova,
Institute of Oil Refining and Petrochemistry FSBEI HE USPTU in Salavat,
Gubkin Str., 22b, Salavat, Republic of Bashkortostan,
453250, Russia

e-mail: FoxindFerra4000@yandex.ru

Аннотация. В рамках рыночной экономики важно следить за информационной безопасностью для повышения конкурентоспособности предприятия и его репутационного уровня. Утечки конфиденциальных данных могут сделать предприятие менее привлекательным для клиентов и инвесторов, а также снизить прибыль из-за устранения ущерба [1, 2]. При этом 52 % компаний по всему миру считают наибольшей угрозой системе корпоративной безопасности самих сотрудников (в России, по данным Group-IB, этот показатель и вовсе достигает 88 %) [3]. Одним из основных средств коммуникаций, используемых на современных предприятиях, являются электронные письма, передаваемые с помощью почтового сервера. Ежедневно через него на предприятиях проходит большой поток электронных писем: на крупных предприятиях их количество оценивается в несколько тысяч. Из-за такого числа становится труднее отслеживать их содержание, чем и могут воспользоваться недобросовестные сотрудники с целью кражи важных конфиденциальных данных, таких как, например, логины, пароли и паспортные данные. По результатам исследования, проведенным в первом полугодии 2020 года, было выявлено, что 40 % утечек данных происходило из-за намеренных действий сотрудников, которые имели доступ к данным, а остальные 60 % – по невнимательности и наивности [4]. Таким образом, защита конфиденциальных данных, передаваемых с использованием электронной почты через электронные письма, является задачей актуальной и имеет, ввиду широкого использования данного сервиса в современных информационных системах, большую практическую значимость. В данной статье рассматривается существующий бизнес-процесс «Проверка электронных писем».

Abstract. In the current economy, it is important to monitor security for the profit of the enterprise and its reputation level. Sensitive data leaks can lead to reduced customer exposure, as well as reduced revenue due to scale up. At the same time, 52 % of companies around the world consider their own employees to be the largest corporate security commission (in Russia, according to Group-IB, this figure and the exception are 88 %). Some of the primary means of communication present in today's enterprises are electronic, accessible through a mail

server. Every day, a large stream of emails passes through it at enterprises: in large enterprises, their number is estimated at several thousand. Due to such a number of persons, it becomes more difficult to take into account their content, which is what unscrupulous employees play with the theft of confidential data, such as, for example, logins, passwords and passport data. According to the results of a study conducted in the first half of 2020, it was revealed that 40 % of data leaks are regular due to intentional actions of employees that are necessary to access data, and the remaining 60 % are due to inattention and naivety. Thus, the protection of confidential data, available using e-mail through e-mails, is relevant and has a wide use of this service in modern information sources, greater practical accessibility. This article replicates the Existing Email Validation business process.

Ключевые слова: электронные письма, электронная почта, почтовый сервер, конфиденциальные данные, бизнес-процесс.

Keywords: emails, e-mail, mail server, confidential data, business process.

На промышленном предприятии существующая модель бизнес-процесса «Проверка электронных писем» основана на ручном способе проверки электронных писем и может иметь такие проблемы, как:

- пропуск электронного письма с конфиденциальными данными из-за «человеческого фактора», который может увеличить своё влияние при проверке большого количества электронных писем;
- проверка выборочных электронных писем в виду невозможности для сотрудника проверить все, что может повлечь за собой пропуск писем с конфиденциальными данными.

На рисунке 1 показана декомпозиция бизнес-процесса «Проверка электронных писем» в нотации IDEF3.

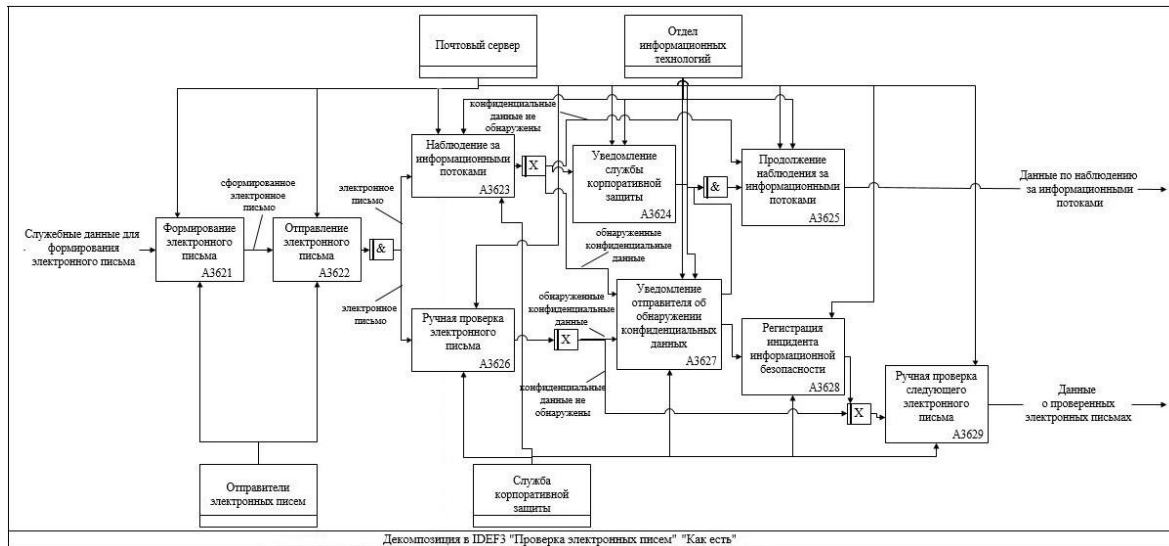


Рисунок 1. Декомпозиция бизнес-процесса «Проверка электронных писем»

На рисунке 2 показана декомпозиция бизнес-процесса «как есть» «Проверка электронных писем» в нотации EPC.

В декомпозиции присутствуют следующие участники:

- отправители электронных писем. Бизнес-процесс начинается с формирования электронного письма сотрудником предприятия, имеющим учётную запись на почтовом

сервере или сторонним человеком вне рассматриваемого предприятия, для дальнейшего его отправления;

– служба корпоративной защиты, осуществляющая проверку электронных писем для предотвращения утечек конфиденциальных данных, которые могли попасть в электронные письма путём намеренных или случайных действий отправителей. В случае обнаружения электронного письма с конфиденциальными данными должны уведомить об этом отправителя и зарегистрировать инцидент информационной безопасности;

– отдел информационных технологий, наблюдающий за информационным потоком предприятия. В ходе своей наблюдательской деятельности могут заметить утечку конфиденциальных данных в электронных письмах. В случае обнаружения электронного письма с конфиденциальными данными должны уведомить отправителя и службу корпоративной защиты.

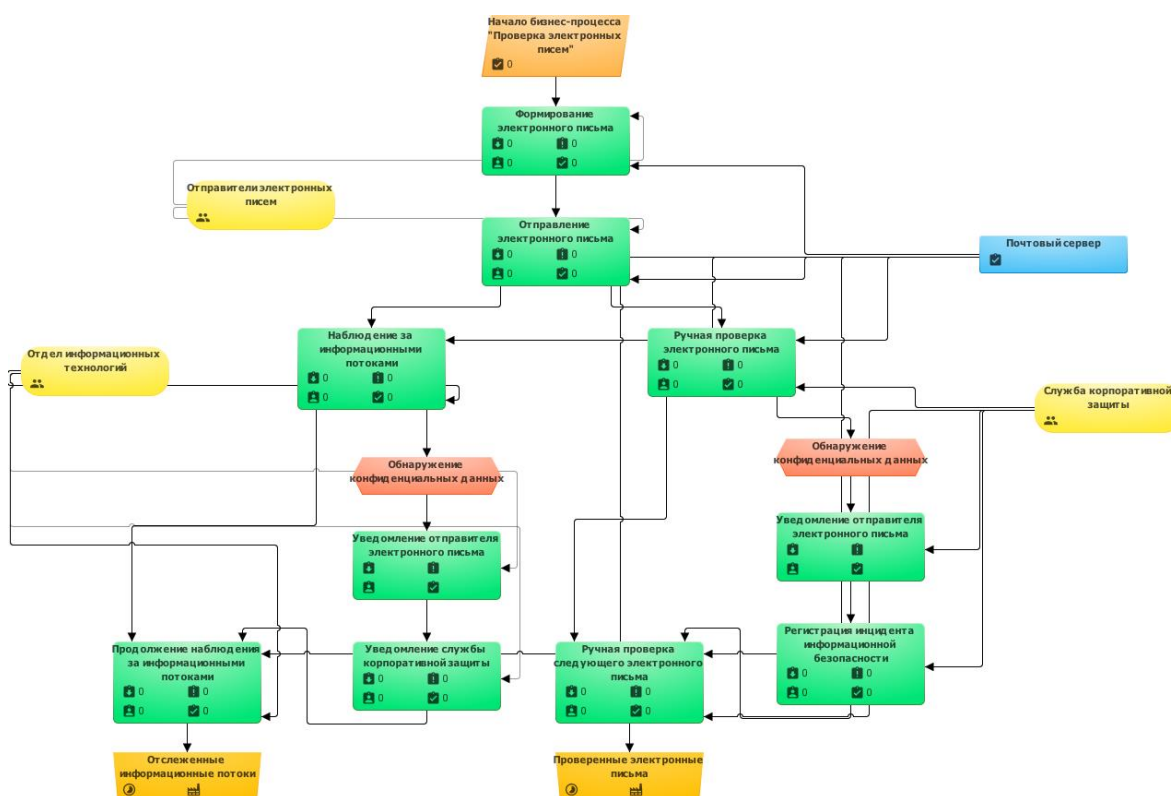


Рисунок 2. Декомпозиция бизнес-процесса «как есть» «Проверка электронных писем» в нотации EPC

Результат логического анализа бизнес-процесса «Проверка электронных писем» приведён в таблице 1. Ценность может быть оценена как увеличивающий потребительскую ценность продукта (УЦ) и не увеличивающий (НУЦ).

Таблица 1 – Оценка шагов бизнес-процесса «Проверка электронных писем»

| Шаг | Ценность | Возможность совершенствования или удаления |
|----------------------------------|----------|--|
| 1 | 2 | 3 |
| Формирование электронного письма | УЦ | Нет |
| Отправление электронного письма | УЦ | Нет |

Продолжение таблицы 1

| Шаг | Ценность | Возможность совершенствования или удаления |
|--|----------|--|
| Наблюдение за информационными потоками | НУЦ | Да |
| Уведомление службы корпоративной защиты | НУЦ | Да |
| Продолжение наблюдения за информационными потоками | НУЦ | Да |
| Ручная проверка электронного письма | НУЦ | Да |
| Ручная проверка следующего электронного письма | НУЦ | Да |
| Регистрация инцидента информационной безопасности | НУЦ | Да |
| Уведомление отправителя электронного письма об обнаружении конфиденциальных данных | НУЦ | Да |
| Предоставление данных о проверенных электронных письмах | НУЦ | Да |

Существующая модель бизнес-процесса «Проверка электронных писем» имеет возможность совершенствования или удаления для нескольких шагов. Предлагается использовать программное средство для понижения количества инцидентов информационной безопасности, связанных с утечкой конфиденциальных данных через электронные письма, пересылаемые с помощью почтового сервера на предприятии.

Составлена и представлена на рисунке 3 стратегическая карта целей по снижению количества инцидентов информационной безопасности, связанных с утечкой конфиденциальных данных через электронные письма.

Снижение количества инцидентов информационной безопасности, связанных с утечкой конфиденциальных данных через электронные письма, повысит уровень информационной безопасности, репутации и конкурентоспособности предприятия.

Для обнаружения конфиденциальных данных в электронном письме в программном средстве применяется алгоритм Вагнера-Фишера для нахождения расстояния Левенштейна с целью определения нечёткой эквивалентности двух множеств (входящего текста и имеющихся ключевых слов в базе данных промышленного предприятия). В электронных письмах сотрудникам промышленного предприятия также иногда приходится прикладывать изображения формата .png, .pdf, .jpeg, .jpg., поэтому имеется необходимость проверять и их. Для оптического распознавания символов из изображений, которые отправители могут приложить к электронному письму, применяется библиотека оптического распознавания символов АBBYU Cloud OCR. На облачный сервер компании АBBYU посылается https-запрос с изображением формата .pdf, .png, .jpg, .jpeg. Полученный результат оптического распознавания в виде файла формата .txt дальше проверяется на схожесть с ключевыми словами из базы данных с помощью расстояния Левенштейна. На рисунке 4 изображена схема работы программного средства в случае обнаружения конфиденциальных данных с вероятностью $\geq 0,5$ в электронном письме. На рисунке 5 изображена схема работы программного средства в случае, если вероятность обнаружения конфиденциальных данных $< 0,5$.

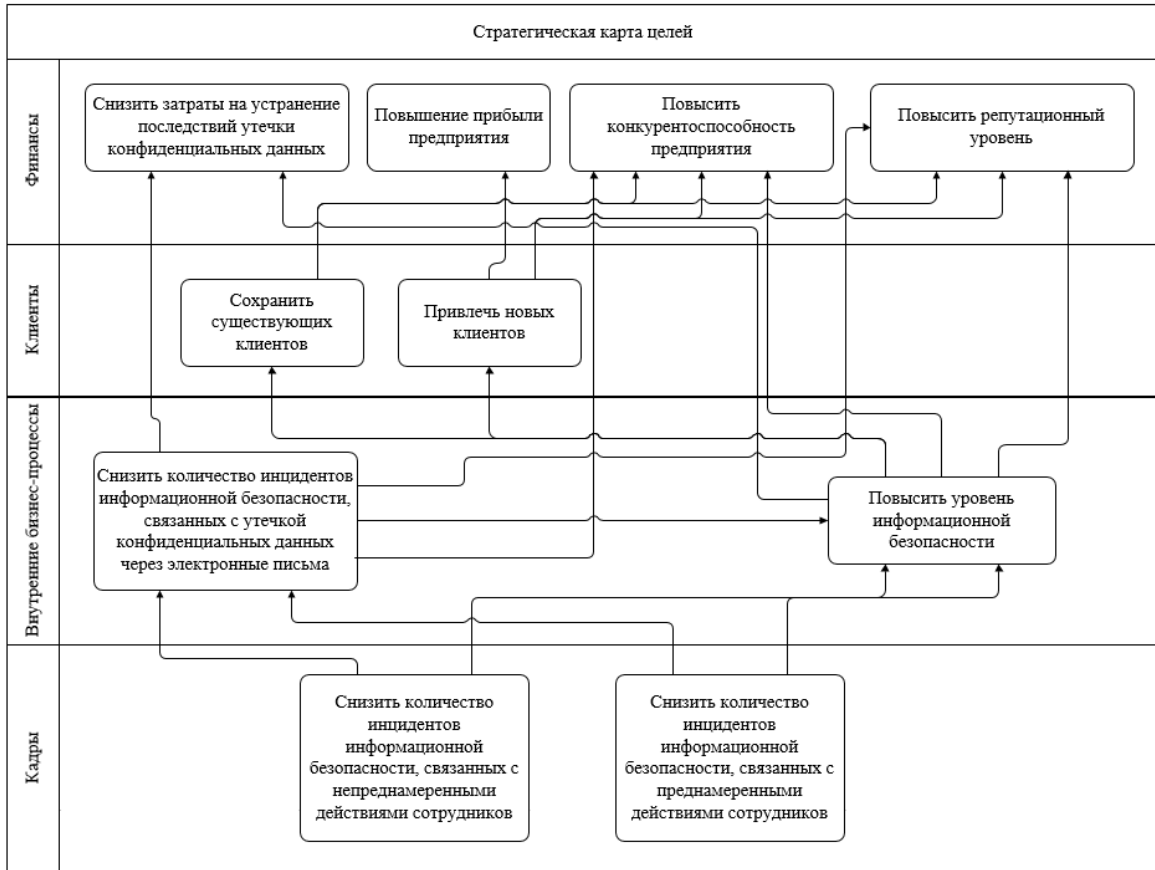


Рисунок 3. Стратегическая карта целей

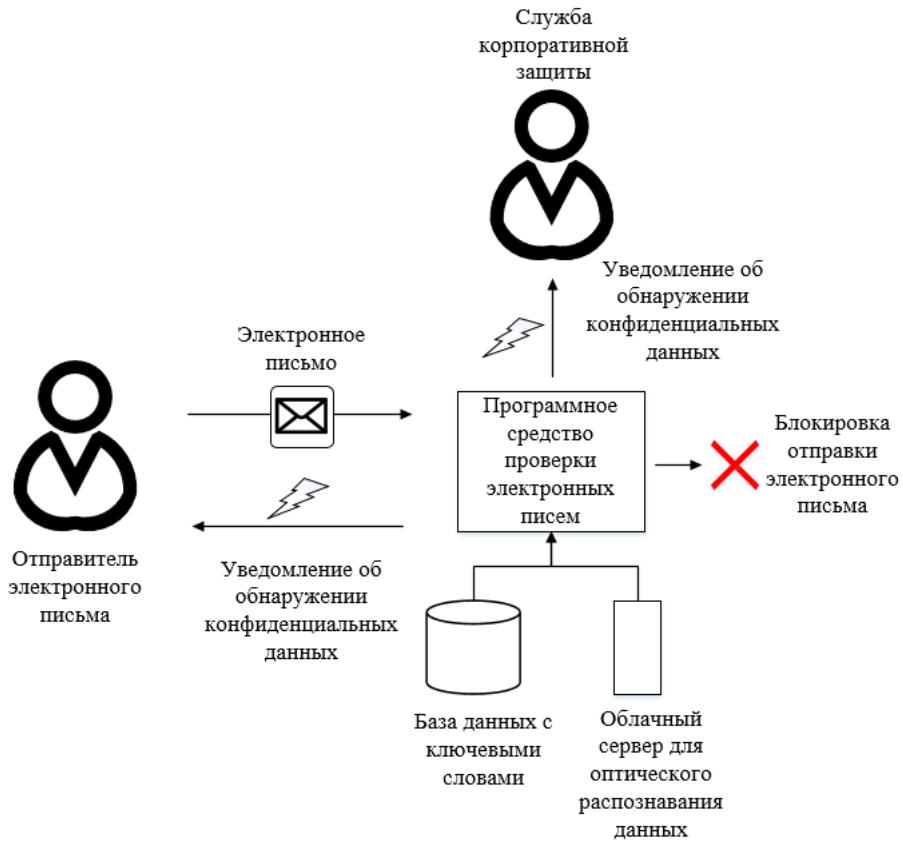


Рисунок 4. Случай обнаружения конфиденциальных данных



Рисунок 5. Без обнаружения конфиденциальных данных

В случае обнаружения программным средством конфиденциальных данных блокируется отправка электронного письма, а также приходит уведомление службе корпоративной защиты и отправителю.

Выводы

Использование программного средства позволит снизить количество инцидентов информационной безопасности, связанных с утечкой конфиденциальных данных через электронные письма, пересылаемые с помощью почтового сервера, повысит уровень репутации и конкурентоспособности предприятия.

Литература

1. Полякова, С.П. Системная интерпретация экономической и информационной безопасности предприятия/С.П. Полякова//Современные вызовы экономики и управления в России в условиях многополярного мира. – СПб: ООО «Скифия-принт», 2021. – С. 132-135.
2. Канищева, Е.А. Методы оценки опыта и деловой репутации предприятия/Е.А. Канищева, С.В. Ходыревская//Качество в производственных и социально-экономических системах. – Курск: Закрытое акционерное общество «Университетская книга», 2018. – С. 253-257.
3. Большинство сотрудников не задумываются о том, что их мессенджеры контролируют [Электронный ресурс]. – URL: <https://www.forbes.ru/tehnologii/411175-bolshinstvo-sotrudnikov-ne-zadumyvayutsya-o-tom-chto-ih-messendzhery-kontroliruyut> (Дата обращения 11.01.22).
4. Исследование уровня информационной безопасности в компаниях России и СНГ за 2020 год [Электронный ресурс]. – URL: <https://static.searchinform.ru/uploads/sites/1/2021/03/issledovanie-urovnya-ib-v-rf-i-sng-za-2020-god.pdf> (Дата обращения 11.01.22).

МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ

УДК 004.8

БИОМЕТРИЧЕСКАЯ СИСТЕМА КОНТРОЛЯ И УПРАВЛЕНИЯ ДОСТУПОМ НА ОСНОВЕ НЕЙРОННОЙ СЕТИ

BIOMETRIC ACCESS CONTROL AND MANAGEMENT SYSTEM BASED ON A NEURAL NETWORK

Журавлев Н.А., Фомичева О.Е.,
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский технологический университет «МИСиС»
г. Москва, Российская Федерация

N.A. Zhuravlev, O.E. Fomicheva,
Federal State Autonomous Educational Institution of Higher Education
“National Research Technological University “MISIS”
Moscow, Russian Federation

e-mail: nikitazhuravl@yandex.ru

Аннотация. В статье представлены результаты исследований и анализа охранных систем контроля и управления доступом (СКУД), рассмотрен теоретический базис для биометрического модуля распознавания лиц. В качестве теоретического базиса были рассмотрены структура нейронной сети, техническое обеспечение, инструменты реализации и потоки данных системы контроля и управления доступом на предприятии. Данные результаты исследований получены в ходе выполнения бакалаврской выпускной квалификационной работы кафедры Автоматизированных систем управления Национального исследовательского технологического университета (НИТУ) «МИСиС». Актуальность данной работы обусловлена необходимостью автоматизации систем контроля и управления доступом персонала на предприятии, в которых применяется устаревшее техническое и программное обеспечение. Также немаловажными факторами являются минимизация общих затрат на человеческие ресурсы и необходимость повышения уровня безопасности на предприятии. Модуль распознавания лиц является частью подсистемы идентификации и допуска персонала на объект, основан на методе идентификации лица по внешним признакам при помощи нейронной сети с использованием блочной структуры анализа объекта, позволяет минимизировать угрозу проникновения нежелательных лиц на объект, а также позволяет повысить эффективность работы охранной системы в целом.

Abstract. The article presents the results of research and analysis of security access control and management systems (ACS), the theoretical basis for the biometric facial recognition module is considered. The structure of the neural network, technical support, implementation tools and data flows of the enterprise access control and management system were considered as a theoretical basis. These research results were obtained during the bachelor's graduation qualification work of the Department of Automated Control Systems of the National Research Technological University (NUST) “MISIS”. The relevance of this work

is due to the need to automate systems for monitoring and controlling personnel access to the enterprise, which use outdated technical and software. Also important factors are minimizing the total cost of human resources and the need to increase the level of security at the enterprise. The facial recognition module is part of the subsystem of identification and admission of personnel to the object, is based on the method of identifying a person by external signs using a neural network using a block structure of object analysis, minimizes the threat of unwanted persons entering the object, and also allows to increase the efficiency of the security system as a whole.

Ключевые слова: СКУД система, биометрическая система, нейронная сеть, распознавание лиц.

Keywords: ACS system, biometric system, neural network, facial recognition.

Большинство современных предприятий используют автоматизированные системы контроля и управления доступом (далее СКУД). СКУД система – это комплекс взаимосвязанных программных и технических средств, с помощью которых реализуется функция охраны и контроля доступа на объект. На сегодняшний день наиболее популярным и практичным методом реализации СКУД систем является биометрическая система контроля доступа. Этот метод значительно повышает уровень безопасности предприятия и его сотрудников.

На начальных этапах реализации распознавания личности сотрудников предприятия устанавливались электронные турникеты со считывателями карт или отпечатков пальцев. Но сегодня, в связи с высоким уровнем развития науки о биометрике и биометрических информационных технологий, большинство компаний предпочитают иные методы контроля и доступа на объект, более точные и удобные в использовании. Например, метод идентификации лиц по видеопотоку в режиме реального времени.

Высокие требования компаний к охранным системам уже не одно десятилетие создают огромный спрос на рынке биометрической видеоаналитики. Безусловно в современных реалиях экономического кризиса в промышленной сфере сложно строить прогноз с учетом стоимости аппаратных средств для видеоидентификации. Но даже учитывая сверх завышенные цены на оборудование аналитическая компания Inkwood Research прогнозирует значительных рост рынка видеоаналитики с 3647 млн долл. в 2018 г. до 20450 млн долл. к 2027 году, со среднегодовым темпом роста CAGR в 20,61%. Больше всего рынок вырастет в Северной Америке: почти на 80%. Европа покажет рост около 50%, чуть меньший рост – страны Юго-Восточной Азии. [1]

Для наглядности текущей проблемы рассмотрим общую структуру стандартной охранной системы предприятия, а также проведем структурный системный анализ потоков данных в системе контроля и управления доступом.

В целом, процесс управления службой безопасности можно схематично описать следующим образом. При входе в проходное помещение сотрудник прикладывает магнитную карту к турникету. Охранник за стойкой следит за несанкционированным доступом, и в случае незаконного проникновения или другой нештатной ситуации вызывает сотрудников с поста охраны при помощи «тревожной кнопки». После прохода через турникет информация с магнитной карты передается через центральный контроллер СКУД в базу данных бюро пропусков, где происходит ведение картотеки и в отдел кадров, где в свою очередь производится учет рабочего времени вошедшего в здание сотрудника. Также информация о входе поступает на центральный пост службы безопасности, где происходит проверка сотрудника охранником-оператором. Если сотрудник перемещается в течение рабочего дня в закрытые служебные помещения, на

центральном посту службы безопасности появляется уведомление о санкционировании доступа. В случае несанкционированного доступа включается сигнал тревоги и с центрального поста службы безопасности оператор вызывает охранников с поста охраны на первом этаже для принятия служебных мер.

Рассмотрим движение потоков данных при помощи DFD диаграммы. DFD диаграмма наглядным образом поможет понять, из чего состоит информационная система охранной организации на основе СКУД и как ее необходимо обработать и структурировать для дальнейшей работы над проектированием модуля распознавания лиц для пропускной охранной системы предприятия (рис. 1).

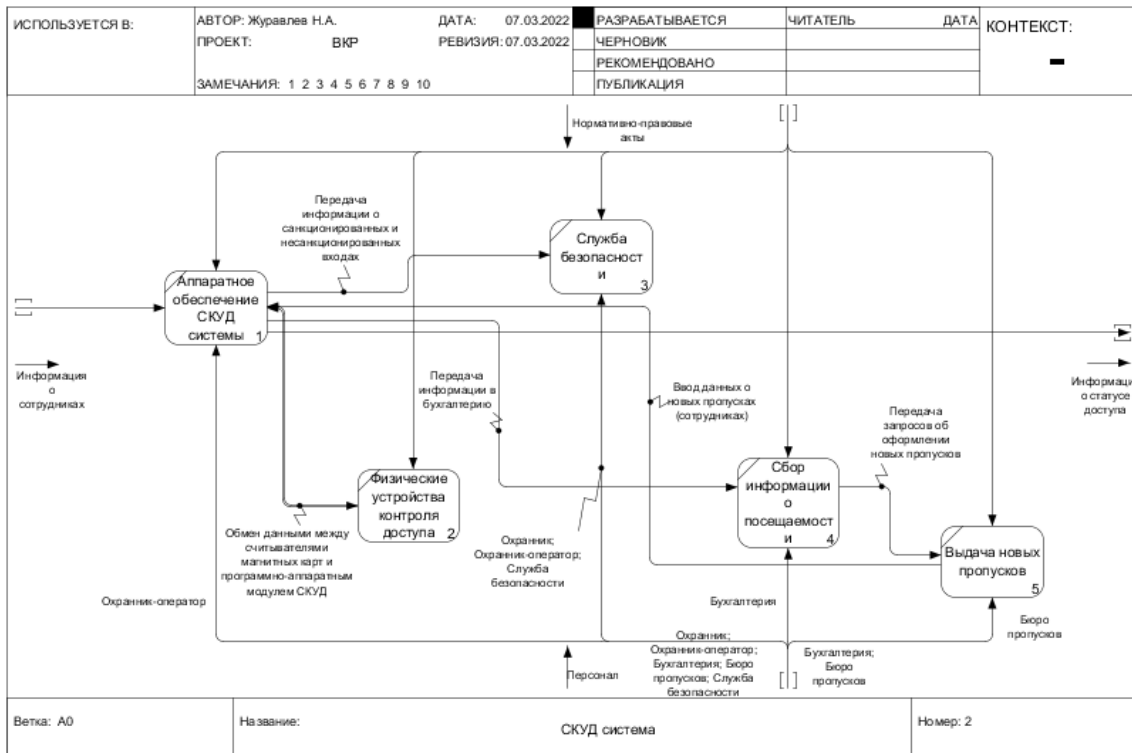


Рисунок 1. Общая диаграмма системы потоков данных СКУД системы

Информация о сотрудниках поступает в аппаратное обеспечение СКУД системы, откуда передается в службу безопасности, на физические устройства контроля доступа, в пункт сбора информации о посещаемости, а также в качестве выходного информационного потока «статус доступа». Периферийные источники и получатели информационных потоков также снабжают информацией аппаратное обеспечение СКУД системы.

Таким образом можно наглядно выделить ряд проблем:

- карту легко потерять – это отразится на потерянном рабочем времени и риске проникновения нежелательных лиц;
- карту легко подделать – на сегодняшний день магнитные ключ – карты дублируются повсеместно, даже в магазине ключей;
- всегда необходим сотрудник охраны, который будет проверять и идентифицировать, является ли человек истинным владельцем магнитной карты – это затрагивает немалые финансовые ресурсы;
- нужен сотрудник, который будет выдавать и менять карты, а также разбираться с возможными ошибками (в нашем случае сотрудник бюро пропусков) – также затрагивает финансовые ресурсы.

Решением данных проблем является неоднократно доказавший свою эффективность метод биометрической идентификации.

Существует множество методов распознавания по геометрии лица. Все они основаны на том, что черты лица и форма черепа каждого человека индивидуальны. Эта область биометрии многим кажется привлекательной, так как мы узнаем друг друга в первую очередь по лицу. Данная область делится на два направления: 2-D распознавание и 3-D распознавание. У каждого из них есть достоинства и недостатки, однако многое зависит еще и от области применения и требований, предъявляемых к конкретному алгоритму [2].

Безусловно, на данный момент наиболее эффективными способами идентификации являются уже старые и проверенные методологии в области биометрии. Технология идентификации по чертам лица появилась сравнительно недавно и стремительно развивается. Из безусловных плюсов стоит отметить относительную дешевизну разработки и установки таких систем, а также их стремительное развитие путем создания новых современных технологий.

На сегодняшний день существует достаточно большое количество возможных методик и алгоритмов в сфере распознавания и идентификации лиц и объектов. Можно выделить общую структуру процесса распознавания лиц (рис. 2).

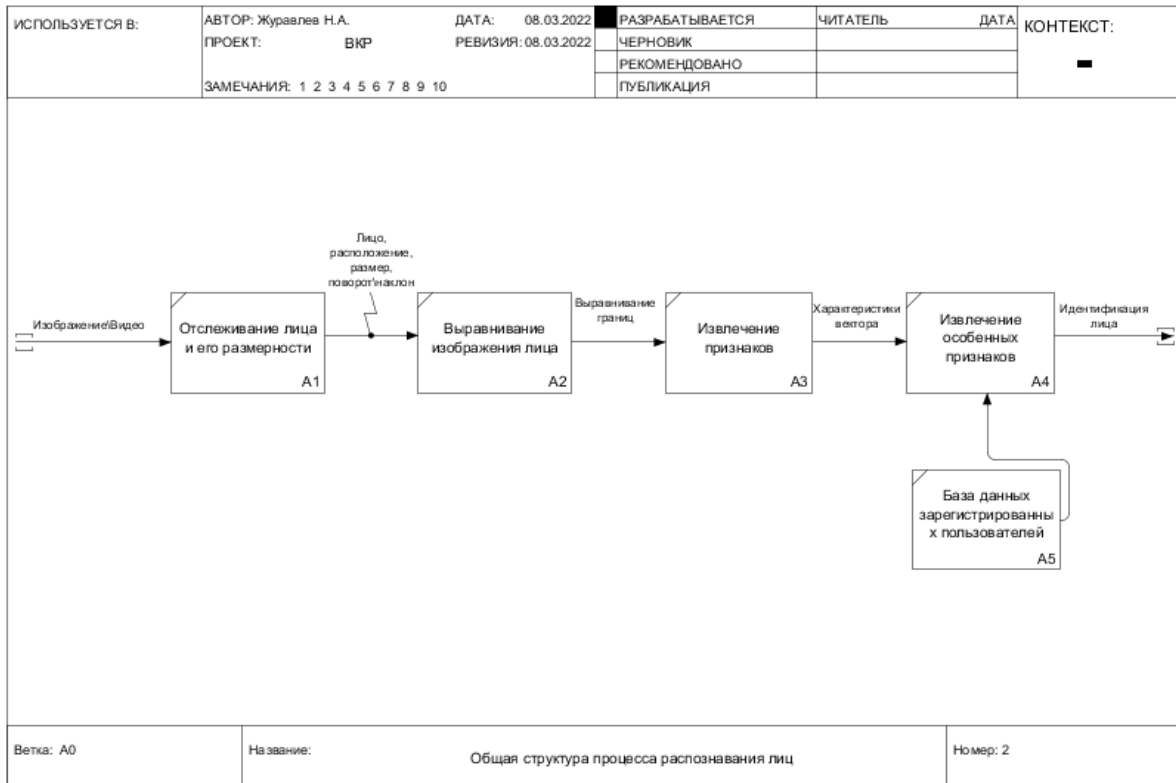


Рисунок 2. Общая структура процесса идентификации лица

Наиболее целесообразным является метод идентификации при помощи нейронной сети. В настоящее время существует около десятка разновидностей нейронных сетей. Одним из самых широко используемых вариантов является сеть, построенная на многослойном персептроне, которая позволяет классифицировать поданное на вход изображение/сигнал в соответствии с предварительной настройкой/обучением сети (рис. 3).

Обучаются нейронные сети на учебном наборе примеров. Суть обучения сводится к настройке весов межнейронных связей в процессе решения оптимизационной задачи

методом градиентного спуска. В процессе обучения нейронной сети происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Предполагается, что обученная нейронная сеть сможет применить опыт, полученный в процессе обучения, на неизвестные образы за счет обобщающих способностей [3].

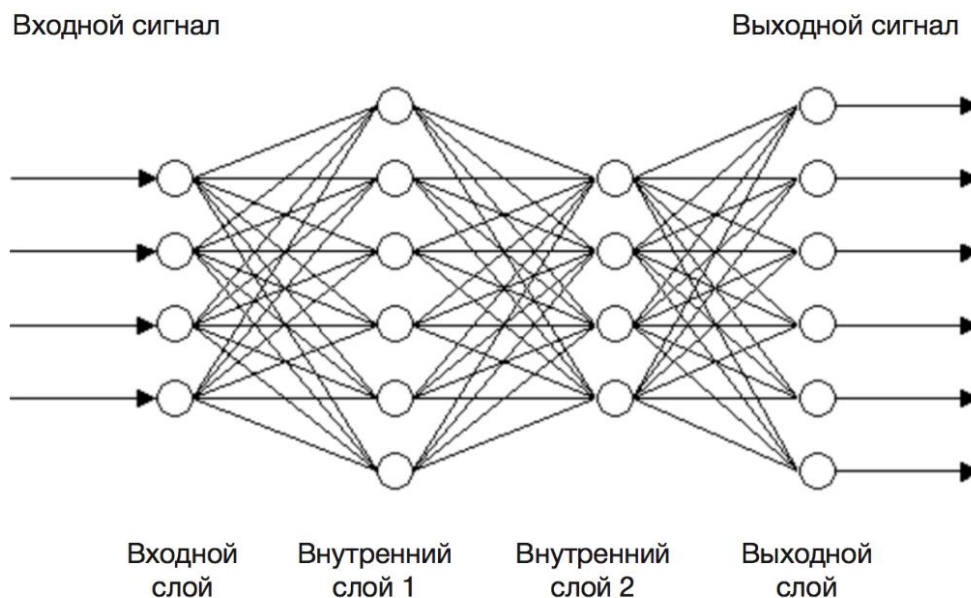


Рисунок 3. Структура нейронной сети на основе многослойного персептрона

Реализация данного метода требует стандартного технического и программного обеспечений, которые используются в СКУД системах. А именно, необходимо серверное оборудование, с помощью которого реализуется процесс обучения и переобучения сети; IP камера для получения входных биометрических данных; а также контроллер для передачи сигнала на открытие дверей или турникетов [4]. Модуль распознавания лиц реализован на языке программирования Python с использованием библиотеки OpenCV [5].

Выводы

Исследование структуры стандартных систем контроля и управления доступом показывает их недостаточный уровень обеспечения безопасности. Магнитные карты, которые являются ключевым методом идентификации, несут в себе ряд проблем, которые сказываются на безопасности охраняемого объекта.

Решением данных проблем является модуль распознавания лица в видеопотоке для СКУД системы. В качестве метода идентификации используется – метод нейронной сети, который является наиболее эффективным и актуальным для решения данных проблем. Результатом является повышение уровня безопасности на предприятии и снижение затрат на человеческие ресурсы.

Литература

1. Архипова Е.А., Поляничко К.С. Тенденции развития видеоаналитики в мире, Архипова Е.А., Поляничко К.С. Тенденции развития видеоаналитики в мире, 2019.

2. Волковицкий В.Д., Волхонский В.В. Системы контроля и управления доступом. СПб.: Университет ИТМО, 2015. 53 с.
3. Брилюк Д.В., Старовойтов В.В. Распознавание человека по изображению лица нейросетевыми методами. Минск, 2002. 54 с.
4. Мащенко Р.Г. Системы охранной сигнализации: основы теории и принципы построения: учебное пособие. М.: Горячая линия – Телеком, 2004.
5. Зайцева Е.В., Степанова А.Л. Классификация современных методов трекинга объектов в интеллектуальных системах видеонаблюдения. Горный информационно-аналитический бюллетень ОВ N5, 2013, Москва, МГГУ, с. 56-63.

УДК 004.932

СПОСОБ ПОДБОРА СВЕРТОЧНЫХ ЯДЕР НА ПРИМЕРЕ MNIST

METHOD OF SEARCHING OF CONVOLUTIONAL KERNELS ON MNIST

Хлыбов А.В., Гиниятуллин В.М., Ермолаев Е.В.,
Уфимский государственный нефтяной технический университет,
ул. Космонавтов, 1, г. Уфа, Республика Башкортостан, 450064, Россия

A.V. Khlybov, V.M. Giniyatullin, E.V. Ermolaev,
Ufa State Petroleum Technological University,
Kosmonavtov Str., 1, Ufa, Republic of Bashkortostan, 450064, Russia

e-mail: brinkinvision@gmail.com

Аннотация. Современные модели нейронных сетей содержат десятки и сотни ядер в сверточных слоях, каждое из которых необходимо обучить. Процесс обучения глубоких нейронных сетей и применение рассчитанных весов на рабочем ходе является ресурсоёмким. Было показано, что сверточные ядра с целочисленными значениями -1, 0 и +1 позволяют заменить обучаемые ядра, состоящие из многозначных чисел, на подбираемые. При этом перебор осуществляется с малой размерности ядер. Был сформулирован критерий оптимальности подбора сверток для прикладной задачи, с помощью которого можно подбирать работоспособные варианты ядер. Желательно, чтобы в сверточном слое ядра были большого размера (предположительный максимум 15*15). Это позволит уменьшить их количество в слое при той же точности модели. В результате проведения полного перебора на наборе данных MNIST оказалось, что ядра должны быть несбалансированными, «скошенными» в положительную область. Работоспособные ядра 3*3 для исследуемого датасета содержат одно значение «-1», одно-три значения «0», остальное – «+1». В случае ядер с худшей точностью оказалось, что для них характерно давать в качестве результата значение класса «1», что связано с большим количеством экземпляров класса «1» в обучающей выборке.

Abstract. Modern neural network models contain tens and hundreds of kernels in convolutional layers, each of which needs to be trained. The process of training deep neural networks and applying the calculated weights on the working pass is resource-intensive. It was shown that convolutional kernels with integer values -1, 0 and +1 make it possible to replace trainable kernels consisting of multi-valued numbers with selected ones. In this case, the enumeration is carried out from a small size of the kernels. A criterion for the optimal searching

of convolutions for an applied problem was formulated, with the help of which it is possible to search efficient variants of kernels. It is desirable that the kernels in the convolutional layer be large (assumed maximum 15*15). This will reduce their number in the layer with the same accuracy of the model. As a result of the full search on the MNIST dataset, it turned out that the kernels should be unbalanced, “skewed” into the positive axis. Efficient 3*3 kernels contain one “-1” value, one-three “0” values, and the rest “+1” values. In the case of kernels with worse accuracy, it turned out that it is typical for them to give the value of class “1” as a result, which is associated with a large number of examples of class “1” in the training sample.

Ключевые слова: сверточное ядро, mnist, троично-сбалансированная система счисления, целочисленные коэффициенты, весовые матрицы.

Keywords: convolutional kernel, mnist, ternary-balanced system, integer, weight matrices.

Базовая архитектура сверточных нейронных сетей основана на использовании двумерных сверточных ядер (2D convolution). Основной областью использования сверточных слоев является обработка изображений. Применение сверточных слоев вместо полносвязных перед выходным слоем нейронной сети обосновывается тем, что в случае только полносвязных слоев количество нейронов на входе будет равно входному вектору изображения. Даже небольшое изображение 10*10 будет требовать 100 нейронов на входе. Также сверточные слои позволяют учитывать зависимости и закономерности изображения в пространстве [1].

Однако задача обработки изображений является ресурсоемкой. Поэтому необходимы способы сокращения и упрощения нейронных моделей, предназначенных для их обработки [2].

Целью работы является разработка формализованной методики поиска работоспособного ядра 3*3 для набора данных MNIST [3].

Исходный набор данных содержит изображения рукописных цифр от 0 до 9 в градациях серого. Размерность каждого изображения 28*28 пикселей. Значения пикселей изменяются в диапазоне от 0 до 255. Обучающая выборка содержит 60 000 изображений, тестовая – 10 000. В таблице 1 представлено распределение количества экземпляров каждого класса.

Таблица 1 – Распределение по классам количества экземпляров

| Класс | Количество в обучающей выборке | Количество в тестовой выборке |
|-------|--------------------------------|-------------------------------|
| 0 | 5923 | 980 |
| 1 | 6742 | 1135 |
| 2 | 5958 | 1032 |
| 3 | 6131 | 1010 |
| 4 | 5842 | 982 |
| 5 | 5421 | 892 |
| 6 | 5918 | 958 |
| 7 | 6265 | 1028 |
| 8 | 5851 | 974 |
| 9 | 5949 | 1009 |

Для исследований была выбрана модель LeNet-5 [4], структура которой представлена в таблице 2.

Таблица 2 – Структура LeNet-5

| Название | Основные параметры |
|-------------------|--|
| Сверточный слой | 6 ядер размерности 5*5 |
| AveragePool | 2*2, шаг=2 |
| Сверточный слой | 16 ядер размерности 5*5 |
| AveragePool | 2*2, шаг=2 |
| Полносвязный слой | 120 нейронов, функция активации=tanh |
| Полносвязный слой | 84 нейрона, функция активации=tanh |
| Полносвязный слой | 10 нейронов, функция активации=softmax |

После обучения точность модели составила по обучающей выборке (train_acc) 99.17 %, по тестовой (test_acc) – 98.8 %.

Исходная структура нейронной сети была изменена следующим образом: имеется всего одно ядро 3*3 в сверточном слое и 10 нейронов в выходном полносвязном слое (таблица 3).

Таблица 3 – Упрощенная структура нейронной сети

| Название | Основные параметры |
|-------------------|--|
| Сверточный слой | 1 ядро размерности 3*3 |
| Полносвязный слой | 10 нейронов, функция активации=softmax |

Точность упрощенной нейронной сети уменьшилась до следующих значений: по обучающей выборке=90.91 %, по тестовой=91.06 %. Полученные в процессе обучения весовые коэффициенты сверточного ядра представлены на рисунке 1.

| | | |
|---------|---------|---------|
| 0.4992 | 0.4859 | -0.2383 |
| -0.1556 | -0.097 | 0.1565 |
| 0.0781 | -0.3754 | -0.0652 |

Рисунок 1. Обученное сверточное ядро на упрощенной структуре (train_acc=90.91 %, test_acc=91.06 %)

Упрощенная структура нейронной сети (таблица 3) была использована для перебора сверточных ядер размерности 3*3, значения коэффициентов которых дискретны и сведены к трехзначности (-1, 0, +1) [5]. Всего ядер $3^9=19683$, ядра нумеруются в троично-сбалансированной системе счисления. Поскольку ячейки матрицы нумеруются с 0, начиная с левого верхнего угла, то из каждого ядра свертки можно сформировать троично-сбалансированное число, где номер ядра – это десятичное представление троично-сбалансированного числа, номер ячейки – знакоместо троично-сбалансированного числа, значение в ячейке – троичное значение получаемого числа.

В результате перебора был получен набор ядер с различной точностью распознавания. Значения точности по обучающей выборке были сведены в график

(рисунок 2). Из рисунка 2 видно, что ядра, лежащие в положительной полуоси X, дают в среднем большую точность, чем ядра из отрицательной полуоси.

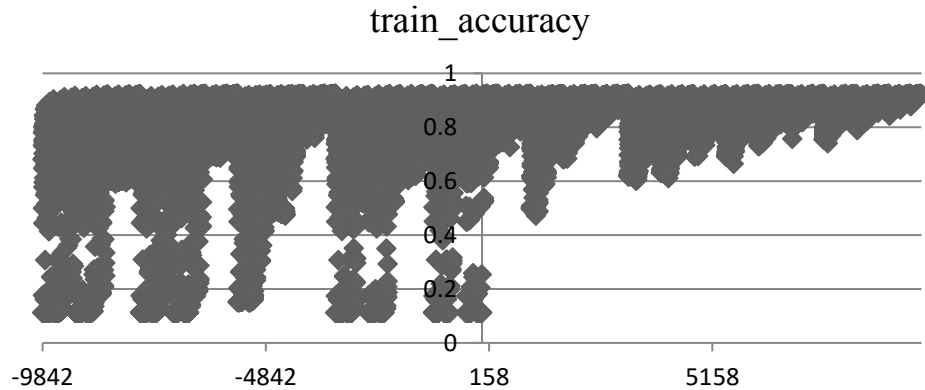


Рисунок 2. Иллюстрация результирующих значений точности по обучающей выборке для всех перебираемых ядер 3*3

В результате полного перебора наибольшая точность нейронной сети по обучающей выборке (train_acc) составила 92.49 %, а по тестовой (test_acc) 92.60 %, что соответствует ядрам №-5468 и №6168 (рисунок 3).

| Ядро №-5468 | | |
|-------------|-------|-------|
| -1^0 | 0^1 | 1^2 |
| 1^3 | 1^4 | 1^5 |
| 1^6 | 1^7 | 1^8 |

| Ядро №6168 | | |
|------------|-------|--------|
| 1^0 | 0^1 | -1^2 |
| 1^3 | 1^4 | 0^5 |
| 1^6 | 1^7 | 0^8 |

Рисунок 3. Ядра с наибольшей точностью распознавания по обучающей (№-5468, train_acc=92.49 %, test_acc=92.57 %) и по тестовой (№6168, train_acc=92.28 %, test_acc=92.60 %) выборке

Ядра 3*3 с максимальной точностью распознавания оказались несбалансированными, «скошенными» в положительную область. Они содержат одно значение «-1», два-три значения «0» и остальное – значения «+1», в отличие от обученного ядра (рисунок 1), где значения распределены примерно равномерно.

Ядра с наихудшей точностью распознавания имеют примерно одинаковую точность (train_acc=11.24 %, test_acc=11.35 %). Примеры таких ядер представлены на рисунке 4.

| Ядро №-9841 | | |
|-------------|--------|--------|
| -1^0 | -1^1 | -1^2 |
| -1^3 | -1^4 | -1^5 |
| -1^6 | -1^7 | -1^8 |

| Ядро №-30 | | |
|-----------|--------|--------|
| 0^0 | 0^1 | 0^2 |
| 0^3 | 0^4 | -1^5 |
| 0^6 | -1^7 | 0^8 |

Рисунок 4. Примеры ядер с худшей точностью распознавания

Для плохих ядер вектор вероятности определения класса примерно равен следующему [0.1; 0.11; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1; 0.1].

Например, для ядра №-9841 при предсказании значения экземпляра под номером 2 он равен:

[0.10057574; 0.11252894; 0.09987152; 0.10223734; 0.09668048; 0.0894477; 0.09821966; 0.10542194; 0.09736338; 0.0976533].

Наибольшее значение вероятности определения класса «1» обусловлено тем, что экземпляров этого класса в обучающей выборке большее количество, и оно равно 6742 (таблица 1).

Выводы

В статье была описана методика подбора сверточного ядра 3×3 на примере набора данных MNIST. Алгоритм состоит в следующем:

- упрощается архитектура существующей сети до такого состояния, чтобы в первом сверточном слое осталось одно ядро;
- перебор происходит на ядрах малого размера (2×2 , 3×3);
- в качестве критерия оптимальности ядра выбраны значения метрики accuracy по обучающей и тестовой выборке;
- подобрав работоспособное (возможно не оптимальное) ядро, можно либо увеличивать его размер, либо подбирать к нему дополнительные ядра.

Таким образом, разработанный алгоритм может использоваться в качестве библиотеки для оптимизации архитектур сверточных нейронных сетей.

Литература

1. Saha S.A Comprehensive Guide to Convolutional Neural Networks – the ELI5 way [Электронный ресурс]. 2018. URL: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (дата обращения: 29.03.22).
2. Pan W., Dong H., Huo Y. DropNeuron: Simplifying the Structure of Deep Neural Networks [Электронный ресурс]. URL: <https://arxiv.org/abs/1606.07326> (дата обращения: 22.03.2022), 2016.
3. Digit Recognizer [Электронный ресурс]. URL: <https://www.kaggle.com/c/digit-recognizer/data> (дата обращения: 22.03.2022).
4. Module05_mnist_conv.ipynb [Электронный ресурс]. URL: https://github.com/SlinkoIgor/Neural_Networks_and_CV/blob/master/module05_mnist_conv.ipynb (дата обращения: 22.03.2022).
5. Filippova A.G., Belozyorov E.S., Filippov V.N., Belozyorov A.E., Sultanova E.A. Development Method of Designing a Typed Unified user Interface [Электронный научный журнал Нефтегазовое дело. 2013. № 4. С. 510-528.

СЕТИ И ТЕЛЕКОММУНИКАЦИИ

УДК 004.057.2

ВЫБОР ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБСЛУЖИВАНИЯ В СОГЛАШЕНИИ ОБ УРОВНЕ ОБСЛУЖИВАНИЯ

SELECTING QUALITY OF SERVICE INDICATORS IN A SERVICE LEVEL AGREEMENT

Фокин А.Б, Шмойлов А.В, Коркин А.Г.,
Академия ФСО России,
г. Орел, Российская Федерация

A.B. Fokin, A.V. Shmoilov, A.G. Korkin,
Academy of the FSO of Russia,
Orel, Russian Federation

e-mail: tatarin57ru@mail.ru

Аннотация. На сегодняшний день сфера телекоммуникации и инфокоммуникации развивается очень стремительно, количество услуг и пользователей растет с каждым днем. Все это ведет к существенному увеличению сложности и размеров сетей связи, что приводит к значительно большей подверженности различного рода авариям и выносит на первый план проблемы, связанные с анализом надежности этих сетей. Операторы конкурируют между собой и дорожат своими клиентами, предлагая им высокое качество, безопасность и постоянную готовность услуги. Однако между поставщиком услуг и потребителем в любой момент времени могут возникнуть споры, связанные с качеством предоставляемых услуг связи. Эффективным инструментом, регулирующим порядок взаимоотношений и ответственности между потребителем и поставщиком услуг (оператор), направленный на обеспечение показателей качества услуг связи в соответствии с уровнем обслуживания является соглашение об уровне обслуживания (SLA). В соглашении об уровне обслуживания оператор решает следующие задачи, выбор показателей качества обслуживания, ответственность поставщика услуг (оператора) за выполнение установленных нормативов, организацию мониторинга установленных показателей качества, контроль выполнения соглашения и доведение информации до пользователей. Из перечисленных задач, выбор показателей качества обслуживания и нормативов для них представляет наибольший интерес, как для операторов, так и для пользователей. Выбираемые показатели делятся на две категории: специальные и общие.

Abstract. Today, the field of telecommunications and infocommunication is developing very rapidly, the number of services and users is growing every day. All this leads to a significant increase in the complexity and size of communication networks, which leads to a much greater susceptibility to various kinds of accidents and brings to the fore the problems associated with the analysis of the reliability of these networks. Operators compete with each other and value their customers, offering them high quality, security and constant availability of services. However, disputes may arise at any time between the service provider and the consumer related to the quality of the communication services provided. The Service level

agreement (SLA) is an effective tool that regulates the relationship and responsibility between the consumer and the service provider (operator), aimed at ensuring the quality of communication services in accordance with the service level. In the service level agreement, the operator solves the following tasks: the choice of service quality indicators, the responsibility of the service provider (operator) for meeting the established standards, the organization of monitoring of established quality indicators, monitoring the implementation of the agreement and communicating information to users. Of the listed tasks, the choice of service quality indicators and standards for them is of the greatest interest for both operators and users. The selected indicators are divided into two categories: special and general.

Ключевые слова: оператор, SLA, качество обслуживания, коэффициент готовности.

Keywords: operator, SLA, quality of service, availability factor.

Совершенствование существующих или проектирование новых телекоммуникационных сетей ведет к существенному увеличению их сложности и размеров, что приводит к значительно большей подверженности различного рода авариям и выносит на первый план проблемы, связанные с анализом надежности этих сетей. От надежности сети связи зависит качество предоставляемых услуг операторами своим потребителям. При этом оператор должен гарантировать качество предоставляемых услуг, требуемое заказчиками, исходя из показателей надежности своей сети, которое он может предоставить.

В настоящее время в отраслях информационных и коммуникационных технологий эффективным инструментом, регулирующим порядок взаимоотношений и ответственности между потребителем и поставщиком услуг (оператор), направленный на обеспечение показателей качества услуг связи в соответствии с уровнем обслуживания является соглашение об уровне обслуживания (Service Level Agreement, SLA). Впервые понятие SLA введено в библиотеке инфраструктуры информационных технологий (ITIL). Согласно [1] соглашение об уровне обслуживания (SLA) это документированное соглашение двух и более сторон, определяющее гарантированные значения показателей качества услуги связи в соответствии с уровнем обслуживания, порядок взаимодействия и ответственность сторон. SLA применяются ведущими операторами связи при оказании услуг корпоративным клиентам и другим операторам (провайдерам).

Основной идеей соглашения об уровне обслуживания является гарантийные обязательства поставщиком услуг по предоставлению определенных требований к уровню обслуживания. Под обязательствами понимается заявленный оператором уровень обслуживания, который он гарантированно обеспечивает на основе соглашения об уровне обслуживания - гарантированный уровень обслуживания [1]. Нарушение которого приводит оператора прежде всего к материальной ответственности, а также потери репутации перед потребителем и даже к расторжению договоров и перехода клиентов к конкурентам.

Для успешного применения SLA оператор решает ряд задач, таких как выбор показателей качества обслуживания и нормативов для них, ответственность поставщика услуг (оператора) за выполнение установленных нормативов, а также организацию мониторинга установленных показателей качества, контроль выполнения соглашения и доведение информации до пользователей.

В рамках данной статьи из всех перечисленных задач, решаемых оператором, наибольший интерес представляет задача по выбору показателей качества обслуживания

и нормативов для них. От выбора этих показателей во многом зависит качество предоставляемых услуг оператором. Выбираемые оператором показатели можно разделить на две категории: специальные показатели, зависящие от технологии и/или услуг и общие показатели, не зависящие от этого. К специальным показателям относят показатели качества, принятые в Рекомендациях МСЭ-Т и других документах междугородных организаций, установленные для соответствующего вида услуг и/или сетей. К ним относятся такие показатели как задержка переноса пакетов (кадров), вариация переноса пакетов (джиттер), коэффициент потери пакетов (кадров) и др. А к числу общих показателей в основном относят показатели надежности (готовности), определяющие количественные характеристики свойств надежности (безотказность, долговечность, ремонтпригодность, сохраняемость). Они бывают единичными – характеризующие одно свойство надежности (вероятность безотказной работы, средняя наработка между отказами, вероятность восстановления, средний ресурс и др.) и комплексными – характеризующие несколько свойств (коэффициент готовности, коэффициент оперативной готовности, коэффициент технического использования и др.).

Согласно [1, 2, 3, 4] одним из важных пунктов, включаемых в состав SLA, являются требования к надежности [5], которое оказывает существенное влияние на качество обслуживания (QoS). В качестве одного из таких показателей используется готовность соединения, на практике в SLA чаще других используют показатель коэффициента готовности. Коэффициент готовности, определяет допустимое суммарное время работоспособности за определенный временной период, без учета периодов предназначенных для выполнения профилактических работ, переключений, измерений и т.п., что соответствует определению в [6]. Определение коэффициента готовности в современных сетях, которые являются сложноразветвленными, является NP-трудной задачей. Получить точные значения коэффициента готовности известными методами, описанными в [7] оказывается не всегда возможным за приемлемое время, однако разработаны современные методы, представленные в [8, 9] применение которых способствует его расчету в приемлемые сроки.

Таким образом, для поставщика услуг выбор нормативных значений параметров для включения в соглашение об уровне обслуживания, в частности коэффициента готовности является «краеугольным камнем». Так как, с одной стороны, желание удовлетворить требования пользователей и опередить конкурентов подталкивает оператора к установлению высоких норм, с другой стороны, необходимо обеспечить реальную достижимость взятых на себя обязательств и минимизировать риск нарушений SLA, влекущих репутационные и финансовые потери [10]. Исходя из этого, оператор обязан внимательно проанализировать свои реальные возможности, по оказанию качественных услуг связи, прежде чем предлагать своим клиентам SLA. В противном случае, придется или понизить требования, или принять меры по повышению качества и надежности сети.

Выводы

SLA являются апробированным и эффективным инструментом регулирования отношений между потребителями и поставщиками услуг, направленным на обеспечение качества услуг. Порядок их применения установлен в ряде международных и российских стандартов. Важным пунктом в соглашении является выбор показателей качества обслуживания и нормативов для них, который оператор обязан выполнять и одним из значимый является показатель готовности, выраженный через коэффициент готовности. На сегодняшний момент практически все ведущие операторы связи Российской

Федерации предлагают своим пользователям соглашения об уровне обслуживания (SLA).

Литература

1. ГОСТ Р 55389-2012. Система национальных стандартов в области качества услуг связи. Соглашение об уровне обслуживания (SLA).
2. ITU-T Recommendation E.860 (06/2002). Framework of a service level agreement.
3. ITU-T Recommendation M.3342 (07/2006). Guidelines for the definition of SLA representation templates.
4. TM Forum GB917. SLA Management Handbook. Rel. 3.1, v. 1.2. 2012.
5. Нетес В.А. Задание требований по надежности в соглашениях об уровне обслуживания//Электросвязь. 2004. № 4. С. 37-39.
6. ГОСТ 27.002–2015. Надежность в технике. Термины и определения.
7. Филин Б.П. Методы анализа структурной надежности сетей связи. – Москва: Радио и связь, 1988. – 208 с.
8. Батенков К.А. Формирование сечений телекоммуникационных сетей для анализа их устойчивости с различными мерами связности./К.А. Батенков, А.А. Батенков, А.Б. Фокин//Информатика и автоматизация. – 2021. – Т. 20. – №2. – С. 371-406.
9. Батенков, К.А. Методы формирования множеств состояний телекоммуникационных сетей для различных мер связности./К.А. Батенков, А.Б. Фокин, А.А. Батенков//Труды СПИИРАН. – 2020. – Т.19. – № 3. – С. 664-673.
10. Нетес В.А. Соглашения об уровне обслуживания и надежность//Надежность. 2017. № 4. С. 27-30.

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ И ГИС-ТЕХНОЛОГИИ

УДК 004.41

ГЕОСОЦИАЛЬНЫЕ СЕТИ И КАРТОГРАФИРОВАНИЕ GEOSOCIAL NETWORKS AND MAPPING IN EMERGENCIES

Набиева К.И., Баймурзин Д.И.,
Уфимский государственный нефтяной технический университет,
ул. Космонавтов, 1, г. Уфа, Республика Башкортостан, 450064, Россия

K.I. Nabieva, D.I. Baimurzin,
Ufa State Petroleum Technological University,
Kosmonavtov Str., 1, Ufa, Republic of Bashkortostan, 450064, Russia

e-mail: nabieva310199@mail.ru

Аннотация. Рост социальных геолокационных сетей и картографирования приводит к более эффективному сбору и распространению точной информации, а также, быстрому реагированию в наиболее пострадавших районах, в результате которого может значительно снизиться ущерб в чрезвычайных ситуациях. Геосоциальные сети позволяют собирать информацию, которая может служить основным источником понимания масштабов аварийных ситуаций. В данной работе рассматривается использование социальных сетей на основе геолокации в чрезвычайных ситуациях, поскольку своевременные, точные и понятные коммуникации являются основным условием для правильного реагирования в экстренных условиях. На сегодняшний день, актуальность геосоциальных сетей стремительно развивается. Их преимущество ещё и в объединении необходимых функций для работы с картой в одном приложении, удобстве и экономии времени. Статья имеет следующую структуру – во втором разделе определены ключевые термины и представлена статистика распространения социальных сетей в мире. Третья часть посвящена влиянию социальных сетей в аварийных ситуациях, уделено особое внимание геосоциальным сетям и сервисам. Рассмотрены примеры таких ситуаций стихийных бедствий в Гаити и Японии. Далее, сформированы выводы.

Abstract. The growth of geolocation-based social networks and mapping leads to more efficient gathering and dissemination of accurate information and rapid response in the most affected areas, which may significantly reduce damage in emergency situations. These services collect information that can serve as a major source of understanding dimensions of an emergency. In this paper, the use of geolocation-based social networks and services in emergency situations is discussed, because timely, accurate, and clear communication is the main prerequisite for a valid response to such conditions. Today, the relevance of geosocial networks is rapidly developing. Their advantage is also in combining the necessary functions for working with the map in one application, convenience and saving time. The paper is organized as follows - the key terms are defined in the second section and the statistics of social networks distribution in the world and in our country are presented. The third part deals with the influence of social networks in crisis situations, with special emphasis on geolocation-based

networks and services. Examples of such situations of natural disasters in Haiti and Japan are considered. A commentary is given in the form of a conclusion.

Ключевые слова: геолокационные социальные сети, картографирование, чрезвычайные ситуации, коммуникация, технологии.

Keywords: geolocation-based social networks, mapping, emergencies, communication, technology.

Введение

Широкое использование и многочисленные возможности, предлагаемые сетью интернет, существенно влияют на современное общество. Учитывая тот факт, что мы живем в эпоху информационных технологий, ежедневное использование данной платформы может отражаться, помимо личного пользования, в экономике, политике, здравоохранении, образовании, туризме и многих других областях. Таким образом, можно сделать вывод, что роль интернета обширна и неизмерима.

Применение инструментов и технологий Web 2.0 позволяет своим пользователям создавать и изменять контент – активно участвовать в создании интернета. Таким образом, концепции, лежащие в основе коммуникаций, сотрудничества и обучения, в современном обществе существенно трансформируются. С помощью социальных сетей, таких как Facebook, Twitter и YouTube, ежедневно происходит обмен большими объемами информации в виде текста, изображений и видео. Считается, что социальные сети являются основой современного электронного общения и делают мир более открытым и взаимосвязанным.

1. Определения, виды и распределение социальных сетей

1.1 Определение социальных сетей

В литературе, в зависимости от точки зрения, можно найти разные определения социальных сетей:

Социальные сети – это совокупность интернет-приложений, платформ и средств массовой информации, которые позволяют осуществлять коммуникацию между людьми, совместно создавать и совместно использовать различный контент [1].

Контент, которым пользователи делятся на различных веб-сайтах и сервисах, облегчающих общение между ними, называется социальными средствами массовой информации, это один из основных способов социального взаимодействия в современном обществе.

Социальные сети можно назвать веб-службами, которые позволяют пользователям создавать общедоступные или приватные личные профили и обмениваться сообщениями с другими пользователями в той же системе. На сегодняшний день, термины «социальная сеть» и «социальное взаимодействие» являются эквивалентными. Социальное общение характеризуется социальным взаимодействием между членами группы, а социальная сеть создает связь между людьми в пределах социальных сетей.

Проще говоря, социальные сети – это сервисы, с помощью которых мы делимся контентом или ищем информацию в нашей виртуальной среде

1.2 Виды социальных сетей

Социальные сети позволяют каждому человеку создавать и просматривать списки друзей, писать комментарии к профилям друзей, отправлять личные сообщения, обмениваться фотографиями, видео и другим мультимедийным контентом, позволяя каждому пользователю активно участвовать в создании контента. Социальные сети можно разделить на несколько основных категорий в зависимости от их назначения и контента, который они продвигают (рисунок 1):

- социальное взаимодействие (Facebook³, Twitter¹, Google+),
- мультимедийные социальные сети (YouTube, Instagram¹, Tumblr),
- информативные социальные сети (doityourself.com, forbesstockpicking.com, hgtv.com),
- профессиональные социальные сети (linkedin.com, classroom20.com, canonprofessionalphotographer.com),
- образовательные (thestudentroom.com, themathforum.com, eLearners.com),
- научные (academia.edu, researchgate.net),
- направленные на определенные области (gardenweb.com, automotiveforums.com, sportspundit.com).

Вышеперечисленная информация показывает, что в настоящее время существуют сайты социальных сетей для любой области интересов – спорта, фотографии, кулинарии, путешествий и т. д. Существуют сайты для обмена фотографиями, видео, информирования других о различных мероприятиях, знакомств с новыми людьми и восстановления связи со старыми друзьями.

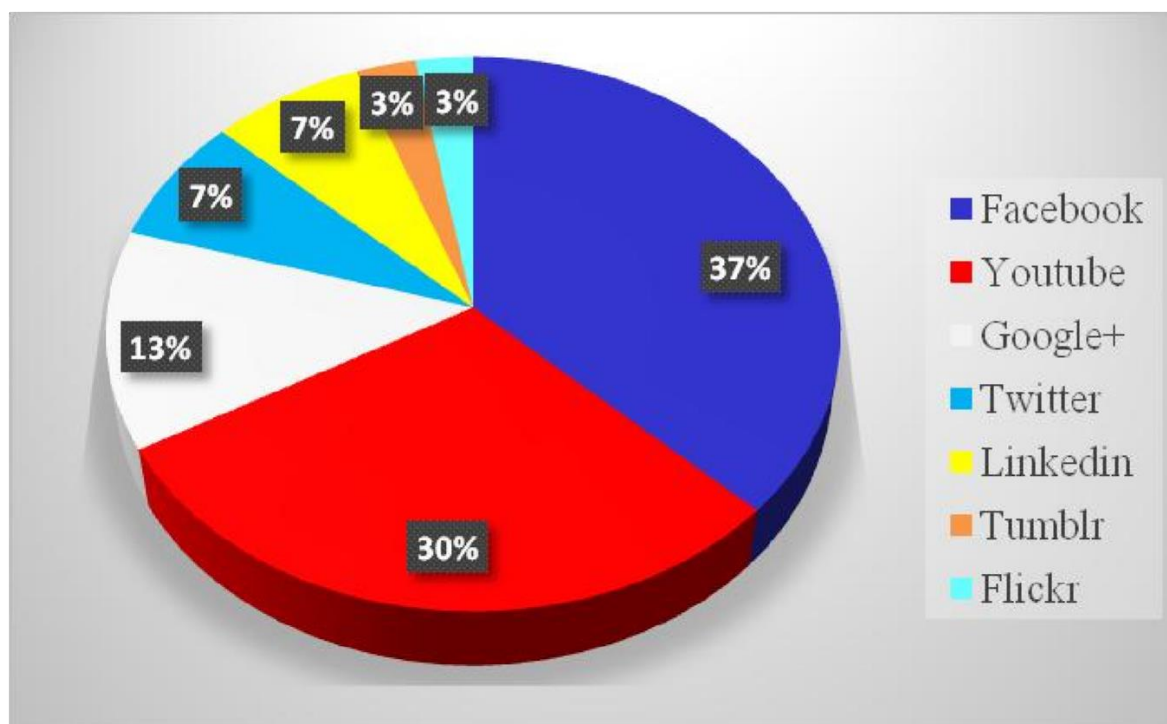


Рисунок 1. Самые популярные социальные сети

Кроме того, социальные сети могут быть полезны для укрепления общественного авторитета, чтобы отдельные лица или организации могли представить свои эксперты в определенных областях во всем мире.

³ Запрещена на территории России

Как показано на рисунке 1, наиболее популярными являются те социальные сети, основная функция которых – простое социальное взаимодействие. Некоторые социальные сети содержат контент различных типов и не принадлежат исключительно к одной категории.

1.2.1 Facebook⁴

Facebook был основан Марком Цукербергом в 2004 году. Сегодня эта социальная сеть насчитывает около 1,4 миллиарда активных пользователей. В России количество его пользователей достигает около 1,6 миллионов – 40,2% мужчин и 59,8% женщин (рисунок 2).

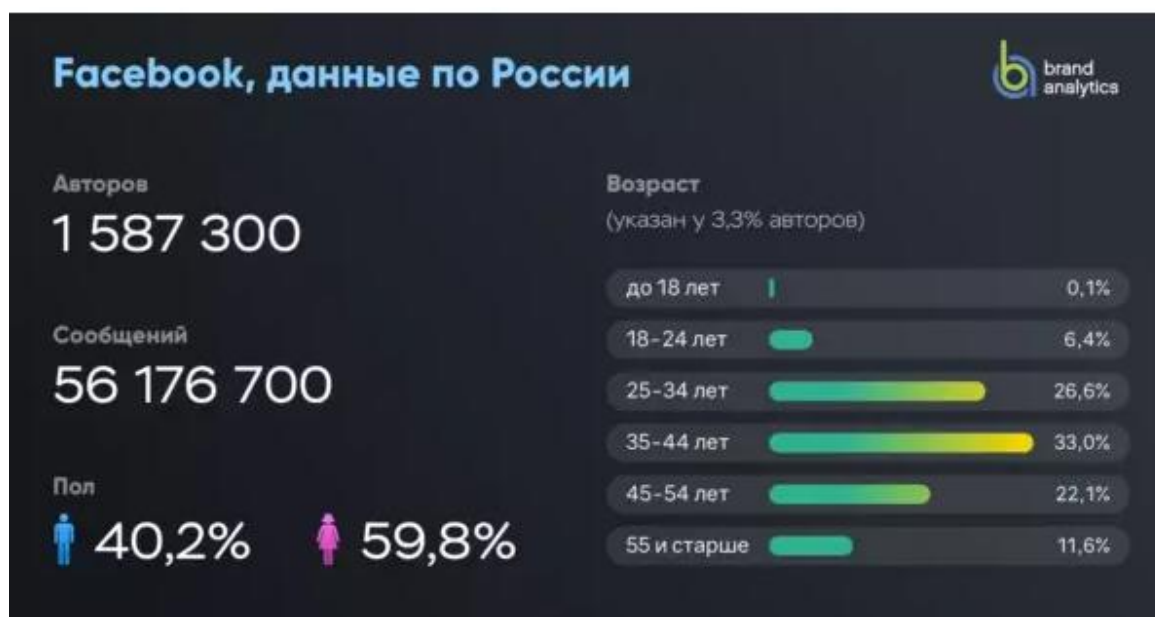


Рисунок 2. Пользователи Facebook в России

Facebook – это социальная сеть, основанная на модели дружбы. Его пользователям доступен широкий спектр инструментов, таких как обмен неограниченным количеством фотографий и видео.

1.2.2. Twitter²

Twitter – это социальная сеть, созданная Джеком Дорси в 2006 году. Для коммуникации пользователи могут создавать сообщения, называемые твитами. Данной социальной сетью пользуются более чем 200 миллионов пользователей, это одна из самых популярных социальных сетей после Facebook. Количество активных пользователей Twitter в России составляет около 690 300.

Один твит ограничен 140 символами, что составляет длину SMS-сообщения. Поэтому Twitter рассматривается как SMS в Интернете. Twitter также можно использовать как блог, чтобы твиты были видны всем.

⁴ Запрещена на территории России

1.2.3. Геолокационные социальные сети и службы

Технология геолокации, использующая данные, полученные с компьютера или мобильного устройства, подключенного к Интернету, позволяет идентифицировать физическое местоположение устройства, с которого отправляются данные. Данные, полученные таким образом, в основном используются в целях определения местоположения (определение географического местоположения, например, с помощью GPS), геокодирования (привязки адреса объекта на карте к его географическим координатам) и геотегирования (добавления географических данных, например, добавления физического местоположения к фотографии объекта) [2].

Геосоциальные сети в первую очередь предназначены для мобильных устройств. Их самая популярная функция – это регистрация (пользователи сообщают нам о своем текущем местонахождении), за которой обычно следует публикация этой информации в Facebook или Twitter. Пользователи могут входить в систему с помощью текстовых SMS-сообщений или приложений, которые можно найти на всех популярных платформах. Местоположение загружается из оборудования GPS на мобильном устройстве или из сетевого местоположения, предоставленного приложением.

Foursquare была первой социальной сетью на основе геолокации, которая позволяла пользователям сообщать друзьям о своем местонахождении. Также популярны Gowalla и геолокационные сервисы в Facebook и Google.

Waze – это геолокационное приложение, разработанное и внедренное израильской компанией с целью улучшения информирования водителей о дорожных условиях в режиме реального времени. Waze превратился в нечто большее, и сегодня это геосоциальный инструмент, объединяющий 50 миллионов пользователей, которые сообщают друг другу об условиях дорожного движения по всей планете - люди делятся полезной информацией о том, где пробки или где проводятся дорожные работы. Благодаря активному участию пользователей в Waze существуют точные карты в реальном времени.

2. Влияние социальных сетей в чрезвычайных ситуациях: геотегинг и картирование

Помимо своих основных функций, социальные сети играют важную роль в оказании помощи при стихийных бедствиях и в их предотвращении. Они также важны для повышения осведомленности общественности о безопасности уязвимых групп населения во время стихийных бедствий.

Социальные сети используются для отслеживания волонтеров, помощи членам семьи в коммуникации и для распространения важной информации. На сегодняшний день существует тенденция к использованию социальных сетей в аналогичных ситуациях (с использованием сотовых телефонов).

Новые формы социальных сетей включают картографирование и геотеги, то есть инструменты для определения местоположения. Геолокационные сервисы – Ushahidi, OpenStreetMap и Google Map Maker – лишь некоторые из них, которые используются для обозначения кризисов и бедствий [3].

С помощью текстовых сообщений, электронной почты и различных интернет-сервисов эти платформы получают большие объемы данных и рисуют их на карте, что помогает в быстром реагировании на наиболее пострадавшие районы. Каждая платформа имеет свои отличия – Ushahidi собирает информацию от людей, чтобы отобразить информацию о возможной чрезвычайной ситуации, создавая интерактивный

картографический проект. OpenStreetMap использует существующие спутниковые изображения и реальную цифровую карту, которая впоследствии может быть изменена.

Эти службы, вместе с другими социальными сетями, могут предоставить такую информацию, как местонахождение ближайшей больницы или пунктов волонтерства. Они способствуют координации распределения гуманитарной помощи. Цифровые карты помогают волонтерам, предоставляя им информацию о проходимых дорогах, лучших маршрутах в определенные районы и т.д.

Социальные сети можно рассматривать как инструмент, помогающий в самом прогнозе возможного возникновения чрезвычайной ситуации, как средство связи во время и после чрезвычайного происшествия. Интеграция социальных сетей в планы коммуникации в чрезвычайных ситуациях должна включать использование социальных сетей во время тренингов и учений организаций быстрого реагирования, а также других лиц.

Каждая организация, участвующая в реагировании на чрезвычайные ситуации, должна вкладывать время, деньги и силы в кампании в социальных сетях, чтобы систематически распространять соответствующую информацию.

В январе 2010 года на Гаити произошло мощное землетрясение силой 7 баллов по шкале Рихтера. Данные показывают, что службы, отвечающие за реагирование на чрезвычайные ситуации и люди, находящиеся в опасности, использовали социальные сети [4]. Это событие является примером того, как картографирование как цифровой инструмент может быть использовано на этапе восстановления и реабилитации после аварии. OpenStreetMap собирает информацию (спутниковые снимки неотмеченных областей) и создает карты сразу после землетрясения. Точная информация о районах, пострадавших от землетрясения, спасла большое количество людей.

Во время катастрофы, случившейся в Японии 11 марта 2011 года, когда землетрясение силой 9,0 баллов по шкале Рихтера, за которым последовало цунами и ядерная авария, привело к огромным потерям и нанесло неопределимый ущерб, использование социальных сетей сыграло решающую роль. Новости и отчеты распространялись через Facebook, Twitter и YouTube быстрее, чем через любые традиционные СМИ [5]. Социальные сети способствовали повышению осведомленности жертв и оказывали помощь в поиске пропавших без вести с помощью таких средств массовой информации, как созданное Google приложение – Google Person Finder.

Выводы

В статье обсуждается применение и влияние геосоциальных сетей в чрезвычайных ситуациях. Приведенные примеры передовой практики определяют способы, с помощью которых соответствующие организации и органы, ответственные за управление чрезвычайными ситуациями, должны оптимизировать их использование во время или после чрезвычайной ситуации. Из вышеизложенного можно сделать вывод, что важно улучшить использование социальных сетей, чтобы эффективно собирать и распространять точную информацию в режиме реального времени.

В связи с тем, что социальные сети растут с экспоненциальной скоростью, будущие исследования должны быть посвящены их подробному анализу, особенно в отношении служб геолокации в предотвращении стихийных бедствий, поскольку они имеют огромное значение на этапе восстановления и реабилитации после аварии.

Литература

1. Социальные сети//SEO Wiki: [Электронный ресурс]. – URL: <https://wiki.rookee.ru/socialnaya-set/> (дата обращения: 10.04.2022).
2. Александр Птица. Геосоциальные сети: найдутся все!//ITCua: [Электронный ресурс]. – URL: https://itc.ua/articles/chto_takoe_geosocialnaya_set_53305/ (дата обращения: 10.04.2022).
3. Геоинформационная система//Википедия. Свободная энциклопедия: [Электронный ресурс]. – URL: https://ru.wikipedia.org/wiki/Геоинформационная_система (дата обращения: 10.04.2022).
4. Pearson, S., & O'Connell, F. Avoiding Social Media Pitfalls. Politics (Campaigns & Elections), 31(294), 43, (2010).
5. Ad Age, S. Global citizens, marketers rely on social media after Japanese quake. Advertising Age, 82(11), 2-21, (2011).

УДК 004.575.4

РАЗРАБОТКА ГЕОИНФОРМАЦИОННЫХ СИСТЕМ УПРАВЛЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОСТРАНСТВ

DEVELOPMENT OF GEOINFORMATION SYSTEMS FOR MANAGEMENT OF EDUCATIONAL SPACES

¹Галпакова А.Б., ²Аннаева С.Д.,
¹Национальный институт образования,
²Туркменский государственный университет им. Махтумкули,
Ашхабад, Туркменистан

A.B. Galpakova¹, S.D. Annayeva²,
¹National Institute of Education,
²Turkmen State University named after Magtymguly,
Ashgabat, Turkmenistan

e-mail: selbi1294@gmail.com

Аннотация. Для развития и усовершенствования общества на основе национальных и общечеловеческих ценностей, необходимо выявлять и воспитывать всесторонне развитую личность. Следовательно, выполнение данной задачи требует получение знаний населением в современных условиях, с применением развитой цифровой системы образования. В связи с этими условиями, одно из важнейших задач концепции развития цифровой системы образования является внедрение образовательной сети, которая позволяет усовершенствование и обучение новейшими информационными методами содержание науки. Для реализации этой задачи необходимо создать образовательную сеть. В создании этой сети нужно учесть основной объект управления – образовательное пространство. Это пространство является совокупностью использования образовательных программ, услуг, технологий и методов их обучения, также определенных подходов управления. Соответственно, оно усовершенствует подход планирования распределения объектов на базе

геоинформационной системы. Этот подход осуществляет анализ с учетом социально-экономических и природных особенностей региона.

Abstract. For the development and improvement of society on the basis of national and universal values, it is necessary to identify and educate a comprehensively developed personality. Therefore, the implementation of this task requires the acquisition of knowledge by the population in modern conditions, using a developed digital education system. In connection with these conditions, one of the most important tasks of the concept of development of the digital education system is the introduction of an educational network that allows the improvement and training of the content of science with the latest information methods. To implement this task, it is necessary to create an educational network. In creating this network, it is necessary to take into account the main object of management - the educational space. This space is a combination of the use of educational programs, services, technologies and methods of teaching them, as well as certain management approaches. Accordingly, it will improve the approach of planning the distribution of objects based on a geographic information system. This approach carries out the analysis taking into account the socio-economic and natural characteristics of the region.

Ключевые слова: цифровая, образование, геоинформационный, управление, регион, данные.

Keywords: Digital, education, geoinformational, management, region, data.

Целью данной работы является оценка возможностей цифровых систем, обеспечивающих принять оптимальные решения управления в среде образования. [1]

Для принятия таких решений управления цифровым системам необходимы следующие исходные данные:

- просканированная карта изучаемого региона для осуществления цифровизации;
- комплекс характеристик с точными координатами изучаемых объектов;
- ранее существующие данные в растровом или векторном формате;
- автоматически оцифрованные изображения.

Для решения задачи цифрового моделирования образовательного пространства оптимальным способом является использование отсканированного картографического изображения региона и последующего ввода на него оцифрованных координат объекта.

Для большинства отсканированных картографических средств считается удобной величина разрешения 300-400 dpi. Использование более высокой степени разрешения снижает процесс оцифровки и требует больших объемов оперативной памяти. При этом, увеличение показателя никак не сказывается на качестве изображения.

В случае, когда сканирование карты целиком невозможно, можно использовать метод сканирования по частям. Соединение их осуществляется по наблюдательским точкам. Для осуществления правильного соединения необходимо по крайней мере 8 опорных точек (4 вертикальных и 4 продольных). Результаты сканирования сохраняются в разных форматах (TIFF, GIF, JPEG, BMP и т.д.).

Технологическая схема векторизации картографических материалов состоит из нескольких операций: сканирование, оцифровка, создание проекта, ввод символьных данных, сетевой анализ, распределение объектов по слоям, увеличение, дифференциация буферной зоны.

Четкое и быстрое выполнение этих операций требует выбора практических цифровых систем.

В первых шагах разработки цифровой системы управления в определенном пространстве проводится выбор программного обеспечения, далее при их использовании предусматривается решение задачи создания проекта размещения объектов образовательного пространства региона.

Программное обеспечение цифровых систем включает в себя несколько технологических особенностей: векторизация бумажного вида карты, надземных и аэрокосмических топоизображений, ввод и изменение GPS-приёмника информации, восстановление рельефа по стереокартинкам с помощью метода фотограмметрии, построение топологической модели по векторным данным, решение задачи картографической алгебры и т.д. [2].

Программное обеспечение системы представляет собой взаимосвязанные комплексы концентрических покрытий. Ядром комплекса является системное программное обеспечение, в том числе операционная система, система управления базой данных и т.д.

В настоящее время существует многочисленные геоинформационные системы, которые относятся к системам владеющими основными функциями такие как создание карты, его визуализация и анализ. Самые продуктивные из них: ArcGIS, MapInfo, Autodesk Map, GeoMedia, ERDAS Imagine [3].

Используемые объекты в системе классифицируются не по значению, а по типам графических примитивов.

Система имеет собственный встроенный язык программирования и набор компонентов для использования в своих приложениях, что позволяет обработку системных операций с помощью встроенного языка программирования.

MapInfo позволяет им создавать тематические карты различными способами, визуализируя трехмерные поверхности. Геоинформационная система ArcGIS 10.1 (продукт американской компании ESRI) в настоящее время является самой мощной и динамично развивающейся системой векторной геолокации в мире. Система ArcGIS – это набор программ, предназначенных для решения различных задач. Он состоит из следующих комплектов программного обеспечения:

- настольная цифровая геолокация – это программный пакет в ArcGIS Desktop, который включает в себя ArcView, ArcEditor и ArcInfo. Они представляют собой общие ядра приложений, основанные на ArcObjects (комплекс компонентов ActiveX), которые позволяют решать различные проблемы геолокации и создавать собственные модули приложений;

- серверная цифровая геолокация – это блок, предназначенный для управления приложениями ArcGIS Server – корпоративными сетями или объектами цифровой геолокации в Интернете; ArcSDE – это мощный сервер пространственных данных, на котором работает ArcIMS, современный сервер на основе СУБД, позволяющий создавать и распространять карты через Интернет;

- встроенная цифровая геолокация – ArcGIS Engine – это библиотека компонентов и инструментов для цифровых систем, с помощью которой можно расширять возможности программы и создавать новые дополнительные приложения;

- мобильная цифровая геолокация – включающая в себя программу ArcPad, оборудована мобильными компьютерами с функцией GPS. Эта программа позволяет просматривать, обновлять и собирать данные в полевых условиях.

Цифровая геосистема поддерживает большинство моделей информации в современном пространстве, включает в себя векторные топологические, растровые модели, а также получение объектно-ориентированной модели новой базы данных.

Выводы

В результате приведенного выше анализа и оценки программного обеспечения, очевидно, что существуют различные типы инструментов цифровой геолокации. Они различаются в зависимости от своих возможностей и проблем, которые они решают. При выборе конкретной цифровой системы необходимо прежде всего рассмотреть поставленные вопросы и требования к системе. Это связано с тем, что каждая система, несмотря на свою универсальность, решает только определенные типы проблем. Следовательно, геоинформационные системы ArcGIS и MapInfo соответствуют требованиям программных инструментов для создания цифровой геолокации образовательного пространства региона.

Литература

1. Концепция цифрового образования в Туркменистане. Постановление Президента Туркменистана от 1 сентября 2017 г. №340.
2. Вайдер Лаан Дренк. Использование современных геоинформационных систем (ГИС) для анализа риска и оценки чрезвычайных ситуаций, для контроля и преодоления последствий аварий//Матер. Междунар. семин. «Радиоактив, отходы: оценка риска, минимизация образ., перераб. и захоронение», Москва, 13-17 сент., 1993.: Тез. докл. – М., 2014. – С. 40-42.
3. Кревский И.Г. «Геоинформационные системы (ГИС). Основные сферы применения информационных технологий в управлении». Журнал «Информационный бюллетень ГИС-Ассоциации» №6, 2009.

УДК 004.575.4

ВОЗМОЖНОСТИ ВНЕДРЕНИЯ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЦИФРОВОЕ ОБРАЗОВАТЕЛЬНОЕ ПРОСТРАНСТВО

POSSIBILITIES OF INTRODUCING GEOINFORMATION TECHNOLOGIES INTO THE DIGITAL EDUCATIONAL SPACE

¹Аннаев М.С., ²Ымамгулыев Ы.Г.,

¹Национальный институт образования

²Центр технологий Академии наук Туркменистана
Ашхабад, Туркменистан

M.S. Annaev¹, Y.G. Ymamgulyev²

¹National Institute of Education

²Technology Center of the Academy of Sciences of Turkmenistan
Ashgabat, Turkmenistan

e-mail: mammetmbi@gmail.com

Аннотация. В настоящее время задача развития и совершенствования общества на основе национальных и общечеловеческих ценностей, формирование целостной стержневой идентичности требует создания условий для современного образования населения. Для создания этих условий в системе образования и ранее

несоответствующих образовательных учреждений необходимо учитывать ряд преобразований в негосударственных организациях (интернет-кафе, библиотеки, учебные центры, центры дополнительного профессионального образования т. д.). Такие преобразования приводят к разнообразным изменениям функций образовательных учреждений, предоставляющих существующие образовательные услуги, а также к усовершенствованиям используемых в них концепций. Одним из основных таких понятий является понятие «система образования». Описанные выше обоснования привели к расширению значения этого понятия и появлению понятия «образовательное пространство». Одним из важнейших вопросов образовательного пространства является удобное размещение новых субъектов обучения, учитывающих транспортные и информационно-коммуникационные, социально-экономические и природно-географические показатели, необходимые для определенного количества учащихся. Решение такой задачи требует более оперативного развертывания образовательных учреждений в регионе и внедрения новых технологий сбора информации о существующих объектах системы образования.

Abstract. At present, the task of developing and improving society on the basis of national and universal values, the formation of a holistic core identity requires the creation of conditions for modern education of the population. To create these conditions in the education system and previously inappropriate educational institutions, it is necessary to take into account a number of transformations in non-governmental organizations (Internet cafes, libraries, training centers, centers for additional professional education, etc.). Such transformations lead to various changes in the functions of educational institutions that provide existing educational services, as well as improvements in the concepts used in them. One of the main such concepts is the concept of “education system”. The justifications described above led to the expansion of the meaning of this concept and the emergence of the concept of “educational space”.

One of the most important issues of the educational space is the convenient placement of new subjects of education, taking into account the transport and information and communication, socio-economic and natural-geographical indicators necessary for a certain number of students. Solving this problem requires a more rapid deployment of educational institutions in the region and the introduction of new technologies for collecting information about existing objects of the education system.

Ключевые слова: геоинформационный, оцифровка, территория, формат, система, объект, модель.

Keywords: Geoinformation, digitization, territory, format, system, object, model.

Внедрение геоинформационных технологий в среду пространственного размещения объектов системы образования сопряжено с особыми задачами, требующими заранее большого труда: внедрение картографических данных, оцифровка карт, проектирование базы данных и первая заливка. Однако, с другой стороны, заранее построенная модель послужит основой для будущих исследовательских и проектных работ.

Важно создать цифровую модель, обеспечивающую оптимальные решения управления в пространстве знаний, которой посвящена данная работа.

Для создания такой модели необходимы следующие условия:

- сканированная карта исследуемой территории для оцифровки изображений;
- набор точных координат и характеристик исследуемых объектов;
- уже существующие данные в растровом или векторном формате;

– автоматически оцифрованные изображения.

Декартову систему координат удобно использовать при создании цифровой пространственной модели региона, а форма объектов модели осуществляется в пространстве R^3 . В результате интеграция сформированного регионального образовательного пространства в глобальное образовательное пространство с использованием геоинформационных технологий послужит основой для создания модели образовательного пространства в масштабах страны.

Данный результат обеспечит реализацию задачи «создания единой образовательной сети среди образовательных учреждений» в составе основных задач Концепции развития системы цифрового образования в Туркменистане [1].

Таким образом, внедрение геоинформационных технологий в контексте автоматизации планирования географического расположения образовательных учреждений и сервисов для решения вопросов управления образованием находится на соответствующем уровне.

Основная цель работы – визуализировать данные, которые имеют особое значение в пространстве цифрового образования, и предоставить им геоинформационные технологии, которые позволят им решать вопросы распределения в соответствии с требуемым пространством, необходимой для ввода, хранения и обработки информации.

Достижение таких целей предполагает решение следующих вопросов:

- анализ методов, алгоритмов, моделей, позволяющих оценить эффективность управления системой образования на региональном уровне;
- финансовый подход к оценке эффективности управления образованием на региональном уровне;
- разработка цифровой информационной системы, модели и методики для автоматизации определения уровня активности образовательных учреждений региона;
- разработка и внедрение нейросетевых моделей для поддержки принятия решений руководителями образовательных учреждений;
- разработка и внедрение имитационной модели для управления системой образования региона.

Анализ существующих методов планирования систем образования региона позволил выявить следующие принципиальные противоречия [2, 3]:

- взаимосвязь между размещением аспектов образования и потребностями населения в образовательных услугах, другими социально-экономическими и природно-географическими показателями, развитием транспорта и коммуникаций;
- между традиционными методами описания, анализа, регистрации, сбора информации по аспектам образования и возможностями управления образовательными учреждениями с помощью цифровых и коммуникационных технологий в образовательном пространстве;
- среди традиционных подходов к планированию комплексного развертывания региональной системы образования и возможностей использования геоинформационных технологий при планировании регионального развертывания аспекты образовательной среды.

Решение проблемы внедрения геоинформационных технологий в образовательное пространство региона будет более эффективным в следующих условиях:

- ясны требования к структуре и элементам образовательного пространства;
- необходимо проанализировать и определить условия, показатели, определяющие оптимальное размещение образовательных учреждений и услуг;

– разработать методику внедрения геоинформационных технологий в цифровое образовательное пространство региона.

В соответствии с объектом, предметом, целью и идеями, представленными в исследовании, будут решены следующие вопросы:

- уточнение концепции образования в регионе;
- определить влияние природных, социально-экономических, образовательных и организационных показателей, условий на структуру и услуги образовательных учреждений, сформировать их уровни;
- обоснование условий информационного моделирования региональных систем, возможностей геоинформационных технологий в управлении образованием;
- разработка геоинформационной модели пространственного размещения новых аспектов образования и реализация на примере конкретного объекта региона.

Выводы

Предусматриваемая здесь инновационная разработка направлена на разработку и внедрение геоинформационных системы, способной включать в себя показатели образовательного пространства региона: интеграция с базой данных ЛВС – это новинка, позволяющая принимать решения в интерактивном режиме. Лицо, принимающее решение, имеет реальную возможность просмотреть актуальную информацию в любой момент (без знания местоположения образовательного пространства, уровня подготовки электронных ресурсов, процесса обработки данных). Это система поддержки принятия решений в области геоинформационных технологий, которая имеет определенные преимущества в улучшении методического обеспечения учащихся и учителей, в поддержании высоких стандартов образования, в повышении эффективности образовательных учреждений и организаций в области пространственной информации.

Литература

1. Концепция развития системы цифрового образования в Туркменистане.//Постановление Президента Туркменистана от 1 сентября 2017 года.
2. Пахомов П.И., Немтинов В.А. Геоинформационная модель территории для поддержки принятия решений по управлению объектами муниципальных систем.//Вестник ТГТУ. 2009. Том 15. №1. С. 199-207.
3. Плисенко О.А. Цифровое моделирование территорий как основа для вычислительных экспериментов ГИС.//ИнтерКарго/ИнтерГИС 11: Устойчивое развитие территорий: теория ГИС и практический опыт. – Ставрополь: Изд-во СГУ, 2005. С. 48-53.

СИСТЕМНОЕ ПРОГРАММИРОВАНИЕ И ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА

УДК 004.28.23.37

ЗАДАЧА ПРОГНОЗИРОВАНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА НЕФТЕПРОДУКТОВ АТМОСФЕРНО-ВАКУУМНОЙ ПЕРЕГОННОЙ УСТАНОВКИ

THE PROBLEM OF PREDICTING THE QUALITY INDICATORS OF PETROLEUM PRODUCTS IN AN ATMOSPHERIC-VACUUM DISTILLATION PLANT

¹Сильнова С.В., ²Мустафина Д.Б.,

¹ООО ИК «СИБИНТЕК»,

²ФГБОУ ВО «Уфимский государственный авиационный технический университет»,
г. Уфа, Российская Федерация

S.V. Silnova¹, D.B. Mustafina²,

¹Ltd. IC “SIBINTEK”,

²FSBEI HE “Ufa State Aviation Technical University”,
Ufa, Russian Federation

e-mail: dinara-10a@mail.ru

Аннотация. Российская нефтяная промышленность является одним из основных секторов топливно-энергетического комплекса страны, поэтому большое внимание уделяется ее развитию. Энергетическая стратегия России на период до 2030 года предусматривает увеличение добычи нефти до 530 млн тонн и достижение коэффициента извлечения нефти с 0,35 до 0,37. Решение этой задачи возможно только с помощью инновационных технологий, которые, в свою очередь, могут эффективно внедряться только в совокупности с современными системами автоматизации. Основной вопрос экономической стратегии государства – ускорение научно-технического прогресса, что позволит добиться значительного повышения производительности труда. Одним из важнейших направлений, повышающих производительность труда, является автоматизация производства. Автоматизация для нефтегазовых предприятий имеет особое значение, так как она является одной из ведущих отраслей и в значительной степени определяет экономическое развитие страны. Одной из важнейших задач, стоящих на сегодняшний день перед нефтяными предприятиями, является управление качеством нефти. Для того, чтобы обеспечить требуемое качество, необходимо поддерживать основные технологические параметры нефти, такие как температура, давление, расход и т.д., значение которых поступают от датчиков, находящихся на технологическом объекте. Задача поддержания требуемых параметров нефти осложняется высокой чувствительностью к нарушению заданного режима, большим числом точек контроля и управления, наличием примесей в сырье, систематическими погрешностями в показаниях расходомеров, плотномеров, и других контрольно-измерительных средств, из-за отложения в трубопроводах твердых продуктов реакции. Актуальным решением для преодоления указанных проблем является разработка нейронной сети для прогнозирования показателей качества атмосферно-вакуумной

перегонной установки, позволяющей снизить корректирующие действия и вмешательства со стороны оператора установки при ведении технологического процесса.

Abstract. The Russian oil industry is one of the main sectors of the country's fuel and energy complex, so much attention is paid to its development. The energy strategy of Russia for the period up to 2030 provides for an increase in oil production to 530 million tons and the achievement of an oil recovery factor from 0.35 to 0.37. The solution to this problem is possible only with the help of innovative technologies, which, in turn, can be effectively implemented only in conjunction with modern automation systems. The main issue of the state's economic strategy is the acceleration of scientific and technological progress, which will make it possible to achieve a significant increase in labor productivity. One of the most important areas that increase labor productivity is production automation. Automation for oil and gas enterprises is of particular importance, as it is one of the leading industries and largely determines the economic development of the country. One of the most important tasks facing oil companies today is oil quality management. In order to ensure the required quality, it is necessary to maintain the main technological parameters of oil, such as temperature, pressure, flow rate, etc., the values of which come from sensors located at the technological facility. The task of maintaining the required oil parameters is complicated by high sensitivity to violation of the set regime, a large number of monitoring and control points, the presence of impurities in the feedstock, systematic errors in the readings of flow meters, density meters, and other control and measuring equipment, due to the deposition of solid reaction products in pipelines. An urgent solution to overcome these problems is the development of a neural network for predicting the quality indicators of an atmospheric-vacuum distillation plant, which makes it possible to reduce corrective actions and interventions on the part of the plant operator during the process.

Ключевые слова: Показатели качества, нефть, бензин, прогнозирование, переработка нефти.

Keywords: Quality indicators, oil, gasoline, prediction, oil refining.

Цель переработки заключается в преобразовании сырой нефти и природного газа в полезную товарную продукцию.

Технология первичной переработки нефти основана на разделении нефти методом ректификации на узкие нефтяные фракции и определяется направлениями использования фракций, выделяемых на установках атмосферно-вакуумной перегонки [3, С. 13-15].

Обезвоженная и обессоленная нефть с установки обезвоживания и обессоливания после подогрева до температуры 195-205 °С за счет рекуперации тепла материальных потоков, отходящих с установки, поступает на разделение в колонну частичного отбензинивания сырья К-1.

Назначение К-1 – отбор из нефти легкого бензина и основной части растворенных газов для нормализации количества бензиновых углеводородов в основной колонне К-2 и стабилизации режима её работы при возможных колебаниях состава сырья

Дистиллятные пары из К-1 конденсируются в аппаратах водяного охлаждения (АВО) и разделяются в сепараторе С-1 на жидкую и газовую фазы. Часть жидкой фазы возвращается в К-1 в качестве флегмы, а балансовый избыток (фракция легкого бензина) отводится с установки.

Свойства товарных нефти и нефтепродуктов оценивают с помощью стандартных лабораторных методов испытаний для контроля качества и проверки соответствия требованиям спецификаций [2, С. 58-59].

Например, по плотности керосина можно судить о его составе. Высокая плотность говорит о высоком содержании ароматических углеводородов, низкая – о преобладании парафиновых углеводородов. Температура начала кипения стабильного бензина является важным показателем качества, так как позволяет отслеживать потери бензина в процессе хранения и транспортировки [4, С. 13-15].

Мониторинг текущих производственных ситуаций – это один из лучших методов управления производством. Данный метод позволяет отслеживать качество нефтепродуктов на начальных стадиях процесса нефтепереработки. Текущее состояние процесса нефтепереработки можно отследить по таким параметрам, как температура, давление, расход и т.д., значения которых поступают от датчиков, находящихся на технологической установке. В настоящее время при оценке качества нефтепродуктов используются лабораторные анализы, при этом имеет место значительный лаг между моментом, когда были отобраны пробы для выполнения анализов и местом принятия управленческого решения, так как проведение многих анализов занимает достаточно большое время. То есть лабораторные анализы отражают архивное значение некоторого показателя качества, а с учетом непостоянства технологического режима (всегда присутствуют некоторые колебания по давлению и температуре) анализируемый показатель уже может измениться. Лабораторные анализы ввиду различных факторов, таких как стоимость проведения, квалификация персонала, сложность отбора пробы с установки, проводятся в соответствии с определенным графиком и два последовательно проводимых анализа могут быть выполнены с разницей по времени менее шести часов [5, С. 121-124]. Поэтому управлять режимом работы установки, опираясь на данные лабораторных анализов, недостаточно эффективно.

Лабораторные испытания для поддержания качества нефтепродуктов проводятся три раза в день [1, С. 2-3]. Но поскольку процесс переработки нефти осуществляется непрерывно, соответственно контроль качества нефти и нефтепродуктов должен также проводиться в режиме реального времени. В следствие указанных причин встает задача прогнозирования показателей качества нефтепродуктов. Существуют различные методы прогнозирования: линейная и нелинейная регрессия, метод скользящей средней, нейронные сети.

Линейная регрессия предназначена для получения прогноза непрерывных числовых переменных.

Нелинейная регрессия – частный случай регрессионного анализа, в котором рассматриваемая регрессионная модель есть функция, зависящая от параметров и от одной или нескольких свободных переменных [6, С. 58-59].

Сглаживание методом скользящей средней заключается в том, что индивидуальные, случайно варьирующие уровни ряда, заменяют средними, типическими уровнями, при расчете которых происходит погашение случайных колебаний [8, С. 178-179].

Искусственные нейронные сети представляют собой математические модели и их аппаратно-программные реализации, которые основаны на принципе организации функционирования биологических нейронных сетей нервных клеток живых организмов. С точки зрения структурного построения искусственная нейронная сеть представляет собой систему взаимодействующих искусственных нейронов [7, С. 153-155].

На сегодняшний день для прогнозирования показателей качества нефтепродуктов атмосферно-вакуумной перегонной установки используется метод линейной регрессии.

Коэффициент детерминации построенной модели составил 0.07, который свидетельствует о неудовлетворительном результате.

В связи с этим предлагается строить искусственную нейронную сеть для прогнозирования показателей качества нефтепродуктов, а также использовать метод скользящего окна, позволяющий сформировать из членов временного ряда набор данных, который будет служить обучающим множеством для построения модели.

Правильное формирование обучающей выборки является ключевым фактором при решении задач машинного обучения. Обучающая выборка сформирована из данных, собранных за 2019-2022 год, из нее были удалены данные за периоды простоя атмосферно-вакуумной перегонной установки. Ширина окна для скользящего среднего будет равна трем. Временной горизонт прогнозирования планируется производить на один шаг вперед, в дальнейшем при глубоком исследовании – от трех до пяти шагов вперед.

Выводы

Рассмотрена одна из важнейших проблем нефтяных предприятий – ускорение научно-технического прогресса. Решением является автоматизация производства. Одной из задач автоматизации является управление качеством нефти. Для того, чтобы обеспечить требуемое качество, необходимо поддерживать плановые значения основных технологических параметров нефтепродуктов. Актуальным решением для преодоления указанной проблемы является прогнозирование показателей качества нефтепродуктов атмосферно-вакуумной перегонной установки. Для решения задачи прогнозирования предложено использовать искусственную нейронную сеть, что позволит осуществить управление технологическим процессом в режиме реального времени и уменьшить активное вмешательство со стороны оператора установки.

Литература

1. ГОСТ 2517-2021. Нефть и нефтепродукты. Методы отбора проб. Введ. 2014-03-01. М.: Изд-во стандартов. 50 с.
2. Расчет технологических установок системы сбора и подготовки скважинной продукции: учеб. пособие//С.А. Леонтьев, Р.М. Галикеев, О.В. Фоминых и др. – Тюмень: ТюмГНГУ, 2010 – 116 с.
3. Сбор, транспорт и хранение нефти, нефтепродуктов и газа: учеб. пособие//Н.Ю. Башкирцева, Р.Р. Рахматуллин, Р.Р. Мингазов, и др. – Казань: КНИТУ, 2016 – 132 с.
4. Технология и установки переработки нефти и газа. Свойства нефти и нефтепродуктов: учеб. пособие//Н.Н. Агибалова – СПб: Лань, 2020. – 124 с.
5. Мустафин Ф.М. Резервуары для нефти и нефтепродуктов: Учебник для вузов. – СПб.: Изд-во Недра, 2010. – 480 с.
6. Древис Ю.Г. Технические и программные средства систем реального времени: Учебник для вузов – 2-е изд. – М.: Изд-во Лаборатория знаний, 2016. – 337 с.
7. Автоматизация технологических процессов и производств в нефтегазовой отрасли: учеб. пособие//С.В. Еремеев – СПб.: Лань, 2018 – 136 с.
8. Искусственные нейронные сети учеб. пособие//В.С. Ростовцев – Санкт-Петербург: Лань, 2021 – 216 с.

СИСТЕМЫ И МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ

UDC 004.05 6.53

ALGORITHMS OF ENCRYPTION ON THE BASIS OF LOCAL STANDARDS

АЛГОРИТМЫ ШИФРОВАНИЯ НА ОСНОВЕ МЕСТНЫХ СТАНДАРТОВ

N.N. Ochilov,
State Testing Center
under the Cabinet of Ministers of the Republic of Uzbekistan,
Tashkent, Uzbekistan

Очилов Н.Н.,
Государственный центр тестирования
при Кабинете Министров Республики Узбекистан,
г. Ташкент, Узбекистан,

e-mail: nizom.ochilov91@gmail.com

Abstract. The article discusses the priorities of the international community to ensure information security. Cooperation between states in this area is still developing, and in the Republic of Uzbekistan, special attention is paid to the protection of state secrets and confidential information. In the Republic of Uzbekistan, there is a encryption/decryption algorithm based on the O'zDSt 1105:2009 standard, which describes a block cipher algorithm. This article presents the theoretical results obtained on the basis of the O'zDSt 1105:2009 standard. In the cryptographic standard O'zDSt 1105:2009, the exchange of a table consists of replacing 256 values, and according to a given formula, a bit is formed on the arguments d , L , R , depending on the extended key k_{se} [2]. During the analysis, the initialization time of the system in encryption modes was also studied. In most cases, the initial time deviations of the encryption depending on the selected mode are not significant. In CBC and CFB modes, the initialization time increases.

Аннотация. В статье рассматриваются приоритеты международного сообщества по обеспечению информационной безопасности. Сотрудничество между государствами в этой сфере все еще развивается, и в Республике Узбекистан особое внимание уделяется защите государственной тайны и конфиденциальной информации. В Республике Узбекистан существует алгоритм дешифрования/дешифрования на основе стандарта O'zDSt 1105:2009, который описывает алгоритм блочного шифрования. В статье представлены теоретические результаты, полученные на основе стандарта O'zDSt 1105:2009. В криптографическом стандарте O'zDSt 1105:2009 обмен таблицей состоит из замены 256 значений, и по заданной формуле формируется бит на аргументах d , L , R в зависимости от расширенного ключа k_{se} [2]. В ходе анализа также изучалось время инициализации системы в режимах шифрования. В большинстве случаев начальные временные отклонения шифрования в зависимости от выбранного режима несущественны. В режимах CBC и CFB время инициализации увеличивается.

Keywords: encryption, decryption, standard, parallel computing, linear, differential, round, key length.

Ключевые слова: шифрование, дешифрование, стандарт, параллельные вычисления, линейный, дифференциальный, раунд, длина ключа.

Introduction

The operation of cryptographic functions is performed in the process of functions that determine the minimum latency of responses to queries. The amount of RAM required to perform such actions must be reserved in advance by the system administrator. Because the computer has a limited amount of RAM, libraries cannot be used anywhere.

Modern operating systems use cryptographic methods of information protection everywhere. A special module is created to unify cryptographic functions into operating systems. This module includes various cryptographic functions.

Material and methods

In the cryptographic standard O'zDSt 1105:2009, the table exchange consists of the replacement of 256 values, and according to the given formula the bit on the arguments d , L , R is formed depending on the extended key k_{se}

$$\{[(i + L) \bmod 256] + 1\}^d \bmod 256 \quad (1)$$

where d is the elevation with the R parameter. The total number of replacement tables generated on the basis of different values of these parameters is 4,161,600. Therefore, our main goal is to analyze the method of automatic checking the quality of switching tables used in the encryption algorithm of the standard O'zDSt 1105:2009 of the Republic of Uzbekistan.

The algorithm used in the cryptographic standard O'zDSt 1105:2009 is relatively new because it is relatively new. The analysis of the cryptographic standard O'zDSt 1105:2009 uses a parallel algorithm that allows you to automatically evaluate its replacement tables.

Allows you to analyze arbitrary replacement tables based on the created algorithm. For example, it will be possible to improve this algorithm when checking the stability of all 4-bit switch tables or specific function classes. Since the S-block is the "heart" of the entire cryptosystem, the results of research on it are of great practical importance [1].

The main object of our analysis is an exchange table consisting of the exchange of natural numbers. A lookup table or substitutionbox is a bijective transformation.

$$F: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$$

where \mathbb{F}_2 is a secondary limited area.

F can also be calculated as a system of logical functions ($f: \{0, 1\}^n \rightarrow \{0, 1\}^n$):

$$F(x_1, x_2, \dots, x_n) = (f_1, f_2, \dots, f_n)$$

$$\begin{cases} f_1 = f_1(x_1, x_2, \dots, x_n) \\ f_2 = f_2(x_1, x_2, \dots, x_n) \\ \dots \\ f_n = f_n(x_1, x_2, \dots, x_n) \end{cases} \quad (2)$$

Therefore, the i -function is called F_i , the i -component of the vector. We express $x = (x_1, x_2, \dots, x_n)$, x_i .

We define the binary vector 1 in position i as $e_i = (0, 0, \dots, 1, 0, \dots, 0)$.

$$\Sigma_i = \{(x', x'') : x', x'' \in X, x'_i + x''_i = e_i\}, i = \overline{1, n}, \Sigma = \bigcup_{i=1}^n \Sigma_i \quad (3)$$

get If n changes one of the incoming bits, then each outgoing bit varies with probability $p = 1/2$, then F satisfies the basic rigid criterion. If F meets the basic strict criterion $P(v_i^j = 1) = \frac{1}{2}, \forall j = \overline{1, n}$

$$v_i = (x', x'') = F(x') + F(x''), (x', x'') \in \Sigma_i \quad (4)$$

Therefore, parallel algorithms with high dimension and accuracy based on parallel computing technology were created through OpenMPI (message interface library). The algorithm automatically adapts to the computing resources given to it – the number of computational processes - without changing the code. In addition, high accuracy ensures that none of the computers in the cluster running the algorithm run idle, even if they have different parameters [1, 2].

There are at least two calculation processes, and they are all numbered (from 0). One of the processes is called the server process (from 0) or simply the server cluster, the rest are called computing or working processes. Server performance algorithms and workflows differ. The server devices are responsible for breaking down the initial task into smaller classes of a certain size (set in the compilation of the program). The algorithm is presented in block diagram 1 (Picture 1).

In order not to lose all data during a power outage or power outage while the program is running, data of a certain frequency is written to temporary files using functions in the log file.

Results

The total number of replacement tables generated on the basis of the different values above was 4,161,600, for each of which the automatic quality evaluation $eval()$ function was worked out using linear and differential method, and the attacks were organized.

Table 1 – The difference between encryption algorithms

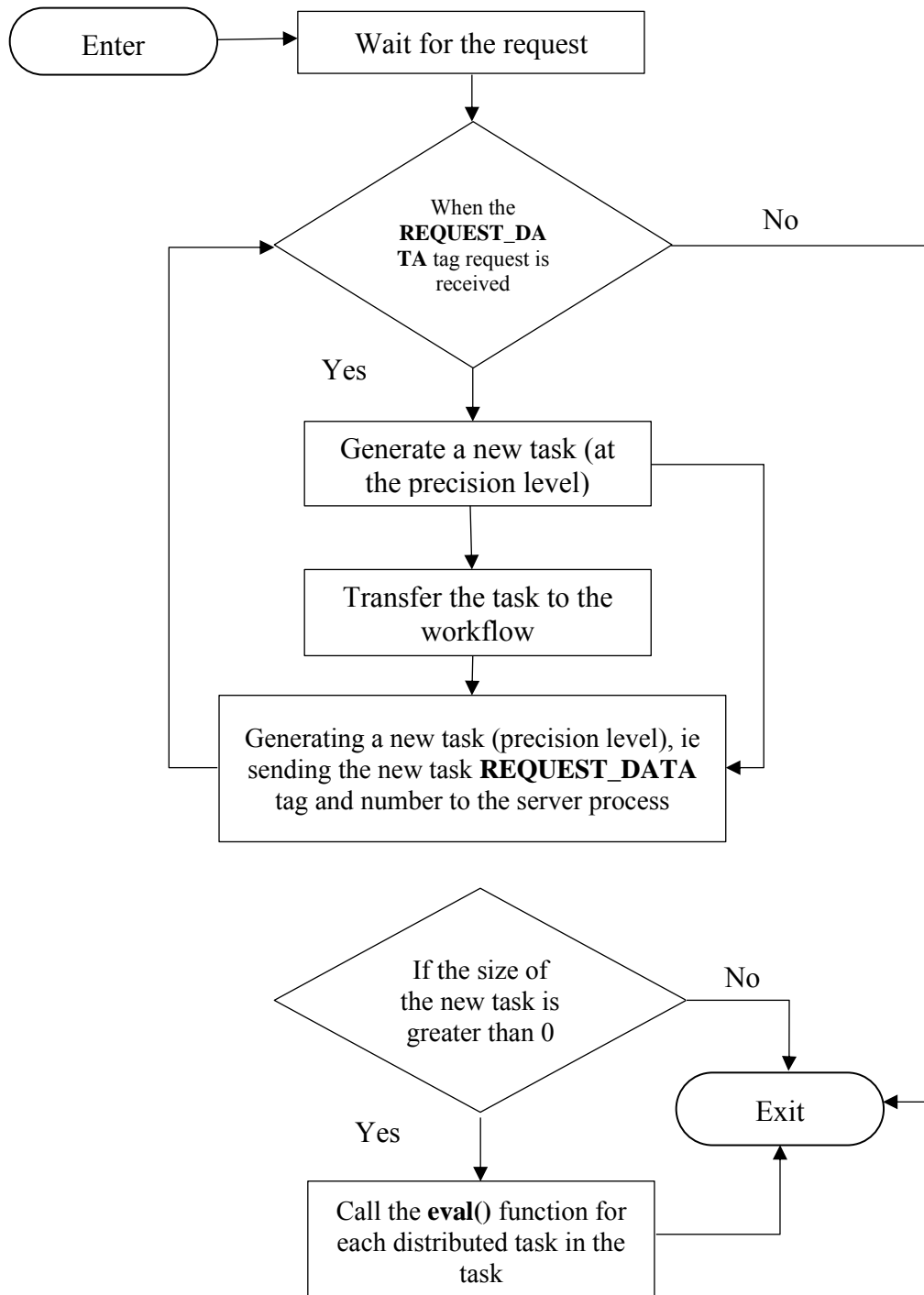
| Algorithm name | Key length, bits | Block length, bits | The number of cycles in the encryption |
|------------------|------------------|--------------------|--|
| Rijndael | 256 | 128 | 14 |
| GOST 28147-89 | 256 | 64 | 32 |
| O'zDSt 1105:2009 | 256 (512) | 256 | 8 |

If tables are identified during the linear analysis process, then the independence criteria for such tables are not met. Based on the results obtained, the replacement tables allow to confirm that they are resistant to linear and differential analysis, which is sufficient for more than 4 rounds.

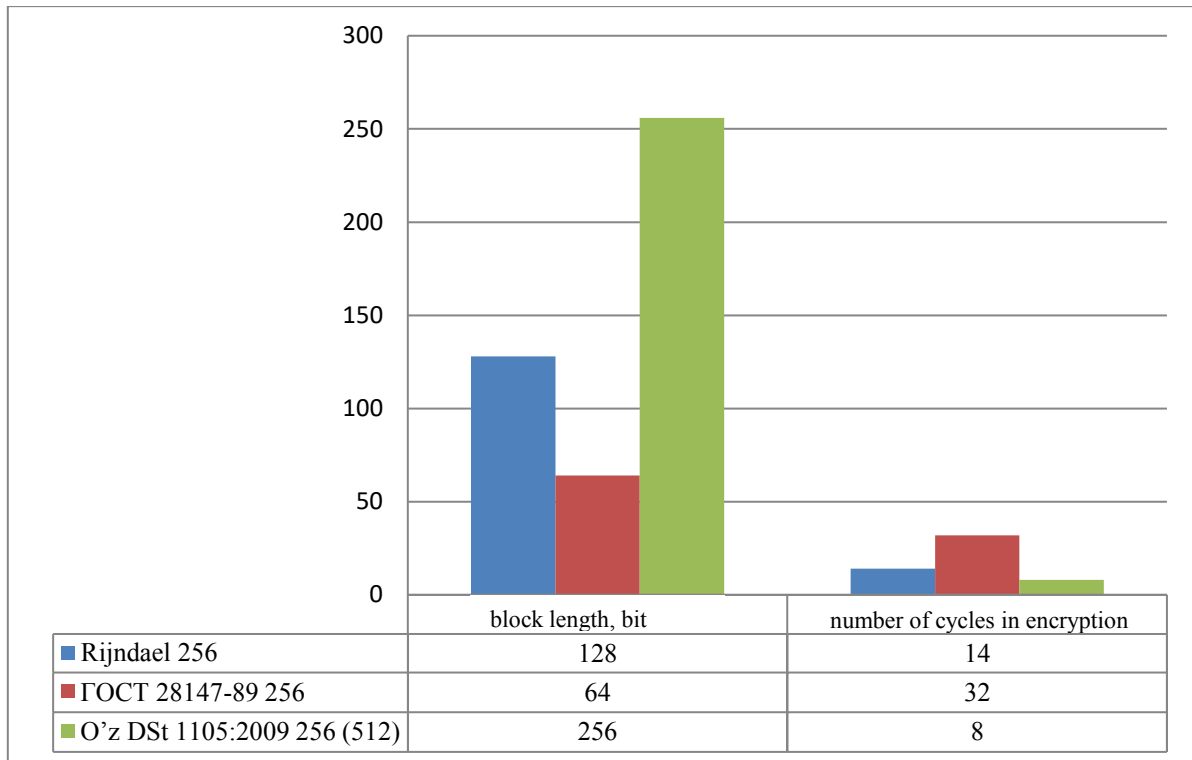
O'zDSt1105:2009 encryption algorithm and the difference between GOST 28147-89 encryption algorithms in Table 1 and other foreign countries (Picture 2).

It is known that in order to create and use high-quality encryption algorithms, programmers need sufficient knowledge and experience in this field. It is recommended to have an idea about optimizing their algorithms when using ready-made cryptographic libraries and testing them in practice. Pre-calculating the values of some functions during the initial

calculation process and storing them in the form of a search table allows the data in it to be exchanged in existing tables, which leads to optimization of processor time. High-level programming languages do not allow to perform some of the operations required to create cryptographic algorithms efficiently. In the Assembler language, the program code provides optimal performance, taking into account the specific features of the processor architecture (Picture 3).



Picture 1. Block diagram.
Algorithm block diagram based on parallel computing technology



Picture 2. Difference diagram of encryption algorithms

However, the use of these methods reduces the portability of the program, complicating modification and maintenance.

Such methods are widely used in open source OpenSSL and LibreSSL cryptographic libraries, including SSL/TLS protocols, symmetric and asymmetric encryption algorithms, and hashing functions.

Modern processors support efficient data set processing functions.

Parallelization of calculations allows different phases of the same algorithms to be performed simultaneously, thus significantly reducing their execution time.

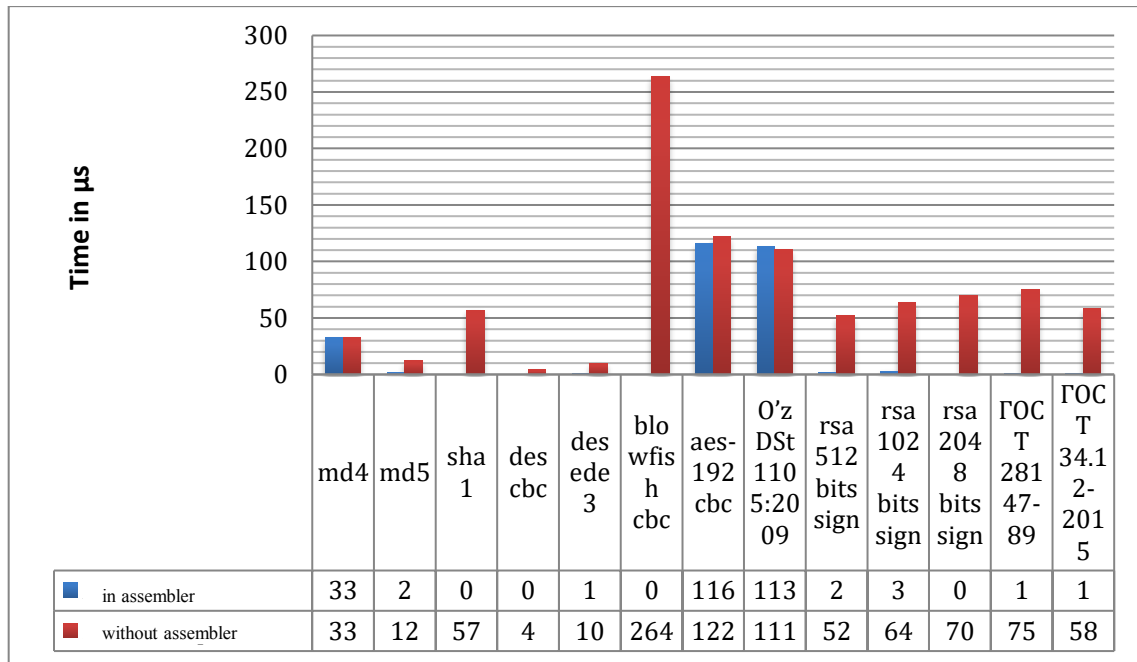
There are many ways to organize parallel computing: performing multi-stream operations on a single processor core, performing parallel operations on multiple processor cores within a single processor, performing parallel operations on all processors available on the network but on different computers.

It is also not possible to effectively parallel all operating modes of block encryption.

The ECB (electronic codebook, simple switching mode) and CTR (counter mode) can parallel simple text blocks for independent processing.

CBC (cipher block chaining mode) and CFB (cipher feedback mode) are only well-parallel decryption procedures.

There is no possibility of parallelization in OFB (output feedback mode), but it is possible to increase the efficiency of encryption and decryption issues by other methods, as shown in Table 2.



Picture 3. Copulation rate in Intel Compiler technology

Table 2 – Modes of operation of different block encryption algorithms

| | ECB | CBC | CFB | OFB | CTR | Cryptographic libraries |
|-------------------------|-----|-----|-----|-----|-----|-------------------------------|
| DES | + | + | + | + | | LibreSSL |
| AES 256 | + | + | + | + | | LibreSSL |
| GOST 28147-89 | + | | + | | + | LibreSSL |
| GOST 34.12-2015 | + | + | + | + | + | Library of GOST algorithm |
| O'zDSt 1105:2009 | | + | | | | Based on additional libraries |

Conclusion

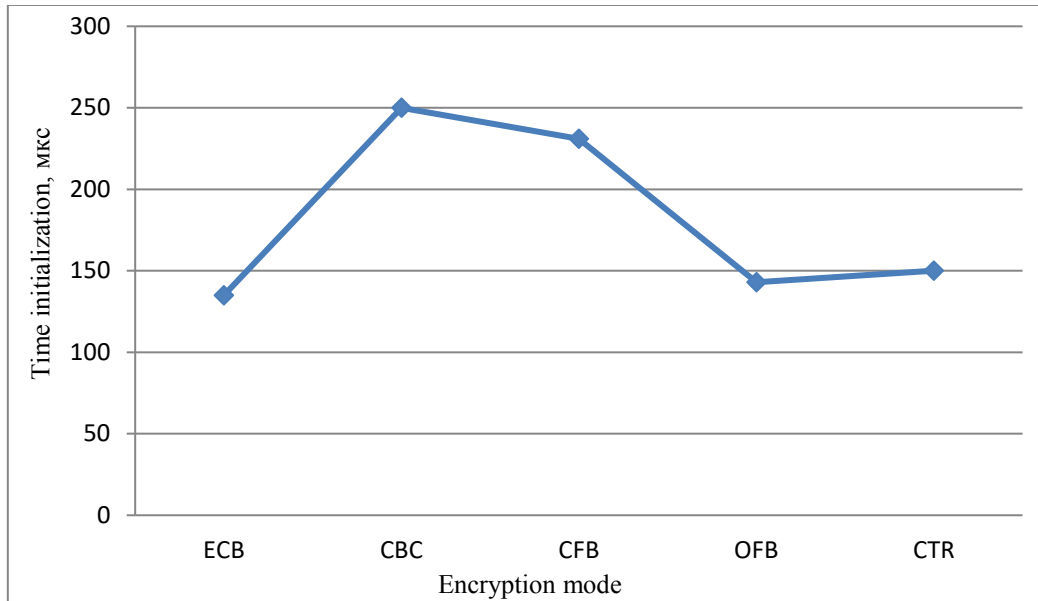
The analysis of the encryption standard O'zDSt 1105:2009 revealed the following:

1. O'zDSt 1105:2009 encryption algorithm uses 2 keys - encryption key and functional key, each of which is a 256-bit sequence. The interaction of these keys is equivalent to the use of a 512-bit encryption key in the encryption algorithm, which in turn prevents the possibility of unauthorized decryption of data;

2. If high-level security elements are used, the functional key is changed in each session;

3. O'zDSt 1105:2009 encryption standard has been confirmed to be resistant to linear and differential analysis, which requires more than 4 rounds.

4. The following classification of the evaluation of the quality indicators of encryption algorithms is proposed and the analysis of encryption modes using parallelization problems was performed. During the analysis, the initialization time of the system in encryption modes was also studied. In most cases, the initial time deviations of the encryption depending on the selected mode are not significant. In CBC and CFB modes, the initialization time increases. In CBC mode, the cryptographic tolerance of the encryption algorithm increases (Picture 4). Parallelization in encryption of the highest index in CBC mode was performed at 0.00025 sec.



Picture 4. Encryption modes initialization time

References

1. Qi Luo, *Advancing Computing, Communication, Control and Management*, Springer Science & Business Media, 2009., p. 290.
2. Algorithm for cryptographic transformation GOST 28147-89, IPK Standards Publishing House, Moscow, 1996.
3. Mao, V. *Modern cryptography: theory and practice: trans. from English/B. Mao.* – M.: Publishing House “Williams”, 2005.
4. Stallings, V. *Cryptography and protection of networks: principles and practice: trans. from English/V. Stallings.* – M.: Publishing House “Williams”, 2001.
5. Ferguson, N. *Practical cryptography: trans. from English/N. Ferguson, B. Schneier.* – M.: Publishing House “Williams”, 2005.
6. Shannon, K. *Communication theory in secret systems//Works on information theory and cybernetics: trans. from English/K. Shannon.* – M.: Publishing house of foreign literature, 1963. S. 333-369.
7. Schneier, B. *Applied cryptography: protocols, algorithms, source texts in C: per. from English/B. Schneier.* – M.: Triumph, 2012.
8. GOST 28147–89 “Information processing systems. Cryptographic protection. Algorithm for cryptographic transformation”.
9. GOST R 34.13—2015 “Information technology. Cryptographic information protection. Modes of Operation of Block Ciphers “ISO/IEC 18033-2:2006 Information technology – Security techniques – Encryption algorithms – Part 2: Asymmetric ciphers.
10. FIPS PUB 197. Federal Information Processing Standards Publication. Advanced Encryption Standard (AES). November 26, 2001.
11. FIPS PUB 198-1. Federal Information Processing Standards Publication. The Keyed-Hash Message Authentication Code (HMAC). July 2008.
12. FIPS PUB 202. Federal Information Processing Standards Publication. SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions. August 2015.

УДК 004.056

**ВОЗМОЖНОСТИ REGEX ДЛЯ РЕШЕНИЯ ПРОБЛЕМ
ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ****REGEX CAPABILITIES FOR SOLVING INFORMATION
SECURITY PROBLEMS**

¹Симомян Р.А., ^{1,2}Корнеев Н.В.,
¹РГУ нефти и газа (НИУ) имени И.М. Губкина,
²Финансовый университет при Правительстве Российской Федерации
г. Москва, Российская Федерация

R.A. Simonyan¹, N.V. Korneev^{1,2},
¹Gubkin Russian State University of Oil and Gas
Moscow, Russian Federation
²Financial University under the Government of the Russian Federation
Moscow, Russian Federation

e-mail: robsimonyan2003@gmail.com

Аннотация. Во многих узких кругах известен Regex – формальный язык поиска. Regex осуществляет поиск с помощью строки образца, у которого достаточно сложный синтаксис, из-за чего многие избегают его. В статье мы попытались изучить возможности Regex, и насколько целесообразно его использование с точки зрения решения проблем информационной безопасности. Regex разрешает выполнять массу действий, таких как: проверка введённого текста, поиск и замена в файле, пакетное переименование, поиск файлов, тест на паттерны в строках, взаимодействие с серверами Apache. Именно это и подтолкнуло на мысль, использования данного мощного инструмента, для решения различных проблем с информационной безопасностью. Изучив разные сферы применения Regex, мы пришли к заключению, что из-за сложности написания, его используют очень редко, но, несмотря на это, если правильно использовать гибкость этого инструмента, можно добиться интересных результатов. Приведены примеры метода «умного поискового запроса с помощью Regex», регулярных выражений в части программ «Microsoft defender» под названием «Defender for Cloud Apps», в Data Loss Prevention (DLP), которое обнаруживает и предотвращает утечки данных.

Abstract. In many narrow circles, Regex is known - the formal search language. Regex searches using a pattern string, which has a rather complex syntax that many people avoid. In the article, we tried to study the possibilities of Regex, and how expedient its use is from the point of view of solving information security problems. Regex allows you to perform a lot of actions, such as: checking the entered text, searching and replacing in a file, batch renaming, searching for files, testing for patterns in strings, interacting with Apache servers. This is what prompted the idea of using this powerful tool to solve various problems with information security. After studying the various uses of Regex, we came to the conclusion that due to the complexity of writing, it is used very rarely, but despite this, if you use the flexibility of this tool correctly, you can achieve interesting results. Examples of “Smart Search with Regex”, regular expressions in the “Defender for Cloud Apps” part of Microsoft's defender, in Data Loss Prevention (DLP), which detects and prevents data leaks, are provided.

Ключевые слова: Regex, регулярные выражения, информационная безопасность, сравнение данных, извлечение информации, пользовательский ввод.

Keywords: Regex, regular expressions, information security, data comparison, information extraction, user input.

Regex чаще всего используют программисты сайтов для проверки правильности введенных телефонных номеров и почтовых адресов, системные администраторы чтобы быстро найти и/или заменить что-то в логах, люди, работающие с большими базами данных, и, нередко, пользователи системы Linux. Regex по сути своей формальный язык с возможностью поиска и манипуляций с подстроками в разнообразных файлах. Регулярные выражения работают с помощью символов и метасимволов, которые для обычного пользователя непонятны заковырки. Для Regex существует шпаргалка (рисунок 1), которая призвана облегчить жизнь программиста в этой сфере [1].

| Assertions | Groups and Ranges | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|-------|-----------|--------------------|---------------------|--------------------|----------|--------------------|----------|--------------------|----------------------|-----------|------------|---------------------|----------|------------------|---|--------------------|----------|---------------------|---|--------------------|--------------------------|--------------------|----------------------------|-------------------------------|-----------------------|--|--|------------------|-----------------------------|------|--|-----|------------------------|--------|----------------------|-------|-------|--|--------|-------------------------------|--|-------|---------------------|--|--------|-------------------|--|-------|-------------------------------|--|-------|-------------------------------|--|-------|-------------------|--|----|-----------------------------|--|
| <table border="1"> <thead> <tr> <th>Assertions</th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>?=</td> <td>Lookahead assertion</td> <td></td> </tr> <tr> <td>?!</td> <td>Negative lookahead</td> <td></td> </tr> <tr> <td>?<=</td> <td>Lookbehind assertion</td> <td></td> </tr> <tr> <td>?!< or ?<!</td> <td>Negative lookbehind</td> <td></td> </tr> <tr> <td>?></td> <td>Once-only Subexpression</td> <td></td> </tr> <tr> <td>?()</td> <td>Condition [if then]</td> <td></td> </tr> <tr> <td>?() </td> <td>Condition [if then else]</td> <td></td> </tr> <tr> <td>?#</td> <td>Comment</td> <td></td> </tr> </tbody> </table> | Assertions | | | ?= | Lookahead assertion | | ?! | Negative lookahead | | ?<= | Lookbehind assertion | | ?!< or ?<! | Negative lookbehind | | ?> | Once-only Subexpression | | ?() | Condition [if then] | | ?() | Condition [if then else] | | ?# | Comment | | <table border="1"> <thead> <tr> <th>Groups and Ranges</th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>.</td> <td>Any character except new line (\n)</td> <td></td> </tr> <tr> <td>(a b)</td> <td>a or b</td> <td></td> </tr> <tr> <td>(...)</td> <td>Group</td> <td></td> </tr> <tr> <td>(?...)</td> <td>Passive (non-capturing) group</td> <td></td> </tr> <tr> <td>[abc]</td> <td>Range (a or b or c)</td> <td></td> </tr> <tr> <td>[^abc]</td> <td>Not (a or b or c)</td> <td></td> </tr> <tr> <td>[a-q]</td> <td>Lower case letter from a to q</td> <td></td> </tr> <tr> <td>[A-Q]</td> <td>Upper case letter from A to Q</td> <td></td> </tr> <tr> <td>[0-7]</td> <td>Digit from 0 to 7</td> <td></td> </tr> <tr> <td>\x</td> <td>Group/subpattern number "x"</td> <td></td> </tr> </tbody> </table> | Groups and Ranges | | | . | Any character except new line (\n) | | (a b) | a or b | | (...) | Group | | (?...) | Passive (non-capturing) group | | [abc] | Range (a or b or c) | | [^abc] | Not (a or b or c) | | [a-q] | Lower case letter from a to q | | [A-Q] | Upper case letter from A to Q | | [0-7] | Digit from 0 to 7 | | \x | Group/subpattern number "x" | |
| Assertions | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?= | Lookahead assertion | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?! | Negative lookahead | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?<= | Lookbehind assertion | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?!< or ?<! | Negative lookbehind | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?> | Once-only Subexpression | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?() | Condition [if then] | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?() | Condition [if then else] | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ?# | Comment | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Groups and Ranges | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| . | Any character except new line (\n) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| (a b) | a or b | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| (...) | Group | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| (?...) | Passive (non-capturing) group | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [abc] | Range (a or b or c) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [^abc] | Not (a or b or c) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [a-q] | Lower case letter from a to q | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [A-Q] | Upper case letter from A to Q | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [0-7] | Digit from 0 to 7 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \x | Group/subpattern number "x" | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Character Classes</th> <th></th> </tr> </thead> <tbody> <tr> <td>\c</td> <td>Control character</td> </tr> <tr> <td>\s</td> <td>White space</td> </tr> <tr> <td>\S</td> <td>Not white space</td> </tr> <tr> <td>\d</td> <td>Digit</td> </tr> <tr> <td>\D</td> <td>Not digit</td> </tr> <tr> <td>\w</td> <td>Word</td> </tr> <tr> <td>\W</td> <td>Not word</td> </tr> <tr> <td>\x</td> <td>Hexadecimal digit</td> </tr> <tr> <td>\O</td> <td>Octal digit</td> </tr> </tbody> </table> | Character Classes | | \c | Control character | \s | White space | \S | Not white space | \d | Digit | \D | Not digit | \w | Word | \W | Not word | \x | Hexadecimal digit | \O | Octal digit | <table border="1"> <thead> <tr> <th>Quantifiers</th> <th></th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>*</td> <td>0 or more</td> <td>{3}</td> <td>Exactly 3</td> </tr> <tr> <td>+</td> <td>1 or more</td> <td>{3,}</td> <td>3 or more</td> </tr> <tr> <td>?</td> <td>0 or 1</td> <td>{3,5}</td> <td>3, 4 or 5</td> </tr> </tbody> </table> <p>Add a ? to a quantifier to make it ungreedy.</p> | Quantifiers | | | | * | 0 or more | {3} | Exactly 3 | + | 1 or more | {3,} | 3 or more | ? | 0 or 1 | {3,5} | 3, 4 or 5 | | | | | | | | | | | | | | | | | | | | | | | | |
| Character Classes | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \c | Control character | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \s | White space | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \S | Not white space | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \d | Digit | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \D | Not digit | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \w | Word | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \W | Not word | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \x | Hexadecimal digit | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \O | Octal digit | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Quantifiers | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| * | 0 or more | {3} | Exactly 3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| + | 1 or more | {3,} | 3 or more | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ? | 0 or 1 | {3,5} | 3, 4 or 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>POSIX</th> <th></th> </tr> </thead> <tbody> <tr> <td>[upper:]</td> <td>Upper case letters</td> </tr> <tr> <td>[lower:]</td> <td>Lower case letters</td> </tr> <tr> <td>[alpha:]</td> <td>All letters</td> </tr> <tr> <td>[alnum:]</td> <td>Digits and letters</td> </tr> <tr> <td>[digit:]</td> <td>Digits</td> </tr> <tr> <td>[xdigit:]</td> <td>Hexadecimal digits</td> </tr> <tr> <td>[punct:]</td> <td>Punctuation</td> </tr> <tr> <td>[blank:]</td> <td>Space and tab</td> </tr> <tr> <td>[space:]</td> <td>Blank characters</td> </tr> <tr> <td>[cntrl:]</td> <td>Control characters</td> </tr> <tr> <td>[graph:]</td> <td>Printed characters</td> </tr> <tr> <td>[print:]</td> <td>Printed characters and spaces</td> </tr> <tr> <td>[word:]</td> <td>Digits, letters and underscore</td> </tr> </tbody> </table> | POSIX | | [upper:] | Upper case letters | [lower:] | Lower case letters | [alpha:] | All letters | [alnum:] | Digits and letters | [digit:] | Digits | [xdigit:] | Hexadecimal digits | [punct:] | Punctuation | [blank:] | Space and tab | [space:] | Blank characters | [cntrl:] | Control characters | [graph:] | Printed characters | [print:] | Printed characters and spaces | [word:] | Digits, letters and underscore | <table border="1"> <thead> <tr> <th>Escape Sequences</th> <th></th> </tr> </thead> <tbody> <tr> <td>\</td> <td>Escape following character</td> </tr> <tr> <td>\Q</td> <td>Begin literal sequence</td> </tr> <tr> <td>\E</td> <td>End literal sequence</td> </tr> </tbody> </table> <p>"Escaping" is a way of treating characters which have a special meaning in regular expressions literally, rather than as special characters.</p> | Escape Sequences | | \ | Escape following character | \Q | Begin literal sequence | \E | End literal sequence | | | | | | | | | | | | | | | | | | | | | | | | |
| POSIX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [upper:] | Upper case letters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [lower:] | Lower case letters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [alpha:] | All letters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [alnum:] | Digits and letters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [digit:] | Digits | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [xdigit:] | Hexadecimal digits | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [punct:] | Punctuation | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [blank:] | Space and tab | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [space:] | Blank characters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [cntrl:] | Control characters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [graph:] | Printed characters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [print:] | Printed characters and spaces | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| [word:] | Digits, letters and underscore | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Escape Sequences | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \ | Escape following character | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \Q | Begin literal sequence | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \E | End literal sequence | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Common Metacharacters</th> <th></th> <th></th> <th></th> </tr> </thead> <tbody> <tr> <td>^</td> <td>[</td> <td>.</td> <td>\$</td> </tr> <tr> <td>{</td> <td>*</td> <td>(</td> <td>\</td> </tr> <tr> <td>+</td> <td>)</td> <td> </td> <td>?</td> </tr> <tr> <td><</td> <td>></td> <td></td> <td></td> </tr> </tbody> </table> <p>The escape character is usually \</p> | Common Metacharacters | | | | ^ | [| . | \$ | { | * | (| \ | + |) | | ? | < | > | | | <table border="1"> <thead> <tr> <th>Pattern Modifiers</th> <th></th> </tr> </thead> <tbody> <tr> <td>g</td> <td>Global match</td> </tr> <tr> <td>i *</td> <td>Case-insensitive</td> </tr> <tr> <td>m *</td> <td>Multiple lines</td> </tr> <tr> <td>s *</td> <td>Treat string as single line</td> </tr> <tr> <td>x *</td> <td>Allow comments and whitespace in pattern</td> </tr> <tr> <td>e *</td> <td>Evaluate replacement</td> </tr> <tr> <td>U *</td> <td>Ungreedy pattern</td> </tr> </tbody> </table> <p>* PCRE modifier</p> | Pattern Modifiers | | g | Global match | i * | Case-insensitive | m * | Multiple lines | s * | Treat string as single line | x * | Allow comments and whitespace in pattern | e * | Evaluate replacement | U * | Ungreedy pattern | | | | | | | | | | | | | | | | | | | | | | | | |
| Common Metacharacters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ^ | [| . | \$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| { | * | (| \ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| + |) | | ? | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| < | > | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pattern Modifiers | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| g | Global match | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| i * | Case-insensitive | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| m * | Multiple lines | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| s * | Treat string as single line | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| x * | Allow comments and whitespace in pattern | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e * | Evaluate replacement | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| U * | Ungreedy pattern | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Special Characters</th> <th></th> </tr> </thead> <tbody> <tr> <td>\n</td> <td>New line</td> </tr> <tr> <td>\r</td> <td>Carriage return</td> </tr> <tr> <td>\t</td> <td>Tab</td> </tr> <tr> <td>\v</td> <td>Vertical tab</td> </tr> <tr> <td>\f</td> <td>Form feed</td> </tr> <tr> <td>\xxx</td> <td>Octal character xxx</td> </tr> <tr> <td>\xhh</td> <td>Hex character hh</td> </tr> </tbody> </table> | Special Characters | | \n | New line | \r | Carriage return | \t | Tab | \v | Vertical tab | \f | Form feed | \xxx | Octal character xxx | \xhh | Hex character hh | <table border="1"> <thead> <tr> <th>String Replacement</th> <th></th> </tr> </thead> <tbody> <tr> <td>\$n</td> <td>nth non-passive group</td> </tr> <tr> <td>\$2</td> <td>"xyz" in /^(abc(xyz))\$/</td> </tr> <tr> <td>\$1</td> <td>"xyz" in /^(?:abc)(xyz)\$/</td> </tr> <tr> <td>\$'</td> <td>Before matched string</td> </tr> <tr> <td>\$'</td> <td>After matched string</td> </tr> <tr> <td>\$+</td> <td>Last matched string</td> </tr> <tr> <td>\$&</td> <td>Entire matched string</td> </tr> </tbody> </table> <p>Some regex implementations use \ instead of \$.</p> | String Replacement | | \$n | nth non-passive group | \$2 | "xyz" in /^(abc(xyz))\$/ | \$1 | "xyz" in /^(?:abc)(xyz)\$/ | \$' | Before matched string | \$' | After matched string | \$+ | Last matched string | \$& | Entire matched string | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Special Characters | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \n | New line | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \r | Carriage return | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \t | Tab | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \v | Vertical tab | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \f | Form feed | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \xxx | Octal character xxx | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \xhh | Hex character hh | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| String Replacement | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$n | nth non-passive group | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$2 | "xyz" in /^(abc(xyz))\$/ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$1 | "xyz" in /^(?:abc)(xyz)\$/ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$' | Before matched string | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$' | After matched string | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$+ | Last matched string | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| \$& | Entire matched string | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Рисунок 4. Шпаргалки для Regex

Как показывает практика, даже опытные программисты в этой сфере нуждаются в шпаргалках, и даже с трудом разбирают свой собственный код через два месяца.

После изучения сложности регулярных выражений, стало понятно, что не каждый решится использовать их в сфере информационной безопасности, так как решения проблем такого вида, должны быть понятными другим специалистам для добавления нового функционала или исправления каких-то выявленных багов.

Вместе с этим очень часто, для добавления нового функционала в Regex приходится полностью переписывать выражение.

Тем не менее функционал регулярных выражений в паре с невиданной гибкостью дает возможность без ограничений работать с любыми текстовыми файлами. Самое распространенное место использование Regex – это первичная проверка на фальшивость номера телефона и почтового адреса для регистрации пользователя (рисунок 2), а также соответствие остальных данных шаблону [2].

Регулярные выражения также используются для сложных задач, таких как поиск и замена каких-то файлов или подстрок внутри файла.

```

6
7 string numbers = @"^(+7|8)?[\s-]?(?[\d]{0-9}{2}\)?[\s-]?[0-9]{3}[\s-]?[0-9]{2}[\s-]?[0-9]{2}$";
8 String mails = @"^A(?:[a-z0-9!#$%&'*/=?^_`{|}~]+(?:\.[a-z0-9!#$%&'*/=?^_`{|}~]+)*@(?:[a-z0-9]?(?:[a-z0-9-]*[a-z0-9])?\.)+[a-z0-9]?(?:[a-z0-9-]*[a-z0-9])?\Z";
9 var vvodi = new string[]
10 {
11     "deseb47355@tonaeto.com",
12     "+12345678999",
13     "reifjer123_effwe.com",
14     "robsimonyan2003@gmail.com",
15     "+13435465566",
16     "kbtak@rambler.ru",
17     "",
18     "89998881658",
19     "899988816-58",
20     "+89998881658",
21     "+7 (999) 808-16-58"
22 };
23 Console.WriteLine("Russian phone numbers");
24 for(int i = 0; i < vvodi.Length; i++)
25 {
26     if (Regex.IsMatch(vvodi[i], numbers, RegexOptions.IgnoreCase))
27     {
28         Console.WriteLine(vvodi[i]);
29     }
30 }
31 Console.WriteLine("");
32 Console.WriteLine("Good emails");
33 for(int i = 0; i < vvodi.Length; i++)
34 {
35     if (Regex.IsMatch(vvodi[i], mails, RegexOptions.IgnoreCase))
36     {
37         Console.WriteLine(vvodi[i]);
38     }
39 }
40 }

```

```

> dotnet run
Russian phone numbers
89998881658
899988816-58
+7 (999) 808-16-58

Good emails
deseb47355@tonaeto.com
robsimonyan2003@gmail.com
kbtak@rambler.ru
>

```

Рисунок 5. Пример проверки

Возможности регулярных выражений не ограничиваются этим, в умелых и опытных руках они применяются в различных сферах. Один из таких примеров: используя возможности ключевого поиска (keyword-based) и регулярных выражений, нигерийские студенты представили метод «умного поискового запроса с помощью Regex». Сопоставив регулярные выражения с образцом, они получили очень полезный инструмент для расширения методологий поиска. Полученный ими инструмент, после дальнейшей доработки, облегчит работу даже неопытного пользователя, которому не будут мешать нерелевантные результаты поиска [3].

При изучении регулярных выражений, как инструмента решения проблем информационной безопасности, мы наткнулись на достаточно интересные возможности и сферы применения.

Как оказалось, сегодня Regex используется в антивирусной программе компании Microsoft – «Microsoft Defender», которую недавно переименовали в «Defender for Cloud Apps».

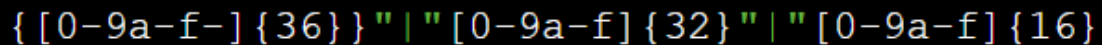
Политика проверки содержимого подприложения использует Regex, для проверки соответствий паттернов [4]. В информационной безопасности важную роль занимает такое решение, как DLP или Data Loss Prevention, которое обнаруживает и предотвращает утечки данных.

DLP работает по методу блокировки извлечения конфиденциальных данных, его используют для обеспечения внутренней безопасности и соблюдения нормативных требований.

Системы DLP работают очень интересным способом: они защищают данные предприятий, сначала выявляя конфиденциальную информацию, а затем используя глубокий анализ контента, обнаруживают и предотвращают потенциальные утечки данных [5]. В таком анализе контента используются разные методы, такие как соответствие ключевых слов, и конечно же регулярные выражения, для распознавания конфиденциального контента. В результате предприятия могут выявлять, отслеживать и автоматически предотвращать кражу или раскрытие защищенных данных [6].

Тем не менее не стоит забывать и об обратной стороне. Regex всего лишь инструмент и как он будет использоваться зависит только от специалиста.

Изучая использование регулярных выражений, мы наткнулись на интересную ситуацию с компанией «Solar Winds». В начале 2020 года хакеры тайно взломали системы SolarWind в Техасе и добавили вредоносный код в систему программного обеспечения компании. Система под названием «Орион» широко используется компаниями для управления ИТ-ресурсами. Согласно документам SEC, у Solarwinds 33 000 клиентов, которые используют Orion [7]. В данном случае, регулярные выражения очень сильно помогли хакерам. Они извлекали команды из тел HTTP-ответов путём поиска hex-строк с помощью регулярного выражения (рисунок 3).



```
{[0-9a-f-]{36}}"|" [0-9a-f]{32}"|" [0-9a-f]{16}"
```

Рисунок 6. Регулярное выражение, использованное для взлома

Код создал бэкдор к информационным системам клиентов, который затем использовали хакеры для установки еще большего количества вредоносного ПО, которое помогало им шпионить за компаниями и организациями. Стоит заметить, что «Backdoor» оставался и работал в системе незаметно на протяжении 7 месяцев [8].

Еще один пример использования Regex в паре с API (предоставленного разработчиками Facebook – в России признана экстремистской и запрещена). Полученный инструмент собирает личные данные в открытом доступе и представляет их в виде временной шкалы или ленты. Безусловно, в руках злоумышленника, «Парсинг в социальных сетях с использованием API разработчиков социальных сетей и регулярных выражений» слишком серьезное и опасное оружие.

Если говорить кратко о работе этого метода, все начинается с анализа потока информации в социальных сетях, с того, как пользователи ищут информацию через социальные сети и как информация представляется пользователю.

Метод извлечения информации заключается в использовании веб-скрапинга, который реализуется через Facebook (в России признана экстремистской и запрещена) Developers API.

Извлеченная информация будет сопоставлена с предпочтениями пользователя с помощью Regex (или регулярного выражения) (рисунок 4), которое является языковой конструкцией и которую можно использовать для сопоставления текста с использованием некоторых шаблонов.

```

string revent = "promotion|event|seminar|concert";
for (int i = 0; i < dpKey.Length; i++)
{
    Match mobjects = Regex.Match(dpKey[i], revent, RegexOptions.IgnoreCase);
    if(mobjects.Success)
    {
        objects.Add("true");
    }
    else
    {
        objects.Add("false");
    }
}

```

Рисунок 7. Сопоставление информации с предпочтениями пользователя с помощью Regex

Этап разработки алгоритма будет включать дизайн фрагмента кода, который можно использовать для парсинга веб-страниц. Это исследование может быть преобразовано в приложение. Веб-приложение, к которому можно получить доступ из любого места и с любых устройств. И для дальнейших исследований алгоритм парсинга веб-страниц может быть разработан с автоматическим или запланированным парсингом, чтобы бизнес или приложение работали автоматически [9].

Выводы

Исследовав Regex для сферы информационной безопасности, мы пришли к следующим выводам:

1. Регулярные выражения могут использоваться в любой сфере, где есть работа с строками и подстроками. Но наиболее предпочтительным считаем сферу информационной безопасности.
2. Регулярные выражения в этой сфере занимают очень неоднозначную позицию, но вместе с тем открывают очень много возможностей, как для защиты, так и для атаки.
3. При выполнении серьезного проекта мало кто применяет Regex (из-за сложности синтаксиса), но, как показал наш анализ, люди, которые все-таки решают использовать Regex получают очень интересные и неповторимые инструменты.

Литература

1. Regular Expressions Cheat Sheet by DaveChild // Cheatography.com. URL: <https://cheatography.com/davechild/cheat-sheets/regular-expressions/> (дата обращения: 28.03.2022).
2. 4 варианта практического использования регулярных выражений//techrocks.ru. URL: <https://techrocks.ru/2019/08/21/4-practical-use-cases-for-regular-expressions/> (дата обращения: 28.03.2022).
3. Developing Smart Web-Search Using RegEx//arxiv.org. URL: <https://arxiv.org/ftp/arxiv/papers/2110/2110.04767.pdf> (дата обращения: 28.03.2022).
4. Working with the RegEx engine (Microsoft) // kbs.bestantivirus.co.uk. URL: <https://kbs.bestantivirus.co.uk/working-with-the-regex-engine-microsoft/> (дата обращения: 28.03.2022).
5. Информационная безопасность или зачем нужна DLP // bnkomi.ru. URL: <https://www.bnkomi.ru/data/relize/133632/> (дата обращения: 28.03.2022).

6. Десять вопросов вендору dlp-системы//infowatch.ru. URL: https://www.infowatch.ru/sites/default/files/analytics/files/10_questions_DLP.pdf (дата обращения: 28.03.2022).

7. The US is readying sanctions against Russia over the SolarWinds cyber attack. Here's a simple explanation of how the massive hack happened and why it's such a big deal//businessinsider.com. URL: <https://www.businessinsider.com/solarwinds-hack-explained-government-agencies-cyber-security-2020-12#:~:text=In%20early%202020%2C%20hackers%20secretly,Orion%2C%20according%20to%20SEC%20documents> (дата обращения: 28.03.2022).

8. Хакеры SolarWinds размазали свои байты в HTTP-трафике через регулярные выражения//pvsm.ru. URL: <https://www.pvsm.ru/informatsionnaya-bezopasnost/359926> (дата обращения: 28.03.2022).

9. Social Media Web Scraping using Social Media Developers API and Regex//sciencedirect.com. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919311561> (дата обращения: 28.03.2022).