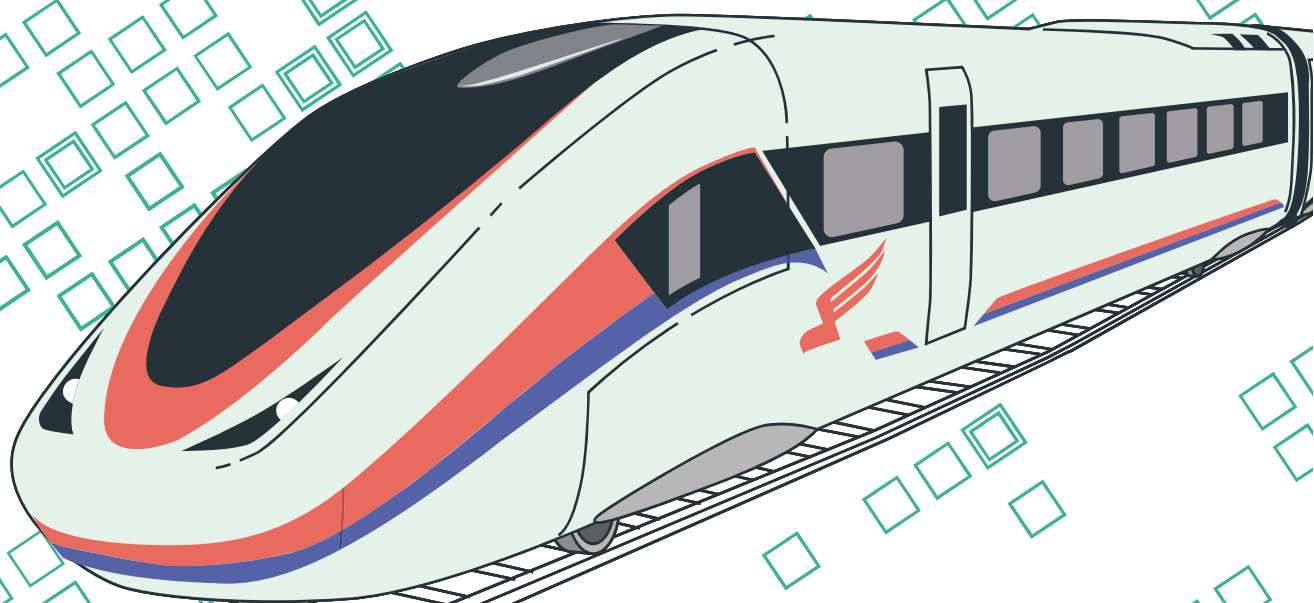


Интеллектуальные технологии на транспорте

Intellectual Technologies
on Transport



Выпуск 3
2023

**Интеллектуальные технологии на транспорте
(сетевой электронный научный журнал)
Выпуск 3 (35), 2023**

ISSN 2413-2527

Сетевой электронный научный журнал, свободно распространяемый через Интернет.
Публикуются статьи на русском и английском языках с результатами исследований
и практических достижений в области интеллектуальных технологий
и сопутствующих им научных исследований.

Журнал основан в 2015 году.

Учредитель и издатель

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Петербургский государственный университет путей сообщения Императора Александра I» (ФГБОУ ВО ПГУПС)

Главный редактор

Хомоненко А. Д., д.т.н., проф., С.-Петербург, РФ

Сопредседатели редакционного совета

Валинский О. С., к.т.н., ректор ПГУПС, С.-Петербург, РФ
Чаркин Е. И., зам. ген. директора по ИТ ОАО «РЖД», Москва, РФ

Редакционный совет

Ададуров С. Е., д.т.н., проф., ВНИИЖТ, Москва, РФ
Дудин А. Н., д.ф.-м.н., проф., БГУ, Минск, Беларусь
Корниенко А. А., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Макаренко С. И., д.т.н., доц., ПАО «Интелтех»,
С.-Петербург, РФ

Меркурьев Ю. А., Dr. Habil., проф., член Латвийской АН,
РТУ, Рига, Латвия
Титова Т. С., д.т.н., проф., первый проректор ПГУПС,
С.-Петербург, РФ
Юсупов Р. М., д.т.н., проф., чл.-корр. РАН, СПб ФИЦ РАН,
С.-Петербург, РФ

Редакционная коллегия

Божко Л. М., д.э.н., доц., ПГУПС, С.-Петербург, РФ –
заместитель главного редактора
Баталов Д. И., к.т.н., ПГУПС, С.-Петербург, РФ –
научный редактор
Александрова Е. Б., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Басыров А. Г., д.т.н., проф., ВКА, С.-Петербург, РФ
Безродный Б. Ф., д.т.н., проф., НИИАС, Москва, РФ
Благовещенская Е. А., д.ф.-м.н., проф., ПГУПС,
С.-Петербург, РФ
Бубнов В. П., д.т.н., проф., С.-Петербург, РФ
Булавский П. Е., д.т.н., доц., ПГУПС, С.-Петербург, РФ
Василенко М. Н., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Глухов А. П., д.т.н., ПГУПС, С.-Петербург, РФ
Гуда А. Н., д.т.н., проф., РГУПС, Ростов-на-Дону, РФ
Ермаков С. Г., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Заборовский В. С., д.т.н., проф., СПбПУ, С.-Петербург, РФ
Канаев А. К., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Котенко А. Г., д.т.н., проф., ВНИИЖТ, Москва, РФ

Куренков П. В., д.э.н., к.т.н., проф., РУТ (МИИТ), Москва, РФ
Лецкий Э. К., д.т.н., проф., РУТ (МИИТ), Москва, РФ
Наседкин О. А., к.т.н., доц., ПГУПС, С.-Петербург, РФ
Никитин А. Б., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Новиков Е. А., д.т.н., доц., ВКА, С.-Петербург, РФ
Охтилев М. Ю., д.т.н., проф., НИО ЦИТ «Петрокомета»,
С.-Петербург, РФ
Привалов А. А., д.воен.н., проф., С.-Петербург, РФ
Соколов Б. В., д.т.н., проф., СПб ФИЦ РАН,
С.-Петербург, РФ
Таранцев А. А., д.т.н., проф., ИПТ РАН, С.-Петербург, РФ
Утепбергенов И. Т., д.т.н., проф., АУЭС, Алматы, Казахстан
Фазылов Ш. Х., д.т.н., проф., НИИ развития цифровых
технологий и ИИ, Ташкент, Узбекистан
Хабаров В. И., д.т.н., проф., СГУПС, Новосибирск, РФ
Ходаковский В. А., д.т.н., проф., ПГУПС, С.-Петербург, РФ
Чехонин К. А., д.ф.-м.н., доц., ХВИЦ ДВО РАН,
Хабаровск, РФ

Адрес редакции:

190031, Санкт-Петербург, Московский пр., 9
e-mail: itt-pgups@yandex.ru
Телефон: +7 (812) 457-86-06

Сетевое издание «Интеллектуальные технологии на транспорте (сетевой электронный научный журнал),
Intellectual Technologies on Transport» зарегистрировано Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций.
Свидетельство Эл № ФС77–61707 от 07 мая 2015 г.

Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).

Периодичность выхода – 4 номера в год. Выпуски журнала доступны на сайте <http://itt-pgups.ru>.

Информация предназначена для детей старше 12 лет.

Intellectual Technologies on Transport Issue 3 (35), 2023

ISSN 2413-2527

Network electronic scientific journal, open access. It publishes articles in Russian and English with the results of research and practical achievements in the field of intelligent technologies and associated research.

Founded in 2015.

Founder and Publisher

Federal State Budgetary Educational Institution of Higher Education
«Emperor Alexander I St. Petersburg State Transport University»

Editor-in-Chief

Khomonenko A. D., Prof., St. Petersburg, Russia

Co-chairs of the Editorial Council

Valinsky O. S., rector of PSTU, St. Petersburg, Russia
Charkin E. I., CIO of JSC «Russian Railways», Moscow, Russia

Editorial Council Members

Adadurov S. E., Prof., VNIIZHT, Moscow, Russia
Dudin A. N., Prof., BSU, Minsk, Belarus
Kornienko A. A., Prof., PSTU, St. Petersburg, Russia
Makarenko S. I., As. Prof., Inteltech, St. Petersburg, Russia

Merkuryev Yu. A., Prof., Academician of the Latvian Academy
of Sciences, RTU, Riga, Latvia
Titova T. S., Prof., First Vice-Rector PSTU, St. Petersburg, Russia
Yusupov R. M., Prof., Corr. Member of RAS, SPC RAS,
St. Petersburg, Russia

Editorial Board Members

Bozhko L. M., As. Prof., PSTU, St. Petersburg, Russia –
Deputy Editor-in-Chief
Batalov D. I., PSTU, St. Petersburg, Russia –
Science Editor
Aleksandrova E. B., Prof., SPbPU, St. Petersburg, Russia
Basyrov A. G., Prof., MSA, St. Petersburg, Russia
Bezrodny B. F., Prof., NIIAS, Moscow, Russia
Blagoveshchenskaya E. A., Prof., PSTU,
St. Petersburg, Russia
Bubnov V. P., Prof., St. Petersburg, Russia
Bulavsky P. E., As. Prof., PSTU, St. Petersburg, Russia
Vasilenko M. N., Prof., PSTU, St. Petersburg, Russia
Glukhov A. P., PSTU, St. Petersburg, Russia
Guda A. N., Prof., RSTU, Rostov-on-Don, Russia
Ermakov S. G., Prof., PSTU, St. Petersburg, Russia
Zaborovsky V. S., Prof., SPbPU, St. Petersburg, Russia

Kanaev A. K., Prof., PSTU, St. Petersburg, Russia
Kotenko A. G., Prof., VNIIZHT, Moscow, Russia
Kurenkov P. V., Prof., RUT (MIIT), Moscow, Russia
Letsky E. K., Prof., RUT (MIIT), Moscow, Russia
Nasedkin O. A., As. Prof., PSTU, St. Petersburg, Russia
Nikitin A. B., Prof., PSTU, St. Petersburg, Russia
Novikov E. A., As. Prof., MSA, St. Petersburg, Russia
Okhtilev M. Yu., Prof., JSC «Petrokometa», St. Petersburg, Russia
Privalov A. A., Prof., PSTU, St. Petersburg, Russia
Sokolov B. V., Prof., SPC RAS, St. Petersburg, Russia
Tarantsev A. A., Prof., IPT RAS, St. Petersburg, Russia
Utebergenov I. T., Prof., AUPET, Almaty, Kazakhstan
Fazilov Sh. X., Prof., AIRI, Tashkent, Uzbekistan
Khabarov V. I., Prof., STU, Novosibirsk, Russia
Khodakovskiy V. A., Prof., PSTU, St. Petersburg, Russia
Chekhonin K. A., As. Prof., Khabarovsk FRC RAS, Khabarovsk, Russia

Editorial address:

190031, St. Petersburg, Moskovsky ave., 9
e-mail: itt-pgups@yandex.ru
Phone: +7 812 457 86 06

The online journal «Intellectual Technologies on Transport» is registered by the Federal Service for Supervision of Communications, Information Technologies, and Mass Media.
EI No. FS77-61707 Testimony from May 7, 2015.

The journal is registered in the Russian Science Citation Index (RSCI).

Frequency of release - 4 issues per year. Issues of the magazine are available at <http://itt-pgups.ru>

The content is for children over the age of 12.

© Federal State Budgetary Educational Institution of Higher Education
«Emperor Alexander I St. Petersburg State Transport University», 2023

Содержание

Интеллектуальные системы и технологии на транспорте

- Охтилев М. Ю., Коромысличенко В. Н., Охтилев П. А., Зянчурин А. Э., Васильев В. И.*
Концепция инженерии знаний в задачах обеспечения интероперабельности
АСУ и информационных систем на основе интеллектуальных технологий 5
- Степанов Ю. Н., Яковлева Н. В., Ермаков С. Г., Баталов Д. И.*
Реализация проекта по роботизации рутинных операций для повышения операционной
эффективности компании 14

Системный анализ и моделирование сложных систем

- Герасименко П. В., Ходаковский В. А.*
Моделирование и оценивание риска достижения планового среднего расстояния перевозки
груза автомобильным транспортом России 22
- Кизляк О. П., Дергачев А. И., Дергачев С. А., Тарбаева Е. А., Сергеева Т. Г.*
К вопросу оценки эффективности деятельности логистических компаний: новая реальность,
концептуальный подход 28
- Хонина О. И., Забродин А. В.*
Задача раскраски графа в контексте оптимизации расписания: программное решение 32
- Аль-Хазраджи С. Х., Халил М. М., Халил И. Х.*
Биомедицинское прикладное программное обеспечение для измерения размеров частиц
на электронных микроскопических изображениях (на англ.) 38

Информационная безопасность и кибербезопасность

- Смагин В. А., Солдатенко В. С.*
Новые горизонты применения в теории надежности высоконадежных систем
гипотезы Н. М. Седякина 44
- Хоршид Ф. Х., Ибрагим Н. Х., Сайхунд К.*
Система прогнозирования дефектов программного обеспечения на основе
хорошо отлаженной техники «случайного леса» (на англ.) 51

Телекоммуникации

- Слободчиков Н. А., Дергачев А. И., Куранова О. Н.*
Перспективное развитие новых услуг сервиса в городском транспортном потоке (на англ.) 60

Электронное обучение и дистанционные образовательные технологии

- Федоров Д. Ю., Забродин А. В.*
Оптимизация образовательного процесса: построение индивидуальных учебных траекторий
с помощью вариативной части учебного плана на примере задачи о рюкзаке 68

-
- Памяти коллеги 74

Contents

Intellectual systems and transport technologies

Okhtilev M. Yu., Koromyslichenko V. N., Okhtilev P. A., Zianchurin A. E., Vasiljev V. I.
Problems of Ensuring Interoperability in the Integration of Existing Automated Control Systems
and Information Systems Based on Intelligent Technologies 5

Stepanov Yu. N., Yakovleva N. V., Ermakov S. G., Batalov D. I.
Implementation of a Project on Robotization of Routine Operations to Improve
the Operational Efficiency of the Company 14

System analysis and modelling of complex systems

Gerasimenko P.V., Khodakovsky V. A.
Modeling and Assessing the Risk of Achieving the Planned Average Distance of Cargo
Transportation by Road Transport in Russia 22

Kizlyak O. P., Dergachev A. I., Dergachev S. A., Tarbaeva E. A., Sergeeva T. G.
On the Issue of Assessing the Effectiveness of Logistics Companies: A New Reality,
Conceptual Approach 28

Khonina O. I., Zabrodin A. V.
Graph Coloring Problem in the Context of Schedule Optimization: Software Solution 32

Al-khazraji S. H., Khalil M. M., Khalil A. I.
A Biomedical Application Software for Measuring Particles Sizes in Electro-Microscopic
Images (in English) 38

Information security and cybersecurity

Smagin V. A., Soldatenko V. S.
New Horizons of Application in the Theory of Reliability of Highly Reliable Systems
of the Hypothesis of N. M. Sedyakin 44

Khorsheed F. H., Ibrahim N. J., Saihood Q.
Software Defect Prediction System Based on Well-Tuned Random Forest Technique (in English) 51

Telecommunication

Slobodchikov N. A., Dergachev A. I., Kuranova O. N.
Prospective Development of New Service Offerings in the Urban Traffic Flow (in English) 60

E-learning and distant educational technologies

Fedorov D. Yu., Zabrodin A. V.
Optimization of the Educational Process: Construction of Individual Educational Trajectories
Using the Variable Part of the Curriculum Using the Example of the Knapsack Problem 68

.....

In Memory of Colleague 74

Концепция инженерии знаний в задачах обеспечения интероперабельности АСУ и информационных систем на основе интеллектуальных технологий

д.т.н. М. Ю. Охтилев, к.т.н. В. Н. Коромысличенко,
к.т.н. П. А. Охтилев, А. Э. Зянчурин
Санкт-Петербургский государственный университет
аэрокосмического приборостроения
Санкт-Петербург, Россия
oxt@mail.ru, v.koromyslichenko@yandex.ru,
pavel.oxt@mail.ru, zwilas@gmail.ru

В. И. Васильев
Санкт-Петербургский государственный
университет телекоммуникаций
Санкт-Петербург, Россия
vasiljev.valentin2014@yandex.ru

Аннотация. Рассматривается специфика инженерии знаний при обеспечении интероперабельности действующих и модернизируемых АСУ и информационных систем. Показывается, что интероперабельность может быть достигнута за счет обеспечения декларативного характера проектирования, основанного на применении концептуального подхода к моделированию и онтологическому описанию стадий жизненного цикла. Декларативное проектирование может быть описано в рамках логико-лингвистической модели, при этом экспертами разрабатываются графы знаний, на базе которых составляются частные онтологии различных стадий жизненного цикла.

Ключевые слова: интероперабельность, модель предметной области, онтология, данные, знания, графы, прикладное программное обеспечение, искусственный интеллект.

ВВЕДЕНИЕ

Обеспечение интероперабельности предполагает объединение источников данных и структурно-функциональных моделей АСУ и информационных систем (ИС), при этом требуемая функциональность АСУ и ИС может быть достигнута путем реализации исполнительных систем, осуществляющих вычисления с опорой на заложенные в систему знания при использовании принципов декларативного проектирования прикладного программного обеспечения (ПО) [1–3]. В этом случае реализованные в АСУ и ИС любые алгоритмы и модели данных должны быть представимы декларативными спецификациями в виде интерпретируемых наборов данных. Реализация указанного подхода предполагает применение унифицированных онтологических полимодельных комплексов, обеспечивающих всю систему едино интерпретируемой терминологией. Задание исполнимых спецификаций в этом случае должно осуществляться экспертами в процессе анализа и разработки модели предметной области (ПрО) в виде базы знаний, на проблемно-ориентированных языках желательно с визуальной (графической) нотацией (с учетом положения, что у экспертов-технологов навыки программирования отсутствуют) [4–6].

ОНТОЛОГИЧЕСКИЕ МОДЕЛИРОВАНИЕ

В ДЕКЛАРАТИВНОМ ПРОЕКТИРОВАНИИ ПРИКЛАДНОГО ПО

Применение интеллектуальных систем обеспечивает решение сложных задач интероперабельности, при этом основная трудность решения связана с использованием слабо формализованных знаний специалистов-практиков, в которых логическая (или смысловая) обработка информации превалирует над вычислительной. Изучение методов проектирования таких систем, как показано в работе [7], находится в поле теории искусственного интеллекта (ИИ, или инженерии знаний) и предполагает разработку моделей и методов извлечения, структурирования, формализации (представления) знаний для их обработки в интеллектуальных АСУ и ИС.

Имеющая место в этом случае цепочка «данные — информация — знания», предполагает, во-первых, целесообразность определения семантики (смысла) данных на основе интерпретации синтаксических выражений, а во-вторых, позволяет говорить о гносеологической цепочке представления знаний при выборе языков моделирования: факт, обобщенный факт, эмпирический закон, теоретический закон [8–10]. Так, например, такой авторитет в области ИИ как Э. Х. Тыгу рассматривал программирование как метод доказательства теорем в предположении, что синтаксис языка программирования ограничивает возможности описания решений математических задач, а корректный и конечный процесс вычислений — собственно доказательство (вывод) теоремы [11].

В процессе разработки модели ПрО и наполнения базы знаний (БЗ) возникают задачи, связанные с извлечением эмпирического знания экспертов, которое характеризуется [7, 9]:

- определением знаний в различных категориях, или модальностью;
- диалектичностью знаний, или противоречивостью;
- невозможностью полного описания ПрО, или неполнотой.

С 80-х годов 20-го века и по настоящее время в попытках формализации проблемно-ориентированных языков пред-

ставления знаний и преодоления проблем извлечения эмпирического опыта экспертов высокими темпами развивается направление, связанное с использованием онтологического моделирования как основы организации БЗ в АСУ и ИС.

Онтологию представляют в виде спецификации концептуализации, где концептуализация — структура реальности, рассматриваемая независимо от словаря предметной области конкретной ситуации (определение Т. Груббера) [7, 9, 12].

Онтологии позволяют формализовать описания предметных сущностей и их взаимосвязей при проектировании АСУ и ИС [7]. Это обстоятельство определяет их важную роль в теории ИИ. С другой стороны, онтологическое моделирование обеспечивает сохранение семантики и возможность последующего использования знаний о ПрО. Таким образом, в рамках концептуального определения терминологического словаря той ПрО, для которой создаются АСУ и ИС, онтологическое моделирование должно предварять разработку любой АСУ и ИС [7, 12]. Следует отметить, что в зависимости от практического приложения онтологии могут применяться в гораздо более широком спектре задач, связанных с обеспечением сквозного взаимодействия разнородных систем с сохранением семантики передаваемой информации.

Сложные ПрО, в том числе для критических приложений, характеризуются структурированными и слабоструктурированными данными, хранимыми, как правило, в разнотипных БД/БЗ. При этом в условиях территориально распределенных существующих и унаследованных АСУ и ИС требуется их совместная обработка. Реализация такой совместной обработки, по существу, и является основной целью обеспечения семантической интероперабельности систем.

В результате проведенного к настоящему времени анализа различных технологий интеграции данных [7, 13–25] выявлено, что перспективным методом семантической интеграции автоматизированных систем является применение программных медиаторов, основанных на онтологиях и выступающих как семантические посредники в обработке данных. Такие семантические медиаторы являются носителями онтологических БД, что обеспечивает унификацию их языковых систем и позволяет обеспечивать семантически целостный диалог между взаимодействующими АСУ и ИС, куда такие программные медиаторы внедрены.

Следует отметить, что онтология определяет понятийную систему ПрО, то есть совокупность ее понятий (концептов) и отношений между ними, в то время как концептуальная модель описывает структуру и поведенческие аспекты ПрО. Концептуальная модель ПрО включает сущности, релевантные целям системы, их определения и связи между ними, в то время как онтология — это понятия ПрО и семантические отношения между ними. Онтология, следовательно, логически дополняет модель ПрО, представленную ее концептуальной схемой. В технологиях БЗ использование онтологии в качестве концептуальных схем ПрО и в качестве основы интерфейсов разнородных систем БЗ позволяет формулировать запросы данных в терминах онтологий и осуществлять рассуждения (Reasoning) на онтологиях. Понятия и отношения базирующегося на онтологии интерфейса в этом случае ста-

новятся посредниками между пользователем с его информационными потребностями. Информационные потребности пользователя выражаются в терминах онтологий и традиционной модели данных. Такие системы называют системами доступа к данным, основанными на онтологиях, или OBDA-системами (Ontology-Based Data Access Systems, OBDA-systems) [41].

Онтологический подход используется в OBDA-системах для онтологического аннотирования концептуальных схем БД/БЗ. Это делается с целью достижения независимости от технических аспектов СУБД и формулирования запросов в терминах ПрО [11, 19, 20, 26–29]. Так как концептуальное моделирование обычно используется для описания поведенческих и функциональных аспектов, онтологический подход может быть использован и в этом случае. Следует отметить, что на текущий момент такой подход еще не получил широкого распространения [9, 29]. Например, в работе [30] определяется понятие «онтологической чистоты» — показателя корректности и полноты концептуальной модели относительно ее онтологии и отмечается, что онтолого-управляемое проектирование становится легче для понимания при использовании онтологий в силу оперирования естественными для пользователя понятиями.

Таким образом, при разработке интеллектуальных технических систем онтологии целесообразно использовать для [9]:

- формирования и фиксации общего разделяемого всеми экспертами блока знаний о ПрО;
 - описания семантики данных с целью явной концептуализации ПрО;
 - обеспечения возможности повторного использования знаний;
 - описания функциональности АСУ и ИС (типов решаемых задач).
- Онтология может использоваться в процессе функционирования АСУ и ИС с целью:
- обеспечения совместного использования разнородных данных и знаний;
 - реализации процессов, составляющих функциональность системы;
 - лучшего понимания ПрО пользователями системы;
 - достижения интеграции и миграции разнородных данных и знаний между территориально распределенными АСУ и ИС;

– обеспечения функциональной совместимости существующих неинтеллектуальных систем путем надстройки поверх них семантически аннотирующих модельно-алгоритмических комплексов, основанных на онтологиях.

Для формализации онтологий в качестве теоретического аппарата зачастую используются дескрипционные логики [7, 29, 31–37]: по существу — семейство формализмов для структурированного представления знаний, которое создано для применения формальной семантики в структурированных, не основанных на логике языках представления знаний — семантических сетях и языках фреймов. Дескрипционные логики реализуют компромисс между выразительными возможностями и вычислительной сложностью рассуждений, обеспечивая сочетание возможностей разрешимых фрагментов логики пер-

вого порядка и модальных логик с некоторыми расширениями, так как не все аспекты реального мира представимы в этих логиках [38].

Почти все обобщения имеют исключения или соблюдают только до определенной степени, что составляет основную сложность (принцип естественных разновидностей) [39]. В этой связи целесообразно разделять типовые сущности и исключительные (индивиды). Следующий важный аспект — использование категорий для суждений о конкретных объектах, которые проводятся с заданием на них отношений (таксономии, партономии, тезаурусы и пр.). Декрипционная логика позволяет сравнительно легко разрешить эти задачи и в ряде случаев имеет более низкую вычислительную сложность, чем, например, выражения в логике с двухместными предикатами. Рассмотренные положения подтверждают перспективность применения онтологий для вычислительно сложных задач, обусловленных совместным манипулированием разнородной информацией, передаваемой из различных источников данных существующих АСУ и ИС.

Такие принципы обобщения и категоричное описание позволяют задавать поверх разнородных моделей данных существующих и унаследованных АСУ и ИС унифицированные модели представления знаний, основанные на онтологиях. Эта особенность позволяет интегрировать разнородную информацию АСУ и ИС, выполненных на различных платформах, и обеспечить создание единой интерпретируемой системы БЗ, которая может стать основой реализации интероперабельности АСУ и ИС.

Кроме того, необходимо отметить, что использование отдельно взятых онтологий или их линейной совокупности (например, в виде множества предметных онтологий БД) характерно для многих существующих интеллектуальных технических систем [7, 12, 16, 40–42]. Иногда используется иерархия онтологий для интеграции гетерогенных источников данных. Онтологии, как формализм представления знаний, могут быть использованы не только для описания статичной модели ПрО, но также и для описания процессных аспектов, иерархий целей и задач [39, 43–45]. При переходе от одной онтологии к другой могут задаваться ограничения, либо допущения для задаваемой системы фактов о ПрО, что обеспечивается определением некоторой совокупности взаимосвязанных онтологий так, что это позволяет более гибко проектировать модели. Подобную совокупность онтологий принято называть онтологической системой. В литературных источниках приводятся различные определения этого понятия, однако на практике соответствующей формальной и реализующей системы найдено не было [7, 21, 28].

ПРАКТИЧЕСКИЕ ВОПРОСЫ

ДЕКЛАРАТИВНОГО ПРОЕКТИРОВАНИЯ ПРИКЛАДНОГО ПО

Декларативное проектирование прикладного ПО базируется на системном анализе ПрО (анализе исходных данных и разработке модели ПрО), разработке спецификаций требований, разработке кода, тестировании и сопровождении в соответствии с концепцией управления жизненным циклом программных систем [40]. Анализ исходных данных и разработка спецификаций требований — это важнейший этап проектирования, опирающийся на экспертные знания. Экспертные знания о ПрО существенно сложнее

современных логических, формальных или иных представлений таких моделей. Существует значительный разрыв между абстрактным, экспертным представлением модели ПрО и ее практическим выражением в виде БД или БЗ. Экспертные знания логически не полностью упорядочены, постоянно дополняются и изменяются. В таких условиях использование традиционных БД/БЗ затруднено, в этом случае рассмотрение возможности применения нетрадиционных подходов к построению модели ПрО — графовых, документарных, ключ-значение и других (NoSQL) — представляется достаточно обоснованным.

В рамках логико-лингвистической парадигмы [10] создание модели ПрО можно представить в виде следующей модели разработки, схема которой представлена на рисунке 1.

Логико-лингвистическая модель разработки предполагает преобразование данных, представленных в графической или лингвистической форме, в логическую форму (по аналогии с логико-лингвистической системой управления Д. А. Поспелова [10]). Представленная модель относится к семиотическим системам и на ее основе может быть разработана модель любой ПрО в виде онтологий с использованием дескрипционных логик или других языковых формализмов.

Выбор модели представления знаний на различных этапах трансформации экспертных знаний в онтологию имеет существенное значение и не может быть произвольным. На первой стадии осуществляется извлечение требований (Requirements Elicitation), которое проводится на основании анализа замысла заказчика, исходных данных, модели предметной области и потребностей заинтересованных в виде интервьюирования, сценарного анализа, моделирования и прототипирования. При этом экспертные знания должны быть представлены в лингвистическом и графическом виде, для того чтобы избежать когнитивных потерь знаний при их передаче на следующую стадию трансформации в соответствии с этапом жизненного цикла программы [46].

Извлечение требований, при онтологическом подходе, предполагает составление перечня терминов, понятий и их определений. Термины и понятия должны располагаться в перечне с учетом отношений и зависимостей, существующих между ними. Таким образом, создается структурированная онтологическая модель — глоссарий/тезаурус, которая естественным образом может быть представлена в виде графа помеченных свойств (Labeled Property Graph, LPG) [47]. Далее, на основании графа глоссария и исследования ПрО, разрабатываются LPG-графы информационных процессов и релевантных им метаданных, соответствующие целям и задачам программной системы, определяются выявленные вычислительные задачи и их взаимосвязи. На базе анализа полученных LPG-графов формируется концептуальная архитектура и спецификация требований. Полученная система изоморфных графов используется как верхний уровень описания требований и архитектуры программной системы, как графы знаний предметной области, на базе которых могут быть разработаны различные частные онтологии: требований заказчика, спецификаций требований, описания архитектуры и прочие.



Рис. 1. Логико-лингвистическая модель формирования схемы предметной области

Следует отметить, что LPG-графы могут быть разработаны экспертами, не имеющими навыков программирования, а формализованное описание онтологий — специалистами инженерии знаний. При этом графы знаний обеспечивают валидацию и верификацию онтологий, легкое добавление и изменение знаний.

ЗАКЛЮЧЕНИЕ

Онтологическое моделирование и концептуализация лежат в основе декларативного проектирования, которым обеспечивается достижение интероперабельности. При этом на начальных стадиях жизненного цикла с помощью экспертных знаний моделируются структурные отношения и связи между информационными понятиями, релевантными терминам и понятиям ПрО, создается граф знаний ПрО. Полученный экспертами граф лежит в основе частных онтологий различных стадий жизненного цикла программных систем.

ЛИТЕРАТУРА

1. Макаренко, С. И. Интероперабельность человеко-машинных интерфейсов: Монография. — Санкт-Петербург: Научные технологии, 2023. — 185 с.
2. UK Ministry of Defence Architecture Framework. URL: <http://www.modaf.com> (дата обращения 21.10.2018).

3. Giachetti, R. Interoperability Analysis Method for Mission-Oriented System of Systems Engineering / R. Giachetti, S. Wangert, R. Eldred // Proceedings of the 2019 IEEE International Systems Conference (SysCon) (Orlando, FL, USA, 08–11 April 2019). — Institute of Electrical and Electronics Engineers, 2019. — 6 p. DOI: 10.1109/SYSCON.2019.8836808.

4. Олейников, А. Интероперабельность — ключевая технология повышения эффективности систем вооружения, управления и связи / А. Олейников, С. Макаренко, С. Козлов // Радиоэлектронные технологии. 2022. № 1. С. 66–73.

5. Макаренко, С. И. Сетевая война — принципы, технологии, примеры и перспективы: Монография / С. И. Макаренко, М. С. Иванов. — Санкт-Петербург: Научные технологии, 2018. — 898 с.

6. NCOIC Interoperability Framework (NIF v2.1) and NIF™ Solution Description Reference Manual (NSD-RM v1.2) / G. Osvalds, M. Bowler, A. Jones, [et al.] — Network Centric Operations Industry Consortium, 2010. — 125 p. URL: <http://documents.tips/documents/nif-solution-description-reference-manual-nsd-rm-reference-manual-nsd-rm.html> (дата обращения 21.10.2018).

7. Охтилев, П. А. Алгоритмы и онтологические модели информационно-аналитической поддержки процессов создания и применения космических средств: дис. на соиск.

учен. степ. канд. техн. наук: 05.13.01 / Охтилев Павел Алексеевич; Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). — Санкт-Петербург, 2019. — 408 с.

8. Гаазе-Рапопорт, М. Г. Структура исследований в области искусственного интеллекта / М. Г. Гаазе-Рапопорт, Д. А. Поспелов // Толковый словарь по искусственному интеллекту / Авт.-сост.: А. Н. Аверкин, М. Г. Гаазе-Рапопорт, Д. А. Поспелов. — Москва: Радио и связь. Редакция литературы по информатике и вычислительной технике, 1992. — С. 5–20.

9. Гаврилова, Т. А. Инженерия знаний. Модели и методы: Учебник / Т. А. Гаврилова, Д. В. Кудрявцев, Д. И. Муромцев. — Санкт-Петербург: Лань, 2016. — 324 с. — (Учебники для вузов. Специальная литература).

10. Искусственный интеллект: Справочник: в 3 кн. Кн. 2. Модели и методы / под ред. Д. А. Поспелова. — Москва: Радио и связь. Редакция литературы по информатике и вычислительной технике, 1990. — 304 с.

11. Тыгу, Э. Х. Концептуальное программирование. — Москва: Наука. Главная редакция физико-математической литературы, 1984. — 256 с. — (Проблемы искусственного интеллекта).

12. Боргест, Н. М. Онтология проектирования. Теоретические основы: Учебное пособие. Часть 1. Понятия и принципы. — Самара: Самарский гос. аэрокосмический ун-т, 2010. — 91 с.

13. Атаева, О. М. Основные понятия формальной модели семантических библиотек и формализация процессов интеграции в ней / О. М. Атаева, В. А. Серебряков // Программные продукты и системы. 2015. № 4 (112). С. 180–187. DOI: 10.15827/0236-235X.112.180-187.

14. Ахметов, Р. Н. Концепция создания и применения перспективной АСУ подготовки и пуска ракеты космического назначения «Союз-2»: новые подходы к интеграции, интеллектуализации, управлению / Р. Н. Ахметов, И. Е. Васильев, В. А. Капитонов, [и др.] // Авиакосмическое приборостроение. 2015. № 4. С. 3–54.

15. Бахмут, А. Д. Модель интеграции информационных ресурсов производственного цикла РН «Союз-2» на основе онтологического доступа к данным / А. Д. Бахмут, М. Ю. Охтилев, П. А. Охтилев // Научная сессия ГУАП: Сборник докладов научной сессии, посвященной Всемирному дню авиации и космонавтики (Санкт-Петербург, Россия, 08–12 апреля 2019 г.): в 3 ч. Часть 2. Технические науки. — Санкт-Петербург: ГУАП, 2019. — С. 295–300.

16. Биряльцев, Е. В. Интеграция реляционных баз данных на основе онтологий / Е. В. Биряльцев, А. М. Гусенков // Ученые записки Казанского университета. Серия Физико-математические науки. 2007. Т. 149, Кн. 2. С. 13–34.

17. Божинский, И. А. Методы и технологии интеграции информационных систем и распределенных баз данных // Радиоэлектроника и информатика. 2015. № 2. С. 30–36.

18. Бова, В. В. Онтологическая модель интеграции данных и знаний в интеллектуальных информационных системах // Известия Южного федерального университета. Технические науки. 2015. С. 225–237.

19. Вовченко, А. Е. Анализ и сравнение систем интеграции неоднородных информационных ресурсов / А. Е. Вовченко, Л. А. Калинин // Электронные библиотеки: пер-

спективные методы и технологии, электронные коллекции (RCDL-2008): Труды 10-й Всероссийской научной конференции (Дубна, Россия, 07–11 октября 2008 г.). — Дубна: Объединенный институт ядерных исследований, 2008. — С. 115–121.

20. Когаловский, М. Р. Методы интеграции данных в информационных системах. 2010. 9 с. URL: http://www.ipgras.ru/old_site/articles/kogalov10-05.pdf (дата обращения 05.09.2022).

21. Левашова, Т. В. Принципы управления онтологиями, используемые в среде интеграции знаний // Труды СПИИРАН. 2002. Т. 2, Вып. 1. С. 51–68.

22. Маслобоев, А. В. Мультиагентная система интеграции распределенных информационных ресурсов инноваций / А. В. Маслобоев, М. Г. Шишаев // Программные продукты и системы. 2007. № 4. С. 30–32.

23. Серебряков, В. А. Семантическая интеграция данных // Просеминар кафедры системного программирования МГУ (Москва, Россия, весна 2012 г.). — Москва, 2012. — 119 с. URL: <http://sp.cmc.msu.ru/proseminar/2012/serebryakov.2012.04.20.pdf> (дата обращения 22.03.2018).

24. Ткаченко, Н. И. Применение сервис-ориентированной архитектуры при интеграции систем управления технологическими процессами / Н. И. Ткаченко, Н. А. Спирин // Известия Томского политехнического университета. Уpravление, вычислительная техника и информатика. 2010. Т. 317, № 5. С. 61–67.

25. Черняк, Л. Интеграция данных: синтаксис и семантика // Открытые системы. СУБД. 2009. № 10. С. 24–29.

26. Когаловский, М. Р. Метаданные в компьютерных системах // Программирование. 2013. Т. 39, № 4. С. 28–46.

27. Когаловский, М. Р. Перспективные технологии информационных систем. — Москва: ДМК Пресс, 2003. — 288 с. — (ИТ-Экономика) (Лекции МГУ).

28. Евгеньев, Г. Б. Интеллектуальные системы проектирования: Учебное пособие для студентов вузов. — Москва: Изд-во МГТУ им. Н. Э. Баумана, 2009. — 335 с. — (Информатика в техническом университете).

29. Когаловский, М. Р. Системы доступа к данным, основанные на онтологиях // Программирование. 2012. Т. 38, № 4. С. 55–77.

30. Montiel-Sanchez, C. The BWV-Model as Method Engineering Theory / C. Montiel-Sanchez, D. Pfeiffer, J. Becker // Proceedings of the 13th Americas Conference on Information Systems (AMCIS) (Keystone, CO, USA, 09–12 August, 2007). — Association for Information Systems, 2007. — Art. No. 83, 10 p.

31. Цуканова, Н. И. Онтологическая модель представления и организации знаний: Учебное пособие для вузов. — Москва: Горячая линия — Телеком, 2016. — 272 с.

32. The Description Logic Handbook: Theory, Implementation, and Applications. Second Edition / F. Baader, [et al.] (eds.). — New York: Cambridge University Press, 2007. — 621 p. DOI: 10.1017/CBO9780511711787.

33. Murthy, S. K. Automatic Construction of Decision Trees from Data: A Multidisciplinary Survey // Data Mining and Knowledge Discovery. 1998. Vol. 2, Is. 4. Pp. 345–389. DOI: 10.1023/A:1009744630224.

34. Application of Decision-Making Support Technology for Management of Space Vehicle Life Cycle / P. A. Okhtilev, A. D. Bakhmut, A. V. Krylov, [et al.] // Proceedings of the II International Conference on Control in Technical Systems (CTS-2017)

(Saint Petersburg, Russia, 25–27 October 2017). — Institute of Electrical and Electronics Engineers, 2017. — Pp. 41–44.
DOI: 10.1109/CTSYS.2017.8109483.

35. Pan, J. Z. Description Logics: Reasoning Support for the Semantic Web: A Thesis for the Degree of Doctor of Philosophy. — University of Manchester, 2004. — 215 p.

36. Linking Data to Ontologies / A. Poggi, D. Lembo, D. Calvanese, [et al.] // Journal on Data Semantics X / S. Spaccapietra (ed.). — Heidelberg: Springer, 2008. — Pp. 133–173. — (Lecture Notes in Computer Science, Volume 4900).
DOI: 10.1007/978-3-540-77688-8_5.

37. Progressions and Innovations in Model-Driven Software Engineering / V. G. Díaz, [et al.] (eds.). — Hershey (PA): IGI Global, 2013. — 388 p. — (Premier Reference Source).
DOI: 10.4018/978-1-4666-4217-1.

38. Вахитов, А. Р. Преимущества дескриптивной логики при обработке знаний / А. Р. Вахитов, В. Б. Новосельцев // Известия Томского политехнического университета. Управление, вычислительная техника и информатика. 2008. Т. 313, № 5. С. 73–76.

39. Рассел, С. Искусственный интеллект: современный подход. Второе издание = Artificial Intelligence: A Modern Approach. Second Edition / пер. с англ. и редакция К. А. Птицына. — Москва: ИД Вильямс, 2006. — 1408 с.

40. Охтилев, М. Ю. Программная инженерия. Инженерный подход: Учебное пособие / М. Ю. Охтилев, В. Н. Коромысличенко, П. А. Охтилев. — Санкт-Петербург: ГУАП, 2021. — 163 с.

41. Гаранина, Н. О. Онтология процессов, ориентированная на верификацию / Н. О. Гаранина, И. С. Ануреев, О. И. Боровикова // Моделирование и анализ информационных систем. 2018. Т. 25, № 6. С. 607–622.
DOI: 10.18255/1818-1015-2018-6-607-622.

42. Горшков, С. В. Введение в онтологическое моделирование. Ревизия 2.4: Методическое пособие. — [Екатеринбург]: ТриниДата, 2018. — 150 с. URL: <http://trinidata.ru/files/SemanticIntro.pdf> (дата обращения 01.10.2018).

43. Фёдоров, И. Г. Адаптация онтологии Бунге — Ванда — Вебера к описанию исполняемых моделей бизнес-процессов // Прикладная информатика. 2015. Т. 10, № 4 (58). С. 82–92.

44. Фёдоров, И. Г. Анализ концептуальной модели бизнес-процесса с использованием онтологии Бунге — Ванда — Вебера // Экономика, статистика и информатика. Вестник УМО. 2014. № 6. С. 216–221.
DOI: 10.21686/2500-3925-2014-6-216-221.

45. Фёдоров, И. Г. Моделирование бизнес-процессов в нотации BPMN2.0: Монография. — Москва: Московский гос. ун-т экономики, статистики и информатики, 2013. — 255 с.

46. Куликова, А. А. Методы и средства формирования и использования онтологий проектов в процессе проектирования автоматизированных систем: дис. на соиск. учен. степ. канд. техн. наук: 05.13.12 / Куликова Анна Александровна; Ульяновский государственный технический университет. — Ульяновск, 2022. — 207 с.

47. Фаулер, М. NoSQL: новая методология разработки нереляционных баз данных = NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence / М. Фаулер, П. Дж. Салададж; Пер. с англ. Д. А. Ключина. — Москва: ИД «Вильямс», 2013. — 192 с.

Problems of Ensuring Interoperability in the Integration of Existing Automated Control Systems and Information Systems Based on Intelligent Technologies

Grand PhD M. Yu. Okhtilev, PhD V. N. Koromyslichenko,
PhD P. A. Okhtilev, A. E. Zianchurin
St. Petersburg State University of Aerospace Instrumentation
Saint Petersburg, Russia
oxt@mail.ru, v.koromyslichenko@yandex.ru,
pavel.oxt@mail.ru, zwilas@gmail.ru

V. I. Vasiljev
St. Petersburg State University
of Telecommunications
Saint Petersburg, Russia
vasiljev.valentin2014@yandex.ru

Abstract. The article discusses the specifics of knowledge engineering in ensuring the interoperability of existing and modernized automated control systems and information systems. It is shown that interoperability can be ensured by declarative design based on conceptual modeling and ontological description of the stages of the life cycle. Declarative design can be described within the framework of a logical-linguistic model, while experts develop knowledge graphs on the basis of which private ontologies of various stages of the life cycle are compiled.

Keywords: interoperability, domain model, ontology, data, knowledge, graphs, application software, artificial intelligence.

REFERENCES

1. Makarenko S. I. Interoperability of human-machine interfaces: Monograph [Interoperabelnost cheloveko-mashinnykh interfeysov: Monografiya]. Saint Petersburg, Naukoemkie Tekhnologii Publishing House, 2023, 185 p.
2. UK Ministry of Defence Architecture Framework. Available at: <http://www.modaf.com> (accessed 21 Oct 2018).
3. Giachetti R., Wangert S., Eldred R. Interoperability Analysis Method for Mission-Oriented System of Systems Engineering, *Proceedings of the 2019 IEEE International Systems Conference (SysCon), Orlando, FL, USA, April 08–11, 2019*. Institute of Electrical and Electronics Engineers, 2019, 6 p. DOI: 10.1109/SYSCON.2019.8836808.
4. Oleynikov A., Makarenko S., Kozlov S. Interoperability — A Key Technology for Improving the Efficiency of Weapons Systems, Control and Communications [Interoperabelnost — klyuchevaya tekhnologiya povysheniya effektivnosti sistem vooruzheniya, upravleniya i svyazi], *Radio Electronic Technologies [Radioelektronnye tekhnologii]*, 2022, No. 1, Pp. 66–73.
5. Makarenko S. I., Ivanov M. S. Network-centric warfare — principles, technologies, examples and prospects: Monograph [Setentsricheskaya voyna — printsipy, tekhnologii, primery i perspektivy: Monografiya]. Saint Petersburg, Naukoemkie Tekhnologii Publishing House, 2018, 898 p.
6. Osvalds G., Bowler M., Jones A., et al. NCOIC Interoperability Framework (NIF v2.1) and NIF™ Solution Description Reference Manual (NSD-RM v1.2). Network Centric Operations Industry Consortium, 2010, 125 p. Available at: <http://dokumen.tips/documents/nif-solution-description-reference-manual-nsd-rm-reference-manual-nsd-rm.html> (accessed 21 Oct 2018).
7. Okhtilev P. A. Algorithms and Ontological Models of Information and Analytical Support for the Creation and Application of Space Facilities [Algoritmy i ontologicheskie modeli informatsionno-analiticheskoy podderzhki protsessov sozdaniya i primeneniya kosmicheskikh sredstv]: diss. on competition of a scientific degree PhD (Engin.). Saint-Petersburg, 2019, 408 p.
8. Gaaze-Rapoport M. G., Pospelov D. A. The structure of research in the field of artificial intelligence [Struktura issledovaniy v oblasti iskusstvennogo intellekta]. In: *Averkin A. N., Gaaze-Rapoport M. G., Pospelov D. A. (eds.) Explanatory dictionary on artificial intelligence [Tolkovyy slovar po iskusstvennomu intellektu]*. Moscow, Radio and Communications Publishing House, 1992, Pp. 5–20.
9. Gavrilova T. A., Kudryavtsev D. V., Muromtsev D. I. Engineering of knowledge. Models and methods: Textbook [Inzheneriya znaniy. Modeli i metody: Uchebnik]. Saint Petersburg, LAN Publishing House, 2016, 324 p.
10. Pospelov D. A. (ed.) Artificial intelligence: Reference book. Book 2. Models and methods [Iskusstvennyy intellekt: Spravochnik. Kniga 2. Modeli i metody]. Moscow, Radio and Communications Publishing House, 1990, 304 p.
11. Tyugu E. Kh. Conceptual programming [Kontseptualnoe programmirovaniye]. Moscow, Nauka Publishers, 1984, 256 p.
12. Borgest N. M. Ontology of design. Theoretical foundations: Study guide. Part 1. Concepts and principles [Ontologiya proektirovaniya. Teoreticheskie osnovy: Uchebnoe posobie. Chast 1. Ponyatiya i printsipy]. Samara, Samara State Aerospace University, 2010, 91 p.
13. Ataeva O. M., Serebryakov V. A. The Basic Concepts of a Semantic Libraries Formal Model and Its Integration Process Formalization [Osnovnye ponyatiya formalnoy modeli semanticheskikh bibliotek i formalizatsiya protsessov integratsii v ney], *Software and Systems [Programmnyye produkty i sistemy]*, 2015, No. 4 (112), Pp. 180–187. DOI: 10.15827/0236-235X.112.180-187.
14. Akhmetov R. N., Vasiliev I. E., Kapitonov V. A., et al. Concept of Creation and Applying Perspective ACS Prepara-

tion and Launch of a Space Rocket «Soyuz-2»: New Approaches to Integration, Intellectualization and Management [Kontseptsiya sozdaniya i primeneniya perspektivnoy ASU podgotovki i puska rakety kosmicheskogo naznacheniya «Soyuz-2»: novye podkhody k integratsii, intellektualizatsii, upravleniyu], *Aerospace Instrument-Making [Aviakosmicheskoe priborostroenie]*, 2015, No. 4, Pp. 3–54.

15. Bakhmut A. D., Okhtilev M. Y., Okhtilev P. A. Information Resources Integration Model of the Launch Vehicle «Soyuz-2» Production Cycle on the Basis of Ontological Access to Data [Model integratsii informatsionnykh resursov proizvodstvennogo tsikla RN «Soyuz-2» na osnove ontologicheskogo dostupa k dannym], *Scientific session of GUAP: Collection of Reports. Part 2. Technical sciences [Nauchnaya sessiya GUAP: Sbornik dokladov nauchnoy sessii. Chast 2. Tekhnicheskie nauki]*. Saint Petersburg, St. Petersburg State University of Aerospace Instrumentation, 2019, Pp. 295–300.

16. Birialtsev E. V., Gusenkov A. M. Based on Ontology Approach to Integration of Relation Databases [Integratsiya relyatsionnykh baz dannykh na osnove ontologiy], *Proceedings of Kazan University. Physics and Mathematics Series [Uchenye zapiski Kazanskogo universiteta. Seriya Fiziko-matematicheskoe nauki]*, 2007, Vol. 149, Book 2, Pp. 13–34.

17. Bozhinsky I. A. Methods and Technologies of Integration of Information Systems and Distributed Databases [Metody i tekhnologii integratsii informatsionnykh sistem i raspredelennykh baz dannykh], *Radioelectronics and Informatics Journal [Radioelektronika i informatika]*, 2015, No. 2, Pp. 30–36.

18. Bova V. V. Ontological Model of Data Integration and Knowledge in Intelligent Information Systems [Ontologicheskaya model integratsii dannykh i znaniy v intellektualnykh informatsionnykh sistemakh], *Izvestiya Southern Federal University. Engineering Sciences [Izvestiya Yuzhnogo federalnogo universiteta. Tekhnicheskie nauki]*, 2015, Pp. 225–237.

19. Vovchenko A. E., Kalinichenko L. A. Analysis and Comparison of Systems for Heterogeneous Information Resources Integration [Analiz i sravnenie sistem integratsii neodnorodnykh informatsionnykh resursov], *Digital Libraries: Advanced Methods and Technologies, Digital Collections (RCDL-2008): Proceedings of the Tenth Anniversary of All-Russian Research Conference [Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kolleksii (RCDL-2008): Trudy 10-y Vserossiyskoy nauchnoy konferentsii]*, Dubna, Russia, October 07–11, 2008. Dubna, Joint Institute for Nuclear Research, Pp. 115–121.

20. Kogalovsky M. R. Methods of data integration in information systems [Metody integratsii dannykh v informatsionnykh sistemakh], 2010, 9 p. Available at: http://www.ipr-ras.ru/old_site/articles/kogalov10-05.pdf (accessed 05 Sep 2022).

21. Levashova T. V. Principles of Ontology Management Used in the Knowledge Integration Environment [Printsiipy upravleniya ontologiyami, ispolzuemye v srede integratsii znaniy], *SPIIRAS Proceedings [Trudy SPIIRAN]*, 2002, Vol. 2, Is. 1, Pp. 51–68.

22. Masloboev A. V., Shishaev M. G. Multi-Agent System of Integration of Distributed Information Resources of Innovations [Multiagentnaya sistema integratsii raspredelennykh informatsionnykh resursov innovatsiy], *Software and Systems [Programmnye produkty i sistemy]*, 2007, No. 4, Pp. 30–32.

23. Serebryakov V. A. Semantic Data Integration [Semanticheskaya integratsiya dannykh], *Proseminar of the Department of System Programming of Moscow State University [Proseminar kafedry sistemnogo programmirovaniya MGU]*, Moscow, Russia, Spring 2012. Moscow, 2012, 119 p. Available at: <http://sp.cmc.msu.ru/proseminar/2012/serebryakov.2012.04.20.pdf> (accessed 22 Mar 2018).

24. Tkachenko N. I., Spirin N. A. Application of Service-Oriented Architecture in the Integration of Process Control Systems [Primenenie servis-orientirovannoy arkhitektury pri integratsii sistem upravleniya tekhnologicheskimi protsessami], *Bulletin of the Tomsk Polytechnic University. Management, Computing and Information Science [Izvestiya Tomskogo politekhnicheskogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika]*, 2010, Vol. 317, No. 5, Pp. 61–67.

25. Chernyak L. Data Integration: Syntax and Semantics [Integratsiya dannykh: sintaksis i semantika], *Open systems. DBMS [Otkrytye sistemy. SUBD]*, 2009, No. 10, Pp. 24–29.

26. Kogalovsky M. R. Metadata in Computer Systems [Metadannye v kompyuternykh sistemakh], *Programming and Computer Software [Programmirovaniye]*, 2013, Vol. 39, No. 4, Pp. 28–46.

27. Kogalovsky M. R. Promising technologies of information systems [Perspektivnye tekhnologii informatsionnykh sistem]. Moscow, DMK Press Publishing House, 2003, 288 p.

28. Evgenov G. B. Intelligent design systems: Study guide for universities [Intellektualnye sistemy proektirovaniya: Uchebnoe posobie dlya studentov vuzov]. Moscow, Bauman Moscow State Technical University, 2009, 335 p.

29. Kogalovsky M. R. Ontology-Based Data Access Systems [Sistemy dostupa k dannym, osnovannye na ontologiyakh], *Programming and Computer Software [Programmirovaniye]*, 2012, Vol. 38, No. 4, Pp. 55–77.

30. Montiel-Sanchez C., Pfeiffer D., Becker J. The BWV-Model as Method Engineering Theory, *Proceedings of the 13th Americas Conference on Information Systems (AMCIS), Keystone, CO, USA, August 09–12, 2007*. Association for Information Systems, 2007, Art. No. 83, 10 p.

31. Tsukanova N. I. Ontological model of knowledge representation and organization: Study guide for universities [Ontologicheskaya model predstavleniya i organizatsii znaniy: Uchebnoe posobie dlya vuzov]. Moscow, Goryachaya Liniya — Telekom Publishing House, 2016, 272 p.

32. F. Baader, et al. (eds.) The Description Logic Handbook: Theory, Implementation, and Applications. Second Edition. New York, Cambridge University Press, 2007, 621 p. DOI: 10.1017/CBO9780511711787.

33. Murthy S. K. Automatic Construction of Decision Trees from Data: A Multidisciplinary Survey, *Data Mining and Knowledge Discovery*, 1998, Vol. 2, Is. 4, Pp. 345–389. DOI: 10.1023/A:1009744630224.

34. Okhtilev P. A., Bakhmut A. D., Krylov A. V., et al. Application of Decision-Making Support Technology for Management of Space Vehicle Life Cycle, *Proceedings of the II International Conference on Control in Technical Systems (CTS-2017), Saint Petersburg, Russia, October 25–27, 2017*. Institute of Electrical and Electronics Engineers, 2017, Pp. 41–44. DOI: 10.1109/CTS2017.8109483.

35. Pan J. Z. Description Logics: Reasoning Support for the Semantic Web: A Thesis for the Degree of Doctor of Philosophy. University of Manchester, 2004, 215 p.

36. Poggi A., Lembo D., Calvanese D., et al. Linking Data to Ontologies. In: S. Spaccapietra (ed.) *Journal on Data Semantics X. Lecture Notes in Computer Science. Volume 4900*. Heidelberg, Springer, 2008, Pp. 133–173. DOI: 10.1007/978-3-540-77688-8_5.

37. Díaz V. G., et al. (eds.) *Progressions and Innovations in Model-Driven Software Engineering*. Hershey (PA), IGI Global, 2013, 388 p. DOI: 10.4018/978-1-4666-4217-1.

38. Vakhitov A. R., Novoseltsev V. B. Advantages of Descriptive Logic in Knowledge Processing [Preimushchestva deskriptivnoy logiki pri obrabotke znaniy], *Bulletin of the Tomsk Polytechnic University. Management, Computing and Information Science [Izvestiya Tomskogo politekhnicheskogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika]*, 2008, Vol. 313, No. 5, Pp. 73–76.

39. Russell S., Norvig P. *Artificial Intelligence: A Modern Approach*. Second Edition [Iskusstvennyy intellekt: sovremennyy podkhod. Vtoroe izdanie]. Moscow, Williams Publishing House, 2006, 1408 p.

40. Okhtilev M. Yu., Koromyslichenko V. N., Okhtilev P. A. *Software engineering. Engineering approach: Study guide [Programmnaya inzheneriya. Inzhenernyy podkhod: Uchebnoe posobie]*. Saint Petersburg, St. Petersburg State University of Aerospace Instrumentation, 2021, 163 p.

41. Garanina N. O., Anureev I. S., Borovikova O. I. Verification Oriented Process Ontology [Ontologiya protsessov, orientirovannaya na verifikatsiyu], *Modeling and Analysis of Information Systems [Modelirovanie i analiz informatsionnykh sistem]*, 2018, Vol. 25, No. 6, Pp. 607–622. DOI: 10.18255/1818-1015-2018-6-607-622.

42. Gorshkov S. V. *Introduction to ontological modeling. Revision 2.4: Methodological guide [Vvedenie v ontologicheskoe modelirovanie. Reviziya 2.4: Metodicheskoe posobie]*. Yekaterinburg, TriniData, 2018, 150 p. Available at: <http://trinidata.ru/files/SemanticIntro.pdf> (accessed 01 Oct 2018).

43. Fedorov I. G. Adaptation of Bunge — Wand — Weber Ontology for Business Process Modeling [Adaptatsiya ontologii Bunge — Vanda — Vebera k opisaniyu ispolnyaemykh modeley biznes-protsessov], *Applied Informatics [Prikladnaya informatika]*, 2015, Vol. 10, No. 4 (58), Pp. 82–92.

44. Fedorov I. G. The Analyses of Business Process Conceptual Model Based on Bunge — Wand — Weber Ontology [Analiz kontseptualnoy modeli biznes-protsessa s ispolzovaniem ontologii Bunge — Vanda — Vebera], *Economics, Statistics and Informatics. Bulletin of Educational Methodical Association [Ekonomika, statistika i informatika. Vestnik UMO]*, 2014, No. 6, Pp. 216–221. DOI: 10.21686/2500-3925-2014-6-216-221.

45. Fedorov I. G. Modeling of business processes in BPMN2.0 notation: Monograph [Modelirovanie biznes-protsessov v notatsii BPMN2.0: Monografiya]. Moscow, Moscow State University of Economics, Statistics and Informatics, 2013, 255 p.

46. Kulikova A.A. *Methods and Means of Forming and Using Project Ontologies in the Process of Designing Automated Systems [Metody i sredstva formirovaniya i ispolzovaniya ontologiy proektov v protsesse proektirovaniya avtomatizirovannykh sistem]: diss. on competition of a scientific degree PhD (Engin.)*. Ulyanovsk, 2021, 207 p.

47. Fowler M., Sadalage P. J. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence [NoSQL: novaya metodologiya razrabotki nerelyatsionnykh baz dannykh]*. Moscow, Williams Publishing House, 2013, 192 p.

Реализация проекта по роботизации рутинных операций для повышения операционной эффективности компании

Ю. Н. Степанов, Н. В. Яковлева

Главный вычислительный центр — филиал ОАО «РЖД»
Нижний Новгород, Россия
ivc_stepanovyun@grw.rzd.ru, ivc_yakovlevanv@grw.rzd.ru

д.т.н. С. Г. Ермаков, к.т.н. Д. И. Баталов
Петербургский государственный университет
путей сообщения Императора Александра I
Санкт-Петербург, Россия
ermakov@pgups.ru, d.i.batalov@yandex.ru

Аннотация. Стратегия цифровой трансформации ОАО «РЖД» была разработана в 2019 году и включает широкий спектр цифровых технологий, которые уже внедрены или рассматриваются к возможному внедрению, в том числе автоматизацию выполнения рутинных операций — роботизированную автоматизацию процессов. В статье рассматривается хронология проекта по внедрению программных роботов в ГВЦ, цели и задачи проекта, выбранные решения и технологии, предпринятые организационные, технические и кадровые меры, достигнутые результаты и планы на будущее. Рассмотрено понятие агрессивной роботизации, направленной на планирование деятельности ГВЦ с целью увеличения объема разрабатываемых программных роботов для снижения нагрузки на персонал при выполнении рутинных операций, выделения большего времени на решение интеллектуальных задач, отладку процессов роботизации.

Ключевые слова: RPA, программные роботы, роботизированная автоматизация процессов, цифровая трансформация, операционная эффективность, Главный вычислительный центр ОАО «РЖД», агрессивная роботизация.

ВВЕДЕНИЕ

Российская транспортная отрасль демонстрирует высокие темпы цифровой трансформации. ОАО «РЖД», являясь крупнейшей государственной компанией, во многих отношениях стала лидером цифровой трансформации.

Стратегия цифровой трансформации РЖД была разработана в 2019 году. Спектр цифровых технологий, которые уже внедрены или рассматриваются к возможному внедрению, подробно расписан и очень широк:

- аналитика на базе машинного обучения;
- «интернет вещей»;
- распределенные реестры;
- управление пользовательским опытом;
- виртуальная и дополненная реальность;
- автоматическая идентификация и отслеживание объектов;
- новые интерфейсы взаимодействия с пользователем;
- речевые сервисы;
- цифровые двойники и моделирование;
- электронные площадки;
- автоматизация рутинных операций (Robotic Process Automation, RPA);
- носимые устройства;
- «большие данные»;

- цифровое моделирование (Building Information Modeling, BIM);
- автономная техника;
- предсказательная диагностика;
- процессное управление организацией (Business Process Management, BPM) [1].

Исследования анализа эффективности использования новой технологии автоматизации показали, что внедрение RPA снижает затраты на FTE (Full-Time Equivalent, Эквивалент полной занятости) на 50 процентов.

Кроме того, с точки зрения затрат, RPA дешевле, чем внедрение BPMS (Business Process Management System, Программное обеспечение для управления бизнес-процессами), так как RPA требует меньших затрат ресурсов и времени [2].

В настоящее время в России используются более 10 отечественных RPA-платформ. Составляются их ежегодные рейтинги [3].

В 2017 году технологии RPA стали распространяться в России. Участники рынка постепенно начали узнавать о возможностях RPA-систем. Проникновению технологии способствовало то, что из-за конкуренции с появившимися на рынке отечественными RPA-платформами, иностранные компании-вендоры оптимизировали ценовую политику. В результате до одного года сократился срок окупаемости внедрений RPA-систем.

В 2018 году некоторые крупные компании перешли от пилотных внедрений к полномасштабным проектам, призванным перевести рутинные операции в ведение программных роботов. Также над внедрением начали задумываться крупный бизнес, финансовый сектор и госструктуры [4].

Дальнейший активный спрос на технологию RPA во многом был обусловлен короткими сроками внедрения, очень быстрой окупаемостью и возвратом инвестиций в проект. Рынок, уставший от больших, сложных и ресурсоемких ИТ-проектов положительно воспринял технологию, где средний срок создания робота на один процесс составляет максимум 1,5 месяца, а за счет высвобождения ресурсов проект окупается в среднем за 5-8 месяцев [5].

При роботизации рабочего процесса автоматически выполняется следующий функционал [6]:

- получение доступа к почте (чтение, загрузка, отправление материалов);
- взаимодействие с корпоративными системами;

- оптическое распознавание символов (OCR);
- обмен данными между приложениями;
- выполнение задач 24 часа в сутки, по расписанию или запросу.

Одна минута работы робота в среднем соответствует 10-15 минутам работы человека. Преимущества программной роботизации перед «стандартными» ИТ-проектами:

- уменьшение ресурсных затрат;
- единый программный интерфейс;
- управление доступами к системам;
- гибкая настройка и изменение роботов.

ХРОНОЛОГИЯ ПРОЦЕССА РОБОТИЗАЦИИ ГВЦ ОАО «РЖД»

В 2018 году группа сотрудников ГВЦ ОАО «РЖД» объединилась для проведения исследовательской работы в области RPA. Члены группы, Степанов Ю. Н., Логинов А. Л., Морозов С. Д., Лелеков В. А., выполнили поиск и подбор материалов в сети Интернет, посетили информационные семинары, которые проводили разработчики платформ программной роботизации.

В 2019 году в ГВЦ был запущен процесс цифровой трансформации, в котором особое внимание уделялось разработке программных роботов. Куратором процесса роботизации был определен Смоляров Дмитрий Аркадьевич, заместитель директора ГВЦ.

Хронология процесса роботизации ГВЦ ОАО «РЖД» приведена в таблице 1.

Таблица 1

Этапы реализации проекта по внедрению программных роботов в ГВЦ ОАО «РЖД»

№ п/п	Этап проекта	Содержание работ	Дата
1.	Старт проекта внедрения технологии RPA в ГВЦ	Создание проектной группы по разработке и внедрению программных роботов. Руководитель проектной группы — Степанов Ю. Н.	4 квартал 2019 г.
2.	План-факт трансформации ГВЦ ОАО «РЖД» на 2020 г.	Выбор платформы Robin RPA	Ноябрь-декабрь 2019 г.
		Разработка сценариев программных роботов ГВЦ	2-3 квартал 2020 г.
3.	Старт проекта «Внедрение фабрики роботизации в ОАО «РЖД»	Разработка плана мероприятий	Февраль 2021 г.
		1. Создание полигонов разработки и эксплуатации программных роботов для функциональных заказчиков (филиалов ОАО «РЖД»). 2. Повышение качества оформления паспортов программных роботов (модификация программных роботов, улучшение качества).	Февраль 2021 г.
		1. Разработка временного регламента взаимодействия ИВЦ/Центров технологического сопровождения/Центров компетенций при разработке и эксплуатации программных роботов для процессов ГВЦ и функциональных заказчиков. 2. Организация ввода в эксплуатацию Оркестратора — робота, управляющего запуском и согласованием работы других программных роботов.	Март 2021 г.
		1. Разработка порядка подключения к АС технологических учетных записей. 2. Разработка карт процессов роботизации и комплектов технико-нормировочных карт. 3. Выработка единых правил разработки программных роботов.	Апрель 2021 г.
		Актуализация ПТР в связи с формированием полигонов роботизации для функциональных заказчиков	Июнь 2021 г.
		Создание механизма учета эффектов от роботов функциональных заказчиков	Сентябрь 2021 г.
		Модернизация Автоматизированной системы оформления заявок для перехода к постоянному решению по ведению учетных записей в информационных системах для подключения программных роботов	Ноябрь-декабрь 2021 г.
4.	Утилизация проекта «Внедрение фабрики роботизации в ОАО «РЖД»	Прекращение работ в рамках проекта «Внедрение фабрики роботизации в ОАО «РЖД» и передача результатов для дальнейшего сопровождения в Центр компетенций по эксплуатации платформы роботизации (ЦКЭ РПА). Обеспечено включение мероприятий в план ЦКЭ РПА на 2021 г.	Март 2021 г.
5.	Формирование ЦКЭ РПА	Единая точка входа и организатор процесса — ЦКЭ РПА	с 2021 г.
6.	Подготовка плана работ по агрессивной роботизации ОАО «РЖД»	Планирование деятельности ГВЦ с целью увеличения объема разрабатываемых программных роботов (с 1 тыс. до 10 тыс. в год)	с 2023 г.

Предпосылками и мотивацией к реализации проекта роботизации явились огромные объемы данных о работе компании и пользователей ИС, которые сотрудникам информационно-вычислительных центров (ИВЦ) ОАО «РЖД» приходилось вводить в электронные документы.

В РЖД работает около 500 информационных систем и 240 тыс. пользователей. 9 500 контрагентов ведут с ОАО «РЖД» электронный документооборот.

Технической поддержкой занимаются 4,5 тыс. ИТ-специалистов ГВЦ, расположенных в 16 ИВЦ от Хабаровска до Калининграда.

Анализ обращений пользователей показал, что более 70 % запросов — повторяющиеся и однотипные. [7]. В год обрабатывается до 7 млн запросов.

Целью проекта роботизации стало повышение операционной эффективности ОАО «РЖД» за счет увеличения

скорости работы и снижения затрат бэк-офиса в связи с роботизацией рутинных операций.

Для достижения указанной цели были определены следующие задачи [8]:

1. Сократить время формирования и заполнения внутренних документов, оптимизировать документооборот с контрагентами и запросы пользователей инфраструктуры ОАО «РЖД».

2. Создать собственную «фабрику роботов» и разработать систему оценки экономической эффективности роботизации.

3. Снизить долю рутинных и однотипных операций, обрабатываемых человеком, за счет внедрения искусственного интеллекта в операционные процессы компании.

РЕШЕНИЕ И ТЕХНОЛОГИИ ДЛЯ РЕАЛИЗАЦИИ ПРОЕКТА

В 2019 году ОАО «РЖД» провело конкурс среди отечественных разработчиков RPA-платформ.

По результатам конкурса была выбрана платформа ROBIN RPA, которая позволяет создавать роботизированные приложения под специфические задачи, решаемые информационными системами компании. Платформа Robin RPA широко использовалась в процессе цифровой трансформации и других транспортных компаний [9].

Также было принято решение о создании собственных центров компетенций для работы на платформе ROBIN RPA на базе Главного вычислительного центра и информационно-вычислительных центров компании, расположенных в 16 городах присутствия ОАО «РЖД».

Обычно RPA-платформы ограничивают пользователей по сроку действия и количеству установок, однако ОАО «РЖД» приобрело право пользования платформой ROBIN без ограничений.

Координация проекта роботизации была возложена на ООО «РЖД-Технологии».

При реализации проекта роботизации в ГВЦ ОАО «РЖД» были востребованы следующие функциональные возможности платформы ROBIN RPA:

- работа с web-страницами (сбор, наполнение, перенос контента);
- работа с приложениями Windows (запуск приложения, имитация работы пользователя);
- работа с базами данных (выполнение SQL-скриптов);
- работа с офисными приложениями (заполнение документов, сбор и сортировка информации из таблиц и текста);
- работа с файловой системой (копирование, удаление, перемещение файлов и папок);
- работа с неструктурированными текстовыми данными (классификация документов и обращений, извлечение сущностей из текста);
- работа с электронной почтой (отправка, получение писем с использованием фильтров);
- машинное зрение (поиск элемента приложения по картинке);
- преобразование данных в заданный формат.

Среди отечественных RPA-платформ ROBIN RPA отличалась следующими преимуществами:

- контроль действий робота за счет ведения записи действий;

– снижение затрат и количества ошибок в результате использования робота;

- высокая гибкость и хорошая масштабируемость;
- круглосуточная работа робота, без выходных и перекуров;
- сокращение времени выполнения задач за счет объединения данных из разрозненных систем;
- простота перенастройки робота по сравнению с переобучением человека;
- отсутствие необходимости сложной системной интеграции роботов в существующую ИТ-инфраструктуру.

ОРГАНИЗАЦИЯ ПРОЦЕССА РОБОТИЗАЦИИ

Для реализации пилотного проекта в 2019 г. в ГВЦ был выполнен комплекс организационных и технических мероприятий и кадровых решений:

1. Разработаны планы и регламенты:

- создана методика разработки и внедрения роботов и прогноза эффективности их использования на основе данных Service Desk;
- разработаны планы роботизации (планы по разработке и внедрению программных роботов в работу Центров технологического сопровождения/ИВЦ);
- разработано архитектурное решение по развертыванию комплекса с платформой роботизации;
- утверждено временное программно-техническое решение;
- разработан комплект документов по информационной безопасности;
- подготовлен временный регламент разработки и приемки программных роботов в эксплуатацию.

2. Введено проектное управление:

- созданы проектные команды;
- введен регулярный контроль результатов;
- проведены коммуникационные площадки с сотрудниками Центров технологического сопровождения/ИВЦ;
- образован мотивационный фонд (с целью поощрения проектных команд по итогу выполнения планов роботизации).

3. Внесены изменения в структуру ГВЦ:

- сформированы Центры технологического сопровождения Нового Поколения (ЦТС НП), скорректирован регламент Центров технологического сопровождения/ИВЦ (в регламенты ЦТС/ИВЦ внесены корректировки с целью формирования групп разработки и проектирования программных роботов);
- сформирован Центр компетенций по эксплуатации платформы роботизации (ЦКЭ РПА).

4. Осуществлено распространение инструментов роботизации:

- создан портал РПА (портал мониторинга программных роботов);
- внедрен робот-оркестратор (программный робот по управлению эксплуатацией — система управления роботизацией рутинных операций).
- организованы учебные курсы для разработчиков и проектировщиков программных роботов.

ПРОМЕЖУТОЧНЫЕ ИТОГИ РЕАЛИЗАЦИИ ПРОЕКТА

По итогам 2020 года в ОАО «РЖД» были подведены промежуточные итоги проекта роботизации [8]:

– 200 специалистов ОАО «РЖД» сфокусировались на работе с искусственным интеллектом и выполнении интеллектуальных задач благодаря реализации проекта роботизации;

– внедрено более 1 100 программных роботов;
– роботизировано более 1 000 рутинных операций;
– в 3-5 раз увеличилась скорость выполнения рутинных операций с повышением качества и исключением ошибок;

– порядка 600 000 запросов пользователей ИС ОАО «РЖД» обработали роботы;

– с 15 до 4 минут сократилось время обработки входящей заявки пользователя ИС в службе поддержки.

Наиболее наглядный эффект продемонстрировали следующие роботы:

1. Робот АСУ ВОП-3 «Администрирование внутренних пользователей: создание, продление и блокировка учетных записей»:

– 23 899 обращений выполнено роботом;
– роботом выполнено 88 % всего количества обращений;

– время выполнения обращений сокращено с 48 до 8 часов;

– 4 сотрудника привлечены к участию в новых проектах.

2. Робот «Заполнение данных в схеме железных дорог и междорожных стыковых пунктов приема груженых вагонов данными из ИС СИС»:

– заполнение данных полностью без участия человека: время заполнения сокращено с 3 часов до 7 минут;

– на 100 % исключено количество ошибочных данных;
– на 90 % повышена частота обновления данных.

Для функциональных заказчиков (филиалов ОАО «РЖД») были сформулированы перспективы получения эффекта от использования программных роботов:

– сокращение времени сотрудников на выполнение рутинных операций (отчеты, внесение данных в автоматизированные системы, сбор статистики, аналитика и др.);

– перевод сотрудников на выполнение более интеллектуальных задач;

– снижение количества ошибок из-за отсутствия человеческого фактора;

– увеличение скорости выполнения рутинных операций ввиду того, что робот работает быстрее человека;

– организация непрерывного бизнес-процесса для компании ввиду того, что робот не отвлекается на отпуск, больничный и т. д.;

– возможность решения новых бизнес-задач, то есть переключение внимания и сил на направления, на которые ранее не хватало времени.

ФАБРИКА РОБОТИЗАЦИИ

Успешные промежуточные итоги проекта роботизации, подведенные по итогам 2020 г., позволили перейти к реализации в 2021 г. следующего проекта, «Облачной фабрики программных роботов» для транспортно-логистической отрасли [10].

«Облачная фабрика» — это отечественная цифровая инфраструктура, на которой клиенты смогли автоматизировать свой бизнес. Речь идет об администрировании информационных систем и баз данных, электронном оформлении услуг, обработке звонков контакт-центра. Проект включил в себя платформу для эксплуатации программных роботов, модуль разработки и магазин готовых роботов, где пользователи смогли покупать программные решения для применения в рамках платформы.

На портале потенциальные заказчики из числа подразделений РЖД и коллег по транспортно-логистическому комплексу могли создать собственный кабинет для покупки и заказа роботов. Им были доступны готовые роботы или типовые блоки, из которых можно собрать робота нужной конфигурации. Там же можно было заказать оригинального программного робота. Для этого нужно было оставить заявку с приложением техзадания, которая уходила к производителям, одобренным ООО «РЖД-Технологии» для работы на платформе. После этого разработчики делали предложения, из которых заказчик выбирал подходящее [11]. При этом планграфик разработки роботов был рассчитан на срок от 25 до 41 дней (рис. 1).

В рамках проекта решались следующие организационные задачи [12]:

1. Популяризация идеи программной роботизации внутренним заказчикам (подразделениям РЖД).

2. Разработка методологии расчета экономической эффективности для обоснования целесообразности роботизации процессов.

3. Разработка и утверждение регламентов и локальных нормативных актов по роботизации.

4. Создание центра обучения по созданию и эксплуатации программных роботов.

5. Создание «Фабрики роботизации» на базе регионально распределенных подразделений ГВЦ.

6. Промышленный выпуск роботов и ускоренный их ввод в эксплуатацию.

7. Создание механизмов и системы управления эксплуатацией парка программных роботов, количественно превышающего 1 000 шт.

8. Создание методик поиска процессов для роботизации и масштабирование использования роботов во всех подразделениях РЖД.

Одной из важных технологических задач проекта являлось внедрение системы управления парком роботов, которая позволила управлять массовым распространением технологии RPA в подразделениях ГВЦ. Данную функцию выполнял модуль диспетчеризации роботов на платформе ROBIN — «ROBIN Orchestrator» и специализированный программный робот, в которых собирается вся информация о работе роботов и об АСУ, с которыми они взаимодействуют. Так же эта система позволяет контролировать работоспособность роботов, вести учет планируемых изменений в АСУ и анализировать необходимые исправления роботов при планируемых изменениях АСУ, контролировать сертификацию эксплуатируемых роботов в ГВЦ [13].

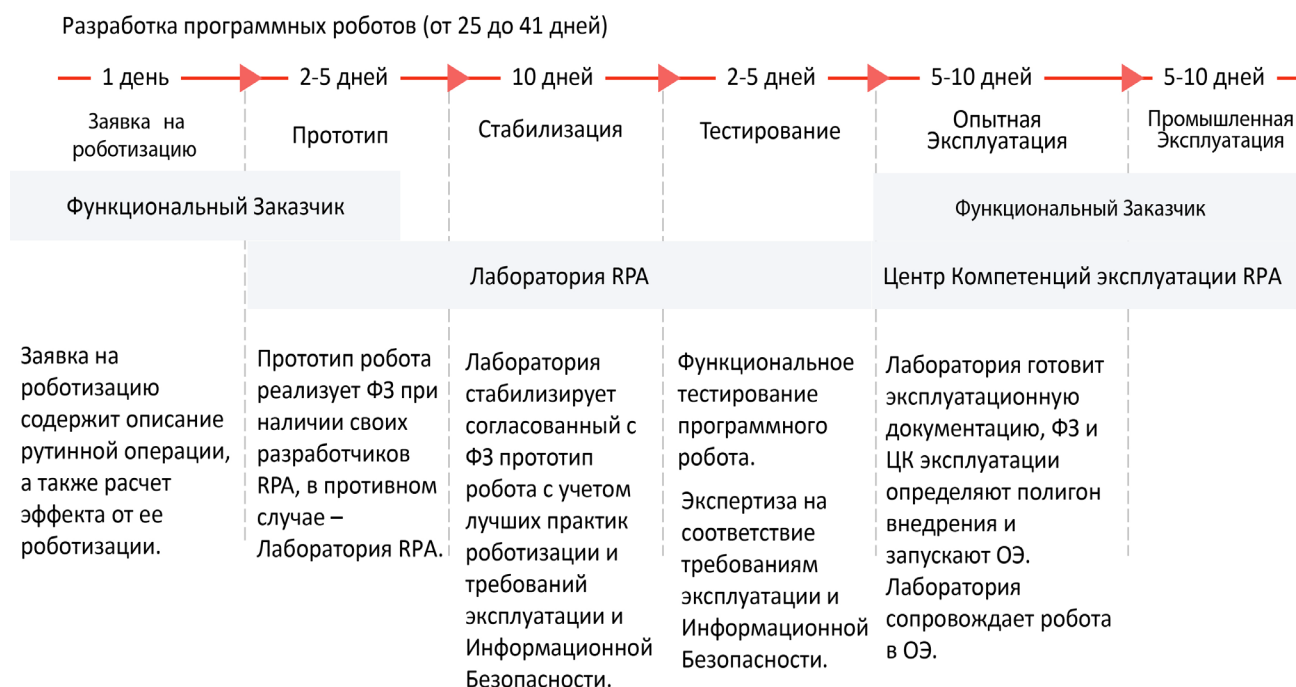


Рис. 1. План-график разработки роботов

ЦЕНТР КОМПЕТЕНЦИЙ

ПО ЭКСПЛУАТАЦИИ ПЛАТФОРМЫ РОБОТИЗАЦИИ

После завершения проекта «Облачной фабрики программных роботов» все результаты (разработанные роботы) были переданы для дальнейшего сопровождения в ЦКЭ РПА — единую точку входа и организатору процесса роботизации в ГВЦ [14].

Одним из итогов этого периода стало формирование и внедрение модели профессиональных компетенций в области RPA.

В рамках реализации проекта для развития корпоративных компетенций в Корпоративном университете РЖД прошли обучение 314 человек.

Для реализации задач импортозамещения обучено 420 человек (на программах за счет бюджета ГВЦ).

Разработаны 3 курса по RPA, по которым через систему дистанционного обучения повысили квалификацию: в ГВЦ — 1 439 человек, в подразделениях РЖД — 2 680 человек.

Прошли оценку: по корпоративным компетенциям — 605 руководителей, 3 156 специалистов, по профессиональным компетенциям — 196 руководителей, 2 331 специалист.

Заключены договоры с шестью вузами железнодорожного транспорта на предоставление доступа в Систему дистанционного обучения к курсам «ЧАС.Знаний».

АГРЕССИВНАЯ РОБОТИЗАЦИЯ ОАО «РЖД»

С 2023 года началась подготовка плана работ по агрессивной роботизации ОАО «РЖД» с целью увеличения объема разрабатываемых программных роботов (с 1 тыс. до 10 тыс. в год).

План предусматривал решение следующих задач:

1. Разгрузить персонал от рутинных операций, выделить время на решение интеллектуальных задач.

2. Наладить процессы роботизации от задумки до эксплуатации.

3. Обеспечить принятие обоснованных экономически выгодных управленческих решений о роботизации.

Для решения указанных задач предусматривалось выполнить мероприятия в нескольких разделах:

1. Организационные вопросы:

– провести расширенные ознакомительные семинары с функциональными заказчиками по технологии роботизации;

– обеспечить регулярные рассылки информационных листов, публикации в «Вестнике цифровой трансформации» с новинками в области внедрения цифровых сервисов.

2. Популяризация технологии:

– дать старт расширенному проекту роботизации;

– включить роботизацию в качестве обязательного мероприятия в программу повышения эффективности.

3. Инструменты оформления заявок на роботизацию для функциональных заказчиков:

– разработать портал роботизации (маркетплейс) для функциональных заказчиков с примерами роботов в разрезе информационных систем, презентациями новинок и достигнутых эффектов, инструментом для расчета эффекта и формирования паспорта.

С целью реализации проекта началась разработка и совершенствование укрупненного цикла роботизации.

Потребовалась синхронизация Регламента учета и эксплуатации роботов с программой цифровизации РЖД.

Для обеспечения такой увязки ГВЦ совместно с Департаментом информатизации РЖД в рамках рабочих групп по внедрению ИТ-услуг RPA был создан единый сквозной процесс — концепция «корзинок».

Был подготовлен Регламент приемки и ввода в эксплуатацию программных роботов в РЖД, который определяет

требования безопасности, лицензирования и создает базу для подачи заявок и распределения лимита финансирования Департамента информатизации РЖД.

Был внедрен режим подконтрольной эксплуатации роботов, что позволяет передавать роботов заказчику сразу после разработки и процедуры тестирования.

Таким образом, укрупненный цикл роботизации совершенствуется, готовится новая форма заявки на разработку робота для дальнейшего сокращения сроков ее рассмотрения и конечной передачи пользователю готового робота.

Использование концепции «корзинок» и методика оценки эффективности позволили создать единый сквозной процесс, имеющий следующие преимущества:

- создание и ввод в эксплуатацию роботов в течение всего года;
- интеграция в действующие регламенты более высокого уровня;

– соблюдение всех норм информационной безопасности и документирования.

Проектом предусматривается создание системы защиты информации по роботизации.

Регламентами предусмотрена необходимость системы защиты информации для каждой автоматизированной системы или автоматизированных систем управления технологическими процессами. При этом рассмотрение на соответствие регламентам и согласование ввода в эксплуатацию происходит один раз в конце года, что не оптимально.

Поэтому проектом рассматривается возможность создания единого защищенного корпоративного контура, внутри которого можно будет проводить роботизацию процессов (рис. 2).

Это даст возможность эффективно с точки зрения информационной безопасности передавать роботов в подконтрольную эксплуатацию заказчику.



Рис. 2. Защищенный контур для систем роботизации

МОДЕЛЬ ГВЦ НОВОГО ПОКОЛЕНИЯ

Реализация проектов роботизации ОАО «РЖД», которая выполнялась в основном силами работников ГВЦ, позволила с новых позиций взглянуть на возможные направления цифровой трансформации самого ГВЦ.

Были сформулированы цели этого процесса:

1. Повышение качества обслуживания пользователей ИТ-услуг. Обеспечение готовности ГВЦ к эксплуатации новых цифровых платформ.

2. Необходимость безусловного выполнения программы по повышению операционной эффективности и оптимизации расходов ОАО «РЖД» до 2025 года.

Достижение указанных целей возможно только при изменении технологического уклада ГВЦ. Направлениями изменений являются:

1. Внедрение новых инженерных технологий:
 - роботизированная автоматизация процессов (РРА);
 - искусственный интеллект (ИИ);
 - платформа управления (MAS).
2. Повышение мотивации:
 - формирование дополнительных материальных фондов ГВЦ/ИВЦ под реализацию проектов роботизации.
3. Внедрение систем управления процессами:
 - вертикализация;
 - система управления на основе ЖЦ ИС;
 - внедрение лучших практик (ITIL 4);
 - коммуникационные площадки.
4. Проведение организационных изменений:
 - формирование ЦТС нового поколения;

– создание офиса проектной координации.

В результате предусматривается новая схема взаимодействия информационных систем и пользователей ГВЦ ОАО «РЖД» (рис. 3).

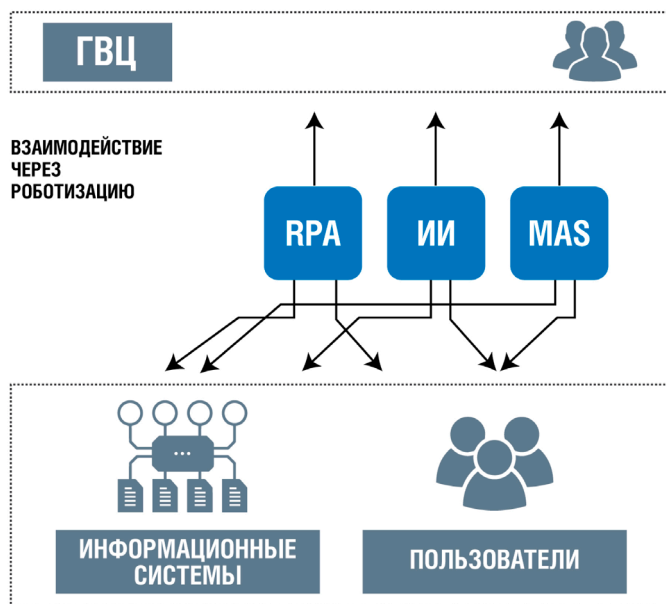


Рис. 3. Модель ГВЦ нового поколения

Были определены следующие целевые установки цифровой трансформации ГВЦ:

- обработка обращений с помощью цифровых помощников;
- сокращение времени ожидания;
- повышение производительности труда.

ЗАКЛЮЧЕНИЕ

С 2018 года в ГВЦ ОАО «РЖД» проводились исследовательские и опытно-конструкторские разработки в области автоматизации выполнения рутинных операций. К 2023 году процесс роботизации в ОАО «РЖД» с использованием RPA-технологий прошел через несколько этапов, в течение которых были реализованы несколько проектов с нарастающей сложностью поставленных целей и задач.

Для каждого из проектов был выполнен подбор и обоснование решений и технологий, предпринимались организационные, технические и кадровые меры. Ежегодно подводились итоги, анализировались достигнутые результаты и формировались планы на будущие периоды. В настоящее время ГВЦ ОАО «РЖД» реализует проект агрессивной роботизации с целью увеличения объема разрабатываемых программных роботов и приступил к реализации проекта цифровой трансформации самого ГВЦ.

В статье впервые авторами, причастными к реализации описываемых проектов, изложена хронология процесса роботизации ГВЦ ОАО «РЖД». Приведенные данные могут быть полезны крупным компаниям при реализации масштабных проектов цифровой трансформации, в том числе с использованием RPA-технологий.

ЛИТЕРАТУРА

1. Суконников, Г. В. О цифровой трансформации ОАО «РЖД» // Экономика железных дорог. 2022. № 8. С. 30–38.

2. Willcocks, L. P. The IT Function and Robotic Process Automation / L. P. Willcocks, M. Lacity, A. Craig // The Outsourcing Unit Working Research Paper Series. 2015. No. 15/05. 39 p. URL: <http://eprints.lse.ac.uk/64519> (дата обращения 11.03.2023).

3. RPA-платформы // Портал про RPA 2.0. URL: http://rpa2.ru/rpa_platformy (дата обращения 12.09.2023).

4. Почему RPA становится технологией года? // TAdviser — Государство. Бизнес. Технологии. — 2016. — 10 июня. URL: <http://www.tadviser.ru/a/460508> (дата обращения 11.03.2023).

5. Вотяков, С. «RPA Connect: Перегрузка» — Все идет по плану! / С. Вотяков, Д. Баталов // Сетевое издание «Промышленность-Сегодня.РФ». — 2022. — 10 августа. URL: <http://промышленность-сегодня.рф/articles/rpa-connect-perezagruzka-vse-idet-po-planu> (дата обращения 11.03.2023).

6. Роботизация бизнес-процессов (RPA). — KPMG International, 2022. — 13 с. URL: <http://assets.kpmg.com/content/dam/kpmg/kz/pdf/2022/02/RPA-Brochure.pdf> (дата обращения 12.09.2023).

7. Каргина, Л. А. Роль технологий RPA в цифровой трансформации ОАО «РЖД» / Л. А. Каргина, Т. В. Ионова, С. Л. Лебедева // Экономика железных дорог. 2022. № 8. С. 62–69.

8. Степанов, Ю. Н. Роботизация рутинных операций для повышения операционной эффективности компании // Роботизация бизнес-процессов для цифровой трансформации: Материалы онлайн-конференции Форума All-over-IP (17 августа 2021 г.). URL: <http://www.all-over-ip.ru/2021/rpa> (дата обращения 12.09.2023).

9. Ермаков, С. Г. Использование платформы Robin RPA в процессе цифровой трансформации транспортных компаний / С. Г. Ермаков, Д. И. Баталов, И. С. Мельников // Интеллектуальные технологии на транспорте. 2023. № 1 (33). С. 5–14. DOI: 10.24412/2413-2527-2023-133-5-14.

10. Фабрика роботов // Железнодорожный транспорт. 2021. № 2. С. 2–3.

11. Вьюгин, И. Роботы выходят в люди // Гудок. — 2021. — № 12 (27106). — 27 января.

12. Robin RPA Day // ROBIN. Robotic Intelligence. URL: <http://www.rpa-robin.ru/novosti/robinrpadaypostrls> (дата обращения 12.09.2023).

13. Фабрика роботизации // Проект Года. URL: http://globalcio.ru/projects/19799/?sphrase_id=24844 (дата обращения 12.09.2023).

14. Зубов, А. В ГВЦ тестируют новую версию программных роботов // Gudok.ru — Транспортный портал. — 2022. — 17 февраля. URL: http://www.gudok.ru/content/science_education/1595595/ (дата обращения 12.09.2023).

Implementation of a Project on Robotization of Routine Operations to Improve the Operational Efficiency of the Company

Yu. N. Stepanov, N. V. Yakovleva

The Main Computing Center
of Russian Railways JSC
Nizhny Novgorod, Russia

ivc_stepanovyun@grw.rzd.ru, ivc_yakovlevanv@grw.rzd.ru

Grand PhD S. G. Ermakov, PhD D. I. Batalov

Emperor Alexander I St. Petersburg
State Transport University
Saint Petersburg, Russia

ermakov@pgups.ru, d.i.batalov@yandex.ru

Abstract. Digital Transformation Strategy Russian Railways JSC was developed in 2019. The range of digital technologies that have already been implemented or are being considered for possible implementation is quite wide, and includes, among other things, automation of routine operations — Robotic Process Automation. The article discusses the chronology of the project on the introduction of software robots in the Main Computing Center, the goals and objectives of the project, the selected solutions and technologies, the organizational, technical and personnel measures taken, the results achieved and plans for the future. The concept of aggressive robotization is considered, aimed at planning the activities of the Main Computing Center in order to increase the volume of software robots being developed to reduce the burden on personnel when performing routine operations, allocate more time to solving intellectual tasks, debugging robotization processes.

Keywords: RPA, software robots, robotic automation of processes, digital transformation, operational efficiency, the Main Computing Center of Russian Railways JSC, aggressive robotization.

REFERENCES

1. Sukonnikov G. V. About the Digital Transformation of JSC «Russian Railways» [O tsifrovoy transformatsii OAO «RZhD»], *Railway Economics [Ekonomika zheleznikh dorog]*, 2022, No. 8, Pp. 30–38.
2. Willcocks L. P., Lacity M., Craig A. The IT Function and Robotic Process Automation, *The Outsourcing Unit Working Research Paper Series*, 2015, No. 15/05, 39 p. URL: <http://eprints.lse.ac.uk/64519> (accessed 11 Mar 2023).
3. RPA platforms [RPA-platformy], *Portal About RPA 2.0 [Portal pro RPA 2.0]*. Available at: http://rpa2.ru/rpa_platformy (accessed 12 Sep 2023).
4. Why RPA Becomes Technology of Year? [Pochemu RPA stanovitsya tekhnologiyey goda?], *TAdviser — Government. Business. IT [TAdviser — Gosudarstvo. Biznes. Tekhnologii]*. Published online at June 10, 2016. Available at: <http://www.tadviser.ru/a/460508> (accessed 11 Mar 2023).
5. Votyakov S., Batalov D. «RPA Connect: Reboot» — Everything Is Going According to Plan! [«RPA Connect: Perezagruzka» — Vse idet po planu!], *Network Publication «Industry-Today.RF» [Setevoe izdanie «Promyshlennost-Segodnya.RF»]*. Published online at August 10, 2022. Available at: <http://промышленность-сегодня.рф/articles/rpa-connect-perezagruzka-vse-idet-po-planu> (accessed 11 Mar 2023).
6. Robotic Process Automation (RPA) [Robotizatsiya biznes-protssesov (RPA)]. KPMG International, 2022, 13 p.

Available at: <http://assets.kpmg.com/content/dam/kpmg/kz/pdf/2022/02/RPA-Brochure.pdf> (accessed 12 Sep 2023).

7. Kargina L. A., Ionova T. V., Lebedeva S. L. The Role of RPA Technologies in the Digital Transformation of Russian Railways [Rol tekhnologiy RPA v tsifrovoy transformatsii OAO «RZhD»], *Railway Economics [Ekonomika zheleznikh dorog]*, 2022, No. 8, Pp. 62–69.

8. Stepanov Yu. N. Robotization of Routine Operations to Improve the Operational Efficiency of the Company [Robotizatsiya rutinnykh operatsiy dlya povysheniya operatsionnoy effektivnosti kompanii], *Robotization of Business Processes for Digital Transformation: Proceedings of the Online Conference of the All-over-IP Forum [Robotizatsiya biznes-protssesov dlya tsifrovoy transformatsii: Materialy onlayn-konferentsii Foruma All-over-IP]*, August 17, 2021. Available at: <http://www.all-over-ip.ru/2021/rpa> (accessed 12 Sep 2023).

9. Ermakov S. G., Batalov D. I., Melnikov I. S. Using the Robin RPA Platform in the Process of Digital Transformation of Transport Companies [Ispolzovanie platformy Robin RPA v protsesse tsifrovoy transformatsii transportnykh kompaniy], *Intellectual Technologies on Transport [Intellektualnye tekhnologii na transporte]*, 2023, No. 1 (33), Pp. 5–14. DOI: 10.24412/2413-2527-2023-133-5-14.

10. Robot Factory [Fabrika robotov], *Railway transport [Zheleznodorozhnyy transport]*, 2021, No. 2, Pp. 2–3.

11. Vyugin I. Robots Become People [Roboty vykhodyat v lyudi], *Gudok*, January 27, 2021, No. 12 (27106).

12. Robin RPA Day, *ROBIN. Robotic Intelligence*. Available at: <http://www.rpa-robin.ru/novosti/robinrpadaypostrls> (accessed 12 Sep 2023).

13. Robotization Factory [Fabrika robotizatsii], *Project of the Year [Proekt Goda]*. Available at: http://globalcio.ru/projects/19799/?sphrase_id=24844 (accessed 12 Sep 2023).

14. Zubov A. A New Version of Software Robots is Being Tested at the Main Computing Center [V GVTs testiruyut novuyu versiyu programmnykh robotov], *Gudok.ru*. Published online at February 17, 2022. Available at: http://www.gudok.ru/content/science_education/1595595/ (accessed 12 Sep 2023).

Моделирование и оценивание риска достижения планового среднего расстояния перевозки груза автомобильным транспортом России

д.т.н. П. В. Герасименко, д.т.н. В. А. Ходаковский

Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
pv39@mail.ru, hval104@mail.ru

Аннотация. Проведено моделирование и прогнозирование ключевого экономического показателя, определяющего стоимость доставки груза автомобильным транспортом. Используемый математический аппарат для проведения исследования позволил построить модель среднего расстояния доставки одной тонны груза и его верификацию, выполнить прогнозирование как точечное, так и интервальное, что позволило определить величину риска несвоевременной доставки груза. Для этого были использованы статистические данные по перевозке грузов автомобильным транспортом, охватывающие с 2004 по 2021 годы. Верификация осуществлена по результатам транспортировки груза в 2022 году. Математические модели были построены с помощью регрессионного анализа, реализация которого осуществлена в Excel.

Ключевые слова: моделирование, прогнозирование, оценивание, регрессия, коэффициент детерминации, точечная оценка, интервальная оценка, риск.

ВВЕДЕНИЕ

Основные грузовые перевозки в России обеспечивает автомобильный транспорт совместно с трубопроводным и железнодорожным.

Как известно, в России, несмотря на необходимость доставки грузов на дальние расстояния, из-за недоступности дорог надлежащего качества автомобильный транспорт используется мало [1]. Однако для доставки определенных товаров (например, продуктов питания) его использование является лидирующим. Так, в 2018 году объем грузоперевозок (тоннаж перевезенных грузов) в России составил 8,3 млрд тонн, при этом наибольшая доля традиционно пришла на автомобильный транспорт — 67,1 %.

ДИНАМИКА ГРУЗОБОРОТА АВТОМОБИЛЬНОГО ТРАНСПОРТА

Исследуя динамику изменения грузооборота за последние 5 допандемийных лет (2015–2019 гг.), установлено, что с каждым годом он увеличивался. Рост грузооборота на автомобильном транспорте произошел в связи с вводом в эксплуатацию большого количества новых высокоскоростных дорог. Увеличилось количество междугородных рейсов на расстояние от 300 км и более [2]. В таблице 1 представлена динамика роста суммарного объема груза, грузооборота и средней дальности доставки одной тонны груза автомобильным транспортом в РФ за период с 2004 по 2021 годы [3, 4].

Среднее расстояние (дальность) доставки одной тонны груза, представленное в таблице 1, определяется в результате деления величины грузооборота на суммарный объем груза. Оно служит в качестве одного из основных параметров работы транспорта, поэтому ее планируют в определенных условиях доставки, в предположении, что она будет с минимальным риском достигнута в реальных условиях транспортирования.

Вместе с тем реальные условия сохраняются неопределенными вплоть до доставки груза. Поэтому на этапе принятия решения на начало транспортирования субъект, принимающий его, по опыту или статистическим данным среднего расстояния устанавливает возможные варианты отклонения его от плановой величины. Приняв решение на начало перевозки, он рискует, что реальное расстояние не будет достигнуто.

Таблица 1

Основные показатели доставки груза автомобильным транспортом по годам

Год доставки груза	2004	2005	2006	2007	2008	2009	2010	2011	2012
Суммарный объем груза, млн т	6 568	6 685	6 753	6 861	6 893	5 240	5 236	5 663	5 829
Грузооборот, млрд т км	182	194	199	206	216	180	199	223	249
Средняя дальность, км	28	29	29	30	31	34	38	39	43
Год доставки груза	2013	2014	2015	2016	2017	2018	2019	2020	2021
Суммарный объем груза, млн т	5 635	5 417	5 357	5 395	5 404	5 544	5 735	5 405	5 491
Грузооборот, млрд т км	250	247	247	248	255	259	275	272	285
Средняя дальность, км	44	46	46	46	47	47	48	50	52

Как известно, риск возникает тогда, когда существует неопределенность в достижении конечной цели функционирования объекта транспортировки в реальных условиях внешней среды, которые могут быть отличными от заданных для объекта условий [5]. Для оценивания величины риска необходимо расчетный аппарат транспортировки груза рассматривать как функционирование сложной системы. В первую очередь субъектом должна быть поставлена цель, которая оценивается предметным показателем или комплексом показателей. Для ее достижения должен функционировать в условиях неопределенности объект, на начало функционирования которого должен принимать решение субъект.

Обычно цель выражается в достижении определенного состояния системы либо определенных значений параметров системы. Неопределенность является неотъемлемым атрибутом перед принятием решения субъектом на функционирование объекта. Это достаточно широкое понятие, поскольку отображает, прежде всего, объективную невозможность получения абсолютных знаний о внутренних и внешних условиях функционирования объекта во внешней среде, а также неоднозначность ее параметров. Обычно под неопределенностью понимают отсутствие полной информации о функционировании объекта, невозможность точного предсказания будущего в достижении конечной цели. Вместе с тем процесс достижения поставленной цели субъект перед принятием решения должен оценить либо на основании опыта, либо путем моделирования.

В большинстве случаев субъект, который принимает решение на функционирование объекта, не знает всей совокупности факторов, влияющих на конечную цель функционирования объекта, поэтому вынужден принимать ряд гипотез, прежде чем строить модель функционирования объекта. Именно здесь закладываются основы неопределенности в достижении конечной цели. Мету неопределенности для некоторого случайного события оценивают вероятностью безопасно достичь поставленную цель системы. Очевидно, безопасность и риск составляют полную группу событий. Следовательно, риск можно измерять вероятностью, вычисленную путем вычитания из единицы вероятности безопасного достижения цели.

Таким образом, риск можно определить, как поступок (действие) субъекта принять решение на функционирование объекта, направленное на достижение планируемой цели системы в условиях неопределенности ее достижения.

Тогда, применительно к решаемой задаче, под показателем риска недостижения среднего расстояния доставки груза автомобильным транспортом следует понимать субъективную характеристику меры отклонения планируемого расстояния от показателя, смоделированного в условиях неопределенности.

Исходя из определения риска и его показателя, методика вычисления вероятности недостижения планового показателя будет включать следующие основные этапы:

- построить по статистическим данным результатов доставки груза математическую модель целевого показателя в виде аналитической функции регрессии зависимости среднего расстояния от года транспортировки груза;
- проверить качество модели с помощью погрешностей аппроксимации, коэффициента детерминации и статистической значимости модельной функции регрессии;
- выполнить точечную и интервальную оценки прогнозной величины показателя цели, которая позволит установить левую и правую интервальную границы показателя цели;
- расположить между доверительными границами среднего расстояния плановое значение, разбив тем самым доверительный интервал на рабочую и критическую области;
- определить показатель риска недостижения планового значения.

В качестве показателя конечной цели в работе принята вероятность недостижения планового значения среднего расстояния, которая определяется путем попадания в критическую область прогнозного интервального значения величины среднего расстояния. В свою очередь, доверительный интервал среднего расстояния строится по ее модели [6].

Для определения вероятности необходимо принять закон распределения достижения величины показателя цели, например нормальный закон, как показано на рисунке 1 для плотности.

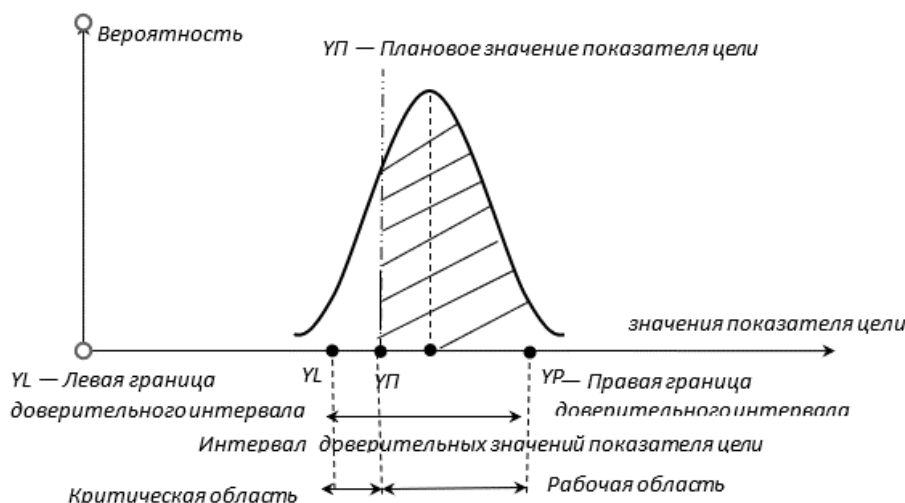


Рис. 1. Схема, поясняющая определение показателя риска

Согласно этапам исследования и по представленным в таблице 1 статистическим данным построена линейная математическая модель изменения средней дальности от года доставки груза. Аналитическая зависимость, ее график и коэффициент детерминации, которые определены по модели, представлены на рисунке 2.

Для проверки качества модели использован инструмент «Регрессия» надстройки «Пакет анализа» Excel [7].

Инструмент «Регрессия» пакета анализа данных Excel позволил по статистическим данным получить кроме значений выборочных коэффициентов линейной регрессии, корреляции и детерминации, также разложения общей суммы квадратов на объясненную и остаточную и расчетное значение *F*-статистики. Основные величины применения инструмента «Регрессия» представлены в таблице 2.



Рис. 2. График линейной аналитической зависимости и коэффициент детерминации

Таблица 2

Выборочные оценки параметров применения инструмента «Регрессия»

Коэффициенты линейной регрессии		Коэффициент корреляции	Коэффициент детерминации	Статистика Фишера
<i>a</i>	<i>b</i>	<i>R</i>	<i>R</i> ²	<i>F</i>
-2 949	1,4856	0,97	0,95	288
Суммы квадратов разностей		Общая	Факторная	Остаточная
		1 129	1 069	60

Из таблицы 2 и рисунка 2 вытекает величина коэффициента детерминации, равная 0,95. Следовательно, связь между результатом реализации модели и действительным значением доставки грузов весьма тесная и соответствует 95 %. Расчет величины статистики Фишера показывает, что ее значение, равное 288, существенно больше табличной, равной 4,5, что свидетельствует о том, что факторная дисперсия значительно превышает остаточную. Тогда можно сделать вывод, что связь между средним расстоянием доставки груза и годом транспортировки достаточно тесная [6].

На основании опытных статистических данных и результатов модельных аналитических расчетов в таблице 3 представлены абсолютные и относительные погрешности.

Анализ таблицы 3 показывает, что ошибка аппроксимации не превышает абсолютного значения среднего расстояния на величину 1 697 км, а относительную — на 10,05 %. Таким образом, можно сделать вывод, что построенная математическая модель применима для проведения прогноза, поскольку 3 качественных показателя (погрешность и коэффициент детерминации) соответствуют требованиям.

Таблица 3

Абсолютная и относительная погрешности модели

Год доставки груза	2004	2005	2006	2007	2008	2009	2010	2011	2012
Расчетные значения	26,6	28,1	29,6	31,1	32,6	34,1	35,6	37,0	38,5
Абсолютная погрешность, км	0,06	0,15	0,71	1,20	1,30	0,48	0,67	0,61	1,61
Относительная погрешность, %	0,010	0,080	1,760	5,090	5,910	0,810	1,610	1,330	9,050
Год доставки груза	2013	2014	2015	2016	2017	2018	2019	2020	2021
Расчетные значения	40,0	41,5	43,0	44,5	45,9	47,4	48,9	50,4	51,9
Абсолютная погрешность, км	1,69	1,56	1,04	0,17	0,03	1,02	1,15	0,68	0,63
Относительная погрешность, %	10,000	8,510	3,770	0,100	0,003	3,630	4,650	1,610	1,380

Как известно, на доставку перевозимых грузов и грузооборот оказывает влияние множество случайных факторов, которые вызывают разброс статистических опытных данных [8]. Этот разброс потребовал осуществить интер-

вальный прогноз в дополнение к точечному прогнозу. Результаты их представлены на рисунке 3 в виде границ доверительных интервалов среднего расстояния относительно точечного прогноза с 2022 по 2025 год.



Рис. 3. Графическое представление результатов точечного и интервального прогнозов

Под интервальным прогнозом в работе понимается множество значений среднего расстояния, которое включено в доверительный интервал возможных значений. Тем самым, в работе на примере статистических годовых данных грузооборота и суммарного объема доставленного груза построена математическая модель средней дальности доставки груза и выполнено прогнозирование как точечного, так и интервального значений среднего расстояния доставки одной тонны груза в 2022–2025 годах в РФ.

Располагая интервальным значением среднего расстояния доставки одной тонны груза, при принятом нормальном законе плотности распределения этой величины в пределах доверительного интервала и заданной плановой величине средней дальности по соотношениям теории вероятностей вычислена вероятность достижения поставленной цели, а именно вероятность превышения средней дальности доставки груза автомобильным транспортом в 2025 году.

Для случая оценки вероятности достижения средних расстояний доставки груза автомобильным транспортом в 2025 году функция плотности вероятности, распределенная по нормальному закону, примет следующие параметры: математическое ожидание, в качестве которого выступает точечный прогноз, равен $\hat{Y}_{2025} = 51,08$ км и среднеквадратическое отклонение, вычисленная как одна шестая длины интервала, $S_Y = 1\,685$ км. Тогда плотность распределения вероятности примет вид:

$$f(Y) = \frac{1}{1,685 \times \sqrt{2\pi}} \exp \left[-\frac{(Y - 51,08)^2}{2 \times 1,685^2} \right]$$

Исходя из нормальной плотности распределения среднего расстояния, легко определить вероятность превышения планового значения среднего расстояния по формуле:

$$P(Y \leq Y_{пл}) = \int_{Y_{пл}}^{Y^P} \frac{1}{1,685 \times \sqrt{2\pi}} \exp \left[-\frac{(Y - 51,08)^2}{2 \times 1,685^2} \right] dY =$$

$$= \frac{1}{1,685 \times \sqrt{2\pi}} \int_{52,2}^{53,33} \exp \left[-\frac{(Y - 51,08)^2}{2 \times 1,685^2} \right] dY = 0,26.$$

Наконец, риск недостижения поставленной целевой задачи равна единице минус вероятность превышения плана. Соответственно, показатель риска как вероятность недостижения плана (вероятность попасть в критическую область) равен 0,74. Таким образом, достижение планового среднего расстояния на 1 км в 2025 году маловероятно, если тренд изменения по годам и принятое в статье плановое значение этого показателя будут сохранены до 2025 года. Дальнейшее развитие методического аппарата целесообразно направить на его развитие применительно к перемещению грузов по международным транзитным транспортным коридорам, а также для определения соответствия моделируемого маршрута транспорта фактическому времени в пути [9, 10].

ЛИТЕРАТУРА

1. Куренков, П. В. Моделирование топологии взаимодействия элементов и потоков различных типов в системе внешнеторговых перевозок // Интеллектуальные технологии на транспорте. 2020. № 2 (22). С. 5–17.
2. Василенко, М. А. Динамика грузооборота по видам грузов и формам транспортного обслуживания / М. А. Василенко, А. М. Васильченко, С. В. Семибратова // Материалы XIII Международной студенческой научной конференции «Студенческий научный форум 2021» (Сочи, Россия, 07–10 ноября 2021 г.). URL: <http://scienceforum.ru/2021/article/2018026220> (дата обращения 11.08.2023).

3. Транспорт в России. 2022: Статистический сборник / Федеральная служба государственной статистики (Росстат). — Москва, 2022. — 101 с.

4. Транспорт в России. 2020: Статистический сборник / Федеральная служба государственной статистики (Росстат). — Москва, 2020. — 108 с.

5. Герасименко, П. В. Оценивание рисков необеспечения своевременной доставки груза железнодорожным транспортом / П. В. Герасименко, Г. Б. Титов // Проблемы экономики и управления на железнодорожном транспорте (ЭКУЖТ 2013): Материалы VIII Международной научно-практической конференции (Судак, Украина, 08–11 октября 2013 г.). — Киев: Гос. экономико-технологический ун-т транспорта, 2013. — С. 293–295.

6. Герасименко, П. В. Метод моделирования риска при повышении стоимости услуг // Известия Международной академии наук высшей школы. 2011. № 2 (56). С. 64–70.

7. Кударов, Р. С. Мониторинг пассажиропотоков, формирующих входной пассажиропоток на станции «Пушкинская» в часы пик / Р. С. Кударов, П. В. Герасименко // Шаг в будущее (Неделя науки-2006): Материалы научно-технической конференции студентов, аспирантов и молодых ученых / под ред. В. В. Сапожникова. — Санкт-Петербург: ПГУПС, 2006. — С. 189–191.

8. Гарбарук, В. В. Оценка рисков отцепки вагонов в пути следования / В. В. Гарбарук, А. Е. Красковский, В. Н. Фоменко // Интеллектуальные технологии на транспорте. 2015. № 1 (1). С. 11–17.

9. Kesting, A. Enhanced Intelligent Driver Model to Access the Impact of Driving Strategies on Traffic Capacity / A. Kesting, M. Treiber, D. Helbing // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2010. Vol. 368, Is. 1928. Pp. 4585–4605. DOI: 10.1098/rsta.2010.0084.

10. Simulation Model of Transport Traceability System for Transit Customs Goods within the EEU / D. Akhmedov, D. Yeryomin, D. Zhaxygulova, S. Trepashko // Transport and Telecommunication. 2019. Vol. 20, No. 2. Pp. 133–141. DOI: 10.2478/ttj-2019-0012.

Modeling and Assessing the Risk of Achieving the Planned Average Distance of Cargo Transportation by Road Transport in Russia

Grand PhD P. V. Gerasimenko, Grand PhD V. A. Khodakovsky
Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia
pv39@mail.ru, hval104@mail.ru

Abstract. Modeling and forecasting of a key economic indicator that determines the cost of cargo delivery by road has been carried out. The mathematical apparatus used for the study made it possible to build a model of the average delivery distance of one ton of cargo and its verification, to perform model point and interval forecasting and assessed the risk of untimely delivery of cargo. The mathematical model and forecast were made on a sample of statistical data from 2004 to 2021, the years of cargo transportation by road in the Russian Federation. Verification was carried out based on the results of cargo transportation in 2022. The model is based on regression analysis and the least squares method. The calculation apparatus is implemented using the Excel.

Keywords: modeling, forecasting, estimation, regression, coefficient of determination, point estimation, interval estimation, risk.

REFERENCES

1. Kurenkov P. V. Modeling the Topology of the Interaction of Elements and Flows of Various Types in the System of Foreign Trade Transportation [Modelirovanie topologii vzaimodeystviya elementov i potokov razlichnykh tipov v sisteme vneshnetorgovykh perevozk], *Intellectual Technologies on Transport [Intellektualnye tekhnologii na transporte]*, 2020, No. 2 (22), Pp. 5–17.

2. Vasilenko M. A., Vasilchenko A. M., Semibratova S. V. The Dynamics of Cargo Turnover by Types of Cargo and Forms of Transport Services [Dinamika gruzooborota po vidam gruzov i formam transportnogo obsluzhivaniya], *Materials of the XIII International Student Scientific Conference «Student Scientific Forum 2021» [Materialy XIII Mezhdunarodnoy studencheskoy nauchnoy konferentsii «Studencheskiy nauchnyy forum 2021»]*, Sochi, Russia, November 07–10, 2021. Available at: <http://scienceforum.ru/2021/article/2018026220> (accessed 11 Aug 2023).

3. Transport in Russia 2022: Statistical collection [Transport v Rossii. 2022: Statisticheskiy sbornik]. Moscow, Federal State Statistics Service (Rosstat), 2022, 101 p.

4. Transport in Russia 2020: Statistical collection [Transport v Rossii. 2020: Statisticheskiy sbornik]. Moscow, Federal State Statistics Service (Rosstat), 2020, 108 p.

5. Gerasimenko P. V., Titov G. B. Risk assessment of Failure to Ensure Timely Delivery of Cargo by Rail [Otse-nivanie riskov neobespecheniya svoevremennoy dostavki gruz-a zheleznodorozhnym transportom], *Problems of Economics*

and Management in Railway Transport (EKUZHT 2013): Proceedings of the VIII International Scientific and Practical Conference [Problemy ekonomiki i upravleniya na zheleznodorozhnom transporte (EKUZHT 2013): Materialy VIII Mezhdunarodnoy nauchno-prakticheskoy konferentsii], Sudak, Ukraine, October 08–11, 2013. Kyiv, State Economic and Technological University of Transport, 2013, Pp. 293–295.

6. Gerasimenko P. V. Method of risk modeling when the cost of services increases [Metod modelirovaniya riska pri povyshenii stoimosti uslug], *Proceedings of the International Higher Education Academy of Sciences [Izvestiya Mezhdunarodnoy akademii nauk vysshey shkoly]*, 2011, No. 2 (56), Pp. 64–70.

7. Kudarov R. S., Gerasimenko P. V. Monitoring of Passenger Flows Forming the Entrance Passenger Flow at Pushkinskaya Station During Rush Hours [Monitoring passazhiropotokov, formiruyushchikh vkhodnoy passazhiropotok na stantsii «Pushkinskaya» v chasy pik], *Step into the Future (Science Week–2006): Materials of the Scientific and Technical Conference of Students, Postgraduates and Young Scientists [Shag v budushchee (Nedelya nauki-2006): Materialy nauchno-tekhnicheskoy konferentsii studentov, aspirantov i molodykh uchenykh]*. Saint Petersburg, St. Petersburg State Transport University, 2006. Pp. 189–191.

8. Garbaruk V. V., Kraskovsky A. E., Fomenko V. N. Assessment of Risks Due to In-Transit Uncoupling of Cars [Otsenka riskov ottsepki vagonov v puti sledovaniya], *Intellectual Technologies on Transport [Intellektualnye tekhnologii na transporte]*, 2015, No 1 (1), Pp. 11–17.

9. Kesting A., Treiber M., Helbing D. Enhanced Intelligent Driver Model to Assess the Impact of Driving Strategies on Traffic Capacity, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2010, Vol. 368, Is. 1928, Pp. 4585–4605. DOI: 10.1098/rsta.2010.0084.

10. Akhmedov D., Yeryomin D., Zhaxygulova D., Trepashko S. Simulation Model of Transport Traceability System for Transit Customs Goods within the EEU, *Transport and Telecommunication*, 2019, Vol. 20, No. 2, Pp. 133–141.

DOI: 10.2478/tj-2019-0012.

К вопросу оценки эффективности деятельности логистических компаний: новая реальность, концептуальный подход

к.т.н. О. П. Кизляк, к.воен.н. А. И. Дергачев, к.э.н. С. А. Дергачев, Е. А. Тарбаева, к.т.н. Т. Г. Сергеева
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
kizlyako@mail.ru, d_ader@mail.ru, debug@mail.ru, elena.tarbaeva57@yandex.ru, sergeeva@pgups.ru

Аннотация. Рассматриваются возможные варианты решения проблемы повышения конкурентоспособности и привлечения новых грузовладельцев. Описана концепция при оценке экономической эффективности компании учитывать совокупные затраты, агрегируемые методами расчета приведенных затрат с учетом объемов возврата на инвестиции. Предлагаемый подход позволяет наиболее полно учитывать несколько целей. Его преимуществом является возможность оценивать эффективность деятельности компании с позиций системного подхода, основываясь на количественных характеристиках и выявленных закономерностях.

Решение данной проблемы позволит оценивать результативность логистического процесса, наметить рациональные пути снижения затрат, оценить качество функционирования компании и выявить уровень конкурентоспособности компании.

Ключевые слова: оптимизация расходов, логистические маршруты, транспортно-логистическая система, спектр услуг.

В настоящее время в мировой логистике наблюдаются серьезные помехи в работе цепей поставок. Транспортным компаниям стало предельно сложно поддерживать свою конкурентоспособность и привлекать новых грузовладельцев. В данной ситуации операторам транспортного рынка приходится менять маршруты и повышать стоимость своих услуг. Российские перевозчики тоже осознают, что рынок больше не останется прежним. В лучшем случае они смогут сохранить своих поставщиков и изменить логистические маршруты в целях оптимизации расходов на складские операции и транспортировку [1]. В худшем случае им придется сократить количество партнеров и искать новые логистические схемы для обхода санкций. Хотя потенциальный ущерб от изменения логистических цепочек очевиден, главным является вопрос, что делать в данной ситуации.

С одной стороны, современные условия конкурентного рынка транспортных услуг устанавливают новые требования к их качеству. В логистике в настоящее время происходят значительные изменения, такие как внедрение роботизации, беспилотных систем доставки, переход на экологически чистые источники энергии, автоматизация транспортных процессов, развитие новых высокотехнологичных направлений передачи данных и цифровизация рынка. Ключевыми технологиями становятся интернет вещей (IoT), распределенные реестры (Blockchain), интеллектуальные системы (AI/ML), виртуальная и дополненная реальность (VR/AR) и т. д.

Все эти инновации играют существенную роль в развитии логистики. Российские компании-операторы также активно совершенствуют свои бизнес-процессы. За последние годы география грузоперевозок значительно изменилась, и наземный транспорт стал основным драйвером роста [2].

Однако несмотря на убедительные цифры, вопрос об эффективности работы транспортно-логистической системы все еще остается открытым, а роль транспорта в логистических цепях остается ключевым фактором рынка. Это подтверждают результаты опроса руководителей предприятий, активно использующих цепи поставок, проведенного агентством РвС в начале текущего года. «Повышение эффективности» и «управление или сокращение затрат» в логистических цепях стали самыми популярными ответами [3, 4].

С другой стороны, в транспортно-логистической системе функционирует конкурентный транспортный сервис, предоставляющий потребителю свободу выбора услуг. Эта система включает в себя множество участников, которые стремятся достичь наибольшего результата и наилучшего удовлетворения интересов клиентуры [5].

Из вышесказанного следует, что в современных условиях логистические провайдеры фокусируются как на решении базовых функций, так и на повышении качества предоставляемых услуг в целом. Следовательно, перед современными бизнесами и компаниями, оказывающими услуги в сфере транспортного экспедирования, возникает сложная задача создания инновационного логистического пространства путем развития высокоэффективных цепей поставок.

При этом важно отметить, что высокое качество логистики могут обеспечить компании, специализирующиеся на предоставлении логистических услуг. Эти компании должны постоянно расширять свой спектр услуг, чтобы наилучшим образом удовлетворить потребности клиентов и выполнить основные функции. Во-первых, они должны оказывать услуги по складированию и транспортировке, для чего им необходимы склады, терминалы и транспортные средства. Во-вторых, компании должны ориентироваться на организацию логистического процесса, соответствующего потребностям каждого конкретного клиента [6, 7].

В настоящее время компании, владеющие реальными активами, уже специализируются на предоставлении складских и транспортных услуг. Другие компании расширяют свой спектр услуг, взяв на себя функции создания

интегрированных логистических цепей с помощью совершенствования логистической инфраструктуры, использования новых средств транспортировки и снижения затрат в рамках SCM-системы управления цепями поставок и т. д.

Переменные издержки в первую очередь определяются процессом грузоперевозок. Чем больше расстояние, на которое требуется перевезти груз, тем выше издержки, и наоборот, чем короче маршрут, тем ниже издержки (расходы на топливо, затраты на труд и др.). Кроме того, объем перевозимого груза является еще одним фактором, определяющим переменные издержки [8].

Общие затраты компании определяются путем суммирования постоянных и переменных издержек. Следовательно, принятие решения о совершенствовании транспортного парка и развитии логистической инфраструктуры должно основываться на строгих критериях, позволяющих провести сравнительную оценку.

В предыдущих исследованиях авторы предлагали оценивать эффективность логистической компании с помощью нескольких принципиальных соотношений:

- 1) результат, отнесенный к издержкам, или прибыль, полученная на каждый затраченный рубль;
- 2) издержки, отнесенные к результату, или удельные затраты на каждую полученную единицу результата;
- 3) результат, скорректированный на издержки, отнесенный к результату, другими словами, удельный эффект на каждую полученную единицу результата.

Некоторые специалисты рекомендуют использовать методы попарного сравнения и определение точки безубыточности для анализа. В практике также используется упрощенный способ оценки, который учитывает только логистические издержки.

Однако стоит подчеркнуть, что все эти методы требуют глубокого аналитического исследования, включая выявление конфликтующих затрат для обоснования и разработки оптимальных решений. Логистические компании стремятся устанавливать тарифы, обеспечивающие их прибыльность. Они также активно инвестируют часть прибыли в транспортную инфраструктуру и транспортные средства, чтобы сохранять свою конкурентоспособность на рынке.

Фактически экономическая эффективность логистических провайдеров может быть определена по соотношению результатов и затрат, связанных с их созданием и функционированием. Результаты могут быть оценены количественными показателями, такими как объем выполненных перевозок, или показателями качества обслуживания, например соблюдение сроков доставки груза. В современных условиях сопоставление показателей качества является предпочтительным, поскольку может отражать их конкурентоспособность.

На наш взгляд, при оценке экономической эффективности компании следует учитывать совокупные затраты, включая возврат на инвестиции. Для этой цели целесообразно использовать методологический подход, описанный в данной работе, и оценивать эффективность компании с помощью векторного критерия:

$$K(x_i) = \{k_1(x_i), k_2(x_i), \dots, k_m(x_i)\} = \{k_j(x_i), j = \overline{1, M}\},$$

где $k_j(x_i)$ — j -й частный критерий, M — число частных критериев.

Функция $K(x_i)$, учитывающая множество исходных вариантов, в свою очередь порождает отношение порядка, зависящее от смыслового содержания функции. Так, если из $K(x_i) > K(x_s)$ следует, что $x_i > x_s$, то имеет смысл критерий типа «эффект». Традиционно на транспорте такими критериями являются достигнутый уровень выполнения установленных сроков доставки грузов, показатель комплексности обслуживания клиентуры и др. Если же $x_s > x_i$, то $K(x_i)$ является критерием типа «затраты». К ним целесообразно отнести затраты на содержание производственной инфраструктуры и др.

Тогда математическая постановка задачи в векторной форме примет вид:

$$X_j^* = x_j \begin{cases} \max k_j(x_i), j = \overline{1, M}, x_i \in X \\ \min k_j(x_i), j = \overline{m+1, M}, x_i \in X \end{cases} \quad (1).$$

Из системы (1) видно, что каждое из уравнений определяет вариант решения, использование которого наиболее предпочтительно только с точки зрения одного критерия. Общее же решение представляется пересечением множеств результатов решений по каждому частному критерию:

$$X^* = X_1^* \cap X_2^* \cap \dots \cap X_M^*.$$

Оценку рассогласования критериев типа «эффект» целесообразно производить на основании безразмерной функции

$$\left(1 - \frac{k_j}{k_{j\text{эф}}}\right),$$

а для критериев типа «затраты» — в виде $(k_j/k_{jз})$.

Предлагаемый подход, на наш взгляд, позволяет наиболее полно учитывать несколько целей. Его преимуществом является возможность оценивать эффективность деятельности компании с позиций системного подхода, основываясь на количественных характеристиках и выявленных закономерностях.

ЗАКЛЮЧЕНИЕ

Текущая геополитическая обстановка, увеличение цен на топливо и недостаток квалифицированного персонала значительно усугубили неопределенность в сфере логистики и таким образом определили пути улучшения устойчивости цепей поставок при помощи оптимизации управления затратами логистическими компаниями.

Авторы предлагают инструментарий, который позволяет оценить эффективность логистического процесса, определить разумные пути сокращения затрат, оценить качество функционирования компании и выявить уровень конкурентоспособности компании. Полученные расчетные данные могут служить основой для стратегии бизнес-планирования и разработки новых цепей поставок.

ЛИТЕРАТУРА

1. Кизляк, О. П. Имитационное моделирование грузовых потоков при мультимодальных перевозках / О. П. Кизляк, Т. Г. Сергеева // IV Бетанкуровский Международный инженерный форум: Сборник трудов (Санкт-Петербург, Россия, 30 ноября–02 декабря 2022 г.). Санкт-Петербург: ПГУПС, 2022. — С. 177–179.

2. Транспорт. Официальная статистика // Федеральная служба государственной статистики (Росстат): официальный сайт. URL: <http://rosstat.gov.ru/statistics/transport> (дата обращения 01.04.2023).

3. Коровяковский, Е. К. Пути повышения эффективности рынка железнодорожных перевозок Республики Узбекистан на основе модернизации состава и структуры собственников вагонного парка и контейнеров / Е. К. Коровяковский, М. Б. Сабуров // Известия Петербургского университета путей сообщения. 2022. Т. 19, Вып. 1. С. 40–48. DOI: 10.20295/1815-588X-2022-1-40-48.

4. Никифорова, Г. И. Анализ логистического маршрута доставки груза в современных условиях // IV Бетанкуровский Международный инженерный форум: Сборник трудов (Санкт-Петербург, Россия, 30 ноября–02 декабря 2022 г.). — Санкт-Петербург: ПГУПС, 2022. — С. 255–257.

5. Кизляк, О. П. К вопросу формирования логистической инфраструктуры в сфере международного товародвижения // Развитие инфраструктуры и логистических технологий в транспортных системах (РИЛТТРАНС-2015): Сборник трудов Международной научно-практической конференции (Санкт-Петербург, Россия, 23–25 сентября 2015 г.). — Санкт-Петербург: ПГУПС, 2016. — С. 90–93.

6. Малахова, Т. А. Алгоритм оценки возможности организации мультимодальной перевозки // Управление эксплуатационной работой на транспорте (УЭРТ–2022): Сборник трудов Международной научно-практической конференции (Санкт-Петербург, Россия, 15–16 марта 2022 г.). — Санкт-Петербург: ПГУПС, 2022. — С. 262–266.

7. Анализ влияния глобальных экономических процессов на изменение спроса в международных высокоскоростных железнодорожных грузовых перевозках / В. М. Саренков, А. Ф. Колос, С. В. Шкурников, [и др.] // Проектирование развития региональной сети железных дорог. 2019. № 7. С. 49–56.

8. Юдникова, Е. С. Клиентоориентированность как важнейший фактор конкурентоспособности и экономической эффективности транспортно-логистического бизнеса / Е. С. Юдникова, А. А. Воронов // Экономика устойчивого развития. 2022. № 1 (49). С. 103–107. DOI: 10.37124/20799136_2022_1_49_103.

On the Issue of Assessing the Effectiveness of Logistics Companies: A New Reality, Conceptual Approach

PhD O. P. Kizlyak, PhD A. I. Dergachev, PhD S. A. Dergachev, E. A. Tarbaeva, PhD T. G. Sergeeva
Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia

kizlyako@mail.ru, d_ader@mail.ru, debug@mail.ru, elena.tarbaeva57@yandex.ru, sergeeva@pgups.ru

Abstract. Possible solutions to the problem of increasing competitiveness and attracting new cargo owners are being considered. The concept is described when assessing the economic efficiency of a company to take into account the total costs aggregated by the methods of calculating the applied costs, taking into account the volume of return on investment. The proposed approach allows for the most complete consideration of several objectives. Its advantage is the ability to evaluate the effectiveness of the company's activities from the standpoint of a systematic approach, based on quantitative characteristics and identified patterns.

The solution of this problem will allow us to assess the effectiveness of the logistics process, outline rational ways to reduce costs, assess the quality of the company's functioning and identify the level of competitiveness of the company.

Keywords: optimization of costs for warehouse operations and transportation, logistics routes, transport and logistics system, range of services.

REFERENCES

1. Kizlyak O. P., Sergeeva T. G. Simulation Modeling of Cargo Flows in Multimodal Transportation [Imitatsionnoe modelirovanie gruzovykh potokov pri multimodalnykh perevozkakh], *Proceedings of the IV Betancourt International Engineering Forum [IV Betankurovskiy Mezhdunarodnyy inzhenernyy forum: Sbornik trudov]*, Saint Petersburg, Russia, November 30–December 02, 2022. Saint Petersburg, St. Petersburg State Transport University, 2022, Pp. 177–179.

2. Transport. Official statistics [Transport. Ofitsialnaya statistika], *Federal State Statistics Service (Rosstat): Official website [Federalnaya sluzhba gosudarstvennoy statistiki (Rosstat): ofitsialnyy sayt]*. Available at: <http://rosstat.gov.ru/statistics/transport> (accessed 01 Apr 2023).

3. Korovyakovskiy V. K., Saburov M. B. Effectiveness Rise Ways of Railway Transportation in Uzbekistan Republic on the Basis of Owner Composition and Structure Modernization of Railcar Park and Containers [Puti povysheniya effektivnosti rynka zheleznodorozhnykh perevozkov Respubliki Uzbekistan na osnove modernizatsii sostava i struktury sobstvennikov vagonnogo parka i konteynerov], *Proceedings of Petersburg Transport University [Izvestiya Peterburgskogo universiteta putey soobshcheniya]*, 2022, Vol. 19, Is. 1, Pp. 40–48. DOI: 10.20295/1815-588X-2022-1-40-48.

4. Nikiforova G. I. Analysis of the Logistic Route of Cargo Delivery in Modern Conditions [Analiz logisticheskogo marshruta dostavki gruzov v sovremennykh usloviyakh], *Proceedings of the IV Betancourt International Engineering Forum [IV Betankurovskiy Mezhdunarodnyy inzhenernyy forum: Sbornik trudov]*, Saint Petersburg, Russia, November 30–

December 02, 2022. Saint Petersburg, St. Petersburg State Transport University, 2022, Pp. 255–257.

5. On the Issue of the Formation of a Logistics Infrastructure in the Field of International Commodity Distribution [K voprosu formirovaniya logisticheskoy infrastruktury v sfere mezhdunarodnogo tovarodvizheniya], *Development of Infrastructure and Logistics Technologies in Transport Systems (RILTTRANS-2015): Proceedings of the International Scientific and Practical Conference [Razvitie infrastruktury i logisticheskikh tekhnologiy v transportnykh sistemakh (RILTTRANS-2015): Sbornik trudov Mezhdunarodnoy nauchno-prakticheskoy konferentsii]*, Saint Petersburg, Russia, September 23–25, 2015. Saint Petersburg, St. Petersburg State Transport University, 2016, Pp. 90–93.

6. Malakhova T. A. Algorithm for Assessing the Possibility of Organizing Multimodal Transportation [Algoritm otsenki vozmozhnosti organizatsii multimodalnoy perevozki], *Management of Operational Work in Transport (UERT-2022): Proceedings of the International Scientific and Practical Conference [Upravlenie ekspluatatsionnoy rabotoy na transporte (UERT-2022): Sbornik trudov Mezhdunarodnoy nauchno-prakticheskoy konferentsii]*, Saint Petersburg, Russia, March 15–16, 2022. Saint Petersburg, St. Petersburg State Transport University, 2022, Pp. 262–266.

7. Sarenkov V. M., Kolos A. F., Shkurnikov S. V., et al. Analysis of the Influence of Global Economic Processes on Change in Demand in International High-Speed Railway Transportation [Analiz vliyaniya globalnykh ekonomicheskikh protsessov na izmenenie sprosa v mezhdunarodnykh vysokoskorostnykh zheleznodorozhnykh gruzovykh perevozkakh], *Designing the Development of a Regional Railway Network [Proektirovanie razvitiya regionalnoy seti zheleznykh dorog]*, 2019, No. 7, Pp. 49–56.

8. Yudnikova E. S., Voronov A. A. Customer Orientation as the Most Important Factor of Competitiveness and Economic Efficiency of the Transport and Logistics Business [Klientoorientirovannost kak vazhneyshiy faktor konkurentosposobnosti i ekonomicheskoy effektivnosti transportno-logisticheskogo biznesa], *Economics of Sustainable Development [Ekonomika ustoychivogo razvitiya]*, 2022, No. 1 (49), Pp. 103–107.

DOI: 10.37124/20799136_2022_1_49_103.

Задача раскраски графа в контексте оптимизации расписания: программное решение

О. И. Хонина, к.и.н. А. В. Забродин

Петербургский государственный университет путей сообщения Императора Александра I

Санкт-Петербург, Россия

olga.khonina.0311@mail.ru, teach-case@yandex.ru

Аннотация. Статья посвящена анализу проблемы раскраски графа, актуальной в различных областях, включая составление учебных расписаний, планирование и управление ресурсами и многих других. В ходе исследования эта задача была адаптирована для оптимизации расписания учебных занятий. Основной целью исследования стало создание алгоритма и программного обеспечения для разработки пользовательского интерфейса расписания в учебных учреждениях. Разработанный софт может быть легко настроен для различных типов учебных заведений: от школ до университетов.

Ключевые слова: граф, расписание, алгоритм, жадный алгоритм, задача раскраски графа, визуальное представление, программа, Python.

ВВЕДЕНИЕ

В последние десятилетия теория графов стремительно развивается, вместе с расширением ее применения в различных областях. Одной из классических задач этой теории является задача раскраски графа, где каждой вершине необходимо присвоить определенный цвет так, чтобы смежные вершины имели разные цвета, и при этом общее количество цветов было минимальным. Эта задача входит в список из 21 NP-трудных задач, сформулированных Ричардом Карпом в 1972 году, и по-прежнему остается вызовом для науки, так как нет полиномиального алгоритма для точного решения.

Теория планирования играет важную роль в обеспечении эффективного управления ресурсами, временем и задачами в различных областях. Она позволяет определять цели, разрабатывать стратегии и следить за выполнением задач. Задача раскраски графа может быть использована для оптимизации распределения ресурсов в различных процессах.

В статье внимание сосредоточено на практическом применении задачи раскраски графов в области планирования, особенно в контексте составления учебных расписаний, оптимизации транспортных маршрутов, времени работы оборудования и многих других аналогичных сценариев. Эта задача актуальна и требует эффективных алгоритмических решений.

Важным аспектом хорошего расписания является его визуальное представление. Четкий и привлекательный дизайн расписания упрощает восприятие информации, что особенно важно в быстро меняющейся и динамичной среде. Визуально привлекательное расписание способствует более эффективному использованию информации и снижает вероятность ошибок и недопонимания.

ОСНОВНЫЕ ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Большинство определений теории графов даются по-разному в различных источниках, в основном, они демонстрируют разные подходы и не противоречат друг другу [1–4].

Графом, в общем случае, называется совокупность двух множеств: $G = (V, E)$, где V – множество вершин, $E \subset V \times V$ – множество связей между ними (множество ребер) (рис. 1).

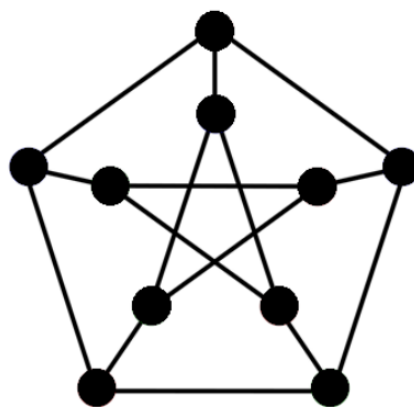


Рис. 1. Пример графа (граф Петерсена)

Две вершины $u, v \in V$ называются смежными, если они связаны ребром: $(u, v) \in E$. Два ребра $e, f \in E$ называются смежными, если они имеют общую вершину: $e \cap f \neq \emptyset$.

Степенью $\deg(v)$ вершины v в графе G называют число вершин, смежных с вершиной v : $\deg(v) = |\{u \in V \mid (v, u) \in E\}|$.

Рассмотрим подробнее теорию раскраски графа [5].

Раскраской вершин графа $G = (V, E)$ является функция $c: V \rightarrow \mathbb{N}$, которая каждой вершине графа ставит в соответствие натуральное число (цвет) так, что любым двум инцидентным вершинам $u, v \in V$ назначаются разные цвета: $\{u, v\} \in E \Rightarrow c(u) \neq c(v)$. Функция c называется функцией раскраски (рис. 2).

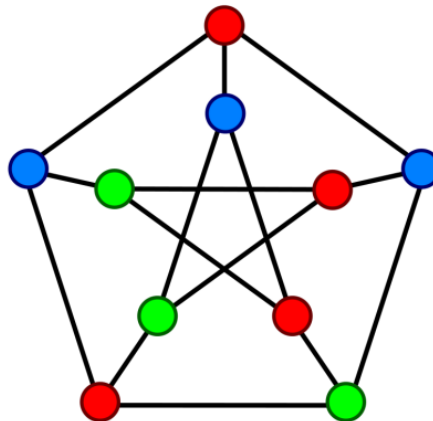


Рис. 2. Пример раскраски графа (граф Петерсена)

Применительно к статье функция раскраски имеет следующий вид: $c: \{0, 1\}^{n \times n} \rightarrow \mathbb{N}n$, то есть по матрице смежности графа мы получаем вектор его цветов.

Граф G , для которого существует раскраска из k цветов, называется k -раскрашиваемым. В этом случае функция раскраски разбивает граф G на независимые подмножества V_1, V_2, \dots, V_k , для которых справедливо: $V_i \cap V_j = \emptyset$ и $V_1 \cup V_2 \cup \dots \cup V_k = V$.

Внутри независимого множества никакие две вершины не связаны ребром.

Наименьшее число k , для которого существует k -раскраска графа G , называется хроматическим числом графа G и обозначается $\chi(G)$. Любая раскраска графа G , для которой необходимо только k цветов и $k = \chi(G)$, называется оптимальной [6].

ОПИСАНИЕ ЗАДАЧИ

Задача раскраски графа представляет собой одну из ключевых в теории графов, так как она относится к классу NP-полных задач, и многие другие задачи этой области сводятся к ней. Раскраска графа может быть вершинной, реберной или тотальной; однако все эти варианты задачи взаимосвязаны, и в рамках нашей реализации ограничимся рассмотрением вершинной раскраски. При такой раскраске графа каждой вершине присваивается определенный цвет таким образом, что никакие две смежные вершины не имеют одного и того же цвета [7, 8]. Суть самой задачи состоит в нахождении минимального количества цветов для раскраски графа.

На основе вышесказанного задачу составления расписания можно интерпретировать как задачу раскраски графа следующим образом:

- преобразование задачи расписания в графовую модель: каждому событию сопоставляется вершина, а ребра между вершинами представляют конфликты времени (события, которые не могут произойти одновременно);
- раскраска графа таким образом, чтобы две смежные вершины (события, которые не могут происходить одновременно) всегда имели разные цвета; в этом контексте «цвет» представляет собой временной слот;
- минимизация количества используемых цветов, что соответствует минимизации общего числа временных слотов.

АЛГОРИТМЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСКРАСКИ ГРАФА

Для решения задачи раскраски графа доступны разнообразные методы, которые подразделяются на точные и приближенные [9, 10].

Точные алгоритмы обеспечивают нахождение оптимального решения, однако их сложность возрастает экспоненциально по мере увеличения размера графа. Сюда входят методы перебора, ветвления и границы, а также алгоритмы, основанные на теории множеств. Однако стоит учитывать, что для больших графов и графов с высоким хроматическим числом вычислительная сложность может быть существенной, и в этом случае стоит рассмотреть использование приближенных или эвристических методов.

Приближенные алгоритмы, в свою очередь, не обещают нахождение оптимального решения, но обладают высокой скоростью работы и могут быть применены к большим графам. К ним относятся: жадные алгоритмы, генетические алгоритмы и другие подобные методы.

При решении рассматриваемой задачи возможно применение различных алгоритмов раскраски графа, однако конкретный выбор алгоритма зависит от размера графа, доступного времени и требуемой точности решения.

ПРИМЕНЕНИЕ АЛГОРИТМОВ РАСКРАСКИ ГРАФА ДЛЯ СОСТАВЛЕНИЯ РАСПИСАНИЯ

Для решения поставленной задачи будем использовать алгоритм жадной раскраски, который относится к приближенным методам.

Жадный алгоритм — это любой алгоритм, который следует эвристике решения проблемы, заключающейся в локально оптимальном выборе на каждом этапе. Во многих задачах жадная стратегия не дает оптимального решения, но жадная эвристика может дать локально оптимальные решения, которые приближаются к глобально оптимальному решению за разумный промежуток времени [11].

Жадный алгоритм для задачи раскраски графа имеет несколько преимуществ:

1. Алгоритм относительно прост в реализации и понимании. Он основан на интуитивно понятной идее выбора цвета для каждой вершины, начиная с первой и продолжая до последней.

2. Он работает быстро для большинства случаев и не требует выполнения сложных вычислений или итераций по всем вершинам графа. Вместо этого он работает локально, выбирая на каждом шаге наиболее подходящий доступный цвет для текущей вершины.

3. Жадный алгоритм обычно дает хорошие результаты раскраски графа и находит раскраску, которая использует не слишком много цветов и удовлетворяет большинству ограничений.

4. Данный алгоритм хорошо масштабируется и может быть применен к большим графам. Он не требует большого количества памяти или вычислительных ресурсов, поэтому может быть использован для решения задач раскраски графа с большим количеством вершин.

5. Жадный алгоритм может быть легко модифицирован или расширен для учета дополнительных ограничений или условий задачи раскраски графа. Это делает его удобным инструментом для решения различных вариантов задачи.

Однако следует отметить, что жадный алгоритм не всегда дает оптимальное решение и может быть подвержен ошибкам. В некоторых случаях он может привести к неправильной раскраске графа или использованию большего количества цветов, чем требуется.

Основные этапы жадного алгоритма:

1. Инициализация — присвоение всем вершинам «бесцветного» состояния.
2. Раскраска — выполнение следующих шагов для каждой вершины графа:
 - а) выбор первой доступной вершины без цвета;
 - б) присвоение вершине первого доступного цвета, который не используется ее соседями;
 - в) повторение шагов а)–б) до тех пор, пока все вершины не получат цвет.

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Разработанная программа, написанная на языке Python, предоставляет возможность чтения и визуализации расписания из файлов Excel (рис. 3), а также создания PDF-доку-

ментов с использованием скриншотов пользовательского интерфейса.

Понедель	1	Схем.	1-206	Гильванои	четная	2023-02-08
Вторник	2	Т.Вер.	7-307	Боровски;	каждая	
Среда	3	САОД	1-206	Забродин	каждая	
Четверг	1	Схем.	4-305	Гильванои	каждая	
Пятница	3	БД	1-218	Молодкин	четная	
Понедель	2	Инж.Гр.	2-314	Леонова	четная	
Понедель	3	Мат.	7-312	Рунев	каждая	
Понедель	4	Мат.	7-312	Рунев	нечетная	
Вторник	1	Т.Вер.	7-513	Боровски;	четная	
Вторник	3	БД	1-217	Хомоненн	каждая	
Понедель	4	Инж.Гр.	4-106	Леонова	четная	
Понедель	5	Т.Вер.	7-312	Боровски;	нечетная	
Понедель	2	Фил.	4-312	Мурейко	нечетная	
Среда	2	Фил.	4-312	Мурейко	каждая	
Среда	4	Т.авт.	1-204	Гильванои	каждая	
Среда	5	Схем	1-204	Гильванои	нечетная	
Четверг	2	Т.авт.	4-305	Гильванои	каждая	
Четверг	3	Физ-ра	ДФК	Сизова	каждая	
Пятница	4	БД	1-218	Молодкин	четная	

Рис. 3. Входные данные

За основу, как указывалось выше, был взят жадный алгоритм, который рассматривает занятия как вершины графа, а ограничения по времени как ребра, которые не могут иметь одинаковый цвет. В начале работы вершины сортируются по убыванию степени, и каждая вершина последовательно раскрашивается в наименьший доступный цвет, который еще не использован среди ее соседей.

Рассмотрим этапы применения алгоритма в данном контексте:

1. Цвета для ячеек выбираются из заранее заданного списка цветов (Colors).

2. Цвета циклически перебираются с помощью *itertools.cycle*. Таким образом, после использования последнего цвета из списка, алгоритм возвращается к первому цвету и начинает перебор снова.

3. Каждый учебный день недели отображается в отдельной колонке таблицы, для него алгоритм сохраняет предыдущий цвет, использованный для окрашивания ячейки с занятием.

4. При заполнении нового дня алгоритм выбирает следующий цвет из цикла. Если этот цвет уже использовался в предыдущем дне, алгоритм продолжает перебирать цвета из цикла до нахождения неиспользованного цвета.

5. Этапы 2–4 повторяются, пока все расписание не будет раскрашено.

Программа представляет собой графический интерфейс для отображения расписания, считываемого из файла Excel, и предоставляет возможность сохранения этого расписания в формате PDF.

ФУНКЦИИ ПРОГРАММЫ

Модель программы представлена на рисунке 4. Рассмотрим используемые в программе функции.

1. *Screenshot_to_pdf(window_title, filename)*. Эта функция принимает два параметра: «window_title» (заголовок окна, которое нужно «захватить» скриншотом) и «filename» (имя файла, в который нужно сохранить PDF).

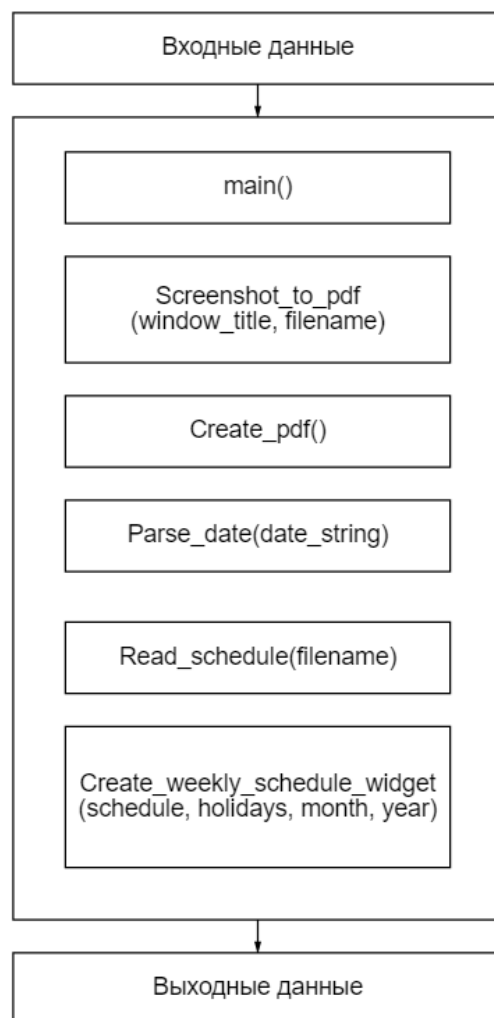


Рис. 4. Модель программы

Сначала функция использует библиотеку *pyautogui* для захвата скриншота окна. Затем скриншот обрезается для удаления оконных рамок. После этого функция создает PDF с использованием библиотеки FPDF и сохраняет его под указанным именем файла.

2. *Create_pdf()*. Данная функция вызывает функцию *screenshot_to_pdf*, передавая ей следующие параметры: название окна «Schedule» и имя файла «output.pdf». Это приводит к созданию PDF-файла с изображением содержимого окна расписания.

3. *Parse_date(date_string)*. Функция принимает строку даты (*date_string*) и преобразует ее в объект *date*, который более удобен для работы, кроме того, данная функция используется при чтении расписания из файла Excel.

4. *Read_schedule(filename)*. Функция открывает файл Excel (*filename*) по указанному пути, считывает данные и сохраняет их в словарь. Возвращает два объекта: расписание и множество дней, считающихся исключениями (праздничные дни).

5. *Create_weekly_schedule_widget(schedule, holidays, month, year)*. Функция создает виджет расписания для определенного месяца и года, используя предоставленное расписание и список выходных дней. Она воспроизводит распределение занятий по дням и времени с использованием

библиотеки *tkinter*. Каждая ячейка окрашивается в соответствующий цвет в зависимости от позиции и группы.

Код программы сначала загружает расписание из файла Excel, затем создает графический интерфейс с видом расписания и кнопкой для создания PDF. После этого программа запускает главный цикл обработки событий *tkinter*. Для реализации используются библиотеки: *tkinter* — для создания пользовательского интерфейса, *openpyxl* — для

чтения данных из Excel, *pyautogui* — для создания скриншотов и *FPDF* для создания PDF-файлов.

Реализованный в программе алгоритм обеспечивает уникальный цвет для каждого занятия в день и гарантирует, что цвета, используемые для раскраски занятий, изменяются от дня к дню. Это делает таблицу расписания более наглядной и легко читаемой.

Результат работы программы приведен на рисунке 5.

Февраль					
		Среда	Четверг	Пятница	
		1	2	3	
			Схем. (4-305, Гильванов)		
		Фил. (4-312, Мурейко)	Т.авт. (4-305, Гильванов)		
		САОД (1-206, Забродин)	Физ-ра (ДФК, Сизова)		
		Т.авт. (1-204, Гильванов)			
		Схем (1-204, Гильванов)			
Понедельник	Вторник	Среда	Четверг	Пятница	
6	7	8	9	10	
9:00-10:30	Схем. (1-206, Гильванов)	Выходной	Схем. (4-305, Гильванов)		
10:45-12:15	Инж.Гр. (2-314, Леонова)		Т.авт. (4-305, Гильванов)		
13:15-14:45	Мат. (7-312, Рунев)		БД (1-217, Хомоненко)	Физ-ра (ДФК, Сизова)	БД (1-218, Молодкин)
15:00-16:30	Инж.Гр. (4-106, Леонова)				БД (1-218, Молодкин)
16:45-18:15					
Понедельник	Вторник	Среда	Четверг	Пятница	
13	14	15	16	17	
9:00-10:30	Т.Вер. (7-513, Боровских)		Схем. (4-305, Гильванов)		
10:45-12:15	Фил. (4-312, Мурейко)	Т.Вер. (7-307, Боровских)	Фил. (4-312, Мурейко)	Т.авт. (4-305, Гильванов)	
13:15-14:45	Мат. (7-312, Рунев)	БД (1-217, Хомоненко)	САОД (1-206, Забродин)	Физ-ра (ДФК, Сизова)	
15:00-16:30	Мат. (7-312, Рунев)		Т.авт. (1-204, Гильванов)		
16:45-18:15	Т.Вер. (7-312, Боровских)	Схем (1-204, Гильванов)			
Понедельник	Вторник	Среда	Четверг	Пятница	
20	21	22	23	24	
9:00-10:30	Схем. (1-206, Гильванов)		Схем. (4-305, Гильванов)		
10:45-12:15	Инж.Гр. (2-314, Леонова)	Т.Вер. (7-307, Боровских)	Фил. (4-312, Мурейко)	Т.авт. (4-305, Гильванов)	
13:15-14:45	Мат. (7-312, Рунев)	БД (1-217, Хомоненко)	САОД (1-206, Забродин)	Физ-ра (ДФК, Сизова)	БД (1-218, Молодкин)
15:00-16:30	Инж.Гр. (4-106, Леонова)		Т.авт. (1-204, Гильванов)		БД (1-218, Молодкин)
16:45-18:15					
Понедельник	Вторник				
27	28				
9:00-10:30	Т.Вер. (7-513, Боровских)				
10:45-12:15	Фил. (4-312, Мурейко)				
13:15-14:45	Мат. (7-312, Рунев)				
15:00-16:30	Мат. (7-312, Рунев)				
16:45-18:15	Т.Вер. (7-312, Боровских)				

Рис. 5. Результат работы программы

ЗАКЛЮЧЕНИЕ

Построение расписаний стало важной частью нашей повседневной жизни. Для ускорения этого процесса требуются эффективные решения, использующие современные технологии.

В статье представлен метод составления не только эффективного, но и визуально привлекательного расписания. Правильно структурированное расписание не только оптимизирует рабочие процессы, но также способствует снижению уровня стресса у персонала. Приоритетное распределение задач и ясная организация информации упрощают

восприятие расписания, делая его более доступным и понятным. Это, в свою очередь, способствует повышению мотивации и общей производительности.

Несомненно, что теория графов предоставляет мощные инструменты для оптимизации расписания, позволяя значительно уменьшить время, необходимое для нахождения оптимального решения. Однако даже с использованием современных методов и алгоритмов задача составления расписания по-прежнему остается сложной и многогранной, требующей дальнейших исследований.

Представленная программа успешно решает проблему раскраски расписания, что делает ее идеальным инструментом для создания расписания в образовательных учреждениях. Путем внесения определенных усовершенствований в код (добавление функций) ее потенциал можно значительно расширить и адаптировать под различные специфические требования.

В будущем можно улучшить данный метод, внедрив более эффективные алгоритмы раскраски графа и адаптируя его для решения специфических задач, таких как управление многозадачностью и учет многокритериальных параметров.

ЛИТЕРАТУРА

1. Харари, Ф. Теория графов = Graph theory / Пер. с англ. В. П. Козырева; под ред. Г. П. Гаврилова. — Москва: Мир. Редакция литературы по математическим наукам, 1973. — 300 с.
2. Оре, О. Графы и их применение = Graphs and their uses / Пер. с англ. Л. И. Головиной; под ред. И. М. Яглома. — Москва: Мир. Редакция литературы по математическим наукам, 1965. — 174 с. — (Современная математика).
3. Оре, О. Теория графов = Theory of graphs / Пер. с англ. И. Н. Врублевской; под ред. Н. Н. Воробьева. — 2-е изд., стер. — Москва: Наука. Главная редакция физико-математической литературы, 1980. — 336 с.
4. Берж, К. Теория графов и ее применения = Théorie des graphes et ses applications / Пер. с фр. А. А. Зыкова; под ред. И. А. Вайнштейна. — Москва: Изд-во иностранной литературы. Редакция литературы по математическим наукам, 1962. — 319 с.
5. Карпов, Д. В. Теория графов. Глава 5. Раскраски. — 2022. — 53 с. URL: http://logic.pdmi.ras.ru/~dvk/ITMO/DM/2022-23/Лекции/5_colorings.pdf (дата обращения 16.09.2023).
6. Белов, В. В. Теория графов: Учебное пособие для втузов / В. В. Белов, Е. М. Воробьев, В. Е. Шаталов. — Москва: Высшая школа, 1976. — 392 с.
7. Татт, У. Т. Теория графов = Graph theory / Пер. с англ. Г. П. Гаврилова. — Москва: Мир. Редакция литературы по математическим наукам, 1988. — 424 с.
8. Раскраска графов // Википедия. — Обновлено 31.01.2023. URL: http://ru.wikipedia.org/wiki/Раскраска_графов (дата обращения 16.09.2023).
9. Родионов, В. В. Методы четырехцветной раскраски вершин плоских графов. — Москва: Ленанд, 2005. — 45 с.
10. Кристофидес, Н. Теория графов. Алгоритмический подход / пер. с англ. Э. В. Вершкова, И. В. Коновальцева; под ред. Г. П. Гаврилова. — Москва: Мир. Редакция литературы по математическим наукам, 1978. — 432 с.
11. Жадная раскраска // Википедия. — Обновлено 06.02.2021. URL: http://ru.wikipedia.org/wiki/Жадная_раскраска (дата обращения 16.09.2023).

Graph Coloring Problem in the Context of Schedule Optimization: Software Solution

O. I. Khonina, PhD A. V. Zabrodin

Emperor Alexander I St. Petersburg State Transport University

Saint Petersburg, Russia

olga.khonina.0311@mail.ru, teach-case@yandex.ru

Abstract. This article is dedicated to the analysis of the graph coloring problem, which is relevant in various fields, including scheduling, resource planning and management, among others. During the research, this problem was adapted for optimizing the scheduling of educational activities. The main goal of the study was to create an algorithm and software for developing a user interface for scheduling in educational institutions. The developed software can be easily customized for different types of educational establishments, from schools to universities.

Keywords: graph, schedule, algorithm, greedy algorithm, graph coloring problem, visual representation, program, Python.

REFERENCES

1. Harary F. Graph Theory [Teoriya grafov]. Moscow, Mir Publishers, 1973, 300 p.
2. Ore O. Graphs and their uses [Grafy i ikh primeneniye]. Moscow, Mir Publishers, 1965, 174 p.
3. Ore O. Theory of graphs [Teoriya grafov]. Moscow, Nauka Publishers, 1980, 336 p.
4. Berge C. Théorie des graphes et ses applications [Teoriya grafov i ee primeneniya]. Moscow, Foreign Literature Publishing House, 1962, 319 p.
5. Karpov D. V. Graph Theory. Chapter 5. Colorings [Teoriya grafov. Glava 5. Raskraski], 2022, 53 p. Available at: http://logic.pdmi.ras.ru/~dvk/ITMO/DM/2022-23/Лекции/5_colorings.pdf (accessed 16 Sep 2023).

6. Belov V. V., Vorobiev E. M., Shatalov V. E. Graph theory: Study guide for universities [Teoriya grafov: Uchebnoe posobie dlya vtuzov]. Moscow, Vysshaya Shkola Publishing House, 1976, 392 p.

7. Tutte W. T. Graph theory [Teoriya grafov]. Moscow, Mir Publishers, 1988, 424 p.

8. Graph Coloring [Raskraska grafov], *Wikipedia [Wikipedia]*. Last update at January 31, 2023. Available at: http://ru.wikipedia.org/wiki/Раскраска_графов (accessed 16 Sep 2023).

9. Rodionov V. V. Methods of four-coloring the vertices of planar graphs [Metody chetyrekhtsvetnoy raskraski vershin ploskikh grafov]. Moscow, Lenand, 2005, 45 p.

10. Christofides N. Graph theory. Algorithmic Approach [Teoriya grafov. Algoritmicheskiy podkhod]. Moscow, Mir Publishers, 1978, 432 p.

11. Greedy Coloring [Zhadnaya raskraska], *Wikipedia [Wikipedia]*. Last update at February 06, 2021. Available at: http://ru.wikipedia.org/wiki/Жадная_раскраска (accessed 16 Sep 2023).

A Biomedical Application Software for Measuring Particles Sizes in Electro-Microscopic Images

M. Sc. S. H. Al-khazraji, PhD M. M. Khalil, PhD A. I. Khalil

University of Diyala

Baqubah, Iraq

samerbaq@yahoo.com, Maadalomar@gmail.com, purecomp.adeel.khalil@uodiyala.edu.iq

Abstract. Nowadays, technological development in the field of microscopy provides an opportunity for obtaining medical images with high resolution. Therefore, more advanced image processing and analysis techniques are required for processing these images. A wide range of programs are available for scientific research. However, the key is to find an open-source application that provides more precise quantitative measurements. In scientific research, open-source software proves valuable due to its ease of modification and redistribution capabilities. The software development process continues to attract the interest of researchers in all fields, particularly in computer science field in order to implement new image processing algorithms. In this work, a new biomedical application software is presented. In this software, image processing techniques were used to separate particles from the background of microscope images. Furthermore, the software gives the researcher the ability to choose and measure nanoparticle elements size in the image. Finally, the proposed software could become a promising tool in the laboratory research compared to other used software.

Keywords: microscopy, Java software, image processing, nanoparticles, SEM.

INTRODUCTION

The introduction of digital-imaging tools in the world of science, revealed the need to use image processing techniques to analyze image data (such as biomedical image data). The goal is to use numerical methods to improve quantitative results, and accelerate repetitive routines because the numerical findings are statistically more convincing than qualitative observations. Nowadays, there is a great interest in the image processing field. Therefore, numerous specialized techniques have been developed specifically for application in biomedical images [1]. The processing of biomedical images is a branch of computer vision science. As a result, the field of computer vision encounters various challenges, including but not limited to low-light conditions and optical properties that need to be addressed. The development in biomedical image processing methods made it possible to confirm the research results by measuring them in a strict statistical way to quantitatively support their studies.

Several studies were applied to the area of biomedical image processing [2, 3]. Furthermore, there are also many books that deal with issues in image processing [4], digital microscopy and digital imaging in optical microscopy [5]. Biomedical image processing tools became one of the most popular instruments for the applications of image processing algorithms [7–9].

In this study, a new application software was introduced in order to compute the size of nanoparticles by processing microscopic images. Furthermore, the proposed software provides the ability to freely choose an area of interest to be measured. This paper is organized as follows: section 2 includes the materials and theory of the proposed software. The execution and Results

are provided in Section 3. Finally, conclusion is presented in section 4.

MATERIALS AND METHOD

SEM Imagery

Scanning electron microscopy (SEM) is an optical-based system for generating high resolution images [10], which could be used to collect information about the surface's features and structures. When an area of interest is illuminated by the imager (with a high energy beam of electrons), the electrons interact with the sample. Therefore, secondary electrons, backscattered electrons and characteristic X-rays are produced. SEM creates images that could be magnified between 10 and 1 000 000 times [11]. This imager is considered as a key tool for scientific research. It is used for monitoring metals, alloys and ceramics, as well as polymers and biological materials [12]. Figure 1 presents the SEM schematic diagram.

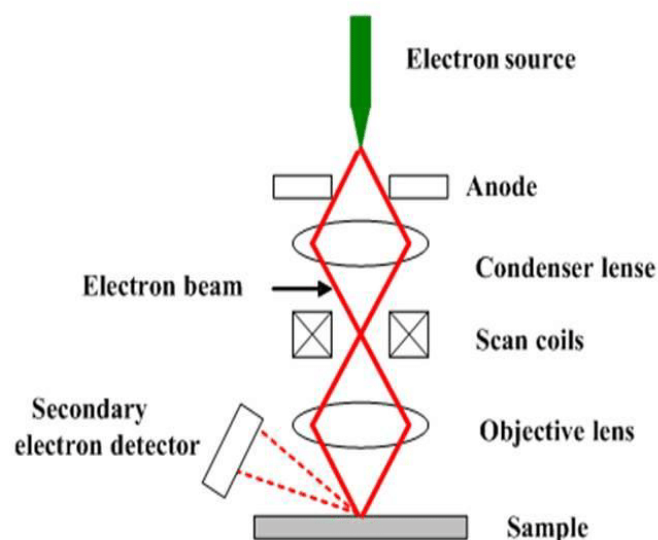


Fig. 1. The schematic diagram of scanning electron microscopy

ImageJ

ImageJ is a software that was developed using java language. The initial release of this software was made available through a collaboration between the National Institutes of Health and the Laboratory for Optical and Computational Instrumentation (LOCI, University of Wisconsin) [13]. ImageJ has been designed with an open architecture, allowing for extensibility through Java plugins and recordable macros. The software facilitates development through its built-in editor and Java compiler. Hence, users have the capability to customize

the ImageJ software to tackle diverse image processing and analysis challenges. This includes applications like three-dimensional live-cell imaging, radiological image processing, and automated haematology, among others. Therefore, these features of ImageJ made it a popular tool for studding image processing applications.

Particle

Particles refer to small, localized objects that possess certain physical or chemical attributes, including shape, volume, or density. These particles exhibit a range of sizes and quantities, spanning from subatomic particles like electrons, to microscopic particles such as atoms and molecules, to macroscopic particles like powders and granular materials. Figure 2 shows an example of SEM images of nanoparticles. Particles serve as building blocks for scientific models that represent larger objects based on their density. These models can range from studying human movement in a crowd to examining the motion of celestial bodies [14]. The term «particle» has a broad definition that varies across scientific disciplines, encompassing any entity composed of particles, thus referred to as particulate.

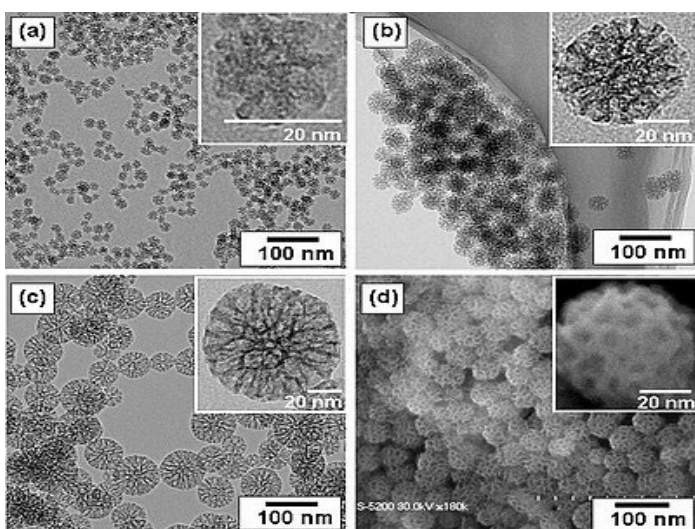


Fig. 2. An example of SEM images of nanoparticles

THE EXECUTION AND RESULTS

The block diagram of the proposed application software is shown in Figure 3.

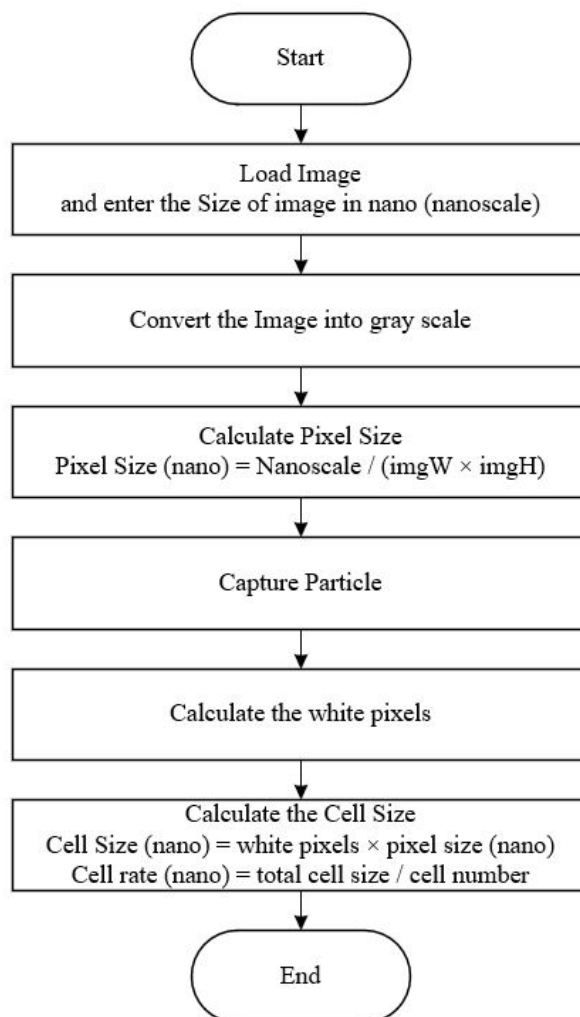


Fig. 3. Flowchart of the proposed application software

The user interface consists of five main action buttons that are aligned in the left side of the interface as shown in Figure 4.



Fig. 4. User interface

First step, the microscope image is loaded using the load image button, where a popup window will appear to help search and load the desired file. If the input image is colored, the program provides a function for converting the colored image to a gray-scale image because it is more efficient to determine the edges of the particles in the image. Once the desired image is loaded, the user can convert the image to a gray-scale image by using the convert button as shown in Figure 5. Weighted method was used in order to convert the colored image to gray-scale as shown in the code below:

```
for (int i = 0; i < img. get Height (); i++) {
for (int j = 0; j < img. get Width (); j++) {
Color c = new Color (result.getRGB(j, i));
int red = (int) (c.getRed() * 0.299);
int green = (int) (c.getGreen() * 0.587);
int blue = (int) (c.getBlue() * 0.114);
int graycolor=red + green + blue;
Color newColor = new Color (graycolor, graycolor, graycolor);
img.setRGB(j, i, newColor.getRGB());
}
```

Our proposed software provides an option to freely highlight an area of particles to be examined as shown in Figure 6.

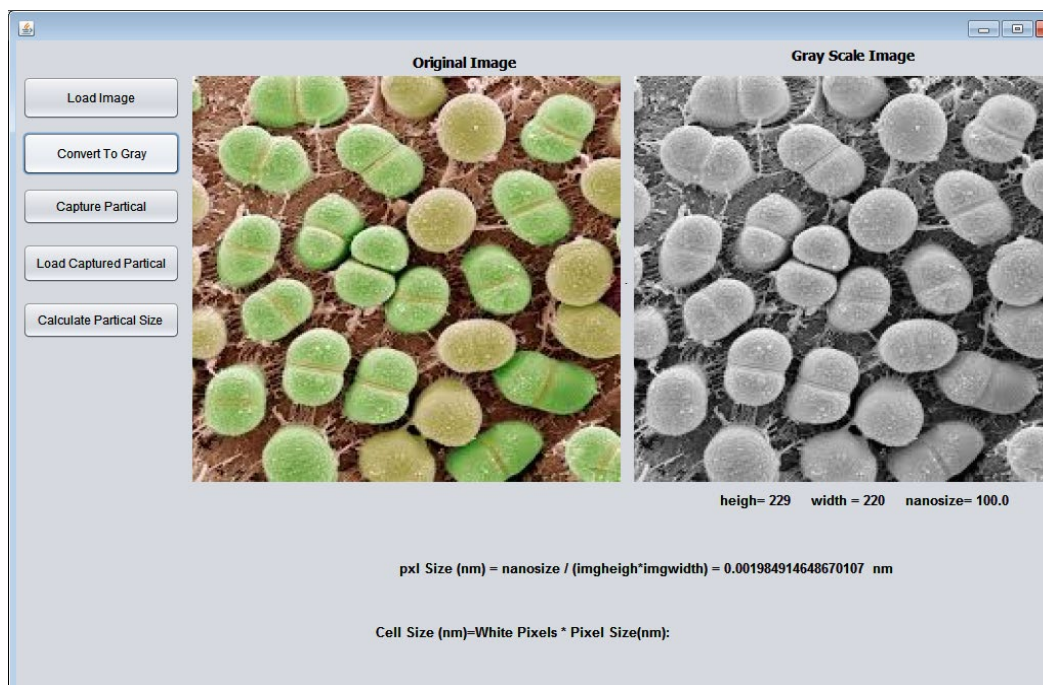


Fig. 5. Converting image to gray-scale image

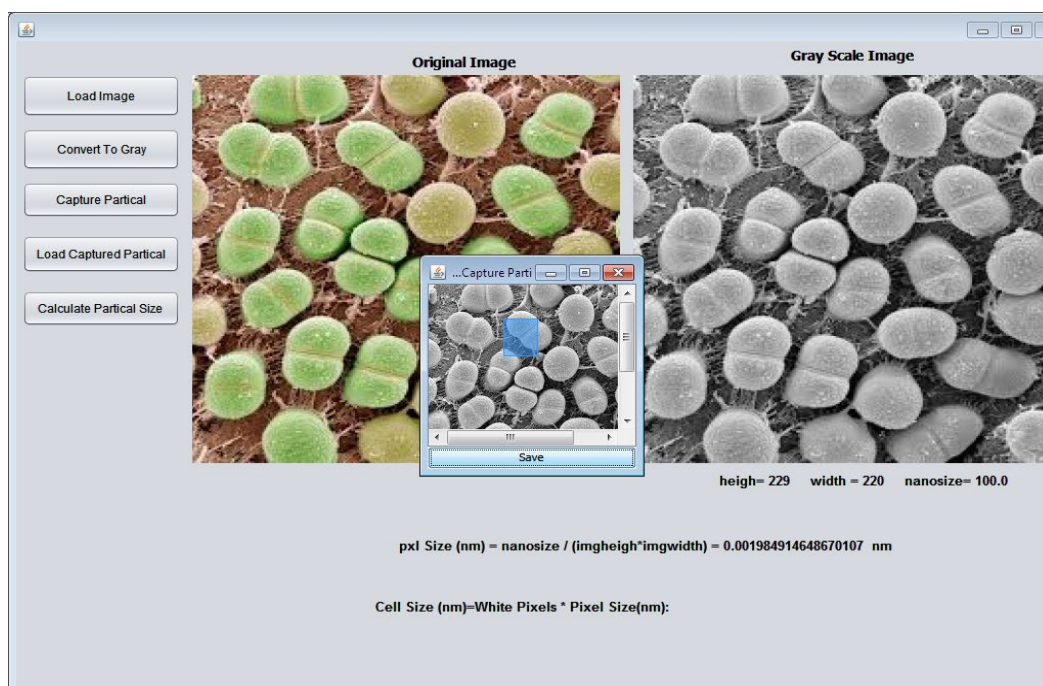


Fig. 6. Choosing an area of particles

After capturing the particle, the white pixels are calculated, and measures are given in nanometers as shown in Figure 7. Furthermore, Table 1 presents the results of 10 measures of different areas in the same image. The results show that our proposed software provides high accuracy results.

The system performs the following steps in order to calculate the nanoparticles size of microscopy images:

1. Calculate the pixel size in nanometers as follows:

If $X = \text{No. of pixels in } Y$ (Y is nanometers in scale bar), then size of 1 pixel (in nanometers): $Z = Y / X$.

2. Calculate the black/white pixels in the particle under investigation.

3. Find the particle's size as follows:

Particle's size (nano) = pixel's size $Z \times N$ (No. of pixels).

4. Find the Average particle's size in the image as follow:

Average particle's size (nano) = total No. of pixels in all measured particles / No. of particles measured.

Table 1

Measures of 10 different areas in the same image

Number of area	Particle size, nm
1	0,0132
2	0,0131
3	0,0140
4	0,0131
5	0,0133
6	0,0134
7	0,0132
8	0,0133
9	0,0135
10	0,0142

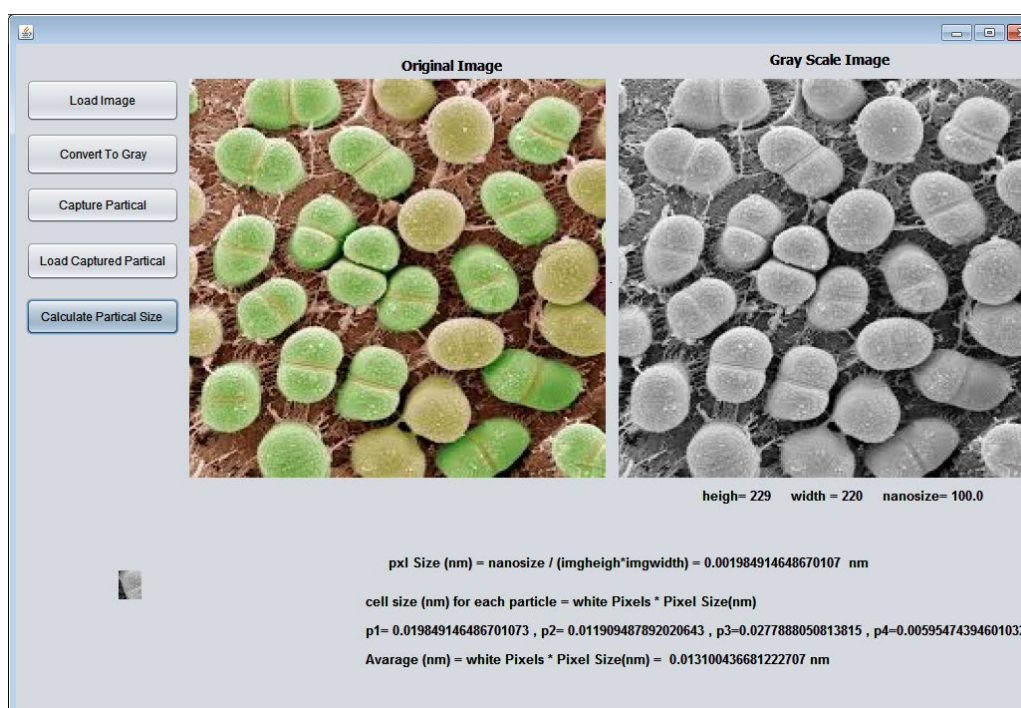


Fig. 7. Measuring nanoparticles size

Furthermore, it is worth mentioning that a pre-processing step was applied to the gray-scale image to remove any noise. Therefore, a non-linear filter (median filter) was used to improve the results of image processing. In fact, the median filter is an appropriate filter for such situations because it preserves edges and removes noise from the image.

CONCLUSIONS

From what has been discussed above, the proposed software provides more freedom in choosing the particles to be examined, collecting sizes of many selected particles from one image or many different images. Furthermore, the program is very simple and easy to use where any researcher can use it with ease and without necessarily having a computer specialist nearby to do the job. The user can examine and calculate as many particles as needed and is not limited to a fixed number of particles as in other available software.

REFERENCES

1. Castleman K. R. Digital image processing. Englewood Cliffs (NJ), Prentice Hall, 1996, 685 p.
2. Eliceiri K. W., Berthold M. R., Goldberg I. G., et al. Biological Imaging Software Tools, *Nature Methods*, 2012, Vol. 9, No. 7, Pp. 697–710. DOI: 10.1038/nmeth.2084.
3. Meijering E., Smal I., Danuser G. Tracking in Molecular Bioimaging, *IEEE Signal Processing Magazine*, 2006, Vol. 23, No. 3, Pp.46–53. DOI: 10.1109/MSP.2006.1628877.
4. Inoué S. Video Image Processing Greatly Enhances Contrast, Quality, and Speed in Polarization-Based Microscopy, *Journal of Cell Biology*, 1981, Vol. 89, Is. 2, Pp. 346–356. DOI: 10.1083/jcb.89.2.346.
5. Sluder G., Wolf D. E. (eds.) Digital Microscopy. Fourth Edition. New York, Academic Press, 2013, 675 p.

6. Konstantinides K., Rasure J. R. The Khoros Software Development Environment for Image and Signal Processing Image Processing, *IEEE Transactions on Image Processing*, 1994, Vol. 3, No. 3, Pp. 243–252. DOI: 10.1109/83.287018.

7. Burger W., Burge M. J. Principles of Digital Image Processing: Core Algorithms. London, Springer-Verlag, 2009, 344 p. DOI: 10.1007/978-1-84800-195-4.

8. Forster B., Van De Ville D., Berent J., et al. Complex Wavelets for Extended Depth-of-Field: A New Method for the Fusion of Multichannel Microscopy Images, *Microscopy Research and Technique*, 2004, Vol. 65, Is. 1-2, Pp. 33–42. DOI: 10.1002/jemt.20092.

9. Delgado-Gonzalo R., Thevenaz P., Seelamantula C. S., Unser M. Snakes with an Ellipse-Reproducing Property. Image Processing, *IEEE Transactions on Image Processing*, 2012, Vol. 21, No. 3, Pp. 1258–1271. DOI: 10.1109/TIP.2011.2169975.

10. Burger W., Burge M. J. Digital Image Processing: An Algorithmic Introduction Using Java. London, Springer-Verlag, 2008, 585 p. DOI: 10.1007/978-1-84628-968-2.

11. Gering E., Atkinson C. T. A Rapid Method for Counting Nucleated Erythrocytes on Stained Blood Smears by Digital Image Analysis, *Journal of Parasitology*, 2004, Vol. 90, No. 4, Pp. 879–881. DOI: 10.1645/GE-222R.

12. Melanthota S. K., Gopal D., Chakrabarti S., et al. Deep Learning-Based Image Processing in Optical Microscopy, *Biophysical Reviews*, 2022, Vol. 14, Is. 2, Pp. 463–481. DOI: 10.1007/s12551-022-00949-3.

13. Ye G., Kaya M. Automated Cell Foreground–Background Segmentation with Phase-Contrast Microscopy Images: An Alternative to Machine Learning Segmentation Methods with Small-Scale Data, *Bioengineering*, 2022, Vol. 9, Is. 2, Art. No. 81, 16 p. DOI: 10.3390/bioengineering 9020081.

14. Hurovitz S., Chan D., Ramele R., Gambini J. Object Detection and Statistical Analysis of Microscopy Image Sequences, *Electronic Letters on Computer Vision and Image Analysis*, 2022, Vol. 21, Is. 1, Pp. 47–58. DOI: 10.5565/rev/elcvia.1482.

Биомедицинское прикладное программное обеспечение для измерения размеров частиц на электронных микроскопических изображениях

М. Sc. С. Х. Аль-Хазраджи, PhD М. М. Халил, PhD А. И. Халил

Университет Диялы

Баакуба, Ирак

samerbaq@yahoo.com, Maadalomar@gmail.com, purecomp.adeel.khalil@uodiyala.edu.iq

Аннотация. В настоящее время технологическое развитие в области микроскопии предоставляет возможность получения медицинских изображений с высоким разрешением. Поэтому для обработки этих изображений требуются более совершенные методы обработки и анализа изображений. Для проведения научных исследований имеется широкий спектр программ. Однако ключом является поиск приложения с открытым исходным кодом, которое обеспечивает более точные количественные измерения. В научных исследованиях программное обеспечение с открытым исходным кодом оказывается ценным благодаря простоте его модификации и возможностей перераспределения. Процесс разработки программного обеспечения продолжает привлекать внимание исследователей во всех областях, особенно в области информатики, в целях реализации новых алгоритмов обработки изображений. В работе представлено новое прикладное программное обеспечение для биомедицинских исследований, в котором методы обработки изображений использовались для отделения частиц от фона изображений микроскопа. Кроме того, программное обеспечение дает исследователю возможность выбирать и измерять размер элементов наночастиц в изображении. И наконец, предлагаемое программное обеспечение может стать перспективным инструментом лабораторных исследований по сравнению с другим используемым программным обеспечением.

Ключевые слова: микроскопия, программное обеспечение Java, обработка изображений, наночастицы, SEM.

ЛИТЕРАТУРА

1. Castleman, K. R. Digital image processing. — Englewood Cliffs (NJ): Prentice Hall, 1996. — 685 p.
2. Biological Imaging Software Tools / K. W. Eliceiri, M. R. Berthold, I. G. Goldberg, [et al.] // Nature Methods. 2012. Vol. 9, No. 7. Pp. 697–710. DOI: 10.1038/nmeth.2084.
3. Meijering, E. Tracking in Molecular Bioimaging / E. Meijering, I. Smal, G. Danuser // IEEE Signal Processing Magazine. 2006. Vol. 23, No. 3. Pp.46–53. DOI: 10.1109/MSP.2006.1628877.
4. Inoué, S. Video Image Processing Greatly Enhances Contrast, Quality, and Speed in Polarization-Based Microscopy // Journal of Cell Biology. 1981. Vol. 89, Is. 2. Pp. 346–356. DOI: 10.1083/jcb.89.2.346.
5. Digital Microscopy. Fourth Edition / G. Sluder, D. E. Wolf (eds.). — New York: Academic Press, 2013. — 675 p. — (Methods in Cell Biology, Vol. 114).

6. Konstantinides, K. The Khoros Software Development Environment for Image and Signal Processing Image Processing / K. Konstantinides, J. R. Rasure // IEEE Transactions on Image Processing. 1994. Vol. 3, No. 3. Pp. 243–252. DOI: 10.1109/83.287018.

7. Burger, W. Principles of Digital Image Processing: Core Algorithms / W. Burger, M. J. Burge. — London: Springer-Verlag, 2009. — 344 p. — (Undergraduate Topics in Computer Science). DOI: 10.1007/978-1-84800-195-4.

8. Complex Wavelets for Extended Depth-of-Field: A New Method for the Fusion of Multichannel Microscopy Images / B. Forster, D. Van De Ville, J. Berent, [et al.] // Microscopy Research and Technique. 2004. Vol. 65, Is. 1-2. Pp. 33–42. DOI: 10.1002/jemt.20092.

9. Snakes with an Ellipse-Reproducing Property. Image Processing / R. Delgado-Gonzalo, P. Thevenaz, C. S. Seelamantula, M. Unser // IEEE Transactions on Image Processing. 2012. Vol. 21, No. 3. Pp. 1258–1271. DOI: 10.1109/TIP.2011.2169975.

10. Burger, W. Digital Image Processing: An Algorithmic Introduction Using Java / W. Burger, M. J. Burge. — London: Springer-Verlag, 2008. — 585 p. — (Texts in Computer Science). DOI: 10.1007/978-1-84628-968-2.

11. Gering, E. A Rapid Method for Counting Nucleated Erythrocytes on Stained Blood Smears by Digital Image Analysis / E. Gering, C. T. Atkinson // Journal of Parasitology. 2004. Vol. 90, No. 4. Pp. 879–881. DOI: 10.1645/GE-222R.

12. Deep Learning-Based Image Processing in Optical Microscopy / S. K. Melanthota, D. Gopal, S. Chakrabarti, [et al.] // Biophysical Reviews. 2022. Vol. 14, Is. 2. Pp. 463–481. DOI: 10.1007/s12551-022-00949-3.

13. Ye, G. Automated Cell Foreground-Background Segmentation with Phase-Contrast Microscopy Images: An Alternative to Machine Learning Segmentation Methods with Small-Scale Data / G. Ye, M. Kaya // Bioengineering. 2022. Vol. 9, Is. 2. Art. No. 81. 16 p. DOI: 10.3390/bioengineering 9020081.

14. Object Detection and Statistical Analysis of Microscopy Image Sequences / S. Hurovitz, D. Chan, R. Ramele, J. Gambini // Electronic Letters on Computer Vision and Image Analysis. 2022. Vol. 21, Is. 1. Pp. 47–58. DOI: 10.5565/rev/elecvia.1482.

Новые горизонты применения в теории надежности высоконадежных систем гипотезы Н. М. Седякина

д.т.н. В. А. Смагин, к.т.н. В. С. Солдатенко
Военно-космическая академия имени А. Ф. Можайского
Санкт-Петербург, Россия
va_smagin@mail.ru, soldatenko_vs@mail.ru

Аннотация. Сформулированы и подтверждены теоретическими соотношениями предложения по ослаблению ряда допущений, на основе которых сформулирована гипотеза Н. М. Седякина. Предложены направления адекватного использования данного научного положения по отношению к новым, ранее не рассматриваемым объектам. К ним относятся информационные и эргатические системы, программное обеспечение вычислительных комплексов. Описанные подходы позволяют обосновать результаты ускоренных испытаний на надежность достаточно широкого круга высоконадежных объектов.

Ключевые слова: надежность объектов, физические принципы надежности, ускоренные испытания, интенсивность отказов, ресурс надежности.

ВВЕДЕНИЕ

В теории надежности технических объектов важное место заслуженно занимает место гипотеза Н. М. Седякина [1]. Внимание к данному научному положению и возможностям его практического применения не ослабевает как в нашей стране [2–7], так и за рубежом [8–10]. Это обусловлено ясным физическим смыслом, достаточной простотой и логической завершенностью полученных теоретических результатов и возможностью их экспериментальной проверки. Интерес к указанной гипотезе возрос в последние десятилетия. Это обусловлено двумя основными причинами. Во-первых, современный уровень развития техники и технологий позволяет достигнуть значения интенсивности $\lambda(t)$ отказов элементов радиоэлектронной аппаратуры, близкие к 10^{-9} час⁻¹, а ресурсов электронных объектов — к десяткам тысяч часов. В этих условиях за практически приемлемое время экспериментально убедиться в обеспечении задаваемых в тактико-техническом задании (ТТЗ) или техническом задании (ТЗ) показателях надежности создаваемых изделий не представляется возможным. Одним из направлений решения этой актуальной задачи является оценивание значений требуемых показателей по результатам анализа безотказности (долговечности) элементов в условиях более тяжелых по сравнению с номинальными. В этом направлении получены значительные частные результаты (их анализ представлен далее). Однако гипотезу Н. М. Седякина отличает более высокий уровень общности сформулированных положений. Во-вторых, выяснилось, что рассматриваемый научный результат имеет более высокий уровень универсальности своего применения относительно той сферы, для которой он был первоначально сформулирован, и может быть успешно использован для приложений, которые ранее не рассматривались ни в

теоретическом, ни в практическом приложениях. Вместе с тем это требует уточнения областей обоснованного применения рассматриваемого научного положения, а также проведения анализа отдельных его положений для обеспечения адекватности теоретического обобщения гипотезы Н. М. Седякина в новых направлениях. Именно на решение этих вопросов нацелена данная статья.

РОЛЬ И МЕСТО ГИПОТЕЗЫ Н. М. СЕДЯКИНА В ТЕОРИИ НАДЕЖНОСТИ ВЫСОКОНАДЕЖНЫХ ОБЪЕКТОВ

Для описания взаимосвязи между внешними нагрузками и показателями надежности в настоящее время широко используются модели процессов разрушения [11]. Подходы, базирующиеся на их основе, позволяют оценивать интервал времени, в течение которого объект может противостоять заданной нагрузке до отказа, вызванного износом. Установлено, что наиболее точными моделями разрушения являются те из них, которые описывают физику процессов, приводящих к отказу элемента на уровне структуры или молекулярном уровне. Их называют структурными моделями закрытой формы, составными моделями или моделями физики процессов, приводящих к отказу [11]. Недостатком указанных моделей является их высокая сложность и по этой причине сравнительно узкая область применения. Отдельным важным видом модели разрушения является эмпирическая модель. Эмпирические модели математически описывают данные, собранные в процессе испытаний и эксплуатации определенных типов элементов. Наиболее широкое распространение нашли формы моделей, представленные на рисунке 1.

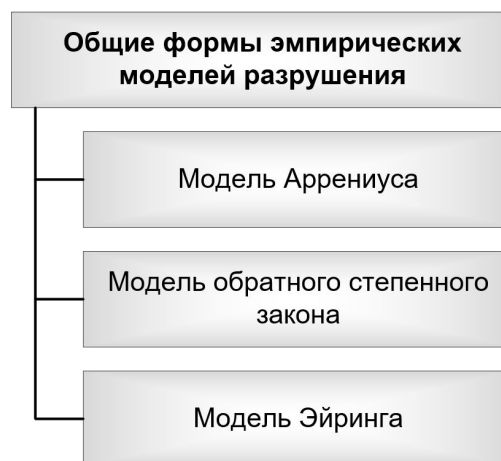


Рис 1. Формы эмпирических моделей разрушения

В основе первой из указанных моделей лежит уравнение зависимости интенсивности реакции инверсии сахарозы от температуры, полученное экспериментально нобелевским лауреатом в области химии С. Аррениусом (S. Arrhenius). Им был введен коэффициент ускорения времени указанной реакции с учетом изменения воздействующей температуры. К настоящему времени известно, что полученная закономерность также применима для оценивания характеристик отказов элементов микроэлектроники и ряда других типов объектов [12].

Следует отметить, что выполнение закона Аррениуса отвечает условиям линейной экстраполяции результатов испытаний с области повышенных температур на нормальные условия функционирования интегральных схем. Модель обратного степенного закона описывает ресурс системы, который, как предполагается, обратно пропорционален воздействующей нагрузке. Различные формы обратного степенного закона разработаны для конкретных применений. Развитием рассмотренных подходов является модель Эйринга. Она связывает с интенсивностью отказов интегральных схем не только повышение температуры, но и напряжение смещения. Недостатком приведенных видов эмпирических моделей является сравнительно узкая область их использования, обусловленная условиями доступности результатов проводимых экспериментальных исследований для конкретных типов объектов.

Значительно более общий характер имеют теоретические построения для обоснования ускоренных и форсированных испытаний. Основными из них являются так называемые физические принципы теории надежности. Наиболее широкое применение нашли принцип (гипотеза) Н. М. Седякина, гипотеза Майнера — Палмгрена и принцип наследственности. Гипотеза, сформулированная профессором Н. М. Седякиным, позволила найти объяснение многим ранее полученным экспериментальным результатам при исследовании электронных компонентов. Вторая из указанных гипотез описывает процесс линейного суммирования повреждений. Данное научное положение было сформулировано в период, когда изучение усталостных явлений только началось. Эта гипотеза не учитывает историю нагружения на объект и последовательность ступеней нагрузки. Позже Г. Д. Карташовым и А. И. Перроте на основе гипотезы Н. М. Седякина было сформулировано более общее научное положение — «принцип наследственности производства» [13]. Указанный принцип является достаточно продуктивным в решении ряда частных практических задач ускоренных испытаний отдельных видов изделий.

Следует отметить, что за время, прошедшее с момента формулирования принципа Н. М. Седякина, были проведены многочисленные экспериментальные исследования по проверке его справедливости. Эти опыты позволили, в целом, определить границы его действия. Было установлено, в частности, что в области небольших изменений нагрузочных факторов, действующих на радиоэлектронную аппаратуру (РЭА), данный физический принцип надежности, как правило, выполняется удовлетворительно, но в области значительных их изменений может выполняться неудовлетворительно. Объяснение этому результату пока полностью не дано. Также следует отметить, что возможности гипотезы Н. М. Седякина к настоящему

времени раскрыты далеко не полностью. Перейдем к анализу горизонтов действия этого важного научного положения.

АНАЛИЗ ОСОБЕННОСТЕЙ ГИПОТЕЗЫ Н. М. СЕДЯКИНА

Рассмотрим предварительно общую формулировку гипотезы Н. М. Седякина. Обозначим через ε конкретные условия применения элементов РЭА при эксплуатации. Под этими условиями понимаются механические, электрические, тепловые и иные нагрузки, воздействующие на объект со стороны внешней для него окружающей среды. Тогда выражение для вероятности $P(t/\varepsilon)$ безотказной работы объекта с учетом условий ε его эксплуатации можно записать в следующем виде:

$$P(t/\varepsilon) = \exp \left[- \int_0^t \lambda(z/\varepsilon) dz \right], \quad (1)$$

где $\lambda(t)$ — интенсивность отказов объекта в момент времени t при условиях ε его эксплуатации.

Значение показателя степени в правой части соотношения (1) получило название ресурса (запаса) надежности $r(t/\varepsilon)$, выработанного объектом за время t в условиях ε эксплуатации [1]. Таким образом, можно записать:

$$r(t/\varepsilon) = \int_0^t \lambda(z/\varepsilon) dz. \quad (2)$$

Профессором Н. М. Седякиным было сформулировано следующее утверждение (гипотеза) [1]: «существует множество условий, в котором надежность объекта в будущем зависит от величины запаса надежности (ресурса), выработанного в прошлом, и не зависит от того, как выработан этот ресурс».

Пусть ε_1 характеризует условия, при которых интенсивность отказов объекта по техническим условиям (ТУ) равна $\lambda_1(t/\varepsilon_1)$. Условия ε_2 соответствуют повышенной нагрузке на объект, интенсивность его отказов при этом равна $\lambda_2(t/\varepsilon_2)$. Очевидно, что при этом справедливо неравенство:

$$\lambda_1(t/\varepsilon_1) < \lambda_2(t/\varepsilon_2). \quad (3)$$

За интервал времени t_1 при установленной ТУ нагрузке ε_1 вырабатывается ресурс надежности $r_1(t_1/\varepsilon_1)$. Такой же ресурс может быть выработан при условиях ε_2 при ином значении времени t_2 . Ясно, что при этом справедливо условие $t_1 > t_2$. Следовательно, определив при повышенной нагрузке ε_2 показатели надежности объекта, можно за гораздо меньшее время определить расход его ресурса надежности в условиях эксплуатации по ТУ. Математически гипотеза Н. М. Седякина может быть представлена в виде соотношения

$$P(t/r) = P_1(t/x_1) = P_2(t/x_2), \quad (4)$$

где $P(t/r)$ — вероятность безотказной работы объекта за интервал времени t при условии, что его израсходованный ресурс надежности равен $r(x, \varepsilon)$; интервалы времени x_1 и x_2 удовлетворяют интегральному соотношению

$$r(x, \varepsilon) = \int_0^{x_1} \lambda(z, \varepsilon_1) dz = \int_0^{x_2} \lambda(z, \varepsilon_2) dz. \quad (5)$$

Зададим условия ε работы объекта в следующем виде:

$$\varepsilon = \begin{cases} \varepsilon_1, & 0 < z \leq t_1; \\ \varepsilon_2, & t_1 < z \leq (t_1 + t_2). \end{cases} \quad (6)$$

Тогда интенсивность $\lambda(z, t_2, t_1)$ его отказа в интервале $(t_1, t_1 + t_2)$ удовлетворяет соотношению

$$\lambda(z, t_2, t_1, \varepsilon_1) = \lambda(z + t_2 - t_1, \varepsilon_2), \quad (7)$$

где t_2 находится в результате решения относительно этого показателя уравнения:

$$\int_0^{t_2} \lambda(z, \varepsilon_2) dz = \int_0^{t_1} \lambda(z, \varepsilon_1) dz. \quad (8)$$

Таким образом, с помощью соотношений (1)–(8) сравнительно просто получить значения оценок надежности исследуемых объектов с учетом действующих на них нагрузок. Однако достигается такая кажущаяся простота использованием достаточно жестких предположений. Ослабление некоторых из них может привести к расширению области применения гипотезы Н. М. Седякина. Рассмотрим основные направления развития данной гипотезы.

НАПРАВЛЕНИЯ ТЕОРЕТИЧЕСКОГО ОБОБЩЕНИЯ ГИПОТЕЗЫ Н. М. СЕДЯКИНА

Основные возможные направления развития гипотезы Н. М. Седякина представлены на рисунке 2.

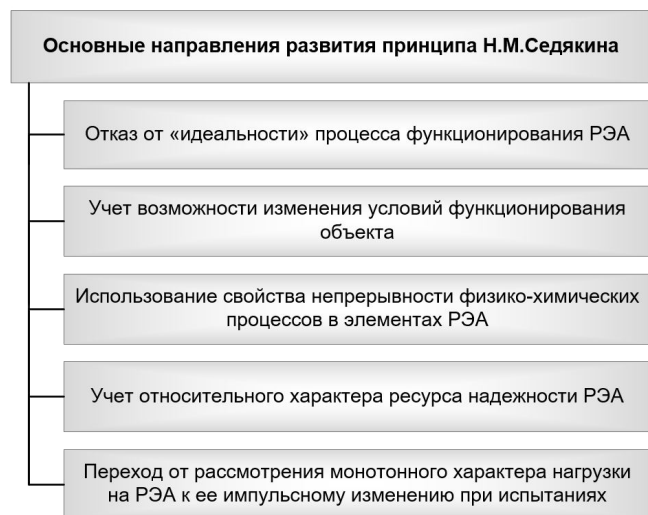


Рис. 2. Направления коррекции гипотезы Н. М. Седякина

Первым направлением теоретического обобщения гипотезы Н. М. Седякина является ослабление допущения о том, что при любом распределении времени до отказа объекта случайный ресурс его надежности распределен экспоненциально. Указанное допущение приводит к выводу о том, что процесс расходования ресурса надежности объекта является марковским. Однако такое допущение соответствует только идеализированным процессам функционирования технических объектов. Следовательно, для определения возможности увеличения области применимости гипотезы Н. М. Седякина, необходимо проанализировать справедливость ее применения для ресурса надежности объектов с рассмотрением более широкого класса распределений.

Второе направление состоит в следующем. Функция ресурса (запаса) надежности объекта, описываемая выражением (2), зависит от двух аргументов — t и ε . Однако при исследовании свойств ресурса надежности ранее этот показатель рассматривался как функция, у которой переменная t изменяется во времени, а нагрузка ε остается неизменной. Это соответствует частному случаю, когда заранее заданы интенсивности отказа исследуемого объекта для рассматриваемых фиксированных значений нагрузки. Для увеличения общности гипотезы в данном случае следует рассмотреть утверждение о том, что ресурс надежности объекта, как реализация случайной величины, не является неизменным при различных нагрузках. Его значение может зависеть как от увеличения нагрузки, так и от характера ее изменения.

Третье направление состоит в том, чтобы использовать свойство непрерывности физико-химических процессов, протекающих в материалах элемента. Это приводит к возможности для расчетов надежности применять гипотезу Н. М. Седякина не только с применением интегральной формы расчета вероятности безотказной работы исследуемого объекта, но и перейти к ее интегрально-дифференциальной форме. Поясним это следующим образом. Пусть для элемента установлено два режима ε_1 и ε_2 работы, отличающиеся уровнями нагрузки для него. Установим, что режим ε_2 характеризуется большим уровнем нагрузки на элемент по сравнению с нагрузкой при режиме ε_1 , соответствующим ТУ на объект. В этом случае ресурс надежности элемента при использовании режима ε_2 может быть выражен через ресурс надежности при использовании режима ε_1 с помощью соотношения [14]

$$\int_0^{x(t)} \lambda(z, \varepsilon_2) dz = \int_0^t \sum_{i=0}^{\infty} \frac{\partial \lambda^i(z, \varepsilon_1)}{i! \partial \varepsilon_1^i} (\varepsilon_2 - \varepsilon_1)^i dz. \quad (9)$$

Выражение (9) нетрудно представить в следующем виде:

$$\int_0^{x(t)} \lambda(z, \varepsilon_2) dz = \sum_{i=0}^{\infty} \frac{(\varepsilon_2 - \varepsilon_1)^i}{i!} \int_0^t \frac{\partial \lambda^i(z, \varepsilon_1)}{\partial \varepsilon_1^i} dz. \quad (10)$$

В дальнейшем при сохранении общности рассуждений и для упрощения выкладок можно ограничиться только двумя первыми членами ряда в правой части соотношения (10). При сделанном допущении получим соотношение для интенсивности отказов $\lambda(t, \varepsilon_2)$ элемента в режиме использования ε_2 . Выполняя операцию дифференцирования для двух частей формулы (10), имеем следующее соотношение [14]:

$$\lambda(t, \varepsilon_2) = \frac{\varepsilon_2}{\varepsilon_1} \times \left(\lambda\left(\frac{\varepsilon_2}{\varepsilon_1} t, \varepsilon_1\right) + (\varepsilon_2 - \varepsilon_1) \frac{\partial \lambda\left(\frac{\varepsilon_2}{\varepsilon_1} t, \varepsilon_1\right)}{\partial \varepsilon_1} \right). \quad (11)$$

Значительный практический интерес имеет переход от более нагруженного режима ε_2 к режиму ε_1 по ТУ. Поэтому целесообразно, используя выражение (11), получить соотношение для интенсивности отказов $\lambda(t, \varepsilon_1)$ элемента в режиме использования ε_1 . Искомая расчетная формула для показателя $\lambda(t, \varepsilon_1)$ имеет следующий вид:

$$\lambda(t, \varepsilon_1) = \frac{\varepsilon_1}{\varepsilon_2} \times \left(\lambda\left(\frac{\varepsilon_1}{\varepsilon_2} t, \varepsilon_1\right) + (\varepsilon_1 - \varepsilon_2) \frac{\partial \lambda\left(\frac{\varepsilon_1}{\varepsilon_2} t, \varepsilon_2\right)}{\partial \varepsilon_2} \right). \quad (12)$$

Выражения (11) и (12) позволяют обоснованно корректировать значения интенсивностей отказов элементов РЭА на основе результатов их испытаний.

Четвертое направление. Выражение (5) для значения ресурса $r(t, \varepsilon)$ надежности исследуемого элемента имеет статистический характер и не может быть абсолютным для любой нагрузки. Этот показатель должен иметь относительный характер по отношению к некоторому (базовому) значению ε_0 интенсивности отказов рассматриваемого типа элементов РЭА. Такое допущение приводит к тому, что показатель $r(t, \varepsilon)$ элемента в режиме ε должен зависеть не только от значения времени t , но и от режима ε_0 , принимаемого за начало отсчета.

Пятое направление. Выражение (5) для значения ресурса $r(t, \varepsilon)$ надежности элемента РЭА предполагает постоянство условий ε на интервале испытаний. Известно, что на практике интенсивность отказов элементов аппаратуры существенно возрастает при резких кратковременных перепадах нагрузки. Это предполагает проведение исследований в направлении учета импульсного характера воздействий на объект исследования.

Несомненно, приведенные направления развития гипотезы Н. М. Седякина потребуют проведения не только теоретических, но и ряда экспериментальных исследований. Рассмотрим теперь возможности использования данной гипотезы, отличающиеся постановкой решаемых задач.

ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ГИПОТЕЗЫ Н. М. СЕДЯКИНА

Основные возможности применения гипотезы Н. М. Седякина в сферах, отличных от надежности элементной базы РЭА, представлены на рисунке 3.



Рис. 3. Новые возможности применения гипотезы Н. М. Седякина

Понятие надежности технических и иных объектов не является застывшим и неизменным. В настоящее время

предлагаются различные подходы по расширению этого сложного свойства [15]. Не остается в стороне от этого процесса и гипотеза Н. М. Седякина. Достаточно новым является ее использование при анализе безопасности, надежности и качества продукции высоких технологий на основе статистики ускоренных испытаний [16].

Достаточно часто данный физический принцип надежности используется в своей исходной трактовке, но применительно к новым видам технических объектов, существенно отличающихся от элементной базы РЭА [17, 18]. Еще более интересным является применение гипотезы Н. М. Седякина для объектов нетехнической природы. Оказалось, что достаточно приемлемые для практики результаты можно получить при анализе надежности человека-оператора с учетом уровня его обучения [19]. Аналогичный подход использовался также для оценивания системы предотвращения нештатных ситуаций в эргатических системах управления на транспорте [20].

Огромные возможности таятся в применении гипотезы Н. М. Седякина в области надежности программного обеспечения (ПО) современных вычислительных систем. Можно выделить два направления этой работы:

- ускоренное определение показателей качества ПО при принудительном форсировании среды функционирования проверяемых программ, например за счет повышения быстродействия технической составляющей вычислительной системы [21];

- оценивание показателей безотказности и долговечности информационных систем различного назначения на основе принципа Н. М. Седякина, когда при расчетах учитываются особенности редящих нестационарных импульсных потоков дефектов ПО этих систем [22, 23].

Таким образом, можно сделать следующие важные выводы:

1. В теоретическом плане у гипотезы профессора Н. М. Седякина имеются значительные резервы, использование которых позволит значительно обогатить теорию и практику ускоренных испытаний высоконадежных объектов.

2. В условиях развития современной техники, прежде всего автоматизированной, рассматриваемый физический принцип надежности может найти достаточно широкое применение к таким объектам исследования, как программное обеспечение информационных систем.

3. Имеющиеся разработки позволяют говорить об определенной адекватности моделей эргатических систем, построенных на основе гипотезы Н. М. Седякина.

ЗАКЛЮЧЕНИЕ

Предложенные в статье направления развития гипотезы Н. М. Седякина позволяют существенно снизить достаточно жесткие требования к исходным допущениям, в рамках которых ранее рассматривалось применение данного научного положения. Это позволит расширить возможности организации и проведения ускоренных испытаний высоконадежных технических объектов.

Обобщенные в материалах статьи возможности применения гипотезы Н. М. Седякина говорят о том, что сфера ее использования намного шире, чем это предполагалось ранее. Несомненно, что следует продолжить работу в данном направлении, определять объекты различной физической природы и принципа действия, для которых окажутся

справедливыми условия применения рассмотренных научных выводов.

Предложенные подходы, как представляется, будут полезны специалистам по надежности технических объектов и программному обеспечению информационных систем, а также специалистам, изучающим свойства эргатических систем.

ЛИТЕРАТУРА

1. Седакин, Н. М. Об одном физическом принципе теории надежности // Известия Академии наук СССР. Техническая кибернетика. 1966. № 3. С. 80–87.
2. Гишваров, А. С. Моделирование ускоренных испытаний технических систем на надежность и ресурс // Вестник Уфимского государственного авиационного технического университета. 2007. Т. 9, № 1 (19). С. 26–40.
3. Ерошкин, А. Л. Оценка надежности полупроводниковых приборов и микросхем / А. Л. Ерошкин, Р. А. Попо // Международный журнал прикладных и фундаментальных исследований. 2015. № 12-2. С. 221–225.
4. Кулибаба, А. Я. Проблемы ускоренных испытаний электронной компонентной базы на надежность / А. Я. Кулибаба, С. П. Прищепова, А. Ю. Штукарёв // Ракетно-космическое приборостроение и информационные системы. 2014. Т. 1, Вып. 2. С. 81–85.
5. Авдеев, В. В. О возможностях ускоренных испытаний медицинского оборудования / В. В. Авдеев, Т. В. Колтакова // Моделирование, оптимизация и информационные технологии. 2017. Т. 5, № 1 (16). 9 с.
6. Дорошевич, К. К. Методология проведения ускоренных испытаний перспективных больших интегральных схем и сверхбольших интегральных схем на надежность / К. К. Дорошевич, П. В. Дорошевич, А. П. Поздеев // Военная электроника и электротехника: научно-технический сборник. 2011. Вып. 63, Ч. 2. С. 39–43.
7. Дорошевич, П. В. Исследования по определению энергии активации сверхбольших интегральных схем // Военная электроника и электротехника: научно-технический сборник. 2013. Вып. 65 (Ф), Ч. 2. С. 85–92.
8. Meeker, W. Q. Statistical Tools for the Rapid Development and Evaluation of High-Reliability Products / W. Q. Meeker, M. Hamada // IEEE Transactions on Reliability. 1995. Vol. 44, Is. 2. Pp. 187–198. DOI: 10.1109/24.387370.
9. Billinton, R. Basic Considerations in Generating Capacity Adequacy Evaluation / R. Billinton, D. Huang // Proceedings of the 2005 Canadian Conference on Electrical and Computer Engineering (CCECE) (Saskatoon, Canada, 01–04 May 2005). — Institute of Electrical and Electronics Engineers, 2005. — Pp. 611–614. DOI: 10.1109/CCECE.2005.1557005.
10. Jiang, K. New Models and Concepts for Power System Reliability Evaluation Including Protection System Failures / K. Jiang, C. Singh // IEEE Transactions on Power Systems. 2011. Vol. 26, No. 4. Pp. 1845–1855. DOI: 10.1109/TPWRS.2011.2156820.
11. Солдатенко, В. С. Надежность и испытания вооружения, военной и специальной техники: Учебное пособие / В. С. Солдатенко, Д. В. Садин, И. О. Голиков. — Санкт-Петербург: ВКА им. А. Ф. Можайского, 2021. — 165 с.
12. Строгонов, А. В. Долговечность интегральных схем и производственные методы ее прогнозирования // Chip News: Инженерная микроэлектроника. 2002. № 6. С. 44–49.
13. Перроте, А. И. Основы ускоренных испытаний радиоэлементов на надежность / А. И. Перроте, Г. Д. Карташов, К. Н. Цветаев. — Москва: Советское Радио, 1968. — 224 с. — (Библиотека инженера по надежности).
14. Смагин, В. А. Обобщение физического принципа теории надежности профессора Н. М. Седякина // Информатика и космос. 2006. № 3. С. 71–78.
15. Расширение понятия «надежность» в современной электроэнергетике / А. Л. Куликов, В. Л. Осокин, Б. В. Папков, Т. В. Шилова // Вестник НИИЭИ. 2018. № 3 (82). С. 88–98.
16. Никулин, М. С. О развитии и внедрении методов статистики ускоренных испытаний для анализа безопасности, надежности и качества продукции высоких технологий / М. С. Никулин, М. В. Сильников, К. А. Дубаренко // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. 2014. Вып. 1 (190). С. 217–223.
17. Оценка надежности пьезоэлектрических вакуумных прецизионных резонаторов, применяемых в радиоэлектронной аппаратуре космических аппаратов / И. В. Архипова, А. В. Батурин, И. А. Голубева, А. И. Митюшов // I-methods. 2018. Т. 10, № 4. С. 5–10.
18. Ефремов, Л. В. Ускоренные испытания стальных образцов на износостойкость методом искусственных баз / Л. В. Ефремов, А. В. Тикалов, А. Д. Бреки // Известия высших учебных заведений. Приборостроение. 2016. Т. 59, № 8. С. 671–676. DOI: 10.17586/0021-3454-2016-59-8-671-676.
19. Смагин, В. А. Модель надежности оператора в условиях обучения // Надежность и контроль качества. Методы менеджмента качества. 1999. № 9. С. 57–64.
20. Смагин, В. А. Эвристическая модель ликвидации нештатных ситуаций в эргатических системах управления // Интеллектуальные технологии на транспорте. 2017. № 1 (9). С. 5–10.
21. Смагин, В. А. Форсированные быстродействием испытания программного обеспечения на надежность // Автоматика и вычислительная техника. 2003. № 5. С. 3–11.
22. Солдатенко, В. С. Надежность информационных систем: Учебное пособие. — Санкт-Петербург: ВКА имени А. Ф. Можайского, 2008. — 86 с.
23. Смагин, В. А. Обобщение понятия ресурса надежности информационно-измерительных систем на основе принципа Н. М. Седякина / В. А. Смагин, Р. О. Лавров, С. Ф. Литвиненко // Надежность и качество сложных систем. 2022. № 4. С. 40–47. DOI: 10.21685/2307-4205-2022-4-5.

New Horizons of Application in the Theory of Reliability of Highly Reliable Systems of the Hypothesis of N. M. Sedyakin

Grand PhD V. A. Smagin, PhD V. S. Soldatenko

Mozhaisky Military Space Academy

Saint Petersburg, Russia

va_smagin@mail.ru, soldatenko_vs@mail.ru

Abstract. Proposals for weakening a number of assumptions on the basis of which the hypothesis of N. M. Sedyakin was formulated and confirmed by theoretical relations. Directions of adequate use of this scientific position in relation to new objects not previously considered are proposed. These include information and ergatic systems, computer complex software. The described approaches will justify the results of accelerated reliability tests of a fairly wide range of highly reliable objects.

Keywords: reliability of objects, physical principles of reliability, accelerated tests, failure rate, reliability resource.

REFERENCES

- Sedyakin N. M. About One Physical Principle of Reliability Theory [Ob odnom fizicheskom printsipe teorii nadezhnosti], *Proceedings of the Academy of Sciences of the USSR. Technical Cybernetics [Izvestiya Akademii nauk SSSR. Tekhnicheskaya kibernetika]*, 1966, No. 3, Pp. 80–87.
- Gishvarov A. S. Modeling of Accelerated Tests of Technical Systems for Reliability and Resource [Modelirovanie uskorennykh ispytaniy tekhnicheskikh sistem na nadezhnost i resurs]. *Vestnik of Ufa State Aviation Technical University [Vestnik Ufimskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta]*, 2007, Vol. 9, No. 1 (19), Pp. 26–40.
- Eroshkin A. L., Popo R. A. Evaluation of Reliability Semiconductor Devices and Microcircuits [Otsenka nadezhnosti poluprovodnikovyykh priborov i mikroskhem], *International Journal of Applied and Fundamental Research [Mezhdunarodnyy zhurnal prikladnykh i fundamentalnykh issledovaniy]*, 2015, No. 12-2, Pp. 221–225.
- Koulibaba A. Ya., Prischepova S. P., Shtukarev A. Yu. Problems of Accelerated Life Tests of Electronic Components [Problemy uskorennykh ispytaniy elektronnoy komponentnoy bazy na nadezhnost], *Rocket-Space Device Engineering and Information Systems [Raketno-kosmicheskoe priborostroenie i informatsionnye sistemy]*, 2014, Vol. 1, Is. 2, Pp. 81–85.
- Avdeev V. V., Koltakova T. V. About the Possibilities of Accelerated Testing of Medical Equipment [O vozmozhnostyakh uskorennykh ispytaniy meditsinskogo oborudovaniya], *Modeling, Optimization and Information Technology [Modelirovanie, optimizatsiya i informatsionnye tekhnologii]*, 2017, Vol. 5, No. 1 (16), 9 p.
- Doroshevich K. K., Doroshevich P. V., Pozdeev A. P. Methodology of Accelerated Testing of Promising Large Integrated Circuits and Ultra-Large Integrated Circuits for Reliability [Metodologiya provedeniya uskorennykh ispytaniy perspektivnykh bolshikh integralnykh skhem i sverkhbolshikh integralnykh skhem na nadezhnost], *Military Electronics and Electrical Engineering [Voennaya elektronika i elektrotehnika: nauchno-tekhnicheskii sbornik]*, 2011, Is. 63, Pt. 2, Pp. 39–43.
- Doroshevich P. V. Research on Determining the Activation Energy of Ultra-Large Integrated Circuits [Issledovaniya po opredeleniyu energii aktivatsii sverkhbolshikh integralnykh skhem], *Military Electronics and Electrical Engineering [Voennaya elektronika i elektrotehnika: nauchno-tekhnicheskii sbornik]*, 2013, Is. 65 (F), Pt. 2, Pp. 85–92.
- Meeker W. Q., Hamada M. Statistical Tools for the Rapid Development and Evaluation of High-Reliability Products, *IEEE Transactions on Reliability*, 1995, Vol. 44, Is. 2, Pp. 187–198. DOI: 10.1109/24.387370.
- Billinton R., Huang D. Basic Considerations in Generating Capacity Adequacy Evaluation, *Proceedings of the 2005 Canadian Conference on Electrical and Computer Engineering (CCECE), Saskatoon, Canada, May 01–04, 2005*. Institute of Electrical and Electronics Engineers, 2005, Pp. 611–614. DOI: 10.1109/CCECE.2005.1557005.
- Jiang K., Singh C. New Models and Concepts for Power System Reliability Evaluation Including Protection System Failures, *IEEE Transactions on Power Systems*, 2011, Vol. 26, No. 4, Pp. 1845–1855. DOI: 10.1109/TPWRS.2011.2156820.
- Soldatenko V. S., Sadin D. V., Golikov I. O. Reliability and Testing of Weapons, Military and Special Equipment: Study guide [Nadezhnost i ispytaniya vooruzheniya, voennoy i spetsialnoy tekhniki: Uchebnoe posobie]. Saint Petersburg, Mozhaisky Military Space Academy, 2021, 165 p.
- Strogonov A. V. Durability of Integrated Circuits and Production Methods of Its Forecasting [Dolgovechnost integralnykh skhem i proizvodstvennye metody ee prognozirovaniya], *Chip News: Microelectronics Engineering [Chip News: Inzhenernaya mikroelektronika]*, 2002, No. 6, Pp. 44–49.
- Perrote A. I., Kartashov G. D., Tsvetayev K. N. Fundamentals of Accelerated Testing of Radio Elements for Reliability [Osnovy uskorennykh ispytaniy radioelementov na nadezhnost]. Moscow, Soviet Radio Publishing House, 1968, 224 p.
- Smagin V. A. Generalization of the Physical Principle of the Reliability Theory by Professor N. M. Sedyakin [Obobshchenie fizicheskogo printsipa teorii nadezhnosti professora N. M. Sedyakina], *Information and Space [Informatsiya i kosmos]*, 2006, No. 3, Pp. 71–78.
- Kulikov A. L., Osokin V. L., Papkov B. V., Shilova T. V. The Extension of the Concept «Reliability» in Modern Electric Power Industry [Rasshirenie ponyatiya «nadezhnost» v sovremennoy elektroenergetike], *Bulletin NGIEI [Vestnik NGIEI]*, 2018, No. 3 (82). Pp. 88–98.

16. Nikulin M. S., Silnikov M. V., Dubarenko K. A. On the Development and Implementation Methods of Statistics of Accelerated Tests for Safety Analysis, Reliability and Quality of Products of High Technologies [O razviti i vnedrenii metodov statistiki uskorennykh ispytaniy dlya analiza bezopasnosti, nadezhnosti i kachestva produktsii vysokikh tekhnologiy], *St. Petersburg Polytechnical University Journal [Nauchno-tekhnicheskie vedomosti Cankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta]*, 2014, Is. 1 (190), Pp. 217–223.

17. Arkhipova I. V., Baturin A. V., Golubeva I. A., Mityushov A. I. Assessment of the Reliability of Piezoelectric Vacuum Resonators Used in Electronic Equipment of Spacecrafts [Otsenka nadezhnosti pyezoelektricheskikh vakuumnykh pretsizionnykh rezonatorov, primenyaemykh v radioelektronnoy apparature kosmicheskikh apparatov], *I-methods*, 2018, Vol. 10, No. 4, Pp. 5–10.

18. Efremov L. V., Tikalov A. V., Breki A. D. Accelerated Testing of Steel Samples on the Durability by the Artificial Bases Method [Uskorennye ispytaniya stalnykh obraztsov na iznosostoykost metodom iskusstvennykh baz], *Journal of Instrument Engineering [Izvestiya vysshikh uchebnykh zavedeniy. Priborostroenie]*, 2016, Vol. 59, No. 8, Pp. 671–676. DOI: 10.17586/0021-3454-2016-59-8-671-676.

19. Smagin V. A. Model of Operator Reliability in Training Conditions [Model nadezhnosti operatora v usloviyakh obucheniya], *Reliability and Quality Control. Quality Management Methods [Nadezhnost i kontrol kachestva. Metody menedzhmenta kachestva]*, 1999, No. 9, Pp. 57–64.

20. Smagin V. A. Heuristic model of liquidation of super-numerary situations in man-machine control systems [Evristsicheskaya model likvidatsii neshtatnykh situatsiy v ergaticheskikh sistemakh upravleniya], *Intellectual Technologies on Transport [Intellektualnye tekhnologii na transporte]*, 2017, No. 1 (9), Pp. 5–10.

21. Smagin V. A. The Tests Forced by Speed of the Software for Reliability [Forsirovannye bystrodeystviem ispytaniya programmnoho obespecheniya na nadezhnost], *Automatic Control and Computer Sciences [Avtomatika i vychislitel'naya tekhnika]*, 2003, No. 5. Pp. 3–11.

22. Soldatenko V. S. Reliability of information systems: Study guide [Nadezhnost informatsionnykh sistem: Uchebnoe posobie]. Saint Petersburg, Mozhaisky Military Space Academy, 2008, 86 p.

23. Smagin V. A., Lavrov R. O., Litvinenko S. F. Generalization of the concept of reliability resource information and measurement systems based on the principle of N. M. Sedyakin [Obobshchenie ponyatiya resursa nadezhnosti informatsionno-izmeritelnykh sistem na osnove printsipa N. M. Sedyakina], *Reliability and Quality of Complex Systems [Nadezhnost i kachestvo slozhnykh sistem]*, 2022, No. 4, Pp. 40–47. DOI: 10.21685/2307-4205-2022-4-5.

Software Defect Prediction System Based on Well-Tuned Random Forest Technique

F. H. Khorsheed, N. J. Ibrahim

University of Diyala
Baqubah, Iraq

farah_hatam@uodiyala.edu.iq, nebras.jalel@uodiyala.edu.iq

Q. Saihood

Ashur University College
Baghdad, Iraq

qusaysaihood@au.edu.iq

Abstract. Software quality is the main criterion for increasing user demand for software. Therefore, software companies seek to ensure software quality by predicting software defects in the software testing phase. Having an intelligent system capable of predicting software defects helps greatly in reducing time and effort consumption. Despite the great trend to develop software defect prediction systems based on Machine Learning techniques in last few years, the accuracy of these systems is still a major challenge.

Therefore, in this study, a software defect prediction system based on three stages is presented to improve the prediction accuracy. First stage, data pre-processing is performed, which includes (data cleaning, data balance, data normalization, and feature selection). Second stage the hyperparameters of ML are tuned using Grid Search technique. Finally, a well-tuned ML technique is implemented to predict software defects.

Performance experiments were carried out on the JM1 dataset where the proposed system achieved promising results in predicting software defects. Among ML techniques used, a well-tuned RF technique outperformed the rest of the used ML techniques, in addition to the techniques mentioned in previous works, with an accuracy of 88,26 %. This study proves that the selection of important features and efficient hyperparameter tuning of ML techniques significantly improve the accuracy of software defect prediction.

Keywords: machine learning, Random Forest, software defects, feature selection, prediction.

INTRODUCTION

The world is witnessing a great development in computer technology, smart phones, and the Internet of Things. This development leads to the development of software and an increase in demand for it because software is a means of connecting humans and electronic devices. Today people live in the age of software where there are millions of programs developed daily [1, 2]. The most important phase of Software Development Life Cycle (SDLC) is the testing phase in which future software defects are checked before exporting the software for use [3]. A software defect is a software error that leads to incorrect results that may be caused by errors in the source code. The occurrence of software defects in software in the future negatively affects the quality and reliability of the software [4]. Also, the process of repairing them is costly and may withdraw all versions from the market, and thus cause costly losses for the company producing the program. These flaws may also cause serious security holes that hackers can exploit as a way to exploit your request and sometimes steal important information or money [5].

The biggest challenge for software companies and programmers is to predict future software defects in real-life scenarios [6]. Therefore, early prediction of software defects is seen as the most important research field since the beginning of the

software era [4]. To ensure software quality many quality attributes and metrics have been developed with many quality assurance techniques, but still the important question of how to ensure that software will possess good quality is an open issue. Finding out which units are prone to failure is closely related to the quality of the program. Defects prediction involves early detection of those dangerous modules of the program that are prone to errors and impair quality [7]. Detecting defective modules at the early stage is vital as the cost of rectification increases in later stages of the development life cycle. Software metrics extracted from historical software data are used to predict defective units [8, 9]. Therefore, defective modules must be detected and removed 100% to ensure high quality software.

In recent years, Artificial Intelligence (AI) methods have shown a great ability to predict hidden patterns in huge data. There have been many studies that dealt with the use of Machine Learning (ML) and Deep Learning (DL) techniques in predicting software defects in the software testing phase. Many ML and DL techniques such as (Logistic Regression (LR) technique, Decision Tree (DT) technique, Support Vector Machine (SVM) technique, Random Forest (RF) technique, K-Nearest Neighbor (KNN) technique, Multi-Layer Perceptron (MLP) technique, Artificial Neural Network (ANN), and Convolutional Neural Network (CNN)) were built to predict software defects [10]. Although there are many studies dealing with the use of AI methods to predict software defects, there are several shortcomings. One of the most important shortcomings is the accuracy of these methods. Until now, the accuracy of predicting software defects using ML and DL techniques is still a great challenge [7, 11–13]. In addition to identifying the features that affect the process of predicting software defects. Efficient selection of important features greatly influences the prediction of software defects with high accuracy [7].

Therefore, this study aims to build a particular system to predict software defects effectively. The proposed system consists of three stages: In the first stage, data pre-processing is implemented, which includes filling in missing values, balancing data categories, data normalization, and selection of important features affecting the prediction process. In the second stage, Grid search technique is applied to effectively adjust the hyperparameters of ML techniques to improve their performance. Finally, six different ML techniques (RF, DT, SVM, LR, KNN, and MLP) are built for predicting software defects. Performance experiments for this study were performed on the JM1 dataset which is one of 12 online open-source software defect datasets provided by the NASA Software Engineering Repository [14]. The performance of the proposed system was evalu-

ated using common ML evaluation metrics (Accuracy, Precision, Sensitivity, Specificity, F Score). In addition to calculating the confusion matrix that shows actual and predictor values.

The rest of the article is organized as follows: In the second section, related work is summarized. In the third section, the general approach of the proposed system is summarized, which includes all steps of data pre-processing and hyperparameter tuning, in addition to mentioning ML techniques used briefly. In the fourth section, evaluation matrices are mentioned to evaluate the performance of ML techniques used in this work. In the fifth section, the results are mentioned and discussed, in addition to comparing them with the results of related work. Finally, the conclusion and future work summarizes the contributions of the work and outlines future work in Section sixth.

RELATED WORKS

Recent years have seen a significant increase in the use of AI methods in predicting software defects. Various ML and DL techniques have been used to explore hidden patterns in the software's source code. Early software defects prediction helps to ensure software quality. In this section, several studies using ML and DL techniques for software defect prediction are summarized.

P. D. Singh and A. Chug [15] analyzed 7 datasets from NASA Promise dataset repository using DT, Naive Bayes (NB), Linear classifier (LC), ANN and Particle Swarm Optimization (PSO) for predicting software defects. 10-fold cross validation was used to evaluate the techniques used. The LC technique achieved a better performance than the rest of the techniques used when it was used on each of the datasets (JM1, KC2, PC1, and AT). While the DT technique outperformed the rest of the techniques with two data sets (CM1 and KC1).

A. Iqbal, et al. [7] introduced a comprehensive study of a variety of ML techniques to predict software defects so that researchers can later use this study as a basic plan in their future research. NB, MLP, Radial Basis Function (RBF), SVM, KNN, kStar (K*), One Rule (OneR), PART, DT, RF, and ensemble methods were applied to 12 datasets from NASA with and without using feature selection techniques. The performance of the techniques used was evaluated using the commonly used performance metrics Accuracy, Precision, Recall, F Score, MCC, and ROC. The results of this study show that the prediction accuracy of the used techniques is still not good enough, in addition to that the used datasets suffer from a large variation in the size of the categories.

A. Iqbal and S. Aftab [13] tried to solve the data imbalance using oversampling technique. In addition to suggesting use feature selection techniques and Multi-Layer Perceptron (MLP) technique for software defect prediction. Performance experiments were conducted on 12 datasets provided by NASA and in two different directions: with and without the over-sampling technique. In addition, the performance of the proposed system was compared with a set of common ML classifiers using common performance measures, as it was noted that the proposed system with the over-sampling technique performs well compared to other ML classifiers.

B. Khan, et al. [11] compared the performance of seven ML techniques (MLP, SVM, J48, RBF, RF, Hidden Markov Model (HMM), Dependency Decision Tree (CDT), KNN, Average Dependency Estimator (A1DE), and NB) based on seven dif-

ferent datasets provided by NASA. Performance of ML techniques was evaluated using various measures such as MAE, RAE, and RMSE. and RRSE, recall, and accuracy. The performance of ML techniques is evaluated using different measures (Accuracy, Recall, Relative Absolute Error (RAE), Root Mean Squared Error (RMSE), Root Relative Squared Error (RRSE), and Mean Absolute Error (MAE)). Among the seven used ML techniques RF technique has the highest accuracy rate of 88.32%. The outputs of this research can be used as a reference point for new research.

R. Vats and A. Kumar [16] explored several supervised and unsupervised ML techniques to determine the best method for predicting software defects. Performance experiments were conducted on 9 datasets from NASA, where experiments showed that supervised ML techniques are more suitable than unsupervised ML techniques for predicting software defects.

R. Shrimankar, et al. [17] proposed to use ensemble methods (XGBoost, AdaBoost, and Gradient Boost) as well as base ML methods (LR, MLP, NB, SVM, RF, DT, and KNN) to analyze the efficiency of ML techniques in software defect prediction. Performance experiments have been conducted on 12 datasets provided by NASA, where XGBoost method slightly outperformed other used methods on some of used datasets.

Previous work indicates that achieving high accuracy in predicting defects is still a major challenge, especially when using the JM1 dataset as shown in table 1. Also, there is no specific ML technique that significantly outperforms other techniques in predicting software defects, as well as the importance of identifying a feature selection technique that is appropriate with the used ML technique. In addition, most of the datasets with software defects suffer from a large disparity between categories.

METHODOLOGY

The proposed general framework includes the following stages: In the first stage, data pre-processing is well implemented, which includes cleaning the data by making sure that the data is free of any inconsistencies, balancing classes within the data set, normalizing the data, and determining the importance of each feature using DT technique as a feature selection technique. In the second stage, using the Grid search technique to effectively adjust the hyper-parameters of the ML technique used to improve its performance. In the third stage, apply six different ML techniques (RF, DT, SVM, LR, KNN, and MLP) to predict software defects. Finally, the use of common performance metrics (Accuracy, Precision, Sensitivity, Specificity, F Score, ROC curve, and Confusion matrix) to evaluate the ML technologies used as shown in Figure 1.

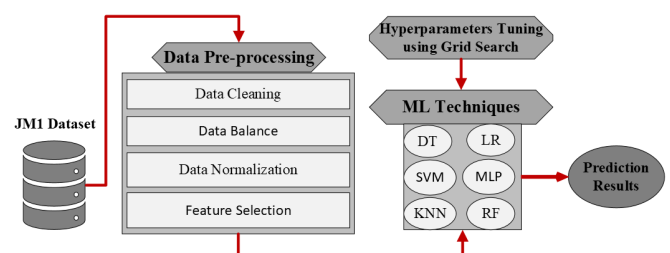


Fig. 1. The Proposed General Framework for Software Defect Prediction

Summary of Related Work

Study	Year	Dataset	FS Techniques	Data Balance	ML techniques	JM1 ac., %
P. D. Singh and A. Chug [15]	2017	CM1, JM1, KC1, KC2, PC1, AT, KC1 LC	No	No	DT, NB LC, ANN, PSO	80,64
A. Iqbal, et al. [7]	2019	CM1, JM1, KC1, KC3, MC1, MC2, MW1, PC1, PC2, PC3, PC4, PC5	Yes	No	NB, MLP, RBF, SVM, KNN, K*, OneR, PART, DT, and RF	80,61
A. Iqbal and S. Aftab [13]	2020	CM1, JM1, KC1, KC3, MC1, MC2, MW1, PC1, PC2, PC3, PC4, PC5	Yes	Yes	NB, MLP, RBF, SVM, KNN, K*, OneR, PART, DT, RF, MLP-FS, and MLP-FS-ROS	80,44
B. Khan et al. [11]	2021	CM1, JM1, KC3, MC1, MC2, PC1, PC2, PC3	No	No	MLP, SVM, J48, RBF, RF, HMM, CDT, KNN, A1DE, NB	82,02
R. Vats and A. Kumar [16]	2021	CM1, JM1, KC3, MC1, MC2, PC1, PC2, PC3, PC4	No	No	Bagging, AdaBoost, RF, K-Mean, and K-harmonic Mean (KHM)	88,00
R. Shrimankar, et al. [17]	2022	CM1, JM1, KC1, KC3, MC1, MC2, MW1, PC1, PC2, PC3, PC4, PC5	No	No	XGBoost, AdaBoost, Gradient Boost, LR, MLP, NB, SVM, RF, DT, KNN	80,00

JM1 DATASET

There are 12 online open-source software defect datasets provided by the NASA promise software engineering repository. In this study performance experiments were performed on one of the toughest datasets provided by NASA known as JM1 [18]. The JM1 dataset consists of 10 885 samples and 21 features in addition to the target category (Defect or Not Defect). The JM1 dataset was developed using the C programming language and is a real-time predictive terrestrial system that uses mimicries to generate predictions. Table 2 describes the details of the JM1 dataset.

Table 2

Description of JM1 dataset features

No.	Feature	Type	Description
1	loc	numeric	Halstead line count of code
2	v(g)	numeric	Halstead's «cyclomatic complexity»
3	ev(g)	numeric	Halstead's «essential complexity»
4	iv(g)	numeric	Halstead's «design complexity»
5	n	numeric	Halstead's total operators + operands
6	v	numeric	«volume»
7	l	numeric	unique «program length»
8	d	numeric	unique «difficulty»
9	i	numeric	total «intelligence»
10	e	numeric	total «effort»
11	b	numeric	module
12	t	numeric	time estimator
13	IOCode	numeric	line count
14	IOComment	numeric	count of lines of comments

15	IOBlank	numeric	count of blank lines
16	IOCodeAndComment	numeric	lines of comments + line count of code
17	uniq_Op	numeric	operators
18	uniq_Opnd	numeric	operands
19	total_Op	numeric	operators
20	total_Opnd	numeric	operands 0,875 branch-Count: numeric % of the flow graph
21	branchCount	numeric	the flow graph
22	defects	{false,true}	has/has not one or more reported defects

DATA PRE-PROCESSING

The performance of ML techniques depends mainly on the quality of the data provided to it. Therefore, data pre-processing is an essential and important stage in ML and data mining processes to ensure data quality. Oftentimes, real world data has many issues such as data inconsistency, noise, may contain missing values, or may not be suitable for ML techniques [19]. If ML techniques are applied directly to this raw data, the performance of the model will be negatively affected, and it may be learned incorrectly [20]. Therefore, data pre-processing is carried out to prepare the data well by cleaning the data and filling in the missing values, as well as transforming the data to a form appropriate with the used ML techniques [21]. Data pre-processing in this work includes four stages: data cleaning, data balancing, data normalization, and feature selection.

DATA CLEANING

The first stage of data pre-processing is the data cleaning process in which the data is addressed from noise, inconsistent data, redundant data, and missing values [19]. The JM1 dataset contains some missing values for some features (uniq_Op, uniq_Opnd, total_Op, total_Opnd, BranchCount) which are addressed by filling them by calculating the mean value of the column to which the missing value belongs.

DATA BALANCING

The JM1 dataset contains 10 885 samples, of which 8 779 are of the non-defect category and 2 106 are of the defect category, which indicates a significant deviation between the two data categories. To address the problem of categories imbalance in the JM1 dataset, The Synthetic Minority Over-Sampling Technique (SMOTE), a technique used to increase small category size [22], was used. The size of the small category was increased by generating additional samples using the SMOTE technique to be both categories equal in size.

DATA NORMALIZATION

The dataset usually contains a discrepancy in the range of numeric data between the different columns. When some columns have a very high range, while others may have a very low range, this can negatively affect the performance of the ML technique [23]. Therefore, data normalization is used to normalize the data to be between a uniform range while maintaining the differences in the ranges of values and not losing information [24]. In this work, the Min-max Scaler method has been applied to numeric features to convert them to range between 0 and 1 by following the equation next [25]. The scale of data for a few features may be essentially distinctive from those of others, which may hurt the execution of our models. It is particularly the case with algorithms that depend on a degree of separation, such as Neural Networks and KNN. It is additionally accommodating for optimizing machine learning processes like gradient descent and empowers convergence to happen faster, and it can offer assistance to improve the execution and speed of the execution of algorithms. Since the data is as of now scaled-down, complex calculations basically required to optimize algorithms are quicker. In expansion to that, it can moreover be accommodating when comparing diverse datasets or models in terms of their performances [22].

$$D = \frac{(X - \min(X))}{(\max(X) - \min(X))},$$

where D is the normalized value, X is the original value of a feature, $\min(X)$ is smallest value of a feature, $\max(X)$ is biggest value of a feature.

FEATURE SELECTION

Feature selection is the process of selecting a set of input data features that have a significant impact on the target. Determining the important and appropriate Features able to improve the performance of the ML technique and reduce the time it takes to train [26]. The DT technique has a feature importance property that can be used to calculate feature importance [27]. The importance of the JM1 dataset Features was calculated using DT technique and then features with less importance were removed.

ML TECHNIQUES

ML is part of the science of AI, which are statistical methods that enable a computer or any device to build its own concept based on the data provided to it in the training phase. There are several types of ML depending on the type of learning: supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning [28]. In this work, 6 supervised ML techniques were used.

DT

It is one of the most popular non-parametric supervised ML techniques that can be used for both classification and regression tasks. The structure of DT technique is like an upside-down tree where the first node represents the root and then it is split into inner nodes. Each inner node refers to the feature, the branches refer to the rules and the leaves refer to the result of the techniques [28].

RF

RF is one of the most powerful and popular ML techniques that can be used for both classification and regression tasks. RF technique is one of the ensemble methods based on bagging method which aggregates a set of decision trees. Each decision tree within the forest is trained on a subset of the data, and thus the final decision is made using tree-majority voting. RF is mostly used to solve some problems of DT technique and some individual ML techniques [7, 11, 28]:

- in most cases, it gives better classification accuracy than DT technique;
- provides an effective way to deal with lost data;
- solving the problem of overfitting in DT technique;
- more stable than DT technique.

Random forest could be a Supervised Machine Learning Calculation that's utilized broadly in Classification and Regression issues. It builds decision trees on distinctive tests and takes their larger part vote for classification and normal in case of regression [25]. A random forest may be a machine learning technique that's won't to solve regression and classification problems. It utilizes ensemble learning, that is a technique that mixes several classifiers to produce solutions to complicated problems.

Features of a Random Forest algorithm:

- it's more accurate than the decision tree algorithm;
- it provides an effective way of handling missing data;
- it can produce a reasonable prediction without hyper-parameter tuning;
- it solves the issue of overfitting in decision trees;
- in every random forest tree, a subset of features is selected randomly at the node's splitting point.

SVM

SVM is one of the most important ML techniques suitable for both small and complex data. SVM can be used to solve both classification and regression problems, but it is usually used to solve classification problems. SVM technique finds the best separation line between data categories by maximizing margins and determining the best separation of data [19, 28].

LR

LR is one of the simplest classification techniques for ML based on statistical background. LR is a statistical method for analyzing a data set in which there are one or more independent variables that determine the result. Every user who uses LR technique needs to know log probabilities, the key concept behind a LR technique. LR estimates the probability that an event, such as a defect or a non-defect, will occur based on a given data set of independent variables. Since the outcome is a probability, the dependent variable is constrained between 0 and 1 [29].

KNN

KNN is a non-parametric ML technique that can be used for classification and regression tasks [19]. The way the KNN technique works is very simple and effective to determine if a new sample is a software defect or not. The distance between the sample to be classified and all points of the training data set of software defects is measured using one of the distance scales such as the Euclidean distance, Manhattan distance, and Minkowski distance. Next, the number K representing the number of nearest neighbors is determined, which determines the nature of the new sample (defect or not) according to the nature of its nearest neighbors [29]. In this work, the distance between the samples of the training set and the samples of the test set was measured using a Minkowski distance.

MLP

MLP is a fully connected feedforward ANN that can be used for both classification and regression tasks. The MLP network consists of three layers, an input layer used to enter the training data and a layer or hidden layers used to process the input data. And the last layer is the output layer, which is used to obtain the required outputs [19].

TUNING HYPERPARAMETERS OF ML TECHNIQUES

ML techniques have hyperparameters that help fine-tune their performance. Fine tuning of hyperparameters of ML techniques helps to stabilize performance, improve accuracy, and reduce model complexity. Usually, hyperparameters of ML techniques are manually tune, which is cumbersome and time consuming. Therefore, the Grid Search technique was used to tune the hyperparameters of the ML techniques used in this work. Grid search technique is fed with the hyperparameters and their potential value, and then the 10-fold validation method is used to train the technique with all the possible values of the hyperparameters. Finally, the best values of the hyperparameters with the best performance of the ML techniques are determined.

PERFORMANCE MEASUREMENT

The performance of used ML techniques was evaluated by calculating the confusion matrix in addition to using six common performance evaluation criteria (Accuracy, Precision, Sensitivity, Specificity, F1 Score, and ROC Curve). The confusion matrix contains the actual values and predicted values from which the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are calculated. The «False» (not defect) class is represented as negative '0', while the «True» (defect) class is represented as positive '1'.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$\text{Sensitivity} = \frac{TP}{TP + FN},$$

$$\text{Specificity} = \frac{TN}{TN + FP},$$

$$\text{F1 Score} = \frac{2 \times TP}{2 \times TP + FP + FN}.$$

RESULTS AND DISCUSSION

The proposed RF technique, as well as other ML techniques used in this study, were built using the scikit-learn library. Performance experiments were conducted on the JM1 dataset, where the dataset was separated 80 % for training ML techniques and 20 % for testing them.

The proposed RF technique achieved the best performance with an accuracy of 88,26%, along with a precision of 87,76 %, Sensitivity of 88,47 %, Specificity of 88,06 %, and F1-Score of 88,11 % compared to the other techniques used as shown in Table 3. Then, KNN technique ranked second with a competitive performance with an accuracy of 85,76 %, along with a precision of 81,60 %, Sensitivity of 91,71 %, Specificity of 80,0 %, and F1-Score of 86,36 %. Also, the DT technique achieved a good performance with an accuracy of 82,17 %, along with a precision of 81,54 %, Sensitivity of 82,39 %, Specificity of 81,96 %, and F1-Score of 81,97 %, while the performance of the MLP, LR, and SVM techniques was modest with an accuracy of 70,50 %, 66,88 %, and 65,51 %, respectively.

Table 3

Performance comparison of ML techniques used in this study

Technique	Accuracy, %	Precision, %	Sensitivity, %	Specificity, %	F1 Score, %
RF	88,26	87,76	88,47	88,06	88,11
DT	82,17	81,54	82,39	81,96	81,97
SVM	65,51	68,14	56,10	74,62	61,54
KNN	85,76	81,60	91,71	80,00	86,36
LR	66,88	68,10	61,43	72,15	64,59
MLP	70,50	70,55	68,67	72,26	69,60

The performance of ML techniques was also compared by drawing the ROC curve as shown in Figure 2. In addition, a confusion matrix was calculated for each of ML techniques used to show the correct and incorrect predictions for both classes as in Figures 3.

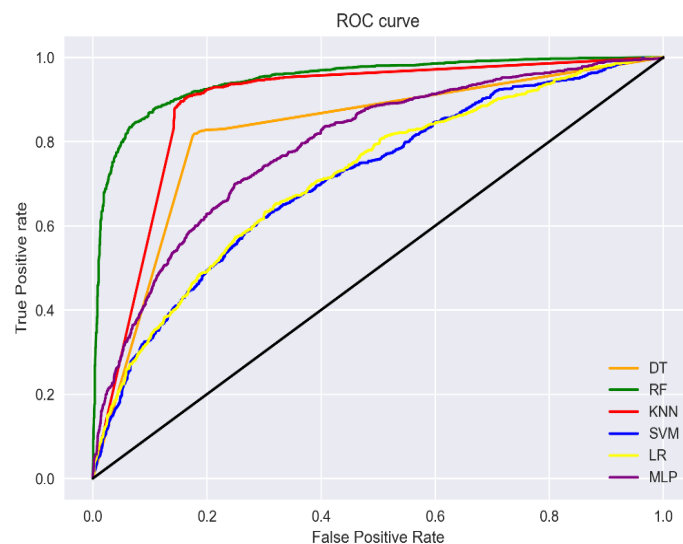


Fig. 2. ROC Curve of used ML techniques

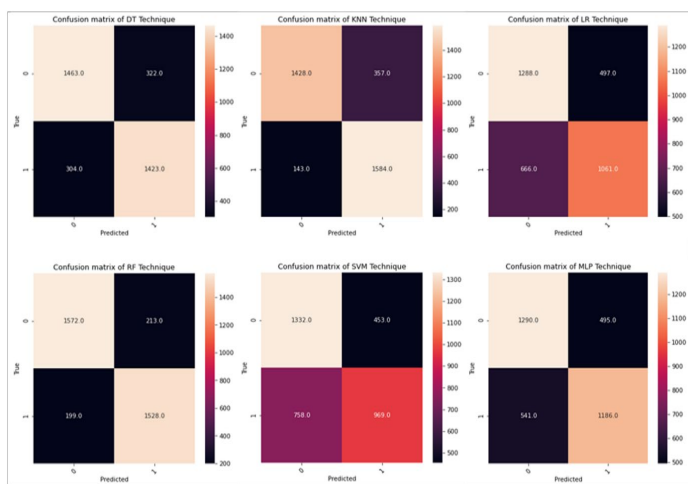


Fig. 3. Confusion matrix of used ML techniques

The proposed system also proved superior when compared with the achieved results of related work as shown in Table 4. The superiority of the RF technique over the RF techniques used in studies [7, 11, 13, 16] as well as over all other techniques used in previous works. Also, the superiority of KNN technique used in this study over the KNN techniques used in the studies [7, 11, 13, 17]. Also, the DT technique used in this study achieved remarkable superiority over the DT techniques used in previous works [7, 15, 17]. The proposed system in this study achieved promising results in predicting software defects, superior to previous works. This study proved that preparing the data well and selecting the effective features helps to improve the prediction accuracy. Also, choosing the appropriate ML technique for this task, in addition to fine tuning its hyperparameters, greatly improves the prediction accuracy.

Table 4

Comparison of the proposed work with related work

Study	Accuracy, %
[15]	80,64
[7]	80,61
[13]	80,44
[11]	82,02
[16]	88,00
[17]	80,00
Our DT	82,17
Our KNN	85,76
Our RF	88,26

CONCLUSION AND FUTURE WORKS

Predicting software defects at the software testing stage is very important to ensure the quality of software before it is offered to users. Therefore, this study aims to build an intelligent system to automatically predict software defects in the software testing phase based on ML techniques. The proposed system consists of three basic stages: data pre-processing, hyperparameter tuning, and application of ML techniques. In the data pre-processing stage, the data was cleaned, balanced, and normalized, in addition to using the DT technique to select the important features. In the second stage, the Grid search technique

is applied to tune the hyperparameters of the ML techniques. In the final stage, ML techniques are applied, and their performance is compared using various scales. Where the RF technique achieved promising results with an accuracy of 88,26 %, superior to other techniques used. RF technique has been shown to be highly effective in predicting software defects when used with an appropriate FS technique. This system was tested only on JM1 dataset, so there is a possibility to test it on other datasets provided by NASA in future works. Individual classifiers can also be combined using ensemble methods to improve performance.

REFERENCES

- Hassan F., Farhan S., Fahiem M. A., Tauseef H. A Review on Machine Learning Techniques for Software Defect Prediction, *Technical Journal*, 2018, Vol. 23, No. 02, Pp. 63–71.
- Ayon S. I. Neural Network Based Software Defect Prediction Using Genetic Algorithm and Particle Swarm Optimization, *Proceedings of the First International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, May 03–05, 2019*. Institute of Electrical and Electronics Engineers, 2019, 4 p. DOI: 10.1109/ICASERT.2019.8934642.
- Jayanthi R., Florence L. Software Defect Prediction Techniques Using Metrics Based on Neural Network Classifier, *Cluster Computing*, 2019, Vol. 22, Suppl. Is. 1, Pp. 77–88. DOI: 10.1007/s10586-018-1730-1.
- Catal C., Diri B. A Systematic Review of Software Fault Prediction Studies, *Expert Systems with Applications*, 2009, Vol. 36, Is. 4, Pp. 7346–7354. DOI: 10.1016/j.eswa.2008.10.027.
- Rashid M., Kaur L. Finding Bugs in Android Application Using Genetic Algorithm and Apriori Algorithm, *Indian Journal of Science and Technology*, 2016, Vol. 9, Is. 23, 5 p. DOI: 10.17485/ijst/2016/v9i23/94572.
- Zhou Y., Shan C., Sun S., et al. Software Defect Prediction Model Based On KPCA-SVM, *Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, United Kingdom, August 19–23, 2019*. Institute of Electrical and Electronics Engineers, 2019, Pp. 1326–1332. DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00244.
- Iqbal A., Aftab S., Ullah I., et al. A Feature Selection Based Ensemble Classification Framework for Software Defect Prediction, *International Journal of Modern Education and Computer Science*, 2019, Vol. 11, No. 9, Pp. 54–64. DOI: 10.5815/ijmecs.2019.09.06.
- Aftab S., Ahmad M., Hameed N., et al. Rainfall Prediction in Lahore City Using Data Mining Techniques, *International Journal of Advanced Computer Science and Applications*, 2018, Vol. 9, Is. 4, Pp. 254–260. DOI: 10.14569/IJACSA.2018.090439.
- Herzig K., Just S., Rau A., Zeller A. Predicting Defects Using Change Genealogies, *Proceedings of the 24th International Symposium on Software Reliability Engineering (ISSRE), Pasadena, CA, USA, November 04–07, 2013*. Institute of Electrical and Electronics Engineers, 2013, Pp. 118–127. DOI: 10.1109/ISSRE.2013.6698911.

10. Pachouly J., Ahirrao S., Kotecha K., et al. A Systematic Literature Review on Software Defect Prediction Using Artificial Intelligence: Datasets, Data Validation Methods, Approaches, and Tools, *Engineering Applications of Artificial Intelligence*, 2022, Vol. 111, Art. No. 104773, 33 p. DOI: 10.1016/j.engappai.2022.104773.
11. Khan B., Naseem R., Shah M. A., et al. Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques, *Journal of Healthcare Engineering*, 2021, Vol. 2021, Art. No. 8899263, 16 p. DOI: 10.1155/2021/8899263.
12. Elsabagh M. A., Farhan M. S., Gafar M. G. Cross-Projects Software Defect Prediction Using Spotted Hyena Optimizer Algorithm, *SN Applied Sciences*, 2020, Vol. 2, Is. 4, Art. No. 538, 15 p. DOI: 10.1007/s42452-020-2320-4.
13. Iqbal A., Aftab S. A Classification Framework for Software Defect Prediction Using Multi-Filter Feature Selection Technique and MLP, *International Journal of Modern Education and Computer Science*, 2020, Vol. 12, No. 1, Pp. 18–25. DOI: 10.5815/ijmecs.2020.01.03.
14. Sayyad Shirabad J., Menzies T. J. The PROMISE Repository of Software Engineering Databases, *School of Information Technology and Engineering, University of Ottawa*. Available at: <http://promise.site.uottawa.ca/SERepository> (accessed 15 Jul 2023).
15. Singh P. D., Chug A. Software Defect Prediction Analysis Using Machine Learning Algorithms, *Proceedings of the 7th International Conference on Cloud Computing, Data Science and Engineering — Confluence, Noida, India, January 12–13, 2017*. Institute of Electrical and Electronics Engineers, 2017, Pp. 775–781. DOI: 10.1109/CONFLUENCE.2017.7943255.
16. Vats R., Kumar A. Software Defects Prediction Using Supervised and Unsupervised Machine Learning Approaches: A Comparative Performance Analysis, *International Journal of Research in Engineering, Technology and Science*, 2021, Vol. 13, Is. 8, 14 p.
17. Shrimankar R., Kuanr M., Piri J., Panda N. Software Defect Prediction: A Comparative Analysis of Machine Learning Techniques, *Proceedings of the 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS), Bhubaneswar, India, August 05–06, 2022*. Institute of Electrical and Electronics Engineers, 2022, Pp. 38–47. DOI: 10.1109/MLCSS57186.2022.00016.
18. Can H., Jianchun X., Ruide Z., et al. A New Model for Software Defect Prediction Using Particle Swarm Optimization and Support Vector Machine, *Proceedings of the 25th Chinese Control and Decision Conference (CCDC), Guiyang, China, May 25–27, 2013*. Institute of Electrical and Electronics Engineers, 2013, Pp. 4106–4110. DOI: 10.1109/CCDC.2013.6561670.
19. Saihood Q., Sonuç E. The Efficiency of Classification Techniques in Predicting Anemia Among Children: A Comparative Study, *Emerging Technology Trends in Internet of Things and Computing (TIOTC 2021): Revised Selected Papers of the First International Conference, Erbil, Iraq, June 06–08, 2021*. Cham, Springer Nature, 2022, Pp. 167–181. DOI: 10.1007/978-3-030-97255-4_12.
20. Naz H., Ahuja S., Nijhawan R., Ahuja N. J. Impact of Data Pre-Processing in Information Retrieval for Data Analytics. In: *Kumar A., et al. (eds.) Machine Intelligence, Big Data Analytics, and IoT in Image Processing: Practical Applications*. Beverly (MA), Scrivener Publishing, 2023, Pp. 197–224. DOI: 10.1002/9781119865513.ch9.
21. Chu X., Ilyas I. F., Krishnan S., Wang J. Data cleaning: Overview and emerging challenges, *SIGMOD '16: Proceedings of the International Conference on Management of Data, San Francisco, CA, USA, June 26–July 01, 2016*. New York, Association for Computing Machinery, 2016, Pp. 2201–2206. DOI: 10.1145/2882903.2912574.
22. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-Sampling Technique, *Journal of Artificial Intelligence Research*, 2002, Vol. 16, Pp. 321–357. DOI: 10.1613/jair.953.
23. Sola J., Sevilla J. Importance of input data normalization for the application of neural networks to complex industrial problems, *IEEE Transactions on Nuclear Science*, 1997, Vol. 44, Is. 3, Pp. 1464–1468. DOI: 10.1109/23.589532.
24. Singh D., Singh B. Investigating the Impact of Data Normalization on Classification Performance, *Applied Soft Computing*, 2020, Vol. 97, Part B, Art. No. 105524, 23 p. DOI: 10.1016/j.asoc.2019.105524.
25. Khaire U. M., Dhanalakshmi R. Stability of Feature Selection Algorithm: A Review, *Journal of King Saud University — Computer and Information Sciences*, 2022, Vol. 34, Is. 4, Pp. 1060–1073. DOI: 10.1016/j.jksuci.2019.06.012.
26. Hota H. S., Shrivastava A. K. Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques. In: *Kundu M. K., et al. (eds.) Advanced Computing, Networking and Informatics — Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014), Kolkata, India, June 24–26, 2014*. Cham, Springer International Publishing, 2014, Pp. 205–211. DOI:10.1007/978-3-319-07353-8_24.
27. Aggarwal S. Machine Learning Algorithms, Perspectives, and Real-World Application: Empirical Evidence from United States Trade Data, *International Journal of Science and Research*, 2023, Vol. 12, Is. 3, Pp. 292–313. DOI: 10.21275/SR23305084601.
28. Zhang W., Huang G., Zheng K. Application of Logistic Regression and Machine Learning Methods for Idiopathic Inflammatory Myopathies Malignancy Prediction, *Clinical and Experimental Rheumatology*, 2023, Vol. 41, Is. 2, Pp. 330–339. DOI: 10.55563/clinexprheumatol/8ievtq.
29. Uddin S., Haque I., Lu H., et al. Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction, *Scientific Reports*, 2022, Vol. 12, Art. No. 6256, 11 p. DOI: 10.1038/s41598-022-10358-x.

Система прогнозирования дефектов программного обеспечения на основе хорошо отлаженной техники «случайного леса»

Ф. Х. Хоршид, Н. Дж. Ибрагим

Университет Диялы
Бакуба, Ирак

farah_hatam@uodiyala.edu.iq, nebras.jalel@uodiyala.edu.iq

К. Сайхунд

Университетский колледж Ашура
Багдад, Ирак

qusaysaihood@au.edu.iq

Аннотация. Качество программного обеспечения является основным критерием для повышения спроса пользователей на программное обеспечение. Поэтому компании, занимающиеся программным обеспечением, стремятся обеспечить качество программного обеспечения путем прогнозирования его дефектов на этапе тестирования. Наличие интеллектуальной системы, способной прогнозировать дефекты программного обеспечения, значительно снижает затраты времени и усилий. Несмотря на широкую тенденцию разработки систем прогнозирования дефектов программного обеспечения на основе техники машинного обучения в последние несколько лет, точность этих систем по-прежнему является серьезной проблемой.

В данном исследовании для повышения точности прогноза представлена система прогнозирования дефектов программного обеспечения, состоящая из трех этапов. На первом этапе выполняется предварительная обработка данных, которая включает в себя очистку данных, баланс данных, нормализацию данных и выбор признаков. На втором этапе гиперпараметры настраиваются по методике Grid Search. Наконец, хорошо отлаженная техника машинного обучения реализована для предсказания дефектов программного обеспечения.

На базе набора данных JM1 были проведены эксперименты, в ходе которых предлагаемая система дала многообещающие результаты в прогнозировании недостатков программного обеспечения. Среди используемых методов хорошо настроенный метод Random Forest с точностью 88,26 % превзошел остальные используемые методы машинного обучения. Проведенное исследование доказывает, что выбор важных особенностей и эффективная гиперпараметрическая настройка методов машинного обучения значительно улучшают точность прогнозирования дефектов программного обеспечения.

Ключевые слова: машинное обучение, случайный лес, дефекты программного обеспечения, выбор признаков, прогнозирование.

ЛИТЕРАТУРА

1. A Review on Machine Learning Techniques for Software Defect Prediction / F. Hassan, S. Farhan, M. A. Fahiem, H. Tauseef // Technical Journal. 2018. Vol. 23, No. 02. Pp. 63–71.
2. Ayon, S. I. Neural Network Based Software Defect Prediction Using Genetic Algorithm and Particle Swarm Optimization // Proceedings of the First International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) (Dhaka, Bangladesh, 03–05 May 2019). — Institute of Electrical and Electronics Engineers, 2019. — 4 p. DOI: 10.1109/ICASERT.2019.8934642.

3. Jayanthi, R. Software Defect Prediction Techniques Using Metrics Based on Neural Network Classifier / R. Jayanthi, L. Florence // Cluster Computing. 2019. Vol. 22, Suppl. Is. 1. Pp. 77–88. DOI: 10.1007/s10586-018-1730-1.

4. Catal, C. A Systematic Review of Software Fault Prediction Studies / C. Catal, B. Diri // Expert Systems with Applications. 2009. Vol. 36, Is. 4. Pp. 7346–7354. DOI: 10.1016/j.eswa.2008.10.027.

5. Rashid, M. Finding Bugs in Android Application Using Genetic Algorithm and Apriori Algorithm / M. Rashid, L. Kaur // Indian Journal of Science and Technology. 2016. Vol. 9, Is. 23. 5 p. DOI: 10.17485/ijst/2016/v9i23/94572.

6. Software Defect Prediction Model Based On KPCA-SVM / Y. Zhou, C. Shan, S. Sun, [et al.] // Proceedings of the 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI) (Leicester, United Kingdom, 19–23 August 2019). — Institute of Electrical and Electronics Engineers, 2019. — Pp. 1326–1332. DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00244.

7. A Feature Selection Based Ensemble Classification Framework for Software Defect Prediction / A. Iqbal, S. Aftab, I. Ullah, [et al.] // International Journal of Modern Education and Computer Science. 2019. Vol. 11, No. 9. Pp. 54–64. DOI: 10.5815/ijmecs.2019.09.06.

8. Rainfall Prediction in Lahore City Using Data Mining Techniques / S. Aftab, M. Ahmad, N. Hameed, [et al.] // International Journal of Advanced Computer Science and Applications. 2018. Vol. 9, Is. 4. Pp. 254–260. DOI: 10.14569/IJACSA.2018.090439.

9. Predicting Defects Using Change Genealogies / K. Herzig, S. Just, A. Rau, A. Zeller // Proceedings of the 24th International Symposium on Software Reliability Engineering (ISSRE) (Pasadena, CA, USA, 04–07 November 2013). — Institute of Electrical and Electronics Engineers, 2013. — Pp. 118–127. DOI: 10.1109/ISSRE.2013.6698911.

10. A Systematic Literature Review on Software Defect Prediction Using Artificial Intelligence: Datasets, Data Validation Methods, Approaches, and Tools / J. Pachouly, S. Ahirrao, K. Kotecha, [et al.] // Engineering Applications of Artificial Intelligence. 2022. Vol. 111. Art. No. 104773. 33 p. DOI: 10.1016/j.engappai.2022.104773.

11. Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques / B. Khan, R. Naseem, M. A. Shah, [et al.] // *Journal of Healthcare Engineering*. 2021. Vol. 2021. Art. No. 8899263. 16 p. DOI: 10.1155/2021/8899263.
12. Elsabagh, M. A. Cross-Projects Software Defect Prediction Using Spotted Hyena Optimizer Algorithm / M. A. Elsabagh, M. S. Farhan, M. G. Gafar // *SN Applied Sciences*. 2020. Vol. 2, Is. 4. Art. No. 538. 15 p. DOI: 10.1007/s42452-020-2320-4.
13. Iqbal, A. A Classification Framework for Software Defect Prediction Using Multi-Filter Feature Selection Technique and MLP / A. Iqbal, S. Aftab // *International Journal of Modern Education and Computer Science*. 2020. Vol. 12, No. 1. Pp. 18–25. DOI: 10.5815/ijmecs.2020.01.03.
14. Sayyad Shirabad, J. The PROMISE Repository of Software Engineering Databases / J. Sayyad Shirabad, T. J. Menzies // *School of Information Technology and Engineering, University of Ottawa*. Available at: <http://promise.site.uottawa.ca/SERepository> (дата обращения 15.07.2023).
15. Singh, P. D. Software Defect Prediction Analysis Using Machine Learning Algorithms / P. D. Singh, A. Chug // *Proceedings of the 7th International Conference on Cloud Computing, Data Science and Engineering — Confluence* (Noida, India, 12–13 January 2017). — Institute of Electrical and Electronics Engineers, 2017. — Pp. 775–781. DOI: 10.1109/CONFLUENCE.2017.7943255.
16. Vats, R. Software Defects Prediction Using Supervised and Unsupervised Machine Learning Approaches: A Comparative Performance Analysis / R. Vats, A. Kumar // *International Journal of Research in Engineering, Technology and Science*. 2021. Vol. 13, Is. 8. 14 p.
17. Software Defect Prediction: A Comparative Analysis of Machine Learning Techniques / R. Shrimankar, M. Kuanr, J. Piri, N. Panda // *Proceedings of the 2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)* (Bhubaneswar, India, 05–06 August 2022). — Institute of Electrical and Electronics Engineers, 2022. — Pp. 38–47. DOI: 10.1109/MLCSS57186.2022.00016.
18. A New Model for Software Defect Prediction Using Particle Swarm Optimization and Support Vector Machine / H. Can, X. Jianchun, Z. Ruide, [et al.] // *Proceedings of the 25th Chinese Control and Decision Conference (CCDC)* (Guiyang, China, 25–27 May 2013). — Institute of Electrical and Electronics Engineers, 2013. — Pp. 4106–4110. DOI: 10.1109/CCDC.2013.6561670.
19. Saihood, Q. The Efficiency of Classification Techniques in Predicting Anemia Among Children: A Comparative Study / Q. Saihood, E. Sonuç // *Emerging Technology Trends in Internet of Things and Computing (TIOTC 2021): Revised Selected Papers of the First International Conference* (Erbil, Iraq, 06–08 June 2021). — Cham: Springer Nature, 2022. — Pp. 167–181. — (Communications in Computer and Information Science, Vol. 1548). DOI: 10.1007/978-3-030-97255-4_12.
20. Impact of Data Pre-Processing in Information Retrieval for Data Analytics / H. Naz, S. Ahuja, R. Nijhawan, N. J. Ahuja // *Machine Intelligence, Big Data Analytics, and IoT in Image Processing: Practical Applications* / A. Kumar, [et al.] (eds.). — Beverly (MA): Scrivener Publishing, 2023. — Pp. 197–224. DOI: 10.1002/9781119865513.ch9.
21. Data cleaning: Overview and emerging challenges / X. Chu, I. F. Ilyas, S. Krishnan, J. Wang // *SIGMOD '16: Proceedings of the International Conference on Management of Data* (San Francisco, CA, USA, 26 June–01 July 2016). — New York: Association for Computing Machinery, 2016. Pp. 2201–2206. DOI: 10.1145/2882903.2912574.
22. SMOTE: Synthetic Minority Over-Sampling Technique / N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer // *Journal of Artificial Intelligence Research*. 2002. Vol. 16. Pp. 321–357. DOI: 10.1613/jair.953.
23. Sola, J. Importance of input data normalization for the application of neural networks to complex industrial problems / J. Sola, J. Sevilla // *IEEE Transactions on Nuclear Science*. 1997. Vol. 44, Is. 3. Pp. 1464–1468. DOI: 10.1109/23.589532.
24. Singh, D. Investigating the Impact of Data Normalization on Classification Performance / D. Singh, B. Singh // *Applied Soft Computing*. 2020. Vol. 97, Part B. Art. No. 105524. 23 p. DOI: 10.1016/j.asoc.2019.105524.
25. Khaire, U. M. Stability of Feature Selection Algorithm: A Review / U. M. Khaire, R. Dhanalakshmi // *Journal of King Saud University — Computer and Information Sciences*. 2022. Vol. 34, Is. 4. Pp. 1060–1073. DOI: 10.1016/j.jksuci.2019.06.012.
26. Hota, H. S. Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques // H. S. Hota, A. K. Shrivastava // *Advanced Computing, Networking and Informatics — Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014)* (Kolkata, India, 24–26 June 2014) / M. K. Kundu, [et al.] (eds.). — Cham: Springer International Publishing, 2014. — Pp. 205–211. — (Smart Innovation, Systems and Technologies, Vol. 27). DOI: 10.1007/978-3-319-07353-8_24.
27. Aggarwal, S. Machine Learning Algorithms, Perspectives, and Real-World Application: Empirical Evidence from United States Trade Data // *International Journal of Science and Research*. 2023. Vol. 12, Is. 3. Pp. 292–313. DOI: 10.21275/SR23305084601.
28. Application of Logistic Regression and Machine Learning Methods for Idiopathic Inflammatory Myopathies Malignancy Prediction / W. Zhang, G. Huang, K. Zheng // *Clinical and Experimental Rheumatology*. 2023. Vol. 41, Is. 2. Pp. 330–339. DOI: 10.55563/clinexprheumatol/8ievtq.
29. Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction / S. Uddin, I. Haque, H. Lu, [et al.] // *Scientific Reports*. 2022. Vol. 12. Art. No. 6256. 11 p. DOI: 10.1038/s41598-022-10358-x.

Russian version of the article © N. A. Slobodchikov, A. I. Dergachev, O. N. Kuranova is published in *Proceedings of Petersburg Transport University*, 2023, Vol. 20, Is. 2, Pp. 279–289.
DOI: 10.20295/1815-588X-2023-2-279-289.

Prospective Development of New Service Offerings in the Urban Traffic Flow

PhD N. A. Slobodchikov, PhD A. I. Dergachev, PhD O. N. Kuranova
Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia
kola_slob@mail.ru, d_ader@mail.ru, olga_kuranova@mail.ru

Abstract. One of the possible solutions to the problem of congestion of the road network in conditions of a high level of motorization of cities is considered. The concept of the service «Mobility as a Service» as one of the progressive methods of urban traffic management is described. The technologies that allow this service to work, while increasing the efficiency of the use of vehicles, are considered. The analysis of carsharing, kicksharing and bikesharing services as a «first and last mile» transport system was carried out.

The solution to this problem can be the integration of all existing transport systems into a single system that can have a greater impact on the efficiency of mobility of urban users.

Keywords: mobility as a service, urban transport, public transport, rental system, carsharing, kicksharing, bikesharing, first and last mile transport.

The high rates of development of transport mobility of the population in modern cities around the world have posed a serious problem to society: first of all, it is the deterioration of the

environmental situation, secondly, congestion of the road network and, as a consequence, a decrease in traffic speed, and thirdly, an increase in the costs associated with the transportation of passengers.

All mankind strives to solve this problem in various ways, of the possible solutions to this problem, it can be noted that the most effective solution is the introduction of progressive methods of urban traffic management. One of these methods is the implementation of the «Mobility as a Service» (MaaS) service in the urban environment, which implies the possibility of an operational choice of various modes of transport for trips, and contributes to the refusal of people to use a personal car in cities.

The MaaS is primarily the integration of various forms of transport services into a «Single mobility Service» available at the request of the client. In an enlarged form, the basic structure of the MaaS can be presented in the following form: (Fig. 1) [1].



Fig. 1. The structure of the MaaS

As can be seen from the figure, the MaaS implies a change in the form of vehicle ownership, in other words, one vehicle should serve as many people as possible. What technologies allow this service to work and improve the efficiency of using vehicles. This is primarily carsharing.

Carsharing is a term that reflects the essence of the modern approach to using a car. It belongs to the emerging class of «mobility services» that use modern technologies to provide access to mobility using a car without the consumer owning a personal vehicle. It acts as a logical addition to the traditional ways of providing people with transportation needs, such as taxis and traditional car rental. Their main difference from carsharing is the absence of the need for modern information and communication technologies for commercial viability, at the same time they also use new technologies to improve customer service [2].

To date, more than 50 000 drivers use carsharing in Russia, the total number of cars involved in the business is about 5 000, it is constantly growing. The fleet consists of several tens of thousands of vehicles.

The carsharing service usually involves access to a car owned by another individual or legal entity for a fixed cost. During the use of the carsharing services, the user is responsible for the car and its condition.

This service is very popular in Russia, so in 2021 the Moscow Traffic Management Center conducted a study of the carsharing system in Moscow for 2021 to obtain a qualitative and quantitative assessment of this service [3].

As a result of its implementation, data were obtained that showed that in 2021 Muscovites used carsharing services 49 million times, which is an average of 135 thousand trips per day, for comparison, in 2021 the metro was used 2,1 billion times, taxis — 328 million times, and rental scooters and bicycles — 13,3 million times. Figure 2 shows the distribution diagram of the average number of trips per day.

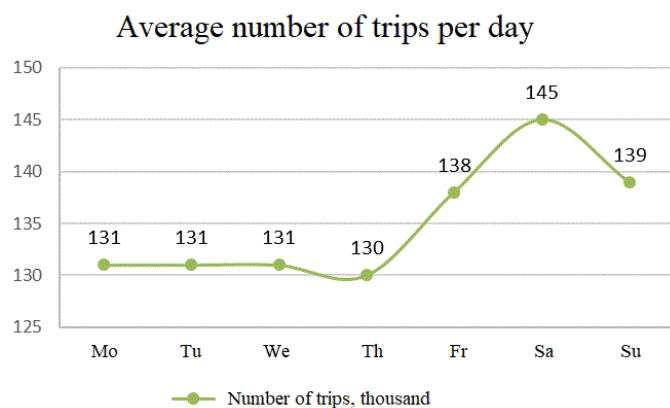


Fig. 2. Trip distribution diagram

As can be seen from the diagram, the greatest demand for carsharing services is observed on Saturday throughout the year, this is due to the fact that more users on this day had weekend trips for personal business. At the same time, the lowest demand was observed on Thursdays and Mondays, this is due to the fact that on weekdays users preferred to use public transport.

When analyzing the distribution of trips by month, data were obtained showing that the greatest increase in demand for carsharing services is observed in the summer months, this is

due to an increase in users during the holidays. At the same time, the lowest demand falls on the winter months of the year, the reason for this may be the use of public transport as a safer alternative to movement, regardless of weather and climatic conditions. This pattern can be seen in Figure 3, which shows a diagram of the distribution of the number of trips by month of the year.

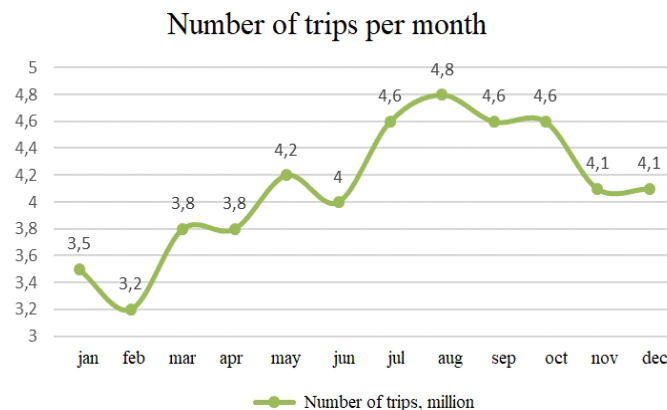


Fig. 3. Diagram of the distribution of the number of trips by month of the year

The difference between working and non-working days of the week is especially noticeable in the morning and afternoon indicators. On weekends during the morning rush hour, the number of trips is significantly lower than the number of trips on weekdays from 5 to 9 o'clock, while on working days there are fewer orders during the day — at this time, service users are at work or at school. The greatest demand for the use of carsharing services falls on the morning and evening rush hours on weekdays, at which time users make 33 % of trips, while the average distance of a carsharing trip is 16 km, and the average car rental time is 54 minutes from the moment of booking in the application. Figure 4 shows a diagram of the distribution of the share of carsharing trips by time of day.

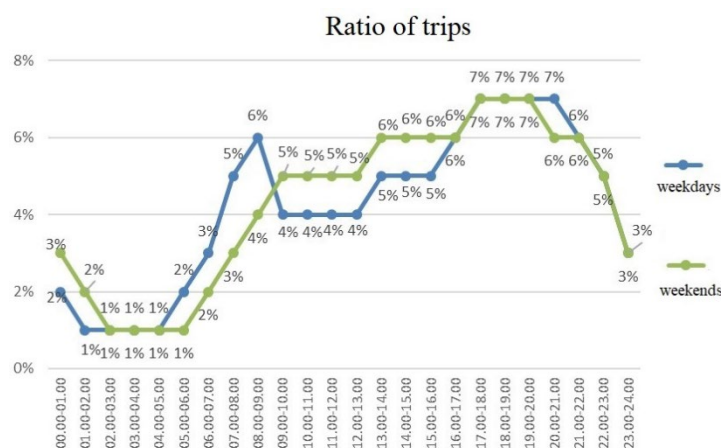


Fig. 4. Diagram of the distribution of the share of carsharing trips by time of day

The obtained results of the analysis allow us to state that there is a positive trend towards a significant increase in the use of carsharing services, first of all, the growth in demand for services may be due to the fact that owning your own car becomes much more expensive due to increased costs for repairs and

maintenance, parking fees, as a result, the maintenance of a personal car becomes not only it's just not profitable, but on the contrary burdensome.

Another rapidly developing way to meet the need for movement for the population is kicksharing, which implies short-term rental of scooters and electric scooters by analogy with carsharing.

There is a trend in the world when users no longer perceive scooters as just entertainment, but began to perceive them as «last mile transport», which is used for short trips inside and between city districts. Moreover, this service is increasingly used not only by young people under 25, but also by mature people from 30-50 years old. To determine measures to improve this type of service, a study was conducted by analysts of the Moscow Traffic Management Center in 2021, during which a qualitative and quantitative assessment of the service was carried out.

During the study, it was revealed that in 2021, the demand for kicksharing increased significantly, 8,6 million times used scooter rental services during the year, whereas in 2020 — about 2 million times. During the year under review, Muscovites rode scooters 8,6 million times — an average of 35 thousand rentals every day, but at the same time such alternatives to personal transport as carsharing and taxis remain more popular than means of individual mobility. Figure 5 shows a diagram of the distribution of the number of trips by day of the week in the study period.

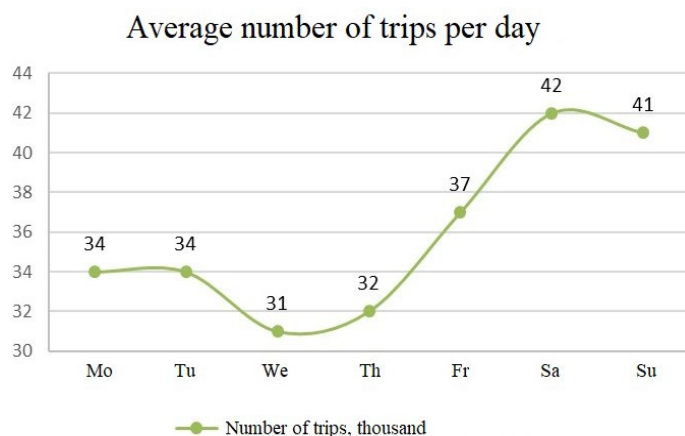


Fig. 5. Diagram of the distribution of the number of trips

As it can be seen from it, there is a greater demand for scooter rental services on weekends, while kicksharing is used in different ways on working and non-working days. On weekends and holidays, demand traffic is gradually increasing from 8 am to 6 pm, while on weekdays there are two peaks: in the morning from 9 to 10 and in the evening from 18 to 20, the appearance of peak demand is due to the fact that users are increasingly using a scooter to make short trips. Figure 6 shows a diagram of the distribution of the share of trips on scooters.

The kicksharing period during the year depends more on weather conditions, so the analysis of the distribution of trips by month was carried out from April to November showed that the growth in demand for scooter rental services is observed in the summer months, this is due to an increase in users during the holidays. In the spring and autumn periods of the year, there is a reduction in the number of trips relative to the summer period, primarily this can be explained by weather conditions,

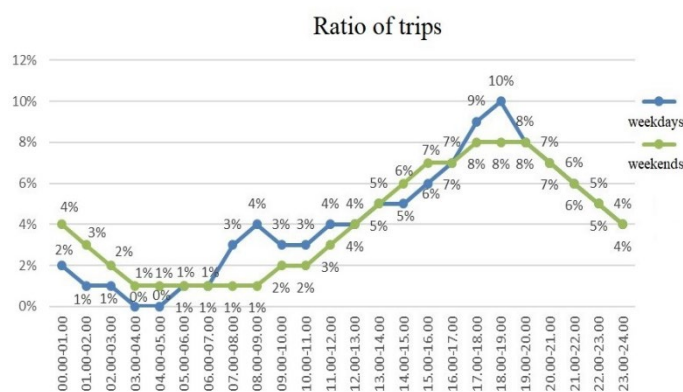


Fig. 6. Diagram of the distribution of the share of trips on scooters

since it was much safer and more comfortable for users to use public transport services. On average, 18 thousand scooters were available to users per day in August, while 8 out of 10 scooters were rented during the day. Figure 7 shows a diagram of the distribution of trips by month of the year.



Fig. 7. Diagram of the distribution of trips by month of the year

Analysts of the Moscow Traffic Management Center compiled a heat map showing the density of the distribution of trips during the day, the richer and warmer the color of the map, the more trips ended in this segment [3]. Figure 8 shows a heat map of Moscow, which clearly shows the distribution of orders for the use of scooters.

Based on the heat map obtained as a result of the conducted research, it can be concluded that scooters become part of multimodal routes — those in which different vehicles are used, with both the initial or final segment of the path separately, and the user can use these services both at the beginning and at the end of his journey.

Scooter rentals were most popular in the immediate vicinity of the metro, as users preferred to use the rental service to get to and from the station to the end point of their route. Figure 9 shows a fragment of a heat map showing the segment with the highest demand for scooter rental services.

Another technology that can meet the need of users to move in urban conditions is a bicycle rental service or bike sharing.

Bikesharing is a mobility service that is based on short-term access to rented bicycles as needed.

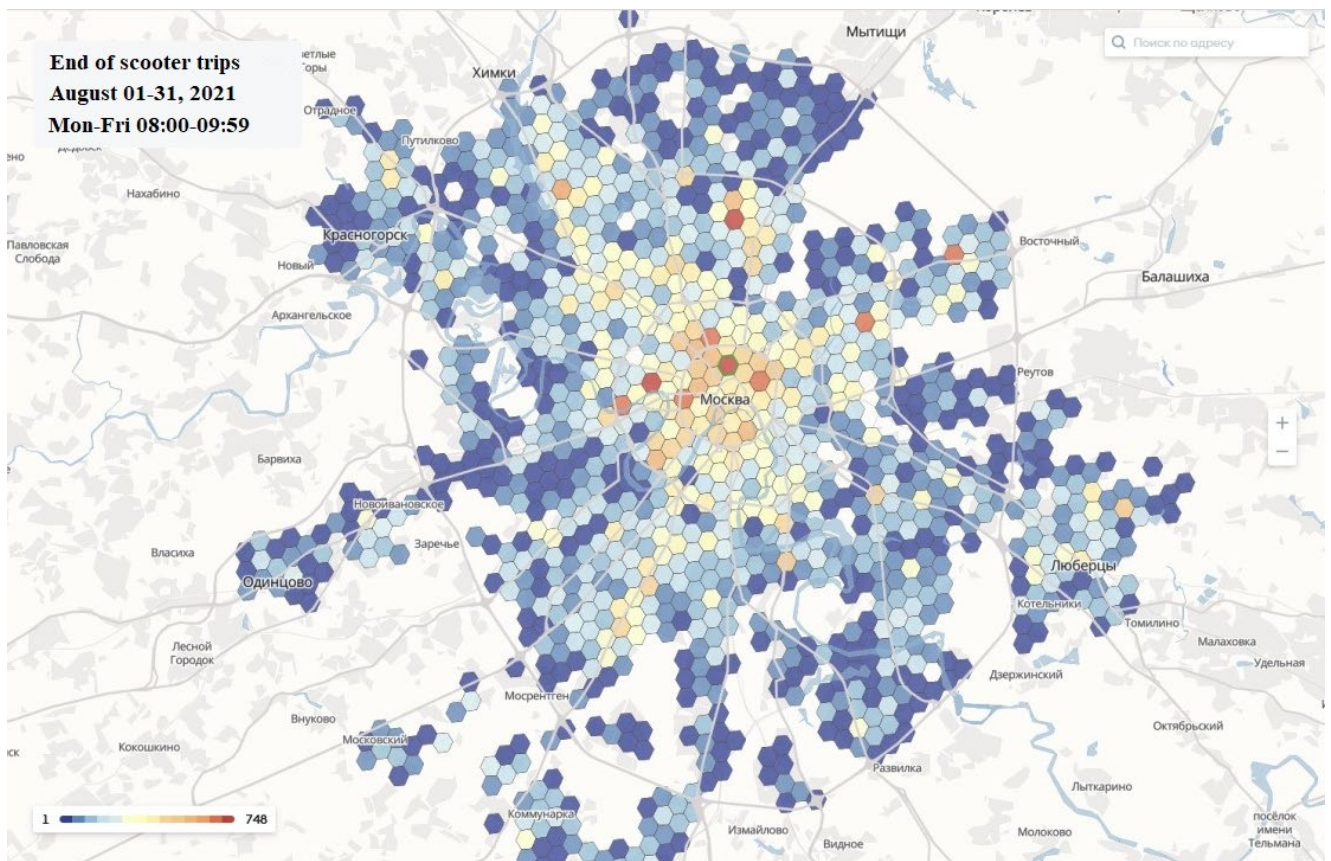


Fig. 8. Heat map of Moscow

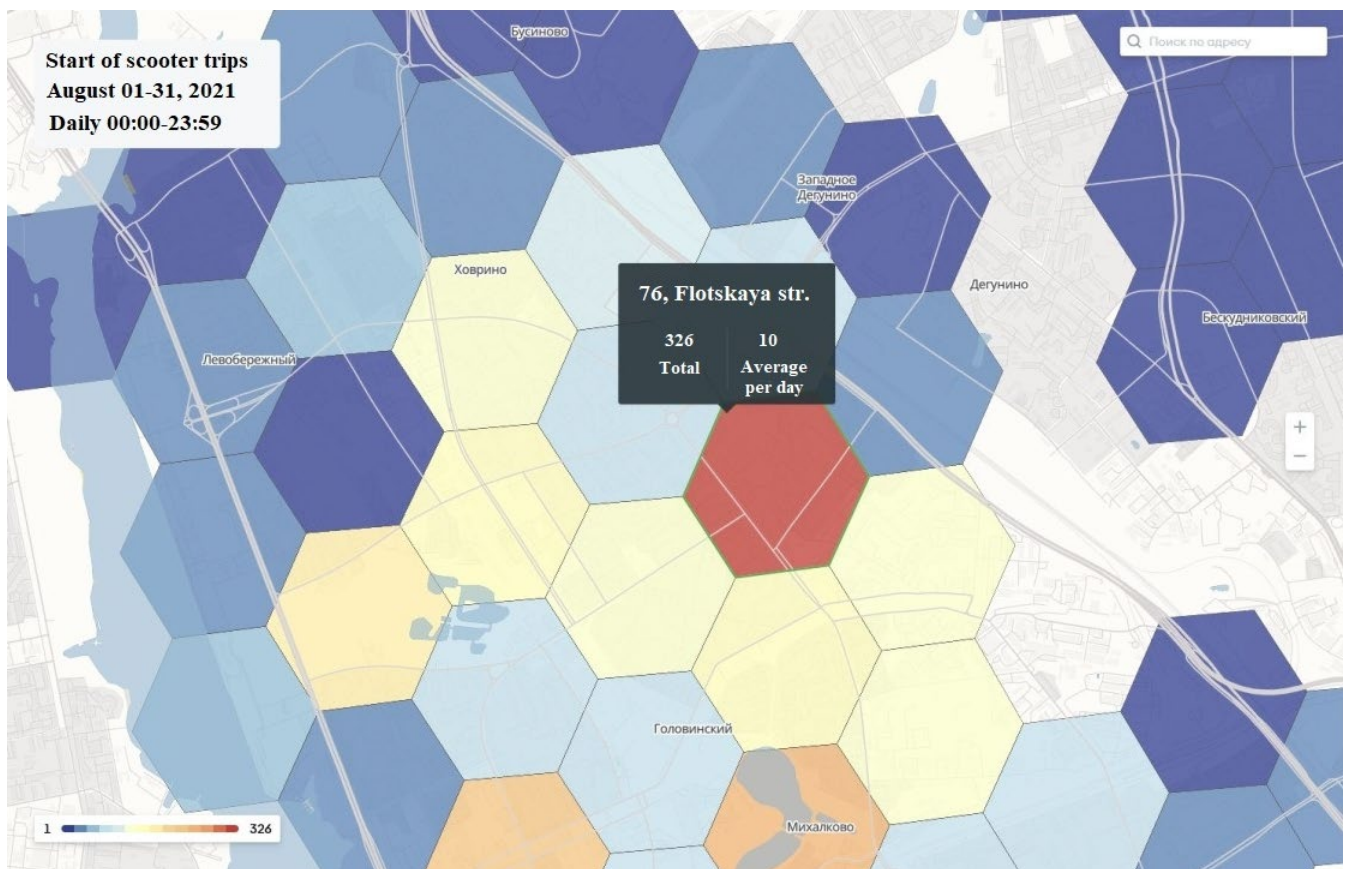


Fig. 9. A fragment of the heat map

In conditions of high motorization of cities, bicycle rental is considered as an integral part of the urban transport infrastructure and serves as an addition to public transport, which is becoming increasingly popular all over the world. Today, there

are more than 1 600 providers offering a total of 18 million bicycles for sharing, while in 2005 there were only 17 bike services in the world. Figure 10 illustrates a diagram of the global expansion of bicycle sharing systems [4].

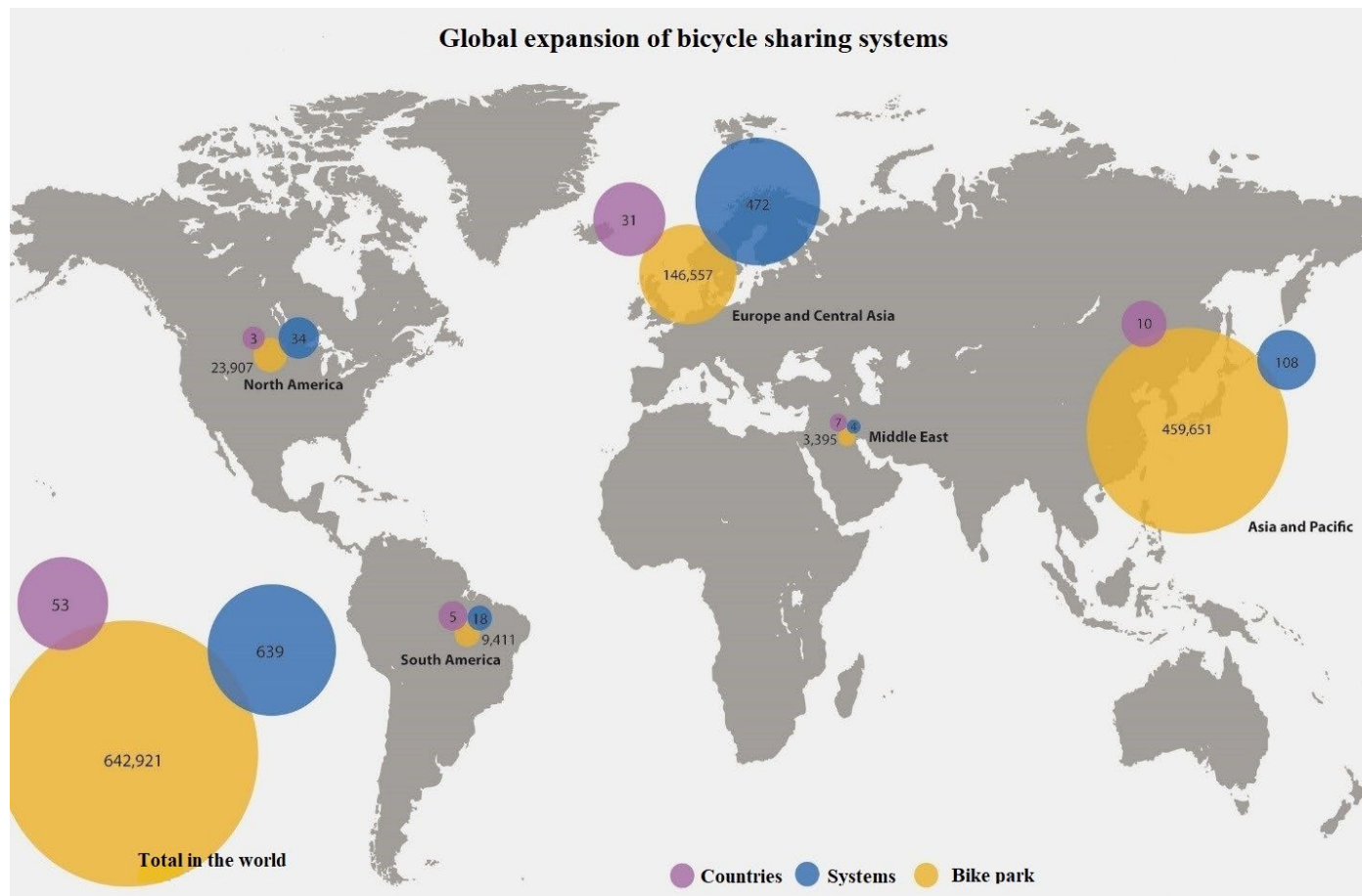


Fig. 10. Diagram of the global expansion of bicycle sharing systems

In Russia, the first city to launch a bicycle rental system was Moscow, the pilot project began in 2013. During the first year of the program, more than 64 thousand people used the bicycle rental service, but by 2015 the number of users had increased 7 times. The bicycle infrastructure in Moscow was created from scratch, but by 2015 the total length of bike paths in Moscow was 250 km, which significantly improved the safety of road users [3].

To date, there are more than 773 km of bike paths, 700 rental stations and 7 000 bicycles in the capital. A significant increase in bike paths and, in general, the development of cycling infrastructure contributed to the fact that, on average, there are more than 6,1 trips per bike per day in the capital, which is more than 2 times more than in London — one of the leading cities in the popularity of bike rental — in the British capital, there are 2,6 trips per bike on average. Such indicators mean that users began to switch to a bicycle in order to make targeted trips, and not just for leisure. Figure 11 shows a map of bike paths in Moscow, Figure 12 shows a map of bicycle rental stations.

New mobility services are changing the transport sector, either by providing completely new mobile solutions, or by changing traditional vehicles through the introduction of progressive methods of urban traffic management [5].

CONCLUSION

In conclusion, would like to note that the growth of new MaaS services was associated with a reduction in the use of personal cars and an increase in the use of public transport. In general, new mobility services replace more trips by private transport than by public transport.

1. The concepts considered, such as kicksharing, bikesharing, carsharing, do not fulfill the same goals as a personal vehicle or public transport; they only partially complement. Rental systems are able to solve the problem of the «first and last mile», which is associated with access to public transport or movement within the area.

2. The services of the MaaS are usually used in combination with public transport, as they allow expanding the coverage area of public transport and, in general, ensuring greater connectivity of the transport network of cities as a result of the development of multimodal trips.

Today, the main problem and difficulty in implementing the MaaS is that all transport systems, including public transport and rental services, are represented by separate segments of demand, and are not parts of a single system. The solution to this problem can be the integration of all existing transport systems into a single system that can have a greater impact on the efficiency of mobility of urban users.

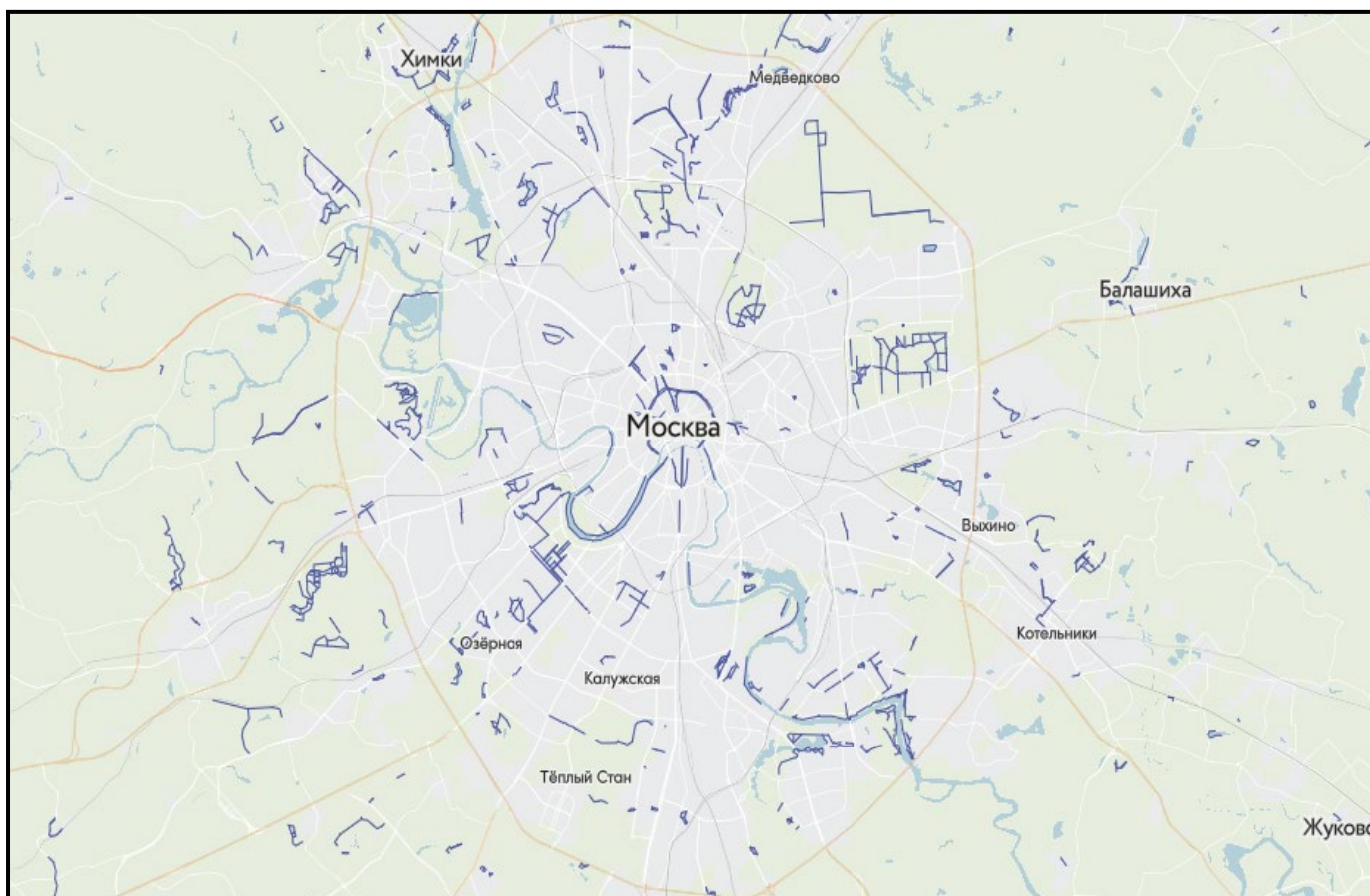


Fig. 11. Map of bike paths in Moscow

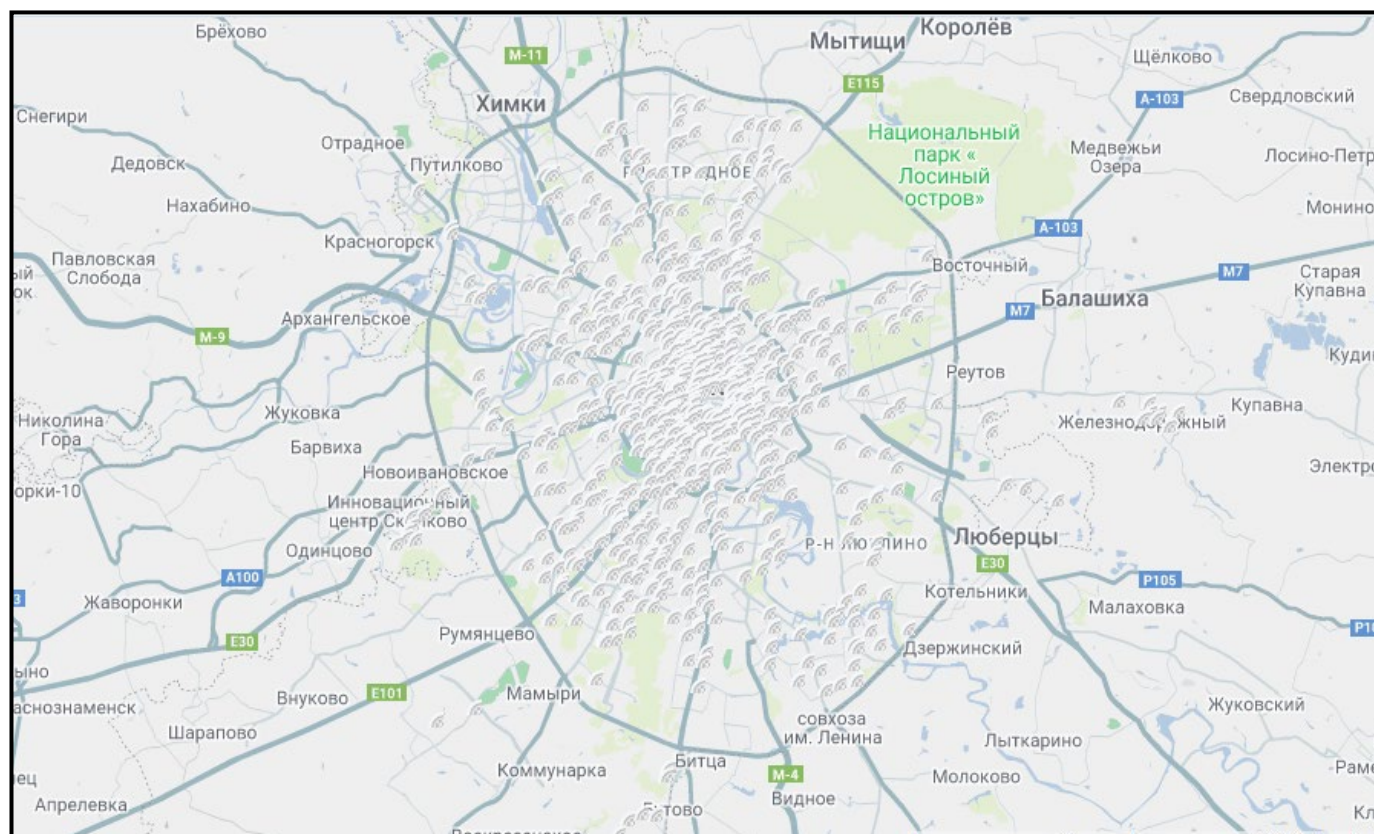


Fig. 12. Map of rental stations

REFERENCES

1. Ready for MaaS? Easier Mobility for Citizens and Better Data for Cities, *International Association of Public Transport (UITP)*. Available at: <http://www.uitp.org/publications/ready-for-maas-easier-mobility-for-citizens-and-better-data-for-cities> (accessed 23 Nov 2022).
2. Andreeva E. A., Belkova E. V., Solodkiy A. I., et al. *Upravlenie transportnymi potokami v gorodakh: Monografiya* [Traffic management in cities: Monograph]. Moscow, Infra-M Publishing House, 2019, 207 p. (In Russ.)
3. Karshering v 2021: kogda, kuda i otkuda ezdili moskvichi [Carsharing in 2021: When, Where and From Where Muscovites Went], *Data-blog «Razvyazka»: sayt Tsentra organizatsii dorozhnogo dvizheniya Pravitelstva Moskvy* [Interchange Data Blog: Website of the Center for Traffic Management of the Government of Moscow]. Available at: <http://i.transport.mos.ru/fly-over/carsharing> (accessed 30 Nov 2022). (In Russ.)
4. Bike-sharing: globalnaya ekspansiya sistem sovместnogo ispolzovaniya velosipedov [Bike-sharing: Global Expansion of Bike Sharing Systems], *Velomesto*. Available at: <http://dev.velomesto.com/magazine/v-mire/bike-sharing-globalnaya-ekspansiya-sistem-sovmestnogo-ispolzovaniya-velosipedov> (accessed 30 Nov 2022). (In Russ.)
5. Recommendations on a User-Centric Approach for MaaS. MaaS Alliance, 2019, 9 p. Available at: <http://maas-alliance.eu/wp-content/uploads/2019/04/Recommendations-on-a-User-Centric-Approach-for-MaaS-FINAL-180419.pdf> (accessed 01 Dec 2022).

Перспективное развитие новых услуг сервиса в городском транспортном потоке

к.воен.н. Н. А. Слободчиков, к.воен.н. А. И. Дергачев, к.т.н. О. Н. Куранова
Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
kola_slob@mail.ru, d_ader@mail.ru, olga_kuranova@mail.ru

Аннотация. Рассматривается один из возможных вариантов решения проблемы загруженности улично-дорожной сети в условиях высокого уровня автомобилизации городов. Описана концепция сервиса «Мобильность как услуга» как одного из прогрессивных методов управления городскими транспортными потоками. Рассмотрены технологии, позволяющие работать данному сервису, при этом повышая эффективность использования транспортных средств. Проведен анализ сервисов каршеринга, кикшеринга и байкшеринга как системы транспорта «первой и последней мили».

Решением данной проблемы может стать интеграция всех существующих транспортных систем в единую систему, способную в большей степени повлиять на эффективность мобильности пользователей городов.

Ключевые слова: мобильность как услуга, городской транспорт, общественный транспорт, система проката, каршеринг, кикшеринг, байкшеринг, транспорт первой и последней мили.

ЛИТЕРАТУРА

1. Ready for MaaS? Easier mobility for citizens and better data for cities // International Association of Public Transport (UITP). URL: <http://www.uitp.org/publications/ready-for-maas-easier-mobility-for-citizens-and-better-data-for-cities> (дата обращения 23.11.2022).

2. Управление транспортными потоками в городах: Монография / Е. А. Андреева, Е. В. Белкова, А. И. Солодкий, [и др.]. — Москва: Инфра-М, 2019. — 207 с. — (Научная мысль).

3. Каршеринг в 2021: когда, куда и откуда ездили москвичи // Дата-блог «Развязка»: сайт Центра организации дорожного движения Правительства Москвы. URL: <http://i.transport.mos.ru/flyover/carsharing> (дата обращения 30.11.2022).

4. Bike-sharing: глобальная экспансия систем совместного использования велосипедов // Velomesto — информационный журнал для велосипедистов. — 2014. — 24 февраля. URL: <http://dev.velomesto.com/magazine/v-mire/bike-sharing-globalnaya-ekspansiya-sistem-sovmestnogo-ispolzovaniya-velosipedov> (дата обращения 30.11.2022).

5. Recommendations on a User-Centric Approach for MaaS. — MaaS Alliance, 2019. — 9 p. URL: <http://maas-alliance.eu/wp-content/uploads/2019/04/Recommendations-on-a-User-Centric-Approach-for-MaaS-FINAL-180419.pdf> (дата обращения 01.12.2022).

Оптимизация образовательного процесса: построение индивидуальных учебных траекторий с помощью вариативной части учебного плана на примере задачи о рюкзаке

Д. Ю. Федоров, к.и.н. А. В. Забродин

Петербургский государственный университет путей сообщения Императора Александра I
Санкт-Петербург, Россия
elfbrus@gmail.com, teach-case@yandex.ru

Аннотация. Исследуется использование комбинаторных оптимизационных задач для формирования вариативной части учебного плана с целью построения индивидуальных образовательных траекторий обучения. Рассматривается конкретная задача о рюкзаке в контексте оптимизации обучения. Представлен метод формирования индивидуальных учебных планов на основе анализа потребностей и способностей учащихся. Особое внимание уделяется процессу адаптации учебных программ под конкретные потребности студентов, что способствует более эффективному и персонализированному обучению. Результаты исследования могут быть полезными для образовательных учреждений и педагогов, стремящихся улучшить качество образования и поддержать индивидуальный образовательный опыт каждого обучающегося.

Ключевые слова: учебный план, дисциплина, алгоритм, задача о рюкзаке, траектория, оптимизация, статистика, сортировка.

ВВЕДЕНИЕ

С увеличением численности студентов современные образовательные учреждения сталкиваются с вызовом поддержания высокого качества обучения.

Особенно это актуально для университетов, где численность студентов часто достигает значительных масштабов. Такие учебные заведения стоят перед задачей организации и планирования обучения для огромного числа обучающихся, каждый из которых имеет свои индивидуальные образовательные потребности. Эта ситуация требует применения более алгоритмичного и автоматизированного подхода к решению организационных задач.

В статье предложен вариант эффективного решения, направленного на автоматизацию процесса подбора дисциплин для вариативной части учебного плана студентами. Этот подход предоставляет студентам большую гибкость в выборе дисциплин, соответствующих их интересам и академическим целям, и при этом не требует активного вмешательства административного персонала или вышестоящего руководства.

Все больше и больше образовательных организаций автоматизируют процесс взаимодействия с обучающимися. В частности, это касается университетов, число студентов в которых составляет несколько тысяч. В таких условиях необходимо более алгоритмичное и автоматизированное взаимодействие по организационным вопросам. В статье

рассмотрен способ автоматизации подбора дисциплин вариативной части студентом, облегчающим процесс выбора направлений, не требующий участия вышестоящего руководства.

ПРЕДМЕТНАЯ ОБЛАСТЬ

Образовательная программа подразделяется на два основных компонента: обязательную и вариативную часть.

Обязательная часть представляет собой набор учебных предметов или курсов, которые студенты обязаны изучать в рамках своей учебной программы. Эти предметы могут включать в себя основные дисциплины, необходимые для получения образовательного стандарта или квалификации. Обязательная часть также включает в себя фундаментальные знания и навыки, необходимые для успешного завершения образовательной программы.

Вариативная часть предоставляет студентам возможность выбирать определенное количество учебных предметов или курсов из предложенного учебным заведением перечня. Эта часть программы позволяет студентам настраивать свою образовательную траекторию с учетом их интересов, целей и специализации. Вариативная часть значительно способствует индивидуализации обучения и позволяет студентам глубже и шире исследовать предметы, которые соответствуют их учебным интересам и будущим карьерным планам.

Таким образом, комбинация обязательной и вариативной части образовательной программы обеспечивает баланс между установленными стандартами обучения и индивидуальной гибкостью в выборе учебных материалов, что способствует более качественному и персонализированному обучению.

Дисциплины в вариативной части (ВЧ) представляют собой дополнительные учебные предметы. Обычно учебное заведение предоставляет ограниченное количество часов или времени, в рамках которого студент может выбрать наиболее интересные для него дисциплины из предложенного перечня.

Рассмотрим пример. Имеется 100 часов, в которые необходимо поместить следующие предметы: «Теория вероятности» — 25 часов, «Операционные системы» — 50 часов, «Информатика» — 50 часов, «Линейная алгебра» — 25 часов. Студенту необходимо выбрать наиболее приоритет-

ные для него направления, укладываемые в установленные временные рамки. В статье рассматривается идея алгоритмизации этого достаточно нетривиального процесса.

Вводя концепцию рейтинга, мы определяем приоритеты различных специальностей, оценивая их на шкале, например от 1 до 3. Здесь «1» обозначает менее привлекательные направления, «2» — более привлекательные, «3» — наиболее интересные. Путем установления приоритетов для перечисленных дисциплин можно выявить наиболее оптимальный набор предметов, который наберет наибольшее количество баллов рейтинга и при этом уложится в заданный временной интервал.

Задача решается в рамках построения индивидуальной образовательной траектории вариативной части учебного плана одного из направлений подготовки Петербургского государственного университета путей сообщения.

МЕТОДОЛОГИЯ РЕШЕНИЯ

Существует алгоритм, позволяющий автоматизировать решение этой проблемы, так называемая задача о рюкзаке (knapsack problem), интенсивное изучение которой началось после из публикации книги математик Джорджа Данцига (G. Dantzig) в 1957 году. Она заключается в поиске оптимального набора предметов из заданного списка с ограничением по суммарному весу при условии максимизации их суммарной стоимости. Задачу о рюкзаке начать следует с решения задачи для меньшего рюкзака («подрюкзака»), а потом на этой основе попытаться решить исходную задачу [1].

Требуется вычислить подмножество предметов с максимально возможной суммарной стоимостью при условии наличия суммарного размера, не превышающего вместимость рюкзака [2].

Условия задачи: дано N предметов, W — вместимость рюкзака, $w = \{w_1, w_2, \dots, w_n\}$ — соответствующий ему набор положительных целых весов, $p = \{p_1, p_2, \dots, p_n\}$ — соответствующий ему набор положительных целых стоимостей. Нужно найти набор бинарных величин $B = \{b_1, b_2, \dots, b_n\}$, где $b_i = 1$, если предмет n_i включен в набор $b_i = 0$, если предмет не включен, и такой, что:

- 1) $b_1w_1 + \dots + b_nw_n \leq W$;
- 2) $b_1p_1 + \dots + b_np_n = \max$.

Наиболее оптимальный метод решения данной задачи — метод динамического программирования, подразумевающий деление задачи на более мелкие подзадачи.

Пусть $A(k, s)$ есть максимальная стоимость предметов, которые можно уложить в рюкзак вместимости s , если можно использовать только первые k предметов, то есть $\{n_1, n_2, \dots, n_{k-1}\}$. Назовем его набором допустимых предметов для $A(k, s)$.

$A(k, 0) = 0, A(0, s) = 0$. Найдем $A(k, s)$. Возможны 2 варианта:

1. Предмет k не попал в рюкзак. Тогда $A(k, s)$ равно максимальной стоимости рюкзака с такой же вместимостью и набором допустимых предметов $\{n_1, n_2, \dots, n_{k-1}\}$.

2. Предмет k попал в рюкзак. Тогда $A(k, s)$ равно максимальной стоимости рюкзака, где вес s уменьшаем на вес k -го предмета и набор допустимых предметов $\{n_1, n_2, \dots, n_{k-1}\}$, прибавляя стоимость k .

$$A(k, s) = \begin{cases} A(k-1, s), & b_k = 0; \\ A(k-1, s-w_k) + p_k, & b_k = 1. \end{cases}$$

То есть:

$$A(k, s) = \max(A(k-1, s), A(k-1, s-w_k) + p_k).$$

Стоимость искомого набора равна $A(N, W)$, так как нужно найти максимальную стоимость рюкзака вместимостью W , в который входят все предметы [3, 4].

Восстановим набор предметов, входящих в рюкзак. Определим, входит ли предмет n_i в искомый набор. Начинаем с элемента $A(i, w)$, где $i = N, w = W$. Для этого сравниваем $A(i, w)$ со следующими значениями:

1. Максимальной стоимостью рюкзака с такой же вместимостью и набором допустимых предметов $\{n_1, n_2, \dots, n_{i-1}\}$, то есть $A(i-1, w)$.

2. Максимальной стоимостью рюкзака с вместимостью на w_i меньше и набором допустимых предметов $\{n_1, n_2, \dots, n_{i-1}\}$ плюс стоимость p_i , то есть $A(i-1, w-w_i) + p_i$.

Заметим, что при построении A мы выбирали максимум из этих значений и записывали в $A(i, w)$. Если $A(i, w) = A(i-1, w)$, то n_i не входит в искомый набор, иначе входит.

Путем применения рассмотренного алгоритма возможно определить оптимальный состав набора дисциплин для вариативной части обучения, в котором [5]:

W — общее количество часов, выделенных на вариативную часть учебной программы;

w — количество часов, требуемых для изучения каждой конкретной дисциплины;

P — рейтинг каждой дисциплины, который варьируется от 1 до 3;

B — искомый оптимальный набор дисциплин, который будет выбран на основе данного алгоритма.

СПОСОБ РЕАЛИЗАЦИИ

Для решения данной задачи разработано графическое приложение WinForms на языке C++.

Это приложение включает в себя следующие компоненты: модуль для выхода, модуль для загрузки базы данных в приложение, модуль для формирования выборки дисциплин, исходную базу данных, конечную базу данных. Доступ к модулям осуществляется через интерфейс приложения с помощью соответствующих кнопок (рис. 1).

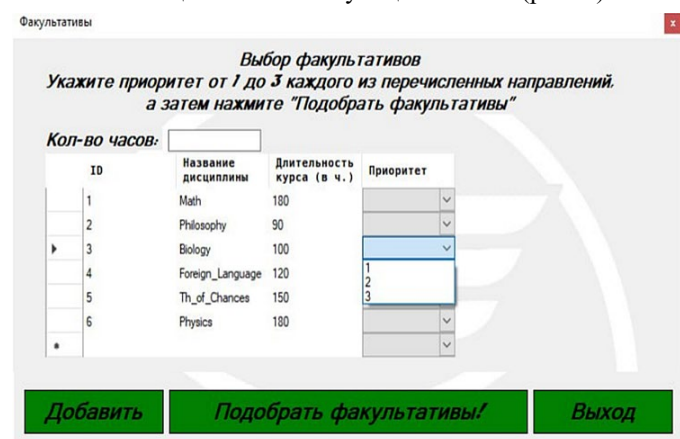


Рис. 1. Интерфейс приложения

Кнопка выхода является вспомогательной и используется для выхода из программы. Кнопка загрузки базы данных («Добавить») после активации обращается к файлу базы данных Access, содержащей информацию о студентах

и результатах их выбора, и подгружает данные из нее в объект DataGridView, который реализует функционал базы данных в приложениях WinForms. Пример программного кода:

```
System.Void Facultatives::MyForm:btn_add_db_Click(System:Object^ sender, System:EventArgs^ e) {
    database->Rows->Clear();
    // Подключение базы данных
    String^ connectionString = "provider = Microsoft.ACE.OLEDB.12.0; Data Source=Facultatives.mdb";
    OleDbConnection^ dbConnection = gcnnew OleDbConnection(connectionString);

    dbConnection->Open();
    String^ query = "SELECT * FROM [Facultatives]";
    OleDbCommand^ dbComand = gcnnew OleDbCommand(query, dbConnection); //
    OleDbDataReader^ dbReader = dbComand->ExecuteReader();

    if (dbReader->HasRows == false) {
        MessageBox::Show("Ошибка считывания данных!", "Ошибка!");
    }
    else {
        while (dbReader->Read()) {
            database->Rows->Add(dbReader["id"], dbReader["name"], dbReader["weight"], dbReader["price"]);
            wl++; // В переменную wl записывается кол-во строк - дисциплин
        }
        dbReader->Close();
        dbConnection->Close();
        return System:Void();
    }
}
```

Во время последовательной загрузки данных из базы, переменная *wl* аккумулирует количество строк, что соответствует доступным дисциплинам для выбора. Это вспомогательная переменная, которая будет использована в процессе создания выборки.

Параллельно загрузке базы данных построчно, в переменную *wl* записывается количество строк, то есть количество дисциплин на выбор. Данная переменная является вспомогательной и будет использована при построении выборки.

```
// Вывод выборки на экран
ges_table->Rows->Clear(); // Обнуляем таблицу результата
j = 0; // И вспомогательные переменные
counter_values = 0;
if (wl == 0) { // Если дисциплины не загружены - ошибка
    MessageBox::Show("Таблица не загружена!", "Ошибка!");
    return System:Void();
}
else if (max_capacity_label->Text == "") { // Если поле часов пустое - ошибка
    MessageBox::Show("Не введено кол-во часов!", "Ошибка!");
    return System:Void();
}
else { // Иначе в max_capacity количество часов для ВЧ
    max_capacity = Convert::ToInt32(max_capacity_label->Text);
}
for (i = 0; i < wl; i++) { // Если есть пустоты в столбце приоритетов - ошибка
if (database->Rows[i]->Cells[3]->Value->ToString() == "") {
    MessageBox::Show("Не все приоритеты заполнены!", "Ошибка!");
    return System:Void();
}
else { // Если пустот нет - формируем массивы весов и стоимостей
string_buffer = database->Rows[i]->Cells[2]->Value->ToString();
weights[i] = Convert::ToInt32(string_buffer);
string_buffer = database->Rows[i]->Cells[3]->Value->ToString();
values[i] = Convert::ToInt32(string_buffer);
}
}
```

После загрузки базы данных необходимо заполнить колонку «Приоритет» (3 — наивысший приоритет). После заполнения указанной колонки необходимо активировать кнопку подбора дисциплин. Обработка данных начинается с проверки ошибок при заполнении полей, а также с обновления счетчика таблицы выборки.

Пример кода программы:

```

    }
    for (i = 1; i < wl; i++) if (values[i] == values[0]) counter_values++; // Сравниваем приоритет первой строки с остальными
    // При каждом совпадении инкрементируем. Если counter_values = кол-ву дисциплин - все приоритеты одинаковые - ошибка
    if (counter_values == wl - 1) {
        MessageBox::Show("Введены одинаковые приоритеты!", "Ошибка!");
        return System::Void();
    } // Иначе - вызов функции построения матрицы весов

```

В рамках данного алгоритма, если нет ошибок, данные подвергаются преобразованию из колонок базы данных DataGridView, а также осуществляется заполнение массивов для дальнейшей обработки в числовом формате. Преобразование данных позволяет работать с ними как с числами в дальнейшем. При этом производится вызов функции *one_item*:

```
else Facultatives::MyForm::one_item(wl, max_capacity, weights, values);
```

```

System::Void Facultatives::MyForm::one_item(int wl, int max_capacity, int weights[], int values[]) {
    int** arr; // Матрица весов
    arr = new int* [wl + 1]; // arr[n+1][m+1], где n - кол-во дисциплин, m - кол-во часов на ВЧ
    for (int i = 0; i < wl + 1; i++)
        arr[i] = new int[max_capacity + 1];
    for (int i = 0; i <= wl; i++) {
        for (int j = 0; j <= max_capacity; j++) {
            if (i == 0 || j == 0) arr[i][j] = 0; // Нулевые строка и столбец заполняются нулями
            else {
                if (weights[i - 1] > j) arr[i][j] = arr[i - 1][j]; // если вес предмета больше текущей вместимости - вписываем
                предыдущее значение в столбце
                else arr[i][j] = max(arr[i - 1][j], values[i - 1] + arr[i - 1][j - weights[i - 1]]); // Иначе записываем значение по формуле:
                максимум из предыдущей ячейки в столбце и суммы стоимости предмета с макс. стоимостью в предыдущей строке, которая помещается в
                остаток места в рюкзаке
            }
        }
    }
    //int answer = arr[wl][max_capacity]; // Макс. стоимость
    find_ans(wl, max_capacity, weights, arr); // Вызов функции формирования искомого массива индексов дисциплин
    for (int i = 0; i < wl + 1; i++) // Очистка памяти
        delete[]arr[i];
    delete[]arr;
    return System::Void();
}

```

Внутри функции *one_item* после формирования матрицы стоимостей происходит вызов функции *find_ans*, формирующей выходную выборку:

```

System::Void Facultatives::MyForm::find_ans(int wl, int max_capacity, int weights[], int** arr) {
    if (arr[wl][max_capacity] == 0) return; // Рекурсивная функция построения искомой выборки. Если значение ячейки = 0 - алгоритм "уперся
    в стенку" и заканчивает обход матрицы
    if (arr[wl - 1][max_capacity] == arr[wl][max_capacity]) find_ans(wl - 1, max_capacity, weights, arr); // Если стоимости предметов равны,
    вызываем ф-цию для wl-1
    else { // Если не равны, значит там однозначно меньшая цена. Записываем текущий предмет в выборку и вызываем функцию для
    предыдущего предмета и цены предметов до предыдущего
        ans[j] = wl-1; // Искомая выборка
        j++; // Счетчик выборки
        find_ans(wl - 1, max_capacity - weights[wl - 1], weights, arr);
    }
    return System::Void();
}

```

Существование функций делает возможным использование такого средства программирования, как рекурсия. Рекурсия позволяет функции вызывать на выполнение саму себя. Это может показаться несколько неправдоподобным, и на самом деле зачастую обращение функции к

Гибкость и расположение проверки условия определяют тип цикла, выбираемый в качестве управляющей структуры [6].

Функция *one_item* принимает в качестве входных аргументов количество дисциплин, количество часов, отведенное на эти дисциплины в текущем семестре, а также массивы, содержащие информацию о каждой дисциплине и их приоритете. Далее реализуется логика, описанная в методологии решения:

самой себе вызывает ошибку, однако при правильном использовании механизм рекурсии может оказаться очень мощным инструментом в руках программиста [7].

Объект называется рекурсивным, если его части определены через него самого. Рекурсия встречается не только в математике, но и в обычной жизни [8].

Функция реализует рекурсивный поиск индексов в матрице способом, также описанным в методологии решения. Индексы записываются в отдельный массив ans. После обхода всей матрицы, функция завершается и возвращает

```

j--; // j в конце рекурсии получился на единицу больше, поэтому сперва уменьшаем его на 1
while (j >= 0) { // В рез. таблицу записываем индекс, название и длительность дисциплины, индекс которой равен ans[j]
    res_table->Rows->Add(database->Rows[ans[j]]->Cells[0]->Value, database->Rows[ans[j]]->Cells[1]->Value, database->Rows[ans[j]]->Cells[2]->Value);
    j--; // Обходим массив до нулевого элемента включительно
}
return System::Void();
}
    
```

По завершении выполнения функции управление возвращается в функцию *main*. Искомый набор дисциплин найден. Оптимальный набор дисциплин выводится на экран (рис. 2).

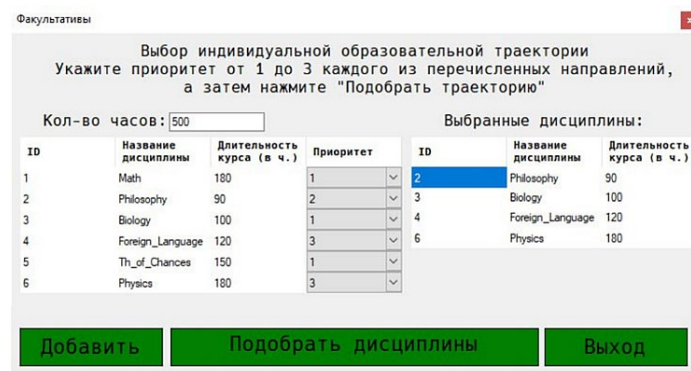


Рис. 2. Вывод результата на экран

ЗАКЛЮЧЕНИЕ

Построение вариативной части учебного плана является автоматизируемым процессом. При большом наборе входных данных эта деятельность может оказаться трудоемкой для ручного выполнения. В работе было продемонстрировано, что автоматизация процесса построения индивидуальных образовательных траекторий студентов в вариативной части учебного плана университета является не только возможной, но и весьма перспективной. Эта автоматизация позволит сократить трудозатраты и вероятность ошибок, что может существенно облегчить работу учебных заведений.

Разработанная система имеет потенциал для дальнейшего усовершенствования, что включает в себя расширение возможностей, таких как чтение данных из таблиц Excel и интеграция в системы дистанционного обучения университета. Это позволит сделать ее более функциональной и удобной для использования в реальном образовательном процессе.

ЛИТЕРАТУРА

1. Бхаргава, А. Грокаем алгоритмы. Иллюстрированное пособие для программистов и любопытствующих = Grokking Algorithms: An illustrated guide for programmers and other curious people / Пер. с англ. Е. Матвеева. — Санкт-Петербург: Питер, 2017. — 288 с. — (Библиотека программиста).

управление в функцию *one_item*, которая также завершается и возвращает управление функции кнопки. Эта функция в свою очередь выводит во вторую базу данных DataGridView полученную выборку, перебирая массив индексов ans и изменяя тип данных для вставки в таблицу:

2. Рафгарден, Т. Совершенный алгоритм. Алгоритмы для NP-трудных задач = Algorithms Illuminated. Part 4: Algorithms for NP-Hard Problems / Пер. с англ. А. Логунова. — Санкт-Петербург: Питер, 2021. — 304 с. — (Библиотека программиста).

3. Баушев, А. Н. Структуры и алгоритмы обработки данных: Учебное пособие. Часть I. — Санкт-Петербург: ПГУПС, 2019. — 54 с.

4. Баушев, А. Н. Структуры и алгоритмы обработки данных: Учебное пособие. Часть II / А. Н. Баушев, А. В. Тырва. — Санкт-Петербург: ПГУПС, 2020. — 44 с.

5. Задача о рюкзаке // Викиконспекты. — Обновлено 04.09.2022.

URL: http://neerc.ifmo.ru/wiki/index.php?title=Задача_о_рюкзаке (дата обращения 16.04.2023).

6. Макконнелл, С. Совершенный код. Мастер-класс. Второе издание = Code Complete. Second Edition / Пер. с англ. под общей ред. В. Г. Вшивцева. — Москва: Русская редакция, 2010. — 896 с.

7. Лафоре, Р. Объектно-ориентированное программирование в C++. Четвертое издание / Пер. с англ. А. Кузнецова, [и др.]. — Санкт-Петербург: Питер, 2016. — 928 с. — (Классика Computer Science).

8. Вирт, Н. Алгоритмы и структуры данных. Новая версия для Оберона / Пер. с англ. под ред. Ф. В. Ткачева. — Москва: ДМК Пресс, 2010. — 272 с. — (Классика программирования).

Optimization of the Educational Process: Construction of Individual Educational Trajectories Using the Variable Part of the Curriculum Using the Example of the Knapsack Problem

D. Yu. Fedorov, PhD A. V. Zabrodin

Emperor Alexander I St. Petersburg State Transport University
Saint Petersburg, Russia

elfbrus@gmail.com, teach-case@yandex.ru

Abstract. The article examines the use of combinatorial optimization problems for the formation of a variable part of the curriculum in order to build individual educational learning trajectories. A specific knapsack problem is considered in the context of learning optimization. A method for forming individual curricula is presented based on an analysis of the needs and abilities of students. Particular attention is paid to the process of adapting curricula to the specific needs of students, which contributes to more effective and personalized learning. The results of the study may be useful for educational institutions and teachers seeking to improve the quality of education and support the individual educational experience of each student.

Keywords: curriculum, discipline, algorithm, knapsack problem, trajectory, optimization, statistics, sorting.

REFERENCES

1. Bhargava A. Grokking Algorithms: An illustrated guide for programmers and other curious people [Grokaem algoritmy. Illyustrirovannoe posobie dlya programmistov i lyubopytstvuyushchikh]. Saint Petersburg, Piter Publishing House, 2017, 288 p.
2. Roughgarden, T. Algorithms Illuminated. Part 4: Algorithms for NP-Hard Problems [Sovershennyy algoritm. Algoritmy dlya NP-trudnykh zadach]. Saint Petersburg, Piter Publishing House, 2021, 304 p.
3. Baushev A. N. Data processing structures and algorithms: Study guide. Part I [Struktury i algoritmy obrabotki dannykh: Uchebnoe posobie. Chast I]. Saint Petersburg, Petersburg State Transport University, 2019, 54 p.

4. Baushev A. N., Tyrva A. V. Data processing structures and algorithms: Study guide. Part II [Struktury i algoritmy obrabotki dannykh: Uchebnoe posobie. Chast II]. Saint Petersburg, Petersburg State Transport University, 2020, 44 p.

5. The Backpack Problem [Zadacha o ryukzake], *Wikinotes [Vikikonspekty]*. Last update at September 04, 2022. Available at: http://neerc.ifmo.ru/wiki/index.php?title=Задача_о_рюкзакe (accessed 16 Apr 2023).

6. McConnell S. Code Complete. Second Edition [Sovershennyy kod. Master-klass. Vtoroe izdanie]. Moscow, Russian Edition Publishers, 2010, 896 p.

7. Lafore R. Object-oriented programming in C++. Fourth Edition [Obektno-orientirovannoe programmirovaniye v C++. Chetvertoe izdanie]. Saint Petersburg, Piter Publishing House, 2016, 928 p.

8. Wirth N. Algorithms and data structure. New version for Oberon [Algoritmy i struktury dannykh. Novaya versiya dlya Oberona]. Moscow, DMK Press Publishing House, 2010, 272 p.

ПАМЯТИ КОЛЛЕГИ



25 октября 2023 года после непродолжительной болезни скончался профессор кафедры «Информатика и информационная безопасность» ЖИГАЛКО Евгений Фаддеевич.

10 апреля 2023 года Евгений Фаддеевич отметил свое 80-летие, он работал в нашем университете с 2001 года. Возглавлял кафедру «Прикладная математика», затем был профессором на кафедре «Математика и моделирование».

Жигалко Е. Ф. — высококвалифицированный педагог, доктор физико-математических наук — с 1997 года являлся действительным членом Академии навигации и управления движением. Евгений Фаддеевич был признанным специалистом по механике сплошных сред.

Все его занятия проводились на высоком научном и методическом уровне. Евгений Фаддеевич — автор многих научных трудов и учебных пособий. Под его руководством защищены две кандидатские диссертации. Имеет большое количество изобретений, и эта его работа отмечена знаком «Изобретатель СССР».

За многолетнюю добросовестную и плодотворную работу в университете он награжден юбилейным знаком «200 лет ПГУПС», юбилейным нагрудным знаком «200 лет транспортному образованию России».

Евгений Фаддеевич пользовался заслуженным уважением сотрудников кафедры и обучающихся. Коллеги и ученики помнят и глубоко скорбят. Светлая память Учителю.

In Memory of Colleague

On October 25, 2023 Professor of the Department of Information Technology and IT Security ZHIGALKO Evgeny Faddeevich, died after a short illness.

On April 10, 2023 Evgeny Faddeevich celebrated his 80th birthday, he worked at our university since 2001. Headed the Department of Applied Mathematics, then was a professor at the Department of Mathematics and Modeling.

Zhigalko E. F. is highly qualified teacher, Doctor of Physical and Mathematical Sciences, since 1997 was valid a member of the Academy of Navigation and Motion Control. Evgeny Faddeevich was a recognized specialist in Continuum Mechanics.

All his classes were held at high scientific and methodological level. Evgeny Faddeevich is the author of many scientific papers and tutorials. Under his leadership, two candidate dissertations were defended. Has a large number of inventions, and this work of his is marked with the «Inventor of the USSR» sign.

For many years of conscientious and fruitful work at the university, he was awarded the anniversary sign «200 years of PGUPS», anniversary badge «200 years of transport education in Russia».

Evgeny Faddeevich enjoyed the well-deserved respect of the department staff and students. Colleagues and students remember and grieve deeply. Bright memory to the Teacher.