

Том 16. №6. 2021

Ноябрь–Декабрь

Сквозной номер выпуска — 96

Университет «Синергия»

ISSN 1993-8314 (Print)

ISSN 2687-0649 (Online)

DOI: 10.37791/2687-0649-2021-16-6

Научно-практический журнал «Прикладная информатика»

Журнал включен в Перечень ведущих периодических изданий, рекомендованных ВАК для публикации результатов диссертационных исследований

РЕДАКЦИОННЫЙ СОВЕТ

Главный редактор

Дли М. И., докт. техн. наук, профессор, зав. кафедрой ИТЭУ, зам. директора филиала НИУ «МЭИ», г. Смоленск

Сопредседатели редакционного совета

Рубин Ю. Б., докт. экон. наук, профессор, чл.-корр. РАО, президент Университета «Синергия», зав. кафедрой теории и практики конкуренции, г. Москва**Мешалкин В. П.**, докт. техн. наук, профессор, академик РАН, директор Международного института логистики ресурсосбережения и технологической инноватики, РХТУ им. Д. И. Менделеева, г. Москва

Члены редакционного совета

Батыршин И. З., д-р, профессор, Мексиканский нефтяной институт, приглашенный профессор Исследовательского вычислительного центра Национального политехнического института Мексики, г. Мехико, Мексика**Борисов В. В.**, докт. техн. наук, профессор, кафедра вычислительной техники, филиал Национального исследовательского университета «МЭИ», г. Смоленск**Брекис Эд.**, д-р, ассоциированный профессор, зав. кафедрой эконометрики и бизнес-информатики, Латвийский университет, г. Рига, Латвия**Волкова В. Н.**, докт. экон. наук, профессор, Институт компьютерных наук и технологий, Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург**Козлов В. Н.**, докт. техн. наук, профессор, Институт компьютерных наук и технологий, Санкт-Петербургский политехнический университет Петра Великого, г. Санкт-Петербург**Краковский Ю. М.**, докт. техн. наук, профессор, кафедра информационных систем и защиты информации, Иркутский государственный университет путей сообщения, г. Иркутск**Крейнович В. Я.**, д-р, профессор Университета Техаса, г. Эль-Пасо, США**Пецольдт К.**, д-р, профессор, проректор по международному сотрудничеству с Восточной Европой, Технологический университет, г. Ильменау, Германия**Росс Г. В.**, докт. техн. наук, докт. экон. наук, профессор, академик РАН, научная лаборатория семантического анализа и интеграции, Российский экономический университет им. Г. В. Плеханова, г. Москва**Стоянова О. В.**, докт. техн. наук, профессор, кафедра информационных систем в экономике, СПбГУ, г. Санкт-Петербург**Сухомлин В. А.**, докт. техн. наук, профессор, зав. лабораторией открытых информационных технологий, факультет ВМК, МГУ имени М. В. Ломоносова, г. Москва**Халин В. Г.**, докт. экон. наук, профессор, кафедра информационных систем в экономике, экономический факультет СПбГУ, г. Санкт-Петербург**Шоринов А. Ф.**, докт. физ.-мат. наук, профессор, ведущий научный сотрудник, Институт экономики УрО РАН, г. Екатеринбург**Штельцер Д.**, д-р, ger. pol., профессор, глава департамента информации и управления знаниями, Технологический университет, г. Ильменау, Тюрингия, Германия**Юсупов Р. М.**, докт. техн. наук, профессор, чл.-корр. РАН, научный руководитель Санкт-Петербургского института информатики и автоматизации РАН, президент Национального общества имитационного моделирования «НОИМ», г. Санкт-Петербург

Заместители главного редактора

Прокимов Н. Н., канд. техн. наук, доцент, департамент цифровой экономики, Университет «Синергия», г. Москва**Власова Е. А.**, научная редакция Университета «Синергия», г. Москва

Журнал выходит с 2006 г. Периодичность издания – 6 раз в год.

Журнал индексируется в российских и зарубежных базах научной периодики
eLIBRARY (РИНЦ), Russian Science Citation Index (RSCI) на платформе Web of Science,
ВИНИТИ, Ulrich's Periodicals DirectoryУчредитель и издатель: Негосударственное образовательное частное учреждение высшего образования
Университет «Синергия»

Адрес редакции и издателя:

129090, Москва, ул. Мещанская, д. 9/14, стр. 1 (юрид.)

125190, Москва, Ленинградский пр-т, д. 80, корпус Г

Тел.: +7 (495) 987-43-74 (доб. 33-04); e-mail: evlasova@synergy.ru; www.appliedinformatics.ru

© Университет «Синергия»

Journal of Applied INFORM@TICS

Peer-reviewed scientific journal

Vol.16. No.6. 2021

November–December

Continuous issue — 96

Synergy University

ISSN 1993-8314 (Print)

ISSN 2687-0649 (Online)

DOI: 10.37791/2687-0649-2021-16-6

Peer-reviewed scientific journal "Journal of Applied Informatics"

EDITORIAL BOARD

Editor-in-Chief

M. Dli, Dr. Sci. (Eng.), Professor, Head of the ITEM Chair, Deputy Director of the National Research University MPEI Branch, the city of Smolensk

Co-Chairs of the Editorial Board

Yu. Rubin, Dr. Sci. (Econ.), Professor, Corresponding Member of the Russian Education Academy, Head of the Theory and Practice of Competition Chair, President of the Moscow University for Industry and Finance "Synergy", the city of Moscow

V. Meshalkin, Dr. Sci. (Eng.), Professor, Academician of Russian Academy of Sciences (RAS), Director of the International Institute of Logistics and Resource Technology Innovation, D. Mendeleyev University of Chemical Technology of Russia, the city of Moscow

Members of the Editorial Board

I. Batyrshin, Dr. Sci. (Habilitation), IFSA Fellow, Titular Professor C, Centro de Investigacion en Computacion, Instituto Politecnico Nacional, the city of Mexico, United Mexican States

V. Borisov, Dr. Sci. (Eng.), Professor, Computer Science Department, National Research University MPEI Branch, the city of Smolensk

Ed. Brēķis, Dr. oec., Assoc. Professor, Head of the Econometrics and Business Informatics Chair, Faculty of Economics and Management, University of Latvia, the city of Riga, Latvia

V. Khalin, Dr. Sci. (Econ.), Professor, Economic Information Systems Department, St. Petersburg State University, the city of St. Petersburg

V. Kozlov, Dr. Sci. (Eng.), Professor, Institute of Computer Science and Technology, Peter the Great St. Petersburg Polytechnic University, the city of St. Petersburg

Yu. Krakovski, Dr. Sci. (Eng.), Professor, Information Systems and Information Security Department, Irkutsk State Railway Transport Engineering University, the city of Irkutsk

V. Kreinovich, PhD, Professor of Computer Science, University of Texas at El Paso, USA

K. Pezoldt, Dr. oec. habil., Professor, Deputy Rector for International Cooperation with Eastern Europe, Ilmenau University of Technology, the city of Ilmenau, Germany

G. Ross, Dr. Sci. (Eng.), Dr. of Economics, Professor, Academician of Russian Academy of Natural Sciences (RANS), Research Laboratory of Semantic Analysis and Integration, Plekhanov Russian University of Economics, the city of Moscow

A. Shorikov, Dr. Sci. (Phys.-Math.), Professor, Leading Researcher, Institute of Economics of the Ural Branch of RAS, the city of Ekaterinburg

O. Stoyanova, Dr. Sci. (Eng.), Professor, Economic Information Systems Department, St. Petersburg State University, the city of St. Petersburg

V. Sukhomlin, Dr. Sci. (Eng.), Professor, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, the city of Moscow

D. Stelzer, Dr., rer. pol., Professor, Head of the Information and Knowledge Management Department of Ilmenau University of Technology, the city of Ilmenau, Germany

V. Volkova, Dr. Sci. (Econ.), Professor, Institute of Computer Science and Technology, Peter the Great St. Petersburg Polytechnic University, the city of St. Petersburg

R. Yusupov, Dr. Sci. (Eng.), Professor, Corresponding Member of Russian Academy of Sciences (RAS), Scientific Leader of the Saint Petersburg RAS Institute of Informatics and Automation, President of NC "National Society for Simulation Modeling", the city of St. Petersburg

Deputy Chief Editors

N. Prokimnov, Cand. Sci. (Eng.), Associate Professor, Digital Economy Department, Moscow University for Industry and Finance "Synergy", the city of Moscow

E. Vlasova, Scientific Edition Department, Moscow University for Industry and Finance "Synergy", the city of Moscow

Published since 2006. Periodicity: bimonthly.

The journal is indexed in

Russian Science Citation Index (RSCI) on Web of Science platform,
VINITI (Russian Academy of Sciences), Ulrich's Periodicals Directory

Publisher: Moscow University for Industry and Finance «Synergy»

Publisher address: 9/14 s. 1, Meshchanskaya st., Moscow, 129090, Russia

Editorial office address: 80G, Leningradskiy Avenue, Moscow, 125190, Russia

Tel: +7 (495) 987-43-74 (ext. 33-04); e-mail: evlasova@synergy.ru; www.appliedinformatics.ru

© Synergy University

ПАМЯТИ УЧЕНОГО

Пресс-релиз

Научный семинар «Емельяновские чтения. Имитационное моделирование и системный анализ в управлении – 2021» 4

ИТ-МЕНЕДЖМЕНТ

Управление эффективностью

А. О. Суворов, А. А. Петренко, А. Д. Неприна
Сравнительная оценка моделей технического и фундаментального анализа при прогнозировании курса акций 6

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА

Модели и методики

Ю. И. Бутенко, Ю. В. Строганов, А. М. Сапожков
Метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов. 21

С. А. Калайда

Model of creating an economic ecosystem in the framework of economic convergence under the influence of digitalization 28

Эффективные алгоритмы

Р. В. Солопов, А. С. Самульченков, В. И. Зирюкин
Genetic algorithm as a tool for modeling calculations of electric power systems 43

А. С. Лошкарев, И. А. Тимофеев

Исследование процесса бинаризации изображений с использованием локальных значений порога. 54

Программная инженерия

В. В. Миронов, А. С. Гусаренко, Н. И. Юсупова
Программное извлечение данных из word-документов на основе ситуационно-ориентированного подхода. 66

А. Н. Чурсин, Н. А. Мамедова, Ю. В. Нефедов
Разработка кроссплатформенных мобильных приложений – перспективные методы и стандартные практики 84

О. В. Тиханычев

Постановка задачи обеспечения автоматизации полного цикла поддержки принятия решения. 103

ЛАБОРАТОРИЯ

Моделирование процессов и систем

О. А. Гурьянова, А. Ю. Гнибеда, Е. В. Филимонова
Исследование цветовых отличий при воспроизведении памятных цветов на устройствах визуализации 113

*Я. М. Иванько, С. А. Петрова,
М. Н. Барсукова, Ю. В. Столопова*

Задача параметрического программирования с моделями прогнозирования урожайности сельскохозяйственных культур 131

IN MEMORY OF SCIENTIST

Press release

Scientific seminar "Emelyanov readings. Simulation and system analysis in management – 2021" 4

IT MANAGEMENT

Performance management

A. Suvorov, A. Petrenko, A. Neprina
Comparative evaluation of technical and fundamental analysis models when predicting stock prices 6

TOOLS

Models and methods

Iu. Butenko, Yu. Stroganov, A. Sapozhkov
Method for the extraction of Russian-language multicomponent terms from scientific and technical texts 21

S. Kalayda

Model of creating an economic ecosystem in the framework of economic convergence under the influence of digitalization 28

Algorithmic efficiency

R. Solopov, A. Samulchenkov, V. Ziryukin
Genetic algorithm as a tool for modeling calculations of electric power systems 43

A. Lozhkarev, I. Timofeev

Investigation of the process of binarization of images using local values of the threshold 54

Software engineering

V. Mironov, A. Gusarenko, N. Yusupova
Software extract data from word-based documents situationally-oriented approach 66

A. Chursin, N. Mamedova, Yu. Nefedov
Development of cross-platform mobile applications – promising methods and standard practices 84

O. Tikhanychev

Formulation of the task of ensuring automation of the full cycle of decision support 103

LABORATORY

Processes and systems modeling

O. Guryanova, A. Gnibeda, E. Filimonova
Investigation of color differences in the reproduction of memorable colors on visualization devices 113

*Ya. Ivanyo, S. Petrova, M.
Barsukova, Yu. Stolopova*

Parametric programming problem with crop yield forecasting models 131

Научный семинар «Емельяновские чтения. Имитационное моделирование и системный анализ в управлении – 2021»

25 ноября 2021 г. на базе филиала Национального исследовательского университета «МЭИ» в г. Смоленске прошел научный семинар «Емельяновские чтения. Имитационное моделирование и системный анализ в управлении – 2021». Семинар проводился впервые и был посвящен памяти Александра Анатольевича Емельянова – известного российского ученого, организатора науки и высшего образования в области информатики. Вел семинар М. И. Дли (докт. техн. наук, профессор, заместитель директора филиала Национального исследовательского университета «МЭИ» в г. Смоленске).

В программе семинара было выделено две части – вначале некоторые участники поделились своими личными впечатлениями о совместной работе с Александром Анатольевичем, а потом с научными докладами по актуальным проблемам науки и высшей школы выступили другие участники семинара.

Воспоминаниями об Александре Анатольевиче поделился Ю. Б. Рубин (докт. экон. наук, профессор, президент Московского финансово-промышленного университета «Синергия»). Он рассказал о совместной работе с профессором А. А. Емельяновым в Университете «Синергия», а также в Московском экономико-статистическом институте. Юрий Борисович подчеркнул важную роль Александра Анатольевича в становлении Университета «Синергия», где тот работал сначала в должности декана факультета информационных технологий, а затем проректора и вице-президента. В выступлении Ю. Б. Рубина были отмечены также значительные усилия А. А. Емельянова по созданию на редакционно-издательской базе Университета «Синергия» научного журнала «Прикладная информатика», главным редактором и душой которого он был в течение 15 лет. А во время своей работы в МЭСИ Александр Анатольевич, будучи директором Института компьютерных технологий, принимал определяющее участие в работе УМО по специальности «Прикладная информатика», в которое входило более 150 вузов страны. Именно благодаря А. А. Емельянову эта специальность сохранилась в образовательном пространстве высшей школы, и при его активном участии был разработан соответствующий образовательный стандарт.

В. Н. Волкова (докт. экон. наук, профессор, Санкт-Петербургский политехнический университет Петра Великого) в своем выступлении отметила весомый вклад Александра Анатольевича в развитие методологии и инструментальной базы имитационного моделирования как важного метода теории систем и системного анализа. А. А. Емельянов является автором учебника по системному анализу, а также инициатором издания и ответственным редактором справочника «Теория систем и системный анализ в управлении организациями», изданного в издательстве «Финансы и статистика» и в этом году переизданного в электронном варианте. Александр Анатольевич также инициировал в созданном им журнале «Прикладная информатика» рубрику об известных ученых, внесших вклад в развитие информатики, что является весьма значимым для сохранения истории информатики для молодого поколения специалистов.

В. Г. Халин (докт. экон. наук, профессор, Санкт-Петербургский государственный университет) в своем выступлении рассказал о том, как они с Александром Анатольевичем Емельяновым впервые встретились в 1998 году в кабинете заместителя министра общего и профессионального образования России В. Д. Шадрикова, где бурно обсуждалась инициатива СПбГУ, МЭСИ и МГУ по

открытию новой специальности «Математическое обеспечение и администрирование информационных систем» на фоне резкого сокращения общего числа специальностей и направлений подготовки специалистов в высшей школе России. Именно профессионализм и аргументы Александра Анатольевича о необходимости и востребованности подготовки таких специалистов для экономики страны способствовали тому, что в 2000 году в российской высшей школе впервые появилась новая специальность «Математическое обеспечение и администрирование информационных систем», были разработаны и утверждены соответствующие федеральные государственные образовательные стандарты. В СПбГУ было открыто соответствующее УМО, в составе которого Александр Анатольевич работал с момента его создания. Профессор А. А. Емельянов являлся заместителем председателя Совета основных образовательных программ «Бизнес-информатика», «Информационная бизнес-аналитика» СПбГУ, руководил работой УМО по образованию в области прикладной информатики МЭСИ, входил в состав профильных УМО в других вузах. Александр Анатольевич внес огромный личный вклад в развитие российской системы подготовки кадров в области прикладной информатики, математических и инструментальных методов экономики.

Г. В. Чернова (докт. экон. наук, профессор, Санкт-Петербургский государственный университет) поделилась впечатлениями о совместной работе с Александром Анатольевичем в составе Государственной аттестационной комиссии СПбГУ кафедры информационных систем в экономике по специальности «Прикладная информатика в экономике» и направлению «Бизнес-информатика» (магистры и бакалавры). Она подчеркнула, что А. А. Емельянов особо выделял дальнейшие перспективы развития тематики диплома. Его эрудиция и знания помогали каждому студенту посмотреть на свою работу под другим, очень значимым углом зрения, увидеть ее скрытые возможности, понять перспективы дальнейшего исследования. Именно эти качества Александра Анатольевича помогли раскрыть свои таланты многим выпускникам кафедры информационных систем в экономике СПбГУ. Некоторые из них продолжили свои исследования в аспирантуре и успешно защитили диссертации по научной специальности 08.00.13 «Математические и инструментальные методы экономики». При этом официальным оппонентом по ним выступал Александр Анатольевич. Его работа в качестве председателя ГАК СПбГУ – для всех нас образец высочайшего профессионализма.

Характеристике главных направлений научной работы Александра Анатольевича были посвящены обстоятельные доклады О. В. Стояновой (докт. техн. наук, доцент, профессор Санкт-Петербургского государственного университета) об истории разработки системы имитационного моделирования Astor Pilgrim, а также О. В. Булыгиной (канд. экон. наук, доцент, филиал Национального исследовательского университета «МЭИ» в г. Смоленске) о перспективных направлениях применения этой системы, которые плавно перекинули мостик к обсуждению конкретных и актуальных проблем науки и высшей школы.

Выступление А. Ф. Шорикова (докт. физ.-мат. наук, профессор, Уральский федеральный университет имени первого Президента России Б. Н. Ельцина) было посвящено вопросам минимаксного программного управления производством продукции предприятия при наличии выпуклой целевой функции.

На примере Проекта «5 в 100» В. Г. Халин остановился на результатах системного анализа реформирования российской высшей школы.

Поднятые при обсуждении выступлений и докладов вопросы носили дискуссионный характер, что позволяет оценивать проведенный семинар как интересный и полезный. Участники семинара высказали ряд предложений по работе журнала «Прикладная информатика» и единодушно пришли к мнению о целесообразности проведения такого семинара на ежегодной основе.

Материал подготовлен оргкомитетом семинара

Сравнительная оценка моделей технического и фундаментального анализа при прогнозировании курса акций

А. О. Суворов^{1*}, А. А. Петренко², А. Д. Неприна³

¹ Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

² Пермский национальный исследовательский политехнический университет, Пермь, Россия

³ Акционерное общество «АЛЬФА-БАНК», Москва, Россия

* AOSuvorov@hse.ru

Аннотация. Работа посвящена проведению сравнительного анализа эффективности применения моделей ARIMA, ARCH, GARCH, многофакторной модели и модели построения дерева решений. На уровне отраслевого анализа производился учет фондовых индексов. Функционал моделей может быть оценен только на практических примерах, которые представлены в статье. Приведены графики моделей с учетом интервала значимости. Получены результаты применения теста Дики – Фуллера по разным данным для проверки наличия нестационарности. Описаны параметрические аргументы для исследуемых моделей. В качестве критериев оценки для каждой модели были взяты: стандартная ошибка, коэффициент детерминации, скорректированный R-квадрат, P-значение, значение F-статистики. Приведены исходные данные, порядок проведения исследования, полученные результаты и графики. С использованием языка программирования R проведено практическое исследование функционала моделей технического и фундаментального анализа для построения прогнозных значений курса акций ПАО «Сбербанк». Каждая из рассматриваемых моделей была реализована на языке программирования R для статистической обработки данных. Процесс программного моделирования показал сильные и слабые стороны каждой из рассмотренных моделей. Наилучшие результаты показала многофакторная модель. В работе приведены количественные показатели прогнозных значений. Приведена сравнительная таблица статистических показателей результатов прогнозных моделей и сделаны выводы о пригодности их моделирования. Данное исследование проводилось с целью выявления моделей технического и фундаментального анализа, дающих наиболее точный прогноз курса акций с возможностью дальнейшей реализации в компьютерной программе.

Ключевые слова: курс акций, статистические показатели, технический анализ, фундаментальный анализ, многофакторная модель, модель прогнозирования

Для цитирования: Суворов А. О., Петренко А. А., Неприна А. Д. Сравнительная оценка моделей технического и фундаментального анализа при прогнозировании курса акций // Прикладная информатика. 2021. Т. 16. № 6. С. 6–20. DOI: 10.37791/2687-0649-2021-16-6-6-20

Comparative evaluation of technical and fundamental analysis models when predicting stock prices

A. Suvorov^{1*}, A. Petrenko², A. Neprina³

¹ HSE University, Moscow, Russia

² Perm National Research Polytechnic University, Perm, Russia

³ Joint Stock Company "ALFA-BANK", Moscow, Russia

* AOSuvorov@hse.ru

Abstract. The article considers a comparative analysis of the effectiveness in use of ARIMA, ARCH, GARCH models, a multi-factor forecasting model, and a decision tree model. Model functionality can be evaluated on the practical examples presented in the article. The results of applying the Dickey-Fuller test according to various data to verify the presence of non-stationarity are obtained. Parametric arguments for the models under study are described. The initial data, the order of the study, the results and charts are presented. Using the R programming language, practical studies focused on the functionality of the technical and fundamental analysis models were carried out to obtain the forecast values of PJSC "Sberbank" stock rate. The software modeling process showed the strengths and weaknesses of each of the models considered. The best results were shown by the multi-factor model. The paper gives quantitative indicators of the forecast values. A comparative table of the statistical indicators showing the results of the forecast models is presented and the conclusions are drawn based on the suitability of their modeling. Current study was carried out to identify models of technical and fundamental analysis that give the most accurate forecast of the stock price with the possibility of further implementation in a computer program.

Keywords: forecasting model, multi-factor model, fundamental analysis, technical analysis, stock price, statistical indicators

For citation: Suvorov A., Petrenko A., Neprina A. Comparative evaluation of technical and fundamental analysis models when predicting stock prices. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.6-20 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-6-20

Введение

Одна из ключевых проблем при торговле на рынке ценных бумаг – отбор финансовых инструментов, способных помочь приумножить капиталовложения и при этом минимизировать вероятные риски. Профессионалы стараются различными способами предусмотреть изменяющуюся ситуацию и спрогнозировать последующую динамику рынка.

В зависимости от сроков и объемов инвестиций используются два наиболее распространенных метода анализа изменения цены – фундаментальный и технический анализ.

Ключевая цель фундаментального анализа – определение реальной стоимости оцениваемого бизнеса и, исходя из этого, справедливости курса акций. Знание и правильное толкование фундаментальных основ фондового рынка ориентирует его участников правильно интерпретировать изменение цен, предугадывать поворотные моменты, выявлять взаимосвязи между различными макроэкономическими показателями и принимать грамотные решения раньше других [1].

С помощью технического анализа изучают динамику рынков, в первую очередь посредством графиков, для прогнозирования

направления изменения цен. Технический подход к инвестициям отражает идею о том, что цена изменяется в соответствии с трендами, определяемыми отношениями инвесторов к различным политическим, экономическим и социально-психологическим факторам. Способность грамотного использования инструментов технического анализа проявляется в том, чтобы на достаточно ранней стадии обнаружить кардинальное изменение тренда и двигаться по этому тренду до момента, когда доказательства покажут изменение тенденции [2].

Определение направления будущего движения цен является задачей как технического, так и фундаментального анализа. Но при решении данной проблемы используются в обоих случаях разные способы.

Анализ моделей проводится далее на примере акций компании ПАО «Сбербанк». Данный выбор обусловлен тем, что на текущее время ПАО «Сбербанк» – крупнейший банк страны и, таким образом, является системообразующим для финансовой системы и экономики в целом.

Для проведения сравнения моделей целесообразнее всего применить комплексный подход к анализу алгоритмов прогнозирования фундаментального и технического анализа и использовать следующий аппарат исследования: теория биржевых систем, теория статистической обработки данных, теория построения основных котировок.

Целью представленной работы является выявление моделей технического и фундаментального анализа, дающих наиболее точный прогноз курса акций с возможностью дальнейшей реализации в компьютерной программе.

Критерии проведения анализа

Для проведения фундаментального анализа необходимо провести анализ экономического положения в целом, анализ отрасли и анализ непосредственно компании.

Анализ экономического положения в целом предусматривает оценку факторов, ко-

торые влияют на стоимость акций, а именно оценку ситуации с различных точек зрения: политической, экономической, кредитной и финансовой политики. Исходным материалом для этого вида анализа является информация о ставках дисконтирования центральных банков, экономическом курсе правительства, изменениях в политической жизни страны, а также всевозможные слухи и ожидания.

На уровне отраслевого анализа учитываются индексы фондового рынка, например индекс развития промышленного сектора DJIA (Dow Jones Industrial Average) или индекс, определяющийся на основе ста крупнейших по капитализации финансовых компаний NASDAQ Financial-100 (National Association of Securities Dealers Automated Quotation).

Изучая деятельность компании, следует принимать во внимание то, что фундаментальный анализ базируется на идее, при которой рынок время от времени может заблуждаться в оценке ее стоимости. Прибыль же при этом может быть получена нахождением некоторых недооцененных запасов, а также ожиданием того, что произойдет корректировка стоимости компании. Получить представление о стоимости разных компаний, понимать ценообразование на рынке акций можно путем анализа финансовых отчетов компаний.

Рассматривая компанию, нужно проявить интерес к следующему: показатели экономического роста; торговый баланс; степень зависимости от внешних источников сырья; изменение объема денежной массы на внутреннем рынке; информационные материалы, публикуемые компанией; имеющиеся документы о судебных разбирательствах, в которые вовлечена компания; выплаты по дивидендам и др.

Проведение фундаментального анализа – довольно непростая задача. Это связано прежде всего с тем, что в разных условиях одни и те же факторы могут оказывать совершенно разное влияние на рынок, кроме того, могут из незначительных вдруг стать определяющими. Поэтому критика фундаментального анализа обычно сводится к следующим утверждениям:

он неосуществим, а если даже осуществим, то излишен и вследствие этого не нужен [3].

Сутью технического анализа является исследование динамики рынков, обычно при помощи графиков, для прогнозирования направления изменения цен [4].

В качестве входных данных технический анализ использует различные цены: самая высокая и самая низкая, цены открытия и закрытия, за определенный период времени и объем транзакций.

Технический анализ основан на трех аксиомах [5]: движения цен на рынке учитывают всю информацию, движения цен подчиняются тенденциям, история повторяется.

В техническом анализе объектом исследования являются графики (charts), которые отражают движение цен. На графиках выделяется определенный набор типовых элементов, с помощью которых уже описывается поведение самих графиков.

Технический и фундаментальный анализ рассматривают обычно как противоположные подходы к анализу ценных бумаг, поскольку они имеют расхождения во многих аспектах. Эти отличия показывают, что фундаментальный и технический анализ являются принципиально разными подходами к последующему принятию инвестиционных решений. Наблюдаемые отличия не всегда подтверждают факт того, что они приводят к разным инвестиционным решениям. По сути, и технический, и фундаментальный анализ имеют свои преимущества и недостатки, объединив которые можно обеспечить оптимальные результаты.

Оба метода имеют свои положительные и отрицательные стороны, поэтому многие инвесторы для успешной торговли на рынке учитывают информацию, которую дает как технический, так и фундаментальный анализ.

В техническом анализе среди методов прогнозирования различают линейный и нелинейный подходы. При линейном подходе используется эконометрический аппарат для

поиска закономерностей, имеющих линейную зависимость.

Анализ проведем на основе следующих моделей линейного подхода:

- модели ARIMA (Autoregressive Integrated Moving Average) – модели Бокса – Дженкинса. В состав модели входит модель с авторегрессией и модель скользящего среднего [6];

- модели ARCH (Autoregressive Conditional Heteroscedasticity) – авторегрессионной условной модели гетероскедастичности, или стохастической модели, которая не генерирует автокорреляцию [7];

- GARCH-модели (Generalized ARCH), аналогично ARCH-модели, применяющей средневзвешенное прошлых квадратов остатков, однако, в отличие от ARCH-модели, применяющей механизм уменьшающихся значений веса, которые никогда не достигают нуля [8].

Нелинейный подход к прогнозированию реализуется посредством искусственных нейронных сетей, с помощью которых моделируются непараметрические и параметрические процессы и выполняется автоматическое преобразование входных данных. Искусственные нейронные сети могут построить оптимальную модель прогнозирования и, адаптируясь к различным условиям, развиваться вместе с рынком [9].

Однако несмотря на преимущества нейронной сети, на практике она используется редко. Причиной этого является то, что компьютерная программа имеет сложную реализацию и нейронная сеть обучается на большом объеме данных. Указанные особенности нейронных сетей, по мнению авторов, выходят за рамки текущего исследования и требуют отдельного изучения.

Среди алгоритмов прогнозирования посредством фундаментального анализа наиболее популярными являются дерево решений и многофакторная модель [10].

Для прогнозирования цены акций с использованием дерева решений требуются данные за предыдущие годы, включая показатели Open, High, Low, Close Prices и Volume (OHLCV: O обозначает Open – цену открытия

интервала, H означает High – максимум цены интервала, L означает Low – минимум цены интервала, C означает Close – цену закрытия интервала и V означает Volume – объем сделок) [10].

Другим популярным методом прогнозирования в фундаментальном анализе является многофакторная модель, в которой используются многочисленные факторы для объяснения рыночных явлений и (или) равновесных цен на активы. Многофакторные модели принято разделять на три категории: макроэкономические модели, фундаментальные модели и статистические модели [11]. Макроэкономические модели включают в себя такие факторы, как занятость, инфляция, цена на нефть и так далее. Фундаментальные модели анализируют взаимосвязь между ценными бумагами и их базовыми финансовыми показателями, такими как прибыль.

Таким образом, на сегодняшний день существует большое количество алгоритмов прогнозирования, которые используются на рынке акций.

У каждой модели имеются свои как сильные, так и слабые стороны. Но оценить функциональность модели можно лишь на практических примерах, которые представлены далее.

Анализ полученных результатов прогнозирования с использованием выбранных моделей

Сравнительный анализ языков и сред разработки показал, что целесообразнее всего для построения моделей воспользоваться процедурой STL (Seasonal Decomposition of Time Series by Loess – сезонная декомпозиция временных рядов) [12] на языке R с использованием пакета Forecast [13].

Анализ алгоритмов моделей прогнозирования далее осуществлялся на основе данных о стоимости акций ПАО «Сбербанк», собранных из открытых источников за период 01.2011–07.2018.

Анализируемые данные представляют собой дату (год, месяц, день) и цену закрытия в долларах США (ежедневная цена).

Определение наличия сезонности в рассматриваемом ряде является основой при построении модели ARIMA.

На рисунке 1 изображен график модели STL, в котором показаны: исходный ряд данных (Data), сезонность (Seasonal), тренд (Trend) и остаток (Remainder). Анализируя полученный график, можно заметить, что исходный ряд имеет сезонность.

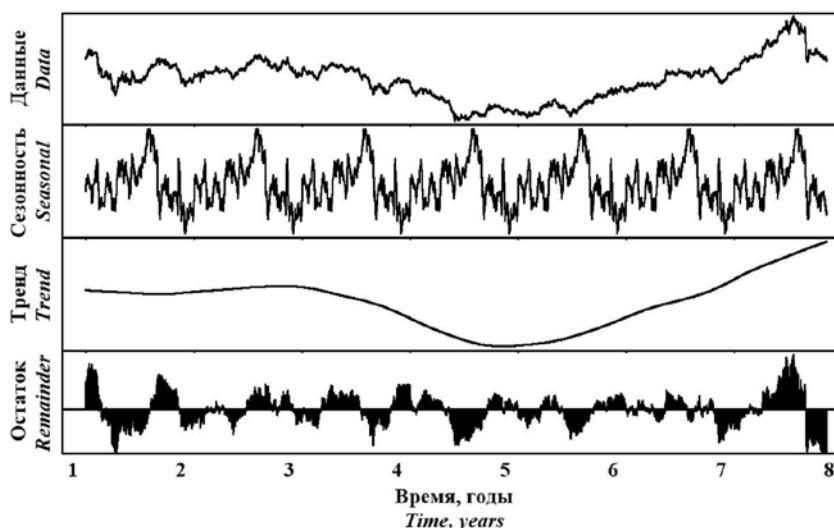


Рис. 1. График модели STL

Fig. 1. STL model chart

Для оценки стационарности наблюдаемого временного ряда применялся тест KPSS (Kwiatkowski – Phillips – Schmidt – Shin Test) [14]. Полученные результаты показывают, что исследуемый ряд является стационарным.

Для определения параметров и построения графиков оценки автокорреляции и частичной автокорреляции функции временного ряда использовались стандартные функции автокорреляции ACF и частной автокорреляции PACF языка R соответственно. Автокорреляционные графики служат визуальным инструментом для определения стационарности ряда и необходимы для выбора параметров порядка модели ARIMA.

Рассматриваемый ряд функции ACF имеет автокорреляции со многими задержками, но связано это может быть с корреляцией переноса с первого или раннего запаздывания, так как на графике PACF присутствует всплеск на лаге 1, а 95%-ый интервал значимости представлен на рисунке 2 как пунктирные линии.

Тест Дики – Фуллера [15] по разным данным отвергает нулевые гипотезы о нестационарности [16]. При построении разрозненных рядов можно увидеть осциллирующий рисунок около нуля без видимого сильного тренда. Это говорит о том, что различие условий порядка единицы является достаточным, а значит, может быть включено в модель.

Чтобы определить параметры модели ARIMA, необходимо оценить выбросы на

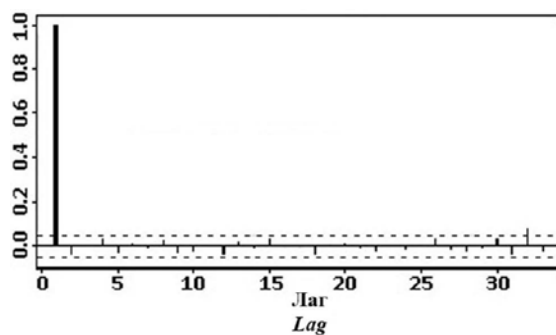


Рис. 2. График PACF

Fig. 2. PACF chart

лагах различных серий. Поэтому для дифференцированных данных необходимо построить графики ACF, PACF (рис. 3а, 3б). На представленных графиках можно увидеть, что в лагах 3 и 8 наблюдаются существенные всплески. Следовательно, можно сделать вывод о том, что один из параметров будет равен 8.

Строим модель ARIMA (автоматическое регрессионное интегрированное скользящее среднее) с тремя параметрами $p=1$, $d=1$, $q=8$ (рис. 4), где p – авторегрессионный компонент, q – компонент скользящей средней, d – порядок интегрирования, который показывает число несезонных различий, необходимое для стационарности. На графиках ACF и PACF (рис. 4) можно увидеть, что в лагах полностью отсутствуют выбросы.

Сравним значения моделей ARIMA (1,1,8) и ARIMA (0,1,0), определяемые с помощью информационного критерия Акаике

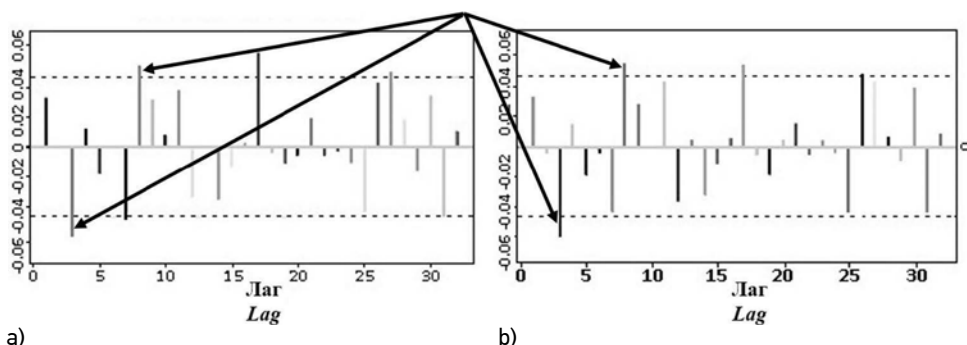


Рис. 3. График для дифференцированных данных: а) ACF; б) PACF

Fig. 3. Chart for differentiated data: a) ACF; b) PACF

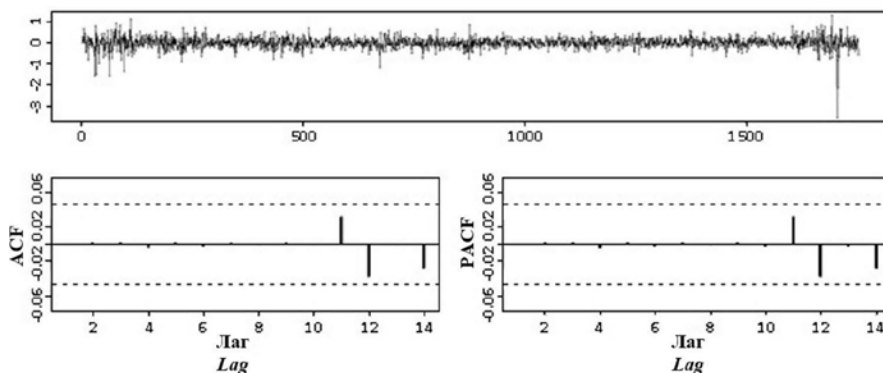


Рис. 4. График сезонности ARIMA (1,1,8)

Fig. 4. ARIMA seasonality chart (1,1,8)

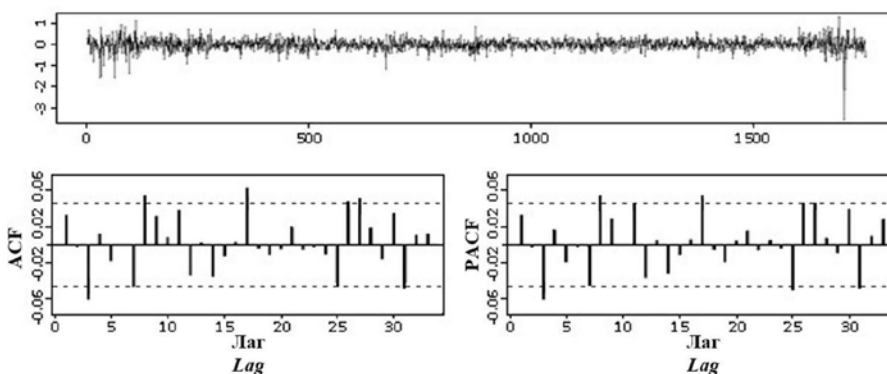


Рис. 5. График сезонности ARIMA (0,1,0)

Fig. 5. ARIMA seasonality chart (0,1,0)

(AIC) языком R автоматически (рис. 5) [13]. Рассчитанное значение критерия AIC для модели ARIMA (1,1,8) получилось равным 393,24, а для модели ARIMA (0,1,0) – 398,52. Исходя из полученных данных, можно сделать вывод о том, что модель ARIMA с параметрами (1,1,8) является лучшей.

По окончании выбора модели можно построить следующие прогнозные графики для цен акций компании ПАО «Сбербанк»: фактический график (Actual), график линии тренда (Trend), график прогноза (Forecast). На рисунке 6 представлен прогнозный график модели ARIMA.

Проверка достоверности прогнозных значений производилась по следующим параметрам: коэффициент детерминации (Multiple R-squared), стандартная ошибка, скорректированный R-квадрат (Adjusted R-squared),

P-значение (p-value), тест значимости (Significant codes), значение F-статистики.

Получены следующие результаты на основе проверки прогнозных значений: коэффициент детерминации – 0,943; стандартная ошибка – 10,1; скорректированный R-квадрат – 0,914; P-значение: p-value меньше 0,05, следовательно, модель является значимой; тест значимости: P-значение от 0 до 0,001, а значит, модель является значимой; значение F-статистики меньше 0,05, следовательно, модель является значимой.

Полученные результаты позволяют сделать вывод о том, что прогнозные значения модели ARIMA достоверны на 90%, и это соответствует допустимой норме.

ARCH является очередной анализируемой моделью. Для применения данной модели необходимо сначала выяснить, обла-

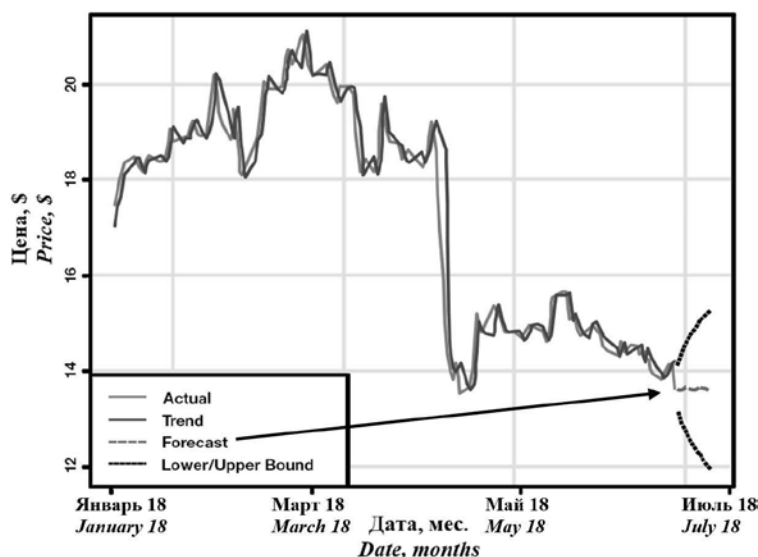


Рис. 6. График модели ARIMA (1,1,8)

Fig. 6. ARIMA model chart (1,1,8)

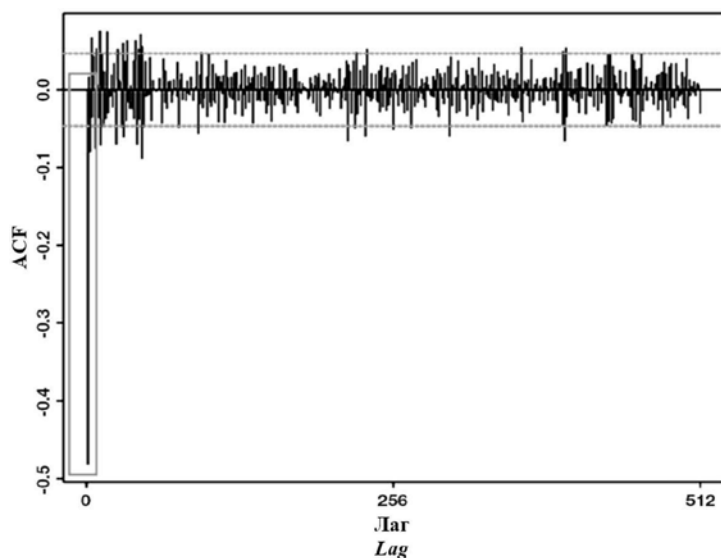


Рис. 7. График ACF для дифференцированных данных

Fig. 7. ACF chart for differentiated data

дают ли данные какой-нибудь тенденцией или сезонностью, для того чтобы затем ее устранить. Для этого реализуется модель STL аналогично описанной выше. Поскольку сезонность в исходном ряде присутствует, она была устранена с помощью декомпозиции ряда [17].

Для определения параметров модели ARCH была использована функция автокор-

реляции (ACF). В рассматриваемом ряде имеется значительная автокорреляция со многими задержками.

График ACF для дифференцированных данных (рис. 7) на основе выбросов на определенных лагах различной серии строится для определения параметров модели ARCH. График показывает, что значительный всплеск имеется на лаге 1. Следовательно, можно сде-

лать вывод о том, что параметр автокорреляции будет равен 1 [18].

При построении графика модели ARCH с параметрами 1,1 можно увидеть, что в лагах полностью отсутствуют выбросы. При сравнении его с моделью язык программирования R с помощью информационного критерия Акаике (AIC) автоматически рассчитал это значение. Значение критерия AIC для модели ARCH (1,1) получилось равным 398,587, у модели ARCH (0,1) – 399,614. Полученные результаты позволяют сделать вывод о том, что лучшей моделью ARCH является модель с параметрами 1,1.

После выбора модели были построены прогнозные графики для цен акций компании ПАО «Сбербанк». Для модели необходимо было построить фактический график (Actual), график линии тренда (Trend), прогноза (Forecast). На рисунке 8 представлена цена акций компании ПАО «Сбербанк» и ее прогнозный график на основе модели ARCH.

Получены следующие результаты на основе проверки прогнозных значений: коэффициент детерминации – 0,528; стандартная

ошибка – 18,15; скорректированный R-квадрат – 0,511; P-значение: p-value меньше 0,05, следовательно, модель значима; тест значимости от 0,001 до 0,01, а значит, модель является умеренно значимой; значение F-статистики меньше 0,05, следовательно, модель является значимой.

Таким образом, можно сделать вывод о том, что прогнозные значения модели ARCH достоверны лишь на 80%, а значит, прогнозные значения могут быть ложными.

Единственная модель GARCH, используемая на практике, – GARCH (1,1) [8]. График прогнозных значений этой модели имеет вид, представленный на рисунке 9.

Получены следующие результаты на основе проверки прогнозных значений: коэффициент детерминации – 0,895; стандартная ошибка – 11,34; скорректированный R-квадрат – 0,884; P-значение: p-value меньше 0,05, следовательно, модель является значимой; тест значимости: P-значение от 0 до 0,001, а значит, модель является значимой; значение F-статистики меньше 0,05, следовательно, модель является значимой.



Рис. 8. Прогнозный график модели ARCH

Fig. 8. ARCH model forecast chart

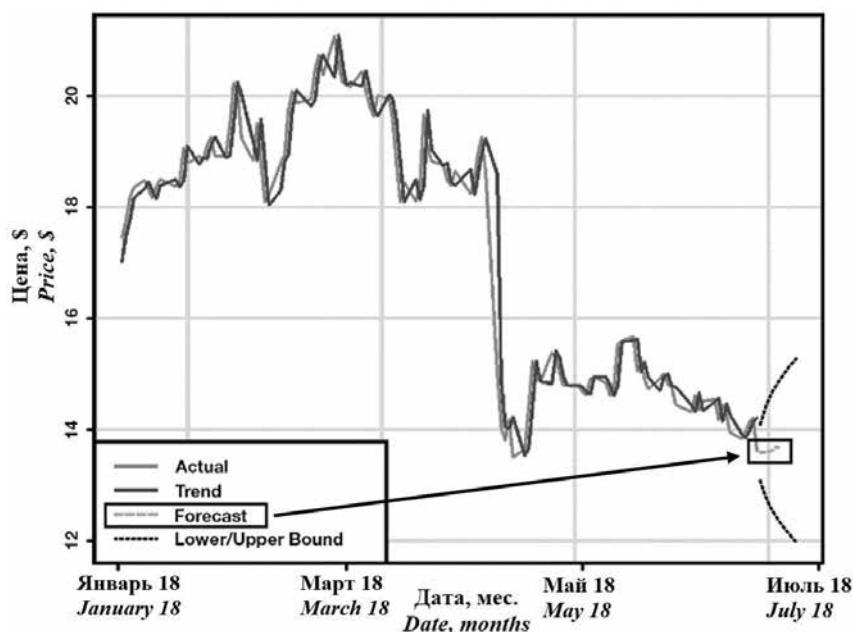


Рис. 9. Прогнозный график модели GARCH

Fig. 9. GARCH model forecast chart

Таким образом, можно сделать вывод о том, что прогнозные значения модели GARCH являются достоверными на 85%, следовательно, им можно доверять.

Реализация моделей прогнозирования фундаментального анализа

Так же как и при техническом анализе, основными данными для анализа алгоритмов моделей прогнозирования являются акции компании ПАО «Сбербанк». Данные о макропоказателях, которые используются при фундаментальном анализе, были получены за период с июня 2011 года по настоящее время с международного финансового портала investing.com.

Входные данные аналогичны используемым при техническом анализе и представляют собой дату (год, месяц, день) и цену закрытия в долларах США (ежедневная цена).

Кроме того, были отобраны два макроэкономических показателя, необходимые для построения многофакторной модели, – ди-

намик официального курса рубля по отношению к доллару США и стоимость нефти марки Brent (за 1 баррель в долларах США).

Для оценки связей факторов модели был использован коэффициент корреляции Спирмена (Spearman correlation coefficient). Когда полученный коэффициент корреляции имеет отрицательное значение, присутствует обратная связь. Если значение положительное, то имеет место быть прямая связь. Чем ближе коэффициент корреляции к единице по модулю, тем сильнее связь между факторами [19].

Во-первых, была оценена связь между стоимостью акции Сбербанка и ценой нефти марки Brent. При вычислении на языке R коэффициент корреляции Спирмена оказался равен 0,542, что соответствует сильной связи между факторами.

Во-вторых, была оценена связь между ценой акции Сбербанка и курсом доллара США. Коэффициент корреляции Спирмена при этом оказался равен 0,52, что подтверждает сильную связь между факторами.

Исходя из этого, можно сделать вывод о том, что показатели для двухфакторной мо-

дели прогнозирования курса акций подобраны правильно и на основе их можно строить прогнозный график (рис. 10).

Получены следующие результаты на основе проверки прогнозных значений: коэффициент детерминации – 0,987; стандартная ошибка – 9,86; скорректированный R-квадрат – 0,991; Р-значение: p-value меньше 0,05, следовательно, модель является значимой; тест значимости: Р-значение от 0 до 0,001, а значит, модель является значимой; значение F-статистики меньше 0,05, следовательно, модель значима.

Таким образом, можно сделать вывод о том, что прогнозные значения многофакторной модели более чем на 90% достоверны.

Дерево решений станет последней из рассматриваемых моделей. Исходными данными для построения дерева решений будет следующая информация об акции: Open (цена открытия); High (максимальная цена); Low (минимальная цена); Close (цена закрытия); Volume (объем).

Для дальнейшего моделирования требуется вычислить необходимые показатели для построения дерева [20]: EMA (Exponential Moving Average) – скользящая средняя, RSI (Relative Strength Index) – индекс относительной силы, Stochastic (Stochastic Oscillator) – стохастический осциллятор, class (Up/Down) – повышение или понижение цены, MACD (Moving Average Convergence/Divergence) – схождение/расхождение скользящих средних.

Для выбора и оценки метрик также могут быть использованы деревья решений. Индикаторы, которые находятся ближе всего к вершине дерева, дают более чистое деление с большим количеством информации, чем расположенные в нижней части.

Для подтверждения неперепределенности данных необходимо использовать усеченное дерево решений. Для этого необходимо определить параметр сложности, который является стоимостью или снижением производительности. После этого добавляется следующий раздел; размер дерева выбирается так, чтобы минимизировать ошибки.

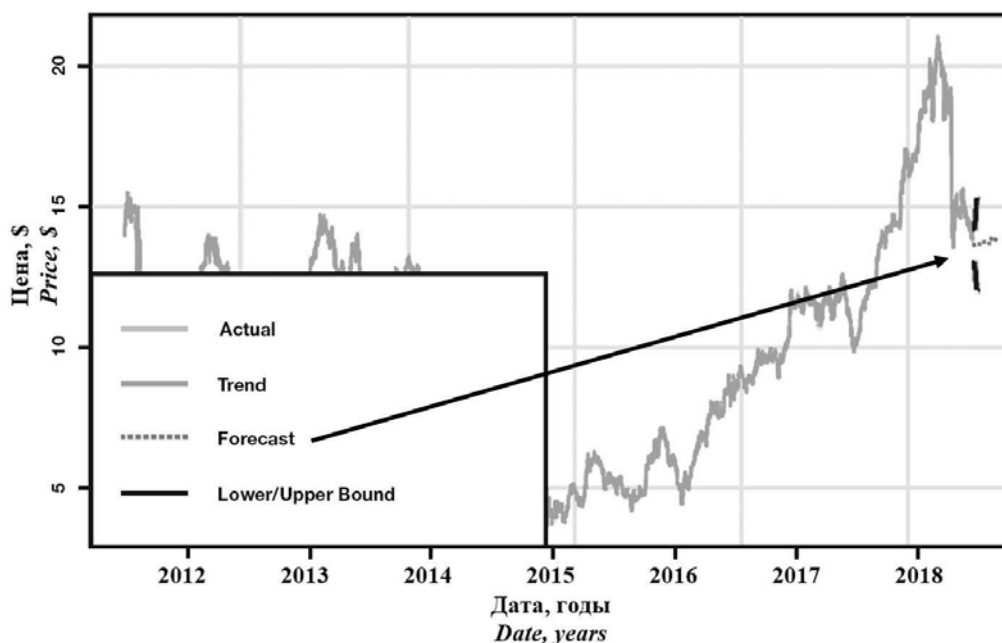


Рис. 10. Прогнозный график двухфакторной модели

Fig. 10. Two-factor model forecast chart

На рисунке 11 приведен график расчета среднего геометрического значения для деревьев каждого размера. Как видно из этого рисунка, лучшим деревом из рассмотренных является дерево с тремя ветвями, так как у него минимальное значение cross-validated error (ошибки перекрестной проверки) и достоверность модели оценивается на уровне 0,996.

Концевые листья итогового дерева решений с тремя ветвями являются прогнозными значениями. В результате получаем следующее условие: если значение индикатора RSI в прогнозный день меньше 44, переходим к левой стороне дерева. Если величина индикатора $EMA \leq -0,047$, то совершаем обход дерева по правой стороне с проверкой стохастического осциллятора. Если величина индикатора $EMA < 30$, то цена акции будет снижаться с вероятностью 0,55, в противном случае цена акции будет увеличиваться с вероятностью 0,64.

Получены следующие результаты на основе проверки прогнозных значений: коэффициент детерминации – 0,535; стандартная

ошибка – 15,23; скорректированный R-квадрат – 0,525; Р-значение: p-value меньше 0,05, следовательно, модель является значимой; тест значимости: Р-значение от 0,001 до 0,01, а значит, модель является умеренно значимой; значение F-статистики меньше 0,05, следовательно, модель является значимой.

Таким образом, можно сделать вывод о том, что достоверность прогнозных значений дерева решений составляет около 80%, а значит, прогнозу можно доверять с осторожностью.

В ходе практической реализации прогнозных моделей технического и фундаментального анализа были протестированы модели ARIMA, ARCH, GARCH, многофакторная модель и дерево решений. Исследование моделей проводилось на основании данных о цене акций ПАО «Сбербанк». Проверка прогнозных значений производилась по ряду статистических параметров, результаты представлены в таблице 1.

Анализ представленной таблицы показывает, что модель ARCH имеет худший про-

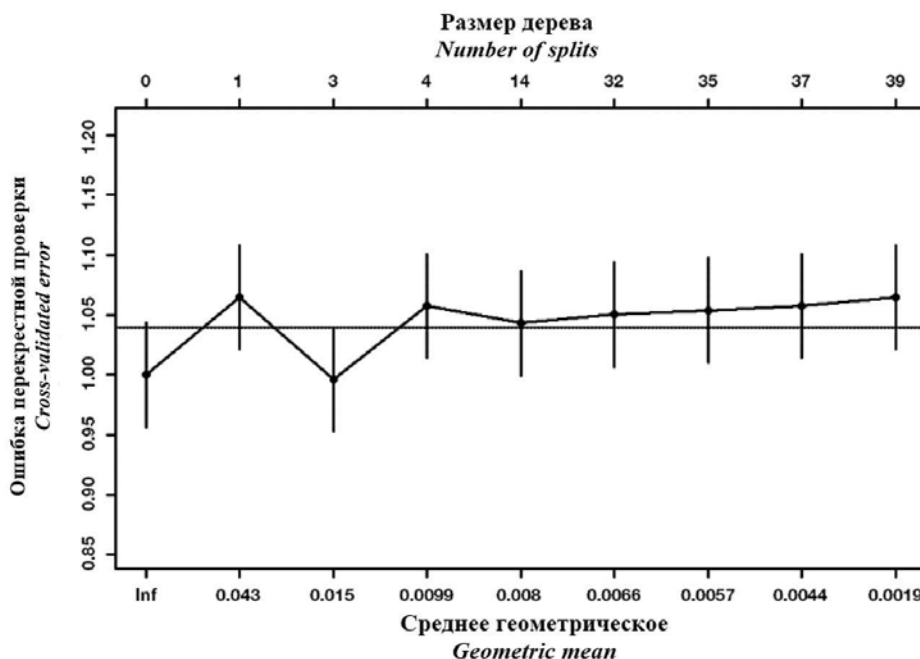


Рис. 11. Среднее геометрическое для деревьев каждого размера

Fig. 11. Geometric mean for trees of each size

Таблица 1. Сравнительная таблица результатов прогнозных моделей**Table 1.** Comparative table of predictive model results

Показатель <i>Indicator</i>	Модель ARIMA <i>Model ARIMA</i>	Модель ARCH <i>Model ARCH</i>	Модель GARCH <i>Model GARCH</i>	Многофакторная Модель <i>Multi-factor forecasting model</i>	Дерево Решений <i>Decision tree model</i>
Коэффициент детерминации <i>Multiple R-squared</i>	0,943	0,528	0,895	0,987	0,535
Стандартная ошибка <i>Standard error</i>	10,1	18,15	11,34	9,86	15,23
Скорректированный R-квадрат <i>Adjusted R-squared</i>	0,914	0,511	0,884	0,991	0,525
P-значение <i>p-value</i>	< 0,05	< 0,05	< 0,05	< 0,05	< 0,05
Тест значимости <i>Significant codes</i>	$0 \leq p \leq 0,001$	$0,001 \leq p \leq 0,01$	$0 \leq p \leq 0,001$	$0 \leq p \leq 0,001$	$0,001 \leq p \leq 0,01$
Значение F-статистики <i>F-statistic</i>	< 0,05	< 0,05	< 0,05	< 0,05	< 0,05

гнозный результат (стандартная ошибка 18,15), что полностью согласуется с исследованиями [16]. Прогнозные результаты модели ARIMA близки к результатам модели GARCH. Это можно объяснить тем, что параметры моделей GARCH и ARIMA похожи. Результаты дерева решений аналогичны прогнозным результатам модели ARCH и не являются достаточно удовлетворительными. Данное обстоятельство объясняется сложностью реализации модели и неправильным выбором коэффициентов. Многофакторная модель показала лучший результат среди всех моделей. Стандартная ошибка модели составила менее 10%, что соответствует допустимой норме.

Заключение

Прогнозные модели фундаментального и технического анализа одинаково важны при

построении графиков прогнозных значений для ценных бумаг. Однако практическое исследование показало, что многофакторная модель, относящаяся к фундаментальному анализу, дает более точную оценку при построении прогноза стоимости акций ПАО «Сбербанк».

Результаты анализа позволят разработать методику и реализовать ее с использованием программных средств для проведения моделирования прогнозных значений в финансовой отрасли для дальнейшего принятия решений. Также возможно применение данных методик в искусственных нейронных сетях.

Перспективой дальнейших исследований можно считать построение алгоритма прогнозирования влияния факторов на цену акции ПАО «Сбербанк» с помощью различных статистических показателей на основе многофакторной модели.

Список литературы

1. Wafi A. S., Hassan H., Mabrouk A. Fundamental analysis models in financial markets – review study // Procedia Economics and Finance. 2015. Vol. 30. P. 939–947. DOI: 10.1016/S2212-5671(15)01344-1.
2. Pring M. J. Technical Analysis Explained. – N.Y.: McGraw-Hill, 2002. – 644 p.
3. Rosenberg M. R., Folkerts-Landau D. The Deutsche Bank guide to exchange-rate determination: A survey of exchange rate forecasting models and strategies. – Business & Economics, 2002. – 160 p.
4. Park C.-H., Irwin S. H. What do we know about the profitability of technical analysis? // Journal of Economic Surveys. 2007. Vol. 21. No. 4. P. 786–826. DOI: 10.1111/j.1467-6419.2007.00519.x.

5. Карпунова С. Ю. Преимущества модели ARIMA для краткосрочного прогнозирования поведения ценовых графиков Forex [Электронный ресурс]. – Донецкий национальный технический университет, 2007. URL: <http://masters.donntu.org/2007/fvti/karpunova/diss/index.html> (дата обращения: 19.09.2020).
6. Bollerslev T. Generalized autoregressive conditional heteroskedasticity // *Journal of Econometrics*. 1986. Vol. 31. No. 3. P. 307–327. DOI: 10.1016/0304-4076(86)90063-1.
7. Dhamija A. K., Yadav S. S., Jain P. K. Forecasting volatility of carbon under EU ETS: a multi-phase study // *Environmental Economics and Policy Studies*. 2016. Vol. 19. P. 299–335. DOI: 10.1007/s10018-016-0155-4.
8. Hill T., Marquez L., O'Connor M., Remus W. Artificial neural network models for forecasting and decision making // *International Journal of Forecasting*. 1994. Vol. 10. No. 1. P. 5–15. DOI: 10.1016/0169-2070(94)90045-0.
9. Ratnaparkhi S., Paradkar M. Use decision trees in machine learning to predict stock movements // *Quantinsti*. Oct. 04, 2017. URL: <https://www.quantinsti.com/blog/use-decision-trees-machine-learning-predict-stock-movements/> (дата обращения: 05.06.2020).
10. Cleveland R. B., Cleveland W. S., McRae J. E., Terpenning I. STL: a seasonal-trend decomposition procedure based on loess // *Journal of Official Statistics*. 1990. Vol. 6. No. 1. P. 3–33.
11. Levin A. U. Stock Selection via nonlinear multi-factor models // *Proceedings of the Advances in Neural Information Processing Systems 8 (NIPS 1995)*. 1995. P. 966–972.
12. Fox J., Andersen R. Using the R statistical computing environment to teach social statistics courses. 2005. P. 1–35. URL: <http://socserv.mcmaster.ca/jfox/Teaching-with-R.pdf> (дата обращения: 11.08.2020).
13. Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? // *Journal of Econometrics*. 1992. Vol. 54. No. 1-3. P. 159–178. DOI: 10.1016/0304-4076(92)90104-Y.
14. STL decomposition // *OTexts*. URL: <https://www.otexts.org/fpp/6/5> (дата обращения: 10.01.2021).
15. Dickey D. A., Fuller W. A. Distribution of the estimators for autoregressive time series with a unit root // *Journal of the American Statistical Association*. 1979. Vol. 74. P. 427–431. DOI: 10.2307/2286348.
16. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice (2nd ed.) // *OTexts*. 2018. URL: <https://otexts.com/fpp2/> (дата обращения: 22.11.2020).
17. Lehman A., O'Rourke N., Hatcher L., J. Stepanski E. JMP for basic univariate and multivariate statistics: a step-by-step guide. – N.C.: Cary: SAS Press, 2005. – 481 p.
18. Tiwari S., Pandit R., Richhariya V. Predicting future trends in stock market by decision tree rough-set based hybrid system with HHMM // *International Journal of Electronics and Computer Science Engineering*. 2013. Vol. 1. No. 3. P. 1578–1857.
19. Meese R. A., Rogoff K. Empirical exchange rate models of the seventies: Do they fit out of sample? // *Journal of International Economics*. 1983. Vol. 14. No. 1-2. P. 3–24. DOI: 10.1016/0022-1996(83)90017-X.
20. Fornari F., Mele A. Sign- and volatility-switching and ARCH models: theory and applications to International Stock Markets // *Journal of Applied Econometrics*. 1997. Vol. 12. No. 1. P. 49–65. DOI: 10.1002/(SICI)1099-1255(199701)12:1<49::AID-JAE422>3.0.CO;2-6.

Сведения об авторах

Суворов Александр Олегович, ORCID 0000-0002-5246-5091, канд. техн. наук, доцент, кафедра информационных технологий в бизнесе, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия, AOSuvorov@hse.ru

Петренко Александр Анатольевич, ORCID 0000-0002-0407-052X, канд. техн. наук, доцент, кафедра информационных технологий и автоматизированных систем, Пермский национальный исследовательский политехнический университет, Пермь, Россия, hawk321@mail.ru

Неприна Анастасия Дмитриевна, ORCID 0000-0003-4926-7235, ведущий системный аналитик, акционерное общество «АЛЬФА-БАНК», Москва, Россия, nastush99@yandex.ru

Статья поступила 04.08.2021, рассмотрена 18.08.2021, принята 06.10.2021

References

1. Wafi A. S., Hassan H., Mabrouk A. Fundamental analysis models in financial markets – review study. *Procedia Economics and Finance*, 2015, vol.30, pp.939-947. DOI: 10.1016/S2212-5671(15)01344-1.
2. Pring M. J. *Technical Analysis Explained*. N.Y., McGraw-Hill, 2002, 644 p.

3. Rosenberg M. R., Folkerts-Landau D. The Deutsche Bank guide to exchange-rate determination: A survey of exchange rate forecasting models and strategies. *Business & Economics*, 2002, 160 p.
4. Park C.-H., Irwin S. H. What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 2007, vol.21, no.4, pp.786-826. DOI: 10.1111/j.1467-6419.2007.00519.x.
5. Karpunova S. Advantages of ARIMA model for short-term forecasting of prices graphics of Forex. Donetsk national technical university, 2007 (in Russian). Available at: <http://masters.donntu.org/2007/fvti/karpunova/diss/index.html> (accessed 19.09.2020).
6. Bollerslev T. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 1986, vol.31, no.3, pp.307-327. DOI: 10.1016/0304-4076(86)90063-1.
7. Dhamija A. K., Yadav S. S., Jain P. K. Forecasting volatility of carbon under EU ETS: a multi-phase study. *Environmental Economics and Policy Studies*, 2016, vol.19, pp.299-335. DOI: 10.1007/s10018-016-0155-4.
8. Hill T., Marquez L., O'Connor M., Remus W. Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, 1994, vol.10, no.1, pp.5-15. DOI: 10.1016/0169-2070(94)90045-0.
9. Ratnaparkhi S., Paradkar M. Use decision trees in machine learning to predict stock movements. *Quantinsti*, Oct. 04, 2017. Available at: <https://www.quantinsti.com/blog/use-decision-trees-machine-learning-predict-stock-movements/> (accessed 05.06.2020).
10. Cleveland R. B., Cleveland W. S., McRae J. E., Terpenning I. STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 1990, vol.6, no.1, pp.3-33.
11. Levin A. U. Stock Selection via nonlinear multi-factor models. *Proceedings of the Advances in Neural Information Processing Systems 8 (NIPS 1995)*, 1995, pp.966-972.
12. Fox J., Andersen R. Using the R statistical computing environment to teach social statistics courses, 2005, pp.1-35. Available at: <http://socserv.mcmaster.ca/jfox/Teaching-with-R.pdf> (accessed 11.08.2020).
13. Kwiatkowski D., Phillips P. C. B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 1992, vol.54, no.1-3, pp.159-178. DOI: 10.1016/0304-4076(92)90104-Y.
14. STL decomposition. OTexts. Available at: <https://www.otexts.org/fpp/6/5> (accessed 10.01.2021).
15. Dickey D. A., Fuller W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 1979, vol.74, pp.427-431. DOI: 10.2307/2286348.
16. Hyndman R. J., Athanasopoulos G. *Forecasting: principles and practice* (2nd ed.). OTexts. 2018. Available at: <https://otexts.com/fpp2/> (accessed 22.11.2020).
17. Lehman A., O'Rourke N., Hatcher L., J. Stepanski E. *JMP for basic univariate and multivariate statistics: a step-by-step guide*. N.C., Cary, SAS Press, 2005, 481 p.
18. Tiwari S., Pandit R., Richhariya V. Predicting future trends in stock market by decision tree rough-set based hybrid system with HHMM. *International Journal of Electronics and Computer Science Engineering*, 2013, vol.1, no.3, pp.1578-1857.
19. Meese R. A., Rogoff K. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 1983, vol.14, no.1-2, pp.3-24. DOI: 10.1016/0022-1996(83)90017-X.
20. Fornari F., Mele A. Sign- and volatility-switching and ARCH models: theory and applications to International Stock Markets. *Journal of Applied Econometrics*, 1997, vol.12, no.1, pp.49-65. DOI: 10.1002/(SICI)1099-1255(199701)12:1<49::AID-JAE422>3.0.CO;2-6.

About the authors

Aleksandr O. Suvorov, ORCID 0000-0002-5246-5091, Cand. Sci. (Eng.), Associate Professor, Information Technologies in Business Department, HSE University, Moscow, Russia, AOSuvorov@hse.ru

Aleksandr A. Petrenko, ORCID 0000-0002-0407-052X, Cand. Sci. (Eng.), Associate Professor, Information Technologies and Computer-Aided Systems Department, Perm National Research Polytechnic University, Perm, Russia, hawk321@mail.ru

Anastasija D. Neprina, ORCID 0000-0003-4926-7235, Senior System Analyst, Joint Stock Company "ALFA-BANK", Moscow, Russia, nastush99@yandex.ru

Received 04.08.2021, reviewed 18.08.2021, accepted 06.10.2021

Метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов

Ю. И. Бутенко^{1*}, Ю. В. Строганов¹, А. М. Сапожков¹

¹ *Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия*
* *iubutenko@bmstu.ru*

Аннотация. В статье представлен метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов на основе структурных моделей терминологических словосочетаний. Описаны существующие подходы к извлечению терминов на основе метода извлечения устойчивых словосочетаний, статистических и гибридных методов, а также отмечены лингвистические аспекты терминоведения, не охваченные перечисленными методами. Охарактеризован лексический состав научно-технических текстов, приведена классификация специальной лексики в научно-технических текстах. Изучены структурные особенности терминологической лексики. Представлены наиболее продуктивные модели многокомпонентных терминологических словосочетаний в русском языке. Предложен метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов, а также описаны его этапы. Показано, что на первом этапе проводится морфолого-синтаксический анализ текста путем приписывания каждому слову его грамматических характеристик. Затем происходит исключение частей речи, которые не могут входить в состав русскоязычных многокомпонентных терминов, а также стоп-слов, которые вместе с термином образуют свободные словосочетания. Полученные цепочки слов далее соотносятся с шаблонами терминологических словосочетаний, имеющихся в базе структурных моделей терминов, а также с терминологическим словарем на предмет наличия исследуемого термина-кандидата. Обоснована необходимость привлечения терминолога для разрешения неоднозначных случаев. Каждый этап метода извлечения русскоязычных многокомпонентных терминов из научно-технических текстов проиллюстрирован примерами. Перечислены перспективы исследования, а также обоснована необходимость усложнения методов извлечения терминов путем дальнейшей классификации терминологической лексики по формальной и семантической структурам, видам антропоморфных терминов, номенклатурным названиям, нормативности/ненормативности терминологических единиц.

Ключевые слова: корпус текстов, научно-технические тексты, извлечение терминов, структура научно-технического текста, многокомпонентный термин

Для цитирования: Бутенко Ю. И., Строганов Ю. В., Сапожков А. М. Метод извлечения русскоязычных многокомпонентных терминов из научно-технических текстов // Прикладная информатика. 2021. Т. 16. № 6. С. 21–27. DOI: 10.37791/2687-0649-2021-16-6-21-27

Method for the extraction of Russian-language multicomponent terms from scientific and technical texts

Iu. Butenko^{1*}, Yu. Stroganov¹, A. Sapozhkov¹

¹ Bauman Moscow State Technical University, Moscow, Russia

*iubutenko@bmstu.ru

Abstract. The article presents a method for extracting Russian-language multicomponent terms from scientific and technical texts based on structural models of terminological collocations. The existing approaches to term extraction on the basis of the method of stable word combination extraction, statistical and hybrid methods are described, and the linguistic aspects of terminology, not covered by the listed methods, are noted. The lexical composition of scientific and technical texts is characterized, the classification of special vocabulary in scientific and technical texts is given. The structural features of terminological vocabulary have been studied. The most productive models of multi-component terminological word combinations in Russian are presented. A method for extracting Russian-language multicomponent terms from scientific and technical texts is offered, and its stages are described. It is shown that the first stage involves morphological and syntactic analysis of the text by attributing to each word its grammatical characteristics. Then there is the exclusion of parts of speech, which can not be part of the Russian multisyllabic terms, as well as stop-words, which together with the term form free word combinations. The resulting word chains are further correlated with the templates of terminological word combinations available in the database of structural models of terms, as well as the terminological dictionary for the presence of the studied candidate term. The necessity of involving a terminologist to resolve ambiguous cases is substantiated. Each step of the method for extracting Russian-language multicomponent terms in scientific and technical texts is illustrated by examples. Further research perspectives are listed, and the necessity of complicating the methods of text extraction, by further classification of terminological vocabulary according to formal and semantic structures, types of anthropomorphic terms, nomenclatural names, normativity/non-normativity of terminological units is substantiated.

Keywords: text corpus, scientific and technical texts, term extraction, structure of scientific and technical text, multi-component term

For citation: Butenko Iu., Stroganov Yu., Sapozhkov A. Method for the extraction of Russian-language multicomponent terms from scientific and technical texts. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.21-27 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-21-27

Введение

Корпуса текстов обычно размечаются для удобства пользования, т. е. текстам и содержащимся в них языковым единицам приписываются специальные метки. Размеченные корпуса обеспечивают специализированными поисковыми системами, реализующими грамматические и лек-

сические виды поиска [4]. Так, для корпуса научно-технических текстов наибольшую значимость приобретает терминологическая разметка, так как именно термины выступают основным средством передачи информации [9].

Работа с корпусами научно-технических текстов требует особого инструментария для выявления устойчивых термино-

логических сочетаний [13]. Среди наиболее распространенных методов выявления многокомпонентных терминов в текстах используют метод выявления устойчивых сочетаний, статистический и гибридный методы [1]. В основе метода выявления устойчивых сочетаний лежит использование грамматики лексико-синтаксических шаблонов, представляющих собой структурные модели лингвистических конструкций [12]. Статистический подход заключается в нахождении n-грамм по заданным частотным характеристикам. Гибридный подход для выделения терминологических сочетаний, объединяющий лингвистический и статистический методы, заключается в предварительном описании моделей, по которым могут быть построены термины, для последующего нахождения их в корпусе [2, 3]. Внутри множеств однотипных синтаксических конструкций выполняется ранжирование в соответствии с той или иной статистической мерой [11]. В рамках указанных методов отсутствует возможность отличить многокомпонентный термин от свободного словосочетания, например *неорганическая химия* и *современная химия* [10]. В текстах часто встречаются и словосочетания вида «общеупотребительное слово и многокомпонентный термин», например *современная неорганическая химия*. В то же время развивающиеся терминологии компьютерных наук, авиации и космонавтики, нанотехнологий и других предметных областей используют новые способы терминообразования, таким образом расширяя структурные модели терминов [8].

Целью исследования является описание метода извлечения русскоязычных многокомпонентных терминологических единиц из корпуса научно-технических текстов на основе структурных моделей терминологических словосочетаний, дополненных грамматическими и лексическими ограничениями на сочетаемость слов в составе многокомпонентного термина.

Формальная структура многокомпонентных терминов

В научно-технических текстах используется специальная лексика следующих видов: общенаучная терминология, узкоспециализированные термины и номенклатуры. Общенаучная терминология представлена собирательными наименованиями, которые могут быть названиями терминологий, например *анализ, синтез, вид, метод, процесс, материал* и др., и отражает абстрактные понятия, зачастую входящие в состав дефиниций. Но с другой стороны, общенаучная лексика входит в состав многокомпонентных терминологических словосочетаний, которые обозначают одно понятие, например *методы химического анализа, оборудование для химического анализа, виды сварки, сварочный материал*. Номенклатурой называют терминологическое обозначение частного специального понятия какой-либо предметной области, например *Ту-156, Уолл Стрит, Захват-22*.

Термин может быть однокомпонентным и состоять из ключевого слова или представлять собой терминологическую группу, в состав которой входит ключевое слово или ядро группы, одно или несколько левых определений и одно или несколько правых или предложных определений, которые уточняют или модифицируют смысл терминологической единицы [5].

Наиболее сложным явлением в процессе автоматического извлечения терминов из научно-технических текстов представляют собой многокомпонентные термины – терминологические словосочетания, образованные лексическим и синтаксическим способами, то есть словосочетания, образованные по определенным моделям [6]. В основе анализа терминологических словосочетаний лежит вычленение исходного терминологического словосочетания и определение последовательности присоединения к нему остальных элементов. Исходным терминологическим словосочетанием, как пра-

вило, является двухкомпонентное субстантивное терминологическое словосочетание, которое в рамках трех-, четырехкомпонентного терминологического словосочетания характеризуется более тесными структурно-семантическими отношениями [5].

Наиболее продуктивным способом номинации является образование составных терминов, состоящих из двух, трех, четырех и пяти компонентов [5]. На практике встречаются термины, состоящие из 10–11 компонентов. В таблице 1 представлены наиболее продуктивные модели многокомпонентных терминологических словосочетаний в русском языке.

Таблица 1. Основные структурные модели многокомпонентных терминологических словосочетаний

Table 1. Basic structural models of multi component term combinations

Количество компонент <i>Number of components</i>	Структурные модели <i>Structural models</i>
1	S_1
2	$(S_6 \wedge S_1) \vee (S_9 \wedge S_1) \vee (S_1 \wedge S_2) \vee (S_6 \wedge S_1)$
3	$(S_6 \wedge S_6 \wedge S_1) \vee (S_6 \wedge S_1 \wedge S_2) \vee (S_6 \wedge S_1 \wedge S_8) \vee (S_6 \wedge S_1 \wedge S_3)$
4	$(S_6 \wedge S_1 \wedge S_{11} \wedge S_5) \vee (S_6 \wedge S_1 \wedge S_{11} \wedge S_4)$

Множество S соответствует морфологическим характеристикам слов русского языка, которые входят в состав многокомпонентного словосочетания. Множество S состоит из 12 элементов: S_1 – имя существительное в именительном падеже, S_2 – имя существительное в родительном падеже, S_3 – имя существительное в винительном падеже, S_4 – имя существительное в творительном падеже, S_5 – имя существительное в предложном падеже, S_6 – имя прилагательное, S_7 – имя числительное, S_8 – наречие, S_9 – причастие, S_{10} – деепричастие, S_{11} – предлог. В связи с тем, что морфологические анализаторы не всегда корректно и полно определяют грамматические характеристики имен прилагательных S_6 ,

в предложенных моделях для них используется только характеристика части речи, а категории рода, числа и падежа будут такие же, как у имени существительного, которое стоит после имени прилагательного.

Добавление структурных моделей терминологических словосочетаний с пятью и более компонентами является нецелесообразным в силу низкой частотности употребления таких терминов на практике.

Метод извлечения многокомпонентных терминов из научно-технических текстов

Предлагаемый метод автоматического извлечения русскоязычных многокомпонентных терминов на основе базы данных структурных моделей терминологических словосочетаний состоит из пяти основных этапов. В качестве примера рассмотрим извлечение терминов для следующего фрагмента текста: *Как наука химия природных соединений возникла одновременно с органической химией.*

1. На первом этапе проводим анализ предложения по частям речи, например:

Как_{союз} наука_{СУЩ,неод,жр ед,им} химия_{СУЩ,неод,жр ед,им}
природных_{ПРИЛ} соединений_{СУЩ,неод,ср мн,рд} ВОЗ-
никла_{ГЛ,сов,неперех жр,ед,прош,изъяв} одновременно_{НАР}
с_{ПРЕДЛ} органической_{ПРИЛ} химией_{СУЩ,неод жр ед,тв}.

2. В состав терминологических словосочетаний не входят глаголы, союзы, местоимения, частицы, знаки препинания, а также некоторые сочетания частей речи, такие как «наречие + предлог» и прочее.

Как_{союз} наука_{СУЩ,неод,жр ед,им} химия_{СУЩ,неод,жр ед,им}
природных_{ПРИЛ} соединений_{СУЩ,неод,ср мн,рд}
возникла_{ГЛ,сов,неперех жр,ед,прош,изъяв} одновременно_{НАР}
с_{ПРЕДЛ} органической_{ПРИЛ} химией_{СУЩ,неод,жр ед,тв}.

Таким образом, остаются следующие цепочки слов:

наука_{СУЩ,неод,жр ед,им}
химия_{СУЩ,неод,жр ед,им} природных_{ПРИЛ} соеди-
нений_{СУЩ,неод,ср мн,рд}
органической_{ПРИЛ} химией_{СУЩ,неод,жр ед,тв}.

3. Проверяем полученные термины-кандидаты на наличие стоп-слов, указанных в специальной зоне словаря, и если они есть, убираем их. Под стоп-словами понимаем слова, которые образуют широко используемые коллокации с терминами, но в совокупности не являются терминами по сути, например *современная химия*, *рассматриваемый метод синтеза о-гликозидов*.

4. Полученные цепочки слов соотносим с шаблонами терминологических словосочетаний, имеющихся в базе структурных моделей терминов:

наука_{СУЩ,неод,жр ед,им} – S_1 принадлежит к моделям терминов;

химия_{СУЩ,неод,жр ед,им} природных_{ПРИЛ} соединений_{СУЩ,неод,ср мн,рд} – $S_1 \wedge S_6 \wedge S_2$ принадлежит к моделям терминов;

органической_{ПРИЛ} химией_{СУЩ,неод,жр ед,тв} – $S_6 \wedge S_1$ принадлежит к моделям терминов.

5. Полученные термины-кандидаты проверяем по словарю корпуса.

5.1. Если термин-кандидат есть в словаре, то извлекаем его как термин.

5.2. Если полученный термин-кандидат отсутствует в словаре, то отправляем его терминологу для обработки вручную.

5.3. Если термин-кандидат состоит из нескольких слов, то пробуем разбить его на несколько терминов.

Исследования терминологических словосочетаний в испанском, английском, немецком и других языках свидетельствуют о наличии структур многокомпонентных терминов, образованных по определенным моделям, отражающим особенности языка, на котором они образованы [7]. Для использования описанного метода извлечения многокомпонентных терминологических словосочетаний из научно-технических текстов на разных языках необходимо только наличие базы данных структурных моделей терминологических словосочетаний обрабатываемого языка.

Перспективой дальнейшего исследования также является усовершенствование метода

извлечения многокомпонентных терминов путем приписывания дополнительных классификационных тегов к термину:

- формальная структура (термин-слово, термин-словосочетание с указанием компонентов);

- содержательная структура (однозначные, многозначные);

- антропоморфные термины (термины-эпонимы, термины-антропонимы, термины-метафоры);

- нормативные и ненормативные терминологические единицы;

- номенклатурные названия и т. д.

Такая классификация терминов позволит решать ряд прикладных задач, например при исследовании тональности текста можно использовать классификацию нормативной и ненормативной терминологии. Последняя может содержать термины с положительной, нейтральной или негативной коннотацией, например: флюшка/флюра (флюорография) используются с нейтральной коннотацией, а херург вместо хирург обладает ярко выраженной отрицательной коннотацией.

Заключение

В статье описан пример применения метода извлечения русскоязычных многокомпонентных терминологических единиц из научно-технических текстов на основе структурных моделей терминологических единиц. Показано, что термин может быть однокомпонентным и состоять из ключевого слова или представлять собой терминологическое словосочетание, а наиболее продуктивным способом номинации является образование составных терминов, состоящих из двух, трех, четырех и пяти компонентов. Основные структурные модели многокомпонентных терминологических словосочетаний состоят из 12 элементов: имя существительное в именительном падеже, имя существительное в родительном падеже, имя существительное в винительном па-

деже, имя существительное в творительном падеже, имя существительное в предложном падеже, имя прилагательное, имя числительное, наречие, причастие, деепричастие, предлог. Метод извлечения русскоязычных многокомпонентных терминологических единиц из научно-технических текстов на основе структурных моделей терминологических единиц состоит из следующих этапов: анализ предложения по частям речи, исключение частей речи, не входящих в состав терминологических словосочетаний, проверка терминов-кандидатов на стоп-слова,

которые образуют широко используемые коллокации с терминами, но в совокупности не являются терминами по семантическому содержанию, проверка терминов-кандидатов по словарю корпуса или передача терминологу на ручную обработку. Перспективой дальнейшего исследования является усовершенствование метода извлечения многокомпонентных терминов путем приписывания дополнительных классификационных тегов к термину, а также исследование возможностей использования описанного метода для других языков.

Список литературы

1. Захаров В. П., Хохлова М. В. Автоматическое извлечение терминов из специальных текстов с использованием дистрибутивно-статистического метода как инструмент создания тезаурусов // Структурная и прикладная лингвистика. 2012. № 9. С. 222–233.
2. Клышинский Э. С., Кочеткова Н. А., Карпик О. В. Метод выделения коллокаций с использованием степенного показателя в распределении Ципфа // Новые информационные технологии в автоматизированных системах. 2018. № 21. С. 220–225.
3. Кочеткова Н. А. Метод извлечения технических терминов с использованием усовершенствованной меры странности // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2015. № 5. С. 25–32.
4. Кружков М. Г. Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов // Системы и средства информатики. 2015. Т. 25. № 2. С. 140–159.
5. Лейчик В. М. Оптимальная длина и структура термина // Вопросы языкознания. 1981. № 2. С. 63–73.
6. Новикова А. А. Сравнение инструментов Sketch Engine и Termostat для извлечения терминологии // International Journal of Open Information Technologies. 2020. Т. 8. № 11. С. 73–79.
7. Becerro F. B. Phraseological variations in medical-pharmaceutical terminology and its applications for English and German into Spanish translations // SciMedicine Journal. 2020. Vol. 2. No. 1. P. 22–29. DOI: 10.28991/SciMedJ-2020-0201-4.
8. Biziukova N. Yu., Tarasova O. A., Rudik A. V. et al. Automatic recognition of chemical entity mentions in texts of scientific publications // Automatic Documentation and Mathematical Linguistics. 2020. Vol. 54. No. 6. P. 306–315. DOI: 10.3103/S0005105520060023.
9. Butenko Iu. I., Garazha V. V. BMSTU Corpus of scientific and technical texts: conceptual framework // Applied Linguistics Research Journal. 2021. Vol. 5. No. 3. P. 76–81. DOI: 10.14744/alrj.2021.15579.
10. Loukachevitch N., Dobrov B. Ontological resources for representing security domain in information-analytical system // Открытые семантические технологии проектирования интеллектуальных систем. 2018. № 8. С. 185–191.
11. Sidnyaev N. I., Butenko I. I., Bolotova E. E. Formal grammar theory in recognition methods of unknown objects // Automatic Documentation and Mathematical Linguistics 2020. Vol. 54. No. 4. P. 215–225. DOI: 10.3103/S000510552004007X.
12. Simon N. I., Kešelj V. Automatic term extraction in technical domain using part-of-speech and common-word features // Proceedings of the ACM Symposium on Document Engineering. 2018. Article 51. P. 1–4. DOI: 10.1145/3209280.3229100.
13. Terryn A. R., Hoste V., Lefever E. In no uncertain terms: a dataset for monolingual and multilingual automatic term extraction from comparable corpora // Language Resources and Evaluation. 2020. No. 54. P. 385–418. DOI: 10.1007/s10579-019-09453-9.

Сведения об авторах

Бутенко Юлия Ивановна, ORCID 0000-0002-9776-5709, канд. техн. наук, доцент, кафедра теоретической информатики и компьютерных технологий, Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия, iubutenko@bmstu.ru

Строганов Юрий Владимирович, ORCID 0000-0001-5628-7395, старший преподаватель, кафедра программного обеспечения ЭВМ и информационных технологий, Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия, stroganovyv@bmstu.ru

Сапожков Андрей Михайлович, ORCID 0000-0001-8424-5686, студент, кафедра программного обеспечения ЭВМ и информационных технологий, Московский государственный технический университет им. Н. Э. Баумана, Москва, Россия, andreysapozhkov535@gmail.com

Статья поступила 04.08.2021, рассмотрена 09.08.2021, принята 22.09.2021

References

1. Zakharov V. P., Khokhlova M. V. Automatic term extraction and statistical analysis in a special text corpus as a tool for thesaurus construction. *Strukturnaya i prikladnaya lingvistika*, 2012, no.9, pp.222-233 (in Russian).
2. Klyshinskii E. S., Kochetkova N. A., Karpik O. V. *Metod vydeleniya kollokatsii s ispol'zovaniem stepennogo pokazatelya v raspredelenii Tsipfa* [Collocation extraction method using the power-law index in the Zipf distribution]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2018, no.21, pp.220-225.
3. Kochetkova N. A. *Metod izvlecheniya tekhnicheskikh terminov s ispol'zovaniem usovershenstvovannoi mery strannosti* [Method for the extraction of technical terms using an improved measure of strangeness]. *Nauchno-tekhnicheskaya informatsiya. Seriya 2: Informatsionnye protsessy i sistemy*, 2015, no.5, pp.25-32.
4. Kruzhkov M. G. Information resources for contrastive studies: electronic text corpora. *Sistemy i sredstva informatiki*=Systems and Means of Informatics, 2015, vol.25, no.2, pp.140-159 (in Russian).
5. Leichik V. M. *Optimal'naya dlina i struktura termina* [Optimal length and structure of the term]. *Voprosy yazykoznanija*, 1981, no.2, pp.63-73.
6. Novikova A. A. Sketch Engine and Termostat tools for automatic term extraction. *International Journal of Open Information Technologies*, 2020, vol.8, no.11, pp.73-79 (in Russian).
7. Becerro F. B. Phraseological variations in medical-pharmaceutical terminology and its applications for English and German into Spanish translations. *SciMedicine Journal*, 2020, vol.2, no.1, pp.22-29. DOI: 10.28991/SciMedJ-2020-0201-4.
8. Biziukova N. Yu., Tarasova O. A., Rudik A. V. et al. Automatic recognition of chemical entity mentions in texts of scientific publications. *Automatic Documentation and Mathematical Linguistics*, 2020, vol.54, no.6, pp.306-315. DOI: 10.3103/S0005105520060023.
9. Butenko Iu. I., Garazha V. V. BMSTU Corpus of scientific and technical texts: conceptual framework. *Applied Linguistics Research Journal*, 2021, vol.5, no.3, pp.76-81. DOI: 10.14744/alrj.2021.15579.
10. Loukachevitch N., Dobrov B. Ontological resources for representing security domain in information-analytical system. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem*, 2018, no.8, pp.185-191.
11. Sidnyaev N. I., Butenko I. I., Bolotova E. E. Formal grammar theory in recognition methods of unknown objects. *Automatic Documentation and Mathematical Linguistics* 2020, vol.54, no.4, pp.215-225. DOI: 10.3103/S000510552004007X.
12. Simon N. I., Kešelj V. Automatic term extraction in technical domain using part-of-speech and common-word features. *Proceedings of the ACM Symposium on Document Engineering*, 2018, article 51, pp.1-4. DOI: 10.1145/3209280.3229100.
13. Terry A. R., Hoste V., Lefever E. In no uncertain terms: a dataset for monolingual and multilingual automatic term extraction from comparable corpora. *Language Resources and Evaluation*, 2020, no.54, pp.385-418. DOI: 10.1007/s10579-019-09453-9.

About the authors

Iuliia I. Butenko, ORCID 0000-0002-9776-5709, Cand. Sci. (Eng.), Associate Professor, Theoretical Informatics and Computer Technologies Department, Bauman Moscow State Technical University, Moscow, Russia, iubutenko@bmstu.ru
Yurii V. Stroganov, ORCID 0000-0001-5628-7395, Senior Lecture, Computer Software and Information Technology Department, Bauman Moscow State Technical University, Moscow, Russia, stroganovyv@bmstu.ru
Andrei M. Sapozhkov, ORCID 0000-0001-8424-5686, Student, Computer Software and Information Technology Department, Bauman Moscow State Technical University, Moscow, Russia, andreysapozhkov535@gmail.com

Received 04.08.2021, reviewed 09.08.2021, accepted 22.09.2021

Model of creating an economic ecosystem in the framework of economic convergence under the influence of digitalization

S. Kalayda^{1*}

¹ Saint Petersburg State University, St. Petersburg, Russia

* s.kalayda@spbu.ru

Abstract. The article contains the study of the capabilities of choosing such a form of joint business organization by an economic convergence initiator which would ensure the maximum growth of such business under the influence of digitalization and thereby the maximum improvement of the initiator's competitive capability. The author has developed a model of creating an economic ecosystem as a really effective ecosystem which takes into account the influence of both positive and possible negative consequences of the impact of digitalization on a joint business and ensures the maximum economic benefits of the joint business. An algorithm has been developed to implement the model in the framework of a certain level and a specific product of digitalization. The main parameters of the model and the algorithm for the implementation of the same are such terms introduced by the author as an economic ecosystem, potentially the most effective economic ecosystem, a real ecosystem, and a really effective ecosystem, which are described by indicators and costs of the economic effect. Following the steps of the algorithm which implements the model will eventually make it possible to create such a version of an economic ecosystem which would bring the highest economic effect to the initiator on a certain convergence level considered when a certain digitalization product is used in the joint business. Comparison of the versions of real ecosystems obtained for each digitalization level and product makes it possible to select the final version which would produce the maximum economic effect for the initiator with respect to convergence levels and digitalization products and hence the maximum growth of the initiator's competitive capacity. The economic ecosystem formed on the basis of the developed model and the algorithm that implements it gives the initiator of the joint business the greatest advantages in a competitive economy.

Keywords: digitalization, economic information system, ecosystem, economic convergence, informatization option

For citation: Kalayda S. Model of creating an economic ecosystem in the framework of economic convergence under the influence of digitalization. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.28-42. DOI: 10.37791/2687-0649-2021-16-6-28-42

Модель формирования экономической экосистемы в рамках экономической конвергенции под влиянием цифровизации

С. А. Калайда^{1*}

¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

* s.kalayda@spbu.ru

Аннотация. Развитие рыночных отношений под воздействием новых тенденций общественного развития приводит к необходимости трансформации экономики, в том числе к поиску новых эффективных форм организации совместного бизнеса. В статье исследуются возможности выбора инициатором совместной деятельности такой формы организации бизнеса, реализуемого в рамках определенного уровня экономической конвергенции, которая под воздействием цифровизации обеспечивает его наибольший рост и, как следствие, максимальное повышение конкурентоспособности инициатора. Автор разработал модель формирования экономической экосистемы как реально эффективной, учитывающей влияние на совместный бизнес как положительных, так и возможных отрицательных последствий цифровизации, которая обеспечивает максимальный экономический эффект совместного бизнеса. Для реализации этой модели в рамках определенного уровня конвергенции и применения в совместном бизнесе конкретного продукта цифровизации разработан алгоритм. Основными параметрами модели и реализующего ее алгоритма являются такие введенные автором понятия, как «экономическая экосистема», «потенциально самая эффективная экономическая экосистема», «реальная экосистема», «реально эффективная экосистема», описываемые показателями экономического эффекта и затрат по его получению. Следование шагам алгоритма, реализующего модель, позволит сформировать такой вариант экономической экосистемы, который обеспечит инициатору на определенном рассматриваемом уровне конвергенции при использовании в совместном бизнесе конкретного продукта цифровизации самый большой экономический эффект. Сравнение полученных по каждому уровню и продукту цифровизации вариантов реально эффективных экосистем дает возможность отбора окончательного варианта экосистемы, обеспечивающего инициатору экономический эффект, максимальный по уровням конвергенции и продуктам цифровизации, и, как следствие, наибольший рост его конкурентоспособности. Сформированная на основе разработанной модели и реализующего ее алгоритма экономическая экосистема дает в условиях конкурентной экономики инициатору создания совместного бизнеса наибольшие преимущества.

Ключевые слова: цифровизация, экономическая информационная система, экосистема, экономическая конвергенция, вариант информатизации

Для цитирования: Калайда С. А. Model of creating an economic ecosystem in the framework of economic convergence under the influence of digitalization // Прикладная информатика. 2021. Т. 16. № 6. С. 28–42. DOI: 10.37791/2687-0649-2021-16-6-28-42

Introduction

The term "ecosystem" was first used in 1935 by British ecologist Arthur Tansley with regard to nature [1]. An ecosystem was and

still is understood as the interrelation of all elements of nature: living beings, natural forces, etc., but such interrelation that manifests not only in the common coexistence but also in the competition of the natural elements. This is why

the application of this term to economy (an economic ecosystem) only becomes possible if the economic ecosystem also is based on the principles of the cooperation and competition of its participants. It was the consideration of these main parameters only which are specific to different types of ecosystems, e.g. biologic and economic ones, that resulted in the fact that an ecosystem is all too often understood in economic literature as a system of interaction of companies – service providers, regulators, and consumers, which supposes both the competence and cooperation for the purpose of providing any given service to the user [2]. Thus, in the end of the 20th century J. F. Moore developed the concept of an entrepreneurial ecosystem. It includes companies, suppliers, consumers, and even competitors the interaction wherewith helps achieve major results [3].

In the context of the impact of digitalization on economic convergence processes, one should note the following: with such approach used, any form of implementation of a joint business in the framework of economic convergence may be considered as an ecosystem characterized by compliance with the principles competition and cooperation principles. Indeed, economic convergence is a mechanism for improvement of competitive capacity of the initiator of creation of a joint business based on convergence of different economic entities' activities. But the starting point of using such a mechanism is the competition and the resulting desire of the convergence initiator to improve its competitive capacity through creation of joint business. A joint business created in the framework of any level of convergence is implemented on the basis of its participants' cooperation established in any manner whatsoever. It does not matter how it was created, whether voluntarily or not. Whatever the case, such cooperation does exist within a joint business.

Economic convergence processes are very strongly influenced by digitalization, a new trend of social development. It is based on digital transformation of information, which more often than not promotes a sharp increase in the

efficiency of the development of economy and society in general. The impact of digitalization on a joint business in the framework of economic convergence processes may differ. It may improve the efficiency of a joint business and thereby improve the competitive capacity of the convergence initiator, or may not promote such improvement.

The objective of a joint business initiator in the framework of any level of economic convergence under the influence of digitalization on such business through any given digitalization product is to select the most effective options of such business organization from among the possible ones, taking into account both the positive and the negative factors of digitalization. This objective is addressed in the article.

The present study is novel in that it makes clear the term "economic ecosystem" and introduces other definitions of the same as applied to economy under the influence of digitalization, the modern trend in social development. The practical significance of the results of the study is the development of a model and an algorithm to implement the same (sequence of certain actions) for choosing the best form of joint business organization in the framework of economic convergence driven by digitalization.

Currently, the development and creation of new business models is given sufficient attention in the scientific literature [4–6]. However, the matters of creation and use (existence) of different forms of joint business are given priority, especially taking into account the existing globalization and integration factors which have become apparent over the last years [3, 7–9]. Such a new form of doing business as an ecosystem has become especially popular. Nevertheless, there is still no unified and generally recognized approach to the definition of this form of joint business organization [10–15]. This determines the relevance of this study. The methodological framework of the study was the works of national and foreign scholars, system and contingency approaches, con-

tent analysis, grouping, comparative and other methods of analysis.

The objective of a joint business initiator in the framework of any level of economic convergence under the influence of digitalization through any given product thereof is to select the most effective options of such joint business organization from among the possible ones.

Earlier it was shown that the highest performance of joint business is achieved when positive impact of digitalization on economic convergence processes ensures [16]:

- existence of the prerequisites for a certain level of convergence;
- management of factors which promote an increase in joint business performance;
- implementation of the capabilities of positive impact of digitalization by the joint business.

This particular result made it possible to formulate the following general idea of creation of a really effective economic ecosystem in which the economic convergence initiator is interested.

General idea of creating a really effective economic ecosystem

For a joint business creation initiator, the matter of the joint business performance is the most important one. For this very reason, the following terms are introduced in this study: *economic ecosystem*, *potentially most effective ecosystem*, *real ecosystem*, and *really effective ecosystem*. As already mentioned, currently the point of view is very popular that any major joint business implemented under the influence of digitalization is presented as an economic ecosystem. To our opinion, this approach requires a closer definition.

First of all, the initiator of any joint business is interested in the economic effect which the initiator can gain from the joint business. However, for an actual joint business to be effective to the greatest possible extent it is necessary to obtain the maximum economic effect from the implementation of all advantages of

digitalization in the joint business. To single out this situation, it is proposed to designate the respective form of doing joint business as an economic ecosystem.

An *economic ecosystem* will be understood as an institutional and organizational form of doing joint business, which is really the most effective one in the framework of the convergence level considered and the certain digitalization product used in such joint business. The maximum effect of a joint business is ensured by fulfillment of the following conditions: all possible positive consequences of the impact of digitalization on all aspects of the business – the prerequisites, the factors, and the very organization of the joint business shall be used, and the costs of reduction and/or mitigation of possible negative consequences of the impact of digitalization on the joint business shall be minimal.

To determine the maximum possible economic effect the initiator is interested in, which is conditioned by the implementation of all positive consequences of digitalization, the term “potentially most effective ecosystem” is introduced. The *potentially most effective economic ecosystem* is, in a certain sense, a conventional ecosystem which only takes into account the positive impact of digitalization on the joint business with respect to all possible areas of such impact – on the prerequisites, the factors, and the very organization of the joint business. The conventionality of this term is determined by the fact that in real life digitalization has not only a positive but also a negative impact on a joint business. Though such an ecosystem gets in line with the highest effectiveness of doing joint business, which the initiator is interested in, it does not take into account the negative consequences of digitalization. Therefore, it cannot be considered to be a real ecosystem, as, apart from the positive consequences, digitalization may be followed by negative consequences for a joint business which involve their elimination or mitigation costs. The concept of a potentially most effective economic ecosystem is only used for the initiator to un-

derstand the maximum effect which it could have obtained from the creation of an ecosystem if digitalization had only positive impact on joint business, i.e. if it were followed by positive consequences only with respect to all possible areas of the joint business (prerequisites, factors, and organization of business).

Taking into account the possibility of occurrence of negative consequences of digitalization, it is necessary to consider real ecosystems which suppose realization of all possible positive consequences of the impact of digitalization on the joint business but which also take into account the costs of reduction or mitigation of possible negative consequences. The initiator is only interested in those real ecosystems where the economic effect of the joint business exceeds the costs related to the reduction or mitigation of possible negative consequences.

A *real ecosystem* is understood as an ecosystem which supposes realization of all possible positive consequences of digitalization and takes into account the costs determined by the developed possible negative consequences reduction and mitigation management program.

Managerial decisions which form the respective management program may be most diverse and therefore may be characterized by various cost amounts. Each version of a possible negative consequences reduction and mitigation management program has its certain costs. As the performance of a real ecosystem being created is determined by the difference between the economic effect resulting from the impact of all positive consequences of digitalization on joint business and the costs of elimination of negative consequences as provided for in a certain management program, a certain version of a real ecosystem will correspond to each version of the management program. The effect from the creation of such a real ecosystem is determined by the aforementioned difference.

Taking into account the aforesaid, the variety of versions of negative consequences management programs is determined by the variety of versions of a real ecosystem the per-

formance of which is differentiated depending on the amount of negative consequences management costs.

It is natural that the convergence initiator is interested in the maximum economic effect. Therefore, it selects the really *effective ecosystem* from a variety of real ecosystems – the one which would produce such maximum effect for the initiator.

A *really effective ecosystem* is understood as a real ecosystem which supposes the positive impact of digitalization on all aspects of joint business (prerequisites, factors, and organization of joint business) and such management of negative consequences which involve minimum costs. It is the really effective ecosystem that the convergence initiator is interested in since it ensures the maximum possible competitive capacity for the initiator.

Implementation of the approach to creating a really effective economic ecosystem

Let us consider the matters of creation of a really effective economic ecosystem in the framework of a separate level of economic convergence in the situation when digitalization has an impact on a joint business through a certain digital product.

Since generally digitalization has both positive and possible negative impact on a joint business, the effect of such joint business may be calculated taking into account the following two situations:

- only all positive results of impact of digitalization on the joint business are taken into account with respect to all areas of such business; the results of impact of digitalization on the prerequisites, factors, and organization of the joint business itself are taken into account (option 1);
- not only all positive results of impact of digitalization on the joint business are taken into account, but also the possible negative consequences of the same (option 2).

Option 1 of accounting of possible consequences of impact of digitalization on joint business defines an ES economic ecosystem as a *potentially most effective ecosystem* EES^+ which only takes into account the influence of all possible positive consequences of the impact of digitalization on the joint business with respect to all aspects thereof, while option 2 defines it as a *real economic system* ES^+ which takes into account the simultaneous influence of all possible consequences of the impact of digitalization on the joint business with respect to all aspects thereof (table 1).

Table 1. Parameters for determination of a really effective economic ecosystem created in the framework of a certain convergence level by utilizing a certain digital product

Type of ecosystem	Potentially most effective ecosystem EES^+	Real ecosystem ES^+	Really effective ecosystem RES^+
Economic effect of creation of an ecosystem	EES^+	EES^+	$ERES^+$
General formula of economic effect calculation	Determined by the total economic effect caused by implementation in the joint business of all possible positive consequences of the impact of digitalization thereon	$EES^+ = EES^+ - ESC^+$	$ERES^+ = \max_{l=1,...,L} \{EES^+_l\}$
Calculation of economic effect according to the 1st cost option	—	$EES^+_1 = EES^+ - ESC^+_1$	—

The EES^+ economic effect of the ES^+ potentially most effective ecosystem considered without regard to the possible negative impact of digitalization will be determined by the cu-

mulative economic effect from the realization of all possible positive consequences of the impact of digitalization on the joint business.

The EES^+ economic effect of an ES^+ real ecosystem considered taking into account both the possible positive and negative impact of digitalization will be calculated as

$$EES^+ = EES^+ - EES^-, \quad (1)$$

where ESC^- denotes the costs (losses) due to the reduction or mitigation of possible negative consequences of the impact of digitalization on the joint business.

Costs of reduction of even an individual possible negative consequence of digitalization may differ. They are determined by the managerial decision(s) which they correspond to. For example, in an ecosystem implemented in the framework of inter-segment economic convergence when digitalization only influences such a core product as information technologies, it is decided to acquire an information technology which would improve the effectiveness of the customer database. However, the size of the customer database itself is so small at the time of joint business creation and for the foreseeable 3-year period that the acquisition of the aforesaid technology at the moment (managerial decision 1) would result in net costs. However, another managerial decision is possible (managerial decision 2) – refuse from acquisition of such technology for a 3-year period. Naturally, the real yearly effect of the ecosystem being created will be higher in the second case: the information technology acquisition costs will be zero.

Since different options of managerial decisions MD^l ($l=1,...,L$) aimed at reduction or mitigation of possible negative consequences of impact of digitalization on business may involve different implementation costs CES^l , ($l=1,...,L$), the set of real ecosystems $\{ES^+_l, l=1,...,L\}$ for an individual economic convergence level considered and a certain digitalization product will be determined by L – the total number of managerial decisions and by implication by the total number of cost options ($l=1,...,L$).

The value of economic effect for each 1st ($l=1, \dots, L$) EES^+_{-l} real ecosystem will be determined

$$EES^+_{-l} = EES^+ - ESC^l_{-l}, \quad (2)$$

where EES^+ is the economic effect produced by the potentially most effective ecosystem ES^+ of a certain convergence level and with a certain digitalization product, considered without regard to the possible negative impact of digitalization, ESC^l_{-l} is the costs related to the 1st managerial decision aimed at reduction or mitigation of possible negative consequences.

Real ecosystems created in the framework of an individual level of economic convergence under the influence of a certain digital product on joint business are used for searching the most effective one, i. e. for searching a RES^+ economic ecosystem which is really effective for such convergence level and digitalization product.

A really effective economic ecosystem RES^+ with a certain convergence level for a certain digitalization product is the real ecosystem which would have the maximum economic effect:

$$ERES^+ = \max \{EES^+_{-l}, l=1, \dots, L\}, \quad (3)$$

where $ERES^+$ is the economic effect produced by the really effective ecosystem.

Table 2 shows an example of identification of a really effective ecosystem.

In the above example, a really effective ecosystem is the ecosystem where the 2nd option of management of possible negative consequences is used – the one which involve minimum costs and therefore produces the maximum economic effect.

When considering costs related to reduction or mitigation of possible negative consequences one should also take into account the following.

They may include e. g. costs directly related to acquisition of digitalization products in the joint business. They may include the cost of the acquired new information technology and the hardware used for creation of a platform, etc. However, they also may include indirect costs involved in the digitalization processes in the joint business. For example, retirement of old equipment due to implementation of a new digital information complex also involves additional costs for the initiator.

It should be noted that the possible negative impact of digitalization on the economic convergence processes can be described by the risk of its negative impact due to the use of a given basic digitalization product associated with the digitalization development stages. The risk is a function of two variables – the probability of the risk event and the extent of possible damages.

Table 2. Example of calculation of an economic effect meeting the criteria of a really effective ecosystem

EES^+ economic effect provided by a potentially most effective ecosystem ES^+ of a certain convergence level for a certain digitalization product (monetary units)	L – the number of options of management of reduction and/or mitigation of possible negative consequences of digitalization impact on joint business	MD^l_{-l} – the 1st option of managerial decision aimed at reduction and/or mitigation of possible negative consequences of digitalization impact on joint business	CES^l_{-l} – costs of the 1st managerial decision aimed at reduction and/or mitigation of possible negative consequences of digitalization impact on joint business (monetary units)	EES^+_{-l} – economic effect for each 1st ($l=1, \dots, L$) real ecosystem (monetary units) (calculation formula (2))	$ERES^+$ – economic effect meeting the criteria of a really effective ecosystem (monetary units) (calculation formula (3))
100	1	1	20	$100 - 20 = 80$	$\max \{80, 85, 70\} = 85$
	2	2	15	$100 - 15 = 85$	
	3	3	30	$100 - 30 = 70$	

The necessity of assessing the real economic effect for each 1st version of ecosystem conditioned upon the 1st managerial decision on the reduction of possible negative consequences supposes:

- identification of its potential opportunities for the growth of efficiency which depend on the maximum compliance with all requirements for the joint business under the conditions of the given level of economic convergence and the use of a specific digital product, including requirements for the prerequisites, the factors, and business organization itself, the fulfillment of which promotes improvement of the joint business performance;
- identification of possible negative consequences which may have an adverse impact on the joint business performance;
- development of managerial decisions aimed at reduction of possible negative consequences;
- assessment of the possible actual result of the joint business, i. e. comparison of improvement of the joint business performance conditioned upon compliance with all requirements placed thereon to ensure improvement of performance, subject to reduction in performance due to possible negative consequences.

When determining any 1st version of a real ecosystem ($l=1,...,L$), the potentially most effective ecosystem ES^+ which supposes the use of all possible positive consequences of digitalization impact only, and the economic effect EES^+ corresponding to such ecosystem are the same for the specific convergence level and the certain digital product used in the joint business.

The content of the potentially most effective economic ecosystem and the value of the respective economic effect only vary for different levels of convergence and different digital products.

Table 3 shows the matrix of possible versions of potentially most effective ecosystems considered when organizing a joint business in the framework of economic convergence under the positive impact of a given digitalization product.

Table 3. Matrix of possible versions of potentially most effective economic ecosystems

Level of convergence \ Digitalization product	Intrasegment	Intersegment	Intersectoral
IT (1st digitalization stage product)	$ES_1^{+IntraS}$	ES_1^{+Isg}	ES_1^{+Isc}
IT and IT platforms (2nd digitalization stage product)	$ES_2^{+IntraS}$	ES_2^{+Isg}	ES_2^{+Isc}
IT, IT platforms, and IT networks (3rd digitalization stage product)	$ES_3^{+IntraS}$	ES_3^{+Isg}	ES_3^{+Isc}

In the notations: lower index is the digitalization product number, the upper index added to "+" is the convergence level index.

Each level of economic convergence and a certain digital product has its own value of the potentially most effective ecosystem. These values will be used to identify a real economic ecosystem in the framework of the respective economic convergence level and the digitalization product used, and to select the best option of joint business implementation among all possible combinations of an economic convergence level and a digital product.

However, one should keep in mind the following regarding the potentially most effective ecosystem.

The condition of application of all possible positive consequences of the impact of digitalization on all areas of joint business which is used in the definition of such an ecosystem requires certain detailing. When determining the economic effect of a potentially most effective ecosystem, all the positive consequences known as of the moment of taking the decision to create a joint business shall be taken into account. Nevertheless, generally such capabilities increase in the course of time either through further utilization of the known areas of improvement of the joint business performance or through occurrence of new additional capabilities of the digital product itself which is used in the ecosystem. An example of the first situation is the involvement of new partners in the joint business for the ecosystem already existing in the framework of in-

tersectoral convergence when the most complex digital product is used: IT, IT platforms and networks combining the same. As of the time of creation of the ecosystem this prerequisite had been realized, but the implementation of this prerequisite (involvement of new partners) may continue for a long period of time in view of the unlimited capabilities of such a digital product. An example of further development of an existing ecosystem through occurrence of new additional capabilities of the digital product itself can be the use of new information products (technologies) created after the creation of the ecosystem, which can improve the performance of the joint business in the framework of the existing ecosystem.

The aforesaid possibility of transformation of the extent of potential economic effect due to a change of the whole list of possible positive consequences over time does not deflate the relevance of the use of the concept of a potentially most effective economic ecosystem and the respective economic effect. This is due to the fact that the indicator the value whereof is used by the initiator when accessing the actual performance of the joint business to be created is the economic effect calculated as of a specific point of time and identified with consideration to the maximum possible utilization of the positive consequence of the impact of digitalization on the joint business and the minimum amount of costs related to the reduction and mitigation of possible negative consequences at the very point of time.

It should be noted that every level of economic convergence (intrasegment, intersegment and intersectoral ones) may suppose the use of one of the core digital products (IT, IT platforms, and networks which combine the same) in a joint business. Therefore, the maximum number of options of construction of really effective ecosystems is limited by 9. However, this does not mean that the initiator will necessarily calculate all the options of building really effective ecosystems, which may be due to the fact that the initiator takes into account other additional factors, e.g. the funds available for the creation of ecosystems.

The combined values of the two parameters which describe the features of a joint business in the framework of a certain level of economic convergence by using a specific digitalization product determine the following classification of ecosystems (table 4).

Table 4. Classification of economic ecosystems

Convergence level – the first parameter of the classification criterion	Intrasegment	Intersegment	Intersectoral
Digitalization product – the second parameter of the classification criterion			
IT (1st digitalization stage product)	RES_1^{IntraS}	RES_1^{Isg}	RES_1^{Isc}
IT and IT platforms (2nd digitalization stage product)	RES_2^{IntraS}	RES_2^{Isg}	RES_2^{Isc}
IT, IT platforms, and IT networks (3rd digitalization stage product)	RES_3^{IntraS}	RES_3^{Isg}	RES_3^{Isc}

Influence of the convergence level and the digitalization product on the joint business performance

Below please find the analysis of the impact of an increased complexity of a digital product and the convergence level on the improvement of an ecosystem's performance.

It should be noted that more complicated digitalization products are more capable of having a positive impact of digitalization on economic convergence processes of any level [17]. This gives grounds to think that the most positive impact on business implemented in the framework of a certain level of economic convergence with respect to the criterion "digitalization product which influences the joint business performance to the greatest extent" is exerted by the most complicated digitalization product: "IT, IT platforms, and networks which combine the same".

Now, when passing from one economic convergence level to another with the same digitalization product used in a joint business, the capabilities of improvement of a joint business performance are determined by the values of the convergence parameters (prerequisites, fac-

tors, and organization of the business) which may change and therefore influence the joint business performance in different ways. For this reason, it cannot expressly be said that the potential performance of ecosystems increase with transfer from one convergence level to another (from intra-segment one to the intersegment and then to the intersectoral one).

In view of the fact that joint business where economic entities from all kinds of economic sectors become the participants of the same and the most complicated digitalization products are used – IT, IT platforms, and networks combining the same, we will briefly describe the features of a joint business implemented in the framework of intersectoral convergence when such a complicated digitalization product is used therein (version ES_3^{+lsc} shown in table 3).

As it was shown above, such a complicated digitalization product generates new prerequisites for a joint business – abandoning the aggressive form of involving partners in the joint business, appearance of representatives of all kinds of segments and sectors of economy, and utilization of all capabilities of the most complicated digital product in the business [18]. The aforesaid prerequisites cover the maximum possible number of positive consequences of the impact of digitalization on a joint business through the above mentioned most complicated digitalization product. That is why the implementation of such prerequisites will secure the potentially most effective ecosystem of the intersectoral convergence level (without regard to the costs) which utilizes the most complicated digitalization product. As compared with potentially most effective ecosystems of other levels and other digitalization products, it guarantees the maximum economic effect.

This being said, the highest performance of such an ecosystem which is only described by the potential economic effect without regard to the costs of such effect is subject to fulfillment of all the following conditions:

- under the influence of an electronic database which includes IT platforms and networks combining the same all positive consequences

of realization of prerequisites and factors of doing the joint business shall be implemented;

- the electronic database shall be used with respect to all areas and aspects of organization of the joint business activities which makes it possible to improve the joint business performance not only through the improvement of the entrepreneurial activity business process itself which is conducted in the framework of intersectoral economic convergence, but also through:

- voluntary involvement of economic entities from all kinds of fields of activity whether technologically related or not to the activity of the ecosystem creation initiator;

- refusal from the aggressive form of involving other participants in the ecosystem;

- implementation of the "win-win" principle with respect to all the ecosystem participants which ensures the congruence of their interests;

- creation of modified, converged, and new products of all participants of the joint business, offered to the joint business customers;

- modification of the convergence initiator's customer-oriented strategy;

- provision of access to all products and services of all the participants of the ecosystem to the customer;

- a joint business implemented in the framework of such an ecosystem shall cover all kinds of segments of social life: infrastructure, communications, marketing, advertising, finance, trade, media, entertainment, government, society, cyber security, startups, investments, education, and human resources¹. When this point is taken into account, some authors use the term "universal ecosystem".

Fulfillment of the aforesaid conditions will mean the use of all sources and factors of perfor-

¹ It should be noted that the fact of compliance with this requirement – coverage of substantially all aspects of social life by the offers of an ecosystem reflects in the practical application of the term "universal ecosystem". When only a certain field of activity is concerned which is related to some niche of social life, the term "niche ecosystem" is used in practice.

mance improvement in a joint business which will result in an improvement of the conversion initiator's competitive capacity [17].

As previously stated, a potentially most effective ecosystem is the one which only takes into account the positive impact of digitalization on all aspects of the joint business. Nevertheless, the possibility of occurrence of negative consequences of the impact of all digitalization stages on the joint business in case of influence of digitalization through a certain product requires comparison of the expected positive and negative results, i. e. identification of the actual economic effect of the creation and operation of the ecosystem within any convergence level.

Model and algorithm of creating a really effective ecosystem

The process of creation of a really effective ecosystem as a result of comparison of all possible real ecosystems of a certain convergence level when a certain digitalization product is used in the joint business may be represented as a model the features of which as determined by implementing the aforesaid approach include the following:

- the ecosystem supposes the use of all possible positive consequences of the impact of digitalization on joint business. Conventionally such an ecosystem where only all possible positive consequences of impact of digitalization on joint business are taken into account can be considered as a *potentially most effective ecosystem* ES^+ . The conventionality of the use of such term is explained by the fact that such ecosystems do not exist in real life: digitalization influences joint business not only favorably but also adversely, which results in a reduced actual economic effect compared with the economic effect corresponding to the potentially most effective ecosystem;
- the model takes into account the influence of both all possible positive and negative consequences of the impact of digitalization on the business through a certain digitalization prod-

uct. Taking this feature of the model into account reflects in the use of the term *real economic ecosystem* $ЭС^+$ which accounts for the simultaneous influence of all possible positive and negative consequences of the impact of digitalization on the joint business;

- reduction and/or mitigation of possible negative consequences of the impact of digitalization is ensured by the development and implementation of certain managerial decisions which involve specific costs. Options of managerial decisions even with respect to the same possible negative consequence may differ, which results in the difference of the related costs;

- each version of a real ecosystem is based upon the common, identical for all versions, potentially most effective economic ecosystem which supposes the use of all positive consequences in joint business. However, the variety of the cost options resulting from different managerial decisions determines a whole set of real economic ecosystems;

- the ultimate purpose of the convergence initiator – to improve the initiator's competitive capacity to the maximum extent – determines the expediency of analysis of all versions of a real ecosystems and selection of the best one. The best version of a real ecosystem will be a really effective economic ecosystem RES^+ which produces the maximum actual economic effect and takes into account both all possible positive consequences of digitalization impact on all areas of the joint business and the losses related to reduction and/or mitigation of possible negative consequences of the same. It is this version of an ecosystem that the initiator is interested in because it would ensure a real growth of economic effect for the initiator and thereby the maximum growth of the initiator's competitive capacity.

To implement the model of creating a really effective ecosystem of a certain level for a specific digitalization product, the author has proposed and developed an algorithm. Figure 1 shows the scheme of implementation of the al-

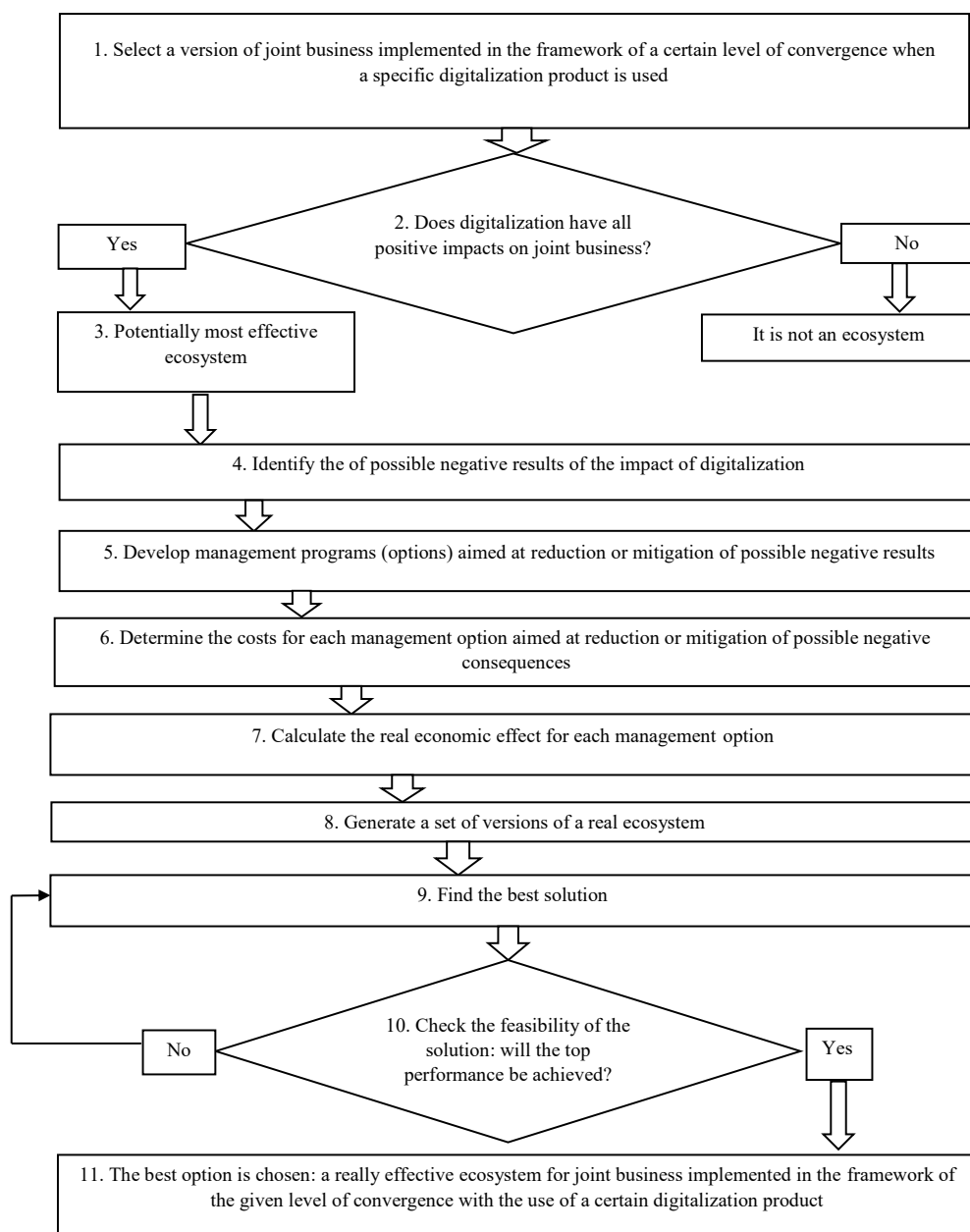


Fig. 1. Algorithm for creating a really effective ecosystem

gorithm of creating a really effective ecosystem to be implemented within a certain level of economic convergence when a specific digitalization product is used in a joint business. A really effective ecosystem built according to this algorithm produces the maximum economic effect of the joint business creation for the initiator, but only in the framework of the given conver-

gence level and the certain digitalization product used in the joint business.

A specific feature of the proposed algorithm is that it implements a model of creating a really effective ecosystem which produces the maximum economic effect determined by the simultaneous allowance for the impact of digitalization on the joint business both through positive

and possible negative consequences. All steps of the algorithm are related to obtaining and comparing the values of economic effect corresponding to the given ecosystem type (potentially most effective ecosystem, real ecosystem, and really effective ecosystem) and to the costs resulting from obtaining of such effect.

The algorithm is used for finding a really effective ecosystem in the framework of a specific digitalization level only with a certain digitalization product used in the joint business. Application of the algorithm for all 9 possible versions of generation of really effective economic ecosystems conditioned upon the combination of an economic convergence level (3 levels exist) and a digitalization product (also 3) makes it possible to select the version of the joint value of a specific convergence level and the certain digitalization product for which the really effective economic ecosystem being generated would produce maximum value of economic effect.

Conclusion

The impact of digitalization on all economic processes, including economic convergence ones, is so substantial that the problem of finding the version of effective influence of the same becomes extremely important.

The article proposes the author's model of creating an economic ecosystem as a really most effective form of doing joint business conducted in the framework of a certain level of convergence with a specific digitalization product used therein, which takes into account both the positive and the possible negative impacts of digitalization on the business.

The model is represented by such parameters introduced by the author as an economic ecosystem, a potentially most effective ecosystem, a real ecosystem, and a really effective ecosystem. Each of the aforesaid parameters is described by indicators of economic effect and costs of obtaining the same.

An *economic ecosystem* as a *really effective ecosystem* is an institutional and organizational

form of doing joint business, implemented in the framework of a certain convergence level under the influence of digitalization on the joint business through a specific digitalization product, which firstly supposes the use of all positive opportunities of digitalization when it has an impact on all the areas of a joint business – the prerequisites, the factors, and the very organization of the business, and secondly ensures the maximum economic effect of joint business creation, taking into account the possible negative consequences of the impact of digitalization on the business.

To show the maximum possible economic effect of creation of an ecosystem, the concept of a potentially most effective ecosystem was introduced – the one which only takes into account the positive impact of digitalization on joint business and thereby produces the maximum economic effect. However, such forms of business do not exist in real life, as the impact of digitalization always involve not only a beneficial, but also an adverse effect. With that said, the introduced concept of a *real ecosystem* supposes that not only all possible positive consequences of digitalization are taken into account, but also the costs determined by the developed program of management of reduction and/or mitigation of the identified possible negative consequences of the same. The version of a real ecosystem which gives the maximum economic effect is a *really effective ecosystem*. It is the one which ensures improvement of the joint business performance for the initiator and achievement of the initiator's objectives related to the improvement of competitive capacity.

To assess different versions of a real ecosystem and select the best one as a really effective economic ecosystem, it is proposed to use the economic effect indicator.

To implement the model of creating a really effective ecosystem in the framework of a certain level and a specific product of digitalization, an algorithm has been developed the use of which with respect to all convergence levels and different digitalization products makes it possible to select the most effective version of an ecosystem.

In addition, the article contains an analysis of the issue of the influence of the economic convergence level on the effectiveness of a joint business in the framework of which such business is conducted; such an influence may differ, as well as the influence of the digitalization product used in such business: a more complicated digitalization product has more capabilities of a positive impact on the business.

The presented model of creating an economic ecosystem ensures achievement of the joint busi-

ness creation initiator's purpose: the improvement of the joint business performance and the competitive capacity of the initiator.

The developed algorithm which implements the model makes it possible to objectively assess the capabilities of the development of a joint business in the form of an ecosystem as the most effective form of doing the same, which takes into account both the maximum positive and the possible negative impact of digitalization on such joint business.

References

1. *Biznes ekosistema brenda* [Brand business ecosystem]. Murat Turgunov. Available at: <https://turgunov.ru/business-ecosystem-brand/> (accessed 08.11.2021).
2. *Ekosistema Sberbank* [Sberbank Ecosystem]. TAdviser. Available at: https://www.tadviser.ru/index.php/Статья:Экосистема_Сбербанка (accessed 08.11.2021).
3. Moore J. F. The Death of competition: Leadership and strategy in the age of business ecosystems. N.Y., Harper Business, 1996, 320 p.
4. Orehova S. V. Formation of sustainable competitive advantages of firm in context of resource conception. *Sovremennaya konkurentsia*=Journal of Modern Competition, 2011, no.4(28), pp.97-105 (in Russian).
5. Chernukhina G. N. Transformation and consumption trends in the modern sharing business model. *Sovremennaya konkurentsia*=Journal of Modern Competition, 2020, vol.14, no.3(79), pp.79-88 (in Russian). DOI: 10.37791/1993-7598-2020-14-3-79-8.
6. Fomchenkova L., Fomchenkov V. Cognitive modelling application to business model choice justify of an economic organization. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.4(94), pp.6-20 (in Russian). DOI: 10.37791/2687-0649-2021-16-4-6-20.
7. Chernova G. V., Pisarenko Zh. V., Kuznetsova N. P. Factors, preconditions and parameters of financial convergence. *Strahovoe delo*, 2017, no.3(288), pp.3-14 (in Russian).
8. Dowrick St., DeLong J. Bradford. Globalization in Historical Perspective. USA, University of Chicago Press, 2003. Available at: <https://www.nber.org/books-and-chapters/globalization-historical-perspective/globalization-and-convergence> (accessed 08.11.2021).
9. Kuznetsova N. P., Chernova G. V. *Konvergentsiya v sfere finansovykh uslug* [Convergence in the field of financial services]. *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika*=Saint Petersburg University Journal of Economic Studies, 2001, no.4, pp.129-136 (in Russian).
10. Gaiisina D. V. *Transformatsiya sovremennykh biznes-modelei v storonu ekosistem: doklad* [Transformation of modern business models towards ecosystems: report]. *VI ezhegodnaya konferentsiya «Proektirovanie biznes-arkhitektury 2017»* [Sixth annual conference "Engineering of business architectures 2017"]. Available at: <https://www.businessstudio.ru/upload/iblock/7e6/%D0%93%D0%B0%D0%B9%D1%81%D0%B8%D0%BD%D0%B0.pdf> (accessed 08.11.2021).
11. Samorodova L. L., Shutko L. G., Yakunina Yu. S. Digital ecosystems and economic complexity of the region as factors of innovative development. *Voprosy innovatsionnoi ekonomiki*=Russian Journal of Innovation Economics, 2019, vol.9, no.2, pp.401-410 (in Russian). DOI: 10.18334/vinec.9.2.40607.
12. Kleiner G. B., Rybachuk M. A., Karpinskaya V. A. Development of ecosystems in the financial sector of Russia. *Upravlenets*=The Manager, 2020, vol.11, no.4, pp. 2-15 (in Russian). DOI: 10.29141/2218-5003-2020-11-4-1.
13. Pisarenko Zh. V. The creation of ecosystem financial conglomerates: case of China. *Strahovoe delo*, 2020, no.5(326), pp.3-13 (in Russian).
14. Voss J. The Financial Ecosystem. The Sarasota Institute, 2017. Available at: <https://sarasotainstitute.global/the-financial-ecosystem> (accessed 08.11.2021).
15. Gudkova T. V. Economic evolution of a company: from the classic concept to a digital ecosystem. *Ekonomicheskoe vrozozhdenie Rossii*, 2019, no.4(62), pp.74-84 (in Russian).

16. Kalayda S. A., Khalin V. G., Chernova G. V. Risks of the negative digitalization impact on the processes of economic convergence. *Strahovoe delo*, 2021, no.9(342), pp.53-64 (in Russian).
17. Kalayda S. A. *Ekosistema kak effektivnaya institutsional'no-organizatsionnaya forma mezhsektornoj finansovoi konvergentsii* [Ecosystem as an effective institutional and organizational form of intersectoral financial convergence]. *VII regional'naya nauchno-prakticheskaya konferentsiya «Aktual'nye voprosy finansov i strakhovaniya Rossii na sovremennom etape»* [VII Regional Scientific and Practical Conference "Topical Issues of Finance and Insurance in Russia at the Present Stage"], 2020, pp.187-191.
18. Chernova G. V., Khalin V. G., Kalayda S. A. Factors and prerequisites of modern economic convergence. *Ekonomika i predprinimatel'stvo*=Journal of Economy and Entrepreneurship, 2020, no.6(119), pp.31-36 (in Russian). DOI: 10.34925/EIP.2020.119.6.003.

About the author

Svetlana A. Kalayda, ORCID 0000-0002-5688-3178, Cand. Sci. (Econ.), Associate Professor, Risk Management and Insurance Department, Saint Petersburg State University, St. Petersburg, Russia, s.kalayda@spbu.ru

Received 10.11.2021, reviewed 15.11.2021, accepted 01.12.2021

Сведения об авторе

Калайда Светлана Александровна, ORCID 0000-0002-5688-3178, канд. экон. наук, доцент, кафедра управления рисками и страхования, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, s.kalayda@spbu.ru

Статья поступила 10.11.2021, рассмотрена 15.11.2021, принята 01.12.2021

Genetic algorithm as a tool for modeling calculations of electric power systems

R. Solopov^{1*}, A. Samulchenkov¹, V. Ziryukin²

¹ The Branch of National Research University "MPEI" in Smolensk, Smolensk, Russia

² Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

* solopov.rv@mail.ru

Abstract. Evolutionary modeling is one of the areas of artificial intelligence, which essence is the computational processes interpretation and the final forms of integral computational algorithms construction from their existence, variability and development the points of view in natural systems. All evolutionary modeling methods are of an optimization nature due to the basic use of the theory of natural selection principles. One of the most common methods of evolutionary modeling is the genetic algorithm (GA). It is the method of adaptive search for solutions based on the principles of the evolution and the natural selection with the preservation of biological terminology in a simplified form theories. Its essence is to determine the most fit individual (solution) by the value of its fitness function during evolution, considering the analysis of the heredity influences and the external environment. Despite the biological terminology, genetical algorithms are a universal computational tool that can be used to solve a wide range of complex problems, including the electric power industry. The authors considered the issue of the genetic algorithm use in the framework of calculating the steady state of the electrical network (SS EN), since the mathematical electrical network model is a system of high-order nonlinear equations, where all the restrictions imposed by the physical properties of the object under consideration are taken into account. Its solution is a rather laborious optimization problem, due to the operating electrical networks complexity. The correct solution of this system is the most critical stage in the calculation of the SS EN. It is the reason for importance and urgency of the search for SS EN calculating optimal methods task. This paper presents the development results of an analytical apparatus that made it possible to search for a solution to the problem of calculating electrical networks steady-state modes using the genetic algorithm based on special software.

Keywords: evolutionary modeling, genetic algorithm, fitness function, electrical networks, mode calculation

For citation: Solopov R., Samulchenkov A., Ziryukin V. Genetic algorithm as a tool for modeling calculations of electric power systems. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.43-53. DOI: 10.37791/2687-0649-2021-16-6-43-53

Генетический алгоритм как инструмент моделирования расчетов электроэнергетических систем

Р. В. Солопов^{1}, А. С. Самульченков¹, В. И. Зирюкин²*

¹ Филиал ФГБОУ ВО «Национальный исследовательский университет "МЭИ"» в г. Смоленске,
Смоленск, Россия

² Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Россия

* *solopov.rv@mail.ru*

Аннотация. Эволюционное моделирование – это одно из направлений искусственного интеллекта, сущностью которого является интерпретация вычислительных процессов и построение конечных форм целостных вычислительных алгоритмов с точки зрения их существования, изменчивости и развития в природных системах. По причине использования в своей основе принципов теории естественного отбора все методы эволюционного моделирования носят оптимизационный характер. Одним из наиболее распространенных методов эволюционного моделирования является генетический алгоритм (ГА) – метод адаптивного поиска решений, основанный на принципах теории эволюции и теории естественного отбора с сохранением биологической терминологии в упрощенном виде, сущностью которого является определение наиболее приспособленной особи (решения) по значению функции ее приспособленности в ходе эволюции с учетом анализа влияний факторов наследственности и внешней среды. Несмотря на биологическую терминологию, ГА являются универсальным вычислительным средством, с помощью которого можно решать широкий круг сложных задач, в том числе и в отрасли электроэнергетики. Авторами был рассмотрен вопрос о применении генетического алгоритма в рамках расчета установившегося режима электрической сети (УР ЭС), так как математическая модель электрической сети представляет собой систему нелинейных уравнений высокого порядка, в которой учтены все ограничения, накладываемые физическими свойствами рассматриваемого объекта, и решение которой ввиду сложности реально действующих электрических сетей – достаточно трудоемкая оптимизационная задача. Корректное решение данной системы является наиболее ответственным этапом расчета УР ЭС. Именно по этой причине поиск оптимальных методов расчета УР ЭС является важной и актуальной задачей. В данной работе представлены результаты разработки аналитического аппарата, позволившего осуществить поиск решения задачи расчета установившихся режимов электрических сетей методом генетического алгоритма посредством специального программного обеспечения.

Ключевые слова: эволюционное моделирование, генетический алгоритм, функция приспособленности, электрические сети, расчет режима

Для цитирования: Солопов Р. В., Самульченков А. С., Зирюкин В. И. Genetic algorithm as a tool for modeling calculations of electric power systems // Прикладная информатика. 2021. Т. 16. № 6. С. 43–53. DOI: 10.37791/2687-0649-2021-16-6-43-53

Introduction

Iterative methods or methods of successive approximations are widely used to calculate the steady-state modes of electrical networks (SS EN). Initially, the calculation of the SS EN was carried out using the simplest methods, such as the Gauss method, the simple iteration method, or the Seidel method. Considering the permanent development of computers and the corresponding computing power growth since the second half of the 20th century it became possible to implement methods using a more cumbersome and complex analytical apparatus, in particular, the family of Newton-Raphson methods of various accuracy orders. These methods made it possible to obtain a result with a fundamentally upper speed and quality.

For a genetic algorithm (GA) the form of the function under consideration is not important in contrast to traditional solution methods, because the GA works with its values at certain points but not with the function itself. Also, the advantage of GA is its loyalty both to the mathematical rigor of the under-consideration function, and to the mechanism for modifying the initial approximation during the search for a solution, namely: the possibility of changing the main operators configuration during the solution, depending on the values of the chromosomes under consideration objective function which as well called vectors of possible decisions.

That is why GA is preferable to solve technical problems that require large computing resources in comparison with the traditional calculation methods, since the absence of imposed restrictions on the under-consideration function and the performance of time-consuming in implementation step-by-step algebraic operations, such as calculating [1, 2].

The genetic algorithm general characteristics

A distinctive GA feature is the idea of using development principles of the of organism populations, which generalization of the main stages makes it possible to form a computational model that takes into account the peculiarities of solving the under-consideration problem from the perspective of this process naturalness. This factor determines the flexibility and variability of the solution process, in contrast to traditional computational calculation methods, which algorithm is unambiguous and provides only a number of modifications [3–8].

The most important property that characterizes GA is its adaptability. Precisely the modeling of the individuum adaptation possibility in the process of evolution, that depends on changes in environmental conditions with the aim of offspring survival and further reproduction, makes GA the most powerful optimization computing apparatus. That is because the solution (individuum) of the considered mathematical problem in within the implemented according to the GA principle computational process (evolution), subjecting to the change in the initially set conditions (environment), will a priori have an optimal character.

GA uses the interaction principles of a set of sought solutions which are represented as the form of chromosomes like the principles of individuum interaction in nature. That's why with the help of GA, having previously correctly formulated condition of the problem (calculating the steady state of the electrical network), you can find a solution that corresponds not only to the normal mode of operation, but to the optimal as well, due to the specifics of the applied technique [9–12].

GA can be conditionally divided into the following stages [9, 10]:

- the initial data type of presentation choice – many problems, where it is advisable to use GA to find a solution, do not allow to present the initial data in an explicit form, due to certain difficulties associated with the subsequent implementation of the algorithm and the specifics of the topic

under consideration. Therefore, the encoding initial data practice by means of a binary code, Gray code, etc. has become widespread. This problem can be leveled by the development of a GA structure, which is aimed at implementing highly specialized tasks;

- the formation of the initial solutions population. A population means a certain number of individuals, which are represented in the form of chromosomes – a set of randomly generated solutions, or the so-called genes, each of which is located at a certain position in the chromosome. In fact, the set of chromosomes characterizes the set of the considered mathematical model solutions – an n -dimensional system of equations, where n is the number of genes in each of the chromosomes. Depending on the topic of the under-consideration problem, it is necessary to include the limitations originally imposed on the initial data (range of variation, etc.), and the possibility of changing the number of initially generated chromosomes, which directly depends on the problem complexity and the required level of solution accuracy;

- the configuration choice of each of the main operators of random changes in GA. The main operators used in the construction of GA are selection operator, crossing over, mutation operator. The selection operator makes it possible to identify in the most rational way and subsequently get rid of less adapted individuals in the population with the purpose of the only best representatives' further modification. Crossing over (the operator responsible for the reproduction of new offspring) is realized through a certain crossing model. It is the exchange of genes located at certain positions in chromosomes, which main function is the modification of chromosomes preceding the population. The mutation operator is designed to prevent the degradation of generations and is implemented by modifying genes with a given probability. The combination of the above operators with various types of configurations makes it possible to build a GA that meets all the task requirements;

- the formation of fitness function. It is an analytical model of the considered physical object that meets the basic requirements of the GA specifics implementation. Genetic algorithm, as noted earlier, involves solving the optimization problem and is inherently an adaptive method for finding the best representatives of the population. Hence, the fitness function should be formed as a function that makes it possible to implement the search for solutions to the optimization problem, regardless of the goals pursued by the solution of the problem (far from always optimization). Also, the type of the considered physical object and the totality of all restrictions imposed on its mathematical description due to the peculiarities characteristic of the separately considered branch of research are essential. That is, this function should provide a correct analytical representation of the considered physical object, taking into account all the characteristic features inherent in it;

- the selection of criteria for stopping the calculation is caused, first of all, by the goals that pursue the solution of various problems, since that they impose a direct restriction on the required accuracy of the results.

All of the above stages of the calculation implementation by the GA method correspond to the classical configuration of this algorithm.

The electric power system representation in the development of the genetic algorithm fitness function

Consider the nonlinear equations of node voltages that describe the SS EN when specifying nonlinear current sources, i. e. generators and loads are set with constant powers that correspond to nonlinear current sources in the equivalent circuit of the electric network [8, 10].

The system of nonlinear equations (1) describing an electrical network with the number of nodes equal to n will have the form:

$$\left\{ \begin{array}{l} \dot{U}_1 \cdot \dot{Y}_{11} + \dot{U}_2 \cdot \dot{Y}_{12} + \dots + \dot{U}_{(n-1)} \cdot \dot{Y}_{1(n-1)} + \dot{U}_n \cdot \dot{Y}_{1n} = \frac{\bar{S}_1}{\bar{U}_1}; \\ \dot{U}_1 \cdot \dot{Y}_{21} + \dot{U}_2 \cdot \dot{Y}_{22} + \dots + \dot{U}_{(n-1)} \cdot \dot{Y}_{2(n-1)} + \dot{U}_n \cdot \dot{Y}_{2n} = \frac{\bar{S}_2}{\bar{U}_2}; \\ \vdots \\ \dot{U}_1 \cdot \dot{Y}_{(n-1)1} + \dot{U}_2 \cdot \dot{Y}_{(n-1)2} + \dots + \dot{U}_{(n-1)} \cdot \dot{Y}_{(n-1)(n-1)} + \dot{U}_n \cdot \dot{Y}_{(n-1)n} = \frac{\bar{S}_{(n-1)}}{\bar{U}_{(n-1)}}; \\ \dot{U}_1 \cdot \dot{Y}_{n1} + \dot{U}_2 \cdot \dot{Y}_{n2} + \dots + \dot{U}_{(n-1)} \cdot \dot{Y}_{n(n-1)} + \dot{U}_n \cdot \dot{Y}_{nn} = \frac{\bar{S}_n}{\bar{U}_n}, \end{array} \right. \quad (1)$$

where \bar{S}_k is a complex conjugate constant three-phase power of the k -node ($k = 1, 2, \dots, n$);

\dot{U}_k – k -node phase-to-phase voltage;

\dot{Y}_{kk} – k -node intrinsic conductivity;

\dot{Y}_{kj} – k -node mutual conductivity ($j = 1, 2, \dots, n; k \neq j$).

Then the problem of calculating the SS EN is reduced to determining the unknown voltages at the considered electrical network nodes.

The voltage module and angle at each node is set by a randomly generated value, each of which corresponds to a specific gene in the chromosome. After through a step-by-step modification, the formation of new chromosomes populations (a set of solutions) is carried out, which compliance with the required result is assessed from the value of their fitness function. Since this function should be of an optimization nature, we will make several transformations with the equations of system (1) for the possibility of this function final form further formation.

In the general case, each of n nodes can be load (2), generator (3), or simultaneously load and generator (4):

$$\dot{U}_k \cdot \dot{Y}_{kk} + \sum_{\substack{j=1 \\ j \neq k}}^n \dot{Y}_{kj} \cdot \dot{U}_j + \frac{\bar{S}_{k(N)}}{\bar{U}_k} = 0; \quad (2)$$

$$\dot{U}_k \cdot \dot{Y}_{kk} + \sum_{\substack{j=1 \\ j \neq k}}^n \dot{Y}_{kj} \cdot \dot{U}_j - \frac{\bar{S}_{k(G)}}{\bar{U}_k} = 0; \quad (3)$$

$$\dot{U}_k \cdot \dot{Y}_{kk} + \sum_{\substack{j=1 \\ j \neq k}}^n \dot{Y}_{kj} \cdot \dot{U}_j + \frac{\bar{S}_{k(N)}}{\bar{U}_k} - \frac{\bar{S}_{k(G)}}{\bar{U}_k} = 0. \quad (4)$$

Formulas (2)–(4) represent current imbalance equations.

In the general case, to describe each of the n nodes of the electrical network, it is advisable to apply formula (4), because of the generator or load absence in the under-consideration node. The variables corresponding to their powers simply take zero value.

After performing algebraic transformations and taking into account the presence of power losses in real electrical networks, we obtain the (5):

$$\bar{U}_k \cdot \dot{U}_k \cdot \dot{Y}_{kk} + \bar{U}_k \cdot \sum_{\substack{j=1 \\ j \neq k}}^n \dot{Y}_{kj} \cdot \dot{U}_j + \dot{S}_{k(N)} - \dot{S}_{k(G)} = \Delta \dot{S}_k, \quad (5)$$

where $\Delta \dot{S}_k$ – the k -node power imbalance equation of under-consideration network ($k = 1, 2, \dots, n$).

Considering all the transformations carried out with the equations included in system (1), we can rewrite it in matrix form (6):

$$\begin{bmatrix} \bar{U}_1 & 0 & \cdots & 0 \\ 0 & \bar{U}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{U}_n \end{bmatrix} \cdot \begin{bmatrix} \bar{Y}_{11} & \bar{Y}_{12} & \cdots & \bar{Y}_{1n} \\ \bar{Y}_{21} & \bar{Y}_{22} & \cdots & \bar{Y}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{Y}_{n1} & \bar{Y}_{n2} & \cdots & \bar{Y}_{nn} \end{bmatrix} \cdot \begin{bmatrix} \dot{U}_1 \\ \dot{U}_2 \\ \vdots \\ \dot{U}_n \end{bmatrix} + \begin{bmatrix} \dot{S}_{1(N)} \\ \dot{S}_{2(N)} \\ \cdots \\ \dot{S}_{n(N)} \end{bmatrix} - \begin{bmatrix} \dot{S}_{1(G)} \\ \dot{S}_{2(G)} \\ \cdots \\ \dot{S}_{n(G)} \end{bmatrix} = \begin{bmatrix} \Delta \dot{S}_1 \\ \Delta \dot{S}_2 \\ \cdots \\ \Delta \dot{S}_n \end{bmatrix}. \quad (6)$$

In terms of the electrical network evolutionary model formation and implementation the most important criterion for the correctness is rightness and compliance with all imposed restrictions on the developed fitness function of the GA. That's why the matrix equation (6) is of a greatest interest [10]. It fully reflects the essence of all physical processes occurring in the electric power system. It is also a correct mathematical model that takes into account the specifics of the applied calculation method, because the goal of solving this matrix equation is to minimize power imbalances in the electrical network.

The final recording form of the fitness function is (7):

$$F = \sum_{j=1}^n |\Delta S_j| \rightarrow \min. \quad (7)$$

Particular attention deserves the fact that while estimating this function, the total imbalance power module is calculated, but not the complex value. This is due to the fact that when the considered function is set in a similar form, the ES SD is optimized by the power losses absolute value.

Thus, the essence of calculating the SS EN by the GA method is to minimize the modules of power imbalances in the electrical network [9, 10].

The genetic algorithm structure development for calculating the steady-state modes of electrical networks

The proposed structure of the algorithm intended for calculating the SS EN by the GA method corresponds to the classical GA configuration.

The modules of the nodes voltage U_i and the corresponding phase shift angles δ_i act as the required parameters when solving the problem of the SS EN calculating. The parameters are represented in real form, where the gene sequence is a vector in which the variables corresponding to each of the nodes are arranged in an orderly manner. Each of the chromosomes corresponding to the set of unknowns, when calculating the SS EN for a scheme with the number of nodes equal to n looks as follows:

$$X_i = (U_1, \delta_1, U_2, \delta_2, \dots, U_n, \delta_n).$$

The dimension of each of the generations except for the zero population is fixed and amounts to ten chromosomes. The number of randomly generated chromosomes in the acceptable ranges for the formation of the first approximation may vary depending on the complexity of the problem being solved.

In order to ensure the possibility of obtaining correct results the architecture of the developed GA (fig. 1), is implemented according to the block principle in a multistage configuration,

regardless of the system dimension. Each of the four stages is an independent GA with settings that consider the characteristics of the results obtained at the previous stage. The choice of representing the GA architecture in a similar form is due to the results of empirical research, during which the optimal model was determined.

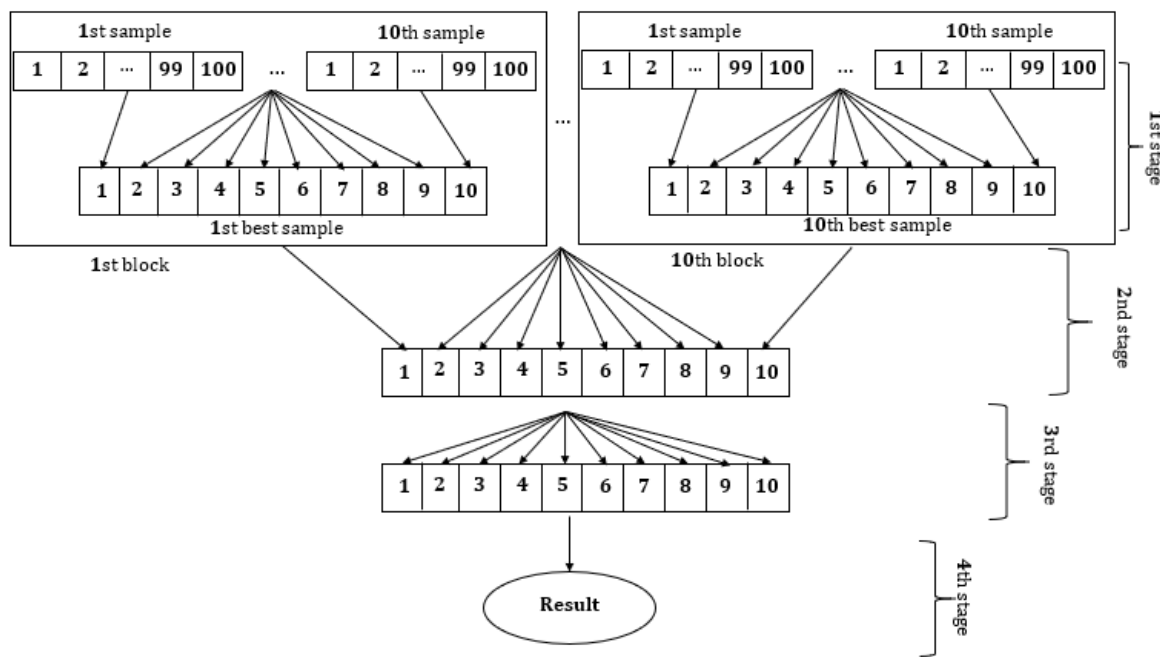


Fig. 1. Four-step architecture of genetic algorithm

The first stage was designed as a block structure for the possibility of more careful representatives' selection from the randomly generated zero generation. Each of the first generation ten chromosomes of a single block is the result of choosing the best one from a hundred randomly generated chromosomes of ten initial generations. Thus, the result of the first stage implementation is ten blocks, each of which consists of ten most adapted zero populations chromosomes.

The goal of the second stage is to form a population of ten chromosomes, i. e., to choose the best of ten chromosomes from each of the first stage blocks. This result is achieved by using GA with certain settings typical for this stage.

At the third stage, the ten best ones from the chromosomes of the previous population are determined by tenfold application of GA with modified settings.

At the fourth stage the final result of the calculation is determined by a single application of GA from the generation formed at the end of the third stage.

Each of the GA structure is used at each of the stages similarly. The differences in their settings lie in the number of iterations performed during each stage and the range of parameter values change during the application of the mutation operator.

Tournament selection is used as a selection operator, i. e., out of ten chromosomes, five pairs are formed, while comparing the values of the fitness function, the five most adapted representatives of the generation are determined.

Recombination is performed using a multipoint crossover. Its separators have fixed positions in each of the chromosomes. To obtain ten offsprings five parents are crossed in such a way that the fittest individuals are the parents of a larger number of new chromosomes than the less fit ones.

The chromosomes mutation is carried out according to the results of determining mutation random probability values for each of the genes of ten chromosomes and their further quantitative modification if this probability turns out to be higher than a certain value. As noted earlier, the range of quantitative changes and their transition to the next stage of architecture likelihood varies (rather decreases) because at the last stages of the computational process implementation it is necessary to make small adjustments to the gene values in order to obtain the final result. On the contrary, at the first steps the mutation probability and the variables change values take on larger values.

The equation of the total power imbalance in the electrical network acts as a fitness function [10].

The criterion for stopping the calculation is the achievement of the fitness function desired value.

Program for calculating steady-state modes of electrical networks using the genetic algorithm

According to this algorithm, a program was developed for calculating the SS EN using the GA method. A screenshot of the program is shown in figure 2.

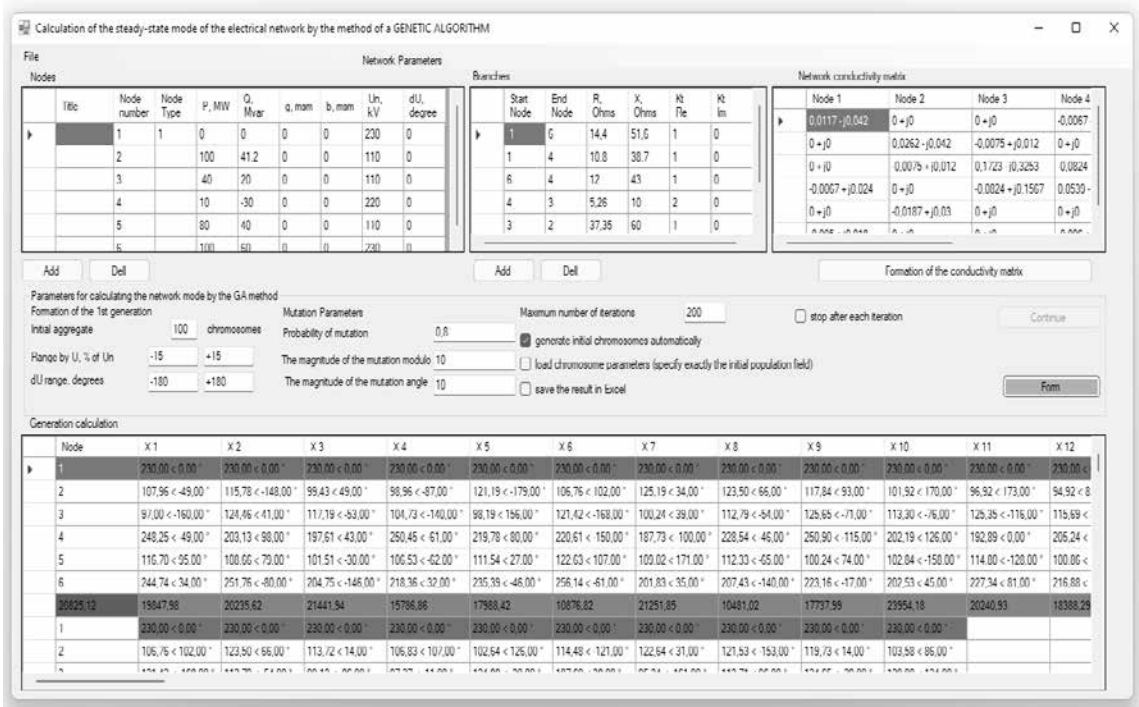


Fig. 2. The network steady-state calculating program using the genetic algorithm method

An example of the three-node closed electrical network steady-state mode calculation (fig. 3) is presented which was carried out using the developed computer program (fig. 2).

The steady-state mode of the considered electrical network was calculated using the RastrWin3 software in order to verify the results obtained. RastrWin3 is a software package designed to solve a wide range of electrical power problems – calculation, analysis, and optimization of various modes of electrical networks and systems. Currently, the main users of this software are: System Operator of the Unified Energy System (SO CDU UES) and its branches [10].

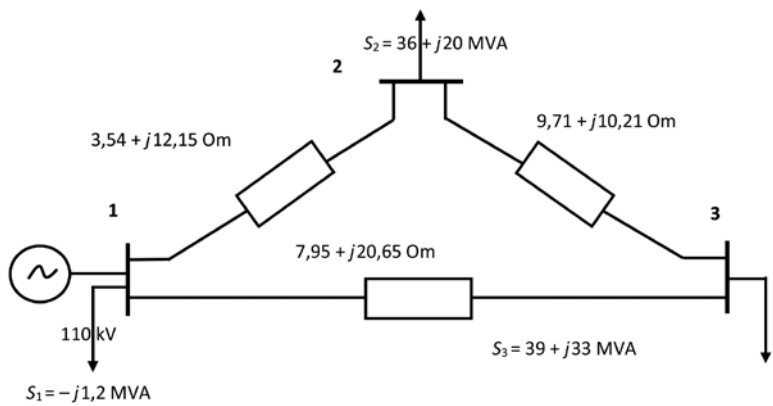


Fig. 3. The considered electrical network scheme

The RastrWin3 software is based on traditional (iterative) calculation methods. Table 1 shows the results of calculating the independent variables (node voltages) of the considered three-node electrical network steady-state mode. The node 1 was taken as the basic and balancing node for calculating the ES SD using the RastrWin3 software, i. e. the its voltage was set initially (nominal value), and the voltages of the other nodes were calculated.

Table 1. The parameter results computing of the considered electrical network steady-state mode

Node number	The genetic algorithm		RastrWin3	
	Voltage module, kV	Phase, degrees	Voltage module, kV	Phase, degrees
1	113,67	0,19	110	0
2	108,59	-1,66	104,18	-2,09
3	107,47	-1,75	102,55	-2,16

After the obtained results of calculated by the GA method node voltages were taken in RastrWin3 as basic to compare the final power losses value in the considered network with the losses values obtained at the initial value of voltages.

Table 2 represents the calculation power losses results in the electrical network, obtained with the initial values of voltages in RastrWin3, the values of voltages calculated by the GA method and values taken as basic in RastrWin3.

Table 2. The computing power losses results of the considered electrical network steady-state mode

Value	RastrWin3 (initial values of voltages)		RastrWin3 (voltage values obtained by the GA method)	
	Active power, MW	Reactive power, MVar	Active power, MW	Reactive power, MVar
Power loss	2,38	6,88	1,77	5,21

The electrical network total power losses received when calculating the SS EN by the GA method, turned out to be quantitatively less than while using RastrWin3 program (table 2).

Conclusion

Currently, information and digital technologies are being actively implemented in the electric power system as part of the smart electric grids development. The process of equipment digita-

lization is actively being carried out at the electrical network substations. Its purpose is to unite the power system into a single automated information and technical network which will be controlled in real time.

Classical methods for calculating the electric power systems modes require significant computing resources which makes it difficult to use them in real time control systems. This is due to a number of significant factors: low accuracy of calculations; restrictions imposed on the power system model by traditional calculation methods, etc. [1, 3–6].

Therefore, the fundamental problem of providing technical calculations in energy systems using innovative computational algorithms which are based on various artificial intelligence methods, particularly evolutionary methods, becomes relevant.

The studies carried out to compare the capabilities of GA and traditional computational methods in relation to the calculation of the SS EN showed the results indicating the prospects for the development of the evolutionary modeling theory of electric power systems [8–12].

References

1. Kureichik V. M., Kaplunov T. G. Adaptive genetic algorithm based on fuzzy rules. *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki*=Izvestiya SFedU. Engineering Sciences, 2018, no.5(199), pp.26-34 (in Russian). DOI: 10.23683/2311-3103-2018-5-26-34.
2. Kureichik V. M., Sinyutin V. G., Kaplunov T. G. Prediction of the state of technical systems using genetic algorithms. *Vestnik ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*=Vestnik of Ryazan State Radio Engineering University, 2018, no.65, pp.107-112 (in Russian). DOI: 10.21667/1995-4565-2018-65-3-107-112.
3. Vlatskaya L. A., Semenova N. G. Application of genetic algorithms in problems of placement optimization of compensating devices. *Elektrotekhnicheskie sistemy i komplekсы*=Electrotechnical Systems and Complexes, 2019, no.4(45), pp.21-28 (in Russian). DOI: 10.18503/2311-8318-2019-4(45)-21-28.
4. Samakwong T., Assawinchaichote W. PID controller design for electro-hydraulic servo valve system with genetic algorithm. *Procedia Computer Science*, 2016, vol.86, pp.91-94. DOI: 10.1016/j.procs.2016.05.023.
5. Jayachitra A., Rajendran V. Genetic algorithm based PID controller tuning approach for continuous stirred tank reactor. *Advances in Artificial Intelligence*, 2014, vol.15, pp.1-8. DOI: 10.1155/2014/791230.
6. Pimenov V. I., Pimenov I. V. Interpretation of a trained neural network based on genetic algorithms. *Information and Control Systems*, 2020, no.6, pp.12-20. DOI: 10.31799/1684-8853-2020-6-12-20.
7. Kychkin A. V., Chudinov A. V. Applying the genetic algorithm for optimizing the location of alternative energy sources in a local section of active-adaptive network. *Sistemy. Metody. Tekhnologii*=Systems. Methods. Technology, 2016, no.1(29), pp.76-82 (in Russian). DOI: 10.18324/2077-5415-2016-1-76-82.
8. Samulchenkov A. S. *Opyt primeneniya geneticheskikh algoritmov dlya rascheta ustanovivshegosya rezhima elektricheskoi seti* [Experience the use of the genetic algorithm to compute steady state response of an electric circuit]. *Sbornik nauchnykh trudov XV Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii studentov i aspirantov «Informatsionnye tekhnologii, energetika i ekonomika»: v 3 t. (Smolensk, 26–27 aprelya)* [Proceedings of the 15th International scientific and technical conference of students and postgraduates "Information technology, Energy and Economics": in 3 parts (April, 26-27, Smolensk, Russia)] , 2018, vol.1, pp.64-67.
9. Solopov R. V., Samulchenkov A. S. Application of genetic algorithm for electric circuit steady-state response calculation. *Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo universiteta*=Proceedings of Irkutsk State Technical University, 2018, vol.22, no.2(133), pp.131-141 (in Russian). DOI: 10.21285/1814-3520-2018-2-131-141.
10. Solopov R. V., Samulchenkov A. S. Development of the fitness function of the electrical network evolutionary model for computing of the settling behavior using the method of genetic algorithm. *Proceedings of the International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)* (March, 25-29, Sochi, Russia), 2019, pp.1-5. DOI: 10.1109/ICIEAM.2019.8742968.
11. Yuancheng Li, Wang Yiliang, Bin Li. A hybrid artificial bee colony assisted differential evolution algorithm for optimal reactive power flow. *International Journal of Electrical Power and Energy Systems*, 2013, vol.52, no.1, pp.25-33. DOI: 10.1016/j.ijepes.2013.03.016.
12. Sahu G., Swarnkar K. Review of reactive power optimization using evolutionary computation techniques. *Advance in Electronic and Electric Engineering*, 2014, vol.4, no.1, pp.73-82.

About the authors

Roman V. Solopov, ORCID 0000-0002-0882-7278, Cand. Sci. (Eng.), Associate Professor, Electrical Power Systems Department, Branch of the National Research University "MPEI" in Smolensk, Smolensk, Russia, solopov.rv@mail.ru
Anton S. Samulchenkov, ORCID 0000-0002-5415-481X, Master's Student, Electrical Power Systems Department, Branch of the National Research University "MPEI" in Smolensk, Smolensk, Russia, samulchenkov@yandex.ru
Vladislav I. Ziryukin, ORCID 0000-0002-5962-9810, Postgraduate, Higher School of High Voltage Engineering, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia, vladocs2014@yandex.ru

Received 31.05.2021, reviewed 01.07.2021, accepted 03.09.2021

Сведения об авторах

Солопов Роман Вячеславович, ORCID 0000-0002-0882-7278, канд. техн. наук, доцент, кафедра электроэнергетических систем, филиал Федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет "МЭИ"» в г. Смоленске, Смоленск, Россия, solopov.rv@mail.ru

Самульченков Антон Сергеевич, ORCID 0000-0002-5415-481X, магистрант, кафедра электроэнергетических систем, филиал Федерального государственного бюджетного образовательного учреждения высшего образования «Национальный исследовательский университет "МЭИ"» в г. Смоленске, Смоленск, Россия, samulchenkov@yandex.ru

Зирюкин Владислав Игоревич, ORCID 0000-0002-5962-9810, аспирант, Высшая школа высоковольтной энергетики, Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия, vladocs2014@yandex.ru

Статья поступила 31.05.2021, рассмотрена 01.07.2021, принята 03.09.2021

Исследование процесса бинаризации изображений с использованием локальных значений порога

А. С. Лошкарев^{1*}, И. А. Тимофеев¹

¹ Поволжский государственный университет телекоммуникаций и информатики,
Самара, Россия

* lozhkarev-as@mail.ru

Аннотация. Использование глобальных порогов бинаризации при обработке изображений не всегда дает корректный результат. Это особенно часто наблюдается при обработке изображений с неравномерной освещенностью. На одних участках изображения автоматически определяемый порог бинаризации позволяет получить достаточно хорошо визуализируемые объекты, тогда как на других участках необходимые для анализа объекты становятся «засвеченными» или, наоборот, «затененными». Бинаризация играет очень важную роль в тех случаях, когда необходимо локализовать на изображении все объекты интереса, особенно когда объект интереса содержит в себе информацию, которая будет использоваться на следующих этапах обработки. Многоградационные изображения могут содержать в себе множество объектов интереса, таких как номерные знаки автомобилей, номера вагонов поезда, лица людей, дефекты производимой продукции. Каждый из этих случаев требует качественной обработки для последующего распознавания. Если на обрабатываемом изображении присутствуют шумы или показатели яркости распределены неравномерно, то процесс бинаризации может привести к потере важной информации – потере части символа, обрыву контура объекта или, наоборот, возникновению новых областей, ошибочно прибавляемых к объекту интереса, – тени других объектов, грязь на номерном знаке. Поэтому процесс бинаризации требует очень точной предварительной калибровки под все возможные условия съемки – светлое и темное время суток, учет возможной зашумленности (помехи при передаче сигнала), экстремальные ситуации (сильный град или дождь). В данной статье авторы исследуют процесс бинаризации изображений с неравномерной освещенностью с использованием нескольких локальных порогов бинаризации вместо одного глобального. Предлагается проверять гистограммы полученных фрагментов на количество пиков или «мод». Если гистограмма бинаризованного фрагмента является одномодальной, то данный фрагмент не подлежит дальнейшей обработке и порог бинаризации на нем определен правильно.

Ключевые слова: бинаризация, обработка изображений, локальный порог, мера средней яркости, метод Отсу, гистограмма

Для цитирования: Лошкарев А. С., Тимофеев И. А. Исследование процесса бинаризации изображений с использованием локальных значений порога // Прикладная информатика. 2021. Т. 16. № 6. С. 54–65.
DOI: 10.37791/2687-0649-2021-16-6-54-65

Investigation of the process of binarization of images using local values of the threshold

A. Lozhkarev^{1*}, I. Timofeev¹

¹ Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia

* lozhkarev-as@mail.ru

Abstract. The use of global binarization thresholds in image processing does not always give the correct result. This is especially common when processing images with uneven illumination. In some areas of the image, the automatically determined binarization threshold makes it possible to obtain sufficiently well visualized objects, while in other areas, the objects necessary for analysis become "overexposed" or, conversely, "shaded". In cases where it is necessary to localize all objects of interest in the image, binarization plays a very important role, especially in cases where the object of interest contains information that will be used in the next stages of processing. Multi-gradation images can contain many objects of interest, such as car license plates, train car numbers, people's faces, and defects in manufactured products. Each of these cases requires high-quality processing for subsequent recognition. If there are noises on the processed image or the brightness indicators are unevenly distributed, then the binarization process can lead to the loss of important information – the loss of a part of the symbol, the breakage of the object's contour, or, conversely, the emergence of new areas that are mistakenly added to the object of interest – shadows of other objects, dirt on the license plate sign. Therefore, the binarization process requires a very accurate preliminary calibration for all possible shooting conditions – daylight and dark hours of the day, taking into account possible noise (interference in signal transmission), extreme situations (strong hail or rain). In this article, the authors investigate the process of binarization of images with uneven illumination using several local binarization thresholds instead of one global one. It is proposed to check the histograms of the obtained fragments for the number of peaks or "modes". If the histogram of a binarized fragment is single-mode, then the given fragment is not subject to further processing and the binarization threshold on it is defined correctly. The study of the relationship between the binarization threshold and such image parameters as dispersion and smoothness has been carried out. On those fragments where the value of the average brightness measure differs from the average for all fragments, the binarization threshold is determined incorrectly. If you set the threshold value higher, closer to the average value for all fragments, then as a result binarization will be carried out correctly.

Keywords: binarization, image processing, local threshold, average brightness measure, Otsu method, histogram

For citation: Lozhkarev A., Timofeev I. Investigation of the process of binarization of images using local values of the threshold. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.54-65 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-54-65

Введение

Н а сегодняшний день важное место в системах видеоаналитики занимает процесс сегментации изображения. При помощи бинаризации можно существенно сни-

зить количество анализируемой информации и повысить скорость работы системы в целом [1]. Однако точность работы системы именно на этапе бинаризации также сильно сказывается на результатах распознавания образов.

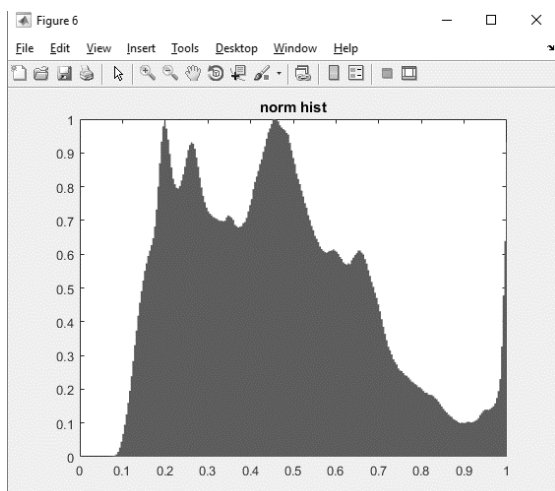


Рис. 1. Гистограмма исходного изображения

Fig. 1. Histogram of source image

В методах глобальной обработки пороговая поверхность является плоскостью с постоянным порогом яркости, т. е. значение порога рассчитывается исходя из анализа гистограммы всего изображения и является одинаковым для всех пикселей исходного изображения, как показано на рисунке 1. Глобальная пороговая обработка имеет существенный недостаток – если исходное изображение имеет неоднородное освещение, то области, которые освещены хуже, целиком классифицируются как передний план.

При автоматическом определении порога бинаризации возможны потери информации на изображении. На тестовом изображении при использовании метода Отсу [2] порог бинаризации получился равным 0,48. Данный результат не позволяет в дальнейшем распознать символы на изображении, как показано на рисунке 2.

Исследование процесса бинаризации изображения при предварительной сегментации на отдельные области

Чтобы решить проблему, связанную с потерей части визуальной информации на бина-



Рис. 2. Бинаризованное изображение с автоматически определенным порогом методом Отсу

Fig. 2. Binarized image with automatically determined threshold by Otsu method

ризованном изображении, предлагается следующий способ анализа гистограмм.

Чем сложнее гистограмма исходного изображения, тем сложнее определить порог бинаризации. Вид гистограммы можно определить исходя из количества пиков яркости пикселей. Гистограмма, показанная на рисунке 3, является многомодовой, т. е. количество выраженных пиков яркости больше одного [3].

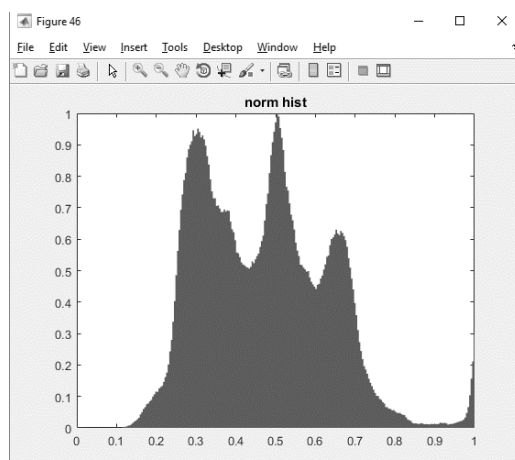


Рис. 3. Гистограмма, отвечающая требованию «многомодовая»

Fig. 3. Histogram meeting the requirement of "single mode"

Предлагается поделить изображение на 16 одинаковых по размеру фрагментов. Для каждого фрагмента вычисляется нормированная гистограмма яркости и порог бинаризации, как показано на рисунке 4. Если гистограмма фрагмента является одномодовой, то фрагмент не подлежит дальнейшей обработке.

Данное количество фрагментов обусловлено тем фактом, что дисперсия распределения яркости пикселей гораздо слабее выражена на меньшем количестве фрагментов, например при делении на 4, 8 или 12 фрагментов. А это означает, что на каждом из этих фрагментов порог бинаризации по методу Отсу практически не будет отличаться от глобального порога бинаризации для всего изображения в целом.

Деление изображения на 16 фрагментов является максимально допустимым. С увеличением числа фрагментов на бинаризованном изображении начинают явно преобладать перепады яркостей соседних фрагментов, и дальнейший анализ визуальной информации становится затруднителен (рис. 5).

На изображении с неравномерным распределением яркости (рис. 6) наблюдаются

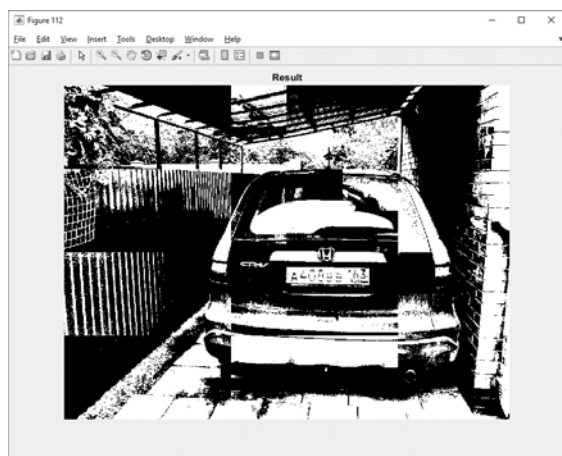


Рис. 4. Итоговая обработка – деление изображения на 16 фрагментов и дальнейшее деление фрагментов на 4 подфрагмента

Fig. 4. Final processing – dividing the image into 16 fragments and further dividing the fragments into 4 sub-fragments

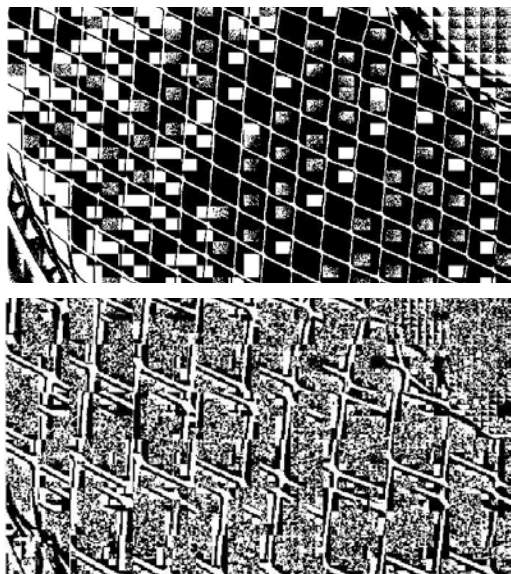


Рис. 5. Примеры бинаризованных изображений, на которых количество исходных фрагментов превышает 16

Fig. 5. Examples of binarized images in which the number of original fragments exceeds 16

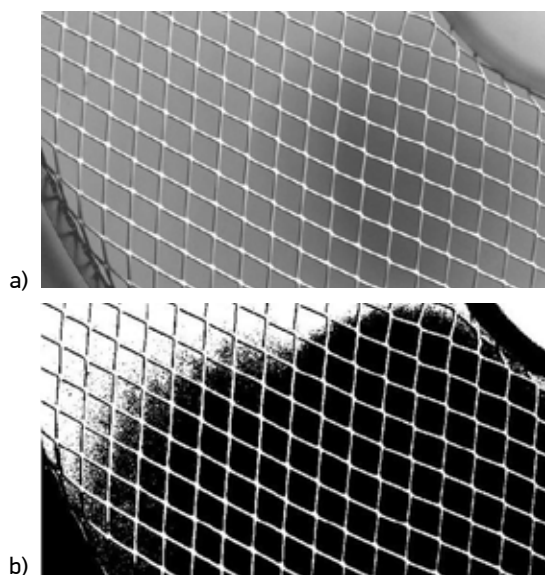


Рис. 6. Изображение с неравномерным распределением яркости: а) исходное изображение; б) бинаризованное изображение с порогом 0,58431

Fig. 6. Image with uneven brightness distribution: а) original image; б) binarized image with a threshold of 0.58431

потери информации при бинаризации. Эти потери связаны с распределением яркости пикселей на изображении (рис. 7). Чем больше значение дисперсии распределения яркости, тем менее точно будет подобран глобальный порог бинаризации. Для одного участка изображения порог будет оптимальным, а для другого, с иным распределением значений яркости, – неподходящим.

При делении на 16 фрагментов и отдельной бинаризации каждого из них остаются проблемные участки (рис. 8а). А также дальнейшее деление каждого фрагмента еще на 4 не дает достоверных результатов (рис. 8б, 9–11). Прямой зависимости значения оптимального порога бинаризации от значений яркости пикселей на обрабатываемом фрагменте не выявлено. Авторы предлагают рассматривать связь между порогом бинаризации и такими параметрами изображения, как дисперсия и гладкость [4]. Эти статистические характеристики изображения используются для текстурного описания изображения и могут быть получены из гистограммы яркости изображения или его области. Тектурный метод описания хорошо подходит для работы с многоградационными изображениями.

Нумерация фрагментов идет из верхнего левого угла построчно: первая строка состоит из фрагментов 1.1, 1.2, 1.3, 1.4, вторая строка – 2.1, 2.2, 2.3, 2.4 и так далее. Как можно заметить, фрагменты 3.1, 4.1, 1.3 и 1.4 являются проблемными, так как порог бинаризации на

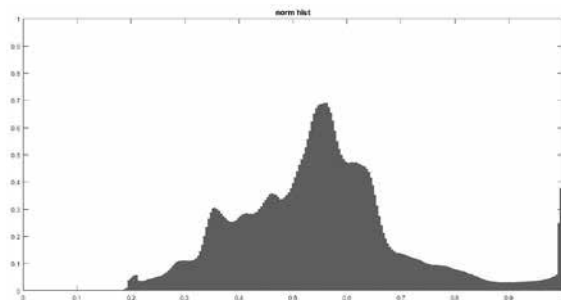


Рис. 7. Гистограмма изображения с неравномерным распределением яркости
Fig. 7. Image histogram with uneven brightness distribution

них определен ниже, чем на остальных фрагментах, которые бинаризованы приемлемо, то есть на них не наблюдается резкого визуального перепада яркости пикселей, границы между соседними фрагментами не выражены столь явно [5].

$$\mu_2(r) = \sum_{i=0}^{L-1} (r_i - m)^2 p(r_i). \quad (1)$$

Это выражение есть не что иное, как дисперсия значений яркости r , которая обычно обозначается $\sigma^2(r)$. (Напомним, что стандартное отклонение σ есть квадратный корень из

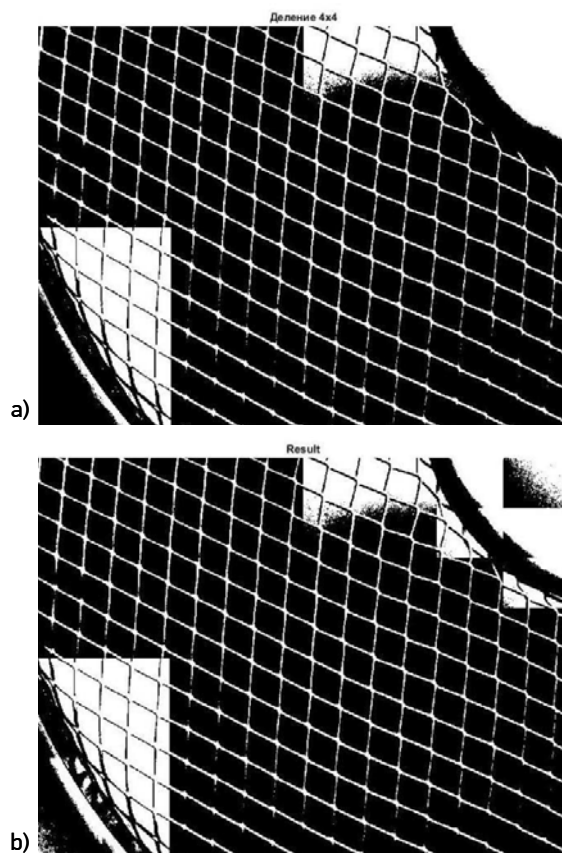


Рис. 8. Бинаризация изображения при делении на фрагменты: а) при делении на 16 фрагментов; б) при последующем делении каждого фрагмента еще на 4 фрагмента

Fig. 8. Binarization of the image when dividing into fragments: а) when divided into 16 fragments; б) with the subsequent division of each fragment into 4 more fragments

дисперсии.) Тогда как математическое ожидание является мерой средней яркости, дисперсия (или стандартное отклонение) является мерой контраста на изображении [6].

Когда требуются лишь среднее и дисперсия, часто их оценивают непосредственно

по значениям отсчетов, без вычисления гистограммы [7]. Эти оценки называются *выборочное среднее* и *выборочная дисперсия* соответственно. Они задаются следующим выражением:

$$m = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \quad (2)$$

для $x = 0, 1, 2, \dots, M-1$ и $y = 0, 1, 2, \dots, N-1$. M и N – число строк и столбцов в изображении, то есть его размер. Подобная интерпретация применима и к уравнению (3):

$$\sigma^2 = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - m]^2. \quad (3)$$

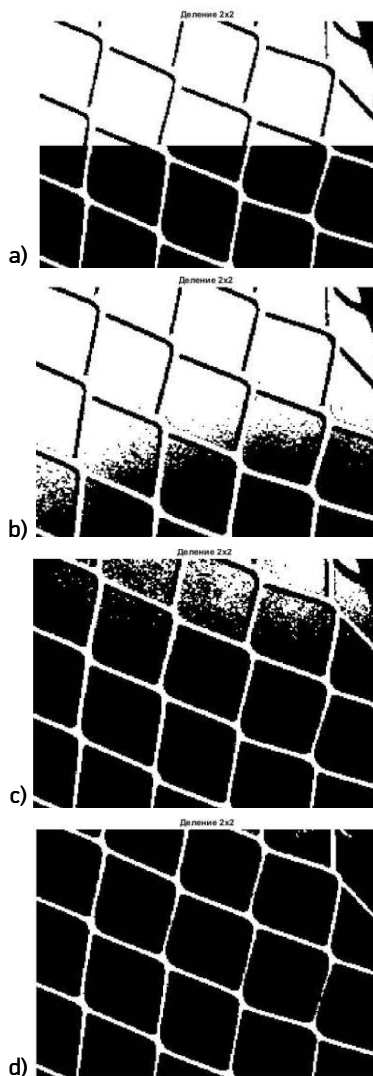


Рис. 9. Фрагмент 1.3 после деления на 4 подфрагмента и их обработки: а) значениями локальных порогов бинаризации; б) значением порога 0,6; в) значением порога 0,7; д) значением порога 0,75

Fig. 9. Fragment 1.3 after dividing into 4 subfragments and processing them: a) values of local binarization thresholds; b) threshold value 0.6; c) threshold value 0.7; d) threshold value 0.75

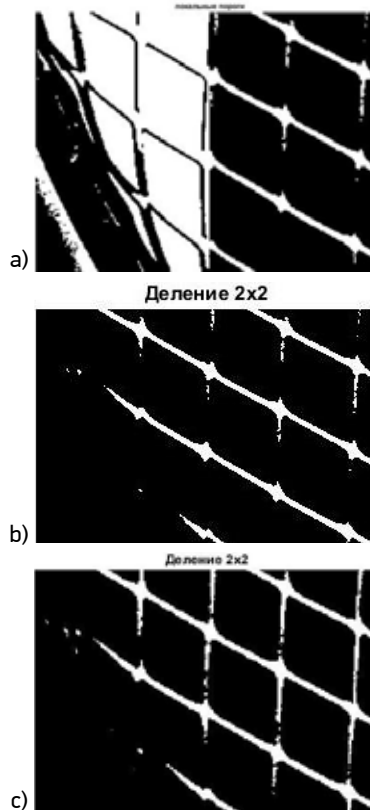


Рис. 10. Фрагмент 3.1 после деления на 4 подфрагмента и их обработки: а) значениями локальных порогов бинаризации; б) значением порога 0,7; в) значением порога 0,67

Fig. 10. Fragment 3.1 after dividing into 4 subfragments and processing them: a) values of local binarization thresholds; b) threshold value 0.7; c) threshold value 0.67

В задаче улучшения визуального качества изображений перед бинаризацией используются, как правило, два типа измерения среднего и дисперсии. *Глобальные* среднее и дисперсия вычисляются по всему изображению и обычно применяются для общего выравнивания интенсивности и контраста изображения в целом [8]. Лучше всего применять эти два параметра в методах локального улучшения, где *локальные* среднее и дисперсия используются как основа для проведения изменений, зависящих от характеристик

окрестности вокруг каждого элемента изображения.

В таблицах 1 и 2 даны статистические характеристики 16 фрагментов изображения. *TR* – порог бинаризации, определенный методом Отсу [9], *S1* – мера средней яркости, *S2* – значение дисперсии (мера средней контрастности), *S3* – значение гладкости, *S4* – третий момент (характеризует асимметричность гистограммы), *S5* – мера равномерности [10–12] (максимальна для постоянной яркости), *S6* – мера случайности (энтропия).

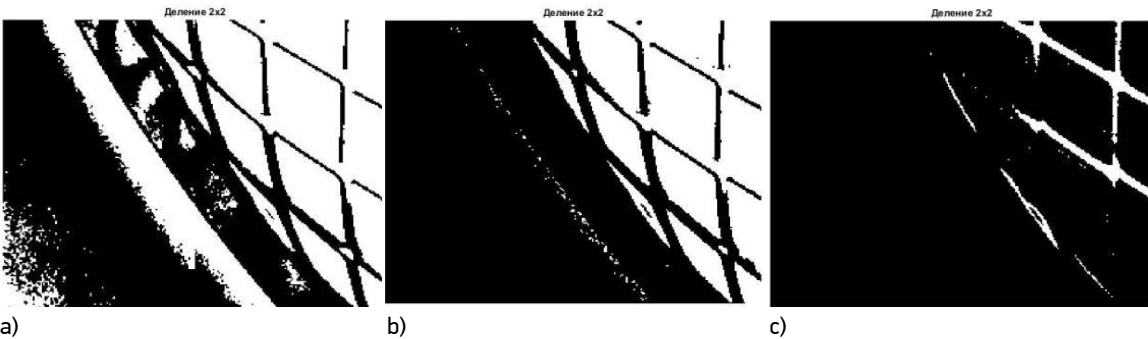


Рис. 11. Фрагмент 4.1 после деления на 4 подфрагмента и их обработки: а) значениями локальных порогов бинаризации; б) значением порога 0,5; с) значением порога 0,6

Fig. 11. Fragment 4.1 after dividing into 4 subfragments and processing them: a) values of local binarization thresholds; b) threshold value 0.5; c) threshold value 0.6

Таблица 1. Статистические данные для 16 фрагментов исходного изображения

Table 1. Statistics for 16 slices of the original image

(1.1) <i>TR</i> =0,7373 <i>S1</i> =164,4248 <i>S2</i> =29,5729 <i>S3</i> =0,0133 <i>S4</i> =0,0133 <i>S5</i> =0,0892 <i>S6</i> =5,6281	(1.2) <i>TR</i> =0,7412 <i>S1</i> =164,3555 <i>S2</i> =29,6354 <i>S3</i> =0,0133 <i>S4</i> =0,1283 <i>S5</i> =0,0266 <i>S6</i> =5,992	(1.3) <i>TR</i> =0,635 <i>S1</i> =160,8418 <i>S2</i> =31,8249 <i>S3</i> =0,0153 <i>S4</i> =0,2235 <i>S5</i> =0,0127 <i>S6</i> =6,6898	(1.4) <i>TR</i> =0,6235 <i>S1</i> =168,922 <i>S2</i> =33,4944 <i>S3</i> =0,017 <i>S4</i> =0,3248 <i>S5</i> =0,0107 <i>S6</i> =6,8209
(2.1) <i>TR</i> =0,7176 <i>S1</i> =155,983 <i>S2</i> =31,2786 <i>S3</i> =0,0148 <i>S4</i> =0,2155 <i>S5</i> =0,0338 <i>S6</i> =5,8348	(2.2) <i>TR</i> =0,7059 <i>S1</i> =147,5321 <i>S2</i> =33,2651 <i>S3</i> =0,0167 <i>S4</i> =0,717 <i>S5</i> =0,0251 <i>S6</i> =6,1029	(2.3) <i>TR</i> =0,6471 <i>S1</i> =122,6785 <i>S2</i> =39,6946 <i>S3</i> =0,0237 <i>S4</i> =1,9426 <i>S5</i> =0,0152 <i>S6</i> =6,4953	(2.4) <i>TR</i> =0,6627 <i>S1</i> =136,0052 <i>S2</i> =34,4619 <i>S3</i> =0,0179 <i>S4</i> =0,9956 <i>S5</i> =0,0131 <i>S6</i> =6,6738

Окончание таблицы 1

(3.1)	(3.2)	(3.3)	(3.4)
TR=0,4549	TR=0,6941	TR=0,6118	TR=0,6549
S1=134,9564	S1=139,5615	S1=110,6898	S1=123,0663
S2=40,9027	S2=35,8683	S2=44,222	S2=37,6822
S3=0,0251	S3=0,0194	S3=0,0292	S3=0,0214
S4=0,0901	S4=1,0905	S4=2,8948	S4=1,7552
S5=0,027	S5=0,025	S5=0,0278	S5=0,0155
S6=6,2873	S6=6,0554	S6=6,0305	S6=6,4189
(4.1)	(4.2)	(4.3)	(4.4)
TR=0,4392	TR=0,7176	TR=0,6745	TR=0,7294
S1=105,0664	S1=141,9051	S1=129,6782	S1=141,541
S2=35,1062	S2=35,9886	S2=39,3711	S2=35,6532
S3=0,0186	S3=0,0195	S3=0,0233	S3=0,0192
S4=0,5193	S4=0,9485	S4=1,501	S4=0,9682
S5=0,0144	S5=0,0333	S5=0,0137	S5=0,0132
S6=6,5557	S6=5,8184	S6=6,6378	S6=6,6501

Таблица 2. Статистические характеристики 16 фрагментов изображения при их последующем делении на 4 фрагмента

Table 2. Statistical characteristics of 16 image fragments with their subsequent division into 4 fragments

(1.1)–(1.1)	(1.1)–(1.2)	(1.2)–(1.1)	(1.2)–(1.2)	(1.3)–(1.1)	(1.3)–(1.2)	(1.4)–(1.1)	(1.4)–(1.2)
TR=0,7451	TR=0,7451	TR=0,7412	TR=0,7569	TR=0,5569	TR=0,5647	TR=0,5961	TR=0,7843
S1=167,64,96	S1=165,9025	S1=167,1325	S1=168,3286	S1=170,9555	S1=170,8	S1=159,4788	S1=200,7873
S2=30,0928	S2=27,936	S2=29,1908	S2=27,5348	S2=27,3245	S2=31,2115	S2=30,0015	S2=8,4099
S3=0,0137	S3=0,0119	S3=0,0129	S3=0,0115	S3=0,0114	S3=0,0148	S3=0,0137	S3=0,0011
S4=0,0925	S4=0,0127	S4=0,0042	S4=–0,005	S4=–0,0471	S4=–0,1213	S4=–0,1802	S4=–0,0017
S5=0,047	S5=0,0581	S5=0,0499	S5=0,0436	S5=0,0326	S5=0,0203	S5=0,0155	S5=0,0333
S6=5,5232	S6=5,2344	S6=5,49	S6=5,4871	S6=5,7173	S6=6,25	S6=6,3857	S6=5,0944
(1.1)–(2.1)	(1.1)–(2.2)	(1.2)–(2.1)	(1.2)–(2.2)	(1.3)–(2.1)	(1.3)–(2.2)	(1.4)–(2.1)	(1.4)–(2.2)
TR=0,7333	TR=0,7294	TR=0,7333	TR=0,7373	TR=0,7294	TR=0,7098	TR=0,5294	TR=0,6235
S1=162,7450	S1=161,3982	S1=162,4671	S1=159,4939	S1=154,2814	S1=147,3304	S1=138,0768	S1=177,3451
S2=29,8031	S2=29,9905	S2=31,5501	S2=29,2686	S2=30,96	S2=30,7396	S2=32,6385	S2=18,0128
S3=0,0135	S3=0,0136	S3=0,0151	S3=0,0130	S3=0,0145	S3=0,0143	S3=0,0161	S3=0,005
S4=0,148	S4=0,1098	S4=0,2388	S4=0,3066	S4=0,5022	S4=0,6547	S4=0,663	S4=0,1448
S5=0,0525	S5=0,048	S5=0,0375	S5=0,0415	S5=0,0201	S5=0,0191	S5=0,0134	S5=0,0236
S6=5,3827	S6=5,4877	S6=5,7233	S6=5,5085	S6=6,2102	S6=6,2347	S6=6,6157	S6=5,7647
(2.1)–(1.1)	(2.1)–(1.2)	(2.2)–(1.1)	(2.2)–(1.2)	(2.3)–(1.1)	(2.3)–(1.2)	(2.4)–(1.1)	(2.4)–(1.2)
TR=0,7255	TR=0,7098	TR=0,7137	TR=0,7137	TR=0,6784	TR=0,6588	TR=0,6863	TR=0,5333
S1=157,7141	S1=156,9901	S1=153,2858	S1=149,4421	S1=133,1489	S1=126,6196	S1=139,8114	S1=136,1348
S2=28,0491	S2=32,2384	S2=31,1089	S2=32,1084	S2=36,6414	S2=36,7146	S2=35,2955	S2=31,5914
S3=0,012	S3=0,0157	S3=0,0147	S3=0,0156	S3=0,0202	S3=0,0203	S3=0,0188	S3=0,0151
S4=0,0566	S4=0,2593	S4=0,483	S4=0,6857	S4=1,4589	S4=1,7037	S4=1,1112	S4=0,4662
S5=0,0547	S5=0,0446	S5=0,0445	S5=0,0333	S5=0,0182	S5=0,0255	S5=0,0204	S5=0,0137
S6=5,3147	S6=5,6228	S6=5,4817	S6=5,7931	S6=6,3126	S6=5,9822	S6=6,3162	S6=6,6524

(2.1)–(2.1)	(2.1)–(2.2)	(2.2)–(2.1)	(2.2)–(2.2)	(2.3)–(2.1)	(2.3)–(2.2)	(2.4)–(2.1)	(2.4)–(2.2)
TR=0,7255	TR=0,7098	TR=0,702	TR=0,6902	TR=0,6353	TR=0,6196	TR=0,651	TR=0,702
S1=155,433	S1=153,7948	S1=147,1763	S1=140,2242	S1=118,7857	S1=112,1598	S1=125,6315	S1=142,4431
S2=30,0276	S2=34,3013	S2=32,7133	S2=35,5967	S2=41,1335	S2=40,8554	S2=37,0408	S2=31,1625
S3=0,0137	S3=0,0178	S3=0,0162	S3=0,0191	S3=0,0254	S3=0,025	S3=0,0207	S3=0,0147
S4=0,0933	S4=0,438	S4=0,6722	S4=1,127	S4=2,2925	S4=2,5356	S4=1,7203	S4=0,8404
S5=0,0446	S5=0,0438	S5=0,0537	S5=0,0235	S5=0,0187	S5=0,0297	S5=0,0224	S5=0,0191
S6=5,5379	S6=5,6521	S6=5,3131	S6=6,063	S6=6,2552	S6=5,7912	S6=6,1006	S6=6,2473
(3.1)–(1.1)	(3.1)–(1.2)	(3.2)–(1.1)	(3.2)–(1.2)	(3.3)–(1.1)	(3.3)–(1.2)	(3.4)–(1.1)	(3.4)–(1.2)
TR=0,4471	TR=0,7216	TR=0,702	TR=0,6784	TR=0,6235	TR=0,5961	TR=0,6275	TR=0,6745
S1=132,4204	S1=150,2692	S1=144,3052	S1=135,2855	S1=113,1126	S1=107,0717	S1=113,7704	S1=130,7896
S2=42,899	S2=33,3617	S2=33,4592	S2=36,9497	S2=43,2643	S2=45,6157	S2=39,9288	S2=34,0917
S3=0,0275	S3=0,0168	S3=0,0169	S3=0,0206	S3=0,028	S3=0,031	S3=0,0239	S3=0,0176
S4=–0,2794	S4=0,5192	S4=0,7743	S4=1,3874	S4=2,7307	S4=3,2598	S4=2,4252	S4=1,3906
S5=0,0296	S5=0,05192	S5=0,0545	S5=0,0257	S5=0,0227	S5=0,049	S5=0,0303	S5=0,0216
S6=6,1869	S6=5,433	S6=5,2969	S6=5,9475	S6=6,0934	S6=5,377	S6=5,744	S6=6,0624
(3.1)–(2.1)	(3.1)–(2.2)	(3.2)–(2.1)	(3.2)–(2.2)	(3.3)–(2.1)	(3.3)–(2.2)	(3.4)–(2.1)	(3.4)–(2.2)
TR=0,4392	TR=0,7216	TR=0,7176	TR=0,6824	TR=0,6275	TR=0,5961	TR=0,6314	TR=0,6941
S1=108,0695	S1=149,0667	S1=143,6807	S1=134,9749	S1=117,4891	S1=105,0859	S1=114,1285	S1=133,5766
S2=38,3417	S2=33,2535	S2=35,0603	S2=36,7877	S2=42,8767	S2=43,9858	S2=38,149	S2=33,6374
S3=0,3943	S3=0,4598	S3=0,0186	S3=0,0204	S3=0,0275	S3=0,0289	S3=0,0219	S3=0,0171
S4=0,3943	S4=0,4598	S4=0,9051	S4=1,3722	S4=2,4787	S4=3,1702	S4=2,212	S4=1,3123
S5=0,0151	S5=0,0563	S5=0,0448	S5=0,0341	S5=0,0192	S5=0,0533	S5=0,0274	S5=0,0242
S6=6,6254	S6=5,3569	S6=5,4726	S6=5,7015	S6=6,2829	S6=5,2258	S6=5,8321	S6=5,9635
(4.1)–(1.1)	(4.1)–(1.2)	(4.2)–(1.1)	(4.2)–(1.2)	(4.3)–(1.1)	(4.3)–(1.2)	(4.4)–(1.1)	(4.4)–(1.2)
TR=0,3725	TR=0,4314	TR=0,7255	TR=0,7020	TR=0,6471	TR=0,6314	TR=0,6627	TR=0,7294
S1=88,433	S1=136,4618	S1=144,4223	S1=137,4308	S1=126,1514	S1=114,0765	S1=124,2848	S1=141,1995
S2=22,5158	S2=36,4618	S2=35,0518	S2=36,7108	S2=40,1824	S2=41,1019	S2=37,4791	S2=32,7921
S3=0,0077	S3=0,02	S3=0,0185	S3=0,0203	S3=0,0242	S3=0,0253	S3=0,0211	S3=0,0163
S4=0,116	S4=–0,1646	S4=0,8883	S4=1,2503	S4=1,9306	S4=2,5181	S4=1,8276	S4=1,0542
S5=0,0165	S5=0,0463	S5=0,0504	S5=0,0376	S5=0,0226	S5=0,0216	S5=0,0197	S5=0,0267
S6=6,2309	S6=5,6629	S6=5,3566	S6=5,6393	S6=6,1951	S6=6,0927	S6=6,1936	S6=5,9104
(4.1)–(2.1)	(4.1)–(2.2)	(4.2)–(2.1)	(4.2)–(2.2)	(4.3)–(2.1)	(4.3)–(2.2)	(4.4)–(2.1)	(4.4)–(2.2)
TR=0,3647	TR=0,4235	TR=0,7176	TR=0,7176	TR=0,7137	TR=0,7137	TR=0,7373	TR=0,7647
S1=86,6765	S1=108,6942	S1=143,8236	S1=141,9435	S1=139,8818	S1=138,6030	S1=144,4931	S1=156,1864
S2=12,3583	S2=36,4271	S2=34,8106	S2=36,9135	S2=35,5981	S2=34,4894	S2=32,9811	S2=31,5555
S3=0,0023	S3=0,02	S3=0,0183	S3=0,0205	S3=0,0191	S3=0,018	S3=0,0165	S3=0,0151
S4=0,0282	S4=0,1581	S4=0,6643	S4=1,0179	S4=1,1496	S4=34,4894	S4=0,08938	S4=0,4418
S5=0,279	S5=0,0153	S5=0,0537	S5=0,0474	S5=0,0304	S5=1,1064	S5=0,0245	S5=0,224
S6=5,47	S6=6,5948	S6=5,3438	S6=5,5055	S6=5,84	S6=6,0477	S6=5,9726	S6=6,0404

Можно заметить, что на проблемных участках значение порога бинаризации ниже глобального порога по всему изображению.

Как видно из таблиц 1 и 2, какой-либо прямой зависимости порогового значения бинаризации от статистических характеристик нет. Единственное, что можно использовать, – мера средней яркости. На тех фрагментах, где ее значение отличается от среднего по всем фрагментам, порог бинаризации определен некорректно.

Если выставить значение порога выше, ближе к среднему значению по всем фрагментам, то в результате бинаризация будет проведена корректно. Таким образом, искомое значение порога бинаризации $TR_{\text{ист}}$ для каждого из 16 фрагментов изображения определяется по следующей формуле:

$$TR_{\text{ист}} = TR_{\text{фрагм}} + \Delta, \quad (4)$$

где $TR_{\text{фрагм}}$ – найденное значение порога бинаризации фрагмента с помощью функции *gray-tresh* по методу Отсу;

Δ – эмпирически получаемая величина, на которую необходимо увеличить значение $TR_{\text{фрагм}}$.

Соответственно, значения $TR_{\text{фрагм}}$ для фрагментов изображения 1.3, 3.1, 4.1 равны 0,635; 0,4549 и 0,4392. Значения $TR_{\text{ист}}$ полученные эмпирически, равны 0,75; 0,67 и 0,6. Следовательно, значения Δ равны 0,115; 0,2151 и 0,1608.

Деление изображения на 16 фрагментов является максимально допустимым. С увеличением числа фрагментов на бинаризованном изображении начинают явно преобладать перепады яркостей соседних фрагментов и дальнейший анализ визуальной информации становится затруднителен.

Список литературы

1. Лошкарев А. С. Сегментация изображений с помощью кластеризации // Инфокоммуникационные технологии. 2017. Т. 15. № 4. С. 388–393. DOI: 10.18469/ikt.2017.15.4.11.
2. Otsu N. A threshold selection method from gray-level histograms // IEEE Transactions on Systems, Man, and Cybernetics. 1979. Vol. 9. No. 1. P. 62–66. DOI: 10.1109/TSMC.1979.4310076.
3. Шапиро Л., Стокман Дж. Компьютерное зрение: учеб. пособие. – М.: Математика, 2013. – 752 с.

Заключение

В данной работе авторами подробно рассмотрены параметры, которые можно использовать для повышения качества бинаризации анализируемого изображения.

При обработке многоградационных изображений или изображений с неравномерным распределением значений яркости пикселей необходимо учитывать возможные потери информации, необходимой для дальнейшего распознавания образов. В процессе бинаризации общим глобальным порогом могут быть утеряны важные элементы, такие как часть номерного знака, часть контура объекта, на которых перепады яркости выражены очень сильно. Либо, наоборот, возможно наращивание областей и их ошибочное прибавление к объектам интереса. Во избежание таких ситуаций необходимо делить изображение на отдельные фрагменты и бинаризовать каждый из них как самостоятельное изображение, тем самым уменьшая значение дисперсии распределения яркости пикселей, что дает более достоверные результаты и способствует корректному распознаванию образов в дальнейшем.

Авторами предложен способ повышения качества бинаризации изображения при предобработке путем деления на фрагменты и их локальной бинаризации по алгоритму Отсу. Количество фрагментов обусловлено тем фактом, что дисперсия распределения яркости пикселей гораздо слабее выражена на меньшем количестве фрагментов, например при делении на 4, 8 или 12 фрагментов. Эмпирическим путем получено выражение поиска значения порога бинаризации для каждого из фрагментов изображения, повышающее достоверность обрабатываемой информации.

4. Бирюков Е. Д., Иванов С. Ю. Способы выделения объекта при компьютерном распознавании статических зрительных образов // Новые информационные технологии в автоматизированных системах. 2010. № 13. С. 163–168.
5. Лошкарев А. С., Новосельцев С. Е. Методы сегментации изображений // Материалы III научно-практической всероссийской конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук». 2017. С. 443–447.
6. Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATLAB / пер. с англ. В. В. Чепыжова. – М.: Техносфера, 2006. – 616 с.
7. Bernsen J. Dynamic thresholding of grey-level images // Proceedings of the International Conference on Pattern Recognition (ICPR'86). 1986. P. 1251–1255.
8. Яне Б. Цифровая обработка изображений. – М.: Техносфера, 2007. – 584 с.
9. Kapur J. N., Sahoo P. K., Wong A. K. C. A new method for gray-level picture thresholding using the entropy of the histogram // Computer Vision Graphics and Image Processing. 1985. Vol. 29. No. 3. P. 273–285. DOI: 10.1016/0734-189X(85)90125-2.
10. Gatos B., Pratikakis I., Perantonis S. J. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information // Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008). 2008. P. 1–4. DOI: 10.1109/ICPR.2008.4761534.
11. Kulyas M., Kulyas O., Loshkarev A. The use of global image characteristics for neural network pattern recognitions // Proceedings of the 14th International Scientific and Technical Conference on Optical Technologies in Telecommunications. 2017. Vol. 10342. Article 1034211. DOI: 10.1117/12.2270475.
12. Методы компьютерной обработки изображений / под ред. В. А. Сойфера. – М.: ФИЗМАТЛИТ, 2003. – 784 с.

Сведения об авторах

Лошкарев Алексей Сергеевич, ORCID 0000-0001-8598-8908, старший преподаватель, кафедра информационных систем и технологий, Поволжский государственный университет телекоммуникаций и информатики, Самара, lozhkarev-as@mail.ru

Тимофеев Илья Алексеевич, ORCID 0000-0003-0476-8378, магистрант, кафедра информационных систем и технологий, Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия, timofeev_ia@mail.ru

Статья поступила 06.08.2021, рассмотрена 18.08.2021, принята 12.10.2021

References

1. Lozhkarev A. S. Image segmentation using clusterization. Infokommunikacionnye tehnologii, 2017, vol.15, no.4, pp.388-393 (in Russian). DOI: 10.18469/ikt.2017.15.4.11.
2. Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 1979, vol.9, no.1, pp.62-66. DOI: 10.1109/TSMC.1979.4310076.
3. Shapiro L., Stockman J. *Komp'yuternoe zrenie: ucheb. posobie* [Computer vision: textbook manual]. Moscow, Matematika Publ., 2013, 752 p.
4. Biryukov E. D., Ivanov S. Yu. *Sposoby vydeleniya ob"ekta pri komp'yuternom raspoznavanii staticheskikh zritel'nykh obrazov* [Methods of object isolation in computer recognition of static visual images]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2010, no.13, pp.163-168.
5. Lozhkarev A. S., Novosetcev S. E. *Metody segmentatsii izobrazhenii* [Image segmentation methods]. *Materialy III nauchno-prakticheskoi vserossiiskoi konferentsii (shkoly-seminara) molodykh uchennykh «Prikladnaya matematika i informatika: sovremennye issledovaniya v oblasti estestvennykh i tekhnicheskikh nauk»* [Materials of the III scientific-practical all-Russian conference (school-seminar) of young scientists "Applied mathematics and informatics: modern research in the field of natural and technical sciences"], 2017, pp.443-447.
6. Gonsales R., Vuds R., Eddins S. *Tsifrovaya obrabotka izobrazheniy v srede MATLAB* [Digital Image Processing using MATLAB]. Transl. from English V. V. Chepyzhova. Moscow, Tekhnosfera Publ., 2006, 616 p.
7. Bernsen J. Dynamic thresholding of grey-level images. Proceedings of the International Conference on Pattern Recognition (ICPR'86), 1986, pp.1251-1255.
8. Yane B. Digital image processing. Moscow, Tekhnosfera Publ., 2007, 584 p.
9. Kapur J. N., Sahoo P. K., Wong A. K. C. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision Graphics and Image Processing, 1985, vol.29, no.3, pp.273-285. DOI: 10.1016/0734-189X(85)90125-2.

10. Gatos B., Pratikakis I., Perantonis S. J. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), 2008, pp.1-4. DOI: 10.1109/ICPR.2008.4761534.
11. Kulyas M., Kulyas O., Loshkarev A. The use of global image characteristics for neural network pattern recognitions // Proceedings of the 14th International Scientific and Technical Conference on Optical Technologies in Telecommunications, 2017, vol.10342, article 1034211. DOI: 10.1117/12.2270475.
12. *Metody komp'yuternoi obrabotki izobrazhenii* [Methods of computer image processing]. Ed. by V. A. Soifer. Moscow, FIZMATLIT Publ., 2003, 784 p.

About the authors

Alexey S. Lozhkarev, ORCID 0000-0001-8598-8908, Senior Lecturer, Information Systems and Technologies Department, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia, lozhkarev-as@mail.ru
Il'ya A. Timofeev, ORCID 0000-0003-0476-8378, Master's Student, Information Systems and Technologies Department, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia, timofeev_ia@mail.ru

Received 06.08.2021, reviewed 18.08.2021, accepted 12.10.2021

Программное извлечение данных из word-документов на основе ситуационно-ориентированного подхода

В. В. Миронов¹, А. С. Гусаренко^{1}, Н. И. Юсупова¹*

¹ Уфимский государственный авиационный технический университет, Уфа, Россия

**gusarenko@ugatu.su*

Аннотация. В статье рассмотрены вопросы применения ситуационно-ориентированного подхода для программной обработки word-документов. Рассматриваемые документы подготавливаются пользователем в среде текстового процессора Microsoft Word или его аналогов и используются в дальнейшем как источники данных. Открытость форматов Office Open XML и Open Document Format позволила применить концепцию виртуальных документов, отображаемых на ZIP-архивы, для программного доступа к XML-компонентам word-документов в ситуационно-ориентированной среде. Обоснована важность выработки предварительных соглашений относительно размещения информации в документе для последующего поиска и извлечения, например, с помощью заранее подготовленных шаблонов-заготовок. Для форматов DOCX и ODT рассмотрено использование ключевых фраз, закладок, элементов управления контентом, пользовательских XML-компонентов для организации извлечения введенных данных. Для каждого варианта построены древовидные модели доступа к извлекаемым данным, а также соответствующие XPath-выражения. Отмечено, что использование того или иного варианта зависит от функциональных возможностей и ограничений текстового процессора и характеризуется различной сложностью разработки шаблона-заготовки, внесения данных пользователем и программирования извлечения данных. Рассмотрен практический пример обработки метаданных научной статьи, подготовленной в среде Microsoft Word для публикации в научном журнале. Примененное решение основано на занесении метаданных в статью с помощью размещенных в шаблоне-заготовке элементов управления контентом, привязанных к элементам пользовательского XML-компонента. Разработанная иерархическая ситуационная модель HSM обеспечивает извлечение XML-компонента, загрузку его в DOM-объект и XSLT-преобразования для получения результирующих данных: отчета об ошибках и JavaScript-кода для последующего использования извлеченных метаданных.

Ключевые слова: ситуационно-ориентированная база данных, иерархическая ситуационная модель, виртуальный документ, открытый текстовый формат, метаданные научной статьи, Open Journal System, DOCX, ODT

Для цитирования: *Миронов В. В., Гусаренко А. С., Юсупова Н. И. Программное извлечение данных из word-документов на основе ситуационно-ориентированного подхода // Прикладная информатика. 2021. Т. 16. № 6. С. 66–83. DOI: 10.37791/2687-0649-2021-16-6-66-83*

Software extract data from word-based documents situationally-oriented approach

V. Mironov¹, A. Gusarenko^{1*}, N. Yusupova¹

¹ Ufa State Aviation Technical University, Ufa, Russia

* gusarenko@ugatu.su

Abstract. The article discusses the use of situation-oriented approach to software processing word-documents. The documents under consideration are prepared by the user in the environment of the Microsoft Word processor or its analogs and are used in the future as data sources. The openness of the Office Open XML and Open Document Format made it possible to apply the concept of virtual documents mapped to ZIP archives for programmatic access to XML components of word documents in a situational environment. The importance of developing preliminary agreements regarding the placement of information in the document for subsequent search and retrieval, for example, using pre-prepared templates, is substantiated. For the DOCX and ODT formats, the article discusses the use of key phrases, bookmarks, content controls, custom XML components to organize the extraction of entered data. For each option, tree-like models of access to the extracted data, as well as the corresponding XPath expressions, are built. It is noted that the use of one or another option depends on the functionality and limitations of the word processor and is characterized by varying complexity of developing a blank template, entering data by the user and programming data extraction. The applied solution is based on entering metadata into the article using content controls placed in a stub template and bound to elements of a custom XML component. The developed hierarchical situational model of HSM provides extraction of an XML component, loading it into a DOM object and XSLT transformations to obtain the resulting data: an error report and JavaScript code for subsequent use of the extracted metadata.

Keywords: situationally-oriented database, hierarchical situational models, virtual document, open text format, the metadata of the scientific article, Open Journal System, DOCX, ODT

For citation: Mironov V., Gusarenko A., Yusupova N. Software extract data from word-based documents situationally-oriented approach. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.66-83 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-66-83

Введение

Программное извлечение тех или иных текстовых данных из офисных документов – информационно-технологическая задача, которая встречается достаточно часто. В том числе применительно к документам, созданным в среде текстового процессора Microsoft Word и его аналогов, таких как OpenOffice Writer, LibreOffice Writer, отечественный МойОфис Текст, китайский WPS Office и другие. Подобные документы будем называть word-документами. В этой статье нас интересуют word-документы, которые имеют «двойное назначение»: во-первых, исходный красиво оформленный текст поли-

графического качества, во-вторых, источник текстовых данных для дальнейшей автоматизированной обработки. В качестве характерного примера можно привести процесс публикации статей в научном журнале. Авторы оформляют в электронном виде статью в соответствии с установленными требованиями и загружают ее на сайт журнала. Там статья проходит процедуру рецензирования, доработки, исправления, доводится до окончательного вида публикации. Кроме того, в ходе редакционного процесса авторы или редакторы вносят в базу данных многочисленные метаданные статьи, такие как заглавие, сведения об авторах, аннотацию, ключевые слова, спи-

сок литературы и другое, которые размещаются на сайте журнала и загружаются в библиографические базы. Как правило, метаданные уже присутствуют в тексте статьи, поэтому их ручное извлечение и занесение в базу представляют собой избыточную работу. Появляется естественное желание автоматизировать процесс извлечения метаданных из электронного документа. Таким образом, возникает задача программной обработки word-документов с целью извлечения текстовых данных. Для решения этой задачи нужно ответить на два вопроса: 1) как осуществить программный доступ к содержимому word-документа; 2) как программным путем найти в word-документе нужные сведения.

Возможность программной обработки word-документов обусловлена тем, что они основаны на открытых форматах, таких как формат DOCX (стандарт Office Open XML) и формат ODT (стандарт Open Document Format) [1–3]. В этих форматах электронный документ организован в виде ZIP-архива, в котором упакованы папки с XML-файлами. Таким образом, программа может открывать архив, извлекать из него нужный XML-файл и обрабатывать его нужным образом. Для этого, в принципе, можно использовать разные подходы: от непосредственной работы с XML-файлами до применения специальных программных средств. В статье данная задача решается на платформе ситуационно-ориентированных баз данных – информационного процессора, обеспечивающего обработку разнородных данных под управлением иерархической ситуационной модели [4, 5].

Поиск и идентификация нужной информации во внутреннем представлении word-документов предполагают некоторые предварительные соглашения о том, как вносится эта информация на этапе создания документа. В зависимости от особенностей структуры word-документа конкретная задача извлечения информации будет формулироваться по-разному. Например, может быть требование, чтобы аннотация научной статьи размеща-

лась в абзаце, который начинается с определенной фразы, скажем, «Резюме:». Тогда при извлечении информации необходимо найти в документе такой абзац и очистить его от лишних данных. В другом случае ключевая фраза в аннотации может отсутствовать, но установлено правило, что аннотация целиком занимает абзац, расположенный сразу после абзаца с заглавием статьи, и тогда задача состоит в нахождении абзаца, следующего за заглавием. Можно представить себе статью в свободном стиле, не предусматривающем строгих правил размещения аннотации. В этом случае требуется отыскать абзац или несколько последовательных абзацев, идентифицировав их по содержанию как аннотацию. Здесь рассматривается задача, предполагающая достаточно строгие априорные требования к размещению в документе извлекаемых данных, позволяющие однозначно эти данные идентифицировать при обработке документа в ситуационно-ориентированной среде.

Анализ публикаций по данной тематике свидетельствует об интересе к общим проблемам применения текстовых процессоров и программной обработке word-документов как в концептуальном плане информационного поиска и обработки текстов [6–8], так и в плане практического применения в разных предметных областях [9–11] и построения программных инструментов для решения конкретных задач [12–14]. Во многих найденных публикациях обсуждаются технические аспекты программной обработки word-документов на основе OLE Automation, VBA, Word Object Model и другого инструментария [15–20]. Ряд публикаций связан с программным заполнением данными шаблонов-заготовок word-документов [21–23] – это задача занесения данных, которая по своей сути является обратной по отношению к рассматриваемой задаче извлечения данных. Что касается ситуационно-ориентированного подхода, то следует отметить работу [24], в которой рассмотрена задача программного занесения данных в заготовки word-документов, а также

работу [25], где схожая задача решается применительно к документам векторной графики. Вместе с тем не удалось найти какие-либо публикации, посвященные извлечению из word-документов требуемой информации путем их программной обработки как текстовых файлов, структурированных XML-разметкой. Отсутствие проработанности соответствующих подходов и моделей обуславливает актуальность исследований в данном направлении.

Итак, объектом исследования в данной работе является внутренняя структура word-документов в плане размещения в них семантических данных, т. е. данных, вносимых пользователем и имеющих смысл для дальнейшей программной обработки. Предмет исследования – возможности организации ручного внесения и последующего программного извлечения этих данных на основе ситуационно-ориентированного подхода. Конечной целью при этом ставится повышение производительности таких информационных систем за счет автоматизации – сокращения ручного труда, связанного с извлечением/занесением данных. Для достижения этой цели ниже рассматриваются задачи:

- применения ситуационно-ориентированного подхода к обработке word-документов как архивов XML-компонентов;
- исследования специфики внутреннего XML-формата word-документов, существенной для идентификации извлекаемых данных;
- демонстрации работоспособности и эффективности предложенного подхода на примере практического извлечения метаданных из научной статьи и загрузки их в издательскую систему.

Обработка word-документов на основе ситуационно-ориентированного подхода

Ситуационно-ориентированные базы данных – это подход, развивающий идею polyglot persistence [26–28] – совместного использо-

вания разнородных данных в одном приложении. В основе этого подхода [27] лежит использование встроенной иерархической ситуационной модели для управления процессом обработки данных (model-driven approach), а также концепция виртуальных документов, отображаемых на реальные разнородные данные (рис. 1).

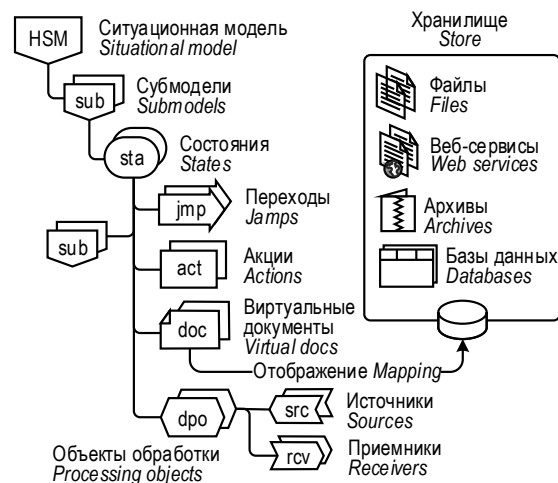


Рис. 1. Иерархическая ситуационная модель ситуационно-ориентированной базы данных [29]

Fig. 1. A hierarchical situational model of a situation-oriented database [29]

Иерархическая ситуационная модель (Hierarchical Situation Model, HSM) [29] задает в декларативной форме процесс обработки данных. Она представляет собой иерархию субмоделей sub, содержащих множество состояний sta, которые могут содержать другие субмодели sub, переходы jmp и акции act. Виртуальные документы doc вместе с обработчиками dpo обеспечивают унифицированную обработку разнородных внешних данных. Виртуальные документы могут отображаться на локальные файлы, удаленные веб-сервисы, архивы, реляционные базы данных. На основе информации, содержащейся в декларациях виртуальных документов, внешние данные загружаются в обработчики – объекты обработки данных dpo, где они подвергаются нужным преобразованиям.

Основные типы обработчиков – это DOM-объекты (Document Object Model – объектная модель документа), ориентированные на обработку XML-данных, и ассоциативные массивы, ориентированные на обработку JSON-данных [30].

В данной статье предлагается применение именно этого подхода, поскольку извлечение данных из word-документов, основанных на различных форматах, и последующее преобразование извлеченных данных в другой формат лежат в русле общей концепции единообразной обработки разнородных данных. Это, с одной стороны, распространит ситуационно-ориентированный подход на word-подобные документы, с другой – упростит саму обработку word-документов за счет применения декларативной иерархической ситуационной модели.

Отображение на word-документы

С учетом того, что в нашей задаче word-документ представляет собой ZIP-архив, важна, в первую очередь, возможность отображения виртуальных документов на архивы. А тот факт, что в папках архива размещены XML-файлы, обуславливает использование в качестве обработчиков DOM-объекты. При этом

нужно знать внутреннюю структуру как ZIP-архива word-документа, так и XML-файлов в нем. На рисунке 2 внутренняя структура задания виртуального документа иллюстрируется для двух форматов word-документов: для формата DOCX (рис. 2a), который используется в среде редактора Microsoft Word, и для формата ODT (рис. 2b), применяемого в редакторах OpenOffice Writer, LibreOffice Writer и других аналогов Word.

HSM-элемент `doc:word-doc` (рис. 2a) задает виртуальный документ, который отображается на файл `wd.docx`, указанный в атрибуте `path`, и который интерпретируется как ZIP-архив в соответствии со значением атрибута `type`. Аналогично HSM-элемент `doc:open-office` (рис. 2b) задает виртуальный документ, который отображается на файл `ood.odt`, который тоже интерпретируется как ZIP-архив. В формате DOCX основная разметка документа упакована в файле `document.xml`, размещенном в папке `word`, отображение на нее задает entry-элемент `ent:document`, вложенный в `doc:word-doc`. В формате ODT основная разметка документа упакована в файле `content.xml`, размещенном в корне архива, отображение на нее задает entry-элемент `ent:content`, вложенный в `doc:open-office`. Ссылаясь на

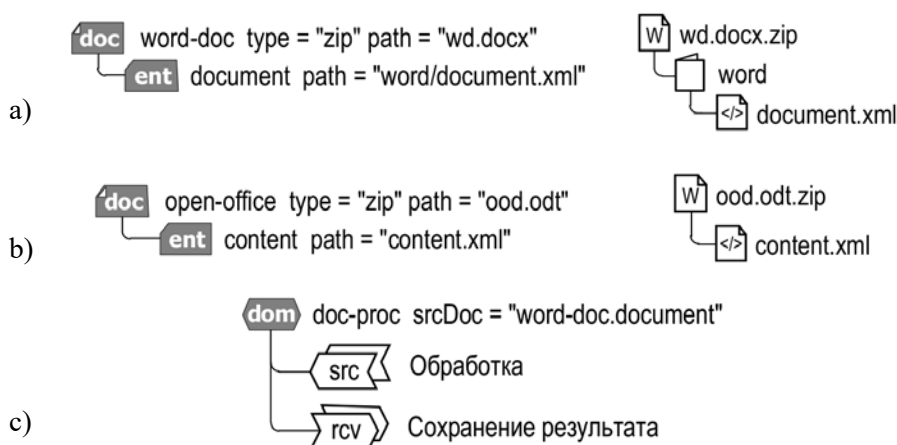


Рис. 2. Задание виртуального документа с отображением на форматы DOCX (a), ODT (b) и обработка в DOM-объекте (c)

Fig. 2. Defining a virtual document with mapping to DOCX (a) and ODT (b) formats and processing in a DOM object (c)

entry-элементы, получаем доступ к соответствующим XML-компонентам архива для обработки. Например, файл `document.xml`, извлеченный из word-документа, загружается в DOM-объект `dom:doc-proc` (рис. 2с). Дальнейшая обработка загруженного файла задается с помощью вложенных элементов-источников `src` и элементов-приемников `rcv`. С их помощью можно задать преобразование XML-документа и сохранить результат в том или ином формате. При этом доступны мощные средства манипуляции XML, предусмотренные в DOM, в том числе язык запросов XPath и язык трансформации XSLT.

Поиск извлекаемых данных в XML-дереве документа

Хотя для внутреннего представления word-документов применяются открытые стандартизованные форматы, которые подробно документированы, поиск данных в них является нетривиальной задачей. Это объясняется чрезвычайно сложной внутренней структурой XML-разметки, в которой собственно семантические сведения расчленены и перемешаны с данными, задающими внешнее представление документа (шрифты, отступы, стили, структурные элементы и т. п.). Поэтому необходимо исследование внутренней структуры word-документов для создания моделей и методов, ориентированных именно на поиск и идентификацию нужных данных. Такое исследование было проведено путем непосредственного изучения внутренней XML-разметки документа при различных способах задания внешнего представления данных. Результаты изложены ниже в этом разделе.

XML-компонент word-документа, загруженный в DOM-объект, доступен для программной обработки в виде XML-дерева, содержащего узлы XML-элементов, XML-атрибутов. Возникает задача отыскания нужных узлов XML-дерева, соответствующих извлекаемым данным word-документа. Для этого извлекаемые данные должны удовлет-

ворять требованию *идентифицируемости*, а именно при внесении данных на этапе создания word-документа пользователем должны соблюдаться требования, позволяющие в дальнейшем отыскать эти данные в процессе программной обработки документа. Предлагаются следующие приемы обеспечения идентифицируемости данных: ключевые фразы, закладки, элементы управления контентом, пользовательские XML-компоненты.

Использование ключевых фраз

Для обеспечения идентифицируемости этим методом в искомые данные вводят определенные ключевые слова или фразы, и тогда задача извлечения сводится к отысканию абзаца (нескольких абзацев), содержащих данную фразу. При этом нужно учитывать особенности внутренней XML-разметки, определяемые используемым форматом word-документа. В формате DOCX каждому абзацу соответствует элемент `W:p`, в который вложен один элемент (или несколько элементов) `W:t` со своим форматированием, который, в свою очередь, содержит элемент `W:t`. Здесь префикс `W` задает принадлежность XML-элемента пространству имен

<http://schemas.openxmlformats.org/wordprocessingml/2006/main>.

Текстовое содержимое абзаца в целом, таким образом, складывается из текстового содержимого всех вложенных элементов `W:t`. В формате ODT каждому абзацу соответствует элемент `text:p`, содержащий текстовые узлы непосредственно или в виде текстового содержимого элементов `text:span` со своим форматированием. Здесь префикс `text` задает принадлежность XML-элемента пространству имен

`urn:oasis:names:tc:opendocument:xmlns:text:1.0`.

Таким образом, чтобы извлечь текстовое содержимое абзаца, нужно собрать текст из всех внутренних текстовых узлов соответствующего XML-элемента `W:p` или `text:p`.

Текстовое содержимое абзаца, содержащего ключевой текст, скажем, «Резюме:», может быть извлечен из XML-дерева, например, с помощью такого XPath-выражения:

```
//p[contains(string(),"Резюме:")]string().
```

Использование закладок

При использовании этого метода пользователь выделяет в документе нужный участок текста с помощью так называемой закладки (bookmark). Закладки могут быть заранее размещены в шаблоне-заготовке word-документа, и тогда пользователь вносит данные в места, отмеченные соответствующими закладками. Поскольку каждой закладке присвоено уникальное имя, при обработке документа программа может найти и извлечь выделенный текст. Во внутренней XML-разметке каждой закладке соответствует два XML-элемента: начальный и конечный, которые вложены в элемент абзаца. В формате DOCX это XML-элементы `w:bookmarkStart` и `w:bookmarkEnd`, причем каждая закладка идентифицируется XML-атрибутом-идентификатором `w:id`, назначаемым автоматически, и атрибутом-именем `w:name`, назначаемым пользователем. В формате ODT применяются элементы `text:bookmark-start` и `text:bookmark-end`, которые идентифицируются атрибутом-именем `text:name`. Для извлечения текста абзаца,

содержащего закладку с именем «Резюме», может быть применено следующее XPath-выражение (формат DOCX):

```
//p[bookmarkStart/@name='Резюме']/string().
```

Набор текстовых узлов, выделенных закладкой внутри абзаца, может быть извлечен следующим выражением:

```
//r[preceding-sibling::bookmarkStart/@name='Резюме'
```

```
and following-sibling::bookmarkEnd]/string().
```

Использование элементов управления содержимым

Элементы управления содержимым (контентом) (content controls) – это объекты – контейнеры данных в документе, задающие определенное структурирование и форматирование. При использовании в шаблонах-заготовках документов такие элементы, как форматированный и неформатированный текст, списки и другие, удобны и при занесении данных пользователями, и при извлечении этих данных программами. На рисунке 3 приведены иерархические модели внутреннего представления элементов управления контентом в форматах DOCX и ODT. Модели построены в результате непосредственного исследования XML-разметки и содержат только те элементы и атрибуты, которые существенны в плане рассматриваемой задачи.

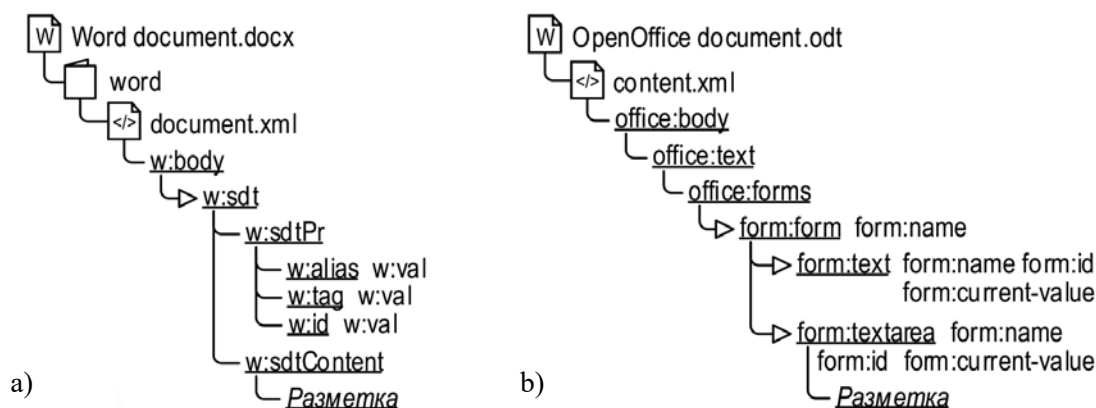


Рис. 3. Внутреннее представление элементов управления контентом в форматах DOCX (a) и ODT (b)

Fig. 3. Internal representation of content controls in DOCX (a) and ODT (b) formats

В формате DOCX (рис. 3а) элементы управления задаются с помощью XML-элементов `w:sdt` (structured document tags – теги структурированного документа), которые размещаются внутри элементов-абзацев `w:p`. Во вложенном элементе `w:sdtPr` размещены идентифицирующие данные: автоматически формируемый идентификатор `w:id` и задаваемые разработчиком имена `w:alias` и `w:tag`. По значениям имен имеется возможность программного извлечения содержимого элементов управления из вложенного XML-элемента `w:sdtContent`. Например, текстовое содержимое элемента с именем «Резюме» можно извлечь с помощью следующего XPath-выражения:

`//sdt[sdtPr/alias/@val='Резюме']/string()`.

В формате ODT (рис. 3b) текстовые элементы управления задаются с помощью XML-элементов `form:text` (текстовое поле без форматирования), `form:textarea` (текстовое поле с форматированием), которые размещаются внутри элементов форм `form:form`. Здесь префикс `form` задает принадлежность XML-элемента пространству имен `urn:oasis:names:tc:opendocument:xmlns:form:1.0`.

Идентифицирующие данные представлены в виде XML-атрибутов – автоматически формируемого идентификатора `form:id` и задаваемого разработчиком документа имени `form:name`. Текстовое содержимое элемента задается без форматирования в атрибуте `form:current-value`, отформатированный текст представлен в виде вложенной разметки элемента `form:textarea`. Таким образом, текстовое содержимое элемента с именем «Резюме» можно извлечь, например, с помощью следующего XPath-выражения:

`//textarea[@name='Резюме']/@current-value`.

Применение пользовательских XML-компонентов

Пользовательские XML-компоненты (custom XML parts – XML-части) – это XML-данные пользователя, внедренные

в документ. Элементы пользовательского XML-компонента могут быть привязаны к элементам управления содержимым документа, и тогда ввод или изменение данных в элементах управления будет синхронно отражаться в пользовательском XML-компоненте, и наоборот. На рисунке 4 приведены иерархические модели внутреннего представления пользовательских XML-

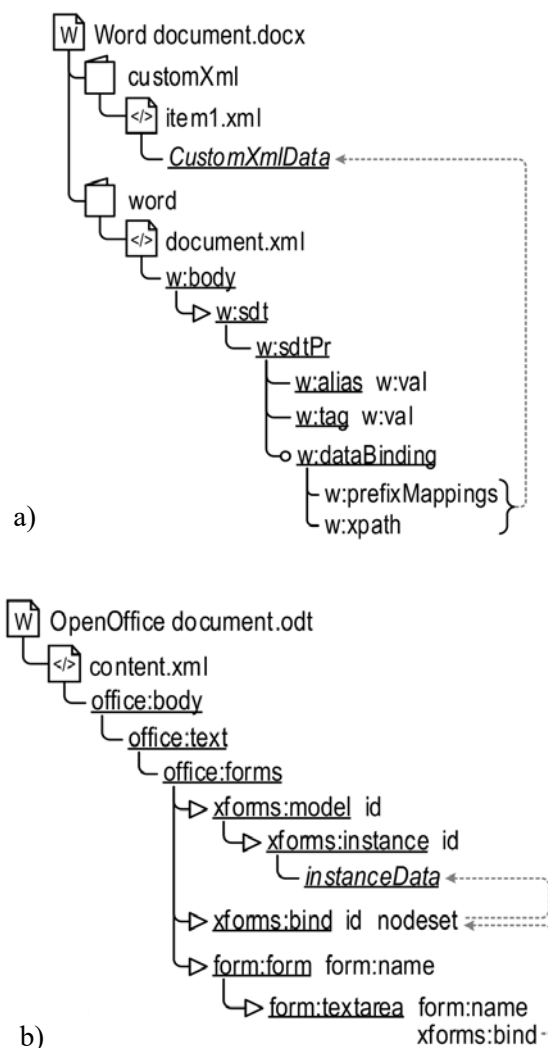


Рис. 4. Внутреннее представление пользовательских XML-частей в форматах DOCX (а) и ODT (б)

Fig. 4. Internal representation of custom XML parts in DOCX (a) and ODT (b) formats

частей в форматах DOCX и ODT, полученные путем непосредственного исследования XML-разметки.

В формате DOCX каждая пользовательская XML-часть упакована в ZIP-архиве word-документа в отдельном файле `itemn.xml` в папке `customXml` (рис. 4а). Связывание элемента управления с XML-элементом пользовательской части задается с помощью элемента `w:dataBinding`, размещенного в элементе `w:sdtPr` (атрибут `w:prefixMappings` ссылается на пользовательский XML-компонент, а атрибут `w:path` – на XML-элемент пользовательского компонента). Например, текстовый элемент управления с именем «Резюме» может быть связан с XML-элементом **Резюме** в пользовательском XML-компоненте. Последний может быть легко извлечен с помощью XPath-выражения `//Резюме`.

В формате ODT похожая функциональность обеспечивается с помощью так называемых XForm, размещаемых в основном файле разметки `content.xml` в элементах `xforms:instance`, вложенных в элементы `xforms:model` (рис. 4b). Здесь префикс `xforms` задает принадлежность XML-элемента пространству имен

<http://www.w3.org/2002/xforms>.

Связывание достигается через элементы `xforms:bind`: атрибут `xforms:bind` элемента управления ссылается на идентификатор `id` в `xforms:bind`, который через атрибут `nodeset` ссылается на вложенный XML-элемент в `xforms:instance`.

Выбор способа идентифицируемости

Таким образом, существует несколько способов обеспечения идентифицируемости извлекаемой текстовой информации в word-документах. Возможность использования того или иного способа зависит от функциональных ограничений применяемого текстового процессора (редактора). Наиболее полно в настоящее время рассмотренные возможности реализованы в Microsoft Word

версии 2013 и выше. Способы характеризуются различной сложностью с трех взаимосвязанных точек зрения: во-первых, с точки зрения *разработчика* шаблона-заготовки; во-вторых, с точки зрения *пользователя* в плане занесения данных в шаблон-заготовку; в-третьих, с точки зрения *программиста* в плане программного извлечения и очистки данных. Так, метод ключевых фраз не требует предварительных усилий в плане подготовки шаблона заполняемого документа. Однако для его реализации потребуются строгие правила для пользователя по заполнению документа, что порождает возможность ошибок. Остальные способы требуют определенных усилий по предварительной разметке заполняемых полей в шаблоне документа – размещению закладок, элементов управления. Использование элементов управления потенциально может создать удобства пользователю при заполнении данных за счет удобного интерфейса и контроля ошибок вводимых данных. Использование пользовательских XML-компонентов – наиболее трудоемкий подход в плане разработки шаблона: необходимо предварительно создать XML-файл с элементами заполняемых данных, внедрить его в шаблон в виде XML-компонента, разместить в отображаемой части документа элементы управления контентом и настроить привязку элементов управления к соответствующим элементам пользовательского XML-компонента. Извлечение данных в этом варианте – самое простое, поскольку нужные данные в виде пользовательского XML-компонента уже отделены от остальных частей документа.

Предложенные способы обеспечения идентифицируемости информации в word-документах, разработанные модели доступа в XML-компонентах и построенные XPath-выражения для извлечения данных характеризуются новизной и составляют значимость данной работы в плане методологии обработки семантической информации.

Практический пример: загрузка статей в издательскую систему

Практическая значимость полученных результатов состоит в том, что они способны служить основой информационной технологии программной обработки word-документов в различных областях. В качестве практического примера извлечения текстовых данных из word-документов на основе ситуационно-ориентированного подхода рассмотрим эту задачу применительно к процессу подготовки к публикации статей в научном журнале на платформе издательской системы Open Journal System (OJS) (рис. 5).

В ходе издательского процесса как сами статьи, так и их метаданные загружаются в систему и при начальной отправке, и в ходе доработки и исправления. При традиционном подходе эти процессы дублируются: на сайт загружается очередная версия статьи, содержащая в своем тексте в том числе метаданные, и, кроме того, метаданные заносятся отдельно (авторами и/или сотрудниками редакции). Занесение метаданных при этом выполняется вручную: очередной элемент метаданных копируется из статьи и вносится в соответствующее поле веб-интерфейса издательской системы. Ввиду большого числа элементов эта процедура достаточно трудоемкая. Поставлена цель: сократить трудоемкость

за счет автоматизированного извлечения метаданных из текста статьи и загрузки в систему. При этом предполагается также автоматизировать проверку извлекаемых метаданных для выявления определенных ошибок.

Пользовательский XML-компонент

Рассматриваемая задача допускает применение заранее подготовленного хорошо структурированного шаблона, в который авторы статьи вносят как основной текст, так и метаданные. Поэтому для удобства авторов при внесении метаданных в документ и для простоты последующего извлечения в шаблоне применяются элементы управления контентом и пользовательский XML-компонент.

На рисунке 6 приведено дерево разработанного пользовательского XML-компонента, внедренного в шаблон в виде файла `item1.xml`. Корневой XML-элемент `metadata` содержит атрибуты `xmlns` и `version`, задающие соответственно идентификатор данного XML-компонента и его версию. XML-компонент содержит однозначные и многозначные элементы и агрегаты, текстовое содержимое которых соответствует тем или иным метаданным. Многозначный агрегат `автор` содержит элементы со сведениями о каждом соавторе статьи, многозначный элемент `ссылка` – библиографическую ссылку из списка литературы.

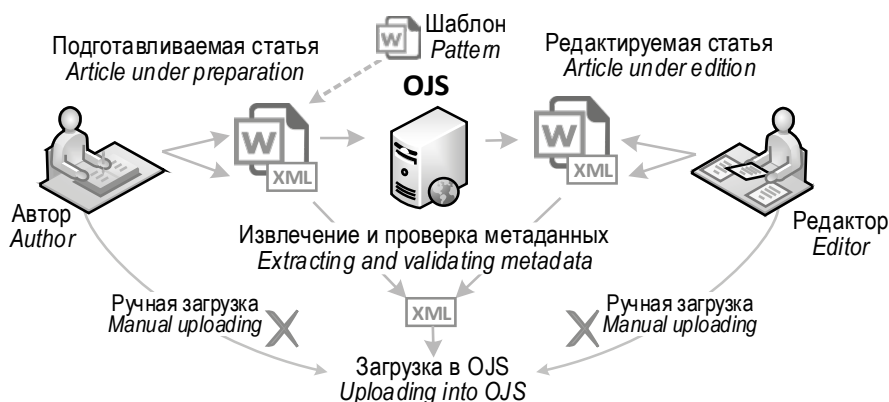


Рис. 5. Загрузка статей в OJS при подготовке и редактировании

Fig. 5. Uploading paperes to OJS during preparation and editing

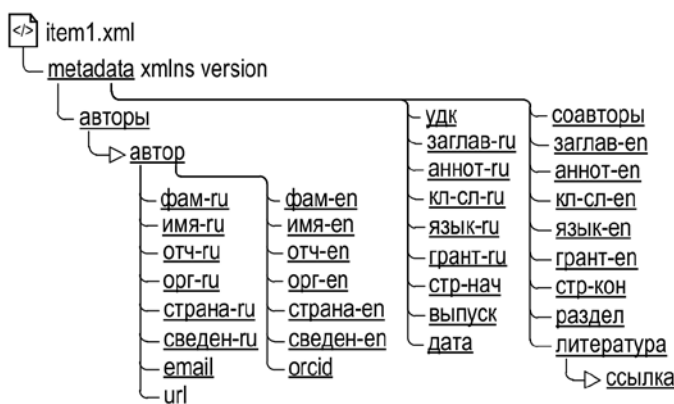


Рис. 6. Дерево пользовательского XML-компонента шаблона статьи

Fig. 6. Custom XML part tree of a paper pattern

Многие элементы метаданных представлены в двух вариантах — русском (имя с суффиксом -gu) и английском (с суффиксом -en).

Каждому элементу метаданных в документе соответствует привязанный к нему элемент управления содержимым, размещенный в соответствующем месте word-документа. Большинство элементов управления — это элементы типа «неформатированный текст», но используются и элементы других типов, такие как «раскрывающийся список» (позволяет выбрать один из фиксированных разделов для статьи) и «выбор даты» (позволяет задать дату подачи статьи, используя календарь). Для многозначных объектов **автор** и **ссылка** применена функциональность повторяемости, позволяющая добавлять и удалять экземпляры одного типа.

Иерархическая ситуационная модель

Процедура извлечения метаданных реализована в виде веб-приложения, серверная часть которого запускается HTTP-запросами, поступающими от клиентов. На рисунке 7 представлена HSM-модель **sub:Metadata-proc**, обеспечивающая извлечение и обработку метаданных из word-документа.

Модель содержит 5 состояний: **sta:begin** — начальная обработка HTTP-запроса; **sta:GET** — обработка запроса типа GET; **sta:POST** — обработка запроса типа POST; **sta:requestErr** —

обработка ситуации ошибочного запроса; **sta:uploadErr** — обработка ситуации ошибки при загрузке обрабатываемого документа. В начальном состоянии с помощью виртуального документа **doc:request** устанавливается доступ к параметрам HTTP-запроса, инициировавшего обработку модели **sub:Metadata-proc**, и далее с помощью переходов **jmp:requestErr** и **jmp:further** выполняется смена состояния для дальнейшей обработки. В состоянии **sta:GET** через виртуальный документ **doc:request** клиенту возвращается ответ на его HTTP-запрос, ответ содержит форму **form.htm** для загрузки word-документа. В состояниях обработки ошибок клиенту возвращаются соответствующие сообщения об ошибках. Собственно обработка word-документа задана в состоянии **sta:POST**, рассмотрим его подробнее.

В состоянии **sta:POST** выполняется обработка файла, загруженного клиентом с помощью формы, полученной в ответ на GET-запрос. Виртуальный документ **doc:word** задает доступ к загруженному файлу как к ZIP-архиву, при этом элемент **ent:customXml** указывает на интересующий нас файл **item1.xml**, упакованный в этом архиве. Атрибут **onErr** обеспечивает переход в состояние **sta:uploadErr** при обнаружении ошибок доступа к архиву (например, если загружен файл, не являющийся ZIP-архивом).

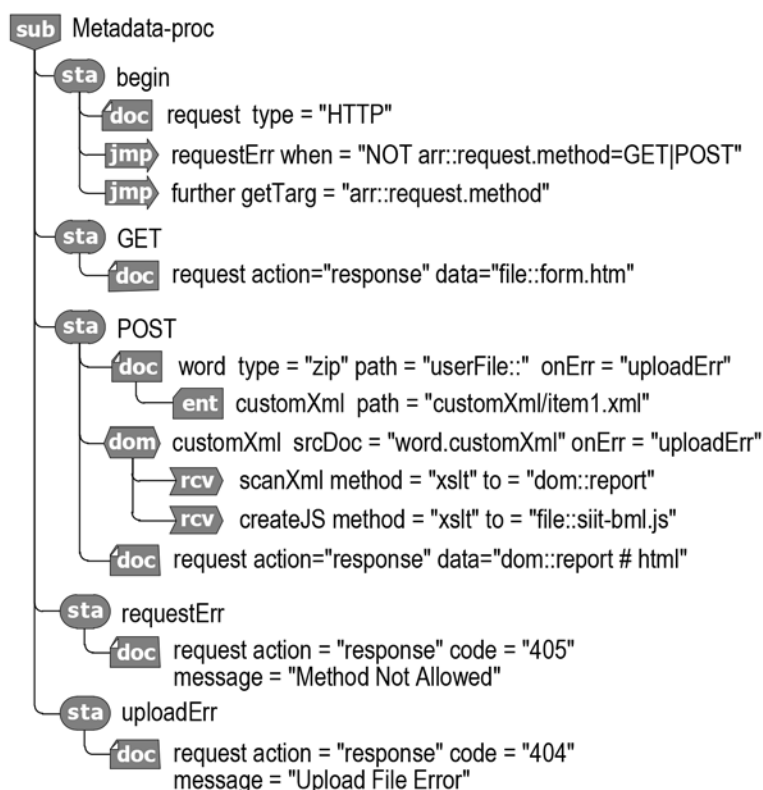


Рис. 7. Иерархическая ситуационная модель для извлечения и обработки метаданных

Fig. 7. Hierarchical situational model for extracting and processing metadata

Элемент `dom:customXml` обеспечивает загрузку файла `item1.xml` в соответствующий DOM-объект, в случае обнаружения ошибки (например, при отсутствии этого файла в архиве) выполняется переход в состояние `sta:uploadErr`.

Элемент-приемник `rcv:scanXml` выполняет формализованную проверку метаданных путем XSLT-преобразования XML-данных, загруженных в DOM-объект, в соответствии с таблицей стилей `scanXml.xsl`. Результат преобразования помещается в неявно создаваемый DOM-объект `dom:report`. Далее с помощью элемента `doc:request` с атрибутом `response` содержимое `dom:report` отправляется клиенту в формате HTML для отображения браузером.

Элемент-приемник `rcv:createJS` путем XSLT преобразует XML-данные, загруженные в DOM-объект, в JavaScript-код, кото-

рый сохраняется на сервере в файле `siit-bml.js`. В дальнейшем этот файл используется для загрузки метаданных на сайт издательской системы.

Использование извлеченных метаданных

Таким образом, результатами интерпретации HSM-модели `sub:Metadata-proc` являются отчет о проверке метаданных, отправляемых клиенту как ответ на HTTP-запрос, и JavaScript-код, сохраненный на сервере. Пример отчета о проверке метаданных представлен на рисунке 8. Отчет содержит сообщения о выявленных нарушениях формальных требований к метаданным и их согласованности между собой.

Сгенерированный JavaScript-код используется для внесения метаданных в формы издательской системы. На рисунке 9 приведен

Статья для СИИТ – извлечение и проверка метаданных

Файл «sample_article_rus_2021.docx» успешно загружен.

Выявленные нарушения:

- ** Число соавторов, указанных на 1-м листе (3), не равно числу авторов в разделе метаданных (4).
- ** Несоответствие 3-го соавтора (С. С. Сидоров) и автора в метаданных (Николай Николаевич Николаев).
- ** В аннотации слов 64, должно быть не менее 200.
- ** В переводе аннотации (abstract) обнаружены русские буквы.
- ** Ключевых слов – 1, должно быть не менее 5, разделенных точкой с запятой.
- ** Language – язык статьи по-английски – указан неверно, должно быть либо «Russian», либо «English».
- ** В списке литературы в 1-й ссылке в блоке транслитерации [...] обнаружены русские буквы.
- ** В списке литературы в 5-й ссылке в блоке транслитерации [...] не обнаружено указание «(in Russian)».

Рис. 8. Пример отчета о проверке метаданных

Fig. 8. Metadata validation report sample

Авторы	
Имя *	И.
Отчество	И.
Фамилия *	Иванов
Адрес (E-mail) *	ivanov@mail.ru
ORCID ID	http://orcid.org/0000-0002-1825-00
URL	elibrary.ru/author_profile.asp?author
Организация	ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)
Страна	Россия
Биографическое описание (Например, подразделение и должность)	Иванов Иван Иванович Проф. каф. автоматизированных систем управления. Дипл. радиофизик (Воронежск. гос. ун-т, 1975). Д-р техн. наук по упр. в техн. системах (защ. в УГАТУ, 1995). Иссл. в обл. иерархических моделей и ситуационного управления

Рис. 9. Фрагмент OJS-формы, заполненной метаданными

Fig. 9. A fragment of an OJS form filled with metadata

фрагмент подобной формы на платформе издательской системы Open Journal System (версия 2.4).

При традиционном заполнении пользователь вынужден вносить данные в каждое поле, набирая их вручную или копируя из word-документа. В предложенном решении заполнение полей формы выполняется автоматически с помощью букмарклета (bookmarklet). Букмарклет запускает для текущей страницы

веб-браузера функцию JavaScript, сгенерированную на основе извлеченных метаданных и сохраненную на сервере. Эта функция выполняет поиск полей формы по их идентификаторам и занесение в поля соответствующих метаданных. Простой расчет показывает, что для статьи с двумя соавторами необходимо вручную занести 48 элементов метаданных, в случае трех соавторов таких элементов будет уже 63, а в случае четырех – 78. Таким

образом, за счет автоматизации достигается существенное снижение трудоемкости этого процесса. Разработанные решения применяются в процессе издания научного журнала «Системная инженерия и информационные технологии» (<http://siit.ugatu.su>) в Уфимском государственном авиационном техническом университете.

Заключение

Итак, в данной статье рассмотрены вопросы применения ситуационно-ориентированного подхода для программной обработки word-документов. А именно документов, которые подготовлены пользователем в среде текстового процессора Microsoft Word или его аналогов и которые используются в дальнейшем как информационные источники данных. Открытость форматов Office Open XML и Open Document Format позволила применить концепцию виртуальных документов, отображаемых на ZIP-архивы, для программного доступа к XML-компонентам word-документов в ситуационно-ориентированной среде.

Обоснована важность выработки предварительных соглашений относительно размещения информации в документе для последующего поиска и извлечения, например, с помощью заранее подготовленных шаблонов-заготовок. Для форматов DOCX и ODT рассмотрено использование ключевых фраз, закладок, элементов управления контентом, пользовательских XML-компонентов для организации извлечения введенных данных. Для каждого варианта построены древовид-

ные модели доступа к извлекаемым данным, а также соответствующие XPath-выражения. Отмечено, что использование того или иного варианта зависит от функциональных возможностей и ограничений текстового процессора и характеризуется различной сложностью разработки шаблона-заготовки, внесения данных пользователем и программирования для извлечения данных.

Рассмотрен практический пример обработки метаданных научной статьи, подготовленной в среде Microsoft Word для публикации в научном журнале. Примененное решение основано на занесении метаданных в статью с помощью размещенных в шаблоне-заготовке элементов управления контентом, привязанных к элементам пользовательского XML-компонента. Разработанная иерархическая ситуационная модель HSM обеспечивает извлечение XML-компонента, загрузку его в DOM-объект и XSLT-преобразование для получения результирующих данных – отчета об ошибках и JavaScript-кода для последующего использования извлеченных метаданных.

Результаты нашли практическое применение при издании научного журнала: на основе программного извлечения метаданных научной статьи обеспечена автоматизированная первичная проверка их правильности и согласованности, а также автоматизированное занесение метаданных в издательскую систему, что существенно снижает трудоемкость по сравнению с традиционным ручным способом. Результаты могут быть легко распространены на другие журналы и издательские системы.

Список литературы

1. Hou X., Li N., Yang H., Liang Q. Comparison of wordprocessing document format in OOXML and ODF // Proceedings of the 6th International Conference on Semantics, Knowledge and Grids. 2010. P. 297–300. DOI: 10.1109/SKG.2010.44.
2. Schubert S. The next millennium document format // Proceedings of the ACM Symposium on Document Engineering. 2019. Article. 40. P. 1–4. DOI: 10.1145/3342558.3345419.
3. Roig J., Ribera M. Implementation of the OOXML standard since its approval until today // Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2020. P. 129–134. DOI: 10.1145/3439231.3440607.

4. Миронов В. В., Юсупова Н. И., Шакирова Г. Р. Ситуационно-ориентированные базы данных: внешние представления на основе XSL // Вестник Уфимского государственного авиационного технического университета. 2010. Т. 14. № 4 (39). С. 200–209.
5. Гусаренко А. С., Миронов В. В. Использование RESTful-сервисов в ситуационно-ориентированных базах данных // Вестник Уфимского государственного авиационного технического университета. 2015. Т. 19. № 1 (67). С. 232–239.
6. Kulkarni A., Shivananda A. Extracting the Data. – In book: Natural Language Processing Recipes. – Berkeley, CA: Apress, 2021. P. 1–29. DOI: 10.1007/978-1-4842-7351-7_1.
7. Болотова Л. С., Данчул А. Н., Новиков А. П., Сурхаев М. А., Никишина А. А. Первичная идентификация в технологии информационного поиска. Часть 1 // Прикладная информатика. 2015. Т. 10. № 4 (58). С. 128–142.
8. Болотова Л. С., Данчул А. Н., Новиков А. П., Сурхаев М. А., Никишина А. А. Первичная идентификация в технологии информационного поиска. Часть 2 // Прикладная информатика. 2015. Т. 10. № 6 (60). С. 128–143.
9. Joun J., Chung H., Park J., Lee S. Relevance analysis using revision identifier in MS Word // Journal of Forensic Sciences. 2020. Vol. 66. P. 323–335. DOI: 10.1111/1556-4029.14584.
10. Jarzabek S., Dan D. Documentation reuse: Managing similar documents // Proceedings of the International Conference on Information Reuse and Integration (IRI). 2017. P. 372–375. DOI: 10.1109/IRI.2017.52.
11. Bešić D. Mikroservis za ekstrakciju teksta iz Word i PDF dokumenata // Zbornik radova Fakulteta tehničkih nauka. 2021. P. 1252–1255. DOI: 10.24867/13BE26Besic.
12. Duretec K., Rauber A., Becker C. A. Text extraction software benchmark based on a synthesized dataset // Proceedings of the Joint Conference on Digital Libraries (JCDL). 2017. P. 1–10. DOI: 10.1109/JCDL.2017.7991565.
13. Karcioğlu A. A., Yaşa A. C. Automatic summary extraction in texts using genetic algorithms // Proceedings of the 28th Signal Processing and Communications Applications Conference (SIU). 2020. P. 1–4. DOI: 10.1109/SIU49456.2020.9302205.
14. Harmata S., Hofer-Schmitz K., Nguyen P. H., Quix C., Bakıu B. Layout-aware semi-automatic information extraction for pharmaceutical documents. – In: Data integration in the life sciences. DILS. Lecture Notes in Computer Science / ed. by M. DA Silveira, C. Pruski, R. Schneider. Springer. Cham, 2017. Vol. 10649. P. 71–85. DOI: 10.1007/978-3-319-69751-2_8.
15. Zhang J., Xie Y., Shen J., Wang L., Lin H. Text information hiding method using the custom components. – In: Cloud computing and security. ICCS 2018. Lecture Notes in Computer Science / ed. by X. Sun, Z. Pan, E. Bertino. Springer. Cham, 2018. Vol. 11066. P. 473–484. DOI: 10.1007/978-3-030-00015-8_41.
16. Lubenets Y. V., Miroshnikov A. I. Software supports for remote examination on mathematical disciplines in higher education // Proceedings of the 1st International Conference on Technology Enhanced Learning in Higher Education (TELE). 2021. P. 274–277. DOI: 10.1109/TELE52840.2021.9482472.
17. Абрамова И. А., Сыркин В. В., Степанов А. П. Расширения стандартного функционала и интерфейса приложений MS Office на основе разработки пользовательских надстроек // Наука и военная безопасность. 2020. № 2 (21). С. 192–199.
18. Мирошникова Е. П., Левоневский Д. К., Мотиенко А. И. Модули импорта/экспорта и аналитики данных в электронной редакции журнала «Труды СПИИРАН» для автоматизированного взаимодействия с глобальными индексами и агрегаторами // Проблемы искусственного интеллекта. 2019. № 3 (14). С. 58–75.
19. Резниченко О. С., Сиваков С. И., Резниченко Т. А. Методика автоматизированного формирования сведений о научных публикациях университета для отчета в системе управления НИР Минобрнауки России // Университетское управление: практика и анализ. 2020. Т. 24. № 2. С. 44–58. DOI: 10.15826/umpa.2020.02.013.
20. Pinto J., Rathod D., Quadros A. Text summarizer for URL and .DOCX files // International Journal of Advanced Research in Computer Science. 2020. Vol. 11. No. 4. P. 18–21. DOI: 10.26483/ijarcs.v11i4.6639.
21. Байнова М. С., Соколов А. М. Инструменты автоматизированного сбора и анализа социологической информации о территориальной идентичности жителей крупных городов // Прикладная информатика. 2021. Т. 16. № 2 (92). С. 92–102. DOI: 10.37791/2687-0649-2021-16-2-92-102.
22. Новиков А. Ю., Кейно П. П. Агрегирование разрозненных данных научных сообществ путем внедрения системы ведения портфолио в сервисе ConfID // Прикладная информатика. 2020. Т. 15. № 2 (86). С. 28–36. DOI: 10.37791/2687-0649-2020-15-2-28-36.
23. Юлаев А. А. Программа для автоматизированного создания и заполнения документов участка выходного контроля: свидетельство о государственной регистрации программы для ЭВМ № 2020610158 Российская Федерация; заявитель Научно-производственная фирма «Микран»; опубл. 09.01.2020.

24. Измайлов В. В., Новоселова М. В. Автоматизированная система формирования вариантов заданий на базе документа MS Word // Программные продукты, системы и алгоритмы. 2017. № 1. С. 1–5. DOI: 10.15827/2311-6749.17.1.1.
25. Неустроев С. А., Беликова М. Ю., Власов М. В. Программа для автоматического формирования шаблона отчета по результатам ЕГЭ: свидетельство о государственной регистрации программы для ЭВМ № 2019610358 Российская Федерация; заявитель Горно-Алтайский государственный университет; опубл. 10.01.2019.
26. Миронов В. В., Гусаренко А. С., Диметриев Р. Р., Сарваров М. Р. Создание персонализированных документов на основе ситуационно-ориентированной базы данных // Вестник Уфимского государственного авиационного технического университета. 2014. Т. 18. № 4 (65). С. 191–197.
27. Миронов В. В., Гусаренко А. С., Тугузбаев Г. А. Ситуационно-ориентированные базы данных: формирование персонализированных графических документов для поддержки учебного проектирования // Моделирование, оптимизация и информационные технологии. 2020. Т. 8. № 2 (29). С. 19–20. DOI: 10.26102/2310-6018/2020.29.2.013.
28. Sadalage P. J., Fowler M. NoSQL distilled: a brief guide to the emerging world of polyglot persistence. – Addison-Wesley, 2013. – 234 p.
29. Миронов В. В., Гусаренко А. С., Юсупова Н. И. Ситуационно-ориентированные базы данных: Polyglot Persistence на основе REST-микросервисов // Прикладная информатика. 2019. Т. 14. № 5 (83). С. 86–97. DOI: 10.24411/1993-8314-2019-10038.
30. Mironov V., Gusarenko A., Yusupova N., Smetanin Yu. JSON documents processing using situation-oriented databases // Acta Polytechnica Hungarica. 2020. Vol. 17. No. 8. P. 29–40. DOI: 10.12700/APH.17.8.2020.8.3.

Сведения об авторах

Миронов Валерий Викторович, ORCID 0000-0002-0550-4676, докт. техн. наук, профессор, кафедра автоматизированных систем управления, Уфимский государственный авиационный технический университет, Уфа, Россия, mironov@list.ru

Гусаренко Артём Сергеевич, ORCID 0000-0003-4132-6106, канд. техн. наук, доцент, кафедра автоматизированных систем управления, Уфимский государственный авиационный технический университет, Уфа, Россия, gusarenko@ugatu.su

Юсупова Нафиса Исламовна, ORCID 0000-0002-7114-7638, докт. техн. наук, профессор, декан факультета информатики и робототехники, заведующая кафедрой вычислительной математики и кибернетики, Уфимский государственный авиационный технический университет, Уфа, Россия, yussupova@ugatu.ac.ru

Поддержка исследований

Выполнено в рамках исследований, поддержанных РФФИ, грант № 19–07–00682.

Статья поступила 09.08.2021, рассмотрена 18.08.2021, принята 22.09.2021

References

1. Hou X., Li N., Yang H., Liang Q. Comparison of wordprocessing document format in OOXML and ODF. Proceedings of the 6th International Conference on Semantics, Knowledge and Grids, 2010, pp.297-300. DOI: 10.1109/SKG.2010.44.
2. Schubert S. The next millennium document format. Proceedings of the ACM Symposium on Document Engineering, 2019, article 40, pp.1-4. DOI: 10.1145/3342558.3345419.
3. Roig J., Ribera M. Implementation of the OOXML standard since its approval until today. Proceedings of the 9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 2020, pp.129-134. DOI: 10.1145/3439231.3440607.
4. Mironov V. V., Yusupova N. I., Shakirova G. R. Situation-oriented databases: external view in the basis of XSL. *Vestnik Ufmskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta*, 2010, vol.14, no.4(39), pp.200-209 (in Russian).
5. Mironov V. V., Gusarenko A. S. Using of RESTful-services in situationally-oriented databases. *Vestnik Ufmskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta*, 2015, vol.19, no.1(67), pp.232-239 (in Russian).
6. Kulkarni A., Shivananda A. Extracting the Data. In book: Natural Language Processing Recipes. Berkeley, CA, Apress, 2021, pp.1-29. DOI: 10.1007/978-1-4842-7351-7_1.

7. Bolotova L. S., Danchul A. N., Novikov A. P., Surkhaev M. A., Nikishina A. A. Initial identification in technology of informational search (part 1). *Prikladnaya Informatika*=Journal of Applied Informatics, 2015, vol.10, no.4(58), pp.128-142 (in Russian).
8. Bolotova L. S., Danchul A. N., Novikov A. P., Surkhaev M. A., Nikishina A. A. Initial identification in technology of informational search (part 2). *Prikladnaya Informatika*=Journal of Applied Informatics, 2015, vol.10, no.6(60), pp.128-143 (in Russian).
9. Joun J., Chung H., Park J., Lee S. Relevance analysis using revision identifier in MS Word. *Journal of Forensic Sciences*, 2020, vol.66, pp.323-335. DOI: 10.1111/1556-4029.14584.
10. Jarzabek S., Dan D. Documentation reuse: Managing similar documents. *Proceedings of the International Conference on Information Reuse and Integration (IRI)*, 2017, pp.372-375. DOI: 10.1109/IRI.2017.52.
11. Bešić D. *Mikroservis za ekstrakciju teksta iz Word i PDF dokumenata* [Microservice for text extraction from Word and PDF documents]. *Zbornik radova Fakulteta tehničkih nauka*, 2021, pp.1252-1255 (in Slovak). DOI: 10.24867/13BE26Besic.
12. Duretec K., Rauber A., Becker C. A Text extraction software benchmark based on a synthesized dataset. *Proceedings of the Joint Conference on Digital Libraries (JCDL)*, 2017, pp.1-10. DOI: 10.1109/JCDL.2017.7991565.
13. Karcioğlu A. A., Yaşa A. C. Automatic summary extraction in texts using genetic algorithms. *Proceedings of the 28th Signal Processing and Communications Applications Conference (SIU)*, 2020, pp.1-4. DOI: 10.1109/SIU49456.2020.9302205
14. Harmata S., Hofer-Schmitz K., Nguyen P. H., Quix C., Bakiu B. Layout-aware semi-automatic information extraction for pharmaceutical documents. In: *Data integration in the life sciences. DILS 2017. Lecture Notes in Computer Science*. Ed. by M. Da Silveira, C. Pruski, R. Schneider. Springer, Cham, 2017, vol.10649, pp.71-85. DOI: 10.1007/978-3-319-69751-2_8.
15. Zhang J., Xie Y., Shen J., Wang L., Lin H. Text information hiding method using the custom components. In: *Cloud computing and security. ICCCS 2018. Lecture Notes in Computer Science*. Ed. by X. Sun, Z. Pan, E. Bertino. Springer, Cham, 2018, vol.11066, pp.473-484. DOI: 10.1007/978-3-030-00015-8_41.
16. Lubenets Y. V., Miroshnikov A. I. Software supports for remote examination on mathematical disciplines in higher education. *Proceedings of the 1st International Conference on Technology Enhanced Learning in Higher Education (TELE)*, 2021, pp.274-277, DOI: 10.1109/TELE52840.2021.9482472.
17. Abramova I. A., Syrkin V. V., Stepanov A. P. Extensions of the standard functionality and interface of MS office applications based on the development of user add-ons. *Nauka i voennaya bezopasnost'*, 2020, no.2(21), pp.192-199 (in Russian).
18. Miroshnikova E. P., Levonevskiy D. K., Motienko A. I. Modules for import, export and data analytics in the electronic journal management system of the "Spiiras Proceedings" journal for automated interaction with global indices and aggregators. *Problemy iskusstvennogo intellekta*=Problems of Artificial Intelligence, 2019, no.3(14), pp.58-75 (in Russian).
19. Reznichenko O. S., Sivakov S. I., Reznichenko T. A. Method of automated generation of information about university's scientific publications for reporting in the research management system of the Russian ministry of science and higher education. *Universitetskoe upravlenie: praktika i analiz*=University Management: Practice and Analysis, 2020, vol.24, no.2, pp.44-58 (in Russian). DOI: 10.15826/umpa.2020.02.013.
20. Pinto J., Rathod D., Quadros A. Text summarizer for URL and .DOCX files. *International Journal of Advanced Research in Computer Science*, 2020, vol.11, no.4, pp.18-21. DOI: 10.26483/ijarcs.v11i4.6639.
21. Baynova M. S., Sokolov A. M. Tools for automated collection and analysis of sociological information on the territorial identity of city residents. *Prikladnaya Informatika*=Journal of Applied Informatics, 2021, vol.16, no.2(92), pp.92-102 (in Russian). DOI: 10.37791/2687-0649-2021-16-2-92-102.
22. Novikov A., Keyno P., Heterogenius data collecting in scientific communities using portfolio management system in ConfID service. *Prikladnaya Informatika*=Journal of Applied Informatics, 2020, vol.15, no.2(86), pp.28-36 (in Russian). DOI: 10.37791/2687-0649-2020-15-2-28-36.
23. Yulaev A. A. *Programma dlya avtomatizirovannogo sozdaniya i zapolneniya dokumentov uchastka vykhodnogo kontrolya* [A program for the automated creation and filling of documents for the final control section]. Certificate of state registration of a computer program No. 2020610158 RF; applicant Scientific and Production Company "Mikran"; published 09.01.2020.
24. Izmailov V. V., Novoselova M. V. *Avtomatizirovannaya sistema formirovaniya variantov zadaniy na baze dokumenta MS Word* [Automated system for generating task options based on MS Word document]. *Programmnye produkty, sistemy i algoritmy*=Software Journal: Theory and Applications, 2017, no.1, pp.1-5. DOI: 10.15827/2311-6749.17.1.1.

25. Neustroev S. A., Belikova M. Yu., Vlasov M. V. *Programma dlya avtomaticheskogo formirovaniya shablona otcheta po rezul'tatam EGE* [A program for automatically generating a report template based on the results of the exam]. Certificate of state registration of a computer program No. 2019610358 RF; applicant Gorno-Altai State University; published 10.01.2019.
26. Mironov V. V., Gusarenko A. S., Dimetrev R. R., Sarvarov M. R. The personalized documents generating using DOM-objects in situation-oriented databases. *Vestnik Ufimskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta*, 2014, vol.18, no.4(65), pp.191-197 (in Russian).
27. Mironov V. V., Gusarenko A. S., Tuguzbaev G. A. Situation-oriented databases: the formation of personalized graphic documents for educational design support. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii*=Modeling, Optimization and Information Technology, 2020, vol.8, no.2(29), pp.19-20 (in Russian). DOI: 10.26102/2310-6018/2020.29.2.013.
28. Sadalage P. J., Fowler M. NoSQL distilled: a brief guide to the emerging world of polyglot persistence. Addison-Wesley, 2013, 234 p.
29. Mironov V. V., Gusarenko A. S., Yusupova N. I. Situation-oriented databases: polyglot persistence based on REST microservices. *Prikladnaya Informatika*=Journal of Applied Informatics, 2019, vol.14, no.5(83), pp.86-97 (in Russian). DOI: 10.24411/1993-8314-2019-10038.
30. Mironov V., Gusarenko A., Yusupova N., Smetanin Yu. JSON documents processing using situation-oriented databases. *Acta Polytechnica Hungarica*, 2020, vol.17, no.8, pp.29-40. DOI: 10.12700/APH.17.8.2020.8.3.

About the authors

Valeriy V. Mironov, ORCID 0000-0002-0550-4676, Dr. Sci. (Eng.), Professor, Automation Control Systems Department, Ufa State Aviation Technical University, Ufa, Russia, mironov@list.ru

Artem S. Gusarenko, ORCID 0000-0003-4132-6106, Cand. Sci. (Eng.), Associate Professor, Automation Control Systems Department, Ufa State Aviation Technical University, Ufa, Russia, gusarenko@ugatu.su

Nafisa I. Yusupova, ORCID 0000-0002-7114-7638, Dr. Sci. (Eng.), Professor, Dean of the Informatics and Robotics Faculty, Head of the Computational Mathematics and Cybernetics Department, Ufa State Aviation Technical University, Ufa, Russia, yussupova@ugatu.ac.ru

Acknowledgments

Completed in the framework of research supported by the Russian Foundation for Basic Research, grant No. 19-07-00682.

Received 09.08.2021, reviewed 18.08.2021, accepted 22.09.2021

Разработка кроссплатформенных мобильных приложений – перспективные методы и стандартные практики

А. Н. Чурсин¹, Н. А. Мамедова^{2*}, Ю. В. Нефедов²

¹ ООО «ЛУКОЙЛ-Интер-Кард», Москва, Россия

² ФГБОУ ВО «Российский экономический университет им. Г. В. Плеханова», Москва, Россия

* mamedova.na@rea.ru

Аннотация. Исследование раскрывает актуальные методические аспекты разработки кроссплатформенных мобильных приложений. Теоретическая значимость исследования заключается в обобщении и изучении программных продуктов и сред IDE, обеспечивающих кроссплатформенную разработку, а также утилит кодогенерации. Представлены результаты сравнительного анализа подходов кроссплатформенной разработки и результаты сравнительного анализа инструментов разработки кроссплатформенных мобильных приложений. Определены факторы, замедляющие разработку кроссплатформенного приложения. Обозначены решения для минимизации функциональных потерь в условиях построения универсального программного продукта. Были приведены данные по использованию Shell-скриптов, что повысит эффективность использования утилиты кодогенерации. Определены шаги по интеграции общей логики в нативные проекты. Практическая значимость результатов исследования представлена методикой разработки кроссплатформенных мобильных приложений с использованием комбинации инструментов для снятия выявленных функциональных ограничений. Исследование содержит комплекс последовательных действий и выводы, к которым пришли исследователи в процессе формирования методики кроссплатформенной разработки мобильных приложений. Методика позволит разработать кроссплатформенный программный продукт, адаптированный для последующих модификаций и расширений. Данный материал может быть оформлен в виде стандарта компании-разработчика в формате ее внутреннего документа (инструкции).

Ключевые слова: мобильное приложение, нативное приложение, кроссплатформенность, проектирование, разработка

Для цитирования: Чурсин А. Н., Мамедова Н. А., Нефедов Ю. В. Разработка кроссплатформенных мобильных приложений – перспективные методы и стандартные практики // Прикладная информатика. 2021. Т. 16. № 6. С. 84–102. DOI: 10.37791/2687-0649-2021-16-6-84-102

Development of cross-platform mobile applications – promising methods and standard practices

A. Chursin¹, N. Mamedova^{2*}, Yu. Nefedov²

¹ LUKOIL-Inter-Card LLC, Moscow, Russia

² Plekhanov Russian University of Economics, Moscow, Russia

* mamedova.na@rea.ru

Abstract. The study reveals the current methodological aspects of the development of cross-platform mobile applications. The theoretical significance of the research lies in the generalization and analysis of software products and IDE environments that provide cross-platform development, as well as code generation utilities. The results of a comparative analysis of cross-platform development approaches and the results of a comparative analysis of cross-platform mobile application development tools are presented. The factors that slow down the development of a cross-platform application are identified. Solutions for minimizing functional losses in the conditions of building a universal software product are indicated. Data on the use of Shell scripts was provided, which will increase the efficiency of using the code generation utility. The steps for integrating common logic into native projects are identified. The practical significance of the research results is represented by the methodology for developing cross-platform mobile applications using a combination of tools to remove the identified limitations. The study contains a set of sequential actions and conclusions that the researchers came to in the process of forming a methodology for the cross-platform development of mobile applications. The methodology will allow us to develop a cross-platform software product adapted for subsequent modifications and extensions. This material can be designed in the standard of the developer company in the format of an internal document (instructions) of the developer company.

Keywords: mobile app, native app, cross-platform, design, development

For citation: Chursin A., Mamedova N., Nefedov Yu. Development of cross-platform mobile applications – promising methods and standard practices. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.84-102 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-84-102

Введение

Цифровые технологии, мобильные устройства и социальные сети стали неотъемлемой частью повседневной жизни людей во всем мире. Согласно данным Digital 2020: Global Overview Report, более 5,19 миллиарда человек в мире пользуются мобильными телефонами, прирост за 2019 год составил 2,4% – 124 миллиона человек [1]. Положительная динамика прироста рынка мобильных приложений наблюдается с 2014 года, согласно данным платфор-

м Hootsuite и GlobalWebIndex. На 2020 год, по данным платформы мобильных данных и аналитики App Annie Intelligence, на мобильные приложения теперь приходится 10 из каждых 11 минут пользования мобильным устройством [2]. За 2019 год пользователи смартфонов загрузили более 200 миллиардов мобильных приложений, потратив совокупно 120 миллиардов долларов на приложения и покупки в них. К 2022 году аналитики App Annie Intelligence прогнозируют рост до 250 миллиардов скачиваний мобильных приложе-

ний и рост доходов с мобильных платформ соответственно.

Такая динамика и активный прогноз обеспечены следующими факторами: рост продаж мобильных устройств, развитие Интернета вещей, развитие инфраструктуры мобильных платежей, развитие приложений по требованию (для заказа конкретных услуг), развитие облачных приложений [3]. Зрелость рынка мобильных операционных систем (далее – ОС), на котором доминируют такие компании, как Google (Android) и Apple (iOS), стимулирует рост магазинов цифровой дистрибуции. Число программистов увеличивается из года в год. Как результат – для охвата увеличивающейся аудитории пользователей мобильных приложений разработчик такого приложения должен разместить его на разных платформах.

Превентивное решение заключается в применении подхода кроссплатформенной разработки мобильных приложений. Во-первых, это снижает стоимость проекта за счет уменьшения расходов на оплату труда и продолжительности проекта. Во-вторых, применение общей логики продукта для разных платформ снижает риск расхождения данных из-за особенностей нативных языков, подходов разработчиков к работе внутренних алгоритмов или банальных ошибок. Однако есть нюансы. Доминирующие на рынке операционные системы Android и iOS имеют между собой массу отличительных черт, начиная от особенностей защиты информации и заканчивая отличиями пользовательских интерфейсов (UI) [4]. На актуальной версии iOS работает около 90% устройств (Apple уделяет большое внимание обновлению своих устройств), на актуальной версии Android – около 12% [5]. Также следует учитывать фактор высокого качества нативных языков и нативных приложений для каждой из платформ в сравнении с приложениями на основе примитивных подходов к кроссплатформенной разработке.

И тем не менее кроссплатформенность как способность программного обеспечения работать с двумя и более аппаратными плат-

формами и (или) операционными системами [6] является очень востребованной характеристикой. Хотя решающим фактором становится даже не сам факт динамического развития рынка мобильных приложений. Обращает на себя внимание то, что мобильные приложения вплотную приблизились по функционалу к программам, написанным для настольных компьютеров, а в определенных аспектах даже превосходили их. За последние несколько лет количество инструментов для написания кроссплатформенных продуктов увеличилось [7, 8]. Однако до сих пор существует свободное пространство перспективных решений, развивающих рынок универсальных приложений и вариативность инструментов и подходов, доступных разработчику [9–11]. С позиции науки такие перспективные решения направлены на развитие области исследования по специальности 05.13.11, в частности пункта специальности «Модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем». Это обстоятельство позволяет сформулировать объект исследования как область перспективных решений для кроссплатформенной разработки мобильных приложений. Предметом исследования определены комбинации инструментов кроссплатформенной разработки, которые имеют преимущества в сравнении со стандартными практиками. Цель данного исследования заключается в разработке методики создания кроссплатформенных мобильных приложений, которая имела бы преимущества относительно существующих методик.

Методическая основа проведенного исследования

Традиционно выделяют несколько подходов к разработке кроссплатформенных приложений: прогрессивное веб-приложение, гибридное мобильное приложение, кроссплатформенное приложение на основе трансляции языка [12]. Каждый из них обеспечи-

вает высокую скорость разработки приложения, однако не может обеспечить качество, сопоставимое с качеством нативного приложения; этот факт даже стал предметом активного обсуждения в сети Интернет. Далее изложено обобщение данных о содержании и развитии подходов для разработки кросс-платформенных приложений.

Основным минусом первых двух подходов является постоянная зависимость от сети Интернет. Прогрессивное веб-приложение по факту работает в окне браузера. Отсюда такие минусы, как невозможность выполнить рекомендации для каждой платформы и невозможность поддерживать все новейшие технологии разработчика платформы. Гибридное приложение хоть и не выполняется в окне браузера, но все вычисления также произво-

дятся на удаленном сервере. Третий подход заключается в написании общего кода (а иногда и общего интерфейса), а далее код транслируется с помощью связывающих библиотек в нативные языки. У такого подхода также имеются недостатки, поскольку трансляция языка все равно не может обеспечить того качества, которое характерно для нативных приложений. Исходя из данных статистики Stack Overflow [13], наиболее популярны гибридные приложения и кроссплатформенные приложения на основе трансляции языка.

Методом сравнительного анализа выявлены предпочтения разработчиков мобильных приложений (табл. 1). Сравнение проведено для определения пула инструментов разработки мобильных приложений и последующего выбора подхода разработки.

Таблица 1. Результаты сравнительного анализа подходов кроссплатформенной разработки
Table 1. Results of a comparative analysis of cross-platform development approaches

Критерии сравнения <i>Comparison criteria</i>	Прогрессивное веб-приложение (выполняется на стороне сервера) <i>Progressive web application (running on the server side)</i>	Гибридное приложение (выполняется на стороне сервера) <i>Hybrid application (running on the server side)</i>	Кроссплатформенное приложение на основе трансляции языка (выполняется на устройстве) <i>Cross-platform application based on language translation (running on the device)</i>
Работа с интерфейсом <i>Working with the interface</i>	Выполняется на сервере – исключает возможность использования без сети <i>Executed on the server – eliminates the possibility of using without a network</i>	Выполняется нативно – позволяет работать с приложением и без сети <i>Executed natively – allows you to work with the application and without a network</i>	Выполняется нативно – позволяет работать с приложением и без сети <i>Executed natively – allows you to work with the application and without a network</i>
Работа с геолокацией <i>Working with geolocation</i>	Работа с геолокацией происходит через браузер на устройстве <i>Working with geolocation is done through the browser on the device</i>	Работает нативно через стандартные библиотеки <i>Works natively through standard libraries</i>	Работает на устройстве через стандартные библиотеки <i>Works on the device via standard libraries</i>
Работа с датчиками и сенсорами <i>Working with detectors and sensors</i>	Возможность отсутствует. Ограничение накладывается браузером <i>This option is not available. The restriction is imposed by the browser</i>	Происходит нативно через стандартные библиотеки <i>The native takes place through the standard library</i>	Происходит на устройстве через стандартные библиотеки <i>Takes place on the device using the standard library</i>

Продолжение таблицы 1

Критерии сравнения <i>Comparison criteria</i>	Прогрессивное веб-приложение (выполняется на стороне сервера) <i>Progressive web application (running on the server side)</i>	Гибридное приложение (выполняется на стороне сервера) <i>Hybrid application (running on the server side)</i>	Кроссплатформенное приложение на основе трансляции языка (выполняется на устройстве) <i>Cross-platform application based on language translation (running on the device)</i>
Работа с базами данных <i>Working with databases</i>	Возможность работать с данными даже в кэше исключена <i>The ability to work with data even in the cache is excluded</i>	Возможность работать с данными в кэше присутствует <i>The ability to work with data in the cache is present</i>	Может происходить как на стороне сервера, так и на устройстве <i>It can occur both on the server side and on the device</i>
Пользовательский интерфейс <i>User interface</i>	Часто не имеют вида нативных iOS-приложений. Инструменты разработки не поддерживают дизайны операционных систем по умолчанию <i>Often they don't have the look of native iOS apps. Development tools do not support default operating system designs</i>	Выполняется нативно или на стороне сервера. Инструменты разработки не поддерживают дизайны ОС по умолчанию при разработке UI на стороне сервера, а значит приложения имеют доступ ко всем UI-фреймворкам (используются «родные» IDE) <i>Executed natively or on the server side. Development tools do not support OS designs by default when developing UI on the server side, which means applications have access to all UI frameworks (native IDE are used)</i>	Работа с пользовательским интерфейсом происходит на устройстве. Приложения не имеют доступа ко всем UI-фреймворкам из-за того, что используются «неродные» IDE <i>Working with the user interface takes place on the device. Applications do not have access to all UI frameworks due to the fact that "non-native" IDE are used</i>
Влияние исполнения кода на размер приложений <i>The effect of the execution of the code size applications</i>	Размер приложений крайне невелик <i>The size of the applications is extremely small</i>	Размер приложений высок (при этом весь исполняемый код работает на стороне сервера) <i>The size of the applications is high (with all executable code running on the server side)</i>	Размер приложений зависит от инструмента разработки и поставленных задач <i>The size of the applications depends on the development tool and the tasks set</i>
Работа с приложениями с повышенными системными требованиями к графике <i>Working with applications with increased system requirements for graphics</i>	Не подходит, так как весь основной код выполняется на сервере <i>Not suitable, since all the main code is executed on the server</i>	Не подходит, так как весь основной код выполняется на сервере <i>Not suitable, since all the main code is executed on the server</i>	Не подходит, так как работа идет на языках, которые не поддерживаются нативно <i>Not suitable, as the work is done in languages that are not natively supported</i>

Продолжение таблицы 1

Критерии сравнения <i>Comparison criteria</i>	Прогрессивное веб-приложение (выполняется на стороне сервера) <i>Progressive web application (running on the server side)</i>	Гибридное приложение (выполняется на стороне сервера) <i>Hybrid application (running on the server side)</i>	Кроссплатформенное приложение на основе трансляции языка (выполняется на устройстве) <i>Cross-platform application based on language translation (running on the device)</i>
Языки разработки <i>Development languages</i>	Выполняется на стороне сервера, вся работа происходит на языках HTML, JavaScript и CSS <i>It is performed on the server side, and all the work is done in HTML, JavaScript, and CSS</i>	Выполняется на стороне сервера, вся работа происходит на языках HTML, JavaScript и CSS <i>It is performed on the server side, and all the work is done in HTML, JavaScript, and CSS</i>	Исполняется на языке разработки, на который нацелен инструментарий <i>It is executed in the development language that the tool-kit is aimed at</i>
Работа с компонентами операционных систем <i>Working with operating system components</i>	Выполняется нативно. Посредником выступает браузер устройства <i>Executed natively. The intermediary is the device browser</i>	Выполняется нативно <i>Executed natively</i>	Выполняется на устройстве <i>Executed on the device</i>
Работа с железом <i>Working with hardware</i>	Выполняется нативно через браузер. Браузер ограничивает возможности работы с железом <i>Executed natively via the browser. The browser restricts the ability to work with hardware</i>	Выполняется нативно <i>Executed natively</i>	Выполняется на устройстве <i>Executed on the device</i>
Управление памятью <i>Memory management</i>	Отсутствует возможность управления памятью на устройстве <i>There is no way to manage the memory on the device</i>	Выполняется нативно. Позволяет тонко распределить ресурсы, необходимые для работы приложения <i>Executed natively. Allows you to fine-tune the resources needed for the application to work</i>	Выполняется на устройстве. Высокая нагрузка на процессор из-за использования сторонних IDE, связующих библиотек <i>Runs on the device. High CPU load due to the use of third-party IDEs, linking libraries</i>
Работа с процессором (например, разбиение задач на потоки) <i>Working with the processor (for example, splitting tasks into threads)</i>	Отсутствует возможность работы с процессором на устройстве <i>There is no possibility to work with the processor on the device</i>	Задачи, не вынесенные на сервер, работают с процессором нативно <i>Tasks that are not submitted to the server work natively with the processor</i>	Выполняется на устройстве. Высокая нагрузка на процессор из-за использования сторонних IDE, связующих библиотек <i>Runs on the device. High CPU load due to the use of third-party IDEs, linking libraries</i>

Критерии сравнения <i>Comparison criteria</i>	Прогрессивное веб-приложение (выполняется на стороне сервера) <i>Progressive web application (running on the server side)</i>	Гибридное приложение (выполняется на стороне сервера) <i>Hybrid application (running on the server side)</i>	Кроссплатформенное приложение на основе трансляции языка (выполняется на устройстве) <i>Cross-platform application based on language translation (running on the device)</i>
Безопасность приложения <i>Application security</i>	Безопасность крайне низкая (основной код выполняется удаленно, использование языков для web-разработки) <i>Security is extremely low (the main code is executed remotely, using languages for web development)</i>	Безопасность крайне низкая (основной код выполняется удаленно, использование языков для web-разработки) <i>Security is extremely low (the main code is executed remotely, using languages for web development)</i>	Выполняется на устройстве. Возникают уязвимости в безопасности связующих библиотек <i>Runs on the device. Possible "holes" in the security of linking libraries</i>

С учетом данных сравнения популярных подходов к разработке кроссплатформенных приложений на следующем шаге исследования было решено уделить внимание инструментам гибридного подхода и подхода на основе трансляции языка. Критический анализ данных подходов направлен на поиск преимуществ и ограничений, препятствующих выявлению перспективной комбинации инструментов кроссплатформенной разработки мобильных приложений. Далее следует обобщение данных о применяемых разработчиками инструментах, популярность которых задает своего рода стандарты профессиональной деятельности.

Одним из самых популярных инструментов на сегодняшний день является Xamarin (инструмент подхода на основе трансляции языка). Xamarin – это фреймворк для кроссплатформенной разработки мобильных приложений (iOS, Android, Windows Phone) с использованием языка C#. В языке кроется первая проблема данного подхода, так как C# нативно не поддерживается ни iOS, ни Android. Разработчикам приходится искать обходные пути. С технической точки зрения Xamarin использует C# и нативные библиотеки, обернутые в .Net-уровень для разработки кроссплатформенных приложений. Такие приложения

часто сравнивают с нативными для платформ iOS и Android с точки зрения производительности и пользовательского опыта. Кроме того, Xamarin может использовать весь доступ к API для использования базовых возможностей платформ в приложениях Xamarin, таких как ARKit на iOS или Android Multi-Window. Стоит напомнить, что масштабные обновления базовых возможностей выходят раз в год, и разработчикам Xamarin приходится адаптировать свой продукт, что требует времени. Код, связанный с бизнес-логикой, доступом к базе данных и сетевым взаимодействием, может быть общим для всех платформ, однако Xamarin позволяет создавать специфичный для платформы уровень кода пользовательского интерфейса.

Таким образом, кроссплатформенные приложения Xamarin выглядят максимально похожими на нативные на любом устройстве, обеспечивая неплохой пользовательский интерфейс по сравнению с универсальными гибридными приложениями¹.

Другим популярным инструментом в списке для разработки гибридных приложений является фреймворк Ionic (инструмент гибридного подхода). AngularJS используется для реализации основных функций Ionic. Эта

¹ <https://medium.com/>

среда разработки мобильных приложений на HTML5 поставляется с элементами пользовательского интерфейса и макетами в нативном стиле, которые можно получить с нативным SDK на iOS или Android. Ionic обладает богатым набором расширенных функций, таких как мобильные компоненты, интерактивные парадигмы, типографика и расширяемая базовая тема, которая облегчает работу. Для работы в качестве нативного приложения требуется встроенная оболочка, такая как Cordova или PhoneGap. Ionic ориентирован на нативные или гибридные приложения вместо мобильных веб-сайтов [14]. Таким образом, приложения Ionic предназначены не для запуска в мобильном браузерном приложении, таком как Chrome или Safari, а в виде низкоуровневой оболочки браузера, такой как iOS UIWebView или Android WebView, которые заключены в такие инструменты, как Cordova/PhoneGap.

Отдельными преимуществами фреймворк Ionic обладает благодаря использованию AngularJS с открытым исходным кодом. AngularJS полностью основан на JavaScript и HTML (меняет статический HTML на динамический HTML). Он расширяет возможности HTML, добавляя встроенные атрибуты и компоненты, а также предоставляет возможность создавать пользовательские атрибуты с использованием JavaScript¹.

Третьим анализируемым инструментом является C++ Builder (инструмент гибридного подхода). Он, в отличие от Xamarin и Ionic, способен компилировать приложения iOS и Android, написанные на языке C++. C++ Builder обладает рядом полезных характеристик. Во-первых, имеет поддержку модульного тестирования DUnitX для мобильных платформ (iOS и Android). Во-вторых, имеет собственную базовую архитектуру визуального управления под названием FireMonkey – содержит элементы управления пользовательского интерфейса в нативном стиле и нативные сервисы. Эти характеристики обеспечивают

высокую производительность и, соответственно, плавную прокрутку, которую ожидают пользователи от нативных приложений. Но при этом свойства, методы и события компонентов будут одинаковыми в разных ОС. Это позволяет сохранять точное соответствие особенностям платформ Windows, Mac, iOS и Android, что влияет на эффективность кроссплатформенной разработки. Актуальная версия C++ Builder на базе RAD Studio 10 обеспечивает производительность CPU, сравнимую с нативной, и более высокую безопасность, чем веб-приложения или приложения на основе трансляции языка (на языке сценариев).

RAD Studio для C++ Builder поддерживает оптимизированный рабочий процесс для разработчика, который автоматизирует развертывание и отладку на физических устройствах. Функционирует одна кодовая база для подключения к различным корпоративным серверам (встроенное многоуровневое корпоративное подключение для Oracle, Sybase, InterBase, MySQL, SOAP), локальным базам данных (SQLite, IBLite и InterBase ToGo) или облачным сервисам. Но есть и объективные недостатки: платная лицензия RAD Studio и отсутствие RAD Studio на Mac OS². И поскольку генерировать сертификаты, необходимые для отладки тестирования и публикации приложения способен только Xcode на Mac OS, IDE (Integrated Development Environment) зачастую устанавливается на Windows, запущенный в виртуальной машине компьютера Mac. Это ведет к удорожанию разработки.

Набирает популярность в UI фреймворк Flutter, с помощью которого можно создавать приложения под Android, iOS, Mac, Web и Windows. Стоит отметить, что Flutter прост в освоении и предлагает подробную и удобную документацию для начинающих. Для создания Flutter-проекта потребуется установка sdk и стороннее приложение по типу VSCode или Idea. Важно знать, что проекты Flutter довольно громоздкие. Во Flutter за-

¹ <https://developer.android.com/>

² <https://www.techopedia.com/>

метно много сходств с фреймворком Swift UI от Apple, так что разработчики, имеющие опыт написания кода на Swift UI, имеют преимущество при разработке.

На данный момент Flutter еще находится в развитии, и многие функции в нем доступны в бета-режиме, однако он уже сыскал популярность в профессиональных кругах. Например, Flutter пользуется успехом во фриланс-сегменте разработки. Как итог – Flutter предлагает интересный подход к созданию кросс-платформы, но за счет масштаба проектов подойдет для небольших команд или разработчиков-одиночек, которым необходимо покрыть максимум платформ с минимальными трудозатратами.

Результаты применения метода сравнительного анализа инструментов разработки

кроссплатформенных мобильных приложений представлены в таблице 2.

Резюмируя результаты сравнения инструментов кроссплатформенной разработки в таблице 2, делаем вывод, что описанные инструменты обладают критическими недостатками по ряду параметров. Среди них такие, как скорость обновлений, доступ к нативным библиотекам и совместимость со сторонними библиотеками, а также размер приложений и особенности тестирования. Это дает основание для предложения в качестве решения собственной методики разработки кроссплатформенных мобильных приложений.

Базой для методики послужила практика разработки кроссплатформенных приложений и ряд ограничений, с которыми столкнулись разработчики, связанных с различиями

Таблица 2. Результаты сравнительного анализа инструментов разработки кроссплатформенных мобильных приложений

Table 2. Results of a comparative analysis of cross-platform mobile application development tools

Инструмент <i>Tool</i>	Xamarin	Ionic	C++ Builder
Критерий <i>Criterion</i>			
Повторное использование кода <i>Code reuse</i>	Возможность использовать до 96% исходного кода <i>Ability to use up to 96% of the source code</i>	Возможность использовать до 80% исходного кода <i>Ability to use up to 80% of the source code</i>	Возможность повторного использования кода <i>The ability to reuse the code</i>
Производительность <i>Productiveness</i>	Производительность небольших приложений близка к нативной <i>The performance of small applications is close to native</i>	Высокая для небольших приложений <i>High for small applications</i>	Высокая <i>Elevated</i>
Использование нативного .api (скорость обновлений) <i>Using the native one .api (update rate)</i>	Поддерживается <i>Supported</i>	Отсутствует <i>Absent</i>	Отсутствует <i>Absent</i>
Открытость исходного кода <i>Open source code</i>	Присутствует <i>Available</i>	Присутствует <i>Available</i>	Отсутствует <i>Absent</i>
Экосистема (симуляторы, справочники) <i>Ecosystem (simulators, reference books)</i>	Присутствует <i>Available</i>	Присутствует <i>Available</i>	Присутствует <i>Available</i>

Инструмент <i>Tool</i>	Xamarin	Ionic	C++ Builder
Критерий <i>Criterion</i>			
Работа с нативным интерфейсом (поддержка изменений дизайна) <i>Working with the native interface</i>	Присутствует несколько интегрированных вариантов <i>There are several integrated options available</i>	Присутствует <i>Available</i>	Присутствует внутри IDE <i>Available inside the IDE</i>
Тестирование <i>Testing</i>	Нет встроенной системы тестирования <i>No built-in testing system</i>	Нет встроенной системы тестирования <i>No built-in testing system</i>	Присутствует <i>Available</i>
Доступ к нативным библиотекам (датчики, сенсоры) <i>Access to native libraries (detectors, sensors)</i>	Ограниченный <i>Limited</i>	Ограниченный <i>Limited</i>	Ограниченный <i>Limited</i>
Язык разработки (скорость компиляции) <i>Development language (compilation speed)</i>	C#	HTML, CSS, Javascript	C++
Размер приложений <i>Application size</i>	Превышает стандарты нативных <i>Exceeds native standards</i>	Соответствует нативным <i>Corresponds to the native ones</i>	Соответствует нативным <i>Corresponds to the native ones</i>
Приложения с повышенными системными требованиями к графике <i>Applications with increased system requirements for graphics</i>	Не подходит <i>Not corresponding</i>	Не подходит <i>Not corresponding</i>	Подходит <i>Corresponding</i>
Совместимость со сторонними библиотеками (в том числе GitHub) <i>Compatibility with third-party libraries (including GitHub)</i>	Значительно ограничена <i>Significantly limited</i>	Ограниченная <i>Limited</i>	Полная <i>Complete</i>
Безопасность <i>Safety</i>	Стандартная <i>Standard</i>	Низкая <i>Low</i>	Стандартная <i>Standard</i>
Сообщество <i>Community</i>	Небольшое <i>Minor</i>	Небольшое, развивается быстро <i>Minor, develops quickly</i>	Большое (разработчики C++) <i>Significant (C++ developers)</i>
Стоимость одной лицензии <i>Cost of one license</i>	Бесплатно <i>Freely</i>	Бесплатно <i>Freely</i>	370 000 руб./лицензия <i>370 000 rubles/license</i>

в архитектурных решениях операционных систем и языках разработки под эти системы.

Во-первых, были обобщены данные об архитектурах целевых ОС, а также языков программирования. Различия в архитектурах ОС, по-разному устроенная работа с файлами, открытость/закрытость систем и многое другое ограничивают кроссплатформенную разработку. Эффективность кроссплатформенных приложений существенно падает из-за того, что разработчики вынуждены использовать большое количество связующих библиотек, а это негативно влияет на производительность, безопасность и размер программных продуктов. Язык C++ нативно поддерживается на нескольких платформах. Это позволило предположить, что логично искать решение с использованием этого языка. Несмотря на появление большого количества новых языков программирования с упрощенным синтаксисом и большим набором «вшитых» возможностей, язык C++ занимает лидирующие позиции в плане скорости, расходуемой памяти и энергоэффективности.

Во-вторых, практика разработки выявила недостатки существующих подходов к разработке кроссплатформенных приложений, среди которых можно выделить повышенную нагрузку на CPU, слабую безопасность итогового продукта и большой размер приложений по сравнению с нативными. Установлены причины возникновения ограничений, накладываемых существующими инструментами кроссплатформенной разработки. В основе этих проблем лежит необходимость использовать неподдерживаемые нативно языки разработки, а также то, что сторонние компании не успевают обновлять свои инструменты в этой динамично развивающейся отрасли.

Указанные выводы были положены в основу собственной методики разработки кроссплатформенных приложений с использованием такой комбинации инструментов, для которой не характерны выявленные ограничения.

Результаты исследования

На основании принятого решения о разработке методики были последовательно совершены следующие действия:

1. Проанализирован и обоснован выбор ОС для разработки кроссплатформенного приложения.

2. Проведена оценка утилиты кодогенерации `djinni` (от Dropbox) и проверка соответствия утилиты всем требованиям разрабатываемой методики.

3. Выбрана валидная поставленная задаче среда разработки.

4. Выбраны вспомогательные утилиты для работы с исходным кодом.

5. Определена необходимость использования интерфейсов в коде.

6. Определен способ повышения эффективности путем использования Shell-скриптов.

7. Установлены особенности разработки общей логики.

8. Установлен способ корректной интеграции логики в нативные компиляторы, а также особенности интеграции в проект базы данных.

Учитывая быстрое развитие инструментов кроссплатформенной разработки, крайне сложно оценить временной горизонт актуальности, однако можно спрогнозировать долгосрочное использование предложенной методики исходя из следующих факторов:

1. Предложение использовать максимально гибкий и универсальный язык C++ для решения задач.

2. Для привлечения аудитории разработчики альтернативных методик стараются максимально упростить работу разработчику, вследствие чего появляются готовые объемные фреймворки и библиотеки, что ведет к замедлению работы готового продукта и увеличению его размера из-за избыточной функциональности. В предложенной методике отсутствует данный недостаток.

3. Методика занимает вполне определенный нишевой сегмент кроссплатформенной разработки. Решение реализуемо в основном в больших проектах, в которых по соображениям информационной безопасности невозможно использование сторонних библиотек и фреймворков и в которых упор делается на быстродействие и небольшой размер итогового продукта. С учетом этого можно сделать вывод о том, что методика будет актуальна продолжительный период времени.

Стоит отметить, что в случае появления на рынке конкурента данной методике, ему, несомненно, потребуется время на развитие. В качестве примера можно привести комплект средств разработки Flutter, который получил первый релиз в мае 2017 года, однако и по сей день сильно подвержен изменениям с каждым новым релизом, что сильно мешает разработке на дистанции и поддержке приложений, написанных при помощи данного инструмента.

Ниже представлено краткое содержание произведенных действий и выводы, к которым пришли исследователи в процессе формирования методики кроссплатформенной разработки мобильных приложений.

1. Операционная система. Разработку было решено производить на операционной системе Mac OS по следующим причинам:

1.1. На ОС Windows и ОС Linux отсутствует возможность полноценно разрабатывать и тестировать ПО под iOS.

1.2. В отличие от Microsoft, Apple разрабатывает свои операционные системы под конкретные сборки «железа». Mac OS настолько эффективно распределяет ресурсы, что способна запускаться на компьютере с оперативной памятью 512 Мб, однокристальным процессором с частотой 1 ГГц и 9 свободными Гб памяти на жестком диске.

1.3. За счет системы Sandbox'ов система от Apple является более защищенной от вредоносного ПО. Кроме того, Mac OS не имеет привычки фрагментировать файлы, оставляя за собой временные файлы, боль-

шое количество встроенных программ и многое другое.

2. *Утилита кодогенерации Djinni*. Это инструмент для создания объявлений типов на разных языках и привязок интерфейса. Утилита дает возможность вынести на сторонний язык только общую логику, оставляя пользовательский интерфейс и особенности систем на нативном языке. Как и в иных консольных приложениях, в ней отсутствует пользовательский интерфейс, и взаимодействие происходит через вводимые в терминах текстовые команды. За счет этого утилита занимает меньше места, а вся производительность уходит только на функциональность. Помимо этого, данная утилита является программой с открытым исходным кодом на ресурсе GitHub. Выбор утилиты обусловлен следующими ее особенностями:

2.1. Создаются параллельные определения типов C++, Java и Objective-C из одного файла описания интерфейса.

2.2. Поддерживается пересечение примитивных типов трех основных языков и пользовательских перечислений, записей и интерфейсов.

2.3. Создается код интерфейса, разрешающий двунаправленные вызовы между C++ и Java (с JNI) или Objective-C.

2.4. Утилита позволяет автоматически генерировать функции внутри компилятора.

3. *Среда разработки*. Для написания нативного кода, касающегося интерфейса, требуются нативные IDE (имеют графический пользовательский интерфейс, который помогает разработчикам создавать программные приложения в интегрированной среде).

3.1. Сравнительный анализ инструментов для приложений под iOS не выявил критических преимуществ между Xcode и AppCode. Под текущий проект выбор Xcode обусловлен тем, что этот инструмент не требует дополнительного ПО, поддерживает тесты и имеет поддержку storyboards. Эти преимущества перекрывают скромные возможности кастомизации интерфейса IDE.

3.2. Сравнительный анализ инструментов для приложений под Android (Android Studio, IntelliJ IDEA и Eclipse) показал, что Android Studio является сегодня наиболее адекватным вариантом для разработки под ОС Android. Вместе с тем следует помнить, что встроенной поддержки Java на Mac OS от Apple нет. Также по умолчанию не устанавливается пакет эмуляторов и ряд инструментов для работы с языком C++. Необходимо скачать последнюю версию Java и Java Developer Kit с официального сайта Oracle, а также установить комплект JDK для работы Android Studio, чтобы снять эти ограничения.

4. *Вспомогательные утилиты.* После выбора IDE необходимо вернуться в консоль, прописать базовые интерфейсы и настроить скрипт для работы утилиты. Логично использовать один из редакторов исходного кода. Выбор инструментов осуществлялся между Microsoft Visual Studio Code, Sublime text и Atom. Результаты сравнительного анализа в основном совпали с данными рейтинга редакторов исходного кода, размещенного на сайте stackshare¹. Итоговым решением стал выбор Visual Studio Code (Sublime text был определен как допустимая альтернатива) – эргономичность интерфейса и удобство использования объединяет эти варианты. Во внимание были приняты следующие характеристики:

4.1. Visual Studio Code – многоязычная IDE со встроенным дебагером; коробочная разработка Front-end; возможность рефакторинга за счет вкладок и некоторых инструментов; встроенный магазин плагинов. Существенным недостатком был определен медленный запуск.

4.2. Sublime text – небольшой вес; дополнительные плагины; высокая скорость; неограниченность во времени пробной версии; развитое комьюнити; модульность. Среди недостатков учтены редкость обновлений, слабый «коробочный» функционал,

крутая кривая обучения, стоимость полной лицензии (около 70 долл. США).

5. *Использование интерфейсов в коде.* Интерфейс взаимодействия описывает поведение или возможности класса C++ без привязки к конкретной реализации этого класса. Интерфейсы C++ реализованы с использованием абстрактных классов (ABC). Абстрактные классы не тождественны абстракции данных, которая представляет собой концепцию хранения деталей реализации отдельно от связанных данных. Класс становится абстрактным путем объявления, по крайней мере, одной из его функций чистой виртуальной функцией. Чистая виртуальная функция указывается путем помещения «= 0» в ее объявление. Целью абстрактного класса является предоставление соответствующего базового класса, от которого могут наследоваться другие классы. Абстрактные классы не могут использоваться для создания объектов и служат только в качестве интерфейса. Попытка создания экземпляра объекта абстрактного класса приводит к ошибке компиляции.

6. *Использование Shell-скриптов.* Shell Script – это компьютерная программа, предназначенная для запуска оболочкой Unix, интерпретатором командной строки. Типичные операции, выполняемые сценариями оболочки, включают манипуляции с файлами, выполнение программ и печать текста. Сценарии оболочки используются для автоматизации административных задач, инкапсуляции сложных деталей конфигурации и получения полной мощности ОС. Сценарий оболочки (Shell Script) необходимо прописать для более эффективной работы утилиты djinni:

6.1. Перед скриптом следует прописать конфигурацию. В нее входят:

A. Djinni IDL file location. Файл, в котором прописаны интерфейсы на языке C++ для дальнейшего использования их утилитой.

B. C++ namespace for generated src. Название директории, в которую будут помещаться исходные C++ файлы.

¹ <https://stackshare.io/stackups/atom-vs-sublime-text-vs-visual-studio-code#stats>

B. Objective-C class name prefix for generated src. Название директории, в которую будут помещаться исходные Objective-C файлы.

G. Java package name for generated src. Название директории, в которую будут помещаться исходные Java-файлы.

6.2. Далее следует текст самого скрипта – указывается базовая директория с установленной утилитой. Следом указываются директории, прописанные в конфигурации, и выходные директории. Выходные директории используются для размещения там сгенерированного кода, который способна воспринять одна из целевых IDE. На всякий случай следует произвести очистку созданных директорий, используя команду «`rm -rf`».

6.3. Строка скрипта начинается со строки «`#!/bin/bash`». Первая строка называется hashbang или shebang. Это сообщает Unix, что данный скрипт должен запускаться через оболочку `/bin/bash`.

6.4. Далее следует написать перечень команд для работы утилиты `djinni`. После запуска утилиты необходимо прописать назначенные выше директории в утилиту. Утилита благодаря этому скрипту создает иерархию проекта.

7. *Разработка общей логики.* Переходим к написанию C++ кода:

7.1. Добавляем к иерархии директорий еще одну с названием «`src`» в папке «`src`» (два файла с расширениями `.h` и `.cpp`). Это позволит разместить весь код C++ вне любых файлов проекта. Впоследствии весь код C++ будет размещаться в этой директории. В более сложном приложении эта директория будет включать кроссплатформенную бизнес-логику, такую как взаимодействие с базой данных SQLite, и вызовы веб-API.

7.2. Далее переходим к описанию работы с нативными компиляторами (на примере IDE Xcode). Необходимо добавить исходные файлы C++ в проект, включая интерфейс, который был прописан ранее, сгенерированные Djinni-файлы, оба файла реализации,

написанные вручную. Важный момент заключается в том, что при добавлении файлов в появившемся диалоговом окне необходимо выбрать «Создать ссылки на папки», а не «Копировать элементы при необходимости», в противном случае Xcode создаст дубликаты файлов внутри папки проекта. Это приведет к тому, что они будут не синхронизированы с кодом C++ в папке «`src`». Кроме того, нужно обязательно убедиться, что установлен флажок рядом с целью развертывания, чтобы файлы были включены при публикации.

7.3. Для портирования на мобильные устройства следует создать новый Target. Target наследует параметры сборки проекта, но любые из них можно переопределить, указав различные параметры на целевом уровне. Схема XCode сама определяет активный Target. Во время создания Target'a Single View Application также можно сразу включить в проект-тесты, которые позволяют проверять и корректировать код и пользовательский интерфейс, написанный на нативных языках, что существенно увеличит эффективность разработки.

7.4. Далее необходимо добавить мост, соединяющий логику iOS с частью проекта. Поскольку исходник C++ уже находится в проекте Xcode, есть возможность просто добавить исходные файлы C++ в новый Target iOS. Далее следует добавить оставшуюся часть кода-моста, сгенерированного Djinni для iOS, для взаимодействия с C++. Также понадобятся некоторые дополнительные исходные файлы Djinni – `deps / djinni / support-lib / objc /`.

7.5. Переименование файла «`main.m`» в «`main.mm`». Это делается для совместимости кода с Objective-C++. После этого остается написать на нативном языке интерфейс View Controller (импортировать нужный класс из файлов «`.h`»), с которым будет взаимодействовать логика.

8. *Интеграция логики в нативные компиляторы.* Наиболее широко используемой современной системой контроля версий в мире

сегодня является Git. Согласно исследованию сайта medium, самыми популярными инструментами для работы с git являются Sourcetree и GitHub Desktop. Сравнение этих утилит показало, что Sourcetree обладает явными преимуществами. Его универсальность подтвердилась при оценке по таким критериям, как графический интерфейс, поддержка GIT-систем, поддержка git-flow и HR-flow, наличие собственного терминала, скорость работы, умение программы решать конфликты в ветках при разработке в команде. Для настройки Sourcetree и репозитория следовало пройти несколько шагов.

8.1. Первый шаг – составление файла «.gitignore». Основная сложность состояла в правильном написании файла «.gitignore». Среди файлов, генерируемых языками и IDE, удалось выявить все, что не требуется хранить в системе контроля версий. В частности, в проекте кроссплатформенного приложения необходимо учесть ненужные файлы, сгенерированные несколькими IDE, а также утилиту кодогенерации и сопутствующие программы. В игнорируемый лист были помещены: файл «.DS_Store»; файлы «build/», «build_ios/» и «DerivedData/»; файл «*.pbxuser» (его альтернативы – «!default.pbxuser», «*.modelv3», «!default.modelv3», «*.modelv2v3», «!default.modelv2v3»); файлы «*.perspectivev3» и «!default.perspectivev3»; «xcuserdata/»; файлы «*.moved-aside», «*.xccheckout», «*.xcscmblueprint»; файлы «*.hmap», «*.ipa», «*.dSYM.zip», «*.dSYM». По этой же логике в игнорируемом листе были учтены файлы для утилиты Djinni и C++, а также файлы для Android Studio, которые не стоило включать в комит.

8.2. Второй шаг – интеграция git в проект. Для этого следовало:

- а) перейти в корень проекта при помощи команды «cd»;
- б) добавить туда файл «.gitignore»;
- в) добавить все файлы (кроме тех, что попадают под правила «.gitignore») с помощью команды «git add -A»;
- г) выполнить первый инициализирующий commit с помощью команды

git commit -m «init», добавить изменения в репозиторий. В данной команде после параметра -m задается описание добавляемых изменений. Стоит давать осмысленные комментарии к commit, чтобы можно было оперативно найти те или иные изменения и т. д.;

- д) добавить в Sourcetree локальный и глобальный репозиторий.

Ниже приведена последовательность действий разработчика по применению методики кроссплатформенной разработки мобильного приложения в формате внутреннего документа (инструкции) компании-разработчика.

Описание методики кроссплатформенной разработки мобильного приложения

1. Установка и настройка ПО для разработки.

1.1. Установка Xcode. Xcode скачивается из магазина Mac App Store, предустановленного на всех компьютерах Mac, или с сайта.

1.2. Создание аккаунта разработчика Apple. Создать аккаунт разработчика можно на сайте <https://developer.apple.com/>. Цена аккаунта составляет 100 долларов США в год.

1.3. Вход в Xcode с аккаунтом разработчика.

1.4. Установка симулятора iOS. Операционная система должна быть актуальной версии как на симуляторе, так и на физическом устройстве для тестирования (iPhone или iPad).

1.5. Установка Java Development Kit. Установка происходит через DMG-файл, скачиваемый с сайта <http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>. Версию Java, установленную на данный момент (если установлена), можно проверить через команду java-version в терминале.

1.6. Установка Android Studio с сайта <https://developer.android.com>.

1.7. Установка вспомогательных инструментов для Android Studio через SDK Manager.

1.7.1. Установка Android SDK Build-Tools и Android SDK Tools.

1.7.2. Установка CMake.

1.7.3. Установка Android Emulator.

1.7.4. Установка Android SDK Platform-Tools.

1.7.5. Установка Intel x86 Emulator Accelerator.

1.7.6. Установка Android NDK.

1.7.7. Установка ConstraintLayout for Android.

1.7.8. Установка Solver for Constraint Layout.

1.7.9. Установка Android Support Repository.

1.7.10. Установка Google Repository.

1.8. Установка устройства для эмулятора Android через AVD Manager.

1.9. Установка Visual Studio Code с сайта <https://code.visualstudio.com>.

1.10. Установка расширений для Visual Studio Code.

1.10.1. Установка расширения для работы с кодом утилиты CMake.

1.10.2. Установка расширения Djinni Syntax Highlighting для работы с синтаксисом утилиты djinni.

1.10.3. Установка расширения Shell-format для работы с кодом Shell-скриптов.

1.11. Установка утилиты djinni.

1.11.1. Создание папки для проекта.

1.11.2. Инициализация git-репозитория через терминал командой «git init».

1.11.3. Установка сабмодуля djinni командой «git submodule add» с сайта <https://github.com/dropbox/djinni.git> deps/djinni.

2. Работа с GIT.

2.1. Написание файла «.gitignore».

2.2. Инициализирующий commit.

2.3. Подключение репозитория к Sourcetree.

3. Разработка кроссплатформенного кода.

3.1. Написание интерфейсов.

3.2. Написание общей логики.

3.3. Написание sh-скрипта.

3.3.1. Присваивание файла с интерфейсами переменной «djinni_file».

3.3.2. Присваивание файла с генерируемым пространством имен C++ переменной «namespace».

3.3.3. Присваивание префикса имени класса Objective-C переменной «objc_prefix».

3.3.4. Присваивание имя пакета Java переменной «java_package».

3.3.5. Получение базовой директории в переменную «base_dir».

3.3.6. Получение Java-директории из имени пакета в переменную «java_dir».

3.3.7. Получение выходных директорий командами «\$base_dir/generated-src/ФАЙЛ» (для файлов «cpp», «objc» и «jni») в переменные «ФАЙЛ_out».

3.3.8. Получение выходной директории командой «\$base_dir/generated-/java/\$java_dir» для Java в переменную «java_out».

3.3.9. Очистка директорий, описанных в пунктах 3.3.7 и 3.3.8, поочередно командами «rm -rf».

3.3.10. Скрипт для запуска djinni прописан разработчиками в официальном репозитории на сайте <https://github.com/dropbox/djinni>.

3.4. Запуск скрипта консольной командой «run_djinni.sh». В случае неудачи из-за недостатка ввести команду «chmod +x ./run_djinni.sh» и повторить запуск скрипта.

3.5. Создание в корне проекта новых директорий (одна вложена в другую) «.../src/cpp».

3.6. Сохранение файлов с логикой в каталог из пункта 1.11.1.

3.7. Интеграция общей логики в Xcode.

3.7.1. Создание нового консольного приложения в Xcode на языке C++.

3.7.2. Интеграция C++ файлов (.hpp и .cpp) из папки из пункта 1.11.1 и папки

«generated_src» путем перетаскивания. Данные необходимо передать по ссылке, указав соответствующий пункт в контекстном меню после перетаскивания. При переносе необходимо указать Target нынешнего проекта.

3.7.3. Написание необходимого кода в файле «main.cpp» (например, вывод промежуточных результатов в консоль).

3.7.4. Создание нового Target для iOS-приложения с определенными параметрами.

3.7.4.1. Указание языком разработки Objective-C.

3.7.4.2. Включение тестов в проект.

3.7.4.3. Указание проекта для интеграции.

3.7.5. Добавление к файлам C++ Target iOS-приложения.

3.7.6. Переименование файла «main.m» в «main.mm».

3.8. Интеграция общей логики в Android Studio.

3.8.1. Создание C++ проекта в Android Studio.

3.8.2. Настройка файла «CMake».

3.8.3. Редактирование файла «Gradle».

3.8.4. Редактирование файла «Main Activity» (например, для вывода промежуточных результатов в консоль).

4. Разработка пользовательского интерфейса iOS-приложения.

5. Разработка пользовательского интерфейса Android-приложения.

Заключение

В результате проведенных работ выявлены валидные условия задачи инструменты кроссплатформенной разработки. Среди них – нативные IDE, редактор исходного кода, программа для работы с GIT и утилита кодогенерации. Была описана методика работы с консольной утилитой кодогенерации djinni. Были проанализированы нативные компиляторы под систему iOS и выбран

рабочий вариант среды разработки – Xcode. В ходе исследования, проведенного с целью выявить наиболее эффективный инструмент разработки, было установлено, что альтернатив Android Studio практически нет, так как все конкуренты либо устарели, либо не отличаются высоким качеством.

Установлены факторы, замедляющие разработку кроссплатформенного приложения, в частности, необходимость прописывать многие команды вручную и необоснованное складирование данных в репозитории. Была проведена работа по рассмотрению Shell-скриптов, что повысит эффективность использования утилиты кодогенерации. Рассмотрена структура скрипта, его синтаксис и обоснована необходимость его использования в проекте, нацеленном на кроссплатформенную разработку.

Определены шаги по интеграции общей логики в нативные проекты. Описано, как правильно включить сгенерированные файлы-мосты между C++ и Objective-C в структуру проекта. Был проведен сравнительный анализ, в ходе которого удалось выявить отсутствие схожей проблемы для системы iOS. Были определены необходимые пакеты SDK для максимально корректной работы приложения, для максимального охвата пользователей. Описаны шаги выбора и настройки системы контроля версий и репозитория. Был разработан файл «gitignore», отвечающий требованиям проекта. Были проанализированы утилиты, помогающие в работе с GIT, и выявлена наилучшая из них.

Предложена методика разработки кроссплатформенного программного продукта. Она наиболее эффективна для приложений, в которых присутствует большой объем внутренней логики и которая может пострадать при переписывании на разные языки. Именно тогда методика проявляет себя с лучшей стороны. Методика представлена в форме, соответствующей внутренней инструкции компании-разработчика мобильных приложений.

Список литературы

1. Digital 2020 // Hootsuite. URL: <https://hootsuite.com/resources/digital-2020> (дата обращения: 17.12.2020).
2. The state of mobile 2020 report // APP ANNIE. URL: <https://www.appannie.com/en/insights/market-data/state-of-mobile-2020/> (дата обращения: 30.12.2020).
3. Zohud T., Zein S. A systematic mapping study of cross-platform mobile apps // Journal of Computer Science. 2019. Vol. 15. No. 4. P. 519–536. DOI: 10.3844/jcssp.2019.519.536.
4. Beranek M., Kovar V. A component-based method for developing cross-platform user interfaces for mobile applications. – In book: Fourth International Congress on Information and Communication Technology. 2020. P. 201–217. DOI: 10.1007/978-981-15-0637-6_17.
5. Menegassi A. A., Endo A. T. Automated tests for cross-platform mobile apps in multiple configurations // IET Software. 2020. Vol. 14. No. 1. P. 27–38. DOI: 10.1049/iet-sen.2018.5445.
6. Crossplatform // TechTerms. URL: <https://techterms.com/definition/crossplatform> (дата обращения: 18.12.2021).
7. Benouda H., Azizi M., Moussaoui M., Esbai R. Automatic code generation within MDA approach for cross-platform mobiles apps // Proceedings of the 1st International Conference on Embedded and Distributed Systems (EDIS). 2017. P. 1–5. DOI: 10.1109/EDIS.2017.8284045.
8. Lachgar M., Lamhaddab K., Abdali A., Elbaamrani K. Pragmatic approach to modeling and generating mobile cross-platform applications // Journal of Computer Science. 2019. Vol. 15. No. 3. P. 416–434. DOI: 10.3844/jcssp.2019.416.434.
9. Ardagna C. A., Bellandi V., Ceravolo P., Damiani E., Finazzo R. A methodology for cross-platform, event-driven Big Data analytics-as-a-service // Proceedings of the International Conference on Big Data. 2019. P. 3440–3448. DOI: 10.1109/BigData47090.2019.9005503.
10. Yahya-Imam M. K., Palaniappan S., Ghadiri S. M. Investigation of methodical framework for cross-platform mobile application development: Significance of Codename One // International Journal of Computer Aided Engineering and Technology. 2019. Vol. 11. No. 4-5. P. 439–448. DOI: 10.1504/IJCAET.2019.100443.
11. Jia X., Ebone A., Tan Y. A performance evaluation of cross-platform mobile application development approaches // Proceedings of the 5th International Conference on Mobile Software Engineering and Systems. 2018. P. 92–93. DOI: 10.1145/3197231.3197252.
12. Ferreira C. M. S., Peixoto M. J. P., Duarte P. A. S., Torres A. B. B., Silva Júnior M. L., Rocha L. S., Viana W. An evaluation of cross-platform frameworks for multimedia mobile applications development // IEEE Latin America Transactions. 2018. Vol. 16. No. 4. P. 1206–1212. DOI: 10.1109/TLA.2018.8362158.
13. Silge J. The cliffs of insanity: dramatic shifts in technologies on Stack Overflow // Stack Overflow. 2017. URL: <https://stackoverflow.blog/2017/11/13/cliffs-insanity-dramatic-shifts-technologies-stack-overflow/> (дата обращения: 24.12.2020).
14. Yang Y., Zhang Y., Xia P., Li B., Ren Z. Mobile terminal development plan of cross-platform mobile application service platform based on Ionic and Cordova // Proceedings of the International Conference on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). 2017. P. 100–103. DOI: 10.1109/ICIICII.2017.28.

Сведения об авторах

Чурсин Артём Николаевич, ORCID 0000-0002-3055-1589, ведущий специалист отдела разработки, ООО «ЛУКОЙЛ-Интер-Кард», Москва, Россия, tema320d@icloud.com

Мамедова Наталья Александровна, ORCID 0000-0002-8934-7363, канд. экон. наук, доцент, базовая кафедра цифровой экономики, Институт развития информационного общества, Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский экономический университет им. Г. В. Плеханова», Москва, Россия, mamedova.na@rea.ru

Нефедов Юрий Викторович, ORCID 0000-0002-9772-2629, канд. экон. наук, доцент, базовая кафедра цифровой экономики, Институт развития информационного общества, Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский экономический университет им. Г. В. Плеханова», Москва, Россия, nefedov.yv@rea.ru

Статья поступила 02.03.2021, рассмотрена 09.03.2021, принята 16.08.2021

References

1. Digital 2020. Hootsuite. Available at: <https://hootsuite.com/resources/digital-2020> (accessed 17.12.2020).
2. The state of mobile 2020 report. APP ANNIE. Available at: <https://www.appannie.com/en/insights/market-data/state-of-mobile-2020/> (accessed 30.12.2020).
3. Zohud T., Zein S. A systematic mapping study of cross-platform mobile apps. *Journal of Computer Science*, 2019, vol.15, no.4, pp.519-536. DOI: 10.3844/jcssp.2019.519.536.
4. Beranek M., Kovar V. A component-based method for developing cross-platform user interfaces for mobile applications. In book: *Fourth International Congress on Information and Communication Technology*, 2020, pp.201-217. DOI: 10.1007/978-981-15-0637-6_17.
5. Menegassi A. A., Endo A. T. Automated tests for cross-platform mobile apps in multiple configurations. *IET Software*, 2020, vol.14, no.1, pp.27-38. DOI: 10.1049/iet-sen.2018.5445.
6. Crossplatform. TechTerms. Available at: <https://techterms.com/definition/crossplatform> (accessed 18.12.2021).
7. Benouda H., Azizi M., Moussaoui M., Esbai R. Automatic code generation within MDA approach for cross-platform mobiles apps. *Proceedings of the 1st International Conference on Embedded and Distributed Systems (EDIS)*, 2017, pp.1-5. DOI: 10.1109/EDIS.2017.8284045.
8. Lachgar M., Lamhaddab K., Abdali A., Elbaamrani K. Pragmatic approach to modeling and generating mobile cross-platform applications. *Journal of Computer Science*, 2019, vol.15, no.3, pp.416-434. DOI: 10.3844/jcssp.2019.416.434.
9. Ardagna C. A., Bellandi V., Ceravolo P., Damiani E., Finazzo R. A methodology for cross-platform, event-driven Big Data analytics-as-a-service. *Proceedings of the International Conference on Big Data*, 2019, pp.3440-3448. DOI: 10.1109/BigData47090.2019.9005503.
10. Yahya-Imam M. K., Palaniappan S., Ghadiri S. M. Investigation of methodical framework for cross-platform mobile application development: Significance of Codename One. *International Journal of Computer Aided Engineering and Technology*, 2019, vol.11, no.4-5, pp.439-448. DOI: 10.1504/IJCAET.2019.100443.
11. Jia X., Ebone A., Tan Y. A performance evaluation of cross-platform mobile application development approaches. *Proceedings of the 5th International Conference on Mobile Software Engineering and Systems*, 2018, pp.92-93. DOI: 10.1145/3197231.3197252.
12. Ferreira C. M. S., Peixoto M. J. P., Duarte P. A. S., Torres A. B. B., Silva Júnior M. L., Rocha L. S., Viana W. An evaluation of cross-platform frameworks for multimedia mobile applications development. *IEEE Latin America Transactions*, 2018, vol.16, no.4, pp.1206-1212. DOI: 10.1109/TLA.2018.8362158.
13. Silge J. The cliffs of insanity: dramatic shifts in technologies on Stack Overflow. *Stack Overflow*, 2017. Available at: <https://stackoverflow.blog/2017/11/13/cliffs-insanity-dramatic-shifts-technologies-stack-overflow/> (accessed 24.12.2020).
14. Yang Y., Zhang Y., Xia P., Li B., Ren Z. Mobile terminal development plan of cross-platform mobile application service platform based on Ionic and Cordova. *Proceedings of the International Conference on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*, 2017, pp.100-103. DOI: 10.1109/ICIICII.2017.28.

About the authors

Artem N. Chursin, ORCID 0000-0002-3055-1589, Leading Specialist of the Development Department, LUKOIL-Inter-Card LLC, Moscow, Russia, tema320d@icloud.com

Natalia A. Mamedova, ORCID 0000-0002-8934-7363, Cand. Sci. (Econ.), Associate Professor, Basic Department of Digital Economy, Institute of the Information Society, Plekhanov Russian University of Economics, Moscow, Russia, mamedova.na@rea.ru

Yuri V. Nefedov, ORCID 0000-0002-9772-2629, Cand. Sci. (Econ.), Associate Professor, Basic Department of Digital Economy, Institute of the Information Society, Plekhanov Russian University of Economics, Moscow, Russia, nefedov.yv@rea.ru

Received 02.03.2021, reviewed 09.03.2021, accepted 16.08.2021

Постановка задачи обеспечения автоматизации полного цикла поддержки принятия решения

О. В. Туханычев^{1}*

¹ *Группа компаний «Техносерв», Москва, Россия*

**tow65@yandex.ru*

Аннотация. Актуальность темы статьи определяется тем, что существующие системы поддержки принятия решений (СППР) обеспечивают автоматизацию только части этапов типового цикла управления, полностью возлагая на пользователя такие важные из них, как целеполагание и управление выполнением разработанных планов. Сложившаяся ситуация определяется объективными сложностями формализации и динамичности указанных этапов, но на практике это разрывает единый процесс управления на автономные, частично автоматизированные участки. Существующие программные компоненты целеполагания и автоматизированного управления, реализованные в системах корпоративного управления (системы типа Enterprise Resource Planning) или средствах SMART-целеполагания (Specific, Measurable, Achievable, Relevant, Time bound), являются не компонентами СППР, а лишь отдельными программными средствами, решающими проблему частично. Более того, исследования в области разработки подобных средств для СППР в настоящий момент не ведутся. С учетом этого целью исследования являлось получение описательной постановки задачи автоматизации процессов целеполагания и управления для реализации в СППР. В статье с использованием общенаучных методов исследования проведен анализ содержания цикла управления и существующих технологий автоматизации его элементов. С применением результатов анализа синтезированы предложения и сформулирована постановка задачи на автоматизацию этапов целеполагания и управления выполнением планов, для реализации в составе СППР, предложены подходы, которые могут быть использованы для реализации поставленной задачи. Проблему автоматизации начального этапа цикла управления, целеполагания, предлагается решить применением специализированных экспертных систем, обеспечивающих формирование ранжированных перечней возможных целей на основе описательной постановки задачи, формируемой пользователем СППР на естественном языке. Решение проблемы поддержки принятия решений на этапе управления предлагается обеспечить за счет использования упрощенных динамических алгоритмов управления.

Ключевые слова: поддержка принятия решений, автоматизация управления, типовой цикл управления, этапность принятия решений

Для цитирования: *Туханычев О. В.* Постановка задачи обеспечения автоматизации полного цикла поддержки принятия решения // Прикладная информатика. 2021. Т. 16. № 6. С. 103–112. DOI: 10.37791/2687-0649-2021-16-6-103-112

Formulation of the task of ensuring automation of the full cycle of decision support

O. Tikhanychev^{1*}

¹ Group of companies «Technoserv», Moscow, Russia

*tow65@yandex.ru

Abstract. The relevance of the topic of the article is determined by the fact that the existing decision support systems (DMSS) provide automation of only part of the stages of a typical management cycle, completely imposing on the user such important ones as goal setting and management of the implementation of developed plans. This is determined by the objective difficulties of formalization and the dynamism of these stages, but breaks the single management process into autonomous, partially automated sections. No research is currently being conducted to solve this problem. Taking this into account, the purpose of the study was to obtain a descriptive formulation of the task of automating the processes of goal-setting and management for implementation in the DMSS. The article, using general scientific research methods, analyzes the content of the control cycle and the existing technologies for automating its elements. Using the results of the analysis, proposals were synthesized and the task was formulated to automate the stages of goal-setting and management of the implementation of plans, for implementation as part of the DMSS, approaches were proposed that can be used to implement the task. The problem of automating the initial stage of the management cycle, goal-setting, is proposed to be solved by using specialized expert systems. To solve the problem of decision support at the control stage, it is proposed to use simplified dynamic control algorithms. In a direct formulation, such a problem is posed for the first time. The article proposes a set of methods and basic algorithms that provide a solution to the problem of automating goal-setting and management processes and, in the future, their implementation in the structure of the DMSS.

Keywords: decision support, control automation, typical control cycle, stages of decision making

For citation: Tikhanychev O. Formulation of the task of ensuring automation of the full cycle of decision support. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.103-112 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-103-112

Введение

Несмотря на достаточно длительное развитие теории, на практике системы поддержки принятия решений (СППР) используются не так активно, как требуется с точки зрения повышения эффективности управления [1, 2].

Сложившаяся ситуация определяется целым набором объективных и субъективных проблем [3, 4], сдерживающим создание и внедрение СППР в практику управления. К таким проблемам можно отнести то, что

существующие СППР обеспечивают автоматизацию только части этапов типового типа управления, в первую очередь планирования и автоматизации рутинной деятельности персонала, например формирования и документирования управляющих воздействий [5, 6]. Именно под такие задачи создано подавляющее большинство существующих ERP-систем (Enterprise Resource Planning), обеспечивающих планирование экономической деятельности [7–9]. Основную долю автоматизируемых функций в специализированных системах автоматизации управления состав-

ляют также задачи планирования [10, 11]. Для ряда автоматизированных систем, управляющих процессами и объектами с высокой инерционностью, поддержка принятия решений на этапе управления обеспечивается периодическим уточнением планов с применением существующих СППР. Но такое решение обеспечивает парирование проблемы поддержки решений на управление только в частных случаях, когда длительность этапа планирования существенно меньше располагаемой продолжительности цикла управления: общего решения оно не дает. Еще одна проблема существующих СППР – отсутствие у них функционала поддержки целеполагания, полностью возлагаемого на пользователей системы.

Учитывая современное состояние предметной области, можно сделать вывод о том, что существующие системы автоматизированного управления обеспечивают поддержку принятия решений в ограниченном объеме, упуская часть этапов цикла управления, что не может не сказаться на качестве принимаемых решений.

Таким образом, решение проблемы организации полного цикла автоматизированной поддержки принятия решений является актуальной задачей, обеспечивающей дальнейшее развитие автоматизации управления.

Выделение проблемных этапов поддержки принятия решений

Принято считать, что типовой цикл управления состоит из нескольких основных этапов:

- целеполагание;
- оценка ситуации;
- выработка общего замысла действий и планирование;
- постановка задач и их выполнение;
- контроль и анализ результатов;
- уточнение управляющих воздействий (корректировка задач).

В цикле управления нет неважных или необязательных этапов. Этапы цикла можно сокращать, объединять, но общая структура

цикла для обеспечения эффективности управления должна сохраняться. А СППР, в современном понимании подобных систем, предназначаются и, соответственно, разрабатываются только для поддержки этапа планирования и отдельных составляющих этапов управления, обеспечивающих выработку управленческих решений: сбор данных и информационное обеспечение оценки ситуации, выполнение рутинных функций по формированию управляющих воздействий и т. п.

Как показывает анализ практики управления, для обеспечения поддержки принятия решений на остальных этапах цикла управления существующие СППР просто не предназначены [12]. В результате важнейшие этапы, определяющие задание цели действий (целеполагание) и управление достижением спланированного результата (контроль выполнения и корректировка задач), аналитической и информационно-расчетной поддержкой не обеспечиваются.

Сложившаяся ситуация порождает противоречие между активным развитием информационных технологий и ограниченным использованием их возможностей для полноценной и всеобъемлющей автоматизации всех этапов цикла управления. Данное противоречие в перспективе может порождать кризис автоматизации управления и должно быть разрешено.

О возможности автоматизации поддержки этапа целеполагания

Этап целеполагания, а именно формирование набора значений, описывающих требуемое конечное состояние системы, учитывая сложность формализации и множество влияющих факторов, разработчики автоматизированных СППР, как правило, оставляют за лицом, принимающим решения (ЛПР). Во многом это вынужденная мера: данный этап настолько слабо формализуем, в нем оценивается такое количество влияющих факторов количественно-качественной природы, что

разработать программные средства его поддержки крайне сложно. По крайней мере на современном этапе развития технологий автоматизации управления. Играет определенную роль и субъективный фактор, определяемый тем, что ЛПП в большинстве ситуаций считает, что способно самостоятельно эффективно сформулировать цель действий, нуждаясь в СППР только для выработки методов ее достижения. В некоторых случаях, для опытных управленцев, это мнение обосновано, но так бывает не всегда.

Более того, существуют ситуации, когда в явном виде, в форме набора количественных критериев, цель действий сформулировать сложно, имеется набор частных целей, выполнение которых в определенных сочетаниях приводит к заданному результату. И определить заранее, какая последовательность предпочтительнее, невозможно или слишком затратно по времени. Иногда цель вообще может быть задана только в качественной форме, непригодной для формирования задачи поиска решения системам автоматизированного управления. Постановка задачи для СППР в подобном контексте может быть сформулирована в следующей форме:

- сформировать предложения о том, какой исходный посыл необходимо выбрать для реализации задачи, поставленной в неявной форме;
- сформулировать неявную задачу в формализованном виде, понятном для «классической» СППР, или сформулировать набор компонентов c_k на основе нестрого описанной цели действий.

Решение сформулированной задачи в том или ином виде обеспечит поддержку ЛПП на этапе целеполагания при разных уровнях неопределенности: когда цель сформулировать проблематично (первый вариант) или когда ее трудно задать конкретно, в формализованной форме (второй вариант).

В настоящее время подобная задача ни в первой, ни во второй постановке не решается ни одной автоматизированной системой.

Анализ существующего инструментария показывает, что для поддержки данного этапа цикла управления имеются частные решения в виде программных средств поддержки целеполагания, обладающие низким уровнем автоматизации: SMART-целеполагание (Specific, Measurable, Achievable, Relevant, Time bound), построение дерева целей и другие [13, 14]. Но, как показывает анализ, их практическая реализация в ряде ERP-систем, например в SAP SuccessFactors, является лишь графическим инструментом для реализации диалогового формирования целей в интересах решения частных задач управления. Например, последняя – для решения задач HR-отдела компании. Считать такие инструменты полноценными компонентами целеполагания для СППР нельзя.

В ряде случаев под инструментом целеполагания разработчики понимают саму СППР в целом, считая, что завершение полного цикла расчетов и позволит обеспечить реализацию принципов SMART-методологии. Теоретически это работает, но этап целеполагания при таком подходе за счет замены понятия «постановка цели» на «определение возможности ее достижения» («Measurable, Achievable, Relevant, Time bound», без реализации функции «Specific») вновь возлагается на оператора. То есть SMART-подход нельзя считать решением проблемы автоматизации целеполагания.

В то же время по крайней мере для второго варианта действий могут быть использованы подходы на основе применения экспертных систем, включаемых в СППР как составная часть. Эти экспертные системы могут быть реализованы на основе автоматизированного использования существующих методов поиска решения: аппарата формальной логики, методов сценарного анализа и им подобных. Но использоваться они должны не в итерационно-поисковом режиме, как это делается существующими программными средствами, а в режиме предписывающего компонента СППР, являясь неотъемлемой составной ча-

стью системы и обеспечивая следующий алгоритм:

- формирование набора альтернатив целей;
- описание их на естественном языке;
- ранжирование по приоритетам;
- представление ЛПР для выбора конкретной цели.

Только при реализации такого подхода можно будет считать, что этап целеполагания автоматизирован и включен в единый цикл поддержки принятия решения.

Обеспечение поддержки принятия решений на этапе управления (реализации решений)

Этап управления в типовом цикле включает следующие важные элементы:

- контроль, то есть сопоставление достигнутых результатов с заданными, анализ критичности отклонений;
- формирование предложений, обеспечивающих достижение поставленной цели с учетом возникших отклонений;
- корректировка планов, а в отдельных случаях и цели действий, что является завершением цепи обратной связи и управленческого цикла в целом.

Данный этап как наиболее динамичный и, соответственно, обладающий жесткими требованиями к средствам автоматизации обычно остается за рамками исследований при разработке СППР, которые сейчас являются преимущественно планирующими системами. Что, учитывая его важность в типовом цикле управления, недопустимо. Для устранения этого недостатка предлагается рассмотреть процесс управления с точки зрения его формализации.

Формально, с точки зрения типовой модели управления, этап управления состоит из непрерывного анализа ситуации и уточнения на его основе управляющих воздействий, обеспечивающих приведение управляемой системы в требуемое состояние с учетом ком-

пенсации ошибок управления и внешних возмущающих воздействий.

В рамках исследовательской модели для формализованного описания процесса управления управляемая система, как правило, представляется в форме векторов, описанных в фазовых координатах состояния.

Под вектором состояния системы понимается семейство множеств вида

$$C = \{\{c_k\}\{c_y\}\{c_s\}\},$$

где C – семейство множеств, описывающих текущее состояние системы;

c_k – множество параметров состояний, соответствующих контрольным параметрам вектора целей управления (параметрам, приводящим управляемую систему к цели);

c_y – множество управляемых параметров;

c_s – множество свободных параметров.

Вектор состояния постоянно меняется по времени и всегда определяется с некоторой ошибкой, порождающей неопределенность для лица, принимающего решения. В результате возникает ошибка управления, которая тоже является векторной величиной. Вектор ошибки управления – это вектор, описывающий отклонение реального процесса, отображаемого вектором состояния контрольных параметров, от вектора целей управления.

В рамках принятой модели процесс управления понимается как последовательность формирования управляющих решений с параметрами c_y , обеспечивающими вывод системы в состояние c_k к установленному времени и с приемлемыми затратами.

Определено, что задачей СППР на этапе планирования является формирование управляющих параметров, реализация которых приведет управляемый объект в требуемую точку фазового пространства, то есть обеспечит достижение цели. В дальнейшем, на этапе управления, СППР должна на основе обработки данных о текущем состоянии управляемой системы непрерывно или с малой дискретностью вырабатывать уточненные управляющие параметры, обеспечивающие достижение цели.

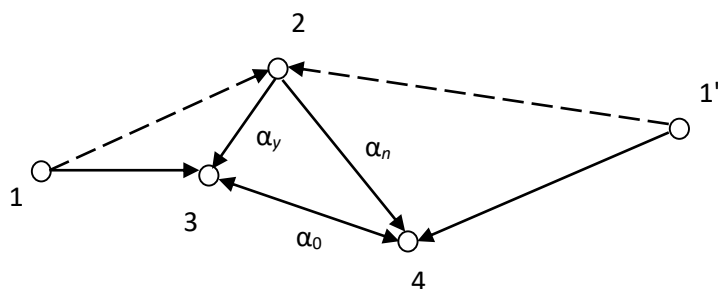


Рис. 1. Логическая схема применения СППР на этапе управления

Fig. 1. DMSS application scheme at the control stage

С учетом такой модели в подавляющем большинстве случаев управляемая система (рис. 1), которая должна была из состояния 1 прийти в состояние 2 (планируемое достижение цели), на самом деле окажется в состоянии 3 (реальный результат). А взаимодействующая (конкурирующая) система или прогнозируемые параметры внешней среды, изначально находящиеся в точке 1', окажутся не в ожидаемом состоянии 2, а в точке реального положения 4. В результате возникает ошибка α_y , являющаяся векторной суммой ошибки отработки управляющих воздействий α_0 и ошибки прогнозирования поведения взаимодействующих (конкурирующих) систем и прогноза влияющих на систему параметров внешней среды α_n , то есть ошибки планирования. Возникает задача реализации управления, обеспечивающего достижение цели действий в условиях, отличных от плановых.

Количество состояний на рисунке 1 описано упрощенно: для одной итерации управления при пошаговом взаимодействии с единственной системой. В реальных условиях матрица состояний будет выглядеть сложнее. Что, впрочем, не меняет формализованный вид описания задачи.

По своей структуре и используемым данным указанная задача во многом похожа на задачу планирования, достаточно успешно решаемую СППР уже на настоящем этапе. Но задача управления имеет ряд отличий, ограничивающих использование обычных, планирующих, СППР в качестве средств поддержки данного этапа:

- необходимость в постоянном и оперативном прогнозировании развития обстановки, как в пассивном, от состояния «как есть» (предиктивная аналитика), так и в активном моделировании, а также в оперативном решении оптимизационных задач для уточнения планов (предписывающая аналитика);

- намного более высокие требования к продолжительности каждого этапа цикла управления – от автоматизированного сбора данных и оценки обстановки до автоматического формирования управляющих (корректирующих) воздействий для уточненного варианта действий;

- ограничение использования алгоритмов итерационного принятия решений, необходимость быстрого нахождения единственно верных вариантов действий.

Иногда, не вполне, впрочем, обоснованно, считая, что на этапе управления решения не принимаются, а контролируются путем мониторинга состояния управляемой системы и среды ее функционирования, а также контроля нахождения параметров в рамках плана, разработчики пытаются решить эту задачу с использованием существующих планирующих СППР в дискретном режиме уточнения планов. Особенно такой подход характерен для разработчиков систем на основе онтологий [11] и при управлении объектами и процессами с высокой инерционностью [15–18]. В то же время создаваемые по классическим принципам СППР, как правило, не ориентированы на высокую динамичность принятия решений, свойственную

процессу управления, а дискретность их применения ограничивается отношением длительности этапа планирования к требуемой продолжительности цикла управления. Но на этапе управления, как указано ранее, достаточно жесткие требования к длительности цикла управления. И обеспечить их выполнение может только реализация специфических требований к управляющим СППР:

- автоматизация сбора и агрегирования (масштабирования) данных обстановки;
- повышение наглядности доведения информации о текущем, целевом и прогнозируемом состоянии управляемой системы;
- реализация проактивных принципов управления;
- автоматизация формирования и доведения управляющих воздействий.

Современная концепция разработки СППР не рассматривает указанные требования как обязательные в связи с тем, что в ее основе лежат принципы, определяемые разработкой планирующих систем [19–21]. Так сложилось исторически как в связи с большей сложностью процесса планирования, так и с недостаточной вычислительной мощностью электронной вычислительной техники (ЭВТ) на начальном этапе становления автоматизации поддержки принятия решений [22]. Но сейчас ситуация изменилась, как и изменились возможности современных информационных технологий. Сформированная задача может и должна быть решена.

Формулировка задачи обеспечения поддержки решений на этапе управления может быть следующей: получение по заданным критериям приемлемого по точности решения в сроки, соответствующие динамичности управляемой системы, с учетом инерционности доведения и выполнения решений.

Вариантов решения такой задачи, как представляется, может быть несколько: от разработки упрощенных алгоритмов (оптимизация по вычислительным ресурсам) до использования растущих вычислительных возможностей современной ЭВТ и средств коммуника-

ции (оптимизация по времени решения за счет снятия ресурсных ограничений).

Простейший из вариантов обеспечивается именно за счет использования возможностей современных информационных технологий: методов обработки больших данных, дополненной (augmented reality, AR) и виртуальной (virtual reality, VR) реальности [23, 24], других подходов, реализующих одну из современных тенденций автоматизации управления – использование вычислительно-емких, но предельно простых и надежных алгоритмов управления за счет существенно возросших возможностей ЭВТ. Такой подход позволит использовать алгоритмы, реализованные и проверенные в существующих планирующих СППР, для обеспечения в них функций поддержки этапа управления.

Насколько эффективен «прямой» подход, может показать практика.

Более сложное решение, с получением строгого математического обоснования и использованием специализированных алгоритмов, оптимизированных как по точности результатов, так и по времени решения, является другим вариантом решения проблемы. Возможность его использования и варианты решения задачи могут быть получены в ходе дальнейших исследований.

Заключение

Таким образом, в статье сформулирована постановка задачи на дополнение недостающего функционала СППР в части автоматизации двух этапов управления: целеполагания и контроля выполнения задач. Научная новизна проведенного исследования заключается в разработке формализованного описания и алгоритмизации этапов целеполагания и управления, обеспечивших постановку ранее в прямом виде не ставившейся задачи выбора и разработки для них математических методов поддержки принятия решений.

В настоящее время необходимый для описания указанных этапов математический ап-

парат в явном виде отсутствует. Но проблема требует решения, и это решение может быть найдено для начала за счет прямого использования возможностей современных информационных технологий, а впоследствии более эффективно – разработкой специализированного математического аппарата. Разработка специализированных целеполагающих и уп-

равляющих СППР или соответствующих блоков в составе существующих систем обеспечит замкнутый цикл автоматизированного управления, полноценное обеспечение поддержки принятия управленческих решений. Есть вероятность, что этому будет способствовать постановка задачи, сформулированная в данной статье.

Список литературы

1. Тиханычев О. В. Об уточнении классификации систем поддержки принятия решений // Информатизация и связь. 2016. № 2. С. 50–53.
2. Тиханычев О. В. О некоторых проблемах предметной области поддержки принятия решений // Программные продукты и системы. 2016. № 3. С. 24–28. DOI: 10.15827/0236-235X.115.024-028.
3. Ларичев О. И., Петровский А. В. Системы поддержки принятия решений. Современное состояние и перспективы их развития. – В книге: Итоги науки и техники. Техническая кибернетика. Т. 21. – М.: ВИНТИ, 1987. С. 115–125.
4. Термины и определения [Электронный ресурс] // Информационные ресурсы ИПИ РАН. URL: http://www.ipiran.ru/niap/index_3.html (дата обращения: 27.12.2020).
5. Eom S. B. Decision support systems research: Reference disciplines and a cumulative tradition // Omega. 1995. Vol. 23. No. 5. P. 511–523. DOI: 10.1016/0305-0483(95)00018-J.
6. Ларичев О. И., Мошквич Е. М. Качественные методы принятия решений. – М.: Наука, 1996. – 272 с.
7. Simonovic S. P. Decision support for sustainable water resources development // Proceeding of the International UNESCO Symposium on Water Resources Planning in a Changing World (Karlsruhe, Germany). 1994. Article 1059. P. 3–13.
8. Ginzberg M. J., Stohr E. A. A decision support system: issues and perspectives. – In book: Processes and tools for decision support. – Amsterdam: North-Holland Publ. Co, 1983. – 136 p.
9. Gorry G. A., Scott-Morton M. S. A framework for management information systems // Sloan Management Review. 1971. Vol. 13. P. 36–43.
10. McDonald J. Designing a decision support system (DSS) for academic library managers using preprogrammed application software on a microcomputer // Library Software Review. 1986. Vol. 5. No. 1. P. 9–15.
11. Skorodid С. П., Голда О. Л. Модель процесу прийняття рішення на застосування резервів (загальновійськових, протидесантних) в операціях угруповань військ // Social Development & Security. 2018. Vol. 4. No. 2. P. 85–92. DOI: 10.5281/zenodo.1237097.
12. Merkert J., Mueller M., Hubl M. A survey of the application of machine learning in decision support systems. – University of Hoffenheim, 2015. – 133 p.
13. Tariq A., Rafi K. Intelligent decision support systems – A framework // Information and Knowledge Management. 2012. Vol. 2. No. 6. P. 12–19.
14. Gibert K., Sánchez-Marré M., Sevilla B. Tools for environmental data mining and intelligent decision support // Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany. 2012. P. 1726–1734.
15. Sánchez-Marré M., Gibert K., Cabello A., Sem F. A methodology for the characterization of intelligent environmental decision support systems // Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany. 2012. P. 153–162.
16. Gibert K., Sánchez-Marré M. A picture on environmental data mining real applications. What is done and how? // Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany. 2012. P. 1612–1619.
17. Gibert K., Conti D., Vrecko D. Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants // Environmental Engineering and Management Journal. Vol. 11. No. 5. P. 931–944. DOI: 10.30638/eemj.2012.117.
18. Ltifi H., Trabelsi G., Ayed M. B., Alimi A. M. Dynamic decision support system based on Bayesian networks. Application to fight against the Nosocomial infections // International Journal of Advanced Research in Artificial Intelligence. 2012. Vol. 1. No. 1. P. 22–29.

19. Микони С. В., Бураков Д. П., Гарина М. И. Инструментальная система для решения задач многокритериального выбора // Программные продукты и системы. 2009. № 4. С. 6–9.
20. Тиханычев О. В. Теория и практика автоматизации поддержки принятия решений: монография. – М.: Эдитус, 2018. – 76 с.
21. Микони С. В. Теория принятия управленческих решений: учебное пособие. – СПб.: Издательство «Лань», 2015. – 448 с.
22. Mikoni S. V. Ordering multiattribute optimization methods from the utility point of view // International Journal Information Technologies & Knowledge. 2013. Vol. 7. No. 1. P. 94–99.
23. Тиханычев О. В. Виртуальная реальность и поддержка принятия решений // Прикладная информатика. 2019. Т. 14. № 4 (82). С. 56–64. DOI: 10.24411/1993-8314-2019-10026.
24. Тиханычев О. В. Технологии «дополненной реальности» как элемент систем поддержки принятия решений // Информатизация и связь. 2017. № 2. С. 63–65.

Сведения об авторе

Тиханычев Олег Васильевич, ORCID 0000-0003-4759-2931, канд. техн. наук, заместитель начальника отдела управления перспективных разработок, ГК «Техносерв», Москва, Россия, tow65@yandex.ru

Статья поступила 27.08.2021, рассмотрена 13.09.2021, принята 06.10.2021

References

1. Tikhanychev O. V. Classification of systems of support of decision-making problems and their solutions. *Informatizatsiya i svyaz*=Informatization and Communication, 2016, no.2, pp.50-53 (in Russian).
2. Tikhanychev O. V. On some problems in the decision support domain. *Programmnye produkty i sistemy*=Software products and systems, 2016, no.3(115), pp.24-28 (in Russian). DOI: 10.15827/0236-235X.115.024-028.
3. Larichev O. I., Petrovsky A. V. *Sistemy podderzhki prinyatiya reshenii. Sovremennoe sostoyanie i perspektivy ikh razvitiya* [Decision support systems. Current state and prospects of their development]. *V knige: Itogi nauki i tekhniki. Tekhnicheskaya kibernetika* [In book: Results of Science and Technology. Technical Cybernetics], vol.21. Moscow, VINITI Publ., 1987, pp.115-125.
4. *Terminy i opredeleniya* [Terms and definitions]. *Informatsionnye resursy IPI RAN*. Available at: http://www.ipiran.ru/niap/index_3.html (accessed 27.12.2020).
5. Eom S. B. Decision support systems research: Reference disciplines and a cumulative tradition. *Omega*, 1995, vol.23, no.5, pp.511-523. DOI: 10.1016/0305-0483(95)00018-J.
6. Larichev O. I., Moshkovich E. M. *Kachestvennye metody prinyatiya reshenii* [Qualitative decision making methods]. Moscow, Nauka Publ., 1996, 272 p.
7. Simonovic S. P. Decision support for sustainable water resources development. *Proceeding of the International UNESCO Symposium on Water Resources Planning in a Changing World* (Karlsruhe, Germany), 1994, article 1059, pp.1-13.
8. Ginzberg M. J., Stohr E. A. A decision support system: issues and perspectives. In book: *Processes and tools for decision support*. Amsterdam, North-Holland Publ. Co, 1983, 136 p.
9. Gorry G. A., Scott-Morton M. S. A framework for management information systems. *Sloan Management Review*, 1971, vol.13, pp.36-43.
10. McDonald J. Designing a decision support system (DSS) for academic library managers using preprogrammed application software on a microcomputer. *Library Software Review*, 1986, vol.5, no.1, pp.9-15.
11. Skorodid S., Golda A. *Model' protsesu pryynyattya rishennya na zastosuvannya rezerviv (zahal'noviyys'kovykh, protydesantnykh) v operatsiyakh uhrupovan' viys'k* [Model of the process of decision-making resolution (general, anti-distinctive) in grouping troops operations]. *Social Development & Security*, 2018, vol.4, no.2, pp.85-92 (in Ukrainian). DOI: 10.5281/zenodo.1237097.
12. Merkert J., Mueller M., Hubl M. A survey of the application of machine learning in decision support systems. *University of Hoffenheim*, 2015, 133 p.
13. Tariq A., Rafi K. Intelligent decision support systems – A framework. *Information and Knowledge Management*, 2012, vol.2, no.6, pp.12-19.
14. Gibert K., Sánchez-Marré M., Sevilla B. Tools for environmental data mining and intelligent decision support. *Proceedings of the International Congress on Environmental Modelling and Software (iEMSs)*, Leipzig, Germany, 2012, pp.1726-1734.

15. Sánchez-Marré M., Gibert K., Cabello A., Sem F. A methodology for the characterization of intelligent environmental decision support systems. Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany, 2012, pp.153-162.
16. Gibert K., Sánchez-Marré M. A picture on environmental data mining real applications. What is done and how? Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany, 2012, pp.1612-1619.
17. Gibert K., Conti D., Vrecko D. Assisting the end-user in the interpretation of profiles for decision support. An application to wastewater treatment plants. Environmental Engineering and Management Journal, vol.11, no.5, pp.931-944. DOI: 10.30638/eenj.2012.117.
18. Ltifi H., Trabelsi G., Ayed M. B., Alimi A. M. Dynamic decision support system based on Bayesian networks. Application to fight against the Nosocomial infections. International Journal of Advanced Research in Artificial Intelligence, 2012, vol.1, no.1, pp.22-29.
19. Mikoni S. V., Burakov D. P., Garina M. I. Instrumental system for solving multi-criteria choice. *Programmnye produkty i sistemy*=Software products and systems, 2009, no.4, pp.6-9 (in Russian).
20. Tikhanychev O. V. *Teoriya i praktika avtomatizatsii podderzhki prinyatiya reshenii: monografiya* [Theory and practice of decision support automation: monograph]. Moscow, *Editus* Publ., 2018, 76 p.
21. Mikoni S. V. *Teoriya prinyatiya upravlencheskikh reshenii: uchebnoe posobie* [Theory of managerial decision making: textbook]. St. Petersburg, *Lan'* Publ., 2015, 448 p.
22. Mikoni S. V. Ordering multiattribute optimization methods from the utility point of view. International Journal Information Technologies & Knowledge, 2013, vol.7, no.1, pp.94-99.
23. Tikhanychev O. V. Virtual reality and decision support. *Prikladnaya Informatika*=Journal of Applied Informatics, 2019, vol.14, no.4(82), pp.56-64 (in Russian). DOI: 10.24411/1993-8314-2019-10026.
24. Tikhanychev O. V. Technology "augmented reality" in decision support systems. *Informatizatsiya i svyaz'*=Informatization and Communication, 2017, no.2, pp.63-65 (in Russian).

About the author

Oleg V. Tikhanychev, ORCID 0000-0003-4759-2931, Cand. Sci. (Eng.), Deputy Head of Department in the Office of Advanced Development, Group of companies "Technoserv", Moscow, Russia, tow@yandex.ru

Received 27.08.2021, reviewed 13.09.2021, accepted 06.10.2021

Исследование цветовых отличий при воспроизведении памятных цветов на устройствах визуализации

О. А. Гурьянова^{1,2}, А. Ю. Гнибеда^{1,3}, Е. В. Филимонова^{1*}

¹ Университет «Синергия», Москва, Россия

² ФГБОУ ДПО «Академия медиаиндустрии», Москва, Россия

³ Московский политехнический университет, Москва, Россия

*elena-gamilton@mail.ru

Аннотация. Данная работа посвящена исследованию изменений цветовых координат на различных устройствах визуализации при воспроизведении цвета, в частности смартфонах как одних из наиболее часто используемых устройств в современном мире, что связано с аппаратной зависимостью цветовоспроизводящей системы. Целью работы является выбор средств визуализации, определение их технических характеристик, определение тестовых цветов для визуализации на различных устройствах, определение допусков в воспроизведении каждого цвета при использовании различных просмотровых устройств. Для реализации цели были поставлены такие задачи, как выбор изображений, содержащих памятные цвета. Данные цвета являются основополагающими при определении допусков в воспроизведении цвета, так как информация о них заложена у каждого человека на основании его жизненного опыта и знаний, и изменение в воспроизведении памятных цветов, нарушение в цветопередаче являются наиболее визуально заметными. Памятные цвета преобразовываются в образцы – тест-объекты, по которым производится определение координат в аппаратно-независимом цветовом пространстве. Производится определение допусков при изменении цветовоспроизведения за счет применения селективной (цветовой) коррекции. При решении задачи было отмечено, что имеются цвета, в которых при небольших изменениях цветовых координат в достаточно сильной степени изменяется визуальная составляющая, при этом другие цвета при численно одинаковом изменении визуально не меняются. Для выбранных цветов на различных средствах визуализации производится расчет разницы цветовоспроизведения и приводятся расчеты цветовых отличий для различных моделей просмотровых устройств при визуальной разнице восприятия. Данное исследование будет полезным при внесении поправки цветовых координат для минимизации цветовой разницы формируемого изображения на различных цветопроизводящих устройствах.

Ключевые слова: цвет, цветовая разница, координаты цвета, цветовое пространство, средство визуализации, селективная коррекция, цветовая коррекция

Для цитирования: Гурьянова О. А., Гнибеда А. Ю., Филимонова Е. В. Исследование цветовых отличий при воспроизведении памятных цветов на устройствах визуализации // Прикладная информатика. 2021. Т. 16. № 6. С. 113–130. DOI: 10.37791/2687-0649-2021-16-6-113-130

Investigation of color differences in the reproduction of memorable colors on visualization devices

O. Guryanova^{1,2}, A. Gnibeda^{1,3}, E. Filimonova^{1*}

¹ Synergy University, Moscow, Russia

² FSBEI CPE "Academy of Media Industry", Moscow, Russia

³ Moscow Polytechnic University, Moscow, Russia

* elena-gamilton@mail.ru

Abstract. This work is devoted to the study of changes in color coordinates on various visualization devices during color reproduction, in particular smartphones, as one of the most commonly used devices in the modern world, which is associated with the hardware dependence of the color reproduction system. The purpose of the work is to select visualization tools, determine their technical characteristics, determine test colors for visualization on various devices, determine the tolerances in reproduction of each color when using various viewing devices. To achieve the goal, such tasks were set as the selection of images containing memorable colors. These colors are fundamental in determining the tolerances in color reproduction, since information about them is inherent in each person on the basis of his life experience and knowledge, and a change in the reproduction of memorable colors, a violation in color rendering, is the most visually noticeable. Memorable colors are converted into samples – test objects, which are used to determine coordinates in a device-independent color space. Determination of tolerances is made when changing color reproduction through the use of selective (color) correction. When solving the problem, it was noted that there are colors in which, with small changes in color coordinates, the visual component changes to a sufficiently strong degree, while other colors, with a numerically identical change, do not visually change. For the selected colors on various visualization tools, the difference in color reproduction is calculated and the calculations of color differences for different models of viewing devices with a visual difference in perception are given.

Keywords: color, color difference, color coordinates, color space, renderer, selective correction, color correction

For citation: Guryanova O., Gnibeda A., Filimonova E. Investigation of color differences in the reproduction of memorable colors on visualization devices. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.113-130 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-113-130

Введение

Областью исследования в данной работе является исследование с применением средств вычислительной техники, в частности обнаружение закономерности в изменении цвета при определенном методе кодирования цветовой информации.

На цифровых устройствах воспроизведение изображения осуществляется с помощью пикселей (точек) трех основных цветов. Каждый пиксель цифрового изобра-

жения несет в себе три основные составляющие информации:

- позиционирование;
- информацию о цвете;
- информацию о яркости.

Информация о цвете передается устройствами благодаря системе (модели) RGB (Red, Green, Blue), или КЗС (красный, зеленый, синий), что является кодированием информации.

Цветовая модель RGB является моделью аддитивного синтеза, основанной на смеше-

нии излучений. В аддитивной RGB-системе все оттенки спектра получаются из различного сочетания трех основных цветов – красного, синего и зеленого (Red, Green, Blue), заданных с разным уровнем интенсивности. Цветовая модель RGB используется при формировании цвета на экранах любых цифровых устройств, воспроизводящих цветную изобразительную информацию.

За счет большого числа элементов зрительная система человека усредняет сигналы, поступающие от пикселей, и формируется представление о цвете [12].

Основным недостатком системы RGB является ее аппаратная зависимость, означающая, что воспроизведение цвета будет зависеть от конкретных технических характеристик цветовоспроизводящего устройства и его текущих настроек. Причина таких отличий в том, что RGB не является стандартизированным пространством, и цвет, отображаемый на различных устройствах, будет зависеть от таких параметров, как:

- соотношение сигналов по трем каналам (RGB);
- интенсивность источника излучения;
- спектральная характеристика выделяющего светочувствительного элемента.

Очень часто можно слышать жалобы пользователей, связанные с тем, что при обработке изображения на компьютере и последующей передаче изображения на телефон цвета выглядят по-разному. Задача данной работы определить разницу в цветопроизведении одних и тех же цветов на различных просмотрных устройствах.

Выбор цветовой модели для проведения измерений

Проблемы, описанные выше, возникли с того момента, когда была разделена обработка изобразительной информации во времени и пространстве. Были выявлены основные задачи воспроизведения, и они послужили основой для создания системы

управления цветом (Color Management System, CMS). Данная система была создана в начале 1990-х годов фирмами KODAK, Microsoft, Heidelberg, Hell, Apple в рамках международного консорциума по цвету (International Color Consortium, ICC).

Основными компонентами CMS являются:

1. Аппаратно-независимое цветовое пространство (Profile Connection Space, PCS – пространство привязки профилей), которое применяется в качестве эталонного, а также в качестве промежуточного звена для перевода изображения из одной цветовой системы в другую.

2. Профили, которые представляют собой таблицы, содержащие информацию о цветовых характеристиках устройств, воспроизводящих цвет.

Управление цветом осуществляется за счет модуля управления цветом (Color Management Module, CMM), основной функцией которого является «расшифровка» информации, содержащейся в профиле цветопроизводящего устройства, и использование ее для перевода из одного цветового пространства в другое. При этом используются различные способы конвертации цветов, которые выходят за пределы цветового охвата того или иного пространства (Rendering Intents, RI).

При формировании аппаратно-независимого цветового пространства используют различные подходы, например привязку цветовых координат к психофизиологическим параметрам ощущения цвета зрительной системой человека либо к четко определенным величинам длин волн видимого диапазона спектра электромагнитного излучения.

Так как телефоны оснащены собственными цветовыми профилями, не осуществляя настройку всех аппаратов под определенные параметры, но можно посчитать цветоразницу с применением аппаратно-независимых пространств.

Для данной работы выбрано цветовое пространство Lab, где: L – светлота; а – хроматический компонент в диапазоне от зеле-

ного цвета до красного; b – хроматический компонент в диапазоне от синего цвета до желтого.

Система Lab применяется для обработки изображений в качестве промежуточного цветового пространства, так как через него осуществляется конвертирование данных. Цветовое пространство Lab, как показали исследования, хорошо согласуется с физиологическим восприятием цвета зрительной системой человека, поскольку учитывает не только изменение длин волн электромагнитного излучения видимого диапазона спектра, но и светлоту. К тому же данное пространство, в отличие от RGB, является равноконтрастным. Данное свойство проявляется в том, что в цветовых моделях аддитивного синтеза незначительное изменение длины волны электромагнитного излучения в одной части пространства, приводящее к визуальному изменению цвета в другой части этого же пространства, может быть незамеченным человеческим глазом. Таким образом, пространство RGB не является равноконтрастным. В свою очередь, пространство Lab, с точки зрения распределения отличий между цветовыми оттенками, является равномерным. При этом возможно определить величину так называемого цветового отличия или цветовых отличий ΔE . Пусть есть две точки: $L_1 a_1 b_1$ и $L_2 a_2 b_2$. Тогда

$$\Delta E = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2 + (L_1 - L_2)^2}. \quad (1)$$

Обычно за величину ΔE принимается значение, приблизительно равное 2. Именно этот коэффициент говорит о том, что разница между оттенками цвета или самими цветами будет визуально заметна. Таким образом, цветовое пространство Lab является одним из наиболее подходящих цветовых пространств для использования в качестве промежуточного в процессе обработки изображения. Это обусловлено следующими свойствами: Lab содержит в себе все цвета, которые может воспринимать зрительный анализатор человека; цветовое пространство является од-

нозначным; имеется возможность оценить цветовые различия при изменении цветовых координат; возможность изменения только чистого цвета, независимо от изменения его светлоты; возможность коррекции светлоты, независимо от цвета; возможность редакционной коррекции, направленной на изменение цвета по известным параметрам цветового тона, насыщенности [9].

Точность при воспроизведении цвета

В работах [7, 10] предлагается использовать три уровня точности для оценки цветовоспроизведения: физическая, физиологическая и психологическая.

Физическая точность, связанная с воспроизведением цвета на оттиске, часто не может быть реализована, так как спектральные характеристики печатных красок очень отличаются от спектральных характеристик красителей изображений.

Физиологическая точность, или колориметрическая точность, означает, что цвета, созданные красителями с разными спектральными характеристиками, визуально будут восприниматься одинаково при одинаковой спектральной характеристике освещения. При изменении спектра освещения цвета становятся визуально различимыми. Достижение физиологической точности возможно только при условии, что цветовой охват изображения оригинала не выходит за пределы цветового охвата применяемых при печатании оттиска красок и бумаги.

Согласно положениям [7], репродукция считается психологически точной, если любое изменение цвета каждого отдельного участка репродукции не сопровождается одновременным изменением цвета в остальных участках.

Иначе говоря, психологическая точность, в отличие от физической и физиологической, является точностью воспроизведения изображения в целом. Психологическая оценка

точности цветовоспроизведения основана на субъективном восприятии. Общеизвестно, что психологически точные репродукции хотя и несут в себе искажения цветов натуральных объектов, но эти изменения находятся в пределах допуска и потому воспринимаются человеком оптимальными (нормальными) по качеству цветопередачи.

Принимая во внимание фундаментальность определения психологической точности [7], становятся нужными конкретные рекомендации для достижения наилучшего воздействия цветом в условиях заданного поля допуска, которое в работе было определено для конкретных изображений, имеющих в своем составе памятные цвета. В данной работе именно психологическая точность была взята как основополагающая, так как она позволяет оценить наиболее частые репродуцируемые объекты – это кожа человека, зелень и голубое небо, а также цвета фруктов и овощей. Цвета могут являться памятными и/или сюжетно-важными. Сюжетно-важные – это цвета, оказывающие наибольшее влияние на зрительное восприятие всего изображения, на впечатление от него. Они содержат воспроизведение наиболее важных семантических участков репродукции. При проведении репродукционного процесса в первую очередь необходимо обеспечить воспроизведение сюжетно-важных цветов внутри допусков психологической точности. Допуски на их воспроизведение минимальны по сравнению с другими цветами изображения:

- в портрете сюжетно-важным является цвет кожи человека;
- в природных пейзажах важен цвет зелени, голубое или закатное красное небо, вода, песок, снег;
- в городских пейзажах – цвет асфальта, красный кирпич, черепица;
- сюжетно-важные цвета в натюрмортах – кожа фруктов и овощей, цветы, продукты питания.

Памятные цвета – цвета природных или любых других объектов, которые хранятся

в памяти человека и которые при отсутствии реальных объектов для сравнения могут отождествляться с цветом предъявляемого ему цветного изображения [2].

Факторы, влияющие на искажение цвета, воспроизводимого различными устройствами

Обработка изображений для обеспечения максимального воздействия на потребителя представляет сложность, так как изображения могут значительно отличаться по своим информативным свойствам. Как уже известно из [10], полное колориметрическое тождество объекта и его изображения в подавляющем большинстве случаев цветного репродуцирования не достижимо и, согласно физической и физиологической точностям, обеспечивающим совпадение по цвету в каждой точке изображения, может задаваться только в случаях, когда репродукция воспроизводится теми же красками и в тех же условиях, что и оригинальное изображение.

Каждое устройство, которое работает с цветом, обладает способностью воспроизводить определенную гамму цветов, т. е. имеет так называемый цветовой охват. Цветовой охват – полный набор всех цветов для данного цветового пространства или устройства. Воспроизводимая гамма зависит от многих факторов, одним из которых является конструкция конкретного устройства, используемого цветового пространства или модели. При этом каждое устройство имеет свой, характерный только для него, цветовой охват [6].

Существует несколько типов матриц: IPS, LED, OLED, TN (TFT), LCD, AMOLED и SuperAMOLED.

В IPS молекулы кристаллов располагаются параллельно, и такое их расположение позволяет добиться широких углов обзора и лучшей цветопередачи, чем при использовании-twisted-молекул [8]. Минусом является то, что при таком расположении кристаллов в матрице увеличивается время от-

клика пикселей. Матрицы IPS показывают высокую контрастность и отличаются большими углами обзора. Этот тип является самым распространенным, часто обозначается как SFT-матрицы.

Технически LED-дисплеев не существует по причине невозможности производства экранов малых диагоналей из светодиодов. Таким термином именуют LCD, которые имеют подсветку из светодиодов.

По мере развития технологий появились OLED. Это те же светодиоды, но органические. OLED-светодиоды подразделяются на активные и пассивные. Различие состоит в том, что в активной OLED-матрице каждым диодом управляет отдельный транзистор. Благодаря этому в такой матрице высокие цветопередача, контрастность изображения и углы обзора [11].

В TN-матрице глубина цвета и контрастности оставляют желать лучшего, углы обзора маленькие. Достаточно повернуть корпус дисплея и цвета станут инвертированными. Современные TN-дисплеи строятся на основе технологии TFT. Различие в том, что TFT-матрица активна. Но даже при помощи этих преобразований данная матрица показывает низкие уровни контраста и цветопередачи.

Главной особенностью AMOLED-матриц является глубина черного цвета.

Технологии не стоят на месте, вследствие чего появился способ удаления воздушной прослойки. Технология OGS является как раз таковой. Стандартно экран телефона состоит из защитного стекла, матрицы и воздушной прослойки между ними. OGS позволяет избавиться от лишнего слоя – воздушной прослойки – и сделать матрицу частью защитного стекла. В результате изображение как будто находится на поверхности стекла, а не скрыто под ним, чем достигается эффект улучшения качества изображения.

В проведенных исследованиях [14] оценка возможностей цветовоспроизведения экрана Samsung Galaxy S7 Edge осуществлялась

при помощи приборов: колориметра X-Rite Display Pro и спектрофотометра i1Pro2. Во избежание некорректных замеров функция «Авторегулировка яркости» в настройках экрана была предварительно отключена. Измерялся стандартный набор, состоящий из 54 контрольных замеров цвета, часто используемый для тестирования и верификации экранов. Сравнение возможностей цветопередачи проводилось относительно стандартного целевого пространства sRGB.

Соответствие возможностей экрана указанным требованиям может считаться хорошим результатом для мониторов и экранов портативных компьютеров при использовании их в среде операционных систем с полноценной поддержкой управления цветом. В случае использования экранов, лишенных этой поддержки, результат в виде удачного сочетания настройки, включающей в себя цветовую температуру, баланс по серому и изначальные возможности экрана, одобренного производителем, будет считаться хорошим [3].

Цветовая температура – один из важнейших параметров при визуальном восприятии изображения. Измеряется цветовая температура в Кельвинах, обозначается буквой «К» и варьируется в зависимости от используемых параметров. Самые низкие показатели у оранжевого и красного свечения, например солнца на закате – до 2000 К, а самые высокие у ярко-синего неба и белоснежных просторов – свыше 7000 К [13].

Освещение делят на 3 основные группы: теплый цвет, имеющий желтоватый оттенок, получают от огня, вольфрамовых ламп и солнца на закате. Он находится в пределах от 1000 до 3500 К: свечка или костер дадут 1000–2000 К, лампы накаливания – около 2400–2800 К, а солнечный свет в начале или в конце дня – 2900–3500 К, ближе к нейтральному. Таким образом, показатели будут зависеть от перечня факторов: погода, время года, облачность и прочие моменты [1].

Методы коррекции для психологически точного воспроизведения цвета

Если говорить о цвете, то с применением программного обеспечения можно повлиять на его изменение. Ниже представлены одни из часто используемых инструментов, а также осуществлен выбор метода коррекции для оценки цвета в проводимом исследовании.

Кривые (Curves) – это важный и незаменимый инструмент для выполнения цветокоррекции по числам. С его помощью мы можем: убедиться, что в изображении задействован весь тональный диапазон, улучшить детализацию в самых важных его областях, а также исправить те цвета, которых в данном изображении быть не должно [9]. Корректировать тональность и цвет изображения можно, меняя форму кривой в диалоговом окне «Кривые». Чем круче части кривой, тем большие области приобретают высокий контраст; лежащие в плоскости части кривой представляют области низкого контраста.

Hue/Saturation (Цветовой тон / Насыщенность) – еще один из инструментов глобальной цветокоррекции, позволяющий настраивать и изменять отдельные цвета, их оттенки, яркость и насыщенность. С осторожностью следует применять его при работе с теми изображениями, для которых замена цвета на другой критична.

Color Balance (Цветовой баланс) – данная коррекция применяется, как правило, если на изображении присутствуют паразитные оттенки, изображение либо слишком холодное, либо слишком теплое и т. д.

Selective Color (Выборочная корректировка цвета) и Replace Color (Замена цвета) используют в случае, если достичь нужного результата методами общей коррекции изображения не получается и необходимо контролировать цвет локально. Смысл использования этих инструментов заключается в точечной коррекции цветов в изображении, как регулируя цветовые составляющие дан-

ного цвета, так и настраивая для него параметры яркости, насыщенности и т. д.

Корректируя параметры цвета, можно добиться психологически точного его воспроизведения [9].

Для проведения исследования нами был выбран метод коррекции Selective Color (Выборочная корректировка цвета) из-за возможности локальной коррекции цвета.

Определение и выбор памятных цветов

Для проведения исследования были выбраны три изображения, в семантике которых содержатся памятные цвета. Каждое изображение имеет следующие информационные свойства:

- тоновое, т. е. имеет неограниченное количество градаций;
- полноцветное, т. е. имеет все многообразие цветов в пределах цветового охвата;
- без шумов, т. е. не имеет никаких дефектов, цифровой шум отсутствует;
- формат цифрового файла – JPG;
- режим – RGB;
- разрешение – 300 ppi;
- размер изображения: ширина – 172 см, высота – 95 см.

Изображение № 1 (рис. 1) – оригинал, содержит цвета фруктов, являющиеся памятными цветами (вишня, виноград, абрикос, киви, персик, клубника).



Рис. 1. Оригинал № 1

Fig. 1. Original no.1

Изображение № 2 (рис. 2) – оригинал, содержит цвета, являющиеся памятными (небо, облака).



Рис. 2. Оригинал № 2
Fig. 2. Original no.2

Изображение № 3 (рис. 3) – оригинал, содержит телесные цвета, цвета природы (девушка, цветы, песок, море).



Рис. 3. Оригинал № 3
Fig. 3. Original no.3

Для проведения измерений необходимо определить точки, являющиеся мишенями, что в дальнейшем позволит определить координаты выбранного цвета без искажений. В качестве примера приведено изображение № 4 на рисунке 4.

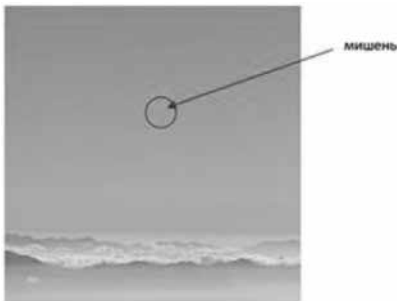


Рис. 4. Мишень для определения координат цвета
Fig. 4. Target for determining color coordinates

В таблице 1 представлены все точки, предназначенные для измерений; они установлены с применением программы и необходимы для того, чтобы измерение производилось в точном месте.

Таблица 1. Точки, предназначенные для точного измерения цветовых координат

Table 1. Points designed to accurately measure color coordinates

Оригинал <i>Original</i>	Точки <i>Points</i>	Изображения с мишенями <i>Target images</i>
Изображение № 1 <i>Image no.1</i>	1	
	2	
	3	
Изображение № 2 <i>Image no.2</i>	1	
Изображение № 3 <i>Image no.3</i>	1	

На каждом изображении выбраны такие мишени:

- Изображение № 1: мишени на вишне (точка 1), персике (точка 2) и киви (точка 3).
- Изображение № 2: мишень на участке неба.
- Изображение № 3: мишень на плече девушки.

После определения всех точек изображения необходимо сравнить цветовые координаты этих точек в цветовом пространстве RGB на просмотрных устройствах.

Средства визуализации

В качестве средств визуализации используются монитор HP Pavilion x360 14-dh0024ur и экран телефона Huawei P30 lite (табл. 2 и 3).

Таблица 2. Технические характеристики монитора

Table 2. Monitor specifications

HP Pavilion x360 14-dh0024ur		
Графический контроллер <i>Graphics controller</i>	Графический контроллер <i>Graphics controller</i>	Intel UHD Graphics 610
Экран <i>Screen</i>	Диагональ/ разрешение <i>Diagonal/ Resolution</i>	14" / 1920×1080 ppi
	Диагональ экрана <i>Screen diagonal</i>	14" (35,5 cm)
Дисплей <i>Display</i>	Технология дисплея <i>Display technology</i>	IPS
	Сенсорный дисплей <i>Touchscreen</i>	Да Yes
Матрица <i>Matrix</i>	Матрица <i>Matrix</i>	1 МПикс

Таблица 3. Технические характеристики телефона

Table 3. Phone specifications

Huawei P30 lite		
Операционная система <i>Operating system</i>	Операционная система <i>Operating system</i>	Android9
Экран <i>Screen</i>	Диагональ/ разрешение <i>Diagonal/ Resolution</i>	6,15" / 2312×1080 ppi
	Диагональ экрана <i>Screen diagonal</i>	6,15" (15,62 cm)
Дисплей <i>Display</i>	Технология дисплея <i>Display technology</i>	IPS
	Количество цветов дисплея <i>Display colors</i>	16,7 млн 16,7 million
	Тип стекла <i>Glass type</i>	2,5D
	Безрамочный <i>Frameless</i>	Да Yes
Фотокамера <i>Camera</i>	Количество основных камер <i>Number of main cameras</i>	3
	Фотокамера <i>Camera</i>	24/8/2 Мпикс 24/8/2 Megapixels
	Стабилизатор изображения <i>Image stabilizer</i>	Цифровой Digital
	Автофокус <i>Autofocus</i>	Да Yes
	Разрешение фронтальной камеры <i>Front camera resolution</i>	32 Мпикс 32 Megapixels

Программное обеспечение для работы с цветом

Для возможности определения цветовых координат на экране монитора использовалась программа Adobe Photoshop, в частности инструмент «Пипетка», который позволяет определить координаты в любом цветовом пространстве. Пример показан на рисунке 5.

Для определения цветовых координат на экране телефона была использована программа Color Lab от компании Gretag Macbeth, функционал которой также позволяет определить координаты изображения в любом цветовом пространстве (рис. 6). Color Lab позволяет просматривать Reference-файлы и выполнять цветовые преобразования, используя систему управления цветом.

Программа, несмотря на свою компактность, очень функциональна, так что перечислим наиболее важные операции [4]:

- составление тестов (так называемых мишеней);

- импорт/экспорт тестовых мишеней;
- цветовые преобразования;
- применение и тестирование профилей;
- просмотр Reference-файлов и файлов измерений.

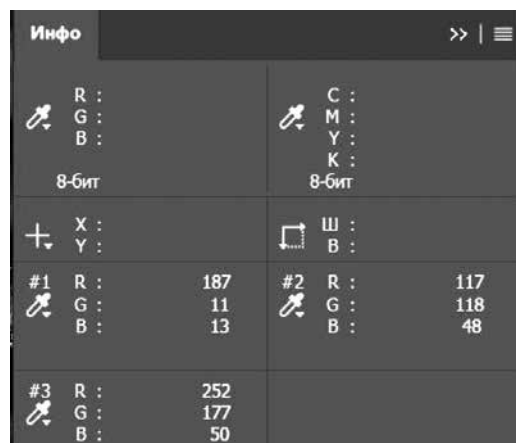


Рис. 5. Инструмент «Пипетка» и координаты цвета в пространстве RGB

Fig. 5. Eyedropper tool and RGB color coordinates

С помощью программного обеспечения проведены измерения цветовых координат

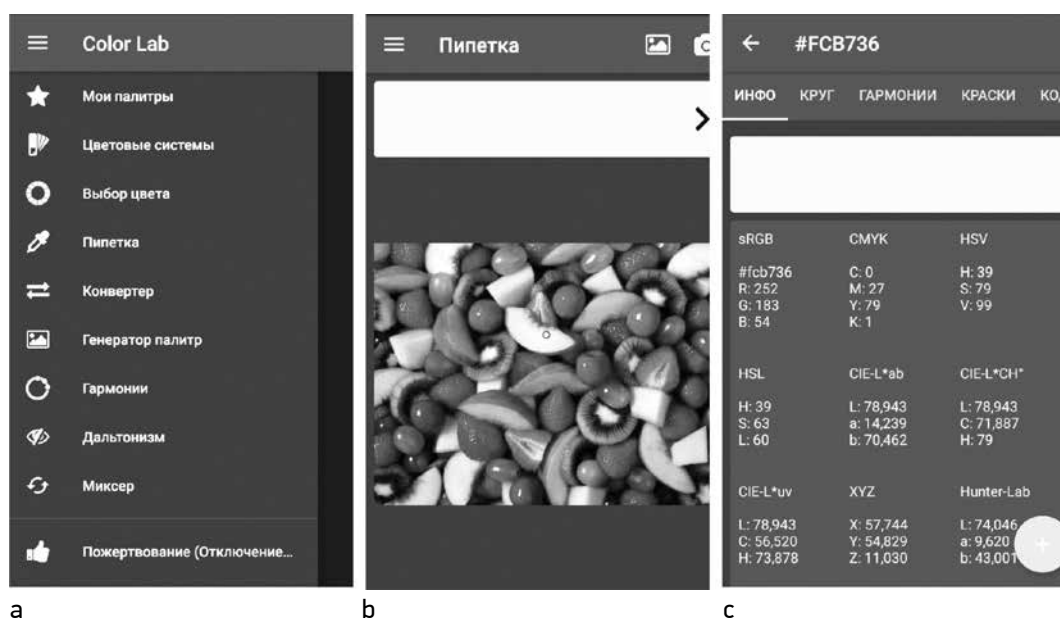


Рис. 6. Функционал программы Color Lab: а – основной инструментарий; б – инструмент «Color Dropper»; в – координаты цвета в различных цветовых пространствах

Fig. 6. Functionality of the Color Lab program: a - basic tools; b - instrument "Color Dropper"; c - color coordinates in different color spaces

оригинальных изображений и выявлено, что, как и предполагалось, изменений нет.

Измерения, проведенные в программе Adobe Photoshop, и измерения, выполненные в программе Color Lab, абсолютно идентичны.

Для определения различного цветовоспроизведения на разных устройствах визуализации необходимо создать условия, в которых возможно определить эту разницу. Поэтому необходимо провести измерения на изображении оригинала и выполненной копии [5].

Измерение цветовых координат в пространствах RGB и Lab

Результаты измерений для трех выбранных точек представлены в таблицах 4 и 5.

С каждого изображения были сделаны копии (копия) и проведены измерения в аппаратно-независимом пространстве.

Таблица 4. Результаты измерений в цветовых пространствах RGB и Lab

Table 4. Measurement results in RGB and Lab color spaces

Оригинал <i>Original</i>	Точки <i>Points</i>	R	G	B	L	a	b
Изображение № 1 <i>Image no.1</i>	1	187	11	13	47	72	61
	2	117	118	48	49	−9	43
	3	252	177	50	81	29	81
Изображение № 2 <i>Image no.2</i>	1	133	181	217	70	−17	−26
Изображение № 3 <i>Image no.3</i>	1	238	182	135	78	17	32

Таблица 5. Результаты измерений копий в цветовых пространствах RGB и Lab

Table 5. Measurement results of copies in RGB and Lab color spaces

<i>Копия</i> <i>Copy</i>	<i>Точки</i> <i>Points</i>	R		G		B		L		a		b	
Изображение № 1 <i>Image no.1</i>	1	216	183	2	11	3	7	54	46	80	71	69	61
	2	121	127	124	129	43	63	51	53	−11	−9	48	39
	3	255	242	179	176	27	53	82	79	29	25	89	78
Изображение № 2 <i>Image no.2</i>	1	115	134	181	221	223	128	70	70	−25	−15	−31	−27
Изображение № 3 <i>Image no.3</i>	1	237	235	180	179	132	132	80	80	24	23	37	36

Определение допусков при изменении цветовоспроизведения за счет проведения селективной коррекции

Для определения допустимых изменений по цвету (цветоразнице) мы воспользовались селективной коррекцией (Color Balance). Изменения проводились по одному цвету, являющемуся дополнительным к красному излучению, – голубому (рис. 7).

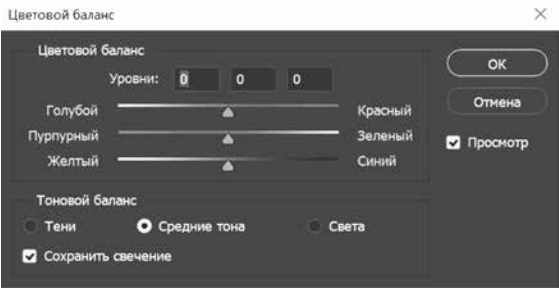


Рис. 7. Цветовой баланс
Fig. 7. Color balance

Измерение цветовых координат выбранных цветов на различных средствах визуализации

Всего выполнено по 5 коррекций для каждой точки. Данные сведены в таблицах 6–10.

Расчет разницы цветовоспроизведения

Из полученных данных видно, что значения координат в цветовом пространстве RGB для оригиналов, открытых в просмотровых устройствах компьютере и телефоне, идентичны. Те малые значения расхождений, которые наблюдаются, связаны с погрешностью

измерения внутривыделенной области (мишени), и они не существенны.

Значения цветовых координат Lab имеют видимые расхождения для различных просмотровых устройств. Расчет разницы различий в воспроизведении цветовых координат позволит понять, в каких цветах имеются ощутимые цветовые отличия и одинаковы ли они в разных цветах.

Расчет цветового отличия ΔE (табл. 11), как было описано выше, можно осуществить по формуле (1), где: $L_1 a_1 b_1$ – значение цветовых координат оригинала; $L_2 a_2 b_2$ – значение цветовых координат копии изображения [5].

Таблица 6. Результаты изменений цветового баланса

Table 6. Results of color balance changes

Изображение № 1 (точка 1) <i>Image no.1 (Points 1)</i>												
Оригинал <i>Original</i>							Скриншоты (монитор / экран телефона) <i>Screenshots (monitor / phone screen)</i>					
№ цв. простр. <i>Color space</i>	R	G	B	L	a	b	R	G	B	L	a	b
1	183	14	16	46	70	59	210/179	2/13	4/23	52/45	78/69	68/59
2	179	17	19	45	69	57	205/175	3/13	6/11	51/44	77/68	66/58
3	174	20	23	44	67	54	200/170	7/14	10/15	50/43	76/67	65/56
4	170	23	27	43	66	51	194/166	9/20	14/20	48/42	74/65	62/53
5	165	26	30	42	64	68	215/161	2/18	3/20	53/40	80/64	69/51

Таблица 7. Результаты изменений цветового баланса

Table 7. Results of color balance changes

Изображение № 1 (точка 2) <i>Image no.1 (Points 2)</i>												
Оригинал <i>Original</i>							Скриншоты (монитор / экран телефона) <i>Screenshots (monitor / phone screen)</i>					
№ цв. простр. <i>Color space</i>	R	G	B	L	a	b	R	G	B	L	a	b
1	110	124	54	49	–17	40	110/116	131/129	53/59	51/52	–22/–17	43/40
2	104	130	60	50	–15	37	96/112	138/141	60/74	52/55	–36/–26	39/34
3	97	136	66	52	–33	35	76/105	140/144	66/77	51/55	–50/–33	34/33
4	91	142	72	53	–41	33	58/99	178/148	71/83	52/55	–64/–38	33/30
5	85	148	78	54	–48	31	121/78	124/139	43/72	51/51	–11/–47	48/30

Таблица 8. Результаты изменений цветового баланса

Table 8. Results of color balance changes

Изображение № 1 (точка 3) <i>Image no.1 (Points 3)</i>												
Оригинал <i>Original</i>							Скриншоты (монитор / экран телефона) <i>Screenshots (monitor / phone screen)</i>					
№ цв. простр. <i>Color space</i>	R	G	B	L	a	b	R	G	B	L	a	b
1	252	181	56	82	27	79	255/242	182/180	34/59	82/80	27/23	87/76
2	252	182	62	83	24	78	255/241	186/185	43/66	83/81	25/19	85/74
3	251	190	68	84	21	76	255/241	194/190	62/72	85/82	21/16	80/72
4	251	193	74	84	20	73	255/240	195/193	62/77	85/83	20/14	73/80
5	251	179	80	85	18	72	255/241	179/195	26/83	82/83	29/14	89/68

Таблица 9. Результаты изменений цветового баланса

Table 9. Results of color balance changes

Изображение № 2 <i>Image no.2</i>												
Оригинал <i>Original</i>							Скриншоты (монитор / экран телефона) <i>Screenshots (monitor / phone screen)</i>					
№ цв. простр. <i>Color space</i>	R	G	B	L	a	b	R	G	B	L	a	b
1	127	185	219	71	-22	-27	93/128	187/184	221/217	69/71	-37/-21	-31/-26
2	121	189	222	72	-26	-27	75/114	192/193	225/223	70/72	-45/-32	-32/-27
3	114	193	224	72	-31	-27	43/123	195/187	226/222	70/71	-54/-24	-33/-28
4	108	197	226	73	-33	-27	34/108	200/198	229/225	71/73	-58/-37	-32/-27
5	102	200	227	73	-40	-27	0/104	202/200	229/224	71/73	-61/-40	-32/-26

Таблица 10. Результаты изменений цветового баланса

Table 10. Results of color balance changes

Изображение № 3 <i>Image no.3</i>												
Оригинал <i>Original</i>							Скриншоты (монитор / экран телефона) <i>Screenshots (monitor / phone screen)</i>					
№ цв. простр. <i>Color space</i>	R	G	B	L	a	b	R	G	B	L	a	b
1	237	186	141	79	15	30	236/232	184/184	138/138	81/80	21/19	35/34
2	236	190	147	80	13	28	234/231	189/186	144/144	82/81	18/18	33/31
3	234	194	152	81	10	27	233/230	192/191	149/148	82/82	16/15	31/30
4	233	198	158	82	9	25	230/229	196/196	157/155	82/83	12/12	27/28
5	231	201	162	83	7	24	229/227	200/199	161/159	84/83	10/9	26/26

Таблица 11. Расчет цветовых отличий между оригиналом и копией на мониторе и экране телефона
Table 11. Calculation of color differences between the original and the copy on the monitor and phone screen

Изображение № 1 <i>Image no.1</i>		
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>	Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>
1	13,30	1,41
2	5,74	5,66
3	8,06	5,39

Выбранные точки соответствуют следующим цветам и оттенкам:

- точка 1 – условно темный оттенок красного цвета;
- точка 2 – оттенок зеленого цвета;
- точка 3 – оттенок оранжевого цвета.

Обозначение оттенков для точек позволит проще отследить изменения цветовых координат по конкретному цвету.

По рассчитанным значениям видно, что воспроизведение цвета на оригинале и копии существенно отличаются, так как отсутствие разницы при зрительном восприятии будет заметно при значении $\Delta E < 1,2$. Это тот допуск, при котором зрительно будет заметна разница в цвете.

Пример различий в цветовоспроизведении показан на рисунке 8 (в градациях серого можно увидеть разницу тона).

В красном цвете разница очевидна, в оттенках зеленого разница слабо заметна, в оттенках оранжевого цвета практически не наблюдается.

Выбранная точка на изображении № 2 соответствует голубому цвету. По голубому цвету цветовые отличия между оригиналом и копией на мониторе и экране телефона не существенна, так же как и в третьей точке изображения № 1, соответствующей зеленому цвету (табл. 12). Отметим, что если рассматривать цветовой спектр, эти цвета являются соседними.

Выбранная точка на изображении № 3 соответствует телесному (бежевому) цвету. Разница в цветовоспроизведении на двух устройствах не критична (табл. 13).

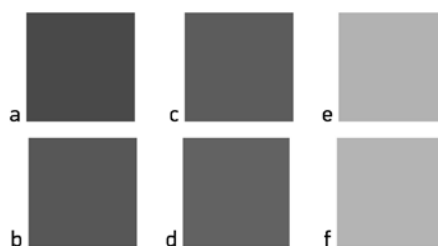


Рис. 8. Различия в цветовоспроизведении:
 а – красный цвет на оригинале, б – красный цвет на копии; с – зеленый цвет на оригинале, d – зеленый цвет на копии; е – оранжевый цвет на оригинале, f – оранжевый цвет на копии
Fig. 8. Differences in color reproduction: a – red color on the original, b - red color on the copy; c – green color on the original, d - green color on the copy; e - orange color on the original, f - orange color on the copy

При проведении селективной коррекции было отмечено, что значения изменения цветового баланса до -40 являлось приемлемым при визуальном восприятии, не нарушалось психологически точное воспроизведение (табл. 14). Это связано с изменением координат только в голубой зоне спектра. Увеличение значения смещения цветовых координат в область голубых тонов изменения приводят к нарушению психологической точности.

При проведении селективной коррекции изменения не были существенными, т. к. проводились только по голубому цвету (табл. 15). Психологически точное восприятие не нарушалось.

Телесный цвет является одним из наиболее сложных при воспроизведении цветов,

Таблица 12. Расчет цветовых отличий между оригиналом и копией на мониторе и экране телефона
Table 12. Calculation of color differences between the original and the copy on the monitor and phone screen

Изображение № 2 <i>Image no.2</i>		
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>	Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>
1	5,74	5,66

Таблица 13. Расчет цветовых отличий между оригиналом и копией на мониторе и экране телефона
Table 13. Calculation of color differences between the original and the copy on the monitor and phone screen

Изображение № 3 <i>Image no.3</i>		
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>	Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>
1	8,06	5,39

Таблица 14. Расчет цветовых отличий после проведения селективной коррекции между оригиналом и копией на мониторе и экране телефона
Table 14. Calculation of color differences after selective correction between the original and the copy on the monitor and phone screen

Селективная коррекция для изображения № 1 <i>Selective correction for image no.1</i>											
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>					Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>					
	Коррекции <i>Corrections</i>					Коррекции <i>Corrections</i>					
	№ <i>No.</i>	1	2	3	4	5	1	2	3	4	5
Значение коррекции <i>Correction value</i>		−20	−40	−60	−80	−100	−20	−40	−60	−80	−100
1	13,45	13,45	15,43	14,49	28,50	1,41	1,73	2,24	2,45	3,61	
2	6,16	11,36	17,06	23,02	40,83	3	5,92	3,61	4,69	3,32	
3	8	7,07	7,87	7,07	20,47	5,39	6,71	6,71	6,78	6	

т. к. в своем составе содержит сочетание всех излучений спектра в разной интенсивности. Небольшое изменение в цветовых координатах приводит к нарушению психологиче- ски точного воспроизведения памятных цве- тов (табл. 16). При значении −20 наблюдалось существенное нарушение в воспроизведении цвета, что видно на рисунке 9.

Таблица 15. Расчет цветовых отличий после проведения селективной коррекции между оригиналом и копией на мониторе и экране телефона

Table 15. Calculation of color differences after selective correction between the original and the copy on the monitor and phone screen

Селективная коррекция для изображения № 2 <i>Selective correction for image no.2</i>										
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>					Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>				
	Коррекции <i>Corrections</i>					Коррекции <i>Corrections</i>				
1	1	2	3	4	5	1	2	3	4	5
	-20	-40	-60	-80	-100	-20	-40	-0	-80	-100
	15,65	19,75	23,85	25,57	21,68	1,41	6	7,14	4	0

Таблица 16. Расчет цветовых отличий после проведения селективной коррекции между оригиналом и копией на мониторе и экране телефона

Table 16. Calculation of color differences after selective correction between the original and the copy on the monitor and phone screen

Селективная коррекция для изображения № 3 <i>Selective correction for image no.3</i>										
Выбранные точки <i>Selected points</i>	Цветовые отличия между оригиналом и копией на мониторе, ΔE <i>Color differences between the original and the copy on the monitor, ΔE</i>					Цветовые отличия между оригиналом и копией на экране телефона, ΔE <i>Color differences between the original and the copy on the phone screen, ΔE</i>				
	Коррекции <i>Corrections</i>					Коррекции <i>Corrections</i>				
1	-20	-40	-60	-80	-100	-20	-40	-60	-80	-100
	1	2	3	4	5	1	2	3	4	5
	8,06	7,35	7,28	3,61	3,74	5,74	5,92	5,92	4,36	2,83



Рис. 9. Различия в цветовоспроизведении при проведении селективной коррекции:
а – цвет на оригинале; б – цвет после изменения цветовых координат (-20)

Fig. 9. Differences in color reproduction during selective correction: a - color on the original; b - color after changing color coordinates (-20)

Заключение

В заключение работы можно сделать следующие выводы:

1. Воспроизведение цвета на оригинале и его визуальной копии на экране телефона отличается в достаточно сильной степени. В исследуемых цветах одинаковые изменения в каждом отдельно взятом цвете дают различные параметры цветоразницы.

2. В воспроизведении красного цвета разница очевидна при небольших изменениях цветовых координат, в оттенках зеленого разница визуально заметна в гораздо меньшей степени. В оттенках оранжевого цвета визуально разница слабоотличима, т. е. все изменения находятся в допустимых пределах.

3. Преобразование цветовых координат с применением селективной коррекции по-разному влияет на восприятие цвета и в разных цветах эта разница также заметна. Наибольшие отличия в цветовоспроизведении наблюдаются в красных и телесных цветах. Даже небольшое изменение критериев в цветовых координатах приводит к нарушению психологически точного воспроизведения.

4. Психологическая точность воспроизведения памятных цветов на различных средствах визуализации нарушается в различных цветах в разной степени.

5. Заранее зная, какая может быть погрешность в воспроизведении цвета, можно смоделировать процесс воспроизведения, минимизируя цветоразницу.

Список литературы

1. ГОСТ Р 52490–2005 (ИСО 7724–3:1984). Материалы лакокрасочные. Колориметрия. Часть 3. Расчет цветовых различий [Электронный ресурс] // Электронный фонд нормативно-технической и нормативно-правовой информации. URL: <https://docs.cntd.ru/document/1200043561> (дата обновления: 03.12.2021).
2. Базыма Б. А. Цвет и психика. – М.: Речь, 2001. – 100 с.
3. Гудилин Д. Гибридные растры, или «Мичуринские» опыты в полиграфии [Электронный ресурс] // КомпьюАрт. 2005. № 4. URL: <https://compuart.ru/article/8704> (дата обращения: 24.11.2021).
4. Гурьянова О. А., Андреев Ю. С. Классификация оригиналов для выбора растровой структуры, обеспечивающей качественное репродуцирование // Известия высших учебных заведений. Проблемы полиграфии и издательского дела. 2014. № 6. С. 20–27.
5. Гурьянова О. А., Филимонова Е. В. Подход к повышению резкости изображения путем выделения контурной информации программными и аппаратными средствами // Прикладная информатика. 2019. Т. 14. № 1 (79). С. 123–141.
6. Заболоцкая М., Андреев Ю. Принципы определения допусков на цветовоспроизведение в полиграфии. – М.: Полиграфист & Издатель, 1998. – 163 с.
7. Лауберт Ю. К., Медовщиков И. А., Нюберг Н. Д. Цветная фотомеханическая репродукция. – М.: Гизлегпром, 1941. – 117 с.
8. Люшер М. Цвет вашего характера. – М.: РИПОЛ классик, 1997. – 240 с.
9. Маргулис Д. Photoshop для профессионалов: классическое руководство по цветокоррекции: пер. с англ. – 5-е изд. – М.: Интерсофтмарк, 2007. – 638 с.
10. Нюберг Н. Д. Теоретические основы цветовой репродукции. – М.: Советская наука, 1948. – 176 с.
11. Одиноких Г. А. Разрешающая способность видеосистем для задач видеоаналитики [Электронный ресурс] // Молодежный научно-технический вестник. 2013. № 12, декабрь. С. 1–9. URL: <http://ainsnt.ru/doc/649755.html> (дата обращения: 24.11.2021).
12. Стефанов С. Полиграфия для рекламистов и не только. Руководство к действию. – М.: Гелла-Принт, 2002. – 352 с.
13. Филимонова Е. В. Информационные технологии в профессиональной деятельности. – М.: КноРус, 2018. – 482 с.
14. Экран смартфона Samsung Galaxy S7 Edge [Электронный ресурс] // Цветовое профилирование устройств. URL: <http://koler.by/index.php?id=samsung-galaxy-s7-edge> (дата обновления: 05.11.2019).

Сведения об авторах

Гурьянова Ольга Александровна, ORCID 0000-0003-4974-3166, канд. техн. наук, доцент, кафедра дизайна и рекламы, Университет «Синергия»; Федеральное государственное бюджетное образовательное учреждение

дополнительного профессионального образования «Академия медиаиндустрии», Москва, Россия, goa0208@yandex.ru

Гнибеда Артём Юрьевич, ORCID 0000-0003-2442-4951, старший преподаватель, кафедра информационного менеджмента и информационно-коммуникационных технологий имени профессора В. В. Дика, Университет «Синергия»; Московский политехнический университет, Москва, Россия, artemgnibeda@yandex.ru

Филимонова Елена Викторовна, ORCID 0000-0002-9791-7610, канд. техн. наук, доцент, департамент цифровой экономики, Университет «Синергия», Москва, Россия, elena-gamilton@mail.ru

Статья поступила 20.10.2021, рассмотрена 06.11.2021, принята 22.11.2021

References

1. GOST R 52490–2005 (ISO 7724–3:1984). *Materialy lakokrasochnye. Kolorimetriya. Chast' 3. Raschet tsvetovykh razlichii* [State Standard R 52490-2005 (ISO 7724-3:1984). Paints and varnishes. Colorimetry. Part 3. Calculation of color differences]. *Elektronnyi fond normativno-tekhnicheskoi i normativno-pravovoi informatsii*. Available at: <https://docs.cntd.ru/document/1200043561> (accessed 03.12.2021).
2. Bazyma B. A. *Tsvet i psikhika* [Color and psyche]. Moscow, Rech' Publ., 2001, 100 p.
3. Gudilin D. *Gibridnye rastry, ili «Michurinskie» opyty v poligrafii* [Hybrid rasters or "Michurin" experiments in printing]. *KompyuArt*, 2005, no.4. Available at: <https://compuart.ru/article/8704> (accessed 24.11.2021).
4. Guryanova O. A., Andreev Yu. S. Classification of originals to select screening structures providing high quality reproduction. *Izvestiya vysshikh uchebnykh zavedenii. Problemy poligrafii i izdatel'skogo dela*, 2014, no.6, pp.20-27 (in Russian).
5. Guryanova O. A., Filimonova E. V. Approach to enhance the sharpness of the image by separating the contour information of software and hardware. *Prikladnaya informatika*=Journal of Applied Informatics, 2019, vol.14, no.1(79), pp.123-141 (in Russian).
6. Zabolotskaya M., Andreev Yu. *Printsipy opredeleniya dopuskov na tsvetovospromozhdenie v poligrafii* [Principles for determining tolerances for color reproduction in printing]. Moscow, Poligrafist & Izdatel' Publ., 1998, 163 p.
7. Laubert Yu. K., Medovshchikov I. A., Nyuberg N. D. *Tsvetnaya fotomekhanicheskaya reproduksiya* [Color photomechanical reproduction]. Moscow, Gizlegprom Publ., 1941, 117 p.
8. Lusher M. *Tsvet vashego kharaktera* [The color of your character]. Moscow, RIPOL klassik Publ., 1997, 240 p.
9. Margulis D. *Photoshop dlya professionalov: klassicheskoe rukovodstvo po tsvetokorreksii* [Photoshop for professionals: classic guide to color correction]. Transl. from English. 5th edition. Moscow, Intersoftmark Publ., 2007, 638 p.
10. Nyuberg N. D. *Teoreticheskie osnovy tsvetovoi reproduksii* [Theoretical foundations of color reproduction]. Moscow, Sovetskaya nauka Publ., 1948, 176 p.
11. Odinokih G. A. The resolving capacity of the optical systems in the intelligent surveillance systems applications. *Molodezhnyi nauchno-tekhnicheskii vestnik*=Youth scientific and technical bulletin, 2013, no.12, December, pp.1-9 (in Russian). Available at: <http://ainsnt.ru/doc/649755.html> (accessed 24.11.2021).
12. Stefanov S. *Poligrafiya dlya reklamistov i ne tol'ko. Rukovodstvo k deistviyu* [Polygraphy for advertisers and not only. Guide to Action]. Moscow, Gella-Print Publ., 2002, 352 p.
13. Filimonova E. V. *Informatsionnye tekhnologii v professional'noi deyatel'nosti* [Information technology in professional activities]. Moscow, KnoRus Publ., 2018, 482 p.
14. *Ekran smartfona Samsung Galaxy S7 Edge* [Samsung Galaxy S7 Edge smartphone screen]. *Tsvetovoe profilirovanie ustroistv*. Available at: <http://koler.by/index.php?id=samsung-galaxy-s7-edge> (accessed 05.11.2019).

About the authors

Olga A. Guryanova, ORCID 0000-0003-4974-3166, Cand. Sci. (Eng.), Associate Professor, Design and Advertising Department, Synergy University; Federal State Budgetary Educational Institution of Continuing Professional Education "Academy of Media Industry", Moscow, Russia, goa0208@yandex.ru

Artem Yu. Gnibeda, ORCID 0000-0003-2442-4951, Senior Lecturer, Information Management and Information and Communication Technologies Department named after Professor V. V. Dik, Synergy University; Moscow Polytechnic University, Moscow, Russia, artemgnibeda@yandex.ru

Elena V. Filimonova, ORCID 0000-0002-9791-7610, Cand. Sci. (Eng.), Associate Professor, Digital Economy Department, Synergy University, Moscow, Russia, elena-gamilton@mail.ru

Received 20.10.2021, reviewed 06.11.2021, accepted 22.11.2021

Задача параметрического программирования с моделями прогнозирования урожайности сельскохозяйственных культур

Я. М. Иваньо¹, С. А. Петрова¹, М. Н. Барсукова^{1*}, Ю. В. Столопова¹

¹ ФГБОУ ВО «Иркутский государственный аграрный университет имени А. А. Ежевского»,
Молодежный, Россия
*margarita1982@bk.ru

Аннотация. В работе рассмотрены факторные модели, позволяющие прогнозировать урожайность сельскохозяйственных культур. Показано, что основными климатическими параметрами, влияющими на результативный признак, являются температура воздуха и осадки в начальный период вегетации. При этом факторы теплообеспечения и влагообеспечения могут представлять собой значения как за месяц, так и иной интервал, близкий к такой продолжительности. Кроме температуры воздуха и осадков на урожайность зерновых культур оказывает влияние время. Модели могут отражать связи результативного признака с факторами на уровне опытных полей, сельскохозяйственных организаций, муниципальных районов. Наличие значимых регрессионных зависимостей, которые могут быть линейными и нелинейными, позволяет уменьшить неопределенность задачи оптимизации производства аграрной продукции путем сокращения случайных и интервальных параметров. Приведена модель параметрического программирования с учетом выражений, характеризующих связи урожайности зерновых культур с метеорологическими параметрами, в двух вариантах для получения оптимальных планов производства аграрной продукции товаропроизводителем. Рассмотрен пример реализации оптимизационной модели для реального хозяйства. Предложенная модель создана для поддержки принятия решения в условиях неопределенности. Работа выполнена согласно статистическим данным об урожайности пшеницы, ячменя и овса по Усольскому, Черемховскому и Иркутскому районам за 1997–2018 гг.; на основе урожайности сортоучастков по Усольскому, Иркутскому, Братскому и Нукутскому районам за 2000–2018 гг. (по данным Госсортокмиссии); на основе урожайности ООО «Сибирская Нива» за период 2005–2018 гг. Кроме того, использованы суточные температуры воздуха и суточные осадки в промежутке май–август за 1997–2018 гг. по метеорологическим пунктам: Усолье-Сибирское, Черемхово, Иркутск и Братск.

Ключевые слова: факторная модель, урожайность, аграрное производство, неопределенность, задача параметрического программирования

Для цитирования: Иваньо Я. М., Петрова С. А., Барсукова М. Н., Столопова Ю. В. Задача параметрического программирования с моделями прогнозирования урожайности сельскохозяйственных культур // Прикладная информатика. 2021. Т. 16. № 6. С. 131–143. DOI: 10.37791/2687-0649-2021-16-6-131-143

Parametric programming problem with crop yield forecasting models

Ya. Ivanyo¹, S. Petrova¹, M. Barsukova^{1*}, Yu. Stolopova¹

¹ Irkutsk State Agricultural University named after A. A. Ezhevsky, Molodezhny, Russia

* margarita1982@bk.ru

Abstract. The paper considers factor models that allow predicting the yield of agricultural crops. It is shown that the main climatic parameters that affect the effective feature are the air temperature and precipitation during the initial growing season. In this case, the factors of heat supply and moisture supply can represent values for both a month and another interval close to this duration. In addition to air temperature and precipitation, the yield of grain crops is affected by time. Models can reflect the relationship of the effective feature with factors at the level of experimental fields, agricultural organizations, and municipal districts. The presence of significant regression dependencies, which can be linear and nonlinear, reduces the uncertainty of the problem of optimizing agricultural production by reducing random and interval parameters. A model of parametric programming is presented, taking into account the expressions that characterize the relationship between the yield of grain crops and meteorological parameters in two variants, in order to obtain optimal plans for the production of agricultural products by the commodity producer. An example of the implementation of an optimization model for a real economy is considered. The proposed model is designed to support decision-making in conditions of uncertainty. The work is carried out according to statistical data on the yield of wheat, barley and oats in the Usolsky, Cheremkhovsky and Irkutsk districts for 1997-2018; based on the yield of variety plots in the Usolsky, Irkutsk, Bratsky and Nukutsky districts for 2000-2018 according to the data of the State Export Commission; based on the yield of LLC "Sibirskaya Niva" for the period 2005-2018. In addition, daily air temperatures and daily precipitation in the period May–August for 1997-2018 were used for meteorological points: Usolye-Sibirskoye, Cheremkhovo, Irkutsk and Bratsk.

Keywords: factor model, yield, agricultural production, uncertainty, parametric programming problem

For citation: Ivanyo Ya., Petrova S., Barsukova M., Stolopova Yu. Parametric programming problem with crop yield forecasting models. *Prikladnaya informatika*=Journal of Applied Informatics, 2021, vol.16, no.6, pp.131-143 (in Russian). DOI: 10.37791/2687-0649-2021-16-6-131-143

Введение

С помощью методов параметрического программирования решаются различные задачи оптимизации в разных отраслях экономики [1, 3, 4]. Необходимость использования подобных задач определена множеством причин. Основная из них обусловлена тем, что исходные данные для численного решения любой реальной задачи оптимизации практически всегда определяются приближенно или могут изменяться под влиянием каких-то фак-

торов, что может существенно сказаться на оптимальности выбираемой программы (плана) действий. При решении прикладных задач параметрической оптимизации в качестве параметров могут выступать время, температура, цена на сырье, трудовые ресурсы, удаленность от поставщиков и другие.

С математической точки зрения параметрическое программирование выступает как одно из средств анализа чувствительности решения к вариации исходных данных, оценки устойчивости решения. Следует отметить,

что использование значимых регрессионных выражений в моделях параметрического программирования позволяет получать оптимальные решения с некоторым упреждением.

Между тем сильные колебания температур воздуха, неравномерность выпадения осадков во времени и пространстве создают значительную неопределенность для ведения сельского хозяйства, планирования и прогнозирования производственно-экономических показателей [6, 7, 11, 13]. При этом изменение климата в значительной степени влияет на биопродуктивность сельскохозяйственных культур [21]. Большое значение в этих условиях имеет оптимизация экономических показателей [20] и управление рисками в аграрном производстве [19].

Ситуация с высоким рассеянием климатических показателей имеет место на территории Восточной Сибири [5, 6, 9], где получение продовольственной продукции осуществляется в условиях высоких рисков [15, 16]. Между тем возможны частные случаи, когда некоторые показатели деятельности сельскохозяйственных товаропроизводителей удастся описать регрессионными зависимостями, снимая часть неопределенности задачи [2, 12, 14, 17, 18]. В этом случае для планирования эффективны методы параметрической оптимизации [1, 10].

В продолжение работ, связанных с разработкой и применением задач параметрического программирования производства продовольственной продукции [3, 4], предлагается оптимальная модель, в которой урожайность сельскохозяйственных культур связана значимыми регрессионными уравнениями с метеорологическими факторами на разных стадиях вегетационного периода.

Исходя из этого, целью данной работы является разработка моделей прогнозирования урожайности сельскохозяйственных культур для решения задач оптимизации производства сельскохозяйственной продукции с использованием методов параметрического программирования.

Для достижения цели были поставлены следующие задачи:

- определение адекватных моделей изменчивости урожайности сельскохозяйственных культур;
- построение задач параметрического программирования для оптимизации производства сельскохозяйственной продукции для хозяйств и муниципальных районов;
- апробация моделей параметрического программирования на сельскохозяйственных предприятиях.

Основные результаты работы

В ряде статей показано, что на территории Иркутской области имеет место тенденция потепления климата на основе анализа температур воздуха по годам и сезонам и безморозного периода за многолетие [8]. Тренд температуры воздуха и безморозного периода проявляется в большей степени на юге. При этом тенденция роста отмечена с середины 40-х годов прошлого века. Следует также отметить, что при исследовании трендов показателей тепла по данным XXI в. тенденция потепления климата в регионе нарушается.

На фоне непредсказуемости изменчивости климатических показателей за последнее двадцатилетие наблюдаются значимые связи между урожайностью зерновых культур и факторами тепла и влагообеспеченности.

На основе исследования влияния суточной температуры воздуха и суточных осадков в период вегетации за многолетие показано их значимое влияние на урожайность пшеницы, ячменя и овса на уровне опытных участков, сельскохозяйственных предприятий и муниципальных образований [8]. Полученные факторные модели имеют линейный и нелинейный вид. В таблице 1 обобщены эмпирические модели, которые значимы по F -критерию Фишера и t -статистикам Стьюдента. Они характеризуют разные территории исследования – сортоучастки и муниципальные районы. При этом полученные

Таблица 1. Факторные модели урожайности зерновых культур, полученные по данным за 2000–2018 гг. для Иркутской области**Table 1.** Factor models of grain yield obtained from data for 2000–2018 for the Irkutsk region

Территория <i>Territory</i>	Культура <i>Culture</i>	Уравнение регрессии <i>Regression equation</i>	Обозначения <i>Designations</i>
Сортоучастки <i>Varietystages</i>	Пшеница, овес <i>Wheat, oats</i>	$y = a_0 + a_1z + a_2r$	z – средняя температура воздуха за май или июнь; r – сумма осадков за май или июнь; t – время (годы); r_m – наибольшие суточные осадки в мае; z_b – сумма средних температур по месяцам за май–август z is the average air temperature for May or June, r is the sum of precipitation for May or June, t is the time (years), r_m is the highest daily precipitation in May, z_b is sum of average temperatures by months for May–August
		$y = a_0 + a_1r + a_2t$	
		$y = a_0 + a_1rz + a_2t$	
		$y = a_0 + a_1r$	
		$y = a_0 + a_1rz$	
Муниципальные районы <i>Municipal districts</i>	Пшеница, овес, ячмень <i>Wheat, oats, barley</i>	$y = a_0 + a_1r + a_2z$	
		$y = a_0 + a_1r_m + a_2z$	
		$y = a_0 + a_1r_m + a_2z_b + a_3t$	
		$y = a_0 + a_1r_m + a_2t$	

регрессионные уравнения применимы для сельскохозяйственных организаций.

Следует иметь в виду, что выбор месяца при оценке температур воздуха и сумм суточных осадков зависит от географических координат территории. Обычно на урожайность зерновых на юге в большей степени влияют показатели тепла и влагообеспечения в мае, а на севере – в июне. В дополнение к факторам z и r на биопродуктивность могут влиять наибольшие суточные осадки в начале вегетационного периода r_m , а также температуры за теплый сезон z_b [8]. Наблюдаются случаи влияния на урожайность зерновых культур времени t , что связано с наличием трендов в многолетних рядах y . В приведенных формулах началом отсчета является 2000 год, за номер которого принята единица. Обращает на себя внимание тот факт, что не всегда удается получить значимые регрессионные выражения. Это, в частности, касается урожайности ячменя на основе данных исследований сортоучастков. Вместе с тем при использовании корреляционно-регрессионного анализа выявлено, что на результативный признак этой культуры влияют те же факторы, что и на другие культуры, – температура воздуха и осадки в начальные месяцы вегетации.

Анализ факторов математических моделей, приведенных в таблице 1, показывает, что они

могут быть описаны законами распределения вероятностей. Это касается температуры воздуха в начальный период вегетации и за весь теплый сезон, а также сумм осадков и наибольших суточных осадков в первые месяцы развития сельскохозяйственных растений. В таблице 2 приведены примеры полученных законов распределения вероятностей на основе свойств вероятностных функций и критерия согласия Колмогорова.

В таблице 2 приведены два закона распределения вероятностей – нормальный и гамма-распределение. В ряде источников литературы [2, 3] показано, что температуры воздуха на различных отрезках времени за многолетие могут быть описаны функцией Гаусса, а осадки – асимметричными аналитическими выражениями, в частности гамма-распределением.

Таким образом, хотя урожайность зерновых культур связана с метеорологическими факторами значимыми зависимостями, возникает проблема предсказания факторов для прогноза результативного признака. Однако в приведенном случае значения урожайности сельскохозяйственных культур (конец вегетационного периода) связаны с предшествующими факторами (начало вегетации), поэтому с определенной точностью могут

Таблица 2. Законы распределения вероятностей температуры воздуха и осадков за оптимальный период начала вегетации (по данным Иркутска за 2005–2018 гг.)
Table 2. The laws of distribution of probabilities of air temperature and precipitation for the optimal period of the beginning of the growing season according to Irkutsk data for 2005–2018

Показатель <i>Indicator</i>	Закон распределения вероятностей <i>The law of probability distribution</i>	Обозначения <i>Designations</i>
Сумма суточных температур воздуха за 25.05–25.06 <i>The sum of daily air temperatures for 25.05-25.06</i>	$p(z_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z_i - \bar{z})^2}{2\sigma^2}}$	z_i – ежегодная сумма суточных осадков за 25.05–25.06; \bar{z} – среднее значение ряда ($\bar{z} = 495^\circ\text{C}$); σ – стандартное отклонение ($\sigma = 41,4^\circ\text{C}$) z_i – the annual sum of daily precipitation for 25.05-25.06; \bar{z} – the average value of the series ($\bar{z} = 495^\circ\text{C}$); σ – the standard deviation ($\sigma = 41,4^\circ\text{C}$)
Сумма суточных осадков за 25.05–25.06 <i>Total daily precipitation for 25.05-25.06</i>	$p(r_i) = \left(\frac{r_i}{b}\right)^{c-1} \frac{e^{-\frac{r_i}{b}}}{b\Gamma(c)}$	$\Gamma(c)$ – гамма-функция; b, c – параметры, связанные со средним значением ряда (\bar{r}) и коэффициентом вариации (c_v); $\bar{r} = 62.9$ мм, $c_v = 0,52$ $\Gamma(c)$ – gamma function; b, c – parameters associated with the average value of the series (\bar{r}) and the coefficient of variation (c_v); $\bar{r} = 62.9$ мм, $c_v = 0,52$

быть получены прогнозы биопродуктивности. Вместе с тем наличие законов распределения вероятностей позволяет оценивать различные варианты развития ситуаций, учитывая вариации сумм осадков и температуры воздуха.

Поскольку урожайность сельскохозяйственных культур входит в левые части моделей оптимизации производства растениеводческой и животноводческой продукции и их сочетания, в этом случае для планирования применимы задачи параметрического программирования. Приведем модель оптимизации сочетания получения сельскохозяйственной продукции с метеорологическими параметрами в левых частях ограничений.

Целевая функция такой модели определяется в виде дохода

$$\sum_{s \in S} d_s x_s + \sum_{h \in H} d_h x_h \rightarrow \max, \tag{1}$$

где d_s – стоимость единицы продукции s -культуры ($s \in S$), тыс. руб.;

d_h – стоимость единицы h -вида животноводческой продукции ($h \in H$), тыс. руб.;

x_s – искомая переменная: объем производства s -культуры, ц;

x_h – искомая переменная: продукция животноводства h -вида;

S – множество видов культур;

H – множество животноводческой продукции.

В приведенной модели использованы следующие ограничения.

Неравенство по распределению производственных ресурсов имеет вид

$$\sum_{s \in S} b_{is}(z, r, t) x_s \leq B_i \quad (i \in I), \tag{2}$$

где $b_{is}(z, r, t)$ – расход i -го ресурса на единицу продукции s -культуры, га, руб.;

z, r и t – метеорологические факторы и время;

B_i – ограничения по использованию ресурсов, га, руб.;

I – множество видов ресурсов.

Ограничение по применению в животноводстве побочной продукции растениеводства, описывающее количество продукции для обеспечения кормами животных записывается так:

$$\sum v_{js} x_s \geq x_j \quad (j \in J), \tag{3}$$

где v_{js} – выход с единицы s -культуры j -вида корма;

x_j – объем производства кормов вида j .

Ограничение по размеру отрасли растениеводства определяется следующим образом:

$$\underline{\psi}_w \leq \sum_{s \in S} (1 + \alpha_s) x_s \leq \bar{\psi}_w \quad (w \in W), \quad (4)$$

где $\underline{\psi}_w$ ($\bar{\psi}_w$) – минимальная (максимальная) площадь культур w -вида (группы);

α_s – коэффициент, учитывающий площадь посевов семян для s -культур;

W – множество агротехнических групп культур.

Для животноводства аналогичное ограничение имеет вид

$$x_h = \lambda_{hh'} x_{h'} \quad (h, h' \in H), \quad (5)$$

где $\lambda_{hh'}$ – коэффициент пропорциональности между поголовьем животных h и их группами h' ;

h' – группы животных.

Сформулируем неравенства, характеризующие производство конечной продукции:

- растениеводства:

$$\sum_{s \in S} u_{q_1 s}(z, r, t) x_s \geq U_{q_1} \quad (q_1 \in Q_1); \quad (6)$$

- животноводства:

$$\sum_{h \in H} u_{q_2 h} x_h \geq U_{q_2} \quad (q_2 \in Q_2), \quad (7)$$

где $u_{q_1 s}(z, r, t)$ и $u_{q_2 h}$ – выход продукции растениеводства s -культуры и животноводства h -вида, ц;

U_{q_1} и U_{q_2} – заданный объем производства продукции растениеводства и животноводства, ц;

q_1 и q_2 – вид товарной продукции;

Q_1 и Q_2 – множества товарной продукции.

Баланс рационов животных по элементам питания имеет вид

$$\sum_{s \in S} a_{is} l_s x_s + \sum_{j \in J} a_{ij} x_j \geq \sum_{h \in H} \beta_{ih} x_h \quad (i \in I), \quad (8)$$

где a_{is} – содержание элемента питания i в единице кормовой продукции, получаемое от культуры s ;

l_s – выход основной кормовой продукции от культуры s ;

a_{ij} – содержание элемента i питания в виде корма j или компоненте кормосмеси;

β_{ih} – минимальная потребность в элементе питания i единицы поголовья вида (группы) h .

По структуре производства кормов имеет место неравенство

$$\begin{aligned} \sum_{h \in H} \varphi_{kh} &\leq \sum_{s \in S} a_{is} l_s x_s + \sum_{j \in J} a_{ij} x_j \leq \\ &\leq \sum_{h \in H} \bar{\varphi}_{kh} x_h \quad (k \in K), \end{aligned} \quad (9)$$

где φ_{kh} , $\bar{\varphi}_{kh}$ – минимально и максимально допустимый нормативный размер потребности в кормах группы k единицы поголовья вида (группы) животных h , выраженный в кормовых единицах.

Искомые переменные должны быть неотрицательными:

$$x_s, x_h \geq 0, \quad (10)$$

В приведенной модели ограничения (2) и (6) содержат в себе регрессионные выражения, которые могут принимать вид, приведенный в таблице 1. При этом суммы осадков и температуры воздуха в начальный период вегетации могут подчиняться законам распределения вероятностей, например нормальному, гамма-распределению и другим (табл. 2).

Приведем пример реализации модели (1)–(10) с учетом регрессионных выражений, характеризующих урожайность зерновых культур, и стохастического описания метеорологических факторов. Такое математическое описание получения сельскохозяйственной продукции позволяет решать следующие задачи:

1) предсказывать биопродуктивность зерновых культур при наличии значимых регрессионных зависимостей результативного признака от метеорологических факторов в начальной стадии вегетации;

2) планировать производство сельскохозяйственной продукции на основе задачи параметрического программирования;

3) выявлять возможные оптимальные планы с учетом вариации метеорологических факторов, соответствующих вероятностным оценкам.

Следует отметить, что связь между урожайностью зерновых культур и суммами осадков и температуры воздуха предлагается оценивать на основе выявления наилучшей продолжительности начальной стадии вегетации [8]. Другими словами, из множества регрессионных выражений, получаемых при вариации исходных и конечных значений продолжительности начальной стадии вегетации, выбирается наилучший вариант по F -критерию Фишера и t -статистикам Стьюдента.

В частности, по данным об урожайности ячменя и овса в ООО «Сибирская Нива» и сумм осадков и температурам воздуха за вегетационный период, полученным на гидрометеорологической станции Иркутска, определены факторные модели, в которых суммы метеорологических показателей вычисляются с 25 мая по 25 июня. Только в этом случае уравнения регрессии и коэффициенты этих уравнений являются значимыми.

Для урожайности ячменя и овса факторные модели имеют вид:

$$y = 62,5 - 0,0829z - 0,00016zr, \quad (11)$$

$$y = -40,7 + 0,125z + 1,87r - 0,00396zr. \quad (12)$$

Здесь z , r – суммы температур воздуха и осадков за период 25 мая – 25 июня.

Для приведенных выражений коэффициенты детерминации равны 0,50 и 0,63 соответственно. Уравнения значимы по F -критерию Фишера и t -статистикам Стьюдента для уровня значимости 0,05. Между тем для рассматриваемой сельскохозяйственной организации не выявлено значимых факторных моделей для урожайности пшеницы.

Алгоритм реализации модели (1)–(12) состоит из следующих операций. Во-первых, моделируются случайные числа, характеризующие вероятности для оценки сумм осадков и температур воздуха. Во-вторых, по получен-

ным вероятностям с помощью законов распределения вероятностей определяются суммы осадков и температур воздуха за расчетный период. В-третьих, по формулам (11) и (12) вычисляются урожайности ячменя и овса. В-четвертых, используя значения коэффициентов при неизвестных и правых частях ограничений, строится задача линейного программирования. В-пятых, определяются оптимальные решения задачи с привлечением специальной прикладной программы (в конкретном случае использована программа PLP). В-шестых, предыдущие процедуры повторяются многократно для получения множества решений, из которых выделяют оптимальные решения, соответствующие заданным вероятностям. Число циклических операций зависит от получаемого диапазона результатов согласно средней вероятности, точности исходных данных и целевого назначения решений.

Очевидно, что для использования рассмотренного алгоритма необходимо предварительно получить значимые регрессионные выражения наподобие формул (11) и (12). Кроме того, эмпирически определяются значения коэффициентов при неизвестных, а также правые части ограничений.

Поскольку долгосрочное планирование аграрного производства осуществляется обычно на 10-летний период, то в качестве опорных вероятностей (функция распределения) использованы значения 0,1; 0,5 и 0,9. Число циклов для решения задачи параметрического программирования применительно к ООО «Сибирская Нива» составило 500. В таблице 3 приведены некоторые результаты моделирования. Так, определены оптимальные объемы продукции, при которых доходы достигают максимального значения для неблагоприятных ($p = 0,1$), усредненных ($p = 0,5$) и благоприятных ($p = 0,9$) метеорологических условий.

Аналогичный вариант модели с корректировкой плана применим для прогностических значений температур воздуха и осадков в текущий год по данным территориального управления по гидрометеорологии. Такой ва-

Таблица 3. Результаты решения задачи параметрического программирования с вероятностными оценками для ООО «Сибирская Нива»

Table 3. The results of solving the problem of parametric programming with probabilistic estimates for the LLC "Sibirskaya Niva"

Вероятность совместных событий p_{ζ} <i>Probability of joint events p_{ζ}</i>	План Plan									Значение целевой функции, млн руб. <i>The value of the target function, million rubles</i>
	Пшеница, ц <i>Wheat, c</i>	Ячмень, ц <i>Barley, c</i>	Овес, ц <i>Oats, c</i>	Кукуруза на силос, ц <i>Corn for silage, c</i>	Картофель, ц <i>Potato, c</i>	Однолетние травы (на сенаж), ц <i>Annual grasses (forhaylage), c</i>	Многолетние травы (на сено), ц <i>Perennial grasses (forhay), c</i>	Молоко, ц <i>Milk, c</i>	Скот в живой массе, ц <i>Livestock in live weight, c</i>	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	
0,1	19 408	3359	3233	36 795	66 493	40 000	10 924	28 000	1900	333,64621
0,5	21 989	2353	1658	33 372	80 000	40 000	9195	28 886	1900	354,75395
0,9	32 153	4774	3575	33 489	80 000	51 084	3000	31 000	1900	376,04053

риант позволяет несколько увеличить продолжительность упреждения, но менее точен по сравнению с использованием реальных метеорологических показателей.

Согласно полученным решениям, расхождение доходов при вероятности совместных событий 0,1 и 0,5 составляет 3,3%, а при 0,5 и 0,9–8,4%. Другими словами, деятельность товаропроизводителя подвержена сильному влиянию метеорологических факторов. При этом имеет место вариация неизвестных модели. Отметим, что коэффициенты экстремальной задачи и правые части, кроме урожайности ячменя и овса, определялись как интервальные показатели с помощью моделирования случайных значений между нижними и верхними оценками. При решении конкретной задачи урожайность пшеницы и картофеля, а также себестоимость молока и мяса представляли собой интервальные оценки. При этом биопродуктивность однолетних, многолетних трав и кукурузы на силос при-

нималась в виде усредненных значений. Эти допущения вызваны особенностями временных рядов. Управление производственными процессами заключается в выборе лицом, принимающим решение, приемлемых для хозяйства вариантов получения продукции в зависимости от метеорологических особенностей начала вегетационного периода. Следует иметь в виду, что мелиорация земель в виде орошения может в значительной степени улучшить деятельность сельскохозяйственного товаропроизводителя, что следует из выражений (11) и (12). Этот параметр и температура воздуха в пределах интервала 25.05–25.06, входящие в левую часть ограничений (2) и (6), влияют в конечном итоге на оптимальное решение предложенных задач параметрического программирования.

Помимо реализации предложенного алгоритма с учетом вероятностных значений сумм температур воздуха и осадков, модель (1)–(12) может быть использована для корректировки

Таблица 4. Результаты решения задачи математического программирования с прогнозными значениями урожайности ячменя и овса за 2016–2018 гг. для ООО «Сибирская Нива»**Table 4.** The results of solving the problem of mathematical programming with the predicted values of the yield of barley and oats for 2016–2018 for LLC "Sibirskaia Niva"

Год Year	План Plan									Значение целевой функции, млн руб. The value of the target function, million rubles
	Пшеница, ц Wheat, c	Ячмень, ц Barley, c	Овес, ц Oats, c	Кукуруза на силос, ц Corn for silage, c	Картофель, ц Potato, c	Однолетние травы (на сенаж), ц Annual grasses (forhaylage), c	Многолетние травы (на сено), ц Perennial grasses (forhay), c	Молоко, ц Milk, c	Скот в живой массе, ц Livestock in live weight, c	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	
2016	5622	16 309	4068	25 682	11 700	19 483	16 434	40 000	1799	285,46952
2017	18 396	4311	3292	25 812	11 340	19 582	16 518	40 000	1808	287,96022
2018	7126	14 806	4068	25 747	11 520	19 532	16 476	40 000	1803	285,73814

планирования производства аграрной продукции. Так, применив формулы (11) и (12) на основе данных осадков и температур текущего года и сформулировав задачу линейного программирования, можно получить план производства аграрной продукции. В таблице 4 приведен пример оптимального решения для трех лет (2016–2018 гг.) с учетом ретроспективного прогноза.

При получении оптимального решения на основе зависимостей (11) и (12) определена прогностическая урожайность ячменя и овса на текущий год. Поскольку для биопродуктивности других культур не выявлено значимых регрессионных выражений, они в примере рассматривались как усредненные значения. Полученные оценки использованы для построения модели параметрического программирования и ее реализации.

Приведенные в таблице 4 результаты моделирования позволяют уточнять запланированные показатели. Следует иметь в виду,

что точность прогностических зависимостей урожайности зерновых культур (11) и (12) понижается при формировании во временных рядах аномальных уровней. В приведенном примере это касается 2016 и 2018 гг., в которые наблюдались неблагоприятные погодные условия. Кроме того, на точность оптимального плана во многом влияет допущение об усредненных значениях большей части сельскохозяйственных культур. Улучшить результаты моделирования можно при использовании во второй задаче случайной вариации урожайности сельскохозяйственных культур, не связанных значимой связью с метеорологическими факторами, наподобие первой задачи.

Отличительной особенностью примера реализации второй модели является уменьшение неопределенности двух показателей – урожайности ячменя и овса.

Таким образом, два варианта моделей параметрического программирования с учетом метеорологических факторов позволяют ре-

шать задачи управления производством в текущий год и оценивать вероятные результаты получения продукции с учетом вариации внешних факторов.

Заключение

Предложены модели параметрического программирования для решения задач оптимизации производства аграрной продукции в двух вариантах. Первый из них позволяет уточнять оптимальные планы в текущем

году благодаря разработанному алгоритму получения прогностических зависимостей урожайности сельскохозяйственных культур от метеорологических факторов в начале вегетационного периода. Согласно второму варианту определяются различные ситуации планирования в зависимости от случайных факторов – температуры воздуха и осадков.

Предложенные модели, реализованные для сельскохозяйственной организации, позволяют повышать эффективность принятия решений.

Список литературы

1. Аветисян А. Г., Гюльзаян Л. С. Метод решения задач параметрического линейного программирования, основанный на дифференциальных преобразованиях // Известия Томского политехнического университета. 2014. Т. 324. № 2. С. 25–30.
2. Астафьева М. Н., Иванько Я. М., Петрова С. А. Пространственно-временные закономерности изменчивости климатических и производственных параметров на юге Восточной Сибири // Экологический вестник. 2013. Вып. 25 (3). С. 13–18.
3. Барсукова М. Н., Иванько Я. М. Приложения параметрического программирования для решения задач оптимизации получения продовольственной продукции // Вестник ИрГТУ. 2017. Т. 21. № 4. С. 57–66. DOI: 10.21285/1814-3520-2017-4-57-66.
4. Барсукова М. Н., Иванько Я. М., Федурин Н. И. Моделирование аграрного производства в условиях изменчивости природно-экономических факторов // Актуальные вопросы аграрной науки. 2012. № 2. С. 57–65.
5. Дмитриев Н. Н., Солодун В. И., Султанов Ф. С., Зайцев А. М. и др. Особенности технологий возделывания сельскохозяйственных культур с учетом влагообеспеченности пашни в Иркутской области // Научно-практические рекомендации ФГБНУ Иркутский НИИИСХ; ФГБОУ ВО «Иркутский государственный аграрный университет». – Иркутск: Изд-во ФГБНУ «Иркутский НИИИСХ», 2018. – 62 с.
6. Житов В. В., Долгополов А. А., Дмитриев Н. Н., Хаданов А. К. Погодные условия и эффективность минеральных удобрений под зерновые культуры в лесостепи Приангарья: монография. – Иркутск: Изд-во ИрГСХА, 2006. – 228 с.
7. Зоркальцев В. И., Полковская М. Н., Федурин Н. И. Моделирование сезонных колебаний цен на сельскохозяйственную продукцию // Актуальные вопросы аграрной науки. 2018. № 28. С. 48–56.
8. Иванько Я. М., Столопова Ю. В. Климатическая изменчивость и агрометеорологические условия Предбайкалья: экспериментальные исследования и моделирование урожайности зерновых культур // Метеорология и гидрология. 2019. № 10. С. 117–124.
9. Серышев В. А., Солодун В. И. Агроландшафтное районирование Иркутской области // География и природные ресурсы. 2009. № 2. С. 86–94.
10. Умнов Е. А., Умнов А. Е. Параметрический анализ в задачах математического программирования // Труды МФТИ. 2014. Т. 6. № 3 (23). С. 73–83.
11. Asseng S., Ewert F., Rosenzweig C. et al. Uncertainty in simulating wheat yields under climate change // Nature Clim Change. 2013. No. 3. P. 827–832. DOI: 10.1038/nclimate1916.
12. Bannayan M., Crout N. M. J., Hoogenboom G. Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom // Agronomy Journal. 2003. Vol. 95. No. 1. P. 114–125. DOI: 10.2134/agronj2003.0114.
13. Challinor A. Towards development of adaptation options using climate and crop yield forecasting at seasonal to multi-decadal timescales // Environmental Science and Policy. 2009. Vol. 12. No. 4. P. 453–465. DOI: 10.1016/j.envsci.2008.09.008.

14. Crane T. A., Roncoli C., Paz K., Breuer N. E. et al. Forecast skill and farmer's skill: seasonal climate forecasts and agricultural risk management in the southeastern United States // *Weather, Climate, and Society*. 2010. Vol. 2. P. 44–59. DOI: 10.1175/2009WCAS1006.1.
15. Ivanyo Ya., Petrova S., Polkovskaya M., Fedurina N. Modeling of the production of agrarian products under the conditions of influence of droughts, rainfall and their combinations // *Proceedings of the 5th International workshop "Critical infrastructures: Contingency management, Intelligent, Agent-based, Cloud computing and Cyber security" (IWCI 2018)*. Advances in Intelligent Systems Research. 2018. Vol. 158. P. 78–84. DOI: 10.2991/iwci-18.2018.14.
16. Ivanyo Ya., Fedurina N., Varanitsa-Gorodovskaya Zh. Mathematical models of agricultural production management in high risk environments // *E3S Web of Conferences*. International Scientific and Practical Conference "Development of the Agro-Industrial Complex in the Context of Robotization and Digitalization of Production in Russia and Abroad" (DAIC 2020). 2020. Article 01018. P. 1–9. DOI: 10.1051/e3sconf/202022201018.
17. Newlands N. K., Zamar David S., Kouadio L. et al. An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty // *Frontiers in Environmental Science*. 2014. P. 1–21. URL: <https://www.frontiersin.org/articles/10.3389/fenvs.2014.00017/full> (дата обращения: 09.11.2021). DOI: 10.3389/fenvs.2014.00017.
18. Ramirez Octavio A. Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, non-normal random variables: the case of corn belt corn, soybean, and wheat yields // *American Journal of Agricultural Economics*. 1997. Vol. 79. №. 1. P. 191–205. DOI: 10.2307/1243953.
19. Porter J. R. et al. Chapter 7. Adaptation and managing risks in agriculture and other food system activities. – In book: *Food security and food production systems*. – Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2014. P. 513–520.
20. Singh A., Panda S. N. Development and application of an optimization model for the maximization of net agricultural return // *Agricultural Water Management*. 2012. Vol. 115. P. 267–275. DOI: 10.1016/j.agwat.2012.09.014.
21. Lobell D. B., Gourdji S. M. The influence of climate change on global crop productivity // *Plant Physiology*. 2012. Vol. 160. No. 4. P. 1686–1697. DOI: 10.1104/pp.112.208298.

Сведения об авторах

Иванько Ярослав Михайлович, ORCID 0000-0003-4118-7185, докт. техн. наук, профессор, проректор по научной работе, Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный аграрный университет имени А. А. Ежевского», Молодежный, Россия, iymex@rambler.ru

Петрова Софья Андреевна, ORCID 0000-0001-9586-583X, канд. техн. наук, доцент, кафедра информатики и математического моделирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный аграрный университет имени А. А. Ежевского», Молодежный, Россия, sofia.registration@mail.ru

Барсукова Маргарита Николаевна, ORCID 0000-0002-1947-0892, канд. техн. наук, заведующая кафедрой информатики и математического моделирования, Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный аграрный университет имени А. А. Ежевского», Молодежный, Россия, margarita1982@bk.ru

Столопова Юлия Владимировна, ORCID 0000-0001-7370-8585, преподаватель, Колледж автомобильного транспорта и агротехнологий, Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный аграрный университет имени А. А. Ежевского», Молодежный, Россия, stolopova.yuliana@yandex.ru

Поддержка исследований

Исследование выполнено при финансовой поддержке РФФИ, грант № 19–07–00322.

Статья поступила 03.03.2021, рассмотрена 18.06.2021, принята 15.07.2021

References

1. Avetisyan A. G., Gyl'zadyan L. S. *Metod resheniya zadach parametriceskogo lineinogo programmirovaniya, osnovannyi na differentsial'nykh preobrazovaniyakh* [Method for solving parametric programming problems]. *Izvestiya Tomskogo politekhnicheskogo universiteta*=Bulletin of the Tomsk Polytechnic University, 2014, vol.324, no.2, pp.25–30.

2. Astaf'eva M. N., Ivanyo Ya. M., Petrova S. A. *Prostranstvenno-vremennye zakonomernosti izmenchivosti klimaticheskikh i proizvodstvennykh parametrov na yuge Vostochnoi Sibiri* [Spatio-temporal patterns of variability of climatic and production parameters in the south of Eastern Siberia]. *Ekologicheskij vestnik*=Ecological Bulletin, 2013, vol.25, no.3, pp.13-18.
3. Barsukova M. N., Ivanyo Ya. M. Applications of parametric programming to solve the problems of food production optimization. *Vestnik IrGTU*=Proceedings of Irkutsk State Technical University, 2017, vol.21, no.4(123), pp.57-66 (in Russian). DOI: 10.21285/1814-3520-2017-4-57-66.
4. Barsukova M. N., Ivanyo Ya. M., Fedurina N. I. Modeling of the agrarian production in the terms of variability of natural and economic factors. *Aktual'nye voprosy agrarnoy nauki*=Actual Issues of Agrarian Science, 2012, no.2, pp.57-65 (in Russian).
5. Dmitriev N. N., Solodun V. I., Sultanov F. S., Zajcev A. M. et al. *Osobennosti tekhnologii vozdeleyvaniya sel'skokhozyaistvennykh kul'tur s uchetom vlogoobespechennosti pashni v Irkutskoi oblasti* [Features of technologies for cultivation of agricultural crops, taking into account the moisture supply of arable land in the Irkutsk region]. *Nauchno-prakticheskie rekomendatsii FGBNU Irkutskii NIISKH; FGBOU VO «Irkutskii gosudarstvennyi agrarnyi universitet»*. Irkutsk, FGBNU Irkutskij NIISKH Publ., 2018, 62 p.
6. Zhitov V. V., Dolgoplov A. A., Dmitriev N. N., Hadanov A. K. *Pogodnye usloviya i effektivnost' mineral'nykh udobrenii pod zernovye kul'tury v lesostepi Priangar'ya: monografiya* [Weather conditions and the effectiveness of mineral fertilizers for grain crops in the forest-steppe of the Angara region: monograph]. Irkutsk, IrGSKHA Publ., 2006, 228 p.
7. Zorkal'cev V. I., Polkovskaya M. N., Fedurina N. I. Modeling of season vibrations of prices for agricultural products. *Aktual'nye voprosy agrarnoy nauki*=Actual Issues of Agrarian Science, 2018, no.28, pp.48-56 (in Russian).
8. Ivanyo Ya. M., Stolopova Yu. V. Long-term variability and agrometeorological conditions in Cisbaikalia: field studies and modeling of grain crops yield. *Meteorologiya i gidrologiya*, 2019, no.10, pp.117-124 (in Russian).
9. Seryshev V. A., Solodun V. I. *Agrolandshaftnoe raionirovanie Irkutskoi oblasti* [Agrolandscape zoning of the Irkutsk region]. *Geografiya i prirodnye resursy*, 2009, no.2, pp.86-94.
10. Umnov E. A., Umnov A. E. *Parametricheskii analiz v zadachakh matematicheskogo programmirovaniya* [Parametric analysis in mathematical programming problems]. *Trudy MFTI*, 2014, vol.6, no.3(23), pp.73-83.
11. Asseng S., Ewert F., Rosenzweig C. et al. Uncertainty in simulating wheat yields under climate change. *Nature Clim Change*, 2013, no.3, pp.827-832. DOI: 10.1038/nclimate1916.
12. Bannayan M., Crout N. M. J., Hoogenboom G. Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom. *Agronomy Journal*, 2003, vol.95, no.1, pp.114-125. DOI: 10.2134/agronj2003.0114.
13. Challinor A. Towards development of adaptation options using climate and crop yield forecasting at seasonal to multi-decadal timescales. *Environmental Science and Policy*, 2009, vol.12, no.4, pp.453-465. DOI: 10.1016/j.envsci.2008.09.008.
14. Crane T. A., Roncoli C., Paz K., Breuer N. E. et al. Forecast skill and farmer's skill: seasonal climate forecasts and agricultural risk management in the southeastern United States. *Weather, Climate, and Society*, 2010, vol.2, pp.44-59. DOI: 10.1175/2009WCAS1006.1.
15. Ivanyo Ya., Petrova S., Polkovskaya M., Fedurina N. Modeling of the production of agrarian products under the conditions of influence of droughts, rainfall and their combinations. *Proceedings of the 5th International workshop "Critical infrastructures: Contingency management, Intelligent, Agent-based, Cloud computing and Cyber security"* (IWCi 2018). *Advances in Intelligent Systems Research*, 2018, vol.158, pp.78-84. DOI: 10.2991/iwci-18.2018.14.
16. Ivanyo Ya., Fedurina N., Varanitsa-Gorodovskaya Zh. Mathematical models of agricultural production management in high risk environments. *E3S Web of Conferences*. International Scientific and Practical Conference "Development of the Agro-Industrial Complex in the Context of Robotization and Digitalization of Production in Russia and Abroad" (DAIC 2020), 2020, article 01018, pp.1-9. DOI: 10.1051/e3sconf/20202201018.
17. Newlands N. K., Zamar David S., Kouadio L. et al. An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in Environmental Science*, 2014, pp.1-21. Available at: <https://www.frontiersin.org/articles/10.3389/fenvs.2014.00017/full> (accessed 09.11.2021). DOI: 10.3389/fenvs.2014.00017.
18. Ramírez Octavio A. Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, non-normal random variables: the case of corn belt corn, soybean, and wheat yields. *American Journal of Agricultural Economics*, 1997, vol.79, no.1, pp.191-205. DOI: 10.2307/1243953.

19. Porter J. R. et al. Chapter 7. Adaptation and managing risks in agriculture and other food system activities. In book: Food security and food production systems. Cambridge, United Kingdom and New York, NY, USA, Cambridge University Press, 2014, pp.513-520.
20. Singh A., Panda S. N. Development and application of an optimization model for the maximization of net agricultural return. Agricultural Water Management, 2012, vol.115, pp.267-275. DOI: 10.1016/j.agwat.2012.09.014.
21. Lobell D. B., Gourdji S. M. The influence of climate change on global crop productivity. Plant Physiology, 2012, vol.160, no.4, pp.1686-1697. DOI: 10.1104/pp.112.208298.

About the authors

Yaroslav M. Ivanyo, ORCID 0000-0003-4118-7185, Dr. Sci. (Eng.), Professor, Vice-rector for Scientific Work, Irkutsk State Agricultural University named after A. A. Ezhevsky, Molodezhny, Russia, iymex@rambler.ru

Sofia A. Petrova, ORCID 0000-0001-9586-583X, Cand. Sci. (Eng.), Associate Professor, Informatics and Mathematical Modeling Department, Irkutsk State Agricultural University named after A. A. Ezhevsky, Molodezhny, Russia, sofia.registration@mail.ru

Margarita N. Barsukova, ORCID 0000-0002-1947-0892, Cand. Sci. (Eng.), Head of Informatics and Mathematical Modeling Department, Irkutsk State Agricultural University named after A. A. Ezhevsky, Molodezhny, Russia, margarita1982@bk.ru

Yuliana V. Stolopova, ORCID 0000-0001-7370-8585, Lecturer, College of Automobile Transport and Agricultural Technologies, Irkutsk State Agricultural University named after A. A. Ezhevsky, Molodezhny, Russia, stolopova.yuliana@yandex.ru

Acknowledgments

The study was carried out with the financial support of the Russian Foundation for Basic Research, Grant No. 19-07-00322.

Received 03.03.2021, reviewed 18.06.2021, accepted 15.07.2021

Подписка-2022

Журнал «Прикладная информатика / Journal of Applied Informatics»

выходит 6 раз в год:

Январь Март Май Июль Сентябрь Ноябрь

Подписка через редакцию

Стоимость подписки на 2022 год:

1 номер – 1190 руб. 3 номера – 3570 руб. 6 номеров – 7140 руб.

Подписку можно оформить с любого месяца.

Тел./факс: +7 (495) 663-93-88 (доб. 33-04)

Руководитель службы маркетинга А. А. Розанцев,
тел. +7 (495) 987-43-74 (доб. 33-04), e-mail: arozantsev@synergy.ru

Подписка на почте

По объединенному каталогу «Пресса России» индекс 88059

По каталогу российской прессы «Почта России» индекс 14241

По каталогу агентства «Урал Пресс» индекс 20497

Доставка осуществляется заказной бандеролью с уведомлением.

Электронный выпуск, а также отдельные статьи журнала можно приобрести на сайтах www.appliedinformatics.ru, www.elibrary.ru и www.litres.ru

К оплате принимаются все виды электронных платежей, банковские карты.

Учредитель и издатель

Негосударственное образовательное частное учреждение высшего образования

«Московский финансово-промышленный университет «Синергия»

Свидетельство о регистрации ПИ № ФС 77–57869 от 25.04.2014 г.

Выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор)

Литературный редактор и корректор Ю. А. Голоперова

Дизайн макета Б. В. Зипунов

Верстка И. Н. Глазунова

Адрес редакции и издателя

129090, Москва, ул. Мещанская, д. 9/14, стр. 1 (юрид.)

125190, Москва, Ленинградский пр-т, д. 80, корпус Г

Тел.: +7 (495) 987-43-74 (доб. 33-04),

e-mail: evlasova@synergy.ru; www.appliedinformatics.ru

Наши реквизиты

Университет «Синергия»

ИНН 7729152149

КПП 770201001

ОГРН 1037700232558

Р/с 40703810338180120073

БИК 044525225

К/с 30101810400000000225

Сбербанк России (ПАО) Вернадское ОСБ

При перепечатке и цитировании материалов ссылка на журнал «Прикладная информатика / Journal of Applied Informatics» обязательна. Редакция не несет ответственности за достоверность информации, опубликованной в рекламных объявлениях.

Мнения авторов и редакции могут не совпадать.

© Университет «Синергия»

Подписано в печать: 17.12.2021

Тираж 100 экз.

Отпечатано в ООО «САМ ПОЛИГРАФИСТ»

129090, г. Москва, Протопоповский пер., д. 6