

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

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

Информационные технологии Проблемы и решения

У ф а
УНПЦ «Издательство УГНТУ»
2 0 2 1

Информационные технологии. Проблемы и решения. – Уфа: УНПЦ «Изд-во УГНТУ», 2021, 3(16). 76 С.

Information technology. – Ufa: ERPC «USPTU Publishing House», 2021, 3(16). 76 p.

Учредитель:

**ФГБОУ ВО Уфимский государственный
нефтяной технический университет**

2021, 3(16)

Издается с 2014 г.

РЕДКОЛЛЕГИЯ

Главный редактор

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

Члены редколлегии

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

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

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

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

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

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

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

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

© ФГБОУ ВО «Уфимский государственный нефтяной технический университет», 2021

© Коллектив авторов, 2021

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

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

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

Отпечатано с готового электронного файла.

Подписано в печать 30.06.2021. Формат 60x80 1/16. Гарнитура «Таймс». Усл. печ. л. 4,42. Тираж 800 экз. Заказ 149.

Учебный научно-производственный центр «Издательство Уфимского государственного нефтяного технического университета»

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

Founder:

**FSBEU NE Ufa State Petroleum
Technological University**

2021, 3(16)

Published since 2014

EDITORIAL BOARD

Editor-in-Chief

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

Editorial Board Members:

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

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

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

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

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

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

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

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

ОГЛАВЛЕНИЕ

ДИСТАНЦИОННОЕ ОБУЧЕНИЕ: КОНЦЕПЦИЯ, МЕТОДОЛОГИЯ, ТЕХНОЛОГИЯ

Петлина Е.М. ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ В ОБРАЗОВАТЕЛЬНОЙ ОРГАНИЗАЦИИ.....	4
--	---

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В НАУКЕ, ОБРАЗОВАНИИ И ПРОИЗВОДСТВЕ

Завгородний М.Г., Майорова С.П., Шлиенкова О.В. ПОШАГОВЫЙ МЕТОД ЧИСЛЕННОГО РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ СЕТЕВЫХ ТЕХНИЧЕСКИХ СИСТЕМ.....	9
--	---

Ермолаев Е.В., Чурилов Д.А. РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ РАСЧЕТА СЕМАНТИЧЕСКОЙ БЛИЗОСТИ В GOOGLE COLAB.....	16
--	----

Хайруллин Р.Н. ДИСТАНЦИОННЫЙ ФОРМАТ ПРОВЕДЕНИЯ НАУЧНО-ОБРАЗОВАТЕЛЬНЫХ МЕРОПРИЯТИЙ: ВОЗМОЖНОСТИ И ПРОБЛЕМЫ.....	22
--	----

Ткаченко А.Л., Кузнецова В.И., Заплатин Г.В. ПРИМЕНЕНИЕ ПРОГРАММНЫХ ПРОДУКТОВ В СФЕРЕ БИЗНЕС АНАЛИТИКИ.....	26
---	----

Белов В.Д., Мошев Е.Р., Ромашкин М.А. ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ СИСТЕМА ДЛЯ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ ТЕПЛОЭНЕРГЕТИЧЕСКОГО ОБОРУДОВАНИЯ.....	32
---	----

Гуртовой Я.А. МЕТОДЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ СОБЫТИЙ В ПИСЬМЕННОМ ТЕКСТЕ	38
--	----

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ, УПРАВЛЕНИИ И БИЗНЕСЕ

Бикметова М.Р., Азбуханов А.Ф. ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЗАЦИИ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА НА ПРЕДПРИЯТИИ.....	45
---	----

МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ

Репьюк Н.С. МОДЕЛЬ ЦИФРОВОЙ ОБРАЗОВАТЕЛЬНОЙ ПЛАТФОРМЫ И ОБЕСПЕЧЕНИЕ ЕЕ БЕЗОПАСНОСТИ.....	52
--	----

Харинов М.В. СУПЕРПИКСЕЛИ В ЦИФРОВОЙ ОБРАБОТКЕ ИЗОБРАЖЕНИЙ.....	58
---	----

СЕТИ И ТЕЛЕКОММУНИКАЦИИ

Жмакина И.Д., Стаценко Л.Г. ОЦЕНКА УРОВНЯ ЭЛЕКТРОМАГНИТНОГО ФОНА, СОЗДАВАЕМОГО БАЗОВЫМИ СТАНЦИЯМИ.....	65
--	----

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ И ГИС-ТЕХНОЛОГИИ

Анфёров М.А. АЛГОРИТМИЗАЦИЯ СИНТЕЗА СТРУКТУРЫ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА В РАМКАХ САПР.....	71
--	----

ДИСТАНЦИОННОЕ ОБУЧЕНИЕ: КОНЦЕПЦИЯ, МЕТОДОЛОГИЯ, ТЕХНОЛОГИЯ

УДК 004.85

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ ДИСТАНЦИОННОГО ОБУЧЕНИЯ В ОБРАЗОВАТЕЛЬНОЙ ОРГАНИЗАЦИИ

USE OF DISTANCE LEARNING TECHNOLOGY IN EDUCATIONAL ORGANIZATION

Петлина Е.М.,
Северо-Кавказский социальный институт,
г. Ставрополь, Российская Федерация

E.M. Petlina,
North Caucasian Social Institute,
Stavropol, Russian Federation

e-mail: 356620@gmail.com

Аннотация. В статье рассматривается вариант использования наработанных методических разработок дистанционного обучения. Авторы предлагают использовать опыт работы образовательной организации во время пандемии. Предлагается рассматривать учебные курсы как модули. Большинство дистанционных технологий основаны на модульной организации учебного процесса. Выделяется четыре направления развития системы модельного обучения на платформе дистанционных технологий. Это конструирование учебного процесса, формирование готовности преподавателя, совершенствование организации самостоятельной работы, внедрение рейтинговой системы. Рассмотрена структура учебного модуля при дистанционных технологиях. Модель системы модульного обучения включает целевой, содержательный, технологический, оценочно-критериальный, результативный блоки. Указан алгоритм создания учебного модуля на образовательной платформе. Предложен алгоритм оценки результатов обучения обучающихся по модулю. Рассматриваемая модель включает три критерия, которые позволяют определять уровень формирования компетенций специалиста. Это когнитивно-познавательный, деятельностно-креативный и личностный критерии. Основные направления развития системы модульного обучения в образовательных организациях позволят сформировать информационную основу для принятия руководством образовательных организаций управленческих решений по повышению качества модульного обучения в образовательных организациях с применением дистанционных технологий; использовать выделенные направления развития системы модульного обучения в образовательных организациях в постпандемический период.

Abstract. The article considers the variability of using the tried and tested methodology developments of distance learning. It is proposed to consider training courses as modules. Most distance learning technologies are based on a modular organization of the educational process. The structure of a training module is considered. The model of the modular training system includes target, content, technology, evaluative-criterial, and effective blocks. The algorithm

for creating a training module on an educational platform is specified. Formation of the content of the training course from training modules in accordance with the logic of the academic subject. Construction of training modules from training elements in accordance with the technology of distance education methods. Consideration of possible options for the relationship between training modules. An algorithm for evaluating learning outcomes is proposed. The model under consideration includes three criteria that allow determining the level of formation of a specialist's competence. These are cognitive-cognitive, activity-creative and personal criteria.

Ключевые слова: дистанционные образовательные технологии, модульное обучение, подготовка кадров, профессиональное образование, учебный курс, образовательная платформа.

Keywords: distance learning technologies, modular training, personnel training, professional education, training course, educational platform.

В современных условиях система образования призвана обеспечивать повышение качества подготовки кадров в условиях кризисных ситуаций, в частности в период пост пандемического развития образовательной организации.

Требования реалий общественной жизни таковы, что эпидемиологическая обстановка расширила возможности образовательных организаций и внедрила широко дистанционные технологии обучения, которые основаны на принципе модульной организации учебного процесса, основанного на компетентностном подходе [2, 5, 7, 9].

Возрастающая роль модульного обучения специалистов, развития его теории и практики в образовательных организациях обусловлена актуальностью инновационного развития системы подготовки обучающихся, их профессионального воспитания и обучения как одного из основных направлений строительства и реформирования системы образования в условиях кризисных ситуаций, особенно в условиях осложненной эпидемиологической обстановки.

Изучение генезиса, исследование и обоснование ключевых положений теории и анализ практики модульного обучения в рамках дистанционного обучения в образовательных организациях свидетельствует о том, что в современных условиях процесс развития системы модульного обучения в образовательных организациях может быть более результативным, особенно с применением информационных технологий в постпандемической образовательной организации [7, 9].

Рассмотрим основные направления развития системы модульного обучения на платформе дистанционных технологий в образовательных организациях в постпандемический период [2].

К ним относятся:

- конструирование учебного процесса с применением модульного обучения дистанционной технологии в образовательных организациях;
- формирование готовности преподавателя к модульному обучению;
- совершенствование организации самостоятельной работы обучающихся в процессе модульного обучения;
- внедрение рейтинговой оценки обучающихся образовательных организаций [1].

Выявление основных направлений развития системы модульного обучения в образовательных организациях рассматривается по следующим направлениям:

- изучение и анализ нормативных документов в системе профессионального образования;

– проведение целенаправленных опросов и бесед с руководством образовательных организаций, преподавателями, сотрудниками факультетов, обучающихся.

В процессе развития системы модульного обучения в образовательных организациях в условиях кризисных ситуаций функция внешнего, административного контроля должна дополняться механизмом внутренней самоорганизации и самоконтроля самих субъектов образовательной деятельности, что предполагает правильный выбор направлений развития системы модульного обучения в образовательных организациях [3, 6, 10].

Первое направление развития системы модульного обучения – конструирование учебного процесса с применением модульного обучения в образовательных организациях [1].

Конструирование учебного процесса с применением модульного обучения в образовательных организациях предопределила необходимость разработки и обоснования модульной программы с позиции проектирования.

Рассматривая модульную программу с позиции проектирования как одного из важных этапов внедрения системы модульного обучения в российском образовании в условиях применения дистанционных технологий, следует обратить внимание на процесс конструирования собственно содержательной части программы, состоящей из набора конкретных модулей и методической части.

При разработке модульной программы необходимо учитывать, чтобы модули были структурно идентичными, что позволяет всем участникам образовательного процесса работать по единым правилам, сохраняя способность гибкой адаптации своих возможностей, так и к профессиональной деятельности [3].

Исходя из характеристики модуля, может быть предложен следующий алгоритм построения учебного курса:

а) формирование содержания учебного курса из обучающих модулей в соответствии с логикой учебного предмета на платформах дистанционного обучения;

б) построение обучающих модулей из учебных элементов в соответствии с технологией дистанционных образовательных методик;

в) проектирование дидактического процесса, т.е. учебно-познавательной деятельности слушателей и управления этой деятельностью со стороны преподавателя, необходимое для достижения поставленных целей по каждому УЭ модуля дистанционной платформы;

г) определение организационных форм обучения и их координация по каждому модулю с учётом его УЭ;

д) разработка системы текущего, промежуточного и итогового контроля и коррекции познавательной деятельности обучающихся;

е) разработка инвариантного состава модуля – учебного текста, методического руководства по обучению, консультаций для преподавателей;

ж) определение временной границы изучения каждого модуля;

з) рассмотрение возможных вариантов взаимосвязи между учебными модулями.

Реализацию данного направления развития системы модульного обучения в образовательных организациях целесообразно осуществлять по следующему алгоритму [4]:

1) входное диагностическое тестирование для определения уровня личностных компонент и готовности работать с элементами дистанционного модульного курса;

2) разработка методических приемов и кейсов содержательной части дистанционного модуля (обучающие семинары, круглые столы, обучение методикам отбора содержания и т. д.);

3) формирование учебных курсов на основе разработанных методических материалов с учетом процессного, системного и личностно-ориентированного подходов;

4) мониторинг, оценка и диагностика уровня сформированности знаний, умений и компетенций на основании разработанного критериального аппарата;

5) анализ и создание условий для работы в удаленном формате преподавателей.

Следующее направление развития системы модульного обучения дистанционной технологии – совершенствование организации самостоятельной работы обучающихся в процессе модульного обучения.

Совершенствовать организацию самостоятельной работы слушателей в процессе модульного обучения – значит значительно повышать её роль в достижении новых образовательных целей, придавая ей проблемный характер, мотивирующий студентов и преподавателей. При этом преподаватель организует, консультирует и контролирует учебную деятельность обучающихся [5], при этом содержанию самостоятельной работы слушателей необходимо придать целенаправленный характер. Для этого необходимо включить самостоятельную работу в структуру учебных модулей, разнообразить заданиями практикоориентированной направленности.

Для оценки уровня освоения учебного модуля предлагается непосредственную оценку результатов обучения определения качества освоенных обучающимися ключевых и профессиональных компетенций проводить с использованием критериально-уровневой модели оценки компетенций специалистов.

Рассматриваемая модель включает три критерия, которые позволяют определять уровень формирования компетенций специалиста.

Знаниевую характеристику отражает когнитивно-познавательный критерий, характеризующий диапазон знаний, имеющихся у будущего специалиста. В процессе овладения знаниями обучающиеся должны уметь правильно давать оценку реальным техническим ситуациям, находить и использовать необходимые сведения в профессиональной деятельности, а также развивать основные компетенции в процессе самообразования [8].

Характеристика сформированных умений раскрывается в деятельностно-креативном критерии, отражающем наличие профессионально-ориентированных умений, которые позволяют реализовать процесс профессиональной деятельности, определять затруднения и находить методы ее совершенствования.

Данным критерием характеризуется направленность профессиональной деятельности на достижение эффективных результатов [4].

Личностный критерий отображает социально значимые качества будущего специалиста как личности: общительность, самостоятельность, рациональность, трудолюбие, предприимчивость, позволяющие молодому специалисту отстаивать позицию в принятии обоснованных профессиональных решений.

Представленное трио критериев позволяет установить уровни сформированности компетенций специалиста: высокий, средний, низкий [4, 5].

Выводы

Обоснованные в статье основные направления развития системы модульного обучения в образовательных организациях позволят:

– сформировать информационную основу для принятия руководством образовательных организаций управленческих решений, направленных на повышение качества модульного обучения в образовательных организациях;

– использовать выделенные направления развития системы модульного обучения в образовательных организациях;

- повысить уровень профессиональной компетентности преподавателей к модульному обучению в условиях образовательной организации;
- в целом оптимизировать процесс профессионального образования студентов в образовательных организациях.

Литература

1. Геворкянц Ж.А. Модульный подход к конструированию учебного процесса современного колледжа: автореф. дис. кан. пед. наук. Владикавказ, 2005. 24 с.
2. Давыдова Н.А. Использование интегративного подхода в процессе формирования ИКТ-компетентности у студентов педагогических профилей бакалавриата // *Инновации в образовании*. 2019. №3. С. 112-121.
3. Петлина Е.М., Горбачев А.В., Мироненко О.Ю. Особенности подготовки будущих преподавателей на основе профессионального стандарта педагога // *Инновации в образовании*. 2019. №10. С. 22-27.
4. Петлина Е.М., Горбачев А.В., Фирсин А.И. Моделирование процесса оценки профессиональных компетенций будущих педагогов // *Общество и личность: гуманистическая идея в современном общественном дискурсе и социальных практиках*. Ставрополь, 2020. С. 85-89.
5. Петлина Е.М. Использование информационных технологий для формирования общих компетенций специалиста // *Актуальные направления научных исследований XXI века: теория и практика*. 2017. Т. 5. №7-1 (33-1). С. 275-279.
6. Петлина Е.М. О выборе метода обучения при формировании компетенций у обучающихся // *Общество и личность: проблемы гуманизации современного социокультурного пространства. Материалы VII Всероссийской научно-практической конференции: сборник научных статей преподавателей, научно-практических работников, обучающихся вузов*. Ставрополь, 2019. С. 135-138.
7. Петлина Е.М. Проблема формирования компетенций специалиста в условиях информационного общества // *Гуманитарно-правовые аспекты развития российского общества: сборник научных трудов по итогам региональной научно-практической конференции*. Ставрополь: Краснодарский университет МВД РФ, 2017. С. 130-132.
8. Петлина Е.М. Формирование организационно-воспитательной компетентности у преподавателя // *Общество и личность: гуманизация в условиях информационной и коммуникационной культуры: сборник научных статей преподавателей, обучающихся вузов, научно-практических работников*. Ставрополь, 2018. С. 192-196.
9. Петлина Е.М., Хатагова С.В. Информатизация образования как основной принцип формирования компетенций специалиста // *Инновации в образовании*. 2017. №3. С. 124-133.
10. Хатагова С.В., Петлина Е.М. Мир классики: методы погружения курсантов // *Социально-экономические, психолого-педагогические, философские и правовые аспекты развития общества: сборник материалов всероссийского научно-практического круглого стола*. Ставрополь, 2016. С. 33-38.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В НАУКЕ, ОБРАЗОВАНИИ И ПРОИЗВОДСТВЕ

УДК 004.02

ПОШАГОВЫЙ МЕТОД ЧИСЛЕННОГО РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ СЕТЕВЫХ ТЕХНИЧЕСКИХ СИСТЕМ

STEPWISE METHOD FOR NUMERICAL SOLUTION OF BOUNDARY VALUE PROBLEMS OF NETWORK TECHNICAL SYSTEMS

¹Завгородний М.Г., ²Майорова С.П., ¹Шлиенкова О.В.,
ФГБОУ ВПО «Воронежский государственный университет»,
ФГБОУ ВПО «Воронежский государственный технический университет»,
г. Воронеж, Российская Федерация

M.G. Zavgorodnij¹, S.P. Majorova², O.V. Shlienkova¹,
Voronezh State University,
Voronezh State Technical University,
Voronezh, Russian Federation

e-mail: mzavgorodnij@yandex.ru

Аннотация. Ряд задач оперативного управления сетевыми техническими системами могут решаться только в классе динамических моделей и, следовательно, сетевые технические системы должны рассматриваться как системы с распределенными параметрами, а процессы, происходящие на отдельно взятом участке, должны описываться системой дифференциальных уравнений. Таким образом, при описании процессов сетевых технических систем приходят к начально-краевым или к краевым задачам, заданным на геометрическом графе (сети, составленной из осевых линий элементов системы).

Рядом зарубежных и российских исследователей, в том числе воронежскими математиками, выполнены постановки начально-краевых и краевых задач (в основном второго порядка), заданных на графе; изучены условия разрешимости таких задач; исследованы вопросы спектральной теории. Численные методы решения краевых задач на графе до настоящего времени практически не разработаны.

В работе предлагается разработанный нами пошаговый метод численного решения краевых задач второго порядка, заданных на графе. Предлагаемый метод не является сеточным. Он аналогичен методам пристрелки и дифференциальной прогонки, которые используются для решения краевых задач на отрезке, и, следовательно, не требует решения систем большой размерности и наличия большого объема памяти. Пошаговый метод был реализован нами в виде эффективных алгоритмов и комплекса хорошо отлаженных программных модулей.

Abstract. A number of problems of operational management of network technical systems can be solved only in the class of dynamic models and, therefore, network technical systems should be considered as systems with distributed parameters, and the processes occurring on a single site should be described by a system of differential equations. Thus, when describing the processes of network technical systems, one comes to initial-boundary value or

boundary value problems set on a geometric graph (a network made up of the center lines of the system elements).

A number of foreign and Russian researchers, including Voronezh mathematicians, have formulated initial-boundary value and boundary value problems (mainly of the second order) given on a graph; the conditions for the solvability of such problems have been studied; the questions of spectral theory have been investigated. Numerical methods for solving boundary value problems on a graph until now have not been practically developed.

In this paper we propose a stepwise method for numerical solution of second-order boundary value problems defined on a graph. The proposed method is not grid-based. It is similar to the shooting and differential running methods, which are used to solve boundary value problems on a segment, and, therefore, does not require the solution of large-dimensional systems and the presence of a large amount of memory. The stepwise method was implemented by us in the form of effective algorithms and a complex of well-debugged software modules.

Ключевые слова: сетевые технические системы, краевая задача, геометрический граф, численные методы.

Keywords: network technical systems, boundary value problem, geometric graph, numerical methods.

Рассмотрим сложную сетевую техническую систему, например, разветвленную трубопроводную гидравлическую систему. Осевые линии элементов этой системы образуют геометрический граф (в дальнейшем – просто граф), который мы будем обозначать через Γ .

Вершинами графа служат технологические узлы сети, а ребрами – отрезки (осевые линии), соединяющие пары вершин.

На каждом ребре задается дифференциальное уравнение, описывающее процессы, происходящие на нем.

В конечных вершинах графа задаются граничные условия (например, параметры подачи или отбора целевого продукта). Во внутренних вершинах графа (в узлах соединения двух и более элементов системы) параметры согласуются на основе общих законов равновесия и непрерывности.

Таким образом, математическая модель, построенная на основе вышеизложенных подходов, представляет собой краевую задачу на графе. Используя понятия и обозначения книг [1, 2], выполним ее формальное описание.

Обозначим через V – множество всех вершин, а через \mathfrak{S} – объединение всех ребер графа Γ .

По определению каждое ребро инцидентно двум различным вершинам, которые самому ребру не принадлежат, и ориентировано.

Обозначим через $\partial\Gamma$ множество граничных вершин, а через I^a – множество ребер, инцидентных вершине a .

Для функций $p(x)$, $g(x)$, $q(x)$ и $f(x)$, принадлежащих пространству $C(\mathfrak{S})$, рассмотрим на \mathfrak{S} дифференциальное уравнение

$$p(x)u''(x) + g(x)u'(x) + q(x)u(x) = f(x) \quad (x \in \mathfrak{S}). \quad (1)$$

Полагаем $\inf_{x \in \mathfrak{S}} |p(x)| > 0$. Дополним уравнение (1) граничными условиями

$$u(b) = u_b, \quad b \in \partial\Gamma \quad (u_b - const), \quad (2)$$

и в каждой внутренней вершине a – условиями согласования

$$u(a_\gamma) = u(a_\eta), \quad \gamma \in I^a \setminus \eta; \quad \sum_{\gamma \in I^a} \alpha_\gamma u'(a_\gamma) = 0, \quad (3)$$

где $u(a_\gamma)$ – односторонний предел функции $u(x)$, вычисленный в вершине a вдоль ребра γ , $\eta \in I^a$ – фиксированное ребро и α_γ , $\gamma \in I^a$ – произвольные константы, отличные от нуля.

Первая группа условий (3) – это условия непрерывности, а последнее условие (3) является аналогом условия Кирхгофа.

Пусть граф Γ является связным графом-деревом.

В предположении, что краевая задача (1)-(3) невырожденная, ставится задача найти ее численное решение. Так как численные методы решения дифференциальных уравнений на отрезке достаточно хорошо разработаны, то основные проблемы численного решения краевой задачи на графе заключаются в локализации условий согласования на то или иное ребро.

Граф-дерево Γ будем задавать матрицей смежности $A = (\alpha_{bd})_{b,d \in V}$ следующего вида: если вершины b и d не являются смежными, то $\alpha_{bd} = 0$; для смежных вершин b и d , инцидентных некоторому ребру γ длины l , задаем $\alpha_{bd} = l$, если ребро γ ориентировано от вершины b к вершине d , и $\alpha_{bd} = -l$, если это ребро ориентировано от d к b .

Прежде чем приступить к решению краевой задачи (1)-(3), рассмотрим вспомогательную задачу и опишем алгоритм построения ее численного решения.

Для этого сначала найдем все внутренние вершины такие, что все смежные каждой из них вершины, кроме одной, являются граничными. Из полученного множества выберем вершины a^1 и a^n так, чтобы маршрут

$$S = \left((b^1, a^1), (a^1, a^2), (a^2, a^3), \dots, (a^{n-1}, a^n), (a^n, b^n) \right)$$

был максимальным по величине $n + 1$ (количеству пройденных ребер).

Здесь b^1 и b^n – граничные вершины, смежные вершинам a^1 и a^n соответственно.

У графа-дерева, отличного от графа-пучка, всегда существует, по крайней мере, две вершины с указанными свойствами. Поэтому всегда возможно построение маршрута S .

Рассмотрим теперь задачу для дифференциального уравнения (1) со всеми граничными условиями (2), кроме условия в вершине b^n , и со всеми условиями согласования (3).

При этом дополним условие в вершине b^1 еще одним условием: $u'(b^1) = u_0^1$ ($u_0^1 - const$). Таким образом, для уравнения (1) на ребре $\gamma_1 = (b^1, a^1)$ будут заданы начальные условия и, следовательно, уравнение будет однозначно разрешимо. Опишем алгоритм построения численного решения полученной задачи, которую назовем *задачей*

с локальными начальными условиями.

На первом этапе на ребре γ_1 с заданным шагом h находим численное решение $u(x)$ задачи Коши для дифференциального уравнения (1) с начальными условиями $u(b^1) = u_{b^1}$ и $u'(b^1) = u_0^1$.

Для этого программный комплекс имеет два модуля: модуль на основе метода Эйлера и модуль на основе метода Рунге-Кутты.

Полученные значения решения $u(x)$ сохраняем во внешней памяти.

Положим $u_1^0 = u(a^1)$.

В силу условий непрерывности (3) на всех ребрах вида $\gamma = (b, a_\gamma^1)$, инцидентных вершине a^1 , кроме ребер, принадлежащих маршруту S , получим краевые задачи для дифференциального уравнения (1) с краевыми условиями $u(b) = u_b$, $u(a_\gamma^1) = u_1^0$.

Для вычисления значений решений полученных краевых задач программный комплекс имеет модуль на основе метода пристрелки.

Полученные значения решений на указанных ребрах также сохраняем.

Итак, мы нашли решения на всех ребрах, инцидентных вершине a^1 , кроме ребра $\gamma_2 = (a^1, a^2)$, и нам известны величины $u'(a_\gamma^1)$, $\gamma \in I^a \setminus \gamma_2$.

Воспользовавшись последним условием согласования (3), вычислим значение u_1^1 одностороннего предела решения $u(x)$ в вершине a^1 вдоль ребра γ_2 .

Найдем на ребре γ_2 численное решение полученной задачи Коши для дифференциального уравнения (1) с начальными условиями $u(a^1) = u_1^0$ и $u'(a^1) = u_1^1$.

Положим $u_2^0 = u(a^2)$. Решение задачи с локальными начальными условиями на пучке ребер, инцидентных a^1 , мы нашли.

На втором этапе удалим из графа Γ вершину a^2 . Тогда граф распадется на компоненты связности.

Сузим нашу задачу на все полученные компоненты связности, кроме двух, которые содержат по одному ребру, принадлежащему построенному маршруту S .

Для каждой компоненты связности в ее вершине a^2 задаем краевое условие $u(a_\gamma^2) = u_2^0$, $\gamma \in I^a \setminus \gamma_2$.

Получим набор краевых задач на графах с меньшим количеством ребер, чем исходный граф Γ .

Найдя (по алгоритму, приведенному ниже) численные решения каждой краевой задачи, обеспечим на ребре $\gamma_3 = (a^2, a^3)$ необходимые для задачи Коши начальные условия.

Проведем вычисления значений решения на ребре γ_3 . Получим решение на всех ребрах, инцидентных двум вершинам a^1 и a^2 .

На следующих этапах, переходя от вершины a^k к следующей вершине a^{k+1} маршрута S , последовательно получим на всем графе Γ решение задачи с локальными начальными условиями.

Найденное решение не обязано удовлетворять условию (2) в вершине b^n и поэтому, как правило, не будет решением рассматриваемой краевой задачи (1)-(3).

Отметим, что предложенный алгоритм может быть как рекурсивным и хранить все необходимые данные в оперативной памяти, так и нерекурсивным. Мы реализовали нерекурсивный алгоритм.

Вернемся к исходной невырожденной краевой задаче (1)-(3).

Пусть $u_f(x)$ и $u_0(x)$ – решения двух задач с локальными начальными условиями.

Первое решение $u_f(x)$ найдено при дополнительном условии $u'(b^1) = 0$ в граничной вершине b^1 .

А второе решение $u_0(x)$ найдено при правой части $f(x) \equiv 0$ дифференциального уравнения (1); при условиях $u(b) = 0$, заданных по всех граничных вершинах $b \in \partial\Gamma$, кроме вершины b^n ; и наконец, при дополнительном условии $u'(b^1) = 1$.

Теорема. Функция $u(x) = u_f(x) + Cu_0(x)$, где $C = \frac{u_{b^n} - u_f(b^n)}{u_0(b^n)}$, является единственным решением невырожденной краевой задачи (1)-(3).

Доказательство. Решение краевой задачи (1)-(3) в силу ее невырожденности единственно и $u_0(b^n) \neq 0$.

Непосредственной проверкой показывается, что приведенная в условии теоремы функция $u(x)$ удовлетворяет дифференциальному уравнению (1), всем краевым условиям (2) и всем условиям согласования (3). **Теорема доказана.**

Таким образом, для численного нахождения решений краевой задачи (1)-(3) необходимо по приведенному выше алгоритму на графе Γ вычислить значения решений $u_f(x)$ и $u_0(x)$ двух задач с локальными начальными условиями.

Затем вычислить константу C и сложить значения решения $u_f(x)$ со значениями решения $u_0(x)$, умноженными на C .

Получим решение краевой задачи (1)-(3).

Описанные методы использовались авторами для анализа распределения давления в теплосети города Смоленска.

Изложенный пошаговый метод позволяет находить собственные значения и массивы значений соответствующих им собственных функций самосопряженного дифференциального оператора

$$(Du)(x) \square -(p(x)u'(x))' + q(x)u(x).$$

Сначала, используя следствие метода «накачки нулей», надо отделить друг от друга собственные значения, не превосходящие по модулю заданного положительного числа.

Затем на каждом интервале, содержащем единственное собственное значение (возможно кратное), подбирать (например, последовательным делением интервала пополам и выбором одной из его половин) число λ так, чтобы значение в вершине b^n решения задачи с локальными начальными условиями для дифференциального уравнения

$$(Du)(x) = \lambda u(x) \text{ при граничных условиях } u(b) = 0, b \in \partial\Gamma \setminus b^n, u'(b^1) = 1$$

не превосходило заданной точности вычислений.

Подобранное число λ будет собственным значением, а найденное решение – собственной функцией.

Проиллюстрируем работу написанных нами программ, реализующих пошаговый алгоритм, на примере графа Γ , заданного следующей матрицей смежности

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -2 & \pi & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & -\pi & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Граф Γ имеет пять ребер, две внутренние вершины степени 3 и четыре граничные вершины. Пусть на объединении ребер графа Γ задана краевая задача для дифференциального уравнения

$$\begin{cases} u''(x) - 2u'(x) + 5u(x) = 5x^2 - 4x + 2 & \text{при } x \in \gamma_1, \\ u''(x) - 3u'(x) + 2u(x) = e^x & \text{при } x \in \gamma_2, \\ u''(x) + u'(x) - 2u(x) = 9\cos x - 7\sin x & \text{при } x \in \gamma_3, \\ u''(x) - 2u'(x) + u(x) = x - 1 & \text{при } x \in \gamma_4, \\ u''(x) + 6u'(x) + 9u(x) = 9x^2 + 12x - 43 & \text{при } x \in \gamma_5 \end{cases}$$

при граничных условиях $u(b_i) = 0, i = \overline{1,4}$, и при условиях согласования (3) с коэффициентами последнего условия $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = -1$ для первой внутренней вершины и $\alpha_3 = 3, \alpha_4 = 1, \alpha_5 = 2$ для второй внутренней вершины.

Были найдены аналитическое и численное решения.

Результаты вычислений сведем в таблицу.

Таблица – Значения численного решения

На первом ребре			На втором ребре			На третьем ребре			На четвертом ребре			На пятом ребре		
x_i	u_i	Δ_i	x_i	u_i	Δ_i	x_i	u_i	Δ_i	x_i	u_i	Δ_i	x_i	u_i	Δ_i
0	0	0	0	0	0	0	-2,82773	0,02863	0	0	0	0	0	0
0,05	-0,158784	0,00246	0,1	-0,08102	0,00013	0,157	-1,4445	0,09481	0,05	0,65998	0,00538	0,05	31,1022	-0,1437
0,1	-0,327445	0,00511	0,2	-0,17560	0,000388	0,314	-0,22708	0,07927	0,1	1,38484	0,01106	0,1	53,5257	-0,1701
0,15	-0,505248	0,00793	0,3	-0,28491	0,000785	0,471	0,947447	0,06822	0,15	2,17973	0,01701	0,15	69,0199	-0,1755
0,2	-0,691193	0,01089	0,4	-0,41007	0,001368	0,628	2,10409	0,06104	0,2	3,04987	0,02321	0,2	79,0763	-0,1707
0,25	-0,884006	0,01393	0,5	-0,55208	0,002175	0,785	3,2602	0,05687	0,25	4,0012	0,03831	0,25	84,921	-0,148
0,3	-1,08212	0,01704	0,6	-0,71173	0,003248	0,942	4,42919	0,05553	0,3	5,03966	0,03657	0,3	87,5452	-0,1014
0,35	-1,28368	0,02013	0,7	-0,88945	0,004645	1,099	5,62171	0,05633	0,35	6,17182	0,04353	0,35	87,6724	-0,1001
0,4	-1,48652	0,02315	0,8	-1,08526	0,00641	1,256	6,84843	0,05908	0,4	7,40488	0,05088	0,4	86,0221	-0,0594
0,45	-1,68814	0,02605	0,9	-1,29849	0,00856	1,413	8,12128	0,06373	0,45	8,7464	0,05868	0,45	83,1361	0,0566
0,5	-1,88576	0,02877	1	-1,52726	0,01151	1,57	9,45407	0,06963	0,5	10,2036	0,0662	0,5	79,2663	0,0998
0,55	-2,07629	0,03121	1,1	-1,76995	0,01418	1,727	10,8653	0,0771	0,55	11,786	0,0744	0,55	74,8256	0,1688
0,6	-2,25633	0,03332	1,2	-2,02122	0,01758	1,884	12,3776	0,0858	0,6	13,3238	-0,0956	0,6	70,076	0,281
0,65	-2,42218	0,03501	1,3	-2,27519	0,0218	2,041	14,0196	0,0956	0,65	15,361	0,0905	0,65	64,9681	0,1992
0,7	-2,56991	0,03621	1,4	-2,52296	0,02562	2,198	15,945	0,224	0,7	17,3743	0,0991	0,7	60,0721	0,3509
0,75	-2,69532	0,03682	1,5	-2,7524	0,02968	2,355	18,0955	0,369	0,75	19,5519	0,1074	0,75	55,0191	0,2611
0,8	-2,79398	0,0368	1,6	-2,94701	0,03341	2,512	20,1834	0,1893	0,8	21,9057	0,1155	0,8	50,2062	0,2497
0,85	-2,86133	0,03603	1,7	-3,08505	0,03614	2,669	22,9334	0,3457	0,85	24,4478	0,123	0,85	45,7173	0,3466
0,9	-2,89265	0,03444	1,8	-3,13771	0,03719	2,826	25,9431	0,3598	0,9	27,1928	0,1312	0,9	41,3564	0,3199
0,95	-2,88306	0,032	1,9	-3,06795	0,03522	2,983	29,2453	0,1747	0,95	30,1534	0,1385	0,95	37,1732	0,1974
1	-2,82773	0,02863	2	-2,82773	0,02863	3,14	33,3458	0,191	1	33,3458	0,1461	1	33,3458	0,1461
$\Delta_{\max} = 0,03682$			$\Delta_{\max} = 0,03719$			$\Delta_{\max} = 0,02863$			$\Delta_{\max} = 0,1461$			$\Delta_{\max} = 0,3509$		

В таблице приведены значения u_i решения в точках x_i и величины Δ_i , показывающие насколько численное решение отличается от аналитического.

Учитывая, что на каждом ребре выбирался шаг h так, чтобы вычислялось всего лишь 21 значение, то полученная точность результатов вычислений хорошая.

Наибольшая погрешность вычислений оказалась на пятом ребре.

Однако, относительная погрешность достаточно мала: $\frac{0,3509}{60,0721} = 0,005841$.

В заключение статьи приведем график численного решения краевой задачи на графе Г.

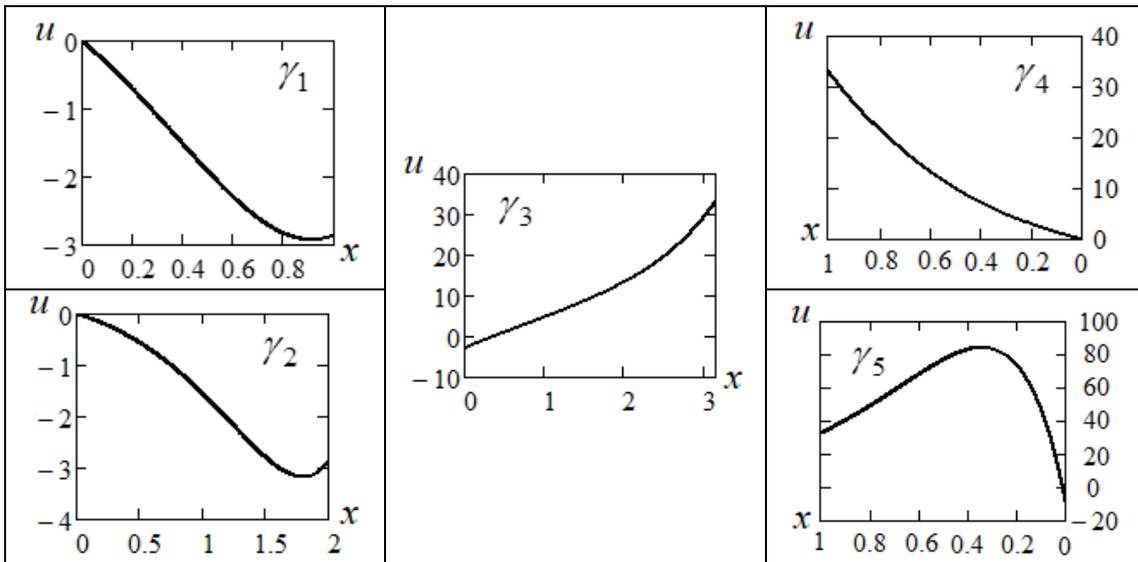


Рисунок 1. График численного решения $u(x)$ на ребрах графа

Литература

1. Дифференциальные уравнения на геометрических графах. / Ю.В. Покорный, О.М. Пенкин, В.Л. Прядиев, А.В. Боровских, К.П. Лазарев, С.А. Шабров. – М.: ФИЗМАТЛИТ, 2004. – 272 с.
2. Завгородний М.Г. Краевые задачи для дифференциальных уравнений на графе. / М.Г. Завгородний, С.П. Майорова. – Воронеж: Издательский дом ВГУ, 2015. – 145 с.

УДК 004.912

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ РАСЧЕТА СЕМАНТИЧЕСКОЙ БЛИЗОСТИ В GOOGLE COLAB

DEVELOPMENT OF AN APPLICATION FOR CALCULATING SEMANTIC PROXIMITY IN GOOGLE COLAB

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

E.V. Ermolaev, D.A. Churilov,
Ufa State Petroleum Technological University,
Kosmonavtov Str., 1, Ufa, Republic of Bashkortostan, 450064, Russia

e-mail: evg.ermo2009@gmail.com

Аннотация. Часто необходимо определить насколько два текста близки между собой. В данном случае не обойтись без такого понятия как мера семантической близости. Это особая мера близости, предназначенная для количественной оценки семантической схожести лексем, например, существительных или многословных выражений. Такая мера показывает высокие значения для пар слов, находящихся в семантических отношениях, и нулевые значения для всех остальных пар. Меры семантической близости успешно применяются в различных задачах обработки текста на естественном языке, таких как определение схожести текстов, расширение поисковых запросов, автоматические ответы на вопросы, снятие семантической неоднозначности слов. Практическое применение мер близости обусловлено пробелом между лексическим покрытием текста и заложенным в него смыслом: одно и то же понятие может быть выражено разными словами. Расчет семантической близости — это трудоемкий процесс, требующая сложного расчета с помощью формул, а также знаний исходных характеристик частотности для каждого термина. В статье предлагается разработать приложение в облачном сервисе GoogleColab для расчета семантической близости с помощью метрик, основанных на нахождении косинусной мера и евклидово расстояния и рассчитать с помощью него данный коэффициент для рабочих программ.

Abstract. It is often necessary to determine how close two texts are to each other. In this case, we can not do without such a concept as a measure of semantic proximity. This is a special measure of proximity, designed to quantify the semantic similarity of tokens, such as nouns or verbose expressions. This measure shows high values for pairs of words that are in semantic relations, and zero values for all other pairs. Semantic proximity measures are

successfully used in various tasks of natural language text processing, such as determining the similarity of texts, expanding search queries, automatic answers to questions, and removing semantic ambiguity of words. The practical application of proximity measures is due to the gap between the lexical coverage of the text and the meaning embedded in it: the same concept can be expressed in different words. Calculating semantic proximity is a time-consuming process that requires complex calculations using formulas, as well as knowledge of the initial frequency characteristics for each term. The article proposes to develop an application in the Google Colab cloud service for calculating semantic proximity using metrics based on finding the cosine measure and Euclidean distance and calculate this coefficient for working programs using it.

Ключевые слова: Google Colab, TensorFlow, ELMo, эмбеддинг, семантическая близость, евклидово расстояние, косинусная мера.

Keywords: Google Colab, TensorFlow, ELMo, Embedding, semantic proximity, Euclidean distance, cosine measure.

Для оценки семантической близости существует ряд метрик. Данные метрики делятся на группы по учитываемым параметрам таким как:

- частота символов или слов (коэффициент Танимото, мера Дайса, коэффициент Симпсона),
- редакционное расстояние (расстояние Ливенштейна, расстояние Джаро),
- семантическая близость слов (косинусная мера, скалярное произведение, евклидово расстояние, манхэттенское расстояние, расстояние Минковского [1-5]).

На основании статьи «Оценка семантической близости между критериями оценивания в рабочих программах вуза» авторов Гиниятуллина В.М., Салиховой М.А., Хлыбова А.В., Чурилова Д.А., был сделан вывод, что для анализа семантической близости достаточно использовать метрики, основанные на нахождении косинусной меры и евклидово расстояния.

Косинусная мера предполагает представление частотностей терминов в двух текстах в виде векторов и определение степени смысловой близости этих текстов на основе вычисления косинуса угла, который образуют два вектора. В соответствии с теоремой косинусов, увеличение значения косинуса указывает на уменьшение угла и большую степень близости между документами.

Косинусная мера вычисляется по формуле:

$$Sim(t1, t2) = \frac{\sum f_{i \in t1} \times f_{i \in t2}}{\sqrt{\sum f_{i \in t1}^2} \times \sqrt{\sum f_{i \in t2}^2}},$$

где Sim – мера смысловой близости текстов t1 и t2,
f – частотность некоторого i-го термина.

Евклидово расстояние трактуется как расстояние между точками или косинус угла между векторами, представляющими документы в пространстве терминов, и рассчитывается по формуле:

$$simEvklid(d, q) = \sqrt{\sum_{i=1}^n ((tf - idf_{di}) - (tf - idf_{qi}))^2}$$

где n – число терминов;

i – переменная – счетчик;
 tf – частота термина в тексте;
 idf – обратная документная частота.

Средой разработки для создания приложения был выбран облачный сервис Google Colab [6, 7].

После выбора среды разработки к ней подключается библиотека TensorFlow. Подключение библиотеки TensorFlow к среде разработки Google Colab представлено на рисунке 1.

TensorFlow – это библиотека или структура программного обеспечения, разработанная командой Google для максимально простой реализации концепций машинного обучения и глубокого обучения. Он объединяет вычислительную алгебру методов оптимизации для легкого вычисления многих математических выражений [8].

```
!pip install tensorflow-gpu==1.15
!pip install tensorflow_hub==0.6

import tensorflow_hub as hub
import tensorflow as tf
```

Рисунок 1. Подключение библиотеки TensorFlow к среде разработки Google Colab

После подключения к Google Colab библиотеки TensorFlow было произведено подключение языковой модели ELMo от DeepPavlov. Процесс подключения языковой модели ELMo к Google Colab представлен на рисунке 2.

Embedding Language Model (ELMo) – языковая модель, генерирующая контекстно-зависимые эмбединги [9, 11].

Эмбединг – векторное представление слова, представленное таким образом, чтобы вектор передавал смысл слова. К примеру: «Есть 4 слова: мужчина, женщина, король, королева. Хорошим примером векторного представления таких слов выглядит так: король – мужчина + женщина = королева» [10, 12].

```
[ ] url = "http://files.deeppavlov.ai/deeppavlov_data/elmo_ru-news_wmt11-16_1.5M_steps.tar.gz"

[ ] elmo = hub.Module(url)

def elmo_vectors(x):
    embeddings=elmo(x, signature="default", as_dict=True)["elmo"]

    with tf.Session() as sess:
        sess.run(tf.global_variables_initializer())
        sess.run(tf.tables_initializer())
        # return average of ELMo features
        return sess.run(tf.reduce_mean(embeddings,1))
```

Рисунок 2. Загрузка модели ELMo от DeepPavlov в Google Colab

После подключения языковой модели была произведена загрузка исходных данных в виде критериев оценивания рабочей программы. Пример исходных данных представлен в таблице 1.

Таблица 1 – Исходные данные

Оценка	Критерий
1. Лабораторная работа	
5	работа выполнена самостоятельно по методическим указаниям, расчеты приведены в полном объеме и без ошибок, сделаны полные выводы, отчет оформлен в соответствии с требованиями, при защите даны развернутые и полные ответы на все вопросы
4	работа выполнена с помощью указаний преподавателя, в расчетах допущены незначительные ошибки, выводы по работе неполные, отчет о лабораторной работе оформлен не в полном соответствии с требованиями, ответы на вопросы неполные
3	работа выполнена с помощью подробных указаний преподавателя, в расчетах допущены существенные ошибки, выводы по работе неполные, отчет о лабораторной работе оформлен без учета требований, ответы на вопросы частично неправильные
2	участие в выполнении лабораторной работе не принималось, в расчетах допущены грубые ошибки, выводы по работе отсутствуют, отчет о лабораторной работе оформлен неаккуратно и без учета требований, ответы на вопросы неправильные или отсутствуют
2. Письменный и устный опрос	
5	знает основные термодинамические закономерности, методы и модели, взаимосвязи между типовыми и механическими процессами, методы моделирования, применяемые в тепловых, механических и химических процессах, дает развернутые и полные ответы на все вопросы
4	знает основные термодинамические закономерности, методы и модели, взаимосвязи между тепловыми и механическими процессами, методы моделирования, применяемые в тепловых, механических и химических процессах, дает ответы на все вопросы, но ответы неполные
3	имеет общие представления об основных термодинамических закономерностях, методах и моделях, взаимосвязи между тепловыми и механическими процессами, методах моделирования, применяемых в тепловых, механических и химических процессах, дает ответы не на все вопросы, ответы на вопросы частично неправильные.
2	не знает основные термодинамические закономерности, методы и модели, взаимосвязи между тепловыми и механическими процессами, методы моделирования, применяемые в тепловых, механических и химических процессах, ответы на вопросы неправильные или отсутствуют.
3. Расчетно-графическая работа	
5	задание выполнено полностью правильно, в отчете приведены необходимые теоретические сведения, ответ полный
4	задание выполнено с небольшими неточностями, к отчету имеются замечания, теоретические сведения приведены не в полном объеме
3	задание выполнено с ошибками, отчет неполный, теоретические сведения приведены в минимальном объеме
2	задание выполнено неверно, отчет неполный, теоретические сведения не приведены, либо задание отсутствует

На рисунке 3 представлена загрузка исходных данных в Google Colab. Исходные данные загружаются для каждой из рабочих программ по группам (corpus), в

каждом corpus первой строкой вводятся критерии для оценки «5», второй для оценки «4», третьей для оценки «3», четвертой для оценки «2».

```
# лабораторная работа
corpus1 = ["работа выполнена самостоятельно по методическим указаниям, расчеты приведены в полном объеме и без ошибок, работа выполнена с помощью указаний преподавателя, в расчетах допущены незначительные ошибки, выводы по работе", "работа выполнена с помощью подробных указаний преподавателя, в расчетах допущены существенные ошибки, выводы по работе", "участие в выполнении лабораторной работе не принималось, в расчетах допущены грубые ошибки, выводы по работе"]

# письменный и устный опрос
corpus2 = ["знает основные термодинамические закономерности, методы и модели, взаимосвязи между тепловыми и механическим", "знает основные термодинамические закономерности, методы и модели, взаимосвязи между тепловыми и механическим", "имеет общие представления об основных термодинамических закономерностях, методах и моделях, взаимосвязи между", "не знает основные термодинамические закономерности, методы и модели, взаимосвязи между тепловыми и механическими"]

# расчетно-графическая работа
corpus3 = ["задание выполнено полностью правильно, в отчете приведены необходимые теоретические сведения, ответ полный", "задание выполнено с небольшими неточностями, к отчету имеются замечания, теоретические сведения приведены не", "задание выполнено с ошибками, отчет неполный, теоретические сведения приведены в минимальном объеме", "задание выполнено неверно, отчет неполный, теоретические сведения не приведены, либо задание отсутствует"]
```

Рисунок 3. Загрузка исходных данных в Google Colab

После загрузки исходных данных производится расчет метрик семантической близости с помощью подгружаемой в систему библиотеки sklearn. Результаты расчета косинусной меры представлены на рисунке 4.

```
+++++
similarity matrix1 =
[[1.      0.83439857 0.8071171  0.76082104]
 [0.83439857 1.      0.9632591  0.92036766]
 [0.8071171  0.9632591  1.0000002  0.93534327]
 [0.76082104 0.92036766 0.93534327 1.0000001 ]]
+++++

similarity matrix2 =
[[1.0000001 0.96713156 0.9273063  0.9422543 ]
 [0.96713156 0.99999994 0.9336955  0.95223874]
 [0.9273063  0.9336955  1.0000001  0.9243821 ]
 [0.9422543  0.95223874 0.9243821  1.0000005 ]]
+++++

similarity matrix3 =
[[0.99999994 0.79298  0.8138788  0.75847745]
 [0.79298  0.99999994 0.91006327 0.7541923 ]
 [0.8138788 0.91006327 1.      0.8010183 ]
 [0.75847745 0.7541923 0.8010183  1.0000002 ]]
```

Рисунок 4. Результаты расчета косинусной меры

На рисунке 4 представлена матрица попарного сравнения. Столбцы матрицы – это расстояния между оценками «5», «4», «3» и «2» в соответствующем порядке. Строки представляют то же самое, что и столбцы. То есть если необходимо значение расстояния между критериями для оценок, например, «5» и «2», то выбирается первый столбец последней строки, либо первая строка последнего столбца.

Далее был проведен расчет евклидова расстояния для исходных данных.

Результаты расчета евклидова расстояния представлен на рисунке 5.

```
similarity matrix1 =  
[[0.0000000e+00 4.7390676e+00 5.1745143e+00 5.8367162e+00]  
 [4.7390676e+00 1.6858739e-07 2.2689378e+00 3.3880594e+00]  
 [5.1745143e+00 2.2689378e+00 0.0000000e+00 3.0733502e+00]  
 [5.8367162e+00 3.3880594e+00 3.0733502e+00 0.0000000e+00]]  
  
similarity matrix2 =  
[[0.0000000e+00 2.1749291e+00 3.2484829e+00 2.9021220e+00]  
 [2.1749291e+00 2.9200194e-07 3.0762053e+00 2.6200259e+00]  
 [3.2484829e+00 3.0762053e+00 0.0000000e+00 3.3134747e+00]  
 [2.9021220e+00 2.6200259e+00 3.3134747e+00 0.0000000e+00]]  
  
similarity matrix3 =  
[[1.6858739e-07 5.6028519e+00 5.2990823e+00 6.4518023e+00]  
 [5.6028519e+00 0.0000000e+00 3.6817579e+00 6.5055599e+00]  
 [5.2990823e+00 3.6817579e+00 2.9200194e-07 5.8613453e+00]  
 [6.4518023e+00 6.5055599e+00 5.8613453e+00 0.0000000e+00]]
```

Рисунок 5. Результаты расчета евклидова расстояния

Выводы

В облачном сервисе было разработано приложение, позволяющее в короткие сроки рассчитать семантическую близость текстов.

Литература

1. Векторное представление слов [Электронный ресурс]. URL:<https://clck.ru/WAFrJ> (дата обращения 14.01.2021)
2. Шумская А.О. Оценка эффективности метрик расстояния Евклида и расстояния Махаланобиса в задачах идентификации происхождения текста // Доклады Томского государственного университета систем управления и радиоэлектроники, Томск: Изд-во ТГУСУиР. 2013, С. 141-145.
3. Хачумов М.В. Расстояния, метрики и кластерный анализ // Искусственный интеллект и принятие решений, Москва: Изд-во Федеральный исследовательский центр «Информатика и управление» РАН. 2012, С. 81-89.
4. Белова К.М., Судаков В.А. Исследование эффективности методов оценки релевантности текстов // Препринты ИПМ им. М.В. Келдыша, М.: Изд-во Институт прикладной математики им. М.В. Келдыша РАН. 2020, С. 1-16.
5. Яцко В.А. Особенности вычисления косинусной меры смысловой близости документов // Актуальные вопросы теории и практики развития научных исследований, Уфа: Изд-во Общество с ограниченной ответственностью «ОМЕГА САЙНС». 2020, С. 69-75.
6. Маринчук А.С. Решение задач с помощью нейронных сетей в Google Colab // Постулат, Биробиджан: Изд-во Приамурский государственный университет им. Шолом-Алейхема. 2020, С. 30.
7. Сахнюк П.А. Возможности Google Colab для изучения технологий машинного обучения и нейронных сетей // Информатизация непрерывного образования – 2018, М.: Изд-во Российский университет дружбы народов (РУДН). 2018, С. 586-588.

8. Семченко Р.В., Еровлев П.А. Программирование нейронных сетей в Python с использованием библиотек Keras и Tensorflow // Постулат, Биробиджан: Изд-во Приамурский государственный университет им. Шолом-Алейхема. 2020, С. 6.
9. Боровков Н.А., Фураев Ф.И. Использование языковой модели ELMO для задачи аспектно-ориентированного анализа тональности // Научная сессия ГУАП, Санкт-Петербург: Изд-во Санкт-Петербургский государственный университет аэрокосмического приборостроения. 2019, С. 301-305.
10. Обзор четырёх популярных NLP-моделей [Электронный ресурс]. URL:<https://clck.ru/WAG3Y> (дата обращения 14.01.2021)
11. Прогнозирование химических соединений с комплексом необходимых свойств. Арсланов И.Г., Дмитриев Г.Ю., Гиниятуллин В.М., Зайцева А.А., Кирлан С.А. Башкирский химический журнал. 2015. Т. 22. №2. С. 80-85.
12. Электронный ключ защиты с функциональностью троичного сопроцессора Габитов Р.Н., Габитова Я.А., Гиниятуллин В.М., Филиппов В.Н. Электронный научный журнал Нефтегазовое дело. 2015. №2. С. 385-396.

УДК 004:371.66

**ДИСТАНЦИОННЫЙ ФОРМАТ ПРОВЕДЕНИЯ
НАУЧНО-ОБРАЗОВАТЕЛЬНЫХ МЕРОПРИЯТИЙ:
ВОЗМОЖНОСТИ И ПРОБЛЕМЫ**

**DISTANT FORMAT OF REALIZATION
OF SCIENTIFIC AND EDUCATIONAL EVENTS:
OPPORTUNITIES AND PROBLEMS**

Хайруллин Р.Н.,
ФГБОУ ВО «Башкирский государственный университет»,
г. Уфа, Российская Федерация

R.N. Khairullin,
Bashkir State University,
Ufa, Russian Federation

e-mail: rimboba@yandex.ru

Аннотация. В данной статье исследуется дистанционный формат проведения научно-образовательных мероприятий.

Организация дистанционного формата проведения научно-образовательных мероприятий является примером проблем, к решению которых ИТ-отрасль приступила не так давно.

Заметный импульс развитию дистанционного формата деятельности придала пандемия коронавирусной инфекции. Тем не менее, оно не сводится лишь к реакции на эпидемиологическую ситуацию.

Дистанционный формат деятельности соответствует текущему уровню технического развития общества, а также имеет известный экономический смысл. Хотя осуществление научных и образовательных мероприятий в дистанционном формате требует дополнительных вложений в инфраструктуру, он позволяет значительно снизить расходы на традиционную материально-техническую базу. Кроме того, он позволяет

увеличить количество участников научно-образовательных мероприятий, открывая перед организациями возможности для увеличения доходов от платных услуг.

Автор останавливается также на двух проблемах, связанных с дистанционным форматом проведения научно-образовательных мероприятий и процессами информатизации и автоматизации в целом.

Первая проблема – проблема качества результатов научно-образовательной деятельности в условиях дистанционного формата.

Вторая проблема, которая сопровождает информатизацию и автоматизацию деятельности с самого начала, – это проблема безработицы.

Abstract. In this article the distant format of realization of scientific and educational events is analyzed.

The organization of a distant format of realization of scientific and educational events is an example of the problems that the IT-industry started to solve not so long ago.

The coronavirus pandemic gave a significant impulse to the development of the distant format of activity. However, it is not just a response to the epidemiological situation.

The distant format of activity corresponds to the current level of technical development of the society, and also has a definite economic meaning. Although the implementation of scientific and educational events in a distant format requires additional investments in infrastructure, it can strongly reduce the expense of traditional material and technical base. Besides, it allows to increase the number of participants in scientific and educational events, opening up opportunities for organizations to get more income from commercial services.

The author also focuses on two problems related to the distant format of scientific and educational events and the processes of informatization and automatization in general.

The first problem is the problem of the quality of the results of scientific and educational activities in the conditions of the distant format.

The second problem connected with informatization and automatization of activity from the very beginning is the problem of unemployment.

Ключевые слова: информатизация, дистанционный формат, научно-образовательные мероприятия, эпидемиологическая ситуация, проблема качества, безработица.

Keywords: informatization, distant format, scientific and educational events, epidemiological situation, problem of the quality, unemployment.

Современное общество нередко именуют информационным и связывают с развитием компьютерной техники и соответствующей инфраструктуры. Данный процесс идет уже в течение нескольких десятилетий и к сегодняшнему дню некоторые его результаты достаточно прочно вошли в повседневную жизнь жителей наиболее развитых стран мира.

Функциональные возможности компьютеров связаны с границами формализации и алгоритмизации, поскольку то, что человек прежде не формализовал, он не может представить в виде компьютерной программы [см.: 2, с. 125].

С другой стороны, сегодня для нас более очевидно, чем это было когда-либо в прошлом, что алгоритмизация применима к весьма широкому кругу явлений, включая, например, таргетированную рекламу, основанную на анализе вкусов и интересов пользователя, или технологию автоматического перевода устной речи в текст.

Все это становится предметом рассмотрения в дискуссиях ученых об искусственном интеллекте [например, см.: 1, с. 375-400; 4, с. 361].

Если обратиться к информатизации научно-образовательной сферы, то целесообразно, на наш взгляд, обратить внимание на следующие моменты.

Есть проблемы, с которыми по большей части новые технические средства уже успешно справились. К ним, на наш взгляд, вполне можно отнести проблему обеспечения учебных занятий иллюстративным и лабораторным материалом, применительно к которой потенциал компьютеров намного превосходит все ранее известные методы.

С другой стороны, существуют проблемы, к решению которых IT-отрасль приступила не так давно. Примером здесь может служить, в частности, организация дистанционного формата проведения научно-образовательных мероприятий, находящаяся в стадии реализации.

Заметный импульс развитию дистанционного формата деятельности придала пандемия коронавирусной инфекции, охватившая в прошедшем 2020 году практически весь земной шар. Как бы то ни было, его развитие отнюдь не сводится лишь к мерам реагирования на наличную эпидемиологическую обстановку. Оно соответствует текущему уровню технического развития общества, а также имеет известный экономический смысл.

Если проекторы или электронная почта известны уже достаточно давно, то веб-камеры и сервисы для проведения видеоконференций, напротив, именно сегодня, в условиях пандемии, получают массовое распространение.

Хотя осуществление научных и образовательных мероприятий в дистанционном формате требует дополнительных вложений в инфраструктуру (обеспечение людей и организаций необходимым количеством компьютеров, подключение их к сети Интернет, соответствующее определенным техническим параметрам, и т.д.), дистанционный формат позволяет значительно снизить расходы на традиционную материально-техническую базу (прежде всего, жилые и служебные помещения, их строительство, ремонт, содержание и так далее).

Во-вторых, дистанционный формат деятельности позволяет значительно увеличить количество участников научных или учебных мероприятий, что открывает перед организациями возможности для увеличения доходов от платных услуг.

Для студента, являющегося потребителем образовательных услуг, дистанционный формат, с одной стороны, также требует некоторых вложений и навыков использования компьютерной техники. С другой стороны, он может создать более комфортные условия для обучения.

В качестве таких условий можно назвать гибкий график учебного процесса, который проще совместить с другими интересами, увлечениями и занятиями, а также возможность учиться в географически отдаленных от места жительства учащегося учреждениях.

Для преподавателей дистанционное обучение означает изменения, во многом аналогичные тем, которые имеют место для студентов. К ним можно добавить автоматизацию многих рутинных работ, от ручного исполнения которых компьютер освобождает человека. К таким работам можно отнести, например, проверку тестовых заданий, которая ранее производилась вручную, отнимая у педагога много времени, а сейчас часто производится автоматически сразу после завершения ввода ответов.

Хотя научная работа сравнительно с сугубо образовательным процессом имеет определенную специфику, в то же самое время эти направления деятельности теснейшим образом переплетаются. Соответственно, в научной области мы можем наблюдать много сходного с областью образования. Становится привычным проведение разного рода конференций и симпозиумов в режиме онлайн в виде видеоконференций и т.д.

Библиотеки и издательства активно переключаются с бумажных документов на электронные документы. Традиционные библиотеки постепенно преобразуются в электронные библиотеки, а традиционные книжные фонды уступают место базам данных [например, см.: 3, с. 205-211]. Издательства тоже переходят к работе с электронными книгами и выпуску бумажных экземпляров книг по индивидуальному заказу потребителя.

Постепенно дистанционная форма взаимодействия охватывает практически все этапы научно-образовательных мероприятий: от информирования о мероприятии и процедуры регистрации участников до рассылок электронных книг, приглашений или сертификатов.

В этом плане дистанционный прием документов от абитуриентов в современных образовательных учреждениях проходит примерно так же, как регистрация участников разного рода научных конгрессов. Аналогично, проведение кратких и длительных учебных курсов, разного рода обучающих семинаров с технической точки зрения мало чем отличаются от научных конференций и консультаций.

Информатизация позволяет автоматически формировать соответствующие базы данных и работать с информацией, которая в них хранится.

Конечно, наряду с упомянутыми выше следствиями, в целом довольно позитивными, обозначенные выше перемены приводят и к определенным трудностям.

В связи с этим остановимся на двух проблемах, первая из которых связана именно с дистанционным форматом проведения научно-образовательных мероприятий, а вторая – с процессами информатизации и автоматизации в целом.

Первая проблема, которая вызывает серьезную озабоченность в условиях дистанционного проведения научно-образовательных мероприятий, – это проблема качества результатов научной и образовательной деятельности.

Конечно, компьютеры помогают проводить вычисления в самых разных областях, совершенствуется механизм таких вычислений, их скорость и точность, создаются многочисленные новые приложения.

В этом плане можно упомянуть активно применяемый при первичной оценке научных и учебных работ контент-анализ в сервисах по определению заимствований. Алгоритмы функционирования таких сервисов постепенно совершенствуются, критерии усложняются, вводятся дополнительные рубрики (например, «самоцитирование»), которые ранее отсутствовали.

Автоматизация подобных проверок и создание для этого соответствующих баз данных, разумеется, необходимы в условиях современного информационного общества. Однако такого рода приложения и сервисы не решают проблемы качества результатов научно-образовательной деятельности полностью.

Вторая существенная проблема, которую можно отметить, сопровождает информатизацию и автоматизацию с самого начала, – это безработица.

Автоматизируя деятельность, повышая производительность, организация зачастую перестает испытывать потребность в прежнем количестве работников. Происходит сокращение персонала, ситуация на рынке труда меняется таким образом, что возникает безработица. Продуманное решение указанной проблемы, вне всяких сомнений, сегодня является одной из ключевых задач общества и государства.

Выводы

Итак, хотя импульс развитию дистанционного формата деятельности придала пандемия коронавирусной инфекции, он не сводится лишь к мерам реагирования на текущую эпидемиологическую обстановку.

Дистанционный формат деятельности открывает перед научно-образовательными организациями возможности снизить расходы на материально-техническую базу и увеличить доходы от платных услуг.

Вместе с тем, использование современных информационных технологий в научно-образовательной сфере связано и с определенными трудностями, к которым можно отнести проблему качества научно-образовательной деятельности, а также проблему безработицы, возникающую вследствие автоматизации деятельности.

Литература

1. Дойч Д. Структура реальности. М.: Альпина нон-фикшн, 2015.
2. Зинченко В.П., Назаров А.И. Размышления об искусственном интеллекте // О человеческом в человеке. М.: Политиздат, 1991.
3. Мотульский Р.С. Общее библиотековедение: Учебное пособие для вузов. М.: ЛИБЕРЕЯ, 2004.
4. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики. М.: Едиториал УРСС, 2003.

УДК 004.89

ПРИМЕНЕНИЕ ПРОГРАММНЫХ ПРОДУКТОВ В СФЕРЕ БИЗНЕС АНАЛИТИКИ

APPLICATION OF SOFTWARE PRODUCTS IN THE FIELD OF BUSINESS ANALYTICS

Ткаченко А.Л., Кузнецова В.И., Заплатин Г.В.,
ФГБУ ВО «Калужский государственный университет им. К.Э. Циолковского»,
г. Калуга, Российская Федерация

A.L. Tkachenko, V.I. Kuznetsova V.I., G.V. Zaplatin,
Kaluga State University named after K.E. Tsiolkovski,
Kaluga, Russian Federation

e-mail: TkachenkoAL@tksu.ru

Аннотация. Востребованность и растущий интерес к программным продуктам становятся отправной точкой для совершенствования бизнес аналитики, потому что основным содержанием этого процесса является анализ большого количества данных, а программный продукт позволяет автоматизировать этот процесс. Инструментами бизнес-аналитики являются аналитические системы, позволяющие обработать, структурировать и предоставить исходную информацию таким образом, чтобы она была удобна для оперирования в принятии управленческих решений, ведь без качественной бизнес-аналитики сегодня практически невозможно обойтись. В статье анализируются как раз такие инструменты из сферы бизнес аналитики. Для анализа взяты программные продукты одного разработчика – Loginom Company ООО «Аналитические технологии» (бывший BaseGroup Labs). Разработчиком программного обеспечения является Российская компания, что важно в настоящее время, когда вопрос перехода на отечественные программные продукты стоит особенно остро, ввиду возможный санкций

и обеспечения информационной безопасности. Рассматриваются возможности технологий анализа Data Mining, Big Data, Data Warehousing, OLAP, Business Intelligence, Scoring, Forecasting, Data Quality, Machine Learning, ETL в рамках одной платформы. Сравняются характеристики и присутствующий функционал. Формируются выводы на основе сравнения. Оба программных продукта можно скачать бесплатно для ознакомления и учебных целей.

Abstract. the demand and growing interest in software products become the starting point for improving business analytics, because the main content of this process is the analysis of a large amount of data, and the software product allows you to automate this process. Business analytics tools are analytical systems that allow you to process, structure and provide the source information in such a way that it is convenient for operating in making management decisions, because it is almost impossible to do without high-quality business analytics today. The article analyzes just such tools from the field of business analytics. For the analysis, the software products of one developer - Loginom Company LLC “Analytical Technologies” (formerly BaseGroup Labs) were taken. The software developer is a Russian company, which is important at the present time, when the issue of switching to domestic software products is particularly acute, due to possible sanctions and information security. The possibilities of Data Mining, Big Data, Data Warehousing, OLAP, Business Intelligence, Scoring, Forecasting, Data Quality, Machine Learning, ETL analysis technologies within a single platform are considered. The characteristics and the present functionality are compared. Conclusions are formed based on the comparison. Both software products can be downloaded for free for reference and training purposes.

Ключевые слова: программные продукты, бизнес аналитика, интерфейс, функционал, анализ данных, операционная система, Loginom, Deductor.

Keywords: software products, business analytics, interface, functionality, data analysis, operating system, Loginom, Deductor.

Основным содержанием бизнес аналитики является сбор и анализ большого количества данных, что помогает принимать более эффективные управленческие решения.

На текущий момент на рынке представлен большой ряд программных продуктов, имеющих схожий между собой функционал и ввиду этого, проблема выбора программного средства является актуальной [1].

Рассмотрим две отечественные аналитические платформы Loginom и Deductor.

Оба данных программных продукта поддерживают русский язык, однако в Loginom также присутствует и английский язык.

Интерфейс Loginom, изображён на рисунке 1, представляет собой веб приложение.

Deductor, представлен на рисунке 2, это приложение для операционной системы Windows.

Оба программных продукта можно использовать на персональном компьютере и сервере предприятия, однако Loginom может работать в облаке (SaaS), что позволяет использовать данное программное обеспечение на серверах поставщика программы.

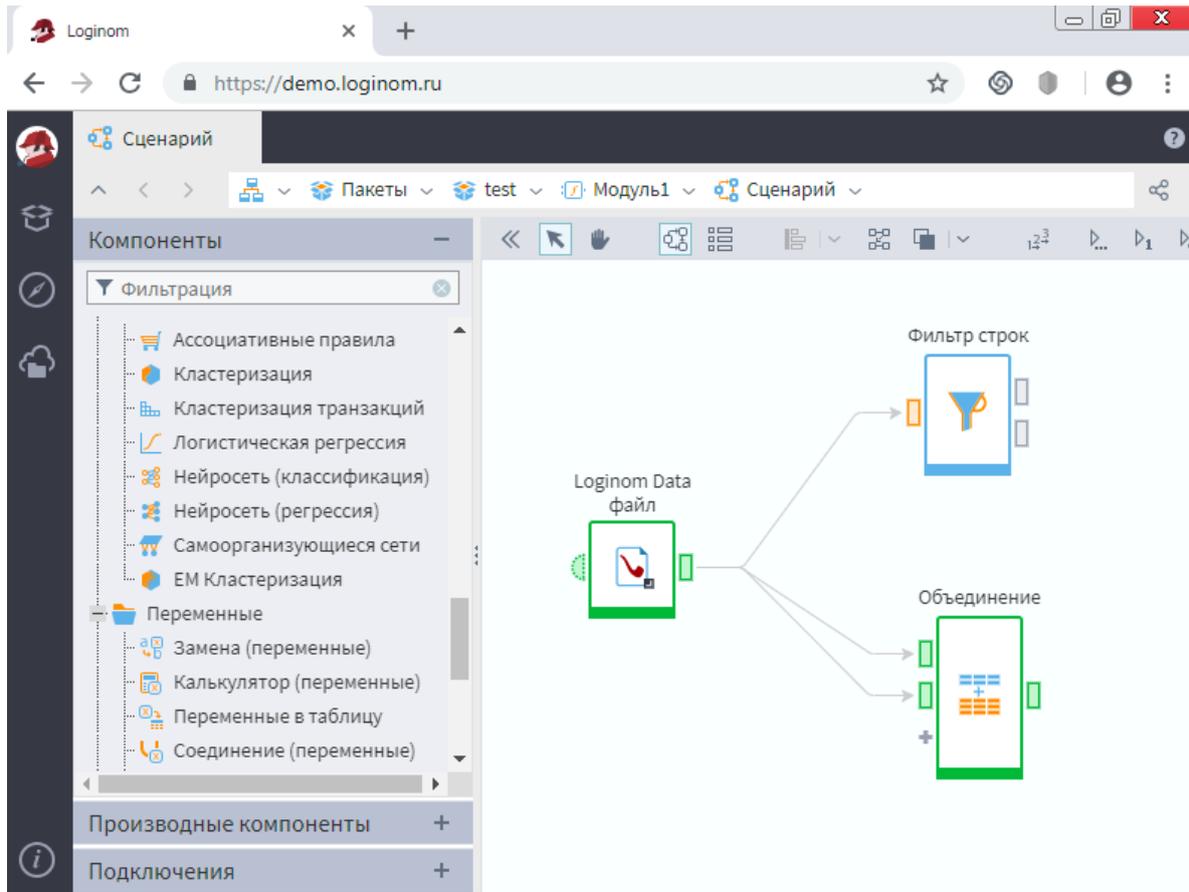


Рисунок 1. Интерфейс Loginom

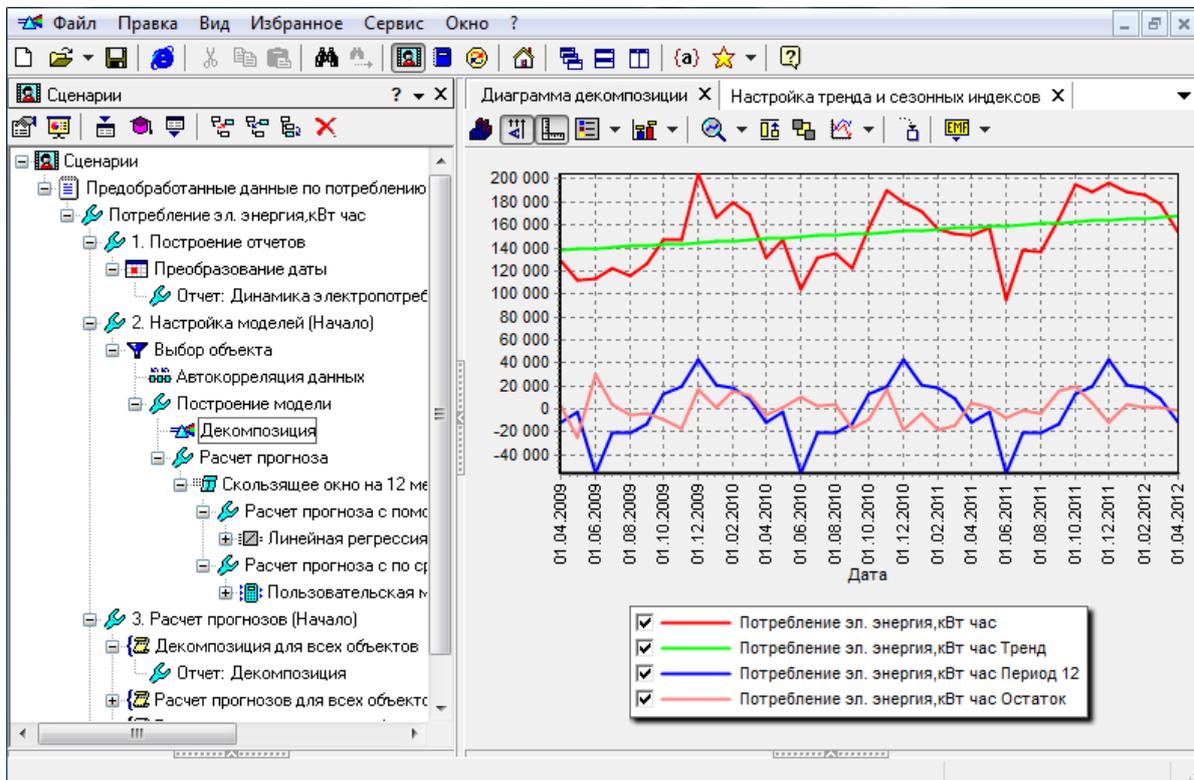


Рисунок 2. Интерфейс Deductor

Сценарий в Deductor представлен в форме дерева, состоящее из узлов-обработчиков данных и визуализаторов. Этот метод удобен, когда нужно реализовать простую логику.

Однако у этого метода имеются свои минусы: постоянно необходимо объединять, разделять, подтягивать данные на различных этапах анализа, декомпозировать большие задачи и объединять логические блоки в подмодели.

В Loginom сценарий отображается в форме графа, который более естественно отображает перемещение данных между узлами, что дает больше возможностей для реализации сложной логики.

Одним из отличий Loginom от Deductor является поддержка четырёх парадигм объектно-ориентированного моделирования: абстракция, инкапсуляция наследование, полиморфизм.

Разработка сценариев в Deductor реализовано по модели «снизу вверх», то есть необходимо наличие данных на входе.

К преимуществам данной модели можно отнести:

- высокую скорость реализации, благодаря наличию информации о структуре данных на входе;
- простоту поиска ошибок и отладки.

Недостатками модели являются:

- ориентированность на отдельных задачах;
- сложность повторного использования в аналогичных задачах;
- необходимость редактирования всего сценария при изменении входных данных.

В Loginom есть возможность проектировать сценарии как «снизу вверх», так и «сверху вниз», то есть при отсутствии входных данных.

Преимущества разработки сценария «сверху вниз» следующие:

- структурированность процесса проектирования;
- описание требований к данным для проектируемой модели;
- возможность повторного использования.

Недостатками модели являются:

- необходимо иметь представление о конечной структуре модели;
- необходимо продумать входы и выходы модели.

Сравнение функционала данных программ представлено в таблице 1.

Таблица 1 – Сравнение Loginom и Deductor

Функции	Программный продукт	
	Loginom	Deductor
Администрирование	Присутствует	Присутствует
Анализ больших данных	Присутствует	Отсутствует
Визуализация данных	Присутствует	Присутствует
Интеллектуальный анализ данных	Присутствует	Присутствует
Машинное обучение	Присутствует	Присутствует
Многопользовательский доступ	Присутствует	Присутствует
Наличие API	Присутствует	Отсутствует
Потоковая аналитика	Присутствует	Отсутствует
Создание собственных компонентов и подключаемые пакеты	Присутствует	Отсутствует

В Loginom администрирование позволяет управлять пользователями, рабочими папками их правами доступа, а также параметрами работы сервера.

В Deductor администрирование позволяет делать тоже самое, за исключением того, что там необходимо устанавливать и настраивать рабочие места, что усложняет работу.

Работа с большими данными в Logiном позволяет в массивах информации выявить тенденции и на основе этого строить поведенческие и предсказательные модели [2].

В Logiном визуализация данных может быть представлена в различных вариантах исходя из того, какой нужен формат отображения данных [3].

Варианты визуализаторов: диаграмма, куб, таблица, статистика, конечные классы, представлены на рисунке 3, отчёт по регрессии, качество бинарной классификации [4].

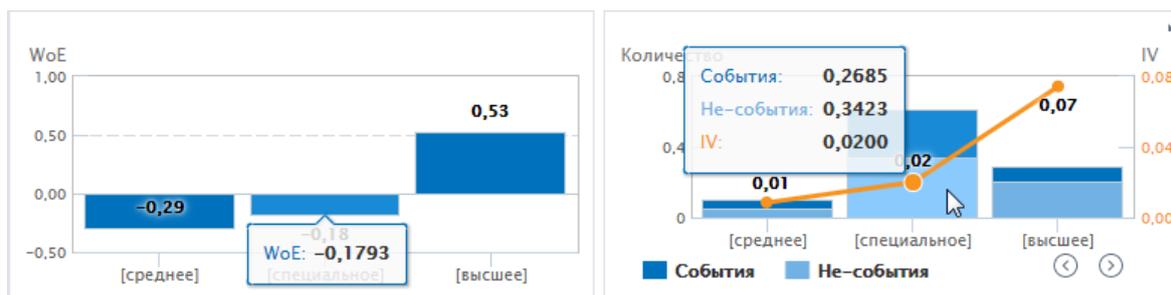


Рисунок 3. WoE-диаграмма и значения информационных индексов IV

Возможные варианты визуализации данных в Deductor: таблица, статистика, диаграмма, куб, матрица корреляции, факторный анализ, конечные классы, граф нейросети, дерево решений, карта Кохонена, отчет по регрессии, связи кластеров, метаданные.

Имеется возможность использовать модели интеллектуального анализа данных в Logiном в следующих сценариях:

- прогнозирование – анализ продаж, прогнозирование нагрузки сервера или времени простоя сервера.
- риск и вероятность - выбор наиболее подходящих клиентов для целевой рассылки, определение точки равновесия, присвоение вероятностей результатам.
- рекомендации – определение товаров, которые с большой долей вероятности могут быть проданы вместе, создание кросс-пакетов.
- поиск последовательностей – анализ выбора клиентов во время совершения покупок, прогнозирование возможных событий.
- группировка – разделение клиентов или событий на кластеры связанных элементов, анализ и прогнозирование общих свойств.

В Deductor интеллектуальный анализ данных позволяет строить следующие:

- решающие деревья – логический алгоритм классификации, основанный на поиске конъюнктивных закономерностей;
- самоорганизующиеся карты Кохенена – самообучающаяся нейронная сеть без учителя, выполняющая задачу визуализации и кластеризации;
- многослойные нейронные сети – нейронные сети, в которых нейроны сгруппированы в слои. В этом случае каждый нейрон предыдущего слоя связан со всеми нейронами следующего слоя, и между нейронами внутри слоев нет никаких связей, представлены на рисунке 4.

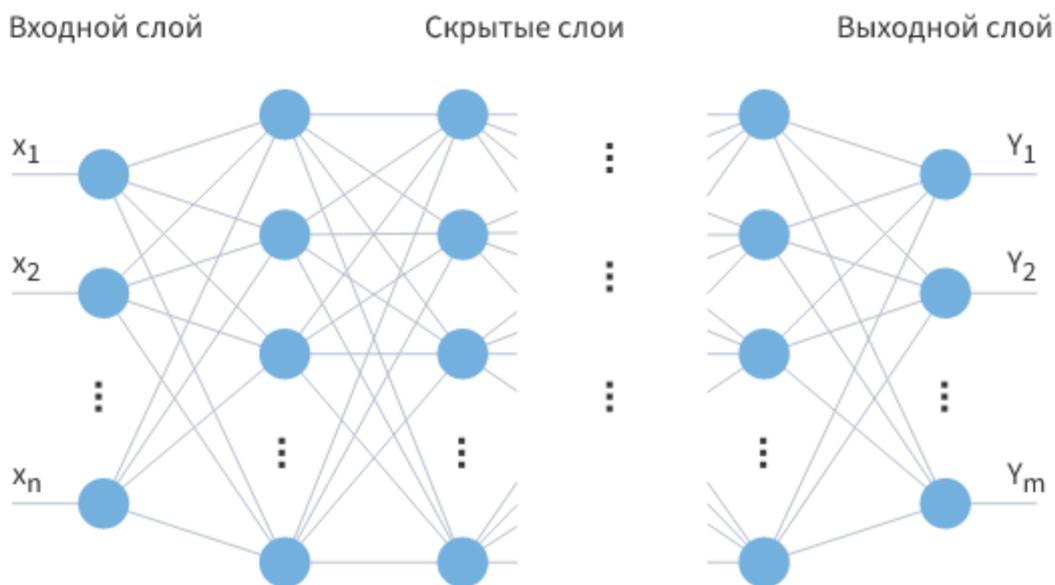


Рисунок 4. Многослойная нейронная сеть

Машинное обучение в обоих программных продуктах позволяет разрабатывать и строить аналитические модели, которые способны автоматически находить в данных скрытые закономерности, а также самостоятельно обнаруживать свойства, необходимые для определения этих закономерностей.

Наличие API в Loginom позволяет автоматически передавать данные из одного программного обеспечения в другое.

Потоковая аналитика в Loginom использует аналитические алгоритмы над данными в режиме реального времени для нахождения ключевых показателей бизнес-процессов.

Создание собственных компонентов и подключаемые пакеты позволяет аналитику создавать собственные компоненты и размещать их в общей палитре, при этом доступ к ним определяет сам автор.

Выводы

Исходя из вышесказанного, можно сделать вывод, о схожести данных аналитических платформ, однако при детальном сравнении Loginom выигрывает у Deductor. Так как у Loginom несколько больше список функций, что является для него плюсом. Deductor в некоторых функциях таких, как визуализация позволяет сделать больше, но в некоторых функциях, как администрирование он попросту усложнён лишними действиями.

Большим плюсом для Loginom является возможность проектирования сценариев «снизу вверх», что сильно расширяет его возможности.

Deductor поддерживается исключительно на операционной системе Windows, Loginom представляет из себя веб приложение, что позволяет ему не зависеть от операционной системы, что является большим плюсом для него.

Также для малого бизнеса больше подойдёт Loginom ввиду своей возможности развёртывания на облаке, что позволяет задействовать вычислительные мощности поставщика программы сэкономив на собственном серверном оборудовании.

Литература

1. Ткаченко А.Л. Business Intelligence: Современный взгляд // В сборнике: Информационные технологии в экономике, бизнесе и управлении. Материалы V Международной научно-практической конференции. 2018. С. 194-202.
2. Кондрашова Н.Г., Русу Я.Ю. Применение программных продуктов в сфере управления бизнес-проектами // Modern Economy Success. 2020. №5. С. 94-99.
3. Ткаченко А.Л., Полпудникова О.В. Анализ и моделирование бизнес-информации с помощью унифицированной программной платформы // В сборнике: Математическое моделирование в экономике, управлении, образовании. Материалы Международной научно-практической конференции. Под редакцией Ю.А. Дробышева и И.В. Дробышевой. 2015. С. 292-297.
4. Волкова Т.А, Сусякова О.Н. Страхование информационных рисков (киберстрахование) Инновационная экономика: перспективы развития и совершенствования. 2018. Т.1. №7 (33), С. 117-122.

УДК 004.9:66.02

ПРОБЛЕМНО-ОРИЕНТИРОВАННАЯ СИСТЕМА ДЛЯ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ ТЕПЛОЭНЕРГЕТИЧЕСКОГО ОБОРУДОВАНИЯ

PROBLEM-ORIENTED SYSTEM FOR TECHNICAL SUPPORT OF HEAT POWER EQUIPMENT MAINTENANCE

Белов В.Д., Мошев Е.Р., Ромашкин М.А.,
ФГАОУ ВО «Пермский национальный исследовательский
политехнический университет»,
г. Пермь, Российская Федерация

V.D. Belov, E.R. Moshev, M.A. Romashkin,
FSBEI HE “Perm national research polytechnic university”,
Perm, Russian Federation

e-mail:vladislav.belov.199@yandex.ru

Аннотация. В статье приведена доля тепловых электростанций в общей выработке тепловой и электрической энергии, а также показана актуальность внедрения интегрированной логистической поддержки оборудования теплоэнергетической отрасли. Показаны недостатки процессов технического обслуживания теплоэнергетического оборудования, выполняемого с помощью неспециализированного программного обеспечения, что снижает рентабельность предприятия и оказывает негативное влияние на уровень промышленной безопасности производства. Показано, что для устранения этих недостатков необходимо разработать и применить проблемно-ориентированную систему. Перечислены функциональные, фреймовые и продукционные модели представления знаний о теплоэнергетическом оборудовании, а также эвристические алгоритмы, которые применялись при разработке указанной системы. Даны результаты разработки проблемно-ориентированной системы,

функционал которой направлен на повышение эффективности процессов технического обслуживания теплоэнергетического оборудования. Показаны диалоговые окна основных графических редакторов системы, которые применяются для разработки изометрических схем теплоэнергетического оборудования, работы с эксплуатационной и ремонтной документацией, а также формирования индивидуальных программ контроля металла оборудования. Описаны основные функции и отличительные возможности редакторов проблемно-ориентированной системы, использование которых на системном уровне позволяет повысить эффективность процессов технического обслуживания теплоэнергетического оборудования, улучшить качество его ремонтной и эксплуатационной документации, а также унифицировать документы по контролю металла.

Abstract. The article shows the share of thermal power plants in the total production of heat and electricity, and also shows the relevance of the introduction of integrated logistics support for equipment in the heat and power industry. The disadvantages of the processes of technical maintenance of heat power equipment executed with non-specialized software are shown. These disadvantages reduce the profitability of the enterprise and have a negative impact on the level of industrial safety. It is shown that in order to eliminate these shortcomings, it is necessary to develop and apply a problem-oriented system. Functional, frame and production models of knowledge representation about heat power equipment are listed as well as heuristic algorithms that were used in the development of this system. The results of the development of a problem-oriented system, the functionality of which is aimed at increasing the efficiency of maintenance processes for heat and power equipment, are given. The dialog boxes of the main graphic redactors of the system are shown. The redactors are used to develop isometric diagrams of heat power equipment, work with operational and repair documentation as well as form individual programs for monitoring equipment metal. The main functions and distinctive capabilities of the redactors are described. Using the redactors at the system level makes it possible to increase the efficiency of maintenance processes for heat power equipment, improve the quality of its repair and operational documentation, and unify documents for metal control.

Ключевые слова: интегрированная логистическая поддержка, жизненный цикл, интерактивный электронный документ, теплоэнергетическое оборудование, проблемно-ориентированная система.

Keywords: integrated logistics support, life cycle, interactive electronic document, heat power equipment, problem-oriented system.

В Российской Федерации основным поставщиком тепловой и электрической энергии являются тепловые электростанции (ТЭС), доля которых в общей выработке указанных видов энергии (по данным Минэнерго за 2019 год) составляет 63,6%. При этом всё больше промышленных предприятий строят электростанции для обеспечения собственных производственных площадок электроэнергией, паром и горячей водой.

Повысить рентабельность теплоэнергетических предприятий, а также обеспечить их эффективную и безопасную эксплуатацию можно, если осуществлять техническое обслуживание технологического оборудования в соответствии с требованиями интегрированной логистической поддержки (ИЛП) [1].

В настоящей работе под ИЛП понимается совокупность видов инженерной и организационной деятельности, осуществляемых субъектами жизненного цикла (ЖЦ) теплоэнергетического оборудования и направленных на формирование системы

технической эксплуатации, обеспечивающей эффективное использование этого оборудования при приемлемой стоимости его ЖЦ[2].

На сегодняшний день полноценная ИЛП изделий осуществляется только в некоторых отраслях промышленности: двигателе- и самолётостроение, военная техника, космонавтика и т.п. В химической, нефтехимической и теплоэнергетической промышленности ИЛП оборудования осуществляется с помощью мало специализированных программных продуктов либо с помощью разрозненного и узкоспециального программного обеспечения (ПО), которое не обеспечивает комплексного подхода к техническому обслуживанию оборудования, а решает только частные задачи. При таком несистемном подходе возникает ряд недостатков, существенно снижающих эффективность ИЛП, как организационно-технологического процесса, а именно:

- несовпадение информации, находящейся в отдельных документах и ПО, применяемом отдельными субъектами ЖЦ оборудования;
- многократное создание разрозненных схем и чертежей одного и того же оборудования на различных этапах ЖЦ;
- низкий уровень автоматизации при работе с паспортно-технической и нормативной документацией;
- сложность обмена данными между субъектами ЖЦ оборудования;
- многократный ввод одних и тех же данных об оборудовании в различные документы, программные продукты;
- высокая сложность выполнения интеллектуальных процедур создания индивидуальных программ контроля, расчётов остаточного ресурса оборудования, расчётов на прочность;
- отсутствие синхронизации информации и изометрических (принципиальных) схем оборудования между отдельными субъектами его ЖЦ.

Указанные недостатки ИЛП существенно снижают эффективность процессов технического обслуживания (ТО) теплоэнергетического оборудования.

Анализ научно-технической литературы показал, что для повышения качества ИЛП необходимо разработать и использовать проблемно-ориентированную систему (ПОС), реализующую методологию интегрированной логистической поддержки применительно к теплоэнергетическому оборудованию.

Ранее по результатам анализа нормативно-технической документации (НТД), а также паспортно-технической и эксплуатационной документации на оборудование тепловых и гидроэлектростанций были разработаны следующие модели и алгоритмы, позволяющие создать специализированное ПО для теплоэнергетической отрасли [3]:

- функциональные модели, формализующие процессы составления индивидуальных программ контроля металла котельных и турбинных агрегатов. Функциональные модели выполнены в соответствии с методологией SADT [4];
- фреймовые модели представления знаний о теплоэнергетическом оборудовании, которые формализуют связи между оборудованием ТЭС на системном уровне и позволяют выделить иерархическую структуру станции в целом;
- производственные модели представления знаний о методах, объёмах и периодичности контроля металла, которые формализуют требования НТД теплоэнергетической отрасли к проведению неразрушающего контроля;
- эвристическо-вычислительные алгоритмы, позволяющие автоматизировать интеллектуальные процедуры составления индивидуальных программ контроля металла.

В настоящей статье представлены результаты разработки ПОС под названием Автоматизированная Система Управления Ресурсом Металла (АС УРМ), реализованной с применением перечисленных выше моделей и алгоритмов.

Основными отличительными особенностями АС УРМ по сравнению с другим ПО являются:

1. представление технологического оборудования в виде иерархической структуры, соответствующей реальной структуре предприятия (рисунок 1), что позволяет работать не только с отдельными единицами оборудования, но и с группами оборудования по установке;

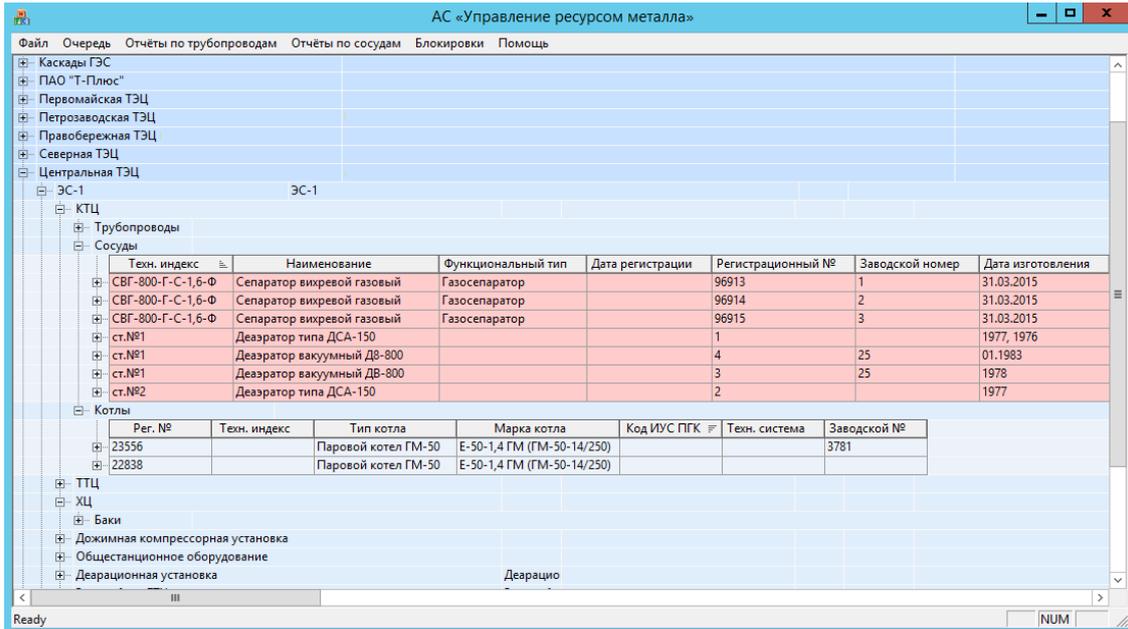


Рисунок 1. Стартовое диалоговое окно АС УРМ, отражающее иерархическую структуру предприятия

2. наличие графических редакторов схем трубопроводов (рисунок 2), сосудов (рисунок 3), а также динамического оборудования (рисунок 4);

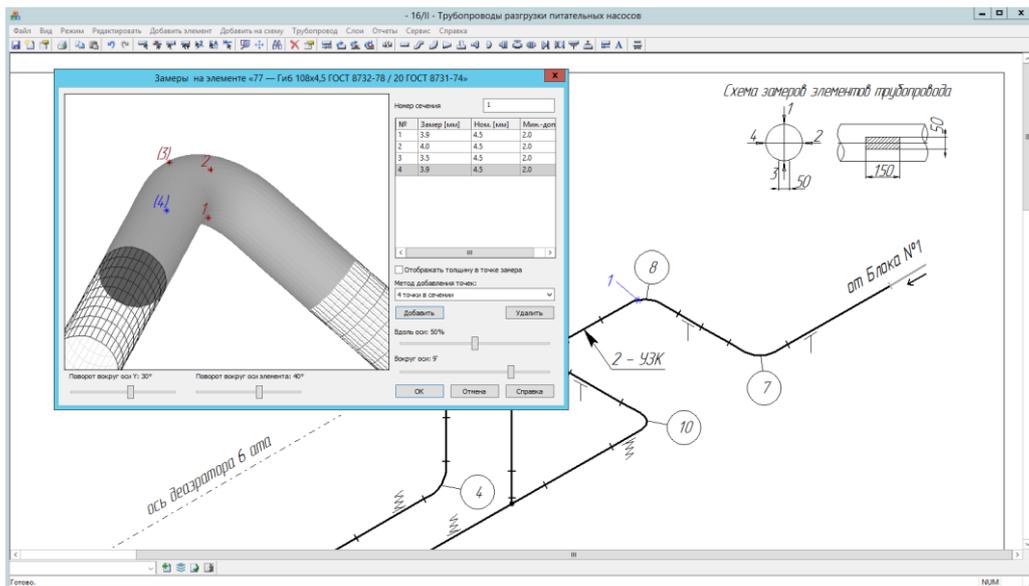


Рисунок 2. Графический редактор схем трубопроводов

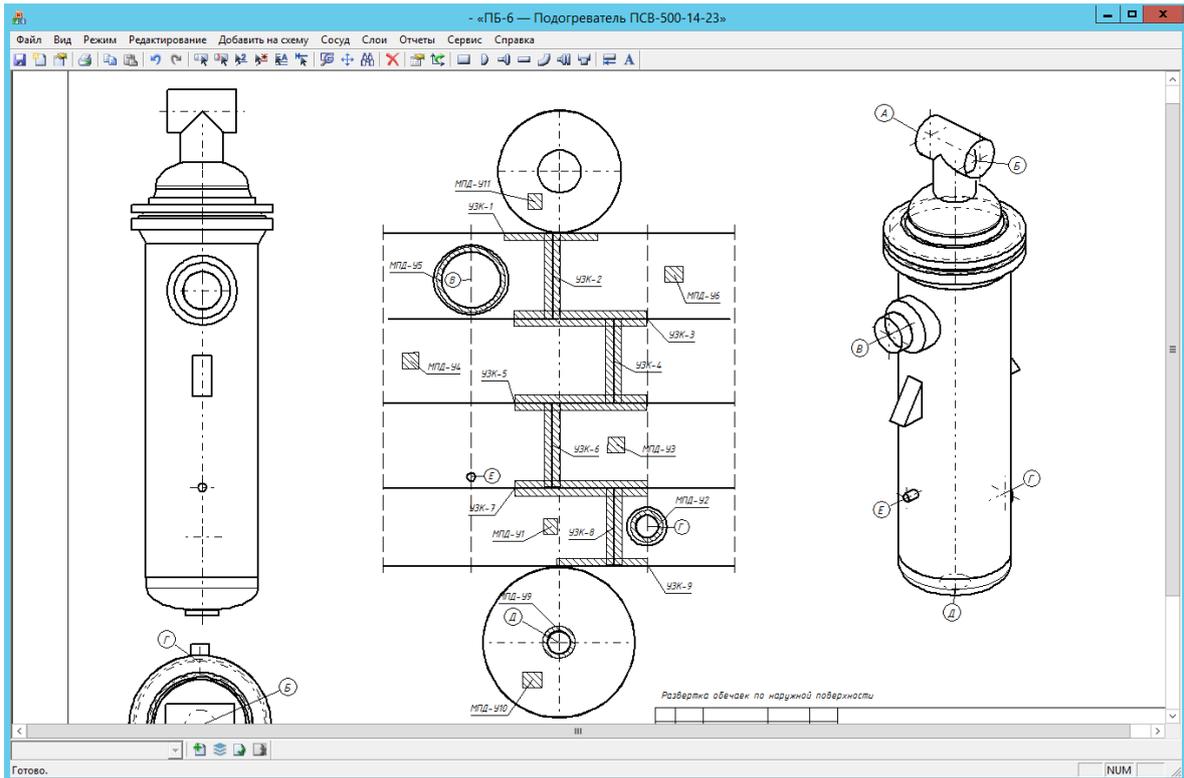


Рисунок 3. Графический редактор схем сосудов

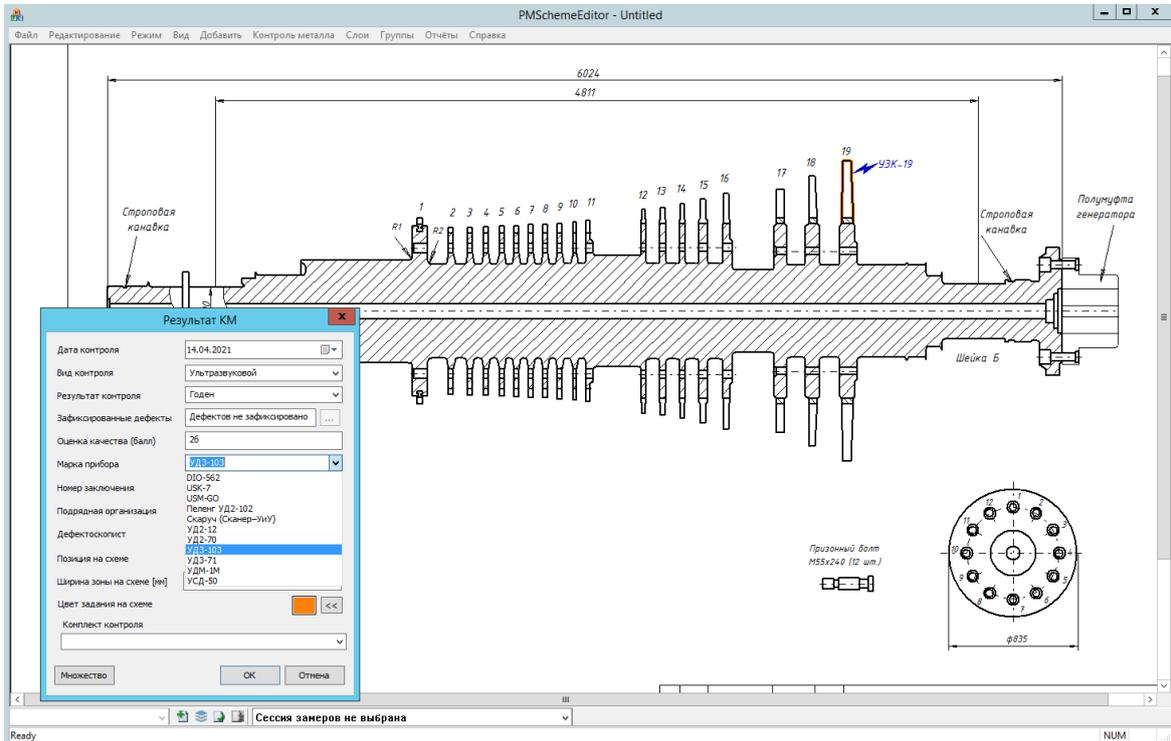


Рисунок 4. Графический редактор схем динамического оборудования

3. учтены требования НТД, регламентирующей процессы ТО теплоэнергетического оборудования;

4. возможность работы с гидромеханическим оборудованием ГЭС (затворы, сороудерживающие решётки);

5. добавлены специфические для теплоэнергетики типы оборудования: котлы,

турбины, генераторы, баки;

6. реализована функция учёта наработки и числа пусков для всех типов оборудования;

7. реализована функция автоматического формирования индивидуальных программ контроля металла оборудования в соответствии с установленными требованиями;

8. имеется возможность внесения замеров остаточных деформаций и тепловых перемещений трубопроводов, работающих в условиях ползучести;

9. реализована функция создания развёрток для обечаек сосудов, а также работы с развёртками: добавления неразрушающего контроля, дефектов, примечаний и т.п.;

10. добавлена функция импорта внешних графических элементов в формате .dxf;

Изометрические схемы оборудования, созданные с помощью редакторов схем трубопроводов, сосудов и динамического оборудования являются интерактивными электронными документами.

Использование определенных ролей при авторизации пользователей позволяет разграничить права доступа к документу. Так, например, начальник установки в соответствии со своей ролью может только утверждать программы контроля металла, вносить записи в эксплуатационный журнал оборудования и редактировать структуру технологической установки.

Роль инженера-дефектоскописта позволяет просматривать схемы, печатать их и вносить результаты контроля металла в соответствии с утверждённой программой.

Благодаря применению нормативных баз данных значительно упрощается процесс создания и редактирования схем сосудов и трубопроводов.

Характеристики всех элементов на схеме пользователь может задать в автоматическом режиме в соответствии с действующей НТД.

Редактор схем динамического оборудования работает не с базами данных элементов, а с внешними файлами данных других графических редакторов. Пользователь переносит существующие схемы в формате .dxf в редактор, после чего работает с графическими примитивами на схеме – группирует, добавляет, назначает элементы, которые необходимо контролировать.

Помимо построения схем сосудов и трубопроводов в АС УРМ возможно создавать схемы узлов котельных агрегатов, деталей паровых турбин и электрогенераторов, а также оборудования гидроэлектростанций – механических затворов, сороудерживающих решёток и водяных турбин. Ниже представлены другие основные функции АС УРМ:

– определение основных характеристик оборудования в соответствии с действующей НТД (категория, группа и подгруппа трубопровода, класс опасности среды и т.п.);

– группировка оборудования по заданным критериям и работа с группами;

– формирование в автоматическом режиме индивидуальных программ контроля металла, заданий на контроль, отчётов по проведённому контролю: заключения, акты и протоколы по всем видам проведённого неразрушающего контроля;

– формирование эксплуатационной и ремонтной документации: паспорт, спецификация, задание на снятие тепловой изоляции и зачистку контрольных зон, акты испытания и приёмки оборудования;

– внесение и обработка результатов неразрушающего контроля с помощью автоматически генерируемых шаблонов;

– расчёты на прочность элементов сосуда или аппарата, определение остаточного ресурса оборудования.

Выводы

Разработанная ПОС за счёт оптимизации процессов технического обслуживания, унификации и ускорения формирования отчётной, ремонтной и эксплуатационной документации и позволяет повысить качество ИПП теплоэнергетического оборудования и как следствие – снизить затраты на его содержание.

Литература

1. Мешалкин В.П. Логистика и электронная экономика в условиях перехода к устойчивому развитию. Москва-Генуя: РХТУ им. Д.И. Менделеева, 2004. 573 с.
2. ГОСТ Р 53394–2017. Интегрированная логистическая поддержка. Основные термины и определения. Введ. 14.09.09. М.: Стандартинформ, 2018. 23 с.
3. Belov, V.D., Moshev, E.R.: Functional Model for the Formation of Individual Metal Control Programs of Boiler Equipment. In: Cyber-Physical Systems: Modelling and Intelligent Control, pp. 323-334. Springer, Cham (2021). URL:https://doi.org/10.1007/978-3-030-66077-2_26
4. Дэвид А. Марка, Клемент Мак Гоуэн. Методология структурного анализа и проектирования SADT. М.: Метатехнология, 1993. 239 с.

УДК 004.81'32, 81'42

МЕТОДЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ СОБЫТИЙ В ПИСЬМЕННОМ ТЕКСТЕ

METHODS FOR AUTOMATIC RECOGNITION OF EVENTS IN WRITTEN TEXT

Гуртовой Я.А.,
УО «Минский государственный лингвистический университет»,
г. Минск, Республика Беларусь

Y.A. Goortovoi,
Minsk State Linguistic University,
Minsk, Republic of Belarus

e-mail: neoyarus@gmail.com

Аннотация. В статье анализируется проблема автоматического распознавания событий в письменном тексте новостных сообщений на английском языке. Логика развития технологий извлечения информации из текстов на естественном языке показывает, что создание эффективной системы извлечения событий и фактов достаточно актуальна, поскольку может значительно облегчить последующую обработку извлеченных структурированных данных. Применительно к распознаванию событий рассмотрены различные виды машинного обучения, такие как обучение с учителем, обучение без учителя и используемые ими алгоритмы (метод Байеса, скрытая марковская модель, логический метод, алгоритмы кластеризации). Показаны преимущества применения машинного обучения к задачам автоматического извлечения событий, которые заключаются в возможности проанализировать исходные данные и

найти в них закономерности без необходимости контролировать каждое действие в программе и даже с минимумом лингвистических знаний. В статье также приведены результаты собственной разработки по извлечению событий методом, основанным на лингвистических шаблонах. В качестве материала исследования выбраны новостные тексты в сети Интернет. Дальнейшее изучение методов автоматического извлечения событий в новостном дискурсе позволит создавать системы их распознавания в медийных текстах для последующего анализа, что в перспективе избавит массовые коммуникации от приемов манипулирования общественным сознанием.

Abstract. The article analyzes the problem of automatic recognition of events in the written text of news messages in English. The logic of the development of technologies for extracting information from natural language texts shows that the creation of an effective system for extracting events and facts is quite relevant, since it can significantly facilitate the subsequent processing of the extracted structured data. With regard to event recognition, various types of machine learning are considered, such as supervised learning, unsupervised learning and the algorithms they use (Bayesian method, hidden Markov model, logical method, clustering algorithms). The advantages of applying machine learning to the tasks of automatic event extraction are shown, which consist in the ability to analyze the initial data and find patterns in it without the need to monitor every action of the program and with only a minimum of linguistic knowledge. The article also presents the results of our own attempt at event extraction using a method based on linguistic templates. News texts on the Internet were selected as the research material. Further study of the methods of automatic extraction of events in news discourse will make it possible to create systems for their recognition in media texts for subsequent analysis, which, in the future, will save mass communications from methods of manipulating public consciousness.

Ключевые слова: событие, автоматическое распознавание событий, машинное обучение, виды машинного обучения, обучение с учителем, обучение без учителя.

Keywords: event, automatic event recognition, machine learning, types of machine learning, supervised learning, unsupervised learning.

Возрастающий объем информации в открытых источниках данных обусловил необходимость разработки способов автоматического распознавания событий в письменном тексте. Автоматическое извлечение событий, возникшее как междисциплинарное прикладное поле на стыке между компьютерной лингвистикой и искусственным интеллектом, в настоящее время не имеет четко установленной терминологической системы. Изучение и совершенствование методов распознавания событий является актуальной задачей для автоматического сентимент-анализа текста. Данные методы позволят создавать системы автоматического распознавания определенных лингвистических паттернов в текстах для их последующего анализа.

В корпусной лингвистике понятие «событие» подразумевает определенную языковую единицу, описывающую некоторое явление во всей его полноте с помощью независимо существующих утверждений. С лексической точки зрения в описании события ключевую роль играет глагол, как основной событийный компонент предложения. Следует отметить, что в настоящий момент нет общепринятых универсальных критериев для анализа англоязычных текстов с целью автоматического определения и оценки событий.

Одним из первых применений систем автоматического анализа англоязычных текстов стало использование корпусной технологии для автоматического извлечения

событий в 1997 году на материале Wall Street Journal. На основе корпуса были проведены исследования семантической ориентации прилагательных. В результате был составлен словарь, который лег в основу большого количества дальнейших разработок. В дальнейшем список слов расширился, были введены новые типы прилагательных, наречий и глаголов, значительно усложнилось расположение разметки корпуса, в дополнение к морфологическим факторам, начали использоваться синтаксически размеченные слова.

Большая часть информации, доступной в электронном виде, представлена текстами на естественном языке. Полезная информация, содержащаяся в них, не структурирована, что означает, что ее нельзя обрабатывать и анализировать с использованием классических методов. Чтобы извлечь информацию из прочитанного фрагмента текста (понять текст), человек должен знать язык, на котором он написан, и обладать «базовыми» знаниями. Аналогично, система извлечения информации из текста должна иметь средства анализа естественного языка и определенный объем знаний в предметной области.

Технология извлечения информации из текстов на естественном языке позволяет автоматически визуализировать относительно большое количество текстов, содержащих относительно небольшое количество искомой информации. Найденная в тексте информация преобразуется в структурированный формат: факты, объекты, отношения должным образом идентифицируются для дальнейшей автоматической обработки (статистическая обработка, визуализация, поиск закономерностей в данных и т. п.) [1].

Information Extraction (IE) – задача извлечения информации – рассматривает вопросы идентификации определенных сущностей и отношений в неструктурированных данных, и прежде всего в текстовых документах [2]. До недавнего времени указанная задача решалась в основном с использованием лингвистических паттернов, разрабатываемых экспертами-лингвистами. В последнее время очень интенсивно стали использоваться методы, основанные на нейронных сетях и машинном обучении [3, с. 9]. Дальнейшая автоматизация машинного обучения является перспективным направлением исследований в области извлечения информации.

К видам машинного обучения относят: обучение с учителем, обучение без учителя, обучение с подкреплением. Традиционно методы машинного обучения используют следующие основные алгоритмы: метод Байеса, марковскую модель, логический метод и др. Каждый из этих алгоритмов выбирается в зависимости от поставленной задачи. *Метод Байеса* используется для определения вероятности события, происходящего в условиях, когда информация доступна частично, например, согласно наблюдаемым сигналам, о принадлежности объекта одному из указанных классов.

Скрытая марковская модель (Hidden Markov chains) [4, с. 257] это статическая модель, которая имитирует некоторые последовательные процессы, в которых скрытые состояния находятся под влиянием переменных состояний. Переход из одного состояния в другое происходит с определенной вероятностью. Таким образом из последовательности наблюдений можно получить информацию о последовательности событий.

Логические методы включают алгоритмы, в которых логические правила организуются в виде дерева решений и строятся в процессе обучения. В узлах проверяется значение атрибутов, выбранных во время обучения, и принимается решение разделить объекты на подклассы. Логические алгоритмы работают как с символическими, так и полными описаниями и в пространстве булевых признаков [3, с. 10].

Наиболее популярным методом машинного обучения является *обучение с учителем* (*supervised learning*), когда модель строится на основе размеченных данных

(обучающего набора), которая затем применяется к новым неразмеченным текстам. Целью обучения с учителем является создание необходимых и достаточных правил, с помощью которых можно произвести классификацию новых объектов, сходных с теми, которые составляли обучающую выборку. При этом каждый обучающий пример (описание объекта) имеет метку, показывающую, к какому классу он принадлежит.

При *обучении без учителя* (*unsupervised learning*) задача состоит в том, чтобы объединить объекты в группы, которые не пересекаются попарно, на основе определенной степени их сходства или различия. Эта задача известна как *кластеризация* объектов. Обучение без учителя обычно используется для анализа структуры данных, а также для формирования обучающей выборки для последующего применения обучения с учителем, чтобы найти правила классификации, описывающие результирующее разделение объектов на группы в пространстве признаков.

Основателем алгоритма машинного обучения без учителя можно считать Питера Терни [5, с. 417]. Используя данный алгоритм и минимум лингвистических знаний, он добился отличных результатов и продемонстрировал принципиальную возможность применения машинного обучения к задачам автоматического извлечения событий [6].

К недостаткам существующих работ можно отнести то, что в них не учитывается степень изменения в оценках сочетаний слов. В то же время часто отмечается, что «способы выражения категории эмфатичности разнообразны и многочисленны и затрагивают все уровни языка» [7, с. 10] и «эмоциональность текста формируется за счет языковых средств всех уровней и воздействует на общую выразительность текста» [8, с. 179]. Однако следует отметить, что нет четких правил для определения оценки произвольных имен прилагательных и имен существительных в одной и той же группе, с учетом их семантических особенностей.

Основными этапами машинного обучения являются:

- 1) выбор и формирование признаков пространства;
- 2) проверка гипотез о различимости сходства объектов и классов объектов; спецификация бинарных операций разности подобия между объектами; спецификация мер сходства/различия для классов объектов;
- 3) формирование обучающей выборки;
- 4) формирование контрольной выборки;
- 5) выбор алгоритма обучения [3, с. 7].

Преимуществом методов машинного обучения без учителя является отсутствие необходимости создавать правила или онтологии. К недостаткам относится высокая трудоемкость маркировки обучающего корпуса.

Основные методы в области машинного обучения можно также отнести к двум классам: «ленивое» обучение (*lazy learning*) и «нетерпеливое» обучение (*greedy learning*). Существенное различие между этими подходами заключается в том, что при «ленивом» обучении извлекаемая информация не обобщается, в то время как при «нетерпеливом» обучении извлекаемая информация обобщается посредством реструктурирования и удаления избыточных и несущественных частей [3, с. 12].

Подход *lazy learning* основан на гипотезе, что решение когнитивных проблем (в частности, преподавания языка) базируется на выводах, основанных на аналогиях, а не на абстрактных правилах, полученных из экспериментов. Этот подход используется в различных направлениях искусственного интеллекта и использует такие методы, как умозаключение на основе сходства, умозаключение на основе примера, умозаключение на основе аналога, умозаключение на основе прецедента (логическое обоснование) и т. д. При «ленивом» обучении примеры обучения без обобщений и реструктуризации добавляются в память. Сходство нового примера с другими рассчитывается на основе показателя сходства, а категория наиболее похожих примеров используется в качестве

основы для прогнозирования категории нового примера. Этот подход используется для фонологических и морфологических проблем, проблем распознавания речи, морфологического и синтаксического анализа, а также для решения проблем морфосинтаксической и семантической неоднозначности.

Исследуемая задача распознавания событий в тексте имеет самое непосредственное отношение к таким актуальным направлениям в общей проблеме автоматической обработки текста, как извлечение информации из текстов (Information Extraction) и извлечение знаний из текстов (Text Mining). В первом случае речь идет о выделении явной информации, доступной в текстах, например, ключевых слов, дат, наименований организаций, названий, описания, аспектов бронирования авиарейсов, гостиниц и т.п. Извлечение информации можно рассматривать как неизбежный предварительный шаг в более серьезных задачах извлечения событий из текстов. На этом этапе в основном используются различные системы классификации и категоризации текста, основанные на методах машинного обучения [3, с. 13].

Самая сложная задача – *извлечение информации о фактах и событиях* из новостных текстов. События обычно включают несколько наименований объектов, которые связаны определенным набором отношений. Информация и типичные конструкции их выражения используются для извлечения фактов и событий. Распознавание события не является проблемой, если все обязательные атрибуты содержатся в одном предложении. Однако извлечение событий может стать сложным процессом, особенно для событий с множеством атрибутов, поскольку составляющие их части могут находиться в разных предложениях текста, а иногда даже не в соседних.

Для общей оценки эффективности методов извлечения событий обычно используется способ, предложенный на конференциях MUC-5 и MUC-6 [9, с. 20]. Для каждого извлеченного факта проверяется, правильно ли заполнены все атрибуты. Если это так, он считается правильно извлеченным. В этом случае, если значение атрибута не было идентифицировано (например, местоположение события), и текст на самом деле не говорит об этом, упущение не считается неправильным (так как оно не может быть заполнено вообще). Таким образом, факт извлекается правильно, если значения всех атрибутов, которые могут быть заполнены, заполнены правильно. Заполненный атрибут считается частично правильным, если не вся информация, связанная с ним, была извлечена, или если была извлечена дополнительная информация.

Метод, основанный на лингвистических шаблонах, является наиболее подходящим методом для извлечения событий благодаря скорости реализации и возможности расширения алгоритмов. Вместо сложного вычисления ядер выполняется процедура сравнения с лингвистическим шаблоном, который является более простым с точки зрения сложности вычислений [10, с. 2].

Основная задача для извлечения фактов с использованием метода шаблона заключается в выделении конкретных событий, упомянутых в тексте. Приведем данные способы в таблице 1.

Таблица 1 – Способы извлечения событий (собственная разработка)

№ пп	Атрибут	Способ извлечения
1	Дата встречи	Словарь и лингвистический шаблон
2	Время встречи	Лингвистический шаблон
3	Название встречи	Лингвистический шаблон
4	Место встречи	Словарь и лингвистический шаблон
5	Отправитель	Лингвистический шаблон

Чтобы извлечь дату и время встречи из входящего сообщения, используется небольшой специализированный словарь, который включает названия месяцев, дней недели и сезонов.

Для определения «названия встречи» был разработан простой алгоритм работы, который был основан на определении в тексте предложений в кавычках («Конференция X») или начинающихся с заглавной буквы. Если кавычек нет, алгоритм ищет слова, начинающиеся с заглавной буквы или аббревиации, и определяет прилагательные, которые к ним относятся. Если конкретное предложение заканчивается прилагательным, существительное, которое следует за ним, также включается в сущность. Чтобы определить атрибут названия встречи, были определены 4 основные лингвистические модели, которые приведены в таблице 2.

Таблица 2 – Лингвистические модели (собственная разработка)

№ пп	Шаблон	Пример
1	Сущ.+прил. / прил.+сущ.	Конференция X
2	Сущ. с заглавной буквы+прил. / прил. с заглавной буквы+сущ.	Европейская конференция
3	Сущ. с заглавной буквы+прил. / прил. с заглавной буквы+прил.+сущ.	Европейская конференция / Европейская научная конференция
4	Аббревиатура+сущ. / сущ.+аббревиатура	АйТи конференция

Для определения атрибута «Место встречи» все слова из текста сообщения приводятся к нормальной форме функцией `normal_form` библиотеки `rumorphy2` для языка Python. Затем преобразованный текст ищет подходящие слова из словаря (например, «университет» или «аудитория»). Если совпадения найдены, найденное слово включается в атрибут, следующее слово или, если следующее слово является прилагательным, оно включается в атрибут, а затем в следующие два слова, включая существительное. Например: «Аудитория №5», «Университет Гамбурга».

Очевидно, что реализация представленного метода требует разработки и задания упомянутых словарей, относительно небольших по объему, а также использования процедур автоматической предобработки текста с целью его лексико-грамматического анализа (POS tagging). Последнее осуществимо, например, с помощью имеющейся в свободном доступе известной библиотеки `Stanford CoreNLT`.

Выводы

Автоматическое извлечение событий, которые несут в себе значимую информацию, является приоритетной задачей в области обработки текстов. Метод, основанный на лингвистических шаблонах, является наиболее подходящим методом для извлечения событий благодаря скорости реализации и возможности расширения алгоритмов, который основывается на составлении и реализации лингвистических шаблонов для выделения необходимых атрибутов и на составлении и применении тематических словарей.

Извлечение событий из неструктурированных данных, таких как новостные сообщения могут быть применимы для практических целей. Например, если система искусственного интеллекта сможет определить событие, это улучшит производительность персонализированных информационных систем, поскольку сообщение может быть выбрано более точно в зависимости от предпочтений пользователя. Таким образом, извлечение событий из текста становится ключевым

компонентом в процессе интеграции текстовых данных, и может рассматриваться как общая составляющая для многих важных задач, таких как распознавание имен объектов, анализ тона текста или извлечение знаний.

Литература

1. Трофимов И.В., Кормалев Д.А., Куршев Е.П., Сулейманова Е.А. Технология извлечения информации из текстов, основанная на знаниях // Программные продукты и системы. 2009. №2. С. 62-66.
2. Klugl P. Context-specific Consistencies in Information Extraction: Rule-based and Probabilistic Approaches. // BoD – Books on Demand, 2015. 208 p.
3. Найденова К.А., Невзорова О.А. Машинное обучение в задачах обработки естественного языка: обзор современного состояния исследований // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. 2008. №4. С. 5-24.
4. Rabiner L.R. A tutorial on hidden Markov models and selected application in speech recognition // Proc. IEEE. 1989. V. 77. No2. P. 257-286.
5. Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Penn, 7-12 July 2002. P. 417-424.
6. Dentith M R. X. The Problem of Fake News //Public Reason. 2017. №8 (1-2). P. 65-79.
7. Жирова И.Г. Лингвистическая категория эмфатичность в антропоцентрическом аспекте: автореф. дис. д-ра филол. наук: 10.02.19. Моск. гос. обл. ун-т. – М., 2007. 40 с.
8. Маслова В.А. Параметры экспрессивности текста // Языковые механизмы экспрессивности; отв. ред. В.Н. Телия. М.: Наука, 1991. С. 179-204.
9. Chinchor N., Sundheim B. MUC-5 Evaluation Metrics. // Proceedings of the Fifth Conference on Messages Understanding. Association for Computational Linguistics, USA. 1993. P. 69-78.
10. Котельников Д.С., Лукашевич Н.В. Итерационное извлечение шаблонов описания событий по новостным кластерам // Труды конференции RCDL-2012, 15-18 октября 2012. С. 292-298.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ЭКОНОМИКЕ, УПРАВЛЕНИИ И БИЗНЕСЕ

УДК 004.89

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЗАЦИИ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА НА ПРЕДПРИЯТИИ

APPLYING MACHINE LEARNING TO AUTOMATE ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS

Бикметова М.Р., Азбуханов А.Ф.,
ООО «РН-БашНИПИнефть»,
г. Уфа, Российская Федерация

M.R. Bikmetova, A.F. Azbukhanov,
RN-BashNIPIneft LLC,
Ufa, Russian Federation

e-mail: BikmetovaMR@bnipi.rosneft.ru

Аннотация. Любая крупная организация сталкивается с необходимостью оперативной обработки входящей корреспонденции. В подавляющем большинстве случаев, поступающие письма регистрируются в системе электронного документооборота вручную. В связи с большим объемом корреспонденции и ограниченной скоростью работы операторов систем электронного документооборота возникает временной лаг между получением письма и началом работ по прилагаемому к нему документу. Автоматизировать анализ, обработку и регистрацию входящих писем в системе электронного документооборота позволяют методы машинного обучения. Использование обученных алгоритмов сокращает трудозатраты на рутинные операции, а также увеличивает оперативность выполнения работ. В статье описаны разработанные алгоритмы для автоматического извлечения реквизитов документов и определения тематики текста. В работе также представлена архитектура интеграции созданных моделей с имеющейся ИТ инфраструктурой предприятия.

Abstract. Any large enterprise is faced with the need for the prompt processing of incoming correspondence. In the vast majority of cases, incoming documents are registered in the electronic document management system manually. Due to the large volume of emails and the limited work rate of the electronic document management system's operators, there is a time lag between the receipt of the letter and the beginning of work on the attached document. Machine learning methods allow automating the analysis, processing, and registration of incoming correspondence in the electronic document management system. The use of trained algorithms reduces labour costs for routine operations, and also increases the efficiency of work. The article describes the developed algorithms for information extraction and document classification. It also presents how the created models are integrated with the existing IT infrastructure of the enterprise.

Ключевые слова: машинное обучение, интеллектуальный анализ данных, классификация документов, извлечение информации, системы электронного документооборота.

Keywords: machine learning, intelligent data analysis, document classification, information extraction, electronic document management systems.

Переход предприятий на электронный документооборот является необходимым шагом на пути автоматизации производственных процессов. Несмотря на разнообразие существующих систем электронного документооборота (СЭД), большинство из них подразумевают наличие оператора, вручную вносящего информацию с бумажных носителей или отсканированных документов. Стоит отметить, что значительную часть рутинных операций можно автоматизировать. Примеры действий в СЭД, поддающихся полной автоматизации:

- создание карточек для регистрации входящих документов.
- внесение реквизитов документа (номер, дата, отправитель и т.п.).
- определение тематики документа.
- выбор ответственного за исполнение документа в соответствии с темой документа.

Мощным инструментом для автоматизации СЭД может выступить внедрение методов машинного обучения. К примеру, алгоритм может распознавать реквизиты входящих документов и автоматически вносить данные в СЭД без вмешательства оператора. Таким образом, сокращается время на рутинные операции, уменьшается временной лаг между получением документа и началом работ по нему. Как следствие, полная автоматизация работы СЭД позволит повысить эффективность работы сотрудников и принесет экономическую выгоду предприятию.

Выше отмечались рутинные операции, поддающиеся автоматизации с помощью алгоритмов машинного обучения. Для реализации подобной системы необходимо разработать и интегрировать с имеющейся ИТ инфраструктурой предприятия следующие компоненты (рисунок 1):

1. сервис распознавания текста со скана документа;
2. сервис выделения реквизитов документа;
3. сервис классификации текста документа для определения тематики.

С помощью планировщика (task scheduler) скрипт-обработчик каждые 5 минут проверяет основной почтовый аккаунт предприятия. При наличии нового письма для него создается новая карточка в СЭД. Вложения письма отправляются в сервисы машинного обучения для анализа и извлечения информации. Распознанные реквизиты используются для автоматического заполнения полей в СЭД. Модель машинного обучения прогнозирует тему документа, после чего назначается ответственный за конкретную тему исполнитель. Далее подробно рассмотрен каждый из этапов обработки входящего документа.

Распознавание текста

В первую очередь вложение из письма поступает в сервис распознавания текста для извлечения текстового слоя из скана документа. Для увеличения точности оптического распознавания символов (англ. optical character recognition, OCR), изображение разбивается на блоки с помощью алгоритма топологического структурного анализа бинарных изображений [1]. В работе используется реализация данного алгоритма, представленная в библиотеке OpenCV [2]. Далее на каждом выделенном блоке производится распознавание текста с помощью библиотеки tesseract [3].

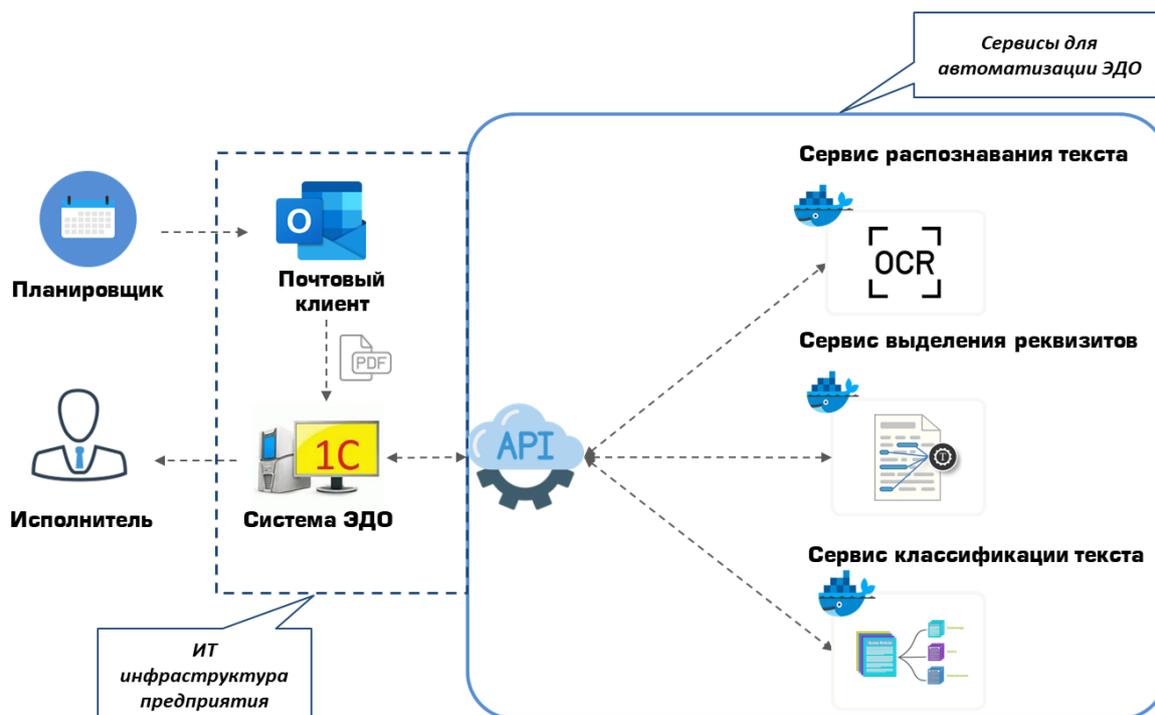


Рисунок 1. Схема взаимодействия сервисов машинного обучения и ИТ инфраструктуры предприятия.

Выделение реквизитов

Координаты выделенного блока (пространственные признаки) и распознанный текст блока (текстовые признаки) отправляются в сервис выделения реквизитов. Проблема извлечения реквизитов рассматривается как задача классификации. Имеется множество блоков X и конечное множество меток классов Y . Существует неизвестная целевая функция $y^*: X \rightarrow Y$, значения которой известны только на конечном подмножестве блоков. Данное подмножество называется обучающей выборкой. Необходимо построить решающую функцию $a: X \rightarrow Y$, которая приближала бы целевую функцию $y^*(x)$ не только на блоках обучающей выборки, но и на всём множестве блоков X [4].

Для построения такого алгоритма в данной работе используется ансамбль моделей – стеккинг. Основная идея подхода заключается в том, что на данных обучаются несколько алгоритмов, их прогнозы объединяют и используют в качестве обучающей выборки для метамодели. Таким образом, модель 2-го уровня делает прогноз, исходя не только из сырых данных, но и на основе предсказаний моделей 1-го уровня (Рисунок 2).

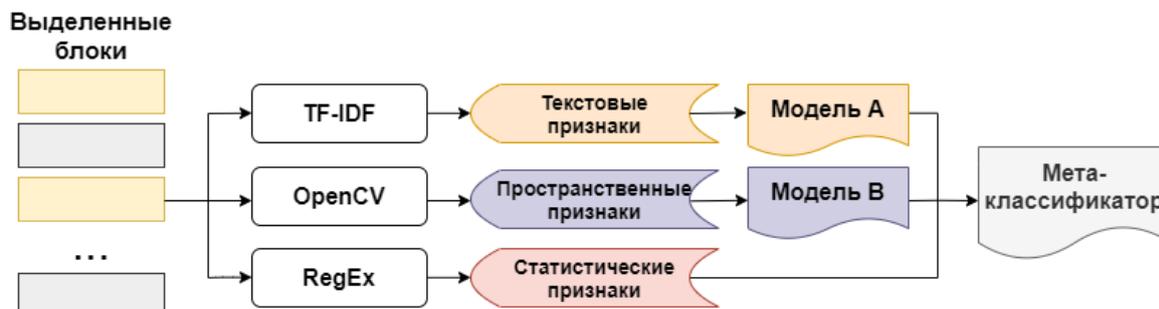


Рисунок 2. Схема разработанной модели машинного обучения

Входом модели А является текст, содержащийся в выделенном блоке. Предварительно производится очистка текста: удаляются знаки препинания и не несущие смысловую нагрузку слова (предлоги, союзы, частицы и т.п.).

Далее осуществляется стемминг – процедура сокращения конкретной словоформы до грамматической основы.

Данный метод позволяет значительно снизить размерность данных за счет исключения различных вариаций одного слова.

Одним из требований алгоритмов машинного обучения является представление подаваемых на вход данных в виде числовых векторов.

Для векторизации текста в работе используется алгоритм TF-IDF (англ. TF – term-frequency, IDF – inverse document frequency) [5].

Для каждого слова рассчитывается вес W так, что слова с высокой частотой употребления в рамках одного документа и с низкой частотой употребления в других документах получают наибольший вес:

$$W_{x,y} = tf_{x,y} \log\left(\frac{N}{df_x}\right), \quad (1)$$

где $tf_{x,y}$ – частота слова x в тексте y ;

df_x – количество текстов, содержащих слово x в выборке;

N – общее количество текстов в выборке.

В качестве обучаемого алгоритма используется логистическая регрессия.

На выходе модель А выдает метку класса, к которому относится представленный текст.

Возможные классы – «Адресат», «Организация-отправитель», «Подписавший» или «Нерелевантно».

Входом модели В являются пространственные признаки: координаты блока, выделенного на этапе распознавания текста.

В качестве обучаемого алгоритма используется градиентный бустинг [6].

Аналогично модели А, на выходе выдается метка прогнозируемого класса.

В качестве входа метамодели, помимо прогнозов модели А и модели В, также используются такие статистические признаки текста, как его длина, количество заглавных букв и использование пунктуации.

Данные характеристики извлекаются из текста с помощью регулярных выражений и встроенных функций языка Python.

В качестве обучаемого алгоритма также используется градиентный бустинг.

Представленный ансамбль моделей обучен на 150 сканах, разбитых на 2500 блоков.

Оценка качества классификатора проводилась на отложенной выборке, состоящей из 50 сканов, разбитых на 1000 блоков.

Для оценки качества работы модели использовались следующие метрики:

– полнота (recall) – доля верно классифицированных блоков относительно всех блоков данного класса.

– точность (precision) – доля верно классифицированных блоков относительно всех блоков, которые модель отнесла к данному классу

– мера F1 – гармоническое среднее между точностью и полнотой.

Результаты работы модели на тестовой выборке приведены в таблице 1.

Таблица 1 – Качество работы модели выделения реквизитов.

Метрика	Значение, %			
	Адресат	Организация-отправитель	Подписавший	Нерелевантно
Полнота	91	81	83	97
Точность	95	82	79	97
Мера F1	93	82	81	97

Такие реквизиты, как номер и дата исходящего документа, а также сроки, указанные в документе, выделяются из текста с помощью регулярных выражений.

Классификация текста

Определение тематики входящего документа также рассматривается как задача классификации. В качестве обучаемого алгоритма используется логистическая регрессия. На вход модели подается векторное представление текста, полученное с помощью алгоритма TF-IDF. Предварительная обработка текста проводится тем же образом, что и при решении задачи выделения реквизитов.

Обучающая выборка состоит из 700 документов, тестовая выборка – из 170 документов. Каждый документ относится к одному из 12 классов. Для оценки качества классификации использовались те же метрики, что и в задаче извлечения реквизитов. Усредненные по всем классам значения метрик, полученные при прогнозировании на тестовой выборке, представлены в таблице 2.

Таблица 2. Качество работы модели по классификации текстов.

Метрика	Значение, %
Полнота	84
Точность	85
Мера F1	84

Выход модели – метка класса документа – используется для определения сотрудника предприятия, ответственного за работу по документам прогнозируемой тематики. Данный сотрудник автоматически назначается исполнителем в системе электронного документооборота.

Машинное обучение как сервис

Для взаимодействия с СЭД разработанные модели машинного обучения разворачиваются в виде сервисов в периметре локальной сети предприятия с помощью Python библиотеки FastAPI [7] и фреймворка контейнеризации Docker [8].

Непрерывная разработка и интеграция кода (Continuous Development/Continuous Integration) осуществляется с помощью системы Jenkins [9].

Полный цикл работы автоматизированной обработки входящих документов подробно представлен на рисунке 3.



Рисунок 3. Цикл работы автоматизированной СЭД

Выводы

Методы машинного обучения являются мощным инструментом для автоматизации рутинных операций при работе с системой электронного документооборота.

В статье рассмотрены разработанные алгоритмы интеллектуального анализа и обработки входящей документации. Описаны способы предобработки данных, используемые инструменты, объемы выборок и полученные в ходе тестирования результаты. Также представлена архитектура взаимодействия имеющейся ИТ инфраструктуры предприятия и сервисов автоматизации СЭД.

С помощью автоматического извлечения реквизитов и классификации текста карточки в СЭД заполняются и направляются ответственному исполнителю без участия оператора.

Результатом подобной автоматизации является уменьшение трудозатрат сотрудников и повышение скорости обработки документации, поступающей в адрес предприятия.

Литература

1. Suzuki S., Abe K. Topological structural analysis of digitized binary images by border following. // Computer vision, graphics, and image processing, 1985. – Т. 30. С. 32-46.
2. Bradski G., Kaehler A. OpenCV // Dr. Dobb's journal of software tools, 2000. – Т. 3.
3. Smith R. An overview of the Tesseract OCR engine // Ninth international conference on document analysis and recognition (ICDAR 2007). IEEE, 2007. – Т. 2. С. 629-633.
4. Воронцов К. В. Математические методы обучения по прецедентам (теория обучения машин) // М.: 2011. – С. 4-5.
5. Jurafsky D., Martin J. H. Speech and Language Processing: An introduction to speech recognition, computational linguistics and natural language processing // Upper Saddle River, NJ: Prentice Hall, 2008. С. 770-771.

6. Friedman J. H. Greedy function approximation: a gradient boosting machine // Annals of statistics, 2001. – С. 1189-1232.
7. FastAPI [Электронный ресурс] – URL:<https://fastapi.tiangolo.com/> (дата обращения: 2.04.2021)
8. Merkel D. Docker: lightweight linux containers for consistent development and deployment // Linux journal, 2014. – Т. 239. С. 2.
9. Jenkins [Электронный ресурс] – URL: <https://www.jenkins.io/> (дата обращения: 2.04.2021)

МОДЕЛИРОВАНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ

УДК 004.42

МОДЕЛЬ ЦИФРОВОЙ ОБРАЗОВАТЕЛЬНОЙ ПЛАТФОРМЫ
И ОБЕСПЕЧЕНИЕ ЕЕ БЕЗОПАСНОСТИDIGITAL EDUCATION PLATFORM MODEL
AND ENSURING ITS SECURITY

Репьюк Н.С.,
Томский государственный университет систем управления
и радиоэлектроники,
г. Томск, Российская Федерация

N. S. Repyuk,
Tomsk state university of control systems
and radioelectronics,
Tomsk, Russian Federation

e-mail: natasha.repyuk1@gmail.com

Аннотация. Сегодня мир массово переходит к цифровизации. Появление разного рода цифровых кроссплатформенных решений позволяет решать ряд задач: образование, подбор и оценка кадров, информатизация и т.п. Цифровая образовательная платформа (ЦОП) – это возможность быстрого обучения для школьников, студентов и сотрудников предприятия. В данной статье приводится структура данной платформы. Архитектура платформы рассмотрена как кроссплатформенное приложение. Необходимость представления модели платформы позволит лучше понять механизм ее работы. Модульная структура данной платформы способствует быстрым коммуникациям, позволяет администрировать процесс, а также собирать и обрабатывать информацию. Цифровая платформа является многопользовательской системой с различными правами доступа пользователей к данным платформы. ЦОП – это взаимодействие администратора, обучающегося и модератора. Для обеспечения безопасности необходимо обеспечить конфиденциальность, доступность и целостность информации данной образовательной платформы. В базе данных хранится персональная информация, которую в соответствии с Федеральным законом №152-ФЗ «О персональных данных» необходимо защищать. Система препятствует несанкционированному доступу к ресурсам платформы. В связи с этим была составлена модель угроз, рассмотрены внешние и внутренние нарушители. Также приведены две основные угрозы автоматизированных систем.

Abstract. Today, the world is massively moving to digitalization. The emergence of various kinds of digital cross-platform solutions allows us to solve a number of tasks: education, recruitment and evaluation of personnel, informatization, etc. A digital educational platform – DEP) is an opportunity for rapid learning for schoolchildren, students and employees of the enterprise. This article describes the structure of this platform. The architecture of the platform is considered as a cross-platform application. The need to present a model of the platform will allow you to better understand the mechanism of its operation. The modular structure of this

platform promotes fast communication, allows you to administer the process, as well as collect and process information. The digital platform is a multi-user system with different user access rights to the platform data. A DEP is an interaction between an administrator, a student, and a moderator. To ensure security, it is necessary to ensure the confidentiality, availability and integrity of the information of this educational platform. The database stores personal information that must be protected in accordance with Federal Law No. 152-FZ "On Personal Data". The system prevents unauthorized access to the platform resources. In this regard, a threat model was compiled, external and internal violators were considered. There are also two main threats of automated systems.

Ключевые слова: цифровая образовательная платформа, модель угроз, безопасность, информация, персональные данные, злоумышленник.

Keywords: digital educational platform, threat model, security, information, personal data, attacker.

В современном мире многие сферы жизнедеятельности человека коснулась автоматизация, исключением не стала и сфера образования. Сейчас большую популярность набирают цифровые образовательные платформы. Они внедряются как в основной процесс обучения школьников, студентов и сотрудников предприятия, так и в иные формы обучения. Можно выделить основные функции ЦОП:

- Сбор, обработка и хранение персональных данных, а также образовательных материалов.
- Моментальный анализ информации, подразумевающий сверку эталонных значений или заданных условий с пользовательскими данными.
- Возможность контролировать действия пользователей на платформе.
- Возможность внедрения криптографических методов защиты информации в ЦОП, таких как цифровая электронная подпись, кэширование данных и т.д.
- Перспектива использования ЦОП на различных устройствах, таких как смартфоны, компьютеры и планшеты.

Цифровая образовательная платформа является автоматизированной системой (АС), которая служит образовательным ресурсом для школьников, студентов и сотрудников предприятия.

Цифровая платформа позволяет наиболее эффективно выстроить образовательный процесс для школьников и студентов, а также обучить новых сотрудников и повысить квалификацию уже работающих на предприятии людей. Создание индивидуальных траекторий развития позволит сделать обучение успешным.

Цифровая образовательная платформа имеет модульную структуру, на базе которой можно создать образовательную систему курсов для решения различных задач. Модульная структура информационной системы имеет следующий вид:

- Комплекс для индивидуальной работы с учебными материалами с использованием различных цифровых решений.
- Онлайн-чат и конференция для дистанционного взаимодействия с преподавателем или экспертом для получения рекомендаций, комментариев и оценок.
- Учебный аналог социальных сетей или мессенджеров.
- Сервис электронной рассылки необходимых учебных материалов.
- Модуль для планирования и формирования курсов.
- Возможность тестирования обучающихся на курсах.
- Модуль для создания онлайн-уроков с возможностью демонстрации загруженных материалов.

- Сервис для загрузки на платформу информации в различных форматах и управления уровнями доступа к данному контенту.
- Модуль для администрирования как отдельных компонентов, так и всей системы в целом.
- Сервис для сбора и анализа статистических данных.

Цифровая образовательная платформа выполняет следующие задачи:

Со стороны неавторизованных пользователей:

1. Просмотр информации о ЦОП;
2. Возможность регистрации.

Со стороны пользователей-обучающихся:

1. Просмотр доступных учебных материалов;
2. Участие в онлайн-чатах и конференциях;
3. Общение с преподавателями (экспертами) и с другими пользователями;
4. Прохождение доступных тестирований;
5. Загрузка и выгрузка необходимых файлов.

Со стороны пользователей-преподавателей:

1. Просмотр доступных учебных материалов;
2. Участие в онлайн-чатах и конференциях;
3. Общение с другими пользователями;
4. Загрузка и выгрузка необходимых учебных материалов;
5. Создание курсов и тестирований в них.

Со стороны администратора:

1. Проверка и одобрение регистрации незарегистрированных пользователей;
2. Разграничение прав доступа пользователей;
3. Управление базой данных платформы;
4. Обеспечение безопасной работы ЦОП.

После описания системы была разработана архитектура цифровой образовательной платформы (рисунок 1). В архитектуре ЦОП в качестве клиента используется веб-браузер на клиентской машине.

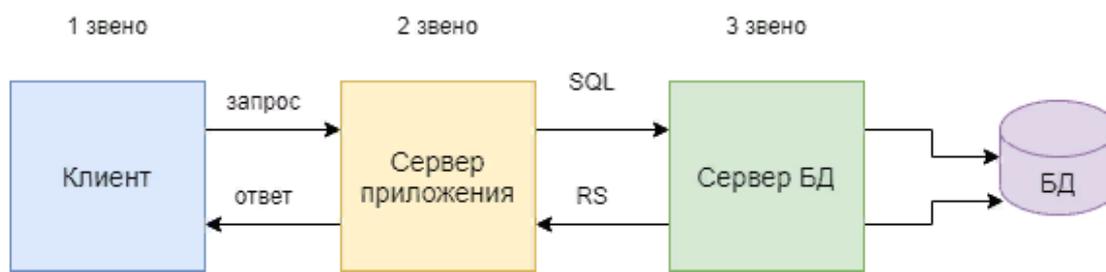


Рисунок 1. Общая архитектура цифровой образовательной платформы

На рисунке 2 изображена модель цифровой образовательной платформы.

В цифровой образовательной платформе присутствует три группы пользователей:

- Администраторы – поддерживают работоспособность системы, управляет разграничением доступа пользователей к модулям платформы, разрабатывают и внедряют в платформу новый функционал.
- Обучающиеся – пользователи системы, которые могут проходить обучение в доступных для них курсах.
- Модератор – чаще всего является преподавателем или экспертом, есть доступ на создание и управление курсов.

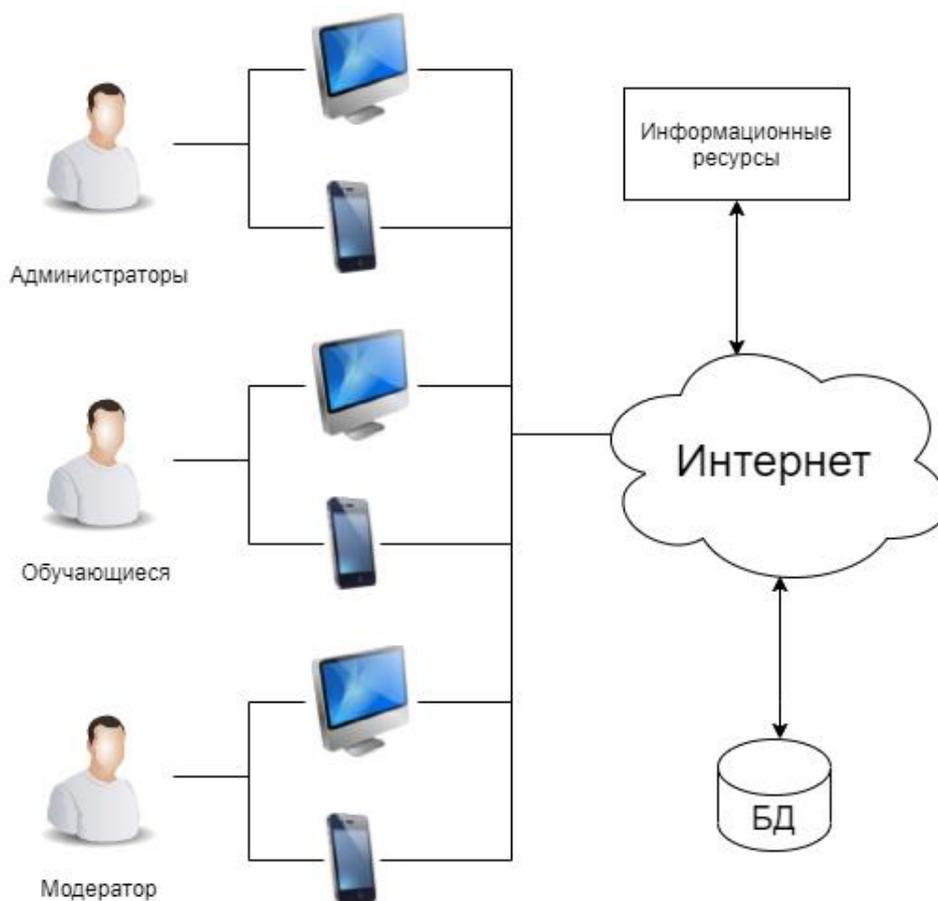


Рисунок 2. Модель цифровой образовательной платформы

Пользователи системы подключаются к платформе через персональный компьютер, планшет или телефон.

К платформе подключена база данных, в которой хранится персональная информация пользователей, а также все материалы курсов. В информационных ресурсах хранятся исходные коды платформы, а также необходимые схемы и модели.

Функционирование цифровой образовательной платформы напрямую зависит от обеспечения безопасности информации. Ее можно определить, как состояние защищенности от всех видов угроз.

Для обеспечения безопасности цифровых платформ необходимо разработать модель угроз. Модель угроз содержит систематизированный перечень угроз безопасности персональных данных (ПДн) при их обработке в информационных системах персональных данных [1].

Необходимо составить модель угроз безопасности персональных данных и вероятного нарушителя для цифровой образовательной платформы, которая является автоматизированной системой.

Хранимая и обрабатываемая в АС информация в соответствии с Федеральным законом Российской Федерации «О персональных данных» [2], относится к информации, составляющей персональные данные.

Цифровая платформа является многопользовательской системой с различными правами доступа пользователей к данным платформы. Разграничение доступа обеспечивается за счет идентификации пользователей. При входе в систему проводится аутентификация пользователей цифровой образовательной платформы. Система препятствует несанкционированному доступу к ресурсам платформы.

Все угрозы безопасности информации АС подразделяются на два класса:

- непреднамеренные угрозы;
- атаки.

Непреднамеренные угрозы не связаны с действиями человека, которые направлены на получение несанкционированного доступа к системе. Несмотря на это, непреднамеренные угрозы могут привести к утечке информации.

Защита от непреднамеренных угроз регламентирована в инструкциях, разработанными и утвержденными на предприятии. Поэтому непреднамеренные угрозы наиболее подробно не рассматриваются в данной статье.

Атаки подготавливаются и осуществляются злоумышленником. Все нарушители делятся на две группы:

- Внешние нарушители – физические лица, не имеющие непосредственного доступа к техническим средствам и ресурсам системы, которые находятся в пределах контролируемой зоны;
- Внутренние нарушители – физические лица, имеющие доступ к контролируемой зоне.

Установим следующие условия функционирования АС:

- Рабочие места разработчиков находятся и функционируют только в контролируемой зоне.
- Серверная часть системы, консолидирующая информационные ресурсы (в том числе ПДн), расположена только в серверных помещениях.
- К ресурсам системы имеют доступ проверенные пользователи, которые авторизовались.

Предполагается, что при выполнении данных условий действия внешнего нарушителя приемлемо нейтрализуются, поэтому далее их действия в данной статье не рассматриваются.

К внутренним нарушителям предположительно относятся:

- администраторы АС;
- технический персонал АС;
- пользователи АС;
- персонал разработчиков АС, который на договорной основе имеет право на техническое обслуживание и модификацию компонентов системы;

В качестве основных категорий знаний нарушителей об АС можно выделить следующие:

- общая информация – информация о назначениях и общих характеристиках платформы;
- эксплуатационная информация – информация, полученная из эксплуатационной документации.

В частности, нарушитель может иметь:

- данные об организации работы, структуре и используемых технических, программных и программно-технических средствах цифровой платформы;
- сведения об информационных ресурсах АС: порядок и правила создания, хранения и передачи информации;
- данные об уязвимостях, включая данные о недеklarированных возможностях технических, программных и программно-технических средств системы;
- данные о реализованных в средствах защиты системы принципах и алгоритмах;
- исходные тексты программного обеспечения АС;
- информацию о способах атак.

Основными объектами атак являются:

- защищаемая информация (в том числе ПДн);
- документация на средства защиты информации и на технические и программные компоненты средств защиты информации;
- каналы связи (внутри контролируемой зоны), не защищенные от несанкционированного доступа к информации организационно-техническими мерами.

Рассмотрим две основные угрозы безопасности информации АС.

Угроза 1.

1. Аннотация угрозы – осуществление несанкционированного доступа к целевой информации, хранимой и обрабатываемой на образовательной платформе.
2. Возможные источники угрозы – пользователи системы.
3. Способ реализации угрозы – осуществление доступа к целевой информации с использованием штатных средств, предоставляемых АС.
4. Используемые уязвимости – недостатки механизмов разграничения доступа к целевой информации.
5. Вид ресурсов, потенциально подверженных угрозе – целевая информация.
6. Нарушаемые характеристики безопасности ресурсов – конфиденциальность, целостность, доступность.
7. Возможные последствия реализации угрозы – несанкционированное ознакомление и разглашение служебной информации и персональных данных пользователей платформы.

Угроза 2.

1. Аннотация угрозы – перехват защищаемой информации в каналах передачи данных с использованием специально разработанных технических средств и программного обеспечения
2. Возможные источники угрозы – пользователи АС; внешние злоумышленники; уполномоченный персонал разработчиков системы, который на договорной основе имеет право на техническое обслуживание и модификацию компонентов АС.
3. Способ реализации угрозы – перехват целевой и технологической информации с использованием специально разработанных технических средств и программного обеспечения, не входящих в состав системы.
4. Используемые уязвимости – недостатки механизмов защиты передаваемой информации, связанные с возможностью ее перехвата из каналов связи и последующего с ней ознакомления.
5. Вид ресурсов, потенциально подверженных угрозе – целевая информация, ПДн пользователей.
6. Нарушаемые характеристики безопасности ресурсов – конфиденциальность, целостность, доступность.
7. Возможные последствия реализации угрозы – несанкционированное ознакомление, разглашение, искажения и удаления персональных данных и целевой информации

Выводы

Исходя из приведенной модели угроз, в АС требуется обеспечить конфиденциальность, целостность и доступность конфиденциальной информации.

Литература

1. «Базовая модель угроз безопасности персональных данных при их обработке в информационных системах персональных данных» (Выписка) (утв. ФСТЭК РФ 15.02.2008)
2. Федеральный закон «О персональных данных» от 27.07.2006 N 152-ФЗ.

UDC 004.932

SUPERPIXELS IN DIGITAL IMAGE PROCESSING

СУПЕРПИКСЕЛИ В ЦИФРОВОЙ ОБРАБОТКЕ ИЗОБРАЖЕНИЙ

M.V. Kharinov,
St. Petersburg Federal Research Center
of the Russian Academy of Sciences,
Saint Petersburg, Russian Federation

Харинов М.В.,
Санкт-Петербургский Федеральный исследовательский центр
Российской академии наук,
Санкт-Петербург, Российская Федерация

e-mail: khar@iias.spb.su

Abstract. The paper relates to the field of artificial intelligence. More specifically, image processing is discussed.

This paper offers the simplest understanding of superpixels as maximal pixel clusters providing error-free obtaining of a limited sequence of accessible *optimal* piecewise constant image approximations with the minimum achievable approximation error (total squared error) E for each number of pixel clusters.

The superpixel understanding and definition contrasts with standard approaches to calculating superpixels. The superpixel hierarchy is generated by the intersection of optimal image partitions into an increasing number of clusters g . It is irregular and differs markedly from the non-hierarchical sequence of optimal approximations, since described by a non-convex dependence of the approximation error E on the cluster number g . Therefore, the task of hierarchical superpixel clustering is posed and solved, which is briefly described in the paper.

The proposed definition of superpixels makes it possible to adequately detect objects in images of various contents. Computational complexity is reduced by processing the image piece by piece.

Аннотация. Статья относится к области искусственного интеллекта. Более конкретно, обсуждается обработка изображений.

В этой статье предлагается простейшее понимание суперпикселей как максимальных кластеров пикселей, обеспечивающих безошибочное получение ограниченной последовательности доступных оптимальных кусочно-постоянных аппроксимаций изображений с минимальной достижимой ошибкой аппроксимации (общая квадратическая ошибка). E для каждого количества кластеров пикселей.

Понимание и определение суперпикселей контрастирует со стандартными подходами к вычислению суперпикселей. Иерархия суперпикселей создается пересечением оптимальных разделов изображений на все большее число кластеров g . Она нерегулярна и заметно отличается от иерархической последовательности оптимальных приближений, поскольку описывается невыпуклой зависимостью ошибки аппроксимации E от числа кластеров g . Поэтому ставится и решается задача иерархической суперпиксельной кластеризации, которая кратко описана в статье.

Предлагаемое определение суперпикселей позволяет адекватно обнаруживать объекты на изображениях различного содержания. Вычислительная сложность снижается за счет обработки изображения по частям.

Keywords: pixel clusters, piecewise constant image approximation, total squared error, convex dependence, definition of superpixels, system of clustering methods.

Ключевые слова: пиксельные кластеры, кусочно-постоянная аппроксимация изображения, общая квадратическая ошибка, выпуклая зависимость, определение суперпикселей, система методов кластеризации.

Introduction

Low-level computer vision is rather a highly desirable project than a complete engineering science about image elements and how to classify these elements as reflections of scene objects. Although there are no generally accepted definitions for image elements, the term has already been coined for them. Wanted image elements are called “superpixels”.

The subject of superpixels becomes especially relevant due to the increase in image resolution and the need to reduce the computational complexity of image processing by replacing operations with pixels by operations with superpixels. Usually superpixels are conceptualized as elements of objects or enlarged pixels, in particular, as image segments matching to the boundaries between objects. Grouping pixels is a very good method to reduce computational complexity at the initial stage of image processing. But this does not neglect the problem of justifying how to use the enlarged pixels in the best way.

Methods for calculating of superpixels are empirically developed, tested, debugged, and compared with each other using scene examples [1]. However, instead of transparent definitions of clearly existing image elements, various heuristic algorithms for obtaining of superpixels are proposed, including those with unproven convergence. Attempts to generalize algorithmic tools for calculating superpixels are carried out by means of a standard search for a unified proximity of pixels, which, on the contrary, most likely depends on the image content. The methods to reduce the standard deviation σ or, equivalently, approximation errors $E \sim \sigma^2$, most often K-means method, are not consciously aimed at real E minimization and obtaining of optimal or at least suboptimal image approximations. Therefore, a transparent definition of superpixels by means of optimal piecewise constant approximations have not yet been considered, as is done in this paper.

Superpixels are defined as the maximal sets of pixels that provide to construct the initial $1, 2, \dots, g_1$ series of optimal piecewise constant image approximations in bottom up merging technique.

The idea of superpixels and approximating of an image by a hierarchy of superpixels is illustrated in Figure 1.

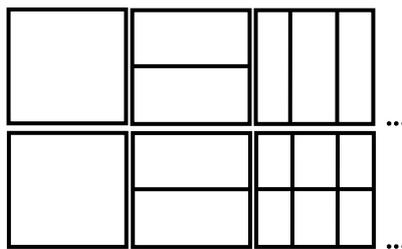


Figure 1. The sequence of optimal image partitions (above) and the hierarchy of superpixel partitions (bottom)

If the optimal image approximations with 1, 2, 3, ... g_1 pixel clusters are determined by the sequence of partitions shown in the upper row of pictures, then the hierarchy of image approximations by superpixels is determined by the hierarchical sequence of partitions into 1, 2, 3,... s superpixels shown in the bottom row of pictures, where $g_1 \leq s \leq (g_1)!$ and the exclamation mark “!” indicates a factorial.

For clarity, the superpixels in Figure 1 are shown as image segments, which is not considered mandatory. Nevertheless, Figure 1 clearly explains that superpixels preserve the boundaries between pixel clusters that are disappeared in the current optimal approximation of the image.

The effect of the disappearance of boundaries between pixel clusters is caused by non-hierarchical optimal approximations. It just means that, starting from the optimal approximation, the computer does not “see” the sharp boundaries between the objects. Therefore, to detect objects, it is preferable to use superpixel clustering with a value of E exceeding the minimum possible one for a given number of superpixels.

Thus, in our definition, superpixels simply express the necessary condition for constructing the series of optimal piecewise constant approximations of the image. On the other hand, superpixels by themselves provide an accurate reproduction of the series of optimal approximations. This independently justifies the use of superpixels to simulate objects in the image.

If the image is described by a hierarchy of optimal approximations, then its superpixel approximations coincide with the optimal ones. Otherwise, the hierarchy of superpixels with some accuracy approaches to the sequence of optimal approximations from above. In order to decide whether it is worth moving from a simple optimization task to a more complex one, an experiment is appropriate. Such an experiment is provided by a simple upgrade of the software developed to create the “Optimal Segmentation Data Base” [2, 3].

Example of Superpixels

Figure 2 presents the optimal and hierarchical superpixel approximations for standard “Lena” image. In the upper row the optimal image approximations with the number of tones from one to six are demonstrated. The corresponding superpixel approximations of the image are placed side by side in the lower column. Pairs of identical approximations in the first and the second column are shown.



Figure 2. The sequence of optimal and superpixel image approximations

Under careful examination, it is not difficult to notice the differences in the approximations being compared, which are effortlessly manifested numerically.

In the columns of Table the data for the sixty first optimal and superpixel approximations are listed. The initial columns are marked with a double bar. They indicate the number of clusters g , i.e. the number of tones in optimal approximations. The following pairs of columns indicate the standard deviation σ and the number of superpixels s generated by intersection of g optimal approximations.

Table – The numbers of superpixels s for reproducing of 1-60 optimal image approximations

g	σ	s									
1	55,8832	1	16	3,92460	46	31	1,87573	79	46	1,29542	106
2	30,6456	2	17	3,70326	50	32	1,79485	79	47	1,26794	107
3	21,2174	4	18	3,50441	55	33	1,72854	80	48	1,23991	109
4	14,9645	7	19	3,32383	60	34	1,68547	82	49	1,21323	110
5	11,6976	11	20	3,15658	63	35	1,64729	84	50	1,18618	111
6	10,0398	16	21	3,0126	66	36	1,61313	87	51	1,16091	112
7	8,46072	18	22	2,87305	67	37	1,57894	89	52	1,13576	112
8	7,51121	24	23	2,73844	67	38	1,54589	90	53	1,11091	114
9	6,81359	27	24	2,60572	68	39	1,51218	92	54	1,08554	115
10	6,14397	30	25	2,49076	70	40	1,4785	94	55	1,06012	118
11	5,57864	33	26	2,37391	70	41	1,44586	96	56	1,03435	119
12	5,11403	36	27	2,25584	74	42	1,414	97	57	1,00904	120
13	4,75689	39	28	2,15389	76	43	1,38327	98	58	0,9852	123
14	4,42306	41	29	2,05655	78	44	1,3534	101	59	0,96397	124
15	4,17825	43	30	1,95565	79	45	1,32287	104	60	0,94304	126

As follows from Table, the hierarchy of superpixels is irregular, in particular, not binary. The number of superpixels s does not grow strictly monotonously with an increase in the current number of clusters g .

Nevertheless, starting from a dozen clusters in the current optimal image approximation, the number of superpixels s is several times greater than the number of clusters g in the current optimal approximation. At the same time, $s - g$ optimal image approximations cannot be accurately obtained by merging of s superpixels.

Figure 3 describes the total sequence of 216 optimal approximations calculated for integer pixel values without an initial pixel enlargement.

The left graph in Figure 3 shows an increase in the number of clusters $s(g)$ in the superpixel image approximation with an increase in the number of clusters g in the current

optimal approximation and a concomitant increment of the number g of the reproducible optimal approximations. The dashed line on the graph corresponds to the case of a hierarchy of optimal approximations, when the superpixel approximations coincide with the optimal ones. A significant deviation of the curve $s(g)$ from the dashed straight line indicates for the current optimal approximation that there is a lack of data about sharp objects in optimal approximations with fewer clusters (Figure 3).

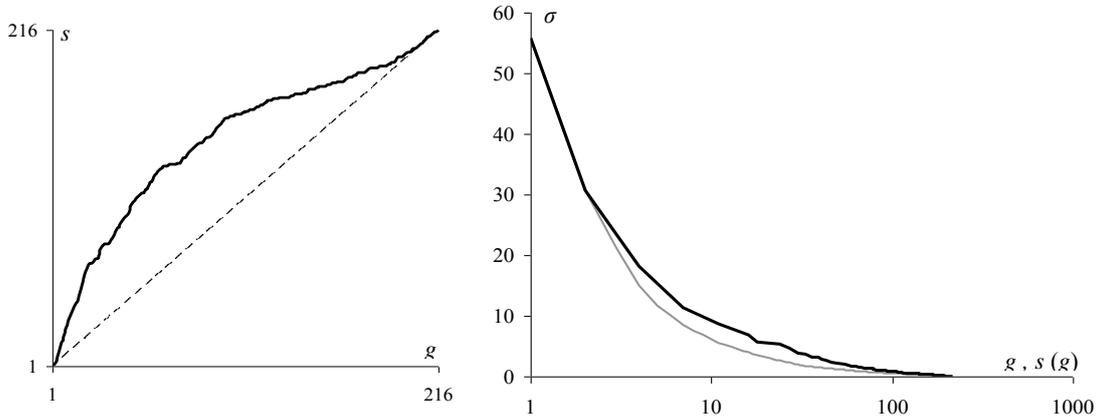


Figure 3. The numbers of superpixels s depending on the number of clusters g (left).

The curves $\sigma(g)$ for optimal and $\sigma(s)$ for hierarchical superpixel image approximations (right)

The right graph in Figure 3 shows the dependence of the standard deviation σ on the number of clusters counted along the abscissa on a logarithmic scale. It expresses the approaching of a non-hierarchical sequence of optimal image approximations by the hierarchy of superpixel approximations. The dependence of the standard deviation $\sigma(s)$ on the number of clusters $s(g)$ in the superpixel approximation is shown by a black curve, and the dependence $\sigma(g)$ on the number of clusters g in the current optimal image approximation is shown in gray.

It should be noted that replacing of s superpixels with the same number of pixel clusters of optimal approximations hinders the reproduction of the source series of optimal approximations, although it reduces the approximation error E .

Hierarchical Superpixel Clustering

The hierarchy of superpixels is much more convenient for computing objects than a non-hierarchical set of clusters of optimal approximations. At the same time, the sequence of optimal approximations has a remarkable advantage, expressed in the fact that it is described by the convex dependence of the approximation errors E_g on the cluster numbers g :

$$E_g \leq \frac{E_{g-1} + E_{g+1}}{2}, \quad g = 2, 3, \dots,$$

which implies proper coordinate transformation for the bottom graph in Figure 3.

To get both advantages of optimal and hierarchical image approximations, it is quite attractive to construct a hierarchy of approximations, which is described by the convex

dependence of the approximation error E on the number of clusters g . To do so, it seems reasonable to start with a sufficient number of superpixels and construct a target hierarchy that satisfies the convexity condition. In this case, the corresponding resulted curve will pass between the black and gray curves in the bottom graph Figure 3.

The idea of superpixel clustering is illustrated in Figure4.

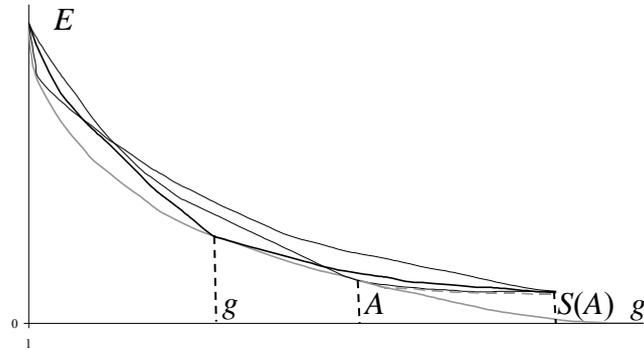


Figure 4. Hierarchical superpixel clustering

The bottom gray line describes the sequence of optimal approximations. To obtain the target superpixel hierarchy only a single approximation of S superpixels is exploited providing the increase of the processing speed and the accurate reproducing of the certain number A of accessible optimal image approximations. This number A is several times smaller than the number of superpixels S . For a more accurate representation of the hierarchical image, a binary hierarchy of approximations is constructed, which is described in Figure 4 by thick black line. It is described by the convex dependence of E on g and contains at least one of the non-trivial optimal image approximations with the number of clusters $g_0 > 1, E(g_0) \neq 0$. The cluster number $2 \leq g_0 \leq A$ from the range of accessible optimal approximations is interpreted as the number of objects in the image that the low-level computer vision software focuses on.

S and g_0 are treated as the tuning parameters that optimize the processing speed and the choice of an appropriate cluster hierarchy to support further adequate object detection .

Conclusion

In our understanding, an image consists of superpixels and has ambiguous content, because it reflects objects of different sizes, as well as various hierarchies of objects. Such a concept is formalized by the statement of the combined approximation and optimization problem, which consists in the hierarchical approaching of the series of optimal piecewise constant image approximations that do not constitute a hierarchy.

The novelty of this paper consists in the transparent definition of superpixels, explained by the simplest Figure 1 and substantiated on a specific example in Figure 2, Figure 3 and Table. This definition logically completes the image superpixel clustering model illustrated in Figure 4, where target superpixel hierarchies designed to fit object hierarchies are described by solid black lines.

From a theoretical point of view, the existence of superpixels and target superpixel hierarchies is crucial. The existence of superpixels defined in this paper is evident. The existence of target superpixel hierarchies is ensured by a system of three classical clustering methods, but only after their certain modernization [4].

External signs of the prescribed modernization of pixel clustering methods are two parameters (Figure 4). Namely, the parameter for the number of objects g_0 , which focuses the software to generate a suitable hierarchy of clusters. Another parameter is the number A of active optimal approximations or the superpixel number $S(A)$, intended to speed up computer calculations.

The internal self-consistency of the hierarchical superpixel clustering model and the operability of the software, as well as a comparison of the model implementation versions, are verified by a simple comparison of the achieved approximation errors for image approximations containing from 2 to A clusters.

This work was performed within the framework of the budgetary theme 0060-2019-0011 (Fundamentals and technologies of big data for sociocyberphysical systems).

References

1. J.E. Vargas-Muñoz, A.S. Chowdhury, E.B. Alexandre, F.L. Galvão, P.A.V. Miranda, A.X. Falcão An iterative spanning forest framework for superpixel segmentation. *IEEE Transactions on Image Processing* 28(7), 3477–3489 (2019)
2. M.V. Kharinov Image Segmentation Method by Merging and Correction of Sets of Pixels. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications* 23(3), 393-401 (2013)
3. D.S. Ostapov Binarization of “background-object” by using pretreatment of an image and fuzzy K-means algorithm. *Bulletin of the Astrakhan State Technical University. Series: Management, Computing and Informatics* No.3, 32-39 (2016)
4. M.V. Kharinov, A.N. Buslavsky Object Detection in Color Image. In: *Proc. of the 14th International Conference on Pattern Recognition and Information Processing (PRIP'2019)*, Minsk, 21-23 May 2019, pp. 43-47. doi:10.13140/RG.2.2.28493.28640

СЕТИ И ТЕЛЕКОММУНИКАЦИИ

УДК 004.629

**ОЦЕНКА УРОВНЯ ЭЛЕКТРОМАГНИТНОГО ФОНА,
СОЗДАВАЕМОГО БАЗОВЫМИ СТАНЦИЯМИ****EVALUATION OF LEVEL OF ELECTROMAGNETIC BACKGROUND
GENERATED BY BASE STATIONS**

Жмакина И.Д., Стаценко Л.Г.,
ФГАОУ ВО «Дальневосточный федеральный университет»,
г. Владивосток, Российская Федерация

I.D. Zhmakina, L.G. Statsenko,
Far Eastern Federal University,
Vladivostok, Russian Federation

e-mail: chadiuk.id@gmail.com

Аннотация. Впервые в 1992 году в центральной печати появилось понятие «электромагнитная экология», в настоящее время термин прочно вошел в лексикон специалистов. Интенсивное использование электромагнитной энергии в современном информационном обществе привело к тому, что сформировался новый значимый фактор загрязнения окружающей среды – электромагнитные поля (ЭМП) техногенного происхождения. В связи с постоянным развитием телекоммуникационных технологий многочисленные радиовещательные и телевизионные станции, системы спутниковой и сотовой связи оказались размещенными на территориях городов и оказывают непосредственное воздействие на здоровье населения. Близость источников ЭМП к жилым районам может негативно воздействовать на здоровье человека и формировать электромагнитную нагрузку на население. Традиционные методы планирования и реализации радиосетей сотовой связи способны привести к увеличению интенсивности электромагнитного фона в местах с высокой плотностью населения. Актуальность приобретает разработка методик практической оценки интенсивности электромагнитного поля, основанных на использовании доступных исходных данных и обладающих высокой предсказательной способностью. В статье рассмотрены методы, позволяющие исследовать электромагнитную обстановку вблизи жилой застройки. Сделан вывод о необходимости внесения зон ограничения застройки в проект градостроительного плана.

Abstract. For the first time in 1992, the concept of “electromagnetic ecology” appeared in the central press; currently, the term has firmly entered the lexicon of specialists. The intensive use of electromagnetic energy in the modern information society has led to the formation of a new significant environmental pollution factor – electromagnetic fields (EMF) of man-made origin. Due to the constant development of telecommunications technologies, numerous broadcasting and television stations, satellite and cellular communication systems have been in cities and have a direct impact on the health of the population. The proximity of EMF sources to residential areas can negatively affect human health and form an electromagnetic load on the population. Traditional methods of planning and implementing cellular radio networks can lead to an increase in the intensity of electromagnetic background

in places with high population density. The development of methods for practical estimation of the intensity of the electromagnetic field, based on the use of available initial data and having high predictive ability, becomes relevant. The article considers methods that allow you to investigate the electromagnetic situation near residential buildings. It was concluded that it is necessary to introduce zones of restriction of development into the draft urban planning plan.

Ключевые слова: электромагнитная экология, электромагнитное поле, электромагнитное излучение, электромагнитное загрязнение, базовые станции сотовой связи.

Keywords: electromagnetic ecology, electromagnetic field, electromagnetic radiation, electromagnetic contamination, cellular base stations.

Основными источниками электромагнитных излучений радиочастотного диапазона в окружающую среду являются передающие радиотехнические объекты (ПРТО) радиосвязи, радиовещания, телевидения, радиолокации.

Близость источников электромагнитного поля (ЭМП) к жилым районам может негативно воздействовать на здоровье человека и формировать электромагнитную нагрузку на население.

В работе [6] проводилось исследование с целью определения воздействия ЭМП в различных средах (рисунок 1).

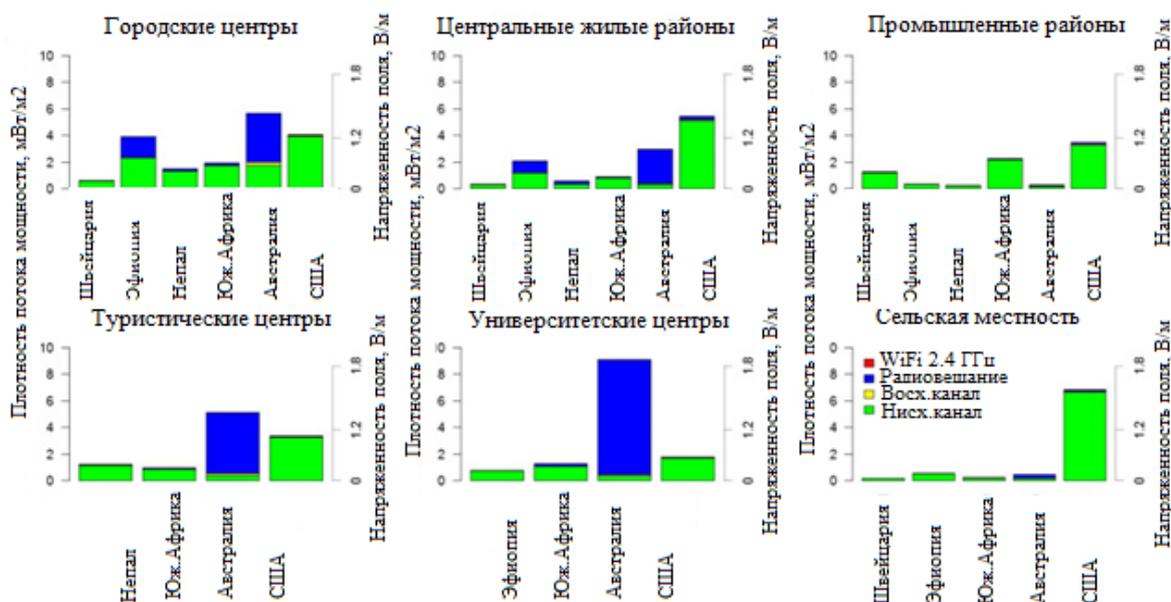


Рисунок 1. Средние уровни воздействия ЭМП для шести сред

Наибольший вклад в электромагнитный фон вносил нисходящий канал. Исследование также демонстрирует, что воздействие ЭМП возрастает в городах.

Жители мегаполисов используют смартфоны для доступа к службам, требующим высокой скорости передачи данных, и ожидают, что их соединение будет поддерживаться везде.

Телекоммуникационные компании отреагировали на это увеличением количества базовых станций.

Таким образом, по мере увеличения числа базовых станций все больше людей живут в непосредственной близости от вышки мобильной связи, чем когда-либо прежде.

Чтобы оценить потенциальные риски для здоровья, необходима надежная оценка воздействия. Геопространственное моделирование воздействия – подход к количественной оценке воздействия РЧ-ЭМП окружающей среды для экологических исследований.

Зарубежные исследователи [7] смоделировали ЭМП для города Амстердама, используя программный комплекс «NISMap».

На рисунке 2 показано распределение напряженности поля по высоте.

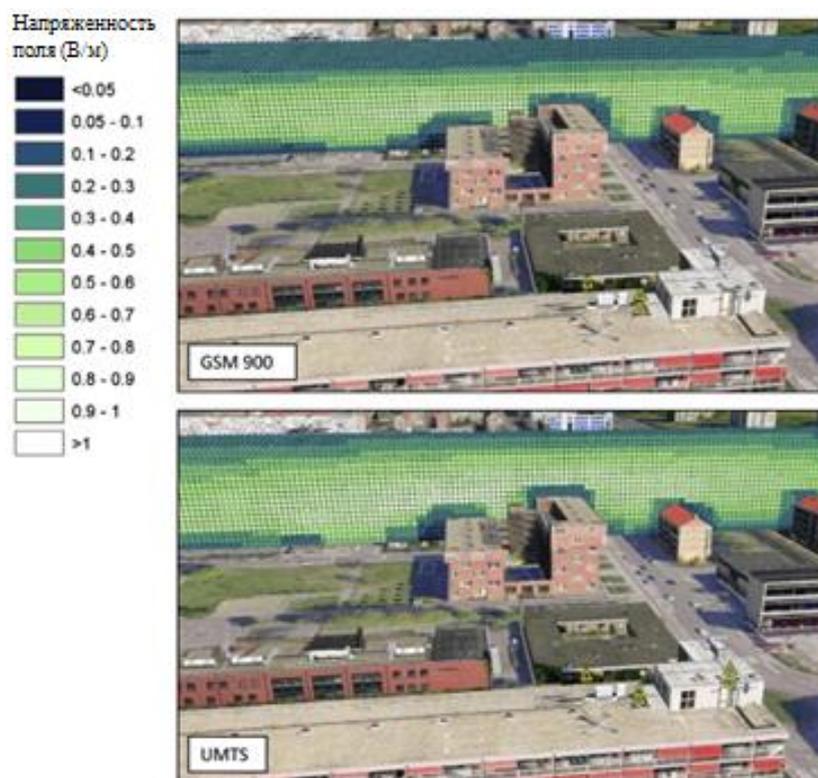


Рисунок 2. 3D-профиль напряженности электрического поля GSM и UMTS (В/м)

В качестве средства расчетного прогнозирования электромагнитного загрязнения на территории Российской Федерации используется Программный комплекс анализа электромагнитной обстановки (ПК АЭМО).

Вопрос электромагнитного загрязнения особенно остро стоит в крупных развивающихся городах.

Авторами [1] был проведен ряд исследований в городе Красноярске.

Как утверждают авторы, наибольший вклад в электромагнитную нагрузку вносит именно сотовая связь – более 80%.

В последующих работах была произведена оценка состояния электромагнитного загрязнения на примере красноярского мегаполиса.

Результат исследования электромагнитной обстановки на территории Красноярска показывает, что плотность потока электромагнитной энергии распределена неравномерно.

Объектом исследования в Перми [2] являлась территория крупного промышленного центра (рисунок 3).

Источники на территории города расположены неравномерно. Наибольшее число источников ПРТО расположено в центральной части города. Практически все источники размещены непосредственно в жилых зонах или максимально приближены к ним.

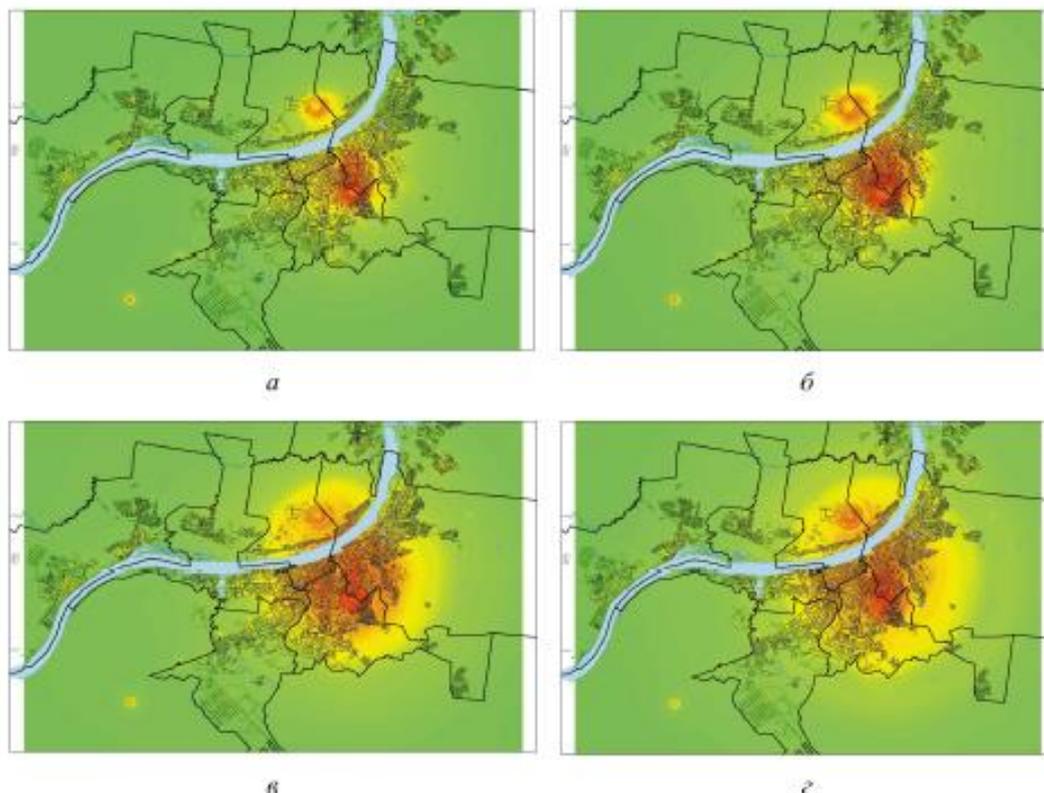


Рисунок 3. Уровень электромагнитного поля на территории Перми на разных высотах от поверхности земли:
а) 3 метра, б) 6 метров, в) 18 метров, г) 30 метров

Инструментальными исследованиями был выявлен ряд превышений критерия безопасности.

Нарушения гигиенических нормативов свидетельствуют об актуальности системного мониторинга уровней ЭМП, оценки рисков для здоровья жителей и обоснования последующих решений по минимизации рисков.

Электромагнитное воздействие в условиях городской застройки достаточно сложное явление, формирующееся за счет различных физических процессов (отражение, дифракция и других), а также когерентных особенностей объектов инфраструктуры, являющихся источниками ЭМП.

Совокупность экспериментальных результатов свидетельствует о высокой пространственной неоднородности электромагнитного загрязнения в условиях урбанизированных сред и необходимости постоянного использования высокочувствительной аппаратуры и использования ГИС-технологий для объективной оценки такой неоднородности. Этому мнению придерживаются авторы [3], которые проводили оценку электромагнитного загрязнения в городе Владивостоке.

Город обладает сложным рельефом, выбор мест размещения ПРТО ограничен, в связи с этим в большинстве случаев размещение базовых станций происходит на малоэтажных нежилых зданиях, окруженных высотными жилыми домами.

В работе [4] выделены факторы опасности передающих радиотехнических объектов по степени риска причинения вреда здоровью (таблица).

Таблица – Факторы риска здоровью населения

Факторы потенциальной опасности ПРТО	Характеристика потенциального риска здоровью населения
Низкая высота подвеса антенн (менее 5 м над землей)	Чрезвычайно высокий
Размещение ПРТО на крышах жилых зданий, образовательных и детских учреждений	Высокий
Большое количество узкополосных передатчиков на одном объекте (более 4)	Высокий
Ориентация жилых, общественных и промышленных зданий наибольшей остекленной поверхностью к источнику ЭМП при расстоянии менее 40 м	Высокий
Отсутствие экранирования и ограждающих конструкций	Высокий
Расположение ПРТО в густонаселенных районах крупных городов и мегаполисов	Высокий

По мнению авторов, факторы, снижающие риск негативного воздействия на здоровье населения, следующие:

- наличие ограждающих конструкций и материалов с радио-экранирующими свойствами;
- использование широкополосных антенн, позволяющих уменьшить количество источников ЭМП и работающих в разных частотных диапазонах;
- размещение антенн ПРТО на отдельно стоящих опорах, мачтах, башнях в малонаселенной местности.

На сегодняшний день представлены результаты разработки композитов, наполненных поглотителями электромагнитных излучений. Исследования этих материалов доказали их способность обеспечить безопасность человека от источников электромагнитных излучений. Радиозащитные материалы снижают уровень электромагнитного поля до уровней, соответствующих санитарным нормам. К примеру, композит из поликарбоната имеет поглощение электромагнитного поля в 25-30 дБ при толщине пластины 3,5 мм, то есть происходит снижение уровня ЭМП в 800-1000 раз [5].

Выводы

Точечное строительство в городах является дополнительным источником переотражения радиосигналов, поэтому на этапе разработки градостроительного плана рекомендуется рассчитывать и включать зоны ограничения застройки для объектов связи.

Литература

1. Жуль Е.Г., Моргулис И.И., Кочемарова Е.В. Формирование электромагнитной нагрузки в условиях городской среды // Вестник КрасГау. 2008. №5.
2. Май И.В., Балашов С.Ю., Вековшина С.А. К оценке уровня электромагнитного поля (300 ГГц-300 МГц) в крупном промышленном центре на базе 3D-моделирования и инструментальных измерений // Анализ риска здоровью. 2017. №3.
3. Стаценко Л.Г., Агеева А.А. Электромагнитная обстановка при формировании городской застройки // Вестник Инженерной школы ДВФУ. 2018. № 4(37).

4. Луценко Л.А., Тулакин А.В., Егорова А.М., Микаилова О.М., Гвоздева Л.Л., Чигряй Е.К. Риск-ориентированная модель контроля уровней ЭМП базовых станций сотовой связи // Гигиена и санитария. 2016. №95(11).

5. Гульбин В.Н., Бибииков С.Б. Композиционные материалы для электромагнитной безопасности // Актуальные вопросы радиобиологии и гигиены неионизирующих излучений: сборник докладов всероссийской научной конференции. 2019. С. 171-173.

6. Comparison of radiofrequency electromagnetic field exposure levels in different everyday microenvironments in an international context / Sanjay Sagar, Seid M. Adem // Environment International. – Volume 114. 2018. P. 297-306.

7. Geospatial modelling of electromagnetic fields from mobile phone base stations / J. Beekhuizen, R. Vermeulen, H. Kromhout, A. Bürgi, A. Huss // Science of The Total Environment. – Volumes 445-446. 2013. P. 202-209.

СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ И ГИС-ТЕХНОЛОГИИ

УДК 004.896:519.237.8

АЛГОРИТМИЗАЦИЯ СИНТЕЗА СТРУКТУРЫ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА В РАМКАХ САПР

ALGORITHMIZATION OF SYNTHESIS OF TECHNOLOGICAL PROCESS STRUCTURE WITHIN CAM SYSTEM

Анфёров М.А.,
ФГБОУ ВО «МИРЭА – Российский технологический университет»,
г. Москва, Российская Федерация

M.A. Anfyorov,
FSBEI HE “MIREA – Russian Technological University”,
Moscow, Russian Federation

e-mail: anfyorov@inbox.ru

Аннотация. Отмечена актуальность повышения эффективности систем автоматизированного проектирования технологических процессов, используемых в рамках жизненного цикла наукоемких изделий. Это определяется необходимостью повышения конкурентоспособности этих изделий.

Производительность современных САПР ТП во многом определяется уровнем автоматизации получаемых проектных решений. В этой связи значимым является разработка моделей и методов формализованного синтеза структуры технологического процесса.

Ввиду сложности данной задачи, отмечена необходимость использования интеллектуальных информационных технологий, позволяющих в условиях информационной неопределенности эффективно находить требуемые проектные решения.

В статье представлены новые положения метода синтеза структуры наукоемких технологических процессов на уровне маршрутного описания. Показано использование информационной технологии кластерного анализа для распределения технологических переходов, принадлежащих типовым проектным решениям, между технологическими операциями. В качестве таких типовых решений используются технологические маршруты, содержащие вышеназванные переходы, определяемые набором атрибутов: конструктивным видом поверхности, параметрами ее качества и эксплуатационных свойств в деталях машин.

В задаче имеет место нечеткость пространства рассматриваемых решений ввиду используемых признаков, содержащих альтернативные значения. Это приводит к нечеткости значения расстояния между объектами кластеризации. Для такого случая используется эволюционный способ кластеризации, предусматривающий представление технологического перехода несколькими объектами кластеризации по числу сочетаний дискретных значений признаков. Реализация данного способа кластеризации обеспечивается использованием генетического алгоритма.

Abstract. The importance of increasing the efficiency of CAM systems used within the life cycle of knowledge-intensive products was noted. This is determined by the need to increase the competitiveness of these products.

The performance of modern CAM systems is largely determined by the level of automation of the resulting design solutions. In this regard, the development of models and methods for the formalized synthesis of the process structure is significant.

In view of the complexity of this task, the need to use intelligent information technologies to effectively find the required design solutions in conditions of information uncertainty was noted.

The article presents new provisions of the method of synthesis of structure of knowledge-intensive technological processes at the level of route description. The use of cluster analysis information technology for the distribution of technological transitions belonging to typical design solutions between technological operations is shown. As such typical solutions, technological routes containing the above transitions are used. They are defined by a set of attributes: the design view of the surface, its quality parameters and operational properties in machine parts.

In the task, there is an illegibility in the space of the solutions in question due to the characteristics used that contain alternative values. This causes the distance between clustering objects to be illegibility. For such a case, an evolutionary clustering method is used.

Ключевые слова: синтез структуры, технологический процесс, кластеризация, САПР ТП, нечеткое множество.

Keywords: structure synthesis, technological process, clustering, CAM systems, fuzzy set.

Практическая значимость результатов выполненных исследований определяется актуальностью повышения эффективности систем автоматизированного проектирования технологических процессов (САПР ТП), используемых в рамках жизненного цикла (ЖЦ) наукоемких изделий. Информационная поддержка процессов ЖЦ в рамках CALS-стратегии способствует снятию противоречия между основными факторами, влияющими на конкурентоспособность данных изделий – повышение их качества (эксплуатационных характеристик) и снижение затрат на реализацию этих процессов.

Системный подход к проектированию технологических процессов (ТП) средствами компьютерных информационных технологий, предполагает моделирование структурных и функциональных свойств проектируемых ТП как основу формализации (алгоритмизации) процесса проектирования. При этом синтез структуры ТП является доминирующим, так как технологический процесс с неудачной структурой уже не может функционировать эффективно. В этой связи формирование структуры в рамках САПР ТП необходимо осуществлять с использованием оптимального выбора проектных решений.

С другой стороны, пространство допустимых решений формируемой структуры чрезвычайно велико – порядка 10^{30} [1]. Данная многовариантность вызвана не только большим разнообразием структурного исполнения элементов ТП на различных иерархических уровнях (см. рисунок 1) – выбором технологического оборудования, методов обработки поверхностей и др., но и порядком следования этих элементов во времени (технологических операций, переходов и др.). Это значительно затрудняет формализацию синтеза структуры ТП, не говоря уже про оптимизацию.

Исторически еще на заре создания первых в нашей стране САПР ТП формировались научные основы формализации синтеза структуры ТП. При этом в качестве типовых проектных решений использовались элементы структуры ТП различных иерархических уровней (см, рисунок 1).

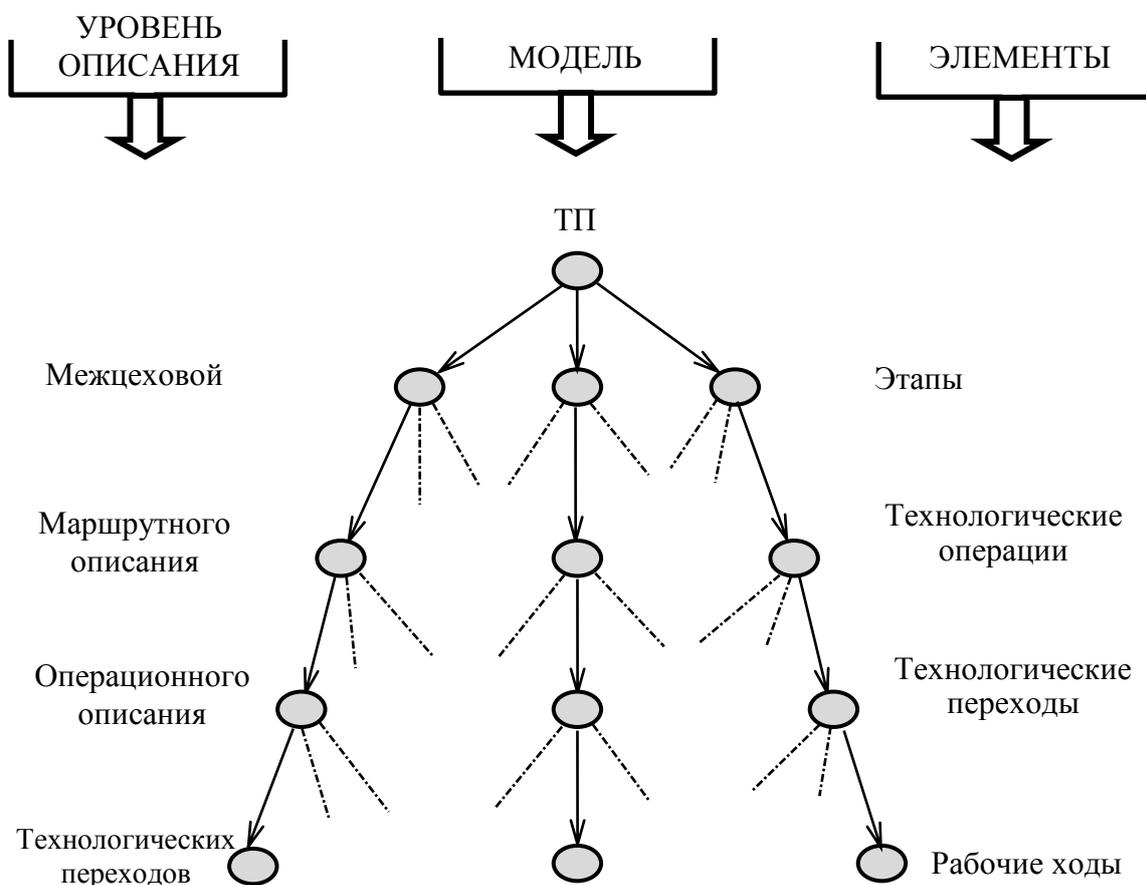


Рисунок 1. Структура технологического процесса

При нисходящем проектировании за основу выбирался типовой технологический процесс [2], при восходящем проектировании – маршруты обработки отдельных поверхностей как совокупность выполняемых технологических переходов [3]. Позднее появился модульный подход к автоматизированной разработке технологических процессов, предполагающий использование комплексных моделей, отображающих модуль (элемент) технологического процесса, привязанный к конкретному конструктивному элементу объекта производства [4].

Современные САПР ТП [5] используют те же методы синтеза структуры проектируемых технологических процессов. Однако это происходит в режиме редактирования маршрутно-операционного описания технологического процесса. Типовые технологические решения, размещаемые в базе данных, используются технологом исходя из его личного опыта. Правда в ряде САПР ТП (например, САПР ТП «Вертикаль» компании «АСКОН») имеется возможность хранения в базе знаний ранее полученных проектных решений на том или ином уровне описания технологического процесса, которые могут использоваться при разработке технологического процесса на аналогичную деталь.

Использование формализованных процедур синтеза готовых структурных технологических решений на уровне маршрутного описания (см. рисунок 1) несомненно повысит производительность работы технолога с САПР ТП.

При этом особенно продуктивным является использование интеллектуальных информационных технологий, частично замещающих интеллект технолога в проектном процессе.

Безаналоговое восходящее проектирование технологических процессов всегда было подвержено затруднениям в части его алгоритмизации.

При этом если создание базы знаний, содержащей типовые решения в виде технологических маршрутов обработки отдельных поверхностей (ТПП), состоящих из технологических переходов, не вызывает затруднений ввиду большого практического опыта по формированию требуемых эксплуатационных свойств поверхностей детали, то формализация синтеза технологических операций из этих переходов (см. рисунок 2) затруднена ввиду вышеназванной многовариантности получаемых решений и необходимости их оценки технологом.

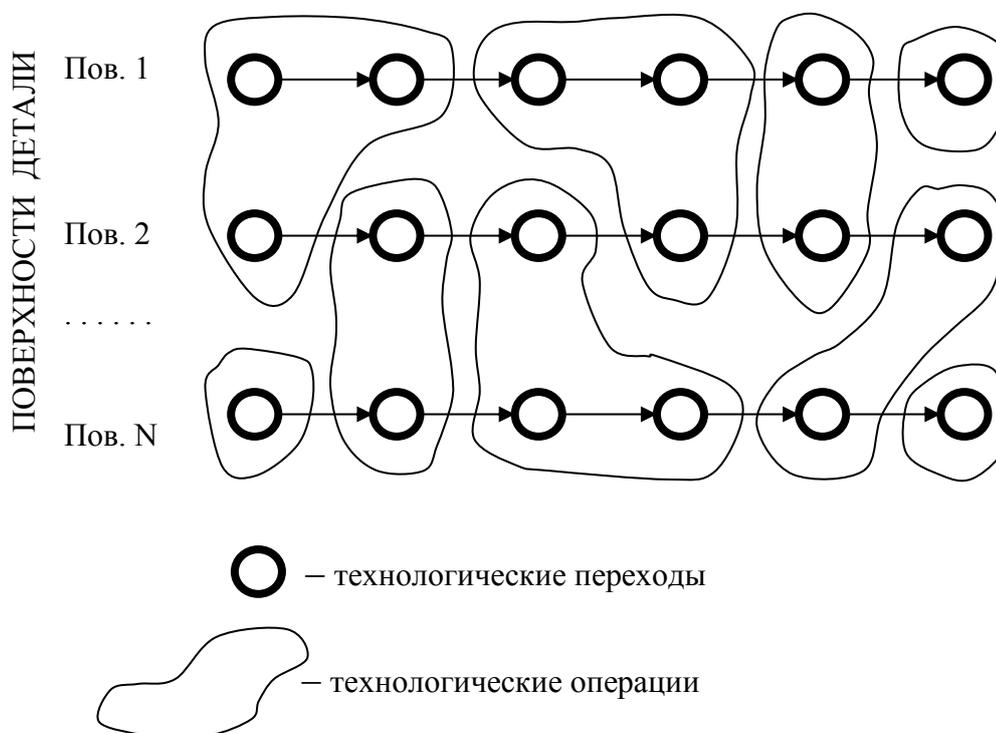


Рисунок 2. Формирование технологических операций при восходящем проектировании

Кроме этого синтез структуры осложняется использованием в наукоемких технологических процессах методов обработки, направленных на модификацию свойств поверхностного слоя ответственных деталей с повышенными эксплуатационными характеристиками: ионной имплантации, нанесения покрытий, электрохимической и др.

Отказ от детерминированных алгоритмов синтеза структуры ТП за счет использования «мягких вычислений» методами искусственного интеллекта позволил повысить эффективность поиска, снижая остроту вышеназванной проблемы.

Проведенное исследование обосновало новый подход к синтезу структуры наукоемкого технологического процесса.

В основу положен принцип восходящего проектирования, предложенный в свое время профессором Цветковым В.Д. [3] (см. рисунок 2).

В качестве информационной технологии используется процедура кластерного анализа [6], реализующая кластеризацию технологических переходов в пространстве характеризующих их признаков.

После декомпозиции конструкции детали на элементарные поверхности для каждой такой поверхности исходя из ее параметров качества (кавалитета точности, микротвердости и др.) и функционально-геометрических характеристик в базе знаний находится ТМП (см. рисунок 2) как типовое проектное технологическое решение. При этом каждому технологическому переходу, содержащемуся в ТМП, сопоставляется множество атрибутов, используемых при принятии решений в процессе формирования технологических операций и всей структуры технологического процесса на уровне маршрутного описания (см. рисунок 1).

Первый атрибут, используемый как отдельное измерение при кластеризации, содержит множество значений – номера поверхностей детали, которые альтернативно могут использоваться в качестве установочных технологических баз.

Второй атрибут, также используемый как отдельное измерение при кластеризации, содержит свое множество значений – коды технологического оборудования, которое альтернативно может использоваться для выполнения данного технологического перехода. Фактически здесь речь идет о методе обработки поверхности. Однако при этом учитывается то обстоятельство, что современное технологическое оборудование позволяют осуществлять различную технологическую обработку за одну установку заготовки.

Третий атрибут представляет собой вектор значений. Каждый элемент вектора образует отдельное измерение, используемое при кластеризации, и описывает отношение выполняемого технологического перехода к вышеуказанным технологическим методам модификации структуры поверхностного слоя детали, направленным на получение повышенных эксплуатационных свойств.

Четвертый атрибут также представляет собой вектор значений. Каждый элемент вектора образует отдельное измерение, используемое при кластеризации, и содержит значение параметра качества поверхности: квалитет точности, параметр шероховатости, этап технологической обработки, параметр микротвердости и др.

В результате получается пространство признаков, в котором осуществляется группирование технологических переходов (см. рисунок 2) в рамках соответствующей операции с использованием математического аппарата кластерного анализа [6].

В задаче имеет место нечеткость пространства рассматриваемых решений ввиду используемых признаков, содержащих альтернативные значения (например, перечень возможных технологических баз). Это приводит к нечеткости значения расстояния между объектами кластеризации (см. рисунок 3).

Для такого случая реализован следующий способ кластеризации. Каждый технологический переход представляется в процедуре кластеризации не одним объектом, а несколькими по числу сочетаний дискретных значений признаков.

Ввиду этого по результатам кластеризации необходимо осуществлять отсев среди объектов, дублирующих технологические переходы.

Теоретически такое удаление объектов кластеризации, представляющих один технологический переход, приводит к изменению картины кластеризации традиционными алгоритмами.

Использование в нашем случае эволюционной кластеризации [6] лишено данного недостатка, так как используемые в генетическом алгоритме хромосомы содержат

только допустимые варианты кластеризации. Для этого используется генетический оператор фильтрации порождаемых хромосом [6].

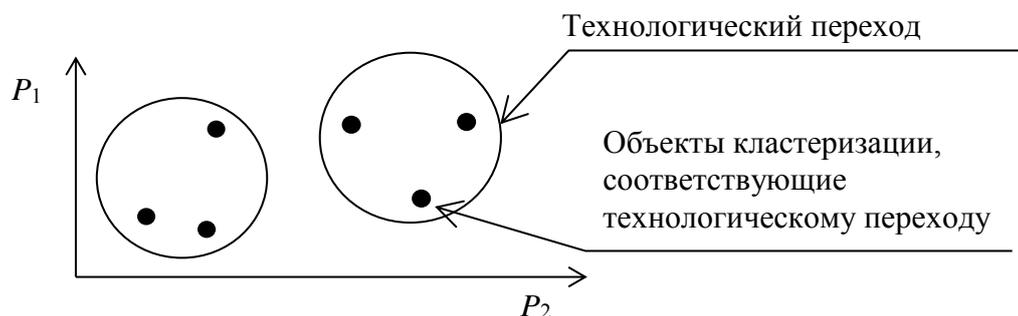


Рисунок 3. Нечеткость расстояния в пространстве дискретных значений признаков кластеризации P_1 и P_2

Для проверки работоспособности предложенного метода синтеза технологических операций с использованием эволюционной кластеризации были проведены вычислительные эксперименты на группе корпусных деталей предприятия среднего машиностроения. Полученные результаты подтвердили эффективность предложенного метода.

Выводы

Во-первых, предложен новый способ синтеза технологических операций средствами САПР ТП, основанный на эволюционной кластеризации, формализующий данный процесс и, соответственно, повышающий производительность автоматизированного проектирования технологических процессов.

Во-вторых, подтвердилась работоспособность предложенного способа по результатам тестовых вычислений для класса корпусных деталей.

Литература

1. Анфёров М.А. Системная оптимизация наукоемких технологий // Авиационная техника: Известия вузов. 2002. №2. С. 57-60.
2. Капустин Н.М. Принципы и методика автоматизированного проектирования технологических процессов обработки деталей в машиностроении: дис. докт. техн. наук: 05.02.08. М., 1976. 447 с.
3. Цветков В.Д. Многоуровневый синтез структурных моделей технологических процессов обработки деталей в автоматизированных системах проектирования: дис. докт. техн. наук: 05.13.12. Минск, 1979. 412 с.
4. Базаров Б.М. Организация проектирования модульных технологических процессов изготовления деталей // Вестн. маш. 1995. №5. С. 23-28.
5. Фролова И.Н., Кутилова О.И. Анализ современных систем автоматизированного проектирования технологических процессов (САПР ТП) // Труды НГТУ им. Р.Е. Алексеева. 2010. №1(80). С. 91-94
6. Анфёров М.А. Генетический алгоритм кластеризации // Российский технологический журнал. 2019. Т.7, №6. С. 134-150.