

ISSN 2949-1215



Российская Академия Наук

ТРУДЫ

Кольского научного центра РАН

3/2024(15)

СЕРИЯ: ТЕХНИЧЕСКИЕ НАУКИ

0+

Российская Академия Наук

ТРУДЫ

3/2024(15)

Научно-информационный журнал
Основан в 2010 году
Выходит 4 раза в год

Кольского научного центра. Серия: Технические науки

Главный редактор
акад. РАН, д. г.-м. н. С. В. Кривовичев

Заместитель главного редактора
к. б. н. Е. А. Боровичев

Редакционный совет:
акад. РАН, д. г. н. Г. Г. Матишов,
чл.-корр. РАН д. б. н. В. К. Жиров,
чл.-корр. РАН д. т. н. А. И. Николаев,
чл.-корр. РАН, д. х. н. И. Г. Тананаев,
д. э. н. Ф. Д. Ларичкин,
д. ф.-м. н. Е. Д. Терещенко,
к. т. н. А. С. Карпов (отв. секретарь)

Редколлегия серии:
чл.-корр. РАН, д. т. н. А. И. Николаев,
чл.-корр. РАН, д. х. н. И. Г. Тананаев,
д. т. н. А. В. Горохов,
д. х. н. С. Р. Деркач,
д. т. н. А. А. Козырев,
д. т. н. Н. В. Коровкин,
д. т. н. С. И. Кривошеев,
д. х. н. С. А. Кузнецов,
д. т. н. С. В. Лукичев,
д. т. н. В. А. Марлей,
д. т. н. М. В. Маслова,
д. т. н. А. В. Маслобоев,
д. т. н. В. А. Маслобоев,
д. т. н. О. В. Наговицын,
д. т. н. А. Г. Олейник,
д. т. н. В. В. Рыбин,
д. т. н. А. Я. Фридман,
д. т. н. М. Г. Шишаев,
к. т. н. Т. Н. Васильева,
к. т. н. И. О. Датъев,
к. х. н. Д. П. Домонов,
к. т. н. О. Г. Журавлева,
к. т. н. А. И. Калашник,
к. т. н. А. С. Карпов,
к. т. н. В. В. Колобов,
к. т. н. Н. М. Кузнецов,
к. т. н. С. М. Маслобоева,
к. т. н. В. А. Минин,
к. т. н. Г. И. Митрофанова,
к. т. н. А. С. Опалев,
к. т. н. В. Н. Селиванов,
к. т. н. И. Э. Семенова,
к. т. н. А. М. Фёдоров

Ответственные редакторы номера:
к. т. н. А. М. Фёдоров

Учредитель — Федеральное государственное бюджетное учреждение науки
Федеральный исследовательский центр «Кольский научный центр Российской
академии наук» (ФИЦ КНЦ РАН)

Свидетельство о регистрации СМИ ПИ № ФС77-83502 от 30 июня 2022 г.
выдано Федеральной службой по надзору в сфере связи, информационных
технологий и массовых коммуникаций.

Научное издание

Технический редактор Е. П. Шинкарук
Редактор Ю. Н. Еремеева
Подписано к печати 27.12.2024.
Дата выхода в свет 29.12.2024.
Формат бумаги 60×84 ½.
Усл. печ. л. 13,71. Заказ № 60. Тираж 500 экз.
Свободная цена.

Адрес учредителя, издателя и типографии:
Федеральное государственное бюджетное учреждение науки
Федеральный исследовательский центр
«Кольский научный центр РАН» (ФИЦ КНЦ РАН)
184209, г. Апатиты, Мурманская область, ул. Ферсмана, 14
Тел.: 8 (81555) 7-53-50; 7-95-95, факс: 8 (81555) 7-64-25
E-mail: ksc@ksc.ru. www.ksc.ru

СОДЕРЖАНИЕ

Олейник А. Г., Федоров А. М., Датьев И. О., Зуенко А. А., Шестаков А. В., Вишняков И. Г.	Использование RAG-технологии для создания интеллектуальной информационной системы поддержки исследовательского поиска.....	5
Быстров В. В., Халиуллина Д. Н., Шишаев М. Г.	Потенциал применения генеративного искусственного интеллекта для имитационного моделирования жизнеспособности региональных социально-экономических систем.....	27
Фридман А. Я., Маслобоев А. В.	Ситуационные цифровые двойники в управлении критическими инфраструктурами: спецификация требований.....	41
Пимешков В. К., Никонорова М. Л., Шишаев М. Г., Вишняков И. Г.	Формирование многослойных графов знаний на базе тематического моделирования текстов.....	50
Шестаков А. В., Зуенко А. А.	Решение задач маршрутизации транспортных средств в рамках парадигмы программирования в ограничениях.....	61
Неупокоева Е. О., Малыгина С. Н., Быстров В. В.	Практика применения нечеткой логики для имитационного моделирования сложных систем.....	69
Зуенко А. А., Фридман О. В.	Обзор методов поиска частых паттернов для интеллектуального анализа данных.....	82
Руденко Н. Н., Вдовиченко Н. А.	Организация управления программными компонентами в системе концептуального моделирования с использованием PYTHON и PYQT5.....	97
Вдовиченко Н. А., Руденко Н. Н.	Разработка ГИС-интерфейса для ситуационной системы вычислительного эксперимента на основе C# и DOTSPATIAL.....	105
Яковлев С. Ю., Шемякин А. С.	Опасные объекты и процессы Арктической зоны Российской Федерации (по материалам конференции «III Юдахинские чтения»).....	113

3/2024(15)

Scientific journal
Published since 2010
Publication frequency — four times a year

Russian Academy of Sciences
TRANSACTIONS

Kola Science Centre. Series: Engineering Sciences

Editor-in-Chief

S. V. Krivovichev, Academician of RAS,
Dr. Sc. (Geology & Mineralogy)

Deputy Editor-in-Chief

E. A. Borovichev, PhD (Biology)

Editorial Council:

G. G. Matishov, Academician of RAS, Dr. Sc. (Geography),
V. K. Zhirov, Cor. Member of RAS, Dr. Sc. (Biology),
A. I. Nikolaev, Cor. Member of RAS, Dr. Sc. (Engineering),
I. G. Tananaev, Cor. Member of RAS, Dr. Sc. (Chemistry),
F. D. Larichkin, Dr. Sc. (Economics),
E. D. Tereshchenko, Dr. Sc. (Physics and Mathematics),
A. S. Karpov, PhD (Engineering) — Executive Secretary

Editorial Board:

A. I. Nikolaev, Cor. Member of RAS, Dr. Sc. (Engineering),
I. G. Tananaev, Cor. Member of RAS, Dr. Sc. (Chemistry),
A. V. Gorokhov, Dr. Sc. (Engineering),
S. R. Derkach, Dr. Sc. (Chemistry),
A. A. Kozirev, Dr. Sc. (Engineering),
N. V. Korovkin, Dr. Sc. (Engineering),
S. I. Krivosheev, Dr. Sc. (Engineering),
S. A. Kuznetsov, Dr. Sc. (Chemistry),
S. V. Lukichev, Dr. Sc. (Engineering),
V. A. Marlej, Dr. Sc. (Engineering),
M. V. Maslova, Dr. Sc. (Engineering),
A. V. Masloboev, Dr. Sc. (Engineering),
V. A. Masloboev, Dr. Sc. (Engineering),
O. V. Nagovitsin, Dr. Sc. (Engineering),
A. G. Oleinik, Dr. Sc. (Engineering),
V. V. Ribin, Dr. Sc. (Engineering),
A. Ya. Fridman, Dr. Sc. (Engineering),
M. G. Shishaev, Dr. Sc. (Engineering),
T. N. Vasileva, PhD (Engineering),
I. O. Datyev, PhD (Engineering),
D. P. Domanov, PhD (Chemistry),
O. G. Zhuravleva, PhD (Engineering),
A. I. Kalashnik, PhD (Engineering),
A. S. Karpov, PhD (Engineering),
V. V. Kolobov, PhD (Engineering),
N. M. Kuznetsov, PhD (Engineering),
S. M. Masloboeva, PhD (Engineering),
V. A. Minin, PhD (Engineering),
G. I. Mitrofanova, PhD (Engineering),
A. S. Opalev, PhD (Engineering),
V. S. Selivanov, PhD (Engineering),
I. E. Semenova, PhD (Engineering),
A. M. Fedorov, PhD (Engineering)

Executive Editors:

A. M. Fedorov, PhD (Engineering)

Founder — Federal State Budget Institution of Science
Federal Research Centre “Kola Science
Centre of the Russian Academy of Sciences”

Mass Media Registration Certificate

ПИ No. ФС77-83502 issued by the Federal Service for Supervision
of Communications, Information Technology and Mass Media
on June, 30 2022

Scientific publication

Technical Editor E. P. Shinkaruk
Editor Yu. N. Yeremeyeva

14, Fersman str., Apatity, Murmansk region, 184209, Russia
Tel.: 8 (81555) 7-93-80. Fax: 8 (81555) 7-64-25
E-mail: ksc@ksc.ru. www.ksc.ru

CONTENTS

Oleynik A. G., Fedorov A. M., Datyev I. O., Zuenko A. A., Shestakov A. V., Vishnyakov I. G.	Using rag technology to design an intelligent information system for support exploratory search.....	5
Bystrov V. V., Khaliullina D. N., Shishaev M. G.	The potential of using generative artificial intelligence to simulate the resilience of regional socio-economic systems.....	27
Fridman A. Ya., Masloboev A. V.	Situational digital twins in management of critical infrastructures: specification of requirements.....	41
Pimeshkov V. K., Nikonorova M. L., Shishaev M. G., Vishnyakov I. G.	Formation of multilayer knowledge graphs based on thematic texts modeling.....	50
Shestakov A. V., Zuenko A. A.	Solving vehicle routing problems in the framework of the constraint programming paradigm.....	61
Neupokoeva E. O., Malygina S. N., Bystrov V. V.	The practice of using fuzzy logic for simulation of complex systems.....	69
Zuenko A. A., Fridman O. V.	Survey of frequent pattern search methods for data mining.....	82
Rudenko N. N., Vdovichenko N. A.	Organizing software components management in a conceptual modeling system using PYTHON and PYQT5.....	97
Vdovichenko N. A., Rudenko N. N.	Development of gis-interface for situational system of computational experiment on the basis of C# and DOTSPATIAL.....	105
Yakovlev S. Yu., Shemyakin A. S.	Hazardous objects and processes of the Arctic zone of the Russian Federation (based on the proceedings of the conference "III Yudakhin readings").....	113

Научная статья
УДК 004.853
doi:10.37614/2949.1215.2024.15.3.001

ИСПОЛЬЗОВАНИЕ RAG-ТЕХНОЛОГИИ ДЛЯ СОЗДАНИЯ ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПОДДЕРЖКИ ИССЛЕДОВАТЕЛЬСКОГО ПОИСКА

Андрей Григорьевич Олейник¹, Андрей Михайлович Федоров^{2✉}, Игорь Олегович Датьев³,
Александр Анатольевич Зуенко⁴, Алексей Владимирович Шестаков⁵, Иван Геннадьевич Вишняков⁶
^{1–6}Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия

¹a.oleynik@ksc.ru, <https://orcid.org/0000-0002-7612-5999>

²a.fedorov@ksc.ru, <https://orcid.org/0000-0002-2862-7994>

³i.datyev@ksc.ru, <https://orcid.org/0000-0002-8372-8704>

⁴a.zuenko@ksc.ru, <https://orcid.org/0000-0002-7165-6651>

⁵a.shestakov@ksc.ru, <https://orcid.org/0000-0002-9052-2579>

⁶i.vishnyakov@ksc.ru, <https://orcid.org/0009-0003-4938-5693>

Аннотация

Работа посвящена практике интеграции большой языковой модели в систему интеллектуальной информационно-аналитической поддержки исследовательского поиска информации в заданных массивах данных. Проведен обзор возможностей современных больших языковых моделей. Особое внимание уделено технологии RAG, используемой для оперативного формирования и управления контекстом поиска. Описаны особенности проектирования и разработки функций бизнес-логики и элементов пользовательских интерфейсов системы. Представлены результаты экспериментов применения разработанной системы к текстам с заданной спецификой. Сделаны выводы о применимости системы для автоматизации процесса заселения онтологий, а также наполнения других хранилищ знаний, построенных с использованием автоматизированного мониторинга больших открытых данных и интеллектуальных средств их обработки.

Ключевые слова:

онтология, RAG-технология, нейронная сеть, большая языковая модель, LLM, информационно-аналитическая система, интеллектуальный анализ данных, проектирование программного обеспечения, разработка программного обеспечения, веб-сервис, веб-приложение

Финансирование:

работа выполнена в рамках НИР «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (регистрационный номер 122022800547-3) и НИР «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер 122022800551-0).

Для цитирования:

Олейник А. Г., Федоров А. М., Датьев И. О., Зуенко А. А., Шестаков А. В., Вишняков И. Г. Использование RAG-технологии для создания интеллектуальной информационной системы поддержки исследовательского поиска // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 5–26. doi:10.37614/2949.1215.2024.15.3.001.

Original article

USING RAG TECHNOLOGY TO DESIGN AN INTELLIGENT INFORMATION SYSTEM FOR SUPPORT EXPLORATORY SEARCH

Andrey G. Oleynik¹, Andrey. M. Fedorov^{2✉}, Igor O. Datyev³, Alexander A. Zuenko⁴,
Aleksey V. Shestakov⁵, Ivan G. Vishnyakov⁶

^{1–6}Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences, Apatity, Russia

¹a.oleynik@ksc.ru, <https://orcid.org/0000-0002-7612-5999>

²a.fedorov@ksc.ru, <https://orcid.org/0000-0002-2862-7994>

³i.datyev@ksc.ru, <https://orcid.org/0000-0002-8372-8704>

⁴a.zuenko@ksc.ru, <https://orcid.org/0000-0002-7165-6651>

⁵a.shestakov@ksc.ru, <https://orcid.org/0000-0002-9052-2579>

⁶i.vishnyakov@ksc.ru, <https://orcid.org/0009-0003-4938-5693>

Abstract

This paper presents the integration of a large language model (LLM) into an intelligent information and analytical system for exploratory search support within specific texts. An overview of the capabilities of modern large language models is provided. Special attention is paid to the Retrieval Augmented Generation (RAG) technology used for efficient formation and control of search contexts. The design and development of the system's business logic functions and user interface elements are detailed. Experimental results applying the developed system to texts with specific characteristics are presented. Conclusions are drawn regarding the system's applicability for automating ontology population. Conclusions are drawn regarding the system's applicability for automating the process of populating ontologies, as well as filling other knowledge repositories built using automated monitoring of open big data and intelligent data processing.

Keywords:

ontology, retrieval augmented generation, RAG, neural network, large language model, LLM, information and analytical system, data mining, software design, software development, web service, web application

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic "Development of theoretical and organizational and technical foundations of information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation" (registration number of the research topic 122022800547-3) and research topic "Methodology for creating information and analytical systems to support regional development management based on formative artificial intelligence and big data" (registration number of the research topic 122022800551-0).

For citation:

Andrey G. Oleynik, Igor O. Datyev, Alexander A. Zuenko, Andrey M. Fedorov, Aleksey V. Shestakov, Ivan G. Vishnyakov Using RAG technology to design an intelligent information system for support exploratory search // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2023. Vol. 15, No. 3. P. 5–26. doi:10.37614/2949.1215.2024.15.3.001.

Введение

Плановые исследования Института информатики и математического моделирования им. В. А. Путилова (далее — ИИММ) включают разработку решений ряда актуальных научных задач, связанных с созданием эффективных методов формирования баз знаний и интеллектуальных информационных систем в условиях сложности, неопределенности, разнородности и динамичности предметной области, а также методов аналитической обработки данных и знаний, ориентированных на проблематику регионального развития и поддержку соответствующего ему управления. Одним из важных направлений является разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации. Эффективное решение задач управления требует повышения уровня согласованности как нормативно-правовой базы, регламентирующей различные виды деятельности, так и «мультидисциплинарного» понятийного аппарата, используемого в процессах принятия решений в различных сферах и на различных уровнях управления. Поэтому в рамках тематики ИИММ предполагается развить методологию создания интеллектуальных проблемно-ориентированных информационных систем на базе концепции «формирующего искусственного интеллекта».

Так, в ходе разработки основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации требует решения задача четкой аргументированной идентификации и классификации объектов, которые могут быть «слабым звеном» включающих их или связанных с ними инфраструктур. К этим объектам прежде всего относятся критически важные объекты (КВО) и потенциально опасные объекты (ПОО) [1]. В качестве инструмента решения данной задачи могут быть использованы формальные онтологии, позволяющие сформировать согласованный понятийный базис для специалистов различных предметных областей [2]. В работе [3] была представлена технология согласования нормативно-правовых документов на основе онтологического подхода, использование которой могло бы способствовать формированию целостной системы норм и правил функционирования различных акторов как на региональном уровне, так и их взаимодействия с уровнем федеральным.

Подобные задачи также возникают при создании информационно-аналитических систем, нацеленных на обеспечение информационной поддержки задач управления региональным развитием. Региональная социально-экономическая система рассматривается как взаимосвязанная триада «бизнес-общество-власть» [4]. Элементом такой системы и ее критической инфраструктуре свойственны

неопределенность и спонтанная динамика в виду наличия факторов риска. Анализировать и управлять объектами и процессами с подобной спецификой позволяют системы информационной поддержки на основе интеллектуализированных компонент в виде семантических моделей, онтологий и других хранилищ знаний, построенных с использованием автоматизированного мониторинга больших открытых данных (например, контента социальных сетей) [5] и интеллектуальных средств их обработки [6].

Несмотря на развитие методов, технологий и инструментальных средств онтологического проектирования, построение конкретных проблемно-ориентированных онтологий остается трудоемкой задачей. Одним из путей технической поддержки создания и наполнения/пополнения онтологий стало использование инструментов искусственного интеллекта. В первую очередь это инструменты работы с текстами на естественном языке. Одной из подзадач обучения онтологии — ее «заселение» (Ontology Population — OP), которая подразумевает расширение существующей онтологии новыми экземплярами без изменения структуры онтологии. Варианту решения этой подзадачи посвящена, в частности, работа [7]. В свою очередь, в исследовании [8], являющемся логическим продолжением работы [7], предлагается вариант решения задачи извлечения отношений из текстов предметной области для их возможного добавления в существующую онтологию. Вопросам извлечения сущностей и отношений посвящена и статья [9], представляющая технологию генеративных многошаговых вопрос-ответов (generative multi-turn question answering). В ней также кратко представлен обзор ряда существующих подходов к решению этой задачи. Отмечено, что недостатком конвейерного режима распознавания сущностей является распространение ошибок. Даются ссылки на публикации по различным стратегиям реализации методов совместного обучения с использованием взаимосвязи задач, подходам к совместному извлечению сущностей и отношений как проблеме заполнения таблиц, а также ряду других технологий.

Как указанные, так и ряд других публикаций представляют технологии использования нейросетевых моделей обработки естественных языковых текстов, предполагающие предварительное обучение, или дообучение этих моделей на объемных текстовых корпусах. Данная процедура сама по себе является довольно ресурсоемкой. Кроме того, такой подход может оказаться неэффективным в случае, когда контекст, с которым работает модель, существенно изменяется со временем. В этом случае начальная база знаний, сформированная на основе контекста, на котором проходило обучение модели, становится неактуальной и полученные ранее «знания» могут повлечь ошибочные для изменившейся ситуации ответы, формируемые моделью. Подобная ситуация может возникнуть, если «актуальный» контекст представляет собой, например, документацию развивающегося программного продукта или изменяющиеся нормативно-правовые документы. Постоянное дообучение модели с целью актуализации ее базы знаний дорого и не эффективно. Более эффективным решением в этом случае представляется использование технологии RAG (Retrieval Augmented Generation) [10]. Данная технология позволяет исключить ресурсоемкий этап дообучения языковой модели при создании проблемно-ориентированных приложений на основе хорошо обученных и уже готовых к использованию языковых моделей таких, как YandexGPT [11], GigaChat [12], OpenAI [13] и др.

Настоящая статья представляет начальный этап исследований коллектива авторов по применению RAG-технологий в работе со справочными и нормативными текстами для создания как автономных интеллектуализированных информационно-справочных систем, так подсистем аналитической поддержки, интегрируемых в сторонние информационные комплексы.

Используемые в работе технологии интеллектуализированной обработки данных

RAG-технология является одним из новых этапов развития технологий работы с текстами на естественном языке, использующих большие языковые модели (Large Language Models, LLM) [14]. LLM представляют собой нейронные сети, основанные на архитектуре трансформер [15] и обученные на больших объемах данных обрабатывать и генерировать текст на естественном языке. Среди задач, для решения которых применяются LLM, наибольший интерес в рамках проводимого исследования представляют обработка запросов и генерация ответов на вопросы пользователя в режиме диалога, а также генерация контекстуализированного текста на основе предоставленных вводных данных. Это позволяет создавать с использованием LLM «интеллектуальные» информационные системы,

взаимодействующие с пользователем на естественном языке, что, в свою очередь, снижает требования к технической подготовке пользователя. Однако априорно обученные LLM обладают существенными ограничениями при обработке запросов, требующих использования «новой» (выходящих за рамки данных, на которых они обучены) информации. В таких ситуациях LLM нередко генерируют «галлюцинации» — ответы, не имеющие отношения к действительности [16].

Во многом преодолеть проблемы, обусловленные ограничениями непосредственного использования LLM, позволяет RAG-технология и ее вариации. В рамках RAG-технологии реализуются механизмы поиска в базе данных информационной системы фрагментов текста, вектора которых наиболее релевантны векторизованному вопросу пользователя. На основе найденных фрагментов с использованием «знаний» LLM генерируется текст ответа. Это позволяет улучшить качество ответа системы на вопрос за счет учета конкретного контекста. Возможность оперативного изменения базы данных системы обуславливает ее адаптивность к изменению данных без необходимости полного переобучения всей модели или ее дообучения, для чего в общем случае требуются значительные вычислительные и временные ресурсы [17, 18].

При этом очевидно, что качество генерируемых системой ответов будет существенно зависеть от полноты и адекватности имеющейся базы данных. Поэтому оценка влияния подготовки данных на качество ответов системы является одной из основных задач начального этапа проводимого исследования.

В статье [10] представлен обзор эволюции RAG-технологии и довольно подробно рассматриваются методы работы трех основных компонентов этих технологий — поиска (Retrieval), дополнения (Augment) и генерации (Generation). Авторы этого обзора выделяют три этапа развития RAG, отличающиеся развитием инструментов тонкой настройки каждого из базовых компонентов.

Первый этап определен как простой, или «наивный» (Naive RAG). В Naive RAG [19] реализуются процедуры индексации исходных данных, их сегментация и векторизация. Векторизованные фрагменты сохраняются в базе данных информационной системы. Аналогичным образом происходит обработка и векторизация поступившего в систему вопроса пользователя. Процедуры извлечения из базы фрагментов, наиболее релевантных поступившему вопросу, основаны на вычислении оценки сходства между вектором вопроса и векторами хранимых фрагментов. Наиболее релевантные фрагменты, извлеченные из базы, совместно с текстом вопроса в дальнейшем включаются в расширенный контекст подсказки для ответа на вопрос, который должна сгенерировать используемая в системе предобученная LLM. Более того, уже и Naive RAG поддерживает и режим «диалога», когда в подсказку могут быть интегрированы вопросы и ответы, предшествующие текущему вопросу.

На втором этапе развития, названном Advanced RAG, вводятся дополнительные процедуры, повышающие качество поиска за счет усовершенствованных механизмов индексации и включения метаданных, позволяющие более точно соотнести извлекаемые данные с вопросом [20].

Третий этап характеризуется вводом в RAG-систему дополнительных компонентов, повышающих ее адаптивность и улучшающих возможности извлечения и обработки целевого контента [10]. Данный этап назван «Модульным» (Modular RAG).

Для реализации инструментальной среды проблемно-ориентированного использования RAG-технологии был применен фреймворк LangChain [21]. Он обеспечивает возможность интеграции и комбинирования различных компонентов обработки естественного языка и возможность создания адаптивных решений, которые могут взаимодействовать с внешними данными и контекстом пользователя. Поддержка LangChain работы с различными предобученными языковыми моделями, в том числе и с локальными, дает возможность исследовать и сравнивать эффективность разных LLM при решении различных задач. К основным компонентам LangChain относятся: API-интерфейсы; шаблоны подсказок (prompt patterns) различных типов; инструменты работы с индексами и инструменты создания и использования цепочек автоматизированного выполнения операций; функции запоминания как текущего «диалога», так и предшествующих. Также LangChain обладает инструментами преобразования, хранения, поиска и извлечения информации, что позволяет создавать RAG-системы.

Организационные, методологические и практические особенности проектирования и разработки системы

Представленная разработка опирается на идею по созданию интеллектуализированной справочной системы, способной выдавать пользователю ответы на основе предварительно заданного

набора документов. На этапе проектирования классический механизм формирования ответов на основе полнотекстового синтаксического поиска был определен как недостаточный. Ответы разрабатываемой системы должны учитывать неявные семантические взаимосвязи между различными фрагментами загружаемого набора документов. С учетом указанных требований для построения системы выбраны технологии, используемые при разработке чат-ботов. В частности, основным рабочим механизмом для реализации системы выбраны большие языковые модели и технология RAG для оперативного доступа к ним.

Указанный набор инструментов и технологий позволил реализовать требуемую базовую функциональность системы. Дополнительно возникла необходимость расширить область применения разработки и использовать ее не только в качестве интеллектуальной справки, но и как исследовательский инструмент для изучения особенностей языковых моделей и эффективных способов взаимодействия с ними.

На рис. 1 в виде UML-диаграммы использования представлены формальные требования к функциональным возможностям информационной системы интеллектуальной справки.

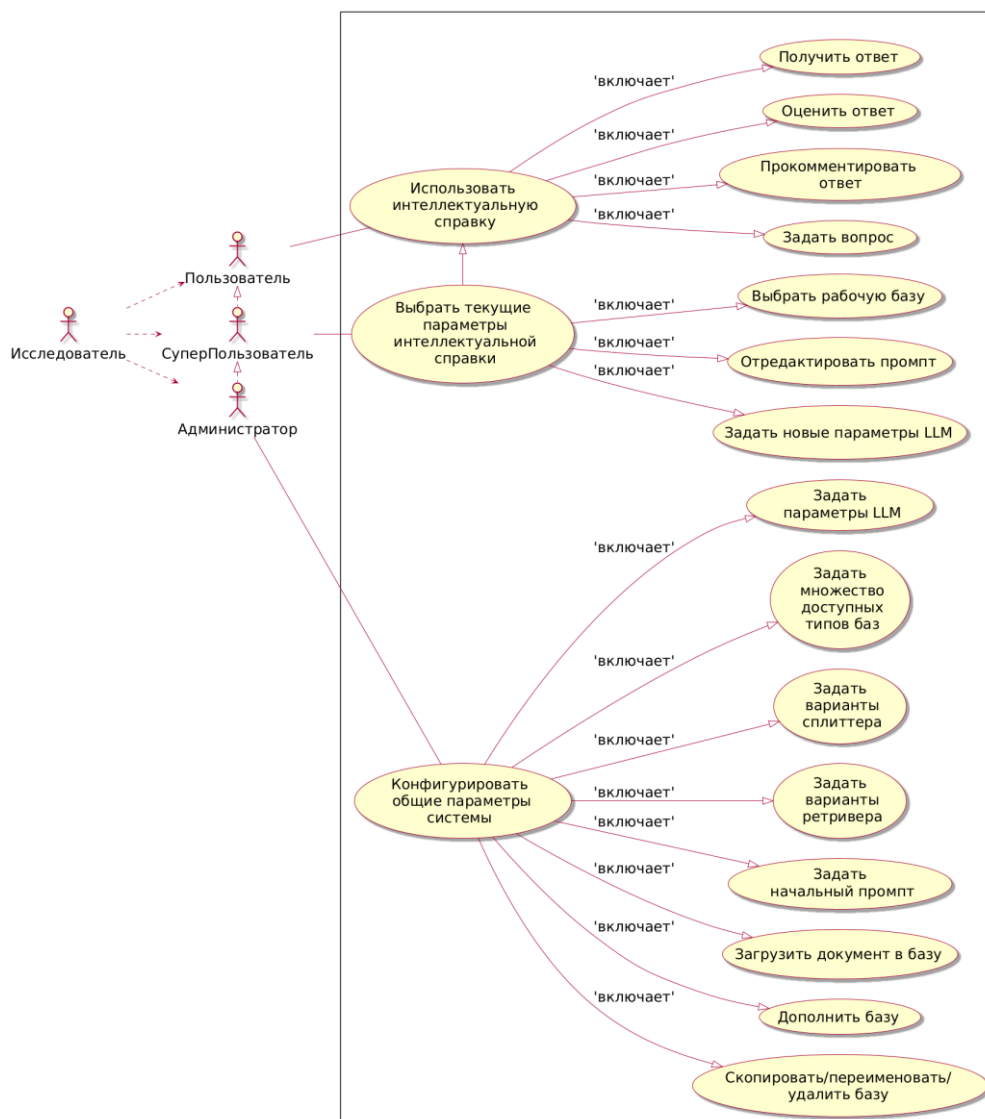


Рис. 1. UML-диаграмма использования — требования к функциональным возможностям информационной системы интеллектуальной справки

Для удобства использования и администрирования в системе реализована классическая ролевая пользовательская модель. Обобщенная роль «Исследователь» специализирована конкретными ролями «Пользователь», «СуперПользователь», «Администратор». Специализация задает уровни доступа пользователей к параметрам и функциям системы. «Администраторам» доступны все инструменты для использования и конфигурирования всех параметров системы. «СуперПользователи», помимо использования, могут изменять текущие параметры интеллектуальной справки. Для «Пользователей» предоставляется доступ к преднастроенным основным функциям справки.

Все функции в системе отнесены к следующим группам: 1) конфигурирование общих параметров системы; 2) настройка текущих параметров конфигурации; 3) непосредственное использование справочной системы.

Первая группа функций предназначена для администраторов системы. В нее входят функции загрузки наборов документов и управления соответствующими базами данных, а также функции для настройки необходимых операций по подготовке документов и их размещению в базах.

Вторую группу функций используют привилегированные пользователи («СуперПользователи») для предварительной настройки системы непосредственно в процессе ее работы. В частности, здесь реализованы действия по выбору рабочей базы из списка ранее загруженных баз данных, а также выбор опций для формирования запросов (промттов) к языковой модели.

Основная группа функций используется обыкновенными пользователями для непосредственной работы со справочной системой. Именно этот набор интерфейсных вызовов был основным требованием к разрабатываемой системе. Здесь сгруппированы действия: сформировать вопрос, отправить его на вход языковой модели, получить и оценить ответ.

В выборе инструментария для разработки предпочтение отдавалось отечественным и платформонезависимым технологиям, программам и системам. В рамках данного проекта для координации и совместной работы команды разработчиков и исследователей использованы:

— GitLab — развернутая на серверной инфраструктуре ИИММ система управления проектами и репозиториями программного кода;

— Яндекс.Диск, Яндекс.Документы — облачная инфраструктура для создания и совместного редактирования документов;

— «Телеграм» — приложение для обмена мгновенными сообщениями.

Для непосредственной разработки активно использовались:

— PyCharm Community — кроссплатформенная интегрированная среда разработки для языка программирования Python;

— VS Codium — свободно лицензируемый дистрибутив кроссплатформенного редактора для разработки различных приложений VS Code (Microsoft);

— Python 3.11.7 — объектно-ориентированный интерактивный язык программирования;

— ALT Linux — отечественный дистрибутив ОС Linux, основанный на ядре Linux и репозитории пакетов Sisyphus. ОС на рабочей станции разработчиков;

— Debian 12 (Bookworm) — операционная система с открытым исходным кодом GNU/Linux. ОС на облачном сервере с конфигурацией 1×3.3 ГГц, CPU 1 Гб, RAM 15 Гб NVMe для развертывания компонентов приложения.

Все используемые в данной работе программные системы, информационные технологии и подходы к коллективной разработке приложений ранее эффективно опробованы авторами как командой разработчиков и успешно зарекомендовали себя на практических этапах [22, 23] в научно-исследовательских работах по поддержке управления региональным развитием [24] и поддержке управления жизнеспособностью региональных критических инфраструктур [25], выполненных в рамках плановых НИР в ИИММ.

Общая архитектура и принципы построения разработанной системы

Архитектура разработанной системы представлена в виде UML-диаграммы развертывания (deployment diagram) на рис. 2. В разработке применена технология докеризации. В настоящее время

в системе используется один докер-компонент, который построен на основе последней версии образа контейнера “langchain/langchain” из репозитория dockerhub. Докеризация позволяет эффективно управлять ресурсами в процессе разработки, гибко разворачивать компоненты проекта в различные рабочие конфигурации, а также масштабировать и планировать дальнейшее развитие полученного решения.

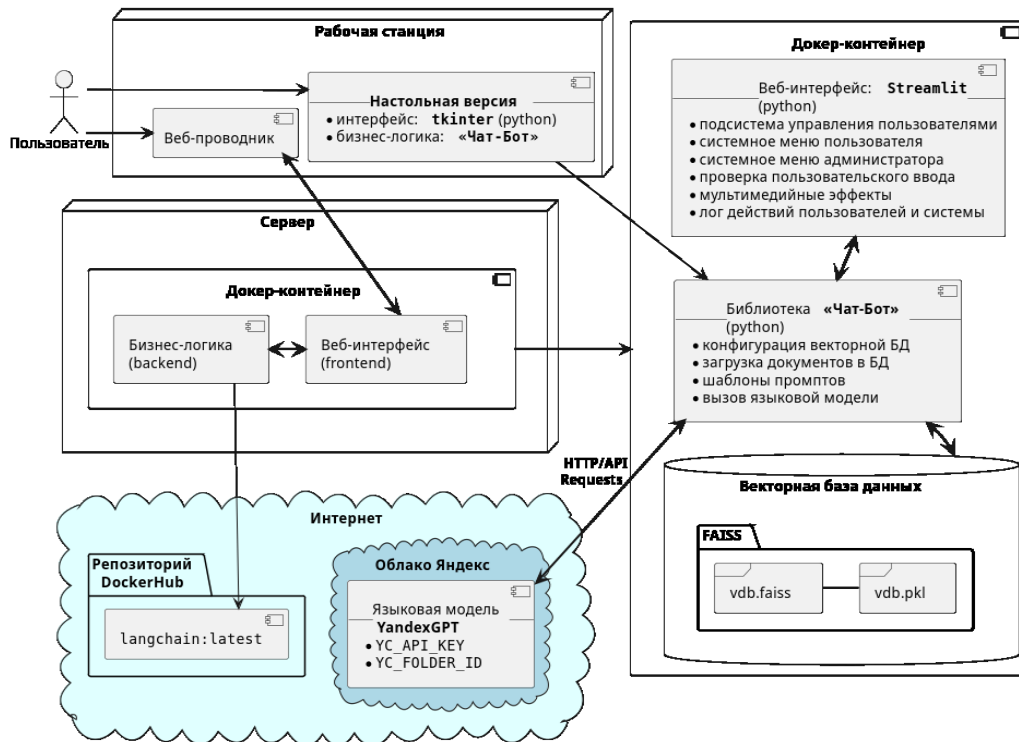


Рис. 2. Архитектура разработанной системы (UML-диаграмма развертывания)

Логически система разделена на два компонента. Реализующий бизнес-логику компонент (серверная часть, бэкенд) выполнен в виде библиотеки на языке Python 3. С помощью функций этой библиотеки производится подготовка всех необходимых структур данных и обращения к языковой модели YandexGPT. Для обеспечения более универсального обращения к выбранной языковой модели и к ее аналогам используется библиотека LangChain, все необходимые компоненты которой предустановлены в используемом образе контейнера. Технология RAG реализуется через использование локальной векторной базы данных FAISS, компоненты которой размещаются в файловой системе докера.

Вторая составляющая системы — это пользовательский веб-интерфейс (фронтенд), также выполненный на языке Python 3 на основе фреймворка Streamlit [26]. Использование языка Python и для бизнес-логики, и для веб-интерфейса позволило естественным образом соединить два этих компонента. Библиотека для работы с языковой моделью импортирована в основной программный модуль, в котором в виде веб-интерфейсов реализована общая логика работы системы.

Дополнительно в разделе «Разработка пользовательского интерфейса» дано описание альтернативной версии интерфейса, разработанной для специального режима запуска системы в формате настольного приложения.

Принципы реализации основных функциональных элементов системы

К основным функциям серверной части системы относятся: 1) загрузка документов в базу данных и 2) обращение к языковой модели для получения ответа на вопрос.

Все функции реализованной библиотеки являются атомарными и работают независимо друг от друга. Каждая функция получает на вход все необходимые для работы данные в виде непосредственных значений из передаваемых параметров и в виде инициализирующих значений из конфигурационных файлов. Полученные в ходе выполнения функции данные либо возвращаются как результат функции, либо сохраняются в заранее определенных структурах — конфигурационных файлах, базах данных. Возникающие в процессе работы функций ошибки и нештатные ситуации обрабатываются и отображаются для пользователя в виде информативных сервисных сообщений. В программном коде широко используется стандартный для языка Python механизм обработки исключительных ситуаций (try...catch). Это позволяет сохранять работоспособность системы на приемлемом для решения поставленных задач уровне.

Сообщения о возникающих в системе событиях, ошибках, штатных и нештатных ситуациях записываются в файл протокола (лог-файл) и доступны администраторам системы для проведения последующих процедур аудита и анализа.

Загрузка документов в базу данных

На рис. 3 в виде UML-диаграммы последовательности (sequence diagram) представлен процесс загрузки документов в векторную базу данных.

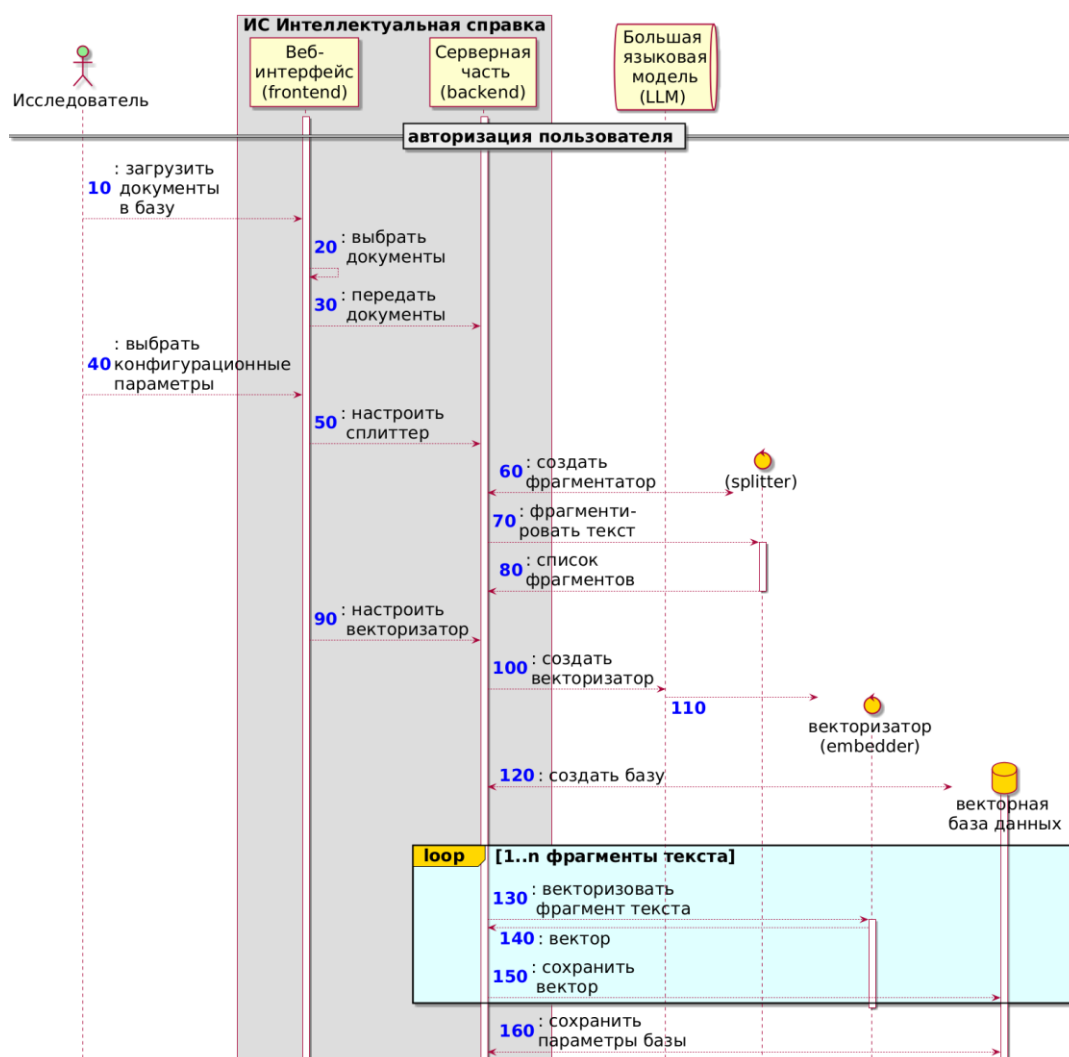


Рис. 3. UML-диаграмма последовательности: загрузка документов в векторную базу данных

В соответствии особенностями технологии RAG неотъемлемой частью разработанной системы является векторная база данных, которая используется для формирования контекстной части запроса (промпта) к языковой модели. Необходимо отметить, что используемые для работы вектора (эмбединги) можно хранить в альтернативных структурах, например, в файлах на диске или в списках в оперативной памяти. В данной работе для удобства реализации технологии RAG, в том числе для обеспечения возможностей адаптивной масштабируемости, гибкой развертываемости и оперативной интеграции, используется векторная база данных FAISS.

Загрузка документов в базы данных производится привилегированным пользователем в рамках предварительной настройки системы. Дизайн пользовательского интерфейса выполнен в классическом стиле, с использованием стандартных интерактивных веб-элементов из фреймворка Streamlit (рис. 4, 5).

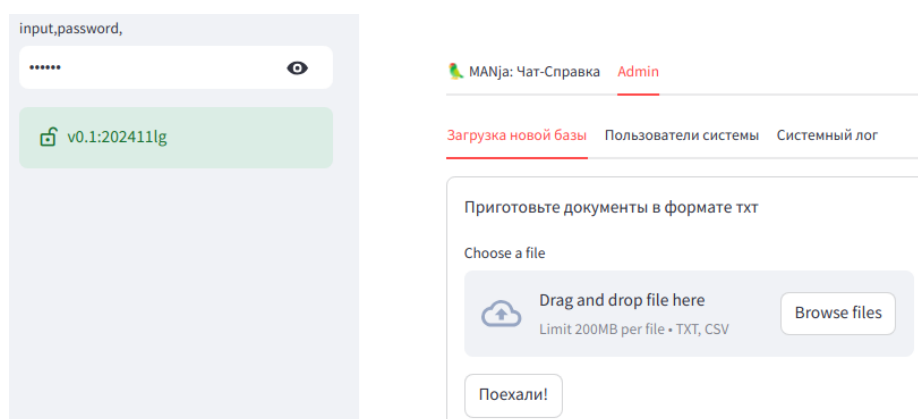


Рис. 4. Интерфейс пользователя: загрузка документов в базу данных

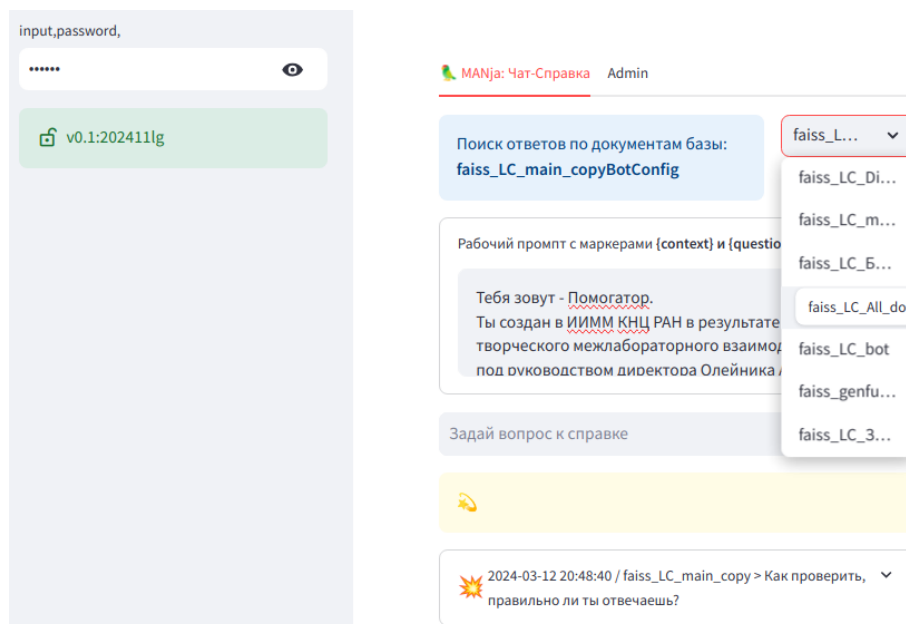


Рис. 5. Интерфейс пользователя: выбор рабочей базы данных

Большая часть представленных на UML-диаграмме операций скрыта от пользователя. После указания загружаемых документов пользователь получает сообщение о статусе загрузки и может продолжать работу с системой. Однако некоторые параметры этих операций представляют собой

исследовательский интерес. Поэтому в системе есть возможность предварительно настраивать эти параметры и сохранять настройки в конфигурационном файле, который и будет в дальнейшем использоваться для загрузки документов.

Подготовка запросов

Процесс работы пользователя с системой представлен UML-диаграммой последовательности (рис. 6).

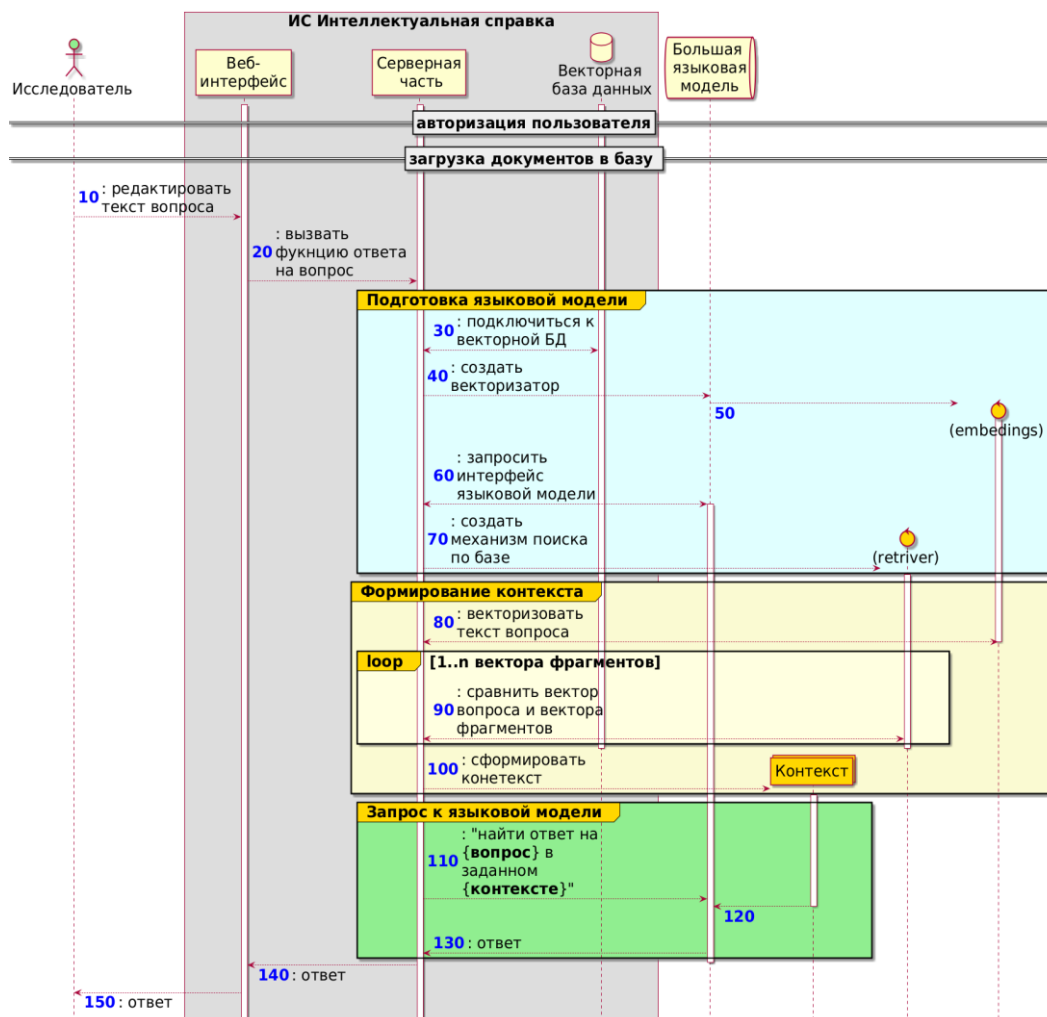


Рис. 6. UML-диаграмма последовательности: запросы пользователя к языковой модели

Для всех пользователей системы рабочим режимом по умолчанию является режим «вопрос-ответ». На первом шаге пользователь формирует в свободной форме вопрос и после подтверждения ввода получает ответ. Список заданных вопросов и полученных на них ответов сохраняется и доступен пользователю в течение активной сессии. Также пользователю предоставляется возможность оценить полученный ответ. Протокол сессии с вопросами, ответами и оценками в дальнейшем можно выгрузить и использовать для оценки качества работы системы.

Для обычных пользователей применяется преднастроенный администратором системы промпт запроса. Пользователям с более высоким уровнем привилегий доступна возможность корректировки промпта. Такая корректировка предполагает формулировку инструкции с использованием в тексте двух обязательных маркеров {context} и {question} (рис. 7). На следующем шаге эти маркеры автоматически заменяются соответственно: а) на контекст, т. е. извлеченный из векторной базы знаний

набор фрагментов; б) и на текст самого вопроса. Объем сформированного таким образом текста промпта не должен превышать допустимого моделью значения. Пример компоновки и наполнения интерфейса на этапе предоставления пользователю результатов запроса представлен на рис. 8.

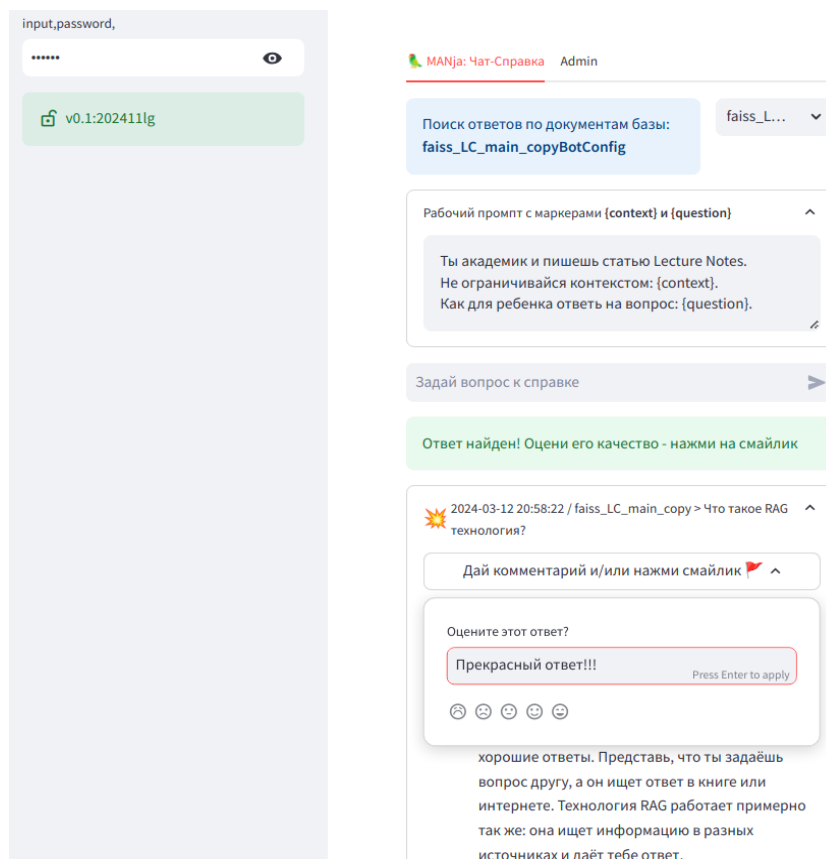


Рис. 7. Интерфейс пользователя: формирование промпта и оценка полученного ответа

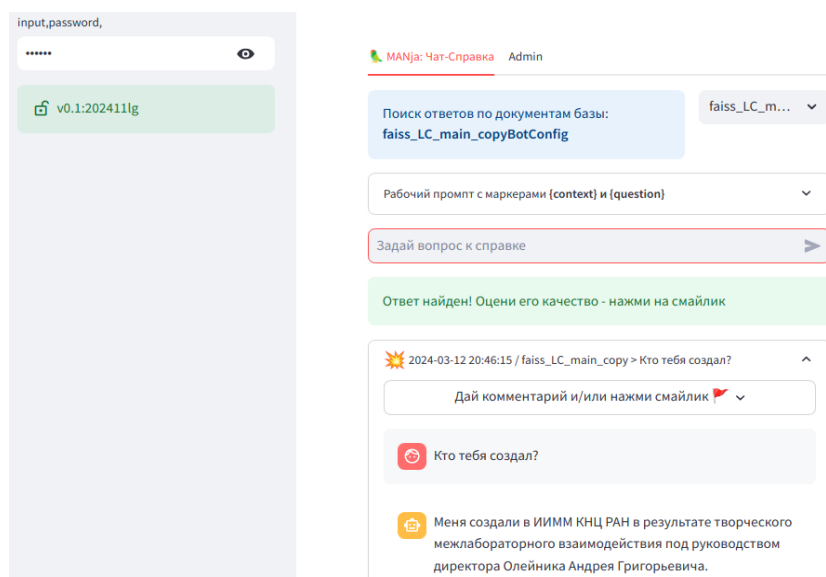


Рис. 8. Интерфейс пользователя: получение ответа на заданный вопрос

Разработка пользовательского интерфейса

Разработанная система построена в соответствии с классической веб-архитектурой и разделена на серверную (бэкенд) и пользовательскую (фронтенд) части. Это позволило эффективно управлять ограниченными ресурсами команды разработчиков, а также заложить в систему возможности по ее оперативному развертыванию, гибкой интеграции и адаптивной масштабируемости.

Использование веб-интерфейса позволяет системе работать в режиме онлайн и обеспечивает обращение к ней одновременно нескольких пользователей. Этот формат работы с системой необходим для решения целевых практических задач, например, при интеграции данной интеллектуальной справочной системы в сторонние предметно-ориентированные информационные системы. Примеры компоновки интерфейса представлены в предыдущем разделе и ниже на рис. 9–12.

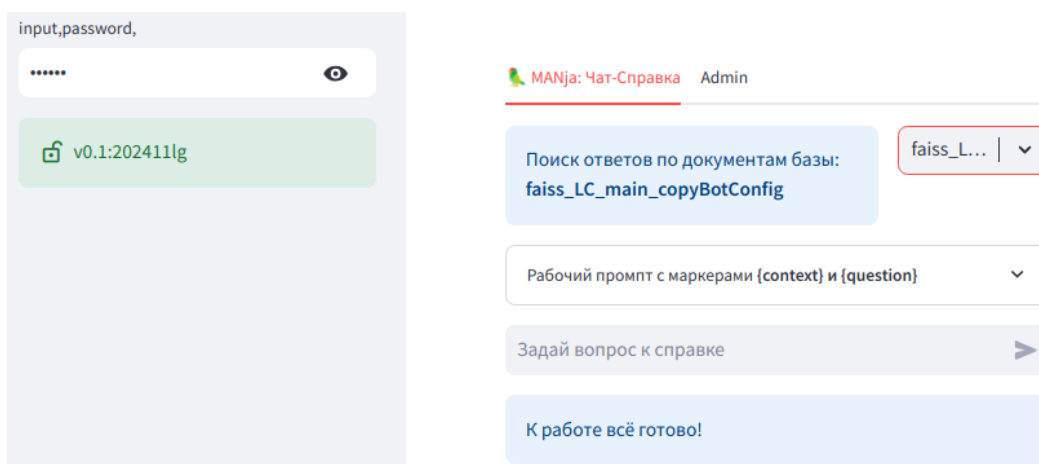


Рис. 9. Пользовательский интерфейс: начало работы

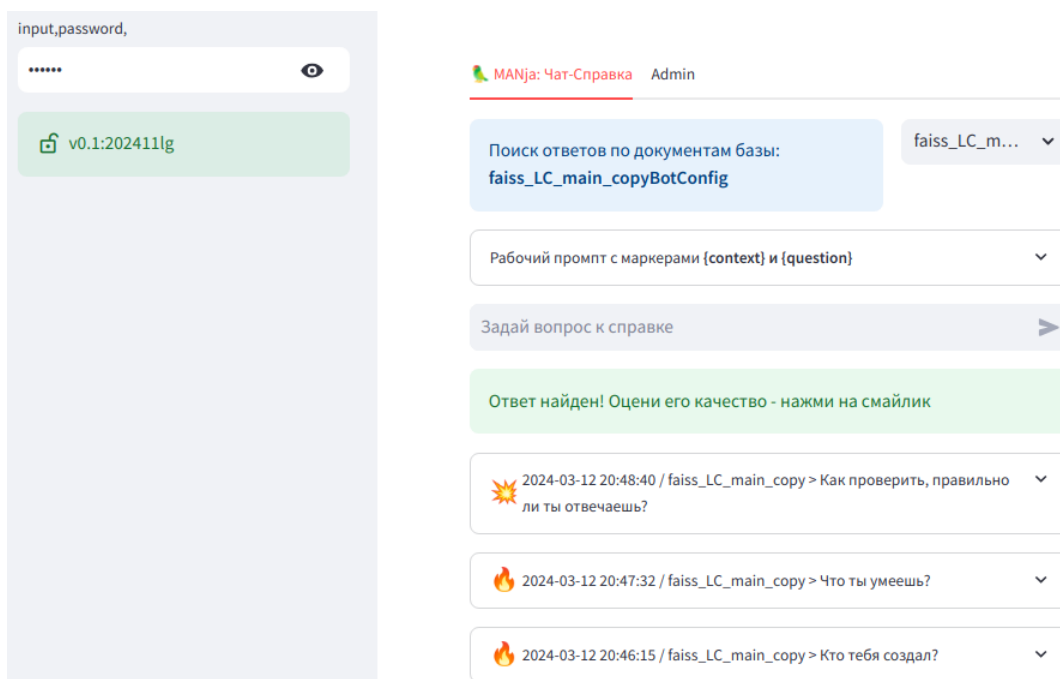


Рис. 10. Пользовательский интерфейс: режим работы «вопрос-ответ»

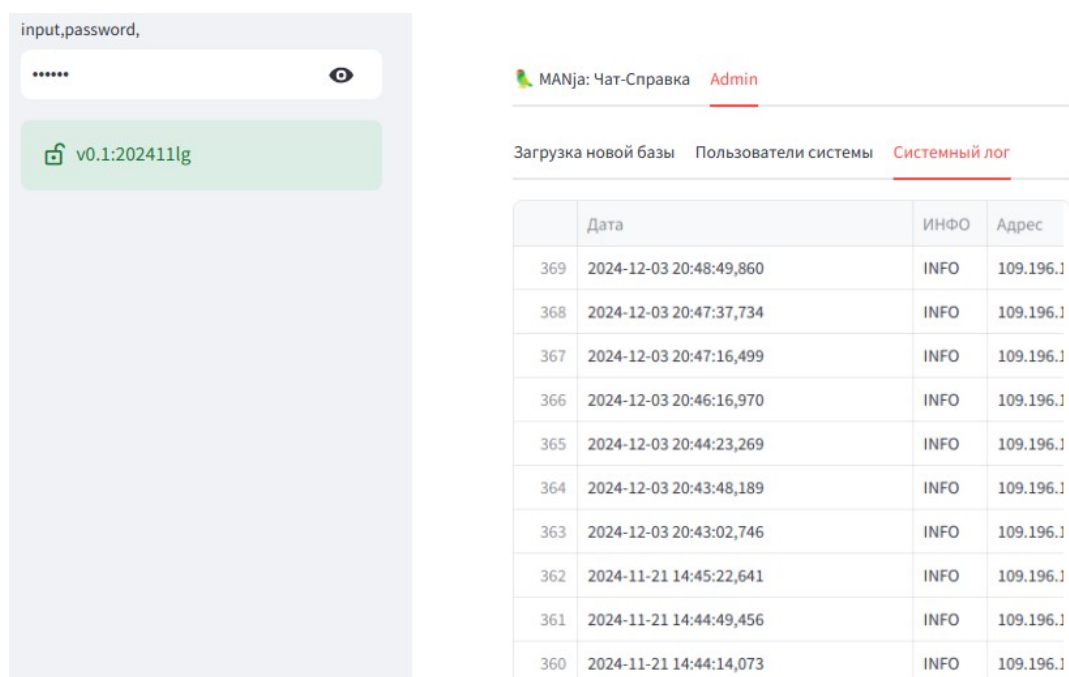


Рис. 11. Пользовательский интерфейс: панель администратора: системный лог

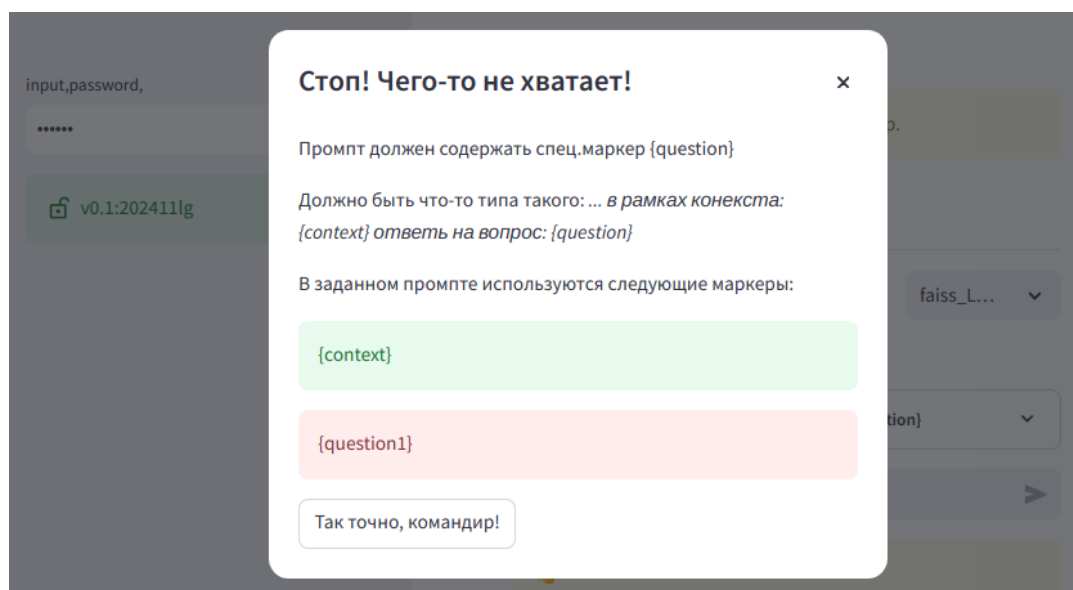


Рис. 12. Пользовательский интерфейс: пример сообщения об ошибке

Необходимо отметить, что большинство современных профессиональных веб-приложений строятся на основе архитектурного стиля REST API [27]. В них, в зависимости от концептуальных и технических требований, для реализации пользовательской части используют фреймворки преимущественно на основе языка JavaScript такие, как Angular, React.js и Vue.js [28]. В данной работе в первую очередь сделана ставка на модульность, возможность гибкой трансформации и адаптивной интеграции. Поэтому на начальных этапах разработки большее внимание уделено базовым функциям бизнес-логики, а пользовательский интерфейс реализован в виде прототипа на основе фреймворка Streamlit. Этот фреймворк, в отличие

от указанных выше и, по сути, признанных в веб-разработке стандартом де-факто, построен на языке Python и имеет открытый исходный код. Streamlit позволяет создавать быстрые прототипы динамических приложений, в том числе для работы с большими данными. На текущем этапе разработки применение данного фреймворка более чем оправдано, т. к. он широко используется исследователями, специалистами по обработке данных и инженерами в области искусственного интеллекта и машинного обучения. Это связано с тем, что самым популярным инструментом решения подобных задач является язык Python. Однако дальнейшее развитие данного проекта предусматривает плановый переход от прототипной реализации фронтенда к более профессиональному варианту на основе фреймворка Vue.js, успешный опыт применения которого был представлен авторами в работе [23].

Другим направлением является использование данной системы в качестве экспериментального стенда для проведения разнопланового тестирования и изучения эффективности применения больших языковых моделей на практике. Для удобства работы в этом режиме в системе дополнительно реализован альтернативный пользовательский интерфейс. Этот вариант интерфейса предоставляет пользователю возможность работать с системой как настольным приложением.

Пользовательский интерфейс настольного приложения реализован на основе кроссплатформенной библиотеки Tkinter. Этот вид интерфейса, по сравнению с веб-вариантом, использует менее выразительные средства (см. рисунки в разделе «Эксперименты»: рис. 13–15). Однако его преимуществом является возможность оперативного и непосредственного взаимодействия пользователя-оператора с программным кодом компонентов системы. Такой подход дает возможность максимально задействовать потенциал хорошо зарекомендовавшего себя в научных исследованиях языка программирования Python, находить новые варианты использования системы, оперативно тестировать экспериментальные решения. Полученные здесь результаты в дальнейшем используются для оптимизации, улучшения и расширения возможностей веб-версии системы.

Эксперименты

Для подготовки вариантов проблемно-ориентированного контекста, с которым должна работать система, реализован многоступенчатый процесс, состоящий из следующих этапов.

Загрузка и считывание документа. Для загрузки содержимого файла используется модуль *UnstructuredWordDocumentLoader* [29] из библиотеки *langchain_community.document_loaders*. Данный модуль имеет несколько режимов настройки загрузки документов.

Разбиение документа. Метод *langchain load_and_split* разделяет документ на части. Метод применяется для упрощения обработки, разбивая его на более управляемые блоки. В данном случае используется разделитель *RecursiveCharacterTextSplitter* [30] для разбиения документов. Разбиение документа дает возможность более простого анализа больших документов и особенно полезен для выделения содержимого таблиц и других структурированных данных.

Создание векторного представления. На этом этапе происходит работа методов класса *YandexGPTEmbeddings*, который отвечает за эмбединг, т. е. преобразование текста в векторные представления.

Создание векторного хранилища. При использовании библиотеки FAISS происходит создание векторного хранилища. На этом этапе обрабатываются фрагментированные документы и их эмбедингов, а также создается индекс для облегчения дальнейшего поиска по тексту.

Варьирование загружаемого контекста позволяет исследовать особенности генерации системой ответов в зависимости как от структуры контекста, так и формулировки запросов.

Анализ качества генерации ответов на пользовательские запросы

Для исследования работы чат-бота на основе модели YandexGPT 3 Lite был проведен эксперимент, в ходе которого пользователи производили запросы и ставили оценки «отлично», «средне», «плохо» полученным ответам. В качестве контекста было использовано руководство пользователя многофункциональной программной системы инженерного моделирования.

В рамках эксперимента был задан 81 запрос. Из всех запросов 31 % (25 запросов) получил оценку «отлично», 44 % (36 запросов) — «средне» и 25 % (20 запросов) — «плохо». Таким образом, в ходе исследования было получено 75 % положительных оценок, что свидетельствует о приемлемом уровне ответов в большинстве запросов к чат-боту. Разберем результаты более детально.

Оценка «отлично» и оценка «средне» для запросов к чат-боту наиболее часто ставилась для запросов с формулировкой вопроса с использованием конструкций: «Где...», «Что такое...», «Какие...», «Чем...».

Также было выявлено, что чат-бот хорошо работает с вопросами на общие темы, для ответа на которые достаточно «знаний» используемой LLM.

Однако в случае вопросов, которые требуют более детального анализа, например, «Что умеет эта программная система?» и «Что нужно для запуска расчета качества?», запросы чаще всего получали оценку «средне». Слабой стороной чат-бота оказались запросы, касающиеся инструкций, т. е. вопросы типа «Как добавить/удалить/изменить...», «Как работает/функционирует...», и «Как правильно...». Такие запросы чаще всего получали пользовательскую оценку «плохо». Разработчики чат-бота предполагают, что данный результат во многом обусловлен тем, что необходимая информация представлена в руководстве в виде изображений.

При оценке работы чат-бота анализировались комментарии пользователей относительно качества ответов на запросы. Среди замечаний к ответам на разные вопросы отмечался как избыток информации, не относящейся к делу, так и недостаточная информативность ответов. Также была выявлена проблема, обусловленная тем, что в различных разделах справки одинаковые по написанию термины могут нести разную смысловую нагрузку.

Проведенный анализ показал, что для повышения качества ответов чат-бота необходимо обеспечить его фокусировку на контексте запроса для более релевантных ответов, а также добавление подробностей в ответы на сложные вопросы для полного понимания пользователем информации по интересуемой процедуре или функции.

Эксперименты с табличными данными

При работе с нормативными и справочными документами для генерации ответа в качестве контекста нередко требуется использовать данные, размещенные во включенных в документы таблицах. Корректная обработка данных, представленных в таблицах, дает возможность более точно выбирать информацию, необходимую для генерации ответов, касающихся как структуры, так и «внутреннего содержания» ячеек таблиц.

Рассмотрим результаты работы чат-бота с таблицей на примере, представляющем фрагмент учебного плана (табл.).

На рис. 13 представлен результат тестового «диалога» пользователя с чат-ботом.

Таблица

Таблица для тестирования чат-бота

Раздел дисциплины	Семестр	Используемые активные/интерактивные образовательные технологии	Количество часов
Теория технологии и комплексной механизации	10	Дискуссионные методы. Проблемное обучение	2 л 4 пр
Технология и комплексная механизация при сплошных системах разработки		Дискуссионные методы. Проблемное обучение	6 пр
Технология и комплексная механизация при углубочных системах разработки		Проблемное обучение	4 л
Перспективное и текущее планирование горных работ. Качество продукции		Дискуссионные методы. Проблемное обучение	2 пр
Итого:			6 л 12 пр

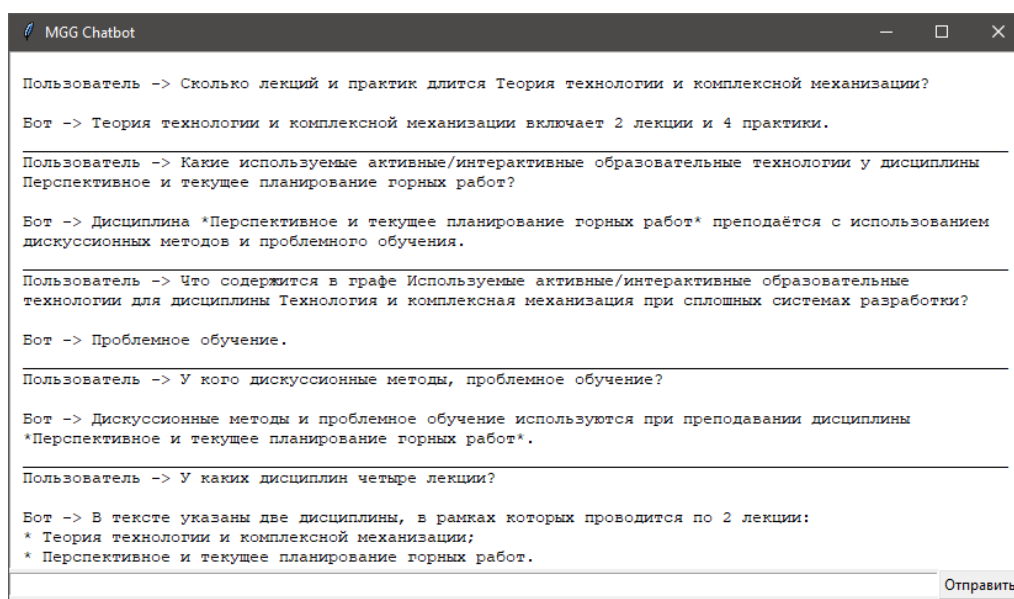


Рис. 13. Примеры запросов по табличным данным

Из представленного на рис. 13 примера видно, что в целом ответы чат-бота с использованием данных из таблицы являются корректными. Тем не менее имеются и неправильные ответы — в случае вопроса о дисциплинах с четырьмя лекциями бот указал на дисциплины, имеющие по две лекции. Выяснение причин появления некорректных ответов требует дополнительного исследования, в том числе и с точки зрения доработки методов интерпретации вопросов.

Эксперименты с анализом частей речи

Одним из направлений использования разрабатываемой системы планируется поддержка построения онтологий предметной области. Онтология, по сути, является концептуальной схемой, представляющей множество сущностей (понятий, объектов) предметной области и связей (отношений) между ними. В текстах на естественном языке сущности, как правило, представляются существительными, а вот для «описания» связи между ними могут использоваться различные части речи. Проведена серия экспериментов, связанных с оценкой возможностей системы идентифицировать и извлекать из текстов определенные части речи, в первую очередь существительных. Ниже представлены некоторые результаты этих экспериментов.

Рис. 14 демонстрирует, что чат-бот успешно идентифицирует различные части речи. Помимо существительных, прилагательных и глаголов, он также определяет более сложные грамматические категории такие, как причастия и деепричастия.

При ответе на запрос по идентификации частей речи (рис. 15) чат-бот приводит список слов, это означает, что система умеет выделять ключевые элементы текста. Тем не менее бот не упомянул такие слова, как «OPMPS» и «Pit», «Mine», «Scheduling», это указывает на наличие проблем в идентификации существительных на латинице и, возможно, аббревиатур.

Пример, представленный на рис. 16, показывает, что чат-бот может провести детальный разбор и выдать характеристику предложения, включая его тип, структуру и синтаксические компоненты. Он правильно определяет подлежащее, сказуемое и выделяет определения и дополнения.

По результатам проведенных экспериментов можно сделать вывод, что чат-бот выдает в большинстве случаев корректные структурированные ответы, демонстрирует способность работать с технической и специализированной лексикой, а также проводить комплексный анализ текста на естественном языке. Анализ текстов выполняется чат-ботом с достаточной степенью точности и детализации.

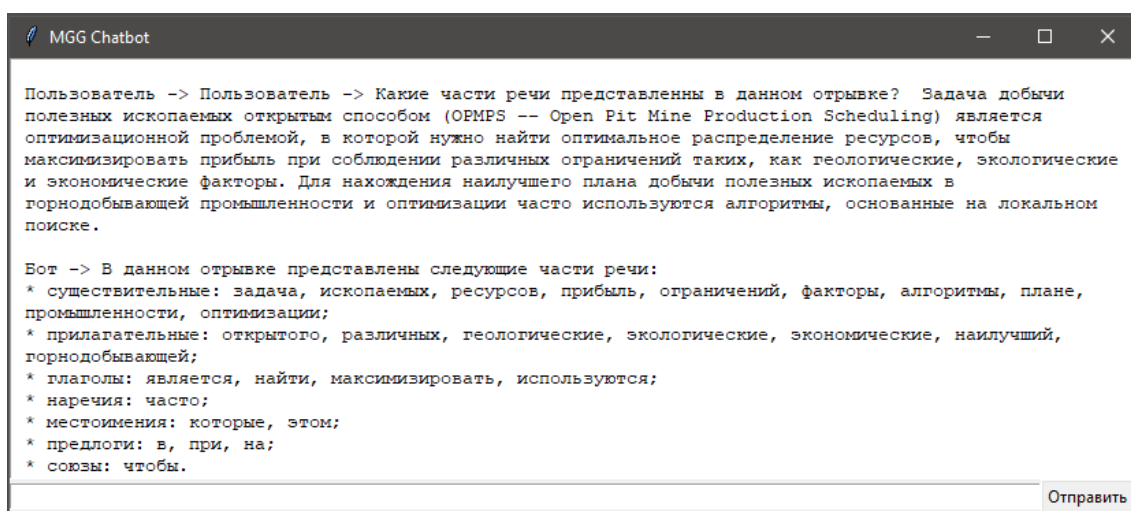


Рис. 14. Пример запроса по определению частей речи

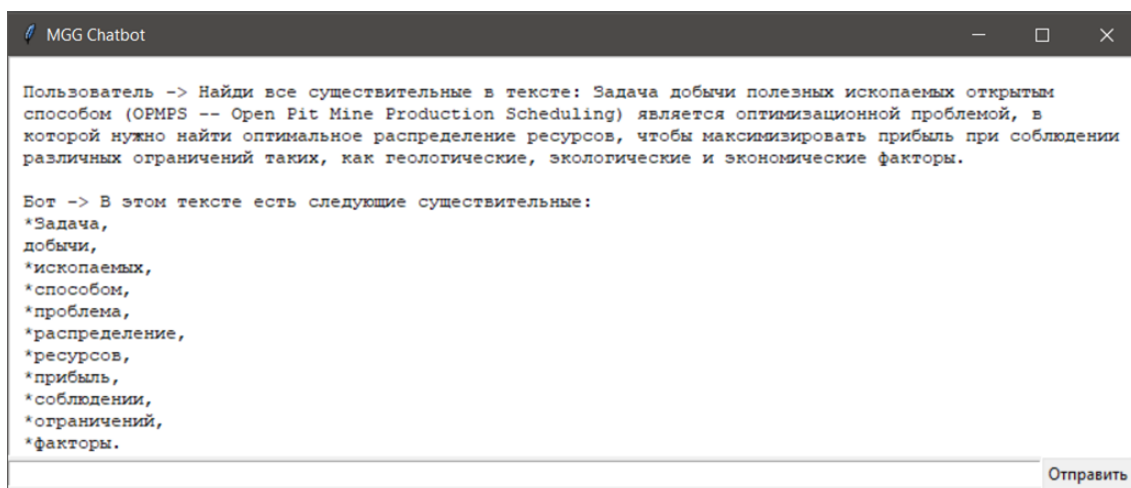


Рис. 15. Пример запроса по идентификации частей речи

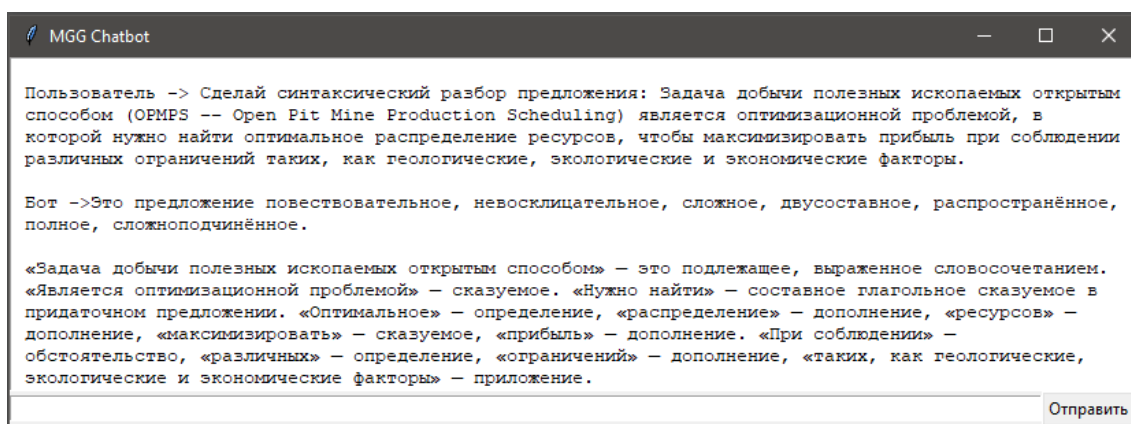


Рис. 16. Пример запроса по синтаксическому разбору предложения

Влияние на релевантность ответа на запрос дробления контекста на разделы

При использовании чат-бота возможны разные подходы к формированию контекста запроса, который напрямую влияет на ответ. Варьирование подготовки контекста позволяет исследовать влияние таких аспектов, как многозначность терминов, контекстуальная зависимость.

На рис. 17 представлены ответы на один и тот же вопрос с использованием разных разделов документации.

Различия в ответах на заданный вопрос обусловлены тем, что в разных разделах документа слово «Модель» интерпретируется различным образом.

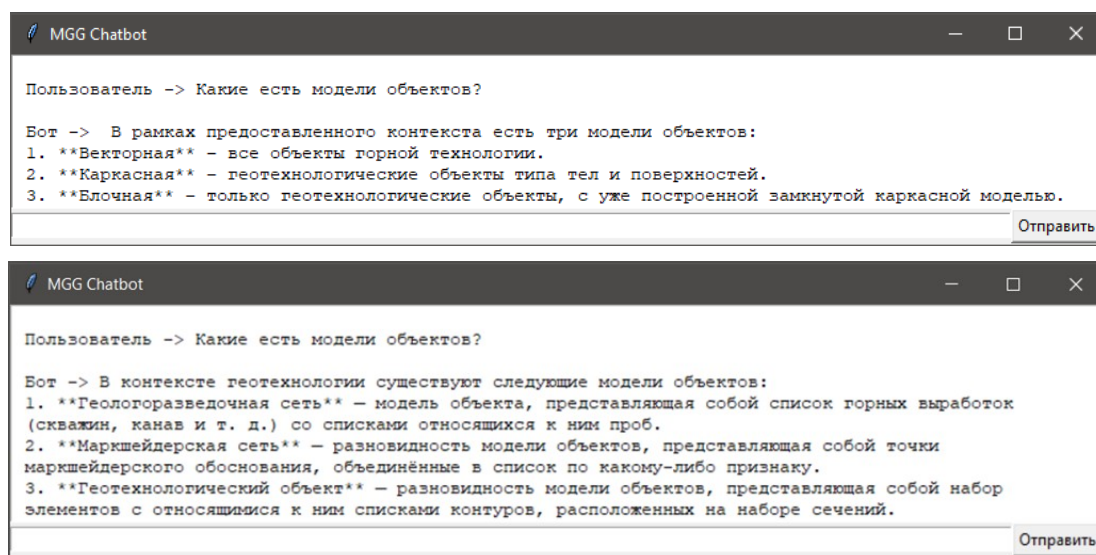


Рис. 17. Варианты ответов при использовании в качестве контекста разных разделов одного документа

Заключение

В публикациях, посвященных использованию больших языковых моделей, отмечается, что, наряду с достижениями LLM в области работы с текстами на естественном языке, имеет место и ряд проблем, не позволяющих полностью «доверять» результатам их работы. Более корректные результаты получаются в случае ограничения (конкретизации) контекста, с которым должна работать LLM в рамках определенной задачи. Обеспечить такую конкретизацию позволяет использование технологии RAG.

В данной статье представлены основные этапы проектирования и реализации интеллектуализированной системы работы с массивами документов. Созданный инструмент ориентирован на интеллектуальную поддержку решения практических задач как в автономном режиме, так и в режиме интеграции в сторонние предметно-ориентированные информационные системы.

Проведенные эксперименты по работе с системой носят предварительный характер. Но и они показали, что разрабатываемая система может не только выполнять функции «подсказчика» для пользователя документации, общаясь с ним на естественном языке, но и существенно упростить работу системных аналитиков и разработчиков онтологий. В рамках этих экспериментов были выявлены подходящие конфигурационные параметры, позволяющие использовать большие языковые модели для выделения в тексте целевых концептов, используемых для заселения онтологий.

Список источников

1. Постановление Правительства РФ от 14 августа 2020 г. № 1225 «Об утверждении Правил разработки критериев отнесения объектов всех форм собственности к критически важным объектам». [Электронный ресурс]. URL: <http://government.ru/docs/all/129416/> (дата обращения: 01.11.2024).

2. Олейник А. Г., Ломов П. А. Разработка онтологии интегрированного пространства знаний // *Онтология проектирования*. 2016. Т. 6, № 4(22). С. 465–474.
3. Ломов П. А., Олейник А. Г. Разработка технологии проверки и согласования нормативно-правовой базы на основе онтологий // *Труды Института системного анализа Российской академии наук*. 2013. Т. 63, № 2. С. 62–69.
4. Bystrov, V. V., Khaliullina, D. N., Malygina, S. N. (2024). Conceptual Modeling of the Resilience of Regional Socio-Economic Systems “Business-Society-Government”. In: Silhavy, R., Silhavy, P. (eds) *Software Engineering Methods in Systems and Network Systems. CoMeSySo 2023. Lecture Notes in Networks and Systems*, vol 934., pp. 179–191.
5. Датьев И. О., Федоров А. М. Аддитивная регуляризация при тематическом моделировании текстов сообществ онлайн-социальных сетей // *Онтология проектирования*. 2022. Т. 12, № 2(44). С. 186–199.
6. Пимешков В. К., Никонорова М. Л., Шишаев М. Г. Комбинированный метод извлечения терминов для задачи мониторинга тематических обсуждений в социальных медиа // *Информатика и автоматизация*. 2024. № 4(23). С. 1110–1138.
7. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population // *Software engineering perspectives in intelligent systems* / ed. Silhavy R., Silhavy P., Prokopova Z. Cham: Springer International Publishing, 2020. P. 919–926.
8. Lomov P., Malozemova M., Shishaev M. Extracting relations from NER-tagged sentences for ontology learning // *Artificial Intelligence Trends in Systems: Proceedings of 11th Computer Science On-line Conference 2022*. Vol. 2. Cham: Springer, 2022. P. 337–344.
9. Wang L. et al. Genre: Generative Multi-Turn Question Answering with Contrastive Learning for Entity–Relation Extraction // *Complex & Intelligent Systems*. 2024. Vol. 10. P. 3429–3443.
10. Gao Y. et al. Retrieval-Augmented Generation for Large Language Models: A Survey // *arXiv:2312.10997 [cs]*. 2023.
11. Yandex Foundation Models [Электронный ресурс]. URL: <https://yandex.cloud/ru/docs/foundation-models/> (дата обращения: 20.11.2024).
12. GigaChat API [Электронный ресурс]. URL: <https://developers.sber.ru/docs/ru/gigachat/api/overview> (дата обращения: 20.10.2024).
13. OpenAI developer platform [Электронный ресурс]. URL: <https://platform.openai.com/docs/overview> (дата обращения: 07.03.2024).
14. Manning C. D. Human Language Understanding & Reasoning // *Daedalus*. 2022. Vol. 151, № 2. P. 127–138.
15. Vaswani A. et al. Attention Is All You Need // *arXiv:1706.03762 [cs]*. 2017.
16. Zhang Y. et al. Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models // *arXiv:2309.01219 [cs]*. 2023.
17. Экономика больших языковых моделей: сколько стоит обучить и сколько — эксплуатировать. [Электронный ресурс]. URL: <https://telegra.ph/ENkonomika-bolshih-yazykovyh-modelej-skolko-stoit-obuchit-i-skolko--ehkspluatirovat-04-07> (дата обращения: 01.11.2024).
18. 30 миллиардов параметров: реально ли обучить русский GPT-3 в «домашних» условиях? [Электронный ресурс]. URL: <https://habr.com/ru/articles/565564/> (дата обращения: 01.11.2024).
19. Ma X. et al. Query Rewriting for Retrieval-Augmented Large Language Models // *arXiv:2305.14283 [cs]*. 2023.
20. Advanced RAG Techniques: an Illustrated Overview [Электронный ресурс]. URL: <https://pub.towardsai.net/advanced-rag-techniques-an-illustrated-overview-04d193d8fec6> (дата обращения: 19.10.2024).
21. Langchain introduction [Электронный ресурс]. URL: <https://python.langchain.com/docs/introduction/>
22. Федоров А. М., Датьев И. О., Вишняков И. Г. Проектирование информационной системы комплексного тематического анализа больших данных социальных медиа // *Онтология проектирования*. 2024. Т. 14, № 1(51). С. 55–70. doi:10.18287/2223-9537-2024-14-1-55-70.
23. Датьев И. О., Федоров А. М., Ревякин А. А. Фокусированный сбор и обработка открытых данных социальных медиа. *Онтология проектирования*. 2024. Т. 14, № 4(54). С. 569–581.

24. Информационная карта научно-исследовательской работы (НИР) «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер: 122022800551-0) [Электронный ресурс]. URL: <https://gisnauka.ru/nioktr/detail/SB2430UAGEM4J2MSBKAK1Q3C> (дата обращения: 01.11.2024).
25. Информационная карта научно-исследовательской работы (НИР) «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (регистрационный номер: 122022800547-3) [Электронный ресурс]. URL: <https://gisnauka.ru/nioktr/detail/IFE0D83R25SX5MYIYNDZQJDL> (дата обращения: 01.11.2024).
26. Streamlit: наиболее быстрый способ создания приложений для обработки данных и обмена ими [Электронный ресурс]. URL: <https://streamlit.io/> (дата обращения: 31.10.2024).
27. Prayogi A. A., Niswar M., Indrabayu and Rijal M. Design and Implementation of REST API for Academic Information System. A A Prayogi et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 875 012047 doi:10.1088/1757-899X/875/1/012047
28. Vue.js: прогрессивный JavaScript-фреймворк [Электронный ресурс]. URL: <https://ru.vuejs.org/> (дата обращения: 01.11.2024).
29. UnstructuredWordDocumentLoader [Электронный ресурс]. URL: https://python.langchain.com/api_reference/community/document_loaders/langchain_community.document_loaders.word_document.UnstructuredWordDocumentLoader.html (дата обращения: 31.10.2024).
30. RecursiveCharacterTextSplitter [Электронный ресурс]. URL: https://python.langchain.com/api_reference/community/document_loaders/langchain_community.document_loaders.word_document.UnstructuredWordDocumentLoader.html (дата обращения: 01.11.2024).

References

1. Postanovlenie Pravitel'stva RF ot 14 avgusta 2020 g. № 1225 "Ob utverzhdenii Pravil razrabotki kriteriev otneseniya ob"ektov vsekh form sobstvennosti k kriticheski vazhnym ob"ektam". [Resolution of the Government of the Russian Federation of August 14, 2020 no. 1225 “On Approval of the Rules for the Development of Criteria for Attributing Facilities of All Forms of Ownership to Critically Important Facilities”]. Available at: <http://government.ru/docs/all/129416/> (accessed 01.11.2024) (In Russ.).
2. Oleynik A. G., Lomov P. A. Razrabotka ontologii integrirovannogo prostranstva znaniy [Development of the ontology of integrated knowledge space]. Ontologiya proektirovaniya [Ontology of designing], 2016, Vol. 6, no. 4, pp. 465–474. (In Russ.).
3. Lomov P. A., Oleynik A. G. Razrabotka tekhnologii proverki i soglasovaniya normativno-pravovoy bazy na osnove ontologii [Development of the technology for verifying and coordinating the regulatory framework based on ontologies], Trudy Instituta sistemnogo analiza Rossiyskoy akademii nauk [Proceedings of the Institute of System Analysis of the Russian Academy of Sciences], 2013, Vol. 63, no. 2, pp. 62–69. (In Russ.).
4. Bystrov V. V., Khaliullina D. N., Malygina S. N. Conceptual Modeling of the Resilience of Regional Socio-Economic Systems “Business-Society-Government”. In: Silhavy, R., Silhavy, P. (eds) Software Engineering Methods in Systems and Network Systems. CoMeSySo. Lecture Notes in Networks and Systems, 2023, Vol. 934, pp. 179–191.
5. Datyev I. O., Fedorov A. M. Additivnaya regularizatsiya pri tematicheskoy modelirovaniy tekstov soobshchestv onlajnovykh social'nykh setej [Additive regularization for topic modeling of social media communities]. Ontologiya proektirovaniya [Ontology of designing], 2022, Vol. 12 no. 2, pp. 186–199. (In Russ.)
6. Pimeshkov V. K., Nikonorova M. L., Shishaev M. G. A Kombinirovannyj metod izvlecheniya terminov dlya zadachi monitoringa tematicheskikh obsuzhdenij v social'nykh media [Combined Term Extraction Method for the Problem of Monitoring Thematic Discussions in Social Media]. Informatika i avtomatizatsiya [Informatics and Automation], 2024, Vol. 23, no. 4, pp. 1110–1138. (In Russ.).

7. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population. Software engineering perspectives in intelligent systems. In: Silhavy R., Silhavy P., Prokopova Z. (eds). Software Engineering Perspectives in Intelligent Systems. CoMeSySo 2020 Advances in Intelligent Systems and Computing, 2020, Vol. 1295, Springer, Cham, pp. 919–926.
8. Lomov P., Malozemova M., Shishaev M. Extracting relations from NER-tagged sentences for ontology learning. Proceedings of 11th Computer Science Online Conference 2022, 2022, Vol. 2, Springer, Cham, pp. 337–344.
9. Wang L. et al. Genre: Generative Multi-Turn Question Answering with Contrastive Learning for Entity–Relation Extraction. Complex & Intelligent Systems, 2024, Vol. 10, pp. 3429–3443.
10. Gao Y. et al. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv:2312.10997 [cs], 2023.
11. Yandex Foundation Models. Available at: <https://yandex.cloud/ru/docs/foundation-models/> (accessed 20.11.2024).
12. GigaChat API. Available at: <https://developers.sber.ru/docs/ru/gigachat/api/overview> (accessed 20.10.2024).
13. OpenAI developer platform. Available at: <https://platform.openai.com/docs/overview> (accessed 07.10.2024).
14. Manning C. D. Human Language Understanding & Reasoning. Daedalus, 2022, vol. 151, no. 2, pp. 127–138.
15. Vaswani A. et al. Attention Is All You Need. arXiv:1706.03762 [cs], 2017.
16. Zhang Y. et al. Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models. arXiv:2309.01219 [cs], 2023.
17. Экономика bol'shih yazykovykh modelej: skol'ko stoit obuchit', i skol'ko — ekspluatirovat' [The economics of large language models: how much it costs to teach, and how much to exploit]. (In Russ.). Available at: <https://telega.ph/EHkonomika-bolshih-yazykovykh-modelej-skolko-stoit-obuchit-i-skolko-ehkspluatirovat-04-07> (accessed 01.11.2024)
18. 30 milliardov parametrov: real'no li obuchit' russkij GPT-3 v «domashnih» usloviyah? [30 billion parameters: is it really possible to teach Russian GPT-3 at home?] (In Russ.). Available at: <https://habr.com/ru/articles/565564/> (accessed 01.11.2024)
19. Ma X. et al. Query Rewriting for Retrieval-Augmented Large Language Models. arXiv: 2305.14283 [cs], 2023.
20. Advanced RAG Techniques: an Illustrated Overview. Available at: <https://pub.towardsai.net/advanced-rag-techniques-an-illustrated-overview-04d193d8fec6> (accessed 19.10.2024).
21. Langchain introduction. Available at: <https://python.langchain.com/docs/introduction/> (accessed 19.10.2024).
22. Fedorov A. M., Datyev I. O., Vishnyakov I. G. Proektirovanie informacionnoj sistemy kompleksnogo tematicheskogo analiza bol'shih dannyh social'nyh media [Designing an information system for integrated topic analysis of social media big data]. Ontologiya proektirovaniya [Ontology of designing], 2024, Vol. 14, no. 1, pp. 55–70. (In Russ.).
23. Datyev I. O., Fedorov A. M., Reviakin A. A. Fokusirovannyj sbor i obrabotka otkrytyh dannyh social'nyh media [Focused collection and processing of open social media data]. Ontologiya proektirovaniya [Ontology of designing], 2024, Vol. 14, no. 4, pp. 569–581. (In Russ.).
24. Informacionnaya karta nauchno-issledovatel'skoj raboty (NIR) «Metodologiya sozdaniya informacionno-analiticheskikh sistem podderzhki upravleniya regional'nym razvitiem, osnovannyh na formiruyushchem iskusstvennom intellekte i bol'shih dannyh» (Registracionnyj nomer: 122022800551-0) [Information card of scientific research "Methodology for creating information and analytical systems to support regional development management based on formative artificial intelligence and big data" (Registration number: 122022800551-0)]. (In Russ.). Available at: <https://gisnauka.ru/nioktr/detail/SB2430UAGEM4J2MSB KAK1Q3C> (accessed 01.11.2024).
25. Informacionnaya karta nauchno-issledovatel'skoj raboty (NIR) «Razrabotka teoreticheskikh i organizacionno-tekhnicheskikh osnov informacionnoj podderzhki upravleniya zhiznesposobnost'yu regional'nyh kriticheskikh infrastruktur Arkticheskoy zony Rossijskoj Federacii» (Registracionnyj nomer: 122022800547-3) [Information card of scientific research "Development of theoretical and organizational and technical foundations of information support for the management of the viability of regional critical infrastructures of the Arctic zone of the Russian Federation" (Registration number: 122022800547-3)]. (In Russ.). Available at: <https://gisnauka.ru/nioktr/detail/IFE0D83R25SX5MYTYNDZQJDL> (accessed 01.11.2024).

26. Streamlit: naibolee bystryj sposob sozdaniya prilozhenij dlya obrabotki dannyh i obmena imi [Streamlit: a faster way to build and share data apps]. (In Russ.). Available at: <https://streamlit.io/> (accessed 31.10.2024).
27. Prayogi A. A., Niswar M., Indrabayu, and Rijal M., Design and Implementation of REST API for Academic Information System, IOP Conf Ser Mater Sci Eng, 2020, Vol. 875, no. 1, P. 012047.
28. Vue.js: The Progressive JavaScript Framework. Available at: <https://vuejs.org/> (accessed 01.11.2024).
29. UnstructuredWordDocumentLoader. Available at: <https://python.langchain.com/apireference/community/documentloaders/langchaincommunity.documentloaders.worddocument.UnstructuredWordDocumentLoader.html> (accessed 31.10.2024).
30. RecursiveCharacterTextSplitter. Available at: <https://python.langchain.com/apireference/community/documentloaders/langchaincommunity.documentloaders.worddocument.UnstructuredWordDocumentLoader.html> (accessed 01.11.2024).

Информация об авторах

А. Г. Олейник — доктор технических наук, директор, главный научный сотрудник;
А. М. Федоров — кандидат технических наук, ведущий научный сотрудник;
И. О. Датьев — кандидат технических наук, старший научный сотрудник;
А. А. Зуенко — кандидат технических наук, ведущий научный сотрудник;
А. В. Шестаков — аспирант, стажер-исследователь;
И. Г. Вишняков — аспирант, системный администратор.

Information about the authors

A. G. Oleynik — Doctor of Science (Tech.), Director, Chief Scientific Officer;
A. M. Fedorov — Candidate of Science (Tech.), Leading Scientific Officer;
I. O. Datyev — Candidate of Science (Tech.), Senior Scientific Officer;
A. A. Zuenko — Candidate of Science (Tech.), Leading Scientific Officer;
A. V. Shestakov — Graduate Student, Intern Researcher;
I. G. Vishnyakov — Graduate Student, System Administrator.

Статья поступила в редакцию 16.10.2024; одобрена после рецензирования 01.11.2024; принята к публикации 07.11.2024.
The article was submitted 16.10.2024; approved after reviewing 01.11.2024; accepted for publication 07.11.2024

Научная статья
УДК 004.89; 004.94
doi:10.37614/2949.1215.2024.15.3.002

ПОТЕНЦИАЛ ПРИМЕНЕНИЯ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ЖИЗНЕСПОСОБНОСТИ РЕГИОНАЛЬНЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМ

Виталий Викторович Быстров^{1✉}, Дарья Николаевна Халиуллина², Максим Геннадьевич Шишаев³

^{1–3}*Институт информатики и математического моделирования имени В. А. Путилова*

Кольского научного центра Российской академии наук, Апатиты, Россия

¹*v.bystrov@ksc.ru[✉], <https://orcid.org/0000-0002-9369-8457>*

²*d.khaliullina@ksc.ru, <https://orcid.org/0000-0002-7273-0649>*

³*m.shishaev@ksc.ru, <https://orcid.org/0000-0001-7070-7878>*

Аннотация

Минимизация участия человека при построении имитационных моделей является актуальной задачей, решение которой востребовано в практике моделирования жизненного цикла сложных динамических систем. Одним из способов ее решения является адаптация существующих решений в области генеративного искусственного интеллекта под специфику разработки компьютерных моделей сложных систем. В статье рассматриваются основные направления автоматизации имитационного моделирования с помощью современных технологий искусственного интеллекта. Авторами предлагаются концепции применения больших языковых моделей для генерации сценариев имитационного моделирования жизнеспособности региональных социально-экономических систем. Рассматриваются теоретические и практические вопросы реализации предложенных решений в современных средах имитационного моделирования. Приводится описание обобщенного алгоритма оценки жизнеспособности региональных систем «Бизнес-Сообщество-Власть», построенного на основе интеграции имитационного моделирования и генеративного искусственного интеллекта.

Ключевые слова:

имитационное моделирование, генеративный искусственный интеллект, большая языковая модель, жизнеспособность, социально-экономическая система

Благодарности:

исследование выполнено в рамках государственного задания Института информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук от Министерства науки и высшего образования Российской Федерации, тема научно-исследовательской работы — «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер 122022800551-0).

Для цитирования:

Быстров В. В., Халиуллина Д. Н., Шишаев М. Г. Потенциал применения генеративного искусственного интеллекта для имитационного моделирования жизнеспособности региональных социально-экономических систем // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 27–40. doi:10.37614/2949.1215.2024.15.3.002.

Original article

THE POTENTIAL OF USING GENERATIVE ARTIFICIAL INTELLIGENCE TO SIMULATE THE RESILIENCE OF REGIONAL SOCIO-ECONOMIC SYSTEMS

Vitaliy V. Bystrov^{1✉}, Darya N. Khaliullina², Maxim G. Shishaev³

^{1–3}*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia*

¹*v.bystrov@ksc.ru[✉], <https://orcid.org/0000-0002-9369-8457>*

²*d.khaliullina@ksc.ru, <https://orcid.org/0000-0002-7273-0649>*

³*m.shishaev@ksc.ru, <https://orcid.org/0000-0001-7070-7878>*

Abstract

Minimization of human involvement in the construction of simulation models is an urgent task, the solution of which is in demand in the practice of modeling the life cycle of complex dynamic systems. One of the ways to solve it is to adapt existing solutions in the field of generative artificial intelligence to the specifics of developing computer models of complex systems. The article discusses the main areas of automation of simulation modeling using

modern artificial intelligence technologies. The authors propose concepts for using large-scale language models to generate scenarios for simulation modeling of the resilience of regional socio-economic systems. Theoretical and practical issues of implementing the proposed solutions in modern simulation environments are considered. A description of a generalized algorithm for assessing the resilience of regional systems "Business-Community-Government" is given, built on the basis of integrating simulation modeling and generative artificial intelligence.

Keywords:

simulation modeling, generative artificial intelligence, large-scale language model, resilience, socio-economic system

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic "Methodology for creating information and analytical systems to support the management of regional development based on formative artificial intelligence and big data" (registration number of the research topic 122022800551-0).

For citation:

Bystrov V. V., Khaliullina D. N., Shishaev M. G. The potential of using generative artificial intelligence to simulate the resilience of regional socio-economic systems // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 27–40. doi:10.37614/2949.1215.2024.15.3.002.

Введение

Имитационное моделирование является одним из популярных методов исследования сложных систем. Данный инструмент рекомендуется использовать в условиях неопределенности (слабая структурированность объекта исследования и/или отсутствие полных данных о его функционировании), а также в случаях, когда эксперименты на реальном объекте в принципе не возможны либо сопровождаются большими ресурсными затратами (времени, финансов, персонала и т. д.). Одним из недостатков имитационного моделирования принято считать то, что качество создаваемой модели напрямую зависит от профессионализма и опыта ее разработчика. Для нивелирования указанного недостатка активно развивается научно-практическое направление, ориентированное на автоматизацию процессов построения имитационной модели, в том числе и на снижение влияния человеческого фактора на результат разработки.

В рамках направления по автоматизации процесса разработки имитационной модели ведутся разнообразные исследования, начиная от формирования исходных данных для настройки модели [1, 2] и заканчивая автоматической генерацией структуры модели в терминах используемой среды компьютерного моделирования [3–7]. В разное время в Институте информатики и математического моделирования им. В. А. Путилова ФИЦ КНЦ РАН также были получены результаты, внесшие вклад в развитие данного направления. В частности, были разработаны технологии автоматизированного синтеза системно-динамической модели на основе концептуальных шаблонов [8] и онтологических описаний [9], технология дистанционного формирования и управления моделями системной динамики [10].

Одним из перспективных и активно развивающихся направлений в области автоматизации имитационного моделирования является применение искусственного интеллекта. В рамках текущей статьи выделяются основные способы интеграции генеративного искусственного интеллекта (ГИИ) с имитационным моделированием, которые, по мнению авторов публикации, могут упростить работу разработчиков компьютерной модели. Также ниже приводится описание концепции применения ГИИ для сценарного моделирования жизнеспособности (resilience) региональных социально-экономических систем. В частности, рассматривается задача автоматизированной генерации сценариев кризисных ситуаций для имитационного моделирования жизнеспособности региональных систем «Бизнес-Сообщество-Власть» (БСВ-систем).

Направления применения генеративного искусственного интеллекта для имитационного моделирования

Проведенный анализ открытых источников информации показал, что выявленные исследования в области применения генеративного искусственного интеллекта совместно с имитационным моделированием можно условно разбить на следующие направления.

Генерация данных для имитационной модели.

В данном случае ГИИ выступает в качестве внешнего программного блока по отношению к модели, основной задачей которого является формирование отдельных значений и временных рядов для определенного набора модельных параметров. Обычно такой подход применяется в условиях отсутствия реальных данных (или их неполноты) о функционировании объекта моделирования. В качестве примера можно привести работу по моделированию кадрового потенциала региона [11], в которой многослойная нейронная сеть применяется для прогнозирования одного из параметров имитационной модели — объема произведенной продукции. Отметим, что может решаться и обратная задача — генерация данных для формирования обучающей выборки нейронной сети на основе результатов вычислительного эксперимента (например, генерация обучающей выборки на основе онтологий и имитационного моделирования [12]).

Интерпретация результатов имитационного моделирования.

В рамках данного направления пытаются анализировать результаты вычислительных экспериментов с помощью генеративного искусственного интеллекта. То есть ГИИ применяется после имитации с целью получения текстового описания результатов на естественном языке, повышая тем самым их интерпретируемость. Например, в работе [13] рассматривается опыт применения больших языковых моделей (LLM) для формирования текстового описания результатов, представленных в виде временных графиков, полученных при симуляции экономического роста страны в условиях пандемии. При этом промпт (текстовый запрос) к LLM расширяют за счет добавления графического файла с результатами моделирования.

Формирование поведения элементов модели.

Данное направление ориентировано на использование генеративного искусственного интеллекта внутри имитационной модели, когда с помощью ГИИ определяется реакция определенных компонентов модели на внешние возмущения. Например, в последних версиях инструментальной среды Anylogic реализована возможность создавать модельные агенты, которые по своей сути являются конвейерами машинного обучения (AutoML) и позволяют получать прогнозные значения параметров агента в ходе имитации [14]. Хотя AutoML в общем случае не относится к ГИИ, при современных тенденциях развития интеллектуальных технологий можно ожидать появления в инструментальных средствах моделирования специальных компонентов, реализованных в виде GAN (Generative Adversarial Network), отвечающих за генерацию линий поведения модельных сущностей.

Генерация имитационной модели или ее структуры.

В рамках данного направления предпринимаются попытки автоматизировать работу разработчика по созданию структуры модели и/или ее программной реализации. Например, в работе [13] применяют большую языковую модель для построения концептуальной модели в виде диаграммы причинно-следственных связей на основе текстового описания постановки задачи моделирования. В исследовании [15] с помощью LLM (ChatGPT) пытаются получить исполнимый код на языке Python, являющийся программной реализацией имитационной модели простой логистической системы. Схема генерации компьютерной модели (рис. 1) включает в себя формирование расширенного запроса к большой языковой модели на основе текстового описания задачи и контекста. При этом к запросу добавляются параметры модели. Эксперименты, представленные в указанной работе, показали, что идея применения больших языковых моделей для создания имитационных моделей жизнеспособна, но опробована она на довольно простых системах. По мере усложнения объекта моделирования трудно будет полностью исключить человека из процесса разработки модели.

Генерация сценариев.

Данное направление ориентировано на определение сценариев развития ситуации в изучаемой системе. При этом сценарии могут определяться разным способом: последовательностью событий, последовательностью состояний системы, текстом в свободной форме и т. д. Например, в работе [16] предлагается применять специализированную GAN для генерации сценария развития чрезвычайной ситуации (стихийного бедствия) в виде 3D-анимации.

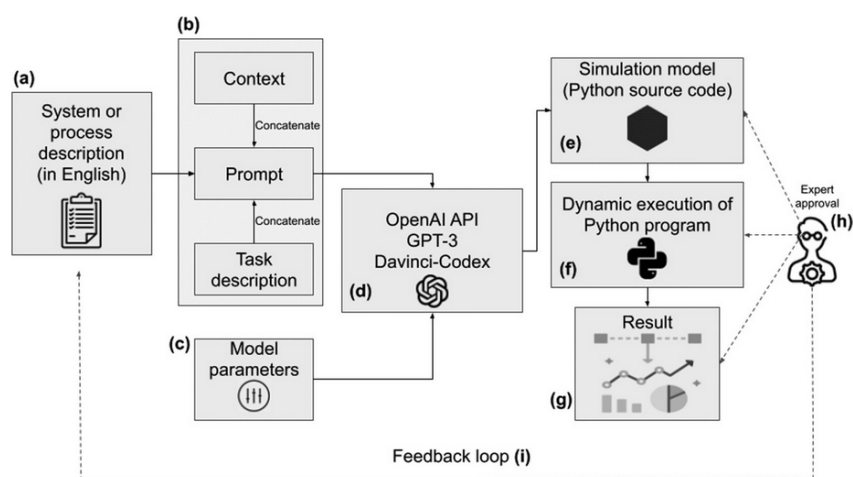


Рис. 1. Генерация имитационной модели логистической системы на основе ее вербального описания [15]

Анализ существующего опыта в области интеграции имитационного моделирования и генеративного искусственного интеллекта позволяет предложить два подхода к применению ГИИ для генерации сценариев при компьютерном моделировании жизнеспособности региональных БСВ-систем.

Генерация сценариев для имитационной модели

Последние несколько лет в Институте информатики и математического моделирования ведутся исследования в области разработки методических и программных средств информационно-аналитической поддержки управления жизнеспособностью (resilience) региональных социально-экономических систем. Одной из ключевых задач проводимого исследования является разработка комплекса имитационных моделей жизнеспособности систем «Бизнес-Сообщество-Власть». Разрабатываемые компьютерные модели используются для оценки жизнеспособности системы в различных сценарных условиях с целью выявления потенциальных кризисных ситуаций.

Традиционно в имитационном моделировании задача определения количества изучаемых сценариев и содержания каждого сценария ложится на плечи разработчика модели. При этом существует два взгляда на понятие сценария моделирования [17]. В первом случае, характерном для точных и технических наук, под сценарием понимается последовательность выполнения действий (операций), описывающая один из вариантов функционирования системы с известной архитектурой (структурой). Вторая точка зрения на понятие сценария применяется в некоторых отраслях естественно-научного знания и утверждает, что сценарий — это некоторая модель (прообраз) системы, который позволяет изучать способ достижения цели с учетом факторов влияния среды. В рамках текущего исследования авторы больше придерживаются первой точки зрения на сценарий для имитационного моделирования, т. к. разработка практических моделей и реализация на их основе сценариев предполагает оперирование конкретными модельными параметрами и их значениями с привязкой их к временной шкале.

В ходе изучения возможностей современных средств генеративного искусственного интеллекта возникла идея об автоматизации деятельности разработчика имитационных моделей за счет привлечения больших языковых моделей (LLM) для создания разнообразных сценариев развития исследуемой системы. Фундаментальной гипотезой применения ГИИ к решению подобного рода задач является то, что LLM может сгенерировать неочевидные (нетиповые) сценарии развития системы, которые разработчик модели (исследователь) мог бы пропустить. Предпосылкой того, что данная гипотеза справедлива, выступает тот факт, что в настоящее время LLM является инструментом, способным аккумулировать факты и знания, представленные в объемных корпусах текстов разной структуры, и синтезировать новые составные объекты (тексты). Сценарий также может быть представлен в виде текстового описания на естественном языке, а значит, его можно попробовать сгенерировать с помощью большой языковой модели.

На рис. 2 представлена концептуальная схема автоматизированного создания сценариев для имитационного моделирования жизнеспособности БСВ-систем. В предложенной авторами концепции генеративный искусственный интеллект выступает в качестве генератора текстовых описаний сценариев развития БСВ-системы. При этом текстовые описания формируются на естественном языке на основе промпта эксперта. Например, в качестве запроса к нейронной сети разработчик может попросить сформулировать описание кризисной ситуации для региональной системы, приведя краткое описание текущего положения дел.

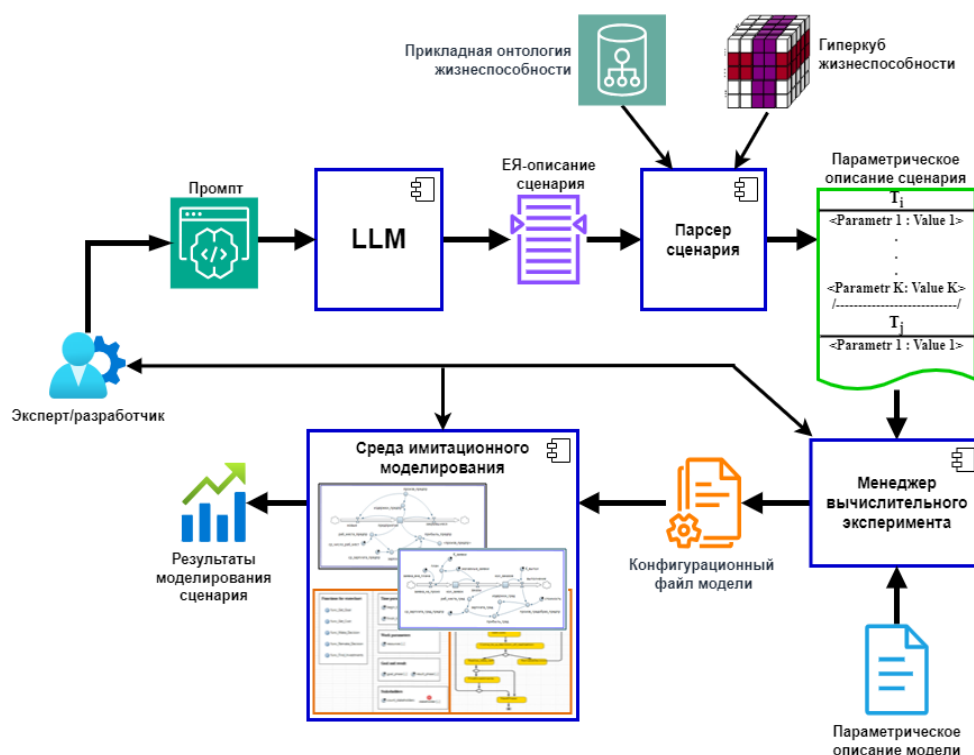


Рис. 2. Концепция генерации сценариев моделирования с помощью большой языковой модели

Согласно приведенной схеме (рис. 2) большая языковая модель на основе промпта разработчика формирует ответ на естественном языке, являющийся описанием сценария развития системы (ЕЯ-описание сценария).

Следующим шагом является формирование из ЕЯ-описания сценария его параметрическое представление. В текущей работе под параметрическим представлением сценария понимается совокупность характеристик/параметров БСВ-системы и ее жизнеспособности, описывающие изменение состояния системы во времени. Одним из способов формального описания параметризованного сценария является представление его в следующем виде:

$$Sc_i = \{ \langle t_j; \overline{Par} \rangle, i = \overline{1, N_{Sc}}, j = \overline{0, K},$$

где N_{Sc} — количество исследуемых сценариев; K — количество временных точек; \overline{Par} — вектор, определяющий состояние БСВ-системы в момент времени t_j .

Элементы вектора \overline{Par} , в свою очередь, можно формально представить в виде пар вида:

$$\langle Par_z(t_j), Value_z(t_j) \rangle,$$

где $z = \overline{1, N_{Par}}$.

Стоит отметить, что количество временных точек, в которых описывается состояние системы, может определяться разным способом. Например, в соответствии с определенной частотой дискретизации либо при возникновении определенных событий, которые приводят к изменению параметров системы.

Задача формирования параметрического описания сценария из его естественно-языкового представления решается специальным программным модулем — парсером сценария. Данный модуль с помощью методов и средств обработки естественного языка (Natural Language Processing, NLP) [18, 19] извлекает именованные сущности, факты, события и данные из сгенерированного текстового описания, соотносит их с информацией, представленной в прикладной онтологии жизнеспособности, и, при необходимости, дополняет ее данными, содержащимися в гиперкубе жизнеспособности БСВ-системы [20].

Сформированное параметрическое описание сценария передается программному компоненту «Менеджер вычислительного эксперимента». Его основной задачей является организовать имитацию в исполнительной среде поданного на вход сценария. Для автоматизации данного процесса предлагается сопоставлять параметрическое описание сценария с параметрическим описанием имитационной модели жизнеспособности БСВ-системы с целью перевода параметров системы, представленных в сценарии, в модельные параметры. Для разрешения возникающих сложностей при трансформации параметров на первоначальных этапах необходимо привлекать эксперта. Однако, с накоплением знаний об особенностях практической реализации данного процесса можно попробовать перенести их в системную онтологию и на ее основе алгоритмизировать процедуру представления сценария в имитационной модели. Результатом функционирования менеджера вычислительного эксперимента является генерация конфигурационного файла для имитационной модели, который позволяет инициализировать модель в исполнительной среде.

Запуск вычислительного эксперимента в среде имитационного моделирования вызывается с помощью команд программного интерфейса (API). Например, для среды Anylogic это можно реализовать с помощью команды «POST /versions/<идентификатор версии>/runs» через REST API, где версия — это идентификатор запускаемой модели. Из-за особенностей организации взаимодействия имитационных моделей, реализованных в Anylogic, со сторонним программным обеспечением для реализации возможности инициализировать модель в соответствии с заданными параметрами сценария необходимо саму модель разрабатывать специфичным образом. Для этого нужно сделать так, чтобы при запуске модели значения ее параметров считывались из внешнего источника информации (электронной таблицы или базы данных), который, по сути, и будет выступать в роли конфигурационного файла. Исполнительная среда проводит вычислительный эксперимент над инициализированной моделью. Результаты имитации предоставляются эксперту и/или разработчику модели для дальнейшего анализа.

Проведенные эксперименты по генерации потенциальных сценариев кризисных ситуаций, которые могут возникнуть в арктическом моногороде (как примере региональной БСВ-системы), с помощью больших языковых моделей (GigaChat от Сбера и ChatGPT от OpenAI) выявили ряд проблем предложенной концепции. Во-первых, вероятность получения подробного описания сценария развития кризисной ситуации без дополнительной информации только на основе запроса эксперта очень низкая. В ходе общения с экспертом LLM пыталась обобщить доступную ей информацию и привести типизацию кризисных ситуаций, которые возможны в региональной БСВ-системе. Суммаризация нейросетью типов кризисных ситуаций и причин их возникновения обладала довольно высокой степенью правдоподобия. Во-вторых, в сгенерированном сценарии могут содержаться параметры, которые не представлены ни в прикладной онтологии жизнеспособности, ни в разработанной имитационной модели. Это потребует реализацию специальных процедур для дополнения указанных базы знаний и модели.

Для устранения первой выявленной проблемы предлагается воспользоваться подходом расширения запроса к большой языковой модели посредством добавления контекста — RAG (Retrieval Augmented Generation) [21]. Основная идея данной технологии заключается в повышении релевантности ответа LLM за счет формирования расширенного запроса на основе соединения

обычного промпта и специальным образом подготовленной дополнительной информации, описывающей проблемно-ориентированный контекст запроса. В текущем исследовании в рамках дополнительного контекста предлагается использовать описание определенной ситуации, которая формируется на основе имитационного моделирования жизнеспособности БСВ-системы.

Основополагающей идеей предоставления на вход большой языковой модели расширенного запроса, содержащего описание определенной ситуации, сложившейся к определенному моменту времени в региональной социально-экономической системе, является то, что в нейронной сети отражен накопленный опыт человечества из разных областей знаний. Значит, с теоретической точки зрения, LLM способна дополнить ограниченный взгляд одного отдельного взятого эксперта на определенные процессы и закономерности. Создатель имитационной модели при ее разработке выступает в роли эксперта и транслирует определенные экспертные знания в модельный язык, закладывая в модель структуру и закономерности, ориентированные на имитацию определенного набора сценариев. При этом он неявно ограничивает множество сценариев, которые способна воспроизводить разрабатываемая модель вследствие своих ограниченных знаний об объекте моделирования и его взаимодействии с окружающей средой.

На рис. 3 приведен график изменения определенного параметра модели (в данном случае жизнеспособности системы R_s) во времени при имитации разных сценариев. Сценарий Sc_b^M является базовым, он закладывается в имитационную модель и характеризует нормальный режим функционирования системы. Множество сценариев $\{Sc_1^M, \dots, Sc_k^M\}$ является набором сценариев, которые потенциально могут воспроизводиться в модели, т. к. это предусмотрел разработчик модели. Сценарий Sc_1^G — это искусственно созданный сценарий развития системы, который не предусматривался разработчиком, но который может появиться, когда исследуемая система BSG будет находиться в определенном состоянии $S(t_p)$ в момент времени t_p . В формировании таких искусственно созданных сценариев может помочь диалог с большой языковой моделью, организованный с использованием RAG-технологии.

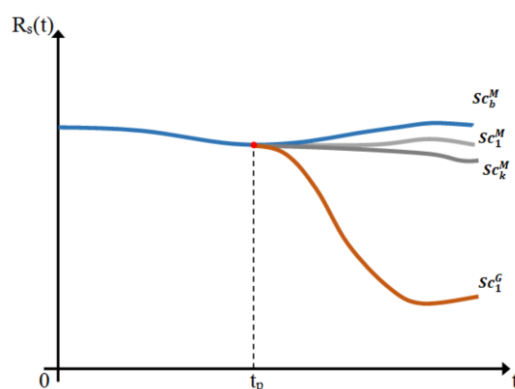


Рис. 3. Изменение жизнеспособности системы при разных сценариях моделирования

На рис. 4 приведено схематичное представление основной идеи организации сценарного моделирования жизнеспособности региональной БСВ-системы на основе генеративного искусственного интеллекта. Разработчик строит модель в имитационной среде традиционным способом, опираясь на экспертные знания об объекте исследования и предметной области и ретроспективных данных, используемых для настройки модели. При этом он ориентируется на воспроизведение определенного набора сценариев функционирования исследуемой системы в соответствии с его представлениями о поведении объекта моделирования, его возможных условиях функционирования и взаимодействии с другими объектами и окружающей средой. При имитации (многократном прогоне) разработанной компьютерной модели получается ансамбль траекторий развития системы, каждой из которых соответствует определенный сценарий. Это множество сценариев ограничено, и их количество определяется числом возможных комбинаций значений управляющих параметров модели.



Рис. 4. Схема генерации сценария развития системы на основе LLM и RAG-технологии

Согласно предложенному подходу разработчик (эксперт) может остановить имитацию модели в определенный момент времени T_p для того, чтобы попытаться изучить альтернативный сценарий развития системы, который может реализовываться, начиная с зафиксированного момента времени. При этом предполагается, что альтернативный сценарий может не входить во множество запланированных траекторий функционирования системы. В зафиксированный момент времени T_p снимается модельное состояние $St^M(T_p)$ и представляется в виде параметрического описания. Например, в среде имитационного моделирования Anylogic технически этот процесс можно реализовать с помощью создания пользовательской функции, которая при возникновении события «фиксация времени» сохраняла бы текущие значения параметров модели и их названия во внешний файл заданной структуры (электронную таблицу или таблицу БД).

Следующим шагом является получение из параметрического описания состояния системы (сформированного из имитационной модели) естественно-языкового описания исследуемого состояния. Данную процедуру осуществляет программный модуль «Генератор контекста». Функционирование генератора контекста может быть реализовано двумя способами:

- на основе онтологий и инженерии знаний (ontology-based);
- на основе больших языковых моделей (neural-based).

В первом случае используется лексическая онтология, содержащая правила формирования естественно-языкового описания из набора пар <Parameter, Value> параметрического описания, совместно с системной онтологией имитационного моделирования. Системная онтология содержит концепты и взаимоотношения между ними, описывающие процесс создания и применения имитационной модели. При необходимости для уточнения понятий (нахождения синонимичных терминов) дополнительно используется граф знаний жизнеспособности системы. Лексическая и системная онтология формируется на основе экспертных знаний, а граф знаний — на основе доступных источников информации, затрагивающие вопросы управления жизнеспособностью больших систем.

Во втором варианте в качестве генератора естественно-языкового контекста выступает большая языковая модель общего назначения. Возможно, для получения наиболее релевантных ответов придется нейросеть «дообучить» на основе подхода fine-tuning за счет предъявления нескольких эталонных примеров ожидаемых результатов работы LLM (Few-shot learning) [22].

После формирования естественно-языкового описания исследуемого состояния системы оно поступает на вход генератора сценария развития кризисной ситуации. В качестве генератора используется большая языковая модель, но, в отличие от ранее рассмотренной концепции генерации сценариев (рис. 2), в данном случае на вход нейронной сети подается расширенный запрос. Расширение запроса реализуется за счет добавления к промпту эксперта естественно-языкового описания исследуемого состояния системы. Разработчик предлагает LLM сформировать сценарий развития кризисной ситуации при заданных начальных условиях. Для технической реализации расширения запроса используется технология RAG. При этом для повышения качества результатов возможно добавление в расширенный запрос дополнительных текстовых фрагментов (chunks), содержащих более детальную информацию о предметной области.

Результатом работы предложенной схемы генерации сценария функционирования исследуемой системы (рис. 4) является сформированное естественно-языковое описание сценария. В дальнейшем данное описание может анализировать эксперт (разработчик модели) и, при необходимости, вносить изменения в имитационную модель. Также возможен вариант комбинирования предложенных схем генерации сценариев: на основе второй схемы создается естественно-языковое описание сценария, которое потом транслируется в сценарий имитационной модели в соответствии с первой схемой.

Стоит отметить, что в данном разделе статьи рассмотрены концептуальные схемы применения генеративного искусственного интеллекта для формирования сценариев, которые могут быть использованы в имитационном моделировании. Реализации на практике предложенных идей может сопровождаться появлением новых научно-практических вопросов, которые на текущем этапе исследования не выявлены.

Оценка жизнеспособности БСВ-системы на основе имитационного моделирования и генеративного искусственного интеллекта

Одной из прикладных задач, при решении которой можно использовать предложенные подходы генерации сценариев на основе интеграции имитационного моделирования и генеративного искусственного интеллекта, является оценка жизнеспособности региональных систем «Бизнес-Сообщество-Власть».

На сегодняшний день существуют разные исследования в области оценки жизнеспособности социально-экономических систем [23]. Большинство из них используют индикаторные системы оценки, включающие в себя разнообразные наборы показателей, отражающих авторский взгляд на понятие жизнеспособности системы. В рамках текущего исследования было предложено попробовать оценить жизнеспособность БСВ-системы как обобщенное свойство, формирующееся из трех составляющих: функциональности, восстанавливаемости и устойчивости (отказоустойчивости). Выбор таких составляющих свойств определен самим понятием жизнеспособности (resilience) — это способность системы при возникновении кризисной ситуации выявлять ее предпосылки, адаптироваться к новым условиям, сопротивляться негативным воздействиям кризиса и изменяться с целью достижения и поддержания приемлемого уровня функционирования как каждого компонента в отдельности, так и системы в целом. При этом под функциональностью понимается свойство системы сохранять свои функции во время своего функционирования; под восстанавливаемостью — свойство системы при возникновении кризисных ситуаций восстанавливать свою функциональность до докризисного уровня; под устойчивостью — свойство системы сохранять свое работоспособное состояние при воздействии внешнего окружения. Такие определения характерны для технических объектов [24], но могут быть использованы и для социально-экономических систем [25].

Имитационное моделирование может помочь оценить указанные свойства при исследовании жизнеспособности БСВ-систем. В частности, в компьютерной модели функциональность системы отражается в виде набора модельных параметров, являющимися элементами заданной индикаторной системы для оценки результатов функционирования объекта моделирования. Свойство восстанавливаемости предлагается оценивать через вычисление таких параметров, как время восстановления и общий объем затрачиваемых ресурсов, определяемых на основе имитации разных сценариев развития системы. С помощью имитационной модели устойчивость оценивается через такие показатели, как: степень снижения функциональности системы; время достижения максимального спада функциональности системы; соотношение успешно «поглощенных» кризисов к общему количеству исследованных кризисных ситуаций и др. Таким образом, для каждого сценария развития системы можно оценить функциональность, восстанавливаемость и устойчивость, опираясь на результаты вычислительных экспериментов. Жизнеспособность БСВ-системы можно вычислить как интегральный показатель, например, через взвешенную сумму значений трех ее составляющих:

$$R_S^{BSG} = w_f * Fc + w_r * Rc + w_s * St,$$

где Fc — функциональность системы; Rc — восстанавливаемость системы; St — устойчивость системы; w_f, w_r, w_s — весовые коэффициенты: $w_f + w_r + w_s = 1$.

Процесс оценки жизнеспособности БСВ-систем на основе имитационного моделирования и генерирующего искусственного интеллекта можно представить в виде следующего обобщенного алгоритма:

1. *Инициализация модели.* На данном этапе на вход разработанной имитационной модели жизнеспособности подаются значения параметров, полученные из различных источников данных (официальная статистика, анкетирование, экспертные оценки и т. д.). В результате, модель настраивается на имитацию базового сценария функционирования системы.

2. *Имитация базового сценария развития БСВ-системы на основе исходных данных.* На данном этапе осуществляется прогон модели (возможно, проведение нескольких экспериментов), а также проводятся расчеты значений свойств исследуемой системы: функциональности (Fc_0), восстанавливаемости (Rc_0) и устойчивости (St_0).

3. *Анализ результатов моделирования.* После проведенного эксперимента с имитационной моделью проводится экспертный анализ полученных результатов симуляции. При этом акцент делается на динамике свойств, связанных с жизнеспособностью системы, и их изменении в соответствии с жизненным циклом жизнеспособности. Выявление моментов времени, когда возможно появление кризиса, реализуется в ручном режиме на основе анализа нормальной траектории развития системы. При этом определение кризисной ситуации в рассматриваемой траектории также осуществляется экспертным методом. Для проверки результатов анализа эксперта о том, содержит ли нормальная траектория кризисную ситуацию, предлагается провести данную процедуру в автоматизированном режиме за счет применения большой языковой модели.

4. *Автоматизированное выявление кризисной ситуации в штатной траектории развития системы.* Для подтверждения или опровержения экспертной оценки о наличии кризисной ситуации в нормальном сценарии предлагается воспользоваться LLM как агрегатором человеческих знаний. Для этого на базе RAG-технологии формируется расширенный запрос, включающий в себя непосредственно промпт эксперта и дополнительный контекст. Контекст определяется как естественно-языковое описание, содержащее начальные условия функционирования системы, графики изменения во времени функциональности, устойчивости и восстанавливаемости системы. В качестве промпта эксперта к большой языковой модели используется запрос вида «необходимо определить, содержится ли в рассматриваемой траектории системы (в первую очередь для функциональности) кризисная ситуация». Возможно, для повышения релевантности ответа LLM придется в запрос добавлять описание основных понятий предметной области, представленных в соответствующих документах. Результатом отработки LLM расширенного запроса будет являться текст, содержащий «мнение» нейросети. Данный вывод можно сопоставлять с экспертной оценкой для ее последующего уточнения/корректировки.

5. *Генерация альтернативных сценариев развития системы.* На данном этапе предусматривается формирование альтернативных траекторий развития моделируемой системы в точках появления потенциального кризиса, выявленных на этапе 3. Данная процедура проводится с помощью предложенной концепции генерации сценариев на основе большой языковой модели и технологии расширения запроса дополнительным контекстом (рис. 4). Естественно-языковые описания моделей, составленные для выбранных состояний системы (предкризисных) за счет фиксации модельного времени, передается в генеративный искусственный интеллект. Результатом выполнения данного этапа является множество естественно-языковых описаний сценариев, которые, по мнению большой языковой модели, являются кризисными.

6. *Имитация альтернативных сценариев.* На данном этапе производится трансформация ЕЯ-описаний сгенерированных сценариев в имитационную модель на основе первой схемы применения LLM для генерации сценариев моделирования (рис. 2). При этом формируется множество конфигурационных файлов, каждый из которых используется для инициализации параметров имитационной модели, определяющих соответствующий сценарий функционирования системы. Далее выполняется ряд вычислительных экспериментов на основе проинициализированной компьютерной модели. Каждый вычислительный эксперимент является имитацией определенного альтернативного сценария развития системы. В каждом эксперименте средствами имитационной модели идет расчет показателей функциональности, восстанавливаемости, устойчивости и жизнеспособности.

7. *Комплексный анализ жизнеспособности БСВ-системы.* На данном этапе формируется сводный отчет о возможных вариантах поведения системы на исследуемом интервале времени. В отчет включается описание начальных условий и базового сценария развития системы; выявленные моменты времени, в которых может появиться потенциальный кризис; описание альтернативных сценариев развития системы и результаты их моделирования. Для каждого проанализированного сценария приводятся полученные количественные оценки показателей функциональности, восстанавливаемости, устойчивости и жизнеспособности. Сформированный отчет может служить основой для анализа потенциальных кризисных сценариев развития системы с позиции оценки ее жизнеспособности. Некоторые отдельные процедуры анализа можно автоматизировать с помощью генеративного искусственного интеллекта. Например, использовать большую языковую модель для аннотирования информации, содержащейся в отчете, с целью формирования тезисных описаний для лица, принимающего решения. Другой пример: на основе отчетных материалов LLM формулирует наиболее (или наименее) предпочтительный сценарий развития БСВ-системы с позиции одного из ее ключевых участников (например, бизнеса).

Предложенный алгоритм оценки жизнеспособности региональных БСВ-систем позволяет сформировать численные оценки трех составляющих жизнеспособности системы на основе результатов имитации разных сценариев развития системы как предусмотренных разработчиком модели, так и сгенерированных большой языковой моделью. Алгоритм базируется на рассмотренных в статье двух концепциях применения генеративного искусственного интеллекта для автоматизации имитационного моделирования, в частности, генерации сценариев кризисных ситуаций.

Заключение

Задача упрощения работы человека по созданию компьютерных моделей все еще является актуальной. Процесс построения модели является субъективным, результаты которого зависят от персональных навыков, знаний и опыта разработчика. При этом на адекватность и эффективность разрабатываемого решения влияет сложность самого объекта моделирования: чем сложнее объект (больше элементов, больше неявных связей между элементами), тем сложнее добиться результатов требуемого качества. Для решения данной задачи привлекаются разные подходы, методы и технологии, ориентированные на полную или частичную автоматизацию действий разработчика имитационной модели.

В текущей статье предлагаются две концептуальные схемы применения генеративного искусственного интеллекта для имитационного моделирования сложных систем. Первая концепция предназначена для генерации сценария развития региональной социально-экономической системы и его последующей трансформации в имитационную модель жизнеспособности такой системы. Вторая схема направлена на генерацию сценариев развития системы на основе результатов имитационного моделирования и реализуется через технологию расширения запроса за счет добавления дополнительного контекста (RAG). В обоих случаях в качестве генератора сценариев предлагается использовать большую языковую модель общего назначения.

В качестве примера применения предложенных концепций совместно с имитационным моделированием рассматривается прикладная задача оценки жизнеспособности региональных «Бизнес-Сообщество-Власть»-систем. В рамках решения данной задачи предложен обобщенный алгоритм, ориентированный на получение количественных показателей свойств жизнеспособности (функциональности, восстанавливаемости, устойчивости) на основе результатов вычислительных экспериментов разных сценариев развития БСВ-системы.

Список источников

1. Шелухин О. И., Шариков А. Ю. Имитация поведения компьютерной системы с помощью искусственных нейронных сетей // Т-Comm: Телекоммуникации и транспорт. 2021. Т. 15, № 5. С. 29–37. doi:10.36724/2072-8735-2021-15-5-29-37.
2. Костромин Р. О., Феоктистов А. Г. Сервис подготовки и запуска имитационных моделей функционирования инфраструктурных объектов в распределенной вычислительной среде // Свидетельство о регистрации программы для ЭВМ RU 2021662946, 10.08.2021. Заявка № 2021662355 от 10.08.2021.

3. Волегов И. С., Замятина Е. Б. Онтологический подход к интеграции компонентов имитационной модели в TRIAD.NET // Открытые семантические технологии проектирования интеллектуальных систем. 2012. № 2. С. 229–236.
4. Prat S. et al. An Automated Generation Approach of Simulation Models for Checking Control/Monitoring System // IFAC-PapersOnLine. 2017. Vol. 50, Iss. 1. pp. 6202–6207. doi:10.1016/j.ifacol.2017.08.1014.
5. Giannakis G. et al. Simulation model generation combining IFC and CityGML data. 2016. [электронный ресурс]. URL: https://www.researchgate.net/publication/307964150_Simulation_model_generation_combining_IFC_and_CityGML_data (дата обращения: 11.10.2024).
6. Pourbafrani M., van Zelst S. J., van der Aalst W. M. P. Supporting Automatic System Dynamics Model Generation for Simulation in the Context of Process Mining // Lecture Notes in Business Information Processing. 2020. vol. 389. pp. 249–263. doi:10.1007/978-3-030-53337-3_19.
7. Huang Y., Verbraeck A., Seck M. (2016). Graph transformation based simulation model generation // Journal of Simulation. 2016. № 10 (4). pp. 283–309. doi:10.1057/jos.2015.21/.
8. Путилов В. А., Горохов А. В., Быстров В. В. Синтез имитационных моделей сложных систем на основе экспертных знаний // Информационные технологии и вычислительные системы. 2008. № 2. С. 27–35.
9. Кудинова О. В., Халиуллина Д. Н. Создание шаблонов имитационной модели прогнозирования трудовых ресурсов с помощью онтологических описаний предметной области // Труды Кольского научного центра РАН. 2013. № 5(18). С. 208–216.
10. Маслобоев А. В., Олейник А. Г., Шишаев М. Г. Информационная технология дистанционного формирования и управления моделями системной динамики // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15, № 4. С. 748–755.
11. Карминская Т. Д., Татьянкин В. М., Тей Д. О., Русанов М. А. Использование кластерного анализа и нейронных сетей в задаче управления региональным рынком труда // Доклады Томского государственного университета систем управления и радиоэлектроники. 2013. № 4(30). С. 205–209.
12. Наместников А. М. Применение онтологического подхода в задаче генерации событийных данных с помощью имитационных моделей // Онтология проектирования. 2023. Т. 13, № 2. С. 243–253. doi:10.18287/2223-9537-2023-13-2-243-253.
13. Akhavan A., Jalali M. S. Generative AI and simulation modeling: how should you (not) use large language models like ChatGPT // Syst. Dyn. Rev. 2024. 40. e1773. doi:10.1002/sdr.1773.
14. Точные прогнозы с имитационным моделированием и машинным обучением от H2O.ai [Электронный ресурс]. URL: <https://www.anylogic.ru/features/artificial-intelligence/h2o-ai/> (дата обращения: 11.10.2024).
15. Jackson I., Saenz M.J., Ivanov D. From natural language to simulations: applying AI to automate simulation modelling of logistics systems // International Journal of Production Research. 2024. 62:4. pp. 1434–1457. doi:10.1080/00207543.2023.2276811.
16. McCormack J., Grierson M. Building Simulations with Generative Artificial Intelligence // In D. Del Favero, S. Thurow, M. J. Ostwald, U. Frohne (Eds.), Climate Disaster Preparedness: Reimagining Extreme Events through Art and Technology. 2024. pp. 137–150. doi:10.1007/978-3-031-56114-6_11.
17. Юдицкий С. А. Сценарный подход к логическому моделированию систем рыночной экономики // Системы управления, связи и безопасности. 2015 [Электронный ресурс]. № 2. С. 147–164. URL: <http://journals.intelgr.com/scs/archive/2015-02/06-Iuditskii.pdf> (дата обращения: 11.10.2024).
18. Kang Y. et al. Natural Language Processing (NLP) in Management Research: A Literature Review // Journal of Management Analytics. 2020. 7(2). pp. 139–172. doi:10.1080/23270012.2020.1756939.
19. Khurana D. et al. Natural language processing: state of the art, current trends and challenges // Multimed Tools Appl. 2023. 82. pp. 3713–3744. doi:10.1007/s11042-022-13428-4.
20. Khaliullina D. N., Bystrov V. V. The Conception of Assessing the Resilience of Critical Infrastructures of Regional Socio-economic Systems // Lecture Notes in Networks and Systems. 2023. vol. 597. pp. 539–553. doi:10.1007/978-3-031-21438-7_43.
21. Khan A. et al. Developing Retrieval Augmented Generation (RAG) based LLM Systems from PDFs: An Experience Report. 2024. doi:10.48550/arXiv.2410.15944. [электронный ресурс]. URL: https://www.researchgate.net/publication/385108254_Developing_Retrieval_Augmented_Generation_RAG_based_LLM_Systems_from_PDFs_An_Experience_Report (дата обращения: 11.10.2024).

22. Gautier I. et al. Atlas: Few-shot Learning with Retrieval Augmented Language Models // *Journal of Machine Learning Research*. 2023. vol. 24. pp. 1-43. [электронный ресурс]. URL: <https://jmlr.org/papers/volume24/23-0037/23-0037.pdf> (дата обращения: 11.10.2024).
23. Халиуллина Д. Н., Быстров В. В. Концептуальные аспекты исследования жизнеспособности региональных социально-экономических систем // *Human Progress*. 2023. Т. 9, № 5. С. 10. doi:10.34709/IM.195.10.EDN RARLIY.
24. Rehak D., Senovsky P., Slivkova S. Resilience of Critical Infrastructure Elements and Its Main Factors // *Systems*. 2018. №6. 21. doi:10.3390/systems6020021.
25. Сушков В. В. Развитие функциональности в технических и бизнес-системах. TRIZ Developers Summit 2020 August 17–21, 2020. Minsk, Belarus. С. 1–8 [Электронный ресурс]. URL: https://r1.nubex.ru/s828-c8b/f3244_06/Souckov-TDS-2020-Functionality.pdf (дата обращения: 11.10.2024).

References

1. Sheluhin O. I., Sharikov A. Yu. Simulation of the behavior of a computer system using artificial neural networks // *T-Comm* 2021. Vol. 15, No. 5, pp. 29-37. doi:10.36724/2072-8735-2021-15-5-29-37. (In Russ.).
2. Kostromin R. O., Feoktistov A. G. Servis podgotovki i zapuska imitacionnyh modelej funkcionirovaniya infrastruktury ob"ektov v raspredelennoj vychislitel'noj srede [Service for the preparation and launch of simulation models of the functioning of infrastructure facilities in a distributed computing environment] // Certificate of registration of a computer program RU 2021662946, 10.08.2021. Bid № 2021662355 from 10.08.2021. (In Russ.).
3. Volegov I. S., Zamyatina E. B. The integration of simulation model components into triad.net using ontological approach // *OSTIS-2012* 2012. No. 2. p. 229–236. (In Russ.).
4. Prat S. et al. An Automated Generation Approach of Simulation Models for Checking Control/Monitoring System. *IFAC-PapersOnLine*, 2017, Vol. 50, Iss. 1, pp. 6202–6207. doi: 10.1016/j.ifacol.2017.08.1014.
5. Giannakis G. et al. Simulation model generation combining IFC and CityGML data, 2016. Available at: https://www.researchgate.net/publication/307964150_Simulation_model_generation_combining_IFC_and_CityGML_data (accessed 11.10.2024).
6. Pourbafrani M., van Zelst S. J., van der Aalst W.M.P. (2020). Supporting Automatic System Dynamics Model Generation for Simulation in the Context of Process Mining. *Lecture Notes in Business Information Processing*, 2020, Vol. 389, pp. 249–263. doi:10.1007/978-3-030-53337-3_19.
7. Huang Y., Verbraeck A., Seck M. Graph transformation based simulation model generation. *Journal of Simulation*, 2016, Vol. 10, No.4, pp. 283–309. <https://doi.org/10.1057/jos.2015.21>.
8. Putilov V. A., Gorokhov A. V., Bystrov V. V. Sintez imitacionnyh modelej slozhnyh sistem na osnove ekspertnyh znaniy [Synthesis of simulation models of complex systems based on expert knowledge] // *Informacionnye tekhnologii i vychislitel'nye sistemy* [Information technology and computing systems] 2008. No. 2, pp. 27–35. (In Russ.).
9. Kudinova O. V., Khaliullina D. N. Sozdanie shablonov imitacionnoj modeli prognozirovaniya trudovyh resursov s pomoshch'yu ontologicheskikh opisaniy predmetnoj oblasti [Creating templates of simulation model for labour force balance prediction based on the ontological descriptions of subject area] // *Trudy Kольского научного центра РАН*. [Transactions of the Kola Science Centre. Information technologies] 2013. Vol. 5, No. 18, pp. 208–216. (In Russ.).
10. Masloboev A. V., Oleynik A. G., Shishaev M. G. Informacionnaya tekhnologiya distancionnogo formirovaniya i upravleniya modelyami sistemnoj dinamiki [Remote synthesis and control information technology of system-dynamic models] // *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2015. Vol.15, No. 4, pp. 748–755. doi:10.17586/2226-1494-2015-15-4-748-755. (In Russ.).
11. Karminskaya T. D., et al. Ispol'zovanie klasternogo analiza i nejronnyh setej v zadache upravleniya regional'nym rynkom truda [Human resources forecasting within economic activities with the use of cluster analysis and neural networks] // *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki* [Reports of Tomsk State University of Control Systems and Radioelectronics] 2013. No. 4(30), pp. 205–209. (In Russ.).

12. Namestnikov A. M. Primenenie ontologicheskogo podhoda v zadache generacii sobytijnyh dannyh s pomoshch'yu imitacionnyh modelej [Application of the ontological approach in the problem of event data generation using simulation models]. *Ontologiya proektirovaniya* [Ontology of designing] 2023; Vol. 13, No. 2, pp. 243–253. doi:10.18287/2223-9537-2023-13-2-243-253. (In Russ.).
13. Akhavan, A., Jalali, M. S. Generative AI and simulation modeling: how should you (not) use large language models like ChatGPT. *Syst. Dyn. Rev.*, 2024, 40: e1773. <https://doi.org/10.1002/sdr.1773>
14. Enhancing the Predictive Capabilities of Simulation with H2O.ai Automatic Machine Learning. Available at: <https://www.anylogic.ru/features/artificial-intelligence/h2o-ai/> (accessed 11.10.2024).
15. Jackson I., Saenz M. J., Ivanov D. From natural language to simulations: applying AI to automate simulation modelling of logistics systems, *International Journal of Production Research*, 2024, 62:4, pp. 1434-1457. doi:10.1080/00207543.2023.2276811.
16. McCormack J., Grierson M. Building Simulations with Generative Artificial Intelligence. In D. Del Favero, S. Thurow, M. J. Ostwald, & U. Frohne (Eds.), *Climate Disaster Preparedness: Reimagining Extreme Events through Art and Technology*. 2024, pp. 137–150. doi:10.1007/978-3-031-56114-6_11.
17. Iuditskii S. A. The Scenary Approach to the Logical Modeling of Systems of a Market Economy // *Sistemy upravleniya, svyazi i bezopasnosti* [Management, communication and security systems], 2015. No. 2, pp. 147–164. Available at: <http://journals.intelgr.com/scs/archive/2015-02/06-Iuditskii.pdf> (accessed 11.10.2024). (In Russ.).
18. Kang Y. et al. Natural Language Processing (NLP) in Management Research: A Literature Review. *Journal of Management Analytics*, 2020. Vol. 7, No. 2, pp. 139–172. <https://doi.org/10.1080/23270012.2020.1756939>.
19. Khurana D. et al. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl.* 2023, Vol. 82, pp. 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>.
20. Khaliullina, D. N., Bystrov, V. V. (2023). The Conception of Assessing the Resilience of Critical Infrastructures of Regional Socio-economic Systems. // *Data Science and Algorithms in Systems. CoMeSySo Lecture Notes in Networks and Systems*, 2022. vol. 597, pp. 539–553. https://doi.org/10.1007/978-3-031-21438-7_43.
21. Khan A. et al. Developing Retrieval Augmented Generation (RAG) based LLM Systems from PDFs: An Experience Report. 2024. Available at: https://www.researchgate.net/publication/385108254_Developing_Retrieval_Augmented_Generation_RAG_based_LLM_Systems_from_PDFs_An_Experience_Report. (accessed 11.10.2024). doi:10.48550/arXiv.2410.15944.
22. Gautier I. et al. Atlas: Few-shot Learning with Retrieval Augmented Language Models // *Journal of Machine Learning Research*, 2023. Vol. 24, pp. 1–43. Available at: <https://jmlr.org/papers/volume24/23-0037/23-0037.pdf> (accessed 11.10.2024).
23. Khaliullina D. N., Bystrov V. V. Conceptual aspects of studying of regional socio-economic systems' resilience. // *Human Progress*, 2023. Vol. 9, No. 5, p. 10. doi:10.34709/IM.195.10. EDN RARLIY (In Russ.).
24. Rehak D., Senovsky P., Slivkova S. Resilience of Critical Infrastructure Elements and Its Main Factors // *Systems*, 2018, No. 6, p. 21. <https://doi.org/10.3390/systems6020021>
25. Sushkov V. V. Development of functionality in technical and business systems. *TRIZ Developers Summit 2020 August 17–21, 2020. Minsk, Belarus*. pp. 1–8. Available at: https://r1.nubex.ru/s828-c8b/f3244_06/Souchkov-TDS-2020-Functionality.pdf. (accessed 11.10.2024). (In Russ.).

Информация об авторах

В. В. Быстров — кандидат технических наук, ведущий научный сотрудник;
Д. Н. Халиуллина — кандидат технических наук, научный сотрудник;
М. Г. Шишаев — доктор технических наук, доцент, главный научный сотрудник.

Information about the authors

V. V. Bystrov — Candidate of Science (Tech.), Lead Research Fellow;
D. N. Khaliullina — Candidate of Science (Tech.), Research Fellow;
M. G. Shishaev — Doctor of Science (Tech.), Associate Professor, Chief Researcher

Статья поступила в редакцию 15.10.2024; одобрена после рецензирования 01.11.2024; принята к публикации 08.11.2024.
The article was submitted 15.10.2024; approved after reviewing 01.11.2024; accepted for publication 08.11.2024.

Научная статья
УДК 004.9
doi:10.37614/2949.1215.2024.15.3.003

СИТУАЦИОННЫЕ ЦИФРОВЫЕ ДВОЙНИКИ В УПРАВЛЕНИИ КРИТИЧЕСКИМИ ИНФРАСТРУКТУРАМИ: СПЕЦИФИКАЦИЯ ТРЕБОВАНИЙ

Александр Яковлевич Фридман^{1✉}, Андрей Владимирович Маслобоев²

^{1, 2}*Институт информатики и математического моделирования имени В. А. Путилова*

Кольского научного центра Российской академии наук, Апатиты, Россия

¹*fridman@iimm.ru[✉], <https://orcid.org/0000-0003-2408-6892>*

²*masloboev@iimm.ru, <https://orcid.org/0000-0002-1231-9225>*

Аннотация

Предложены и обоснованы средства формализации основных технических требований к компьютерным программам-исполнителям (цифровым двойникам) элементов модели критических инфраструктур на примере промышленно-природных комплексов в рамках ранее разработанной ситуационной системы моделирования, ориентированные на ее децентрализованную реализацию в специализированном секторе Интернета, в частности, для решения задач управления жизнеспособностью этих критически важных объектов. Такая спецификация может позволить избежать типичной для компьютерных задач начальной стадии хаотического развития программно-аппаратного обеспечения новых структурных проектов в этой области и ускорить процесс саморазвития, необходимый для эффективного объединения разработок цифровых двойников критически важных объектов и инфраструктур различных типов в рамках единой системы распределенных ситуационных центров.

Ключевые слова:

критическая инфраструктура, промышленно-природный комплекс, ситуационная система моделирования, цифровой двойник, превентивная аналитика безопасности

Благодарности:

исследование выполнено в рамках государственного задания Института информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук от Министерства науки и высшего образования Российской Федерации, тема научно-исследовательской работы — «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (проект № FMEZ-2022-0023).

Для цитирования:

Фридман А. Я., Маслобоев А. В. Ситуационные цифровые двойники в управлении критическими инфраструктурами: спецификация требований // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 41–49. doi:10.37614/2949.1215.2024.15.3.003.

Original article

SITUATIONAL DIGITAL TWINS IN MANAGEMENT OF CRITICAL INFRASTRUCTURES: SPECIFICATION OF REQUIREMENTS

Alexander Ya. Fridman^{1✉}, Andrey V. Masloboev²

^{1, 2}*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences, Apatity, Russia*

¹*fridman@iimm.ru[✉], <https://orcid.org/0000-0003-2408-6892>*

²*masloboev@iimm.ru, <https://orcid.org/0000-0002-1231-9225>*

Abstract

The main means to formalize technical requirements for software executors (digital twins) of the elements of the critical infrastructure model by the example of industrial-natural complexes within the previously developed situational modeling system, oriented towards its decentralized implementation in a specialized sector of the Internet, specifically, for the problem-solving of resilience management of these critical entities, are proposed and substantiated. Such a specification can help to avoid the initial stage of the chaotic development of software and hardware for new structural projects in this area, which is typical for computer problems, and accelerate the process of self-development necessary to effectively combine the developments of digital twins for critical entities and infrastructures of various types within a unified system of distributed situational centers.

Keywords:

critical infrastructure, industrial-natural complex, situational modeling system, digital twin, preventive safety analytics

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic "Development of theoretical and organizational-technical foundations of information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation" (project No. FMEZ-2022-0023).

For citation:

Fridman A. Ya., Masloboev A. V. Situational digital twins in management of critical infrastructures: specification of requirements // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 41–49. doi:10.37614/2949.1215.2024.15.3.003.

Введение

На сегодняшний день актуальным и перспективным направлением исследований в области ситуационного управления безопасностью и устойчивым функционированием критических инфраструктур является разработка цифровых моделей-двойников гетерогенных групп взаимосвязанных критически важных объектов (комплексов), образующих эти инфраструктуры, для повышения их надежности, защищенности и жизнеспособности. Своевременность данных исследований подтверждается директивными документами [1–3], принятыми на государственном уровне, а также результатами обзорных исследований [4–8], анализирующих российский и зарубежный опыт создания и применения цифровых двойников в различных сферах деятельности. Отечественный аналитический обзор [8] не является исчерпывающим, но обобщает широкий круг вопросов, касающихся внедрения цифровых двойников в критические секторы экономики и промышленности. Согласно работе [8] под цифровым двойником в широком смысле понимается любая цифровая копия (концептуальная, математическая, имитационная или киберфизическая модель, виртуальная или электронная копия и т. д.) живого или неживого объекта, процесса или системы (как правило, с невысоким уровнем адекватности), отражающая по определению наиболее существенные для заданной цели и условий характеристики исследуемого физического объекта или процесса. В реальной практике, особенно в критических приложениях таких, как превентивная аналитика безопасности и управление устойчивостью критических инфраструктур, цифровые двойники представляют собой специализированные модельные или программно-технические решения, которые должны удовлетворять жестким требованиям и критериям в части обеспечения высокого уровня адекватности виртуального объекта реальному физическому объекту.

В исследовании [4] предлагается оригинальная классификация цифровых двойников, основанная на уровнях интеграции данных и взаимодействия цифрового и реального физического объектов. Выделяют три основных типа цифровых двойников [8]: цифровая модель (Digital Model), в рамках которой не существует автоматизированного потока данных между физической и виртуальной составляющей объекта; цифровая тень (Digital Shadow), в рамках которой автоматизирован односторонний поток данных от физического объекта к виртуальному объекту; цифровой двойник (Digital Twin), который предполагает двусторонний обмен информацией между цифровым и реальным объектами.

В данной работе, основываясь на типологии [4], рассматривается спецификация требований к цифровым двойникам для ситуационного управления критическими инфраструктурами на примере промышленно-природных комплексов с использованием ситуационной системы моделирования. Разработанные для формализации и автоматизации управления отдельными компонентами критических инфраструктур ситуационная концептуальная модель и ее расширения, являющиеся ядром системы ситуационного моделирования, находятся на стыке типов цифровая модель (Digital Model) и цифровая тень (Digital Shadow), однако в случае успешной практической реализации могут быть позиционированы, как полноценный цифровой двойник (Digital Twin) для исследуемого класса объектов.

Материалы и методы

В соответствии с работой [4] и аналитическими и экспертными материалами, систематизированными в исследовании [8], цифровая модель — это цифровое представление существующего или проектируемого реального объекта, в рамках которого не реализуется

автоматический обмен данными между физическим объектом и его виртуальным аналогом. Цифровая модель должна включать достаточно точное описание характеристик реального объекта как на концептуальном, так и на формальном (математическом) уровне. Настройка параметров и управление реальным объектом или процессом посредством виртуальной модели осуществляется в ручном режиме, т. е. данные о состоянии физического объекта вводятся в модель вручную. В таком случае изменение параметров состояния реального объекта не имеет прямого эффекта на изменение соответствующих характеристик виртуального объекта. Другой принцип взаимодействия реализован в цифровой тени, который предполагает односторонний автоматический обмен информацией между реальным объектом и виртуальным объектом (данные передаются от реального к виртуальному объекту). Изменение состояния реального объекта ведет к соответствующему изменению виртуального объекта, однако обратная связь отсутствует (данные могут передаваться в ручном режиме). Наконец в истинном цифровом двойнике интегрированные данные передаются как от физического объекта к виртуальному объекту, так и в обратном направлении. В этом случае цифровой объект становится контрольной точкой или образцом для реального объекта. Изменение состояния реального объекта ведет к изменению виртуального объекта и наоборот. Информационное взаимодействие в системе «реальный объект — цифровой объект» для разных типов цифровых двойников схематично проиллюстрировано на рис. 1.

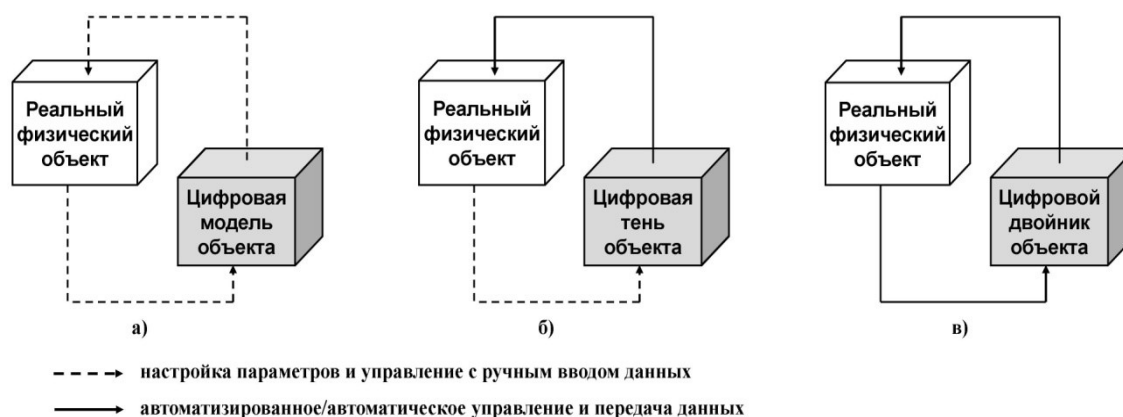


Рис. 1. Поток данных и управления в системе «реальный объект – цифровой объект» [4]:
а — цифровая модель; б — цифровая тень; в — цифровой двойник

Одна из ключевых возможностей ситуационной системы моделирования (ССМ) [9] состоит в предоставлении средств анализа динамики поведения объектов моделирования, в том числе критических инфраструктур, к которым относятся промышленно-природные комплексы (ППК) — в имитационном режиме, который с математической точки зрения обеспечивает решение задачи Коши для векторного разностного уравнения [10] (уравнения состояния ППК), образованного некоторым набором взаимосвязанных программных модулей. Сами по себе такие модули могут иметь самую различную природу: имитаторы процессов, происходящих внутри объектов — составных частей ППК; поднаборы правил специальной структуры, которые хранятся в экспертной системе (ЭС), встроенной в ССМ; процедуры ввода в имитаторы процессов требуемых для расчетов графических характеристик объектов ППК посредством встроенной ГИС ССМ; процедуры определения приоритетов вызова имитаторов процессов в ходе имитации и т. д. В целях удобства формализации всем элементам модели ППК в ССМ (далее она называется ситуационной концептуальной моделью — СКМ) и моделирует ППК в виде взаимодействующего набора трех типов сущностей: объектов — организационных составных частей ППК; ресурсов, формализующих временные ряды величин потоков информации и материалов внутри СКМ, и процессов, которые содержат модели преобразований ресурсов) вручную или автоматически, в ходе конструирования СКМ, назначаются исполнители (с недавних пор их стали именовать цифровыми двойниками [11]),

которые задают особенности компьютерного моделирования этих элементов модели. Для процессов СКМ исполнители определяют интерфейс программного модуля — имитатора процесса — с подсистемой организации имитационного режима; у ресурсов с помощью исполнителя устанавливается способ получения последовательностей значений этих ресурсов во времени (обращение к соответствующим БД или генераторам значений); исполнитель некоторого объекта агрегирует цифровые двойники процессов, приписанных к этому объекту, и конкретизирует интерфейсы их взаимодействия с ГИС соответственно для использования графических характеристик объекта в расчетах или представления результатов расчетов на электронной карте. Если какие-либо ресурсы или процессы не имеют отдельных исполнителей, их быстрый прототип в рамках парадигмы искусственного интеллекта [12] создается назначением им исполнителя «ЭС», тогда значения этого ресурса или всех выходных ресурсов такого процесса должны вычисляться вызовом некоторого поднабора правил ЭС ССМ.

Каждый шаг имитации состоит в подготовке исходных значений ресурсов и контроле их достаточности для срабатывания динамически формируемого набора исполнителей процессов, вызов которых согласно их текущим приоритетам моделирует одну итерацию решения упомянутого выше уравнения Коши и переводит СКМ из текущего состояния в следующее. Корректно сформированная последовательность шагов имитации на заданном интервале времени реализует некоторый сценарий поведения ППК на этом интервале.

Имитация и интерпретация ее результатов происходит в обобщенном экспертном пространстве моделирования [13], допускающем любые типы переменных состояния, в котором опасные и критические (аварийные) значения переменных состояния располагаются по обе стороны диапазона допустимых значений каждой переменной, что обеспечивает интегральную экспертную оценку степени опасности состояния ППК и позволяет унифицировать методы обработки допустимых и опасных ситуаций, складывающихся в изучаемом ППК, т. е. строить модели нештатных и чрезвычайных ситуаций как расширения моделей нормального функционирования ППК. В общем случае при выходе значения некоторой переменной состояния за ее диапазон безопасных значений может потребоваться замена исполнителя ресурса или процесса, генерирующего значения этой переменной.

Как следует из изложенного, в ССМ применяются специализированные цифровые двойники (далее будем их называть ситуационными цифровыми двойниками — СЦД [14]), технические требования к которым отличаются от общепринятых [11] и кратко сформулированы ниже.

Естественно, основное требование состоит в адекватности СЦД некоторому процессу или объекту заданного ППК по тернарному соответствию «вход-состояние-выход». Здесь особую важность приобретает проблема идентификации начального состояния СЦД для исследуемого сценария. Методы ее решения зависят от типа оператора (передаточной функции) СЦД и должны включаться в программный модуль каждого двойника. Для ядра ССМ более существенна задача интеграции баз данных СЦД. С этой целью предлагается использовать расширенную концептуальную модель (РКМ), разработка которой начата в исследовании [15].

Основное отличие РКМ от СКМ состоит в том, что в РКМ детально проработан аппарат сопровождения характеристик, изменяющихся при проведении имитации. РКМ, как и СКМ, опирается на системные типы элементарных данных {integer, float, string}, но дополнительно описывает форматы представления данных на всех этапах моделирования, что создает основу для автоматической генерации БД СЦД как части системы баз данных моделирования (БДМ). Кроме того, в схеме РКМ явно указаны дополнительные отношения «вход-выход» по всем процессам и ресурсам, задействованным в СЦД, которые наиболее существенны для анализа корректности модели, а значит, и для анализа БДМ.

Далее рассмотрим структуру разработанной РКМ более детально.

Результаты и обсуждение

Как и в СКМ [15], схема РКМ имеет следующий вид:

$$S_{PKM} ::= \langle O, P, R, RP, PR, RO, OR, OA, H \rangle, \quad (1)$$

где O, P — множества элементов (объектов, процессов), определяемые аналогично соответствующим элементам СКМ; R — множество ресурсов (соответствует D^{CM} в СКМ); $H \subseteq O \times B(O)$ — отношения

иерархии объектов (B^*) — булеан множества с именем *); $RP \subseteq B(R) \times P$ — отношения «процесс — входные данные»; $PR \subseteq P \times B(R)$ — отношения «процесс — выходные данные»; $RO \subseteq B(R) \times O$ — отношения «объект — входные данные»; $OR \subseteq O \times B(R)$ — отношения «объект — выходные данные»; $OA \subseteq O \times V(P)$ — отношения «объект — приписанные к нему процессы».

Отношения, определенные в модели, можно представить в форме следующих функций:

— функция h определяется аналогично соответствующей функции в СКМ;

— $gr: P \rightarrow B(R)$, $\{r_j\} = gr(p_i)$, где p_i — некоторый процесс; $\{r_j\}$ — множество входных данных процесса p_i ($r_j \in gr(p_i)$);

— $pr: P \rightarrow B(R)$, $\{r_j\} = pr(p_i)$, где p_i — некоторый процесс; $\{r_j\}$ — множество выходных данных процесса p_i ($r_j \in pr(p_i)$);

— $go: O \rightarrow B(R)$, $\{r_j\} = go(o_i)$, где o_i — некоторый объект; $\{r_j\}$ — множество входных данных объекта o_i ($r_j \in go(o_i)$);

— $or: O \rightarrow B(R)$, $\{r_j\} = or(o_i)$, где o_i — некоторый объект; $\{r_j\}$ — множество выходных данных объекта o_i ($r_j \in or(o_i)$);

— $oa: O \rightarrow V(P)$, $\{p_j\} = oa(o_i)$, где o_i — некоторый объект, $\{p_j\}$ — процессы, приписанные объекту o_i .

Множества значений функций, соответствующие сечениям областей значений отношений КМПО (ее фрагментов) по некоторому элементу областей их определения, обозначаются жирным цветом. Аналогично обозначаются сечения вводимых отношений по некоторым подмножествам их областей определения, которые определяются как объединения всех сечений по элементам этих подмножеств. Например, $h(O_i)$, где $O_i \subseteq O$ есть множество объектов, доминируемых данным подмножеством объектов $o_j \in O_i$.

Отношения RP , PR , OR , RO служат для назначения входных, а также выходных ресурсов объектам и процессам. Функции gr , pr , or , ro соответствуют функциям $list_in$ и $list_out$ в СКМ, которые можно применить как к объектам, так и к процессам, но не входят в схему КМПО.

Атрибуты, характеризующие элементы КМПО, описываются моделью атрибутов, определяющей структуру таблиц БД КМ. Модель атрибутов РКМ образуется кортежем:

$$A_{PKM} ::= \langle N, T, n, t, E, T_e, n_e, t_e, T_h, t_h, V, rv \rangle, \quad (2)$$

где атрибуты N, T, E, T_e, T_h и функции n, t, n_e, t_e, t_h определяются аналогично соответствующим атрибутам и функциям СКМ; $V ::= \langle V_{int}, V_{float}, V_{str} \rangle$ — множество возможных значений ресурсных связей; V_{int} — множество целочисленных значений; V_{float} — множество вещественных значений; V_{str} — множество строковых значений; функции rv сопоставляют каждому ресурсу множество его возможных значений: $rv ::= \langle rv_{int}, rv_{float}, rv_{str} \rangle$, $rv_{int}: R \rightarrow B(V_{int})$, $rv_{float}: R \rightarrow B(V_{float})$, $rv_{str}: R \rightarrow B(V_{str})$.

Поскольку процесс построения модели включает не только ее формирование, но и выбор интересующего фрагмента, а также его доопределение, то БДМ обязана хранить не только данные, относящиеся ко всей модели, но и данные, специфичные для отдельных ее фрагментов. Далее рассмотрим схемы фрагментов РКМ. Использование фрагментов РКМ, а также их названия совпадают с соответствующими фрагментами СКМ [15].

Концептуальную модель полной ситуации можно представить кортежем:

$$S_{ПС} ::= \langle O', P', R', BV \rangle, \quad (3)$$

где O' — множество объектов полной ситуации ($O' \subseteq O$); P' — множество процессов полной ситуации ($P' \subseteq P$); R' — множество ресурсов полной ситуации ($R' \subseteq R$); $BV \subseteq R' \times rv(R')$ — отношение «ресурс — множество его начальных значений».

Отношение BV представимо в виде набора функций $bv ::= \langle bv_{int}, bv_{float}, bv_{str} \rangle$, где $bv_{int}: R' \rightarrow rv_{int}(R')$, $bv_{float}: R' \rightarrow rv_{float}(R')$, $bv_{str}: R' \rightarrow rv_{str}(R')$.

Концептуальная модель достаточной ситуации имеет вид:

$$S_{ДС} ::= \langle O'', P'', R'', TSet, RS \rangle, \quad (4)$$

где O'' — множество объектов достаточной ситуации ($O'' \subseteq O'$); P'' — множество процессов достаточной ситуации ($P'' \subseteq P'$); R'' — множество ресурсов достаточной ситуации ($R'' \subseteq R'$); $TSet = \{tset_i\}$ — множество временных рядов; $RS \subseteq R'' \times TSet$ — отношение «ресурс достаточной ситуации — временной ряд».

$TSet \subseteq T_1 \times T_2 \times R'' \times rv(R'')$, где $T_1 \subseteq Time$ — моменты фиксации значений; $T_2 \subseteq Time$ — моменты окончания действия значения, зафиксированного в момент T_1 ; $Time$ — множество моментов времени.

Потребность в модификации отношений схемы концептуальной модели при переходе от СКМ к РКМ обусловлена следующим. Для удобства динамического анализа модели (в частности, разрешимости модели) в схеме СКМ присутствуют отношения ОР и РО (соответственно, отношение «объект — порождающие его выходные данные процессы» и отношение «процесс — создающие его входные данные объекты»). Одна из основных задач настоящего подраздела — не разработка новых алгоритмов проверки модели исследуемого объекта, а создание такого представления схемы концептуальной модели, ее фрагментов и ограничений на их отношения (т. е. создание онтологии), которое могло бы лечь в основу процедур семантического анализа БДМ и запросов, адресованных к ней. В связи с этим схема должна содержать все отношения, существенные с точки зрения структурных ограничений. Для анализа структурных ограничений (статический анализ) модели и ее фрагментов в рамках систем на основе концептуального подхода используются шаблоны, проверка корректности которых опирается на анализ ограничений на входы — выходы процессов (объектов) владельцев шаблонов и подчиненных им элементов. Таким образом, представляется целесообразным в РКМ в качестве базовых явно задавать отношения типа «вход — выход» (в частности, RP, PR, OR, RO), а не отношения ОР и РО, как в СКМ.

Несмотря на изменение отношений в схеме КМПО, все процедуры анализа, применимые к СКМ, можно использовать и в РКМ, поскольку отношения ОР и РО выводимы из отношений RP, PR, OR, RO, что подтверждают следующие теоремы.

Теорема 1. Сечения отношения «объект — порождающие его выходные данные процессы» по каждому объекту из области определения этого отношения есть сечения отношения PR^{-1} по подмножеству множества R, представляющего собой сечение отношения OR по этому объекту:

$$op(o_i) = pr^{-1}(or(o_i)).$$

По определению для любого объекта $o_i \in O$ $or(o_i)$ есть множество ресурсов, являющихся выходными по отношению к рассматриваемому объекту. С другой стороны, зная код ресурса r_j , выходного по отношению к объекту o_i , можно определить процесс, для которого ресурс r_j будет выходным, с помощью функции, обратной функции pr . Пусть в результате получен процесс p_k : $p_k = pr^{-1}(r_j)$.

Если взять сечение отношения PR^{-1} по всему множеству $or(o_i)$, то результирующее множество будет требуемым сечением. Поэтому требуемое сечение можно найти по следующему соотношению:

$$op(o_i) = pr^{-1}(or(o_i)) \blacksquare. \quad (5)$$

Теорема 2. Сечения отношения «процесс — создающие его входные данные объекты» по каждому процессу из области определения этого отношения есть сечения отношения OR^{-1} по подмножеству множества R, представляющего собой сечение отношения RP по этому процессу:

$$po(p_i) = or^{-1}(rp(p_i)).$$

Для любого процесса $p_i \in P$ $rp(p_i)$ есть множество входных ресурсов. Зная код ресурса r_j , являющегося входным по отношению к p_i , можно определить объект, который производит данный ресурс, с помощью функции, обратной функции or . Пусть в результате получен объект o_k : $o_k = or^{-1}(r_j)$.

Если взять сечение отношения OR^{-1} по всему множеству $rp(p_i)$, то требуемое по условию теоремы множество определится по следующему соотношению:

$$po(p_i) = or^{-1}(rp(p_i)) \blacksquare. \quad (6)$$

Таким образом, полученные соотношения (1–6) связывают сечения дополнительных отношений РКМ с параметрами основных отношений, которые строятся при конструировании СКМ, что необходимо

для подготовки и выполнения имитации. Эти соотношения определяют структуру полей БДМ, которые необходимо автоматически генерировать в ходе имитации, и должны учитываться при разработке СЦД компонентов ППК, чтобы обеспечить синхронизацию структур СКМ и баз данных системы моделирования.

Заключение

В настоящей работе предложены и обоснованы основные средства формализации требований к компьютерным программам-исполнителям (ситуационным цифровым двойникам) элементов модели критических инфраструктур на примере промышленно-природных комплексов в рамках ранее разработанной ситуационной системы моделирования, ориентированные на ее децентрализованную реализацию в специализированном секторе Интернета для решения задач управления безопасностью и жизнеспособностью этих критически важных объектов. Такая спецификация может позволить избежать типичной для компьютерных задач начальной стадии хаотического развития программно-аппаратного обеспечения новых структурных проектов в этой области и ускорить процесс саморазвития, необходимый для эффективного объединения разработок цифровых двойников критически важных объектов и инфраструктур различных типов в рамках единой системы распределенных ситуационных центров. Полученные результаты нашли применение при разработке агрегированных цифровых моделей многоуровневой оценки и анализа надежности и устойчивости критически важных объектов [16, 17], а также виртуального когнитивного центра управления региональной безопасностью [18]. Это позволило количественно оценить общие риски и ситуационные факторы, влияющие на жизнеспособность критических инфраструктур различных категорий, полезные эффекты и возможные потери при реализации превентивных мер на всех стадиях управления инфраструктурными объектами.

Список источников

1. Стратегия научно-технологического развития Российской Федерации (Указ Президента РФ № 145 от 28.02.2024 г.) [Электронный ресурс]. URL: http://static.kremlin.ru/media/events/files/ru/HHNAzT11_guvX9Y00_yaFA4KkMWPуYcWS8.pdf (дата обращения: 10.10.2024).
2. Стратегия национальной безопасности Российской Федерации (Указ Президента РФ № 400 от 02.07.2021 г.) [Электронный ресурс]. URL: <http://www.kremlin.ru/acts/bank/47046/page/1> (дата обращения: 10.10.2024).
3. Федеральный закон от 26.07.2017 г. № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» [Электронный ресурс]. URL: <http://static.kremlin.ru/media/acts/files/0001201707260023.pdf> (дата обращения: 10.10.2024).
4. Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. Digital Twin in manufacturing: A categorical literature review and classification // IFAC-PapersOnLine. — 2018. — Т. 51. — № 11. — С. 1016–1022.
5. Lampropoulos G., Larrucea X., Colomo-Palacios R. Digital Twins in Critical Infrastructure // Information. — 2024. — Vol. 15, Iss. 8. — Article no.: 454.
6. Tundis A., Ramírez-Agudelo O. H. Safeguarding Critical Infrastructures with Digital Twins and AI // Proceedings of the 23rd International Conference on Modeling and Applied Simulation (MAS 2024). — 2024. — Article no.: 010.
7. Tran M. Q., Sousa H. S., Teixeira E., Matos J. C., Dang H. T. Digital Twin - solution in the digital age for improving critical infrastructure resilience to extreme events // Life-Cycle of Structures and Infrastructure Systems. 1st Edition. In Biondini and Frangopol (Eds). — CRC Press, 2023. — pp. 4147–4154
8. Цифровые двойники в высокотехнологичной промышленности: монография / под. ред. А. И. Боровкова. СПб.: ПОЛИТЕХ-ПРЕСС, 2022. 492 с.
9. Fridman Alexander. Situational Modeling: Definitions, Awareness, Simulation. USA: Nova Science Publishers, Inc., 2023. — 331 p.

10. Некрасова Т. И. Задача Коши для многомерного разностного уравнения в конусах целочисленной решетки // Журнал Сибирского федерального университета. Математика и физика. 2012. № 5 (4). С. 576–580.
11. Digital Twin: Transforming How We Make Sense of Data // Электрон. дан. — 2022. [Электронный ресурс]. URL: <https://www.ptc.com/ru/industry-insights/digital-twin> (дата обращения: 10.10.2024).
12. Фридман А. Я. Опыт интеллектуализации методов ситуационного моделирования дискретных нестационарных пространственных объектов // Автоматика и телемеханика. 2022. № 6. С. 151–168.
13. Фридман А. Я. Экспертное пространство для ситуационного моделирования промышленно-природных систем // Вестник Московского университета им. С. Ю. Витте. 2014. № 1 (4). С. 233–245.
14. Фридман А. Я. Интеграция концептов цифровых двойников и ситуационной осведомленности в системе моделирования критических инфраструктур // Информационные и математические технологии в науке и управлении. 2022. № 1 (25). С. 70–78.
15. Фридман А. Я. Ситуационное управление структурой промышленно-природных систем. Методы и модели. Saarbrücken, Germany: LAMBERT Academic Publishing, 2015.
16. Masloboev A. V. An index-based method for integral estimation of regional critical infrastructure resilience using fuzzy calculations. Part 1. Problem statement and method generic structure // Reliability and quality of complex systems. — 2024. — no. 1 (45). — pp. 124–141.
17. Masloboev A. V. An index-based method for integral estimation of regional critical infrastructure resilience using fuzzy calculations. Part 2. Resilience capacity models and backbone capabilities // Reliability and quality of complex systems. — 2024. — no. 3 (47).
18. Маслобоев А. В. Виртуальные когнитивные центры как интеллектуальные системы для информационной поддержки управления региональной безопасностью // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 167–170.

References

1. Strategy for scientific and technological development of the Russian Federation (Decree of the President of the Russian Federation No. 145 of February 28, 2024). Available at: <http://static.kremlin.ru/media/events/files/ru/HHNAzTI1guvX9Y00yaFA4KkMWPYcWS8.pdf>. (accessed 10.10.2024)
2. National Security Strategy of the Russian Federation (Decree of the President of the Russian Federation No. 400 of July 2, 2021). Available at: <http://www.kremlin.ru/acts/bank/47046/page/1>. (accessed 10.10.2024)
3. Federal Law of July 26, 2017 No. 187-FL “On the security of critical information infrastructure of the Russian Federation”. Available at: <http://static.kremlin.ru/media/acts/files/0001201707260023.pdf>. (accessed 10.10.2024)
4. Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*. 2018. Т. 51, № 11. С. 1016–1022.
5. Lampropoulos G., Larrucea X., Colomo-Palacios R. Digital Twins in Critical Infrastructure. *Information*. 2024. Vol. 15, Iss. 8. Article no.: 454.
6. Tundis A., Ramírez-Agudelo O. H. Safeguarding Critical Infrastructures with Digital Twins and AI. *Proceedings of the 23rd International Conference on Modeling and Applied Simulation (MAS 2024)*. 2024. Article no.: 010.
7. Tran M. Q., Sousa H. S., Teixeira E., Matos J. C., Dang H. T. Digital Twin - solution in the digital age for improving critical infrastructure resilience to extreme events. *Life-Cycle of Structures and Infrastructure Systems. 1st Edition. In Biondini and Frangopol (Eds)*. CRC Press, 2023. pp. 4147–4154
8. Digital Twins in the High-Technology Manufacturing Industry : monograph / ed. by A. I. Borovkov. St. Petersburg: POLYTECH-PRESS, 2022. 492 p.
9. Fridman A. *Situational Modeling: Definitions, Awareness, Simulation*. USA: Nova Science Publishers, Inc., 2023. 331 p.

10. Nekrasova T. I. The Cauchy Problem for a Multidimensional Difference Equation in Cones of an Integer Lattice. *Journal of the Siberian Federal University. Mathematics and Physics*. 2012. no. 5 (4). P. 576–580.
11. Digital Twin: Transforming How We Make Sense of Data. *Electronic data*. 2022. Available at: <https://www.ptc.com/ru/industry-insights/digital-twin>. (accessed 10.10.2024)
12. Fridman A. Ya. Experience of intellectualization of methods of situational modeling of discrete non-stationary spatial objects. *Automation and Telemekhanics*. 2022. no. 6. P. 151–168.
13. Fridman A. Ya. Expert space for situational modeling of industrial and natural systems. *Bulletin of the Moscow University named after S. Yu. Witte*, no. 1(4), 2014, pp. 233–245.
14. Fridman A. Ya. Integration of Digital Twin and Situational Awareness Concepts in a Critical Infrastructure Modeling System. *Information and Mathematical Technologies in Science and Management*. 2022. no. 1(25). P. 70–78.
15. Fridman A. Ya. Situational Management of the Structure of Industrial-Natural Systems. Methods and Models. Saarbrücken, Germany: LAMBERT Academic Publishing, 2015.
16. Masloboev A. V. An index-based method for integral estimation of regional critical infrastructure resilience using fuzzy calculations. Part 1. Problem statement and method generic structure. *Reliability and quality of complex systems*. 2024. no. 1(45). pp. 124–141.
17. Masloboev A. V. An index-based method for integral estimation of regional critical infrastructure resilience using fuzzy calculations. Part 2. Resilience capacity models and backbone capabilities. *Reliability and quality of complex systems*. 2024. no. 3(47).
18. Masloboev A. V. Virtual cognitive centers as intelligent systems for management information support of regional security. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2014. no. 2(90). pp. 167–170.

Информация об авторах

А. Я. Фридман — доктор технических наук, ведущий научный сотрудник;
А. В. Маслобоев — доктор технических наук, ведущий научный сотрудник.

Information about the authors

A. Ya. Fridman — Doctor of Science (Tech.), Leading Research Fellow;
A. V. Masloboev — Doctor of Science (Tech.), Leading Researcher.

Статья поступила в редакцию 01.10.2024; одобрена после рецензирования 10.10.2024; принята к публикации 21.10.2024.
The article was submitted 01.10.2024; approved after reviewing 10.10.2024; accepted for publication 21.10.2024.

Научная статья
УДК 004.912, 004.822
doi:10.37614/2949-1215.2024.15.3.004

ФОРМИРОВАНИЕ МНОГОСЛОЙНЫХ ГРАФОВ ЗНАНИЙ НА БАЗЕ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ТЕКСТОВ

**Вадим Константинович Пимешков^{1✉}, Марина Леонидовна Никонорова²,
Максим Геннадьевич Шишаев³, Иван Геннадьевич Вишняков⁴**

¹⁻⁴Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия

¹v.pimeshkov@ksc.ru✉, <https://orcid.org/0000-0001-7010-230X>

²m.nikonorova@ksc.ru, <https://orcid.org/0000-0002-4358-2683>

³m.shishaev@ksc.ru, <https://orcid.org/0000-0001-7070-7878>

⁴i.vishnyakov@ksc.ru, <https://orcid.org/0009-0003-4938-5693>

Аннотация

В работе рассматривается проблематика формирования многослойных графов знаний для использования при интеллектуальном анализе данных социальных медиа. Технология основана на «мягкой» структуризации имеющегося пула документов в соответствии с построенной тематической моделью. Проведено экспериментальное опробование технологии и сравнение формальных свойств графовых моделей знаний, полученных с помощью тематического моделирования и традиционным способом. Результаты сравнения указывают на потенциальную эффективность использования многослойных графов знаний, построенных на базе тематической модели текстов, при реализации интеллектуальных процедур обработки данных в условиях множественности и динамичности предметных областей.

Ключевые слова:

граф знаний, тематическая модель, социальные медиа

Благодарности:

исследование выполнено в рамках государственного задания Института информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук от Министерства науки и высшего образования Российской Федерации, тема научно-исследовательской работы «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер 122022800551-0).

Для цитирования:

Пимешков В. К., Никонорова М. Л., Шишаев М. Г., Вишняков И. Г. Формирование многослойных графов знаний на базе тематического моделирования текстов // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 50–60. doi:10.37614/2949.1215.2024.15.3.004.

Original article

FORMATION OF MULTILAYER KNOWLEDGE GRAPHS BASED ON THEMATIC TEXTS MODELING

Vadim K. Pimeshkov^{1✉}, Marina L. Nikonorova², Maxim G. Shishaev³, Ivan G. Vishnyakov⁴

¹⁻⁴Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia

¹v.pimeshkov@ksc.ru✉, <https://orcid.org/0000-0001-7010-230X>

²m.nikonorova@ksc.ru, <https://orcid.org/0000-0002-4358-2683>

³m.shishaev@ksc.ru, <https://orcid.org/0000-0001-7070-7878>

⁴i.vishnyakov@ksc.ru, <https://orcid.org/0009-0003-4938-5693>

Abstract

The paper considers the problem of forming multilayer knowledge graphs for use in intelligent analysis of social media data. The technology is based on the «soft» structuring of the existing document pool in accordance with the constructed topic model. Experimental testing of the technology and comparison of the formal properties of knowledge graph models obtained using topic modeling and the traditional method were conducted. The results of the comparison indicate the potential effectiveness of using multilayer knowledge graphs built on the basis of a topic model of texts when implementing intelligent data processing procedures in the conditions of multiple and dynamic subject areas.

Keywords:

knowledge graph, topic model, social media

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic “Methodology for creating information and analytical systems to support the management of regional development based on formative artificial intelligence and big data” (registration number of the research topic 122022800551-0).

For citation:

Pimeshkov V. K., Nikonorova M. L., Shishaev M. G., Vishnyakov I. G. Formation of multilayer knowledge graphs based on thematic texts modeling // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 50–60. doi:10.37614/2949.1215.2024.15.3.004.

Введение

Графы знаний (ГЗ) являются эффективным способом формализованного представления знаний, подразумевающих использование в автоматизированных интеллектуальных информационных системах различного назначения. Привлекательной стороной ГЗ является тот факт, что аккумулируемые и систематизируемые в их рамках знания извлекаются из источников данных, в противовес общепринятому подходу к построению онтологий, где система знаний формируется априорно экспертами и затем используется при формировании и обработке наборов данных. То есть в случае ГЗ данные играют роль источника эталонных структур понятий и отношений в системе знаний. Это обуславливает высокий уровень актуальности содержащихся в ГЗ знаний и их адаптацию к динамике пользовательских представлений о предметной области.

Проблемным местом ГЗ является то, что при извлечении знаний из данных формируется мультипредметная система знаний, охватывающая в общем случае множество предметных областей. Это порождает проблему множественности возможных (в том числе противоречивых) интерпретаций знаний, а также усложняет техническую задачу оперирования системой знаний в силу ее большого объема. Решением этой проблемы является построение тематизированных (предметно-ориентированных) ГЗ, однако это, в свою очередь, сужает область применимости формируемого ГЗ, что является неприемлемым в некоторых прикладных задачах. Компромиссом между специфичностью и универсальностью ГЗ является ГЗ, имеющий внутреннее разбиение на подграфы, ориентированные на ту или иную предметную область (в общем случае — на специфику интерпретации знаний, требуемую для эффективного решения прикладных задач в предполагаемой сфере применения ГЗ). Такие ГЗ именуется многослойными.

В настоящей работе рассматривается проблематика построения многослойных ГЗ в контексте прикладной задачи мониторинга социальных медиа с целью анализа контента информационных коммуникаций пользователей. Эта задача является характерным примером проблемной области, описываемой множественностью и динамичностью систем семантических понятий и отношений, используемых пользователями. Для автоматизации и интеллектуализации обработки данных в таких условиях требуются адекватные инструменты. Главной идеей работы является использование тематической модели корпуса документов в качестве основы для секционирования ГЗ в соответствии с наиболее адекватными коммуницирующим сообществам стереотипными интерпретациями понятий из различных предметных областей. Цель исследования — оценка эффекта от использования тематического моделирования при построении многослойного ГЗ в сравнении с ГЗ, построенным традиционным способом.

Графы знаний как инструмент интеллектуальной обработки данных предметной области

ГЗ описывает концепты или сущности реального мира и отношения между ними способом, близким к представлению (мышлению) человека, и в виде, пригодном для компьютерной обработки. Формально ГЗ можно представить в виде [1]:

$$G := (V, E, r, \Sigma_V, \Sigma_E, l_V, l_E),$$

где V — множество вершин, выраженное уникальными концептами или терминами; E — множество ребер, заданное выделенными между концептами отношениями; $r: E \rightarrow \{\{x, y\}: x, y \in V\}$ — функция, присваивающая каждому ребру неупорядоченную пару вершин; Σ_V — множество меток вершин; Σ_E — множество меток ребер; $l_V: V \rightarrow \Sigma_V$ — функция, определяющая метки вершин; $l_E: E \rightarrow \Sigma_E$ — функция, определяющая метки ребер.

Иногда о ГЗ говорят как о семантических сетях или сетях знаний, извлеченных из данных реального мира. В настоящее время ГЗ используются в задачах поиска информации, вопросно-ответных системах, системах поддержки принятия решений, рекомендательных системах [2]. ГЗ, как и другие модели знаний, могут быть двух типов — общего назначения и специфичные для предметной области. Первый тип, как правило, представляет собой довольно обширные графы, содержащие знания из множества предметных областей (например, Wikidata, DBpedia и др.), в то время как второй тип ориентирован на более узкую область или отрасль или на решение конкретной задачи.

ГЗ неразрывно связаны с технологией их построения и использования, включающей этапы извлечения знаний (knowledge extraction), слияния знаний (knowledge fusion), обработки знаний (knowledge processing) [3].

Извлечение знаний — первостепенная и основная задача при построении ГЗ, в рамках которой из необработанных данных извлекаются различные сущности, атрибуты и отношения. Слияние знаний предполагает объединение одинаковых объектов из различных источников для получения более точной и согласованной информации, что способствует поддержанию ГЗ в актуальном состоянии. Данный этап включает в себя подзадачу согласования объектов (entity alignment) для обнаружения семантически одинаковых объектов в разных источниках и подзадачу устранения неоднозначности объектов (entity disambiguation) для сопоставления объектов из входных данных с соответствующими уникальными объектами в целевом ГЗ. Обработка знаний предполагает обработку простых фактов и формирование структурированных систем знаний с данными высокого качества. Обычно включает в себя построение онтологий, оценку качества и иногда логический вывод.

Говоря о применении ГЗ для решения различных задач, так или иначе связанных с анализом данных социальными медиа, можно выделить следующие работы. Авторы работы [4] представили ГЗ, построенный на основе данных социальных сетей, для обнаружения фейковых новостей. В основе графа лежит онтология Fandet, предназначенная для представления сложных и часто неполных данных социальных сетей, а также облегчения их анализа. В работе [5] авторы используют ГЗ в задаче деанонимизации группы людей. Они моделируют начальные знания злоумышленника при помощи ГЗ, и в дальнейшем используют его для моделирования двух этапов атаки — деанонимизации и выведение конфиденциальной информации (privacy inference). Такая модель позволяет лучше описать процесс атаки и количественно оценить степень раскрытия конфиденциальной информации. Авторы исследования [6] строят ГЗ на основе новостных статей, посвященных COVID-19. Такой граф обеспечивает инфраструктуру для анализа данных социальных медиа, связанных с COVID-19, которая может быть полезна для исследователей, специалистов по обработке данных и организаций. В работе [7] предлагают архитектуру и прототип платформы, включающей ГЗ для поддержки работы журналистов некоторой исходной информацией в рамках вычислительной журналистики.

Тематические ГЗ, в отличие от обычных, как правило, отражают конкретную узкую тематику, например, аккумуляторы электромобиля, и содержат подробные знания предметной области, включая специализированные сущности и тройки. Авторами работы [8] предлагается фреймворк для автоматизированного построения таких «тематизированных» (theme-specific) ГЗ. Данный фреймворк принимает на вход необработанный тематический корпус и создает ГЗ, который включает в себя значимые сущности и отношения между ними в рамках заданной темы. Создание графа начинается с построения онтологии сущностей темы из «Википедии», на основе которой затем генерируется отношения-кандидаты с помощью больших языковых моделей (LLM) для построения онтологии отношений. Далее, чтобы проанализировать документы из тематического корпуса, авторы сначала сопоставляют извлеченные пары сущностей с онтологией и извлекают отношения-кандидаты. Наконец, авторы учитывают контекст употребления сущностей совместно с онтологией отношений для окончательного определения отношений между сущностями и, соответственно, построения ГЗ.

По мере появления более сложных данных и, соответственно, необходимости разработки более сложных систем знаний, свое развитие получила концепция многослойных ГЗ. Ключевой особенностью таких графов является представление знаний с помощью нескольких слоев, уровней или измерений. Каждый такой слой в графе может представлять различные типы отношений между

сущностями. Например, один слой может представлять социальные связи, другой — профессиональные отношения, а третий — взаимодействия в определенном контексте. Известным примером многослойного ГЗ является KnowWhereGraph [9]. На текущий момент данный граф содержит около 13 миллиардов триплетов и более 30 слоев, включающих пространственные данные (места, регионы), данные о населении, об экстремальных явлениях, инфраструктуре, сельскохозяйственные данные и др. Таким образом, он обеспечивает достаточную полноту и актуальность данных, необходимых для решения различных прикладных задач. В этой работе, в отличие от KnowWhereGraph, многослойность отражает разные контексты совместной встречаемости терминов. Предполагается, что такие контекстные слои зависят от характера исходных данных и решаемой задачи.

Технология формирования многослойного графа знаний на основе тематического моделирования

В рамках настоящей работы рассматривается разбиение на контексты в соответствии с тематическим распределением документов, полученным на основе тематической модели. Такое распределение позволяет посмотреть на данные с точки зрения некоторых скрытых тем, выделенных в наборе данных. Вершинами полученного графа будут являться термины, а отношения между ними будут выражаться через взвешенную направленную ассоциативную связь, полученную на основе статистики совместного употребления данных терминов, принимая во внимание результаты тематического моделирования. Предполагается, что разбиение графа на контексты, которые могут рассматриваться как слои графа или подграфы, на основе тематического распределения документов позволит выделять значимые связи, даже невзирая на малый вес темы в наборе данных. Справедливо заметить, что эффективность такого подхода будет в том числе зависеть от свойств и качества полученной тематической модели.

Формально предлагаемую технологию, использующую тематическую модель в качестве основы для разбиения на контексты (рис. 1), можно представить следующим образом.

Дано: D — набор документов различной тематики; $L \subseteq D \times D$ — асимметричное транзитивное отношение «является откликом», определяет на множестве документов структуру коммуникации. $d_1 L d_2$ означает, что документ (пост или комментарий) d_2 является откликом на документ d_1 . Структура коммуникации состоит из веток обсуждения, представляющих собой последовательности Br документов из D ($Br \subseteq D$), удовлетворяющие условию: $Br = \{d_1, \dots, d_N\}; \forall i < j, d_i L d_j$.

Полная структура коммуникации (полное дерево) задается корневым документом d и всеми транзитивно связанными с ним документами $Br(d)$.

Найти:

$$G := (V, E, K, W_K, w_E, s, t, r),$$

где V — множество вершин — уникальные термины из набора документов D ; E — множество ребер — взвешенные направленные ассоциативные отношения между терминами; K — множество меток контекстов, заданное темами тематической модели; W_K — множество весов ребер, заданное для каждого контекста; $w_E: E \rightarrow W_K$ — функция, присваивающая вес каждому ребру; $s: E \rightarrow V$ — функция, присваивающая каждому ребру начальную вершину; $t: E \rightarrow V$ — функция, присваивающая каждому ребру конечную вершину; $r: E \rightarrow K$ — функция, присваивающая каждому ребру метку контекста.

Первым этапом извлекаются термины из D следующим образом: $TermExtrMethod(D, L, \dots) \rightarrow Br^T$ — структура коммуникации Br с выделенными в ней терминами T .

Вторым этапом строится тематическая модель на основе Br^T и заданного количества тем K :

$$TopicModelConst(Br^T, K) \rightarrow \{\Theta, \Phi\},$$

где Θ — матрица, задающая распределение документов по темам; Φ — матрица, задающая распределение терминов по темам.

Третьим этапом производится построение и заполнение тематизированных матриц совместного употребления терминов: $f(Br^P, \Theta) \rightarrow CM$ — набор матриц совместного употребления, где элемент матрицы совместного употребления по k -й теме вычисляется по формуле:

$$cm_{ij}^k = \sum_{Br^P=1}^M \theta_{Br^P}^k \cdot Br_{ij}^P \cdot \left(\frac{\phi_i^k}{\max(\phi_i)} + \frac{\phi_j^k}{\max(\phi_j)} \right), \quad (1)$$

где $\theta_{Br^P}^k$ — значение принадлежности структуры Br^P теме k ; Br_{ij}^P — значение совместного употребления терминов i и j в рамках Br , получаемое функцией подсчета совместного употребления терминов в структуре коммуникации $fp(Br^T) \rightarrow Br^P$ — вида $\langle i, j, c_{ij} \rangle$, где i, j — термины, c_{ij} — значение совместного употребления данных терминов в конкретной структуре Br ; ϕ_i^k — значение принадлежности термина i теме k ; $\max(\phi_i)$ — максимальное значение принадлежности термина i теме.

В результате получается набор матриц со значениями совместной встречаемости соответствующих терминов. Набор уникальных терминов интерпретируется как вершины графа, а ненулевые значения матриц совместной встречаемости — как взвешенные ассоциативные отношения между ними.

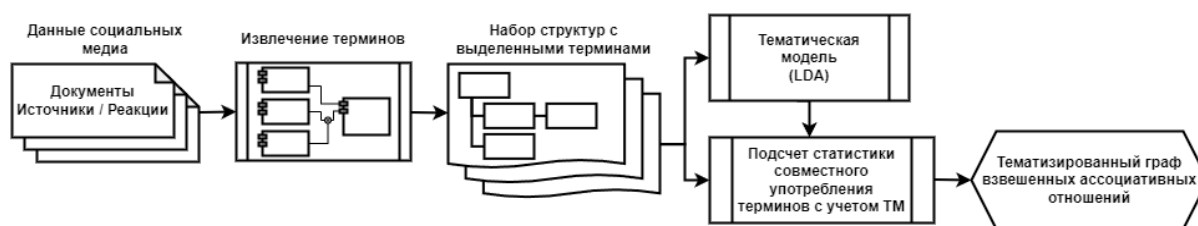


Рис. 1. Технология формирования многослойного ГЗ на основе тематического моделирования

Оценка эффективности применения тематического моделирования для формирования ГЗ

Для проверки предлагаемой технологии формирования многослойного ГЗ был использован датасет публикаций с комментариями социальной сети «ВКонтакте» из 8 групп за приблизительно 2 года. Исходный датасет содержал 52 375 публикации и 235 365 комментариев, из которых содержат текст 50 067 и 221 834 соответственно.

Очистка текста в датасете от нежелательных символов таких, как эмодзи, повторения различных знаков и конструкций, например, телефонных номеров, ссылок или идентификаторов пользователей, проводилась в два этапа. На первом этапе для очистки от нежелательных конструкций использовался набор регулярных выражений. На втором — белый список символов, состоящий из кириллицы, латиницы, цифр и ряда специальных символов.

Для дальнейшей обработки датасет был преобразован в набор структур коммуникации, где каждая структура является представлением публикации и ее откликов (комментариев). Вместе с этим преобразованием проводилось выделение терминов в этих структурах с помощью авторского комбинированного метода извлечения терминов [10]. В результате работы метода в документах были выделены термины, входящие в заранее заданный словарь, именованные сущности, осмысленные биграммы и триграммы, а также контекстно важные униграммы-существительные. Полученные структуры дополнительно были очищены от редко встречающихся терминов. В результате был получен набор из 45 363 структур, содержащих хотя бы два уникальных термина, а также список уникальных терминов датасета. Для обучения тематической модели и дальнейшего ее использования при построении матриц совместного употребления был выполнен препроцессинг терминов: приведение в нижний регистр, стемминг (для приведения в единую словоформу), замена пробелов на нижние подчеркивания (в многословных терминах).

На основе полученных документов с извлеченными терминами был обучен набор из 11 тематических моделей (ТМ) (количество тем варьировалось от 5 до 15) на основе латентного размещения Дирихле (Latent Dirichlet allocation, LDA), реализованной в библиотеке Gensim [11]. При обучении всех моделей использовались следующие параметры: `update_every = 1`, `chunksize = 1000`, `passes = 6`. Для полученных моделей была вычислена метрика согласованности (coherence), которая показывает, насколько выявленные темы значимы и интерпретируемы. Результаты подсчета метрики приведены на рис. 2. Таким образом, в дальнейшей работе было принято решение использовать модель с 7 темами, т. к. она имеет наибольший показатель согласованности тем.

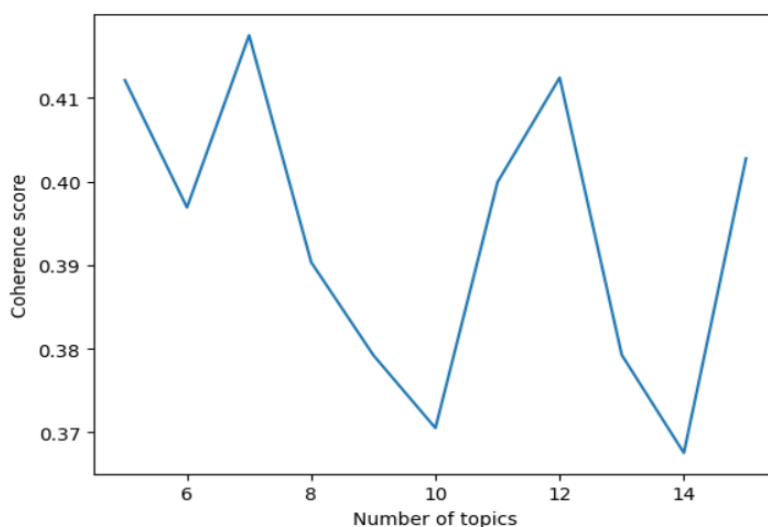


Рис. 2. Изменение метрики согласованности в зависимости от количества тем

Далее на основе набора структур и выбранной тематической модели был произведен подсчет статистики совместного употребления терминов. Значения совместного употребления рассчитывались отдельно для каждой структуры методом, в рамках которого предполагалось, что, во-первых, термины связаны между собой в рамках каждого документа, а, во-вторых, термины источника приводят к появлению термина в реакции и, соответственно, термины источника связаны с каждым термином в реакциях. При этом пары идентичных терминов игнорировались для исключения петель в графе. Пример расчета совместного употребления таким методом приведен на рис. 3. Далее значения совместного употребления (Br_{ij}^P), в соответствии с формулой (1), умножались на соответствующий компонент вектора распределения документа по темам ($\theta_{Br,P}^k$) и на сумму коэффициентов влияния терминов. Полученные значения фиксировались в наборе матриц размером $N \times N$, где N — это число уникальных терминов в датасете (13 171), в ячейке, соответствующей текущей теме и терминам. Таким образом, мы получили набор матриц, в котором каждая матрица отражает определенный слой графа, соответствующий одной из тем, выделенных тематической моделью.

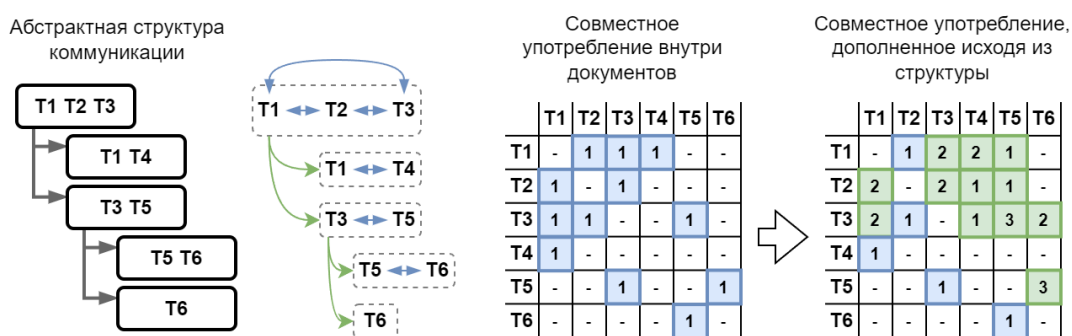


Рис. 3. Пример работы метода подсчета совместного употребления терминов

Для того, чтобы оценить эффект от разбиения графа на темы, был построен аналогичный граф, но без применения тематической модели. Он строился без учета распределения документов по темам, соответственно, в ячейках матрицы фиксировались просто значения совместного употребления терминов без дополнительных коэффициентов.

В обоих случаях, с применением ТМ и без, было замечено, что в матрицах преобладают значения совместного употребления близкие к 0. Соответствующие отношения, со значениями близкими к нулю, говорят о том, что совместное употребление значительного количества терминов наблюдалось всего лишь один или пару раз во всем датасете, следовательно, эта связь не является значимой в рамках рассматриваемой модели знаний. Поэтому было принято решение исключить из матриц такие незначимые связи путем выделения значимых связей.

Для выделения значимых связей из тематизированной матрицы на первом этапе для каждой темы был выбран процент самых сильных связей, значение которого соответствует доле терминов в конкретной теме (значения получены на этапе обучения тематической модели). На втором этапе была проведена оценка значимости в соответствии с эвристически выведенным порогом значимости:

$$Br_k^P(i, j) > \bar{x} \left(\bar{x} \left(Br_k^P(i, \cdot) \right), \bar{x} \left(Br_k^P(\cdot, j) \right), \bar{x} \left(Br_k^P \right) \right),$$

где $\bar{x} \left(Br_k^P(i, \cdot) \right)$ — среднее значение совместного употребления для термина i ; $\bar{x} \left(Br_k^P(\cdot, j) \right)$ — среднее значение совместного употребления для термина j ; $\bar{x} \left(Br_k^P \right)$ — среднее значение совместного употребления для всех терминов в соответствующей матрице (теме) k .

При этом среднее значение совместного употребления для термина рассчитывалось как среднее суммы векторов входящих и исходящих отношений. Для матрицы без ТМ применялась только фильтрация по порогу значимости. В результате такой фильтрации отношений их количество значительно сократилось — с 754 993 до значений, указанных в табл. 1. Там же отражено изменение количества уникальных терминов с хотя бы одной ненулевой связью для каждой матрицы соответственно. До фильтрации в каждой матрице было 11 889 таких терминов.

Таблица 1

Количество отношений и терминов в полученных матрицах после фильтрации

Матрица	Количество отношений	Количество терминов
Без ТМ	117716	6592
Тема № 1	7475	520
Тема № 2	25215	2536
Тема № 3	14555	1875
Тема № 4	16356	1912
Тема № 5	27987	2832
Тема № 6	12004	1390
Тема № 7	9827	968

Стоит заметить, что без фильтрации вовсе мы могли получить практически полностью связанный граф, что не позволит эффективно отслеживать какие-либо связи в такой структуре. С другой стороны, при слишком строгой фильтрации мы однозначно рискуем удалить нужные связи и, соответственно, термины. В данном случае способ фильтрации полагался на специфику анализируемых данных, поэтому оценка его эффективности остается открытой задачей.

По факту наличия отношений между терминами полученные тематические матрицы в значительной степени пересекаются с матрицей без ТМ (рис. 4). Это говорит о том, что после удаления незначимых отношений предложенным порогом фильтрации матрицы все еще близки по структуре, и большая часть связей, присутствующих в тематизированном наборе матриц, также присутствуют и в матрице без ТМ. При этом в некоторых из них все же есть отношения, которые были удалены при фильтрации из матрицы без ТМ. Этот факт, возможно, подтверждает гипотезу о том, что тематизированные ГЗ способствуют выделению некоторых связей на общем фоне.

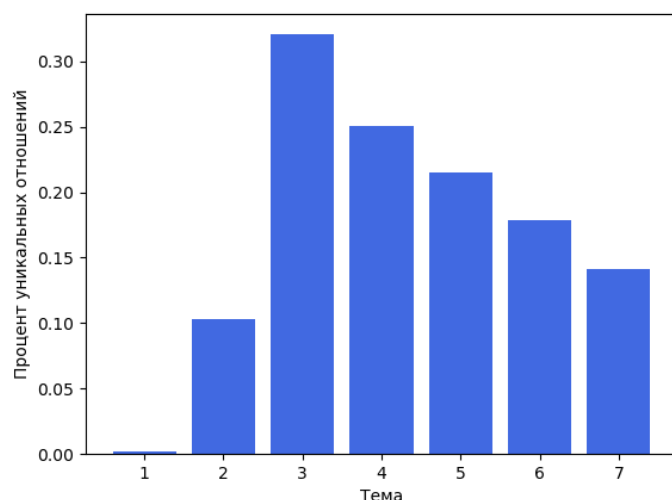


Рис 4. Процент уникальных связей относительно матрицы без тематической модели

Полученные тематизированные матрицы в среднем по факту наличия связи пересекаются не более чем на 5–6 %, а, значит, вероятнее всего, отражают разные ассоциации пользователей. Построенная матрица корреляций (рис. 5) показывает процент общих связей между тематическими слоями матрицы.

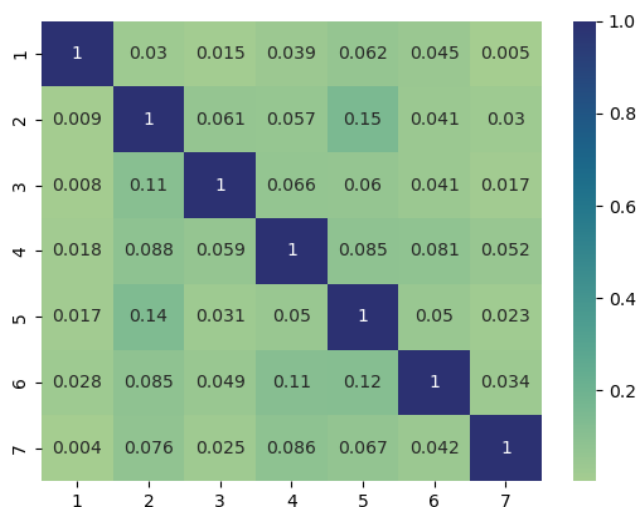


Рис 5. Пересечение отношений между темами (относительно строки)

Полученные матрицы далее были преобразованы в направленные мультиграфы с помощью библиотеки NetworkX [12] следующим образом: вершинами стали уникальные термины, у которых имеется хотя бы одна связь, а ребрами — ненулевые значения матриц по слоям соответственно. Следовательно, данный набор мультиграфов представляет собой тематизированный или многослойный ГЗ.

Для оценки полученных мультиграфов были рассчитаны взвешенные степени их вершин и плотности с помощью методов, реализованных в упомянутой ранее библиотеке. Степень вершины соответствует количеству ребер, инцидентных данной вершине, а взвешенная степень вершины — это сумма весов ребер, инцидентных данной вершине. В табл. 2 по каждому мультиграфу, с применением ТМ и без, представлены топ-5 вершин по их взвешенной степени, на основе которых можно предположить об основных терминах, вокруг которых ведется дискурс в социальной сети в рамках определенной темы.

Таблица 2

Топ-5 вершин по их взвешенной степени

Без ТМ	Апатиты	Мурманск	Кировск	Город	Мурманская область
	259 387	133 795	131 710	97 951	96 640
Слой 1	Средств	Действ	Антибиотик	Препарат	Фильм
	36 901,47	34 609,02	22 508,20	21 022,29	19 672,85
Слой 2	Апатиты	Город	Кировск	Кола	Улиц
	216 725,28	85 487,75	85 300,10	51 679,57	35 160,40
Слой 3	Квартир	Ремонт	Дом	Комнатн квартир	Апатиты
	29 639,01	16 936,36	11 908,89	11 181,37	10 144,00
Слой 4	Работ	Светл памя	Кировск	Апатиты	Вечн памя
	16 446,07	3 964,45	3 423,50	3 339,41	3 090,55
Слой 5	Мурманск	Мурманская область	Област	Росс	Мест
	82 211,37	63 065,60	54 506,15	22 862,24	18 227,96
Слой 6	Покупател	Препарат	Работник	Прав	Организац
	3 083,60	2 959,94	2 699,43	2 543,58	2 202,78
Слой 7	Имандра	Озер	Магазин	Водоем	Приток
	25 387,04	18 612,45	10 106,22	8 140,36	6 797,95

Плотность графа — это отношение числа ребер к максимально возможному. На основе вычисленных плотностей (рис. 6), где максимальное значение соответствует всего лишь 0,013, можно предположить, что полученные графы представляют собой леса.

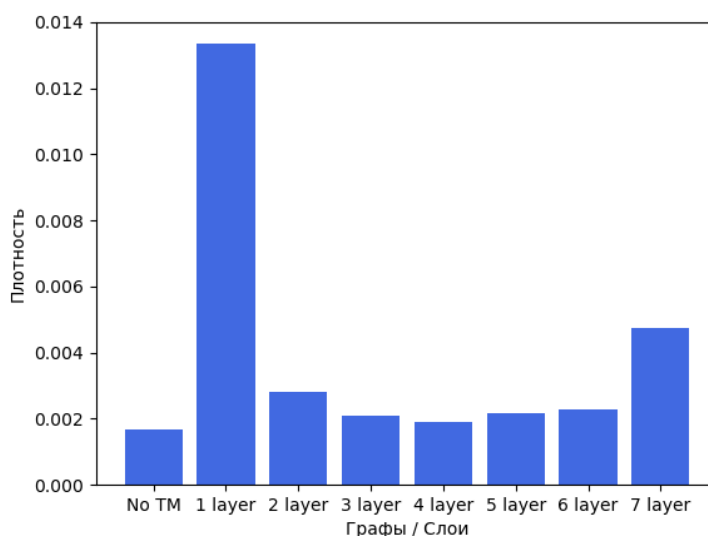


Рис. 6. Показатели плотности полученных графов

Заключение

В результате проделанной работы была разработана и опробована на практике технология формирования многослойных ГЗ на основе данных социальных медиа. В качестве базы для экспериментов использовался датасет публикаций с комментариями социальной сети «ВКонтакте», включающий 52 375 публикации и 235 365 комментариев.

Сравнение ГЗ, полученных по предложенной методике с ГЗ, сформированными традиционным способом, показывает, что в первом случае знания имеют лучшую структуризацию. Это создает предпосылки для реализации более эффективных процедур обработки данных социальных медиа с точки зрения объемов вычислений и корректности результатов.

Список источников

1. Waclawski K. et al. Ontology summit 2020 communiqué: Knowledge graphs // *Applied Ontology*. 2021. Vol. 16, № 2. P. 229–247.
2. Zou X. A Survey on Application of Knowledge Graph // *J. Phys.: Conf. Ser. IOP Publishing*, 2020. Vol. 1487, № 1. P. 012016.
3. Tian L. et al. Knowledge graph and knowledge reasoning: A systematic review // *Journal of Electronic Science and Technology*. 2022. Vol. 20, № 2. P. 100159.
4. Hani A.B. et al. Fane-KG: A Semantic Knowledge Graph for Context-Based Fake News Detection on Social Media // *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*. 2020. P. 1–6.
5. Qian J. et al. Social Network De-Anonymization and Privacy Inference with Knowledge Graph Model // *IEEE Transactions on Dependable and Secure Computing*. 2019. Vol. 16, № 4. P. 679–692.
6. Al-Obeidat F. et al. Cone-KG: A Semantic Knowledge Graph with News Content and Social Context for Studying Covid-19 News Articles on Social Media // *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*. 2020. P. 1–7.
7. Berven A. et al. A knowledge-graph platform for newsrooms // *Computers in Industry*. 2020. Vol. 123. P. 103321.
8. Ding L. et al. Automated Construction of Theme-specific Knowledge Graphs: arXiv:2404.19146. arXiv, 2024.
9. Janowicz K. et al. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence // *AI Magazine*. 2022. Vol. 43, № 1. P. 30–39.
10. Пимешков В. К., Никонорова М. Л., Шишаев М. Г. Комбинированный метод извлечения терминов для задачи мониторинга тематических обсуждений в социальных медиа // *Информатика и автоматизация*. 2024. Т. 23, № 4. С. 1110–1138.
11. Gensim: Topic modelling for humans [Электронный ресурс]. URL: <https://radimrehurek.com/gensim/> (дата обращения: 05.10.2024).
12. NetworkX [Электронный ресурс]. URL: <https://github.com/networkx/networkx> (дата обращения: 05.10.2024).

References

1. Waclawski K. et al. Ontology summit 2020 communiqué: Knowledge graphs. *Applied Ontology*, 2021, vol. 16, no. 2, pp. 229–247.
2. Zou X. A Survey on Application of Knowledge Graph. *Journal of Physics: Conference Series*, 2020, vol. 1487, no 1, p. 012016.
3. Tian L. et al. Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, 2022, vol. 20, no. 2, p. 100159.
4. Hani A.B. et al. Fane-KG: A Semantic Knowledge Graph for Context-Based Fake News Detection on Social Media. *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–6.
5. Qian J. et al. Social Network De-Anonymization and Privacy Inference with Knowledge Graph Model. *IEEE Transactions on Dependable and Secure Computing*, 2019, vol. 16, no. 4, pp. 679–692.
6. Al-Obeidat F. et al. Cone-KG: A Semantic Knowledge Graph with News Content and Social Context for Studying Covid-19 News Articles on Social Media. *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2020, pp. 1–7.
7. Berven A. et al. A knowledge-graph platform for newsrooms. *Computers in Industry*, 2020, vol. 123, pp. 103321.
8. Ding L. et al. Automated Construction of Theme-specific Knowledge Graphs. arXiv:2404.19146, 2024.
9. Janowicz K. et al. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence. *AI Magazine*, 2022, vol. 43, no 1, pp. 30–39.

10. Pimeshkov V.K., Nikonorova M.L., Shishaev M.G. Kombinirovannyj metod izvlecheniya terminov dlya zadachi monitoringa tematicheskikh obsuzhdenij v social'nyh media [A combined term extraction method for the problem of monitoring thematic discussions in social media]. *Informatika i Avtomatizaciya* [Informatics and Automation], 2024, vol. 23, no. 4, pp. 1110–1138.
11. Gensim: Topic modelling for humans. Available at: <https://radimrehurek.com/gensim/> (accessed 05.10.2024).
12. NetworkX. Available at: <https://github.com/networkx/networkx> (accessed 05.10.2024).

Информация об авторах

В. К. Пимешков — аспирант, стажер-исследователь;
М. Л. Никонорова — аспирант, инженер-исследователь;
М. Г. Шишаев — доктор технических наук, главный научный сотрудник;
И. Г. Вишняков — аспирант, системный администратор.

Information about the authors

V. K. Pimeshkov — PhD student, Research Assistan;
M. L. Nikonorova — PhD student, Research Engineer;
M. G. Shishaev — Doctor of Science (Tech.), Chief Research Fellow;
I. G. Vishnyakov — PhD student, System Administrator.

Статья поступила в редакцию 14.10.2024; одобрена после рецензирования 28.10.2024; принята к публикации 06.11.2024.
The article was submitted 14.10.2024; approved after reviewing 28.10.2024; accepted for publication 06.11.2024.

Научная статья
УДК 004.832
doi:10.37614/2949.1215.2024.15.3.005

РЕШЕНИЕ ЗАДАЧ МАРШРУТИЗАЦИИ ТРАНСПОРТНЫХ СРЕДСТВ В РАМКАХ ПАРАДИГМЫ ПРОГРАММИРОВАНИЯ В ОГРАНИЧЕНИЯХ

Алексей Владимирович Шестаков¹, **Александр Анатольевич Зуенко²**

^{1,2}*Институт информатики и математического моделирования имени В. А. Пutilова
Кольского научного центра Российской академии наук, Апатиты, Россия*

¹*a.shestakov@ksc.ru*, <https://orcid.org/0000-0002-9052-2579>

²*a.zuenko@ksc.ru*, <https://orcid.org/0000-0002-7165-6651>

Аннотация

Статья посвящена аналитическому обзору возможностей решения задач маршрутизации транспортных средств в рамках парадигмы программирования в ограничениях. Сначала в статье приводится обзор различных постановок задачи маршрутизации транспортных средств. Затем данные постановки, для облегчения понимания, формулируются на языке целочисленного линейного программирования. После чего дается описание технологии программирования в ограничениях и приводятся типовые глобальные ограничения, полезные в различных постановках задачи маршрутизации транспортных средств. Сделан вывод о том, что гибкость в формулировке ограничений, возможность интеграции с другими методами оптимизации, а также способность учитывать неопределенности делают программирование в ограничениях одним из наиболее востребованных средств для решения задач маршрутизации.

Ключевые слова:

программирование в ограничениях, задача удовлетворения ограничений, задача маршрутизации транспортных средств

Финансирование:

работа выполнена в рамках НИР «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (регистрационный номер 122022800547-3).

Для цитирования:

Шестаков А. В., Зуенко А. А. решение задач маршрутизации транспортных средств в рамках парадигмы программирования в ограничениях // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 61–68. doi:10.37614/2949.1215.2024.15.3.005.

Original article

SOLVING VEHICLE ROUTING PROBLEMS IN THE FRAMEWORK OF THE CONSTRAINT PROGRAMMING PARADIGM

Aleksey V. Shestakov¹, **Alexander A. Zuenko²**

¹*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia*

¹*a.shestakov@ksc.ru*, <https://orcid.org/0000-0002-9052-2579>

²*a.zuenko@ksc.ru*, <https://orcid.org/0000-0002-7165-6651>

Abstract

The article is devoted to an analytical survey of the possibilities of solving vehicle routing problems within the framework of the constraint programming paradigm. First, the article provides an overview of the various formulations of the vehicle routing problem. Then, to facilitate understanding, these statements are formulated in the language of integer linear programming. After that, a description of the constraint programming technology is given and typical global constraints are given that are useful in various formulations of the vehicle routing problem. It is concluded that flexibility in the formulation of constraints, the possibility of integration with other optimization methods, as well as the ability to take into account uncertainties make programming in constraints one of the most popular tools for solving vehicle routing problems.

Keywords:

Constraint programming, constraint satisfaction problem, vehicle routing problem

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic “Development of theoretical and organizational and technical foundations of information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation” (registration number of the research topic 122022800547-3).

For citation:

Shestakov A. V., Zuenko A. A. Local search in open pit mining planning searching // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 61–68. doi:10.37614/2949.1215.2024.15.3.005.

Введение

Проблема маршрутизации транспортных средств (Vehicle Routing Problem, VRP) является одной из центральных задач в области логистики и оптимизации распределения ресурсов [1]. VRP заключается в нахождении оптимальных маршрутов для группы транспортных средств, которые должны обслуживать множество клиентов с учетом различных ограничений таких, как грузоподъемность, временные окна и множество других факторов. Неправильное решение данной задачи может привести к значительным экономическим потерям, увеличению времени доставки и снижению уровня обслуживания клиентов.

На данный момент разработано множество методов для решения VRP, включая классические оптимизационные техники такие, как линейное целочисленное программирование [2], а также различные эвристические и метаэвристические подходы [3]. Однако развитие методов программирования в ограничениях (Constraint Programming, CP) [4] значительно расширило инструментарий для решения данной задачи. Методы CP позволяют формулировать задачи в виде систем ограничений, что делает их особенно эффективными для комплексных и нестандартных сценариев, где традиционные методы могут испытывать трудности. Далее приводится аналитический обзор возможностей решения задач маршрутизации в форме задач удовлетворения ограничений.

Постановка и классификация задач маршрутизации

Задачи маршрутизации (Vehicle Routing Problems, VRP) относятся к классу комбинаторных задач оптимизации, которые связаны с поиском оптимальных маршрутов для парка транспортных средств, обслуживающих определенные объекты [5]. Основная цель VRP — минимизировать общие затраты на перевозку при выполнении всех необходимых заданий таких, как доставка товаров клиентам, сбор грузов, соблюдение временных ограничений и т. д.

Типовая VRP-задача включает в себя:

- парк транспортных средств (грузовиков, фургонов, курьеров и т. д.) с определенными характеристиками (вместимость, расход топлива, скорость);
- набор пунктов назначения (клиенты, склады, точки сбора) с различными требованиями к доставке/сбору (погрузке/выгрузке товара);
- дорожную сеть с известными расстояниями и/или временем в пути между пунктами.

Задача заключается в построении оптимальных маршрутов, по которым транспортные средства должны посетить все пункты назначения, удовлетворяя разнообразным ограничениям, при этом минимизируя общие издержки (расстояние, время, расход топлива и т. п.).

Классическая задача маршрутизации (Capacitated Vehicle Routing Problem, CVRP) [6] заключается в определении оптимальных маршрутов для обслуживания группы клиентов с помощью одного или нескольких транспортных средств, отправляющихся из общего депо.

В CVRP выделяют следующие ограничения:

- каждое транспортное средство имеет ограниченную грузоподъемность, которую оно не может превышать;
- суммарный спрос клиентов, обслуживаемых одним транспортным средством, не должен превышать его грузоподъемность;
- все клиенты должны быть обслужены в точности один раз. Нельзя пропустить ни одного клиента;
- все маршруты начинаются и заканчиваются в одном общем депо;

— все параметры задачи (спрос клиентов, расстояния/время между узлами) являются детерминированными и известными заранее.

В задаче *маршрутизации с временными окнами* (Vehicle Routing Problem with Time Windows, VRPTW) [7] к основным ограничениям CVRP также добавляется, что каждый клиент должен быть обслужен в определенный промежуток времени, называемый «временным окном». Таким образом, добавляются следующие ограничения:

- транспортное средство должно прибыть к клиенту в пределах его временного окна;
- время прибытия транспортного средства к клиенту плюс время его обслуживания не должно превышать время закрытия временного окна;
- время в пути между клиентами и депо должно также учитываться при планировании маршрутов;
- маршрут транспортного средства должен быть завершен до закрытия депо.

Для задачи *маршрутизации с ограничениями по ресурсам* (Resource-Constrained VRP, RCVRP) добавляются ограничения на различные ресурсы, необходимые для выполнения маршрутов:

- топливо/энергия — ограниченное количество топлива или энергии (заряд аккумулятора) для транспортных средств;
- время — ограничения на общее время работы, время вождения, время нахождения в пути;
- персонал — ограниченное количество водителей/экипажей для обслуживания маршрутов;
- оборудование — ограниченное количество погрузочно-разгрузочных механизмов, холодильных установок и т. д.

Задача маршрутизации транспортных средств со сбором и доставкой товаров (Vehicle Routing Problem with Pickups and Deliveries, VRPPD) [8]. В ней для каждого клиента требуется обеспечить не только доставку, но и вывоз товаров. В VRPPD необходимо определить оптимальные маршруты транспортных средств, которые обеспечивают сбор товаров у одних клиентов и их доставку другим клиентам, при этом соблюдая ограничения на грузоподъемность транспорта. Основная цель — минимизация общих затрат на перевозку. Ключевая особенность VRPPD — необходимость обеспечить соответствие между взятыми (собранными) и доставленными грузами: количество товаров, взятых у клиентов, должно соответствовать количеству товаров, доставленных другим клиентам.

Суммарная загрузка транспортного средства на маршруте (сбор + доставка) не должна превышать его грузоподъемность.

Роль программирования в ограничениях в решении VRP

В последние годы методы программирования в ограничениях (Constraint Programming, CP) становятся все более популярными в решении задач VRP. CP позволяет формулировать задачи как системы ограничений, предоставляя гибкие и мощные инструменты для обработки сложных комбинаторных задач, к которым относятся задачи VRP. Основной особенностью CP является возможность интуитивного задания и моделирования ограничений, что позволяет адаптировать подход к конкретным условиям задачи [9].

Согласно CSP (Constraint Satisfaction Problem) [10], задача представляет собой набор переменных с определенной областью возможных значений (домен) и набор некоторых ограничений. Таким образом, ограничения представляют собой отношения между переменными, в то время как задача CSP формализует, какие именно отношения должны соблюдаться для заданного множества переменных решения.

Исходя из вышеизложенного, задача удовлетворения ограничений может быть формализована в рамках инкрементного подхода следующим образом:

- исходное состояние характеризуется отсутствием значений для всех переменных;
- функция, обеспечивающая последовательность действий, требует, чтобы процесс присвоения значений исключал возможность возникновения конфликтов между переменными;
- проверка достижения цели подразумевает, что присвоенные значения должны формировать полное и согласованное решение;
- каждый этап этого процесса оценивается с фиксированной стоимостью, что упрощает анализ и оптимизацию алгоритмических процедур.

Программирование в ограничениях в первую очередь направлено на сокращение множества допустимых значений переменных решения, которые соответствуют всем установленным ограничениям. Когда происходит установление факта невозможности присвоения определенных значений переменной решения, данная информация распространяется через ограничения, что может способствовать формированию дополнительных выводов.

Для достижения окончательного решения применяются различные стратегии поиска, которые продолжаются до тех пор, пока для каждой переменной не будет назначено значение. По завершении первого успешного поиска продолжается работа над нахождением новых решений, которые обладают более оптимальными целевыми значениями.

Ключевым элементом СР является возможность представления задач в компактной и понятной форме. Однако по мере увеличения сложности проблем, связанных с множеством переменных и ограничений, возникает необходимость в более высокоуровневых конструкциях, которые могут облегчить процесс моделирования и упростить решение. Одним из наиболее эффективных инструментов, используемых для этого, являются *глобальные ограничения* [11].

Глобальные ограничения представляют собой специальные классы ограничений в рамках программирования в ограничениях, которые фиксируют связь между заранее не известным числом переменных. В отличие от простых ограничений, которые обычно накладываются на одну или две переменные и имеют специфическую форму, глобальные ограничения охватывают более широкий спектр взаимосвязей, позволяя формулировать более сложные условия.

Такие ограничения характеризуются высокой степенью абстракции, представляя сложные отношения между группами переменных в компактной форме. Они могут быть заданы различными способами. Примером может служить ограничение *alldifferent*(x_1, \dots, x_n), которое определяет, что значения, присваиваемые переменным x_1, \dots, x_n , должны быть попарно различными.

Многие глобальные ограничения могут быть связаны с внутренними структурами, которые представляют возможные значения и правила для переменных (например, конечные автоматы в случае регулярных ограничений).

Ключевыми характеристиками глобальных ограничений являются:

возможность моделировать сложные взаимоотношения, которые не могут быть адекватно выражены с помощью простых ограничений;

оптимизированные решатели могут использовать глобальные ограничения для сокращения пространства поиска и устранения неэффективных подмножеств.

Обычные ограничения, как правило, ограничиваются парами переменных или незначительными группами, и их выражение может потребовать большого количества отдельных условий для описания более сложных взаимосвязей. Например, для обеспечения уникальности значений в группе из 10 переменных с помощью отдельных ограничений потребуется 45 простых ограничений (поскольку необходимо убедиться в уникальности каждой пары), в то время как с помощью глобального ограничения *alldifferent* эта задача может быть выражена одной строкой.

Представление VRP в виде задачи целочисленного линейного программирования (ILP)

Рассмотрим формулировку ILP из [12]. Пусть задано множество клиентов $C = \{1, \dots, n\}$ и транспортных средств $M = \{1, \dots, m\}$. Для индексации склада в начале маршрута используется 0, а для индексации склада в конце маршрута — $n+1$. $N = C \cup \{0, n+1\}$.

В рассматриваемой постановке используются следующие обозначения:

a) в качестве переменных выступают x_{ijk} ($i, j \in N, k \in M$), причем x_{ijk} равна 1, если автомобиль k едет напрямую от клиента i к клиенту j , 0 — в противном случае;

b) c_{ij} — стоимость путешествия из i в j , $i, j \in N$;

c) τ_{ij} — время в пути от i до j , $i, j \in N$, включающее в себя время обслуживания у клиента i ;

d) δ_{ij} — расстояние от i до j , $i, j \in N$;

e) r_i — спрос для клиента $i \in N$;

f) Q_k — вместимость транспортного средства $k \in M$;

g) a_i, b_i — временные окна клиентов $i \in N$;

h) K — целое число.

Главная цель задачи VRP — минимизировать сумму затрат (1):

$$\text{minimise } z_{VRP} = \sum_{k \in M} \sum_{i \in N} \sum_{j \in N} c_{ij} x_{ijk}. \quad (1)$$

Определим ограничения для задачи VRP.

Каждый клиент должен быть посещен ровно один раз (2):

$$\sum_{k \in M} \sum_{j \in N} x_{ijk} = 1 \quad \forall j \in C. \quad (2)$$

Пропускная способность (в предположении неоднородного парка транспорта):

$$\sum_{j \in C} r_i \sum_{j \in N} x_{ijk} \leq Q_k \quad \forall k \in M. \quad (3)$$

Ограничение (4–6) для обеспечения потока транспортных средств от начального склада, через 0 или более клиентов к конечному складу:

$$\sum_{j \in N} x_{0jk} = 1 \quad \forall k \in M, \quad (4)$$

$$\sum_{i \in N} x_{ihk} - \sum_{j \in N} x_{hjk} = 0 \quad \forall h \in C, \forall k \in M, \quad (5)$$

$$\sum_{i \in N} x_{i(n+1)k} = 1 \quad \forall k \in M. \quad (6)$$

Обеспечение отсутствия подтуров (циклов, не включающих депо) описывается ограничениями (7–8):

$$\sum_{i,j \in S} x_{ijk} \leq |S| - 1 \quad \forall S \subseteq C, \quad (7)$$

$$x_{ijk} \in \{0,1\} \quad \forall i, j \in N, \forall k \in M. \quad (8)$$

Ограничение для определения времени прибытия к каждому клиенту (9):

$$t_{ik} + \tau_{ij} - K(1 - x_{ijk}) \leq t_{jk} \quad \forall i, j \in N, \forall k \in M. \quad (9)$$

Обеспечение соблюдения временных окон происходит при помощи ограничения (10):

$$a_i \leq t_{ik} \leq b_i \quad \forall i \in N, \forall k \in M. \quad (10)$$

Формулирование задач маршрутизации в форме задач удовлетворения ограничений

В данной формулировке [13], аналогично ситуации с ILP, рассматривается n заказов от клиентов и парк из m транспортных средств. Термин «визит» используется для обозначения момента, когда транспортное средство осуществляет остановку. Каждому клиенту соответствует одно посещение, в то время как на каждое транспортное средство приходится два специальных визита для моделирования начальной и конечной точек маршрута транспортного средства.

Пусть $C = \{1, \dots, n\}$ — множество клиентов, $M = \{1, \dots, m\}$ — множество транспортных средств и $V = \{1, \dots, n + 2m\}$ — множество визитов. Первый и последний визит транспортного средства k обозначается f_k и l_k соответственно. Визит $f_k = n + k$ соответствует первому визиту транспортного средства k , а $l_k = n + m + k$ — последнему. $F = \{n + 1, \dots, n + m\}$ и $L = \{n + m + 1, \dots, n + 2m\}$ — множества первых и последних визитов соответственно. $C = \{1, \dots, n\}$ — клиенты. Также введем обозначения:

- $q_i \geq 0$ — количество товаров в транспортном средстве после выполнения визита i ;
- Q_k — грузоподъемность транспортного средства k ;
- v_i — транспортное средство, выполняющее визит i ;
- $r_i \neq 0$ — количество товаров, которое нужно забрать во время визита i ;
- s_i — прямой приемник визита i ;
- $t_i \geq 0$ — время начала обслуживания визита i ;
- p_i — целочисленная переменная, моделирует прямого предшественника каждого визита $i \in V$.

Каждый «первый» визит транспортного средства имеет в качестве предшественника «последний» визит транспортного средства ($\forall k \in M p_{jk} = l_k$). Таким образом, переменные-предшественники образуют перестановку V и подчиняются глобальным ограничениям *alldiff constraint* (11):

$$p_i \neq p_j \forall i, j \in V \wedge i < j \quad . \quad (11)$$

Переменные-преемники поддерживаются «когерентными» (согласованными) с переменными-предшественниками с помощью глобальных ограничений *element constraint* (12):

$$s_{p_i} = i \forall i \in V - F \quad p_{s_i} = i \forall i \in V - L. \quad (12)$$

Для первого и последнего визитов накладываются ограничения: $\forall k \in M v_{fk} = v_{lk} = k$.

На маршруте все визиты выполняются одним и тем же транспортным средством (13):

$$v_i = v_{p_i} \forall i \in V - F \quad v_i = v_{s_i} \forall i \in V - L. \quad (13)$$

Ограничения пути (*path constraint*) (14) гарантирует, что количество товаров на борту транспортного средства остается в допустимых пределах:

$$q_i = q_{p_i} + r_i \forall i \in V - F \quad q_i = q_{s_i} - r_i \forall i \in V - L. \quad (14)$$

Ограничение вместимости транспортного средства (15):

$$q_i \leq Q_{v_i} \forall i \in V. \quad (15)$$

Для задачи VRPTW добавляется ограничение времени движения транспортных средств по маршрутам (16):

$$t_i \geq t_{p_i} + \tau_{p_i, i} \quad i \in V - F \quad t_i \leq t_{s_i} - \tau_{i, s_i} \quad i \in V - L. \quad (16)$$

Временные окна для клиентов задаются путем добавления ограничений на переменные t . Ограничение $a \leq t_i \leq b$ означает, что клиент i должен быть посещен в промежуток времени между a и b . Для моделирования нескольких временных окон используется ограничение (17) [4]:

$$a \leq t_i \leq d \wedge b_i \leq b \vee t_i \geq c. \quad (17)$$

Ограничения (17) означают, что клиент i должен быть посещен либо в промежуток времени между a и b , либо в промежуток времени между c и d ($a \leq b \leq c \leq d$).

Различные транспортные средства могут иметь различные доступные временные окна. Введем следующие обозначения:

- а) O_k — самое раннее время маршрута начала движения транспортного средства k ;
- б) H_k — его самое позднее время окончания движения, или горизонт;
- с) δ_{ij} — расстояние от визита i до j ;
- д) C_k — это стоимость единицы расстояния для транспортного средства k .

Ограничения на время доступности транспортных средств (18):

$$O_{v_i} \leq t_i \leq H_{v_i} \quad \forall i \in V. \quad (18)$$

Функция затрат обычно представляет собой общее пройденное расстояние d (19):

$$d = \sum_{i \in V - F} \delta_{p_i, i} \quad d = \sum_{i \in V - L} \delta_{i, s_i}. \quad (19)$$

Если стоимость километра зависит от используемого транспортного средства, функция затрат определяется как (20):

$$d = \sum_{i \in V - F} C_{v_i} \delta_{p_i, i} \quad d = \sum_{i \in V - L} C_{v_i} \delta_{i, s_i}. \quad (20)$$

Для задачи VRPPD:

- а) $r_i \leq 0$ — для доставки товара;

- b) $r_i \geq 0$ — для вывоза товара;
- c) o — заказ;
- d) o_p — визит для доставки товара;
- e) o_d — визит для вывоза товара.

Вывоз и доставка должны осуществляться одним и тем же транспортным средством (21):

$$v_{o_p} = v_{o_d} \quad t_{o_p} < t_{o_d}. \quad (21)$$

Прямая связь от пункта самовывоза до пункта доставки не допускается, вывоз производится после доставки (22):

$$s_i \neq j \quad \forall i \in P \quad \forall j \in D \quad \text{где } P = \{i | r_i > 0\} \text{ и } D = \{i | r_i < 0\}. \quad (22)$$

Заключение

В данной статье рассмотрена одна из ключевых задач в области логистики и эффективного управления ресурсами — задача маршрутизации транспортных средств (VRP). Приводится аналитический обзор возможностей решений задач VRP как задач удовлетворения ограничений.

Методы CP предлагают мощный инструментарий для масштабируемого и гибкого моделирования VRP, обеспечивая возможность эффективно управлять множеством ограничений, характерных для данной задачи. Гибкость в формулировке ограничений, возможность интеграции с другими методами оптимизации, а также способность учитывать неопределенности делают CP уникальным решением для повышения качества маршрутизации.

Практические примеры применения CP демонстрируют его эффективность в решении различных постановок задач VRP. Эти примеры подтверждают, что применение CP позволяет значительно улучшать производительность логистических процессов и снижать операционные затраты.

Список источников

1. Laporte G. Fifty Years of Vehicle Routing // *Transportation Science*. 2009. Vol. 43, № 4. P. 408–416.
2. Çam Ö. N., Sezen H. K. The formulation of a linear programming model for the vehicle routing problem in order to minimize idle time // *Decision Making: Applications in Management and Engineering*. 2020. Vol. 3, № 1. P. 22–29.
3. Gendreau M., Potvin J. Metaheuristics in Combinatorial Optimization // *Annals of Operations Research*. November 2005. Vol. 140, № 1. P. 189–213.
4. Pesant G., Gendreau M., Potvin J., Rousseau J. An exact constraint logic programming algorithm for the travelling salesman with time windows // *Transportation Science*. 1998. Vol. 32, № 1. P. 22–29.
5. Toth P., Vigo D. An Overview of Vehicle Routing Problems // *The Vehicle Routing Problem*. 2002. Vol. 9. P. 1–26.
6. Paraskevopoulos D. C., Laporte G., Repoussis P. P., Tarantilis C. D. Resource constrained routing and scheduling: Review and research prospects // *European Journal of Operational Research*. 2017. Vol. 263, № 3. P. 737–754.
7. Braysy O., Gendreau M. Vehicle routing problem with time windows, part I: Route construction and local search algorithms // *Transportation Science*. 2005. Vol. 39, № 1. P. 104–118.
8. Lu Q., Dessouky. M. An exact algorithm for the multiple vehicle pickup and delivery problem // *Transportation Science*. 2004. Vol. 38, № 4. P. 503.
9. Desrosiers J., Dumas Y., Solomon M. M., Soumis F. Time constrained routing and scheduling // *Network Routing*. Vol. 8. 1995. P. 35–139.
10. Montanari U. Networks of constraints: fundamental properties and applications to picture processing // *Information Sciences*. 1974. Vol. 7. P. 95–132.
11. Regim J. C. Global Constraints: A Survey // *Hybrid Optimization*. Springer Optimization and Its Applications. 2011. Vol. 45. P. 13–32.
12. Kallehauge B., Larsen J., Madsen O. B. Lagrangean duality applied on vehicle routing with time windows // *Computers and Operations Research*. 2001. Vol 33. P. 1464–1487.
13. De Backer B., Furnon V., Prosser P., Kilby P., Shaw P. Solving vehicle routing problems using constraint programming and metaheuristics // *Journal of Heuristics*. 2000. Vol. 6. №4. P. 501–523.

References

1. Laporte G. *Fifty Years of Vehicle Routing*. Transportation Science, 2009, vol. 43, no. 4, pp. 408–416.
2. Çam Ö. N., Sezen H. K. *The formulation of a linear programming model for the vehicle routing problem in order to minimize idle time*. Decision Making: Applications in Management and Engineering, 2020, vol. 3, no. 1, pp. 22–29.
3. Gendreau M., Potvin J. *Metaheuristics in Combinatorial Optimization*. Annals of Operations Research, 2005, vol. 140, no. 1, pp. 189–213.
4. Pesant G., Gendreau M., Potvin J., Rousseau J. *An exact constraint logic programming algorithm for the travelling salesman with time windows*. Transportation Science, 1998, vol. 32, no. 1, pp. 22–29.
5. Toth P., Vigo D. *An Overview of Vehicle Routing Problems*. The Vehicle Routing Problem, 2002, vol. 9, pp. 1–26.
6. Paraskevopoulos D.C., Laporte G., Repoussis P.P., Tarantilis C.D. *Resource constrained routing and scheduling: Review and research prospects*. European Journal of Operational Research, 2017, vol. 263, no. 3, pp. 737–754.
7. Braysy O., Gendreau M. *Vehicle routing problem with time windows, part I: Route construction and local search algorithms*. 2005. Transportation Science, vol. 39, no. 1, pp. 104–118.
8. Lu Q., Dessouky M. *An exact algorithm for the multiple vehicle pickup and delivery problem*. Transportation Science, 2004, vol. 38, no. 4, pp. 503–514.
9. Desrosiers J., Dumas Y., Solomon M. M., Soumis F. *Time constrained routing and scheduling*, 1995, Network Routing, vol. 8, pp. 35–139.
10. Montanari U. *Networks of constraints: fundamental properties and applications to picture processing*. Information Sciences, 1974, vol. 7, pp. 95–132.
11. Regin J. C. *Global constraints: A survey*. Hybrid Optimization. Springer Optimization and Its Applications, 2011, vol. 45, pp. 13–32.
12. Kallehauge B., Larsen J., Madsen O. B. *Lagrangian duality applied on vehicle routing with time windows*. Computers and Operations Research, 2001, vol. 33, pp. 1464–1487.
13. De Backer B., Furnon V., Prosser P., Kilby P., Shaw P. *Solving vehicle routing problems using constraint programming and metaheuristics*. Journal of Heuristics, 2000, vol. 6, no. 4, pp. 501–523.

Информация об авторах

А. В. Шестаков — стажер-исследователь;

А. А. Зуенко — кандидат технических наук, ведущий научный сотрудник.

Information about the authors

A. V. Shestakov — Intern Researcher at the IIMM of the KSC RAS

A. A. Zuenko — Candidate of Science (Tech.), Leading Researcher of the Putilov Institute for Informatics and Mathematical Modeling Kola Science Centre of the Russian Academy of Sciences.

Статья поступила в редакцию 17.10.2024; одобрена после рецензирования 01.11.2024; принята к публикации 08.11.2024.
The article was submitted 17.10.2024; approved after reviewing 01.11.2024; accepted for publication 08.11.2024.

Научная статья
УДК 004.94, 510.644.4
doi:10.37614/2949.1215.2024.15.3.006

ПРАКТИКА ПРИМЕНЕНИЯ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ СЛОЖНЫХ СИСТЕМ

Елена Олеговна Неупокоева^{1✉}, Светлана Николаевна Малыгина², Виталий Викторович Быстров³

^{1–3}Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия

²Филиал МАУ в г. Апатиты, Апатиты, Россия

¹e.neupokoeva@ksc.ru, <https://orcid.org/0000-0003-4059-8724>

²s.malygina@ksc.ru, <https://orcid.org/0000-0001-6010-5662>

³v.bystrov@ksc.ru, <https://orcid.org/0000-0002-9369-8457>

Аннотация

Имитационное моделирование является одним из популярных методов изучения сложных систем разной природы. В работе анализируется существующий опыт применения аппарата нечеткой логики для имитационного моделирования процессов и систем. Рассматриваются основные направления интеграции нечеткой логики и имитационного моделирования. Приводятся примеры компьютерных моделей из разных предметных областей, построенных на базе нечеткой логики, в разрезе методов имитационного моделирования. Отмечаются способы технической реализации интеграции двух рассматриваемых инструментов. Авторы делают вывод о возможности применения нечеткой логики для моделирования жизнеспособности региональных социально-экономических систем.

Ключевые слова:

имитационное моделирование, нечеткая логика, системная динамика, нечеткие когнитивные карты, агентное моделирование

Благодарности:

исследование выполнено в рамках государственного задания Института информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук от Министерства науки и высшего образования Российской Федерации, тема научно-исследовательской работы — «Методология создания информационно-аналитических систем поддержки управления региональным развитием, основанных на формирующем искусственном интеллекте и больших данных» (регистрационный номер 122022800551-0).

Для цитирования:

Неупокоева Е. О., Малыгина С. Н., Быстров В. В. Практика применения нечеткой логики для имитационного моделирования сложных систем // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 69–81. doi:10.37614/2949.1215.2024.15.3.006.

Original article

THE PRACTICE OF USING FUZZY LOGIC FOR SIMULATION OF COMPLEX SYSTEMS

Elena O. Neupokoeva^{1✉}, Svetlana N. Malygina², Vitaliy V. Bystrov³

^{1–3}Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia

²Apatity branch of Murmansk Arctic State University, Apatity, Russia

¹e.neupokoeva@ksc.ru, <https://orcid.org/0000-0003-4059-8724>

²s.malygina@ksc.ru, <https://orcid.org/0000-0002-0924-0188>

³v.bystrov@ksc.ru, <https://orcid.org/0000-0002-9369-8457>

Abstract

Simulation modeling is one of the popular methods for studying complex systems of different nature. The paper analyzes the existing experience of using the fuzzy logic apparatus for simulation modeling of processes and systems. The main directions of integration of fuzzy logic and simulation modeling are considered. Examples of computer models from different subject areas, built on the basis of fuzzy logic, in the context of simulation modeling methods are given. The ways of technical implementation of the integration of the two considered means are noted. The authors conclude that fuzzy logic can be used to model the resilience of regional socio-economic systems.

Keywords:

simulation modeling, fuzzy logic, system dynamics, fuzzy cognitive maps, agent-based modeling

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science

and Higher Education of the Russian Federation, research topic “Methodology for creating information and analytical systems to support the management of regional development based on formative artificial intelligence and big data” (registration number of the research topic 122022800551-0).

For citation:

Neupokoeva E.O., Malygina S.N., Bystrov V.V. The practice of using fuzzy logic for simulation of complex systems // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 69–81. doi:10.37614/2949.1215.2024.15.3.006.

Введение

В наше время многие ученые отмечают, что долгосрочное планирование становится недостаточно гибким для быстро меняющихся параметров реальности [1]. На передний план среди методов прогнозирования выдвигается имитационное моделирование, обладающее множеством достоинств, например, возможность:

- исследовать слабоструктурированные системы, о функционировании которых доступен существенно ограниченный объем информации;
- изучить свойства систем, которые невозможно анализировать в реальных условиях или данные действия потребует большого объема ресурсов;
- детализировать модель, выделив различные агенты для реально существующих объектов и придав каждому уникальные свойства;
- задать различное поведение для отдельных агентов;
- отразить взаимосвязи между агентами;
- использовать случайно возникающие события и т. д.

В Институте информатики и математического моделирования им. В.А. Путилова ФИЦ КНЦ РАН ведется исследование по разработке методических и инструментальных средств информационно-аналитической поддержки управления жизнеспособностью региональных социально-экономических систем. Блок задач этого исследования связан с вопросами имитационного моделирования данного класса сложных систем. В частности, разрабатывается комплекс компьютерных моделей для оценки жизнеспособности (resilience) региональных «Бизнес-Сообщество-Власть»-систем (БСВ-систем). Предложенный метод оценки БСВ-систем базируется на вычислительных процедурах над многомерными структурами данных — гиперкубом жизнеспособности [2]. Эта структура формируется на основе сбора и обработки статистических данных, результатов анкетирования, экспертных мнений и др. Элементами гиперкуба жизнеспособности могут являться как количественные, так и качественные показатели. Одним из формальных аппаратов оперирования качественными показателями, зарекомендовавшим себя в разных предметных областях, является теория нечетких множеств и нечеткая логика. Чтобы осуществить поставленные задачи текущего исследования, возникла идея интегрировать имитационное моделирование и нечеткую логику для реализации возможности работать с качественными оценками. Для этого сперва необходимо проанализировать имеющейся практический опыт других исследовательских команд в области совместного использования двух этих инструментов для построения адекватных моделей сложных систем.

Имитационное моделирование позволяет подстраиваться под быстро меняющиеся исходные данные, обновляющиеся в результате социальных, политических и экономических изменений. Для расширения возможностей традиционных имитационных моделей с целью исследования поведения сложных объектов используется нечеткая логика (fuzzy logic).

Классическая четкая логика дает однозначный ответ «да» или «нет». Нечеткая логика оперирует более расширенным диапазоном понятий — «с какой вероятностью этот ответ правильный». В том числе это позволяет более естественно, чем формально-логические системы, описывать ход человеческого мышления. Существуют различные инструменты, основанные на нечеткой логике [3]:

- нечеткие нейронные сети;
- адаптивные нечеткие системы;
- нечеткие запросы к базам данных;
- нечеткие ассоциативные правила;
- нечеткие когнитивные карты;
- нечеткая кластеризация.

Нечеткая логика находит свое применение в разных видах имитационного моделирования. В случае с агентным моделированием она используется для задания поведения агентов, например, для имитации поведения человека или отдельных его аспектов, а в системной динамике — для вычисления нечетких параметров, например, плавающих показателей торговли и промышленности. Сочетание имитационного моделирования и нечеткой логики используется при моделировании сложных систем из разных предметных областей: складской системы [4], в создании модели управляющих действий летчика [5], управлении установками уличного освещения [6], выбора метода проращивания риса [7], анализе неисправностей в энергосистеме [8] и т. п. Также подобная связка инструментов часто применяется для моделирования социальных, экономических и социально-экономических процессов.

Для того чтобы выявить, как именно применяется нечеткая логика при разработке компьютерных моделей, ниже будут рассмотрены некоторые работы в разрезе применяемых методов имитационного моделирования.

Реализация нечеткой логики на базе агентных моделей

Довольно большой кластер имитационного моделирования с использованием нечеткой логики занимают агентные модели. Тематика этих моделей довольно обширна. Например, в работе [9] описывается инструмент моделирования, который позволяет имитировать социальные процессы в многоагентной системе, где агенты принимают решения в соответствии с правилами нечеткой логики. Инструмент предлагает тороидальный, непрерывный мир, в котором могут существовать два вида растений (пряность и сахар) и произвольное количество агентов разных видов. Агенты имеют потребности и могут быть наделены различными правилами и стратегиями, среди которых они выбирают действия для удовлетворения своих потребностей. В работе [10] описывается применение нечеткой логики для моделирования динамики дружбы в агентной модели. Моделирование было сконфигурировано с использованием популяции из 3 тыс. статических агентов, случайным образом распределенных в пространстве. Имитация осуществлялась в течение 20 лет (1 тыс. шагов агента). Агенты способны взаимодействовать, устанавливая дружеские и супружеские отношения, а также размножаться. Агент сможет общаться примерно с 50 другими «возможными друзьями» на протяжении всей своей жизни. Из них каждый агент сможет выбрать, с кем из них он будет общаться, будет дружить. Этот выбор задается правилами нечеткой логики.

Модель виртуального поселка

В работе [11] рассмотрено создание модели виртуального поселка, наполненного игровыми персонажами. Каждый игровой персонаж при этом оснащен системой принятия решений на основе нечеткой логики. Авторы утверждают, что при рассмотрении системы принятия решений с использованием нечеткой логики для одного персонажа результаты предсказать довольно просто ввиду небольшого количества конечных состояний, но при увеличении количества систем задача становится нетривиальной. Целью создания модели является классификация и описание активностей персонажей, руководствующихся нечеткой логикой при принятии решений. При этом персонажи взаимодействуют в рамках модели и оказывают влияние на свойства друг друга. Набор активных действий и свойств персонажа представлен на рис. 1.

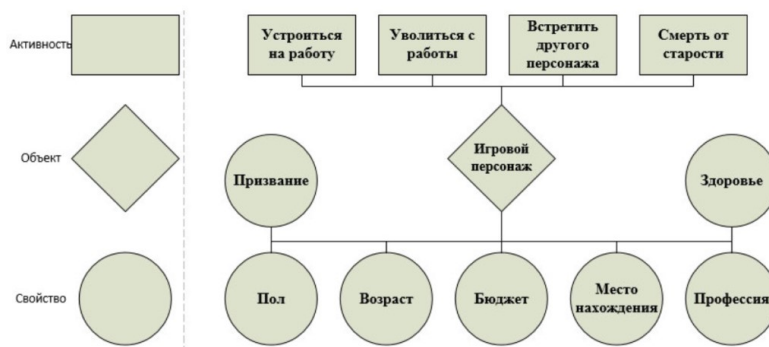


Рис. 1. Активности и свойства игрового персонажа [11]

Всего в модели используется 70 уникальных агентов. Нечеткой логикой персонажи руководствуются в случаях принятия решений, связанных со здоровьем, бюджетом и доверием к другим персонажам. Например, в случае, когда игровой персонаж увольняется с работы, используются три лингвистические переменные, отображающие следующие свойства:

- возраст персонажа (четырепараметрическая криволинейная трапеция);
- бюджет (четырепараметрические и двухпараметрические функции);
- расстояние (трапециевидные и пирамидальные функции). Математическое представление переменной «расстояние» [11] показано на рис. 2:

$$\mu(x) = \begin{cases} 0, & -\infty \leq x \leq -b; \\ 0,5 + 0,5 * \sin \left\{ \pi \left[x + \frac{a+b}{2} \right] / (b-a) \right\}, & -b \leq x \leq -a \\ 1, & -a \leq x \leq a \\ 0,5 - 0,5 * \sin \left\{ \pi \left[x - \frac{a+b}{2} \right] / (b-a) \right\}, & a \leq x \leq b \\ 0, & b \leq x \leq \infty \end{cases}$$

Рис. 2. Математическое представление лингвистической переменной «расстояние» [11]

Модель трудовой миграции населения

В модели [12] рассматриваются миграционные процессы населения. В качестве параметров используются социально-экономические показатели региона. Регионом исследования является Сибирский федеральный округ в составе 10 субъектов. Источник данных — ресурсы Федеральной службы государственной статистики за 2010–2021 гг. Язык разработки — C#, для реализации механизма нечеткой логики используется библиотека Accord.Fuzzy. Агентами модели выступают объекты типа «Человек», обладающие набором различных параметров, например, пол, возраст, фактическая заработная плата (рассчитывается из прожиточного минимума и уровня образования агента) и т. п. Продукционные правила нечеткой логики сформированы с помощью механизма «Если-То» и устанавливают зависимость между желаемым уровнем жизни агента, уровнем образования и неудовлетворенностью заработной платой. Пользовательский интерфейс включает возможность манипулировать параметрами трудового возраста населения, а также просматривать различные показатели результатов моделирования с разделением по территориальным субъектам. Графический интерфейс представлен на рис. 3.

Погрешность динамики численности населения в сравнении с реальными показателями не превысила 10 %, следовательно, модель адекватна. Предполагается использовать ее для анализа демографической обстановки и контроля миграционной динамики.

Модель взаимовлияния миграционной активности населения и экономического развития региона

Авторы работы [13] утверждают, что в связи с событиями последних лет уровень миграции за границу сильно возрос, особенно заметен отток высококвалифицированных специалистов. Для выявления взаимовлияния миграции населения и социально-экономических факторов была создана имитационная агентно-ориентированная модель, которая была протестирована на данных Сибирского федерального округа.

В разработанной модели используется два вида агентов: «Человек» и «Регион». «Регион» объединяет в себе 47 субъектов Российской Федерации, которые внесли наибольший вклад в миграцию из Республики Башкортостан. В качестве исходных данных для заполнения атрибутов агента «Регион» используются данные Федеральной службы государственной статистики.

Количество агентов типа «Человек» высчитывается в масштабе 1:1000 от населения Республики Башкортостан. Агент обладает различными параметрами (пол, возраст, заработная плата и т. п.), которые заполняются на основе статистических данных, и подчиняется общим правилам (рождение,

миграция, смерть и т. п.) Также для каждого агента производится расчет предпочтительного уровня жизни с помощью нечеткой логики, для этого в модели сформированы 34 правила вида «Если-То». Расчеты реализованы в расширении Fuzzy Logic Toolbox в среде Matlab.



Рис. 3. Графический интерфейс модели динамики трудовой миграции населения [12]

Реализация нечеткой логики на базе системно-динамических моделей

Кроме использования в агентных моделях, нечеткая логика может выступать частью системно-динамических моделей (СД-моделей), подменяя один или несколько параметров лингвистическими переменными. Нечеткая логика помогает имитировать уровень удовлетворенности товаром [14], изменение урожайности риса от температуры [7], функционирование строительных систем [15] и т. п.

Модель продаж и обслуживания

В работе [14] предложен метод включения нескольких лингвистических или нечетких переменных в структуру модели системной динамики. Предложенная методика реализована в среде моделирования VENSIM на примере модели продаж и обслуживания (рис. 4). В модели две лингвистические переменные (удовлетворенность клиента в отношении обслуживания и время выполнения заказа, связанное с продуктом) влияют на преобразование потенциальных клиентов в клиентов.

Для определения лингвистических переменных используются треугольные функции принадлежности. Для дефаззификации используется понятие наибольшего из максимума, чтобы перевести нечеткое представление объединенного эффекта в четкое значение.

Моделирование производительности землеройных работ

В работе [15] предложена методика реализации применения нечетких данных для моделирования строительных систем на основе моделей системной динамики. Авторами данного исследования представлен гибридный метод, заключающийся в интеграции нечеткой системной динамики (fuzzy system dynamics, FSD) с гибридными нейро-нечеткими системами, названный нейро-нечеткая системная динамика (neuro-fuzzy system dynamics, N-FSD). Для определения нелинейных, сложных и многомерных отношений между системными переменными в N-FSD используются гибридные нейро-нечеткие системы, что, по словам авторов метода, повышает точность моделей FSD в строительных приложениях. Для определения нечетких переменных используются гауссовские функции принадлежности. Применимость метода N-FSD проверяется путем моделирования производительности землеройных работ (рис. 5).

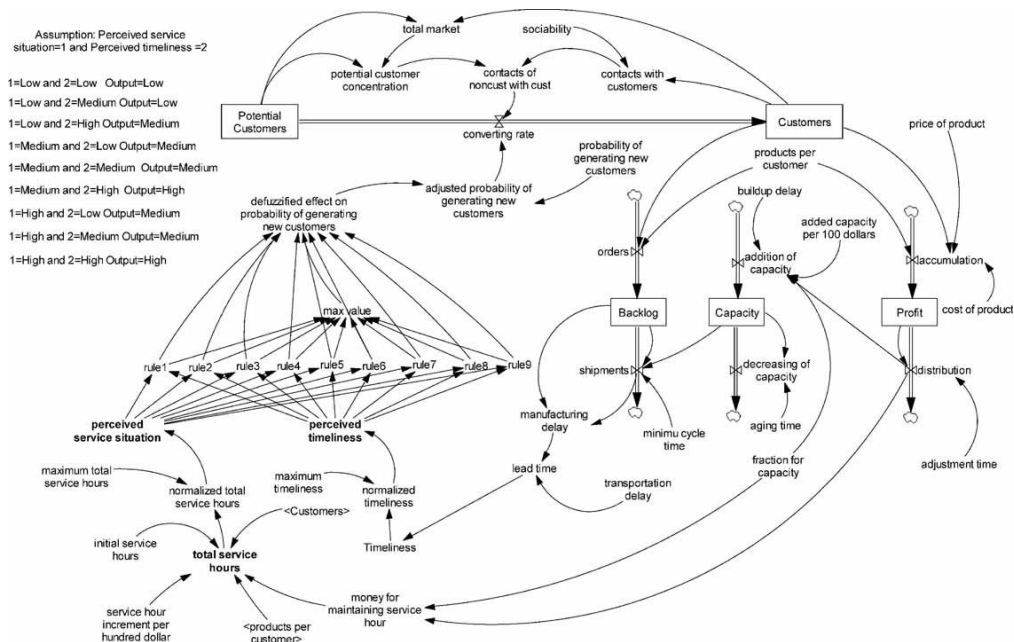


Рис. 4. СД-модель, демонстрирующая комбинированный эффект двух лингвистических переменных [14]

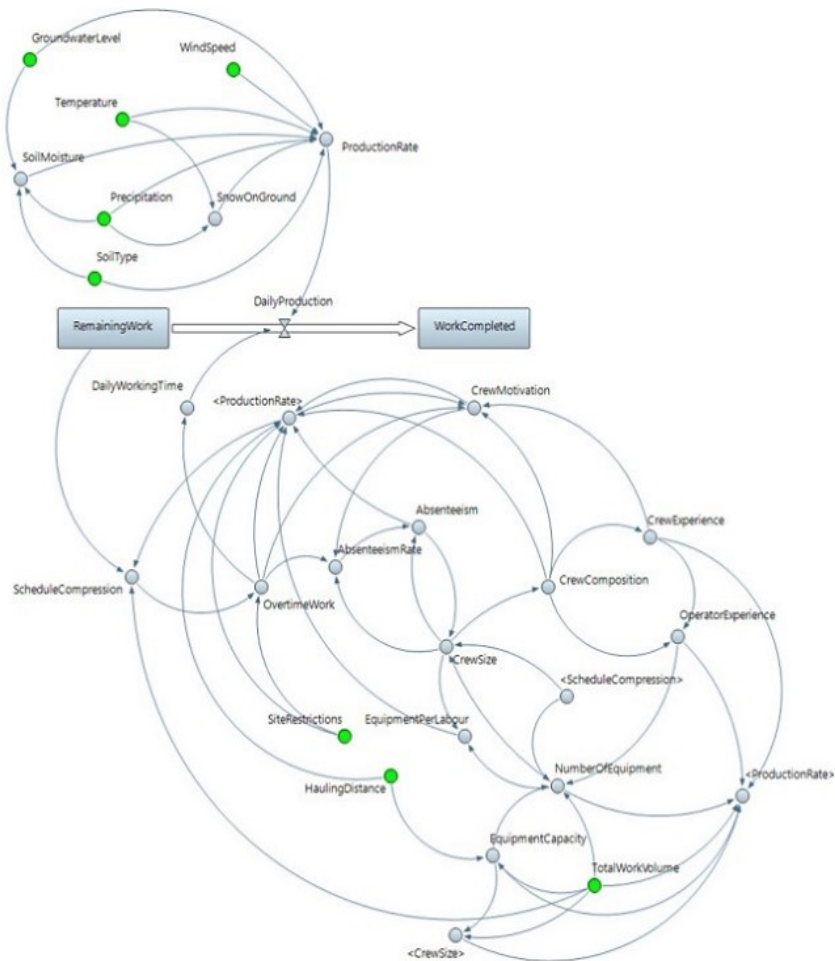


Рис. 5. Качественная модель N-FSD производительности земляных работ [15]

Модель N-FSD землеройных работ, представленная в этой работе, была разработана в AnyLogic с онлайн-подключением к MATLAB. В этой архитектуре модели на каждом шаге времени моделирования AnyLogic отправляет значение входных переменных для каждой системной связи в MATLAB, а MATLAB определяет выходные данные системных связей и возвращает результаты в AnyLogic.

Модель производства риса

В работе [7] предложен метод для реализации СД-модели, основанной на нечеткой логике. Этот метод был протестирован с помощью нечеткого СД-моделирования для системы производства риса, в которой нечеткая логика была применена для учета изменений урожайности риса в зависимости от температуры. На рис. 6 представлена простая модель динамики нечеткой системы для системы производства риса. Вспомогательная переменная с именем 'Fuzzy_Rice_yield' является переменной, выведенной нечеткой логикой.

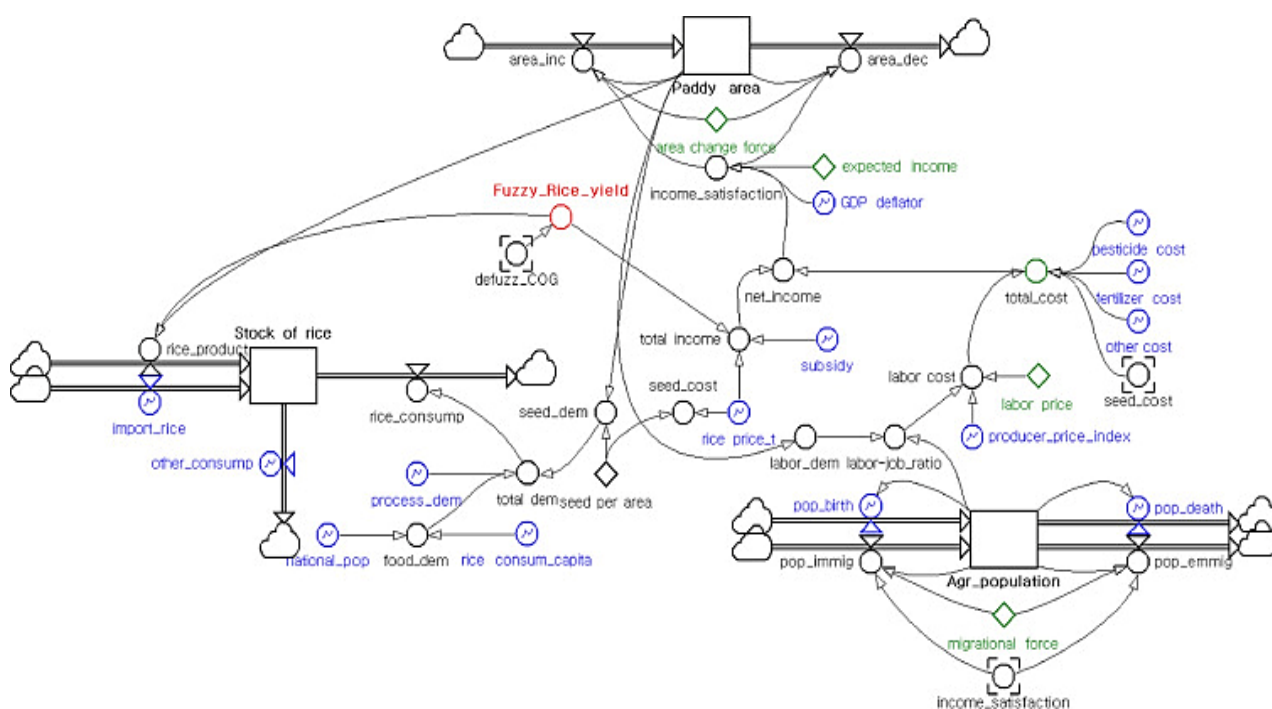


Рис. 6. Модель динамики нечеткой системы для системы производства риса [7]

Урожайность риса, рассчитанная с помощью нечеткой логики в рамках модели, соответствовала историческим данным. Разработанная методология включает процедуры построения нечеткой логики: фаззификацию, нечеткий вывод и оптимизацию на платформе SD. На рис. 7 представлена нечеткая логика для связи между средней температурой и урожайностью риса, реализованная в системно-динамической среде моделирования.

Использование когнитивных карт

Одним из методов нечеткой логики являются нечеткие когнитивные карты. Они используются для моделирования причинно-следственных взаимосвязей и позволяют наглядно представить систему, подвергаемую анализу. Нечеткие когнитивные карты представляют собой ориентированный граф, состоящий из узлов в виде нечетких множеств. Связи между узлами показывают, насколько сильно те или иные показатели влияют друг на друга.

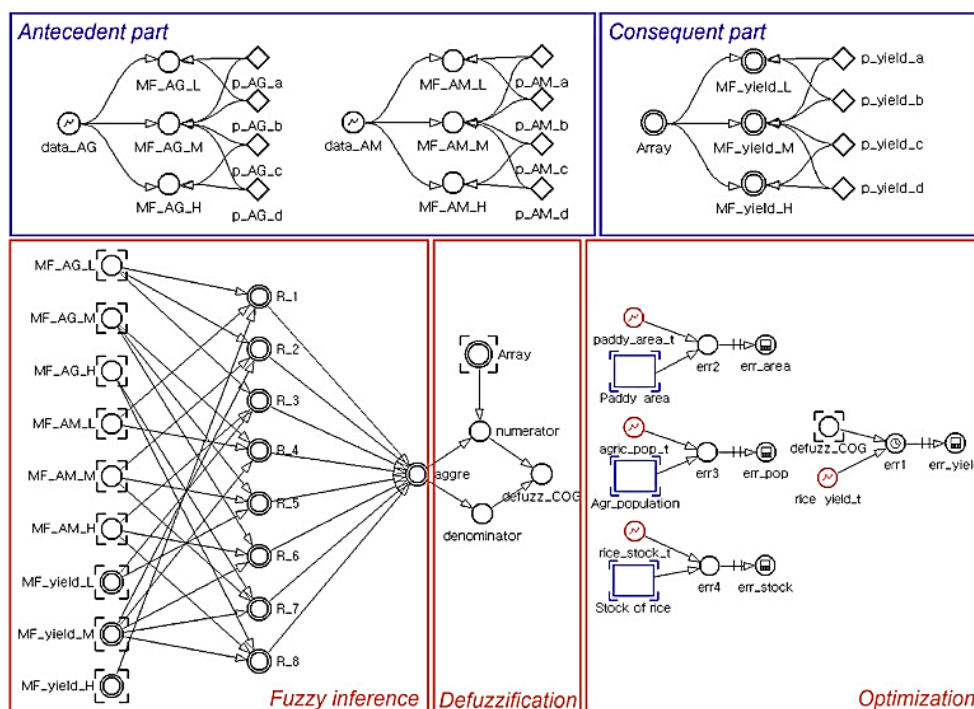


Рис. 7. Модель нечеткой логики [7]

Модель распространения болезней

В работе [16] описывается разработка модели пространственно-временного распространения болезней на примере городского округа Бангалор, Индия. Была реализована агентная модель (Agent-Based Modeling, ABM) с применением нечетких когнитивных карт (Fuzzy Cognitive Map, FCM) (рис. 8) для моделирования индивидуальной поведенческой логики агентов.

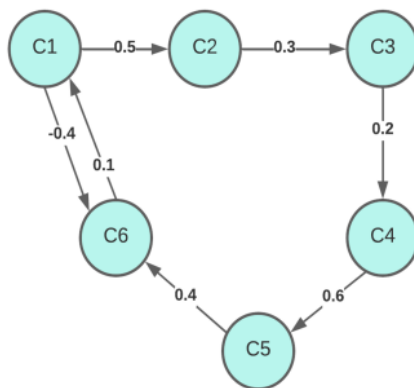


Рис. 8. Пример когнитивной карты [16]

С помощью когнитивных карт моделируется процесс «Восприимчивый — подвергшийся воздействию — заразившийся — удаленный» (SEIR) независимого индивидуального поведения агента, которое влияет на общее изменение тенденции. Когнитивная карта представляет собой граф, вершинами которого являются понятия (например, осознание болезни, страх, профилактические меры), а ребрами обозначают причинно-следственные связи между этими понятиями. Каждому ребру присваивается вес, указывающий силу и направление влияния.

В работе представлена высокоуровневая архитектура модели, имитирующая процесс SEIR (рис. 9).

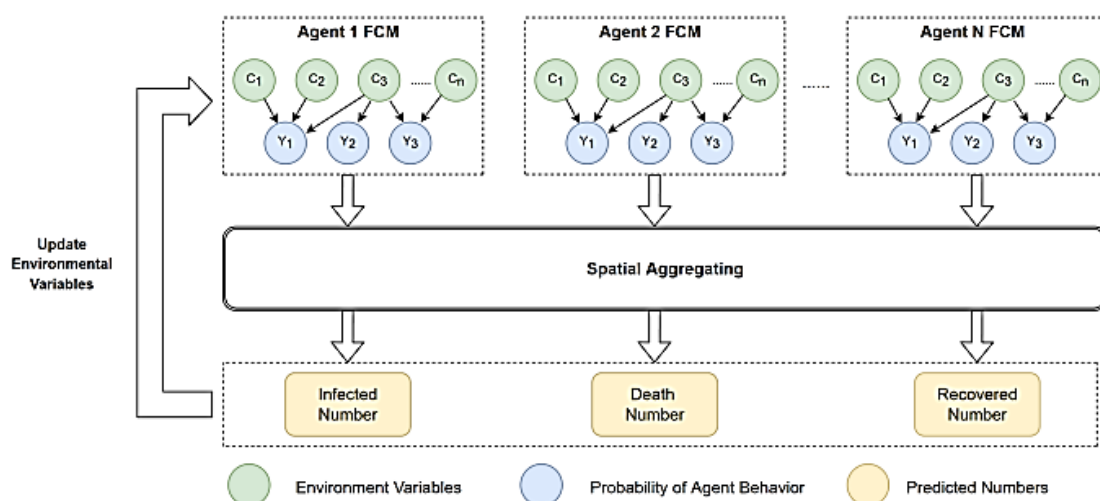


Рис. 9. Высокоуровневая архитектура гибридной модели FCM-ABM, имитирующая процесс SEIR [16]

Прогнозирование развития экономической ситуации

В основе модели, предложенной Заграновской и Эйсером [1], лежат нечеткие когнитивные карты. Авторы утверждают, что в условиях быстро меняющейся ситуации традиционная разработка долгосрочной стратегии не приносит желаемых результатов, поэтому необходимо моделировать сценарии развития экономической ситуации предприятия, и убедительно доказывают, что наиболее предпочтительным методом их создания выступает имитационное моделирование. При этом нечеткая логика хорошо подходит для описания экспертных оценок, поскольку эксперт не всегда абсолютно уверен в определенном результате. Нечеткая логика позволяет добавить такой параметр, как степень уверенности эксперта.

В качестве инструмента для работы с нечеткими когнитивными картами авторы предлагают использовать программное средство FCMapper, созданное на основе электронных таблиц Excel. Внешний вид нечеткой когнитивной карты представлен на рис. 10.



Рис. 10. Когнитивная карта экономической ситуации [1]

С помощью имитационного моделирования предполагается рассматривать различные сценарии. Для этого в первую очередь выявляется стабильное, долгосрочное состояние системы с помощью матрицы смежности. Далее определяются три сценария с различными значениями параметров, для которых производится расчет. Из результатов видно, какие параметры оказывают наиболее сильное влияние на систему, как модель себя поведет, если один из параметров превысит свое значение в несколько раз, и так далее. Авторы считают, что использование нечетких когнитивных карт в совокупности с имитационным моделированием помогает увеличить возможности масштабирования модели и улучшить качество сценариев.

Проведенный анализ открытых источников информации по изучаемой теме показал, что применение нечеткой логики при имитационном моделировании сложных систем довольно распространено и используется для исследования технических, биологических, техногенных и социально-экономических объектов. Вне зависимости от применяемого метода имитационного моделирования нечеткая логика в большинстве случаев используется как механизм выработки неоднозначного решения и способ оперирования экспертными оценками.

Заключение

В настоящее время российские и зарубежные ученые широко применяют нечеткую логику при реализации имитационных моделей для изучения систем различной природы. В зависимости от применяемого метода имитационного моделирования можно выделить основные способы применения данного формального аппарата. Так, для большинства системно-динамических моделей характерно использование нечеткой логики для вычисления внутренних переменных, обладающих разной степенью неопределенности. Для моделей, разработанных с помощью агентного подхода, сложилась практика имитировать поведение отдельных агентов на основе нечетких продукционных правил. Если же рассматривать вопрос программной реализации способа интеграции нечеткой логики и имитационного моделирования, то в реализованных проектах можно встретить разные варианты ее осуществления. В целом, интеграция выполняется двумя основными способами: аппарат нечеткой логики реализуется непосредственно внутри компьютерной модели с использованием элементов конкретного инструментального средства моделирования либо он функционирует на стороне внешних специализированных пакетов (например, MATLAB), а модель взаимодействует с ними через программный интерфейс.

Проведенный анализ источников показал, что первоначальная идея о применении нечеткой логики для имитационного моделирования жизнеспособности региональных социально-экономических систем состоятельна. Опыт совместного использования нечеткой логики и имитационного моделирования обширен и затрагивает вопросы изучения самых разнообразных систем и объектов. При этом использование аппарата нечеткой логики как способа оперирования качественными оценками при имитационном моделировании также реализуемо на практике.

Список источников

1. Заграновская А., Эйсснер Ю. Моделирование сценариев развития экономической ситуации на основе нечетких когнитивных карт // Современная экономика: проблемы и решения. 2017. № 10. С. 33–47.
2. Халиуллина Д. Н., Быстров В. В. Теоретические основы оценки жизнеспособности региональных социально-экономических систем // Труды Кольского научного центра РАН. Серия: Технические науки. 2022. Т. 13, № 2. С. 78–92.
3. Лебедева М. Е. Нечеткая логика в экономике — формирование нового направления // Идеи и идеалы. 2019. Т. 11, № 1–1. С. 197–212. doi:10.17212/2075-0862-2019-11.1.1-197-212.
4. Сажина Ю. В., Липинский Л. В., Свиридова А. С. Применение имитационного моделирования при разработке системы, основанной на нечеткой логике // Актуальные проблемы авиации и космонавтики. 2018. Т. 2, № 4(14). С. 202–206.

5. Верещиков Д. В., Волошин В. А., Ивашков С. С., Васильев Д. В. Применение нечеткой логики для создания имитационной модели управляющих действий летчика // Труды МАИ. 2018. № 99. С. 13.
6. Марушко В. А. Имитационное моделирование системы управления установками уличного освещения в среде MATLAB с помощью нечеткой логики // Инновационное развитие науки и образования: сб. ст. XVII Междунар. науч.-практич. конф. (Пенза, 30 янв. 2023 г.). Пенза: Наука и Просвещение (ИП Гуляев Г. Ю.), 2023. С. 33–36.
7. Kim Y, Lee M., Hong J., Lee Yun-Sik, Wee J., Cho K. Development of a fuzzy logic-embedded system dynamics model to simulate complex socio-ecological systems // Ecological Modelling. — Elsevier, 2024. Vol. 493. 110738. <https://doi.org/10.1016/j.ecolmodel.2024.110738>.
8. Md. Aminur Rahman, Kazi Main Uddin Ahmed, Md. Rayhanus Sakib modeling of a novel fuzzy based overcurrent relay using Simulink // International Journal of Scientific & Technology Research — 2012. Vol. 1, Issue 4. P. 24–29.
9. Epstein, J. G., Möhring M., Klaus G. T. Fuzzy-logical rules in a multi-agent system // Social-Economic Phenomena and Processes. 2006. No. 1-2(1-2). P. 35–39.
10. Hassan, S., Salgado, M., Pavon, J. Friends Forever: Social Relationships with a Fuzzy Agent-Based Model // Hybrid Artificial Intelligence Systems. HAIS 2008. Lecture Notes in Computer Science. — Springer, Berlin, Heidelberg. 2008. Vol. 5271. P. 523–532. https://doi.org/10.1007/978-3-540-87656-4_65.
11. Коваленко, Н. Е., Новиков С. П. Разработка имитационной модели виртуального поселка с использованием системы принятия решений на основе нечеткой логики // Научный альманах. 2019. № 5–2(55). С. 35–41.
12. Корчевская, О. В., Бобровский, В. Е. Разработка агент-ориентированной модели и программного обеспечения трудовой миграции населения // Научно-технический вестник Поволжья. 2023. № 8. С. 74–77.
13. Низамутдинов, М. М., Ахметзянова, М. И. Подход к моделированию взаимовлияния миграционной активности населения и экономического развития региона с применением нечеткой логики, геоинформационных технологий и агент-ориентированного моделирования // Системное моделирование социально-экономических процессов: тр. XLVI Междунар. науч. школы-семинара (Уфа, 09–15 окт. 2023 г.). Воронеж: Издательство «Истоки», 2024. С. 91–97. doi:10.5281/zenodo.10938968.
14. Liu, S., Triantis, K.P., Sarangi, S. Representing qualitative variables and their interactions with fuzzy logic in system dynamics modeling // Systems Research and Behavioral Science. 2011 Vol. 28. P. 245–263. <https://doi.org/10.1002/sres.1064>.
15. Seresht, N.G., Fayek, A.R. Neuro-fuzzy system dynamics technique for modeling construction systems. // Applied Soft Computing. 2020. Vol. 93, 106400. <https://doi.org/10.1016/j.asoc.2020.106400>.
16. Song, Z.; Zhang, Z.; Lyu, F.; Bishop, M.; Liu, J.; Chi, Z. From Individual Motivation to Geospatial Epidemiology: A Novel Approach Using Fuzzy Cognitive Maps and Agent-Based Modeling for Large-Scale Disease Spread. // Sustainability, 2024. Vol. 16. 5036. <https://doi.org/10.3390/su16125036>.

References

1. Zagranovskaya A., Ejssner YU. Modelirovanie scenariy razvitiya ekonomicheskoy situatsii na osnove nechetkikh kognitivnykh kart [Modeling scenarios for the development of the economic situation based on fuzzy cognitive maps]. *Sovremennaya ekonomika: problemy i resheniya* [Modern economy: problems and solutions], 2017, No 10, pp. 33–47. (In Russ.).
2. Khaliullina D.N., Bystrov V.V. Teoreticheskie osnovy ocenki zhiznesposobnosti regional'nykh social'no-ekonomicheskikh sistem [Theoretical foundations for assessing the viability of regional socio-economic systems]. *Trudy Kol'skogo nauchnogo centra RAN. Seriya: Tekhnicheskie nauki* [Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences]. 2022. Vol. 13. No. 2. Pp. 78–92. (In Russ.).
3. Lebedeva, M. E. Nechyotkaya logika v ekonomike - formirovaniye novogo napravleniya [Fuzzy logic in economics - formation of a new direction]. *Idey i Idealy* [Ideas and ideals], 2019, vol. 11, No. 1–1, pp. 197–212. doi:10.17212/2075-0862-2019-11.1.1-197-212. (In Russ.).

4. Sazhina, Yu. V., Lipinskiy, L. V., & Sviridova, A. S. Primenenie imitatsionnogo modelirovaniya pri razrabotke sistemy, osnovannoy na nechyotokoy logike [Application of simulation modeling in the development of a fuzzy logic-based system]. *Aktual'nyye Problemy Aviatsii i Kosmonavtiki* [Current problems of aviation and cosmonautics]. 2018, vol. 2, No. 4(14), pp. 202–206. (In Russ.).
5. Vereshchikov, D. V., Voloshin, V. A., Ivashkov, S. S., & Vasilyev, D. V. Primenenie nechyotokoy logiki dlya sozdaniya imitatsionnoy modeli upravlyayushchikh deystviy letchika [Applying fuzzy logic for developing simulation model of pilot's control actions]. *Trudy MAI* [Proceedings of MAI], 2018, No 99, P. 13. (In Russ.).
6. Marushko, V. A., & Semenova, N. G. Imitatsionnoe modelirovanie sistemy upravleniya ustanovkami ulichnogo osveshcheniya v srede MATLAB s pomoshch'yu nechyotokoy logiki [Simulation modeling of street lighting control systems in MATLAB using fuzzy logic]. *Innovacionnoe razvitie nauki i obrazovaniya: sbornik statej XVII Mezhdunarodnoj nauchno-prakticheskoy konferencii, Penza, 30 yanvarya 2023 goda* [Innovative development of science and education: collection of articles of the XVII International Scientific and Practical Conference]. Penza, Publ. Nauka i Prosveshchenie, 2023, pp. 33–36. (In Russ.).
7. Kim Y, Lee M., Hong J., Lee Yun-Sik, Wee J., Cho K. Development of a fuzzy logic-embedded system dynamics model to simulate complex socio-ecological systems. *Ecological Modelling*. Publ. Elsevier, 2024, Vol. 493, 110738. <https://doi.org/10.1016/j.ecolmodel.2024.110738>.
8. Md. Aminur Rahman, Kazi Main Uddin Ahmed, Md. Rayhanus Sakib modeling of a novel fuzzy based overcurrent relay using Simulink. *International Journal of Scientific & Technology Research*, 2012, Vol. 1, Issue 4, pp. 24–29.
9. Epstein J. G., Möhring M., Klaus G. T. Fuzzy-logical rules in a multi-agent system. *Social-Economic Phenomena and Processes*, 2006, No. 1-2(1-2), pp. 35–39.
10. Hassan S., Salgado M., Pavon J. Friends Forever: Social Relationships with a Fuzzy Agent-Based Model. *Hybrid Artificial Intelligence Systems. HAIS 2008. Lecture Notes in Computer Science*. Publ. Springer, Berlin, Heidelberg, 2008, Vol. 5271, pp. 523–532. https://doi.org/10.1007/978-3-540-87656-4_65.
11. Kovalenko N. E., & Novikov S. P. Razrabotka imitatsionnoy modeli virtual'nogo poselka s ispol'zovaniem sistemy prinyatiya resheniy na osnove nechyotokoy logiki [Development of a simulation model of a virtual village using a fuzzy logic-based decision-making system]. *Nauchnyy Almanakh* [Scientific Almanac], 2019, No 5–2 (55), pp. 35–41. (In Russ.).
12. Korchevskaya, O. V., & Bobrovskiy, V. E. Razrabotka agent-orientirovannoy modeli i programmnoy obespecheniya trudovoy migracii naseleniya [Development of agent-oriented model and software of labor migration of the population]. *Nauchno-tekhnicheskij vestnik Povolzh'ya* [Scientific and technical volga region bulletin], 2023, No 8, Pp. 74–77. (In Russ.).
13. Nizamutdinov, M. M., Ahmetzyanova, M. I. Podhod k modelirovaniyu vzaimovliyaniya migracionnoj aktivnosti naseleniya i ekonomicheskogo razvitiya regiona s primeneniem nechetkoj logiki, geoinformacionnyh tekhnologij i agent-orientirovannogo modelirovaniya [An approach to modeling the interaction of migration activity of the population and economic development of the region using fuzzy logic, geoinformation technologies and agent-oriented modeling]. *Sistemnoe modelirovanie social'no-ekonomicheskikh processov : trudy 46-oy Mezhdunarodnoj nauchnoj shkoly-seminara, Ufa, 09–15 oktyabrya 2023 goda* [System modeling of socio-economic processes : proceedings of the 46th International Scientific School Seminar]. Voronezh: Istoki Publishing House, 2024, Pp. 91–97. doi:10.5281/zenodo.10938968. (In Russ.).
14. Liu S., Triantis K.P., Sarangi S. Representing qualitative variables and their interactions with fuzzy logic in system dynamics modeling. *Systems Research and Behavioral Science*, 2011, Vol. 28, pp. 245–263. <https://doi.org/10.1002/sres.1064>.
15. Seresht N.G., Fayek A.R. Neuro-fuzzy system dynamics technique for modeling construction systems. *Applied Soft Computing*, 2020, Vol. 93, 106400. <https://doi.org/10.1016/j.asoc.2020.106400>.

16. Song Z.; Zhang Z.; Lyu F.; Bishop M.; Liu J.; Chi Z. From Individual Motivation to Geospatial Epidemiology: A Novel Approach Using Fuzzy Cognitive Maps and Agent-Based Modeling for Large-Scale Disease Spread. *Sustainability*, 2024, Vol. 16, 5036. <https://doi.org/10.3390/su16125036>.

Информация об авторах

Е. О. Неупокоева — стажер-исследователь;

С. Н. Малыгина — кандидат технических наук, научный сотрудник;

В. В. Быстров — кандидат технических наук, ведущий научный сотрудник.

Information about the authors

E. O. Neupokoeva — Trainee Researcher;

S. N. Malygina — Candidate of Science (Tech.), Research Fellow;

V. V. Bystrov — Candidate of Science (Tech.), Lead Research Fellow

Статья поступила в редакцию 18.10.2024; одобрена после рецензирования 05.11.2024; принята к публикации 11.11.2024.
The article was submitted 18.10.2024; approved after reviewing 05.11.2024; accepted for publication 11.11.2024.

Научная статья
УДК 004.832
doi:10.37614/2949-1215.2024.15.3.007

ОБЗОР МЕТОДОВ ПОИСКА ЧАСТЫХ ПАТТЕРНОВ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Александр Анатольевич Зуенко^{1✉}, Ольга Владимировна Фридман²

^{1,2}*Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия*

¹*zuenko@iimm.ru[✉], <https://orcid.org/0000-0002-7165-6651>*

²*ofridman@iimm.ru, <https://orcid.org/0000-0003-1897-4922>*

Аннотация

В статье рассматривается одна из задач интеллектуального анализа данных, а именно: задача поиска особого вида зависимостей в данных — частых паттернов. На основе частых паттернов могут строиться ассоциативные правила между признаками. Приводится обзор наиболее популярных методов решения данной задачи. Также рассматривается тесно связанная с ней задача генерации формальных понятий на основе формального контекста и один из методов ее решения — метод «замыкай по одному». К недостаткам рассматриваемых методов относится трудоемкость их модификации при наличии дополнительных (помимо частоты встречаемости в обучающей выборке) требований к искомому паттерну. Делается вывод о необходимости развития существующих методов извлечения паттернов, а также целесообразности разработки новых подходов для решения задач поиска «интересных паттернов», а также поиска паттернов с дополнительными требованиями к их виду.

Ключевые слова:

интеллектуальный анализ данных, частые паттерны, ассоциативные правила, анализ формальных понятий

Благодарности:

работа выполнена в рамках НИР «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (регистрационный номер 122022800547-3).

Для цитирования:

Зуенко А. А., Фридман О. В. Обзор методов поиска частых паттернов для интеллектуального анализа данных // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 82–96. doi:10.37614/2949-1215.2024.15.3.007.

Original article

SURVEY OF FREQUENT PATTERN SEARCH METHODS FOR DATA MINING

Aleksandr A. Zuenko^{1✉}, Olga V. Fridman²

^{1,2}*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre
of the Russian Academy of Sciences, Apatity, Russia*

¹*zuenko@iimm.ru[✉], <https://orcid.org/0000-0002-7165-6651>*

²*ofridman@iimm.ru, <https://orcid.org/0000-0003-1897-4922>*

Abstract

The article discusses one of the problems of data mining, namely: the problem of searching for a special type of data dependencies - frequent patterns. Associative rules between features can be built on the basis of frequent patterns. A survey of the most popular methods for solving this problem is provided. The closely related problem of generating formal concepts based on a formal context is also considered, and one of the methods for solving it — the "close-by-one" method. The disadvantages of the considered methods include the complexity of their modification in the presence of additional (in addition to the frequency of occurrence in the training sample) requirements for the desired pattern. It is concluded that it is necessary to develop existing methods of pattern discovery, as well as the expediency of developing new approaches to solve the problems of searching for "interesting patterns", as well as searching for patterns with additional requirements for their type.

Keywords:

data mining, frequent patterns, associative rules, formal concept analysis

Acknowledgments:

the study was carried out within the framework of the Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences state assignment of the Ministry of Science and Higher Education of the Russian Federation, research topic "Development of theoretical and organizational and technical foundations of information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation" (registration number of the research topic 122022800547-3).

For citation:

Zuenko A. A., Fridman O. V. Survey of frequent pattern search methods for data mining // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 82–96. doi:10.37614/2949-1215.2024.15.3.007.

Введение

В настоящей статье рассматриваются методы решения задачи поиска частых паттернов при обработке данных, которая впервые была поставлена в работе [1].

Пусть задан набор транзакций (транзакционная база данных, обучающая выборка, набор объектов), где каждая транзакция представляет собой набор элементов (предметный набор, набор признаков). Паттерном называется любое подмножество элементов. Частый паттерн — это такое множество элементов (признаков), которое достаточно часто встречается среди объектов обучающей выборки. Обычно задается некий порог θ , и все паттерны, которые встречаются не менее, чем в θ объектах обучающей выборки считаются частыми.

Задача поиска частых паттернов используются как составляющая многих более сложных задач интеллектуального анализа данных, в частности, при поиске ассоциативных правил.

Ассоциативное правило – это выражение вида $X \Rightarrow Y$, где X и Y – наборы элементов. Интуитивный смысл такого правила заключается в том, что транзакции в базе данных, содержащие элементы из X , как правило, также содержат элементы из Y . Ассоциативное правило обычно характеризуется двумя числами, которые называются *поддержка* и *достоверность*. Поддержка правила $X \Rightarrow Y$ – это процент транзакций из транзакционной базы данных, которые содержат как X , так и Y . Достоверность правила – это условная вероятность $P(Y|X)$, т. е. вероятность того, что если в транзакции присутствуют элементы из множества X , то в них будут присутствовать элементы из Y . Пример ассоциативного правила: 97 % клиентов, покупающих, например, творог, также покупают сметану. Здесь 97 % — это достоверность правила. Проблема поиска ассоциативных правил состоит в том, чтобы найти все правила, которые удовлетворяют заданной пользователем минимальной поддержке и минимальной достоверности.

Известны следующие приложения методов поиска ассоциативных правил:

- анализ потребительской корзины;
- размещение предметов;
- обнаружение мошенничества;
- медицинские исследования;
- реинжиниринг процесса и др.

Далее приводится краткое описание нескольких алгоритмов, предназначенных для поиска ассоциативных правил, и на простом примере транзакционной базы данных рассматривается их работа.

Описание примера

Пусть имеются транзакции купленных товаров, представленные в табл. 1.

Таблица 1

Транзакции купленных товаров

№ транзакции	Купленные товары
1	Карандаши, кнопки, клей
2	Кнопки, бумага
3	Кнопки, скрепки
4	Карандаши, кнопки, бумага
5	Карандаши, скрепки
6	Кнопки, скрепки
7	Карандаши, скрепки
8	Карандаши, кнопки, скрепки, клей
9	Карандаши, кнопки, скрепки

В соответствии с базой транзакций создадим объектно-признаковую таблицу (табл. 2).

Таблица 2

Объектно-признаковая таблица

Обозначение	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
№ транзакции	Карандаши	Кнопки	Скрепки	Бумага	Клей
1	1	1			1
2		1		1	
3		1	1		
4	1	1		1	
5	1		1		
6		1	1		
7	1		1		
8	1	1	1		1
9	1	1	1		

Зададим значение минимальной поддержки (*minsup*), равное двум. Требуется найти все частые паттерны с данной поддержкой.

Алгоритм Априори

Первым из наиболее известных примеров практической реализации поиска ассоциативных правил является алгоритм Априори (англ. Apriori) [2], который позволяет находить частые паттерны в данных.

Введем некоторые обозначения: *k*-множеством будем называть множество, состоящее из *k* элементов. Обозначим через L_k множество всех часто встречающихся *k*-множеств. Объединение L_k по всем *k* дает все искомое множество частых паттернов.

Построение L_k выполняется по шагам. Сначала находится L_1 (множество одноэлементных частых паттернов). Затем для каждого фиксированного $k \geq 2$, используя найденное множество L_{k-1} , определяется L_k . Процесс завершается, когда *k* станет больше максимального количества элементов. Определение L_k при известном L_{k-1} выполняется в два шага:

- 1) генерируются множества — кандидаты C_k ;
- 2) затем из этого множества исключаются лишние элементы. Полученное таким образом множество и будет равно L_k .

Шаг 1.

Находим все 1-элементные частые паттерны с заданным значением *minsup* (т. е. записи, представленные в БД минимум 2 раза):

$$L_1 = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}\}.$$

Шаг 2.

Задаем $k = 2$ и запускаем процедуру `apriori_gen`, которая для *i*-элементных частых множеств признаков порождает (*i* + 1)-надмножества и возвращает только множество потенциально частых кандидатов. Таким образом, из множества L_1 формируется множество кандидатов C_2 , отсортированных по алфавиту. Результат:

$$C_2 = \{\{a, b\}, \{a, c\}, \{a, d\}, \{a, e\}, \\ \{b, c\}, \{b, d\}, \{b, e\}, \\ \{c, d\}, \{c, e\}, \\ \{d, e\}\}.$$

Шаг 3.

Инициализируем счетчик нулевым значением (обозначим его как `c.count`).

Шаг 4.

Перебираем все записи БД и находим те, которые содержатся в C_2 .

Первая запись $r_1 = \{a, b, e\}$, а $Cr_1 = \{\{a, b\}, \{a, e\}, \{b, e\}\}$.

Вторая запись $r_2 = \{b, d\}$, а $Cr_2 = \{\{b, d\}\}$.

Третья запись $r_3 = \{b, c\}$, а $Cr_3 = \{\{b, c\}\}$.

Четвертая запись $r_4 = \{a, b, d\}$, а $Cr_4 = \{\{a, b\}, \{a, d\}, \{b, d\}\}$.

Пятая запись $r_5 = \{a, c\}$, а $Cr_5 = \{\{a, c\}\}$.

Шестая запись $r_6 = \{b, c\}$, а $Cr_6 = \{\{b, c\}\}$.

Седьмая запись $r_7 = \{a, c\}$, а $Cr_7 = \{\{a, c\}\}$.

Восьмая запись $r_8 = \{a, b, c, e\}$, а $Cr_8 = \{\{a, b\}, \{a, c\}, \{a, e\}, \{b, c\}, \{b, e\}, \{c, e\}\}$.

Девятая запись $r_9 = \{a, b, c\}$, а $Cr_9 = \{\{a, b\}, \{a, c\}, \{b, c\}\}$.

Шаг 5.

Проверяем, для каких множеств-кандидатов значения счетчика $s.count$ не ниже заданного уровня поддержки $minsup$ — т. е. ≥ 2 :

$L_2 = \{\{a, b\}, \{a, c\}, \{a, e\}, \{b, c\}, \{b, d\}, \{b, e\}\}$.

Шаг 6.

Осуществляется переход к шагу 2 алгоритма, $k = 3$ и снова выполняется процедура $apriori_gen$ по данным L_2 . Формируются следующие пары значений:

$\{a, b\}:\{a, c\}, \{a, c\}:\{a, e\}, \{a, e\}:\{b, c\}, \{b, c\}:\{b, d\}, \{b, d\}:\{b, e\}$.

Далее получим множества из трех элементов:

$C_3 = \{a, b, c\}, \{a, c, e\}, \{a, e, b\}, \{b, c, d\}, \{b, d, e\}$.

Проверяем, для каких множеств-кандидатов значение счетчика не ниже заданного уровня поддержки — т. е. ≥ 2 : $L_3 = \{\{a, b, c\}, \{a, b, e\}\}$.

Затем осуществляется очередной переход к шагу 2, $k = 4$ и выполняется процедура $apriori_gen$ по данным L_3 . В данном случае $C_4 = \{a, b, c, e\}$, но такой набор имеет поддержку, равную 1, поэтому $L_4 = \{\}$ и выполнение алгоритма завершается.

Результат:

$L_S = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a, b\}, \{a, c\}, \{a, e\}, \{b, c\}, \{b, d\}, \{b, e\}, \{a, b, c\}, \{a, b, e\}\}$.

Из данного набора частых паттернов можно получить следующие ассоциативные правила (табл. 3).

Таблица 3

Полученные ассоциативные правила

Купленные товары	Рекомендованные товары
Карандаши	Кнопки
Карандаши	Скрепки
Карандаши	Клей
Кнопки	Скрепки
Кнопки	Бумага
Кнопки	Клей
Карандаши и кнопки	Скрепки
Карандаши и кнопки	Клей

Перечислим достоинства и недостатки алгоритма Априори [3].

Достоинства:

— простота;

— быстрое уменьшение числа сгенерированных кандидатов при установке высокой минимальной поддержки или относительно разреженном базовом наборе.

Недостатки:

- многократное сканирование базового набора;
- большое число сгенерированных кандидатов при слишком большом наборе данных или при слишком низкой поддержке.

Алгоритм эффективен только для небольших наборов либо при высоком уровне минимальной поддержки. Для улучшения работы используются его модификации, направленные на уменьшение числа сканирований входного набора, числа сгенерированных кандидатов, распараллеливание [4, 5]. К сожалению, далеко не во всех случаях эти модифицированные алгоритмы позволяют исправить ситуацию.

Далее рассмотрим еще два популярных алгоритма.

Алгоритм FP-GROWTH

Один из самых эффективных алгоритмов. Позволяет избежать не только затратной процедуры генерации кандидатов, но и многократного сканирования входного набора. Метод базируется на предварительной обработке входного набора и преобразование его в специальную компактную древовидную структуру FP-дерева (frequent pattern tree), и лишь затем происходит вычисление частых наборов [3].

В общем виде алгоритм можно представить следующей последовательностью шагов [4]:

Шаг 1. Производится сканирование входного набора, и все элементы каждой транзакции сортируются в порядке убывания поддержки этих элементов во всем базовом наборе.

Шаг 2. Фильтрация. Производится удаление тех элементов, для которых значение поддержки меньше заданного пользователем значения минимальной поддержки.

Шаг 3. Построение префиксного FP-дерева из оставшихся элементов.

Шаг 4. Извлечение частых паттернов.

Узлом FP-дерева является структура, которая хранит значение узла, ссылки на все дочерние элементы и его значение поддержки для текущего узла. Построение префиксного дерева происходит в несколько этапов:

Этап 1. Построение корневого узла.

Этап 2. Для каждого элемента каждой отсортированной транзакции из входного набора строятся узлы по следующему правилу: если для очередного элемента в текущем узле есть потомок, содержащий этот элемент, то новый узел не создается, а поддержка этого потомка увеличивается на 1, в противном случае создается новый узел-потомок с поддержкой 1. Текущим узлом при этом становится найденный или построенный узел.

Для извлечения частых паттернов необходимо построить условные деревья для каждого элемента. Условное поддереве множества A — это FP-дерево, содержащее только транзакции, в которые входит множество A . Приведем описание алгоритма извлечения частых паттернов [3]. Для каждого элемента в дереве, начиная с элемента с наименьшей поддержкой, необходимо выполнить следующее.

Шаг 1. Добавить этот элемент во множество A .

Шаг 2. Построить условное дерево по этому элементу. В случае, если такое дерево оказывается пустым, то записать в результат элементы множества A (они и будут очередным популярным набором), иначе выполнить этот алгоритм для построенного условного дерева.

Шаг 3. Исключить элемент из множества A .

Шаг 4. Исключить элемент из дерева. Таким образом, дерево проходится рекурсивно снизу вверх полностью, и при этом генерируются все возможные популярные наборы.

Рассмотрим работу этого алгоритма на простом примере, представленном выше.

Выпишем элементы в порядке уменьшения поддержки (см. табл. 2):

$(b:7)$, $(a:6)$, $(c:6)$, $(d:2)$, $(e:2)$. Все элементы имеют поддержку большую или равную заданному значению минимальной поддержки.

Анализируя транзакции (см. табл. 1), получаем начальные наборы:

(b, a, e) , (b, d) , (b, c) , (b, a, d) , (a, c) , (b, c) , (a, c) , (b, a, c, e) , (b, a, c) .

Получаем префиксное дерево (FP-дерево), приведенное на рис. 1.

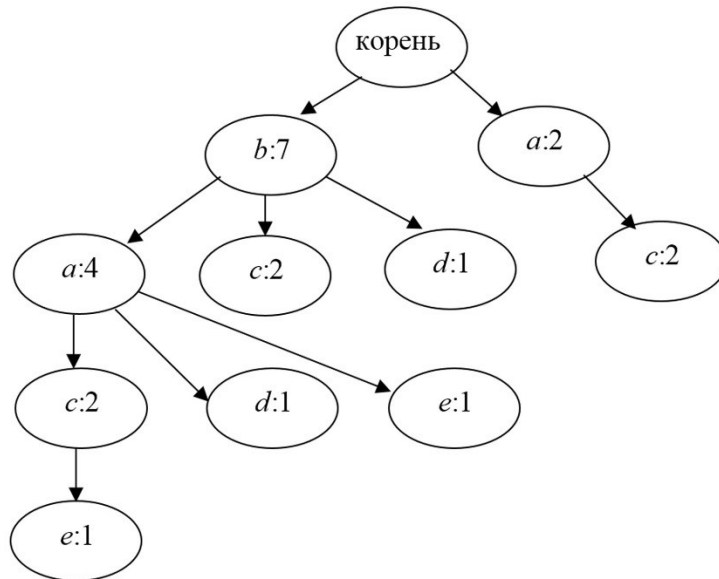


Рис. 1. Префиксное дерево (FP-дерево)

Теперь построим условные деревья для каждого элемента, начиная с того, у которого наименьшая поддержка, — $e:1$ (см. рис. 2).

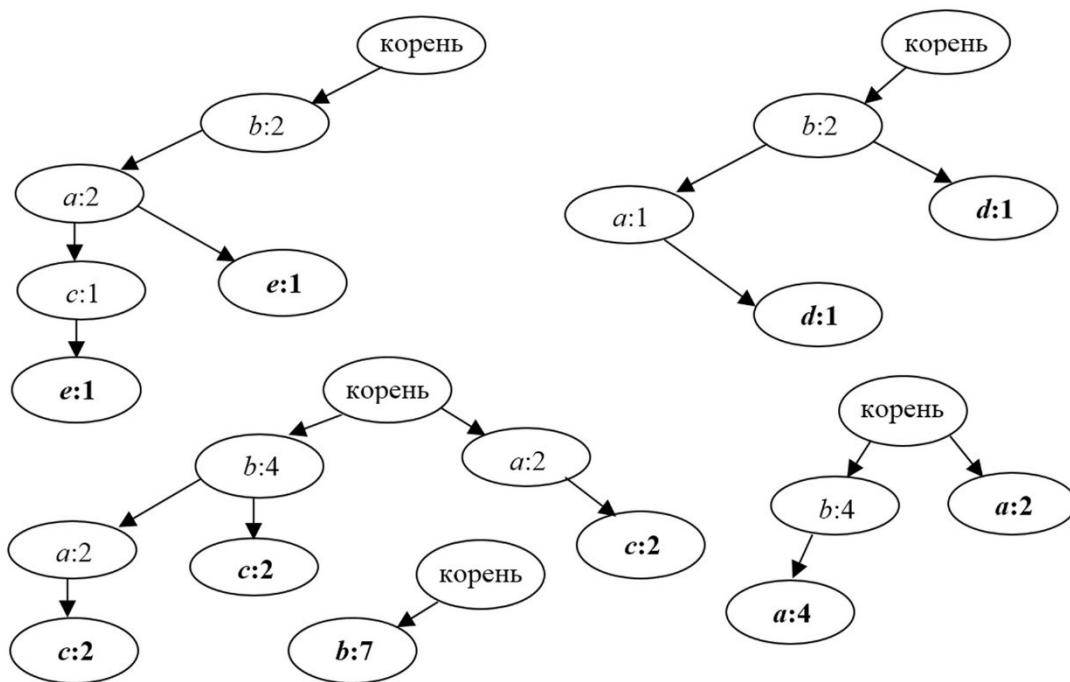


Рис. 2. Условные деревья

Получаем набор частных элементов $\{b\}$, $\{a\}$, $\{c\}$, $\{d\}$, $\{e\}$. Отсюда получаем двухэлементные частые наборы: $\{\{b, a\}, \{b, c\}, \{b, d\}, \{b, e\}, \{a, c\}, \{a, e\}\}$, а также два трехэлементных набора — $\{\{b, a, c\}, \{b, a, e\}\}$.

Достоинства алгоритма [3]:

- позволяет избежать затратной процедуры генерации кандидатов, характерной для алгоритма Априори;
- сжатие базового набора в компактную структуру, обеспечивающее быстрое и полное извлечение предметных наборов;
- число сканирования входного набора сокращено до двух;
- размер дерева обычно меньше размера входного набора данных.

Недостатки алгоритма:

- построение дерева – затратная по времени операция;
- в некоторых случаях, вследствие большого числа узлов и связей, размер FP-дерева может намного превышать размер входного набора данных.

Существует множество модификаций алгоритма, направленных на улучшение тех или иных его свойств. Одна из наиболее быстрых реализаций алгоритма описана в работе [6], а вариант реализации, направленный на уменьшение объема занимаемой памяти, — в исследовании [7]. Далее рассмотрим еще один алгоритм, который также, как и предыдущий, является усовершенствованием алгоритма Априори.

Алгоритм ECLAT

Алгоритм ECLAT (Equivalence Class Clustering and bottom-up Lattice Traversal) является одним из популярных методов майнинга ассоциативных правил и представляет собой наиболее эффективную и масштабируемую версию алгоритма Априори. Рассматриваемые выше алгоритмы Априори и FP-growth используют так называемое горизонтальное представление множеств. Алгоритм ECLAT [8] на первом шаге своей работы преобразует горизонтальное представление в вертикальное (так называемое TID-представление) и в дальнейшем ведется именно его обработка.

Шаги этого алгоритма аналогичны соответствующим шагам алгоритма Априори, кроме функции вычисления поддержки кандидата, которая теперь не требует сканирования базы. Основная идея состоит в том, чтобы использовать пересечения наборов идентификаторов транзакций (TID-множества) для вычисления значения поддержки кандидата и избежать создания подмножеств, которые не существуют в дереве префиксов. При первом вызове функции используются все отдельные элементы вместе с их наборами данных. Затем функция вызывается рекурсивно, и при каждом рекурсивном вызове каждая пара «элемент — TID-множество» проверяется и сопрягается с другими парами «элемент — TID-множество». Этот процесс продолжается до тех пор, пока удастся получать новые пары «элемент — TID-множество».

Рассмотрим задачу поиска частых паттернов при помощи алгоритма ECLAT на том же примере (исходные данные см. в табл. 1, 2).

В табл. 4 представлена база транзакций (табл. 1) с учетом введенных обозначений.

Таблица 4

Транзакции купленных товаров

№ транзакции	Элемент
1	<i>a, b, e</i>
2	<i>b, d</i>
3	<i>b, c</i>
4	<i>a, b, d</i>
5	<i>a, c</i>
6	<i>b, c</i>
7	<i>a, c</i>
8	<i>a, b, c, e</i>
9	<i>a, b, c</i>

Преобразуем исходную таблицу в TID-представление (табл. 5).

Таблица 5

Вертикальное представление (TID) базы транзакций

Элемент	№ транзакции	Поддержка
<i>a</i>	1, 4, 5, 7, 8, 9	6
<i>b</i>	1, 2, 3, 4, 6, 8, 9	7
<i>c</i>	3, 5, 6, 7, 8, 9	6
<i>d</i>	2, 4	2
<i>e</i>	1, 8	2

Теперь рассмотрим попарное соединение оставшихся элементов (табл. 6).

Таблица 6

Вертикальное представление (TID) базы транзакций

Элемент	№ транзакции	Поддержка
<i>a, b</i>	1, 4, 8, 9	4
<i>a, c</i>	7, 8, 9	3
<i>a, d</i>	4	1
<i>a, e</i>	1, 8	2
<i>b, c</i>	3, 6, 8, 9	2
<i>b, d</i>	2, 4	2
<i>b, e</i>	1, 8	2
<i>c, d</i>	—	0
<i>c, e</i>	8	1
<i>d, e</i>	—	0

Удалим пары, поддержка которых меньше заданной минимальной поддержки (помечены цветом в табл. 6). Полученный результат см. в табл. 7.

Таблица 7

Вертикальное представление (TID) базы транзакций

Элемент	№ транзакции	Поддержка
<i>a, b</i>	1, 4, 8, 9	4
<i>a, c</i>	7, 8, 9	3
<i>a, e</i>	1, 8	2
<i>b, c</i>	3, 6, 8, 9	2
<i>b, d</i>	2, 4	2
<i>b, e</i>	1, 8	2

Теперь сформируем трехэлементные множества (табл. 8).

Таблица 8

Вертикальное представление (TID) базы транзакций

Элемент	№ транзакции	Поддержка
<i>a, b, c</i>	8, 9	2
<i>a, b, d</i>	4	1
<i>a, b, e</i>	1, 8	2
<i>b, c, d</i>	—	0
<i>b, c, e</i>	8	1
<i>c, d, e</i>	—	0

Аналогично предыдущему шагу удаляем лишние строки (помечены цветом в табл. 8).
Полученный результат см. в табл. 9.

Таблица 9

Вертикальное представление (TID) базы транзакций

Элемент	№ транзакции	Поддержка
a, b, c	8, 9	2
a, b, e	1, 8	2

На следующем шаге получаем единственный четырехэлементный паттерн (a, b, c, e) с поддержкой 1, тогда как заданное значение минимальной поддержки равно 2.

Работа алгоритма закончена. Получаем набор частых элементов $\{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}\}$, а также получаем двухэлементные частые наборы: $\{\{a, b\}, \{a, c\}, \{a, e\}, \{b, c\}, \{b, d\}, \{b, e\}\}$, и два трехэлементных набора: $\{\{a, b, c\}, \{a, b, e\}\}$.

Таким образом, получаем ассоциативные правила, представленные в табл. 3. Полученные правила аналогичны тем, которые были получены при помощи алгоритмов Априори и FP-GROWTH.

Достоинства алгоритма [3]:

- простота;
- поддержка для любого элемента рассчитывается без сканирования базового набора;
- число сканирований базового набора сокращено до одного раза.

Преимущества перед алгоритмом Априори:

- 1) требования к памяти: алгоритм ECLAT использует меньше памяти, чем алгоритм Априори;
- 2) скорость: алгоритм ECLAT обычно быстрее, чем алгоритм Априори;
- 3) количество вычислений: алгоритм ECLAT не предполагает многократного сканирования данных для расчета отдельных значений поддержки.

Недостатки алгоритма:

- TID-множества могут оказаться слишком большими, поэтому операции с ними могут занимать длительное время;
- большое число сгенерированных кандидатов при малом уровне минимальной поддержки.

Описание одной из оптимальных программных реализаций можно найти в работе [9].

Далее рассмотрим алгоритм, также позволяющий получать ассоциативные правила, но основанный на анализе формальных понятий.

Алгоритм «замыкай по одному» (Close by One – CbO)

Анализ формальных понятий (АФП, Formal Concept Analysis — FCA) — это прикладная ветвь теории решеток, математической дисциплины, которая возникла в начале 1980-х гг. За последние три десятилетия АФП стал популярным, ориентированным на человека инструментом для представления знаний и анализа данных с многочисленными приложениями, в областях поиска информации с акцентом на аспекты визуализации, машинного обучения, интеллектуального анализа данных и обнаружения знаний, интеллектуального анализа текста и др.

Важным достоинством разбираемого в разделе алгоритма CbO является то, что он позволяет получать не просто частые паттерны, а частые замкнутые паттерны (для замкнутых паттернов не существует паттернов меньшей размерности с такой же поддержкой). Обычно среди замкнутых паттернов осуществляется поиск интересных зависимостей, поскольку они представляют собой своеобразный базис в пространстве паттернов.

При использовании математического аппарата АФП [10, 11] для решения обозначенного класса задач интеллектуального анализа данных термину «замкнутый паттерн» соответствует термин «содержание формального понятия» (*intent*), а термину «окрытие замкнутого паттерна» — термин «объем формального понятия» (*extent*). Транзакционная база данных в АФП называется контекстом. Формальные понятия определяются с помощью соответствия Галуа и представляют собой пары

множеств вида: (объем, содержание) [10, 11]. Контекстом в АФП называют тройку $K = (G, M, I)$, где G — множество объектов; M — множество признаков, а отношение $I \subseteq G \times M$ говорит о том, какие объекты какими признаками обладают. Для произвольных $A \subseteq G$ и $B \subseteq M$ определены операторы Галуа: $A' = \{m \in M \mid \forall g \in A (g I m)\}$, $B' = \{g \in G \mid \forall m \in B (g I m)\}$.

Оператор " \prime " (двукратное применение оператора \prime) является оператором замыкания.

Множество объектов $A \subseteq G$ такое, что $A'' = A$ называется замкнутым. Пара множеств (A, B) таких, что $A \subseteq G$, $B \subseteq M$, $A' = B$ и $B' = A$ называется формальным понятием контекста K . Для множества объектов A множество их общих признаков A' служит описанием сходства объектов из множества A , а замкнутое множество A'' является кластером сходных объектов (с множеством общих признаков A'). Отношение "быть более общим понятием" задается следующим образом: $(A, B) \geq (C, D)$ тогда и только тогда, когда $A \supseteq C$.

В работах [10, 11] рассматриваются несколько алгоритмов, генерирующих множество всех формальных понятий и графов-диаграмм решеток понятий. Различные алгоритмы отличаются условием выхода из цикла, методом выбора подмножеств для вычисления замыканий, методом проверки того, было ли понятие сгенерировано ранее (тест на каноничность) и методом вычисления замыканий. Смысл теста на каноничность состоит в том, что для любого понятия существует уникальная каноническая генерация, которая указана в каждом конкретном алгоритме. Интенционал D , полученный из B , является каноническим, если B и D согласуются по всем атрибутам до текущего атрибута j . Если же в D есть атрибут, идущий до j , которого нет в B , то понятие считается неканоническим.

Все алгоритмы можно разделить на две категории: инкрементные алгоритмы [12–14], которые на i -м шаге создают набор понятий или граф-диаграмму для i первых объектов контекста, и пакетные, которые строят набор концептов и его граф-диаграмму для всего контекста с нуля [11, 15]. Инкрементные алгоритмы достраивают решетку посредством постепенного добавления объектов и пересечения с имеющимися понятиями. В противовес этому принципу «пакетные» алгоритмы выполняют все построение в один проход. Любой пакетный алгоритм обычно придерживается одной из двух стратегий: сверху вниз (от максимального объема к минимальному) или снизу вверх (от минимального объема к максимальному).

Алгоритм, предложенный в работе [16], продемонстрировал хорошую производительность для баз данных с очень большим количеством объектов. Аналогичные алгоритмы применялись для машинного обучения и анализа данных [14, 17]. Инкрементный алгоритм, предложенный в работе [18], использует «специальные» операции с объектами. Алгоритм, представленный в работе [19] использует так называемые бинарные диаграммы и работает в очень плотных контекстах.

Выбор алгоритма построения решетки понятий должен основываться на свойствах входных данных. Алгоритм [13] следует использовать для небольших и разреженных контекстов; для плотных контекстов следует использовать алгоритмы, основанные на критерии каноничности, линейные по количеству входных объектов [11]. Алгоритм, представленный в работе [9], хорошо работает в контекстах средней плотности, особенно когда необходимо построить граф-диаграммы. Эксперименты с реальными данными [10] показывают, что, когда нужны только понятия, лучшим выбором будет простой и интуитивно понятный алгоритм, представленный в работе [20].

В исследовании [10] рассматривается семейство алгоритмов SbO , основанных на вычислении замыканий для подмножеств G . Они следуют следующей схеме:

- 1) выбрать одно из условий, и пока это условие верно:
- 2) для некоторого множества $A \subseteq G$; вычислить $(A''; A')$;
- 3) если понятие $(A''; A')$ генерируется первый раз (или, как в некоторых алгоритмах, понятие генерируется в последний раз), добавить его в набор концептов.

Самый простой (наивный) алгоритм, соответствующий этой схеме, вычисляет замыкания всех подмножеств G , кроме пустого. Он выполняет проверку каноничности, просматривая все сгенерированные на данный момент понятия. Цикл выполняется 2^n раз, где $n = |G|$. Таким образом, количество итераций равно

или превышает количество понятий. На каждом этапе цикла сгенерированное понятие проверяется на каноничность, что требует времени, линейного по количеству понятий.

Кроме инкрементных и пакетных алгоритмов, также выделяют упорядоченные. Алгоритмы обходят решетку в некотором заданном порядке. Например, Close by One [10, 11] использует лексикографический порядок, чтобы определить, порождилось ли данное понятие в первый раз. Полученное понятие считается каноничным, если оно не предшествовало текущему по порядку.

Алгоритм Close by One (CbO) использует понятие каноничности и метод выбора подмножеств, представленный в работе [11], и является родоначальником для семейства алгоритмов, предложенных позднее (подробный обзор и сравнительный анализ этих алгоритмов приведен в исследовании [10]). Он использует промежуточную структуру, которая помогает более эффективно вычислять замыкания с помощью созданных понятий. Алгоритм CbO получает каждое новое замыкание из понятия, сгенерированного им на предыдущем шаге, путем пересечения его объема с объемом признака, который не принадлежит его содержанию.

Исходная версия алгоритма CbO использует дерево в качестве промежуточной структуры.

Дерево понятий можно построить следующим образом:

- 1) создать фиктивный корень, соответствующий понятию с пустым содержанием;
- 2) исследовать признаки из M и для каждого понятия дерева проверить, обладают ли рассматриваемым признаком все объекты понятия;
 - если да, добавить его в набор признаков понятия;
 - в противном случае сформировать новый узел и объявить его дочерним узлом текущего;
- 3) содержание соответствующего понятия равно содержанию родительского узла плюс исследуемый признак;
- 4) объем формируемого понятия является пересечением объема, соответствующего рассматриваемому признаку, и объема родительского узла;
- 5) проверить новый узел на каноничность;
- 6) если тест не пройден, удалить новый узел из дерева.

Алгоритм вычисляет понятия в соответствии с лексикографическим порядком, определенным на подмножествах M .

При помощи алгоритма CbO на основе объектно-признаковой таблицы (табл. 2) построим дерево понятий, которое позволит получить замкнутые частые паттерны, необходимые для формулирования ассоциативных правил.

На рис. 3 представлено дерево понятий. В левой части записи, обозначающей узел, перечислены объекты (номера транзакций), а в правой части — обозначения признаков (свойств объектов). Признаки при построении дерева упорядочены в лексикографическом порядке.

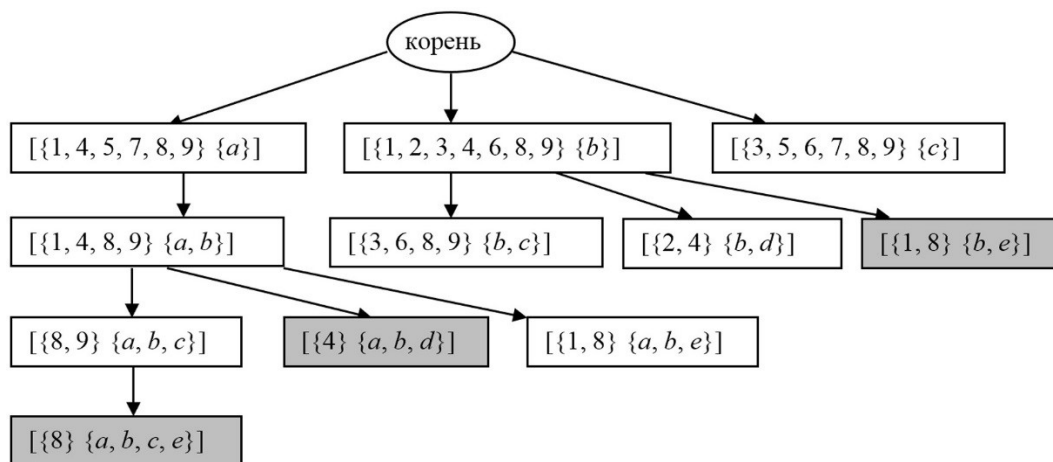


Рис. 3. Дерево понятий

В дереве отражены только те понятия, которые представляют собой частые замкнутые паттерны, поэтому некоторые узлы изначально отсутствуют (например, d и e таковыми не являются, поскольку не порождают понятий). Кроме того, некоторые понятия имеют поддержку меньше заданной минимальной поддержки, равной 2, а узел $\{b, e\}$ отображает понятие, которое входит в понятие $\{a, b, e\}$ и без понятия $\{a\}$ встречается только один раз. Понятие $\{a, b, c, e\}$ встречается всего один раз. Эти узлы в дереве понятий помечены цветом.

Таким образом, получаем решение в виде совокупности частых замкнутых паттернов:

$$\{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{b, d\}, \{a, b, c\}, \{a, b, e\}\}.$$

Сравнивая результаты с результатами, полученными при помощи алгоритма Априори и его модификаций, заключаем, что данное решение содержит меньшее количество паттернов, соответственно, будет получено меньшее число ассоциативных правил.

Используя полученное решение, сформулируем ассоциативные правила (табл. 10):

Полученные правила являются подмножеством правил, сгенерированных при помощи алгоритма Априори и его модификаций.

Таблица 10

Полученные ассоциативные правила

Купленные товары	Рекомендованные товары
Карандаши	Кнопки
Кнопки	Скрепки
Кнопки	Бумага
Карандаши и кнопки	скрепки
Карандаши и кнопки	Клей

Достоинства алгоритма [11]:

— алгоритм СвО позволяет производить постепенную обработку новых данных путем обновления и расширения полученных результатов без выполнения всех вычислений с нуля;

— выбор алгоритма построения решетки понятий должен основываться на свойствах входных данных. Для больших и плотных контекстов самыми быстрыми алгоритмами являются восходящие алгоритмы (стратегия «снизу вверх» — от минимального объема к максимальному), основанные на каноничности;

— алгоритм СвО основан на критерии каноничности и является линейным по числу входных объектов, поэтому используется для плотных контекстов.

Недостатком алгоритма является то, что он хорошо работает с плотными контекстами, но для небольших и разреженных контекстов, контекстов средней плотности требуются другие алгоритмы (см. работу [10]). Эта особенность алгоритма СвО побудила исследователей создавать его различные модификации [12–20].

Заключение

Существует мнение, что с появлением алгоритма FP-GROWTH проблема эффективного поиска частых паттернов снята, поскольку найден самый эффективный алгоритм решения данной задачи. Однако, как правило, поиск всех частых паттернов бессмысленен, поскольку пользователю нужны только наиболее «информативные» («интересные») зависимости на данных. Кроме того, при поиске паттернов требуется учитывать дополнительные требования, помимо частоты, например: замкнутость паттерна; ограничения на подпаттерны и супертерны; ограничения, связанные с агрегатными функциями, возможность поддержки иерархии на множестве элементов (признаков) и т. п. Необходимость учета дополнительных требований к виду паттерна и его «информативности» обычно приводит к довольно трудоемким модификациям базовых методов поиска частых паттернов, разобранных в настоящей статье. В связи с этим возникает необходимость в разработке новых

подходов к созданию алгоритмов поиска частых паттернов, позволяющих гибко настраиваться на дополнительные ограничения. Таким образом, пока рано ставить точку в исследованиях по разработке эффективных алгоритмов извлечения паттернов.

Список источников

1. Agrawal R., Imielinski T. and Swami A. Mining association rules between sets of items in large databases // Proceedings of the ACM SIGMOD Conf on Management of Data. Washington, DC, 1993. P. 207–216.
2. Agrawal R., Srikant R. Fast algorithms for mining association rules in large databases // Proceedings of the 20th International Conference on Very Large Data Bases. VLDB, Santiago, Chile, 1994. P. 487–499.
3. Кириченко Д. О., Артемов М. А. Оптимизация входных данных в задаче поиска шаблонов и ассоциативных правил // Вестник ВГУ, серия: Системный анализ и информационные технологии, 2014. № 4. С. 63–70.
4. Pol U. Design and Development of Apriori Algorithm for Sequential to concurrent mining using MPI // International journal of Computers & Technology. 2013. Vol. 10. № 7. P. 1785–1790.
5. Han J., Pei J., Yin Y., Mao R. Mining of frequent patterns without candidate generation: a frequent-pattern tree approach // Data mining and analysis discovery. 2004. Vol. 8. № 1. P. 53–87.
6. Borgelt C. An Implementation of the FPgrowth Algorithm [Электронный ресурс] // Workshop Open Source Data Mining Software. New York: ACM Press. 2005. URL: <http://www.osdm.ua.ac.be/papers/p1-borgelt.pdf> (дата обращения: 15.09.2024).
7. Стокипный А. Л. Способ эффективного представления исследуемого набора данных в методах поиска ассоциативных правил // Кибернетика та системний аналіз. Харьков. 2009. № 3. С. 153–161.
8. Zaki M. Scalable Algorithm for association mining // IEEE Transactions on Knowledge and Data Engineering. 2000. № 12. P. 372–390.
9. Borgelt C. Efficient Implementations of Apriori and Eclat [Электронный ресурс] // Workshop on Frequent Itemset Mining Implementations. New York: ACM Press. 2003. URL: www.intsci.ac.cn/shizz/fimi.pdf (дата обращения: 11.09.2024).
10. Kuznetsov S. O. and Obiedkov S. A. Comparing performance of algorithms for generating concept lattices // Journal of Experimental & Theoretical Artificial Intelligence. 2002. Paper 100241. P. 1–28.
11. Kuznetsov S. O. A fast algorithm for computing all intersections of objects in a finite semilattice // Automatic Documentation and Mathematical Linguistics, 1993. 27 (5). P. 11–21.
12. Dowling C. E. On the irredundant generation of knowledge spaces // Math J. Psych. 1993. 37 (1). P. 49–62.
13. Godin R., Missaoui R. and Alaoui H. Incremental concept formation algorithms based on Galois lattices // Computation Intelligence. 1995. 11 (2). P. 246–267.
14. Carpineto C. and Romano G. A lattice conceptual clustering system and its application to browsing retrieval. Machine Learning, 1996. 24. P. 95–122.
15. Lindig C. Algorithmen zur begriffsanalyse und ihre anwendung bei softwarebibliotheken (Dr.- Ing.) Dissertation, Techn. Univ. Braunschweig. 1999.
16. Stumme G., Taouil R., Bastide Y., Pasquier N. and Lakhal L. Fast computation of concept lattices using data mining techniques // Proceedings of the 7th Int. Workshop on Knowledge Representation Meets Databases (KRDB 2000), Berlin, Germany, 2000. P. 129–139.
17. Mephu Nguifo E. and Njiwoua P. Using lattice-based framework as a tool for feature extraction, in Feature Extraction. / H. Liu and H. Motoda (eds) // Construction and Selection: A Data Mining Perspective (Boston, MA: Kluwer). 1998. P. 205–216.
18. Van Der Merwe F.J. and Kourie D.G. AddAtom: an incremental algorithm for constructing concept lattices and concept sublattices / Technical report, Department of Computer Science, University of Pretoria. 2002.

19. Yevtushenko S. BDD-based algorithms for the construction of the set of all concepts. Foundations and applications of conceptual structures // Foundations and Applications of Conceptual Structures. Contributions to ICCS. 2002. P. 61–73.
20. Norris E. M. An algorithm for computing the maximal rectangles in a binary relation. // Revue Roumaine de Mathématiques Pures et Appliquées. 1978. 23 (2). P. 243–250.

References

1. Agrawal R., Imielinski T. and Swami A. Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD Conf on Management of Data*. Washington, DC, 1993, pp. 207–216.
2. Agrawal R., Srikant R. Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases*. VLDB, Santiago, Chile, 1994, pp. 487–499.
3. Kirichenko D. O., Artemov M. A. Optimizacija vhodnyh dannyh v zadache poiska shablonov i asociativnyh pravil [Optimization of input data in the problem of finding patterns and association rules] *Vestnik VGU, serija: Sistemnyj analiz i informacionnye tehnologii* [VSU Bulletin, series: System Analysis and Information Technologies], 2014, no 4, pp. 63–70, (In Russ.).
4. Pol U. Design and Development of Apriori Algorithm for Sequential to concurrent mining using MPI. *International journal of Computers & Technology*. 2013, vol. 10, No 7, pp. 1785–1790.
5. Han J., Pei J., Yin Y., Mao R. Mining of frequent patterns without candidate generation: a frequent-pattern tree approach. *Data mining and analysis discovery*. 2004, vol. 8, no 1, pp. 53–87.
6. Borgelt C. An Implementation of the FPgrowth Algorithm. *Workshop Open Source Data Mining Software*. New York: ACM Press, 2005. Available at: <http://www.osdm.ua.ac.be/papers/p1-borgelt.pdf> (accessed: 15.09.2024).
7. Stokipnyj A. L. Sposob effektivnogo predstavlenija issleduemogo nabora dannyh v metodah poiska asociativnyh pravil [Method of effective presentation of the studied data set in methods of searching for associative rules] *Kibernetika ta sistemnyj analiz, Har'kov* [Cybernetics and system analysis, Kharkov]. 2009, no 3. pp. 153–161, (In Russ.).
8. Zaki M. Scalable Algorithm for association mining. *IEEE Transactions on Knowledge and Data Engineering*. 2000, no 12, pp. 372–390.
9. Borgelt C. Efficient Implementations of Apriori and Eclat. *Workshop on Frequent Itemset Mining Implementations*. New York: ACM Press. 2003, Available at: www.intsci.ac.cn/shizz/fimi.pdf (accessed: 11.09.2024).
10. Kuznetsov S. O. and Obiedkov S. A. Comparing performance of algorithms for generating concept lattices. *Journal of Experimental & Theoretical Artificial Intelligence*. 2002, Paper 100241. pp. 1–28.
11. Kuznetsov S. O. A fast algorithm for computing all intersections of objects in a finite semilattice. *Automatic Documentation and Mathematical Linguistics*. 1993, no 27 (5), pp. 11–21.
12. Dowling C. E. On the irredundant generation of knowledge spaces. *Math J. Psych*. 1993, no 37 (1), pp. 49–62.
13. Godin R., Missaoui R. and Alaoui H. Incremental concept formation algorithms based on Galois lattices. *Computation Intelligence*. 1995, no 11 (2), pp. 246–267.
14. Carpineto C. and Romano G. A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 1996, no 24, pp. 95–122.
15. Lindig C. Algorithmen zur begriffsanalyse und ihre anwendung bei softwarebibliotheken (Dr.- Ing.) Dissertation, Techn. Univ. Braunschweig. 1999.
16. Stumme G., Taouil R., Bastide Y., Pasquier N. and Lakhal L. Fast computation of concept lattices using data mining techniques. *Proceedings of the 7th Int. Workshop on Knowledge Representation Meets Databases (KRDB 2000)*, Berlin, Germany, 2000, pp. 129–139.
17. Mephu Nguifo E. and Njiwoua P. Using lattice-based framework as a tool for feature extraction, in Feature Extraction. H. Liu and H. Motoda (eds). *Construction and Selection: A Data Mining Perspective*, Boston, MA: Kluwer, 1998, pp. 205–216.

18. Van Der Merwe F.J. and Kourie D.G. AddAtom: an incremental algorithm for constructing concept lattices and concept sublattices. Technical report, Department of Computer Science, University of Pretoria. 2002.
19. Yevtushenko S. BDD-based algorithms for the construction of the set of all concepts. foundations and applications of conceptual structures. *Foundations and Applications of Conceptual Structures. Contributions to ICCS, 2002*, pp. 61–73.
20. Norris E. M. An algorithm for computing the maximal rectangles in a binary relation. *Revue Roumaine de Mathématiques Pures et Appliquées*. 1978, no 23 (2), pp. 243–250.

Информация об авторах

А. А. Зуенко — кандидат технических наук, ведущий научный сотрудник;

О. В. Фридман — кандидат технических наук, ведущий инженер.

Information about the authors

A. A. Zuenko — Candidate of Science (Tech.), Leading Research Fellow;

O. V. Fridman — Candidate of Science (Tech.), Leading Engineer.

Статья поступила в редакцию 15.09.2024; одобрена после рецензирования 01.10.2024; принята к публикации 08.11.2024.
The article was submitted 15.09.2024; approved after reviewing 01.11.2024; accepted for publication 08.11.24

Научная статья
УДК:004.457
doi:10.37614/2949.1215.2024.15.3.008

ОРГАНИЗАЦИЯ УПРАВЛЕНИЯ ПРОГРАММНЫМИ КОМПОНЕНТАМИ В СИСТЕМЕ КОНЦЕПТУАЛЬНОГО МОДЕЛИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ PYTHON И PYQT5

Никита Николаевич Руденко¹, Никита Александрович Вдовиченко²

^{1, 2}*Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия*

¹*nikirudi1988@mail.ru, <https://orcid.org/0000-0003-0061-0397>*

²*claysthree@yandex.ru, <https://orcid.org/0000-0002-8150-0773>*

Аннотация

Данная разработка системы концептуального моделирования (СКМ) на основе PyQt5 является актуальной для визуализации и проектирования сложных систем, что важно в таких областях, как инженерия, программирование. Визуальные модели позволяют пользователям эффективно взаимодействовать с объектами, упрощая процесс построения и анализа сложных взаимосвязей. Такая СКМ полезна в образовательных целях, а также для прототипирования и демонстрации концепций наглядным образом. Возможность создавать, связывать и изменять объекты интерактивно повышает гибкость и удобство работы с концептуальными моделями.

Ключевые слова:

СКМ, визуальные модели, проектирование

Для цитирования:

Руденко Н. Н., Вдовиченко Н. А. Организация управления программными компонентами в системе концептуального моделирования с использованием Python и PyQt5 // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, No 5. С. 97–104. doi:10.37614/2949.1215.2024.15.3.008.

Original article

ORGANIZING SOFTWARE COMPONENTS MANAGEMENT IN A CONCEPTUAL MODELING SYSTEM USING PYTHON AND PYQT5

Nikita N. Rudenko¹, Nikita A. Vdovichenko²

^{1, 2}*V. A. Putilov Institute of Informatics and Mathematical Modeling
Kola Scientific Center of the Russian Academy of Sciences, Apatity, Russia*

¹*nikirudi1988@mail.ru, <https://orcid.org/0000-0003-0061-0397>*

²*claysthree@yandex.ru, <https://orcid.org/0000-0002-8150-0773>*

Abstract

This development of a conceptual modeling system based on PyQt5 is relevant for visualization and design of complex systems, which is important in such areas as engineering, programming. Visual models allow users to interact effectively with objects, simplifying the process of constructing and analyzing complex relationships. Such SCM is useful for educational purposes, as well as for prototyping and demonstrating concepts in a visual way. The ability to create, link, and modify objects interactively increases the flexibility and convenience of working with conceptual models.

Keywords:

SCM, visual models, prototyping

For citation:

Rudenko N. N., Vdovichenko N. A. Organization of software component management in the conceptual modeling system using Python and PyQt5 // Proceedings of the Kola Science Center of the Russian Academy of Sciences. Series: Technical Sciences. 2024. Vol. 15, No. 3. Pp. 97–104. doi: 10.37614/2949-1215.2024.15.3.008.

Введение

Система концептуального моделирования — это программа или система, которая позволяет моделировать связи между объектами или процессами в определенной предметной области. Она предназначена для формализации сложных систем и процессов в виде декларативных концептуальных моделей [1], которые помогают понять, спроектировать или проанализировать работу этих систем. Чтобы обеспечить доступ к базе данных модели в системе концептуального моделирования

с использованием Python и PyQt5, необходимо разработать архитектуру, которая связывает пользовательский интерфейс [2] с базой данных, где хранятся данные о моделях [3]. Эта архитектура должна включать несколько компонентов: интерфейс, слой доступа к данным, логику обработки данных и саму базу данных.

Перечень инструментов, предоставляемых этой системой моделирования:

- создание и редактирование концептуальных моделей. Пользователь должен иметь возможность создавать новые модели и редактировать уже существующие;

- просмотр и визуализация моделей. Интерфейс должен отображать модели в понятной форме, включая визуальные представления связей и объектов;

- обработка данных моделей. Система должна выполнять операции с данными, такие как валидация, обработка зависимостей, выполнение сценариев и анализ;

- хранение и управление данными. Модели и их данные должны храниться в базе данных для обеспечения надежного доступа, редактирования и сохранности;

- интерактивный доступ к базе данных. Пользователю требуется интерфейс для удобного взаимодействия с базой данных для создания, редактирования и поиска объектов;

В СКМ представлены следующие основные визуальные компоненты:

- формы ввода (QLineEdit, QTextEdit, QComboBox): используются для ввода информации о моделях (название модели, описание, атрибуты и т.д.);

- таблицы и списки (QTableWidget, QListView): отображение списка доступных моделей и их свойств в табличной форме;

- диаграммы и графы (QGraphicsView, QGraphicsScene): для визуализации и редактирования графических моделей, их связей и взаимодействий. Это особенно важно для отображения концептуальных моделей в виде графов;

- кнопки и панели инструментов (QPushButton, QToolBar): используются для выполнения действий: создание новой модели, редактирование, удаление, сохранение и т. д.;

- диалоговые окна (QDialog, QFileDialog): для выполнения специфических задач, таких как сохранение файла, выбор модели для загрузки, подтверждение действия и т. д.;

- графические редакторы (QCanvas, QPainter): для создания и редактирования графических объектов, таких как элементы модели (объекты, связи между ними).

К основным невидимым компонентам системы относятся следующие:

- слой доступа к данным (Data Access Layer, DAL): отвечает за взаимодействие с базой данных, управление запросами на получение, изменение, удаление и добавление данных. DAL изолирует приложение от конкретной реализации базы данных (например, SQL или NoSQL) и предоставляет интерфейс для работы с данными;

- логика управления моделями (Business Logic Layer, BLL): компонент управляет логикой взаимодействия с моделями и их атрибутами. Он обрабатывает бизнес-правила, осуществляет валидацию данных, контролирует корректность связей между моделями и выполняет вычисления, связанные с моделями;

- контроллер взаимодействия с пользователем (Controller/Presenter): управляет связью между пользовательским интерфейсом (визуальными компонентами) и бизнес-логикой системы. Контроллер принимает действия пользователя, передает данные в слой бизнес-логики, а затем обновляет интерфейс на основе полученных данных;

- модуль взаимодействия с внешними системами (API, интеграция): если система концептуального моделирования должна взаимодействовать с внешними системами или обмениваться данными через API (например, для интеграции с другими инструментами моделирования), этот компонент будет отвечать за такие операции;

- модуль валидации данных (Data Validation): компонент проверяет правильность данных, вводимых пользователем или получаемых из внешних источников, перед их сохранением в базу данных.

В ходе разработки системы концептуального моделирования было принято решение использовать язык программирования Python и библиотеку PyQt5 для реализации как визуальных, так и невидимых компонентов. Это решение было обосновано следующими причинами:

- простота и гибкость Python:
 - является высокоуровневым языком программирования, известным своей простотой и читаемостью кода. Это позволило ускорить процесс разработки, сосредоточив внимание на реализации логики системы, а не на сложных технических деталях;
 - широкий выбор библиотек и фреймворков для различных задач (работа с базами данных, API, валидация данных и т. д.) сделал Python идеальным инструментом для создания гибкой и расширяемой архитектуры системы;
- PyQt5 как мощный инструмент для создания графических интерфейсов:
 - предоставляет богатый набор готовых визуальных компонентов (кнопки, формы, таблицы, графические виджеты и др.), которые позволяют быстро создавать функциональные и удобные пользовательские интерфейсы;
 - поддерживает возможность создания сложных графических интерфейсов с использованием визуальных элементов для работы с графами, диаграммами и другими компонентами, необходимыми для визуализации концептуальных моделей;
 - кроссплатформенность PyQt5 дает возможность разработанной системе работать на различных операционных системах (Windows, macOS, Linux) без необходимости внесения значительных изменений в код;
 - интеграция с базами данных и внешними сервисами;
 - Python предоставляет простой и удобный способ взаимодействия с различными базами данных (SQL, NoSQL) через библиотеки, такие как SQLite, SQLAlchemy и др., что сделало возможным быстрое и удобное внедрение слоя доступа к данным (DAL);
 - PyQt5 также хорошо сочетается с библиотеками для работы с сетевыми запросами и API (например, requests), что упростило интеграцию системы с внешними сервисами и модулями;
 - быстрое прототипирование и тестирование;
 - благодаря своей интерпретируемой природе Python позволяет быстро прототипировать и тестировать различные решения, что сделало его идеальным выбором для гибкой разработки и итерационного процесса создания системы.

Связь СКМ и ГИС

Связь СКМ с геоинформационными системами (ГИС) [4] открывает новые возможности для анализа пространственных данных и их визуализации. В рамках такой интеграции СКМ может использоваться для создания моделей объектов, которые имеют пространственную привязку, например, инфраструктурных элементов, транспортных сетей, зданий и природных объектов. В ГИС данные этих объектов могут быть дополнены их географическим положением, координатами и пространственными характеристиками. Таким образом, концептуальные модели могут быть отображены на карте и связаны с реальными геоданными, что позволит визуализировать их влияние в контексте реальной местности.

Это объединение может использоваться, например, для моделирования логистических или транспортных систем, где важно учитывать географическое расположение элементов. СКМ позволит легко моделировать отношения между объектами (дороги, станции, маршруты), а ГИС предоставит платформу для анализа пространственных аспектов, таких как расстояния, топология или доступность ресурсов.

Использование сокетов (Sockets) для взаимодействия между распределенными компонентами

Python поддерживает взаимодействие с использованием сокетов через стандартную библиотеку socket, которая предоставляет все необходимые инструменты для реализации сетевого обмена

данными с использованием протоколов TCP/IP и UDP. Этот подход часто используется для организации взаимодействия между компонентами, расположенными на разных серверах или устройствах.

Что такое сокеты: сокет — это конечная точка в сетевом соединении, через которую компоненты системы могут отправлять и получать данные. Компоненты могут обмениваться сообщениями или передавать данные через сеть, используя сокеты.

Пример использования сокетов в Python: один из компонентов может выступать в роли сервера, который слушает определенный порт и обрабатывает запросы, поступающие от других компонентов (клиентов). Клиенты подключаются к серверу по указанному адресу и обмениваются данными через сокеты.

Основной идеей разработки системы является использование открытой архитектуры, что позволяет расширять функциональность комплекса, добавляя новые визуальные и невидимые компоненты, а также компоненты, взаимодействующие с внешними приложениями через стандартные протоколы.

- Menu Supervisor: в системе предусмотрен центральный компонент Menu Supervisor, который управляет компонентами системы и координирует их взаимодействие. Этот компонент отвечает за:
 - мониторинг изменений в данных, с которыми работают компоненты (в том числе внешние приложения);
 - вызов и управление опциональными компонентами, которые могут быть активированы при изменении данных;
 - управление потоком данных между компонентами, обеспечивая динамическое обновление визуальных элементов при изменении данных в системе.
- ExecAgents: для обеспечения запуска компонентов на удаленных узлах используются агенты выполнения (ExecAgents). Каждый агент отвечает за запуск тех или иных компонентов на локальном устройстве, с учетом переданных параметров таких, как:
 - IP-адрес устройства, на котором должен быть запущен компонент;
 - путь к компоненту в файловой системе.

ExecAgents запускаются на удаленных компьютерах и обеспечивают вызов необходимых компонентов через сети. Коммуникация между агентами и Menu Supervisor происходит через TCP/IP протоколы с использованием стандартных механизмов обмена данными (например, через sockets или Ruro4 для удаленных вызовов процедур).

Компонент DataAccess отвечает за доступ к данным и информирование Menu Supervisor о любых изменениях в данных, с которыми работают компоненты. DataAccess периодически передает обновления об изменении данных через сообщения, что позволяет системе автоматически обновлять те компоненты, которые зависят от этих данных.

Компоненты (визуальные и опциональные): визуальные компоненты (разработанные на PyQt5) представляют собой интерфейсы, которые отображают данные в системе. Эти компоненты могут включать в себя различные типы данных, такие как геоинформационные данные или результаты имитационного моделирования. Опциональные компоненты могут быть запущены по запросу в зависимости от типов данных, с которыми работает система, или изменений в этих данных. Эти компоненты выступают в роли индикаторов изменений и могут быть динамически подключены или отключены в зависимости от конфигурации проекта.

Процесс добавления компонентов

На начальной стадии работы с проектом оператор указывает, какие компоненты должны быть активированы в системе. Это включает:

- определение списка визуальных компонентов, которые будут отображать данные пользователю;
- задание типов данных, с которыми эти компоненты работают (например, геоинформационные данные, данные симуляций и т. д.);
- указание идентификаторов для каждого компонента и его соответствие типам данных.

Взаимодействие компонентов

- Menu Supervisor обеспечивает информационно-управляющие потоки данных между компонентами через:
- TCP/IP протоколы, что позволяет компонентам взаимодействовать даже при их размещении на удаленных машинах;
 - Sockets или Pyro4, которые могут использоваться для организации обмена данными между распределенными компонентами системы.

Каждый визуальный компонент или компонент данных уведомляется о любых изменениях в данных через Menu Supervisor, который получает информацию об изменениях от компонента DataAccess. Это позволяет поддерживать актуальность данных во всех частях системы и автоматически обновлять интерфейсы.

Гибкость и динамичность системы

Одной из ключевых особенностей системы является возможность опционального добавления или исключения компонентов. На начальной стадии работы с проектом оператор может определить, какие компоненты необходимы, а какие можно исключить. Это позволяет легко адаптировать систему под текущие задачи, без необходимости вносить серьезные изменения в архитектуру программного комплекса.

В отличие от технологий DCOM и OLE, подход, реализованный на Python с использованием TCP/IP, не требует дополнительной настройки клиентов и является кроссплатформенным. Это позволяет компонентам взаимодействовать через Интернет или локальные сети с высокой степенью надежности передачи данных.

Пример использования сокетов для взаимодействия компонентов

Для организации информационно-управляющих потоков данных между компонентами можно использовать Python-библиотеку socket. Например:

Листинг 1

Сервер для Menu Supervisor

```
import socket

def start_server():
    server_socket = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    server_socket.bind(('localhost', 8080))
    server_socket.listen(5)
    print("Menu Supervisor запущен и ожидает соединений...")

    while True:
        client_socket, addr = server_socket.accept()
        print(f"Соединение установлено с {addr}")
        data = client_socket.recv(1024).decode()
        print(f"Получены данные: {data}")
        client_socket.send("Данные получены".encode())
        client_socket.close()

if __name__ == "__main__":
    start_server()
```

Клиент для ExecAgent

```
import socket

def send_data():
    client_socket = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
    client_socket.connect(('localhost', 8080))
    client_socket.send("Запуск компонента".encode())
    response = client_socket.recv(1024).decode()
    print(f"Ответ от Menu Supervisor: {response}")
    client_socket.close()

if __name__ == "__main__":
    send_data()
```

Технология OLE Automation и аналогичные подходы в Python

Рассмотрение OLE Automation параллельно альтернативами на языке Python важно для автоматизации взаимодействия с разными операционными системами. OLE Automation позволяет управлять этими программами из других приложений, что полезно для задач, связанных с интеграцией и обменом данными между системами. Однако Python предлагает другие подходы, такие как REST API, и других сетевых технологий для межпроцессного взаимодействия, что может быть более гибким и кроссплатформенным решением. Таким образом, рассмотрение альтернатив помогает выбрать наиболее подходящий метод для конкретной задачи.

В экосистеме Python нет точного аналога OLE Automation из Windows, однако схожие функции можно реализовать с помощью технологий удаленного вызова процедур (Remote Procedure Call, RPC) или через использование библиотек для межпроцессного взаимодействия (Inter-Process Communication, IPC). В Python есть несколько способов для обеспечения такого взаимодействия между компонентами.

Использование библиотеки xmlrpc для RPC

Технология XML-RPC позволяет компонентам взаимодействовать друг с другом по сети, вызывая удаленные методы. Один из компонентов системы может быть сервером, который предоставляет определенные функции (методы), а другой компонент может выступать в роли клиента, который вызывает эти методы удаленно. Пример простого XML-RPC-сервера и клиента:

XML-RPC сервер:

```
from xmlrpc.server import SimpleXMLRPCServer

def add_numbers(x, y):
    return x + y

server = SimpleXMLRPCServer(('localhost', 8000))
print("Сервер запущен, ожидает вызовов...")
server.register_function(add_numbers, 'add')
server.serve_forever()
```

XML-RPC клиент:

```
import xmlrpc.client

client = xmlrpc.client.ServerProxy('http://localhost:8000')
result = client.add(5, 3)
print(f"Результат вызова удаленной функции: {result}")
```

В этом примере клиент вызывает удаленную функцию `add` на сервере, передавая параметры, а сервер выполняет сложение и возвращает результат клиенту (Листинг 3,4).

Модуль Pyro4 для распределенных компонентов

Pyro4 — это библиотека, которая реализует механизм удаленного вызова процедур, позволяя компонентам системы взаимодействовать друг с другом по сети. Pyro4 позволяет определить один компонент как сервер (который предоставляет удаленные методы), а другой как клиент (который вызывает эти методы). Пример Pyro4:

Листинг 5

Pyro4 сервер:

```
import Pyro4

@Pyro4.expose
class ModelManager:
    def get_model_data(self, model_id):
        return f"Данные модели {model_id}"

def start_server():
    Pyro4.Daemon.serveSimple(
        {
            ModelManager: "model.manager"
        },
        ns = False)

if __name__ == "__main__":
    start_server()
```

Pyro4 позволяет создать распределенную систему, где один компонент (сервер) предоставляет данные или методы для управления моделями, а другой компонент (клиент) обращается к этим методам удаленно (Листинг 5, 6).

Листинг 6

Pyro4 клиент:

```
import Pyro4

def fetch_model_data(model_id):
    model_manager = Pyro4.Proxy("PYRO:model.manager@localhost:9090")
    return model_manager.get_model_data(model_id)

if __name__ == "__main__":
    data = fetch_model_data(1)
    print(f"Полученные данные: {data}")
```

Заключение

В результате проведенного анализа была предложена модульная архитектура программного комплекса для системы концептуального моделирования, основанная на использовании языка программирования Python и библиотеки PyQt5. Эта архитектура предоставляет несколько возможностей реализации, каждая из которых может быть адаптирована в зависимости от специфики задач и требований пользователей.

Данная система имеет потенциал для дальнейшего развития для автоматизации анализа данных и расширение функциональности за счет подключения дополнительных модулей и API.

Такие направления могут значительно повысить ценность программного комплекса и его конкурентоспособность. В дальнейшем планируется внедрить в СКМ методы ГИС, которые позволят расширить функциональность системы концептуального моделирования, добавив анализ пространственных данных и интеграцию с реальными картографическими объектами. Это даст возможность моделировать процессы с учетом географических факторов, улучшая точность планирования и анализа в таких сферах, как логистика, городское планирование и управление ресурсами.

Список источников

1. Павлецов А., Фридман А. Организация управления программными компонентами в системе концептуального моделирования // Информационные технологии в региональном развитии. Апатиты, 2003. Вып. III. С. 63–68.
2. Матченко И., Павлецов А., Фридман А., Фридман О. Основы пользовательских и внутренних интерфейсов системы ситуационного моделирования // Математические методы описания и исследования сложных систем. Апатиты: КНЦ РАН, 2001. С. 119–132.
3. Матченко И., Фридман А. Компонент доступа к данным для пакета ситуационного концептуального моделирования // Информационные технологии в региональном развитии. Апатиты, 2003. Вып. III. С. 86–93.
4. Фридман А., Курбанов В. Формальная концептуальная модель промышленно-природного комплекса как инструмент управления вычислительными экспериментами // Труды СПИИРАН. 2014. № 6(37). С. 424–453.

References

1. Pavletsov A., Fridman A. Organization of software component management in a conceptual modeling system // Information Technologies in Regional Development. Apatity, 2003. Issue III. P. 63–68.
2. Matchenko I., Pavletsov A., Fridman A., Fridman O. Fundamentals of user and internal interfaces of the situational modeling system // Mathematical Methods for Describing and Studying Complex Systems. Apatity: KSC RAS, 2001. P. 119–132.
3. Matchenko I., Fridman A. Data access component for the situational conceptual modeling package // Information Technologies in Regional Development. Apatity, 2003. Issue III. P. 86–93.
4. Fridman A., Kurbanov V. Formal conceptual model of an industrial-natural complex as a tool for managing computational experiments // Proceedings of SPIIRAS. 2014. No. 6(37). P. 424–453.

Информация об авторах

Н. Н. Руденко — студент 1 курса аспирантуры;

Н. А. Вдовиченко — студент 1 курса аспирантуры.

Information about the authors

N. N. Rudenko — 1st year student of the Postgraduate School of the Kola Science Center of the Russian Academy of Sciences;

N. A. Vdovichenko — 1st year student of the Postgraduate School of the Kola Science Center of the Russian Academy of Sciences.

Статья поступила в редакцию 01.10.2024; одобрена после рецензирования 10.10.2024; принята к публикации 21.10.2024.

The article was submitted 01.10.2024; approved after reviewing 10.10.2024; accepted for publication 21.10.2024.

Научная статья
УДК:004.4
doi:10.37614/2949-1215.2024.15.3.009

РАЗРАБОТКА ГИС-ИНТЕРФЕЙСА ДЛЯ СИТУАЦИОННОЙ СИСТЕМЫ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА НА ОСНОВЕ C# И DOTSPATIAL

Никита Александрович Вдовиченко¹, Никита Николаевич Руденко²

^{1, 2}*Институт информатики и математического моделирования имени В. А. Путилова
Кольского научного центра Российской академии наук, Апатиты, Россия*

¹*claysthree@yandex.ru, <https://orcid.org/0000-0002-8150-0773>*

²*nikirudi1988@mail.ru, <https://orcid.org/0000-0003-0061-0397>*

Аннотация

В данной статье рассматривается разработка ГИС-интерфейса для ситуационной системы вычислительного эксперимента с использованием языка программирования C# и библиотеки DotSpatial. Основной целью исследования является создание системы, которая позволяет проводить анализ пространственных данных и визуализацию объектов и процессов с возможностью их категоризации. Описаны концептуальная модель предметной области (КМПО), алгоритмы категоризации объектов и процессов, а также методы визуализации данных с использованием ГИС. Результаты разработки могут быть применены в вычислительных задачах, включая моделирование и анализ ситуаций в различных предметных областях.

Ключевые слова

ГИС-интерфейс, C#, DotSpatial, ситуационная система, вычислительный эксперимент, пространственные данные

Для цитирования:

Вдовиченко Н. А., Руденко Н. Н. Разработка ГИС-интерфейса для ситуационной системы вычислительного эксперимента на основе C# и DotSpatial // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, No 3. С. 105–112. doi:10.37614/2949.1215.2024.15.3.009.

Original article

DEVELOPMENT OF GIS-INTERFACE FOR SITUATIONAL SYSTEM OF COMPUTATIONAL EXPERIMENT ON THE BASIS OF C# AND DOTSPATIAL

Nikita A. Vdovichenko¹, Nikita N. Rudenko²

^{1, 2}*V. A. Putilov Institute of Informatics and Mathematical Modeling
Kola Scientific Center of the Russian Academy of Sciences, Apatity, Russia*

¹*claysthree@yandex.ru, <https://orcid.org/0000-0002-8150-0773>*

²*nikirudi1988@mail.ru, <https://orcid.org/0000-0003-0061-0397>*

Abstract

This paper deals with the development of GIS-interface for situational system of computational experiment using C# programming language and DotSpatial library. The main purpose of the study is to create a system that allows for spatial data analysis and visualization of objects and processes with the possibility of their categorization. The conceptual model of the subject area (CMPO), algorithms for categorization of objects and processes, and methods of data visualization using GIS are described. The results of the development can be applied in computational tasks, including modeling and analysis of situations in various subject areas.

Keywords

GIS-interface, C#, DotSpatial, situational system, computational experiment, spatial data

For citation:

Vdovichenko N. A., Rudenko N. N. Development of a GIS interface for a situational system of a computational experiment based on C# and DotSpatial // Proceedings of the Kola Science Center of the Russian Academy of Sciences. Series: Technical Sciences. 2024. Vol. 15, No. 3. Pp. 105–112. doi:10.37614/2949.1215.2024.15.3.009.

Введение

В настоящее время системы геоинформационного анализа (ГИС) играют важную роль в решении задач, связанных с обработкой и визуализацией пространственных данных. Их применение необходимо в различных областях: от экологии и геологии до урбанистики и логистики. Разработка эффективных ГИС-интерфейсов, способных автоматизировать обработку данных и предоставить наглядные визуализации, является важной задачей в сфере научных и прикладных исследований.

Целью настоящей работы является разработка ГИС-интерфейса для ситуационной системы вычислительного эксперимента. Система, реализованная с использованием C# и библиотеки DotSpatial, предоставляет возможности для анализа и категоризации пространственных данных, а также их визуализации в контексте вычислительных моделей.

Для достижения поставленной цели были сформулированы следующие задачи:

- 1) разработка концептуальной модели предметной области (КМПО) для ситуационной системы вычислительного эксперимента;
 - 2) определение категорий данных и объектов, а также их автоматизация для анализа и визуализации;
 - 3) реализация алгоритмов категоризации объектов и процессов;
 - 4) формирование представления объектов в ГИС и визуализация пространственных данных;
- проведение тестирования разработанного интерфейса на примерах вычислительных задач.

Новизна работы заключается в разработке интегрированного ГИС-интерфейса, построенного на базе современных программных технологий (C# и DotSpatial), позволяющего автоматизировать процесс обработки и категоризации пространственных данных, а также визуализировать результаты вычислительных экспериментов.

Результаты разработки могут быть использованы в различных областях, где требуется пространственный анализ данных и их визуализация. Система может быть адаптирована для решения специфических задач в таких областях, как экология, логистика, геология, планирование инфраструктуры, моделирование чрезвычайных ситуаций и т.д. В дальнейшем планируется интеграция системы с системой принятия решений (СПР), что будет особенно полезно в ситуациях, связанных, например, с наводнениями. Это позволит более эффективно анализировать данные, прогнозировать последствия и предлагать оптимальные меры по снижению рисков и ущерба.

Описание схемы КМПО в рамках ситуационной системы вычислительного эксперимента (ССВЭ)

Концептуальная модель предметной области (КМПО) представляет собой формализованное описание объектов и процессов, участвующих в ситуационном анализе [1]. В рамках ССВЭ КМПО выполняет роль связующего звена между моделируемыми процессами и системой ГИС, обеспечивая правильную интерпретацию пространственных данных и их визуализацию. Модель описывает объекты, их атрибуты, ресурсы, а также отношения между ними, что позволяет детализировать анализ и управление процессами [2].

Основными компонентами КМПО при работе с ГИС являются:

- объекты: элементы предметной области, такие как точки, линии, полигоны, которые представляют реальный мир;
- отношения: связи между объектами, отражающие их взаимодействие в пространстве и времени;
- атрибуты: описательные параметры объектов (например, название, координаты, тип);
- ресурсы: дополнительные данные, необходимые для моделирования процессов (например, временные и числовые характеристики).

Для оптимизации обработки данных в рамках КМПО были введены категории, которые назначаются объектам и процессам автоматически. Категоризация упрощает процесс анализа, позволяя группировать объекты по их типу, важности или другим критериям. Например, объекты могут быть разделены на природные и антропогенные, а процессы — на статические и динамические.

В системе выделяются два типа данных:

- параметры — это строковые значения, которые характеризуют объекты, но не изменяются в ходе эксперимента (например, названия мест, описания);
- переменные — числовые значения, которые могут изменяться в зависимости от условий и временных интервалов (например, температура, уровень воды).

Использование категорий для данных и объектов необходимо для упрощения структурирования информации и ее последующего анализа. Категоризация позволяет более эффективно визуализировать данные на карте, отображая их по определенным признакам, таким как цвет, форма и размер. Например, объекты с высоким уровнем риска можно выделить цветом, а динамические изменения — с помощью анимации или стрелок.

Алгоритм категоризации объектов и процессов

Для эффективной работы ситуационной системы вычислительного эксперимента разработан алгоритм автоматического присвоения категорий объектам КМПО. Алгоритм предполагает следующие шаги:

— инициализация объекта: при добавлении нового объекта в систему ему присваиваются уникальные идентификаторы и параметры. Каждый объект уникален и идентифицируется с помощью ключевых параметров, которые могут включать географические координаты, физические характеристики и временные данные;

— определение атрибутов: каждому объекту назначаются атрибуты на основе исходных данных. Эти атрибуты могут быть различными, включая географические (координаты), физические (размеры, материалы), временные (дата создания или изменения), а также специализированные для конкретной предметной области (например, тип использования территории или экологические параметры);

— присвоение категории: в зависимости от атрибутов объекта и его местоположения в системе он автоматически помещается в одну из заранее определенных категорий. Примеры категорий могут включать «здания», «водоемы», «дороги», «растительность», «промышленные зоны» и др. Категории могут быть ограничены спецификой проекта или предметной области, однако можно создать дополнительные категории для более детализированного анализа. Например, в рамках экологического мониторинга могут быть добавлены категории «заповедные зоны» или «участки леса». Для анализа в ГИС используются команды пространственной выборки (например, по атрибутам или координатам), буферного анализа (по радиусу от объекта), а также анализ наложений слоев для оценки взаимосвязи между объектами;

— контроль корректности: на последнем этапе система проверяет соответствие объекта своей категории. Проверка организована через анализ введенных атрибутов и их сравнение с предустановленными критериями для каждой категории. Например, если объект находится в пределах городской территории и его размеры соответствуют параметрам зданий, система подтвердит его принадлежность к категории «здания». В случае несоответствия (например, если объект с географическими координатами водоема был ошибочно отнесен к категории «дороги»), система предложит корректировку.

На основе описанного алгоритма был реализован прототип системы управления объектами в рамках ГИС-интерфейса с использованием языка программирования C# и библиотеки DotSpatial. Каждый этап алгоритма — инициализация объектов, присвоение атрибутов, автоматическая категоризация и проверка корректности — был воплощен в коде и представлен ниже.

Листинг 1

```
// Создаем новый объект – точку с географическими координатами
Feature newObject = new Feature(FeatureType.Point);
newObject.Coordinates.Add(new Coordinate(30.3158, 59.9386)); // Координаты объекта
(например, Санкт-Петербург)

// Добавляем атрибуты объекту
newObject.DataRow["Name"] = "Здание 1";
newObject.DataRow["Height"] = 50; // физический атрибут
newObject.DataRow["DateCreated"] = DateTime.Now; // временной атрибут

// Присвоение категории на основе атрибутов
string category;
if ((double)newObject.DataRow["Height"] > 40)
    category = "Здания";
else
    category = "Малые постройки";

// Сохранение категории как атрибута объекта
newObject.DataRow["Category"] = category;
```

Для каждой категории процессов в системе предусмотрены свои методы реализации и визуализации. Внутренние процессы (происходящие внутри объекта) отображаются с использованием простых графических изменений объекта (например, изменение цвета или прозрачности), внутриуровневые процессы (происходят между объектами одного уровня иерархии) могут быть показаны с помощью стрелок или линий, соединяющих объекты, а межуровневые процессы (взаимодействие объектов, находящихся на разных уровнях иерархии) требуют более сложной визуализации, включающей динамические карты и временные изменения [3].

Одним из ключевых аспектов разработки ГИС-интерфейса является работа с пространственными данными, хранящимися в dBase-файлах. Эти файлы используются для хранения атрибутов объектов таких, как названия, категории, координаты и другие данные.

Для работы с пространственными данными в системе используются функции библиотеки DotSpatial. Она предоставляет инструменты для загрузки, отображения и анализа данных, что позволяет визуализировать объекты на карте и взаимодействовать с ними.

Процедуры визуализации и выборки объектов в ГИС

Визуализация объектов в ГИС требует не только их отображения, но и предоставления возможностей для выборки по различным критериям [4]. В данной системе для выполнения таких запросов применяются специальные функции на основе инструментария, аналогичного Avenue, который был адаптирован для работы с библиотекой DotSpatial.

Основные методы выбора объектов включают:

— запрос по атрибутам: выбор объектов на карте на основе их характеристик (например, можно выбрать все объекты категории «здания» или все объекты с определенным значением параметра «высота»);

— запрос по пространственным параметрам: выбор объектов в зависимости от их положения на карте (например, можно выделить объекты, находящиеся в пределах определенной области или на определенном расстоянии от точки);

— запрос по времени: выбор объектов на основе временных данных, например, для анализа изменения состояния объектов в зависимости от временного промежутка;

— запросы обеспечивают гибкость визуализации объектов и позволяют детализировать анализ пространственных данных.

Для упрощения анализа данных в ГИС-системе реализованы функции создания графических справочников и выборок объектов. Эти справочники включают в себя визуальные метки, соответствующие различным категориям объектов, с использованием цвета, формы и размера для выделения ключевых характеристик.

Одним из ключевых элементов ГИС-интерфейса является картографическое окно (рис. 1), которое предназначено для отображения пространственных данных. Оно поддерживает динамическое масштабирование и панорамирование, что позволяет пользователям изменять масштаб карты и перемещаться по ней для детального изучения интересующих областей. В системе предусмотрена работа с несколькими слоями (рис. 2), что позволяет одновременно визуализировать различные категории объектов такие, как здания, дороги или водные объекты, на отдельных слоях. Для межуровневых процессов предусмотрена возможность анимации, что дает пользователю возможность отслеживать изменения объектов во времени. Для удобства работы с системой создано меню, которое позволяет выбирать, какие объекты и категории данных отображать на карте. Пользователь может фильтровать объекты по категориям, выбирать их отображение за определенные временные промежутки и настраивать графические параметры такие, как цвет и толщина линий, для лучшей визуализации и анализа [5].

Для разработки и описания вычислительных экспериментов с использованием C# и DotSpatial можно рассмотреть пример эксперимента, которые помогут продемонстрировать работу системы.

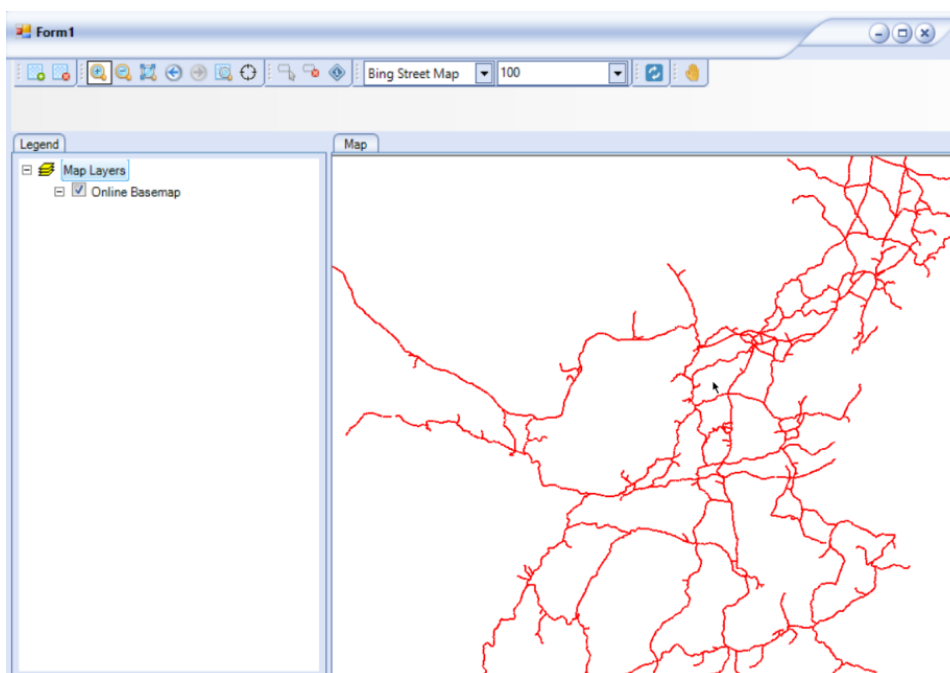


Рис. 1. Интерфейс программы

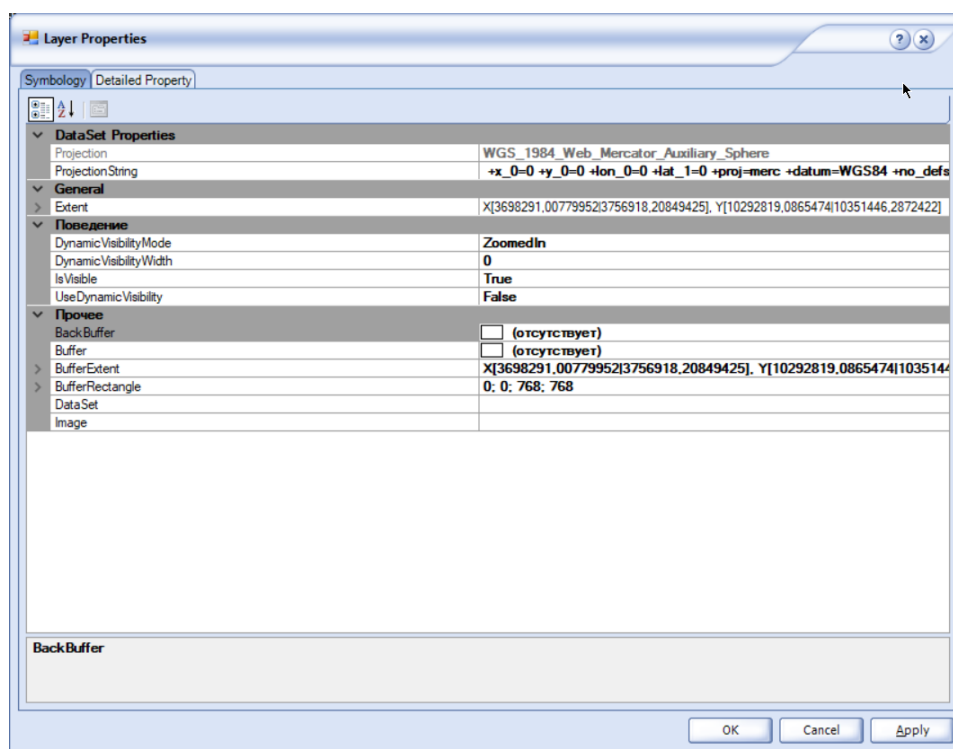


Рис. 2. Настройки слоев

В данном примере осуществляется моделирование наводнений в определенной местности с использованием данных о рельефе и гидрологических параметрах. Целью данного эксперимента является предсказание затопленных областей и анализ их воздействия на ключевые объекты

инфраструктуры такие, как дороги, здания и электросети [6]. Моделирование основывается на динамике изменения уровня воды в реках и озерах, что позволяет учитывать различные сценарии наводнения. Программный код для реализации данного компонента представлен в листинге 2.

Листинг 2

```
// Загрузка данных о рельефе и объектах
string demPath = @"C:\path\to\your\dem.tif";
string infrastructurePath = @"C:\path\to\your\infrastructure.shp";
IMapRasterLayer demLayer = map.Layers.Add(demPath) as IMapRasterLayer;
IMapLayer infrastructureLayer = map.Layers.Add(infrastructurePath);

// Моделирование наводнения (например, увеличение уровня воды на 5 метров)
double floodLevel = 5.0;

// Выявление затопленных объектов
foreach (IFeature feature in infrastructureLayer.DataSet.Features)
{
    double elevation = GetElevation(feature, demLayer);
    if (elevation < floodLevel)
    {
        feature.DataRow["Flooded"] = true;
        // Визуализация затопленных объектов
        feature.Symbolizer = new PointSymbolizer(Color.Blue, PointShape.Circle, 10);
    }
}

// Анализ инфраструктуры и предложения
AnalyzeFloodImpact(infrastructureLayer);
```

Шаги эксперимента:

- 1) загрузка данных о рельефе и объектах инфраструктуры;
- 2) моделирование наводнения с учетом повышения уровня воды;
- 3) выявление затопленных объектов на основе анализа высоты;
- 4) визуализация результатов на карте;
- 5) оценка ущерба для инфраструктуры и предложения по изменению маршрутов или защиты объектов.

На этапе подготовки данных для моделирования наводнений был осуществлен процесс геокодирования, позволяющий преобразовать текстовую информацию об объектах инфраструктуры (например, названия или адреса) в точные географические координаты. Геокодирование представляет собой метод преобразования текстовых данных в координаты, которые обозначают местоположение объектов на поверхности Земли [7, 8]. Эти координаты, выраженные в широте (ϕ) и долготе (λ), определяют положение объектов на сферической или эллипсоидальной модели Земли. Геокодирование можно представить в виде следующей формулы:

$$P = \text{GeoCode}(A),$$

где P — это пара координат (широта, долгота); A — текстовая информация об адресе.

Для моделирования наводнений используется уравнение уровня воды, которое учитывает приток, отток и изменение объема воды в заданной области. Уравнение можно записать следующим образом:

$$\frac{dV}{dt} = Q_{in} - Q_{out},$$

где V — объем воды в резервуаре; Q_{in} — приток воды (например, от дождя или таяния снега); Q_{out} — отток (например, через реку или водоотвод).

После получения координат объектов и выполнения моделирования наводнений, необходимо определить, какие объекты подвержены затоплению. Для этого можно использовать алгоритм анализа высоты:

$$Flooded(O) = \{ 1, \text{если } H_O < H_{water} ; 0, \text{иначе} \},$$

где H_O — высота объекта O ; H_{water} — уровень воды в данной области.

В контексте моделирования наводнений предполагается, что применение геокодирования и использование описанных выше формул способствует повышению точности, поскольку интеграция данных о рельефе и использовании моделей потока (например, моделей для расчета стока воды) может повысить точность предсказаний наводнений на 20–40 %. Это достигается благодаря более точному моделированию физических процессов.

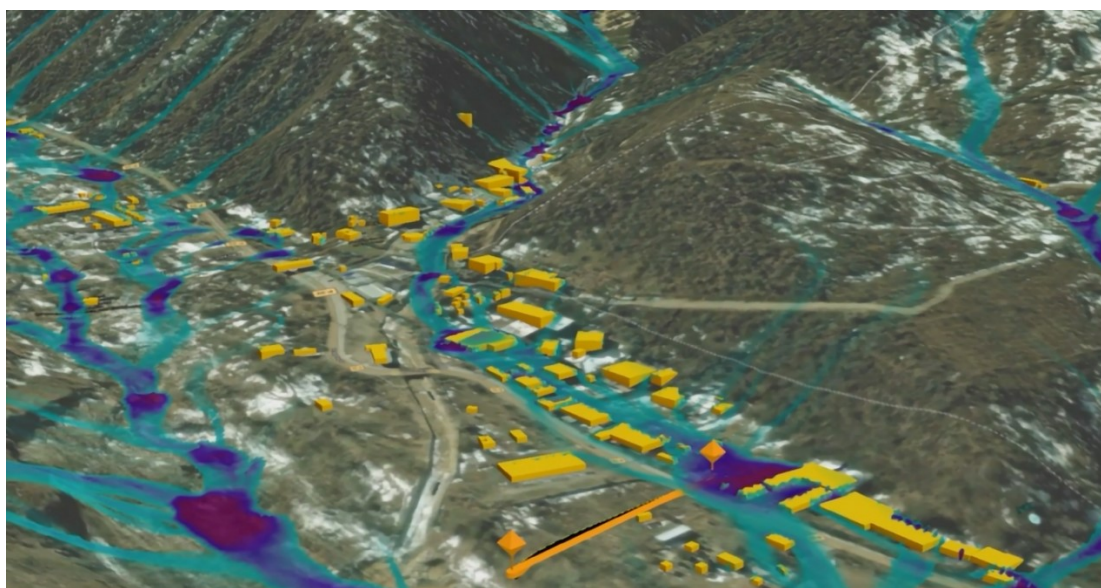


Рис. 3. Вычислительный эксперимент

Заключение

В заключение можно отметить, что разработанный прототип ГИС-системы, построенный на основе C# и библиотеки DotSpatial, успешно реализует базовые задачи ситуационного анализа данных в вычислительных экспериментах. Эта система позволяет проводить простые вычислительные эксперименты с пространственными данными. ГИС-интерфейс обеспечивает визуализацию с использованием нескольких слоев картографической информации и анализ взаимодействий объектов и процессов в статике. Применение формата dBase и поддержка Shapefile обеспечивают надежное хранение и обработку данных, что делает систему устойчивой и гибкой. Данный прототип позволяет проводить анализ базовых ситуационных задач, оценивая пространственные и временные данные. В перспективе возможны дальнейшие улучшения, направленные на расширение функциональности, интеграцию с более сложными моделями анализа данных и разработку новых инструментов для поддержки вычислительных экспериментов.

Список источников:

1. Сагидова М. Л., Фридман А. Я. Основы ГИС-интерфейса в ситуационной системе вычислительного эксперимента // Теоретические и прикладные модели информатизации региона. Апатиты: КНЦ РАН, 2000. С. 108–112.
2. Сагидова М. Л., Фридман А. Я. Реализация интерфейса ГИС и концептуальной модели в ситуационной системе вычислительного эксперимента // Модели социальных, технологических и образовательных процессов. Апатиты: КНЦ РАН, 2001. С. 93–98.

3. Сагидова М. Л., Фридман А. Я., Фридман О.В. Интеллектуальная обработка данных в ГИС-приложениях // Информационные технологии в региональном развитии. Апатиты, 2008. Вып. VIII. С. 78–81.
4. Сагидова М. Л., Фридман А. Я. Метод ГИС-представления картографических задач исследования и обеспечения безопасности // Управление безопасностью природно–промышленных систем. Апатиты: КНЦ РАН, 2000. Вып. III. С. 6–20
5. Фридман А. Я., Курбанов В. Г. Информационная технология интеграции пространственных данных в ситуационную систему моделирования // Труды СПИИРАН. 2016. Вып. 4(47). С. 163–189. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн) www.proceedings.spiiras.nw.ru. doi:10.15622/sp.47.9 17 с.
6. Alexander Fridman. Situational Modeling: Definitions, Awareness, Simulation. USA: Nova Science Publishers, Inc., 2023, 331 pp. <https://novapublishers.com/shop/situational-modeling-definitions-awareness-simulation/> doi:<https://doi.org/10.52305/XIKU5849>.
7. Сагидова М. Л., Фридман А. Я., Олейник А. Г. Алгоритм геокодирования для моделирования протяженных объектов энергетической системы Мурманской области // Системы информационной поддержки регионального развития. Апатиты: КНЦ РАН, 1998. С. 33–36.
8. Сагидова М. Л., Фридман А. Я. Представление результатов вычислительного эксперимента при моделировании природно-технических комплексов средствами ГИС // Информационные технологии в региональном развитии. Апатиты: ИИММ КНЦ РАН, 2005. Вып. V. С. 39–41.

References

1. Sagidova M. L., Fridman A. Ya. Fundamentals of GIS interface in the situational system of computational experiment // Theoretical and applied models of informatization of the region. — Apatity: KSC RAS, 2000. Pp. 108–112.
2. Sagidova M. L., Fridman A. Ya. Implementation of GIS interface and conceptual model in the situational system of computational experiment // Models of social, technological and educational processes. — Apatity: KSC RAS, 2001. — Pp. 93–98.
3. Sagidova M. L., Fridman A. Ya., Fridman O. V. Intelligent data processing in GIS applications // Information technologies in regional development. — Apatity, 2008. — Issue VIII. Pp. 78–81.
4. Sagidova M. L., Fridman A. Ya. GIS-representation method of cartographic research and safety tasks // Safety management of natural-industrial systems. Apatity: KSC RAS, 2000. Issue III. P. 6–20
5. Fridman A. Ya., Kurbanov V. G. Information technology of spatial data integration into situational modeling system // Proceedings of SPIIRAS. 2016. Issue 4(47). P. 163–189. ISSN 2078-9181 (printed), ISSN 2078-9599 (online) www.proceedings.spiiras.nw.ru. doi:10.15622/sp.47.9 17 p.
6. Alexander Fridman. Situational Modeling: Definitions, Awareness, Simulation. USA: Nova Science Publishers, Inc., 2023, 331 pp. <https://novapublishers.com/shop/situational-modeling-definitions-awareness-simulation/> doi:<https://doi.org/10.52305/XIKU5849>
7. Sagidova M. L., Fridman A. Ya., Oleynik A. G. Geocoding algorithm for modeling extended objects of the energy system of the Murmansk region // Systems of information support for regional development. Apatity: KSC RAS, 1998. Pp. 33–36.
8. Sagidova M. L., Fridman A. Ya. Presentation of the results of a computational experiment in modeling natural and technical complexes by means of GIS // Information technologies in regional development. Apatity: IIMM KSC RAS, 2005. Issue V. Pp. 39–41.

Информация об авторах

Н. А. Вдовиченко — студент 1 курса аспирантуры;
Н. Н. Руденко — студент 1 курса аспирантуры.

Information about the authors

N. A. Vdovichenko — 1st year student of the Postgraduate School of the Kola Science Center of the Russian Academy of Sciences;
N. N. Rudenko — 1st year student of the Postgraduate School of the Kola Science Center of the Russian Academy of Sciences.

Статья поступила в редакцию 01.10.2024; одобрена после рецензирования 10.10.2024; принята к публикации 21.10.2024.
The article was submitted 01.10.2024; approved after reviewing 10.10.2024; accepted for publication 21.10.2024.

Научная статья
УДК 004.9
doi:10.37614/2949.1215.2024.15.3.010

ОПАСНЫЕ ОБЪЕКТЫ И ПРОЦЕССЫ АРКТИЧЕСКОЙ ЗОНЫ РОССИЙСКОЙ ФЕДЕРАЦИИ (ПО МАТЕРИАЛАМ КОНФЕРЕНЦИИ «III ЮДАХИНСКИЕ ЧТЕНИЯ»)

Сергей Юрьевич Яковлев¹, Алексей Сергеевич Шемякин²

^{1, 2}*Институт информатики и математического моделирования имени В. А. Путилова Кольского научного центра Российской академии наук, Апатиты, Россия*

¹*Федеральное государственное автономное образовательное учреждение высшего образования «Мурманский арктический университет»*

¹*s.yakovlev@ksc.ru, <https://orcid.org/0000-0001-6433-2096>*

²*a.shemyakin@ksc.ru, <https://orcid.org/0000-0001-5308-5456>*

Аннотация

Статья посвящена вопросам обеспечения устойчивого развития и экологической безопасности территорий Арктической зоны Российской Федерации. Особое внимание уделено проблемам оценки безопасности промышленно-природных комплексов и критических инфраструктур. Статья написана на основе опыта участия в конференции «III Юдахинские чтения», которая состоялась в г. Архангельске с 25 по 28 июня 2024 г.

Ключевые слова:

Арктическая зона Российской Федерации, опасные объекты, оценка безопасности, критические инфраструктуры

Благодарности:

работа выполнена в рамках темы НИР 122022800547-3 «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации».

Для цитирования:

Яковлев С. Ю., Шемякин А. С. Опасные объекты и процессы Арктической зоны Российской Федерации (по материалам конференции «III Юдахинские чтения») // Труды Кольского научного центра РАН. Серия: Технические науки. 2024. Т. 15, № 3. С. 113–118. doi:10.37614/2949.1215.2024.15.3.010.

Original article

HAZARDOUS OBJECTS AND PROCESSES OF THE ARCTIC ZONE OF THE RUSSIAN FEDERATION (BASED ON THE PROCEEDINGS OF THE CONFERENCE "III YUDAKHIN READINGS")

Sergey Yu. Yakovlev¹, Alexey S. Shemyakin²

^{1, 2}*Putilov Institute for Informatics and Mathematical Modeling of the Kola Science Centre of the Russian Academy of Sciences, Apatity, Russia*

¹*Federal State Autonomous Educational Institution of Higher Education "Murmansk Arctic University"*

¹*s.yakovlev@ksc.ru, <https://orcid.org/0000-0001-6433-2096>*

²*a.shemyakin@ksc.ru, <https://orcid.org/0000-0001-5308-5456>*

Abstract

The article is devoted to the issues of ensuring sustainable development and environmental safety of the territories of the Arctic zone of the Russian Federation. Particular attention is paid to the problems of assessing the safety of industrial and natural complexes and critical infrastructures. The article is based on the experience of participating in the conference "III Yudakhin Readings", which was held in Arkhangelsk from June 25 to 28, 2024.

Keywords:

Arctic zone of the Russian Federation, hazardous objects, safety assessment, critical infrastructure

Acknowledgments:

The work was carried out within the framework of the research topic 122022800547-3 "Development of theoretical and organizational-technical foundations for information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation"

For citation:

Yakovlev S. Yu., Shemyakin A.S. Hazardous objects and processes of the Arctic zone of the Russian Federation (based on the proceedings of the conference "III Yudakhin readings") // Transactions of the Kola Science Centre of RAS. Series: Engineering Sciences. 2024. Vol. 15, No. 3. P. 113–118. doi:10.37614/2949.1215.2024.15.3.010.

Введение

В конце июня 2024 г. в г. Архангельске состоялась Всероссийская конференция с международным участием «III Юдахинские чтения». Один из авторов принял участие в пленарном заседании, а также в заседаниях ряда секций. По итогам конференции выпущен сборник [1].

Выступления и материалы сборника позволяют получить обзорное представление о современном этапе, планах и перспективах развития Арктической зоны РФ (АЗРФ).

В соответствии с тематикой планов НИР Института информатики и математического моделирования им. В. А. Путилова Кольского научного центра Российской академии наук (ИИММ КНЦ РАН) и Федерального государственного автономного образовательного учреждения высшего образования «Мурманский арктический университет» (МАУ) особое внимание уделено вопросам оценки безопасности объектов и комплексов АЗРФ.

Общие впечатления

Конференцию традиционно проводил Федеральный исследовательский центр комплексного изучения Арктики имени академика Н. П. Лавёрова Уральского отделения РАН (ФИЦКИА УрО РАН). Она посвящена памяти Феликса Николаевича Юдахина (1934–2011) — выдающегося советского и российского ученого, специалиста в области глубинного строения и геодинамики литосферы, сейсмичности, геоэкологических проблем северных территорий, чл.-корр. АН Киргизской ССР, чл.-корр. РАН, одного из основателей архангельской академической науки.

Пленарное заседание началось с приветственных слов Правительства Архангельской области, РАН и других организаций, а также Л. А. Юдахинной — вдовы Ф. Н. Юдахина. Далее прозвучали доклады общего характера академиков РАН В. А. Черешнева (по истории РАН), А. Д. Гвишиани (по нечетким множествам и большим данным в анализе природных катастроф), представителей Киргизской АН. Выступления академиков и членов-корреспондентов РАН К. В. Лобанова (о Кольской сверхглубокой скважине), В. А. Конторовича (характеристика осадочных бассейнов Карского моря), В. А. Румянцева. (о состоянии поверхностных пресных вод) и других докладчиков были посвящены более конкретным проблемам. Приятной неожиданностью пленарного заседания было выступление Северного Русского народного хора. И в целом культурная программа конференции была весьма интересной: прогулка на теплоходе по Северной Двине, поездка в Музей-заповедник деревянного зодчества «Малые Корелы».

Секционные заседания были частично объединены и укрупнены. Во многих докладах приводились результаты экспедиций участников на объекты и территории, в том числе данные о методах и итогах измерений. Не все выступления равноценно отражены в сборнике [1]. В дальнейшем мы будем ссылаться на этот сборник, указывая авторов и страницы, а также, в ряде случаев, отражать впечатления от выступлений на секциях. Основное внимание будет уделяться описанию и характеристикам опасности объектов АЗРФ.

Секция «Геолого-геофизические исследования на Северном морском пути»

В исследовании [Беленович Т. Я., Неверов Н. А.; с. 14–17] проводился анализ состояния земной коры в Холмогорском тектоническом узле. Было отмечено, что разломы и зоны ослабленной прочности такие, как тектонические узлы, имеют большое значение для твердой среды внешних геосфер. В Холмогорском тектоническом узле было обнаружено резкое увеличение амплитудно-частотных характеристик короткопериодических колебаний магнитного поля во время магнитных бурь. Это приводит к образованию диссипативных структур земной коры, которые создают аномальные условия для развития компонентов окружающей среды.

Работа [Данилов К. Б., Афонин Н. Ю.; с. 18–21] представляет результаты микросейсмического зондирования вмещающей среды вблизи кимберлитовой трубки Поморская Архангельской алмазонасной провинции. По результатам обработки данных сформирована диаграмма распределения относительной интенсивности микросейсм вдоль профиля и по глубине.

Статья [Кутинов Ю. Г., Чистова З. Б., с. 26–30] подчеркивает важную роль дистанционного зондирования Земли (ДЗЗ) при освоении месторождений углеводородного сырья в Арктике. Особое внимание уделяется предотвращению возможных аварий и чрезвычайных ситуаций, связанных с разливами нефтепродуктов в арктических водах. Основной сложностью борьбы с чрезвычайными ситуациями подобного рода является то, что загрязняющие вещества попадают на поверхность льдов и со льдами могут переноситься на значительные расстояния. Эта проблема особенно актуальна в водах Северной Атлантики. При таких условиях методы ДЗЗ становятся особенно актуальными.

Секция «Влияние опасных природных процессов и явлений на безопасность хозяйственной деятельности в прибрежных районах Арктической зоны РФ»

В исследовании [Кузякин Л. П. и др.; с. 49–53] проводится сравнительный анализ состояния Новопортовского мерзлотника в контексте изменяющегося климата и усиления техногенного воздействия. Эти подземные хранилища, расположенные в криолитозоне, широко представлены в арктических регионах Евразии и Северной Америки. Анализ состояния мерзлотника и сравнение с историческими данными позволили выявить, что климатические изменения и социально-экономические факторы представляют угрозу стабильности этого подземного сооружения.

В работе [Миронюк С. Г., Донцова Г. Ю.; с. 59–64] рассматриваются опасные природные процессы, происходящие в шельфово-береговой зоне Печорского моря. Эта зона перспективна для добычи нефти и газа, но освоение месторождений связано с технологическими, экологическими и природными рисками. Глобальное потепление, повышение уровня моря и таяние вечной мерзлоты создают дополнительные угрозы. Прогноз включает определение видов, масштабов, скоростей развития и времени возникновения опасных геологических процессов и явлений. В отсутствие точных данных для количественного прогнозирования эксперты могут составить качественный прогноз изменений инженерно-геологических условий. Для повышения надежности прогнозов рекомендуется использовать метод аналогий. В качестве примера приведены данные об опасных процессах на участках строительства скважин на шельфе, прилегающем к Варандейскому району. В ходе строительных работ и эксплуатации месторождений ожидаются изменения в интенсивности и направленности опасных процессов таких, как миграция наносов, изменения рельефа дна и побережья под воздействием льда, криогенные процессы и газопроявления.

В исследовании [Наход В. А., Малов А. И.; с. 64–69] анализируется воздействие горнодобывающей промышленности на окружающую среду на примере гипсового месторождения в Холмогорском районе Архангельской области. В рамках экспедиции были собраны образцы поверхностных вод, донных отложений и почвы. Основываясь на анализе и классификации собранных данных, основные источники загрязнения окружающей среды были разделены на пять групп, включая микроэлементы и макроэлементы, а также выделены естественные и техногенные причины.

В статье [Николаева Е. С. и др.; с. 69–74] исследуется влияние образования наледей на территории Южной Якутии, которое оказывает значительное негативное воздействие на инженерные объекты. Были рассмотрены методы определения наледей с использованием разностного индекса NDSI на основе снимков Sentinel-1 и Sentinel-2. Результаты дешифровки спутниковых снимков сравнивались с результатами полевых исследований. В рамках полевых исследований на ключевых участках Алданского нагорья и Станового хребта была проведена проверка результатов дешифровки. Результаты данного комплексного исследования позволят получить более точную информацию о процессе образования наледей в регионе и могут быть использованы для создания каталога наледей и улучшения методов их определения с помощью дистанционного зондирования.

В исследовании [Огородов С. А., Ермолов А. А.; с. 74–79] анализируются опасности, связанные с рельефообразующими процессами, для инженерных и хозяйственных объектов, экосистем и жизни людей в прибрежной и шельфовой зонах морей Российской Арктики. Береговая зона моря формируется под воздействием двух основных групп рельефообразующих процессов: абразионно-аккумулятивных и ледово-экзарационных. Подводные трубопроводы и кабели связи, пересекающие

береговую линию, являются наиболее уязвимыми к опасным береговым процессам. Однако, зачастую, строительство осуществляется без учета данных процессов, что, впоследствии, приводит к возникновению аварий и сопутствующему финансовому ущербу.

В работе [Старцев В. О., Лаптев А. Б.; с. 83–87] анализируется состояние и развитие тестирования авиационных и других материалов для удовлетворения требований к материалам, необходимым для безопасной работы самолетов и инфраструктуры, зданий и сооружений в арктическом климате. В Национальном исследовательском центре «Курчатовский институт» — ВИАМ регулярно проводятся натурные климатические испытания материалов во всех климатических зонах — от тропиков до Арктики (Мурманск) и холодного климата (Якутск).

Секция «Влияние изменений климата и антропогенных нагрузок на ландшафты, биоразнообразие и биологические ресурсы»

В статье [Куликов К. Н., Ермаков А. П.; с. 236–238] представлены результаты радиационного анализа кормового блока плавучей технической базы «Лепсе» на этапе завершения вывода из эксплуатации, а также технологические подходы, разработанные для обеспечения радиационной безопасности.

Автор [Лисичкин Г. В.; с. 238–242] изучает проблему ликвидации аварийных разливов нефти в Арктическом регионе. Ледяной и снежный покров значительно усложняют процесс по сравнению с южными широтами. Нефть может оказаться как на поверхности льда, так и под ним после разлива. Устранить нефтяное пятно в ледовых акваториях арктических морей и побережий крайне сложно, особенно в зимний период полярной ночи. В таких условиях сорбционная очистка неприменима. Большие проблемы возникают при попытке очистить от разлитой нефти болота и скалистые берега водоемов. В этом случае сбор использованных сорбентов невозможен.

В заметке [Погожева М. П., Чекменева Н. А., Долгова А. О.; с. 258–259] рассматривается государственный природный заповедник «Остров Врангеля», включенный в список Всемирного природного наследия ЮНЕСКО с 2004 г. Экосистемы Арктики очень чувствительны к загрязняющим веществам из-за низкой интенсивности массообмена и энергообмена, медленного процесса самоочищения и коротких пищевых цепочек, которые способствуют быстрому распространению токсичных веществ до конечных потребителей. Отмечается, что исследование загрязнения Арктики тяжелыми металлами становится все более актуальным из-за растущих рисков как локального, так и удаленного характера, связанных с расширением хозяйственной деятельности в регионе.

Исследование [Румянцев И. С.; с. 262–265] посвящено влиянию нефтегазодобычи на окружающую среду в Арктике на примере Кумжинского газоконденсатного месторождения. Во время разведочного бурения на месторождении происходили различные аварии. Очень известна авария, произошедшая в 1980 г. на скважине № 9, также хорошо запомнились и неудачные попытки ее устранения, включая ядерный взрыв. На ликвидацию аварии ушло 7 лет, ее последствия для окружающей среды были серьезными. Исследования, проведенные в 2000-х гг. в районе скважины, показали превышение предельно допустимых концентраций: в атмосфере — в 44 раза, углеводородов в поверхностных водах — в 180 раз, углеводородов в почве — в 540 раз. Был сделан вывод о том, что негативное воздействие на окружающую среду при бурении газовых и газоконденсатных скважин проявляется в основном в аварийных ситуациях на скважинах.

Работа [Сидорова Т. А., Гребенец В. И., Юров Ф. Д.; с. 266–269] оценивает влияние региональных климатических изменений и микроклиматических условий, формирующихся в городской среде, на температурное поле вечномерзлых пород в Арктическом регионе. Прогнозирование динамики снежного покрова в городах сталкивается с рядом сложностей, связанных с интенсивным антропогенным воздействием на перераспределение и уплотнение снежных масс на городских территориях.

Секция «Социально-экономическое развитие Российской Арктики в новых условиях: проблемы и пути решения»

В исследовании [Пляскина Н. И.; с. 291–295] предлагается динамическая имитационная модель для оценки эффективности инвестиционных проектов в области низкоуглеродной энергетики на основе метода дисконтирования денежных потоков. Углеродное регулирование предполагается

осуществлять через введение трансграничного углеродного налога. Апробация этого подхода проводится на примере нефтегазовой компании в альтернативной энергетике. Исследование анализирует переход нефтегазовых компаний к углеродной нейтральности.

Работа [Постникова У. С.; с. 296–297] рассматривает обеспечение социально-природно-техногенной безопасности регионов холодного климата как одну из приоритетных задач устойчивого развития АЗРФ. Промышленное освоение и техническое оснащение территорий Крайнего Севера приводит к противоречиям между развитием инфраструктуры, повышением занятости и жизненного уровня населения, улучшением социально-экономических показателей региона и его конкурентных преимуществ, с одной стороны, и опасностью для окружающей природной среды и человека, с другой. Работа направлена на анализ опасностей и рисков северных и арктических территорий в условиях многофакторного воздействия.

В статье [Хадько А. И.; с. 303–304] представлены основные подходы к управлению территориями, их преимущества и недостатки: функциональный, системный, процессный, проектный и ситуационный.

Заключение

Отметим некоторые основные направления исследований по безопасности, отраженные на конференции (рис. 1).

1. Изучение опасных геологических процессов и явлений, инженерно-геологических условий, опасных природных процессов (тектонические узлы, кимберлитовые трубки, закарстованные территории, шельфово-береговые зоны, многолетнемерзлые породы, наледи).

2. Опасности и риски, связанные с добычей нефти и газа (нефтегазогеологическое районирование, аварии и чрезвычайные ситуации в условиях АЗРФ, роль ДЗЗ).

3. Проблемы жизнеобеспечения (транспортно-логистические системы, города, жилые дома, сооружения, мониторинг состояния, устойчивость, водообеспечение).

4. Техногенно-природные риски (горнодобывающие производства, арктические полярные станции, заповедные и охраняемые территории, трубопроводы и кабели связи, радиационное воздействие, состояние воды, воздуха, почвы).

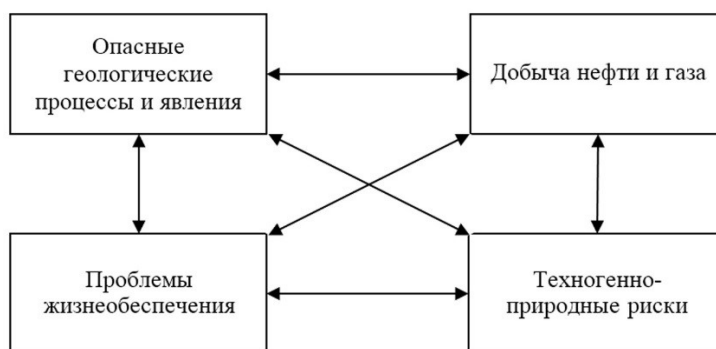


Рис. 1. Опасные объекты и процессы АЗРФ

В целом можно констатировать взаимосвязь направлений (они имеют отношение одновременно к различным сферам: техногенной, природной, социальной), разнообразие объектов исследований, а также методов и способов анализа. В каждом из блоков (рис. 1) можно выделить свои составляющие, отражающие более узкие, частные виды опасностей.

Список источников

1. III Юдахинские чтения: сб. науч. мат-лов, Архангельск, 25–28 июня 2024 г. Архангельск: КИРА, 2024. 468 с. ISBN 978-5-98450-859-9.

References

1. III Yudakhinskiye chteniya: sbornik nauchnykh materialov, Arhangel'sk, 25–28 iyunja 2024 goda. . [III Yudakhinsky readings: collection of scientific materials] — Arhangel'sk: KIRA, 2024. — 468 s. — ISBN 978-5-98450-859-9. 2.

Информация об авторах

С. Ю. Яковлев — кандидат технических наук, старший научный сотрудник;

А. С. Шемякин — младший научный сотрудник.

Information about the authors

S. Yu. Yakovlev — Candidate of Science (Tech.), Senior Researcher;

A. S. Shemyakin — Junior Researcher.

Статья поступила в редакцию 23.10.24; одобрена после рецензирования 29.10.24; принята к публикации 08.11.

The article was submitted 23.10.24; approved after reviewing 29.10.24; accepted for publication 08.11.24.

