

Геоинформатика. 2024. № 3. С. 53–63.  
*Geoinformatika*. 2024;(3):53–63.

### Информационные системы в геологии и геофизике

Научная статья  
 УДК 004.85  
<https://doi.org/10.47148/1609-364X-2024-3-53-63>

## Цифровые технологии будущего — современные решения в науках о Земле

© 2024 г. — В.В. Наумова<sup>1, а)</sup>, М.И. Патук<sup>1, б)</sup>, А.С. Еременко<sup>1, 2, 3, с)</sup>, А.А. Загумённых<sup>1, 2, 3, д)</sup>, В.С. Ерёменко<sup>1, е)</sup>

<sup>1</sup>Государственный геологический музей им. В.И.Вернадского РАН; Москва, Россия

<sup>2</sup>Институт автоматизации и процессов управления ДВО РАН; г. Владивосток, Россия

<sup>3</sup>Дальневосточный федеральный университет; г. Владивосток, Россия

<sup>а)</sup>naumova\_new@mail.ru, <sup>б)</sup>patuk@mail.ru, <sup>с)</sup>academy21@gmail.com, <sup>д)</sup>trueepikvic@gmail.com, <sup>е)</sup>vitaer@gmail.com

**Аннотация:** В статье описаны технологии, способные радикально изменить развитие сразу многих областей: искусственный интеллект, квантовые технологии, большие данные, технологии беспроводной связи, системы распределенного реестра. Авторы рассмотрели ряд перспективных технологий ближайшего будущего, которые и в настоящее время имеют перспективы применения в науках о Земле.

**Ключевые слова:** науки о Земле; искусственный интеллект; квантовые технологии; большие данные; технологии беспроводной связи; системы распределенного реестра.

*Для цитирования:* Наумова В.В., Патук М.И., Ерёменко А.С., Загумённых А.А., Ерёменко В.С. Цифровые технологии будущего — современные решения в науках о Земле // Геоинформатика. — 2024. — № 3. — С. 53–63. <https://doi.org/10.47148/1609-364X-2024-3-53-63>.

### Information systems in geology and geophysics

Original article

## Digital technologies of the future — modern solutions in Earth sciences

© 2024 — V.V. Naumova<sup>1, а)</sup>, M.I. Patuk<sup>1, б)</sup>, A.S. Eremenko<sup>1, 2, 3, с)</sup>, A.A. Zagumennov<sup>1, 2, 3, д)</sup>, V.S. Eremenko<sup>1, е)</sup>

<sup>1</sup>Vernadsky State Geological Museum RAS; Moscow, Russia

<sup>2</sup>Institute of automatic and control processes of FEB RAS; Vladivostok, Russia

<sup>3</sup>Far East Federal University; Vladivostok, Russia

<sup>а)</sup>naumova\_new@mail.ru, <sup>б)</sup>patuk@mail.ru, <sup>с)</sup>academy21@gmail.com, <sup>д)</sup>trueepikvic@gmail.com, <sup>е)</sup>vitaer@gmail.com

**Abstract:** The article discusses technologies that can radically change the development of many areas at once: artificial intelligence, quantum technologies, big data, wireless communication technologies, distributed registry systems. The authors consider a number of promising technologies of the near future that currently have prospects for application in Earth sciences.

**Key words:** Earth sciences; artificial intelligence; quantum technologies; big data; wireless communication technologies; distributed ledger systems.

*For citation:* Naumova V.V., Patuk M.I., Eremenko A.S., Zagumennov A.A., Eremenko V.S. Digital technologies of the future — modern solutions in Earth sciences. *Geoinformatika*. 2024;(3):53–63. <https://doi.org/10.47148/1609-364X-2024-3-53-63>. In Russ.

### Цифровые технологии будущего

В статье описаны технологии, способные радикально изменить развитие сразу многих областей. Эксперты Всемирного экономического форума выпустили рейтинг самых перспективных технологий ближайшего будущего [1]. Авторы рассмотрели ряд перспективных технологий ближайшего будущего, которые и в настоящее время имеют перспективы применения в науках о Земле.

#### 1. Искусственный интеллект (ИИ)

Искусственный интеллект — это отдельное направление компьютерных наук. Специалисты по ИИ разрабатывают системы, которые анализируют информацию и решают задачи аналогично тому, как это делает человек. ИИ использует алгоритмы, которые позволяют компьютеру обрабатывать большие

объемы данных и находить в них закономерности. На основе этих закономерностей ИИ может делать выводы, предсказывать события или принимать решения.

Основные области применения ИИ: голосовые помощники, рекомендательные системы, распознавание образов, автопилоты и автономные транспортные системы, финансовые аналитические системы, языковые переводчики, игровая индустрия, медицинская диагностика, робототехника и др.

В настоящее время наиболее перспективными представляются генеративные нейросети — искусственный интеллект, способный создавать новый и оригинальный контент с помощью анализа паттернов в данных, сложных алгоритмов и методов машинного обучения [2]. Генеративные нейросети

фокусируются преимущественно на создании текста, программного кода, изображений и звука, однако эту технологию можно применять в различных сферах.

## **2. Виртуальные ассистенты**

Виртуальный ассистент — программный агент, который может выполнять задачи для пользователя на основе информации, введенной пользователем, а также информации, полученной из различных интернет-ресурсов.

Появилось множество сервисов и чат-ботов на основе ChatGPT и других нейросетей, которые помогают в самых разных сферах и задачах. Чаще всего ChatGPT как персональному помощнику поручают: поиск и обобщение информации из Интернета; проверку кода; роль персонального репетитора английского; интерпретацию и пересказ фильмов, книг и сериалов; составление индивидуальных подборок по интересам; разработку программы домашних тренировок; мониторинг вакансий, авиабилетов и т. д.; генерацию текстов, ответов на электронные письма, поздравлений; разъяснение сложных или обширных тем. Согласно прогнозам, число используемых цифровых помощников в ближайшее время превысит общее население Земли. От простых чат-ботов, созданных для решения конкретных задач, они эволюционируют в советчиков, способных помочь в решении почти любой проблемы.

## **3. Технологии беспроводной связи**

На сегодняшний день большое развитие в области передачи данных получили беспроводные сети. Это объясняется удобством их использования, дешевизной и приемлемой пропускной способностью. Исходя из текущей динамики развития, можно сделать вывод о том, что по количеству и распространности беспроводные сети в скором времени превзойдут проводные сети.

Беспроводные сенсорные сети получили широкое развитие в последнее время. Такие сети, состоящие из множества миниатюрных узлов, оснащенных маломощным приемопередатчиком, микропроцессором и сенсором, могут связать воедино глобальные компьютерные сети и физический мир. Большой интерес к изучению таких систем обусловлен широкими возможностями применения сенсорных сетей. Из-за своей способности к самоорганизации, автономности и высокой отказоустойчивости такие сети активно применяются в системах безопасности.

Технология ретранслируемой ближней радиосвязи 802.15.4/ZigBee, известная как WSN — Wireless Sensor Network, является одним из современных направлений развития самоорганизующихся отказоустойчивых распределенных систем наблюдения и управления ресурсами и процессами. Сегодня технология беспроводных сенсорных сетей является единственной беспроводной технологией, с помо-

щью которой можно решить задачи мониторинга и контроля, которые критичны ко времени работы датчиков. Объединенные в беспроводную сенсорную сеть, датчики образуют территориально распределенную самоорганизующуюся систему сбора, обработки и передачи информации. Основными областями применения являются контроль и мониторинг измеряемых параметров физических сред и объектов.

## **4. Системы распределенного реестра**

Система распределенного реестра — это база цифровых транзакций, записей о событиях, содержащих критически важную управленческую, юридическую, финансовую и иную информацию, которые хранятся, одновременно создаются и обновляются на всех носителях у всех участников реестра на основе заданных алгоритмов, обеспечивающих ее тождественность у всех пользователей реестра [3].

Множественное копирование данных среди участников гарантирует сохранность и неизменность внесенной в блокчейн информации. Файлы зашифрованы закрытыми ключами, что делает невозможным доступ к ним без ключа. Файлы также делятся на части для хранения на нескольких узлах, исключая единую точку хранения. Из-за специфики устройства блокчейна информацию невозможно подменить, отредактировать или удалить. Кроме того, в отличие от централизованных БД, на блокчейн гораздо сложнее провести хакерскую атаку. Благодаря этим свойствам блокчейн в том числе позволяет сэкономить на хранении данных.

Однажды записанная информация сохраняется в блокчейне навсегда. Соответственно, исчезает возможность манипуляции данными. Кроме того, каждый участник сети имеет доступ ко всей истории транзакций, вплоть до самой первой. Это повышает доверие к компании и к ее бизнес-процессам. Узлы сети блокчейна постоянно сверяют информацию между собой и в случае подмены данных на некоторых из них восстанавливают эти данные.

За счет того, что хранение данных в блокчейне максимально надежно, блокчейн позволяет отказаться от бумажных документов и поместить всю информацию в сеть.

## **5. Большие данные и квантовые вычисления**

Большие данные в широком смысле — это информация столь большого объема и сложного состава, что обычные методы ее поиска, структуризации, анализа или визуализации перестают быть эффективными. В рамках научного направления, называемого «наукой о данных» (data science), ученые разрабатывают подходы массово-параллельной обработки данных с помощью современных алгоритмов.

Квантовые компьютеры обещают революционизировать область вычислений. Они смогут решать

задачи, которые сегодня считаются невозможными для классических компьютеров.

### 6. Виртуальная и дополненная реальность (VR и AR)

VR- и AR-технологии изменяют способ восприятия информации и взаимодействия с миром. Они найдут применение в образовании, популяризации научных знаний. С помощью искусственного интеллекта создаются интерактивные миры, в которых с различной степенью сходства имитируется человеческая реальность.

### 7. Digital-инструменты

Цифровые инструменты — это программное обеспечение, приложения и устройства, используемые для работы с цифровыми данными. Они позволяют пользователям создавать, редактировать, хранить, обрабатывать и передавать информацию в цифровом формате.

Уже сегодня для широкого применения доступны digital-инструменты, которые еще несколько десятилетий назад казались недостижимым будущим.

#### Современные решения в науках о Земле

Современный мир изменяется со все ускоряющимся темпом, активно внедряя в свою структуру и повседневную жизнь все больше и больше цифровых технологий. В современном мире цифровые технологии стали незаменимым инструментом в процессе обработки и передачи информации. Все больше процессов происходит в онлайн-режиме, и цифровые инструменты помогают упростить и автоматизировать множество задач, улучшая таким образом эффективность работы и повышая продуктивность в различных отраслях.

Цифровые технологии начали активно внедряться в науки о Земле, в том числе и в геологию.

#### Искусственный интеллект

Выделяют семь направлений:

- компьютерное зрение;
- обработка естественного языка;
- распознавание и синтез речи;
- рекомендательные системы и интеллектуальные системы поддержки принятия решений;
- перспективные методы и технологии в ИИ;
- нейропротезирование;
- нейроинтерфейсы, нейростимуляция и нейросенсинг.

В последние десятилетия наблюдается стремительный рост интереса к нейронным сетям, которые успешно применяются в различных областях. Нейронные сети используются всюду, где нужно решать задачи прогнозирования, классификации, нелинейной регрессии или управления.

Одной из разновидностей нейронных сетей, предназначенной для обработки изображений и

других точечных форматов, являются сверточные нейронные сети. В геологоразведке они применяются для выявления объектов (например, решения Orefox — URL: <https://orefox.com/>), обработки и интерпретации сейсмических данных (Geolearn — URL: <https://www.geolearn.ai/>), определения минералов-индикаторов в пробах ледниковых отложений (IOS Services Geoscientifiques — URL: <https://www.iosgeo.com/en/>), а также количественного и качественного описания кернов по их фотоснимкам (Geolearn — URL: <https://www.geolearn.ai/>) или гиперспектральным данным (Solve Geosolutions — URL: <https://datarock.com.au/platform/>).

Последовательность входных данных анализируется с помощью такой разновидности нейронных сетей, как рекуррентные нейронные сети. Они адаптированы для анализа временных наборов данных, таких как временные последовательности или текстовая информация. В геологоразведке рекуррентные нейронные сети используют для выявления перспективных участков на основе отчетов, находящихся в свободном доступе (например, решения Goldspot Discoveries [4]), или для геологического документирования данных бурения на основании измерений физических свойств пород (CGG).

В геологии большую часть времени работы исследователей занимают визуальная диагностика и описание горных пород. В работе [5] предложен новый подход к автоматической классификации пород при описании кернов. Описаны доступные методы автоматической классификации пород, предложен новый подход применения сверточной нейронной сети.

В работе [6] авторы воспользовались специальными методами обучения с нулевым результатом и обучения с малым числом примеров. В результате стало возможно предложить открытый критерий для распознавания необработанных минералов, которые отсутствуют в обучающей выборке. Также предоставлены дополнительные наборы образцов для сегментации, определения размеров образцов и классификации с малым числом примеров.

Бурное развитие в последние годы методов искусственного интеллекта, связанное с обработкой и генерацией текстов, открывает замечательные возможности для извлечения новых знаний из потока научной информации, которую очень трудно, а зачастую и невозможно переработать традиционным методом чтения.

В области наук о Земле с помощью методов обработки естественного языка (NLP) решаются следующие задачи: выделение геологических и географических именованных сущностей (NER), извлечение пространственных и временных взаимосвязей, классификация, кластеризация, реферирование геологических отчетов и публикаций, ответы на вопросы.

Разработан широкий спектр алгоритмов машинного обучения, подходящих для выполнения специализированного геологического анализа [2]. Исходный материал для их обучения обычно либо уже имеется, либо может быть получен самостоятельно. Машинное обучение можно использовать с целью выявления геолого-разведочных объектов в условиях избытка данных (например, решения Goldspot Discoveries, SRK Consulting), автоматического выявления геологических зон залегания полезных ископаемых (Maptek), оценки твердости руды на основе результатов анализа (неопубликованные работы), распознавания частиц золота по фотоснимкам пробы ледниковых отложений (IOS Services Geoscientifiques).

Достаточно подробный анализ применения методов искусственного интеллекта для решения геологических задач приведен в статье Патука и Наумовой [2].

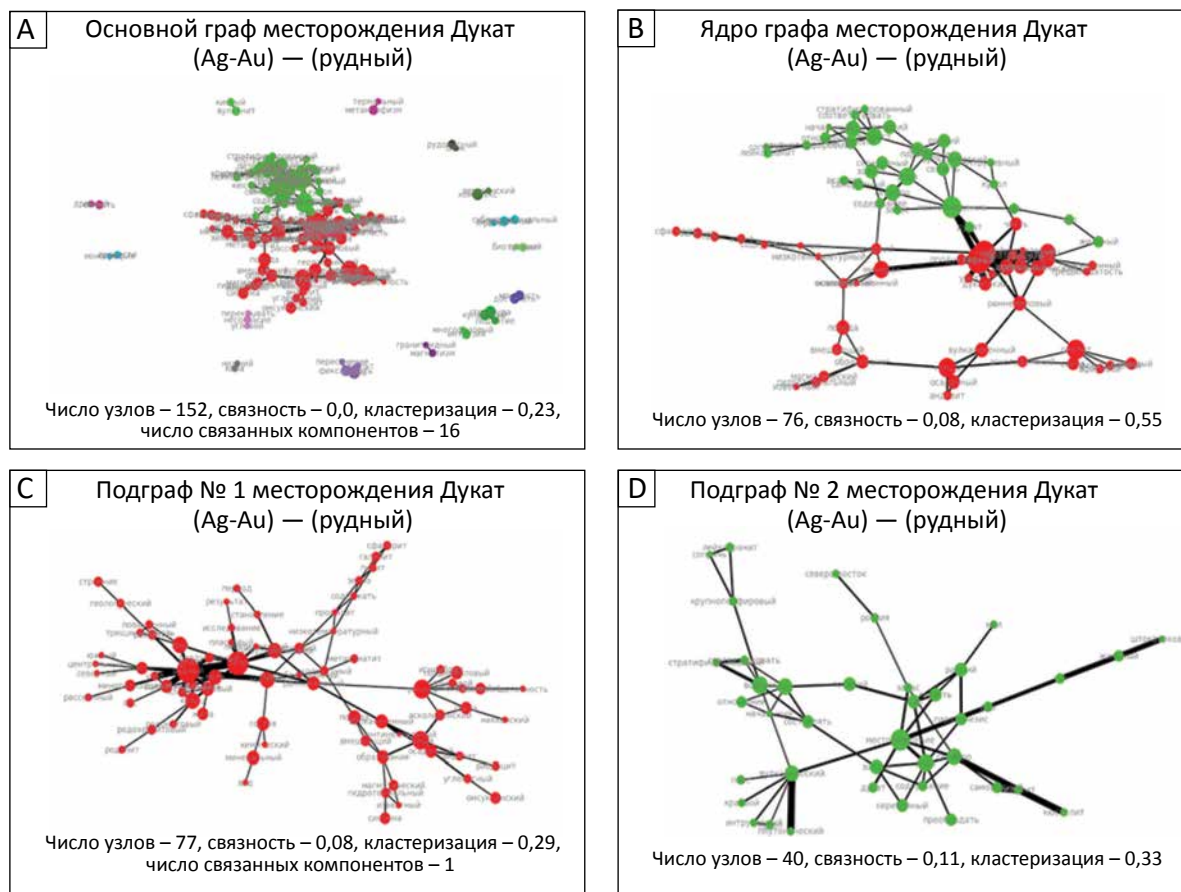
Бурное развитие методов искусственного интеллекта, связанное с обработкой и генерацией текстов, открыло замечательные возможности для извлечения новых знаний из потока научной геологической информации, которую очень трудно, а за-

частую и невозможно переработать традиционным методом чтения.

Авторы предлагают воспользоваться относительно новым подходом с использованием одного из методов искусственного интеллекта — семантического поиска [7]. Данный метод основан на применении больших языковых моделей и сильно зависит от качества их предварительного обучения. Авторами статьи было проведено дообучение нескольких языковых моделей на собранном датасете, который содержал наименования и абстракты статей, посвященных описаниям месторождений твердых полезных ископаемых. Проведен сравнительный анализ полученных результатов. Показаны преимущества семантического поиска по сравнению с традиционными методами. Произведены тестовый поиск и ранжирование публикаций на основе семантической близости с представленным фрагментом текста. Выполнен анализ зависимости качества поиска в зависимости от параметров используемой языковой модели и величины обучающего набора данных. Создан web-сервис поиска информации в архиве публикаций ГГМ РАН <https://repository.geologyscience.ru> (рис. 1).

**Рис. 1.** Подходы для разработки методов анализа текстовой геологической информации с использованием методов искусственного интеллекта и теории графов, предложенные в ГГМ РАН

*Fig. 1. Approaches for developing methods for analyzing textual geological information using artificial intelligence methods and graph theory are proposed at the SGM RAS*



### Виртуальные ассистенты

Цифровой диалоговый ассистент (ЦДА) — это коллега будущего. Он будет уметь самостоятельно выполнять заданный процесс и взаимодействовать с разными информационными системами: например, запрашивать, обогащать или вносить данные. И общаться с ЦДА можно, как и с любым другим коллегой: голосом или текстовыми сообщениями, через почту, мессенджеры, онлайн-чаты или виджеты на сайте. Отчасти цифровые диалоговые ассистенты похожи на уже привычных чат-ботов, но с интеллектом, превосходящим начинку устройств с технологией RPA (Robotic Process Automation). Виртуальные ассистенты взяли лучшее от этих технологий.

В самое последнее время широкое развитие получили чат-боты. Большие языковые модели, такие как BERT, ChatGPT и GPT-4, достигли большого успеха в применении к общеупотребительным языковым областям, таким как новостные, социокультурные, медийные. В области наук о Земле их применение не столь впечатляющее, в первую очередь, из-за отсутствия в их арсенале специфической геологической терминологии, поскольку она отсутствует в тех языковых корпусах текста, на которых обучались данные модели [8]. В этой публикации описана большая языковая модель для наук о Земле — K2 [9] — на основе обучения GPT подобной языковой модели LLaMA. Для целей тренировки модели создан большой набор текстов, содержащий более 6 млн записей, включающих статьи, метаданные статей и данные из Wikipedia. K2 — генеративная модель. Она может создавать тексты по теме наук о Земле и отвечать на соответствующие вопросы.

В ГГМ РАН разработан интеллектуальный помощник по поиску данных и методов их обработки с постановкой задач на естественном языке. Он реализован с помощью ряда web-сервисов и в ближайшее время будет доступен на портале <https://geologyscience.ru/>.

### Технологии беспроводной связи

При горных работах для оценки поведения этих сооружений при воздействии природных процессов и явлений во время строительства и эксплуатации применяется геотехнический мониторинг: измерение смещений, трещин и других признаков развития оползней, уровней воды и т. д. на основе данных от множества датчиков — как поверхностных, так и встроенных в массив грунта (последние измеряют изменения на разной глубине от поверхности в течение более длительных периодов времени) [10]. В качестве коммуникации между датчиками используются беспроводные сенсорные сети.

В наше время цифровизация имеет тенденцию к подключению устройств беспроводных сенсорных сетей к так называемому Интернету вещей (IoT-системе взаимосвязанных устройств, которые могут собирать данные, обрабатывать их и обменивать-

ся ими между собой, с человеком и серверами) для реального мониторинга процессов. Системы WSN и IoT основаны на новейших лазерных или ультразвуковых технологиях и могут проводить измерения с точностью до нескольких миллиметров или выше.

### Технология распределенного реестра

Технология распределенного реестра — технология хранения данных, которая распределяет информацию среди множества узлов связи или между вычислительными устройствами. Такая схема распределения имеет несколько ключевых особенностей: отсутствие центрального администратора; совместное использование с синхронизацией по заданному алгоритму; децентрализованное географическое распределение копий базы данных между всеми узлами связи.

По своей сути это первая база данных, которая лишает необходимости задействовать центральный сервис, распределяет базу по всем узлам связи, возлагая на них ответственность за поддержку системы и проверку информации.

Распределительный реестр стал известен широкому кругу людей в основном благодаря его применению в блокчейне криптовалют, но так систематизировать можно любые данные, в том числе и в науках о Земле.

### Большие данные

В настоящее время обработка больших геологических данных производится преимущественно статистическими методами. Интерпретация данных происходит с помощью стандартных инструментов визуализации.

**Вычислительный узел «Многомерные методы анализа данных» ГГМ РАН** [11] позволяет выполнять обработку табличных данных различными методами анализа данных с настройкой их параметров и визуализацией результатов. Вычислительный узел включает в себя такие группы методов, как предобработка данных, описательная статистика, кластерный анализ, факторный анализ, корреляционный анализ, регрессионный анализ и др. Перечень предлагаемых пользователю методов постоянно пополняется (рис. 2).

Выбор конкретных методов для включения в вычислительный узел обуславливается ориентированностью узла, в первую очередь, на решение геологических задач. Работа с геологическими данными часто опирается на статистические методы анализа данных, проверку гипотез, поиск зависимостей и корреляций. Результаты применения этих методов позволяют сделать выводы об исследуемых данных и принять решения об использовании других методов: кластеризации, классификации и др. При этом важно иметь возможность настройки параметров методов для более детального анализа геологических данных.

Вычислительный узел состоит из следующих компонентов:

- web-сервис для получения запросов от внешних пользователей или систем на обработку данных выбранным методом с указанными параметрами, проверку статуса обработки и получения результатов работы метода;
- реестр методов, содержащий перечень методов, их параметров, а также непосредственную реализацию методов в виде скриптов на языке Python;
- очередь задач, в которую помещаются поступающие запросы на обработку данных, имеющая возможность отслеживания статуса выполнения задач;
- исполнители задач, которые при получении новой задачи из очереди обращаются к реестру методов и выполняют соответствующий скрипт;
- хранилище временных файлов, в котором для каждой задачи хранятся входные файлы и файлы результатов.

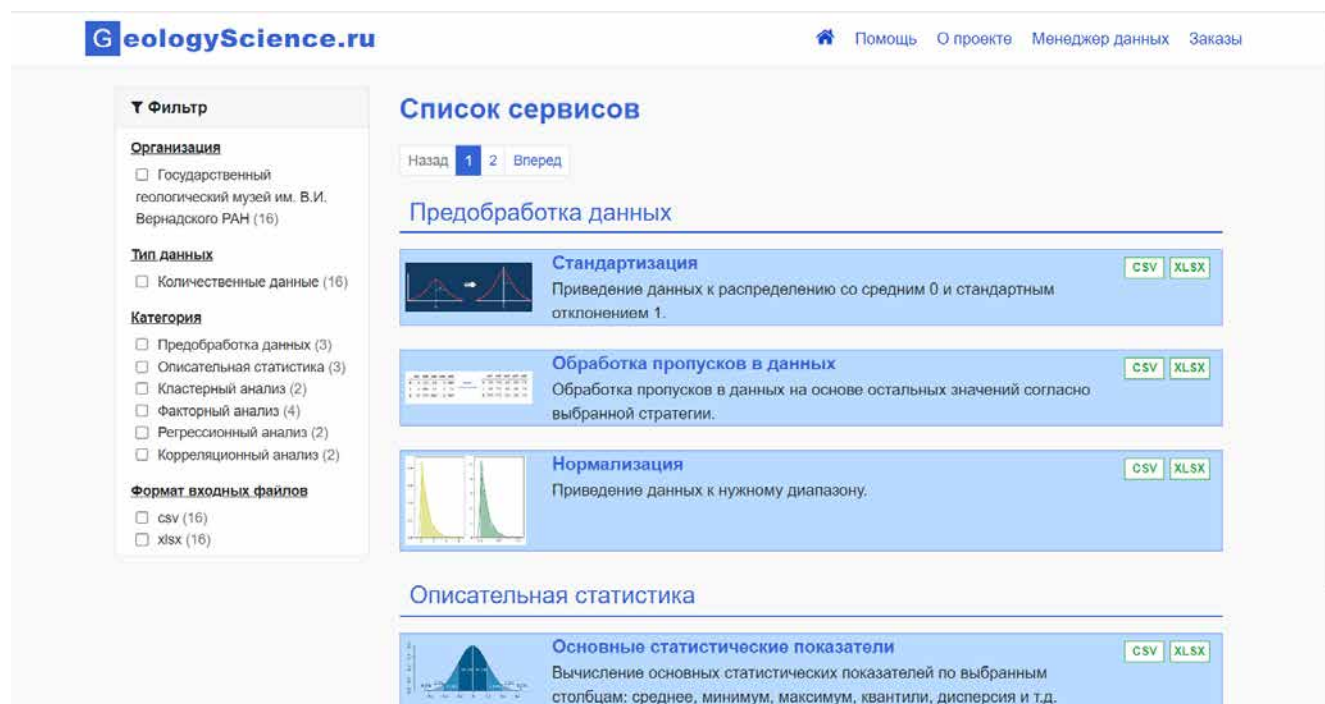
Вычисления производятся в среде Python с использованием известных открытых пакетов обработки данных: Scikit-learn, Pandas, Matplotlib и др. Обработка поступающих запросов на обработку данных производится фреймворком FastAPI посредством REST API, реализованного по стандарту OpenAPI (<http://spec.openapis.org/oas/v3.0.3>), с использованием очереди вычислительных задач, реализованной на основе NoSQL базы данных Redis.

При этом отдельные компоненты системы — web-сервис, база данных с очередью задач, исполнители задач — разворачиваются с применением контейнеров Docker. Такая архитектура позволяет разделить обработку запросов и тяжелые вычисления больших объемов данных, что обеспечивает отказоустойчивость и масштабируемость узла.

Вычислительный узел «Многомерные методы анализа данных» является самостоятельным облачным web-сервисом, что позволяет использовать его отдельно, посредством REST API, реализованного по стандарту OpenAPI 3.0. Это позволяет обращаться к многомерным методам анализа данных, размещенным на вычислительном узле, широкому кругу пользователей, в том числе интегрировать его в другие информационные системы как стороннее приложение для обработки табличных данных. Кроме того, использование стандарта OpenAPI делает узел самодокументированным, при этом при изменении списка методов документация обновляется автоматически (<https://analysis.geologyscience.ru/docs>). Согласно классификации облачных сервисов, вычислительный узел «Многомерные методы анализа данных» реализует модель «Программное обеспечение как сервис» (Software as a Service, SaaS), т. е. пользователи сервиса получают доступ к возможностям сервиса с использованием облачных технологий без необходимости устанавливать и конфигурировать программное обеспечение самостоятельно.

Рис. 2. Вычислительный узел ГГМ РАН «Многомерные методы анализа данных»

Fig. 2. Computing node "Multidimensional methods of data analysis" SGM RAS



Для решения задач обработки больших геологических данных можно использовать машинное обучение без учителя. Данный метод наилучшим образом подходит для геолого-разведочной отрасли, так как при его использовании ответы не задаются и требуется искать зависимости между объектами.

### Digital-инструменты

Анализ геологических данных — трудоемкий и сложный процесс.

Геологи могут использовать ChatGPT для автоматизации многих задач преобразования и очистки данных для лучшей организации и управления геологическими данными.

#### Современное геологическое программное обеспечение с открытым исходным кодом

(обзор CorePlan — <https://www.coreplan.io/blog/top-12-geological-databases-for-mineral-exploration-in-2024>):

**QGIS** — это надежная и удобная в использовании географическая информационная система (ГИС), пользующаяся популярностью у геологов по всему миру. Обширная библиотека плагинов делает его невероятно универсальным для широкого спектра геологических приложений.

**GemPy** — 3D геологическое моделирование с использованием Python Power. Для геологов, знакомых с Python, GemPy предлагает мощный способ создания сложных 3D геологических моделей. Подход к неявному моделированию и вероятностные особенности делают его идеальным для решения сложных геологических сценариев.

Разработанная при поддержке Geoscience Australia, Loop представляет собой комплексную платформу для построения, визуализации и анализа 3D геологических и геофизических моделей.

**Acquire GIM Suite** — программное обеспечение для управления геологическими данными, разработанное для эффективного сбора, управления геологическими данными и доступа к ним.

**Datamine Fusion** объединяет все данные по разведке и добыче в безопасную и доступную базу данных. Он поддерживает комплексное управление данными по геологии, геохимии и геотехнике.

**Plexer** — это платформа управления геологическими данными, предназначенная для централизации и управления всеми данными, относящимися к геологоразведке и добыче полезных ископаемых. Основное внимание уделяется доступу к данным в режиме реального времени и удобным интерфейсам.

**Leapfrog Geo** — это программное обеспечение для 3D геологического моделирования, разработанное Seequent. Оно предназначено для предоставления геологам интуитивно понятных и мощных инструментов для построения и интерпретации сложных геологических моделей.

Для анализа спутниковых данных одним из лидеров среди облачных сервисов является An Open Source Geospatial Data Management & Analysis Platform (<https://www.opendatacube.org>). **Open Data Cube (ODC)** — это некоммерческий проект с открытым исходным кодом, который был создан для обеспечения доступа, управления и анализа больших объемов данных ГИС по мониторингу Земли. Он представляет собой общий аналитический фреймворк, содержащий наборы структурированных данных и инструментов, которые позволяют проводить анализ больших коллекций пространственных данных. ODC был разработан для анализа пространственных данных на больших временных промежутках, однако его можно использовать на любых наборах пространственных данных. Данные могут представлять собой модели высот, геофизические сетки, интерполированные поверхности и т. д. Ключевой особенностью ODC является возможность сохранения каждого уникального элемента набора пространственных данных, в отличие от многих других методов работы с большими коллекциями пространственных данных.

На сегодняшний день наибольший интерес представляют большие языковые модели и задачи, связанные с обработкой естественного языка. Работа авторов статьи в ГТМ РАН посвящена разработке современных сервисов на основе методов искусственного интеллекта в рамках экосистемы данных и сервисов их обработки для поддержки научных исследований ГТМ РАН (<https://geologyscience.ru>) (рис. 3) [12]. Эти сервисы реализуются в форме проблемно-ориентированных чат-ботов, в основе которых лежат большие языковые модели для решения следующих задач: анализа публикаций (на примере информации Репозитория научных статей по геологии РФ ГТМ РАН — (<https://repository.geologyscience.ru>), интеллектуального поиска и анализа информации (на примере Энциклопедии месторождений России ГТМ РАН — <http://wiki.geologyscience.ru>), интеллектуального помощника по поиску данных и методов их обработки с постановкой задач на естественном языке. Вышеперечисленные сервисы являются самостоятельными веб-сервисами, имеющими возможности по их интеграции в вычислительные среды и информационные системы.

В настоящее время ИИ чат-боты, например ChatGPT от AI, умеют писать простой программный код, что очень важно для научных исследований, в том числе и для наук о Земле.

Что может ChatGPT в этом направлении: создавать техническую документацию; оптимизировать код, искать баги; генерировать тесты к коду; переводить с одного языка программирования на другой; осуществлять Frontend-верстку, сгенерировать API; работать с базами данных и SQL-запросами; работать с JSON-данными; нести функцию виртуального компьютера на базе ChatGPT с файловой системой Linux; разрабатывать парсеры; разрабатывать ре-

шения для Python разработчиков; работать с LaTeX-формулами [13].

Языковая модель, используемая ChatGPT, является расширенной (LLM), которая помогает оптимизировать деятельность по программированию.

### Современные цифровые системы популяризации геологических знаний

Для решения вопроса популяризации научных знаний по истории Земли в Государственном геологическом музее им. В.И. Вернадского РАН были разработаны и реализованы две различные версии интерактивного научно-популярного портала

«История Земли: геологический ракурс» [14] <https://populargeology.ru/>.

Прикладные исследования в данном направлении посвящены разработке и реализации системы управления сайтами (CMS) для создания научно-популярных порталов по геологии на примере адаптивной версии научно-популярного портала «История Земли: геологический ракурс». В качестве базовых сущностей разрабатываемого движка были выбраны и реализованы следующие: «главная страница», «статья», «галерея», «видео», «3D-Земля», «временная линия», «компонент времени» и «3D-Экспонат». В результате проделанной работы был создан научно-популярный портал, позволяю-

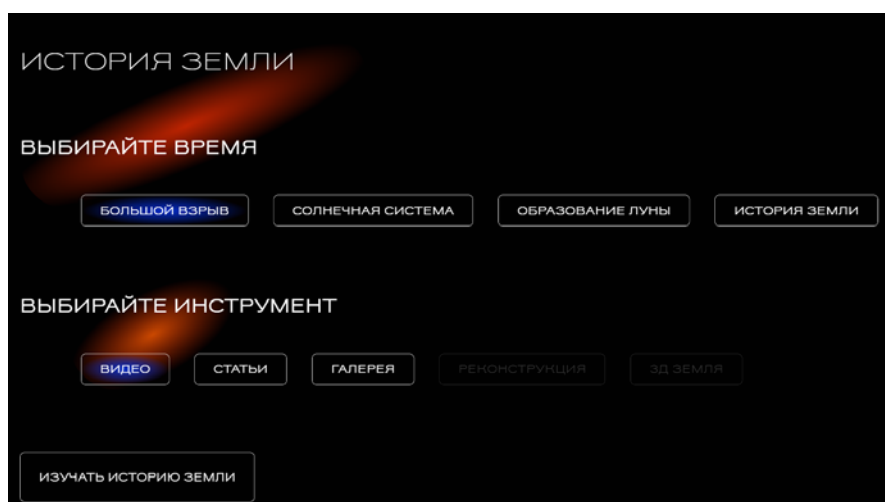
Рис. 3. Рабочее пространство <https://geologyscience.ru> [12]

Fig. 3. Workspace <https://geologyscience.ru> [12]



Рис. 4. Новая версия портала «История Земли: геологический ракурс»

Fig. 4. New version of the Portal “History of the Earth: Geological Perspective”



щий изучать нашу планету в разрезе временных интервалов, событий и артефактов (<https://new.populargeology.ru>) (рис. 4).

Данная система может быть использована при создании научно-популярных интернет-ресурсов в области геологии.

В ГГМ РАН в этом направлении проводятся следующие работы:

- проведено исследование существующих решений на тему популяризации научного знания в области геологии в интерактивной игровой форме;
- осуществлена разработка базовой концепции интерактивной игры по геологии Земли на основе накопленного научного знания при разработке научно-популярных порталов.

Популяризацию знаний о Земле в дальнейшем предлагается осуществлять в интерактивной игровой форме, а также с помощью виртуальных общих пространств. Это цифровые среды, в которых люди могут как взаимодействовать по образовательным и научным поводам, так и просто находить собесед-

ников по интересам и общаться. Их называют метавселенными [15], в которых применяются AR- и VR-технологии.

Под «метавселенной» понимается виртуальное пространство, основанное на нашей физической вселенной, позволяющее делать все то, что и в жизни, плюс то, что возможно только в виртуальной реальности, например, мгновенное перемещение в любую точку пространства и времени.

Чтобы реализовать возможность полного виртуального погружения в метавселенную, для создания трехмерного мира необходимо использовать такие передовые технологии, как виртуальная реальность, 3D-реконструкция, искусственный интеллект и Интернет вещей.

Метавселенные предлагают поистине неограниченные возможности: прогуляться с динозавром и прикоснуться к археоптериксу, окунуться в расплавленную лаву вулкана — для метавселенной нет ничего невозможного.

В качестве примера можно привести мир юрского периода в игре Back to Dinosaur Island [16].

*Работа выполняется в рамках гостемы «Цифровая платформа интеграции и анализа геологических и музейных данных», тема № 1021061009468-8-1.5.1.*

## Список источников

1. *Нозаева К.* Управляемые вирусы и бум нейросетей: названы самые перспективные технологии будущего [Электронный ресурс] // Hi-Tech Mail. – 11.07.2023. – Режим доступа: <https://hi-tech.mail.ru/review/101274-perspektivnyye-tehnologii-budushchego> (дата обращения: 29.07.2024).
2. *Патук М.И., Наумова В.В.* Методы искусственного интеллекта для научных исследований в геологии // Электронные библиотеки. – 2023. – Т. 26. – № 5. – С. 673–696. DOI: 10.26907/1562-5419-2023-26-5-673-696.
3. *Блокчейн-решения для бизнеса и государства* [Электронный ресурс] / Web3 Tech. – 2019–2024. – Режим доступа: <https://web3tech.ru> (дата обращения: 30.07.2024).
4. *Consulting & analytics* [Электронный ресурс] / ALS GoldSpot. – 2024. – Режим доступа: <https://www.alsglobal.com/en/consulting-and-analytics> (дата обращения 30.07.2024).
5. *Baraboshkin E.E., Ismailova L.S., Orlov D.M., Zhukovskaya E.A., Kalmykov G.A., Khotylev O.V., Baraboshkin E.Yu., Koroteev D.A.* Deep convolutions for in-depth automated rock typing // Computers & Geosciences. – 2020. – Vol. 135. – 104330. DOI: 10.1016/j.cageo.2019.104330.
6. *Nesteruk S., Agafonova J., Pavlov I., Gerasimov M., Latyshev N., Dimitrov D., Kuznetsov A., Kadurin A., Plechov P.* MinerallImage5k: A benchmark for zero-shot raw mineral visual recognition and description // Computers & Geosciences. – 2023. – Vol. 178. – 105414. DOI: 10.1016/j.cageo.2023.105414.
7. *Патук М.И., Наумова В.В.* Семантический поиск и ранжирование геологических публикаций для научных исследований // Марчуровские научные чтения 2024 : материалы международной конференции (7–11 октября 2024 г., Новосибирск). – В печати.
8. *Deng C., Zhang T., He Z., Chen Q., Shi Y., Zhou L., Fu L., Zhang W., Wang X., Zhou C., Lin Z., He J.* Learning a foundation language models for geoscience knowledge understanding and utilization [Электронный ресурс]. – 2023. – arXiv:2306.05064 [cs.CL]. – Режим доступа: <https://arxiv.org/abs/2306.05064v1> (дата обращения: 30.07.2024 г.). DOI: 10.48550/arXiv.2306.05064.
9. *Deng C., Zhang T., He Z., Chen Q., Shi Y., Zhou L., Fu L., Zhang W., Wang X., Lin Z., He J.* K2 (GeoLLaMA) Large Language Model for Geoscience [Электронный ресурс] // GitHub. – 2023. – Режим доступа: <https://github.com/davendw49/k2?ysclid=lmwxywt6i750905070> (дата обращения 30.07.2024).
10. *Hristova T., Mitev I., Balev V.* Monitoring of geotechnical facilities through DLT solution // IOP Conference Series: Earth and Environmental Science. – 2021. – Vol. 970. III International Conference "Essays of Mining Science and Practice" (6–8 October 2021, Dnipro). – 012011. DOI: 10.1088/1755-1315/970/1/012011.

11. Загуменнов А.А., Наумова В.В., Еременко В.С. Облачный сервис многомерной обработки количественных данных для решения геологических задач // Вестник НГУ. Серия: Информационные технологии. – 2021. – Т.19. – № 3. – С. 40–49. DOI: 10.25205/1818-7900-2021-19-3-40-49.
12. Загуменнов А.А., Наумова В.В. Сервисы на основе методов искусственного интеллекта для поддержки научных исследований в геологии // Марчуковские научные чтения 2024 : материалы международной конференции (7–11 октября 2024 г., Новосибирск). – В печати.
13. Воловик К. 20 примеров использования ChatGPT — что умеет популярная нейросеть [Электронный ресурс] // Партнеркин. – 26.01.2023. – Режим доступа: <https://partnerkin.com/blog/articles/primery-ispolzovaniya-chatgpt?ysclid=lyze7cy2hu77461277> (дата обращения 30.07.2024).
14. Еременко А.С., Гринев М.В., Одновил Е.А. Разработка системы управления сайтами для создания научно-популярных порталов по геологии (на примере портала «История Земли: геологический ракурс») // Электронные библиотеки. – 2023. – Т. 25. – № 5. – С. 600–628. DOI: 10.26907/1562-5419-2023-26-5-600-628.
15. Панин В.О. Виртуальные метавселенные как этап развития информационных технологий // Материалы международной конференции Российского национального комитета по истории и философии науки и техники РАН, посвященной 90-летию Института истории естествознания и техники им. С.И. Вавилова РАН (28 марта – 1 апреля 2022 г.). – М. : ИИЕТ РАН, 2022. – С. 478–481.
16. Виртуальная реальность и динозавры. Игра Back to Dinosaur Island (DEMO trailer) [Электронный ресурс]. – 2015. Режим доступа: <https://www.youtube.com/watch?v=M6Bj6u2UC70> (дата обращения 30.07.2024).

## References

1. Nogaeva K. Upravlyayemye virusy i bum neirosetei: nazvany samye perspektivnyye tekhnologii budushchego [Controlled viruses and the boom of neural networks: the most promising technologies of the future have been named]. *Hi-Tech Mail*. 11.07.2023. Available at: <https://hi-tech.mail.ru/review/101274-perspektivnyye-tekhnologii-budushchego> (accessed 29.07.2024).
2. Patuk M.I., Naumova V.V. Artificial intelligence methods for scientific research in geology. *Russian digital libraries journal*. 2023;26(5):673–696. DOI: 10.26907/1562-5419-2023-26-5-673-696.
3. Enterprise-grade Blockchain for E-government and Businesses. 2019–2024. Available at: <https://web3tech.ru/en/> (accessed 30.07.2024).
4. Consulting & analytics by ALS GoldSpot. 2024. Available at: <https://www.alsglobal.com/en/consulting-and-analytics> (accessed 30.07.2024).
5. Baraboshkin E.E., Ismailova L.S., Orlov D.M., Zhukovskaya E.A., Kalmykov G.A., Khotylev O.V., Baraboshkin E.Yu., Koroteev D.A. Deep convolutions for in-depth automated rock typing. *Computers & Geosciences*. 2020;135:104330. DOI: 10.1016/j.cageo.2019.104330.
6. Nesteruk S., Agafonova J., Pavlov I., Gerasimov M., Latyshev N., Dimitrov D., Kuznetsov A., Kadurin A., Plechov P. MineralImage5k: A benchmark for zero-shot raw mineral visual recognition and description. *Computers & Geosciences*. 2023;178:105414. DOI: 10.1016/j.cageo.2023.105414.
7. Patuk M.I., Naumova V.V. Semanticheskii poisk i ranzhirovanie geologicheskikh publikatsii dlya nauchnykh issledovaniy [Semantic search and ranking of geological publications for scientific research]. In: Marchukovskie nauchnye chteniya 2024: materialy mezhdunarodnoi konferentsii (7–11 October 2024, Novosibirsk). In print.
8. Deng C., Zhang T., He Z., Chen Q., Shi Y., Zhou L., Fu L., Zhang W., Wang X., Zhou C., Lin Z., He J. Learning a foundation language models for geoscience knowledge understanding and utilization. 2023. arXiv:2306.05064 [cs.CL]. Available at: <https://arxiv.org/abs/2306.05064v1> (accessed 30.07.2024). DOI: 10.48550/arXiv.2306.05064.
9. Deng C., Zhang T., He Z., Chen Q., Shi Y., Zhou L., Fu L., Zhang W., Wang X., Lin Z., He J. K2 (GeoLLaMA) Large Language Model for Geoscience. *GitHub*. 2023. Available at: <https://github.com/davendw49/k2?ysclid=lmswxywt6i750905070> (accessed 30.07.2024).
10. Hristova T., Mitev I., Balev V. Monitoring of geotechnical facilities through DLT solution. *IOP Conference Series: Earth and Environmental Science*. 2021;970:012011. DOI: 10.1088/1755-1315/970/1/012011.
11. Zagumenov A.A., Naumova V.V., Eremenko V.S. Cloud service for multidimensional processing of quantitative data for solving geological problems. *Vestnik NSU. Series: Information Technologies*. 2021;19(3):40–49. DOI: 10.25205/1818-7900-2021-19-3-40-49.
12. Zagumenov A.A., Naumova V.V. Servisy na osnove metodov iskusstvennogo intellekta dlya podderzhki nauchnykh issledovaniy v geologii [AI-based services to support scientific research in geology]. In: Marchukovskie nauchnye chteniya 2024: materialy mezhdunarodnoi konferentsii (7–11 October 2024, Novosibirsk). In print.
13. Volovik K. 20 primerov ispol'zovaniya ChatGPT — chto umeet populyarnaya neiroset' [20 examples of using ChatGPT — what the popular neural network can do]. *Partnerkin*. 26.01.2023. Available at: <https://partnerkin.com/blog/articles/primery-ispolzovaniya-chatgpt?ysclid=lyze7cy2hu77461277> (accessed 30.07.2024).
14. Eremenko A.S., Grinev M.V., Odnovil E.A. Implementation of an engine for creating popular science portals on geology (using the example of the portal “History of the Earth: geological perspective”). *Russian digital libraries journal*. 2023;25(5):600–628.
15. Panin V.O. Virtual metaverse as a stage in the development of information technologies. In: Proceedings of the International Conference of Russian National Committee for the History and Philosophy of Science and Technology of the Russian Academy of Sciences, devoted to the 90th anniversary of S.I. Vavilov Institute for the History of Science and Technology of the Russian Academy of Sciences. Moscow: IIET RAN; 2022. pp. 478–481.
16. Virtual'naya real'nost' i dinovavry. Igra Back to Dinosaur Island (DEMO trailer) [Virtual Reality and Dinosaurs. Game Back to Dinosaur Island (DEMO trailer)]. 2015. Available at: <https://www.youtube.com/watch?v=M6Bj6u2UC70> (accessed 30.07.2024).

Статья поступила в редакцию 29.07.2024 г., одобрена после рецензирования 31.07.2024 г., принята к публикации 01.08.2024 г.  
The article was submitted 29.07.2024; approved after reviewing 31.07.2024; accepted for publication 01.08.2024.

## Информация об авторах

### Наумова Вера Викторовна

Доктор геолого-минералогических наук  
 Главный научный сотрудник, зав. Научным отделом  
 Государственный геологический музей им. В.И. Вернадского РАН  
 125009 Москва, Моховая ул., д. 11, стр. 11  
 e-mail: naumova\_new@mail.ru  
 ORCID: 0000-0002-3001-1638  
 AuthorID: 1968  
 SPIN: 3919-4475

### Патук Михаил Иванович

Кандидат геолого-минералогических наук  
 Научный сотрудник Научного отдела  
 Государственный геологический музей им. В.И. Вернадского РАН  
 125009 Москва, Моховая ул., д. 11, стр. 11  
 e-mail: patuk@mail.ru  
 ORCID: 0000-0003-3036-2275

### Ерёменко Александр Сергеевич

Кандидат технических наук  
 Программист (внештатный сотрудник)  
 Государственный геологический музей им. В.И. Вернадского РАН  
 125009 Москва, Моховая ул., д. 11, стр. 11  
 Старший научный сотрудник  
 Институт автоматизации и процессов управления ДВО РАН  
 690041 Владивосток, ул. Радио, д. 5  
 e-mail: academy21@gmail.com  
 ORCID: 0000-0003-1923-8417

### Загумёнов Алексей Андреевич

Программист (внештатный сотрудник)  
 Государственный геологический музей им. В.И. Вернадского РАН  
 125009 Россия, Москва, Моховая ул., д. 11, стр. 11  
 Младший научный сотрудник  
 Институт автоматизации и процессов управления ДВО РАН  
 690041 Владивосток, ул. Радио, д. 5  
 e-mail: truepikvic@gmail.com  
 ORCID: 0000-0002-0501-5362  
 AuthorID: 641587  
 SPIN: 2196-4572

### Ерёменко Виталий Сергеевич

Научный сотрудник Научного отдела  
 Государственный геологический музей им. В.И. Вернадского РАН  
 125009 Москва, Моховая ул., д. 11, стр. 11  
 e-mail: vitaer@gmail.com  
 ORCID: 0000-0002-5250-5743  
 AuthorID: 984069  
 SPIN: 3939-5998

## Information about authors

### Vera V. Naumova

Doctor of Geological and Mineralogical Sciences,  
 Chief Researcher, Head of Scientific Department  
 Vernadsky State Geological Museum RAS  
 11/11, Mokhovaya str., Moscow, 125009, Russia  
 e-mail: naumova\_new@mail.ru  
 ORCID: 0000-0002-3001-1638  
 AuthorID: 1968  
 SPIN: 3919-4475

### Michail I. Patuk

Candidate of Geological and Mineralogical Sciences  
 Researcher of SGM Scientific Department  
 Vernadsky State Geological Museum RAS  
 11/11, Mokhovaya str., Moscow, 125009, Russia  
 e-mail: patuk@mail.ru  
 ORCID: 0000-0003-3036-2275

### Alexander S. Eremenko

Candidate of Technical Sciences  
 Contract programmer of SGM Scientific Department  
 Vernadsky State Geological Museum RAS  
 11/11, Mokhovaya str., Moscow, 125009, Russia  
 Senior Researcher  
 Institute of automatic and control processes of FEB RAS  
 5, Radio str., Vladivostok, 690041, Russia  
 e-mail: academy21@gmail.com  
 ORCID: 0000-0003-1923-8417

### Aleksei A. Zagumennov

Contract programmer of SGM Scientific Department  
 Vernadsky State Geological Museum RAS  
 11/11, Mokhovaya str., Moscow, 125009, Russia  
 Junior Researcher  
 Institute of Automation and Control Processes of FEB RAS  
 5, Radio str., Vladivostok, 690041, Russia  
 e-mail: truepikvic@gmail.com  
 ORCID: 0000-0002-0501-5362  
 AuthorID: 641587  
 SPIN: 2196-4572

### Vitaliy S. Eremenko

Researcher of SGM Scientific Department  
 Vernadsky State Geological Museum RAS  
 11/11, Mokhovaya str., Moscow, 125009, Russia  
 e-mail: vitaer@gmail.com  
 ORCID: 0000-0002-5250-5743  
 AuthorID: 984069  
 SPIN: 3939-5998