

ПОСТРОЕНИЕ ЦИФРОВОЙ СИСТЕМЫ УПРАВЛЕНИЯ ГЕОЛОГИЧЕСКИМИ ЗНАНИЯМИ ДЛЯ ПОДДЕРЖКИ НАУЧНЫХ ИССЛЕДОВАНИЙ

М. И. Патук¹, [0000-0003-3036-2275], В. В. Наумова², [0000-0002-3001-1638]

^{1, 2} ФГБУН Государственный геологический музей им. В.И. Вернадского РАН
Москва, Россия

¹patuk@mail.ru, ²naumova_new@mail.ru

Аннотация

Описаны новые подходы к сбору данных о научных публикациях из систем открытого доступа с тематикой «Науки о земле». На основе разработанных и адаптированных подходов созданы архив научных публикаций (репозиторий) и комплекс программ доступа к научным публикациям для сбора, поиска, фильтрации, каталогизации и управления публикациями и их метаданными. Для улучшения доступности публикаций и других связанных с ними данных, находящихся на сайтах Государственного геологического музея им. В.И. Вернадского РАН, разработана система Wiki – Геология России. Эта система является тематическим рубрикатором по направлению «Месторождения полезных ископаемых России», с дополнительной тематикой «Минералогия». Все статьи имеют ссылку на источник информации из архива научных публикаций и, опционально, дополнительные ссылки по сходной тематике. Wiki – Геология России являются первым шагом в создании базы знаний по месторождениям полезных ископаемых.

Ключевые слова: Wiki – Геология России, системы управления знаниями, репозиторий

В Государственном геологическом музее (ГГМ) им. В.И. Вернадского РАН накоплено большое количество открытых данных по наукам о земле. Эти данные, в основном, находятся на двух сайтах – Геология России (<http://geologyscience.ru/>) [1] и Портал открытых данных ГГМ РАН (<http://data.sgm.ru/>) [2]. Большой объем данных имеет свою негативную сторону – с увеличением объема растут и затраты на поиск нужной информации. Один из путей для повышения доступности информации – использовать тематическую группировку данных [3]. Такие тематически

сгруппированные данные, представленные в сжатой форме, могут служить объединяющим началом, синтезом разнородной информации, позволяющим формировать новое знание. С другой стороны, предлагаемый подход можно рассматривать как систему «легкого доступа» в электронную библиотеку и связанные с ней, возможно в неявной форме, данные в виде карт, фото и другой текстовой информации. Такой подход частично перекликается с развивающимися в последнее время системами управления знаниями, использующими в качестве базы знаний накопленный опыт организации и, в частности, научно-технические библиотеки [4].

В качестве платформы такой группировки нами была выбрана технология *wiki* – технология построения веб-систем, предназначенных для коллективной разработки, хранения, структуризации текста, гипертекста, файлов, мультимедиа. В качестве программной платформы была выбрана *MediaWiki* как самая популярная, свободно распространяемая платформа. В настоящее время версия этой платформы – 1.36.1 [5]. Созданный сайт (*wiki* – Геология России) расположен по адресу <http://wiki.geologyscience.ru>.

В интернете имеется огромное число сайтов, созданных с помощью названной или подобных ей платформ. Самой известной, безусловно, является Википедия. Принятый в ней принцип энциклопедичности принес ей заслуженную популярность. Однако не все статьи в ней одинаково равнозначны; имеет место неавторитетность и ненадежность информации, зачастую отсутствуют ссылки на первоисточник.

Более близкий нам по своему подходу, отношению к надежности информации и указанию её источников является сайт *Геовикипедия* (<https://wiki.web.ru>), созданный на Геологическом факультете Московского государственного университета им. М.В. Ломоносова. Тематический каталог (научный рубрикатор) сайта показывает широкий охват терминологии предметной области. Авторы проекта акцентируют внимание прежде всего на всестороннем представлении научной терминологии и корректном определении терминов. Можно сказать, что статьи в *Геовикипедии* являются конечным пунктом в определении геологических терминов. Мы же в своем подходе хотим сделать такие статьи отправной точкой в построении базы знаний по выбранным тематикам.

В качестве исходных тематик для группировки данных были выбраны «Месторождения полезных ископаемых» и «Минералогия». Естественно, что эти тематики не покрывают всего объема данных, имеющихся в распоряжении на наших сайтах. Их можно рассматривать в качестве теста для отработки технологии группировки информации. Основной тематикой в нашем подходе является «Месторождения полезных ископаемых», а «Минералогия» служит уточняющим и изобразительным дополнением к основной тематике, по крайней мере в текущей реализации системы.

Информация в MediaWiki хранится в виде статей с заголовками. Заголовками в нашем случае служат наименования месторождений и минералов. Как отмечалось [6], главным недостатком Wiki-технологии является отсутствие явно выраженной и удобно представленной структуры. При увеличении объема информации в системе это будет негативно влиять на ее доступность. Одним из возможных подходов к решению этой проблемы является добавление семантической информации в статьи, публикуемые в Wiki. Такой информацией является система категорий MediaWiki, которая позволила организовать алфавитный доступ к данным (вкладка «Месторождения по алфавиту») (рис. 1).

Кроме алфавитного доступа по наименованию месторождений, аналогичный подход был использован для организации доступа по полезным ископаемым (вкладка «Месторождения по полезным ископаемым»). Для обеспечения этого функционала также были использованы тэги категорий (например – «Полезные ископаемые–Железо»). Использование возможностей категорий в MediaWiki позволило осуществить доступ к информации по алфавиту (по наименованиям страниц) и извлекаемым полезным ископаемым (по наименованию полезных ископаемых).

Вся информация, вносимая в «Wiki – Геология России», берется из научных статей либо из монографий с обязательным указанием авторов и ссылки на статью или монографию. Сами первоисточники располагаются в архиве научных публикаций ГГМ им. В.И. Вернадского (<https://repository.geologyscience.ru/>) [7]. По ссылкам на первоисточник можно перейти на сайт архива научных публикаций и ознакомиться с полным текстом источника информации.

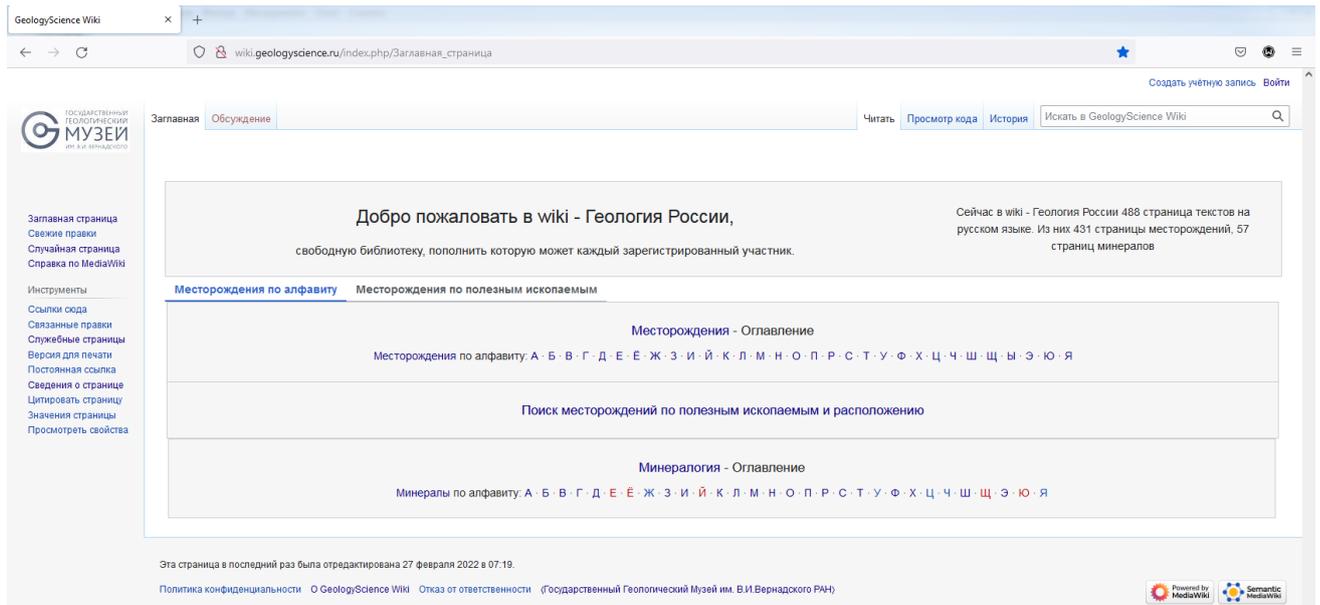


Рис. 1. Главная страница «Wiki-Геология России»

Кроме ссылки на первоисточник на странице месторождения находятся ссылки на описание месторождения в Государственном кадастре месторождений (<https://www.rfgf.ru>) и паспорт месторождения (данные находятся на сайте <http://geologyscience.ru/>) (рис. 2).

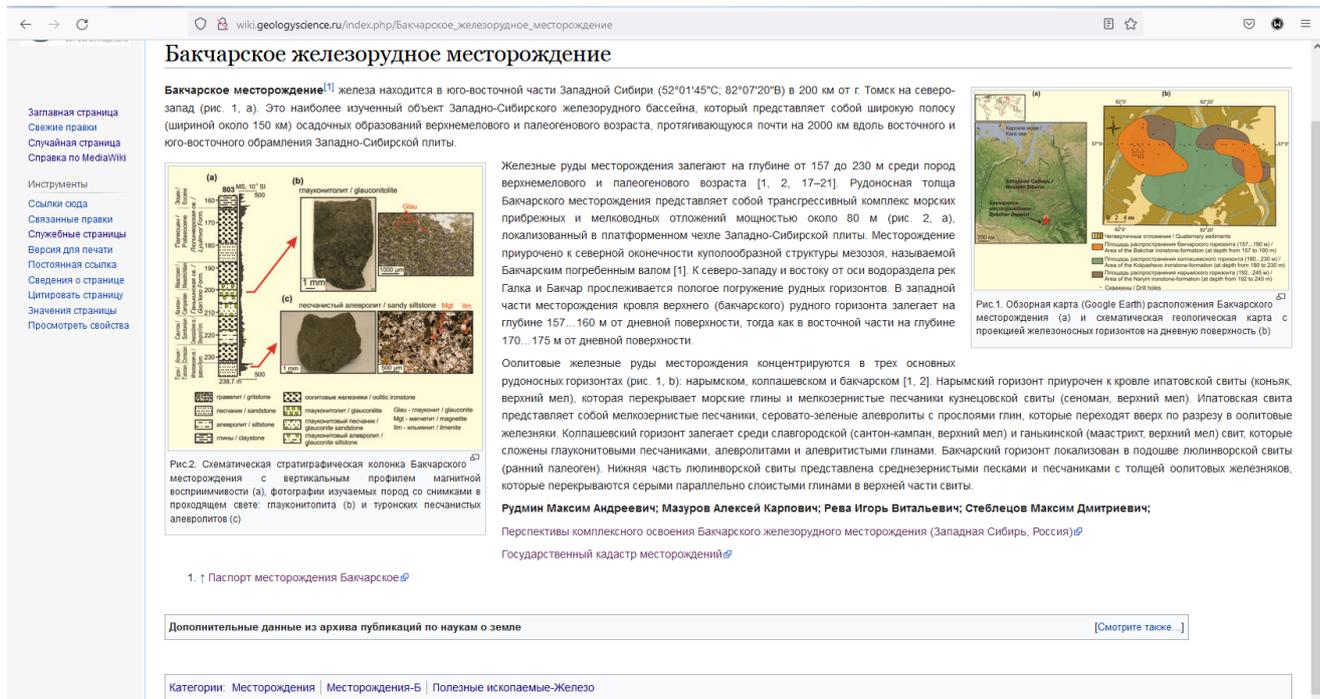
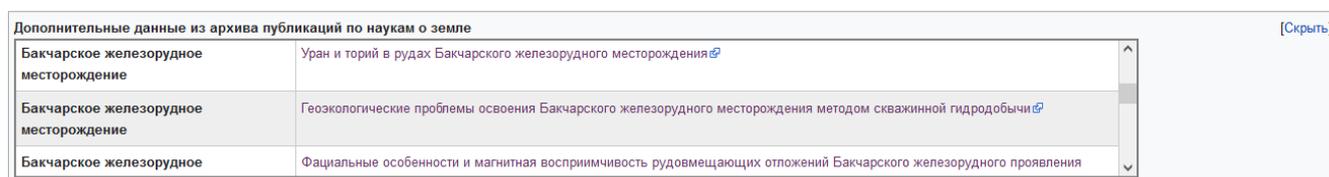


Рис. 2. Страница с описанием месторождения полезного ископаемого

Как правило, данные по конкретному месторождению не исчерпываются одной публикацией. Получить доступ к дополнительной информации можно по ссылке «Дополнительные данные из архива публикаций по наукам о земле» (рис. 3).



Дополнительные данные из архива публикаций по наукам о земле		[Скрыть]
Бакcharское железорудное месторождение	Уран и торий в рудах Бакcharского железорудного месторождения	
Бакcharское железорудное месторождение	Геоэкологические проблемы освоения Бакcharского железорудного месторождения методом скважинной гидродобычи	
Бакcharское железорудное	Фациальные особенности и магнитная восприимчивость рудовмещающих отложений Бакcharского железорудного проявления	

Рис. 3. Окно «Дополнительные данные из архива публикаций по наукам о земле»

Для извлечения дополнительных ссылок на публикации используется расширение MediaWiki – ExternalData – для доступа к базе данных репозитория, работающего на СУБД PostgreSQL. Хранение ссылок на публикации в базе MediaWiki (MySQL) достигается с использованием расширения Cargo.

Увеличение числа страниц месторождений полезных ископаемых выявило необходимость создания отдельного модуля поиска месторождений, с использованием информации об их территориальном расположении. Для реализации такой возможности было установлено дополнительное расширение MediaWiki – Semantic MediaWiki [8]. Это расширение позволяет наряду с категориями добавлять на страницы различные свойства и выполнять семантические запросы на основе внесенной информации. С помощью этого инструментария был реализован блок «Поиск месторождений по полезным ископаемым и их расположению».

Выбор месторождений происходит на основе одного или двух значений полезных ископаемых и территориального размещения (края, области, республики). Можно заполнить как одно поле (т. е. осуществлять поиск только по полезному ископаемому или территории), так и все три поля (рис. 4).

В разделе «Минералогия» представлены краткие описания конкретных минералов с фотографией указанного минерала из Портала открытых данных ГГМ РАН (<http://data.sgm.ru/>) или Портала Минералогического музея им. А.Е. Ферсмана РАН (<https://fmm.ru>). По ссылке под фотографией можно перейти на сайт источника и ознакомиться с более полной информацией по конкретному образцу

минерала. На странице минерала, как и на странице месторождения, можно получить доступ к дополнительной информации о выбранном минерале по ссылке «Дополнительные данные из архива публикаций по наукам о земле». Дополнительно указаны ссылки на данный минерал на 3-х популярных ресурсах – Wikipedia, энциклопедии GeoWiki, базе catalogmineralov.ru (рис. 5).

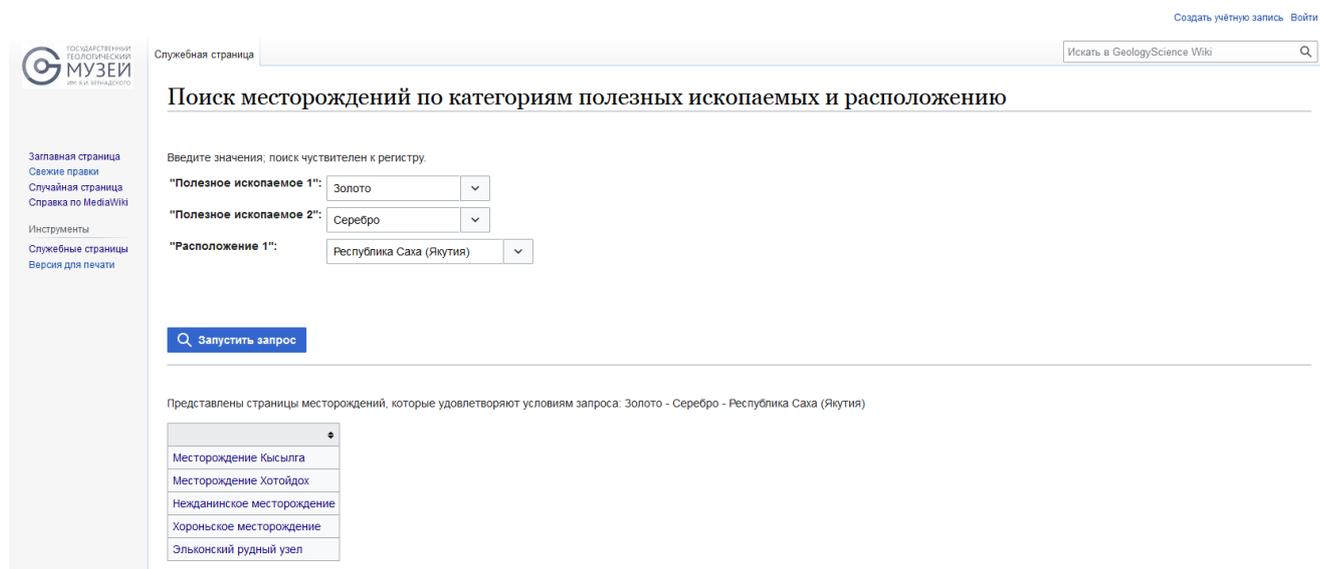


Рисунок 4. Страница поиска месторождений полезных ископаемых.

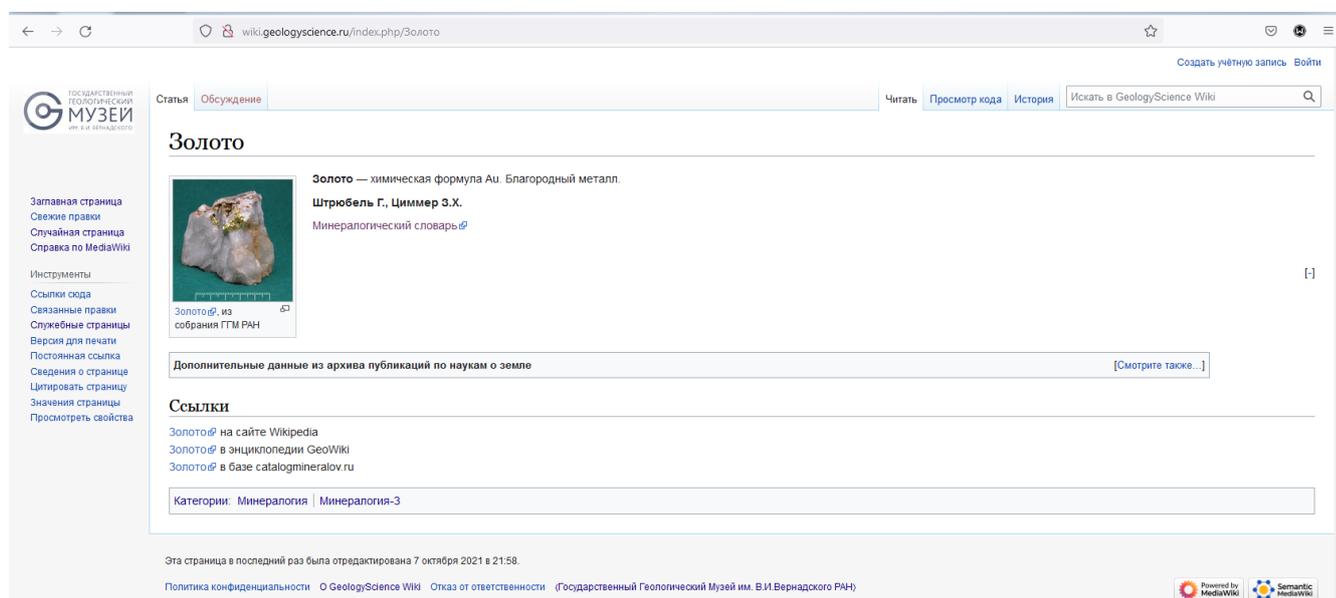


Рисунок 5. Страница с описанием минерала

На страницах описаний месторождений используются внутренние ссылки на другие страницы месторождений и страницы минералов. Они отображаются синим цветом.

Таким образом, создана первая версия сайта «Wiki – Геология России», содержащая 483 страницы описаний месторождений и 57 страниц описаний минералов. Каждая страница содержит ссылку на источник информации. Кроме обязательной ссылки страница может содержать перечень дополнительной литературы, относящейся к выбранному объекту. Ссылки ведут на репозиторий научных публикаций ГГМ им. В.И. Вернадского РАН. (<https://repository.geologyscience.ru/>). Кроме этого, присутствуют ссылки на паспорт месторождения и государственный кадастр месторождения, расположенные на сайте (<http://geologyscience.ru/>). Дополнительно указаны категории полезных ископаемых и территориальное размещение месторождений. Достигнутое объединение разнородной информации, связанной одной тематикой (месторождения полезных ископаемых), делает этот сайт первым шагом к созданию базы знаний в выбранном направлении.

Работы выполняются в рамках Государственного задания ГГМ РАН по теме № 0140-2019-0005 «Разработка информационной среды интеграции данных естественнонаучных музеев и сервисов их обработки для наук о Земле», а также темы государственного задания № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музейных данных».

СПИСОК ЛИТЕРАТУРЫ

1. *Naumova V.V., Eremenko V.S., Platonov K.A., Dyakov S.E., Patuk M.I., Eremenko A.S.* Development of geographically distributed information-analytical geological environment // *Russian Journal of Earth Sciences*. 2019. Vol. 19, ES6012. <https://doi.org/10.2205/2019ES000696>, 2019
2. *Naumova V.V., Platonov K.V., Dyakov S.E., Eremenko V.S., Starodubtseva I.A.* Basic principles of development of open access to data of the Vernadsky State Geological Museum of RAS // *Information Technologies in Earth Sciences and Applications for Geology, Mining and Economy. ITES&MP-2019: Proceedings of the V International Conference, Moscow (Russia), 14–18 October 2019*. Moscow: VNIIGeosystem, 2019. P. 31.

3. Ахмадеева И.Р., Загорулько Ю.А., Саломатина Н.В., Серый А.С., Сидорова Е.А., Шестаков В.К. Подход к формированию тематических коллекций текстов на основе интернет-ресурсов // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, Вып. 4. С. 5–15.

4. Жмайло С.В., Ульянин О.В. Научно-техническая библиотека как составная часть системы управления знаниями организации: взгляд информационного работника // Научные и технические библиотеки. 2020. № 2. С. 9–23. <https://doi.org/10.33186/1027-3689-2020-2-9-23>

5. MediaWiki. <https://www.mediawiki.org/wiki/MediaWiki>

6. Шестаков В.К. Разработка и сопровождение информационных систем, базирующихся на онтологии и Wiki-технологии // Труды 13-й Всерос. науч. конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2011, Воронеж, Россия, 2011. С. 299–306.

7. Патук М.И., Наумова В.В., Еременко В.С. Цифровой репозиторий "GeologyScience.ru": открытый доступ к научным публикациям по геологии России // Электронные библиотеки. 2020. Т. 23, №6. С. 1324–1338. <https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

8. Semantic MediaWiki. https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

BUILDING A DIGITAL GEOLOGICAL KNOWLEDGE MANAGEMENT SYSTEM TO SUPPORT SCIENTIFIC RESEARCH

M. I. Patuk¹, V. V. Naumova²

^{1, 25} *State Geological Museum named after Vladimir Vernadsky, Moscow*

¹patuk@mail.ru, ²naumova_new@mail.ru

Abstract

The paper describes new approaches to collecting data on scientific publications from open access systems with the subject of Earth Science. Based on the developed and adapted approaches, an archive of scientific publications (repository) and a set of programs for accessing scientific publications for collecting, searching, filtering, cataloging and managing publications and their metadata have been created. In order to improve the availability of publications and other related data on the websites of the SGM RAS, the Wiki – Geology of Russia system has been developed. This system is a thematic rubric in the direction of "Mineral deposits of Russia", with an additional topic "Mineralogy". All articles must have a link to the source of information from the archive of scientific publications and, optionally, additional links on similar topics. Wiki – Geology of Russia is the first step in creating a knowledge base on mineral deposits.

Keywords: Wiki – Geology of Russia, knowledge management systems, repository.

REFERENCES

1. Naumova V.V., Eremenko V.S., Platonov K.A., Dyakov S.E., Patuk M.I., Eremenko A.S. Development of geographically distributed information-analytical geological environment // Russian Journal of Earth Sciences. 2019. Vol. 19, ES6012. <https://doi.org/10.2205/2019ES000696>, 2019
2. Naumova V.V., Platonov K.V., Dyakov S. E., Eremenko V.S., Starodubtseva I.A. Basic principles of development of open access to data of the Vernadsky State Geological Museum of RAS // Information Technologies in Earth Sciences and Applications for Geology, Mining and Economy. ITES&MP-2019: Proceedings of the V International Conference, Moscow (Russia), 14–18 October 2019. Moscow: VNIIGeosystem, 2019. P. 31.

3. *Ahmadeeva I.R., Zagorul'ko Yu.A., Salomatina N.V., Seryj A.S., Sidorova E.A., Shestakov V.K.* Podhod k formirovaniyu tematicheskikh kollekcij tekstov na osnove internet-resursov // Vestn. Novosib. gos. un-ta. Seriya: Informacionnye tekhnologii. 2013. T. 11, vyp. 4. S. 5–15.

4. *Zhmajlo S.V., Ul'yanin O.V.* Nauchno-tekhnicheskaya biblioteka kak sostavnaya chast' sistemy upravleniya znaniyami organizacii: vzglyad informacionnogo rabotnika, Nauchnye i tekhnicheskie biblioteki. 2020, № 2. S. 9–23.

<https://doi.org/10.33186/1027-3689-2020-2-9-23>

5. MediaWiki. <https://www.mediawiki.org/wiki/MediaWiki>

6. *Shestakov V.K.* Razrabotka i soprovozhdenie informacionnyh sistem, baziruyushchih na ontologii i Wiki-tekhnologii, Trudy 13j Vserossijskoj nauchnoj konferencii "Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kollekcii" – RCDL'2011, Voronezh, Rossiya, 2011. S. 299–306.

7. Patuk M.I., Naumova V.V., Eryomenko V.S. (2020). Cifrovoy repozitorij "geologyscience.ru": otkrytyj dostup k nauchnym publikacijam po geologii Rossii // Elektronnye biblioteki. 2020. T. 23, No. 6. S. 1324–1338.

<https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

8. Semantic MediaWiki. https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki

СВЕДЕНИЯ ОБ АВТОРАХ



ПАТУК Михаил Иванович – к. г.-м. н., и. о. н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail I. PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

Email: patuk@mail.ru

ORCID: 0000-0003-3036-2275



НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera V. NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

Email: naumova_new@mail.ru,

ORCID: 0000-0002-3001-1638

Материал поступил в редакцию 6 апреля 2022 года