

УДК 004:504.45:504.738

Практические аспекты цифровизации лимнологических и палеолимнологических исследований

Гулаков В.О., Масленникова А.В.

Южно-Уральский федеральный научный центр минералогии и геоэкологии УрО РАН

Гулаков Василий Олегович, инженер, Южно-Уральский федеральный научный центр минералогии и геоэкологии УрО РАН, vasiliygulakov@yandex.ru — ответственный за переписку.

Масленникова Анна Валерьевна, к.г.-м.н., научный сотрудник, Южно-Уральский федеральный научный центр минералогии и геоэкологии УрО РАН, adenophora@inbox.ru, <https://orcid.org/0000-0003-0434-4372>

Аннотация

Статья посвящена описанию программных продуктов, созданных с помощью информационных технологий для исследования озер Урала. База данных «LaDa», содержит информацию о 107 озерах Урала и позволяет проводить многовариантный поиск, используя морфометрические параметры и данные диатомового, гидрохимического и геохимического анализа и визуализировать полученные результаты на карте. База данных-определитель диатомовых водорослей «DiatomDB» представляет собой совокупность собранных и систематизированных авторами данных, позволяющих идентифицировать вид диатомеи по набору наблюдаемых признаков, используя строгий и нестрогий поиск, а также визуализировать полученную информацию для подтверждения точности определения.

Ключевые слова: база данных, цифровизация, информационные технологии, озера, определитель диатомовых водорослей.

Работа выполнена при поддержке Российского научного фонда (грант No. 21-17-00071, <https://rscf.ru/project/21-17-00071/>).

Введение

На Урале находится большое количество озер, донные отложения которых содержат информацию об изменении палеосреды в голоцене и об истории развития озерных экосистем. Поскольку палеолимнологические и лимнологические исследования основаны на значительном объеме аналитических данных, информатизация процесса изучения озер является особенно актуальной. Для реализации этой задачи часто используются базы данных [1, 2, 3, 4, 5, 6]. К недостаткам существующих баз данных относятся их реализация в виде

таблиц Microsoft Excel [1, 3] и/или отсутствие многовариантного поиска, позволяющее оперировать результатами комплексного анализа [5, 6].

Поэтому первой задачей цифровизации палеолимнологических и лимнологических исследований озер Урала было создание базы данных на основе системы управления базами данных (СУБД), позволяющей получать результаты поиска по запросу, применять инструменты для реструктуризации данных и проводить многовариантный поиск без необходимости экспорта результатов в сторонние программы.

Кроме необходимости оперировать результатами комплексного анализа, палеолимнологические исследования включают ряд методов, наиболее трудоемкой частью которых является идентификация микрофоссилий. К ним относится диатомовый анализ, являющийся очень информативным методом изучения истории озер. В настоящее время доступны электронные ресурсы, позволяющие определять диатомовые водоросли на основе микрофотографий [7, 8]. Одним из главных недостатков данных определителей является полное отсутствие или недостаточное описание экологии и распространения видов, что для палеолимнологического исследования является решающим. Экологическую информацию можно получить на других электронных ресурсах, которые, однако, не содержат микрофотографии [9, 10]. Электронный определитель диатомей Северной Америки, кроме высококачественных фотографий, содержит подробные описания морфологии, а для некоторых видов и экологии, а также подобие ключа, основанного на главных особенностях диатомовых водорослей, таких как симметрия, наличие и тип шва [11]. Полноценного ключа, позволяющего идентифицировать диатомею на основе расширенного набора признаков нет ни в одном из найденных нами электронных ресурсов.

Поэтому второй задачей в достижении цели информатизации палеолимнологических исследований была разработка базы данных диатомовых водорослей, содержащей подробное описание морфологии, экологии, распространения вида, а также электронный диатомовый ключ, позволяющий идентифицировать диатомовую водоросль на основе выбора из большого количества признаков.

Материалы и методы

Базы данных «LaDa» и «DiatomDB» разрабатывались на основе СУБД MariaDB в среде операционной системы Linux. Области определения исследовались на основе проводимых исследований [12] и определителей диатомей [13, 14]. «LaDa» имеет древовидную структуру данных, благодаря авторской организации данных позволяющую использовать единым образом разнородную информацию о результатах анализа. Для визуализации результатов поиска использовался открытый API сервиса Яндекс.Карты. «DiatomDB», в свою очередь, реализована в виде двоичного сквозного поискового ключа, для чего авторами был

разработан специальный способ хранения определений диатомей. Управление и работа с базами осуществляется через веб-приложение. Это обеспечивает хорошую масштабируемость и полную кроссплатформенность системы.

Результаты и обсуждение

Разработанная база данных озёр Южного урала «LaDa» хранит информацию о гидрохимии озёр, результатах диатомового и силикатного анализа поверхностных слоев озерных отложений, содержания микроэлементов в воде и донных отложениях, а также морфометрические параметры озёр и климатические характеристики мест отбора проб [15]. База данных спроектирована с возможностью масштабирования для будущего увеличения количества типов результатов анализов и исследований. Это дает возможность не только надёжно хранить результаты анализов, но и ускорить их обработку и математический анализ. База данных создана с упором на масштабируемость с простым добавлением любых методов обработки результатов запросов. Проектирование схем данных с упором на масштабирование позволило упростить процесс работы с различными типами анализов и исследований, облегчить добавление новых типов данных и создать инструмент конструирования запросов к базе данных на основе результатов комплексного анализа. Также это позволило сделать базу данных компактной, что положительно отражается на её быстродействии. Кроме того, выдача результатов запросов в виде таблиц дает возможность легко экспортировать эти результаты в сторонние программные пакеты, такие как, например, Statistica. Многовариантный поиск с конструктором запросов и визуализацией получаемых результатов на карте существенно экономит время на подборку результатов анализов для статистической обработки, а также позволяет эмпирически определять закономерности с последующим статистико-математическим их обоснованием. В целом база данных «LaDa» существенно сократила время, требуемое на обработку результатов различных анализов и выявление закономерностей в седиментогенезе. В настоящее время в базе данных «LaDa» открыт доступ к конструктору запросов в базе данных (<http://paleolimnology.info/DB/construct.php>).

Разработанная база данных-определитель диатомовых водорослей «DiatomDB» [16] представляет собой организованную по авторскому алгоритму совокупность собранных и систематизированных авторами данных, позволяющих идентифицировать вид диатомовой водоросли по набору наблюдаемых признаков, используя строгий и нестрогий поиск, и визуализировать полученную информацию для подтверждения точности определения. В отличие от дихотомических определителей, на каждом шаге уменьшающих пул потенциальных результатов вдвое, электронный диатомовый ключ, включенный в базу данных, использует принцип сита, просеивая все виды по заданному набору признаков за

один шаг. База данных хранит характерный набор признаков для однозначной идентификации, а также фотографии диатомей, сделанные с помощью оптической и электронной микроскопии. В «DiatomDB» заложена неограниченная возможность масштабирования как по количеству видов, так и по количеству признаков, поэтому при последующем дополнительном внесении данных она может применяться для определения диатомей любого ареала обитания. Кроме диатомового ключа, важным отличием нашей базы данных от других электронных определителей диатомовых водорослей, является наличие не только микрофотографий, но и подробной информации о распространении и экологии диатомей. Кроме того, в диатомовом ключе реализована возможность перехода в конструктор запросов базы данных «LaDa». В конструкторе запросов можно изучить распространение и экологию вида на Южном и Среднем Урале, задав необходимые гидрохимические и морфометрические параметры и получив табличные и картографические данные. Благодаря ограничениям области определения данных, авторы смогли разработать схему данных с высоким быстродействием и сравнительно малым объёмом, что благоприятно сказывается на работе «DiatomDB». База данных имеет реляционную структуру. Для удобства работы с базой данных используется веб-приложение (<http://paleolimnology.info/DDB/start.php> или с сайта проекта РНФ <http://uralspaleolakes.info>, вкладка «DiatomDB»). В настоящее время «DiatomDB» продолжает заполняться новой информацией.

Заключение

Таким образом, на основе применения информационных технологий к совокупности данных гидрохимии озёр, результатов диатомового и силикатного анализа поверхностных слоев озерных отложений, содержания микроэлементов в воде и донных отложениях, а также морфометрических параметров озёр и климатических характеристики озёр Урала, созданы базы данных «LaDa» и «DiatomDB», позволяющие работать с большим объёмом аналитических данных об озёрах Урала и качественно идентифицировать диатомовые водоросли, получая подробную информацию о распространении и экологии диатомовых водорослей. Созданные базы данных дают возможность изменять и добавлять информацию.

Работа выполнена при поддержке Российского научного фонда (грант No. 21-17-00071, <https://rscf.ru/project/21-17-00071/>).

Список литературы

1. Разработка палеолимнологической базы данных PaleoOnego / В. А. Гурбич [и др.] // Палеолимнология северной Евразии. Опыт, методология, современное состояние. 2016. С. 171–174.
2. Свидетельство о государственной регистрации базы данных № 2014621709. База данных «Озёра Центральной Якутии» / Пестрякова Л. А., Ушницкая Л. А., Субетто Д. А., Жирков И. И. Оpubл. 10.12.2014.
3. Свидетельство о государственной регистрации базы данных № 2012621208. База данных SPBLAKES для озёр Ленинградской области Российской Федерации / Рянжин С. В., Субетто Д. А., Кочков Н. В., Нестерова Л. А., Малоземова О. В. Оpubл. 23.11.2012.
4. Свидетельство о государственной регистрации базы данных № 2014621070. База палеолимнологических данных «PaleoLake» / Субетто Д. А., Сырых Л. С. Оpubл. 31.07.14.
5. Pienitz R., Cournoyer L. Circumpolar Diatom Database (CDD): a new database for use in paleolimnology and limnology // Journal of Paleolimnology. 2017. № 57. P. 213–219. <https://doi.org/10.1007/s10933-016-9932-0>
6. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource / J. W. Williams [et al.] // Quaternary Research. 2018. V. 89. P. 156–177. <https://doi.org/10.1017/qua.2017.105>
7. Freshwater Diatom Flora of Britain and Ireland / I. Jüttner [et al.] // Amgueddfa Cymru. National Museum Wales, 2023. Доступно в сети: <https://naturalhistory.museumwales.ac.uk/diatoms>.
8. Charles N. Oamaru diatoms. 2018. Доступно в сети: <https://www.oamarudiatoms.co.uk/index.html>
9. Lecointe C., Coste M., Prygiel J. Omnidia: software for taxonomy, calculation of diatom indices and inventories management // Hydrobiologia. 1993. V. 269, № 270. P. 509–513. Доступно в сети: <https://omnidia.fr/presentation/>
10. Schmidt-Kloiber A., Hering D. www.freshwaterecology.info - an online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences // Ecological Indicators. 2015. № 53. P. 271–282. <https://doi.org/10.1016/j.ecolind.2015.02.007>
11. Diatoms.org: supporting taxonomists, connecting communities / S. A. Spaulding [et al.] // Diatom Research. 2021. V 36, № 4. P. 291–304. <https://doi.org/10.1080/0269249X.2021.2006790>

12. Maslennikova A. V. Southern and Middle Urals lakes diatom abundance and hydrochemistry. PANGAEA. 2019. <https://doi.org/10.1594/PANGAEA.904675>
13. Определитель диатомовых водорослей России / М. С. Куликовский [и др.]. Ярославль: Филигрань, 2016. 803 с.
14. Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessment. English edition with updated taxonomy and added species / H. Lange-Bertalot [et al.]. Schmitten-Oberreifenberg: Koeltz Botanical Books, 2017. 942 с.
15. Свидетельство о государственной регистрации базы данных № 2018621834. Комплексная база данных озер Урала (LaDa) / Масленникова А. В., Гулаков В. О. Оpubл. 19.11.2018. Доступно в сети: <http://paleolimnology.info/DB/lada.php>
16. Свидетельство о государственной регистрации базы данных № 2023620230. База данных определитель диатомей «DiatomDB»/Diatoms database-key «DiatomDB» / В. О. Гулаков, А. В. Масленникова, Н. А. Исакова. Оpubл 16.01.2023. Доступно в сети: http://paleolimnology.info/DDB/show_all.php

References

1. Razrabotka paleolimnologicheskoy bazy dannyh PaleoOnego [Development of paleolimnological database PaleoOnego] / V. A. Gurbich [et al.] // *Paleolimnologiya severnoj Evrazii. Opyt, metodologiya, sovremennoe sostoyanie*. 2016. P. 171–174.
2. Certificate of state registration of the database № 2014621709. Baza dannyh «Ozyora Central'noj YAkutii» [Database "Lakes of Central Yakutia"] / Pestryakova L. A., Ushnickaya L. A., Subetto D. A., ZHirkov I. I. Published 10.12.2014.
3. Certificate of state registration of the database № 2012621208. Baza dannyh SPBLAKES dlya ozer Leningradskoj oblasti Rossijskoj Federacii [SPBLAKES database for lakes of the Leningrad region of the Russian Federation] / Ryanzhin S. V., Subetto D. A., Kochkov N. V., Nesterova L. A., Malozemova O. V. Published 23.11.2012.
4. Certificate of state registration of the database № 2014621070. Baza paleolimnologicheskikh dannyh «PaleoLake» [Paleolimnological data base "PaleoLake"] / Subetto D. A., Syryh L. S. Published 31.07.14.
5. Pienitz R., Cournoyer L. Circumpolar Diatom Database (CDD): a new database for use in paleolimnology and limnology // *Journal of Paleolimnology*. 2017. № 57. P. 213–219. <https://doi.org/10.1007/s10933-016-9932-0>
6. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource / J. W. Williams [et al.] // *Quaternary Research*. 2018. V. 89. P. 156–177. <https://doi.org/10.1017/qua.2017.105>

7. Freshwater Diatom Flora of Britain and Ireland / I. Jüttner [et al.] // Amgueddfa Cymru. National Museum Wales, 2023. Available online: <https://naturalhistory.museumwales.ac.uk/diatoms>.
8. Charles N. Oamaru diatoms. 2018. Available online: <https://www.oamarudiatoms.co.uk/index.html>
9. Lecointe C., Coste M., Prygiel J. Omnidia: software for taxonomy, calculation of diatom indices and inventories management // *Hydrobiologia*. 1993. V. 269, № 270. P. 509–513. Available online: <https://omnidia.fr/presentation/>
10. Schmidt-Kloiber A., Hering D. www.freshwaterecology.info - an online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences // *Ecological Indicators*. 2015. № 53. P. 271–282. <https://doi.org/10.1016/j.ecolind.2015.02.007>
11. Diatoms.org: supporting taxonomists, connecting communities / S. A. Spaulding [et al.] // *Diatom Research*. 2021. V 36, № 4. P. 291–304. <https://doi.org/10.1080/0269249X.2021.2006790>
12. Maslennikova A. V. Southern and Middle Urals lakes diatom abundance and hydrochemistry. PANGAEA. 2019. <https://doi.org/10.1594/PANGAEA.904675>
13. *Opredelitel' diatomovykh vodoroslej Rossii [Identification book of diatoms from Russia]* / M. S. Kulikovskij [et al.]. YAroslavl': Filigran', 2016. 803 p.
14. *Freshwater Benthic Diatoms of Central Europe: Over 800 Common Species Used in Ecological Assessment. English edition with updated taxonomy and added species* / H. Lange-Bertalot [et al.]. Schmittgen-Oberreifenberg: Koeltz Botanical Books, 2017. 942 p.
15. Certificate of state registration of the database № 2018621834. Kompleksnaya baza dannykh ozer Urala (Lada) [Integrated database of Ural Lakes (Lada)]/ Maslennikova A. V., Gulakov V. O. Published 19.11.2018. Available online: <http://paleolimnology.info/DB/lada.php>
16. Certificate of state registration of the database № 2023620230. Baza dannykh opredelitel' diatomej «DiatomDB»/Diatoms database-key «DiatomDB» / V. O. Gulakov, A. V. Maslennikova, N. A. Isakova. Published 16.01.2023. Available online: http://paleolimnology.info/DDB/show_all.php