

Водные биологические ресурсы

УДК 004.9:639.2

Программно-информационное обеспечение исследований водных биоресурсов Атлантике*Ф.В. Коломейко, А.Г. Васильев*

Атлантический научно-исследовательский институт рыбного хозяйства и океанографии (ФГБНУ «АтлантНИРО»), г. Калининград

E-mail: fed@atlantniro.ru, vasilyev@atlantniro.ru

Исследование водных биоресурсов Атлантического океана невозможно без соответствующего информационного обеспечения. В основе лежит использование баз данных, которые содержат результаты многолетних рыбохозяйственных исследований этой части Мирового океана, а для обработки применяется специализированное программное обеспечение, автоматизирующее сбор, обработку, анализ и хранение данных. Базы данных включают промысловую и биологическую информацию о гидробионтах и среде их обитания. Для обработки, анализа и визуализации информации в АтлантНИРО созданы и активно применяются информационно-справочные (ИСС), географические информационные системы (ГИС). Повышение эффективности исследований водных биоресурсов возможно за счёт создания автоматизированной системы поддержки принятия решений (АСППР), объединяющей базы знаний, ИСС, ГИС и применяющей методы интеллектуального анализа данных. В настоящее время в разрабатываемой АтлантНИРО АСППР используются несколько методов интеллектуального анализа данных, среди которых: построение деревьев решений и поиск ассоциативных правил.

Ключевые слова: информационное обеспечение, водные биоресурсы, Атлантический океан, базы данных, базы знаний, информационно-справочная система, географическая информационная система, система поддержки принятия решений.

ВВЕДЕНИЕ

Информационное обеспечение современных исследований водных биоресурсов (ВБР) основывается на данных, полученных в результате многолетнего сбора сведений о среде обитания и состоянии ВБР. Во время исследований ВБР собирается информация о пространственно-временной изменчивости атмосферных и гидросферных факторов. К ат-

мосферным факторам относятся температура воздуха, ветры (интенсивность, скорость, направление), атмосферное давление. К гидросферным факторам следует отнести солёность, содержание кислорода, фосфора, азота и других биогенных элементов в воде; температуру воды на поверхности и на определённых горизонтах, расположение гидрологических фронтов и т. п. Необходимо учитывать и биоти-

ческие факторы — распределение и видовой состав фито- и зоопланктона, биологические характеристики ВБР. Фиксируемые значения перечисленных факторов добавляются в базы данных (БД), которые обеспечивают хранение информации, её обработку и эффективное разграничение прав доступа к ней. БД используются в процессе оценки текущего и прогнозируемого биологического состояния ВБР, их численности, пространственного распределения и прогноза промысловой обстановки.

В настоящее время структура и качественные характеристики информационного обеспечения исследований ВБР претерпевают значительные изменения. С одной стороны, это связано с сокращением возможности активного использования специализированных научно-исследовательских судов, ведущих комплексный сбор океанологической и биологической информации, а с другой — с быстрым развитием коммуникационной техники, значительным расширением возможностей получения и передачи данных посредством спутниковых каналов связи, сети Интернет и появлением открытых международных информационных ресурсов с соответствующим наполнением.

С увеличением объёма информации, сокращением времени для её обработки и анализа традиционными методами становится актуальным применение технологий многомерного (технология OLAP — англ. OnLine Analytical Processing) и интеллектуального анализа (англ. — Data mining) данных. В результате применения программно-технических средств повышается уровень автоматизации получения, накопления, обработки и анализа информации. При этом снижается доля участия человека в перечисленных процедурах. Поэтому качество таких программных комплексов непосредственно влияет на точность оценок состояния запасов ВБР и достоверность прогнозирования.

Базы данных и программное обеспечение

Необходимой частью процесса автоматизации и информатизации научных исследований является разработка и использование баз данных и соответствующего программного обеспечения. В связи с этим в АтлантНИРО

создаются и постоянно пополняются промыслово-биологические, гидрологические и гидробиологические базы данных по всем районам Атлантического океана. Ведутся работы по переводу в электронный вид ретроспективной информации (50–80-е гг. XX в.) с бумажных носителей и обработке информации современных рейсов научных наблюдателей и научно-исследовательских судов АтлантНИРО, пополняются соответствующие базы данных. Базы формируются в средах систем управления базами данных (СУБД) Microsoft SQL server (MS SQL) [Станек, 2008] и PostgreSQL [PostgreSQL ..., 2018]. Доступ к ним, выборка, обработка информации и разграничение прав пользователей осуществляются с помощью специально созданного программного комплекса — «Информационно-справочная система АтлантНИРО» [Коломейко, Перевертнюк, 2016]. Функционирование ИСС основывается на клиент-серверной технологии.

С момента своего создания ИСС активно развивается. В настоящее время основные массивы информации состоят из оперативной и ретроспективной промысловой статистики, океанографических рейсовых массивов, данных тралового, кошелькового, ярусного научного и промыслового лова, данных биологии промысловых объектов, ихтио- фито- и зоопланктона, бентоса, данных, полученных с океанографических учётных станций и ряда другой научной информации. Всего банк данных содержит промыслово-биологическую информацию по 2490 рейсам с 50-х гг. XX в. В числе активно используемых в работе можно отнести следующие БД: биология ихтиофауны Атлантики (данные с 1956); биология кальмаров (данные с 1965); биология криля (данные с 1963); ихтио- (данные с 1979), фито- (данные с 1979), зоо- (данные с 1960) планктонные станции океанических районов; эпизоотологические данные рыб (данные с 1995); океанографические станции (данные с 1957); промысловая статистика по районам Атлантики (данные с 1977); тралово-акустические съёмки Балтийского моря (данные с 1992); биоанализы рыб в заливах Балтийского моря (данные с 1966) и ряд других.

Исследования, проводимые в Атлантике, описаны в одной из самых больших и востре-

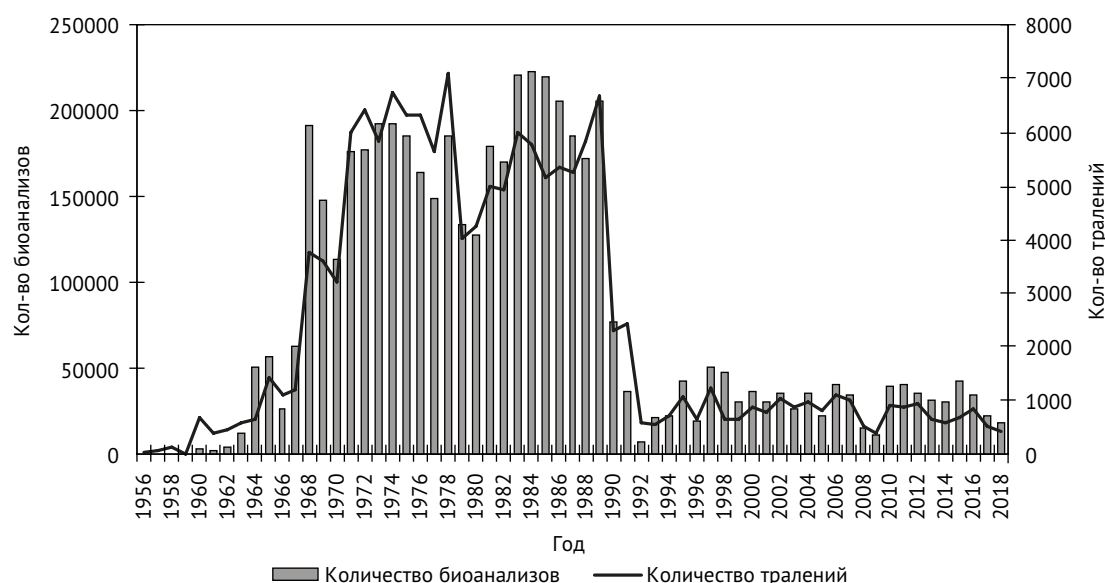


Рис. 1. Распределение по годам количества биоанализов и тралений, информация о которых содержится в БД «Биология океанических районов» (Атлантический океан)

бованных учёными АтлантНИРО БД — «Биология океанических районов». На рис. 1 представлено общее количество биологических анализов и траловых учётных станций, данные о которых в ней содержатся.

Для работы с БД требуются соответствующие программы. Программный комплекс ИСС АтлантНИРО является одной из таких разработок. Он реализован на языках C++, Delphi, Java, Javascript, PHP, SQL и др. В настоящее время начата работа по разработке приложений не только для персональных настольных ПК, но и для мобильных устройств (смартфонов, планшетных компьютеров).

Для работы на мобильных платформах возможно использование веб-ГИС [GIS-Lab ..., 2018], разработка которой также ведётся и в настоящее время. Её элементы уже сейчас доступны научным сотрудникам в локальной вычислительной сети АтлантНИРО. Для работы наблюдателей создан отдельный модуль ИСС (под названием «Enterform»). Он предназначен для перевода промыслово-биологической информации (траловых карточек, биологического анализа, массовых промеров ВБР) в электронный вид непосредственно в рейсе. Модуль может функционировать отдельно от ИСС, что делает его легко переносимым без необходимости установки СУБД на ноутбуки научных наблюдателей. «Enterform» формиру-

ет данные в формате DBASE(dbf), который является универсальным, не требует установки в операционную систему дополнительных надстроек. Работать с форматом.dbf возможно в большинстве программ обработки данных. Интерфейс модуля показан на рис. 2.

На сегодняшний день основные направления работы по информатизации научных исследований ВБР это, во-первых, проектирование, разработка, внедрение, сопровождение и администрирование электронных биологических, промысловых, океанографических баз данных (БД); во-вторых, создание и развитие информационно-справочных и географических информационных систем (ГИС), программ расчёта и анализа указанных данных.

В ИСС созданы входные формы ввода данных для реализации возможности пополнения БД новыми сведениями непосредственно в рейсах научно-исследовательских судов АтлантНИРО. Собранный таким образом информация во время рейса по электронным каналам связи с требуемой периодичностью может отправляться в адрес АтлантНИРО. Поступающие с судна данные (результаты учётных тралений, проведения биоанализов, океанографических станций) проходят верификацию и заносятся в БД сервера, и первичная информация сразу становится доступной по локальной вычислительной сети научным

Рис. 2. Интерфейс модуля ИСС «Enterform», предназначенный для перевода промыслово-биологической информации в электронный вид

сотрудникам, находящимся на берегу. В результате анализа состояния ВБР и среды их обитания, становится возможным уточнить популяционную структуру, закономерности распределения, и образования скоплений ВБР в оперативном режиме.

Интерфейс ИСС для работы с БД «Биология океанических районов» показан на рис. 3.

В промыслово-биологических БД содержится информация, включающая пространственно-временные характеристики параметров среды и состояния ВБР. Одним из наиболее эффективных способов работы с такой информацией являются географические информационные системы (ГИС). В АтлантНИРО ГИС применяются для визуализации результатов научных исследований ВБР, преобразования картографических проекций, пространственного анализа и моделирования, работы с растровой (прежде всего это данные дистанционного зондирования земли — спутниковые данные) и векторной графикой.

Для этих задач используются коммерческие (проприетарные) ГИС Surfer и ArcGIS, но в последнее время ведутся разработки собственной ГИС на основе открытого программного обеспечения (англ. open-source software). При выборе или создании ГИС для исследования ВБР на современном уровне следует учитывать следующее:

- необходимо решать сложные задачи с использованием информации из нескольких БД;
- требуется централизованная унифицированная рабочая среда;
- исследование ведётся неограниченным числом специалистов одновременно.

Указанные потребности могут быть успешно удовлетворены с помощью ГИС с применением веб-технологий. Принимая во внимание наглядность и удобство обмена данными с помощью веб-браузеров и имеющуюся возможность делать это внутри локальной вычислительной сети организации, оптимально

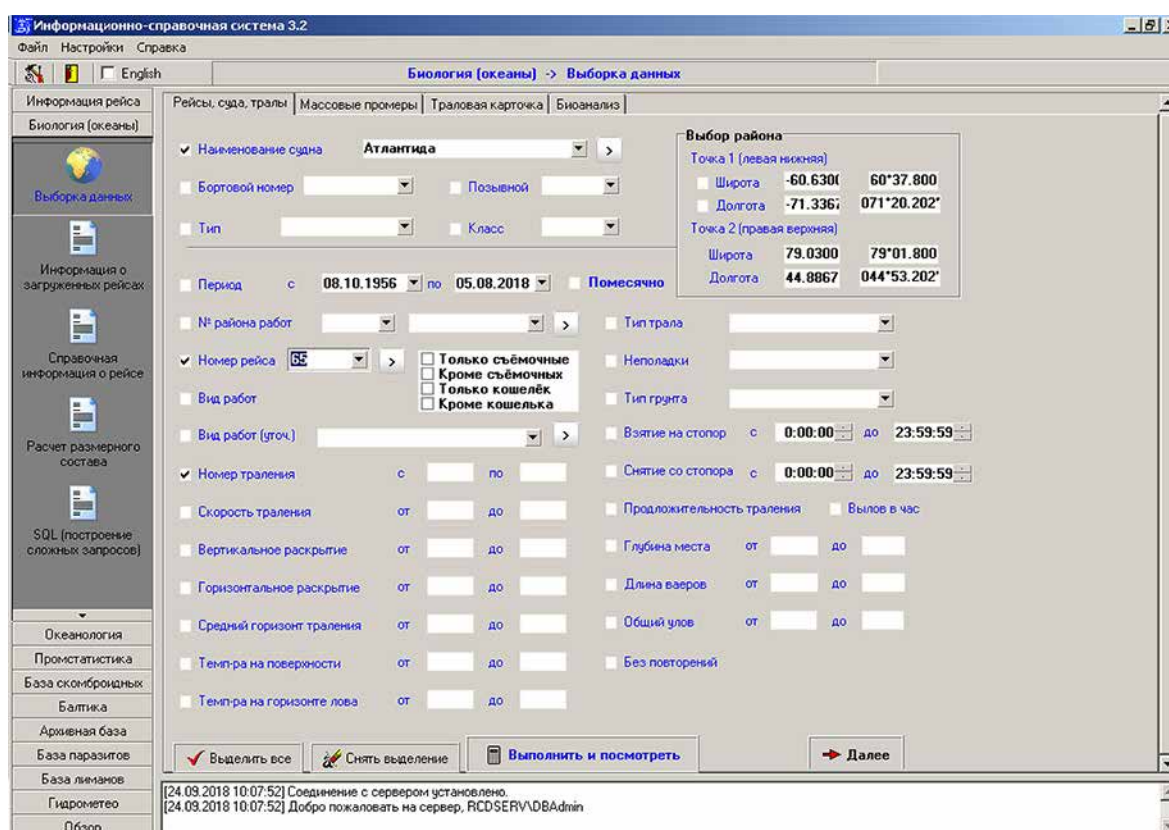


Рис. 3. Главное окно интерфейса ИСС АтланНИРО

в качестве ГИС организации выбрать веб-ГИС.

Открытое программное обеспечение (ПО) может без ограничений устанавливаться, применяться, дорабатываться под свои нужды и распространяться согласно, например, лицензионному соглашению GNU (GPL) [The GNU General ..., 2018]. Это является определяющим для задач, где важна скорость внедрения решений и возможность быстро откорректировать используемые алгоритмы и методы в ПО. Поэтому для исследования водных биоресурсов представляется рациональным использование открытого программного обеспечения (ПО).

В рамках такого подхода в АтланНИРО разрабатывается ГИС [Коломейко, 2013]. Эта система доступна на любом компьютере, подключённом к локальной вычислительной сети. Для реализации проекта выбраны следующие программные средства:

1. Веб-сервер Geoserver для размещения и обработки пространственных данных в сети [GeoServer..., 2018];

2. Программная библиотека OpenLayers для создания веб-карт [Openlayers..., 2018];

3. СУБД PostgreSQL и надстройка PostGIS, предназначенные для хранения в базе географических данных [PostGIS..., 2018];

Система оптимизирована для работы в веб-браузерах Mozilla Firefox и Google Chrome. Интерфейс гибко настраивается, расположение элементов может меняться пользователем и сохраняться для восстановления в других сеансах работы с системой. Предусмотрено добавление новых источников информации.

Главное окно интерфейса ГИС ФГБНУ «АтланНИРО» показано на рис. 4.

Создаваемая веб-ГИС позволяет осуществлять выборки из централизованных БД АтланНИРО, хранить, обрабатывать, преобразовывать векторные (shapefile, KML, GML и т. п.) и растровые файлы (png, jpeg, tiff, arcgrid и т. п.) и выдавать эту информацию по запросу пользователя в картографической форме или в виде таблиц и графиков. В целом, веб-ГИС помогает решать в области исследо-

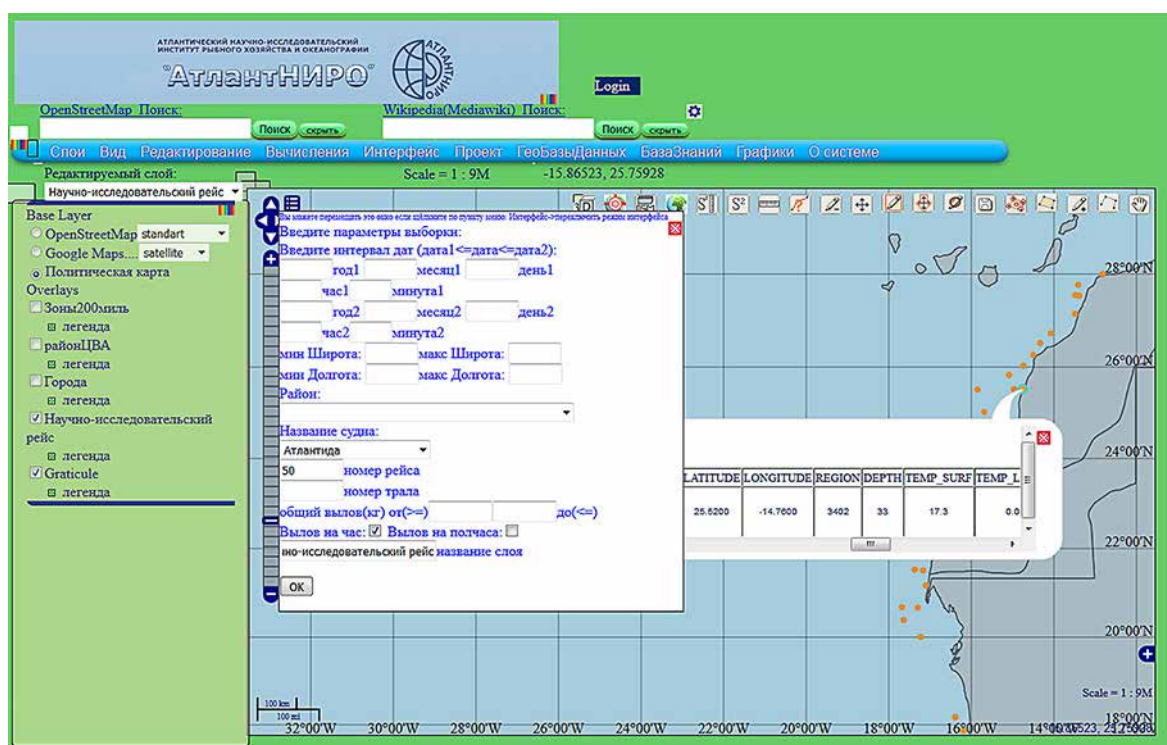


Рис. 4. Главное окно интерфейса ГИС АтлантНИРО

ваний ВБР разнообразные задачи по следующим направлениям:

1. Пространственно-временной мониторинг промысла, гидрологической ситуации и т. п.
2. Создание тематических карт и интерактивных атласов, оперативное картографирование.

Система поддержки принятия решений и интеллектуальный анализ данных

Указанные ИСС и ГИС являются частью разрабатываемой автоматизированной системы поддержки принятия решений (АСППР) в рыбном промысле и научных исследованиях водных биоресурсов на основе пространственно-временного мониторинга. Разработка систем поддержки принятия решений является одним из актуальных направлений развития современных информационных технологий. Использование таких систем в рыбохозяйственных исследованиях позволяет направить одновременно на анализ научных данных всю функциональность ИСС, ГИС с применением не одного метода или модели, а множества ма-

тематических, интеллектуальных методов, баз данных, баз знаний и систем управления базами данных.

Данные, полученные в результате научных исследований ВБР и рыбного промысла, неоднородны и нестационарны. Значения параметров, характеризующих состояние ВБР, зависят от множества стохастических факторов, поэтому классических математических методов (прежде всего методов прикладной статистики) для анализа и достоверного прогноза этих данных не всегда достаточно. В подобных ситуациях необходимо не только использовать первичную информацию в виде баз данных, но и формировать и использовать базы знаний (БЗ), которые содержат результат смысловой, логической обработки информации, опыт специалистов-практиков, экспертов. БЗ в области исследований ВБР можно создавать на основе продукционной модели представления знаний [Гаврилова, Хорошевский, 2000], которая представлена набором правил вида: ЕСЛИ <условие> И <условие>, ТО <вывод, результат или действие>. БЗ на основе продукционной модели представления знаний

может содержать типовые ситуации в интересующем исследователя районе, описание основных особенностей поведения и распределения рыб в зависимости от значений параметров среды. Рекомендации по маневрированию промыслового флота также могут быть включены в БЗ. Таким образом, БЗ, используемая в исследованиях ВБР, должна содержать возможные сценарии изменения состояния среды и распределения биоресурсов в зависимости от анализируемых факторов. Пример выдачи рекомендаций в АСППР из БЗ с учётом отклонения температуры поверхности океана от среднеемноголетних значений показан на рис. 5. Рекомендации, добавленные в БЗ, сформированы на основе характерных периодов промысла, описанных в методическом пособии по краткосрочному прогнозированию промысла в ЦВА [Лукацкий, Маслянкин, 2010].

Для повышения гибкости информационного обеспечения исследований БЗ может находиться как в среде СУБД, так и представлять собой набор отдельных текстовых файлов. При этом БЗ может создаваться самим исследователем (группой исследователей) или является результатом интеллектуального анализа данных, содержащихся в первичных БД. На современном этапе развития информацион-

ных технологий термин «интеллектуальный анализ данных» часто используется в виде англоязычного синонимичного термина — Data Mining. Это процесс автоматизированного поиска в информации, как правило, больших БД скрытых закономерностей (шаблонов информации), тенденций, знаний, которые ранее были не очевидны или требовали научного обоснования. Data Mining может объединять методы прикладной статистики, распознавания образов, искусственного интеллекта, теорию баз данных и др.

Классические статистические методы анализа данных в основном направлены на проверку предварительно заданных предположений или гипотез. Технология Data Mining прежде всего направлена на поиск новых заранее неопределённых и неявных закономерностей. Методы Data Mining помогают выявлять подобные закономерности и определять взаимосвязи параметров. В исследованиях ВБР именно формулировка гипотезы о взаимозависимости интересующих параметров часто является сложной задачей, поэтому применение Data Mining в ряде случаев целесообразней по сравнению с другими методами анализа. Подавляющее большинство методов прикладной статистики осуществляют усреднение данных

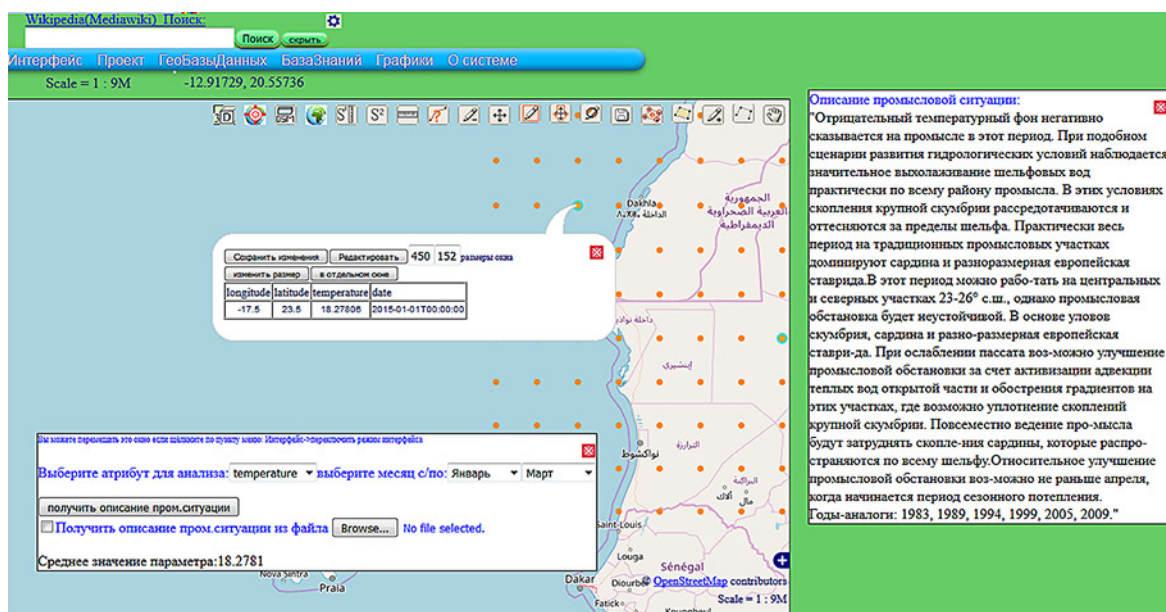


Рис. 5. Пример интерфейса АСППР, выдающей рекомендации из БЗ о ведении промысла с учётом отклонения температуры поверхности океана от среднеемноголетних значений

для выявления взаимосвязей между параметрами, но Data Mining использует алгоритмы, которые, как правило, работают с исходными, реальными значениями.

В настоящее время в разрабатываемой АтлантНИРО АСППР используются несколько методов интеллектуального анализа данных (Data Mining) среди которых: построение дерева решений [Шахиди, 2006] с помощью алгоритма ID3 (Iterative Dichotomiser 3) и поиск ассоциативных правил с помощью алгоритма Apriori [Шахиди, 2004].

Для исследований ВБР дерева решений могут давать наглядное представление выявленных зависимостей (правил) между параметрами в виде иерархической структуры. Правило в данном случае строится в виде логического утверждения «если ..., то ...», как показано выше.

Ассоциативные правила позволяют находить закономерности между связанными событиями и явлениями. Примером такого правила, полученного по результатам анализа промысловых данных 2013–2018 гг., может являться утверждение, что суточный вылов скумбрии судном типа РТМКС в атлантической рыболовной зоне Марокко в августе с вероятностью 63% будет находиться в интервале от 60 до 140 т. Таким образом, целью анализа данных с помощью этого метода является установление следующих зависимостей: если в БД встретился набор значений некоторых параметров, то другие параметры будут принимать определённые значения с заданной вероятностью.

В дальнейшем для развития и увеличения функциональности АСППР планируется расширение набора методов интеллектуального анализа данных, доступных в системе пользователю.

ЗАКЛЮЧЕНИЕ

Использование в рыбохозяйственных исследованиях Атлантического океана постоянно пополняемых больших баз данных, создание и совершенствование справочных и географических информационных систем, разработка и применение систем поддержки принятия решений способствуют повышению информа-

тизации научных исследований водных биоресурсов. А обработка научной информации методами многомерного и интеллектуального анализа данных ведёт к повышению их эффективности.

Сравнительно недавно, сотрудники АтлантНИРО при проведении исследований применяли программы для ЭВМ общего назначения, входящие, как правило, в пакеты офисных программ. Теперь, по мере развития специализированного программного обеспечения — ИСС и АСППР АтлантНИРО, а также использования локальных вычислительных сетей — процессы сбора и обработки биологической, промысловой и океанологической информации унифицируются, уровень автоматизации обработки повышается, время с момента сбора первичных научных данных до их многофакторного автоматизированного анализа сокращается.

Комплексное использование баз данных, современных информационных систем, OLAP и Data Mining технологий приводит к интенсификации научных исследований, способствует принятию оптимальных, научно обоснованных решений по управлению промыслом и исследованиями ВБР.

ЛИТЕРАТУРА

- Гаврилова Т.А., Хорошевский В.Ф. 2000. Базы знаний интеллектуальных систем. СПб.: Питер. 384 с.
- Коломейко Ф.В. 2013. Географические информационные системы на основе программного обеспечения с открытым исходным кодом (open source) и базы данных в научных исследованиях водных биоресурсов // Труды науч. конф. «Водные биоресурсы, аквакультура и экология водоёмов». КГТУ. Калининград. С. 69–72.
- Коломейко Ф.В., Перевертнюк М.В., Бутovich Я.Ф., Щукина Е.В. 2016. Программа для ЭВМ «Информационно-справочная система АтлантНИРО»: А.с. 2016660553 Россия.
- Лукацкий В.Б., Маслянкин Г.Е. 2010. Методическое пособие по краткосрочному прогнозированию промысла в Центрально-Восточной Атлантике. Калининград: Атлант-НИРО. 42 с.
- Станек У.Р. Microsoft SQL Server 2005. 2008. Справочник администратора. М.: Русская Редакция. 544 с.
- Шахиди А. 2004. Введение в анализ ассоциативных правил. Доступно через: <https://basegroup.ru/community/articles/intro> 24.09.2018

Шахиди А. 2006. Деревья решений — общие принципы работы. Доступно через: <https://basegroup.ru/community/articles/description> 24.09.2018

GeoServer is an open source server for sharing geospatial data. Accessible via: <http://geoserver.org> 19.09.2018

GIS-Lab: Веб-ГИС. Доступно через: <http://gis-lab.info/qa/webgis.html> 24.09.2018

OpenLayers — Welcome. Accessible via: <https://openlayers.org> 19.09.2018

PostGIS — Spatial and Geographic Objects for PostgreSQL. Accessible via: <https://postgis.net>. 20.09.2018.

PostgreSQL: The World's Most Advanced open source relational database. Accessible via: <https://www.postgresql.org> 20.09.2018.

The GNU General Public License v3.0 — GNU Project — Free Software Foundation. Accessible via: <http://www.gnu.org/licenses/gpl.html> 24.09.2018.

Поступила в редакцию 18.10.2018 г.
Принята после рецензии 24.10.2018 г.

Trudy VNIRO

2018. Vol. 174

Aquatic biological resources

Program-information support of water biological resources research in the Atlantic Ocean

F.V. Kolomeyko, A.G. Vasilyev

Atlantic Fisheries Research Institute (FSBSI «AtlantNIRO»), Kaliningrad

Water biological resources research in the Atlantic Ocean requires corresponding informational software. It is based on data bases usage that contain the results of many years exploring and software that automates their collecting, handling, analyses and storage. Data bases include fishery and biological, hydrological and hydro-biological information about water biological resources and their natural environment. For this data handling, analyses and visualization the information and reference systems (IRS), geographic information systems (GIS) are created and applied. The efficiency growth of water biological resources researches is possible thanks to the automated decision support system creation that integrates knowledge bases, IRS, GIS and uses data mining methods. The efficiency growth of water biological resources researches is possible thanks to the automated decision support system creation that integrates knowledge bases, IRS, GIS and uses data mining methods. Nowadays several methods of data mining are used in the automated decision support system that is being developed in AtlantNIRO, among them there are the following effective ones: decision trees building and association rules finding.

Keywords: Information support, water biological resources, Atlantic Ocean, data bases, knowledge base, information and reference system, geographic information system, decision support system.

REFERENCES

- Gavrilova T.A., Horoshevskij V.F. 2000. Bazy znanij intellektual'nyh sistem [Knowledge bases of intelligence systems]. SPb.: Piter. 384 s.
- Kolomejko F.V. 2013. Geograficheskie informatsionnye sistemy na osnove programmno obespicheniya s otkrytym iskhodnym kodom (open source) i bazy dannyh v nauchnyh issledovaniyah vodnyh bioresursov [Geographic information systems based on Open Source Software and data bases in water biological resources scientific researches] // Trudy nauch. konf. «Vodnye bioresursy, akvakul'tura i ehkologiya vodoemov». KGTU. Kaliningrad. S. 69–72.
- Kolomejko F.V., Perevertnyuk M.V., Butovich YA. F., Shchukina E.V. 2016. Programma dlya EHVM «Informatsionno-spravochnaya sistema AtlantNIRO» [Program for computer «Information and reference systems by AtlantNIRO »]: A.s. 2016660553 Rossiya.
- Lukatskij V.B., Maslyankin G.E. 2010. Metodicheskoe posobie po kratkosrochnomu prognozirovaniyu promysla v Tsentral'no-Vostochnoj Atlantike [Methodical manual on short-term forecasting in Eastern Central Atlantic]. Kaliningrad: Atlant-NIRO. 42 s.
- Stanek Ul'yam R. Microsoft SQL Server 2005. 2008. Spravochnik administrator [Microsoft SQL Server 2005. Administrator's Pocket Consultant]. M.: Russkaya Redaktsiya. 544 s.
- Shahidi A. 2004. Vvedenie v analiz assotsiativnyh pravil [Introduction in association rules analyses]. Accessible via: <https://basegroup.ru/community/articles/intro> 24.09.2018
- Shahidi A. 2006. Derev'ya reshenij — obshchie printsiipy raboty [Decision trees — general principles]. Accessible via: <https://basegroup.ru/community/articles/description> 24.09.2018
- GeoServer is an open source server for sharing geospatial data. Accessible via: <http://geoserver.org> 19.09.2018
- GIS-Lab: Veb-GIS. Accessible via: <http://gis-lab.info/qa/webgis.html> 24.09.2018
- OpenLayers — Welcome. Accessible via: <https://openlayers.org> 19.09.2018
- PostGIS — Spatial and Geographic Objects for PostgreSQL. Accessible via: <https://postgis.net>. 20.09.2018.
- PostgreSQL: The World's Most Advanced open source relational database. Accessible via: <https://www.postgresql.org> 20.09.2018.
- The GNU General Public License v3.0 — GNU Project — Free Software Foundation. Accessible via: <http://www.gnu.org/licenses/gpl.html> 24.09.2018.

FIGURE CAPTIONS

Fig 1. The number of biological analysis and trawling according to years, information about which is contained in data base «Biology of ocean regions» (The Atlantic Ocean)

Fig. 2. The interface of IRS module «Enterform», intended for fishery-biological information converting in digital form

Fig. 3. The interface main window of AtlantNIRO information and reference system

Fig. 4. The interface main window of AtlantNIRO geographic information system

Fig. 5. The example of automated decision support system interface, when the system gives recommendations about fishery from the knowledge base taking into account the deviation of the sea surface temperature from the mean annual values