

КОЛЛЕКЦИЯ ЗНАНИЙ ОБ ОНЕЖСКОМ И ЛАДОЖСКОМ ОЗЕРАХ

Н.Н. Филатов

Институт водных проблем Севера,

Петрозаводск, пр. А.Невского, 50,

nfilatov@nwpi.karelia.ru,

В.А. Лебедев

Институт прикладных математических исследований

Карельского научного центра РАН,

Петрозаводск, ул. Пушкинская, 11,

lebedev@krc.karelia.ru

Описывается архитектура коллекции знаний об Онежском и Ладожском озерах. Рассмотрены структура и содержание базы метаинформации, функциональный состав и интерфейс взаимодействия пользователей с коллекцией.

Онежское и Ладожское озера являются крупнейшими пресноводными водоемами Европы [1,2].

Их исследование важно с точки зрения использования их богатых природных, рекреационных, транспортных, энергетических и других ресурсов. Однако до сих пор не созданы необходимые информационные базы данных для решения разнообразных научных и практических проблем. Для понимания процессов воздействия человеческой деятельности на экосистемы крупнейших озер Европы, развитие систем управления экосистемой озера и водосбором, а также для экологического и социальноэкономического развития региона, необходимо разработать соответствующие базы данных и знаний. Это необходимо для лучшего управления ресурсами, озер их сохранения и решения практических вопросов. Развитие математических моделей также сдерживается отсутствием этих баз.

В настоящее время Онежское озеро в целом является олиготрофным водоемом, хотя отдельные заливы и губы, подверженные антропогенному воздействию, становятся мезотрофными. Наибольшее воздействие, в первую очередь, оказывают непосредственные сбросы промышленных предприятий, а также речной, ливневой и дренажный сток крупных промцентров: Петрозаводска, Кондопоги и Медвежьегорска. Открытая часть озера до настоящего времени все еще сохраняет олиготрофный характер. Она бедна микрофлорой и характеризуется небольшим развитием планктонных водорослей с доминированием диатомово-золотистого комплекса. Слабое протекание процессов синтеза органического вещества - (фотосинтез - хемосинтез) определяет невысокое развитие последующих звеньев трофической цепи. Биопродукционные процессы всех трофических звеньев вследствие холодноводности водоема выражены слабо. Увеличение за последнее десятилетие в открытом районе озера, особенно в заливе Большое

Онего, количественных показателей почти всех трофических звеньев свидетельствует о проникновении сюда загрязненных вод.

Ладожское озеро является олиготрофно-мезотрофным водоемом. К эвтрофным можно отнести отдельные заливы на юге озера, например, Волховскую губу. Выполненные за последние 10 лет мероприятия, такие как закрытие Приозерского ЦБК, совершенствование технологического процесса на Волховском алюминиевом заводе, Питкярантском и Ляскельском предприятиях, а также падение производства в бассейне озера и уменьшение использования удобрений для сельского хозяйства почти в пять раз привели к замедлению скорости процесса эвтрофирования озера, уменьшили антропогенную нагрузку на водоем.

Цели создания коллекции состоят в обеспечении поддержки исследований и выработки рекомендаций по рационализации водного хозяйства и охране крупнейших озер Европы, обучения студентов и аспирантов на основе систематизации цифрового представления знаний по изученности этих водоемов, проблемам и средствам их решения, т.е. коллекция является электронной энциклопедией знаний об этих водоемах. Она должна: обеспечить полноту, непротиворечивость и доступность знаний в локальной, корпоративной и отчасти в глобальной сети; учитывать особенности предметной области, как-то: междисциплинарность, сложность и разветвленность системы понятий, большую мощность и разнотиповость множеств объектов, распределенных в пространстве, и динамический характер знаний [3].

Полнота коллекции определяется степенью охвата понятий объектов, процессов, отношений, проблем, характеризующих изученность предметной области, совокупность названий которых является каркасом или структурой знаний. Последняя может быть представлена в виде графа, вершины которого обозначены терминами и названиями, а дуги отображают структурообразующие отношения. В качестве последних приняты отношения типа “род-вид”, “целое-часть”, “синонимия”, свойства которых соответствуют свойствам структур, т.е. каждая пара элементов обладает нижней и верхней гранями и подчиняется нестрогому порядку. Одно из представлений структуры – ациклический граф [4]. Таким образом, полнота коллекции определяется полнотой структуры знаний, которая может быть обеспечена посредством инвентаризации классификаций, агрегаций и равнозначностей, применяемых в предметной области, и последующего построения из них структуры. Эта задача не тривиальная и требует создания некоторых технологических средств поддержки. Основа ее решения – это возможность различать объемы понятий так, чтобы верхняя грань структуры (корни) включала понятия наибольшего объема из рассматриваемых, а нижняя грань (листья) – наименьшего объема.

Большая общность ациклического графа позволяет включать в состав структуры знаний все множество классификаций, агрегаций и равнозначностей предметной области, в виде их композиции.

Чтобы обеспечить полноту коллекции, необходимо привлечь знания по геологии, климатологии, географии, гидрологии, гидрохимии, гидробиологии, ихтиологии, токсикологии и др. научным дисциплинам. Содержанием коллекции будут понятия, объекты, их классы и взаимодействия, множества которых и их отношения целесообразно отобразить в структуре знаний в виде ациклического графа.

Верхние уровни структуры знаний – климат, водоемы, водотоки, водосборы, подземные воды, водное хозяйство. В частности, абиотическая среда водоемов будет характеризоваться физико-географическим описанием, водным балансом, уровнем и их изменчивостью, термическим режимом, течениями, поступлением и выносом химических веществ, динамикой химсостава вод по акваториям (зонам, заливам), физико-химическим описанием донных отложений, влиянием климатических и антропогенных воздействий.

В состав структуры включен словарь, охватывающий основные классы названий: видов, систематических и экологических групп и классов растений, животных и микроорганизмов (лат. и рус.), территорий, водных объектов и акваторий, поселений, лесных и сельских хозяйств, охраняемых территорий и акваторий, предприятий, горных пород, минералов, грунтов, химических элементов и соединений, классов вод, типов и форм рельефа, типов экосистем, биоценозов, местообитаний и некоторые другие названия.

Знания в коллекцию вносятся посредством “навешивания” их на структуру. При этом на каждую из вершин могут быть “навешены” знания, поясняющие смысл понятия: пояснения, толкования, исторические справки, описания методик и т.п. Кроме того, к вершинам-листьям должны быть отнесены конкретные знания: свойства, состояния, отношения объектов и их классов, процессы, их модели и т.п.

Формы представления знаний в коллекции – это тексты (возможно иллюстрированные), географические карты, базы данных, программы решения задач, космические снимки, библиографии и др. Количество информационных объектов (документов) может исчисляться тысячами, поэтому возникает проблема доступности знаний, которая может быть решена созданием базы метаинформации, обеспечивающей доступ к конкретным документам.

В процессе изучения сложных предметных областей вероятно возникновение противоречий в знаниях, которые разрешаются обычно нетривиальными неформальными методами и средствами. Здесь речь идет о непротиворечивости структуры знаний, в понятие которой включаются отсутствие омонимии и полисемии терминов и названий и запреты логических циклов (часть не может включать в себя целого). Таким образом, формально непротиворечивость структуры знаний контролируется на отсутствие циклов в графе, представляющем ее, а также на отсутствие в нем одноименных вершин-предков с разным составом вершин-потомков. Одновременно должна контролироваться связность графа [5].

Под доступностью знаний коллекции здесь понимается наличие средств, обеспечивающих эффективное нахождение и визуализацию интересующих фрагментов знаний за время необременительного ожидания.

Граф структуры знаний может быть использован не только для обеспечения полноты и непротиворечивости коллекции, но и для ускорения доступа к знаниям как при наполнении коллекции, так и при просмотре и обработке их, т.е. в качестве базы метаинформации. С этой целью для реализации принято представление этого графа в виде пары: множества помеченных вершин и множества дуг. Таким образом, база метаинформации коллекции реализована как реляционная база данных, структура которой представлена на рис. 1. Она состоит из следующих таблиц: таблицы дуг (композиция множеств всех вершин и дуг), таблицы вершин-листьев, таблицы списков ключевых слов, таблицы карт, таблицы задач и словарей: названий, синонимов и ключевых слов. В качестве помет вершин используются понятия типа вершины, типа знания, комментарии (пояснения), связи с картой, с темой, со слоем карты и с заданиями.

В зависимости от типа и формата документа после его вызова выбирается та или иная программа манипулирования и визуализации. Все это дает следующие преимущества по сравнению с традиционным представлением графов в виде двоичных или троичных деревьев:

- Открыт прямой доступ к любой вершине графа (т.е. не требуется прохождение путей только от корневых вершин, облегчена доступность).
- Достаточно просто контролируется целостность базы данных (в частности, связанность графа).
- Достаточно просто вносятся изменения и дополнения в структуру знаний и в коллекцию, т.е. обеспечен импорт новых понятий и знаний и удаление устаревших (т.е. динамичность).

Часть информации коллекции распределена в пространстве географической карты. Это дает возможность организовать доступ к знаниям через карту. Для этого соответствующие документы связаны с тематическими слоями карты (идентификаторы объектов карты равны соответствующим идентификатором документов). База метainформации содержит таблицу карт, а в таблице листов указаны номера карт, тем и слоев, к которым относится данный документ (рис. 1). Таким образом, после вызова на экран соответствующей тематической карты щелчком мыши по объекту вызывается интересующий фрагмент знаний об этом объекте. Так как для каждого объекта существует множество аспектов его описания (зачастую десятки), то введено понятие темы (тематической карты) так, чтобы количество аспектов на объект слоя в ней не превышало 10-15.

Архитектура коллекции упрощенно показана на рис. 2. В соответствии с требованиями коллекция должна импортировать фрагменты знаний в обусловленных форматах, надежно их хранить и обеспечивать доступ к ним для использования и модификации в локальной, корпоративной (интранет) и глобальной (интернет) сетях, разделяя права доступа на три зоны: внутреннюю, корпоративную и внешнюю. Во внутреннюю зону включается весь объем знаний коллекции с доступом к ней в локальной сети. В корпоративной сети часть знаний, относящихся к незавершенным работам, закрыта. В глобальной сети большая часть знаний второй зоны может быть доступна на договорной основе (закрыта паролями). В качестве хранилища знаний (и метainформации) используется основная память сервера, обслуживаемая файловой системой. Обработка данных в локальной сети осуществляется на клиентских компьютерах при помощи пользовательского программного модуля. Он представляет из себя exe-модуль, созданный по объектно-реляционной технологии средствами FoxPro, Map Object LT и C++, который выполняет функции навигации в базе метаданных, вызывает отобранные фрагменты, выполняет геоинформационные функции, вычислительную обработку и визуализирует результаты. Такое решение позволило достичь эффекта распределенности коллекции между клиентскими компьютерами локальной сети.

В частности, каждый пользователь имеет возможность создавать на своем компьютере личные коллекции и использовать их совместно с общей, например, создавая виртуальные базы данных. Личные коллекции недоступны в локальной сети, однако они могут быть по желанию пользователя открыты через ODBC и зарегистрированы в общей базе метainформации, т.е. станут доступными для других пользователей локальной сети.

Для открытия коллекции в интернет/интранет на Web-сервере объявляется сокращенный вариант базы метainформации. Взаимодействие с ним удаленных пользователей осуществляется через РСУБД и CGI – шлюз посредством динамических форм Web-сайта коллекции. Возможность взаимодействия этих пользователей с цифровыми картами определяется комплектацией их броузеров соответствующими скриптами. В отдельных случаях на основе карт могут быть подготовлены для демонстрации требуемые HTML –документы. В состав коллекции могут включаться удаленные информационные ресурсы из интернет/интранет посредством регистрации их названий и URL в базе метainформации.

Рассмотренная архитектура позволяет реализовать широкий спектр функций взаимодействия с коллекцией знаний. Все функции подразделяются на администраторские и пользовательские. Администраторские функции должны обеспечить импорт, сохранность и модификацию знаний, распределение ресурсов и прав пользователей, мониторинг использования коллекции и перераспределение ресурсов. Администраторские функции подробно здесь не рассматриваются. Часть этих функций может быть делегирована локальным пользователем, и они будут рассмотрены как пользовательские.

Пользовательские функции должны обеспечивать доступ, отбор релевантных знаний, их модификацию и обработку, передачу и визуализацию.

В рамках локальной сети совокупность пользовательских функций подразделена на группы: импорта, поиска, связей, исключения. В состав функций импорта входят функции включения в коллекцию географических карт и отдельных тематических слоев, баз данных, текстов и иллюстраций в форматах txt, doc, rtf, html и др., а также программ решения задач. Кроме того имеется функция связывания документов с географической картой. Следует отметить, что все импортируемые документы должны быть заранее оформлены и размещены в хранилище информации, должны быть известны их названия и адреса. На рис. 3 показано меню функции импорта карт.

Поиск документов в коллекции может осуществляться тремя способами: по названиям документов, по ключевым словам, указанием объектов на карте. На рис. 4 представлено меню поиска документа по ключевым словам. При выборе соответствующего названия пользователю предлагается список функций воздействия на этот документ: просмотр, изменение, добавление, печать, вычисление, диаграммы (конечно, если пользователь имеет соответствующие права, а функция определена на документе).

При поиске документов по карте открывается карта и меню функций: увеличение-уменьшение масштаба, подбор слоев, поиск объекта по названию, классификации объектов, определение длин и географических координат, печать карты. После указания объекта щелчком мыши выдается список аспектов описания объектов этого класса. Выбрав конкретный аспект из списка, вызываем соответствующий документ. На рис. 5 показано меню действий с картой и вызов описания объекта.

Группа функций “связи” предоставляет возможность создания виртуальных БД посредством формирования SQL – задания, а также задания гиперссылок между HTML – документами.

Функция “исключение” позволяет убрать из коллекции устаревший документ или заменить его новым.

Особо отметим предусмотренные в коллекции вычисления на базах данных, которые определяются при импорте в таблице задач. Кроме того могут быть импортированы задачи на математических моделях.

Что касается состава функций для Internet/Intranet, то пока предусмотрен поиск документов по названным и по ключевым словам.

Для удобства пользования коллекцией разработана система интерфейса, включающая около 50 оконных меню, охватывающих все функциональные возможности системы.

В настоящее время в коллекцию включена информация по гидрографии рек и озер, климатологическая БД в привязке к карте масштаба 1:1 000 000. Готовятся данные по гидрологии и гидрохимии, а также карта масштаба 1:200 000.

Литература

1. Онежское озеро. Экологические проблемы. Ред. Филатов Н.Н. Петрозаводск, 1999.
2. Ладожское озеро. Ред. Филатов Н.Н. Петрозаводск, 2000.

3. Лебедев В.А., Филатов Н.Н. Коллекции знаний для Интернет о крупнейших озерах Европы: Ладожское и Онежское. //Тезисы докладов Международной конференции “Интернет и общество”. С.-Петербург, 2000.
4. Лебедев В.А. О методике построения электронных коллекций знаний. //Труды Института прикладных математических исследований, вып. 2. Петрозаводск, 2000.
5. Лебедев В.А., Старкова В.Г., Брагин С.В. Система доступа к коллекции знаний по экологии региона. //Труды Второй Всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. Протвино, 2000.

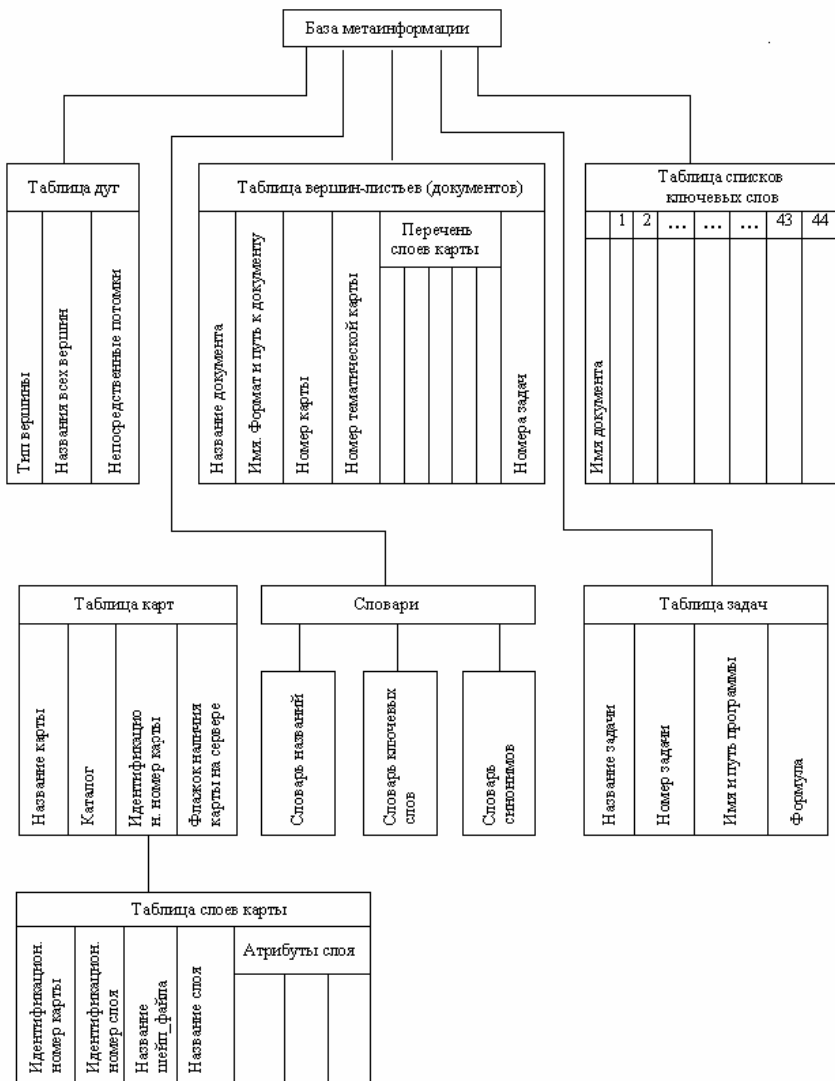


Рис. 1. Структура базы метainформации

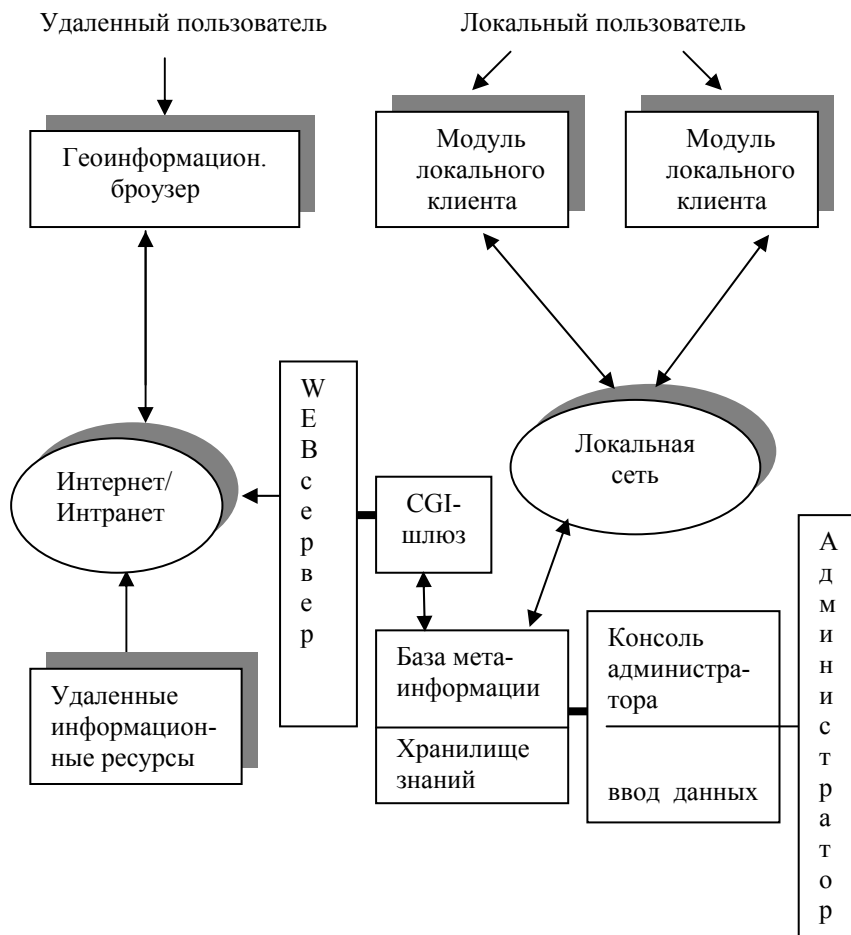


Рис 2. Архитектура коллекции

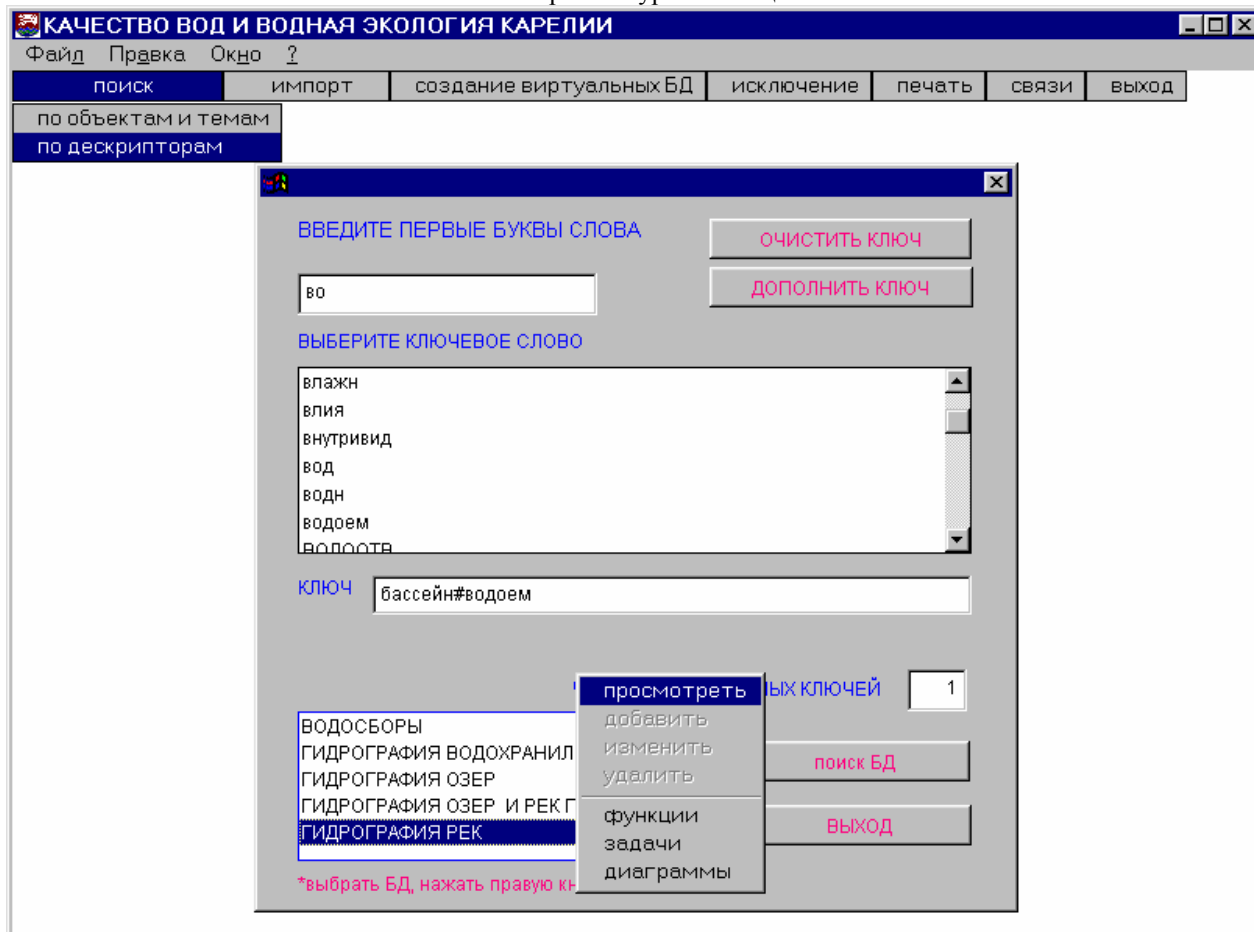


Рис. 4. Меню поиска документа по ключевым словам

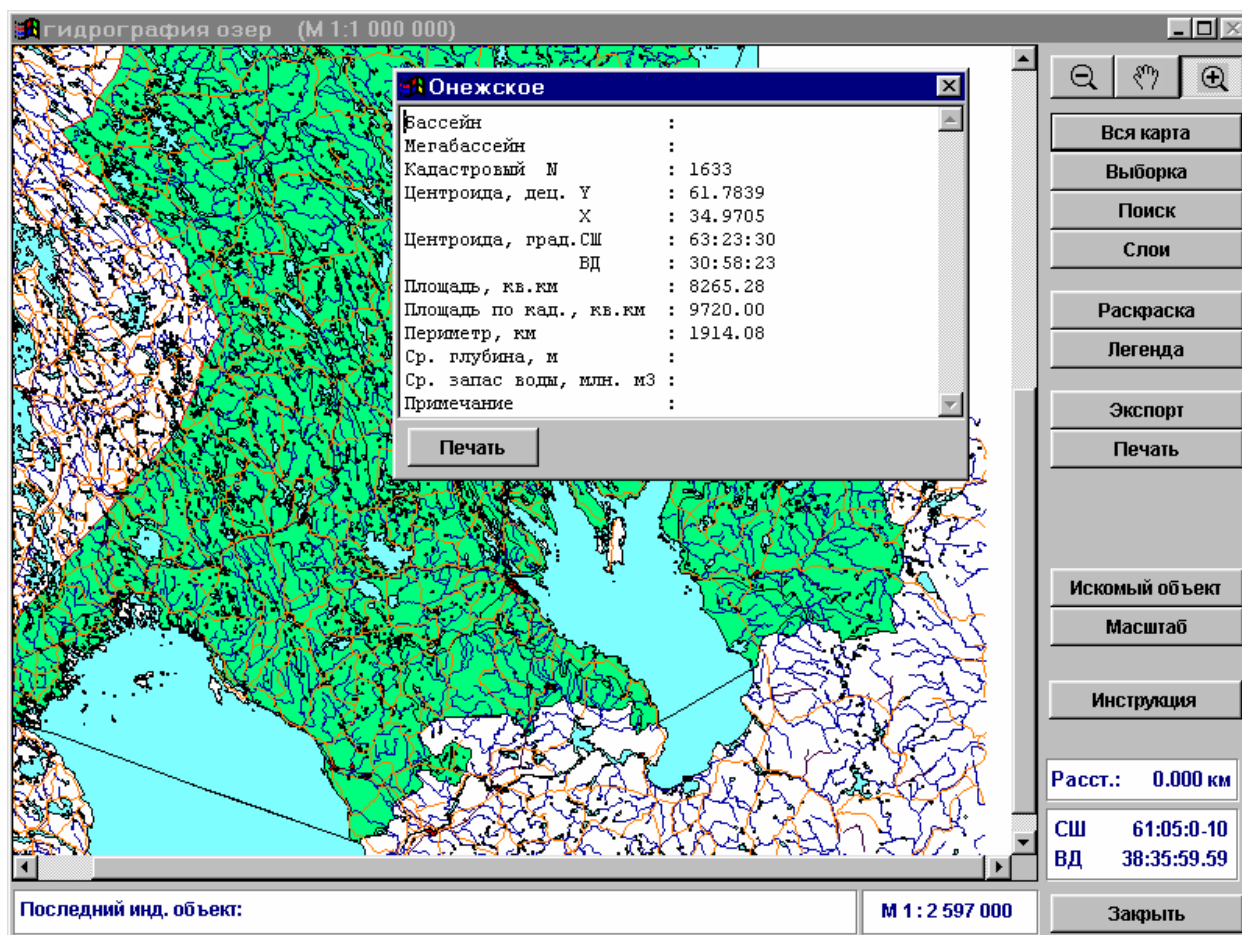


Рис. 5. Меню действий с картой и вызов описания объекта

KNOWLEDGE COLLECTION ON LAKES ONEGA AND LADOGA

N. Filatov

Northern Water Problems Institute, Karelian Research Centre of RAS

50 A. Nevskogo St., Petrozavodsk

nfilatov@nwpi.karelia.ru

V. Lebedev

Institute of Applied Mathematical Research, Karelian Research Centre of RAS

11 Pushkinskaya St., Petrozavodsk

lebedev@krc.karelia.ru

Architecture of the knowledge collection on Lakes Onega and Ladoga is described. The structure and content of the metainformation base, functions and user/collection interface are considered.