

ПУБЛИКАЦИЯ В ИНТЕРНЕТ КОЛЛЕКЦИЙ ТИПА СПРАВОЧНИКОВ

В.А. Лебедев, С.В. Брагин, В.Г. Старкова

Институт прикладных математических исследований Карельского научно-го центра РАН, г. Петрозаводск, Пушкинская, 11

E-mails: lebedev@krc.karelia.ru, bragin@krc.karelia.ru, tarkova@krc.karelia.ru.

Описывается прототип технологии публикации в интернет класса справочников об изученности сложных географических объектов (территорий, озер, каналов), обеспечивающей интерактивный доступ за небольшое число шагов к статьям, базам данных, картам по оглавлению, по предметному и именовому указателям.

ONLINE PUBLICATION OF REFERENCE COLLECTIONS

V. Lebedev, S. Bragin, V. Starkova

Institute of Applied Mathematical Research, Karelian Research Centre of RAS
11 Pushkinskaya St., Petrozavodsk

E-mails: lebedev@krc.karelia.ru, bragin@krc.karelia.ru, tarkova@krc.karelia.ru.

The pilot technology for online publication of reference collections on complex geographical objects (areas, lakes, channels) providing interactive access to entries, databases, maps in a few steps via the table of contents, subject and name indexes is described.

Разработан прототип технологии публикации в Интернет/интранет и произвольного доступа к коллекциям типа справочников об изученности и проблемах сложных географических объектов и областей. Коллекции этого типа отличаются сравнительно большим объемом структурированных, связанных сведений о предметной области, представленных в различных формах. Примером коллекций этого типа может служить создаваемая коллекция об Онежском и Ладожском озерах [1]. Традиционно к справочникам предоставляется произвольный доступ по оглавлению, по предметному и именовому указателям, кроме того, допускаются перекрестные ссылки между статьями. К электронным коллекциям этого типа необходимо сохранить эти виды доступа, обеспечивая дружелюбный интерфейс для пользователя.

Оглавление справочника представляет собой многоуровневую систему рубрик типа дерева. Включая в состав этого графа предметный и именной указатели, получаем систему, которую будем называть структурой знаний (структурная метаинформация). Она представляется ацикличес-

ским графом, в котором вершинами являются названия рубрик, термины и имена людей, а дугами – связи между ними.

Доступ к структурированным информационным ресурсам в Интернет/интранет обычно обеспечивается системой страниц, отображающих дерево рубрик. В рассматриваемом случае структура для доступа сложнее дерева и обладает значительным объемом. Уже это требует разработки специальных решений. Кроме того, следует учесть разнообразие форм представления сведений в виде текстов, иллюстрированных снимками, рисунками, графиками и таблицами, в виде баз данных с произвольным доступом (в частности, библиографий, представляемых как базы данных), в виде электронных географических карт со связанными базами данных.

Граф структуры знаний представляется в виде композиции пары: множества всех вершин и множества дуг, связывающих пары вершин. Вершины помечены, пометы указывают тип вершины и тип (формат) знаний, которые она представляет. Конкретно структура оформлена в виде базы данных с произвольным доступом [1, 2, 3], которая помимо структурных содержит семантические и адресные метаданные информационных ресурсов справочника.

В основу методики публикации справочников положена идея доступа к его информационным ресурсам через реляционную базу метаданных (БМД) [1] за небольшое число шагов посредством динамических форм. В целом схема публикации представлена на рис. 1.

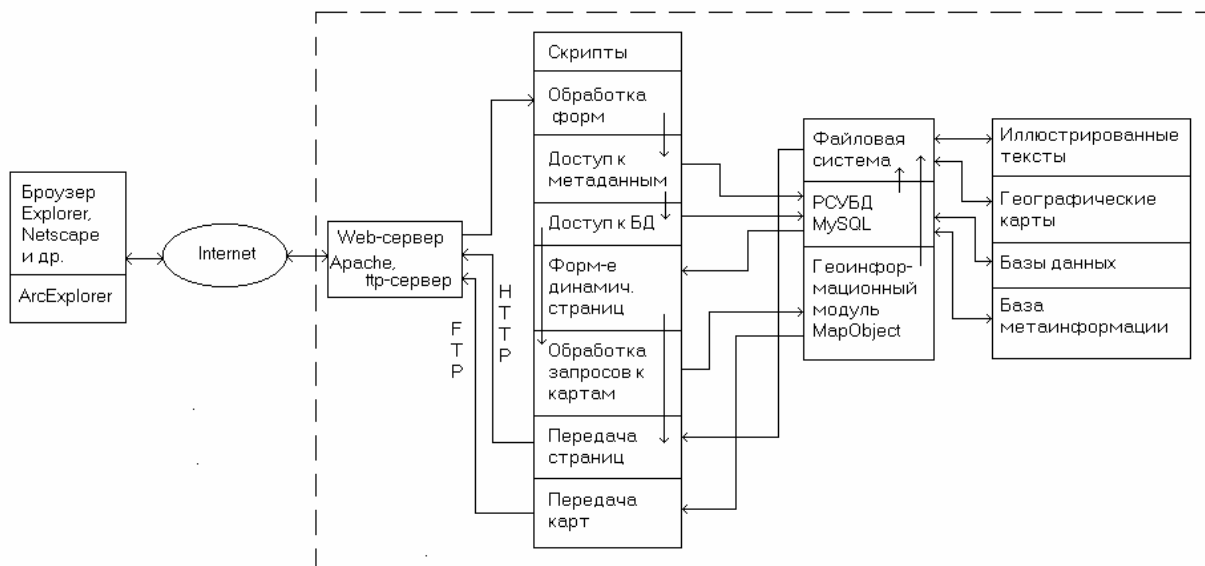


Рис. 1. Схема публикации справочника

Совокупность данных справочника в хранилище данных подразделяется на три секции: тексты в формате HTML, базы данных в формате MySQL и электронные карты в формате Arcview. Каждая из секций содер-

жит множество информационных файлов, соответствующих вершинам-листьям структуры знаний. Запись и чтение файлов, а также контроль их целостности выполняются файловой системой выбранной платформы (Linux, Windows NT или др.). Доступ к реляционным базам данных осуществляется при помощи СУБД MySQL, а к картам – при помощи геоинформационного модуля на базе пакета MapObject LT. База метаданных представляет из себя реляционную базу данных и состоит из ряда таблиц, содержащих структурные, семантические и адресные сведения. В частности, таблицы статей (документов), карт, ключевых слов и словарь названий предназначены для хранения семантических, форматных и адресных сведений.

Для публикации в Интернет необходимо обеспечить удобный для пользователей интерфейс и его интерпретацию сервером для управления MySQL, MapObject и файловой системой, формированием и передачей страниц отклика на запросы. В качестве Web-сервера установлен сервер Apache, в котором объявлена совокупность динамических форм, составляющих интерфейс доступа к справочнику и обеспечивающих интерактивность сеанса доступа. Для обработки информации форм разработан ряд скриптов Web-сервера. Для передачи карты установлен ftp-сервер.

Доступ к информации выполняется за ряд шагов. Причем для каждого последующего шага форма должна содержать информацию, определяемую на предыдущем шаге, что обеспечивает разветвление путей доступа по секциям информации и приближение к требуемому ресурсу.

В соответствии с этим определен состав скриптов и взаимодействие между ними (см. рис. 1). Скрипт обработки форм интерпретирует полученные данные и передает их скрипту доступа к метаданным и доступа к БД, который формирует оператор Select. Отклик БМД в зависимости от типа и формата ресурса направляется либо непосредственно файловой системе для передачи статических страниц или карт, либо на отбор данных из БД для формирования следующего оператора Select. Выборка из БД передается скрипту формирования динамических страниц, либо скрипту обработки запросов к карте. Последний обрабатывается модулем на базе MapObject, а результат передается клиенту.

Компьютер клиента должен быть укомплектован стандартным браузером (MS Explorer, Netscape), а также браузером карт ArcExplorer.

Важное значение в технологии публикации справочников придается сервисным функциям по формированию БМД, дополнению и изменению информационных ресурсов справочника, в частности функции формирования словарей для отбора данных, формирования гиперссылок для предметного и именного указателей и др.

Отметим способ создания предметного и именного указателей. Для коллекции об Онежском и Ладожском озерах в силу междисциплинарности предметной области списки терминов и названий указателя формиру-

ются по каждой дисциплине в отдельности специалистами этих дисциплин. Эти списки объединяются и упорядочиваются по алфавиту. Для создания системы гиперссылок разработана специальная программа. Она просматривает тексты статей, сравнивая выделяемые слова с основой каждой лексемы указателя. В случае совпадения термина с одним из слов текста оно отмечается, и происходит переход к следующему слову текста. По окончании просмотра очередной статьи к отмеченным терминам присоединяются гиперссылки на данную статью. По окончании просмотра всех статей и баз данных получаем указатель в виде базы данных, в которой каждому термину указателя соответствует список гиперссылок. Выбирая некоторый термин в структуре знаний, пользователь получает отклик в виде списка гиперссылок. Отмечая их по очереди, он может получить доступ к каждой из статей, содержащих данный термин.

Доступ к коллекции по Интернет организован следующим образом. В начальной странице (см. рис. 2) пользователь выбирает озеро и способ доступа. При выборе доступа по оглавлению вызывается список корневых рубрик структуры знаний. После выбора любой из них на стороне Web-сервера инициируется скрипт, выполняющий поиск в базе данных структуры знаний.

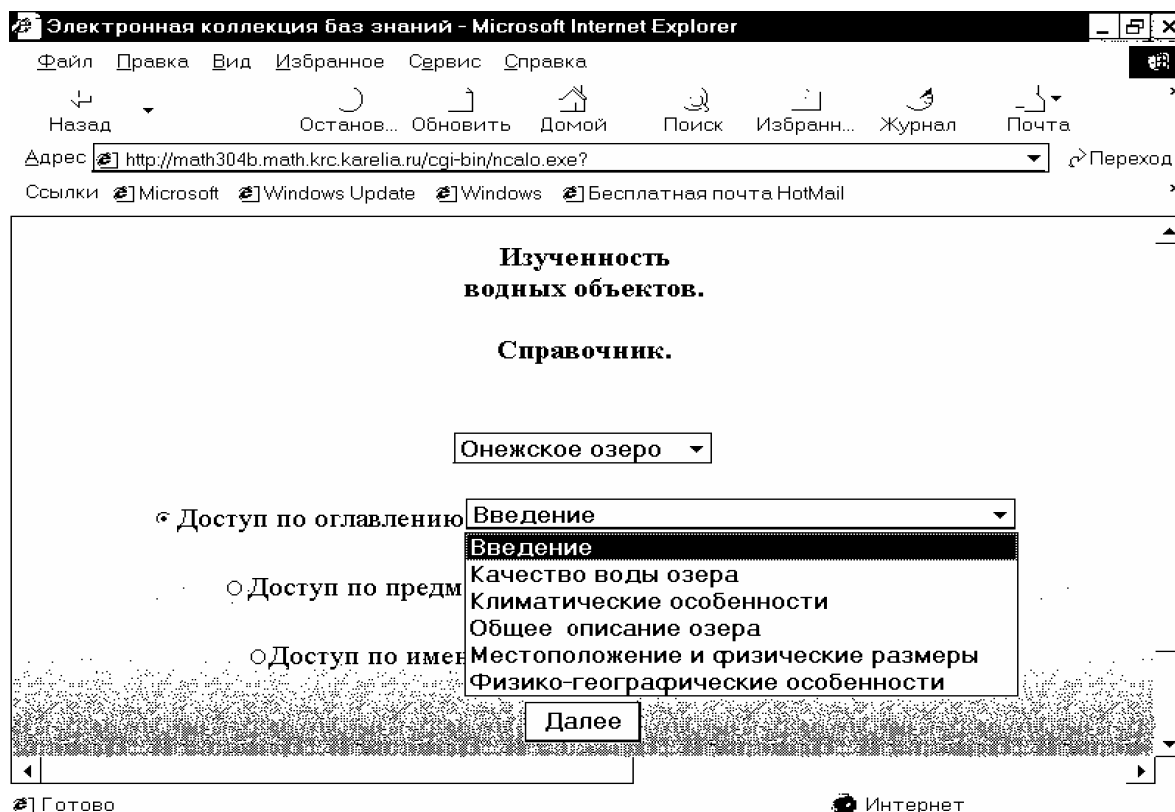


Рис. 2.

Результат поиска представляет собой перечень висячих вершин (листьев), связанных с выбранным корнем. Он динамически присоединяется к форме в виде списка для окончательного выбора интересующего клиента фрагмента знания. На рис. 3 показан пример формы для окончательного выбора фрагмента знаний. Выбранный фрагмент может быть статьей (текстом в формате HTML), базой данных или географической картой в формате ArcView. Формат выбранного фрагмента определяется скриптом по структуре знаний, и в зависимости от этого он формирует страницу отклика. Статьи передаются клиенту в виде обычных статических страниц. Если выбранный фрагмент - база данных, возможны два случая: либо база данных читается полностью, либо из нее выбирается некоторый фрагмент. В первом случае доступ к ней и формирование ответа выполняются почти по обычному CGI-сценарию. Дополнительно формируется колонтитул таблицы с русскими названиями колонок. Во втором случае пользователю выдаются две динамические формы. Сначала форма со списком названий колонок, в которой пользователь должен отметить колонки для формирования условия отбора записей и колонок, помещаемые в результирующий фрагмент. После этого ему выдается форма для задания условия отбора записей. Она содержит названия колонок, отобранных на предыдущем шаге, и соответствующие окна для задания значений (см. рис. 4).

Озера водосбора

Формирование условия отбора записей

Поля БД	Значение	Связь
ПЛОЩАДЬ ЗЕРКАЛА (КВ. КМ)	> 500	И
ПЛОЩАДЬ ЗЕРКАЛА (КВ. КМ)	< 1000	
И ДЛИНА БЕРЕГОВОЙ ЛИНИИ (КМ)	> 500	ИЛИ
ДЛИНА БЕРЕГОВОЙ ЛИНИИ (КМ)	>	

База Данных Карта

Рис. 3.

Значения числового типа набираются пользователем в соответствующих окнах, а для указания символьных значений колонки сопровождаются словарями, из которых он выбирает нужные значения. Количество отбираемых значений в одной колонке за один сеанс отбора – не более 10. При этом они должны быть разделены знаком союза «ИЛИ». Условия по колонкам соединяются между собой знаками «И» и «ИЛИ» по смыслу. Форма сопровождается необходимой инструкцией по формированию условия отбора.

Некоторые базы данных (а также некоторые карты) могут быть открыты в Интернет условно. Для таких фрагментов предусмотрена форма для задания регистрационного имени пользователя и пароля. Необходимость передачи этой формы определяется скриптом по типу знания, указанному в базе данных структуры знаний.

Для чтения и визуализации географических карт используется модуль ArcExplorer, который вызывается из специальной страницы. Возможны два входа для вызова карты. Если карта зарегистрирована в структуре знаний, то после выбора соответствующего названия скрипт Web-сервера сразу передает клиенту страницу для запуска ArcExplorer с необходимыми параметрами чтения карты.

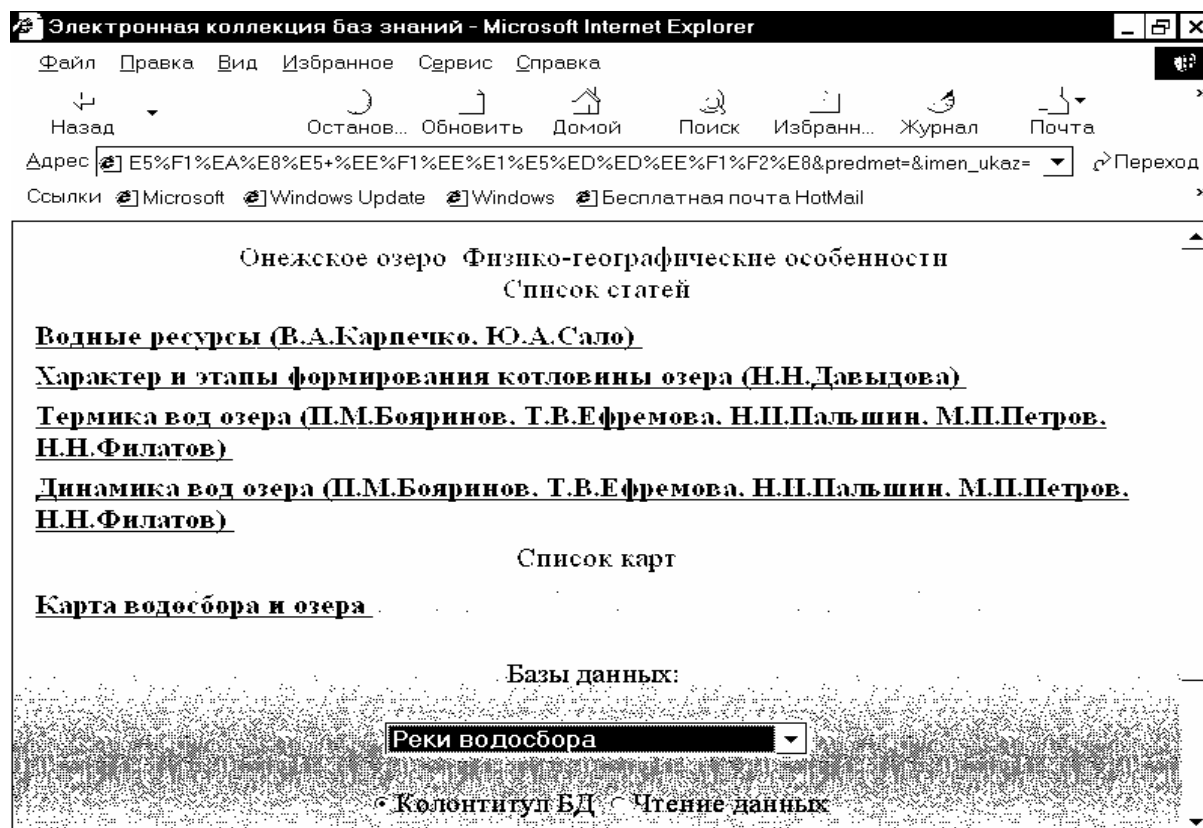


Рис. 4.

Другой вход вызова карты предоставлен из страницы отбора записей из базы данных, если последняя пространственно связана с картой. В этом случае после сформирования условия отбора записей и нажатия кнопки «карта» запускается специальный скрипт web-сервера, который по условию отбора создает временный тематический слой карты, содержащий объекты, релевантные условию отбора, и окрашивает их контрастным цветом. По окончании этой работы, как и в предыдущем случае, клиенту передается страница для запуска ArcExplorer, который читает и показывает карту с выделенными объектами.

Если пользователь выбирает доступ по предметному или именному указателю, он может набрать первые буквы термина или фамилии. В этом случае на следующем шаге ему выдается соответствующий фрагмент словаря указателя. Иначе ему придется делать выбор из словаря большого объема. Далее ему выдается список ссылок на документы, релевантные его выбору, которые он может просмотреть по очереди.

Таким образом, разработан прототип технологии публикации и произвольного Интернет-доступа к электронному справочнику, предоставляющей доступ к любой статье, базе данных или карте за два-четыре шага как по оглавлению, так и по предметному или именному указателю. При этом предметный и именной указатели обеспечивают доступ внутрь статей. Выбор записей баз данных производится по произвольному составу колонок, кроме того, по условию отбора записей из баз данных, связанных с картой, образуется временный тематический слой, содержащий релевантные объекты, который демонстрируется удаленному пользователю на фоне основной карты.

Литература

- [1] Филатов Н.Н., Лебедев В.А. Коллекция знаний об Онежском и Ладожском озерах. //Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Петрозаводск, 2001, с. 202-208.
- [2] Лебедев В.А. О методике построения электронных коллекций знаний. //Методы математического моделирования и информационные технологии. Вып. 2. Петрозаводск, 2000, с. 55-65.
- [3] Лебедев В.А., Брагин С.В., Старкова В.Г. Система доступа к коллекции знаний по экологии региона // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Протвино, 2000, с. 185-187.