

ОПЫТ КОНВЕРСИИ ТРАДИЦИОННЫХ КАРТОТЕК В ЭЛЕКТРОННУЮ ФОРМУ НА ПРИМЕРЕ КАРТОТЕКИ МОСКОВСКОГО ОБЩЕСТВА ИСПЫТАТЕЛЕЙ ПРИРОДЫ

A STUDY OF CONVERSION OF TRADITIONAL CARDS CATALOGS INTO ELECTRONIC FORM, WITH THE CATALOG OF THE MOSCOW SOCIETY OF NATURALISTS AS AN EXAMPLE

Леонов Михаил Васильевич

Кандидат биологических наук, доцент, зав. лабораторией вычислительного практикума и информационных систем факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова
E-mail: Leonow_M_W@cs.msu.su

Mikhail V. Leonov

Киселева Елена Александровна

Инженер лаборатории вычислительного практикума и информационных систем факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова
E-mail: lenakisseleva@gmail.com

Elena A. Kisseleva

Орлов Александр Сергеевич

Кандидат исторических наук, профессор, заведующий Музеем истории МГУ им. М. В. Ломоносова
E-mail: AS-Orlov@bk.ru

Alexander S. Orlov

Рассматривается создание информационной системы на основе мемориальной картотеки членов Московского общества испытателей природы с трудно распознаваемым содержимым карточек. Предложенное решение может быть использовано для вовлечения в научный оборот и других исторически важных рукописных картотек. Разработанное на основе технологии HTML5 программное обеспечение функционирует как на компьютерах под управлением операционной системы (ОС) Windows, так и на мобильных устройствах (планшетах) под управлением ОС Android.

Ключевые слова: информационная система, конверсия традиционных картотек, МОИП, HTML5, WebSQL

The article is about the development of an information system based on the memorial catalog of the Moscow Society of Naturalists, many cards in which are difficult to read. The proposed solution can also be used in the case of any other historically important handwritten material. Our software is based on HTML5 technology; it can operate on Windows-compatible computers as well on mobile devices under the Android operating system.

Keywords: informational system, conversion of traditional catalogs, Moscow Society of Naturalists, HTML5, WebSQL

В XIX и первой половине XX в. основным инструментом для сбора и интеграции тематически связанных данных были картотеки. Некоторые из них вошли в «золотой фонд» истории культуры. Известный пример — картотека С. А. Венгерова с ее двумя миллионами карточек о 80 тысячах деятелей русской культуры. Существуют и другие картотеки, более скромные по количеству карточек, но не потерявшие свое значение и поныне и сохранившиеся в первозданном виде. Доступ к ним по естественной причине ограничен, например, картотека библиотеки Д. П. Сырейщикова¹ на биологическом факультете МГУ находится в подвале служебного помещения библиотеки. К таким важным картотекам принадлежат и картотеки Московского общества испытателей природы (МОИП). Наиболее известная из них содержит данные по членам МОИП за 1806–1855 гг.

Задача переноса данных со старых карточек в компьютерное представление весьма трудоемка. При этом иногда по различным причинам трудно или неразумно осуществлять с них набор текстов. Как правило, эта работа требует квалифицированного труда по расшифровке рукописных пометок, что часто просто невозможно осуществить. Поэтому обычно используют метод так называемых имидж-картотек, или имидж-каталога. При этом карточки сканируются и хранятся в памяти компьютера с сопровождающей (обычно краткой) информацией в виде обычного компьютерного текста, по которому можно проводить поиск. Этот метод был применен и во многих крупных библиотеках страны, например, в Научной библиотеке МГУ.

На факультете вычислительной математики и кибернетики МГУ такой подход был использован в проекте компьютеризации так называемой дореволюционной картотеки членов МОИП. Московское общество испытателей природы — старейшее из ныне существующих научных обществ России. На протяжении многих десятилетий оно объединяло в своих рядах научную элиту и представителей просвещенных слоев России. История этого общества — важная составляющая истории отечественной и мировой науки. Г. Ю. Любарский в своей замечательной книге² приводит мысль философа А. А. Зиновьева о том, что история отдельной организации часто дает больше информации об эпохе и стране, нежели дорогостоящие социологические исследования. Г. Ю. Любарский по этому поводу пишет следующее. «... любая организация «выбалтывает» тайны общества, кирпичиком которого она является. Лучше, однако, чтобы организация была не слишком уж мимолетной, тогда на ее примере можно изучать большой период жизни общества...». Это в полной мере относится к МОИП.

К сожалению, до настоящего времени архив общества в основном существует в виде рукописных материалов — протоколов, писем, картотек. До сих пор нет полного списка членов общества. В работе³ наиболее известного историка общества С. Ю. Липшица, посвященной 135-й годовщине МОИП, высказана надежда, что такой список появится хотя бы к 150-летию... Так вот, списка не появилось ни к 200-летию, ни к 205-летию.

Отсутствие такого списка приводит к неприятным казусам. Приведем только один. В 2012 г. появилась работа⁴, в которой, в частности, перечислялись известные люди, якобы являвшиеся членами МОИП. Это ученые Лейбниц, А. Эйнштейн, К. Гаусс, композитор М. И. Глинка, немецкие профессора из знаменитой «геттингенской семерки». На самом же деле ни один из этих людей не был членом МОИП. Были два однофамильца геттингенских профессоров, т. е. не филологи братья Гримм, а русский зоолог Оскар Андреевич Гримм, не геттингенец Эдуард Альбрехт, а немецкий анатом Карл-Мартин-Пауль Альбрехт, не великий композитор Михаил Глинка, а его однофамильцы, также оставившие след в истории России, и т. д.

Остановимся немного на выборе стратегии подготовки данных для процесса компьютеризации картотеки. Самый «прямой» с абстрактно научной точки зрения способ — это перевести в текстовую форму содержимое карточек. Тогда удобно было бы проводить всевозможные поиски и т. д. Но специфика этих конкретных карточек в том, что они хранят следы нескольких поколений секретарей, технических и научных работников, имевших доступ к этим карточкам. Некоторые фамилии трудно разобрать полностью, есть противоречивые данные на карточках, явно относящихся к одному и тому же лицу, есть кляксы прямо на фамилии, очень много специфичных сокращений, расшифровать которые не могут нынешние сотрудники библиотеки и Архива МОИПа. Да и вообще, канонический метод разработки баз данных (БД) для данного случая не подходит: карточки содержат библиографию, и по программистской науке надо бы создать таблицы, соответствующие этим библиографическим источникам, и устанавливать связи «многие-ко-многим» между персонами и библиографическими источниками. Кроме того, на карточках есть и многочисленные сокращенные ссылки на протоколы, письма, хранящиеся в архиве и т. п. Отсюда следует, что здесь «с ходу» классические методы применять нельзя.

Поэтому было принято компромиссное, но, как выяснилось, вполне адекватное решение. Мы начали сканировать карточки и готовить к вводу в информационную систему, которой еще не было. Параллельно набирались следующие краткие све-

дения по каждому члену МОИП: фамилия, имя, отчество, примечание к имени, дата рождения, дата смерти, дата принятия в члены или члены-корреспонденты (с пометкой о статусе, т. е. «д.ч», «чл-корр», «ученик» и т. д.), «специализация или секция». Почему были выбраны только эти поля? Любое решение было бы компромиссным, как уже было сказано, полную информацию набирать было бы неразумно из-за невозможности правильно расшифровать почерк и сокращения, что потребовало бы гораздо больше времени. Ограничиваться только фамилиями и именами было бы тоже неразумно, по нашему мнению. Поэтому-то и было принято решение вводить эти данные в MS Excel. Пакетный ввод данных был нами опробован еще в эпоху перфокарт при создании БД по кариологии растений семейства Зонтичных в Ботаническом саду МГУ. Когда разработка информационной системы идет параллельно с подготовкой большого количества данных для ввода, в процессе «додумывания» запросов к базе данных и других ее свойств может потребоваться модификация структуры БД, и тогда почти неизбежен повторный начальный ввод данных. Поэтому пакетный ввод представляется весьма удобным средством. Кроме того, уже подготовленные Excel-таблицы можно использовать для некоторых статистических прикидок, расчетов, проверок правильности подготовленных материалов и т. д. В MS Excel есть возможность экспорта данных в формате csv, что, в свою очередь, поможет при вводе в таблицу под управлением системы управления данными.

Вся картотека занимала двенадцать плотно заполненных карточками стандартных библиотечных ящиков. Первым был обработан ящик, который начинался с карточек под буквами «Го» — нас тогда очень интересовал Георг Франц Гофман — первый директор Ботанического сада Московского университета. Не сразу, но довольно быстро стало понятно, что надо правильно выбрать способ именования изображений карточек, чтобы облегчить контроль правильности ввода, и вообще возможность некоторого «распараллеливания» работы.

Грубая прикидка количества всех карточек, точнее, их поверхностей, дала число около 20000. Мы решили отдельно нумеровать персоны внутри каждой буквы, т. е. g0001_1 — для лицевой карточки первой персоны на букву «Г», g0001_1a — для тыльной стороны той же карточки, g0001_2 — для второй карточки для той же персоны, g0001_2a — для ее тыльной стороны и т. д. Было решено укладывать в определенном порядке на поле сканирования по 4 карточки, сканировать их, соответствующим образом именовать файлы. Например, l0235_1_l0236_1_l0236_2_l0236_3.jpg. Эти файлы-изображения потом «разрезались»

по карточкам. Техническая работа заняла примерно год. За это время была запрограммирована и отлажена и несложная система просмотра изображений карточек с поиском по фамилии. Одной из очень полезных функций первоначального варианта системы была возможность распечатки карточек, упорядоченных по алфавиту фамилий (рис. 1).

Итак, на первом этапе была создана «электронная копия» всей картотеки, чем в значительной степени гарантирована сохранность содержащейся в ней информации.

Второй этап решения задачи — разработка информационной системы для анализа полученного архива изображений карточек. В этом архиве оказалось 12087 карточек по 5641 персоне, и если учитывать и тыльные стороны карточек, то более 17000 изображений.

Учитывая современные тенденции развития аппаратного обеспечения, а также отчасти разделяя опасения проф. М. Талера⁵ о проблемах, которые может принести так называемая мобильная революция, мы остановились на реализации оболочки системы средствами технологии HTML5 в среде браузера GoogleChrome. Это означает, что оболочка без какой-либо адаптации сможет работать как на персональных компьютерах под управлением пока самой распространенной операционной системы (ОС) MS Windows, так и на планшетах под управлением ОС Android.

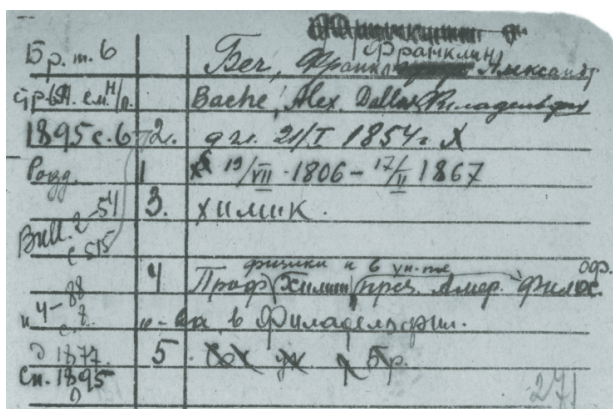
АРХИТЕКТУРА ИНФОРМАЦИОННОЙ СИСТЕМЫ

Система представляет собой приложение для браузера GoogleChrome, использует библиотеки jquery и jqueryUI и технологию WebGL. Предусмотрено два режима работы: административный и читательский (пользовательский).

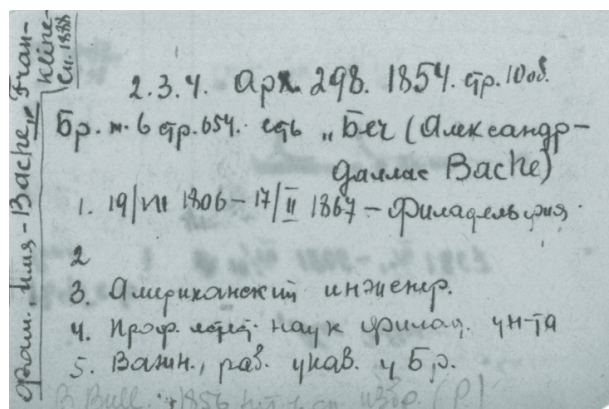
В читательском режиме (рис. 2) в правом верхнем углу находится форма поиска по любой комбинации фамилии, имени, отчества и году рождения. Можно также просматривать список фамилий, указав первую букву фамилии. В левой части окна программы расположены четыре вкладки: «Основная информация», «Дополнительная информация», «Карточки», «Источники». На рисунке 2 отражено состояние окна с открытой первой вкладкой и значениями следующих полей: фамилия, имя, отчество, дата рождения, дата смерти, основные занятия члена МОИП, год принятия в общество и признак почетного члена. Предусмотрены также поля для фамилии и имени на иностранном языке.

К вкладке «Дополнительная информация» отнесены в основном те поля, значения которых на первом этапе обработки не были внесены в БД,

Беч Александр-Даллас 19.07.1806–17.02.1867 химик, инженер b0195

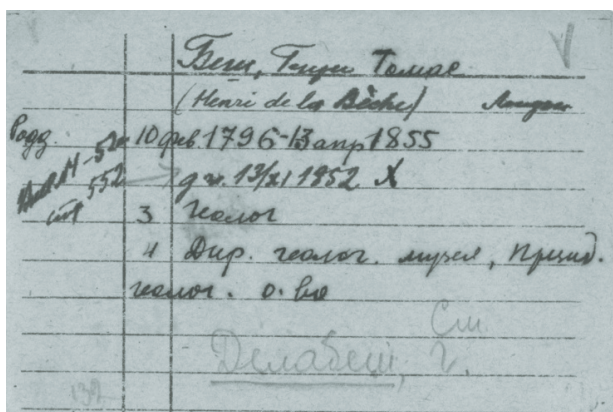


images\195\1_0.jpg, 1, 0, b0195_1.jpg

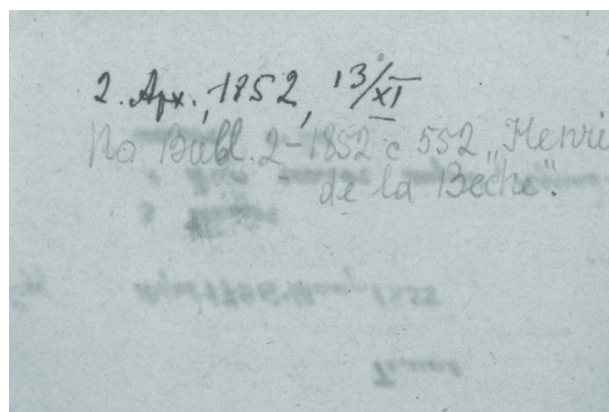


images\195\1_1.jpg, 1, 1, b0195_1a.jpg

Беш Генри Томас 10.02.1796–13.04.1855 геолог b0196

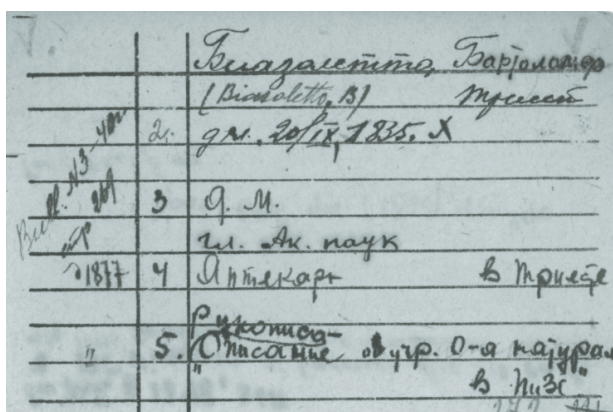


images\196\1_0.jpg, 1, 0, b0196_1.jpg

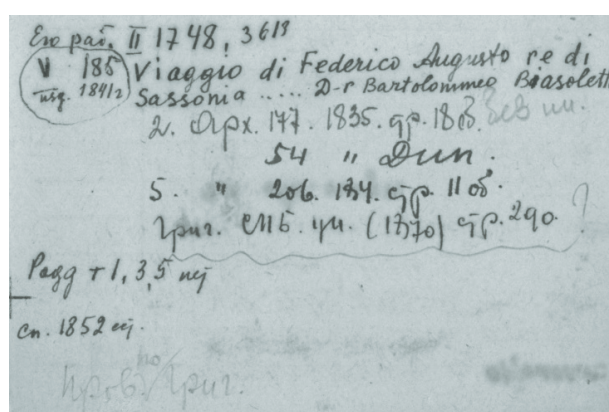


images\196\1_1.jpg, 1, 1, b0196_1a.jpg

Биазолетто Бартоломео **.*.*.* — **.*.*.*.* д.м. b0197



images\197\1_0.jpg, 1, 0, b0197_1.jpg



images\197\1_1.jpg, 1, 1, b0197_1a.jpg

Рис. 1. Фрагмент вывода изображений карточек членов МОИП на букву «Б»

а также те, которые в явном виде не присутствуют в оригинале карточки, но могут пригодиться при анализе данных. Это страна и город, который представляет член МОИП, признак иностранного члена, примечания к имени и фамилии (когда известны несколько форм его написания), поле для краткой биографической справки. Предусмотрены три признака для фильтрации списков, которыми пользователь может пометать отдельных членов, и т. д. Эти поля впоследствии помогут при-

годиться в исследовательской работе специалиста, в том числе для формирования списков для печати.

Раскрыв вкладку «Карточки», пользователь увидит в режиме предварительного просмотра уменьшенные изображения лицевых и тыльных сторон карточек исходной картотеки (рис. 3). Карточек может быть от одной до нескольких десятков. Щелчок по уменьшенному изображению карточки приводит к ее визуализации в увеличенном масштабе, вполне достаточном для чтения (рис. 4).

Мемориальная картотека МОИП 1805-1955

ВМК МГУ имени М.В. Ломоносова
Лаборатория вычислительного практикума и информационных систем
версия 19.06.13

Имя:

Фамилия:

Отчество:

Год рождения:

Всего карточек : 458 Персон: 17544

Выберите букву:

- Основная информация

Фамилия :	Щуровский
Имя :	Григорий
Отчество :	Ефимович
Name :	
Vorname :	
Дата * :	30.01.1803
Дата † :	20.03.1884
Деятельность :	геолог, минералог
Признак почетного члена :	☐
Год принятия в МОИП :	
- Дополнительная информация
- Карточки
- Источники

Щербатов Григорий Алексеевич
Щербатов Григорий Алексеевич
Щербатов Дмитрий Михайлович
Щербатов Дмитрий Михайлович
Щербачев Дмитрий Михайлович
Щербачев Дмитрий Михайлович
Щировский
Щировский
Щировский Владимир Алексеевич
Щировский Владимир Алексеевич
Щодро Николай Ксаверьевич
Щодро Николай Ксаверьевич
Шукин Иван Семенович
Шукин Иван Семенович
Шукин Семен Семенович
Шукин Семен Семенович
Шукина Антонина Васильевна
Шукина Антонина Васильевна
Шукина Елена Николаевна
Шукина Елена Николаевна
Шупаков Иван Григорьевич
Шупаков Иван Григорьевич
Щуровский Григорий Ефимович

Рис. 2. Основное окно системы с основными данными по Г. Е. Щуровскому

Мемориальная картотека МОИП 1805-1955

ВМК МГУ имени М.В. Ломоносова
Лаборатория вычислительного практикума и информационных систем
версия 19.06.13

Имя:

Фамилия:

Отчество:

Год рождения:

Всего карточек : 12087 Персон: 5641

- Основная информация
- Дополнительная информация
- Карточки

0/1 1/1 Карточка №1 Лицевая сторона 1 3/1 3/2 4/1 5/1 6/1 7/1 8/1 9/1 10/1

Щеголев Сергей Федорович
Щелкачев Владимир Николаевич
Щепкин Павел Степанович
Щепотьев
Щербаков Александр Владимирович
Щербаков Алексей Павлович
Щербаков Анатолий Петрович
Щербаков Сергей Васильевич
Щербакова Антонина Алексеевна
Щербакова Нина Александровна
Щербаков Алексей Григорьевич
Щербатов Григорий Алексеевич
Щербатов Дмитрий Михайлович
Щербачев Дмитрий Михайлович
Щировский
Щировский Владимир Алексеевич
Щодро Николай Ксаверьевич
Шукин Иван Семенович
Шукин Семен Семенович
Шукина Антонина Васильевна
Шукина Елена Николаевна
Шупаков Иван Григорьевич
Щуровский Григорий Ефимович

Рис. 3. Окно системы с открытой вкладкой «Карточки» на примере карточек, относящихся к Г. Е. Щуровскому

Вкладка «Источники» должна содержать перечень ссылок, которые указаны для соответствующего члена МОИП. Это не только библиографические ссылки, но и ссылки на письма из архива МОИП и другие рукописные материалы. Заполнение таблицы «Источники» предполагается осуществить на одном из будущих этапов проекта.

Для дополнительных запросов к БД предназначена вкладка, которая вызывается щелчком по гиперссылке «Расширенный поиск» в правой верхней части экрана. В текущей версии формы расширенного поиска можно указать диапазон фамилий, признак почетного члена, город и поль-

зовательские фильтры (отбор по фамилиям, вручную отмеченным пользователем). В дальнейшем предполагается увеличить число параметров поиска, а также автоматизировать подготовку формата выводимого ответа на запрос. Нажатие кнопки «Показать результаты поиска» приводит к появлению всплывающего окна с текстом ответа. Результаты расширенного поиска можно сохранить в файл, выбрав из меню от правой кнопки мыши «Сохранить как». Кроме того, результаты можно скопировать в буфер обмена стандартным образом (ctrl+a, ctrl+c, ctrl+v).

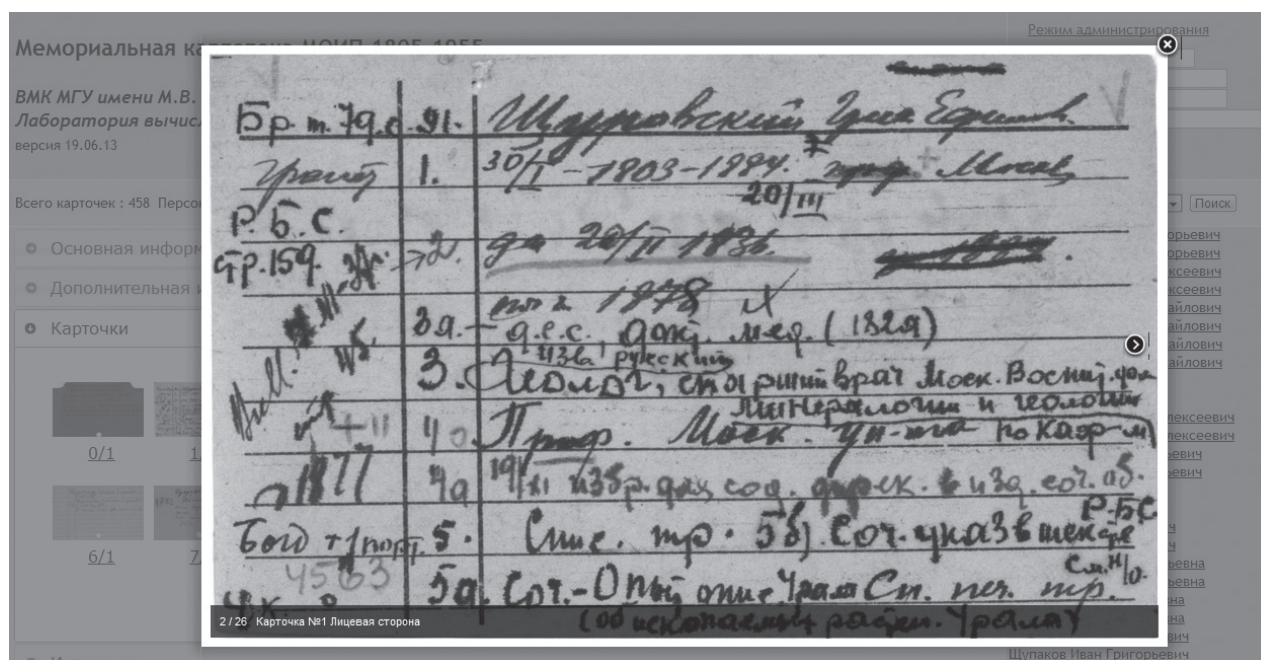


Рис. 4. Карточка в режиме просмотра одной из карточек

АДМИНИСТРАТИВНЫЙ РЕЖИМ

В административном режиме возможно редактирование значений всех полей, удаление карточек, удаление персон, загрузка данных методом пакетного ввода, управление интерфейсом для пользовательского режима (выбор видимых полей), а также управление форматами вывода на запросы пользователя.

ПАКЕТНЫЙ ВВОД

Без пакетного ввода трудно было представить заполнение БД таким большим количеством данных в обозримое время.

Формат пакетного ввода следующий:

[код персоны]; [фамилия]; [имя]; [отчество]; [примечание к имени]; [дата рождения]; [дата смерти]; [дополнительная информация]; [деятельность]; [файлы изображений карточек через запятую]

Пример строки файла пакетного ввода:

a0001; Абовян; Хачатур;;; 1805; 1848; д.ч. 1836; писатель, этнограф, просветитель; a0001_1.jpg, a0001_1a.jpg, a0001_2.jpg, a0001_2a.jpg, a0001_3.jpg, a0001_3a.jpg

Название файла изображения карточки состоит из кода персоны, номера карточки. Буква «а» вслед за номером обозначает тыльную сторону карточки. Текст параметров подготовлен в формате utf-8, чтобы наряду с русскими и стандартными латинскими использовать немецкие умляути и т. п.

РЕЗУЛЬТАТЫ, ВЫВОДЫ И ПЕРСПЕКТИВЫ

Создана «электронная копия» картотеки членов МОИП, тем самым в значительной степени гарантирована сохранность содержащейся в ней информации и возможность ее дублирования для исследователей. Разработана информационная система для управления уже содержащимися в БД основными данными, обеспечивающая достаточно удобные механизмы для структурирования и внесения в БД остальных данных по членам МОИП. Продемонстрирована эффективность использования технологии WebSQL для разработки БД с более чем 17000 изображениями общим объемом более 6 гигабайт. Среди историко-архивных результатов — возможность получения быстрых справок по членам МОИП с 1805 г., довольно точной оценки численности общества за различные годы, использование данных электронной картотеки для ряда работ, в частности, по первому директору Ботанического сада МГУ Георгу-Францу Гофману⁶. Но главное — сделан существенный шаг к созданию общего биографического списка МОИП. Для достижения этой

цели необходим еще значительный объем работы по выяснению отсутствующих данных, устранению противоречий в имеющейся информации. В качестве перспективы развития проекта отметим еще возможность использования нашей системы в качестве ядра при создании электронного архива общества. Несомненно, после надлежащей содержательной обработки возможна организация доступа к этой электронной картотеке через Интернет.

В заключение авторы считают своим приятным долгом выразить свою благодарность за помощь, содействие в работе и поддержку заведующей библиотекой МОИП Ирине Викторовне Грачевой, а также декану факультета ВМК МГУ академику РАН Евгению Ивановичу Моисееву и заместителю декана Борису Ивановичу Березину. Отметим, что идея начать компьютеризацию архива МОИП с компьютеризации мемориальной картотеки принадлежит Сергею Михайловичу Завьялову, бывшему в 2008 г. заведующим архивом МГУ. Неоценим вклад в нашу работу бывшей сотрудницы факультета ВМК, техника Надежды Александровны Пахомовой (Орловой), аккуратно и добросовестно подготовившей несколько тысяч строк для пакетного ввода в систему.

ПРИМЕЧАНИЯ

- ¹ Ермакова Л. Р. Библиотека Д. П. Сырейщикова в фонде отдела биологии. Рукопись доклада на Ломоносовских чтениях 1987 года. Секция библиотковедения.
- ² Любарский Г. Ю. История Зоологического музея МГУ: идеи, люди, структуры. М., 2009. 744 с.
- ³ Липшиц С. Ю. Московское общество испытателей природы за 135 лет его существования (1805–1940). М., 1940.
- ⁴ Доклады Московского Общества Испытателей Природы. Т. 4. М., 2012. С. 29–51.
- ⁵ Таллер М. Дискуссии вокруг Digital Humanities // Историческая информатика. 2012. № 1. С. 5–13.
- ⁶ Леонов М. В. О малоизвестных источниках к биографии Георга-Франца Гофмана (к созданию биографической базы данных) // XII Московское совещание по филогении растений, посвященное 250-летию со дня рождения Георга-Франца Гофмана : материалы. М., 2010. С. 21–24.