

Создание интернет-ресурса по консервации документов

Е.Н. Пименов, В.А. Солдатов

Библиотека Российской академии наук
pen48@list.ru, Vadim.A.Soldatov@gmail.com

Аннотация

Описывается техника составления путеводителя по сетевой информации в области консервации документов. Рассматривается назначение путеводителя, потенциальные пользователи, работа по составлению рубрик, комплектование путеводителя, обработка запросов и некоторые прагматические основания путеводителя.

В настоящее время в Сети все большую популярность получает использование в качестве инструмента информационного поиска систематизированных ссылок по каким-либо темам. Данные средства информационного поиска называют по-разному и часто - веб-библиографиями или путеводителями по сетевой информации [1-6]. В тематической области консервации документов имеется более 50 зарубежных путеводителей, как, например, Conservation OnLine, Handmade paper on the net, Consortium of Academic and Research Libraries in Illinois, Preservation Readiness Webliography, Preserving Access to Digital Information, Audio preservation/restoration WebGuide. Аналогичных отечественных информационных ресурсов пока не имеется, и описываемый в этой работе продукт является первым в России.

1. Назначение и основные параметры путеводителя

Ресурс адресован специалистам по обеспечению сохранности документов, библиотекарям, хранителям, музееведам, биологам, химикам, технологам бумажного производства и реставраторам. Путеводитель может найти применение также в практике преподавания и в научной работе.

Путеводитель содержит две части: зарубежную информацию на английском языке и отечественную информацию на русском. Ссылки в путеводителе отражают более 50 тем, относящихся к наиболее обсуждаемым и актуальным проблемам в области консервации документов. Общее количество сайтов, представленных в путеводителе можно оценить в 10 тыс. по состоянию на сентябрь 2014 г.

В ресурсе имеются ссылки на библиографическую и полнотекстовую информацию. Библиографическая информация представлена ссылками на полные тексты библиографических указателей и запросами к 3 профильным базам данных и одной универсальной БД. Не библиографические данные в настоящем ресурсе имеют характер исследовательской, справочной и отчасти коммерческой полнотекстовой информации.

Ниже приводится первая часть обсуждаемого путеводителя – это ссылки на зарубежную (англоязычную) информацию:

People & Organizations (20)

Finding Conservation Organizations
Finding People in Conservation
Database of Conservators & Bookbinders

Paper and Cardboard (1760)

Types of paper and paperboard
Paper & Cardboard Manufacture
Paper Acidity & Deacidification
Paper Aging

Paper & Book Conservation (2650)

Reference Information (230)
General & Overview Information (160)
Professional Standards (190)
Preventive Conservation (170)
Bookbinding (470)
Book Repair & Restoration (304)
Manuscripts & Rare Books Conservation (325)
Leather Conservation (480)
Photographs Conservation (307)

Environment Management (730)

Environment Control (250)
Indoor Air Quality, Dust (190)
Lighting, UV & IR Radiation (170)
Temperature & Humidity (120)

Pest Management

Pest Control (310)
Integrated Pest Management (110)
Mold & Mildew (60 +)
Mass Treatment
Biocides

Disasters & Emergency

Disaster planning
Disaster recovery
Incidents (theft, bleachings)

Такие же рубрики выделяются и в русскоязычной части путеводителя. Близкие по содержанию классы связаны отношением род-вид (широкое/узкое понятие). Цифры в скобках обозначают количество сайтов, к которым в данном разделе имеются ссылки.

2. Предметы и аспекты рубрик. Списки тематических рубрик

Подготовка путеводителя включает в себя сбор ссылок и их форматирование. Сбор ссылок ведется путем проведения поиска по рассматриваемым рубрикам и оценку их адекватности и полезности для включения в путеводитель. Форматирование рубрик проводится по предметно-аспектной модели. В соответствии с этой моделью принято различать предметную часть (заголовок) предметной рубрики и аспектную часть ее содержания [7, С. 47-48]. При том, что понятие предмет информации может иметь очень разное содержание, в путеводителе оно семантически ограничено, и в этой функции выступают слова с предметно-вещественным категориальным значением.

В странице путеводителя, раскрывающей понятийную схему предметной области и представленной выше, предмет тематических рубрик (S) составляют слова, называющие документы и библиотечные материалы, как paper, cardboard, leather, parchment; books, bookbinding, photographs, manuscripts. В семантической роли аспекта (P) выступают слова со значением ‘консервация документов’ и слова, называющие нижестоящие по отношению к данному термину, или видовые понятия, как Document Conservation и Environment Management, Pest Management и Disaster Management. Предметы и аспекты информации могут иметь „нулевое“ словесное выражение [Ø] [8]. Вне связи с предметной областью нулевые позиции имеют значение “всякий, любой” S или P, как, например, в содержании тематической рубрики Environment Control. Нулевой предмет в этой рубрике (S [Ø] - Environment Control), означает ‘все об’ - Environment Control. Актуальный, предмет информации, стоящий за нулевыми позициями и эксплицируется в контексте предметной области, имеет значение Documents [Books, Collection, Stock] – Environment Control или Repositories [Libraries, Archives, Museums] – Environment Control. Примеры подготовки запросов с нулевыми позициями приводятся в разделе 3. «Комплектование путеводителя и подготовка запросов для комплектования».

Файлы тематических ссылок, являющиеся главным поисковым инструментом в ресурсе, также основываются на предметно-аспектной модели. Форматирование тематических рубрик здесь заключается в том, что предметами служат названия организаций и сайтов, к которым относится ссылка, а аспектами – то, что о них сообщается. В результате описанного представления информации получают списки рубрик, имеющие следующий вид: Abbey Publication – Temperature & RH Guidelines, Aegis monitoring – Wireless Temperature Monitoring Humidity, AJPerrri HVAC – Whole House Humidification.

Списки тематических рубрик разбиваются на подрубрики, и такая их декомпозиция может являться регулярной и иррегулярной. Регулярный характер имеют декомпозиции рубрик по признаку ‘жанрово-видовая характеристика информации’, когда ссылки по каждой теме путеводителя разбиваются на 6 категорий: библиографии, словари, базы данных, учебники и руководства, основная информация и директории. Иррегулярные декомпозиции рубрик не имеют одного основания и могут являться самыми разными по содержанию, как, например, Mold & Mildew – M old Removal, Mold Identification, Fungicides; Pests Control - Integrated Pest Management, Pesticides.

В результате описанной декомпозиции получаются следующие списки хорошо структурированных тематических рубрик:

Pest Control (310)

Bibliographies

Dictionaries

Databases

Guidelines & Manuals

Directories

1300 Pest Control

A and H Pest Control - Pest Library

AGAD Pest Control - Pest Library

AAA Pest Control - Pest Services and Products

About.com - 10 Top Tips for Control of Firewood Insects

About.com - Insect Zoos and Museums across the US

About Pests - Intelligent Insect Control for Homeowners

Absolute Pest Control - Northeast Ohio's Insect

Management Experts

Absolute Serenity Pest Control - Resource Library

Action Pest Control - Commercial Pest Control Services

Action Recovery on Line - Pest Control Debt Recovery & Collection

Active Pest Control - Insects Archives

Adscot Pest Control - Insect/Pest Library

All Experts - Pest Control: Mystery Insect Larvae

Allianz Australia - Pest control to protect your home and health

Alltop - Bats working overtime in Portugal library as pest control

American Institute for Conservation

American Museum of Natural History (Peabody Museum)

- Pest Management Policy

American Pest Management - Insect Library

AMIGOS Preservation Service - Pest Control Seminar,

Austin: A report / by Tom Claeson

Anderson Pest Solutions - Pest Identification

Articlesbase.com - Pest Control

.....

Все выше описанные декомпозиции, регулярные и иррегулярные, (обязательному разделению подлежат излишне широкие рубрики) служат для минимизации наполнения рубрик и удобства работы с путеводителем, так как декомпозиция сокращает число документов, стоящих за рубриками и которые просматривает пользователь. Декомпозиция является также одним из приемов подготовки запросов для проведения сетевого информационного поиска. Поскольку при поиске информации в Интернет предпочтительны узкие по

содержанию запросы, декомпозиция содержания запросов и представление его в виде серии подзапросов обеспечивает многие поисковые преимущества, описанные в публикациях [9-10].

3. Комплектование путеводителя и подготовка запросов для комплектования

Комплектование путеводителя производится в два этапа. На первом просматриваются наиболее интересные профильные сайты. В качестве авторитетных и подлежащих обязательной обработке зарубежных источников были приняты: American Institute for Conservation (AIC), Canadian Conservation Institute (CCI), Getty Conservation Institute, Institute for Conservation (ICON), International Institute for Conservation (IIC), National Park Service (NPS), Northeast Document Conservation Center (NEDCC), Conservation OnLine (CoOL). Русскими авторитетными источниками признаны те, что участвуют в российской национальной программе сохранения библиотечных фондов: Российская государственная библиотека (РГБ), Российская национальная библиотека (РНБ), Всероссийская государственная библиотека иностранной литературы (ВГБИЛ). Государственная публичная историческая библиотека (ГПИБ) и Академия переподготовки работников искусства, культуры и туризма (АПРИКТ). После поиска информации на указанных сайтах комплектование ведется с использованием свободного поиска в Интернет.

Особенность комплектования заключается также в том, что отбираются главным образом веб-страницы (ссылки на них), в которых тема запроса описывается в их заголовках. Данное правило обеспечивает высокую релевантность информационного поиска и одновременно дает уменьшение трудоемкости составления путеводителя. Так, обычный объем информации по узким запросам в рассматриваемой тематической области составляет 3-5 миллионов ссылок (при поиске в Google), а по широким - может оцениваться в десятки миллионов ссылок. Информации, содержащейся лишь в заголовках страниц, имеется относительно и мало, и объем ее в редких случаях превышает 100-150 ссылок. Вследствие этого правила комплектования путеводителя трудоемкость его составления сокращается до разумных пределов за счет уменьшения числа просматриваемых в Интернет документов.

При поиске информации в Интернет с целью комплектования путеводителя запрос разбивается на подзапросы. Основанием для расчленения содержания запроса может являться:

- синонимия поисковых терминов, устанавливаемая по тезаурусу по сохранности документов [11],
- жанрово-видовая характеристика информации,

- уточнение (сужение) содержания темы запроса привязкой его к тематической области консервации документов,
- большой объем информации, стоящей за рубрикой.

Так, например, информация о плесени отбирается по термину mold и по синонимам данного термина mould, mildew, fungi и fungus, то есть по четырем подзапросам. Вторая декомпозиция данной темы проводится по признаку 'жанрово-видовая характеристика информации'. На этой основе выделяется 15 подзапросов, как Mold Bibliography, Mould Bibliography, Fungi Bibliography, Mold Glossary, Mould Glossary, Fungi Glossary, Mold Dictionary, Mould Dictionary, Fungi Dictionary, Mold Database, Mould Database, Fungi Database, Mold "Data base", Mould "Data base", Fungi "Data base" и т.д. Третья декомпозиция темы направлена на привязку запроса к тематической области, представления темы в виде серии узких запросов и таким образом на уменьшение шумовой информации, не относящейся к области консервации документов. Данная декомпозиция эксплицирует упоминавшийся ранее нулевой (актуальный) предмет, который при подготовке запросов индексируется в явном виде как S (documents, books, libraries, museums, archive)/

Результатом описанной декомпозиции являются следующие подзапросы (приводится только часть подзапросов):

Moldy Paper
Moldy Book
Photograph Mold
Photograph "Mould"
Photograph Fungi
Document Mold
Document "Mould"
Document Fungi

Mold Library
"Mould" Library
Mold Library Fungi
Mold Museum
Mold Archives

Mold Bibliography
"Mould" Bibliography
Fungus Bibliography
Mold Dictionary
"Mould" Dictionary
Fungus Dictionary
Dictionary, Glossary
Mold Guidelines
"Mould" Guidelines
Fungus Guidelines
Fungus Directory
Mold Directory
Mould Directory

Mold Removal House +
.....

Mold Testing library +
.....

Описанная техника подготовки запросов дает высокую полноту отбираемой для путеводаителя информации, но отличается большой трудоемкостью. Для поиска информации по рассмотренной теме потребовалось 92 подзапроса и просмотр до 1 тыс. веб-страниц, полученных в результате информационного поиска.

4. Программное обеспечение путеводаителя

Первоначальная версия путеводаителя была реализована в виде набора html-страниц с гиперссылками, организованными по иерархическому принципу. Таким образом была сформирована первая (зарубежная) часть информации в путеводеителе по вопросам сохранности. Он был размещен на основном сайте БАН как статический ресурс (без возможностей поиска и сортировки). Затем он стал разрабатываться как отдельный независимый сетевой ресурс (сайт) на хостинге БАН. В настоящее время закончено формирование русскоязычной части и заканчивается формирование зарубежной.

Сетевой ресурс реализован на базе виртуальной машины с ОС GNU/Linux Debian 6.0. Для создания веб-сервера использовалось ПО Apache HTTP Server v.2.2, а также система управления контентом (CMS – content management system) Plone 4.1. Система Plone является программным обеспечением с открытыми исходными кодами, построена с использованием объектно-ориентированного сервера приложений Zope, написанного на языке программирования Python. Выбор CMS Plone для создания веб-справочника был обусловлен тем, что в ней предусмотрена возможность работы с основным требуемым типом контента – ссылками (адресами, URL) на веб-страницы в Интернете. После установки и настройки CMS на сервере авторизованный пользователь имеет возможность вводить новые ссылки в систему непосредственно из браузера с помощью веб-форм, что существенно ускоряет работу по наполнению базы данных. Кроме того, пользователь может редактировать или удалять существующую информацию, создавать разделы/подразделы в соответствии с логикой предметной области, перемещать при необходимости информацию из одного раздела в другой и т.п. Использование ПО Plone позволило создать сетевой ресурс с возможностью поиска как по всему содержимому сайта, так и по каждому из его разделов. В настоящее время путеводаитель доступен в сети Интернет по адресу: <http://91.151.182.200:8083>. В дальнейшем предполагается интеграция ресурса в основной сайт БАН.

5. Нелингвистические (прагматические) основания путеводаителя

При составлении путеводаителя не ставилась цель собрать в нем всю имеющуюся в Интернет информацию по вопросам сохранности документов. Такая задача являлась бы заведомо невыполнимой, так как количество ссылок по многим из представленных в путеводеителе тем является слишком большим для рассматриваемого сетевого ресурса. Задача работы ставилась по-другому - создать представительную коллекцию ссылок по данной тематике, отражающую информационные потребности главным образом специалистов-исследователей.

Описываемый информационный ресурс создавался в условиях крупной научной библиотеки и в основном соответствует информационным потребностям указанной категории пользователей. “Одно дело, – пишет по поводу этих потребностей О.А. Лавренова, – практик, для которого требуются точные данные об интересующем его предмете и ничего лишнего... Иное – теоретик, готовый “перелопатить” горы литературы, чтобы натолкнуться на интересную для него мысль” [12, С.32]. Этой направленностью путеводаителя объясняется большой объем ссылок, стоящих за рубрикой и в некоторых случаях достигающих 200 ссылок. Поскольку исследователи в своей ежедневной работе обычно имеют дело с большими объемами обрабатываемой информации, работа с объемными списками рубрик у них не должна вызывать затруднений.

Логика научных исследований требует обращения к библиографиям и словарям и выделения этих рубрик в путеводеителе. Ориентация на определенный тип пользователей имеет место и в ходе отбора просматриваемых веб-страниц для их включения в ресурс. При решении вопроса о том, нужна ли определенная ссылка в путеводеителе, оценивается степень информативности ссылки и насколько она интересна с точки зрения исследовательской проблематики. В то же время, как нам представляется, путеводаитель учитывает и информационные потребности пользователей, для которых желательно небольшое число веб-страниц по конкретным проблемам. Для данной категории пользователей, по нашему мнению, больше всего предназначен раздел Guidelines & Manuals, содержащий учебные материалы и руководства по темам путеводаителя.

Путеводаитель по сохранности документов на бумажной работе составлялся силами одного человека в течение 22 месяцев. Время обновления путеводаителя (в экспериментальном порядке пополнялась только его английская часть), составило около семи месяцев. Имея в виду, что русская часть путеводаителя сравнительно невелика по объему и составляет около 3 тыс. ссылок, на регулярное обновление путеводаителя, как русской, так и англоязычной его составляющей, потребуется немногим более одного года. Такую периодичность мы считаем приемлемой.

Литература

- [1] Alimohammadi D. Designing Webliographies in an Effective and Simple Manner: A Step by Step Process [Электронный ресурс] // Webology. Vol. 1, N 1, August, 2004. URL: <http://www.webology.org/2004/v1n1/a2.html>
- [2] Alimohammadi D. Are webliographies still in use? [Электронный ресурс] // The Electronic Library, Vol. 22. Iss: 2. P. 154 – 157. URL: <http://www.emeraldinsight.com/journals.htm?articleid=862050&show=abstract>
- [3] FAO Web guide, Version 1.5: Design. [Электронный ресурс] URL: <http://www.fao.org/webguide/design/ru>
- [4] Бондаренко Я.Л. Веблиография как средство ориентации читателей детской библиотеки в ресурсах Интернет // Материалы 12-й Международной Конференции КРЫМ 2005 «Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса. URL: <http://www.gpntb.ru/win/inter-events/crimea2005/disk/trud.html>
- [5] Ткачук Л.Ф. Особенности вебблиографических пособий: теоретический аспект // Научные и технические библиотеки. 2012. № 4. URL: <http://intranet.gpntb.ru/subscribe/?journal=ntb&year=2012&num=4&art=>
- [6] Петяскина О.В. Веблиографические пособия: методика и особенности создания. [Электронный ресурс] URL: <http://wap.ogiik.forum24.ru/?1-6-0-00000000-000-0-0-1351711946>
- [7] Кругликова В.П. Предметизация произведений печати: общая методика. М.,1967. – 176 с.
- [8] Пименов Е.Н. Предметно-аспектный подход к индексированию информации: актуальный и нулевой предмет индексирования // НТИ. Сер. 2. – 2001. – N 7. – С. 18-25.
- [9] Пименов Е.Н. Информационные поиски в Интернет: проведение адресных поисков // НТИ. Сер. 2. - 2008. - № 5. - С. 1-8.
- [10] Пименов Е.Н. Информационные поиски в Интернет: проведение тематических поисков / Е.Н. Пименов, А.Н. Ильин // НТИ. Сер. 2. - 2010. - № 10. - С. 7-12
- [11] Тезаурус информационно-поисковый по сохранности документов / [сост.: Е.Н. Пименов, Л.Г. Левашова, В.Б. Никитин]; Б-ка РАН. – СПб. URL: http://www.rasl.ru/e_resours/tezaurus/lat-A.htm
- [12] Лавренова О.А.. Есть такая запись в электронном каталоге! Ч.2 // Библиотека. – 2000. – N4. – С.31-36.

Developing Document Conservation Webguide

E.N. Pimenov, V.A. Soldatov

The process of document conservation webguide development is described. We consider the purpose of the guide, its potential users, the work on drawing up lists of keywords and thematic groups, guide acquisition, query processing used for acquisition, and some pragmatic grounds of the guide.