# Новая система презентации электронных книг в системе «Научное наследие России»

*Погорелко К.П.*
*(Библиотека Математического института им. В.А. Стеклова РАН — отдел БЕН РАН)*

Программное обеспечение электронной библиотеки «Научное наследие России» в настоящее время состоит из нескольких взаимодействующих между собой подсистем. Две из них обеспечивают технологические процессы подготовки метаданных и полнотекстовых электронных публикаций. Третья подсистема обеспечивает поиск и презентацию данных пользователям в Интернет. Эта структура программного обеспечения сложилась в 2007 г. и соответствовала требованиям, предъявляемым к системе на тот момент. Стандартом графических файлов в системе являлся формат tiff. Черно-белые изображения представлены в кодировке CCITT Group 3 или Group 4. Цветные изображения и изображения, отсканированные в градациях серого, рассматривались в то время как исключения, и хранятся без сжатия. В настоящее время такая организация тормозит дальнейшее развитие системы и предъявляет избыточные требования к тестированию качества введенных электронных книг. Возможные направления развития программного обеспечения обсуждались автором в [1]. В данной публикации приводится описание реализации новой системы презентации электронных книг, которая должна заменить часть существующей подсистемы презентации.

Система презентации электронных книг реализована как web-приложение на платформе Microsoft asp.net с использованием системы MVC-3. Дизайн внешнего вида электронной книги оставлен без изменения. Добавлены новые кнопки перехода на первую и последнюю страницы публикации. Организована поддержка управления просмотром публикации с клавиатуры. Реализованная система обладает следующими возможностями:

- Преобразование изображений «на лету» в соответствии с запросом пользователя;

- Работа с различными типами изображений (bmp, emf, exif, gif, jpeg, png, tiff);
- Стабильность ссылок на электронные издания (отсутствие технологических подробностей реализации в ссылках — .asp, .aspx, .php и т.д.), что дает возможность индексирования электронных публикаций в Интернет-поисковиках;
- Возможность отслеживания и блокировки массового скачивания изображений.

Сканирование документов в электронной библиотеке «Научное наследие России» происходит, в основном, с разрешением 600 dpi. Поскольку изображения страниц на экране монитора чаще всего соответствуют разрешению 100 dpi, то возникает проблема преобразования изображений. Существующая система презентации данных в Интернет использует механизм кэширования предварительно отмасштабированных изображений. Такой подход требует дополнительных затрат на хранение вариантов изображений и сложного администрирования при изменении изображений в основной базе, например, при исправлении выявленных ошибок. В новой версии системы используется принцип трансформации изображений в требуемый вид по каждому запросу пользователя. Этот подход позволяет обходиться без кэширования и связанных с ним проблем, но предъявляет повышенные требования к эффективности процесса трансформации изображений.

Для преобразования графических изображений в системе презентации используется библиотека работы с графикой, встроенная в Microsoft Framework 4.0. Эта библиотека позволяет читать изображения в большом спектре графических форматов, производить изменения масштаба изображения, осуществлять повороты изображения на углы 90, 180 и 270 градусов и экспортировать изображения в требуемый графический формат.

Для оценки эффективности реализации было проведено измерение временных затрат при пробной эксплуатации системы более чем на 1500 запросах. В результате были получены следующие значения.

Общее время выдачи изображения по запросу изменяется в диапазоне от 265 до 998 миллисекунд и в сред-

нем составляет 538,3 миллисекунды. Сюда входят такие процедуры, как обращение к SQL серверу для получения данных о выдаваемом изображении, поиск и чтение файла в системе хранения, трансформация файла в соответствии с запросом и выдача готового результата. Основное время при обработке изображения составляет поиск и чтение файла из системы хранения и изменяется в диапазоне от 202 до 826 миллисекунд со средним значением 441 миллисекунда. Эти задержки обусловлены аппаратной архитектурой размещения системы «Научное наследие России» и не могут быть улучшены за счет повышения эффективности программной реализации системы презентации. Само преобразование и выдача результата в требуемом формате занимает от 46 до 312 миллисекунд со средним значением 97,2 миллисекунды. Большие времена преобразования (более 200 миллисекунд) получены для графических файлов большого объема (более 25 Мб.), доля которых в системе не превышает 1%.

Таким образом, можно сделать вывод, что реализованная система преобразования изображений «на лету» добавляет к времени простого чтения файла около 0,1 сек., что, по мнению автора, является приемлемым результатом.

За счет конфигурации фильтра поступающих запросов ссылка на электронную публикацию, обрабатываемую данной системой презентации, содержит только адрес размещения web-приложения и внутренний номер публикации в системе. Таким образом, эта внешняя ссылка будет оставаться неизменной при дальнейших реализациях системы на других технологиях, что позволит производить индексацию электронных публикаций Библиотеки Интернет-поисковиками.

Отдельной проблемой при размещении электронных публикаций в свободном доступе в Интернет является проблема авторского права. Как показала практика, размещенные публикации подвергаются массовому скачиванию и размещению на других ресурсах, зачастую даже без указания первоисточника. В данной реализации этой проблеме уделяется особое внимание. Прежде всего, каждое выдаваемое пользователю изображение снабжается «экслибрисом» электронной библиотеки. Кроме того

ведется протокол обращений к системе. Если при анализе частоты обращений возникает подозрение на систематическое скачивание, то в системе предусмотрена возможность проверки, не робот ли это. Пользователю вместо изображения очередной страницы выдается изображение, содержащее тот или иной вопрос, на который надо ответить в текстовом окне. В случае неправильного ответа адрес пользователя будет на некоторое время заблокирован.

Внедрение данной системы презентации позволит приступить к дальнейшей модернизации программного комплекса подготовки электронных публикаций для электронной библиотеки «Научное наследие России».

*Литература*

1. *Погорелко К.П. Эволюция программного обеспечения системы подготовки материалов для электронной библиотеки «Научное наследие России» // Информационное обеспечение науки: новые технологии: Сб. науч. Тр. Под ред. Н.Е. Калёнова — М.: БЕН РАН, — 2011. — С. 260-263.*