

УДК 004.91::[027.021+027.7]

А. Н. Гребнев

НАУЧНОЕ ОБЩЕНИЕ ЧЕРЕЗ ЭЛЕКТРОННЫЙ ФОРУМ И ЭЛЕКТРОННЫЕ ПУБЛИКАЦИИ

1 Введение

Со времен Сократа известно, что диалог рождает новое знание. Только научная дискуссия и конструктивная критика позволяют приблизиться к истине. Современные темпы развития науки требуют более быстрых способов коммуникации между исследователями, в том числе и географически удаленными. Интернет технологии имеют возможности значительно ускорить процессы обмена научной информацией. Необходим программный комплекс, который бы реализовывал эти возможности, отображая устоявшиеся идеологические научные традиции в электронную форму, учитывая российскую специфику (особенности русского языка, низкую финансируемость научно-образовательных учреждений).

Исторически сложились два основных способа научного общения: конференция (форум) и публикация. В настоящей работе предлагаются способы отображения классической научной конференции в научный Интернет-форум, и классической публикации в электронную Интернет-публикацию, обеспечивающие при этом более плотную интеграцию между ними.

2 Электронный форум

В условиях отсутствия возможности живого общения классическую конференцию следовало бы заменить на видео- или хотя бы аудио-конференцию. Однако наиболее дешевым и соответственно приемлемым является текстовая электронная конференция т.е. Интернет-форум.

Выделяют два вида Интернет-форумов: иерархический (древовидный) и линейный (плоский, нитевидный)¹. Каждый из видов применяется в зависимости от аудитории пользователей и их целей. Иерархический форум обычно реализуется в виде дерева, где сообщения можно добавлять к любому из узлов, дерево отображается на странице, чтобы просмотреть тело сообщения нужно перейти по ссылке (заголовку сообщения). В линейный форум можно добавлять сообщения только в конец, причем тела самих сообщений обычно отображаются полностью на одной странице. Однако ни тот, ни другой тип не подходит для научного диалога, с одной стороны необходимо наличие иерархии, чтобы можно было отвлекаться на подтемы, с другой стороны основная тема не должна теряться. Главная же отличительная особенность научного форума от обычного в том, что он должен обеспечивать возможность получения всей беседы в виде линейного текста для последующего создания на его основе научной публикации.

В качестве решения предлагается гибридный вариант форума в основе структуры которого будет бинарное дерево. Вершины дерева содержат сообщения, причем полустепени исхода каждой вершины (кроме вершины принадлежащей стволу) ветки меньше либо, равны единице.

Ствол дерева – путь из корня к листу, все вершины которого помечены как ствольные (включая корень и лист). Стволовая вершина может быть смежна с не более чем двумя ствольными вершинами.

Ветка дерева – путь из ствольной вершины, в лист, через вершины, не помеченные как ствольные.

Тела сообщений (вместе с заголовками) ствола должны быть отображены на одной странице. Ветка, например, может представлять собой обсуждение тонкостей темы не обязательных для прочтения, исторически принято такие отступления помечать на странице вертикальной линией слева. Если из узла сообщения отходит ветка, она должна отображаться как

¹ <http://usabilist.de/seeking/issue.php?article=40719&lang=RU>

список заголовков сообщений (а если их нет, то первых строк тел сообщений) этой ветки, который можно раскрыть на этой же странице или перейти на просмотр ветки на отдельной странице, опять же отображая тела сообщений полностью.

Предложенный подход позволит решить главные проблемы Интернет-форумов, настраивая пользователей на заданную тематику.

3 Электронные публикации

В процессе жизненного цикла классической бумажной публикации участвуют два института: издательство и библиотека, которые обеспечивают создание, распространение, использование и сохранение публикаций. В случае электронной формы принято считать, что все функции выполняет электронная библиотека. Принципы построения информационной системы обеспечивающей функции эл. библиотеки в научно образовательной среде уже рассмотрены в нескольких работах [1][2]. Принципы требуют, чтобы информационная система была основана на Будапештской инициативе 2001 года «Открытый доступ» (*Budapest Open Access Initiative*)² посредством самоархивирования (*self-archiving*) и предоставления открытого доступа (*open access*) (общего бесплатного и оперативного).

Электронный форум, описанный выше, также можно считать подсистемой электронной библиотеки участвующей в процессе создания научной публикации. Как уже было указано, жизненный цикл создания публикации может начинаться с общения в форуме. Многие процессы, в том числе и развитие публикации можно отобразить в виде спирали (закон отрицания отрицания). Выделим 6 основных состояний научной публикации (на примере статьи):

1. Начальное состояние – в ней статья находится с начала появления идеи у авторов (возможно на основе дискуссии в эл. форуме) и доступна только им для совместного чтения и редактирования.

² <http://www.soros.org/openaccess/ru/read.shtml>

2. Состояние ограниченного препринта – статья становится доступной для чтения большому, но ограниченному самими авторами, кругу лиц.
3. Состояние препринта – статья становится доступной всем желающим для чтения.
4. Состояние постпринта в оригинале – статья уже отрецензирована и сдана в печать издателю (в не зависимости от формы: электронная или бумажная), в случае если издатель требует, чтобы статья больше нигде не была опубликована, то это состояние является конечным.
5. Состояние измененного постпринта – статья дополнена и переработана на основе полученных отзывов коллег.
6. Конечное состояние – идея первоначальной статьи изменяется до такой степени, что может претендовать на звание новой статьи, которая начинает новую ветку спирали с состояния 1.

Статья в каждом состоянии (за исключением 4) может иметь несколько версий отличающихся содержанием, причем каждая версия датирована и может быть в любой момент доступна читателю. Именно наличие различных версий публикации позволит сохранить стройность замысла первоначальной идеи, единообразную и элегантную структуру, на всей протяженности ее развития.

Параллельно с оригинальной вариантом статьи может существовать несколько ее переводов на другие естественные языки, они так же могут иметь множество версий.

Главной задачей библиотеки, в том числе и электронной, является задача сохранения информации на максимально длительный срок. Возникает задача определения формата хранения (Word, HTML, TeX, XML, PDF, ...) материалов. С точки зрения форматов хранения, можно выделить два способа: гетерогенный и гомогенный.

Гетерогенный способ подразумевает хранение публикаций в исходном входном авторском формате, т.е. в том формате, в котором автор поместил

свою работу в системе. По запросу читателя исходный формат может быть конвертирован в необходимый выходной формат.

Гомогенный способ подразумевает хранение публикаций в едином формате, при помещении в систему авторский входной формат преобразуется в формат хранения, а при запросе читателя преобразуется в необходимый выходной. В качестве формата хранения должен быть выбран бесплатный открытый, проверенный временем и уже широко используемый. Предлагается XML формат DocBook³. Это наиболее популярный формат (его разработка начата 1991 году) для написания книг статей и особенно технической документации от Организации по Продвижению Стандартов Структурированной Информации (Organization for the Advancement of Structured Information Standards OASIS⁴).

Однако чтобы формат хранения был жизнеспособным необходимо обеспечить взаимное преобразование в наиболее часто используемый формат. На сегодняшний момент наиболее часто используемым форматом в мире, в котором подготавливаются публикации, является Microsoft Word Document. До некоторого времени преобразование из Word в XML было весьма проблематичным. Оно осуществлялось через сохранение документа в открытом формате RTF или в HTML и их последующим преобразованием в необходимый формат. Ввиду ограниченности форматов RTF и HTML и недостаточной проработанности бесплатных конвертеров, преобразование осуществлялось с потерей качества и некоторой информации. Например, после преобразования документ мог несколько отличаться от авторской версии по внешнему виду, а информация об истории изменения терялась вообще. В связи с этим гетерогенный способ был наиболее предпочтительным. Однако с появлением новой версии Microsoft Office 2003, Word имеет возможность сохранения в формате XML (WordprocessingML или WordML) на основе схемы от Microsoft [3]. Возможности языка XSL

³ <http://docbook.org/>

⁴ <http://www.oasis-open.org/>

(eXtensible Stylesheet Language) позволяют осуществить преобразование из формата WordML и обратно, минимизировав потерю качества и информации.

Предлагается следующее решение задачи. Преобразование осуществляет два XSLT (Transformation) конвертера. Первый конвертер преобразует WordML в DocBook содержащий структуру документа и XSL-FO (Format-Object) содержащий авторское оформление документа (стили, примененные шрифты, цвета и т.п.). Второй конвертер выполняет обратное преобразование из DocBook и XSL-FO почти полностью восстанавливая как структуру, так и оформление документа. Разделение на структуру и оформление с помощью пространства имен (*namespace*) позволяет применить альтернативный способ отображения документа в случае необходимости. Все преобразования выполняются непосредственно в приложении Word и прозрачны для пользователя. Word Template, содержащий стили DocBook, позволяет ограничить пользователя в использовании набора стилей, но не свойств форматирования.

Обсуждение в электронном форуме может быть началом создания публикации, однако и сама публикация может быть причиной обсуждения в форуме. Наличие двунаправленных ссылок между сообщениями в форуме и частями текста в публикации, с использованием стандарта XPointer⁵, позволит интегрировать друг в друга два, исторически сложившихся, основных способа научного общения в электронном виде.

4 Заключение

Проектируемый программный комплекс позволит не только реализовать многие функции традиционных институтов: конференция, издательство, библиотека на основе информации представленной в электронной форме, но и добавить новые существенные возможности не доступные классическим конференциям и бумажным публикациям. Создание и внедрение комплекса обеспечит более быструю коммуникацию между

⁵ <http://www.w3.org/XML/Linking>

исследователями, в том числе и географически удаленными, что существенно обогатит науку и образование.

5 Литература

1. А.Н. Гребнев. Научные информационные системы // Вестник УдГУ: Математика.- Ижевск: Изд-во УдГУ, 2003. С.99-106.
2. А.Н. Гребнев. Коммуникационная научно-образовательная среда // Вестник УдГУ: Математика.- Ижевск: Изд-во УдГУ, 2004. С.29-38.
3. Evan Lenz, Mary McRae, Simon St. Laurent. Office 2003 XML. Integrating Office with the Rest of the World. O'Reilly. 2004.

A.N. Grebnev

This paper is continuation of series of articles about design principles of software communicational system in scientific-educational environment. It covers design principles of electronic discussion, some solution aspects of creating, distributing, using and saving electronic publications.

А.Н. Гребнев Научное общение через электронный форум и электронные публикации // Проблемы теоретической и прикладной математики. Труды 36-й региональной молодежной конференции. Екатеринбург: УрО РАН, 2005. С. 327-331.

Гребнев Андрей Николаевич

Удмуртский Государственный Университет,

Кафедра Математического Обеспечения ЭВМ,

Ижевск, ул. Университетская, 1 (корп. 4)

426034, Россия

E-mail: ag@blandware.com

<http://fox.blandware.com>