2004 - Том 7 - Выпуск 1

FictionBook - библиотека и формат на основе XML. Краткая характеристика формата и обзор библиотеки на его основе.

Д.П. Грибов

Сетевая библиотека FictionBook.lib

- Устройства для чтения - переход количества в качество.
- Обзор состояния сетевых библиотек и используемых в них форматов.
- Сравнительная характеристика различных используемых в библиотеках форматов с точки зрения поддержки электронных библиотек.
- Критика PDF как основы электронной библиотеки.
- Сравнительный обзор особенностей основанного на XML формата FictionBook.
- Обзор FictionBook.lib - библиотеки, основанной на FictionBook.

Электронные книги - переломный период?

Количество людей, читающих с экрана, пока мизерно. Единицы предпочитают экран бумаге, чаще все заканчивается банальной распечаткой даже в случае объемных электронных текстов. Единственное достаточно удобное устройство (Rocket eBook, 72 dpi) для чтения появилось раньше своего времени что, в сочетании с неадекватным маркетингом, обеспечило его провал на рынке. Ситуация стала меняться с появлением доступных LCD дисплеев с разрешением 150dpi и выше (http://www.eink.com/, http://www.computerra.ru/today/ferra/27777/). Ряд устройств уже сейчас можно считать превосходящими бумажные книги по всем параметрам, кроме времени автономной работы и цене. ClearType втрое увеличивает горизонтальное разрешение цветных дисплеев, и без того впечатляющее. Массовые устройства для чтения электронных книг появятся на рынке уже в 2004-м году (Sony планирует продажи новой eBook на апрель), и новые устройства, наконец, смогут реально конкурировать с бумагой по цене и качеству отображения текста. В дальнейшем развитие устройств-заменителей бумаги обещает принять лавинообразный характер, инвестиции в этой сфере, помимо частного капитала, активно проводит, к примеру, правительство США.

Для образования, науки и книгоиздания переоценить такое развитие событий невозможно, на ум приходит только сравнение с Гуттенбергом. К этому рубежу индустрия электронных текстов подошла в состоянии разброда, разработчикам новых устройств будет, над чем поломать голову. Электронные тексты сейчас встречаются:

- В одном из фирменных (часто закрытых) форматов (MS lit (далее просто lit), Aportis, iSilo, pdf…)
- В формате программ подготовки текстов (MSWord, Quark, PDF, LaTeX…)
- Как простой текст (библиотека Мошкова)
- html/xhtml
- XML: DocBook, OEB, FictionBook

Что сейчас лежит в электронных библиотеках?

Закрытые и фирменные форматы

Ряд закрытых форматов (в первую очередь это PDF от Adobe и lit), обеспечивающих (фактически не обеспечивающих) защиту контента, лидируют в сфере продажи электронных текстов. Широкое распространение этих форматов объясняется желанием продавцов защитить свою интеллектуальную собственность, а иллюзию защищенности рискуют предлагать только материально заинтересованные стороны. Постепенно к защите контента подключается и Open Source сообщество, но пока на электронных книгах это не сказалось. О причинах распространенности PDF так же можно прочитать здесь.

В основном компании основывают свои форматы на html (lit, iSilo, Aportis, rb), но в итоге ни один из этих форматов не обеспечивает приемлемой индексации текстов и поддержки нелатинского алфавита. У каждого формата - собственные схемы сжатия и шифрования разной степени закрытости. Средства просмотра таких документов доступны не для всех платформ и, как ни странно, зачастую некачественны. Обратная конвертация затруднена, невозможна и/или является противозаконной (например, lit->xhtml).

Подобные ограничения не мешают энтузиастам иметь мини-библиотеки с книгами в фирменных форматах, экономя читателям время на самостоятельную подготовку книг для мобильного устройства, но невозможность автоматизации обычно убивает такие библиотеки даже до устаревания формата.

Хранение текстов в «родном» формате программы верстки практикуется, но лишь в силу дешевизны. Документ, ушедший в печать, в неизменном виде

выкладывается в общий доступ. Многие производители «железа» поддерживают у себя библиотеки инструкций и руководств в формате PDF именно по этой причине. За исключением простоты организации, такая библиотека не имеет в себе ничего положительного - каталогизация затруднена, доступ к документам требует специфического ПО, как правило, мало приспособленного к нуждам читателя и зачастую небесплатного. Хранение в MSWord *.DOC формате является относительно удобным, но документы MSWord по простоте обработки и выбору инструментария сильно уступают даже html.

PDF - подробная критика

PDF, являющийся одним из лидеров, в качестве основного формата хранения книг оказывается ниже всякой критики. Его практически невозможно читать на малом дисплее. ПО от Adobe для немногих платформ, феноменально некачественно для столь старой и популярной программы. PDF невозможно адекватно преобразовать во что-либо, кроме бумажной копии, часто невозможно даже проиндексировать...

Фактически, PDF-документ обычно является закрытым объектом, не содержащим данных о структуре и не предоставляющим информации о содержании, годным лишь для получения твердой копии. Основная цель создателей, дублирование бумажного документа, привела к переносу всех ограничений твердой копии в цифровую версию.

В дополнение к подробной критике PDF как средства представления электронной информации от Якоба Нильсена (1, 2) мы изложили свое видение причин внедрения PDF в мир электронных текстов.

Краткое резюме неутешительно - **PDF можно и нужно применять, но только имея в виду получение твердой копии**. С учетом ожидаемого бума электронных средств для чтения, на наш взгляд, следует однозначно отказаться от использования PDF как базового формата для библиотеки. Это не мешает готовить pdf-файлы, например, из XML, для тех пользователей, кому по тем или иным причинам нужна распечатка.

Открытые форматы - не-XML

В отличие от вышеозначенных форматов, хранение книг в виде простого текста (без разметки html/xml/sgml/...) позволяет одним махом решить проблемы с доступом с разных платформ, индексацией и т.п., но имеет очевидные недостатки - ограниченные возможности форматирования и использования графики, полное отсутствие метаинформации. При всей своей ограниченности, такой подход очень широко распространен. Самая известная библиотека рунета (и одна из самых полных в сети), lib.ru, хранит файлы именно в виде простого текста (фактически, там используется гибрид html-разметки и простого текста, но основной инструмент разметки - пробелы, разрывы строк и несколько спецсимволов). Использование в качестве разметки символов перевода строки и пробелов идеально с точки зрения простоты и лаконичности, но возможности такой разметки ограничены.

html как формат хранения электронных книг обладает весьма неплохим потенциалом, а xhtml - и того более, упрощая обработку и отображение до вполне разумных границ. Подготовка документов, индексация и преобразование в другие форматы для html являются задачами если не тривиальными, то решаемыми. Однако html не обладает достаточной поддержкой метаданных, с чем не всегда можно смириться (например, затруднительно указать развернутую аннотацию с гиперссылками и форматированием), а подготовка xhtml из html не всегда проста. Все это усугубляется довольно богатым синтаксисом (вольно используемым, если мы говорим об html), полная реализация которого является весьма непростой задачей, подчас не решаемой на устройствах с ограниченной памятью и вычислительной мощностью.

Реализация программы чтения с поддержкой даже небольшого подмножества тегов html, как правило, занимает много времени и никак не может считаться тривиальной задачей. Если говорить об достаточно примитивных устройствах для чтения с оперативной памятью меньше 1MB (ряд таких устройств готовится к выпуску, одно уже продвигается на Украине в сферу образования) проблематична даже полноценная поддержка самых распространенных тегов.

Очевидное решение - использование подмножества html, с чем мы и имеем дело в большинстве случаев. Rocket eBook (первое специализированное устройство для чтения, ОС на основе Linux) поддерживает 29 тегов и png-графику. REB1100 (следующая модель) поддерживает то же плюс таблицы. Лидирует по количеству поддерживаемых тегов MSReader (заявлена поддержка OEB), работающий на hi-end карманных компьютерах, но эксперименты с тегами и CSS на деле и здесь обычно заканчиваются неадекватным отображением, медленной работой и ошибками программы.

Можно с полной уверенностью утверждать, что все выходящие на рынок устройства для чтения электронных книг будут отображать html или (в обозримом будущем, более вероятно) его подмножество. Вероятно также, что html будет упаковываться и/или шифроваться.

Стоит отметить, что, несмотря на все вышеизложенное, библиотеки в формате html относительно редки. Обычно выкладываются тексты в «трудоемких» форматах, подготовка которых требует специального инструментария (см. «Закрытые и фирменные форматы» выше) либо просто текст, который обрабатывать несравненно проще, чем html.

Открытые форматы, основанные на XML

При всем богатстве средств форматирования и широте распространения, html ограничен в передаче метаданных. Непростой задачей является и его отображение и обработка, что для электронной библиотеки является критичным. Возможность оперативно предоставить одну главу из 50-и мегабайтного текста в затребованном формате, возможность автоматически внести в каталог книгу с аннотацией и обложкой (что важно для художественной литературы), индивидуально уведомить подписчиков о появлении новой книги в интересующем их разделе и приложить к уведомлению аннотацию, возможность получить список

изменений в последнем издании - реализация такого рода библиотечных сервисов на основе html или простого текста представляется малореальной.

Очевидным решением при хранении упорядоченных текстовых данных сегодня является XML, и решения на его основе появились не вчера. Наиболее интересны для библиотек три основанных на xml формата: Open eBook (OEB), DocBook, FictionBook.

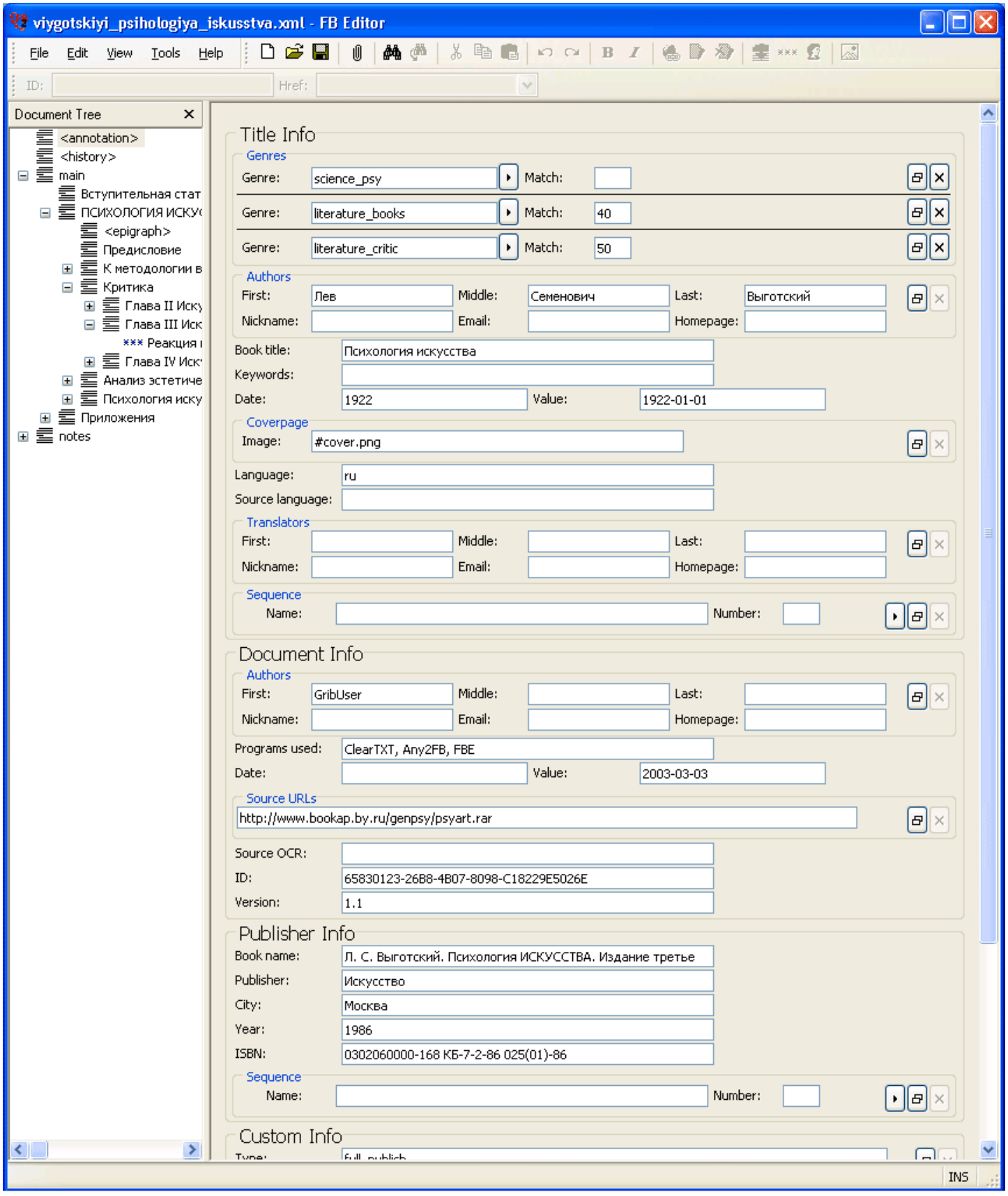
OEB - детище Microsoft и Adobe, давних (и не очень удачливых) игроков на рынке электронных книг. OEB призван компенсировать ограниченность html в хранении метаданных и закрепить за xhtml доминирование в книжном деле. По сути, OEB-книга это пакет из xhtml и графических документов, снабженных метафайлом, хранящим ISBN и другие выходные данные. Сама книга представлена в xhtml, а весь пафос OEB, по сути, сводится к 15-и тегам (используется Dublin Core 1.1 с небольшими расширениями), описывающим автора, название, язык книги и т.п.

Аннотация в OEB не предусмотрена. Вся спецификация OEB - это пересказ xhtml, и ничего, кроме недоумения, ознакомление с OEB не вызывает, хотя желание срезать углы и воспользоваться готовой технологией вполне понятно.

OEB-пакет обычно помещается в ZIP-архиве. Фактически, в сети OEB не встречается. Судя по статистике библиотеки FictionBook.lib, OEB пользователи не скачивают в принципе (0.2% загрузок, судя по отзывам - результат любознательности).

DocBook - очень развитый и продуманный стандарт, разработанный и сопровождаемый OASIS (Organization for the Advancement of Structured Information Standards). Формат идеально подходит для технических текстов, и FictionBook не возник бы, если бы DocBook был минимально приспособлен для работы с художественной литературой. Но DocBook, при всем богатстве, не способен адекватно представить художественную книгу. Например, DocBook не предусматривает разметку стихов, не позволяет описать книжные серии, не имеет данных о переводчике. Для относительно несложных, в большинстве своем, художественных текстов, формат так же явно избыточен.

FictionBook - разработан нами (инициативная группа, возглавляемая Грибовым Дмитрием и Михаилом Мацневым) специально для художественных текстов. В отличие от OEB, книга хранится в одном файле. Графика, метаданные, текст сносок и аннотация - все хранится в виде единого XML-документа, что существенно упрощает администрирование, распределенную обработку и репликацию. В художественной литературе используется ограниченный набор элементов - стихи, аннотация, жирный/наклонный, иллюстрации, эпиграф, сноски. В результате, FictionBook очень прост (см. схема fb2 и комментарии к схеме), но, как показывает практика, обладает всем необходимым для оформления художественных книг. Развитая структура хранения метаданных (подробно см. схема fb2) позволяет полностью автоматизировать работу библиотеки и легко развернуть ряд уникальных сервисов (например, дифференцированную подписку).



FictionBook - эволюция формата

Использование незатейливого XML для хранения и обработки книг в рунете стало развиваться по инициативе Марка Липсмана (http://www.marklipsman.addr.com/index_r.htm). XML использовался для упрощения

создания MS lit и prc документов (штатное П.О. для создания lit далеко от совершенства). Первоначально использовался минимум метаинформации, не было опубликовано DTD. Несмотря на это, подход быстро приобрел популярность. С появлением программ для подготовки и обработки текстов и формализацией требований к документу (программа ClearTXT устанавливается с DTD и поддерживает валидацию) встал вопрос и об их хранении и замаячила перспектива создания полноценной электронной библиотеки. Это привело к введению в документ развернутого узла с метаданными (добавлена информация о языке оригинала, переводчике, названии серии, введена информация о создателе документа и т.д. - подробней см. комментарии к схеме). В дальнейшем изменения носили косметический характер, но одна из особенностей формата заслуживает особого упоминания, так как всегда вызывает недоумение при первом знакомстве.

Изначально сноски в документе вставлялись по месту ссылки, типа A hero<note>Герой (англ.)</note> , но по мере их осложнения и вынужденного внесения в них элементов форматирования, мы приходили к довольно запутанной с точки зрения структуры и обработки конструкции. Невозможность организовать перекрестные ссылки в таком виде так же не внушала оптимизма. В итоге от in-line сносок пришлось отказаться в пользу вынесения в специальный раздел (<body name=”notes”/>). Xlink используется для связывания (<a xlink:href=”#note1” type=”note”>1). Помимо упрощения обработки, мы получаем полную свободу в перекрестных ссылках и оставляем один общий механизм связывания в документе, стандартный и гибкий.

Текущее состояние стандарта (версия 2.0) доступно на официальной странице FictionBook, развернутые комментарии к схеме, примеры документов и важные замечания доступны на странице комментариев к схеме.

FictionBook - перспективы

Практическое использование FictionBook версии 2.0 не выявило значительных проблем, на очереди незначительные расширения синтаксиса (<sub>, <sup>, id для <image/> и т.п., подробно здесь). Единственным действительно узким местом стала вынужденно принятая жанровая классификация, которая на момент создания стандарта была вопиюще не готова.

Сейчас ведется активная проработка нового, достаточно простого, но более функционального, жанрового классификатора. В версии 2.0 используется одноуровневый список, которого явно недостаточно (даже если отвлечься от его содержания). Новое предложение - стандартный список жанров плюс 5-и мерное пространство возраст-страна-эпоха-форма-прием (текущее предложение по жанрам в FB3). Работы в этом направлении только начинаются. К сожалению, в проекте пока не участвуют профессиональные литературоведы (напротив, в основном проект развивают «технари»), и мы вынуждены в меру своих сил изучать имеющиеся наработки и адаптировать их к нашим нуждам.

Так же в документ планируется внести условия его распространения. К примеру, издатель готовит книгу и помечает первую главу как доступную бесплатно (<section share=”free”>..), остальные - за деньги (<section share=”pay-per-

download”/>). В результате, автоматически будут подготовлены платные и бесплатные версии документов для различных платформ, помещены в каталог и мгновенно доступны пользователям для ознакомления и приобретения.

Библиотека, таким образом, может стать и магазином электронных текстов. Соединение в одном месте платных и бесплатных текстов представляется естественным.

FictionBook.lib - библиотека на основе формата FictionBook

Обзор возможностей библиотеки

Как уже отмечалось, формат создавался во многом с прицелом на централизованную обработку и хранение в библиотеке. Обширный заголовок содержит ID документа, информацию об авторе и переводчике, аннотацию, ссылку на обложку, язык документа, а так же другую информацию, интересную пользователям и/или роботам-обработчикам. В настоящий момент на основе формата работает несколько библиотек (1, 2, 3), самая технически развитая - FictionBook.lib.

Книга «Психология искусства» (Лев Семено...

Файл Правка Вид Избранное Сервис Справка

Назад Поиск Избранное

Адрес: http://www.fictionbook.ru/ru/author/viygotskiyi Ссылки

Google Search Web Pagef

Щ Э Ю Я

Искать в библиотеке

Go!

Сложный поиск

Библиотека/ Книги и чтение, История, критика, Психология

Психология искусства

Лев Семенович Выготский - 1986г.

Обращений/мес.: **145**, оценка: **3.2 из 5**

Читать книгу - 1280kb
html | html.zip | fb2 | fb2.zip | rtf.zip
txt.zip | isilo3 | rb | doc.prc.zip | oeb

Аннотация

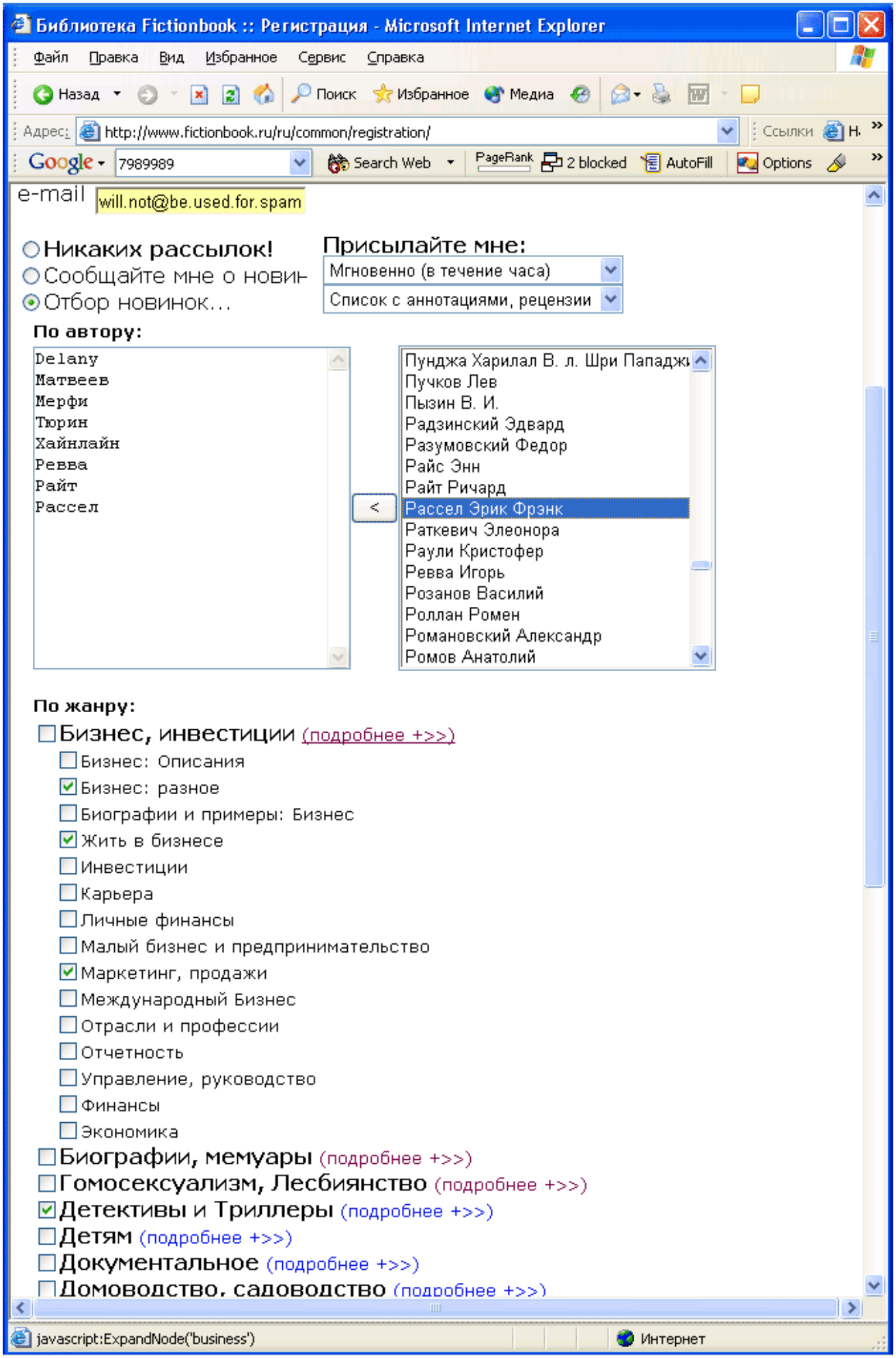
Книга выдающегося советского ученого Л. С. Выготского «Психология искусства» вышла первым изданием в 1965 г., вторым – в 1968 г. и завоевала всеобщее признание. В ней автор резюмирует свои работы 1915-1922 годов и вместе с тем готовит те новые психологические идеи,

Интернет

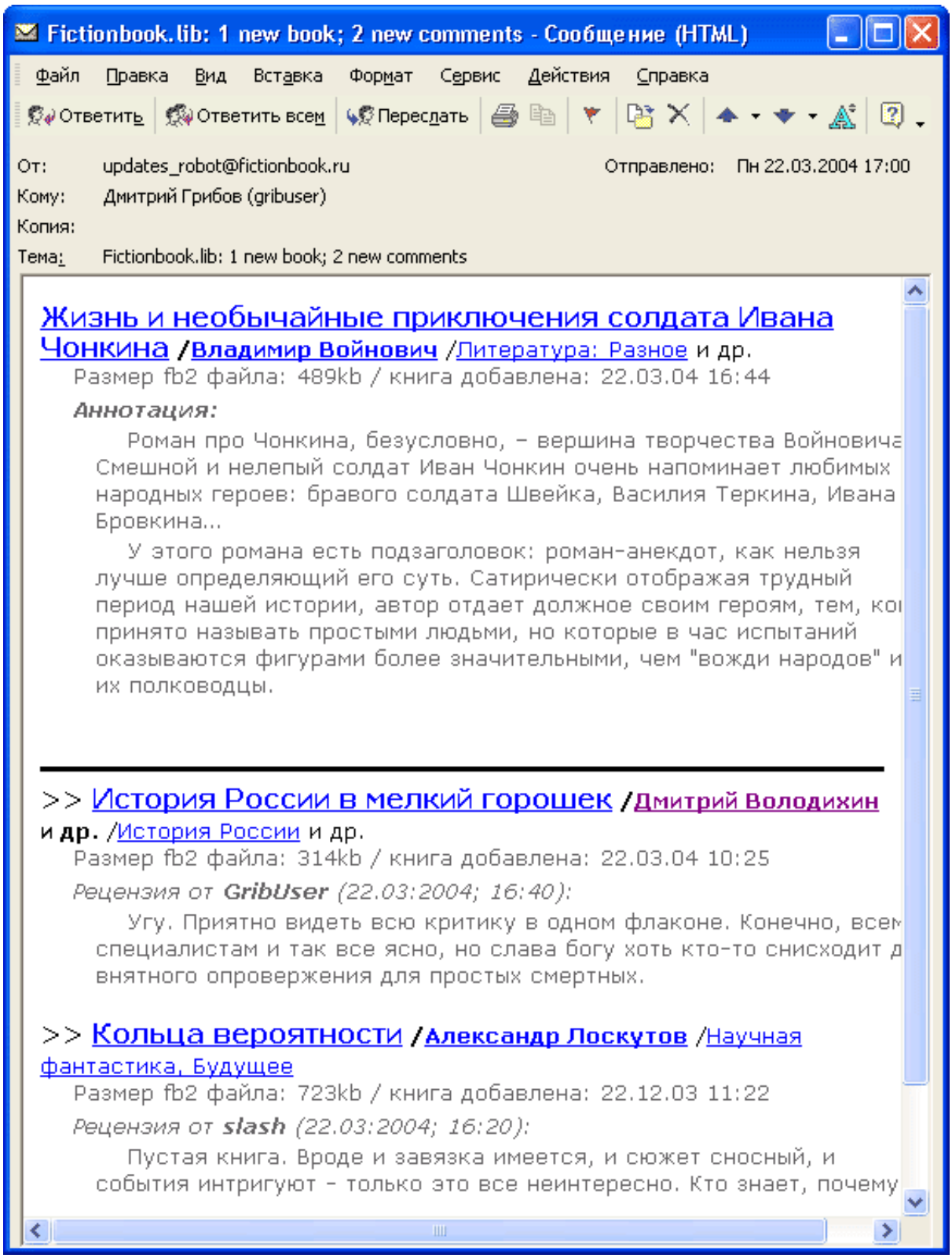
Разработка FictionBook.lib ведется нами уже около года. Фактически, библиотека создана автором этой статьи в одиночку, в свободное время, что, при ее впечатляющей функциональности, лишний раз доказывает перспективность нашего, основанного на XML, подхода к хранению, обработке и распределенному управлению документами. Сейчас програмная часть библиотеки состоит из приблизительно 1.5MB кода (Perl, xsl).

Подготовленный документ может быть отправлен создателем в FictionBook.lib (в настоящий момент - только через web-форму, почтовый робот отключен). Полученный библиотекой документ проверяется на соответствие схеме и регистрируется в библиотеке в соответствующем жанре у соответствующего автора. Из xml-документа автоматически готовятся специальные форматы (html, txt, rtf, iSilo, OEB, rb, lit, Aportis). Как правило, самостоятельная подготовка таких документов - процесс трудоемкий и требующий квалификации и подобная автоматизация экономит много времени и сил.

В библиотеке так же имеется робот, рассылающий подписчикам информацию об обновлениях.

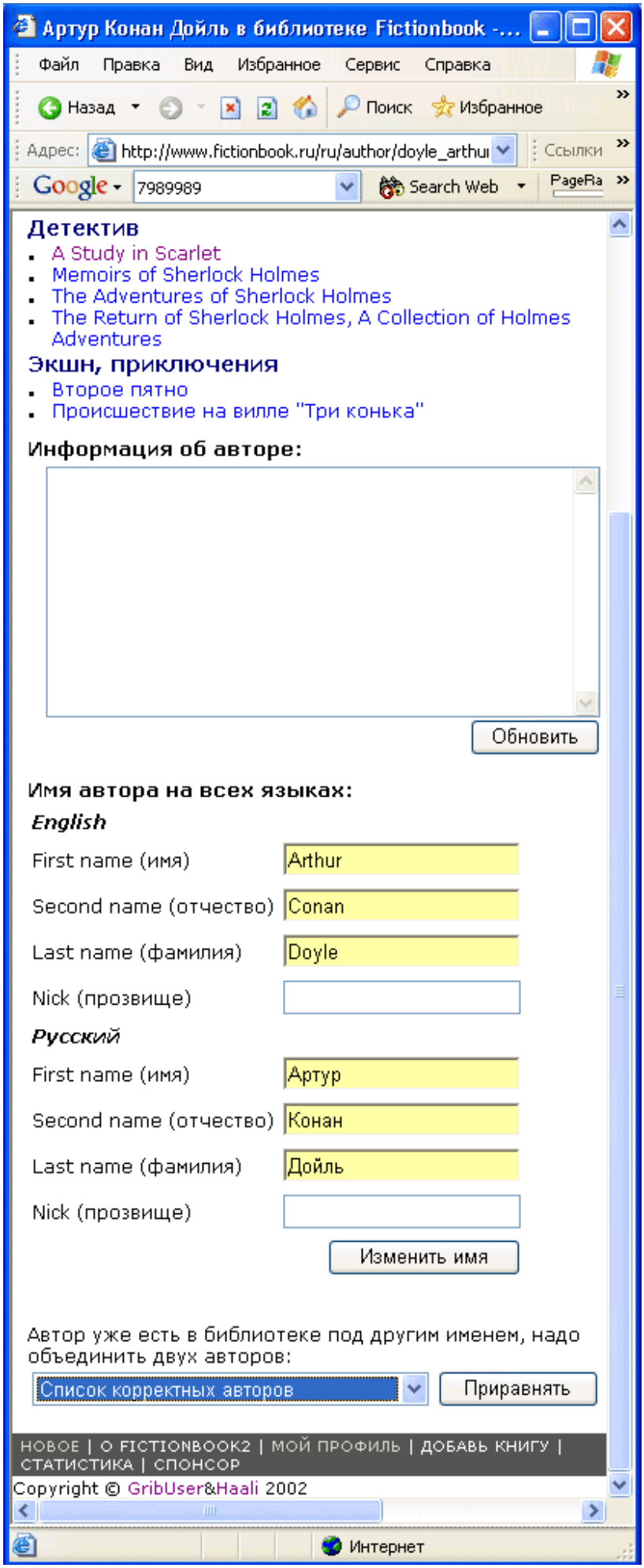

Библиотека Fictionbook :: Регистрация - Microsoft Internet Explorer
Файл Правка Вид Избранное Сервис Справка
Назад Поиск Избранное Медиа
Адрес: http://www.fictionbook.ru/ru/common/registration/
Ссылки
Google 7989989 Search Web PageRank 2 blocked AutoFill Options
e-mail will.not@be.used.for.spam
Никаких рассылок!
Сообщайте мне о новин-
Отбор новинок...
Присылайте мне:
Мгновенно (в течение часа)
Список с аннотациями, рецензии
По автору:
Delany
Матвеев
Мерфи
Тюрин
Хайнлайн
Ревва
Райт
Рассел
<
Пунджа Харилал В. л. Шри Пападжи
Пучков Лев
Пызин В. И.
Радзинский Эдвард
Разумовский Федор
Райс Энн
Райт Ричард
Рассел Эрик Фрэнк
Раткевич Элеонора
Раули Кристофер
Ревва Игорь
Розанов Василий
Роллан Ромен
Романовский Александр
Ромов Анатолий
По жанру:
Бизнес, инвестиции (подробнее +>>)
Бизнес: Описания
Бизнес: разное
Биографии и примеры: Бизнес
Жить в бизнесе
Инвестиции
Карьера
Личные финансы
Малый бизнес и предпринимательство
Маркетинг, продажи
Международный Бизнес
Отрасли и профессии
Отчетность
Управление, руководство
Финансы
Экономика
Биографии, мемуары (подробнее +>>)
Гомосексуализм, Лесбиянство (подробнее +>>)
Детективы и Триллеры (подробнее +>>)
Детям (подробнее +>>)
Документальное (подробнее +>>)
Домоводство, садоводство (подробнее +>>)
javascript:ExpandNode('business')
Интернет

Каждый пользователь может тонко настроить рассылку под себя и получать, к примеру, только информацию о новых книгах интересующих его авторов и/или жанров. Рассылаемые уведомления включают имена авторов книги, название, жанр и, опционально, аннотация. Так же можно получать и все новые рецензии на книги.

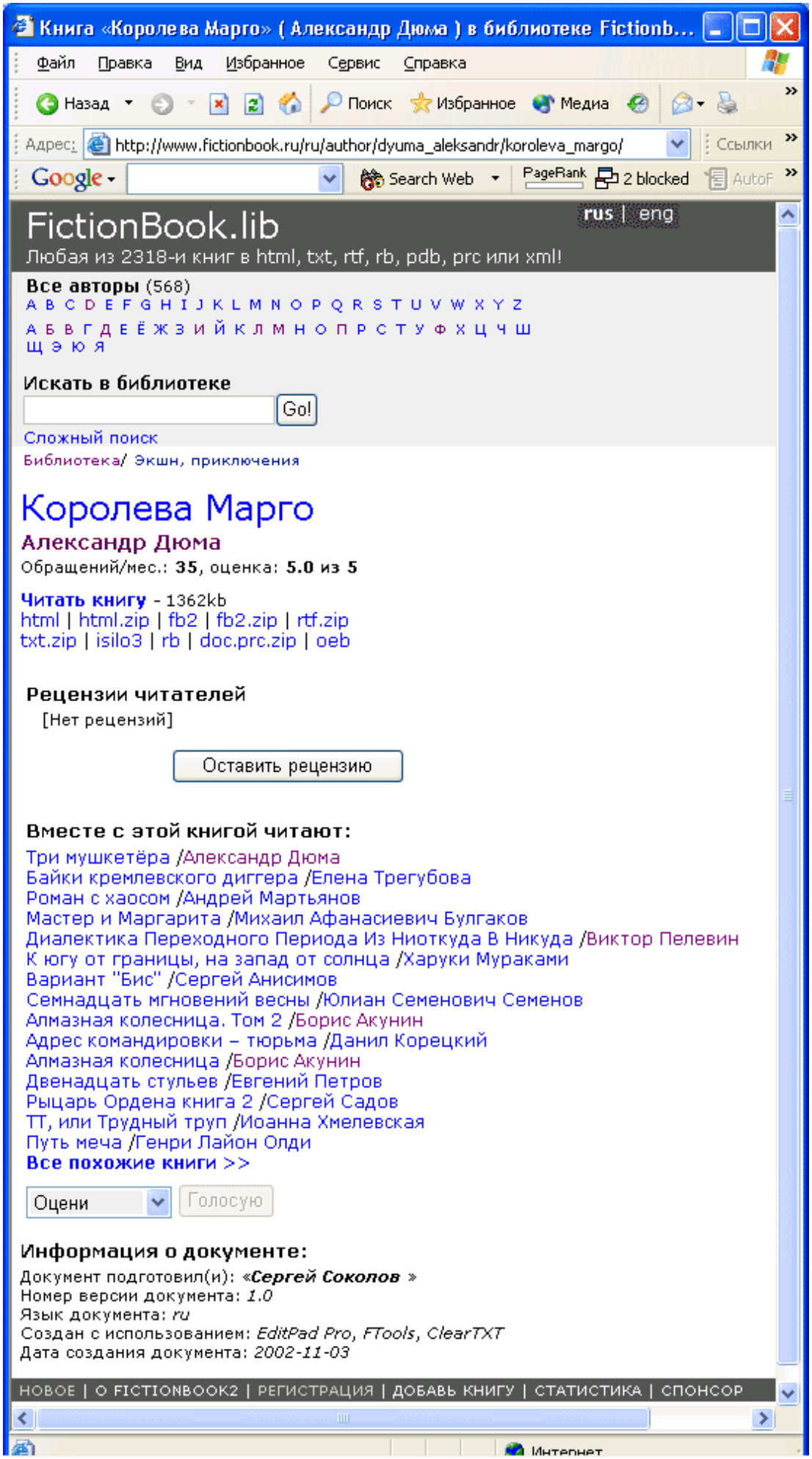


Система администрирования позволяет, в частности, объединить двух авторов. Часто это требуется, если у автора есть книги на нескольких языках и библиотека

не понимает, что «Bujold» идентично «Буджолд» (единожды каждого «нового имени» автора). Фактически, это единственное достаточно трудоемкое неавтоматизируемое действие, все остальное библиотека делает сама на основе метаданных документа. При появлении документа со знакомым ID, библиотека может обновить документ (проверив полномочия отправителя новой версии). Таким образом, большинство задач по управлению документами решаются путем редактирования документа, что позволяет сколь угодно распылить управление библиотекой. Такая форма делегирования функций отлично себя зарекомендовала, а ее последующая доработка обещает еще повысить ее эффективность.


Артур Конан Дойль в библиотеке Fictionbook -...
Файл Правка Вид Избранное Сервис Справка
Назад Поиск Избранное
Адрес: http://www.fictionbook.ru/ru/author/doyle_arthur
Ссылки
Google 7989989 Search Web PageRa
Детектив
A Study in Scarlet
Memoirs of Sherlock Holmes
The Adventures of Sherlock Holmes
The Return of Sherlock Holmes, A Collection of Holmes Adventures
Экшн, приключения
Второе пятно
Происшествие на вилле "Три конька"
Информация об авторе:
Обновить
Имя автора на всех языках:
English
First name (имя) Arthur
Second name (отчество) Conan
Last name (фамилия) Doyle
Nick (прозвище)
Русский
First name (имя) Артур
Second name (отчество) Конан
Last name (фамилия) Дойль
Nick (прозвище)
Изменить имя
Автор уже есть в библиотеке под другим именем, надо объединить двух авторов:
Список корректных авторов
Приравнять
НОВОЕ | О FICTIONBOOK2 | МОЙ ПРОФИЛЬ | ДОБАВЬ КНИГУ | СТАТИСТИКА | СПОНСОР
Copyright © GribUser&Haali 2002
Интернет

При работе над библиотекой много внимания уделено системе рейтингов и подобий. На основе статистики по скачиваниям (140 тыс. записей пользователь-книга) библиотека находит список подобных книг для каждого документа и выделяет популярные книги для каждого жанра. Статистика скачиваний используется и в сортировке результатов поиска. Благодаря комплексной оценке (учитываются скачивания и отзывы) адекватность рейтингов и подобий весьма впечатлила даже самого автора.


Книга «Королева Марго» (Александр Дюма) в библиотеке Fictionb...
Файл Правка Вид Избранное Сервис Справка
Назад Поиск Избранное Медиа
Адрес: http://www.fictionbook.ru/ru/author/dyuma_aleksandr/koroleva_margo/
Ссылки
Google Search Web PageRank 2 blocked AutoF
FictionBook.lib
rus | eng
Любая из 2318-и книг в html, txt, rtf, rb, pdb, prc или xml!
Все авторы (568)
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш
Щ Э Ю Я
Искать в библиотеке
Go!
Сложный поиск
Библиотека/ Экшн, приключения
Королева Марго
Александр Дюма
Обращений/мес.: 35, оценка: 5.0 из 5
Читать книгу - 1362kb
html | html.zip | fb2 | fb2.zip | rtf.zip
txt.zip | isilo3 | rb | doc.prc.zip | oeb
Рецензии читателей
[Нет рецензий]
Оставить рецензию
Вместе с этой книгой читают:
Три мушкетёра /Александр Дюма
Байки кремлевского диггера /Елена Трегубова
Роман с хаосом /Андрей Мартьянов
Мастер и Маргарита /Михаил Афанасиевич Булгаков
Диалектика Переходного Периода Из Ниоткуда В Никуда /Виктор Пелевин
К югу от границы, на запад от солнца /Харуки Мураками
Вариант "Бис" /Сергей Анисимов
Семнадцать мгновений весны /Юлиан Семенович Семенов
Алмазная колесница. Том 2 /Борис Акунин
Адрес командировки – тюрьма /Данил Корецкий
Алмазная колесница /Борис Акунин
Двенадцать стульев /Евгений Петров
Рыцарь Ордена книга 2 /Сергей Садов
ТТ, или Трудный труп /Иоанна Хмелевская
Путь меча /Генри Лайон Олди
Все похожие книги >>
Оцени
Голосую
Информация о документе:
Документ подготовил(и): «Сергей Соколов »
Номер версии документа: 1.0
Язык документа: ru
Создан с использованием: EditPad Pro, FTools, ClearTXT
Дата создания документа: 2002-11-03
НОВОЕ | О FICTIONBOOK2 | РЕГИСТРАЦИЯ | ДОБАВЬ КНИГУ | СТАТИСТИКА | СПОНСОР
Интернет

В планах стоит и «предполагаемая оценка» для пользователя - предсказание того, понравится ли пользователю книга, и насколько, на основе поиска пользователей с аналогичными вкусами. Библиотека хранит оценки каждого пользователя и благодаря этому может строить гипотезы, отыскивая пользователей с похожими оценками. Но эта задача, помимо программирования, потребует и весьма значительных вычислительных ресурсов, которых пока, увы, нет.

Техническая реализация FictionBook.lib

Библиотека сделана на основе mod_perl, PostgreSQL, LibXML+LibXSLT. Проверка поступающих документов на соответствие схеме реализована с xerces-c. Для повышения быстродействия и снижения требований к ресурсам (экономия памяти, которую mod_perl в купе с LibXSLT и PostGreSQL используют очень активно) используется squid. Новая книга анализируется, и вся метаинформация заносится в БД. XML используется в библиотеке повсеместно, извлеченные из БД данные форматируются в XML и передаются XSLT-обработчику. Фактически, библиотека имеет трехуровневую структуру - База данных SQL, mod_perl обработчик, в котором реализована основная логика, и набор XSLT, отвечающий за представление подготовленных в Perl данных. Близки к завершению работы по внедрению движка библиотеки (в основном изменения коснулись XSLT) в библиотеку Альдебарана.

Поскольку в библиотеке большое внимание уделяется сбору и анализу статистики, загрузка процессора становится узким местом. Эта проблема решена через создание внутреннего кэша с разделением по уровню доступа пользователей и другими особенностями, позволяющими персонифицировать библиотеку без излишнего пересчета статистики. Благодаря этому библиотека свободно обслуживает до 50 тыс. запросов в сутки при загрузке процессора порядка 10-60% (два Pentium III 1GHz, 1GB RAM, ATA66 HDD). Низкоскоростной и неустойчивый канал в данный момент является узким местом в плане производительности.

Об авторе

Грибов Дмитрий Петрович - 1975-го года рождения. С отличием окончил в 1999-м году МГСУ по специальности Психология (социальная психология, психология управления), но психологом не работал.

Известен, в частности, благодаря своей программе ClearTXT - первая достаточно мощная программа для подготовки разноформатных электронных текстов к чтению на мобильных устройствах и ПК. Программа номинировалась на премию «Идущий с книгой».

Активно участвует в создании различного П.О. для электронных книг (ClearTXT, Any2FB, FB2Any) и развитии формата FictionBook. Разработчик и администратор сетевой библиотеки FictionBook.lib. Уделяет много внимания простоте

использования своих продуктов.

Любимые инструменты - Perl, xslt, Oracle. Любимые занятия - чтение, программирование.

© *Д.П. Грибов, 2004*