
Трансформация FictionBook2 — DocBook — XSL-FO — PDF

KiR

История переиздания		
Издание 0.1.bzr322	2008.09.03	
Changed for package 0.1.bzr320		
Издание 0.1.bzr255	2007.10.15	
Changed for package 0.1.bzr253		
Издание 0.1.bzr252	2007.10.13	
Changed for package 0.1.bzr252		
Издание 0.1.bzr236	2007.08.30	
Initial revision		

Содержание

Преобразование FictionBook2 — DocBook	1
Описание XSL-трансформации	1
Сопоставление FictionBook2 и DocBook	3
Скрипт преобразования FictionBook2 — DocBook — XSL-FO — PDF	4
Описание	4
Использование	4
Конфигурационный файл	5
Шаблоны	5
Необходимые программные пакеты и библиотеки	6
Замечания по генерации PDF	7
Download	7
Бинарный пакет для Win32	7
Исходный код	8
Ссылки	8
Changelog	9

Преобразование FictionBook2 — DocBook

Описание XSL-трансформации

Преобразование FictionBook2 — DocBook написано на XSLT v1.0 с расширениями: элементом `xsl:document` и функцией `document()`. Данный XSLT-сценарий выполняет следующие действия:

- Переводит FictionBook 2 в DocBook
- Для полей без соответствия в DocBook из `fb:description` пишет соответствующие приложения в конце документа (опционально, см. Таблица 1, «Шаблонные параметры трансформации FictionBook2 — DocBook»);

- Создает индексный файл, в котором находятся:
 - список файлов с их `mime`-типом;
 - метainформация о документе (авторы, название);
 - размеры и разрешение (в dpi) вывода.
- Выводит все вложения в `.fb2` (тег `binary`) в отдельные файлы

Об интернационализации. В `fb2docbook/110n` находятся файлы для поддержки языков (сейчас русский и английский). Если файл под какой-то язык отсутствует, то выбирается язык по умолчанию (см. Таблица 1, «Шаблонные параметры трансформации FictionBook2 — DocBook»).

Входные параметры для xslt-трансформации.

1. `conv_info_idx` — индексный файл. См. transformation output index [2].
2. `document-element` — название элемента типа `xsl:document`. Зависит от используемого процессора.

Шаблонные параметры трансформации. Находятся в шаблонах параметров, используемых для конвертации DocBook — XSL-FO, см. Таблица 6, «Шаблоны основных страниц». Одинаковы для всех шаблонов (кроме начинающихся с `output.`), см. включаемый файл `params/p_fb2_common-default.xsl`.



Внимание

Обязательно прочитайте про параметры `output.*`, `page.width` и `page.height`. В скрипте `btransformer.py` пока обрабатываются только такие варианты записи.

Таблица 1. Шаблонные параметры трансформации FictionBook2 — DocBook

Название параметра	Описание
<code>fb2.book-parts</code>	Указывает, как вложенные <code>fb:section</code> должны заменяться на соответствующие структурные элементы DocBook (например, на первом (верхнем) уровне -- это <code>part</code> , а идущие на том же уровне <code>fb:epigraph</code> и <code>fb:image</code> идут в <code>partintro</code>).
<code>fb2.image.align</code>	Горизонтальное расположение изображения (в FB2 такое указание для рисунков отсутствует).
<code>fb2.image.valign</code>	Вертикальное расположение изображения (в FB2 такое указание для рисунков отсутствует).
<code>fb2.default.language</code>	Язык по умолчанию, если он не указан в FB2-документе или отсутствует его интернационализационная поддержка.
<code>fb2.print.infos</code>	Булев параметр, выводить или нет информацию в виде приложений о FB2-документе.
<code>fb2.not_in_toc</code>	Значение для элемента <code>role</code> , в случае отсутствия названия (тег <code>fb2:title</code>) у элемента <code>fb2:section</code> . Используется при дальнейших преобразованиях из Docbook для исключения пустых элементов в содержании. См. файл параметров <code>params/p_fb2_common-default.xsl</code> .

Таблица 2. Шаблонные параметры трансформации FictionBook2 — DocBook, записываемые в индексный файл и используемые обработчиком бинарных файлов

Название параметра	Описание
<i>output.images_mode.resize</i>	Булев параметр. Определяет, преобразовывать ли изображения. Возможные варианты записи: 1, 0, True, true, False, false. Если истинен, обязательно определять все нижеследующие параметры!
<i>output.dpi.width</i>	Количество точек на дюйм по горизонтали. Записывается просто числом в строке!
<i>output.dpi.height</i>	Количество точек на дюйм по вертикали. Записывается просто числом в строке!
<i>output.max_image_margin.width</i>	Если умножить на 2 и полученное число вычесть из ширины страницы в пикселях, то получим максимальный горизонтальный размер изображения. В пикселях! Записывается просто числом в строке!
<i>output.max_image_margin.height</i>	Если умножить на 2 и полученное число вычесть из высоты страницы в пикселях, то получим максимальный вертикальный размер изображения. В пикселях! Записывается просто числом в строке!

Таблица 3. Шаблонные параметры трансформации FictionBook2 — DocBook, записываемые в индексный файл и используемые обработчиком бинарных файлов, и трансформацией DocBook — XSL-FO

Название параметра	Описание
<i>page.width</i>	Ширина страницы. В миллиметрах. Записывается как действительное число (разделитель — точка), далее без пробела 'mm'. Например, '88.184mm'
<i>page.height</i>	Высота страницы. В миллиметрах. Записывается как действительное число (разделитель — точка), далее без пробела 'mm'. Например, '113.854mm'

Сопоставление FictionBook2 и DocBook

TODO

Скрипт преобразования FictionBook2 — DocBook — XSL-FO — PDF

Описание

Скрипт написан на Python. Служит "обвязкой" для xsl-трансформаций и преобразований с помощью других инструментов, скрывая детали низкоуровневой реализации. Переводит вложенные в .fb2 файлы обратно в бинарный вид и выполняет трансформацию изображений (изменение размера и цветности).

Использование

```
btransformer.py {--config config_file} [ options ...] { input_file } [output_file]
```



Замечание

Входной .fb2-файл может быть сжат:

- **gzip** (расширение .gz)
- **bzip2** (расширение .bz2)
- **zip** (расширение .zip), в данном случае обрабатывается единственный файл в архиве, иначе с именем, совпадающим с названием архива, но без суффикса '.zip'

Таблица 4. Опции командной строки

Название опции (с возможными синонимами)	Параметр опции	Описание
<code>--version</code>		Вывод версии программы и завершение ее выполнения.
<code>-h, --help</code>		Вывод описания опций программы и завершение ее выполнения.
<code>-c, --config</code>	<i>FILE</i>	Обязательный параметр. Указывает конфигурационный файл <i>FILE</i> .
<code>--wh</code>		Включение win32hack — закрытие консольного окна в системе Win32 происходит только после нажатия любой клавиши.
<code>--no-remove_tmp</code>		Отключить удаление временных файлов.
<code>- - n o - update_metadata</code>		Отключить обновление метаданных в выходном pdf-файле.
<code>--xslt_proc</code>	<i>XSLT_PROC</i>	Использовать <i>XSLT_PROC</i> из конфигурационного файла как XSLT процессор.
<code>--fo_proc</code>	<i>FO_PROC</i>	Использовать <i>FO_PROC</i> из конфигурационного файла как XSL-FO процессор.
<code>--pdf_proc</code>	<i>PDF_PROC</i>	Использовать <i>PDF_PROC</i> из конфигурационного файла как процессор, обновляющий метаданные в pdf-файле.
<code>--pdf_optimizer</code>	<i>PDF_OPTIMIZER</i>	Использовать <i>PDF_OPTIMIZER</i> из конфигурационного файла как pdf-оптимизатор.
<code>--xslt_params</code>	<i>FILE</i>	Использовать шаблонный файл <i>FILE</i> для преобразования из DocBook.
<code>--xslt_titlepages</code>	<i>FILE</i>	Использовать файл с описанием форматирования титульных страниц <i>FILE</i> для преобразования из DocBook.
<code>--translify^a</code>		Использовать транслитерацию на латиницу в метаданных pdf-файла.

^aПараметр `--translify` доступен, если установлены модули `pyutils` и `Beautiful Soup`.

Конфигурационный файл

TODO

Шаблоны

В пакеты входят шаблоны, приспособленные для вывода pdf из формата FB2 и для устройства Sony Reader. Находятся в директории `params` дистрибутива.

Таблица 5. Шаблоны для заглавных страниц

Название файла шаблона	Формат или устройство	Примечание
titlepage.templates_prs-default.xml	Sony Reader PRS-50X	Уменьшены размеры шрифтов заглавий
titlepage.templates_a4-default.xml	Бумага, размер A4	Изменений не вносилось (копия из дистрибутива Docbook XSL Stylesheets)
titlepage.templates_a5-default.xml	Бумага, размер A5	Для распечатки по две страницы на A4 Изменений не вносилось (копия из дистрибутива Docbook XSL Stylesheets)

Таблица 6. Шаблоны основных страниц^a

Название файла шаблона	Формат или устройство	Шрифты	Примечание
params_prs-font.liberation.xsl	Sony Reader PRS-50X	Шрифты Liberation [https://www.redhat.com/promo/fonts/]	
params_prs-font.vista.xsl	Sony Reader PRS-50X	Шрифты из Vista основной — Cambria без засечек — Candara моноширинный — Consolas	
params_a4-font.liberation.xsl	Бумага, размер A4	Шрифты Liberation [https://www.redhat.com/promo/fonts/]	
params_a4-font.vista.xsl	Бумага, размер A4	Шрифты из Vista основной — Cambria без засечек — Candara моноширинный — Consolas	
params_a5-font.liberation.xsl	Бумага, размер A5	Шрифты Liberation [https://www.redhat.com/promo/fonts/]	Для распечатки по две страницы на A4
params_a5-font.vista.xsl	Бумага, размер A5	Шрифты из Vista основной — Cambria без засечек — Candara моноширинный — Consolas	Для распечатки по две страницы на A4

^aВо всех шаблонах параметры для преобразования FictionBook2 — DocBook одинаковы.

Необходимые программные пакеты и библиотеки

- XSLT-процессор, поддерживающий спецификацию XSLT v1.0 и инструкцию типа `xsl:document` из спецификации XSLT v1.1. Один из:
 - Libxslt [http://xmlsoft.org/XSLT/] (как в виде модуля для языка Python, так и в виде отдельной программы **xsltproc**)
 - Saxon v6 [http://saxon.sourceforge.net/]
 - 4suite [http://4suite.org/]

- Xalan [<http://xml.apache.org/xalan-j/index.html>] ¹
- XSL Stylesheets for Docbook [<http://wiki.docbook.org/topic/DocBookXslStylesheets>] v1.72.0 и выше
- XSL-FO-процессор. Один из:²
 - FOP [<http://xmlgraphics.apache.org/fop/>] v0.93 и выше
 - XEP [<http://www.renderx.com/>]
- Python [<http://www.python.org>] v2.4 и выше
- Python imaging library (PIL) [<http://www.pythonware.com/products/pil/index.htm>]
- Pdftk [<http://www.pdfhacks.com/pdftk>]

Замечания по генерации PDF

Генерация в FOP. FOP пишет в .pdf только содержание (bookmarks), другую метаинформацию (автор, название) приходится дописывать с помощью pdftk. К сожалению в таком случае Acrobat Pro/Adobe Reader (по крайней мере версии 8) показывает только поле Title (название), при этом вьюер в Sony Reader показывает все поля. Не поддерживает вывод линейаризованного PDF.

Генерация в XEP. XEP имеет расширение к XSL-FO, которое позволяет дописывать всю метаинформацию, и в .pdf, полученных с использованием этого процессора, в Adobe Acrobat виден весь description (описание). Поддерживает линейаризацию при выводе .pdf-файла (опция конфигурации LINEARIZE, отключена по умолчанию).

Линейаризация. Для комфортного просмотра pdf-документов при медленном доступе к файлу или при малом количестве оперативной памяти, необходимо проводить линейаризацию документа (так называемый 'fast web view' в Acrobat). Для этого можно использовать либо Adobe Acrobat (File - Save as (Adobe pdf files, optimized, *.pdf) и выставить соответствующую опцию в Settings), либо скрипт **pdfopt** из пакета Ghostscript [<http://en.wikipedia.org/wiki/Ghostscript>]. Начиная с версии 0.1.bzr252 в fb2docbook включена поддержка оптимизатора **pdfopt**. Это позволяет сразу получать файлы, пригодные к комфортному чтению на устройствах Sony Reader.

Download

Бинарный пакет для Win32

Download.

- Текущая версия: fb2d.pkg.bin.win32-0.1.bzr320.zip
[<http://kir666.ru/fb2docbook/download/fb2d.pkg.bin.win32-0.1.bzr320.zip>]
- fb2d.pkg.bin.win32-0.1.bzr253.zip
[<http://kir666.ru/fb2docbook/download/fb2d.pkg.bin.win32-0.1.bzr253.zip>]
- fb2d.pkg.bin.win32-0.1.bzr252.zip
[<http://kir666.ru/fb2docbook/download/fb2d.pkg.bin.win32-0.1.bzr252.zip>]
- fb2d.pkg.bin.win32-0.1.bzr236.zip
[<http://kir666.ru/fb2docbook/download/fb2d.pkg.bin.win32-0.1.bzr236.zip>]

Требования текущей версии. Установленная JVM.

¹Не поддерживает вывод вложенных файлов, т.к. xalan:write всегда выводит в выходном файле xml-заголовок.

²Оба процессора написаны на Java и требуют установленной JVM

Установка и запуск. Распаковать в директорию по выбору. Далее, находясь в данной директории, обязательно запустить **confsetup.exe** для генерации конфигурационных файлов с указанием абсолютных путей к компонентам пакета! Для запуска использовать скрипт **btransformer.cmd**.



Внимание

Начиная с версии с версии 0.1.bzr320 при установке необходим запуск **confsetup.exe**.

Состав пакета текущей версии.

- Собственно fb2docbook [<http://kir666.ru/fb2docbook/>] v0.1.bzr320
- XSL Stylesheets for Docbook [<http://wiki.docbook.org/topic/DocBookXslStylesheets>] v1.74
- Python [<http://www.python.org>] v2.5 — интерпретатор, не требующий установки, со всеми необходимыми библиотеками (получено с помощью <http://cheeseshop.python.org/pypi/bbfreeze/>)
- Python imaging library (PIL) [<http://www.pythonware.com/products/pil/index.htm>] v1.1.6
- Libxslt [<http://xmlsoft.org/XSLT/>] v1.1.20 — модуль для Python
- Pytils [<http://www.pyobject.ru/projects/pytils/>] v0.2.1
- Beautiful Soup [<http://www.crummy.com/software/BeautifulSoup/>] v3.0.4
- FOP [<http://xmlgraphics.apache.org/fop/>] v0.95
- Набор правил переносов для FOP [<http://offo.sourceforge.net/>]
- Pdftk [<http://www.pdfhacks.com/pdftk>] v1.41
- GPL Ghostscript [<http://www.ghostscript.com/awki>] v8.63, без шрифтов
- Набор шрифтов Liberation [<https://www.redhat.com/promo/fonts/>]³

Исходный код

Download.

- Текущая версия: **fb2docbook.src-0.1.bzr320.zip**
[<http://kir666.ru/fb2docbook/download/fb2docbook.src-0.1.bzr320.zip>]
- **fb2docbook.src-0.1.bzr253.zip**
[<http://kir666.ru/fb2docbook/download/fb2docbook.src-0.1.bzr253.zip>]
- **fb2docbook.src-0.1.bzr252.zip**
[<http://kir666.ru/fb2docbook/download/fb2docbook.src-0.1.bzr252.zip>]
- **fb2docbook.src-0.1.bzr236.zip**
[<http://kir666.ru/fb2docbook/download/fb2docbook.src-0.1.bzr236.zip>]

Ссылки

1. FictionBook 2.0 Schema - комментарии
<http://www.fictionbook.org/2.0/schema/>

³Метрики для FOP'a в архиве отсутствуют, так как начиная с версии 0.94 больше не требуются.

2. DocBook: The Definitive Guide [<http://www.docbook.org/tdg/en/html/docbook.html>]

3. DocBook XSL: The Complete Guide [<http://www.sagehill.net/docbookxsl/>]

Changelog

v0.1.bzr320.

- Секции (тег *fb2:section*) без заголовий fb2-документа в выходном файле более не выводятся в содержание (это не относится в метаинформации в PDF-файле!).
- Исправлены ошибки в XSL преобразовании из FB2.
- В версию пакета под Win32 добавлен скрипт, генерирующий конфигурационные и стартовые файлы. В связи с чем запуск преобразователя возможен из любой текущей директории, так как он больше не завязан на относительные пути компнент.
- В шаблонах PRS-50X включена генерация печатного содержания, так как GPL Ghostscript, начиная с версии 8.63, корректно линеаризирует pdf-файлы.
- Обновлено инструменты в пакете под Win32.
- Исправления в документации.

v0.1.bzr253.

- Улучшена работа с упакованными zip'ом файлами: единственный .fb2 внутри может носить любое имя, иначе выбирается файл, совпадающий с именем архива, но без суффикса '.zip'.
- Добавлена опциональная поддержка транслитерации метаинформации в PDF (автор/название).

v0.1.bzr252.

- Добавлена поддержка оптимизатора (линеаризатора) pdfopt из пакета Ghostscript.
- Тег fb:subtitle заменяется теперь на docbook:bridgeahead.
- Другие несущественные изменения в XSLT-сценарии.
- Обновлено инструменты в пакете под Win32.
- Обновлено шаблоны трансформации Docbook.
- И другое.