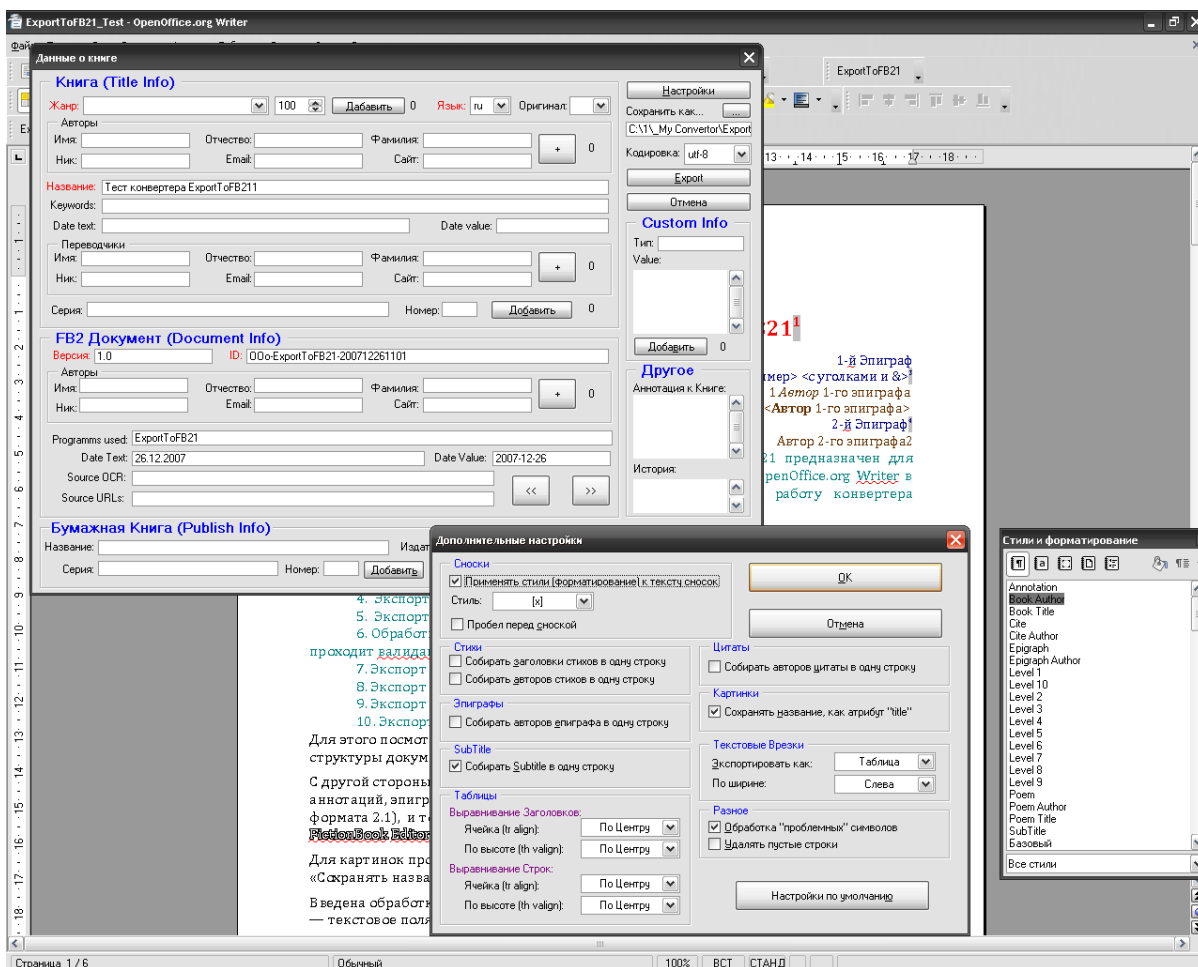


Описание работы конвертера ExportToFB21

1. Описание



1.1. Назначение

Кроссплатформенный конвертор ExportToFB21 предназначен для конвертации документов из форматов, поддерживаемых **OpenOffice.org Writer** в формат **fb2.1**.

Т.о. входные форматы документов для ExportToFB21 следующие: .doc, .dot, .rtf, .txt, .html, .htm, .oth, .odt, .ott, .stw, .sdw, .xml, .vor, .psw, .wps, .wpd, .fodt, .sxw, Microsoft Word 2003XML .xml, и другие, с которыми работает OpenOffice.org Writer.

Конвертор работает на тех платформах, где работает OpenOffice.org Writer, т.е. MS Windows 98/2000/XP/Vista, MacOS X, GNU/Linux, FreeBSD и т.д.

Конвертор не использует DOM технологию – задействован свой парсинг документов и печать в текстовый файл информации, собранной после анализа.

Если документ отформатирован стилями (см. [п.3.](#)) правильно (см. [п.5.](#)), то после его конвертации практически не требуется дальнейшая правка (за редким исключением, о чем см. [п.7.](#)) в редакторах fb2 формата! После проведения десятков тестов с документами различной структурой, полученные fb2 книги у меня всегда проходили валидацию.

1.2. Происхождение и история конвертора

1.2.1. ExportToFB2.1 Version 1.0

Конвертор ExportToFB2.1 - это переработка и развитие конвертора ExportToFictionBook (web: <http://ifolder.ru/3048108>), автором которого является **Thunder**. Это прекрасный конвертор, с интересной идеей его реализации! К сожалению, этот конвертор, поддерживал только 2 уровня вложений заголовков. Поэтому я решил переписать его «под себя», а сначала и для себя, изменив полностью парсинг заголовков, введя возможность использование практически бесконечного числа уровней заголовков. Хотелось подогнать конвертор под формат fb2.1 как можно «ближе» и многое другое. Очень многое здесь переделано, расширено, модернизирован парсинг абзацев, сносок, введены новые алгоритмы, некоторые процедуры остались от исходного кода. Но идея - от **Thunder`a**, за что ему огромная благодарность! Я списался с ним, и он дал добро на использование его исходных кодов для создания других программных продуктов.

1.2.2. ExportToFB2.1 Version 1.1

1. Дополнения: В документах без единого Уровня (Заголовка) если в начале идет Эпиграф, то он не включается в <section> (раньше – включался, хотя результирующий документ оставался валидным).

2. Исправления: В некоторых сложных структурах документа на смежных Уровнях секция не закрывалась (</section>), в результате чего fb2 документ не проходил валидацию. Теперь все работает.

1.2.3. ExportToFB2.1 Version 2.0

Полностью «с нуля» переписан весь парсер документа, в результате чего появились новые возможности конвертора. О них – ниже.

1. Экспорт таблиц с заголовками. В связи с разными возможностями OOo Writer 2.3 и версий ниже экспорт таблиц происходит по-разному. Для версий, ниже 2.3 таблицы экспортируются только с одним заголовком (если он есть), для версии 2.3 – со всеми заголовками, сколько их есть. Конвертор сам автоматически определяет версию OOo Writer`a, и в соответствии с этим конвертирует таблицы. Работу с таблицами в документе см. [п. 4.1](#).

2. Экспорт текста сносок, состоящих из нескольких параграфов (вместе с форматированием). В Version 1.0 и Version 1.1 экспортировался только первый параграф каждой сноски.

3. Экспорт в нужные места тэгов картинок и их названия – в атрибут "title" (сами картинки пока не экспортируются). Картинки могут быть 3-х видов: Привязанные к абзацу, Привязанные к символу и Привязанные как символ (Буквица). Подробности работы с картинками см. [п. 4.3](#).

4. Экспорт текстовых врезок (без экспорта их форматирования) как просто текста, цитат или таблицы (см. [п. 4.2](#)).

5. Экспорт текстовых полей.

1.2.4. ExportToFB2.1 Version 2.1

1. Добавлена возможность указывать % жанра.

2. Изменен список жанров – выбор не в виде кода, а по названию жанра. Так удобнее.

3. Исправил обработку текстовых врезок – раньше, если попадались Линии – конвертор аварийно завершал работу.

4. Отглавливается нумерация параграфов или маркеры и экспортируется, как символ «•» перед параграфом.

5. Ввел сохранение настроек диалога «Настройки» и возможность установок «по умолчанию».

6. Ввел возможность сохранять и загружать профиль автора документа и Source OSR (кнопки «>>» и «<<» соответственно).

7. При открытии диалога экспорта профиль автора документа, если он был создан, автоматически загружается, заполняя соответствующие поля.

1.3. Что реализовано

Конвертор писался для конвертации документов именно в формат **fb2.1**. Поэтому в нем реализованы почти все основные возможности этого формата.

1.3.1. Экспорт Description:

1. <title-info> - весь
2. <document-info> - весь
3. <publish-info> - весь
4. <custom-info> - весь
5. аннотация к документу
6. история документа

1.3.2. Экспорт основных элементов структуры документа

1. Экспорт **структуры** текста (слов и абзацев).
2. Экспорт **сносок и гиперссылок** (на web странички).
3. Экспорт **стилей** абзацев, аннотаций, эпиграфов, цитат, поэм, заголовков, подзаголовков (т.е. теги , <emphasis>, <strikethrough>, <sub>, <sup>, <code>).
4. Экспорт **стилей** авторов цитат, авторов поем, авторов эпиграфов. Для **поэм** реализованы заголовки - <title>
5. Экспорта **стилей текста сносок** (т.е. все курсивы, болды и т.д. отображаются и в тексте сносок).
6. Для поем (стихов) реализован заголовок <title>
7. Реализована возможность использования Аннотации на любом уровне (части книги, глав...).
8. Экспорт **таблиц** с экспортом *заголовков, выравниванием* текста заголовков и ячеек.
9. Экспорт **тегов картинок и картинок-буквиц** (сами картинки пока не экспортируются).
10. Число уровней вложений <section> - 10. При желании его можно увеличить хоть до бесконечности, внося некоторые изменения в код конвертера (подсказка: ищите строки проверки на "Level 1", "Level 2" и т.д. и дополняйте эти строки до нужного вам числа уровней. Всего процедур таких 3-4). Тогда нужно будет создать и соответствующие стили для новых уровней.
11. Самая главная фишка конвертора при работе с <section> - корректный экспорт самой разной структуры документа, без лишних, или наоборот, недостающих <section> или </section>! Подробности см. ниже. Так что не нужно будет мучиться, отлавливая, где чего не хватает, чтобы документ fb2 прошел валидацию. Конечно, все это - при условии *правильного задания стилями уровней заголовкам!*

1.3.3. Дополнительные возможности

1. Реализовано 10 уровней вложения заголовков (<section><title>Заголовок</title></section>). Но, алгоритм построен таким образом, что путем внесения в 3 процедуры нескольких строк кода, число вложений - уровней можно довести хоть до бесконечности! Мне больше 10 никогда не требовалось.

2. Реализована обработка «проблемных» символов (&, <, >), и замена их в fb2 файле на & < и > соответственно. Иначе книга не проходит валидацию! Эта опция хороша для экспорта *технической литературы*. Для другого вида книг, если вы уверены, что этих символов там нет, опцию лучше отключить – скорость экспорта увеличится в несколько раз!

3. Возможность вставки **перед сноской пробела** (опция в **Настройках**).

4. Возможность выбора **стиля сносок**: [x], [x], {x}, { x }, (x), (x) (опция в **Настройках**).

5. Возможность собирать авторов поемы, цитат или эпитафий в одну строку, или оставить авторов одного под другим (опции в **Настройках**).

6. Возможность собирать **Subtitle в одну строку**, или оставить подзаголовки один под другим (опция в **Настройках**).

7. Возможность удаления пустых строк (опция в **Настройках**).

8. Экспорт **текстовых врезок** как просто текст, как цитат или как таблиц.

9. Экспорт **текстовых полей**, вставленных из меню “Вставка” -> “Поля”.

1.4. Что пока не реализовано

1. Экспорт картинок (пока просто их теги).

2. Экспорт *форматирования* текста врезок.

3. Множественный стиль абзацев, текста сносок, цитат, поем, эпитафий, аннотаций, заголовков, подзаголовков (subtitle). Т.е. фраза, где одновременно некоторое слово было выделено и как bold, и как italic будет только как bold (). *Приоритет* выделения (алгоритм конвертора): HyperLink, , <emphasis>, <sup>, <sub>, <code>, <strikethrough>. Одновременно "пометить" слово или фразу несколькими тэгами (например <emphasis><code>пример</code></emphasis> текста с <emphasis>множественным</emphasis> стилем) пока нельзя - алгоритм не доведен до конца, занимает уйму времени - надо парсить не форматированные блоки текста, а каждое слово по нескольку раз.

4. Теги, обеспечивающие разный доступ к прочтению отдельных мест книги - платные книги, свободные книги...

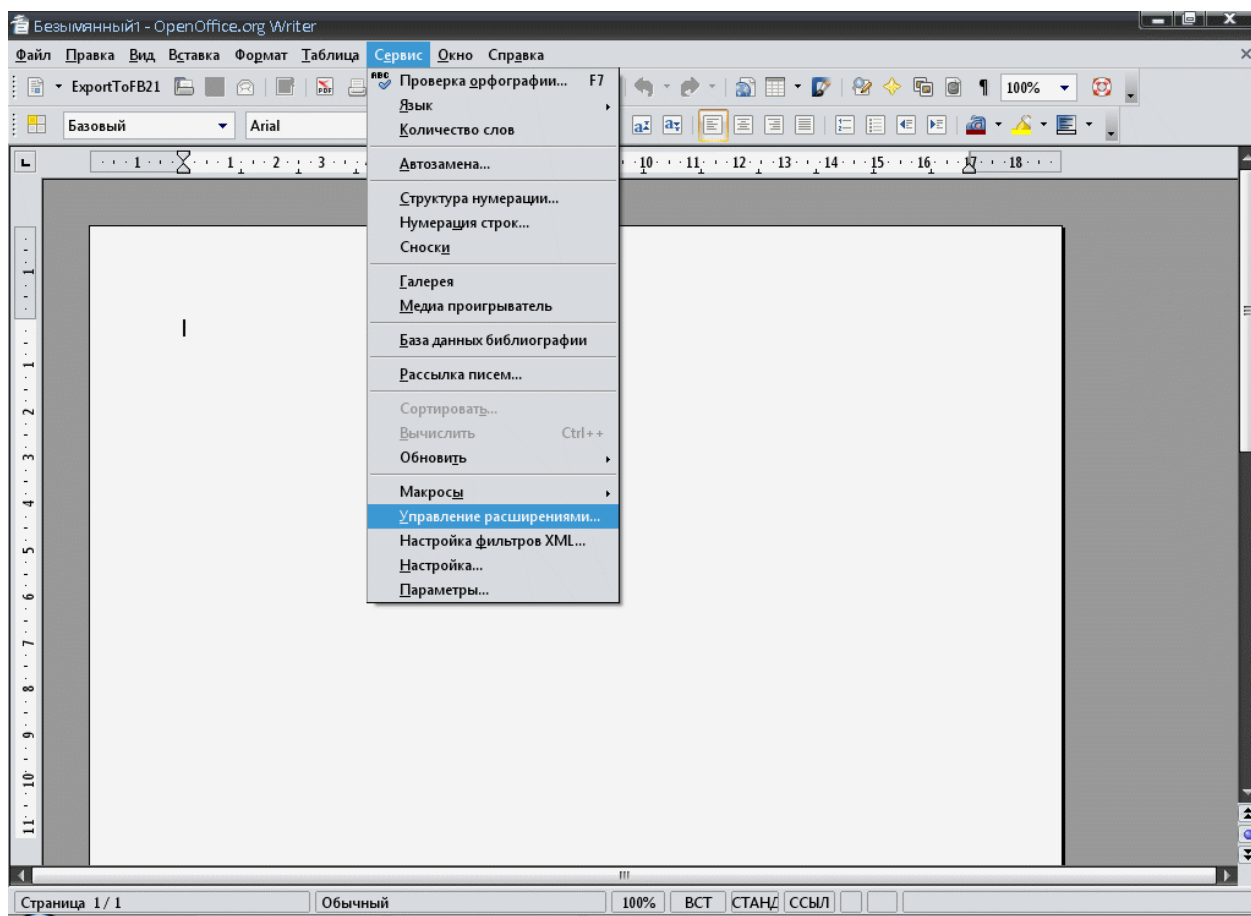
5. Из Description - <source-title-info>

2. Установка конвертора

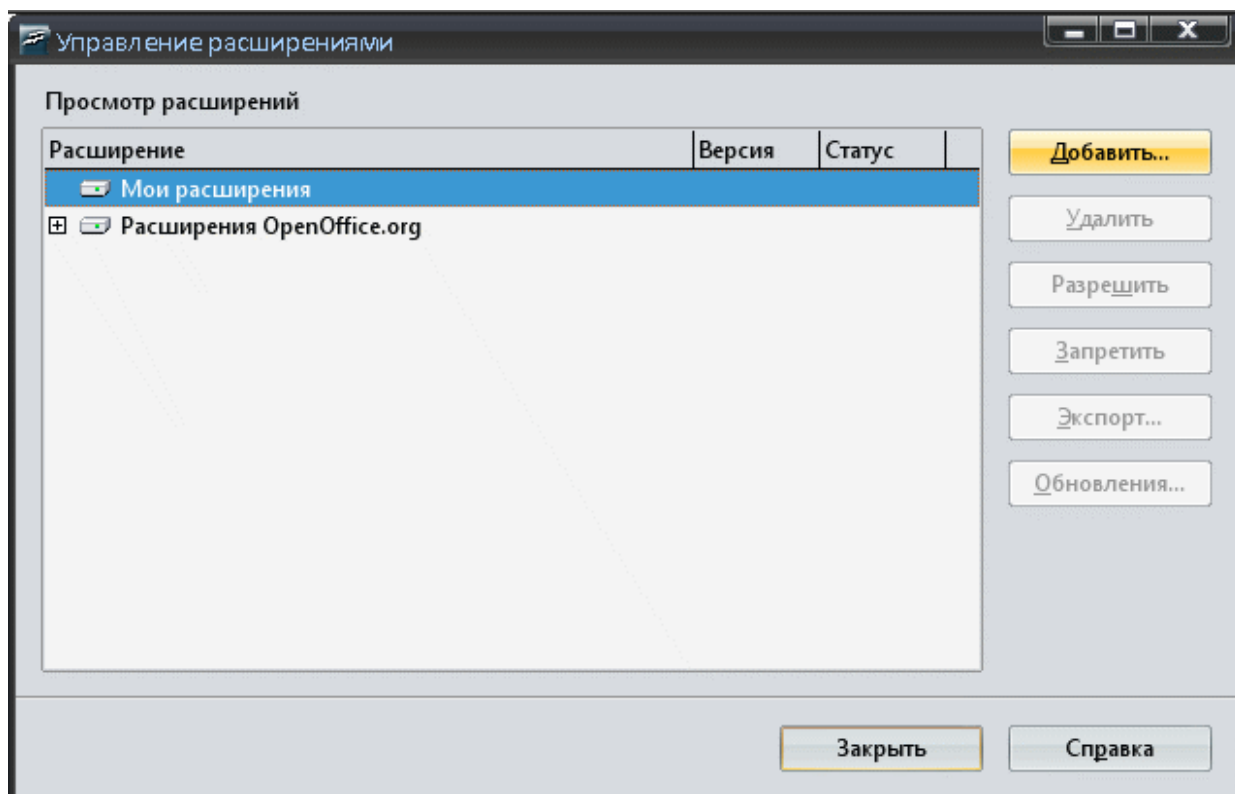
2.1. Установка расширения или .zip архива с конвертором

2.1.1. Первый способ

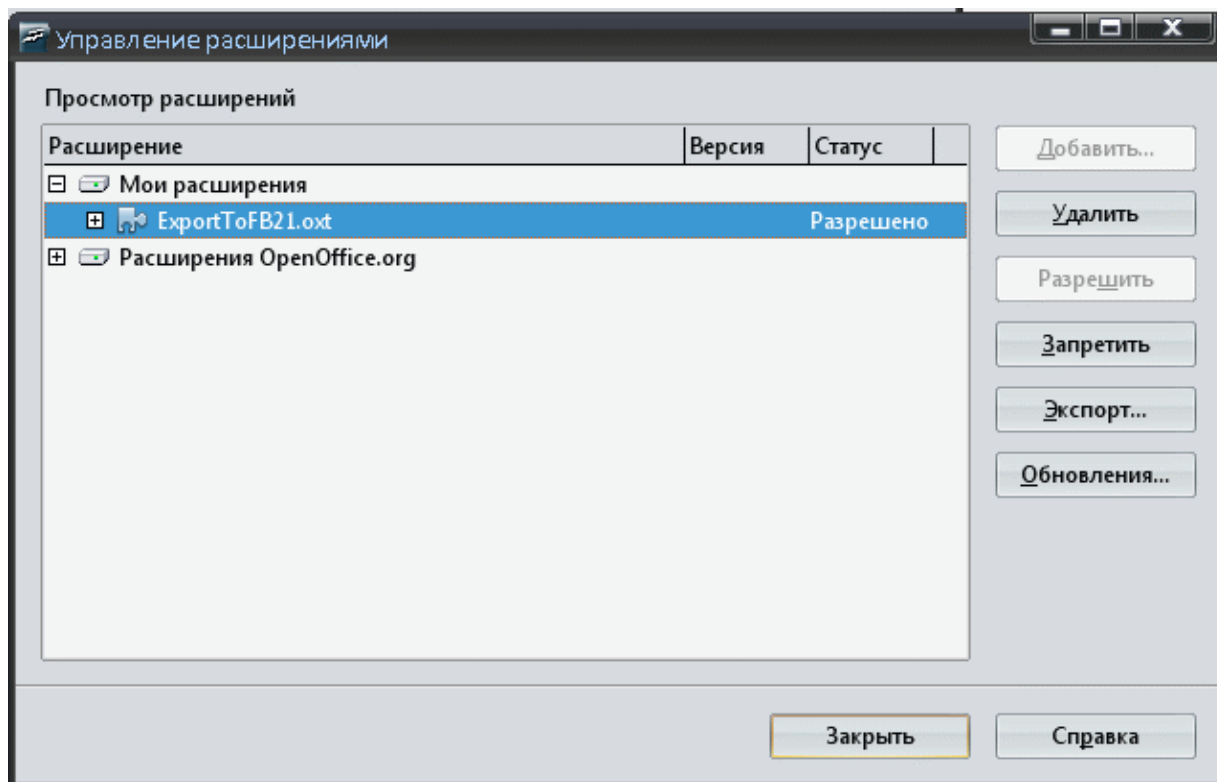
Вызываем диалог «Управление расширениями» через меню Сервис -> Управление расширениями:



В открывшемся диалоге выделяем "Мои расширения", и нажимаем кнопку "Добавить..."



Выбираем расширение **ExportToFB21.oxt**. В результате получаем:



То же самое делаем, если загружаем не расширение .oxt, а .zip архив **ExportToFB21.zip**, в котором находятся файлы диалогов и кодом конвертора.

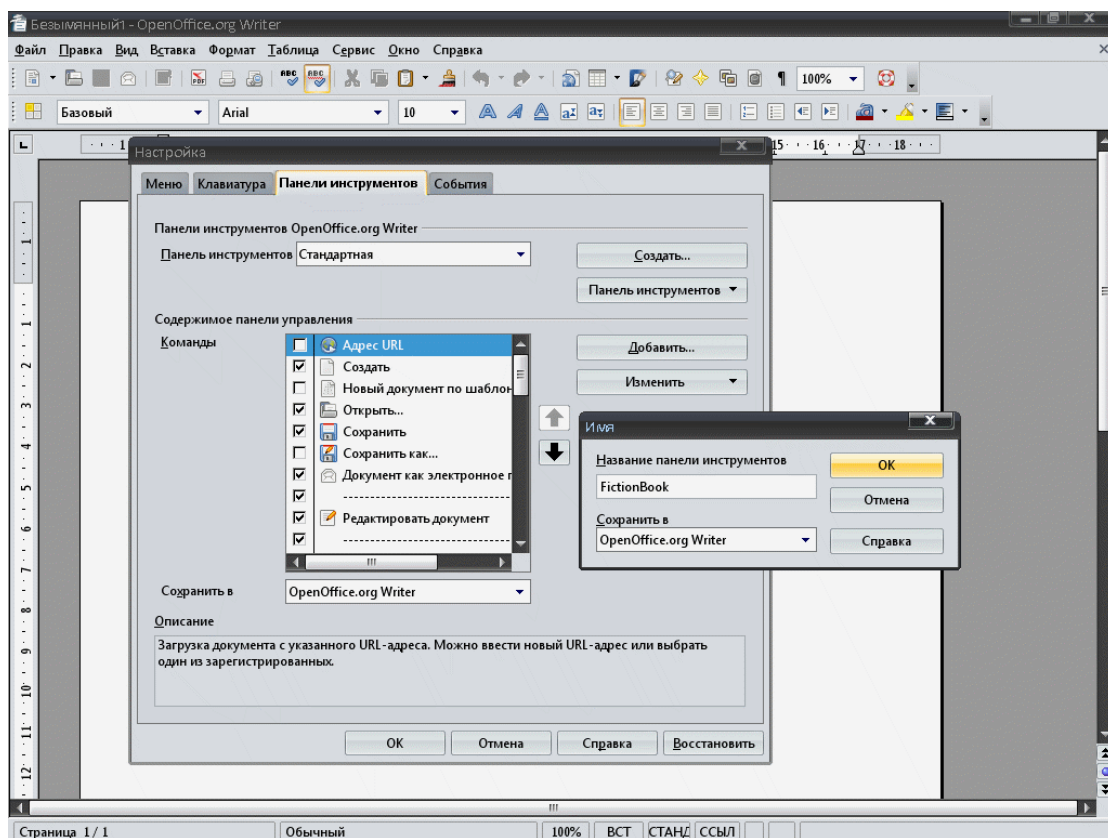
2.1.2. Второй способ

Второй способ применим *только* к установке любого расширения **.oxt** - просто из проводника или файлового менеджера запустите это расширения (по Enter или двойному клику мышки). К .zip это не относится - расширения в архиве .zip устанавливаются так, как описано в первом способе ([п. 2.1.1.](#)).

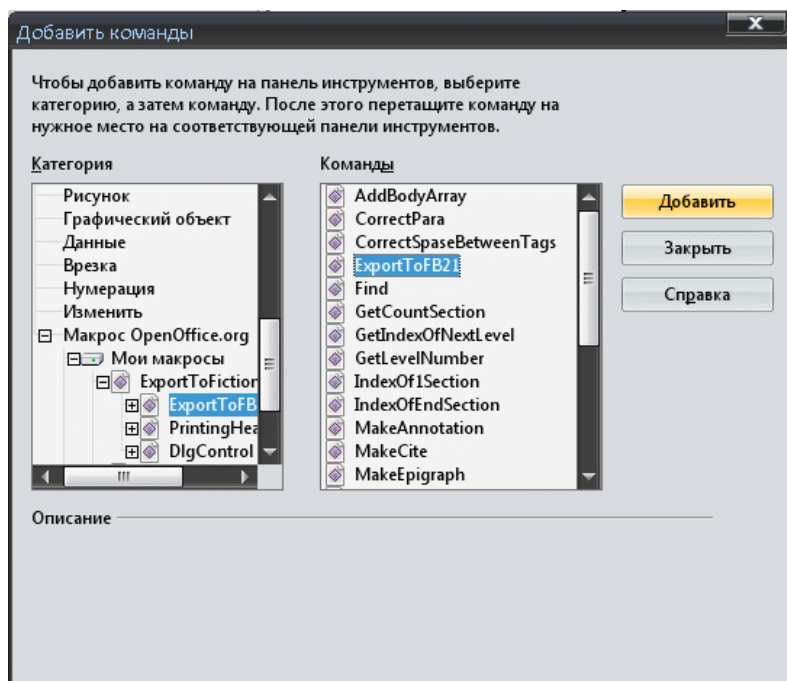
2.2. Вынос кнопки на панель

Для удобства работы с конвертором можно создать панель инструментов с кнопкой экспорта. Делается это так.

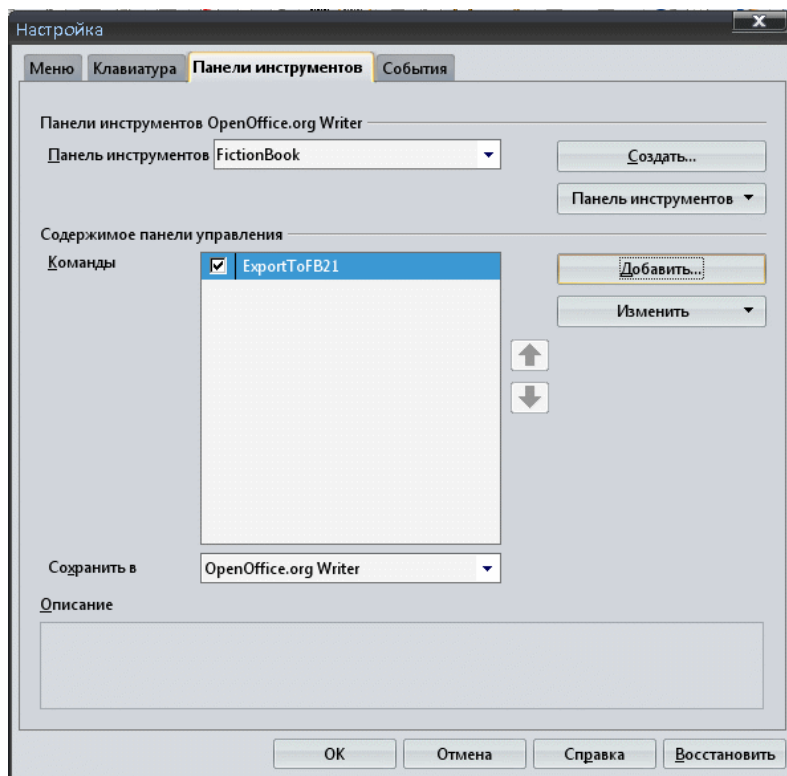
Открываем диалог "Настройка" через меню Сервис -> Настройка. Переходим на вкладку Панели инструментов, нажимаем на кнопку "Создать..." и в открывшемся диалоге вводим название новой панели (в моем случае FictionBook), в "Сохранить в" выбираем OpenOffice.org Writer, нажимаем ОК:



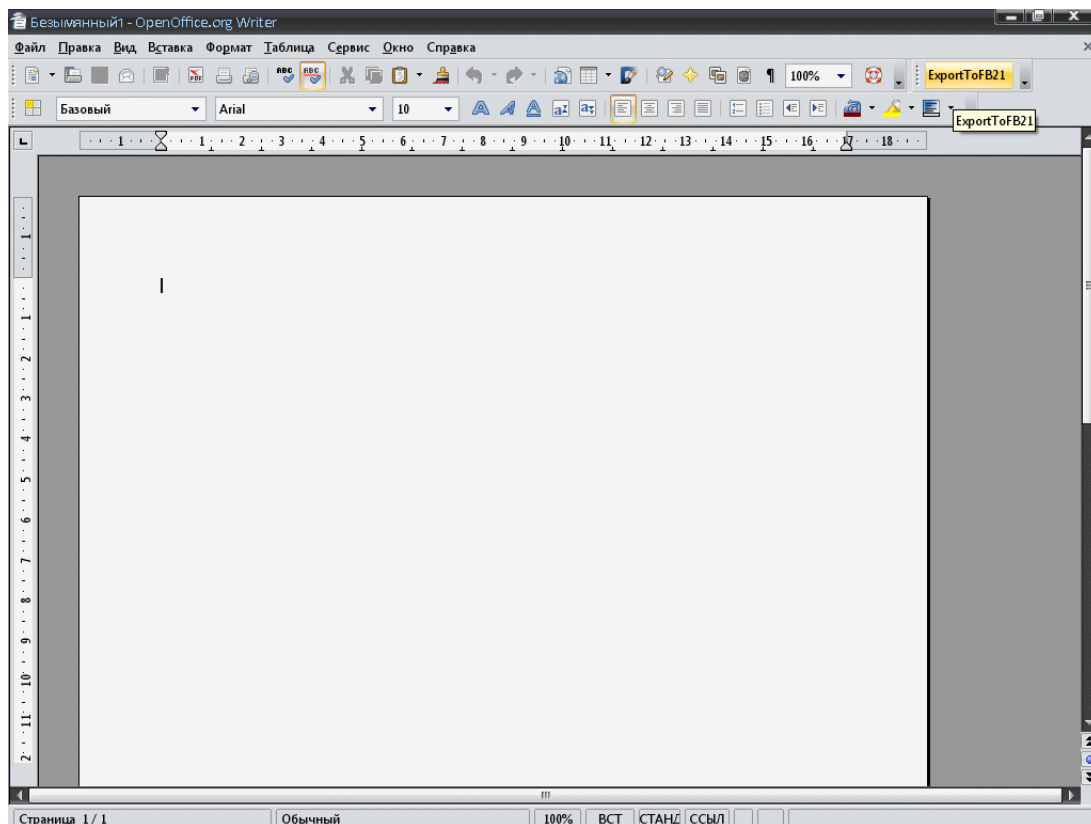
Теперь нажимаем кнопку "Добавить"... диалога "Настройка" и в открывшемся диалоге "Добавить команды" в древовидном списке "Категория" находим элемент Макрос OpenOffice.org / Мои макросы / ExportToFictionBook / ExportToFB21. В списке "Команды" отобразится список макросов. Выбираем ExportToFB21 и нажимаем кнопку "Добавить".



Нажимаем "Закреть". В результате получаем следующее:



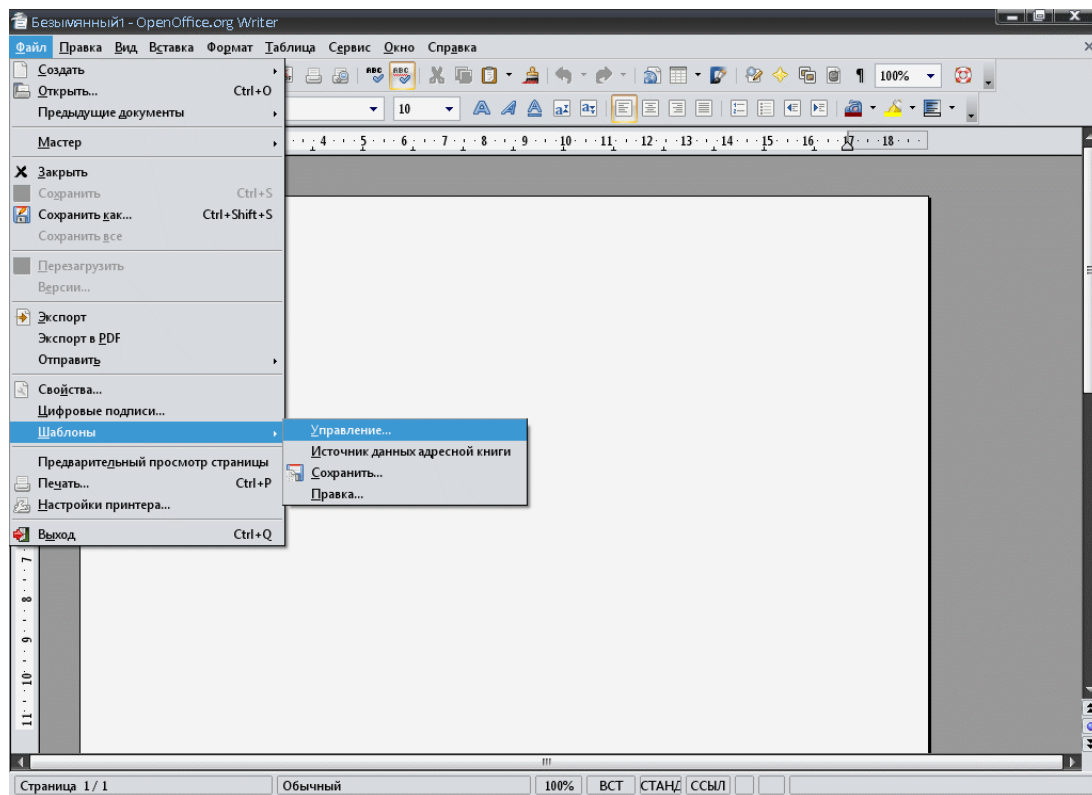
После нажатия "OK" мы увидим новую панель инструментов с нашей кнопкой "ExportToFB21"



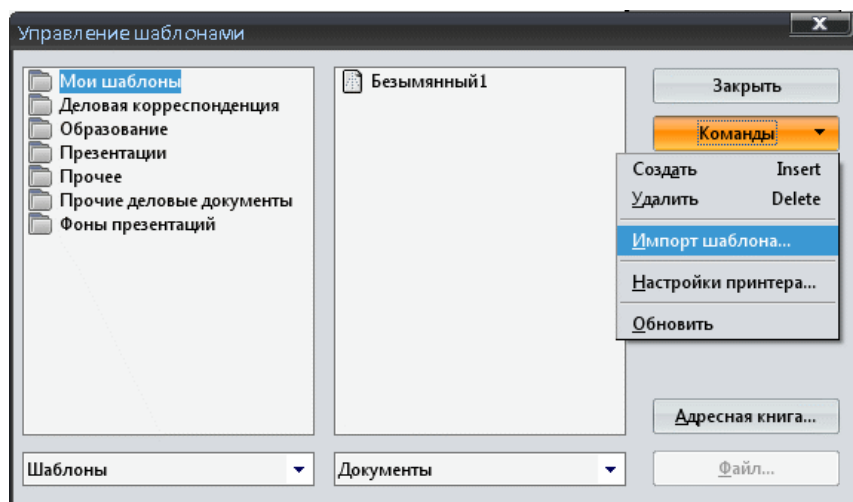
2.3. Установка шаблона стилей fb21_styles.ott

Для экспорта в формат fb2.1 необходимо "разметить" документ специальными стилями.

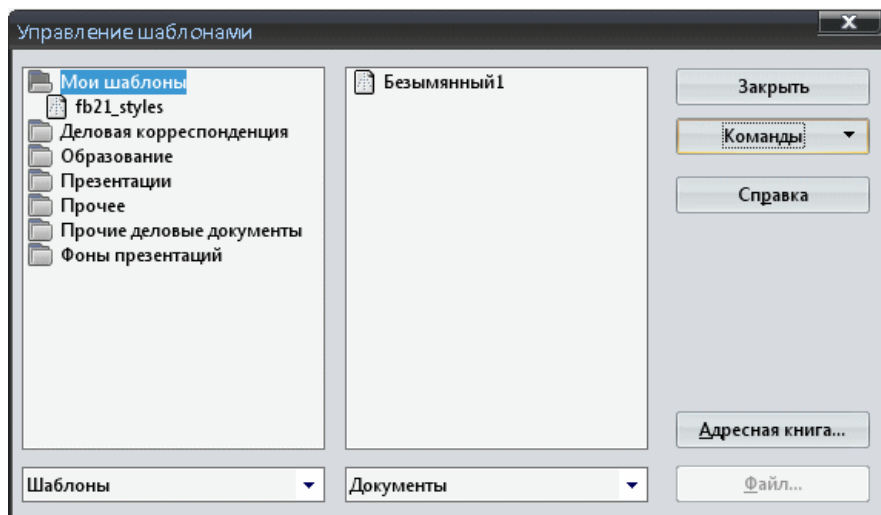
Для того, чтобы их установить, надо запустить диалог "Управление шаблонами" через меню Файл -> Шаблоны -> Управление



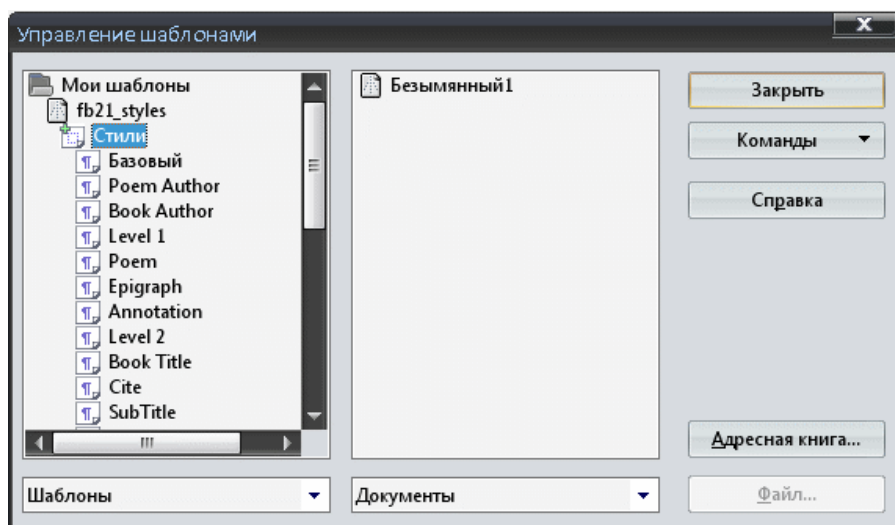
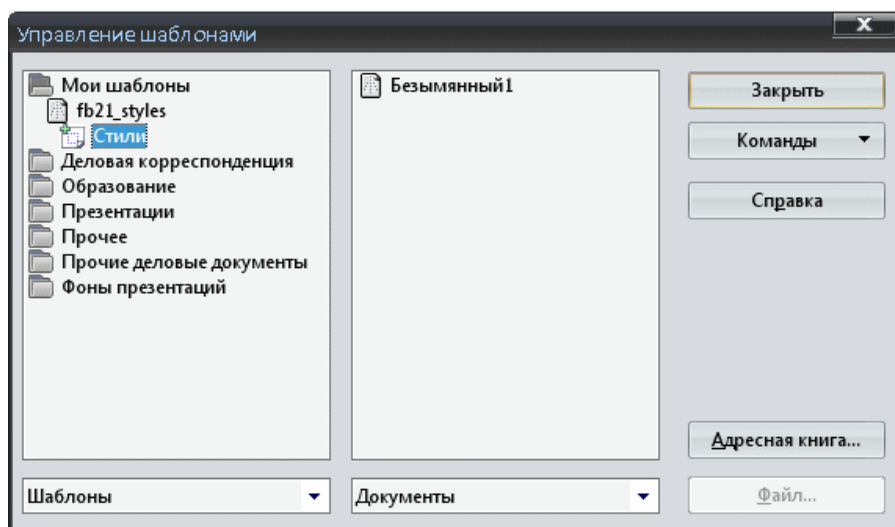
В открывшемся диалоге выделяем в списке шаблонов "Мои шаблоны", нажимаем кнопку "Команды" и выбираем "Импорт шаблона..."



Выбираем и устанавливаем шаблон **fb21_styles.ott**. Результат:



Кликнув два раза по fb21_styles, а потом – два раза по «Стили» мы можем убедиться, что стили действительно установлены:



Нажимаем кнопку "Заккрыть" - все, мы готовы к работе.

3. Работа со стилями структуры документа

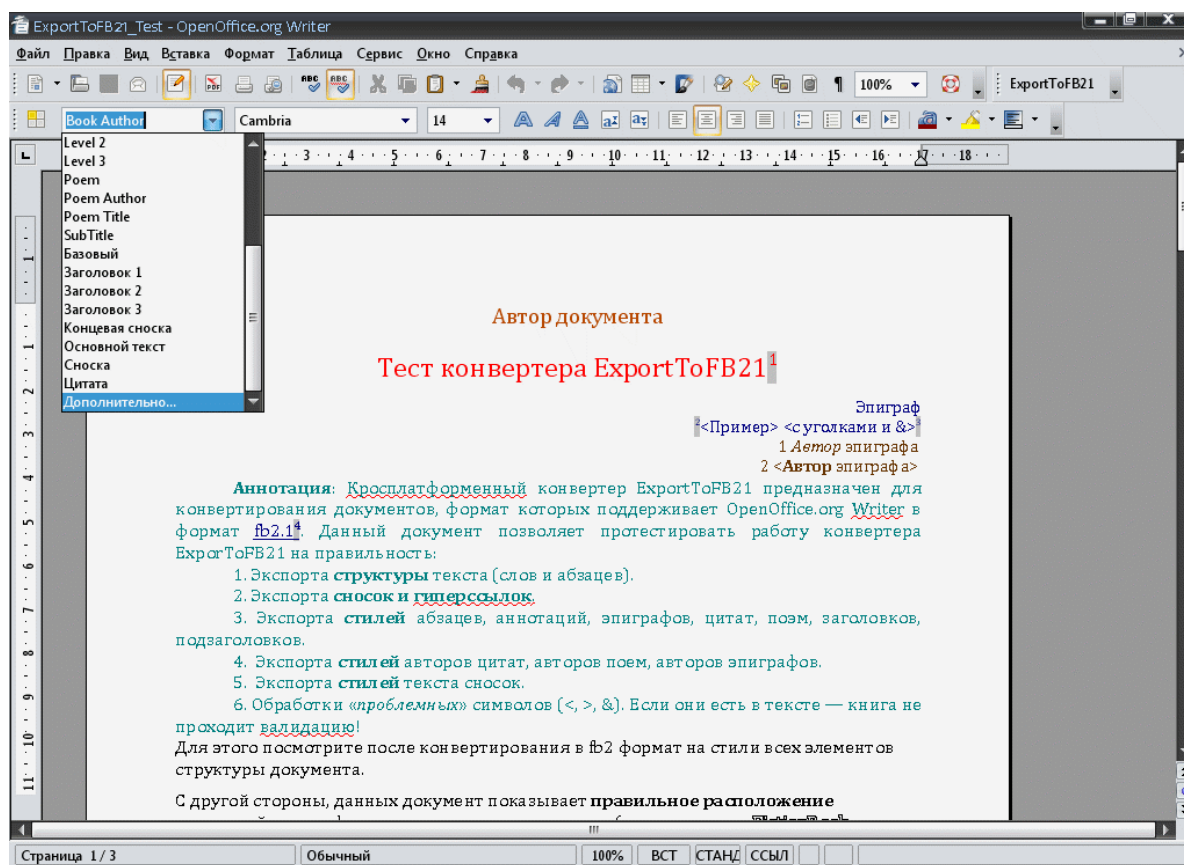
3.1. Загрузка стилей

При экспорте конвертор ищет в документе определенные стили и «обрамляем» абзацы, помеченные этими стилями в соответствующие им теги формата fb2.1 (см. [п.3.3.](#)). Стили придают абзацам определенный вид - цвет, шрифт и т.д. Сами стили можно менять, как вам угодно, изменяя цвет, шрифт. Главное - **название стилей** должно **остаться, как есть!**

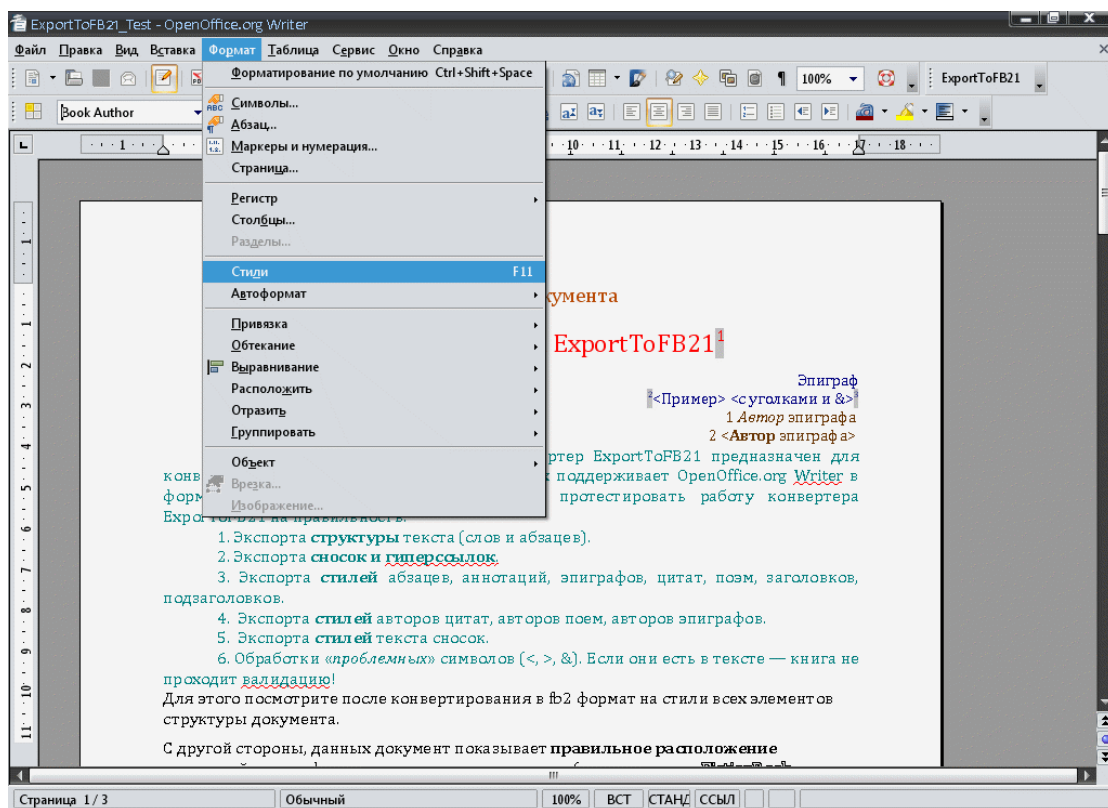
Для этого надо открыть окно «Стили и форматирование», если оно еще не открыто.

3.1.1. Открытие окна стилей

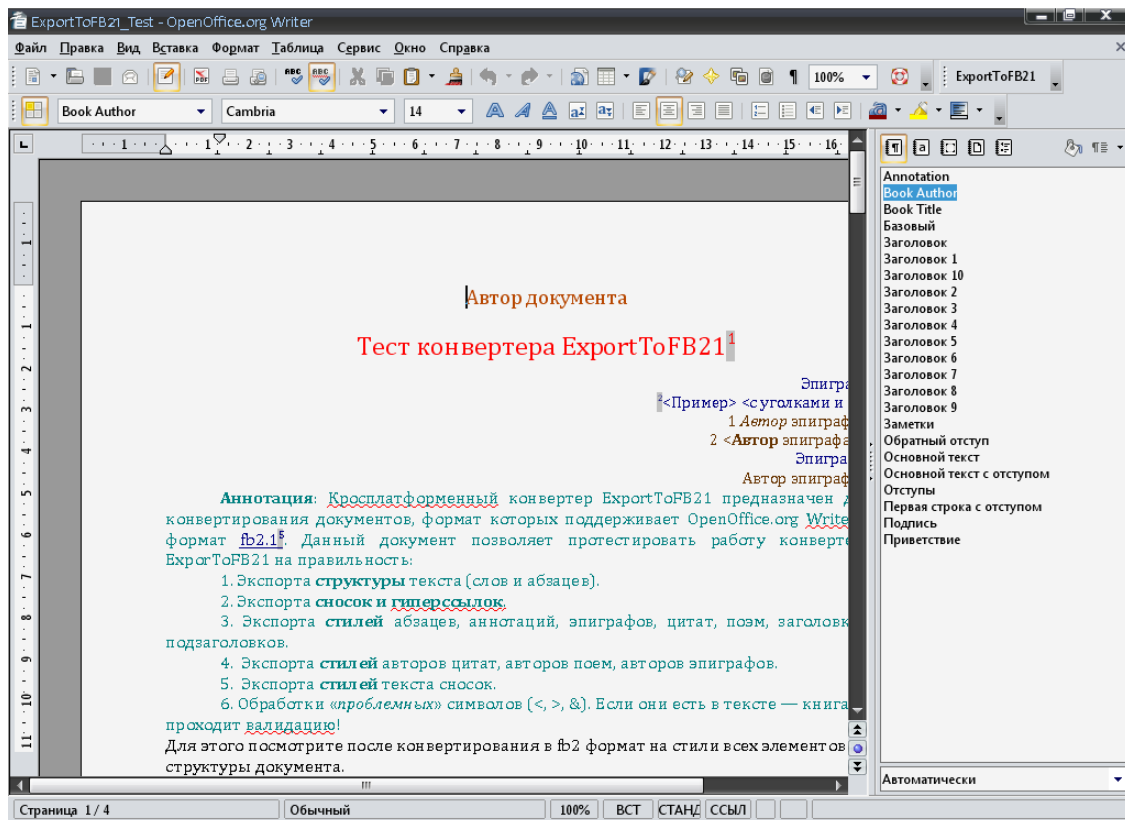
В комбобоксе «Стиль» главного окна редактора выбираем самый нижний элемент «Дополнительно».



Или - в меню Формат выберите «Стили». А можно просто нажать F11.



В результате:

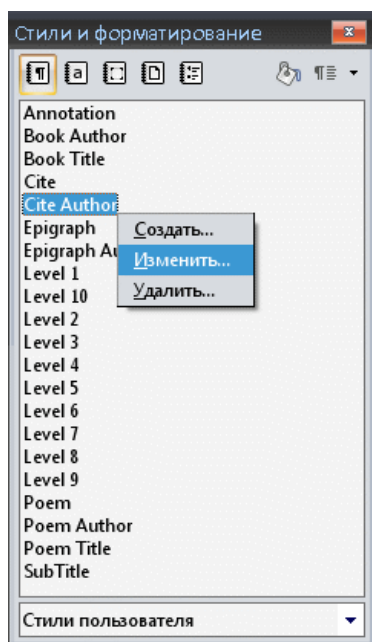


На картинке окно «Стили и форматирование» уже пристыковано к главному окну редактора.

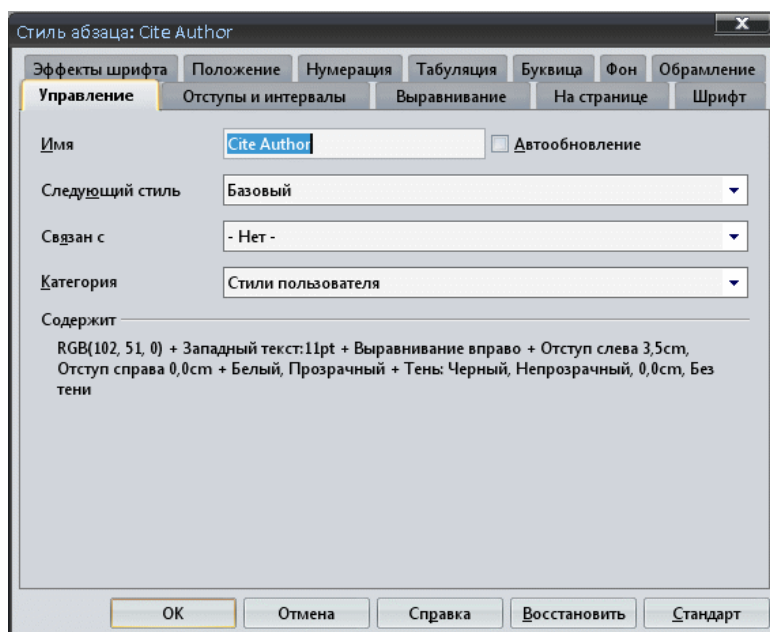
В этом окне выбираем внизу «Стили пользователя». Это как раз те стили, которые нам необходимы для разметки структуры документа.

3.2. Редактирование стилей

Чтобы изменить шрифт, цвет и другие атрибуты выбранного стиля, нажмите на нем правую клавишу мышки, и выберите в контекстном меню «Изменить...»



В результате появится диалог «Стиль абзаца», где можно отредактировать стиль. Категорию лучше оставить как есть – «Стили пользователя» - удобно работать со стилями, когда они в одной категории (см. [п. 3.4.](#)).



Главное: *Нельзя менять название стиля!*

3.3. Соответствие стилей с тегами формата fb2.1

3.3.1. Inline стили, или стили слов и фраз в пределах одного абзаца

1. Полужирный - `` - **Bold**
2. Курсив - `<emphasis>` - *Italic*
3. Зачеркивание - `<striketrough>` - ~~Зачеркивание~~ (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Зачеркивание")
4. Код программ - `<code>` - `Тень` или `Контур` (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Тень" или "Курсив")
5. Верхний индекс - `<sup>` - "Верхний индекс"
6. Нижний индекс - `<sub>` - "Нижний индекс"

3.3.2. Стили, применимые для всего абзаца

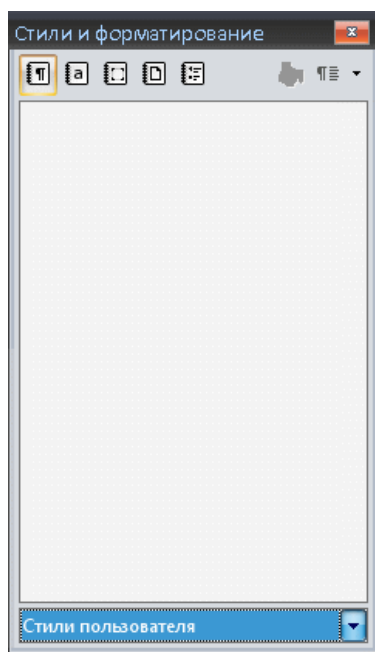
1. Автор книги - `Book Author`
2. Название книги - `Book Title`
3. Аннотация - `<annotation>` - `Annotation`
4. Стих - `<poem>` - `Poem`
Заголовок стиха - `<poem><stanza><title>...</title></stanza></poem>` - `Poem Title`
Автор стиха - `<text-author>` - `Poem Author`
5. Эпиграф - `<epigraph>` - `Epigraph`
Автор эпиграфа - `<text-author>` - `Epigraph Author`
6. Цитата - `<cite>` - `Cite`
Автор цитаты - `<text-author>` - `Cite Author`
7. Подзаголовок (* * *) - `<subtitle>` - `SubTitle`
8. Заголовки (Уровни) - `<section>` - `Level 1 ... Level 10`

3.4. Использование стилей в документе

Книга готовится к экспорту в формат fb2.1 двумя способами: просто открывается готовый документ формата, поддерживаемого OOo Writer или печатается «с чистого листа» - создается новый документ. В любом случае - надо загрузить уже установленные стили, с которыми вы будете работать. Делается это так:

3.4.1. Загрузка стилей

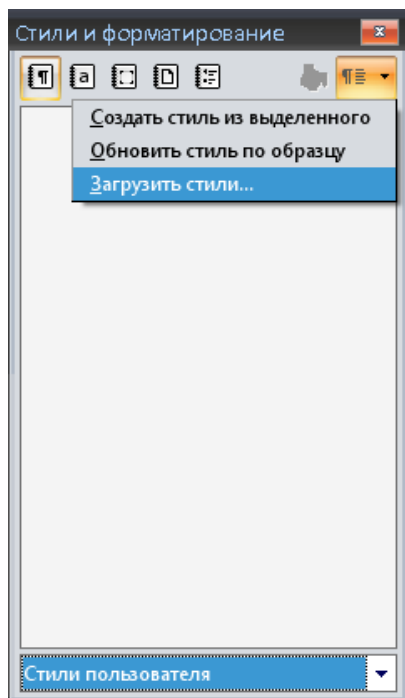
1. Открываем диалог «Стили и форматирование», если он еще не открыт. В комбобоксе «Стиль» выбираем самый нижний элемент «Дополнительно». Или - в меню Формат выберите «Стили», или просто нажмите F11 (картинки см. [п. 3.1.](#)).
2. В открывшемся диалоге выбираем «Стили пользователя»



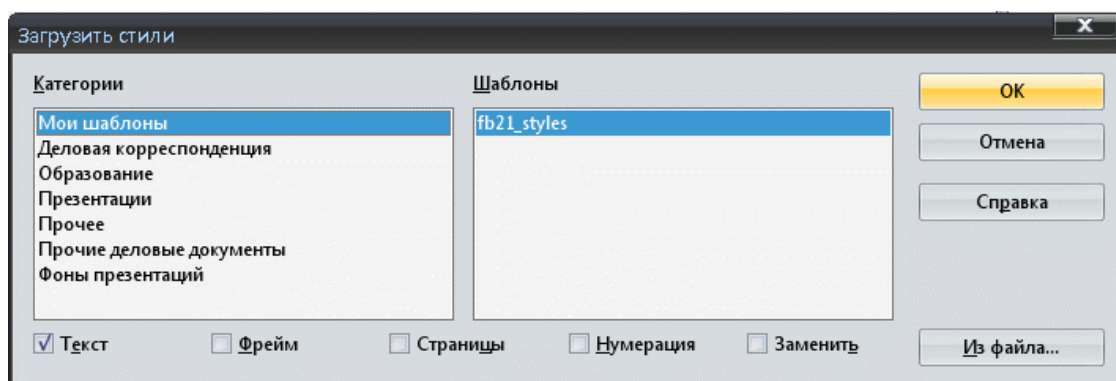
Нажать на самую правую кнопку



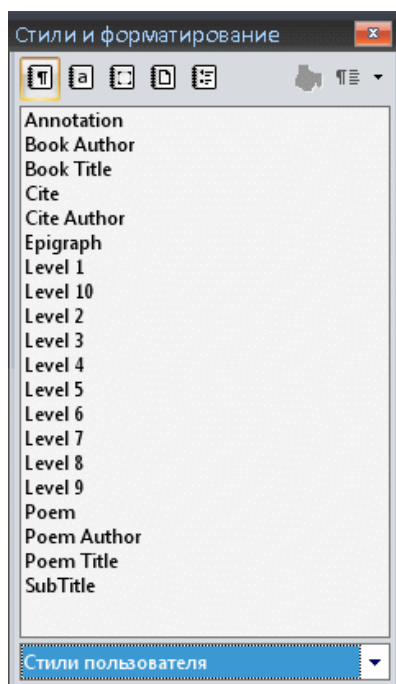
и выбрать "Загрузить стили..."



В открывшемся диалоге для Категории "Мои шаблоны" выбираем **fb21_styles.ott**



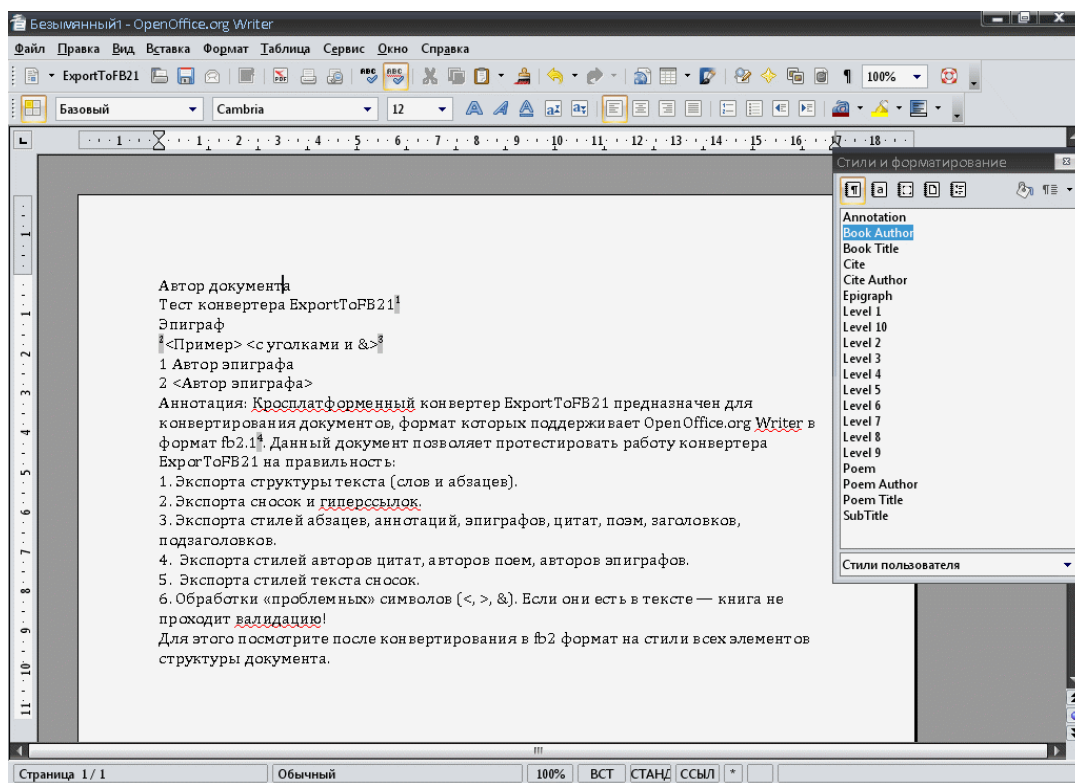
В результате наши стили загружены



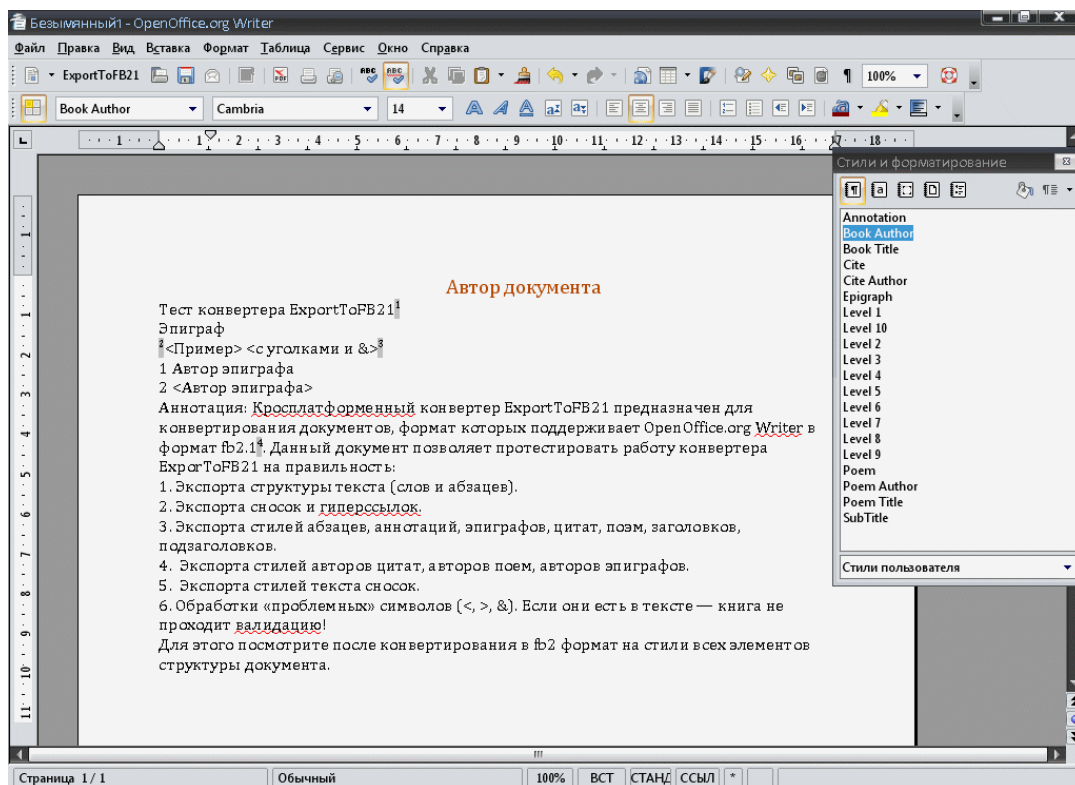
3.4.2. Работа со стилями в документе

Теперь ставим курсор на нужный абзац и помечаем его нужным стилем, кликнув по его названию дважды мышкой:

Было



Стало



И так далее, пока весь документ не будет отформатирован нужными стилями и мы не получим структуру документа, которая будет конвертироваться в формат fb2.1.

4. Особенности работы с текстом

4.1. Работа с таблицами

4.1.1. Виды таблиц

Существует два вида таблиц – простые и сложные. Простые – прямоугольная матрица Row x Col. Сложные – в ячейке могут быть вложены другие строки или таблицы.

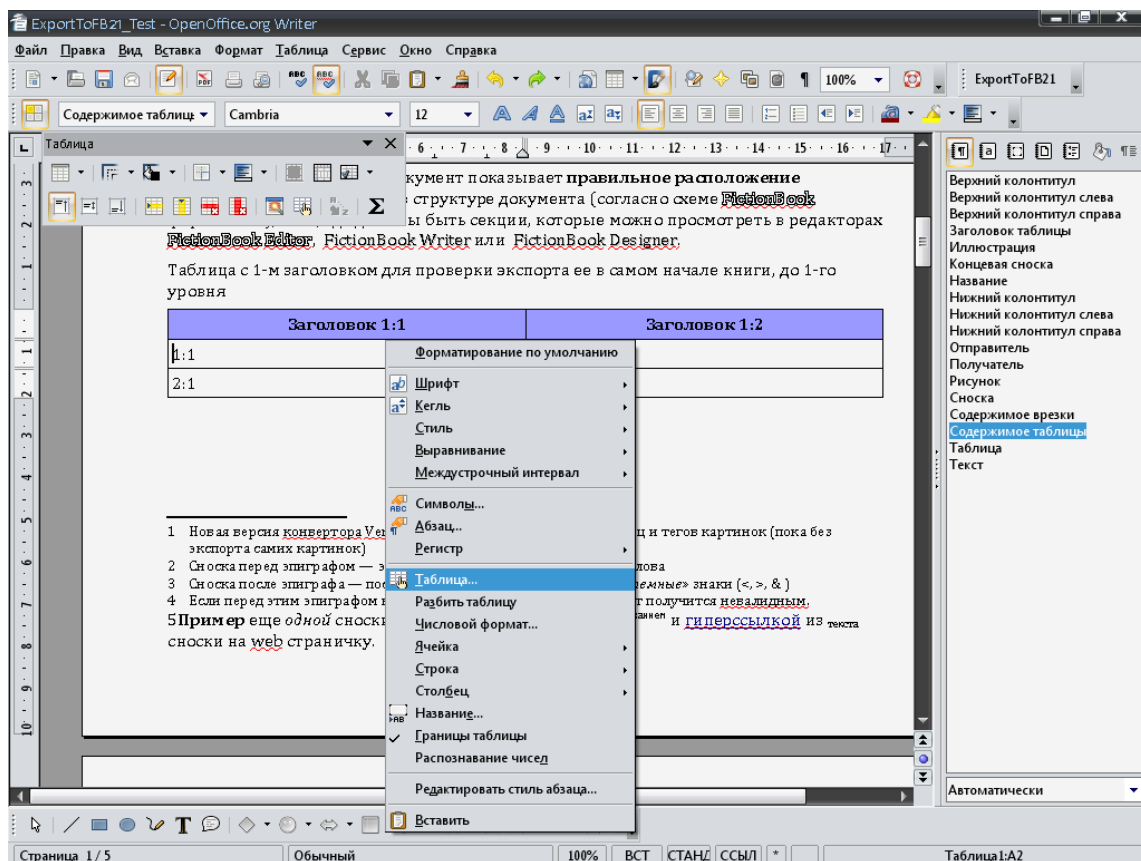
Схема FictionBook поддерживает только простые таблицы. Поэтому конвертор обрабатывает только их. Если в тексте встречается сложная таблица, то она игнорируется.

4.1.2. Название таблицы

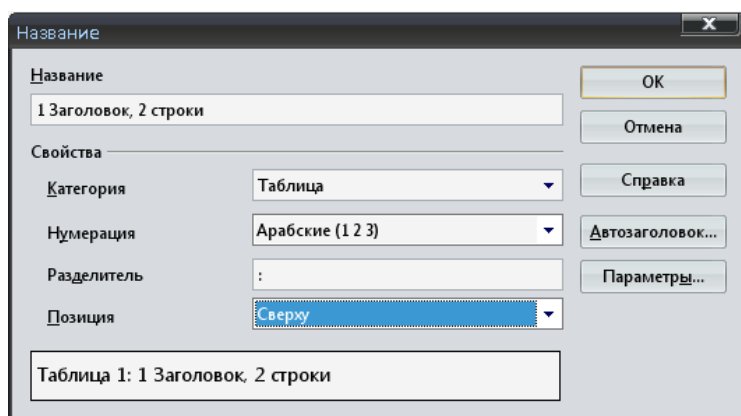
При желании можно дать таблице название (сверху или снизу таблицы). Это название также будет экспортироваться, с учетом форматирования.

Делается это так.

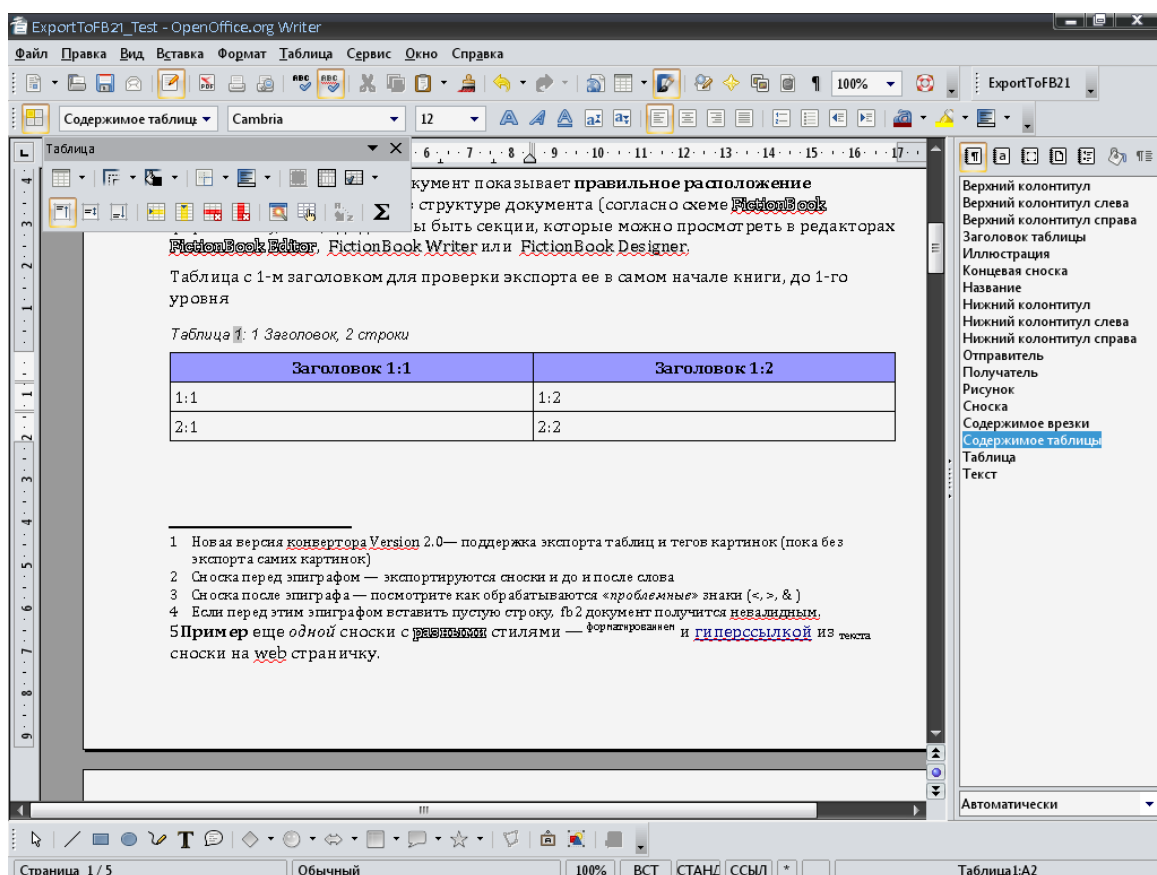
На нужной таблице кликните правой клавишей мышки и выберите «Название».



В появившемся диалоге «Название» введите название таблицы и выберите позицию этого названия – сверху или снизу:



В результате получим следующее:

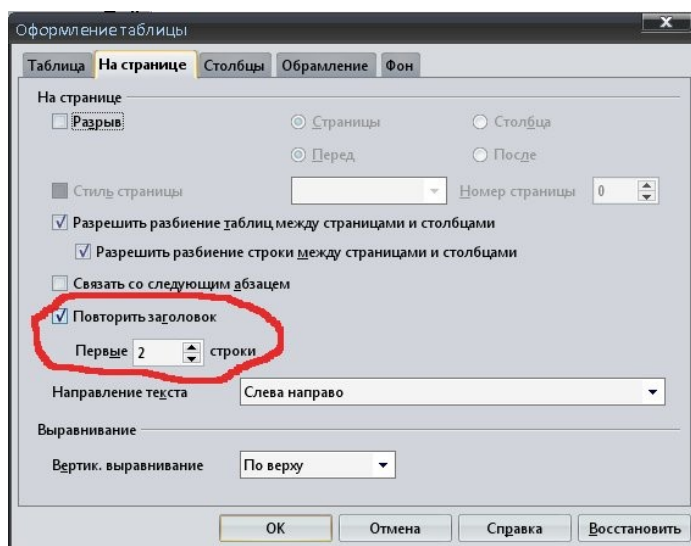


В названии имеется текстовое поле с сквозной нумерацией таблицы — оно тоже экспортируется.

4.1.3. Задание числа заголовков

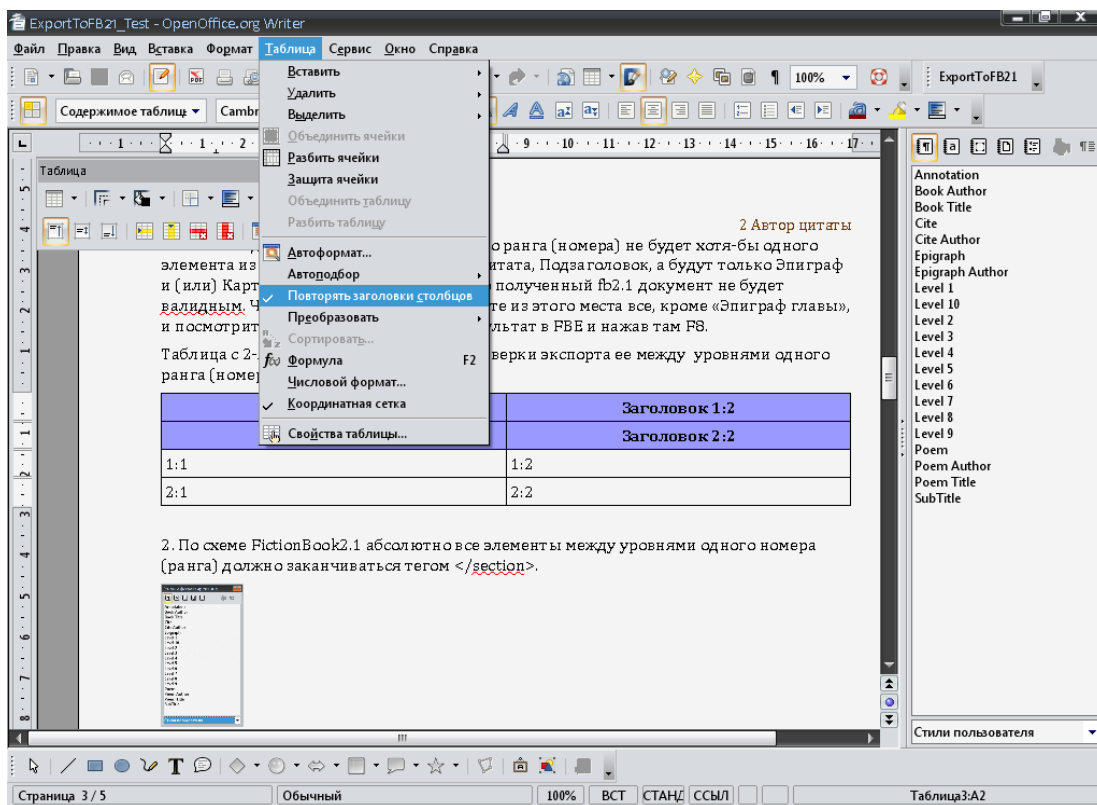
Если наша таблица содержит заголовок, или несколько, то надо задать число строк, из которых он будет состоять, и включить опцию, сигнализирующую, что заголовок есть.

Для этого вызываем диалог «Оформление таблицы» через меню «Таблицы», нажав пункт меню «Свойства таблицы...». Либо через контекстное меню по клику правой клавиши мышки — пункт «Таблица»



На рисунке показано, что таблица имеет Заголовок из 2-х верхних строк.

Галочка “Повторять заголовок” стоять должна обязательно! Это можно сделать и по-другому, встав на таблицу, и выбрав пункт меню «Повторять заголовки столбцов» в меню «Таблица». Только в этом случае можно будет определить, что в таблице действительно есть заголовки.



4.1.4. Важен порядок вставки таблиц в документ

В связи с особенностью алгоритма конвертора очень важно, в какой последовательности вставляются таблицы в документ. Если их вставлять не по-порядку, по

ходу текста, то после конвертации таблицы могут быть *перепутаны местами!*

4.1.5. Выравнивание строк ячеек таблицы

Можно выравнивать строки в каждой ячейке **по ширине**, как вам угодно – конвертор автоматом выравнивает строки для каждой ячейки в результирующем файле. По высоте – пока задается в «Настройках» (пока), и применяется ко всем ячейкам одинаково.

Конвертор экспортирует и форматирование ячеек, но только для текста всей ячейки одновременно, считывая с курсора таблицы данные (по 1-му символу). Поэтому, если первый символ или слово в ячейке было *курсив*, то в fb2 файле весь текст этой ячейки будет выделен *курсивом*.

С другой стороны, все абзацы в ячейке собираются в одну строку, что согласуется со схемой FictionBook2.1.

4.1.6. «Пустые» ячейки таблицы

Иногда требуется, чтобы в таблице были ячейки с «пустым» содержанием. После экспорта в fb2 документ на месте этой «пустой» ячейки будет `<td></td>`, что вызовет ошибку валидации. Если нужны пустые ячейки в таблице – поставьте там просто точку «.».

4.2. Работа с врезками

4.2.1. Виды текстовых врезок

Всего существует 4 вида текстовых врезок:

1. С привязкой к странице.
2. С привязкой к абзацу.
3. С привязкой к символу.
4. С привязкой как символ.

Конвертор «не отлавливает» 1-й вид врезок – с привязкой к странице. Если такая врезка есть в документе, то она будет проигнорирована.

4-й вид врезок вряд ли имеет смысл применять в тексте. Лучше использовать 2-й или 3-й вид врезок.

4.2.2. Возможности

Текстовые врезки экспортируются, пока, без экспорта форматирования, просто как текст.

Возможны три варианта экспорта врезок:

1. Просто, как параграфы текста (в этом случае в fb2 файле их сложно будет отличить от основного текста).
2. Как цитату (весь текст врезки будет представлен в fb2 файле, как цитата `<cite>`).
3. Как таблица (весь текст врезки будет представлен в fb2 файле, как таблица).

Вид экспорта врезок можно выбрать в диалоге «Настройки» (см. [п.6.2.](#)). По умолчанию – «Как таблица».

4.3. Работа с картинками

4.3.1. Допустимые виды картинок

Всего существует 4 вида картинок:

1. С привязкой к странице.
2. С привязкой к абзацу.
3. С привязкой к символу.
4. С привязкой как символ.

При этом картинка каждого вида привязки может быть помещена в текстовую врезку с названием картинки (если на картинке нажать правую клавишу мышки и из контекстного

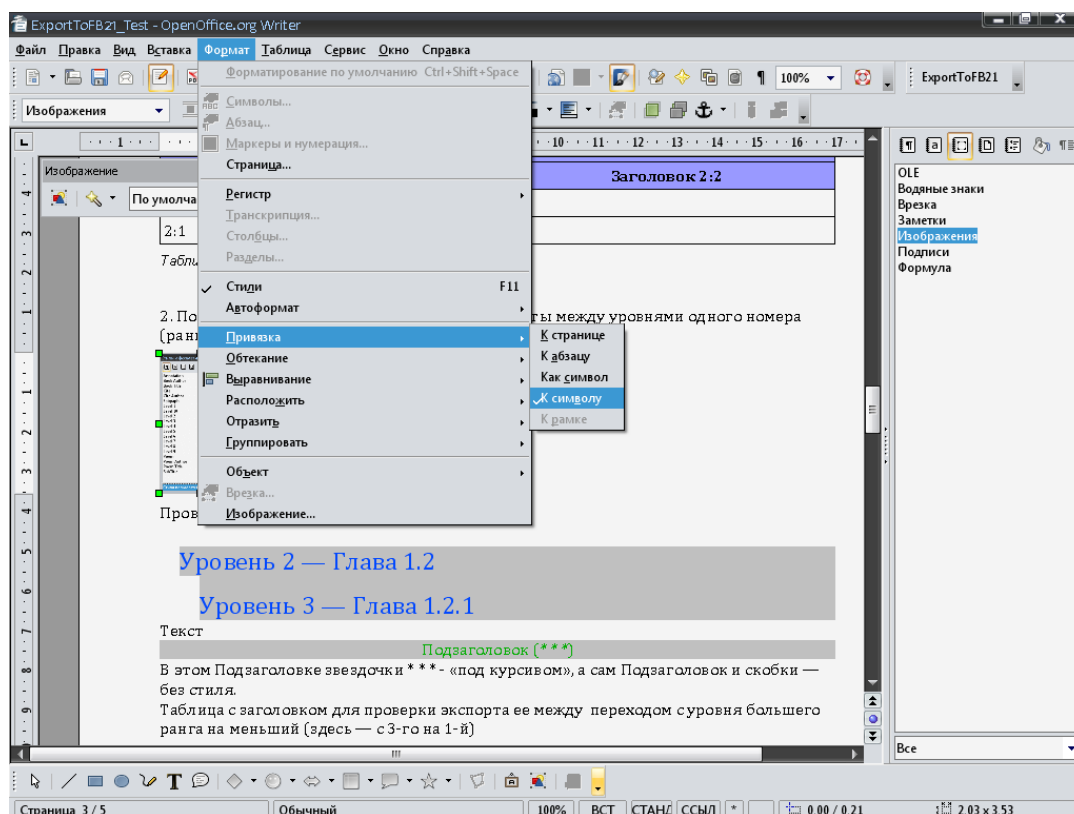
меню выбрать «Название», и диалоге ввести название картинки).

1-й вариант – картинка с привязкой к странице «не отлавливается» конвертором, и картинка будет проигнорирована. Картинки с названиями во врезках тоже пока «не отлавливаются», и экспортировано будет только их название из врезок в виде текста. Поэтому не используйте эти два вида картинок! Если надо задать строку с названием картинки, то просто напишите текстовую строку. С другой стороны, в атрибуты “title” тега <image> будет заноситься (при включенной опции в настройках) имя картинки. Задать его можно так.

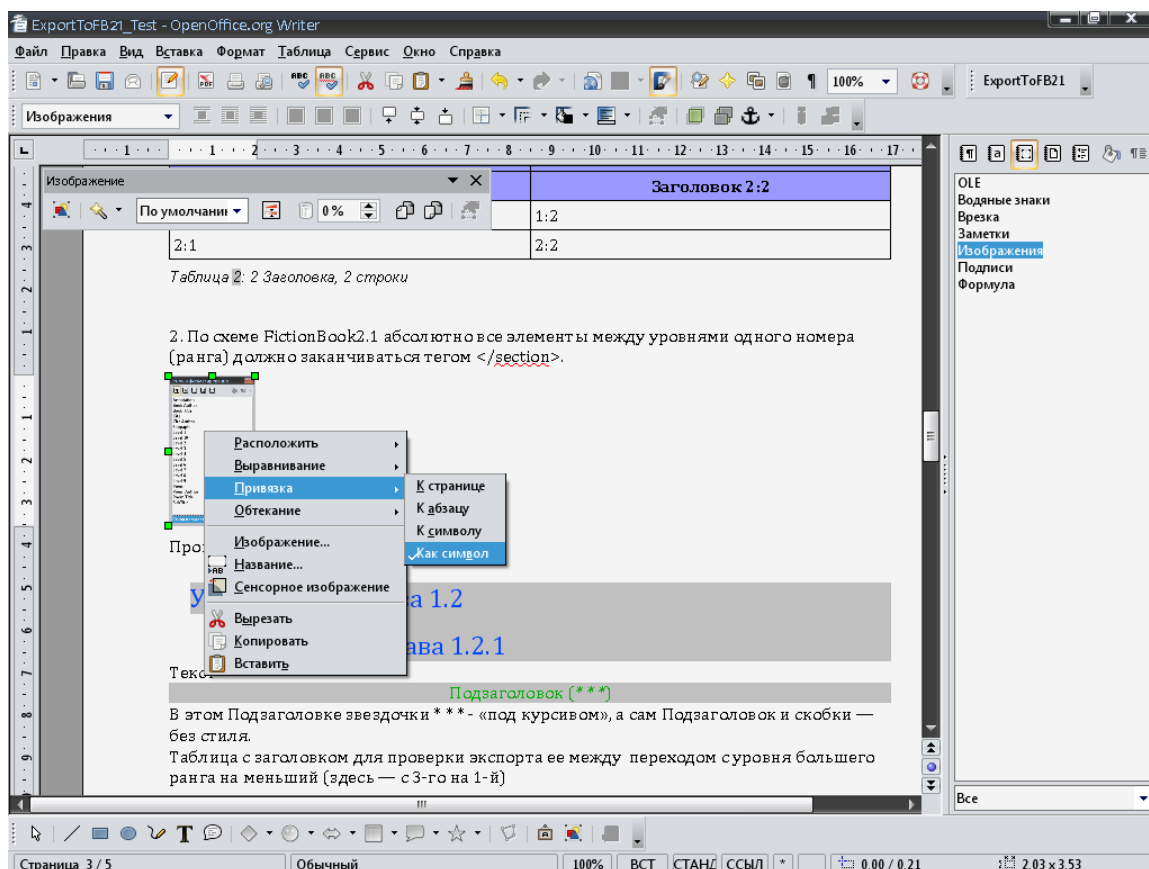
4.3.2. Привязка картинок к тексту

Нужную привязку картинки можно осуществить несколькими способами.

4.3.2.1. Через меню «Формат» -> «Привязка»

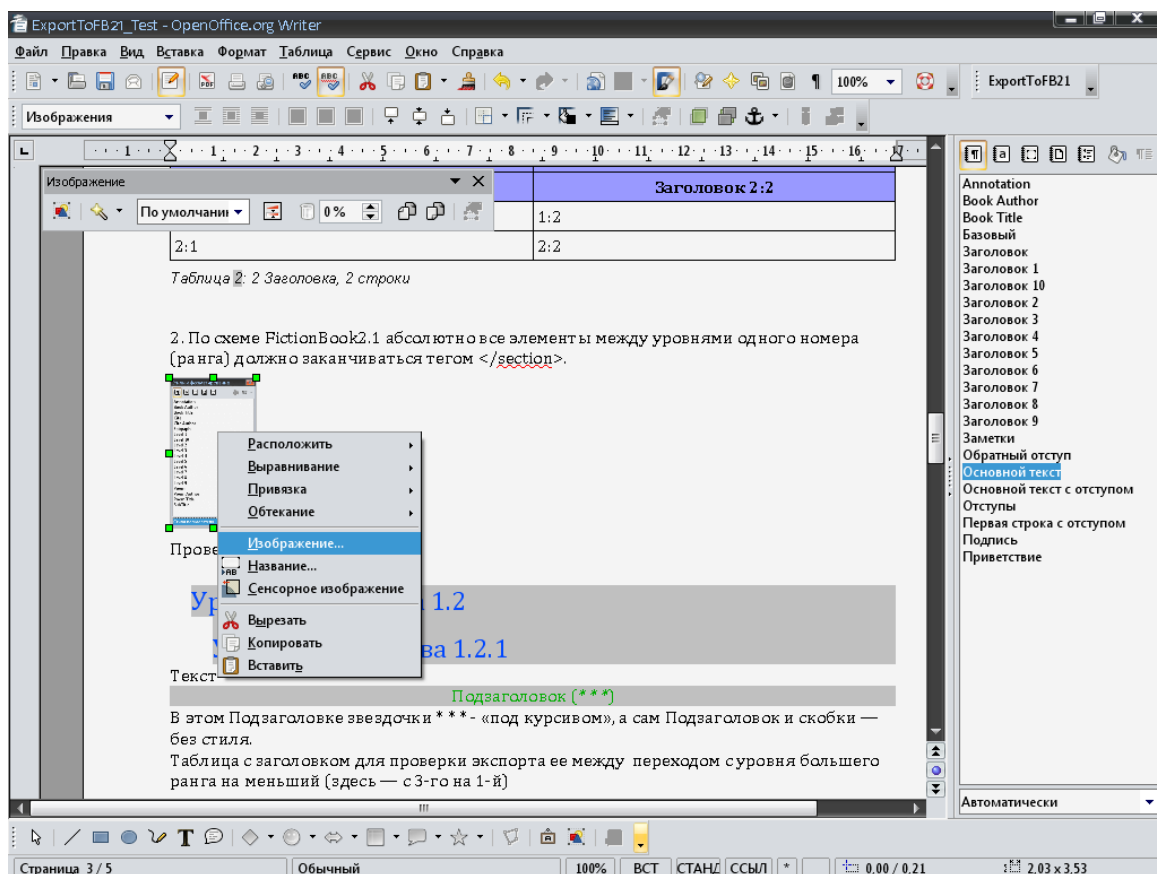


4.3.2.2. Через контекстное меню «Привязка»

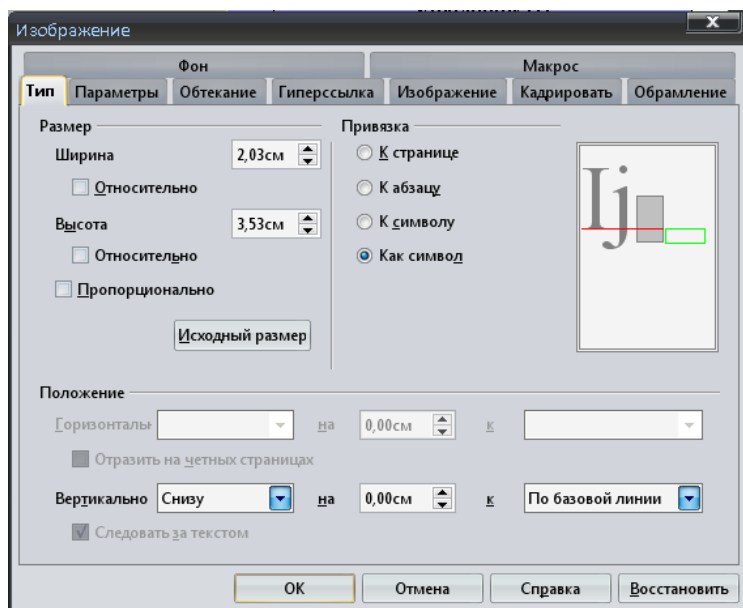


4.3.2.3. Через диалог «Изображение»

Для этого выделите нужную картинку, нажмите на ней правую клавишу мышки и выберите «Изображение...»

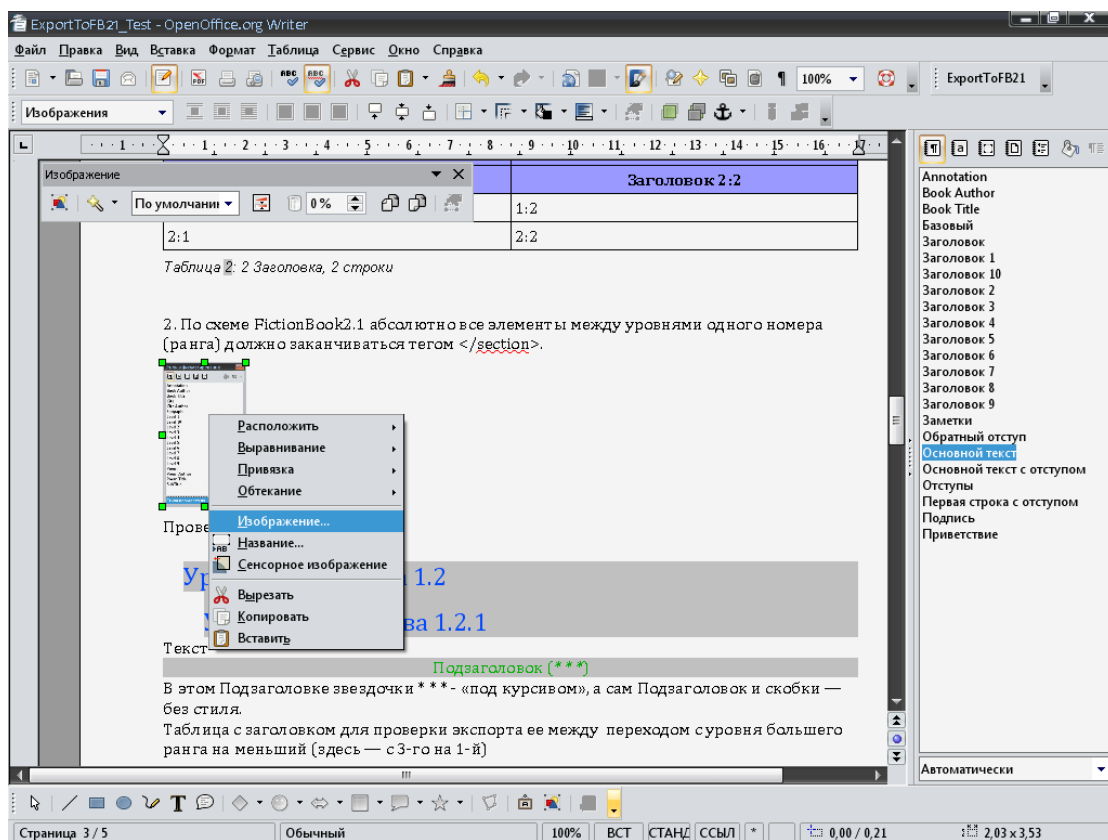


В появившемся диалоге на вкладке «Тип» выберите Привязку картинки:

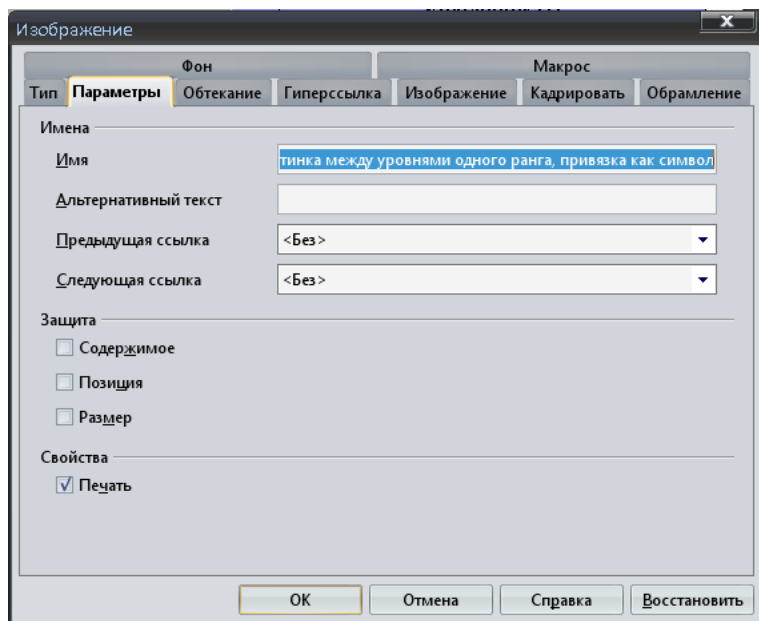


4.3.3. Задание имени картинке

Выделите нужную картинку, нажмите на ней правую клавишу мышки и выберите «Изображение...»



В появившемся диалоге на вкладке «Параметры» в поле «Имя» введите имя картинки:



4.3.4. Картинки в качестве символов (Буквицы)

Картинка с привязкой «как символ» используется или как Буквица, или для отображения небольшой по периметру картинки в границах параграфа между словами.

Привязать картинку, как символ можно, используя несколько способов (см. [п.4.3.2](#)).

Конвертор экспортирует такие картинки в соответствии со схемой FictionBook2.1,

«вживляя» их в тег <p>.

Например, есть такой параграф:

Нажмите на кнопку , потом на .

После конвертации получим следующее:

<p>Нажмите на кнопку <image l:href="#img_1"/>, потом на <image l:href="#img_2"/>.</p>

При этом имя таких картинок просто игнорируется и не заносится в атрибут "title".

Не все читалки отображают такие inline картинки (только Haalii Reader, Cool Reader 2 и All Reader 2.3), тем не менее, конвертор позволяет экспортировать такую структуру.

5. Корректная структура документа (тонкости)

Мы не будем описывать здесь всю схему FictionBook2.1, а только те моменты, которые очень важны для **правильной разметки структуры** конвертируемого документа. Лишь в случае правильной разметки структуры экспорт произойдет без единой ошибки и править в полученном fb2 документе вообще ничего не надо будет. за исключение моментов, описанных в п. 5.

Итак, что такое "правильная" разметка структуры? Когда мы придаем отдельным абзацам определенные стили (поэма, цитата, заголовок (уровень), аннотация и т.д.) мы формируем структуру документа. А эта структура должна соответствовать схеме FictionBook2.1. Иначе документ сконвертируется, но валидацию не пройдет. Например: в книгах часто идет *аннотация, а потом - эпиграф*. Но, согласно схеме FictionBook2.1 должно быть наоборот - *сначала эпиграф, а потом - аннотация*. Если вы сделаете как в книге, а не по схеме - документ будет *невалидный*, и online-библиотеки, если не ошибаюсь, его *не примут*.

А теперь чуть подробнее.

Структуру документа, условно, можно разбить на два блока: 1. От названия книги и до самого первого заголовка (части, раздела) - мы называем это "Уровнем". 2. От этого самого первого заголовка и до конца документа. В этих двух блоках обрамление текста тэгами <section> несколько отличается.

5.1. Блок от начала книги до первого заголовка (уровня)

В 1-м блоке *самым первым* элементов должен быть Эпиграф, если он нужен. За ним могут идти еще несколько Эпиграфов, если они нужны, потом - Картинка, если она нужна, потом - Аннотация, если она нужна. При этом **между** всеми этими элементами не должно быть **ни одной пустой строки**! Иначе документ получится невалидным.

Теперь могут идти следующие элементы структуры в **любом порядке**: Просто абзац текста, пустая строка, Поэма, Цитата, Подзаголовок (SubTitle), Таблица, Картинка. Схематично все это можно представить так (в квадратные скобки [] взяты *необязательные* элементы, в круглые () - названия стилей для создания структуры документов, **жирным** выделены элементы, которых может быть *любое количество*), **синим** цветом - элементы, которые могут идти *в любом порядке*:

[**Эпиграф**] - (Epigraph)

[Картинка]

[Аннотация] - (Annotation)

[**Абзац**] - (Любой стиль, кроме predetermined для структуры)

[Пустая строка] - (Любой стиль, кроме предопределенных для структуры)
[Картинка]
[Таблица]
[Цитата] - (Cite)
[Подзаголовок] - (Subtitle)
[Стих] - (Poem)

5.2. Блок от первого заголовка (уровня) до конца документа

Во 2-м блоке дело обстоит немного по-другому.

2-й блок начинается с самого первого Уровня - Части книги, Раздела книги или просто заголовка. Уровней заголовков может быть 10 (стиль Level 1 ... Level 10) (хотя программно можно очень легко их увеличить хоть до бесконечности – особенность алгоритма). В книге может быть такая структура:

Уровень 1 (Раздел I)
Уровень 2 (Глава 1)
Уровень 2 (Глава 2)
Уровень 3 (Глава 3.1) и т.д.

Наши элементы (абзацы, стихи и т.д.) могут быть как между Уровнями разного ранга - 1-й случай, например:

Уровень 2
Цитата
Уровень 3

так и между Уровнями одного ранга (номера) - 2-й случай, например:

Уровень 2
Цитата
Уровень 2

5.2.1. Смежные Уровни разного ранга (номера)

Для 1-го случая с Уровнями *разного* ранга схема немного другая, чем для блока от начала книги до первого заголовка (уровня). Отличие состоит лишь в том, что между уровнями *может вообще не быть никаких элементов* (стихов, абзацев...). Т.е. схема расположения элементов структуры такая же, как и для 1-го блока:

[Эпиграф] - (Epigraph)
[Картинка]
[Аннотация] - (Annotation)
[Абзац] - (Любой стиль, кроме предопределенных для структуры)
[Пустая строка] - (Любой стиль, кроме предопределенных для структуры)
[Картинка]
[Таблица]
[Цитата] - (Cite)

[Подзаголовок] - (Subtitle)
[Стих] - (Poem)

5.2.2. Смежные Уровни одного ранга (номера)

Для 2-го случая между Уровнями *одного* ранга должен быть *хоть один какой-нибудь элемент*, кроме необязательных Эпиграф и (или) Картинка и (или) Аннотация! Схема расположения элементов структуры такая (в двойные скобки { [] } взяты элементы, которые могут и не быть, но *хоть один из них обязательно* должен присутствовать):

[Эпиграф] - (Epigraph)
[Картинка]
[Аннотация] - (Annotation)
{ [Абзац] } - (Любой стиль, кроме предопределенных для структуры)
{ [Пустая строка] } - (Любой стиль, кроме предопределенных для структуры)
{ [Картинка] }
{ [Таблица] }
{ [Цитата] } - (Cite)
{ [Подзаголовок] } - (Subtitle)
{ [Стих] } - (Poem)

5.2.3. Ошибки при разметке структуры смежных заголовков (уровней)

С другой стороны, если ранги Уровней перепутаны при разбивке структуры, то полученный fb2 документ не будет валидным! Например, нужна такая структура книги:

Уровень 1 (Раздел I)
 Уровень 2 (Глава 1)
Уровень 1 (Раздел II)
 Уровень 2 (Глава 1) и т.д.
 Уровень 3 (Глава 1.1)
Уровень 1 (Раздел III) и т.д.

А мы ошибочно перепутали стили Level 2 и Level 3, и получили такую структуру:

Уровень 1 (Раздел I)
 Уровень 3 (Глава 1)
Уровень 1 (Раздел II)
 Уровень 2 (Глава 1)
 Уровень 3 (Глава 1.1) и т.д.
Уровень 1 (Раздел III) и т.д.

то полученный fb2 документ не будет валидным! Номер следующего (смежного) Уровня по возрастанию не может отличаться от предыдущего более, чем на 1. По убыванию - как угодно. Поэтому, при разметке структуры будьте очень внимательны!

Таким образом, если документ отформатирован стилями в соответствии с этими правилами, диктуемыми схемой fb2.1 будет сконвертирован без единой ошибки и пройдет валидацию.

Для проверки всего этого, а также других возможностей данного конвертора создан файл **ExportToFB21_Test.odt**. Сконвертируйте его в fb2 документ.

6. Конвертация документа в формат fb2.1

6.1. Описание диалога конвертации

Данные о книге

Книга (Title Info)

Жанр: 100 0 Язык: ru Оригинал:

Авторы

Имя: Отчество: Фамилия: 0

Ник: Email: Сайт:

Название: Тест конвертера ExportToFB21

Keywords:

Date text: Date value:

Переводчики

Имя: Отчество: Фамилия: 0

Ник: Email: Сайт:

Серия: Номер: 0

FB2 Документ (Document Info)

Версия: 1.0 ID: 000-ExportToFB21-2007122611041

Авторы

Имя: Отчество: Фамилия: 0

Ник: Email: Сайт:

Programms used: ExportToFB21

Date Text: 26.12.2007 Date Value: 2007-12-26

Source OCR:

Source URLs: <<"/>

Бумажная Книга (Publish Info)

Название: Издатель: Год:

Серия: Номер: 0 Город: ISBN:

Настройки

Сохранить как...

C:\1\My Converter\Export

Кодировка: utf-8

Custom Info

Тип:

Value:

0

Другое

Аннотация к Книге:

История:

Диалог может поместиться на экране с разрешением 800 x 600. Поэтому элементы в правой части диалога имеют несколько уменьшенные размеры.

В этом диалоге задаются данные о книге, издательстве, об авторе fb2 документа (т.е. о себе), история этого fb2 документа и аннотация на книгу, если нужно.

Некоторые поля отмеченные **красным цветом**. Это поля, где данные должны быть обязательно! Этого требует схема FictionBook2.1. Остальные поля заполнять желательно, но не обязательно. В случае их заполнения ваша книга будет содержать много дополнительной ценной информации.

Обязательное поле ID документа заполняется автоматически. При желании данные в нем его можно исправить, как вам угодно.

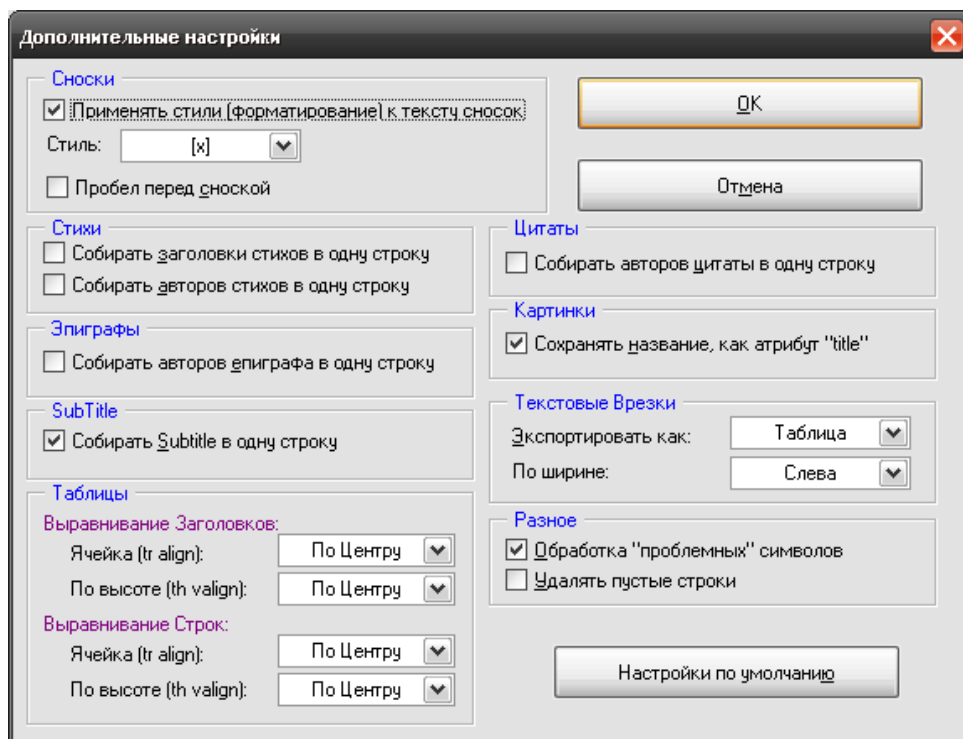
Один момент требует особого разъяснения. Около некоторых полей для ввода данных есть кнопка «Добавить» или «+» и счетчик числа добавленных элементов. Если в эти поля вы ввели данные, но не нажимали кнопку «Добавить» или «+», то в сконвертированном файле fb2 эти данные будут в единственном экземпляре. На счетчике будет 0 (ничего страшного). Когда этих данных должно быть несколько (несколько авторов, несколько переводчиков, несколько custom info, книга входит в несколько серий, несколько человек делали конвертируемый документ), то вводите данные в соответствующие поля, нажмите на

кнопку «Добавить» или «+». Значение счетчика увеличится на 1, поля очистятся. Можно вводить новые значения, если это требуется.

Если вам надо загрузить свой профиль (данные об Авторе Документа и Source OSR), то нажмите кнопку «<<». Если надо сохранить свой профиль – то нажмите кнопку «>>».

По умолчанию fb2 создается в той же папке, где расположен исходный документ. При желании месторасположение результирующего файла можно изменить, нажав кнопку «. . .» - Сохранить как – или введя вручную нужный путь в поле.

6.2. Дополнительные настройки



Диалог «Дополнительные настройки» обеспечивает ряд «удобств» при работе с конвертором.

1. Можно выбрать - вставлять или нет пробел между словом и сноской (многим удобнее читать, когда после слова и перед номером сноски стоит пробел).

2. Можно выбрать стиль сноски: [x], [x], {x}, { x }, (x) или (x). Лучше не вводить стиль сноски руками - конвертор обрабатывает только эти стили. В случае любого другого стиля сносок по умолчанию будет использоваться [x].

3. Опция «Применять стили (форматирование) к тексту сносок» позволяет конвертировать курсивы, болды гиперссылки и т.д. в тексте сноски!

Включение этой опции увеличивает время экспортирования документа.

4. Для стихов предусмотрена возможность собирать несколько заголовков стиха в одну строку (по умолчанию опция отключена)

5. Для стихов, эпиграфов и цитат введена возможность собирать нескольких авторов в одну строку через запятую (по умолчанию опция выбрана) или оставить все, как есть - авторы размещаются на разных строках (по умолчанию опция отключена).

6. Та же возможность и для подзаголовков (subtitle).

7. «Обработка "проблемных" символов» позволяет корректно экспортировать символы

&, < и >, как & , < и >

Включение этой опции также увеличивает время экспортирования документа.

8. Опция «Удаление пустых строк» говорит сама за себя - удаляются все пустые строки из документа в результирующем fb2 файле.

9. Для Картинок предусмотрена опция «Сохранять название картинки, как атрибут “title”». Для inline картинок – с привязкой «как символ» эта опция игнорируется, чтобы не было нарушено соответствие схеме FictionBook2.1.

10. Для Таблиц предусмотрено выравнивание и Заголовков, и Строк в Ячейках по высоте (по ширине выравнивание считывается автоматически). Используйте только “зашитые” в комбобоксы виды выравнивания!

В связи с тем, что текущий алгоритм обработки строк таблиц не позволяет считывать выравнивание по высоте для каждой строки, то выбранные значения будут применимы ко всем таблицам ко всем Заголовкам, Ячейкам и Строкам соответственно (хотя схема FictionBook2.1) позволяет применять свое выравнивание по высоте для каждой строки каждой ячейки отдельно. По ширине – все нормально – автоматом выравнивается каждая ячейка.

11. Для Врезок можно выбрать один из видов их экспорта – как просто текст, как цитату или как таблицу (стоит по умолчанию). Если выбран режим экспорта как таблицу, то появляется возможность задать выравнивание строк таблицы.

7. Когда требуется редактирование результата в fb2-редакторах

1. Когда вы хотите внести в полученный fb2 документ тэги и возможности, пока отсутствующие в конверторе (см. [п.1.4.](#))

2. Когда вы хотите придать отдельным словам и (или) фразам *множественный* стиль.

3. Когда вы конвертировали документ с выключенной опцией «Обработка «проблемных» символов»: &, > и <. Такой невалидный документ, правда, не откроется и в FBE, и в FBW (хотя многие читалки позволяют читать такие книги, просто вырезая эти символы, правда, библиотеки вряд ли примут такие книги!). Править придется только в текстовых редакторах, поддерживающих кодировку, в которую вы конвертировали свой документ.

4. Когда вы отформатировали стилями элементы находятся в документе не в том порядке, в каком они должны идти, согласно схеме FictionBook2.1 Т.е. разбивка структуры документа произведена некорректно (см. [п.5.](#)). Документ будет экспортирован, но невалидным.

5. Когда в документе есть таблица и в ней есть ячейки с «пустым» содержанием. После экспорта в fb2 документе на месте этой «пустой» ячейки будет <td></td>, что вызовет ошибку валидации. Совет – если нужны пустые ячейки в таблице – поставьте там просто точку «.».

6. Все предусмотреть невозможно. Поэтому в конверторе вполне могут быть какие-нибудь багги, которые пока не известны, и которые за долгие часы тестирования документов с разной структурой так и не проявили себя (или я просто не заметил).

8. Планы по развитию конвертора ExportToFB21 (TODO)

1. Сделать то, что не реализовано (см. [п.1.4.](#)), по мере возможности.

2. Оптимизация скорости парсинга.

3. Минивалидатор – выдает сообщение о соответствии/не соответствии помеченного стилями документа схеме FictionBook2.1.
4. Редактор структуры уровней документа – изменения номера уровня и т.д.
5. Просмотрщик элементов структуры (поэмы, цитаты...) в виде дерева и списка элементов, с возможностью перехода на нужный элемент в документе.
6. Обработку текста – многоточия, кавычки-елочки, дефисы...
7. Подогнать конвертор под новую схему FictionBook
8. Что-нибудь еще...