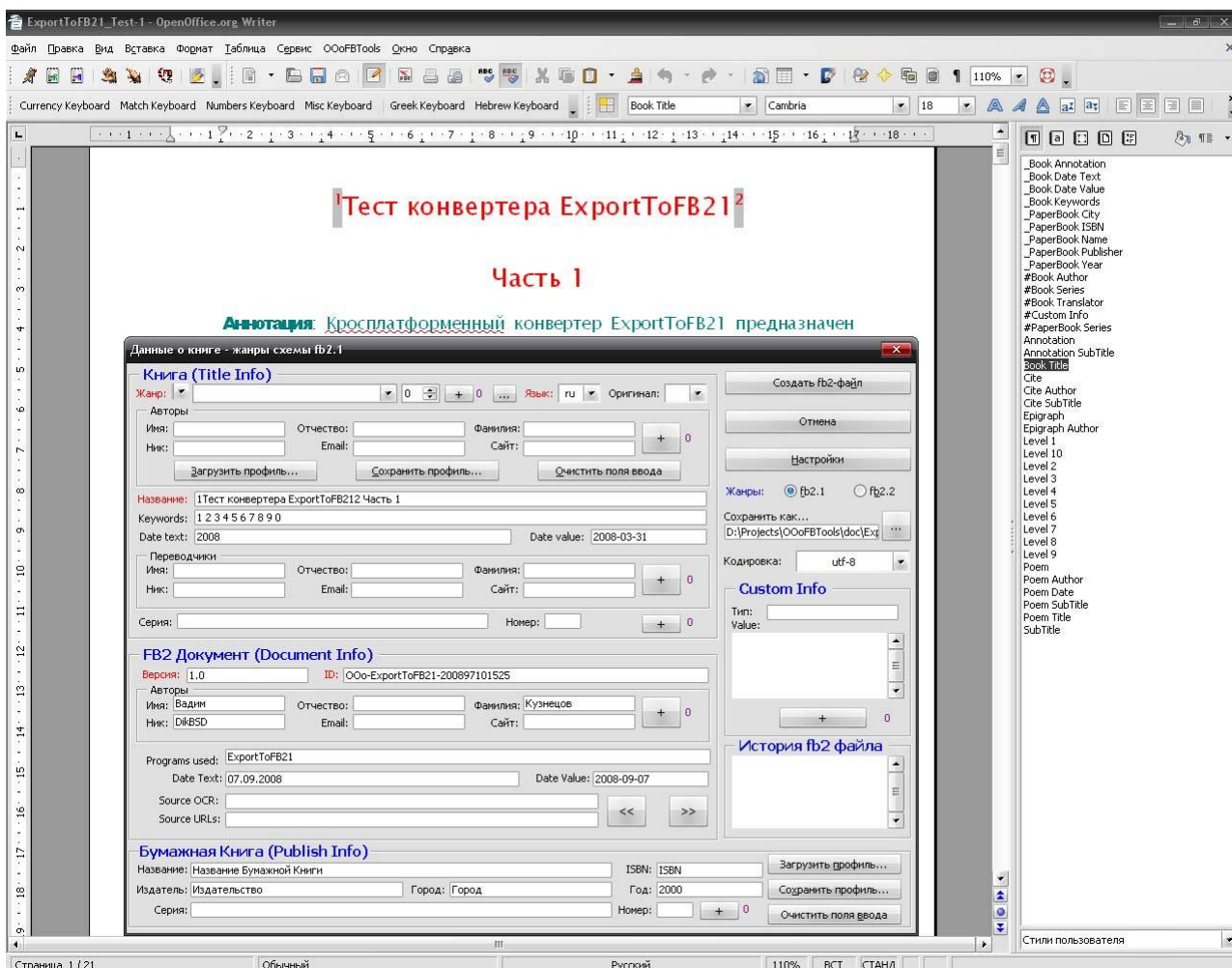


Описание работы пакета OOoFBTools

I Конвертор ExportToFB21

1. Описание



1.1. Назначение

Кроссплатформенный конвертер 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.

Конвертер 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.

Конвертер 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. Что реализовано

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

1.2.1. 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

1.2.2. 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, полужирные начертания, верхние и нижние индексы 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 **картинок и картинок-буквиц**.
9. 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.
10. Экспорт простых и сложных **Таблиц** с экспортом *заголовков, выравниванием* текста заголовков и ячеек, а так же *форматирования стилями* (полужирный, курсив, верхний/нижний индексы и т.д.), *гиперссылок и переходов* из строк таблицы в другие места файла. Работает экспорт *сносок* для строк ячеек таблиц. Экспорт *нумерованных списков-абзацев, маркированных списков*, а также списков с *маркерами-картинки* в ячейке. Сделан экспорт Картинок из ячеек Таблиц.
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!*

1.2.3. 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), (x) (опция в **Настройках**).

5. Возможность собирать авторов поэмы, цитат или эпитафия 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, как таблиц или как структуру (опция в **Настройках**). Экспорт всех стилей (и символьных и абзацевых) и нумерованных списков во Врезках. Экспорт Таблиц и Картинок из Текстовых Врезок. Распознается привязка Картинок только как «К Абзацу», «К символу» и «Как символ». Другие – игнорируются и Картинка не будет экспортирована! *Вложенные Врезки* во Врезки игнорируются!

9. Экспорт **Нумерованных** и **Маркированных** Списков (в том числе и с **Маркеро-Картинок**) из **Текста** документа и **Текстовых Врезок**.

10. Экспорт **текстовых полей**, вставленных из меню “Вставка” – > ”Поля”.

11. Экспорт **ссылок внутри документа** (опция в **Настройках**).

12. Сохранение профиля пользователя и его автоматическая загрузка для полей Автора fb2-документа, OCR и истории fb2-документа.

13. Сохранение и загрузка **профиля** для Автора Книги и его жанра.

14. Сохранение и загрузка **профиля** для Издательства.

15. Автоматическое заполнение полей формы конвертора «Аннотация на Книгу», всех полей группы «Бумажная книга (Publish-Info)» из текста документа с помощью т.н. «Служебных стилей»

16. Есть **Корректор Текста** – замена кавычек «елочкой», удаление лишних пробелов и многое другое...

17. Виртуальные клавиатуры.

18. **Валидатор fb2-файлов** (пока только для системы Windows).

19. Локализация всего пакета на Русский и Английский языки.

20. Экспорт **Нумерованных** и **Маркированных** Списков (в том числе и с **Маркеро-Картинок**) из **Текста** Сносок (Примечаний).

1.3. Что пока не реализовано

1 Для таблиц не реализован экспорт перехода *из* документа на определенный абзац таблиц, а так же экспорт стихов, цитат, аннотаций, эпитафий, подзаголовков и заголовков (Уровней) из таблиц. Не знаю – надо ли это?

2 Для Текстовых Врезок не сделан экспорт перехода *из документа* на определенный абзац Врезки.

3. Экспорт обложек из Врезок (расположены выше Названия Книги – стиль Book Title).

4. Экспорт Картинок, Формул и Диаграмм из текста сносок (Примечаний).

1.4. Лицензионное соглашение

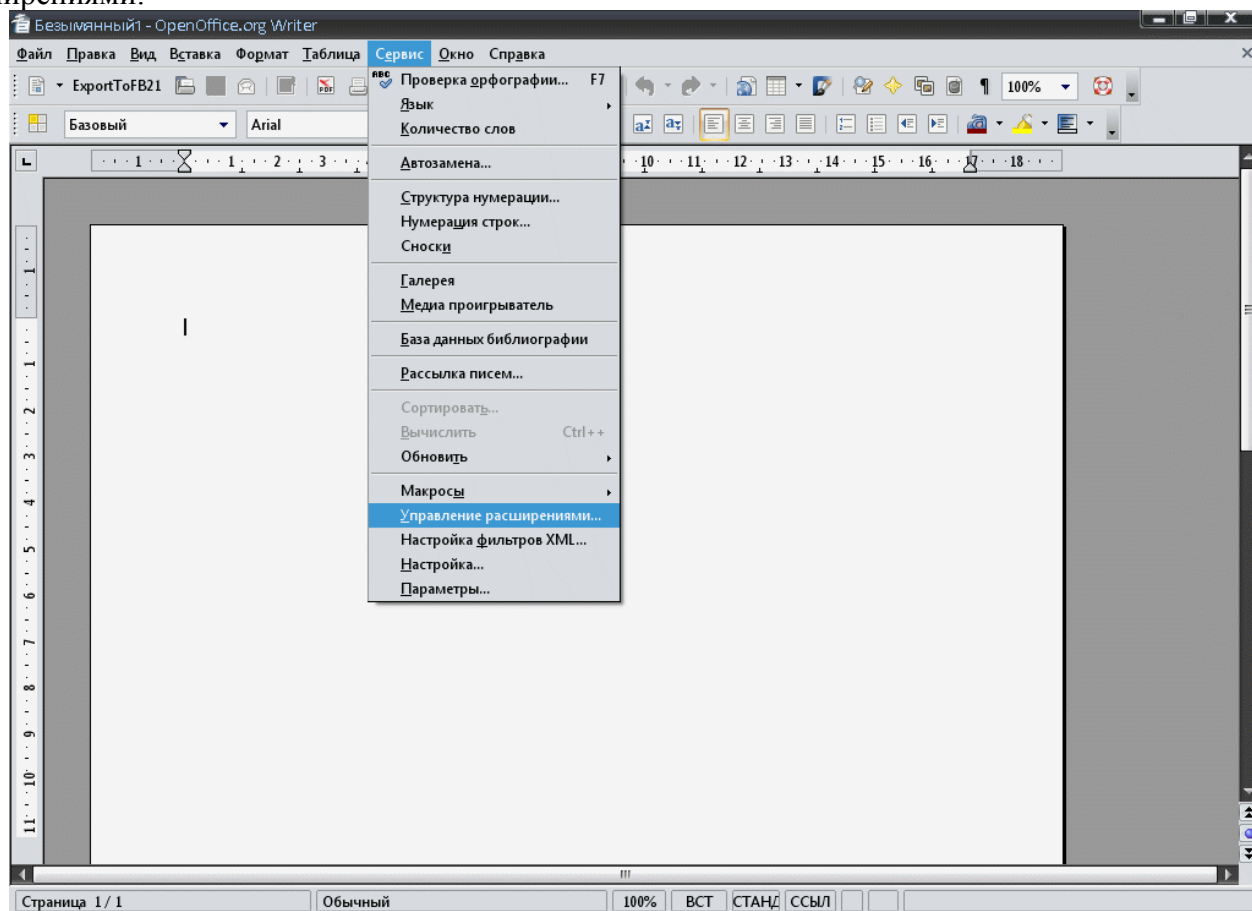
Конвертор создается и распространяется под лицензией GPL. Т.е. его можно свободно использовать, распространять, на базе его кода создавать свои проекты, тоже под лицензией GPL.

2. Установка конвертора

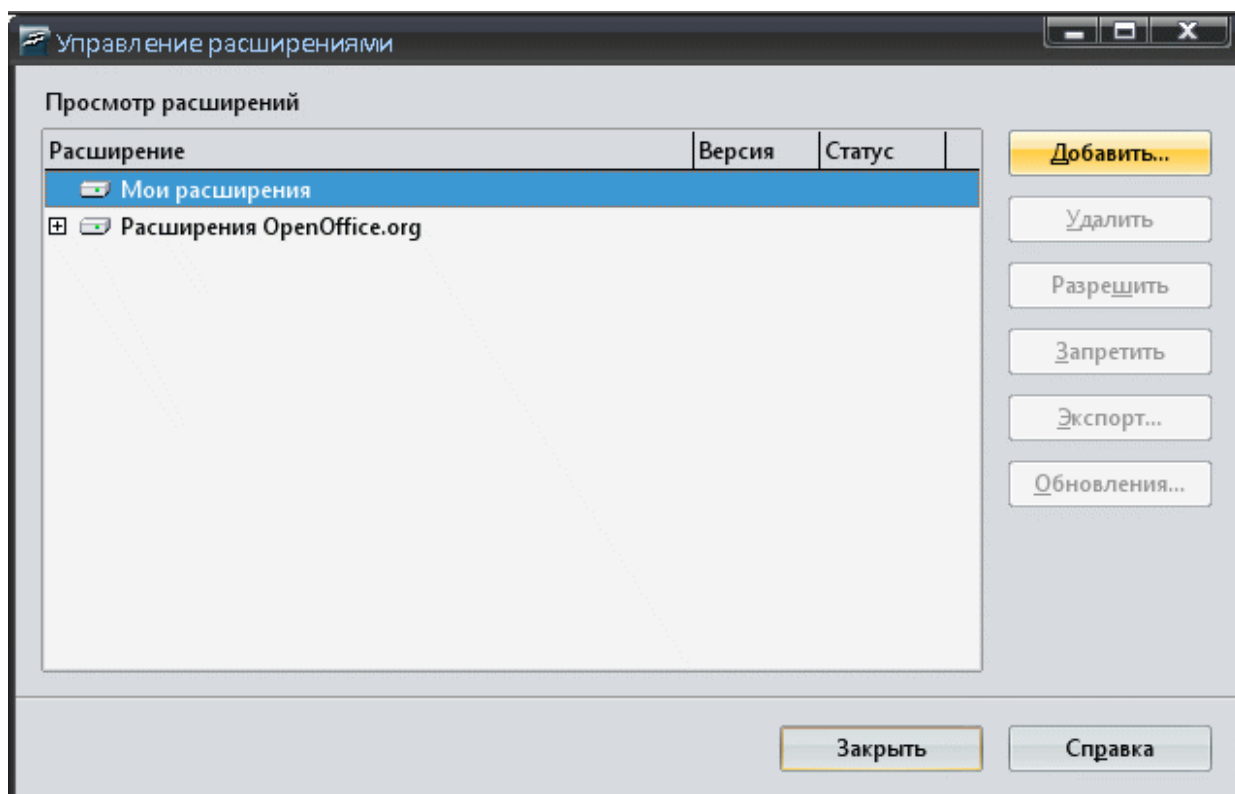
2.1. Установка расширения .oxt или .zip архива с конвертором

2.1.1. Первый способ (для всех операционных систем)

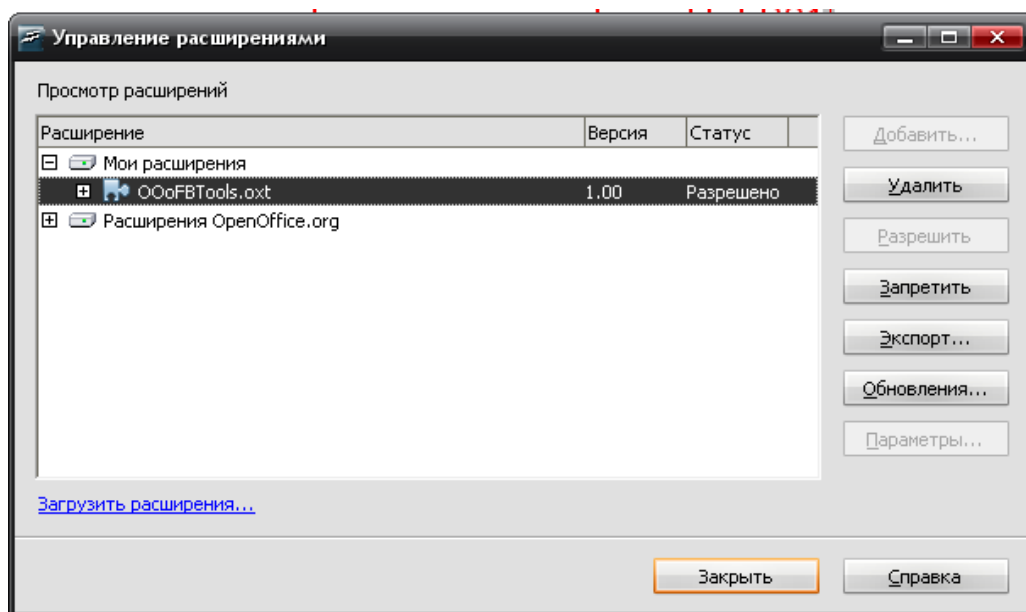
Вызываем диалог «Управление расширениями» через меню Сервис -> Управление расширениями:



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



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

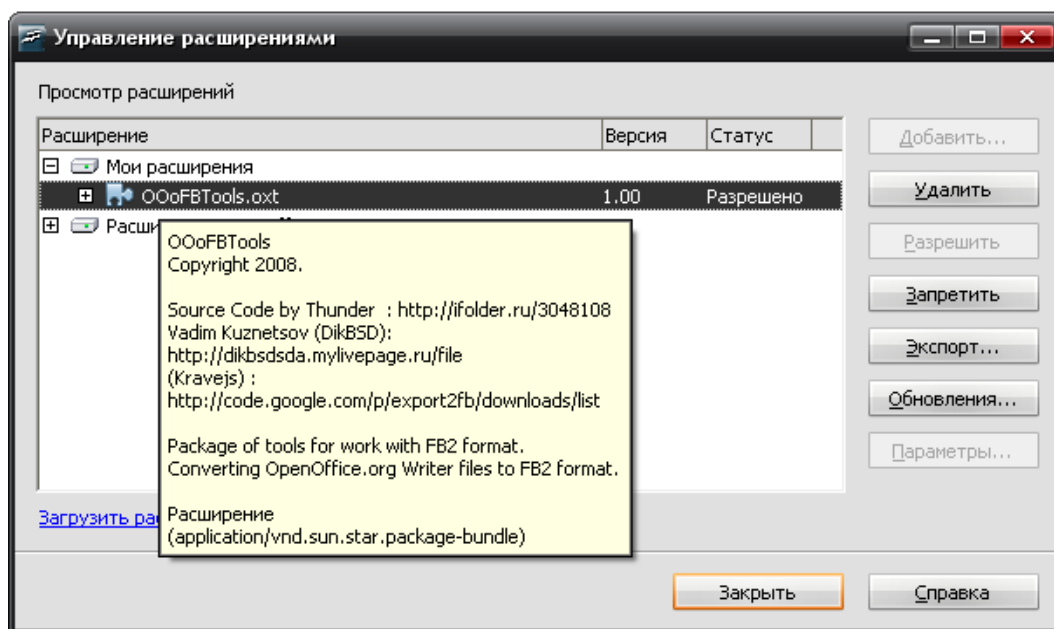


2.1.2. Второй способ (только для Microsoft Windows)

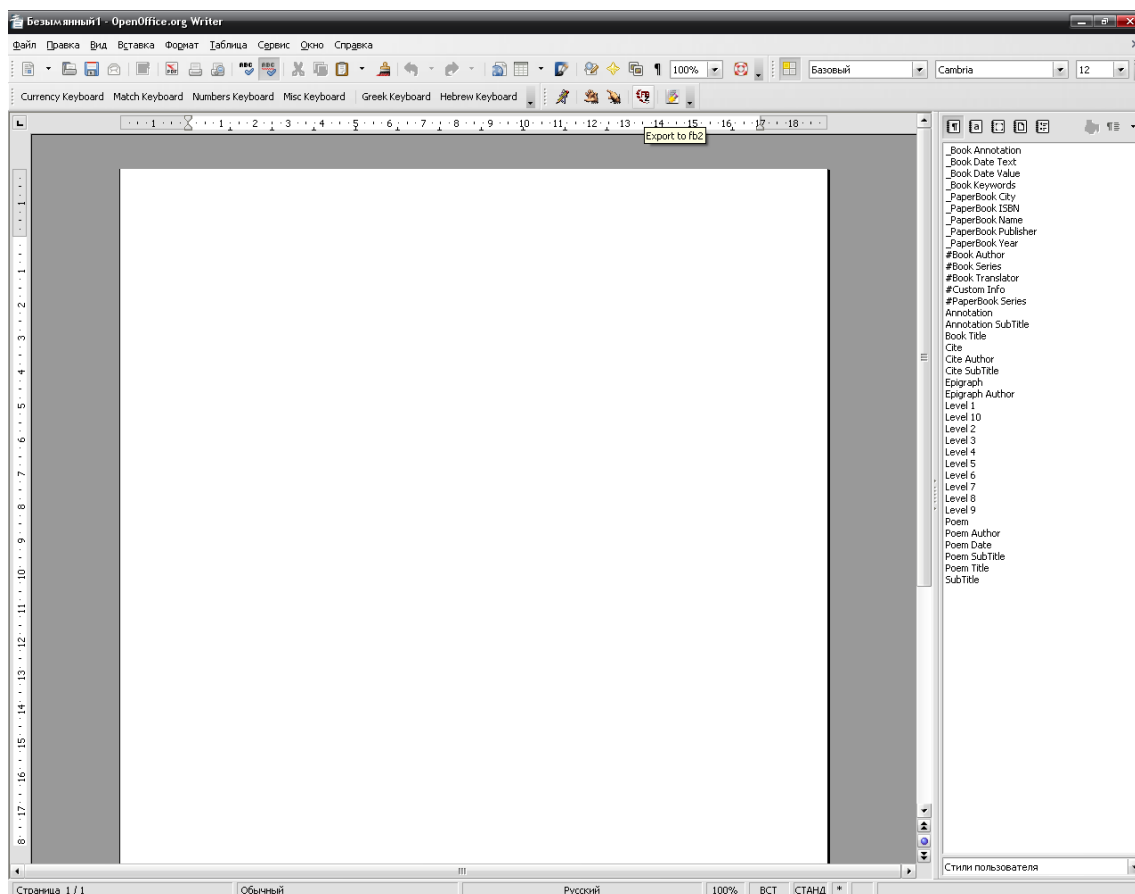
Второй способ применим *только* к установке любого расширения **.oxt** - просто из проводника или файлового менеджера запустите это расширения (по Enter или двойному клику мышки).

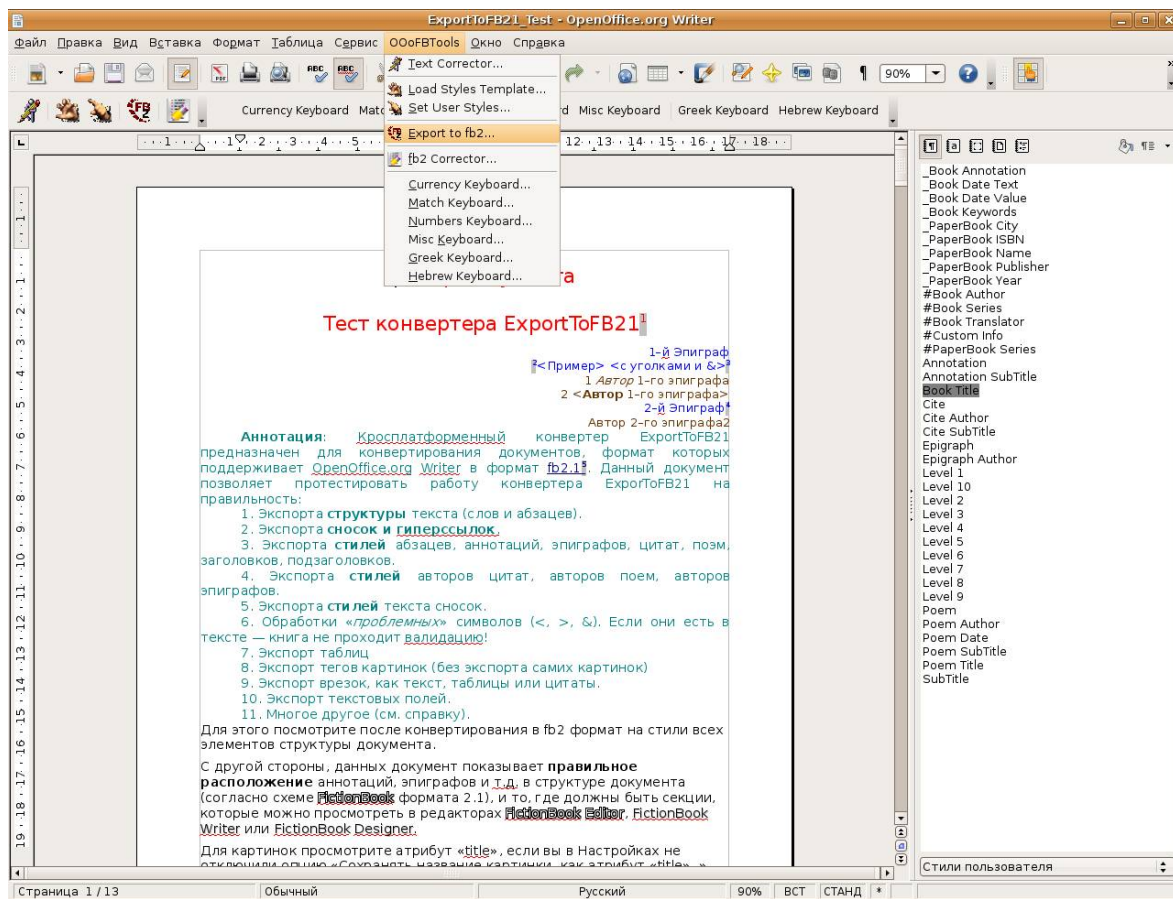
2.1.3. Дополнительная информация, панели инструментов, пункты меню...

При загрузке расширения показывается его версия и другая информация:



СОВЕТ: После установки пакета **OOoFBTools** **ПЕРЕЗАГРУЗИТЕ** программу, чтобы отображались панели инструментов пакета и его пункты меню:

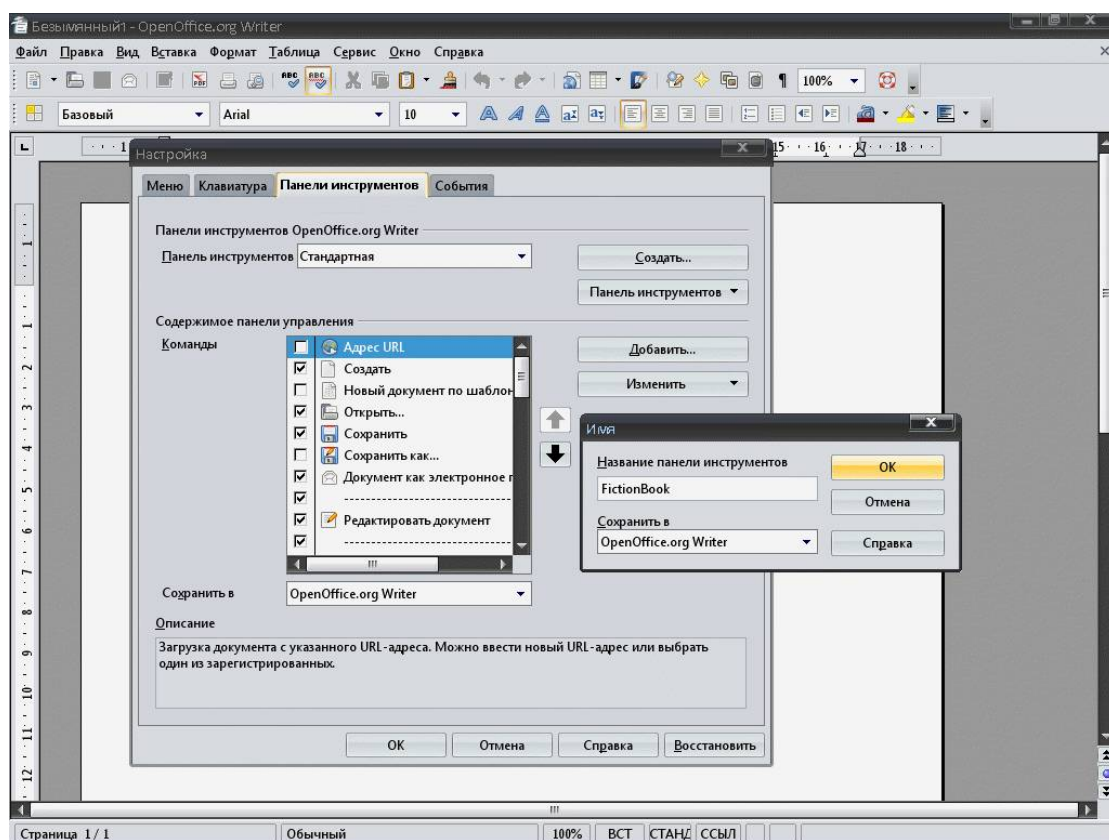




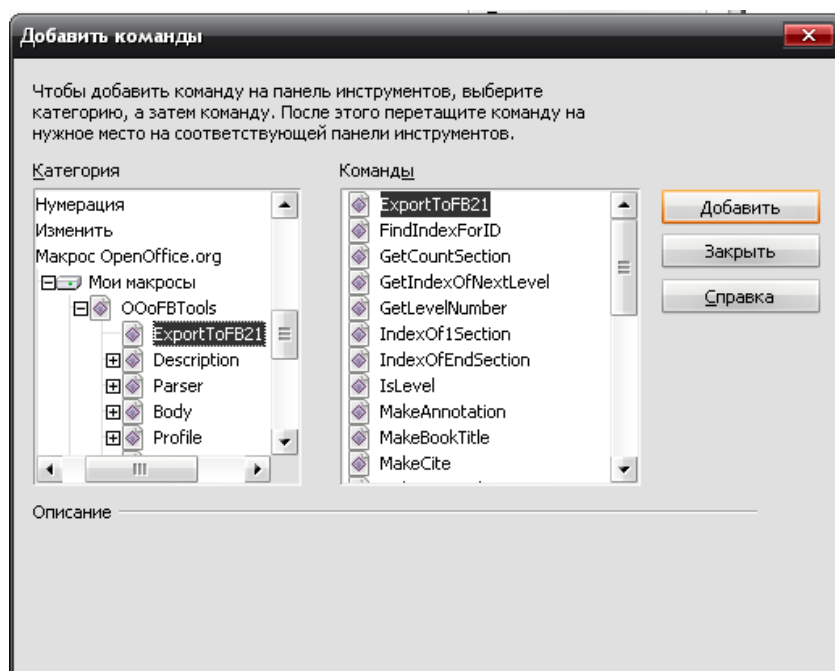
2.2. Вынос кнопок на панель

Если вас по каким-либо причинам не устраивают автоматически созданные панели инструментов конвертора, то вы можете создать свои. Делается это так.

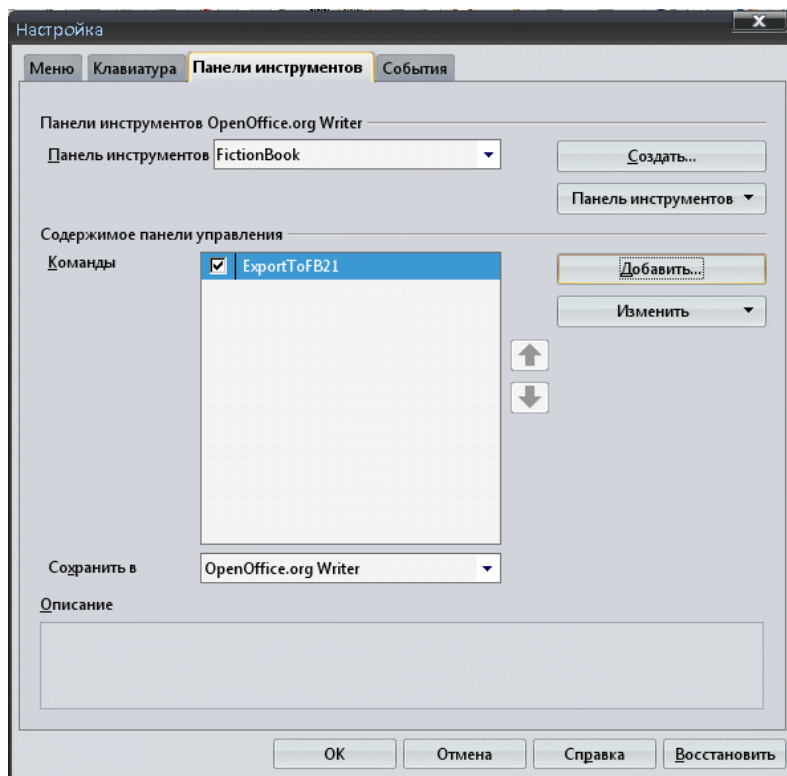
Открываем диалог "Настройка" через меню Сервис -> Настройка. Переходим на вкладку Панели инструментов, нажимаем на кнопку "Создать..." и в открывшемся диалоге вводим название новой панели (в моем случае FictionBook), в "Сохранить в" выбираем OpenOffice.org Writer, нажимаем ОК:



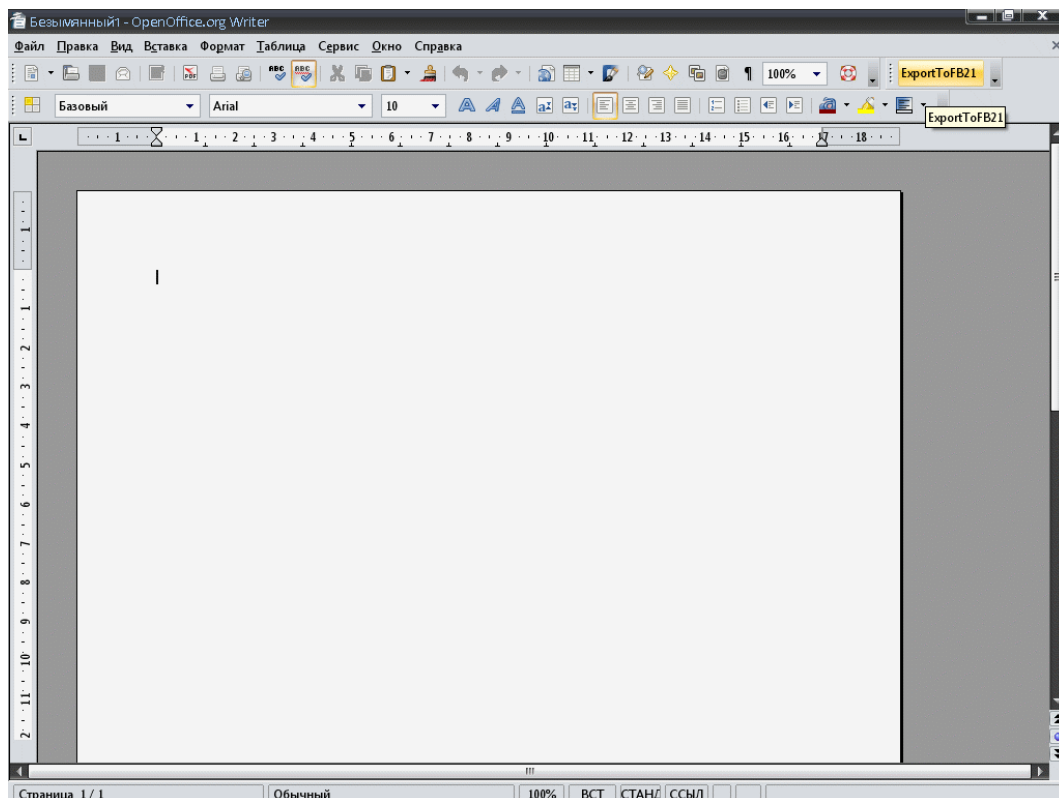
Теперь нажимаем кнопку "Добавить"... диалога "Настройка" и в открывшемся диалоге "Добавить команды" в древовидном списке "Категория" находим элемент Макрос OpenOffice.org / Мои макросы / OoFBTools / ExportToFB21. В списке "Команды" отобразится список макросов. Выбираем ExportToFB21 и нажимаем кнопку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



Нажимаем "Закреть". В результате получаем следующее:



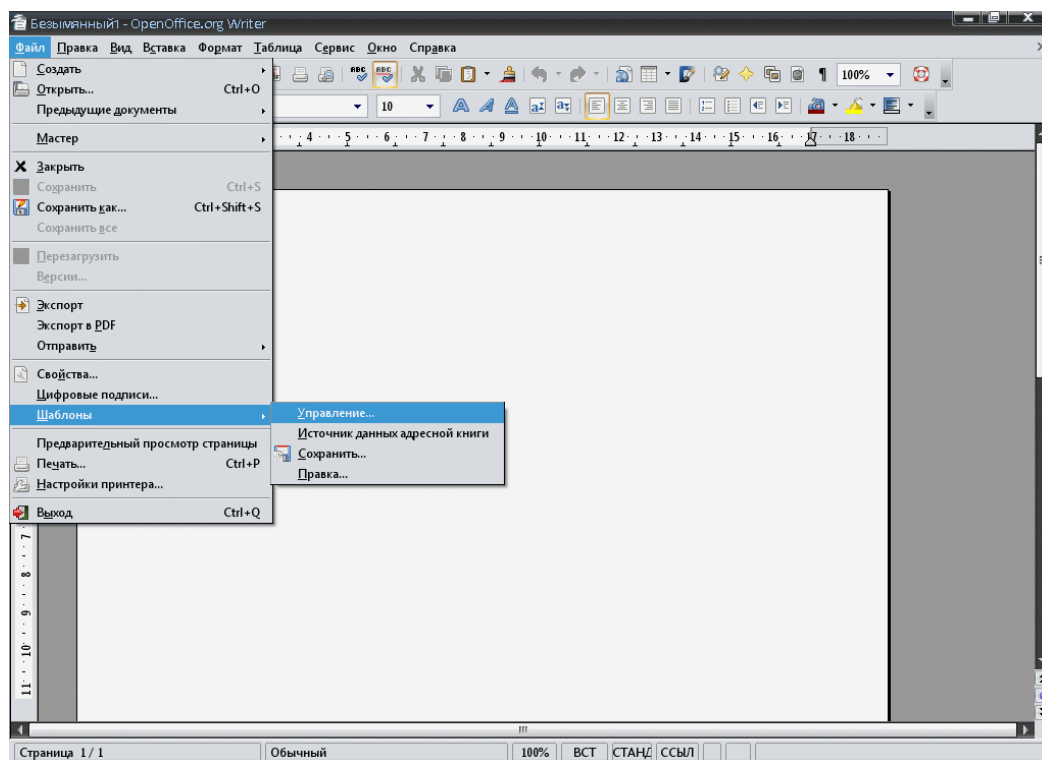
После нажатия "OK" мы увидим новую панель инструментов с нашей кнопкой "ExportToFB21"



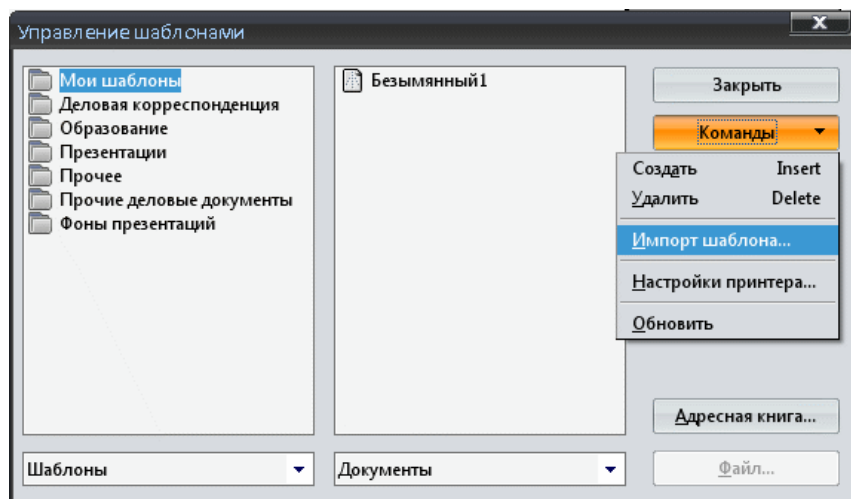
2.3. Установка шаблона стилей fb21_styles.ott

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

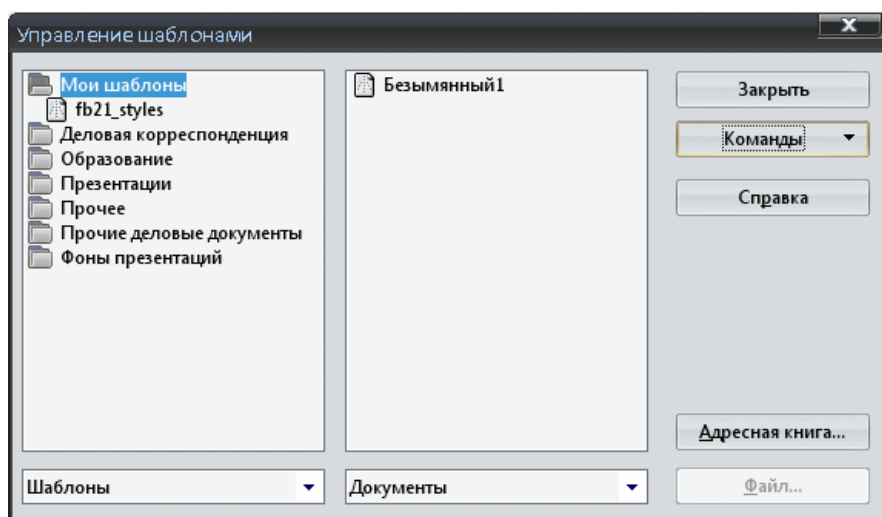
Для того, чтобы их установить, надо запустить диалог "Управление шаблонами" через меню Файл -> Шаблоны -> Управление:



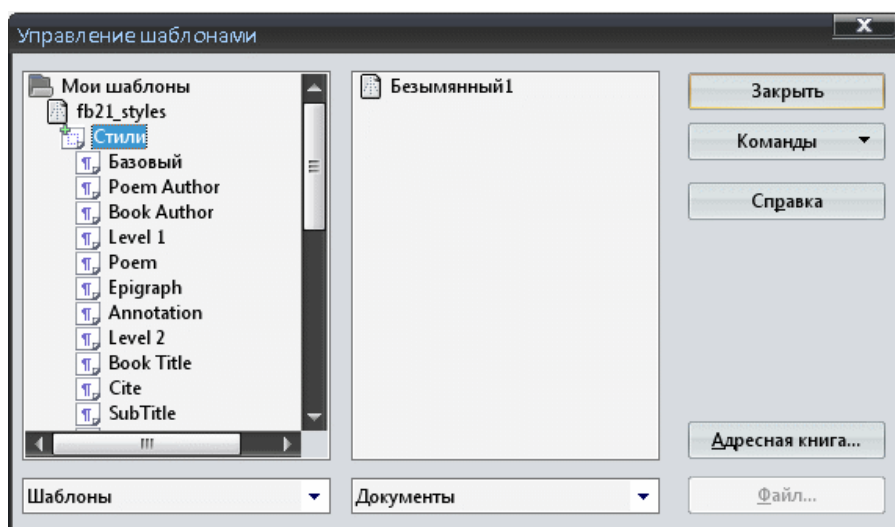
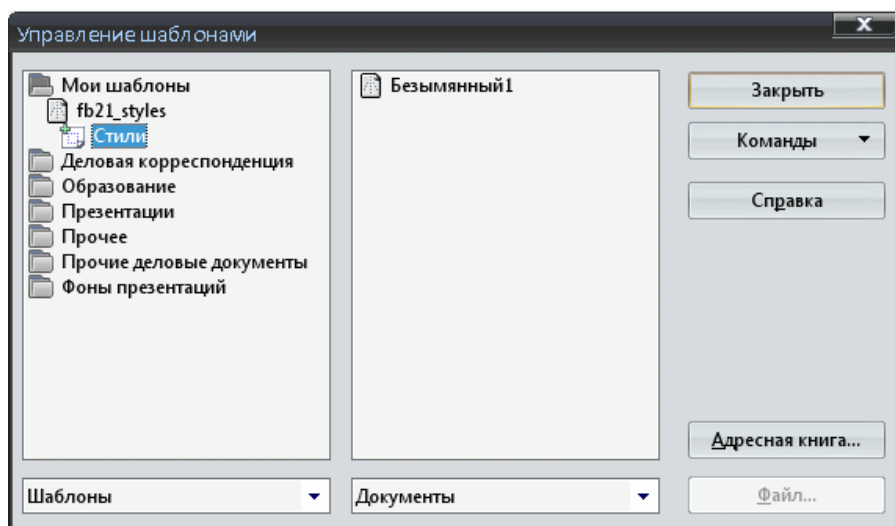
В открывшемся диалоге выделяем в списке шаблонов "Мои шаблоны", нажимаем кнопку "Команды" и выбираем "Импорт шаблона"



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



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



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

2.4. Установка дополнительных платформозависящих пакетов и приложений

2.4.1. Для Linux и других Unix-подобных систем

Для кодирования картинок используется утилита *base64*. Пакет с ней обычно сразу устанавливается при установке дистрибутива Linux. С *Ubuntu* – это так. Если же на вашем дистрибутиве Linux почему-то не установлен этот пакет, то его необходимо доустановить. Чтобы проверить, установлен ли пакет *base64* или нет, выполните в терминале команду: *base64 -?* Если пакет установлен, то вы увидите краткую справку, если нет – то терминал выдаст соответствующее сообщение.

2.4.2. Для системы Mac Os X

Для кодирования картинок используется возможности OpenSSL (установлено по умолчанию!).

2.4.3. Для системы Windows

Для работы Валидатора fb2-файлов в системе должен быть установлен **msxml4** или выше. Его можно скачать с официального сайта корпорации Microsoft.

3. Работа со стилями структуры документа

3.1. Виды стилей

Стили, с которыми работает конвертор, делятся на 2 группы:

1. Inline (применяются в пределах одного абзаца – bold, italic...) и
2. Стили абзаца – их действие распространяется на весь абзац.

В свою очередь стили абзаца и Inline стили делятся на 2 группы: основные и «служебные».

Без всех этих стилей невозможно экспортировать различные структурные элементы книги (стихи, цитаты)!

3.1.1. Inline стили конвертора

Стили слов и (или) фраз используются для выделения в тексте полужирного, курсивного, моноширинного начертания, слов и фраз в верхнем и нижнем индексе, зачеркивания.

Перечень Inline стилей: Bold, Italic, Strikethrough, Тень или Контур, Верхний индекс, Нижний индекс.

Для использования этих стилей достаточно выделить слово (фразу) и нажать соответствующую кнопку панели инструментов или выбрать соответствующий пункт меню или контекстного меню. Здесь все просто.

3.1.2. Основные стили конвертора

Основные стили применяются для «выделения» в тексте документа (книги) таких структурных единиц, как аннотация, цитата, поэма, заголовков, и т.д.

Перечень основных стилей: Annotation, Annotation SubTitle, Book Title, Cite, Cite Author, Cite SubTitle, Epigraph, Epigraph Author, Poem, Poem Author, Poem Title, Poem SubTitle, Poem Date, SubTitle, Level 1 ... Level 10.

Как использовать основные стили см. [п.3.6.2](#).

3.1.3. «Служебные стили» конвертора

«Служебные стили» применяются для «выделения» в тексте документа (книги) Аннотации Книги, всех данных на Авторов Книги, Переводчиков Книги, Дату написания Книги, Ключевых слов, Бумажную книгу (Название, Издатель, Город, Год издания, ISBN, Серия и Номер серии), Жанр и его match (если нужно). Т.е. почти всех тех данных, которые можно заносить вручную в соответствующие поля основной формы конвертора.

«Служебные стили» начинаются с символа «#» или «_», чтобы их легче было отличать от основных и сгруппировать (намного легче и быстрее найти нужный стиль в списке). Конечно, вы можете переименовать их, как вам угодно (см. [«Переопределение стилей по умолчанию на свои стили»](#)).

Стили, начинающиеся с «_» относятся только ко всему абзацу, и не требуют символьных стилей.

Стили, начинающиеся с «#» относятся ко всему абзацу, но для выделения его отдельных частей требуются символьные стили, тоже начинающиеся с «#». Тем самым, символ «#» в начале названия стиля «подсказывает» - «ищи связанные с ним символьные стили, тоже начинающиеся с «#»».

Как использовать основные стили см. [п.3.6.3](#).

3.1.3.1. Таблица совместно используемых «служебных стилей»:

Назначение	Стили Абзаца	Символьные стили
Аннотация на Книгу	_Book Annotation	НЕТ
Ключевые слова	Book Keywords	НЕТ
Жанр	#Genre	#Genre Name, #Genre Match
Язык	#Lang	#Lang Book, #Lang Original
Автор	#Book Author	#Name, #Family, #Patronymic, #Nick, #Email, #WebCite
Переводчик	#Book Translator	#Name, #Family, #Patronymic, #Nick, #Email, #WebCite
Дата написания	#Book Date	#Book Date Text, #Book Date Value
Серия Электронной Книги	#Book Series	#Sequence, #Sequence Number
Серия Бумажной Книги	#PaperBook Series	#Sequence, #Sequence Number
Данные Бумажной Книги	#PaperBook	#PaperBook Name, #PaperBook Publisher, #PaperBook City, #PaperBook Year, #PaperBook ISBN
Информация пользователя	#Custom Info	#Custom Info Type, #Custom Info Value

При создании fb2-файла книги часто используются документы, где нет информации о бумажной книге и издательстве. В этом случае все эти данные заносят в соответствующие поля формы конвертора вручную, считывая с обложки книги.

Другое дело, если вы отсканировали книгу, распознали текст. В этом случае информация о бумажной книге и реквизиты издательства уже есть в тексте. Конечно, можно просто вырезать эти данные в какой-нибудь текстовый редактор, и потом вставляю их в соответствующие поля формы конвертора. А можно поступить проще и удобнее – воспользоваться «служебными стилями». При этом, текст, выделенный «служебными стилями» **не будет экспортирован в fb2-файл**, а *автоматически загрузится в*

соответствующие поля основной формы конвертора. Поэтому **будьте внимательны**, чтобы случайно не отформатировать «служебными стилями» какой-нибудь абзац!

Как использовать основные стили см. [п.3.6.3](#).

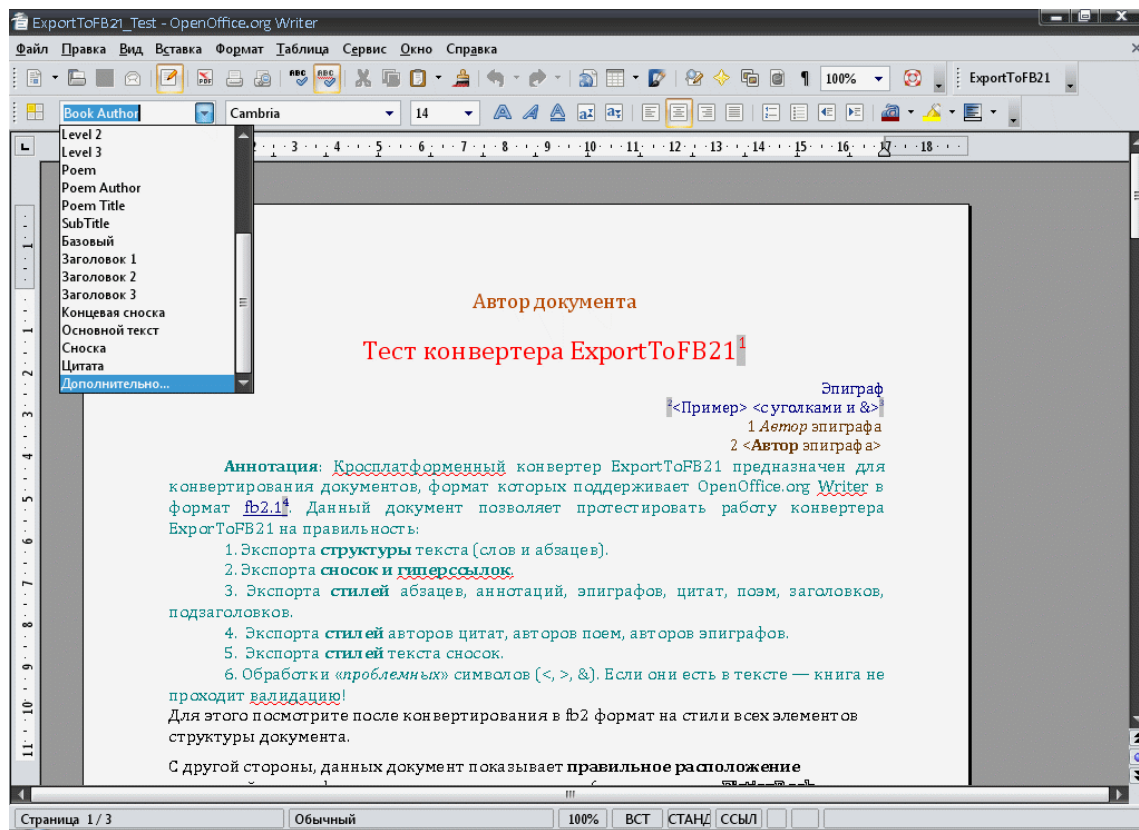
3.2. Загрузка стилей

При экспорте конвертор ищет в документе определенные стили и «обрамляем» абзацы, помеченные этими стилями в соответствующие им теги формата fb2.1 (см. [п.3.5](#)). 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!** Если же вы изменили у стилей и их название, то обязательно переопределите эти стили, как сказано в разделе [«Переопределение стилей по умолчанию на свои стили»](#)!

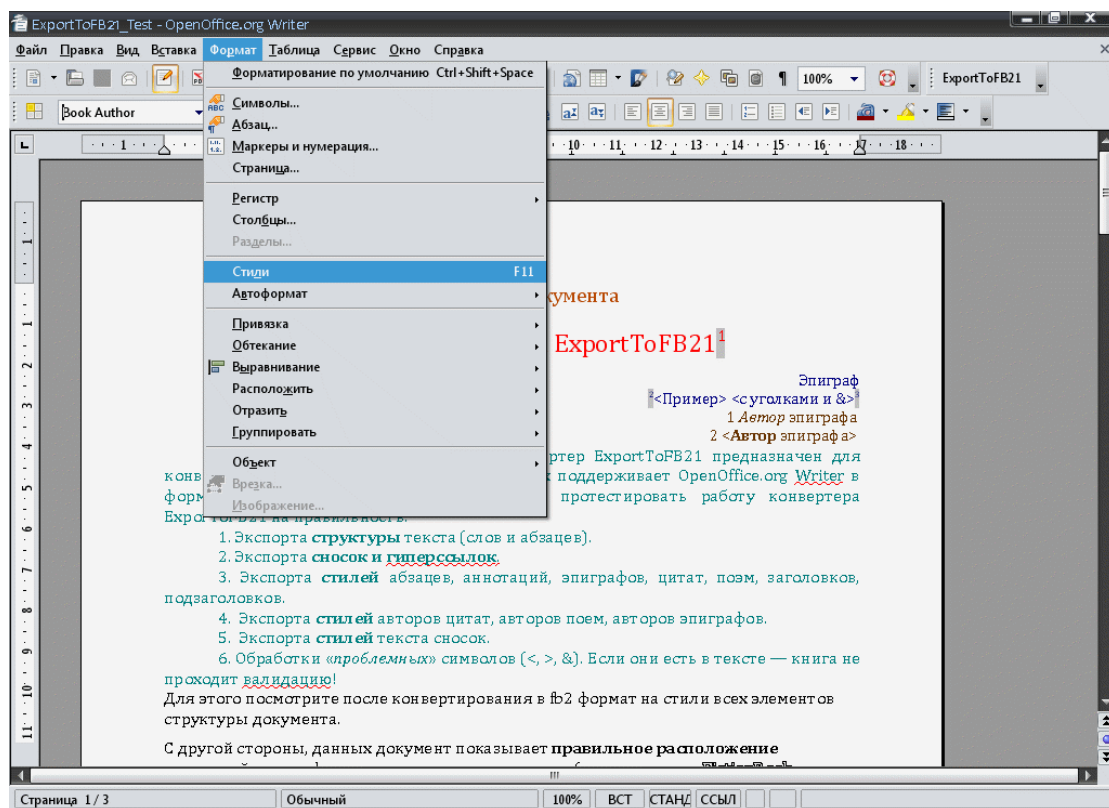
Для работы со стилями надо открыть окно «Стили и форматирование», если оно еще не открыто.

3.2.1. Открытие окна стилей

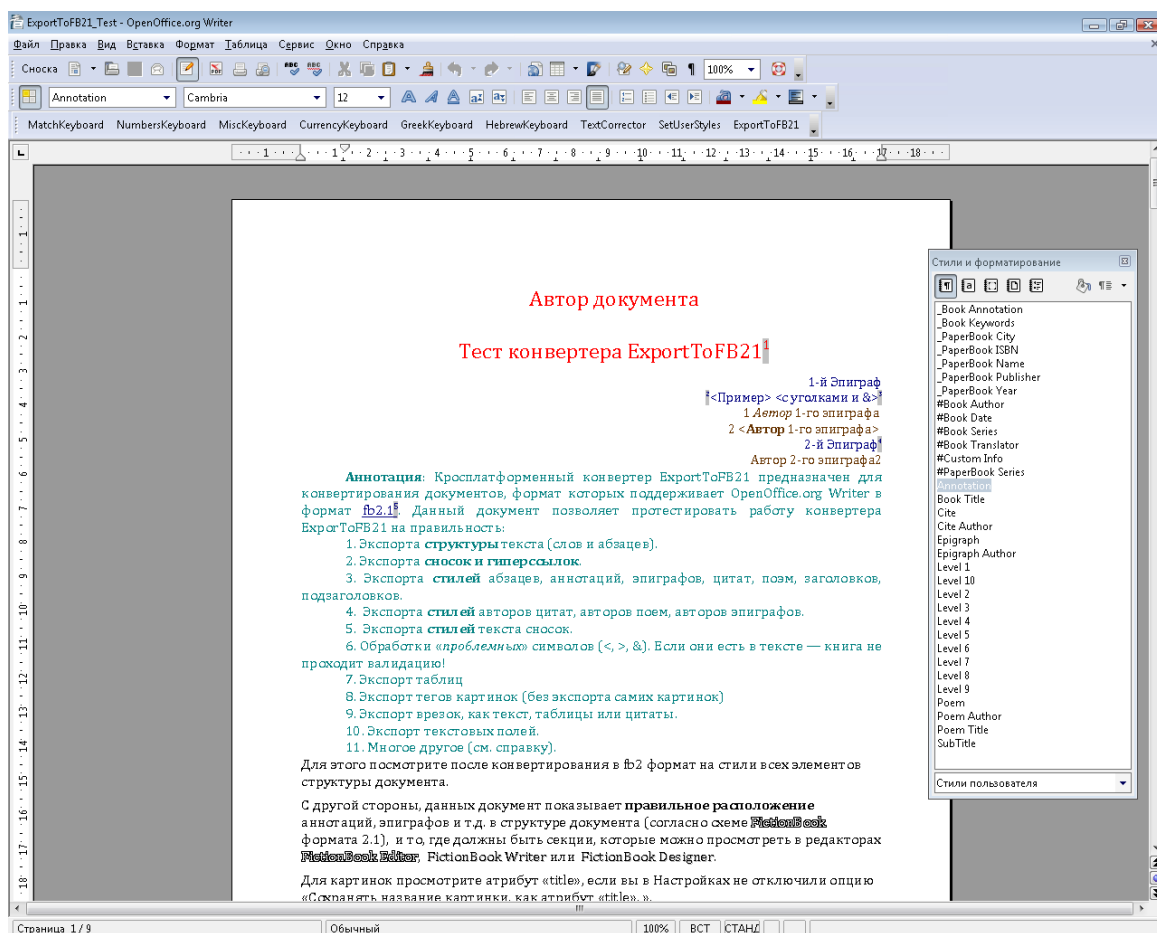
В комбобоксе «Стиль» главного окна редактора выбираем самый нижний элемент «Дополнительно».



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



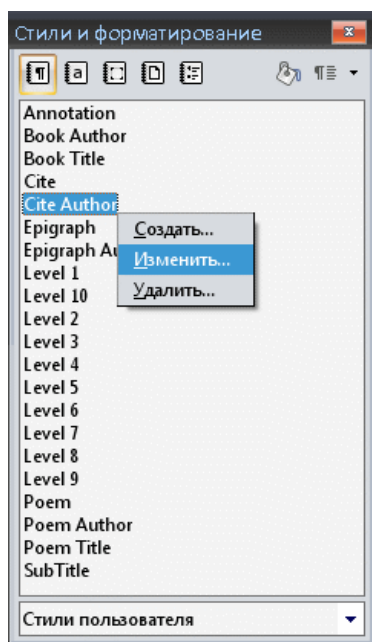
В результате:



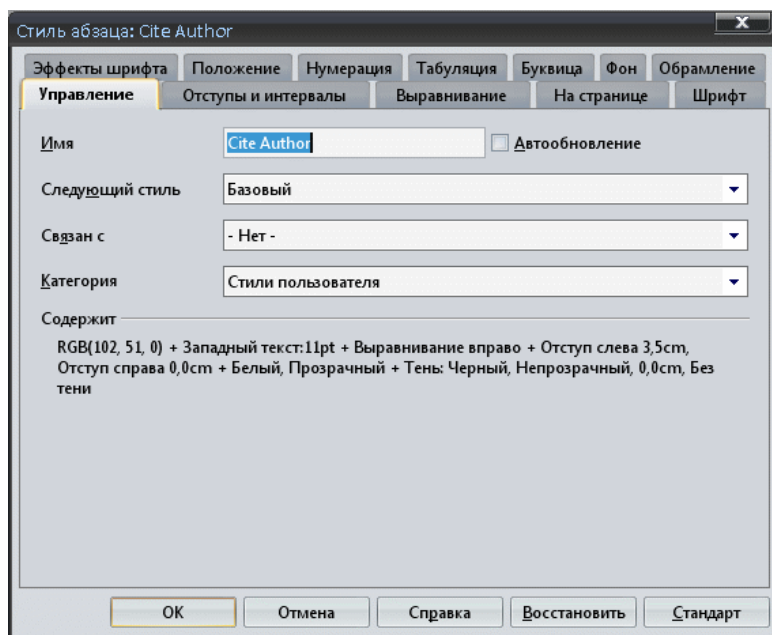
В этом окне «Стили и форматирование» выбираем внизу «Стили пользователя». Это как раз те стили, которые нам необходимы для разметки структуры документа.

3.3. Редактирование стилей

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



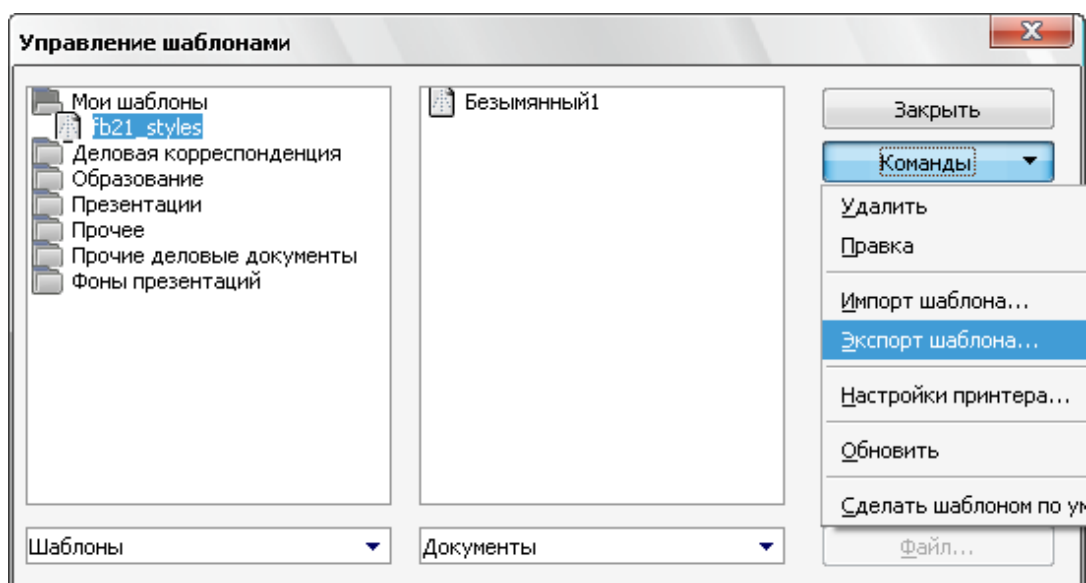
В результате появится диалог «Стиль абзаца», где можно отредактировать стиль. Категорию лучше оставить как есть – «Стили пользователя» – удобно работать со стилями, когда они в одной категории (см. [п. 3.5](#)).



3.4. Сохранение откорректированного шаблона стилей fb21_styles.ott

Для сохранения своего откорректированного шаблона стилей надо запустить диалог "Управление шаблонами" через меню Файл → Шаблоны → Управление (картинку см. [п.2.3](#)).

В открывшемся диалоге кликаем два раза мышкой по "Мои шаблоны", выделяем шаблон стилей **fb21_styles**, нажимаем кнопку "Команды" и выбираем "Экспорт шаблона":



3.5. Соответствие стилей с тегами формата fb2.1

3.5.1. Inline стили (символьные) «по умолчанию»

Назначение	Стиль
Полужирный 	Bold
Курсив <emphasis>	<i>Italic</i>
Зачеркивание <strikethrough>	Зачеркивание (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Зачеркивание ")
Код программ – <code>	Тень или Контур (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Тень" или "Курсив"), а также символный стиль «code»
Верхний индекс – <sup>	"Верхний индекс"
Нижний индекс – <sub>	"Нижний индекс"

3.5.2. Стили «по умолчанию», применимые для всего абзаца

Назначение	Стиль
Название книги	Book Title
Аннотация	Annotation
Подзаголовок аннотации	Annotation SubTitle
Стих	Poem
Заголовок стиха	Poem Title
Подзаголовок стиха	Poem SubTitle
Эпиграф стиха	Epigraph
Автор эпиграфа стиха	Epigraph Author
Автор стиха	Poem Author
Дата написания стиха	Poem Date
Эпиграф	Epigraph
Автор эпиграфа	Epigraph Author
Цитата	Cite

Подзаголовок цитаты	Cite SubTitle
Автор цитаты	Cite Author
Подзаголовок (* * *)	SubTitle
Заголовки (Уровни)	Level 1 ... Level 10

3.5.3. “Служебные стили” «по умолчанию»

Служебные стили по-умолчанию показаны в [таблице](#) совместно используемых «служебных стилей». Вы можете изменить их название, как вам будет удобнее, только в этом случае необходимо будет их переопределить и с помощью диалога задания пользовательских стилей (см. [здесь](#)).

3.6. Использование стилей в документе

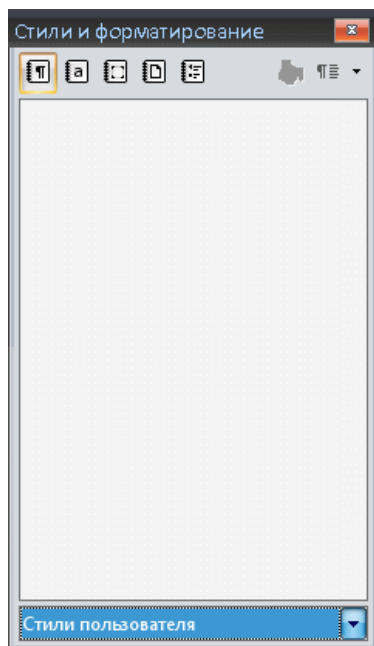
Книга готовится к экспорту в формат fb2.1 двумя способами: просто открывается готовый документ формата, поддерживаемого OOo Writer или печатается «с чистого листа» – создается новый документ. В любом случае – надо загрузить уже установленные стили, с которыми вы будете работать. Делается это так:

3.6.1. Загрузка стилей в документ

3.6.1.1. Первый способ

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](#)).

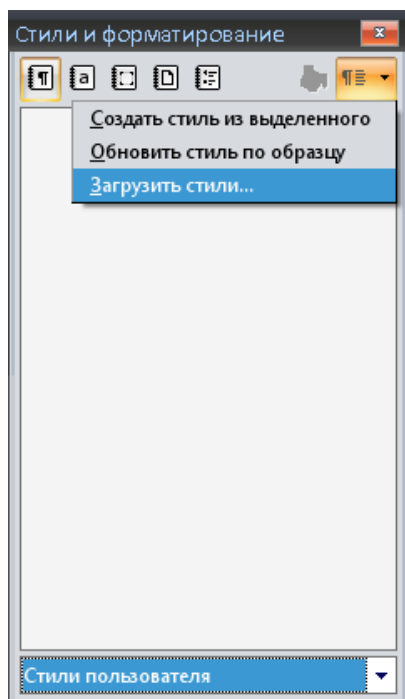
2. В открывшемся диалоге выбираем «Стили пользователя»



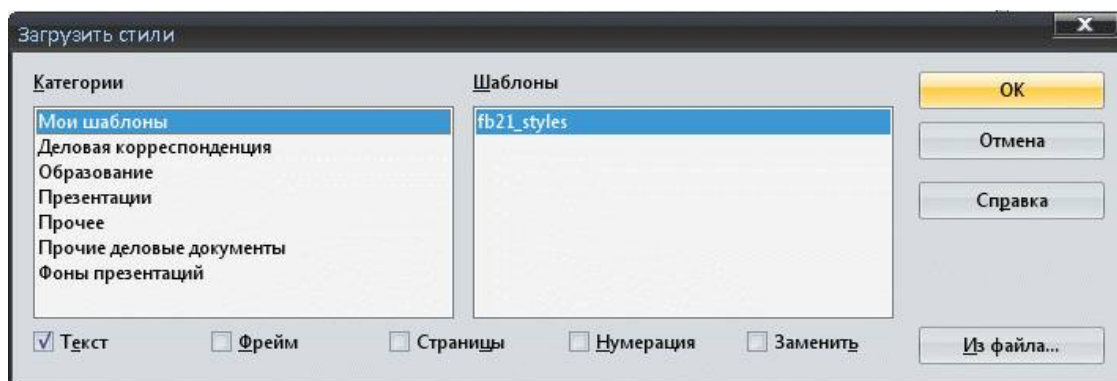
Нажать на самую правую кнопку



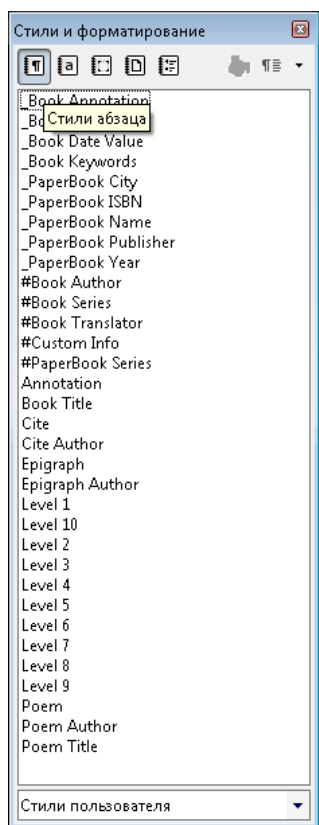
и выбрать "Загрузить стили..."



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

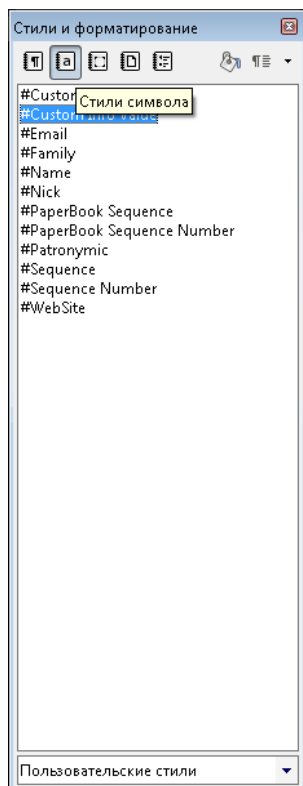


В результате наши стили загружены:



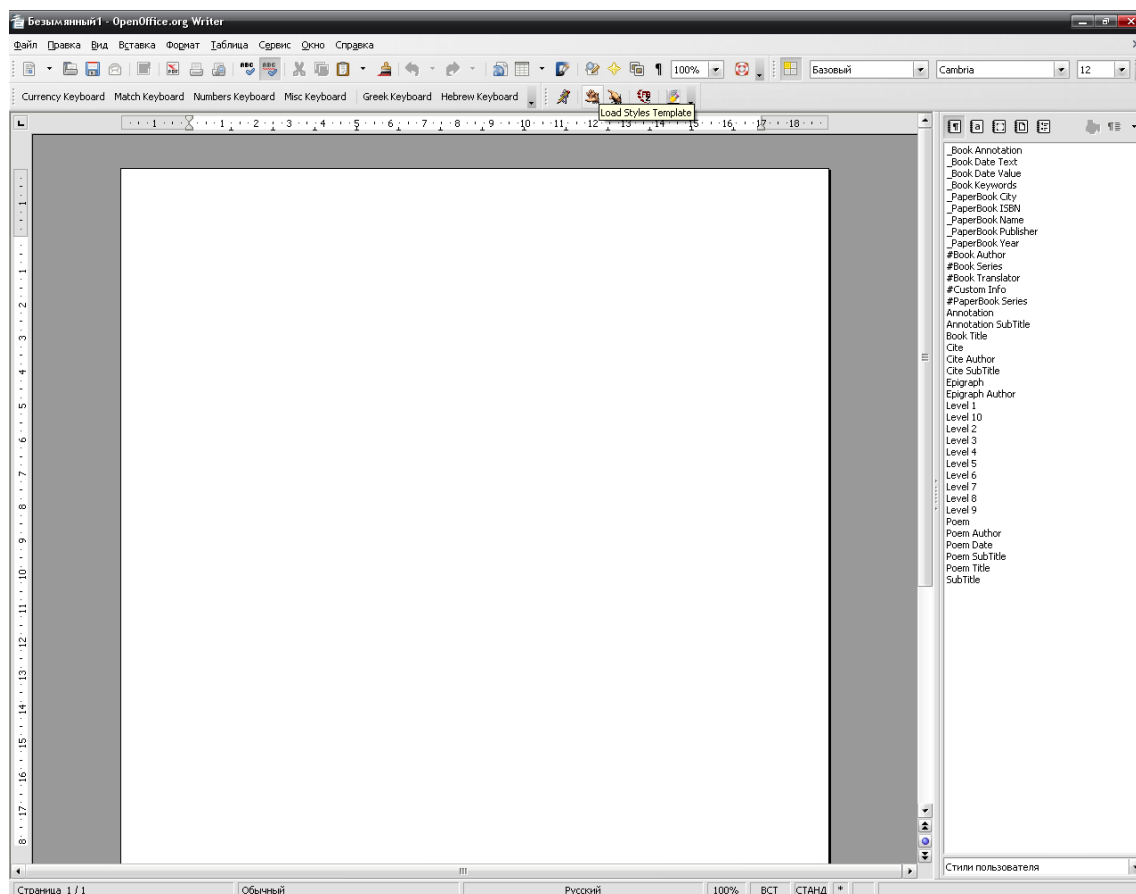
Здесь на картинке представлены стили абзаца (нажата кнопка ¶).

Чтобы выбрать Inline стили (символьные), нажмите кнопку «a»:



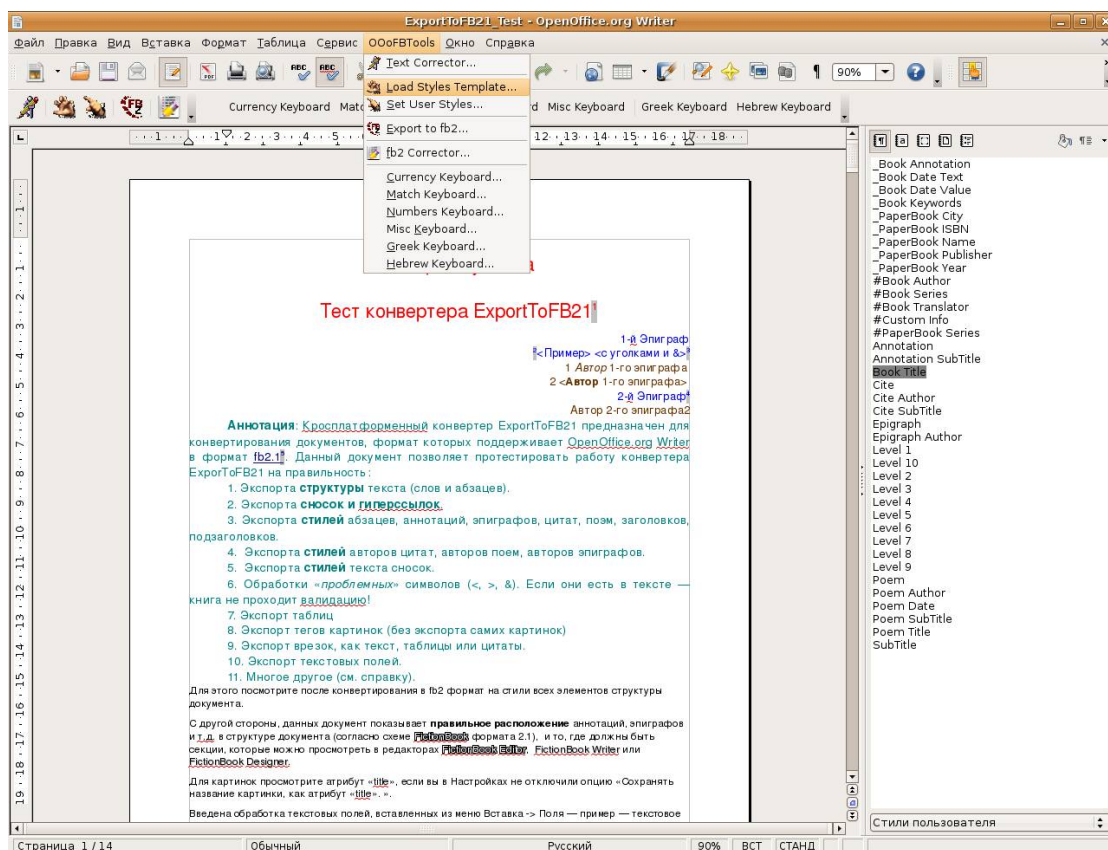
3.6.1.2. Второй способ

Второй способ – самый легкий. При запуске редактора автоматически загружаются панели инструментов конвертора. На панели **FictionBook Tools** есть кнопка **Load Styles Template**. Просто нажимаем на нее, и стили загружены:



3.6.1.3. Третий способ

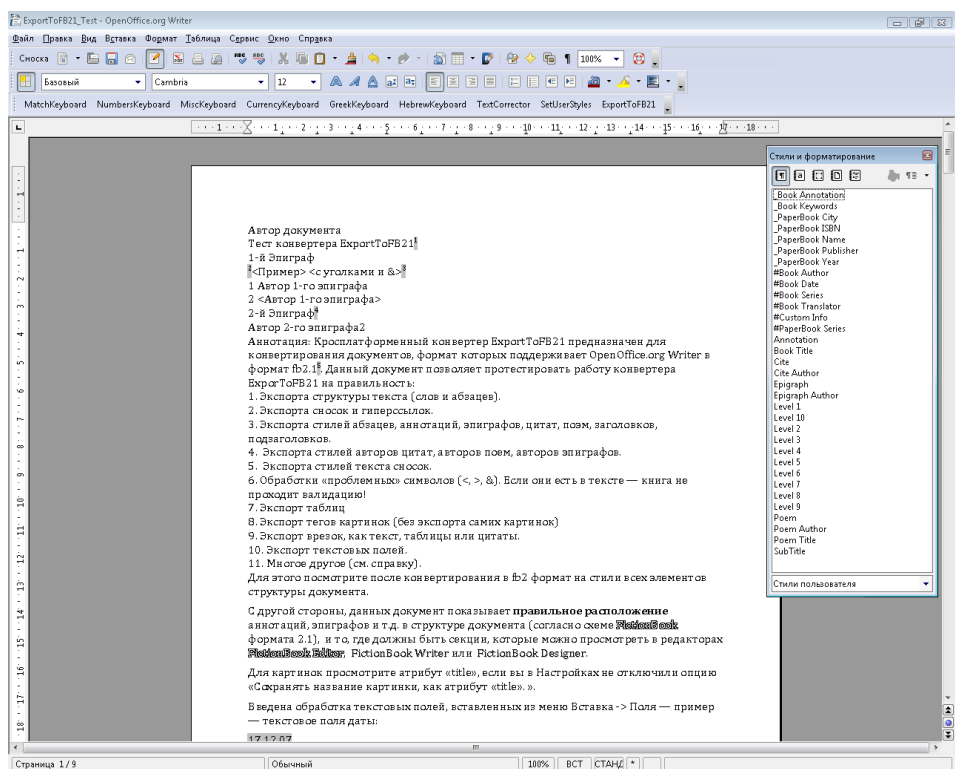
Можно загрузить стили в текущий документ и через меню редактора: OOoFBTools -> Load Styles Template. Результат – тот же:



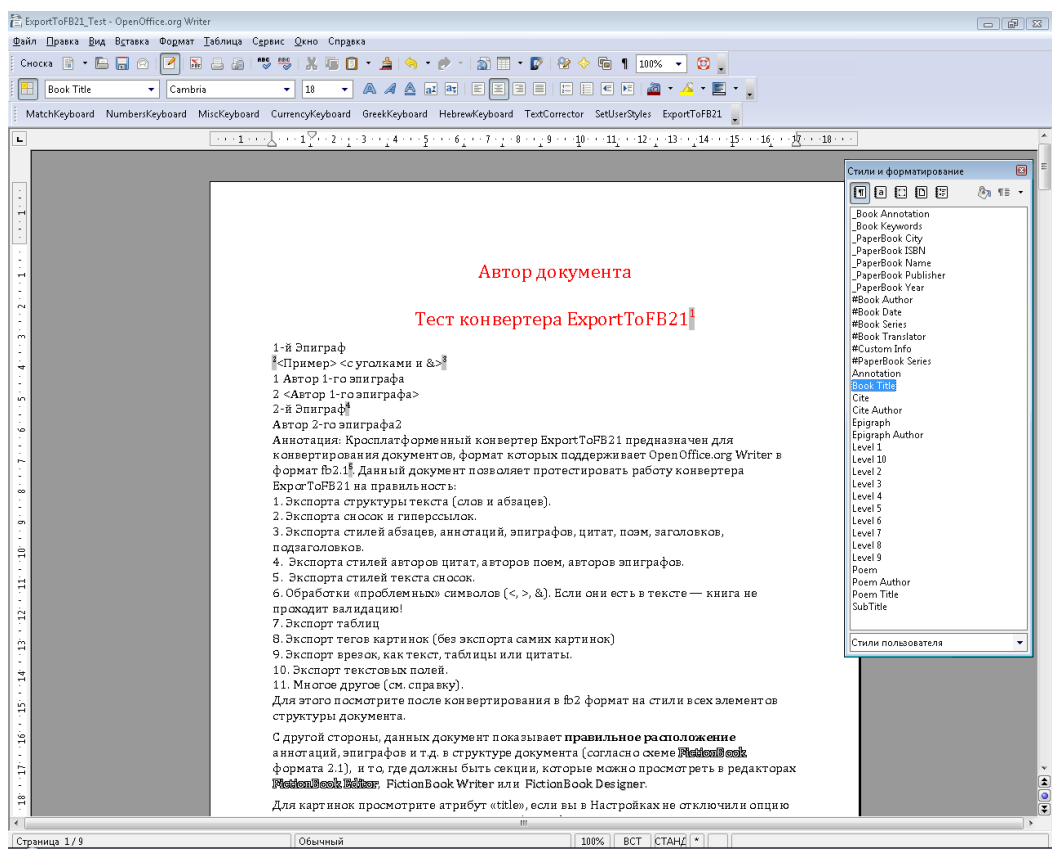
3.6.2. Работа с основными стилями в документе

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.

3.6.3. Работа со «служебными стилями» в документе

Поскольку конвертор изначально поддерживает экспорт для **нескольких** Авторы, **нескольких** Переводчиков и т.д. (списки), то это сказывается на работе со «служебными стилями».

1. Данные без списков

Для данных, которые экспортируются только в единичном экземпляре - Аннотация на Книгу, Ключевые Слова:

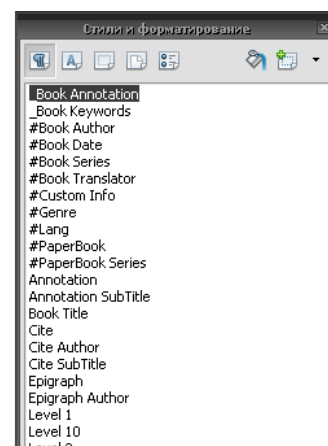
Тут все просто: придаем «их» абзацам «их» стили - _Book Annotation и _Book Kyewords соответственно:

Для Аннотации на Книгу:

Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием». Он состоит из 2-х абзацев с ссылкой на Часть 7, [гиперссылкой](#) и [сноской](#)¹.

Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 [Email](#) Автор2 [HomePage](#)
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 [Email](#) Переводчик2 [HomePage](#)
1 2 3 4 5 6 7 8 9 0
Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001
Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN, Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
[Custom editor](#) ФИО Редактора
[Custom edition](#) 5

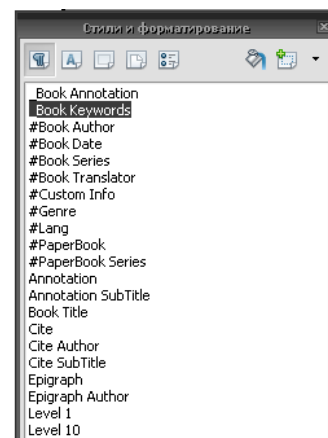


Для Ключевых Слов:

Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием». Он состоит из 2-х абзацев с ссылкой на Часть 7, [гиперссылкой](#) и [сноской](#)¹.

Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 [Email](#) Автор2 [HomePage](#)
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 [Email](#) Переводчик2 [HomePage](#)
1 2 3 4 5 6 7 8 9 0
Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001
Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN, Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
[Custom editor](#) ФИО Редактора
[Custom edition](#) 5



2. Данные для списков

Другая информация на Книгу предполагает списки данных: списки Авторы, Переводчиков, Custom Info, Серий Книги, Серий Бумажной Книги, Данных на Бумажную Книгу.

В этом случае нужно использовать и стили абзаца, и символьные стили для отдельных данных (имя, фамилия, дата...).

Для данных на Бумажную Книгу (Название, Издательство, Год, город, ISBN):

Сначала помечаем нужный абзац стилем #PaperBook:

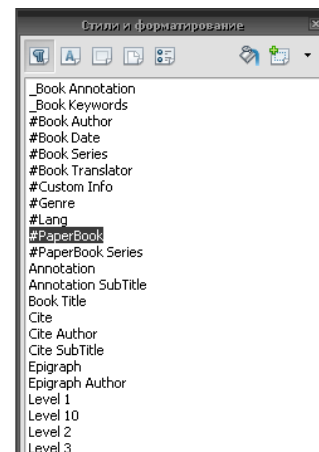
Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием».
Он состоит из 2-х абзацев с ссылкой на Часть 7, гиперссылкой и сноской¹.

Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 Email Автор2 HomePage
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 Email Переводчик2 HomePage
1 2 3 4 5 6 7 8 9 0

Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001

Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN,
Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
Custom editor ФИО Редактора
Custom edition 5



а потом символьными стилями отмечаем отдельные данные в этом абзаце:

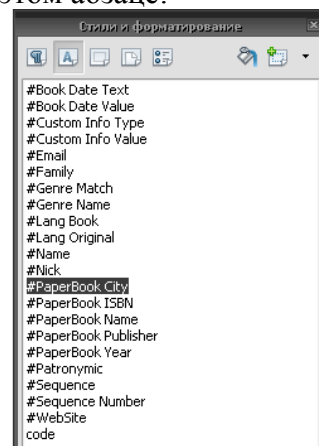
Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием».
Он состоит из 2-х абзацев с ссылкой на Часть 7, гиперссылкой и сноской¹.

Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 Email Автор2 HomePage
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 Email Переводчик2 HomePage
1 2 3 4 5 6 7 8 9 0

Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001

Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN,
Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
Custom editor ФИО Редактора
Custom edition 5



Аналогично делаем и для Даты Написания Книги, Серии Электронной Книги, Серии Бумажной Книги, Custom Info:

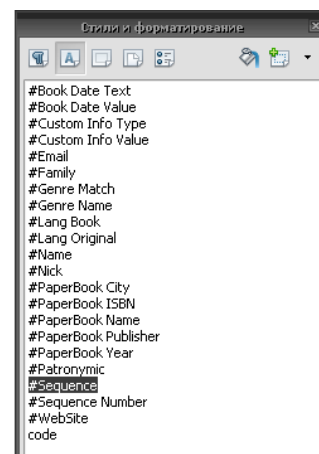
Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием».
Он состоит из 2-х абзацев с ссылкой на Часть 7, гиперссылкой и сноской¹.

Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 Email Автор2 HomePage
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 Email Переводчик2 HomePage
1 2 3 4 5 6 7 8 9 0

Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001

Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN,
Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
Custom editor ФИО Редактора
Custom edition 5



Если Автор (Переводчик) Книги один, то «отведите» для его данных один абзац. Если их несколько, то для каждого Автора (Переводчика) – свой отдельный абзац.

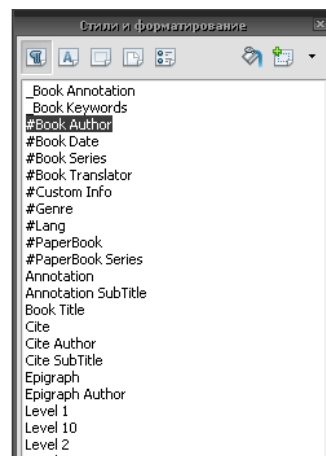
Рассмотрим примеры:

Для 2-х Авторы Книги:

Отмечаем «его» абзац с данными о нем стилем #Book Author

Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием».
Он состоит из 2-х абзацев с ссылкой на Часть 7, гиперссылкой и сноской.
Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 Email Автор2 HomePage
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 Email Переводчик2 HomePage
1 2 3 4 5 6 7 8 9 0
Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001
Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN,
Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
Custom editor ФИО Редактора
Custom edition 5

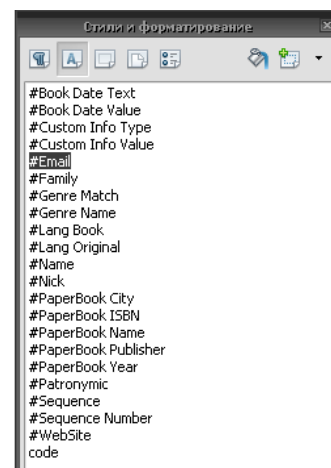


Теперь переключаемся на работу с символьными стилями, нажав кнопку «а» на диалоге «Стили и форматирование»

Имена всех Авторы отмечаем символьным стилем #Name, фамилии - #Family, отчества - #Patronymic и т.д.:

Часть 7. «Служебные стили» для описания свойств документа (секция Description)

Небольшой пример аннотации на книгу с «форматированием».
Он состоит из 2-х абзацев с ссылкой на Часть 7, гиперссылкой и сноской.
Автор1 Имя Автор1 Фамилия Автор1 Отчество Автор1 Ник
Автор2 Имя Автор2 Email Автор2 HomePage
Переводчик1 Имя Переводчик1 Фамилия Переводчик1 Отчество Переводчик1 Ник
Переводчик2 Имя Переводчик2 Email Переводчик2 HomePage
1 2 3 4 5 6 7 8 9 0
Дата написания Книги: Text: 2008, Value: 2008-03-31
Серия Электронной Книги: Серия Название 001
Данные Бумажной Книги: Название Бумажной Книги, Издательство, 2000, ISBN,
Город
Серия Бумажной Книги: Серия БК1 2
Серия Бумажной Книги: 1
Custom editor ФИО Редактора
Custom edition 5



Для Переводчиков Книги все делается аналогично.

Ну, и так далее...

3. Реальный пример

Например, есть текст отсканированного и распознанного документа (небольшой фрагмент с данными о бумажной книге и издательстве):

ББК 42.34 У 242

У 242 **Угарова Т. Ю.** Семейное овощеводство на узких грядках. 3-е изд., перераб. и доп.— М.: ИВЦ «Маркетинг», 1999.— 216 с.

ISBN 5-7856-0092-7

В книге отражен опыт использования метода Митлайдера в России. Подробно изложена технология овощеводства на узких грядках — высокопродуктивная, простая для освоения, предельно унифицированная для разных культур, требующая минимум времени и трудозатрат, эффективная на любых почвах и при любых погодных условиях. Технология адаптирована к почвенно-климатическим условиям России, выявлены особенности ее использования в мелком личном хозяйстве, апробированы эффективные удобрительные смеси, составленные на основе отечественных удобрений. Описаны особенности выращивания отдельных овощных культур на узких грядках, большое внимание уделено правильному подбору сортов и срокам посадки. Подчеркнуты элементы технологии, позволяющие снимать с одной гряды по два урожая овощей в год.

Для опытных огородников и новичков, всех тех, кто хочет обеспечить свою семью высококачественными овощами с 1,5-2 соток огорода, а также тех, кто хочет развернуть рентабельное товарное овощеводческое хозяйство на столь малой площади, как 10—15 соток.

ISBN 5-7856-0092-7 ББК 42.34

©Т. Ю. Угарова, 1999

Нам многое здесь не требуется (только то, что поддерживает форма fb2.1). Нужная нам информация выделена цветом (синим — аннотация на книгу, красным — данные о бумажной книге и издательстве):

ББК 42.34 У 242

У 242 **Угарова Т. Ю.** Семейное овощеводство на узких грядках. 3-е изд., перераб. и доп.— М.: ИВЦ «Маркетинг», 1999.— 216 с.

ISBN 5-7856-0092-7

В книге отражен опыт использования метода Митлайдера в России. Подробно изложена технология овощеводства на узких грядках — высокопродуктивная, простая для освоения, предельно унифицированная для разных культур, требующая минимум времени и трудозатрат, эффективная на любых почвах и при любых погодных условиях. Технология адаптирована к почвенно-климатическим условиям России, выявлены особенности ее использования в мелком личном хозяйстве, апробированы эффективные удобрительные смеси, составленные на основе отечественных удобрений. Описаны особенности выращивания отдельных овощных культур на узких грядках, большое внимание уделено правильному подбору сортов и срокам посадки. Подчеркнуты элементы технологии, позволяющие снимать с одной гряды по два урожая овощей в год.

Для опытных огородников и новичков, всех тех, кто хочет обеспечить свою семью высококачественными овощами с 1,5-2 соток огорода, а также тех, кто хочет развернуть рентабельное товарное овощеводческое хозяйство на столь малой площади, как 10—15 соток.

ISBN 5-7856-0092-7 ББК 42.34

©Т. Ю. Угарова, 1999

Поэтому «лишнюю» информацию можно удалить, а можно и оставить — она мешать не будет. «Нужную» информацию распределяем отдельно на каждой строке: Данные на бумажную Книгу — на своем абзаце, На Автора — на своем, на Электронную Книгу — на своем и т.д. При этом порядок расположения данных значения не имеет.

Вот что у нас получилось:

Данные на Бумажную Книгу:

Угарова Т. Ю. Семейное овощеводство на узких грядках. 3-е изд., перераб. и доп.— М.: ИВЦ «Маркетинг». 1999, ISBN 5-7856-0092-7

Серия Бумажной Книги: Овощеводство на узких грядках

Данные на Электронную Книгу:

Автор Книги: Угарова Татьяна Юрьевна

Серия Электронной Книги: Овощеводство на узких грядках

В книге отражен опыт использования метода Митлайдера в России. Подробно изложена технология овощеводства на узких грядках — высокопродуктивная, простая для освоения, предельно унифицированная для разных культур, требующая минимум времени и трудозатрат, эффективная на любых почвах и при любых погодных условиях. Технология адаптирована к почвенно-климатическим условиям России, выявлены особенности ее использования в мелком личном хозяйстве, апробированы эффективные удобрительные смеси, составленные на основе отечественных удобрений. Описаны особенности выращивания отдельных овощных культур на узких грядках, большое внимание уделено правильному подбору сортов и срокам посадки. Подчеркнуты элементы технологии, позволяющие снимать с одной гряды по два урожая овощей в год.

Для опытных огородников и новичков, всех тех, кто хочет обеспечить свою семью высококачественными овощами с 1,5-2 соток огорода, а также тех, кто хочет развернуть рентабельное товарное овощеводческое хозяйство на столь малой площади, как 10—15 соток.

Теперь применяем «служебные стили»:

К абзацу с данными бумажной книги — стиль абзаца #PaperBook, потом применяем символьные стили к Названию бумажной книги — #PaperBook Name, к Городу, где издали книгу — #PaperBook City, к Издателю — #PaperBook Publisher, к Году издания — #PaperBook Year, к ISBN — #PaperBook ISBN.

К Аннотации на книгу — стиль абзаца _Book Annotation.

Названия стилей — по умолчанию. Вы можете их назвать, как вам нравится, переопределив их (см. [ниже](#)).

Добавим еще абзацы с ФИО Автора, с Серией Книги и Серией Бумажной Книги. Применим к ним «их» стили.

Вот что у нас получилось:

Данные на Бумажную Книгу:

Угарова Т. Ю. Семейное овощеводство на узких грядках. 3-е изд., перераб. и доп. — М.: ИВЦ «Маркетинг», 1999, ISBN 5-7856-0092-7

Серия Бумажной Книги: Овощеводство на узких грядках

Данные на Электронную Книгу:

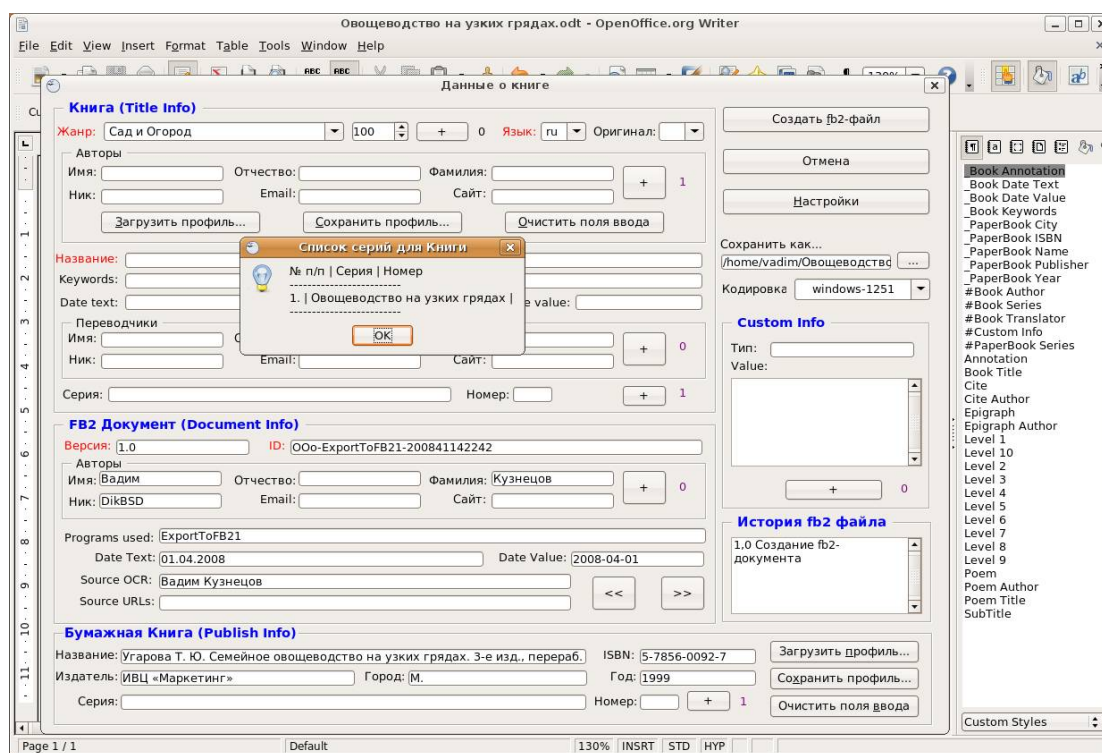
Автор Книги: Угарова Татьяна Юрьевна

Серия Электронной Книги: Овощеводство на узких грядках

В книге отражен опыт использования метода Митлайдера в России. Подробно изложена технология овощеводства на узких грядках — высокопродуктивная, простая для освоения, предельно унифицированная для разных культур, требующая минимум времени и трудозатрат, эффективная на любых почвах и при любых погодных условиях. Технология адаптирована к почвенно-климатическим условиям России, выявлены особенности ее использования в мелком личном хозяйстве, апробированы эффективные удобрительные смеси, составленные на основе отечественных удобрений. Описаны особенности выращивания отдельных овощных культур на узких грядках, большое внимание уделено правильному подбору сортов и срокам посадки. Подчеркнуты элементы технологии, позволяющие снимать с одной гряды по два урожая овощей в год.

Для опытных огородников и новичков, всех тех, кто хочет обеспечить свою семью высококачественными овощами с 1,5-2 соток огорода, а также тех, кто хочет развернуть рентабельное товарное овощеводческое хозяйство на столь малой площади, как 10—15 соток.

Теперь, после запуска конвертора, вся эта информация будет скопирована в соответствующие поля формы конвертора, при этом в *результатирующем fb2-файле* эти данные «окажутся в своих» тэгах, а в основном тексте документа их не будет:

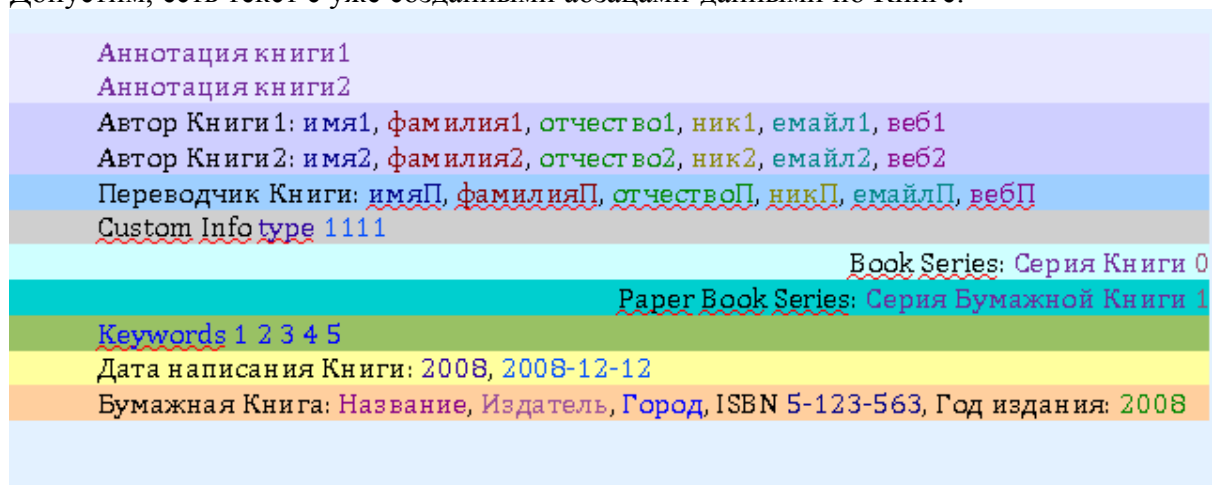


Здесь на картинке показано еще и окно сообщение «Список Серий для Книги», вызванное щелчком мышки по метке-счетчику для серий книги (справа от кнопки «+») — просто для контроля правильности работы инициализации конвертора.

Некоторые инструменты для «служебных стилей»

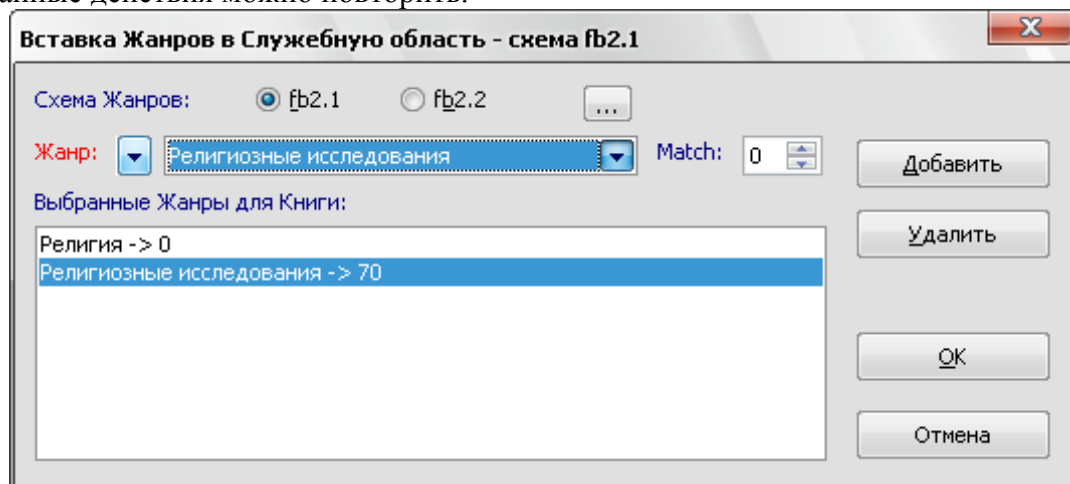
При задании некоторых служебных стилей очень легко сделать ошибку (Жанр Книги и Язык Книги). Конвертор игнорирует неверные данные – иначе fb2-файл будет невалидным (библиотеки не примут). Поэтому в пакете есть несколько инструментов для автоматической вставки 1 или нескольких Жанров и Языка Книги и Оригинала.

Допустим, есть текст с уже созданными абзацами-данными по Книге:



1. Вставка Жанра Книги

Для этого ставим курсор на новом абзаце и выбираем пункт меню OOoFBTools -> Insert Description Genre... Или нажимаем на одноименную кнопку на панели инструментов FictionBook Tools. В появившемся диалоге выбираем из списка Жанр книги, если надо – то и его match. Жмем кнопку «Добавить». Если Книга «подпадает под несколько жанров», то описанные действия можно повторить.



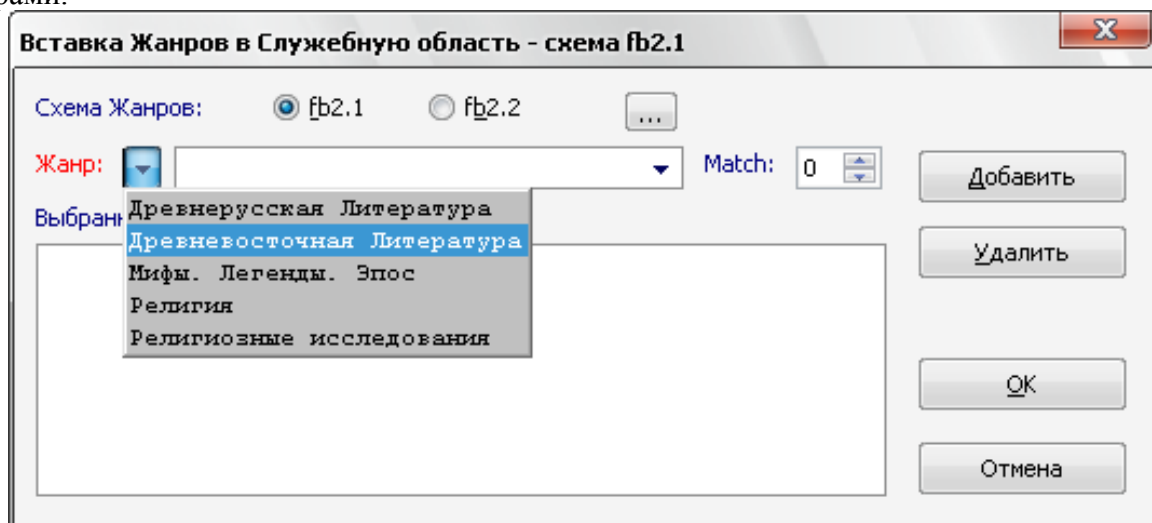
Теперь нажимаем «Ок» и вот что у нас получилось:

Аннотация книги1
Аннотация книги2
Автор Книги1: имя1, фамилия1, отчество1, ник1, email1, веб1
Автор Книги2: имя2, фамилия2, отчество2, ник2, email2, веб2
Переводчик Книги: имяП, фамилияП, отчествоП, никП, emailП, вебП
Custom Info type 1111
Book Series: Серия Книги 0
Paper Book Series: Серия Бумажной Книги 1
Keywords 1 2 3 4 5
Дата написания Книги: 2008, 2008-12-12
Бумажная Книга: Название, Издатель, Город, ISBN 5-123-563, Год издания: 2008
Genre of Book: Религия, Match: 0
Genre of Book: Религиозные исследования, Match: 70

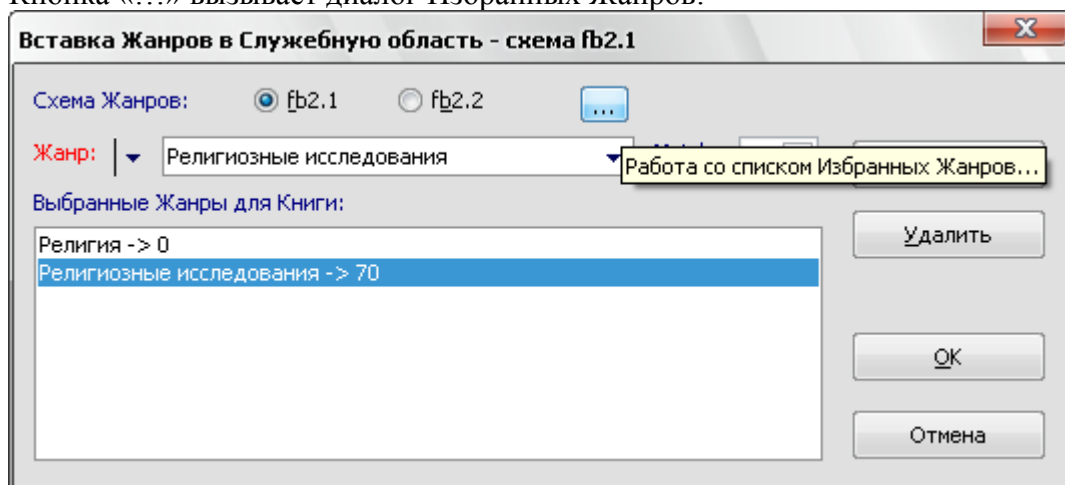
Match = 0 при экспорте будет конвертером игнорироваться.

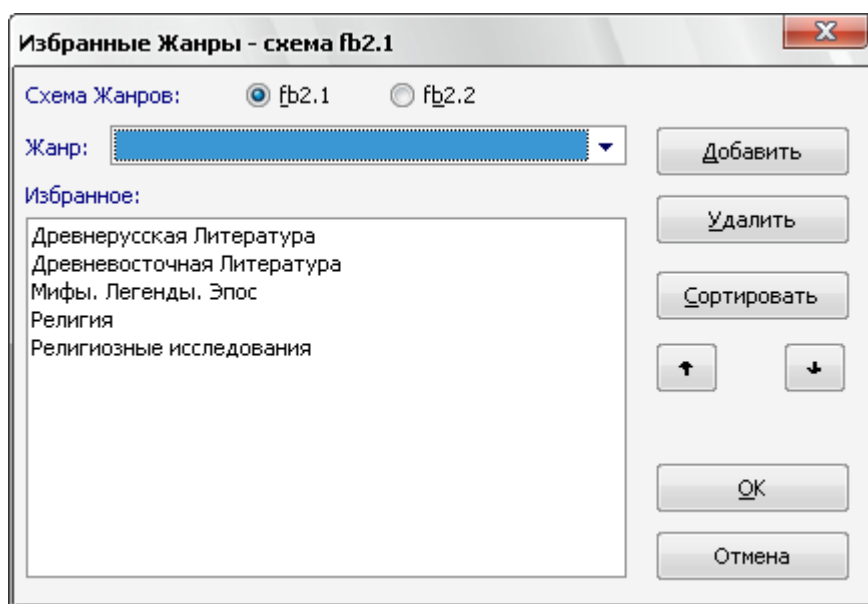
Слова “Genre of Book:” и “Match:” не отмечены никаким символьным служебным стилем и просто игнорируются конвертером. Они добавлены для информативности.

В Инструменте Вставки Жанра Книги сделана возможность работы с Избранными Жанрами.



Кнопка «...» вызывает диалог Избранных Жанров:

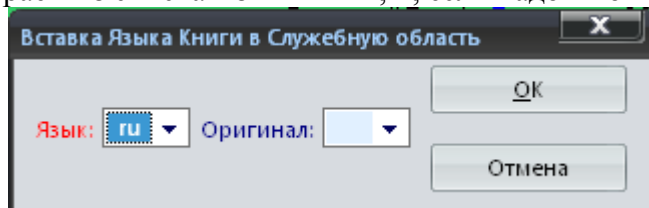




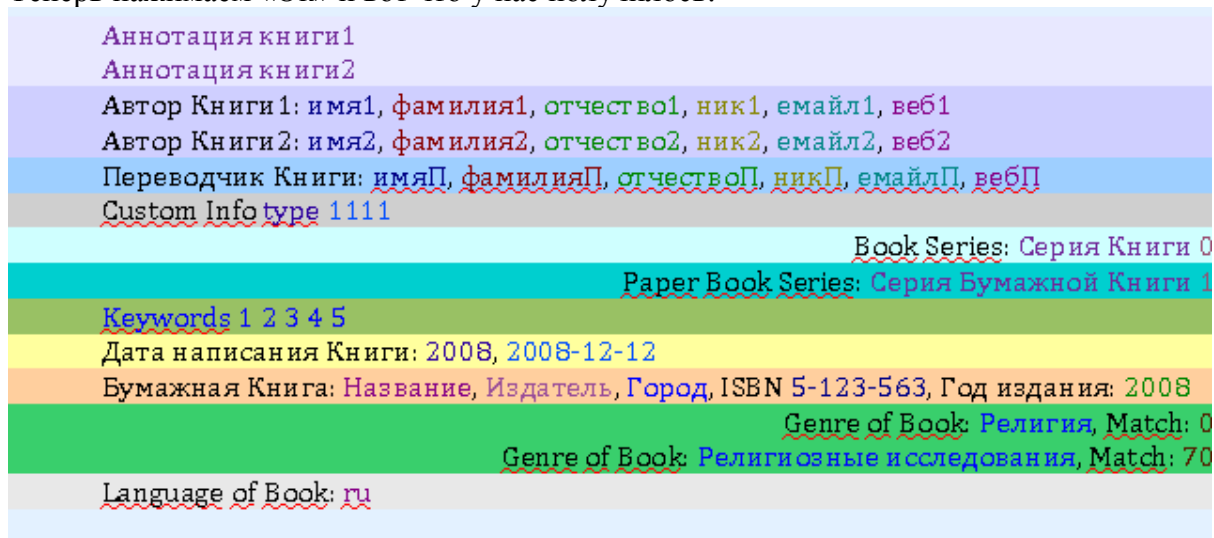
Как работать с Избранными Жанрами см. [здесь](#).

2. Вставка Языка Книги

Для вставки Языка Книги почти то же самое. Ставим курсор на новом абзаце и выбираем пункт меню OOoFBTools -> Insert Description Lang... Или нажимаем на одноименную кнопку на панели инструментов FictionBook Tools. В появившемся диалоге выбираем из списка Язык Книги, и, если надо – то и Язык оригинала Книги.



Теперь нажимаем «Ок» и вот что у нас получилось:



Если был выбран и Язык оригинала, то результат будет немного другой:

Аннотация книги1	
Аннотация книги2	
Автор Книги1: имя1, фамилия1, отчество1, ник1, email1, веб1	
Автор Книги2: имя2, фамилия2, отчество2, ник2, email2, веб2	
Переводчик Книги: имяП, фамилияП, отчествоП, никП, emailП, вебП	
Custom Info type 1111	
	Book Series: Серия Книги 0
	Paper Book Series: Серия Бумажной Книги 1
Keywords 1 2 3 4 5	
Дата написания Книги: 2008, 2008-12-12	
Бумажная Книга: Название, Издатель, Город, ISBN 5-123-563, Год издания: 2008	
	Genre of Book: Религия, Match: 0
	Genre of Book: Религиозные исследования, Match: 70
Language of Book: ru, Original: en	

Слова “Language of Book:” и “Original:” не отмечены никаким символьным служебным стилем и просто игнорируются конвертером. Они добавлены для информативности, чтобы пользователю не запутаться и не забыть, где Язык оригинала, а где – электронной книги.

3.7. Переопределение стилей по умолчанию на свои стили

Если вам удобнее иметь свои стили с **другими** названиями, нежели заложенные в стилях «по умолчанию», то эти стили легко можно переопределить на свои стили (со своими названиями). Например, вместо названия стилей «Level 1»... «Level 10» вы хотите использовать «Заголовок 1» ... «Заголовок 10», и т.д.

С другой стороны, переопределение стилей может пригодиться и в такой ситуации: Вы открыли в ООo Writer документ «из-под» Microsoft Word. А он отформатирован своими стилями (Заголовок 1... Заголовок N и т.д.). Конечно, можно заменить все стили с помощью диалога поиска и замены. Но можно просто *переопределить* стили на нужные (не производя никаких замен внутри самого документа) и запустить конвертор.

Как запускать диалог Определения стилей пользователя?

Создайте на уже созданной ранее панели инструментов (в моем случае FictionBook) или на любой удобной для вас кнопку для переопределения стилей из макросов по следующему пути: Макрос **OpenOffice.org / Мои макросы / OOoFBTools / StylesControls**. Нужный здесь макрос – **SetUserStyles**. Как вытаскивать макросы на кнопки подробно описано в этой справке в [п.2.2](#). Сделайте по аналогии.

Диалог выглядит так – для «Основных» стилей:

Определение пользовательских стилей

☒ Основные стили ☐ "Служебные" стили

Данные о книге
 Название Книги: Book Title

Анотация
 Анотация: Annotation
 Подзаголовок: Annotation SubTitle

Эпиграф
 Эпиграф: Epigraph
 Автор эпиграфа: Epigraph Author

Стих
 Заголовок поэмы: Poem Title
 Поэма: Poem
 Автор поэмы: Poem Author
 Дата поэмы: Poem Date
 Подзаголовок: Poem SubTitle

Подзаголовок
 Subtitle: SubTitle

Заголовки (Уровни)
 Заголовок 1: Level 1
 Заголовок 2: Level 2
 Заголовок 3: Level 3
 Заголовок 4: Level 4
 Заголовок 5: Level 5
 Заголовок 6: Level 6
 Заголовок 7: Level 7
 Заголовок 8: Level 8
 Заголовок 9: Level 9
 Заголовок 10: Level 10

Цитата
 Цитата: Cite
 Автор цитаты: Cite Author
 Подзаголовок: Cite SubTitle

Принять

По умолчанию

Отмена

Загрузить профиль...

Сохранить профиль...

Отдельно для каждого вида стилей

Для «Служебных стилей»:

Определение пользовательских стилей

☐ Основные стили ☒ "Служебные" стили

Для всего абзаца
 Аннотация Книги: _Book Annotation
 Ключевые слова (Keywords): _Book Keywords
 Автор Книги: #Book Author
 Переводчик Книги: #Book Translator
 Серия Книги: #Book Series
 Серия Бумажной Книги: #PaperBook Series
 Дата Написания Книги: #Book Date
 Бумажная Книга: #PaperBook
 Custom Info: #Custom Info
 Жанр: #Genre
 Язык: #Lang

Inline стили (символьные)
Для Бумажной книги
 Название: #PaperBook Name
 Издательство: #PaperBook Publisher
 Город: #PaperBook City
 Год: #PaperBook Year
 ISBN: #PaperBook ISBN

Inline стили (символьные)
Авторы и Переводчики Книги
 Имя: #Name
 Отчество: #Patronymic
 Фамилия: #Family
 Ник: #Nick
 Email: #Email
 Номерpage: #WebSite

Для Серии
 Серия (Название): #Sequence
 Серия (№): #Sequence Number

Для даты Написания Книги
 Book Date (Text): #Book Date Text
 Book Date (Value): #Book Date Value

Для Custom Info
 Custom Info (Type): #Custom Info Type
 Custom Info (Value): #Custom Info Value

Для Жанра
 Жанр: #Genre Name
 Match: #Genre Match

Для Языка Книги
 Язык: #Lang Book
 Оригинал: #Lang Original

Принять

По умолчанию

Отмена

Загрузить профиль...

Сохранить профиль...

Отдельно для каждого вида стилей

Сохраненные стили автоматически загружаются с запуском конвертора, и экспорт документа происходит в соответствии с ними. Их не нужно каждый раз загружать вручную!

Кнопка «По умолчанию» работает для каждого вида стилей отдельно. Т.е., если вы находитесь на вкладке «Основные стили» и нажали «По умолчанию», то поля только для этих стилей заполнятся данными по умолчанию. Поля для стилей «Служебные стили» останутся неизменными. И наоборот. Так удобнее.

3.7.1. Создание набора пользовательских стилей

В диалоге Определения пользовательских стилей есть две кнопки: «Загрузить профиль...» и «Сохранить профиль...» для загрузки и сохранения стилей в поля формы.

Часто работаешь с уже «готовыми», т.е. отформатированными стилями документами.

Можно вручную в тексте документа заменять эти стили на fb2 стили «по умолчанию». Можно в этом диалоге переопределить эти стили на «ваши». А можно сделать еще проще и удобнее: сделать разные профили стилей для разных типов документов, уже отформатированных «своими» стилями – для Microsoft Word, для html файлов и т.д. Когда открываешь определенный документ, просто загружаешь для него «его родные» стили, и все!

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.

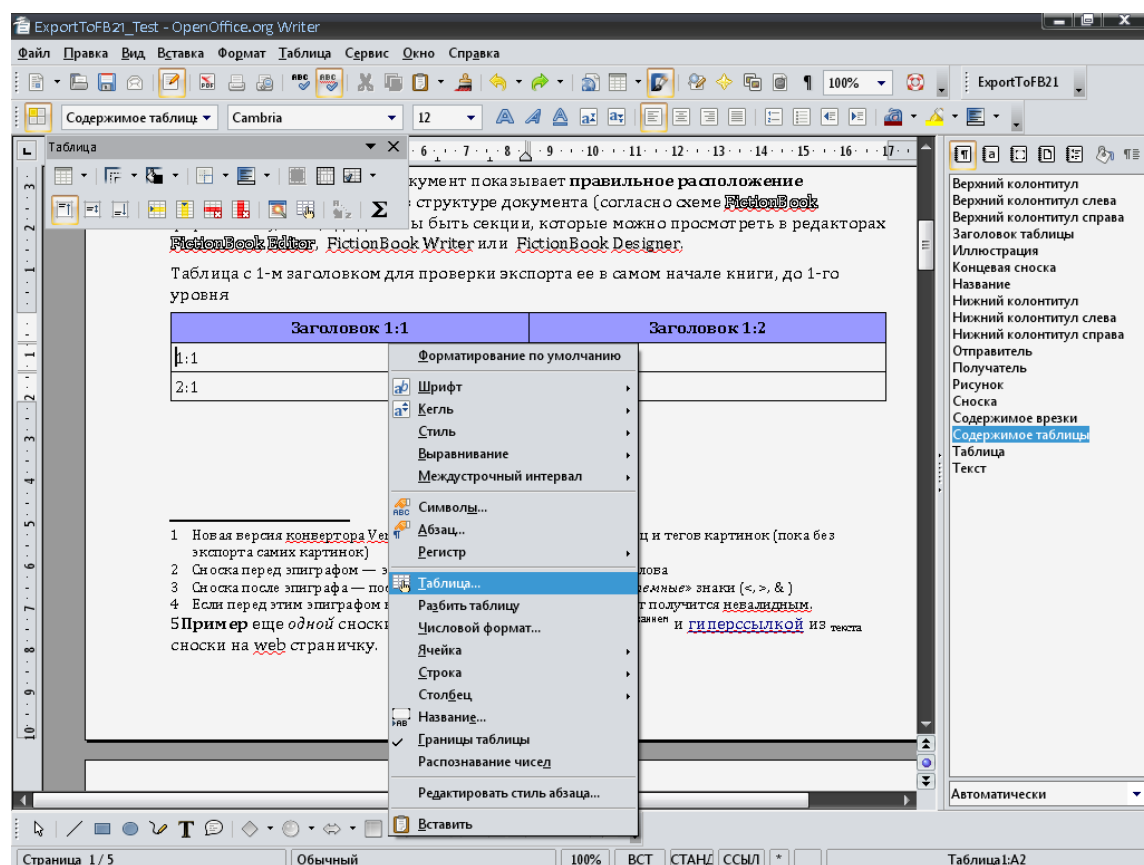
Конвертор корректно экспортирует **только простые** таблицы. Экспортируются и сложные таблицы, но не все сложные таблицы экспортируются согласно их «геометрии». Сложные таблицы, где в ячейку вложена другая таблица – игнорируются (ввиду чрезвычайной сложности создания алгоритма...).

4.1.2. Название таблицы

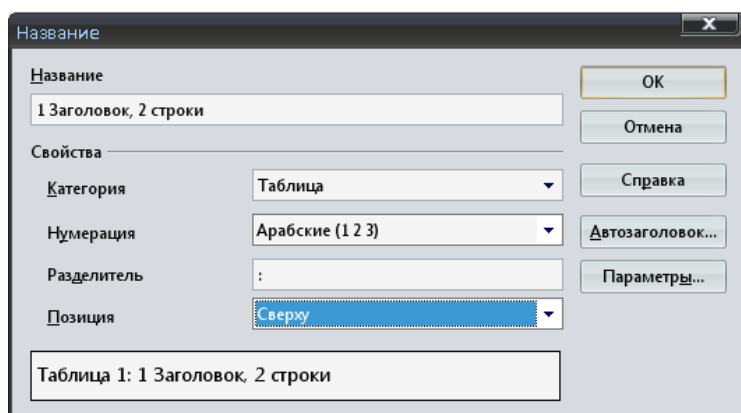
При желании можно дать таблице название (сверху или снизу таблицы). Это название также будет экспортироваться, с учетом форматирования.

Делается это так.

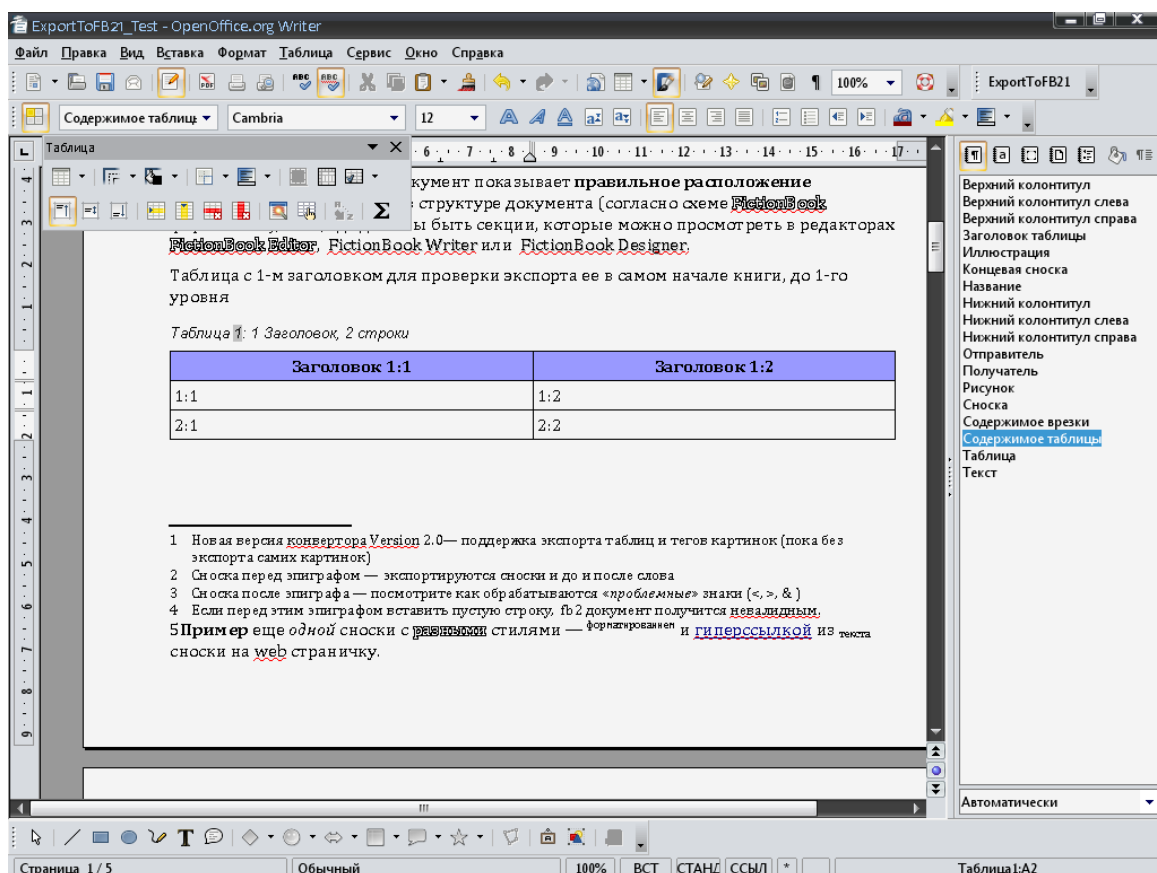
На нужной таблице кликните правой клавишей мышки и выберите «Название».



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



В результате получим следующее:



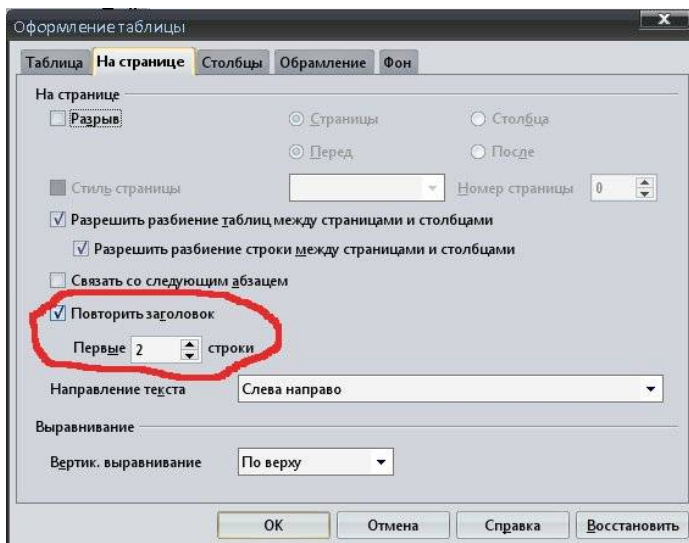
В названии имеется текстовое поле с сквозной нумерацией таблицы – оно тоже экспортируется.

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.

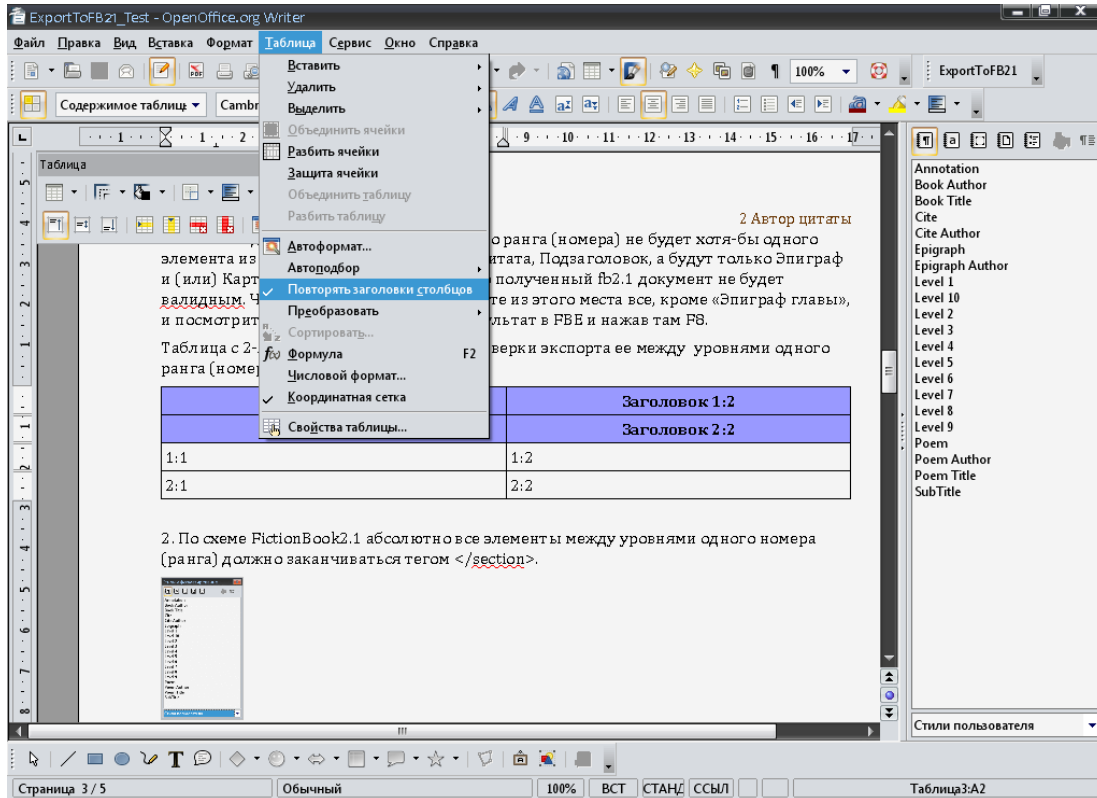
Для этого вызываем диалог «Оформление таблицы» через меню «Таблицы», нажав

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.

Атрибуты align, valign для тегов tr, th, td таблицы сделаны «не обязательными». Т.е. в Настройках конвертора их можно отключить (выбрав из списков пустое значение). В этом случае эти атрибуты в fb2 файл заноситься не будут. Это же реализовано и для «пустых» ячеек таблицы (нет ни одного абзаца), даже при выбранном выравнивании строк в Настройках. Выравнивание **по ширине** в любом случае «считывается» со строк таблиц автоматически.

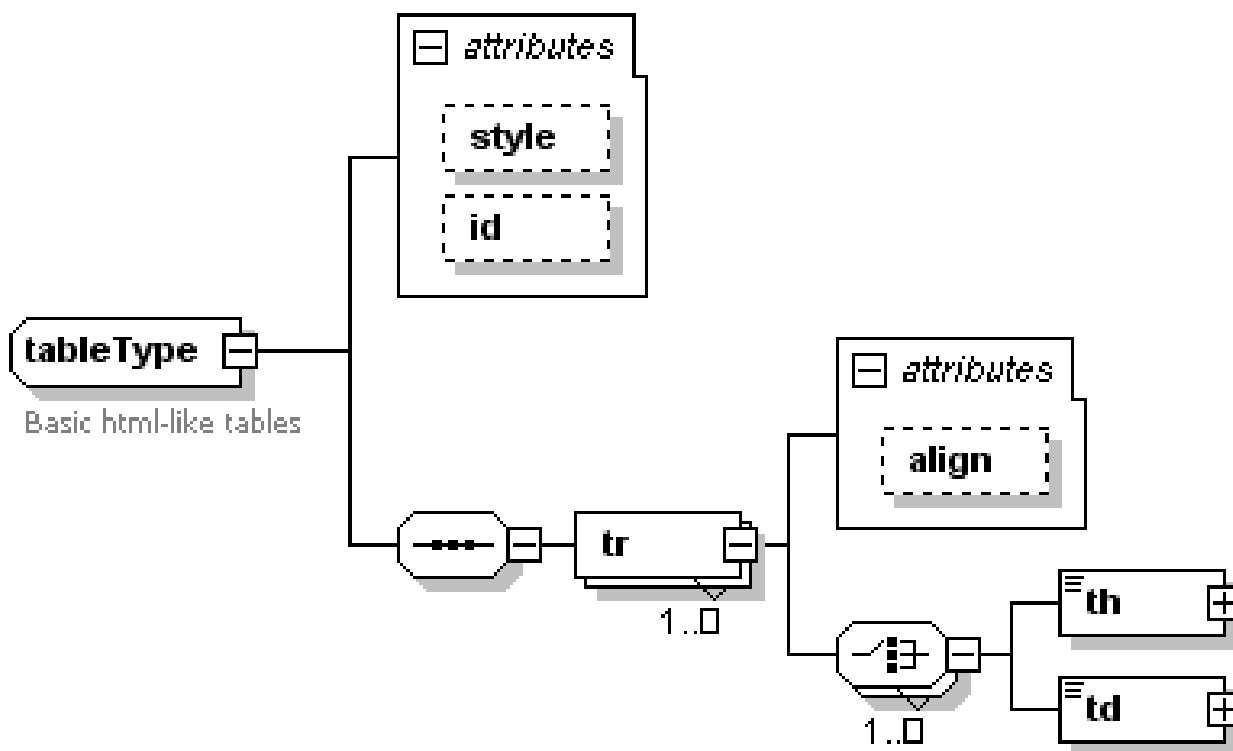
4.1.6. «Пустые» ячейки таблицы

Иногда требуется, чтобы в таблице были ячейки с «пустым» содержанием. Конвертор позволяет экспортировать и такие ячейки таблицы. После экспорта в fb2 документ на месте этой «пустой» ячейки будет <th/> (для «пустого» заголовка) или <td/> (для «пустой» ячейки).

4.1.7. Структура элементов таблицы

Таблица формата fb2 имеет следующую структуру:

4.1.7.1. Тэг table



В тэг *table* вложен тэг *tr* - «контейнер» для строк заголовков (тэг *th*) или данных (тэг *td*) всех столбцов ячеек таблицы. Тэг *table* может содержать от 1 до бесконечности

обязательных тэгов *tr*. В свою очередь, тэг *tr* может содержать от 1 до бесконечности обязательных тэгов *th* или *td*.

Например, есть такая таблица:

Заголовок 1:1	Заголовок 1:2	Заголовок 1:3
2:1	2:2	2:3

В fb2 формате это выглядит так:

```
<table>
  <tr>
    <th align="left">Заголовок 1:1</th>
    <th align="center">Заголовок 1:2</th>
    <th align="right">Заголовок 1:3</th>
  </tr>
  <tr>
    <td align="left">2:1</td>
    <td align="right">2:2</td>
    <td align="center">2:3</td>
  </tr>
</table>
```

4.1.7.2. Тэг *th*

Тэг *th* так же имеет «свои» атрибуты (см. схему ниже) и содержит «отформатированные» символьными (inline) стилями данные-заголовки ячейки таблицы.

Тэг *th* может содержать от 0 до бесконечности «отформатированных» символьными стилями заголовков ячеек таблицы. Т.е. Тэг *th* может быть «пустым», закрытым:

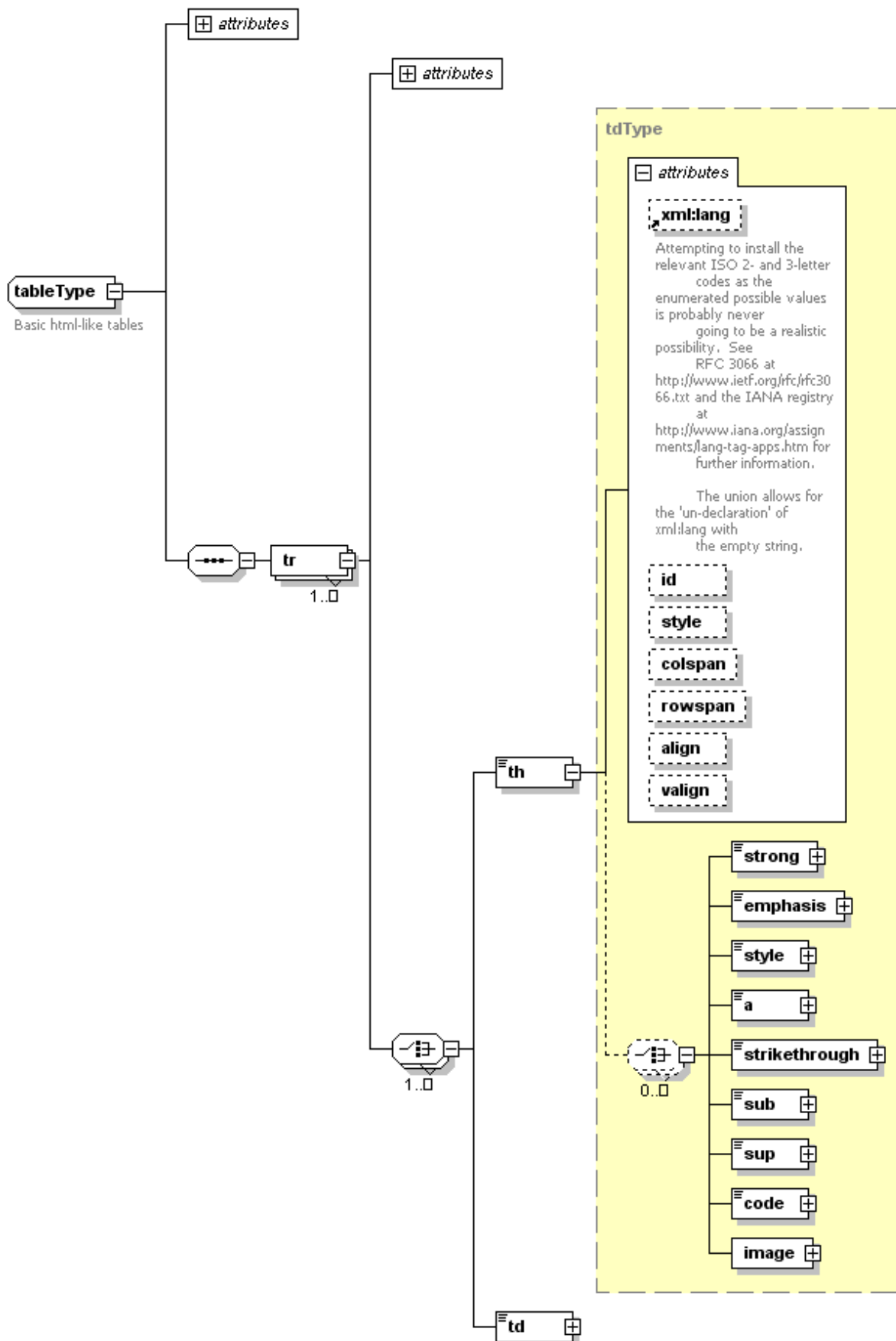
```
<table>
  <tr>
    <th/>
    <th> </th>
  </tr>
</table>
```

А может и содержать заголовки ячеек:

```
<table>
  <tr>
    <th align="left">Заголовок 1:1</th>
    <th align="center">Заголовок 1:2</th>
  </tr>
</table>
```

Атрибуты *align* и *valign* служат для выравнивания заголовка *по ширине* и *по высоте* соответственно. Тэг имеет еще и другие атрибуты.

Схема тэга дана *th* ниже:



4.1.7.3. Тэг `td`

Тэг `td` так же имеет «свои» атрибуты (см. схему ниже) и содержит «отформатированные» символьными (inline) стилями данные-строки ячейки таблицы.

Тэг `td` может содержать от 0 до бесконечности «отформатированных» символьными стилями строк ячеек таблицы. Т.е. Тэг `td` может быть «пустым», закрытым:

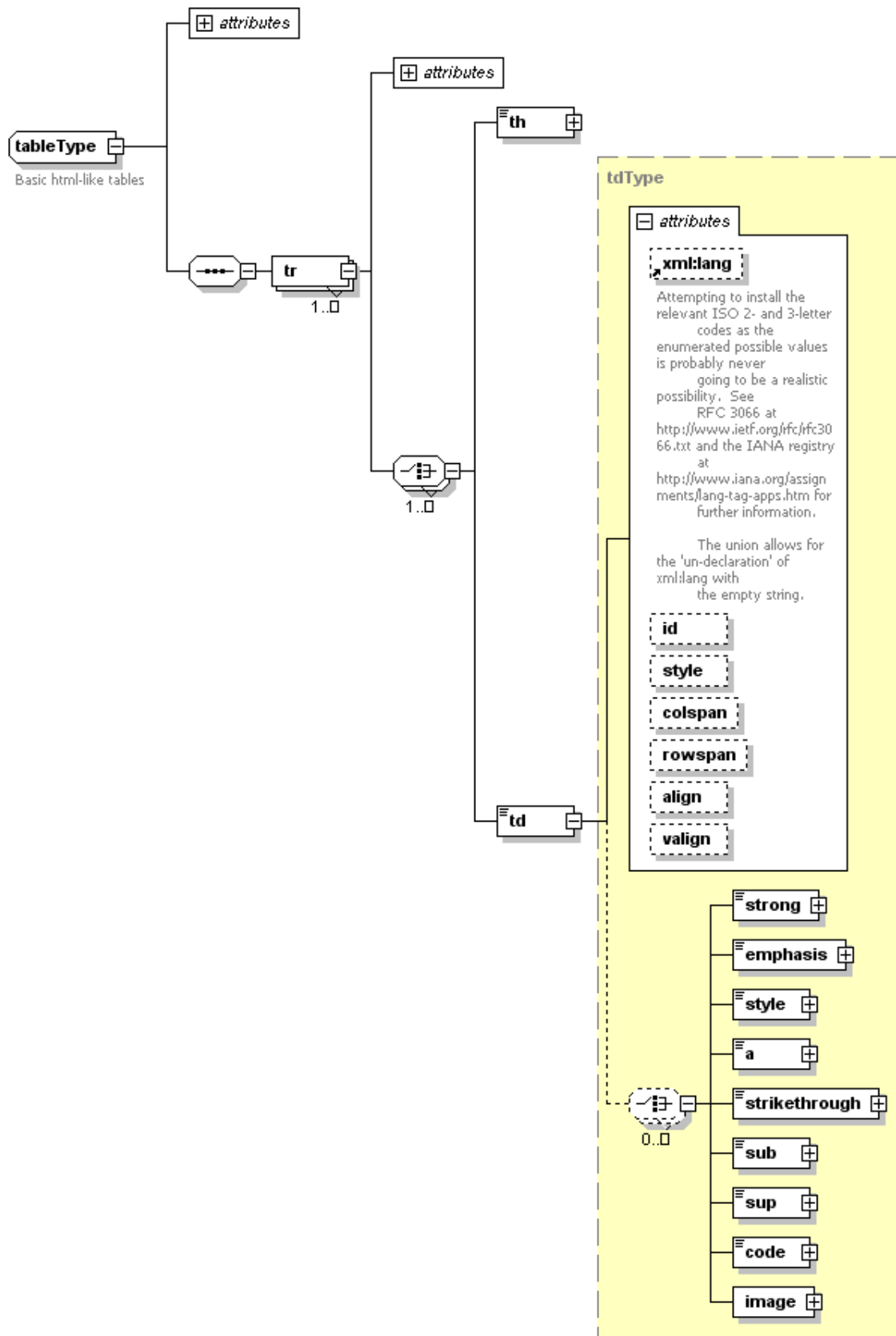
```
<table>
  <tr>
    <td/>
    <td> </td>
  </tr>
</table>
```

А может и содержать заголовки ячеек:

```
<table>
  <tr>
    <td align="left">2:1</td>
    <td align=" right">2:2</td>
  </tr>
</table>
```

Атрибуты *align* и *valign* служат для выравнивания строки *по ширине* и *по высоте* соответственно. Тэг имеет еще и другие атрибуты.

Схема тэга `td` дана ниже:



4.1.8. Экспорт таблиц, ячейки которых могут содержать разное число абзацев

Формат FictionBook поддерживает и простые таблицы, по типу «прямоугольных» матриц, и некоторые сложные таблицы, где строки или столбцы могут быть разбиты. Простая таблица будет экспортирована «на ура» (число строк и колонок – не ограничено):

1:1	1:2
2:1	2:2
3:1	3:2

Более сложные таблицы тоже экспортируются:

1:1		
2:1	2:2	2:3
3:1	3:2	

Формат fb2 **не поддерживает** таблицы, **вложенные в ячейки другой таблицы** (что практически всегда можно увидеть на html файлах, скаченных с сайтов, форумов):

1:1	1:2	
2:1	2:2	
A		
C		
3:1	3:2	

С другой стороны, согласно формату fb2 каждый тэг *th* и *td* может содержать *только один абзац*. Т.е. такая таблица:

1:1(1) 1:1(2)	1:2
2:1	2:2
3:1	3:2(1) 3:2(2)

«превратится» в fb2 формате в такую (абстрактно):

1:1(1) 1:1(2)	1:2
2:1	2:2
3:1	3:2(1) 3:2(2)

Т.е. все абзацы в ячейке «сливаются» в одну строку.

А таких таблиц можно встретить довольно часто (техническая, научная и др. литература). Как же сконвертировать ТАКИЕ «прямоугольные» таблицы в fb2 файл?

Конвертор **позволяет** это сделать за счет добавления так называемых «виртуальных» строк и ячеек. Т.е. все абзацы в ячейке (например 5 абзацев) становятся строками (5 строк).

Например, у нас есть такая таблица с **несколькими абзацами в ячейках**, а некоторые ячейки **пустые**:

Заголовок 1:1(1) Заголовок 1:1(2) Заголовок 1:1(3)	Заголовок 1:2
2:1(1) 2:1(2)	
3:1	3:2(1) 3:2(2) 3:2(3)

Конвертор создаст fb2 код таблицы, которая выглядеть в читалках будет так:

Заголовок 1:1(1)	Заголовок 1:2
Заголовок 1:1(2)	
Заголовок 1:1(3)	
2:1(1)	
2:1(2)	
3:1	3:2(1)
	3:2(2)
	3:2(3)

Т.е., вместо 3 строк мы получили 8, разбив «многоабзацевые» ячейки. Это и есть так называемые «виртуальные» строки.

Чтобы все это «пощупать», сконвертируйте тестовый файл **ExportToFB21_Test.odt** и посмотрите результат (fb2 код в редакторе и «внешний вид» таблиц в читалке). Для Windows лучше всего (на сегодняшний день) отображает таблицы с выравниванием ячеек **Cool Reader 2.66**. Она корректно «отбивает» колонки таблиц. Выберите в ней в настройках скин *default* для того, чтобы увидеть «бордюры» ячеек.

4.1.8.1. Ячейки: экспорт абзацев, как строк и экспорт абзацев как одну строку

Конвертор по умолчанию экспортирует многоабзацевые ячейки, как абзацы. Но, иногда надо, чтобы абзацы определенной ячейки сконвертировались, как одна строка (слились через пробел).

Например, есть таблица:

Tip Совет	Чтобы найти то-то, пойдите туда-то. Если это не помогло, то пойдите туда-то. Ну, если и это не помогло, то мне вас жаль...
---------------------	--

В 1-й ячейке два абзаца по сути дела выражают одно и то же. После экспорта мы получим такую таблицу:

Tip	Чтобы найти то-то, пойдите туда-то.
Совет	Если это не помогло, то пойдите туда-то.
	Ну, если и это не помогло, то мне вас жаль...

Эта таблица выглядит хорошо. Но вот если во втором столбце будет всего один абзац, то тогда после экспорта таблица будет выглядеть уже так:

Tip	Чтобы найти то-то, пойдите туда-то.
Совет	

Такая таблица уже «не очень»: Сразу непонятно, к чему относится «Совет» - напротив него – пустая ячейка. Было бы значительно лучше, если бы таблица выглядела так:

Tip Совет	Чтобы найти то-то, пойдите туда-то.
------------------	-------------------------------------

Как раз для этих целей в конверторе и предусмотрен режим экспорта всех нужной ячейки как слившиеся через пробел абзацы в одну строку. Для этого надо «заблокировать» нужную ячейку: клик правой клавишей мышки на нужной ячейке, выбираем **Ячейка -> Защищено**.

В этом случае наша первоначальная таблица будет выглядеть так:

Tip Совет	Чтобы найти то-то, пойдите туда-то.
	Если это не помогло, то пойдите туда-то.
	Ну, если и это не помогло, то мне вас жаль...

4.1.9. Экспорт сложных таблиц

Пока конвертор может экспортировать не все виды сложных таблиц, из тех, которые поддерживаются форматом fb2.

Такая таблица экспортируется корректно:

1:1	
2:1	2:2
3:1	
4:1	
5:1	5:2

Т.е. сложные таблицы, где *строки* могут содержать *разное число столбцов* экспортируются.

Если в такой таблице не все строки будут иметь одинаковое число столбцов, то она экспортируется не совсем «геометрически» правильно. Т.е такая таблица:

1:1		
2:1	2:2	2:3
3:1		
4:1		
5:1	5:2	

после экспорта в CoolReader 2.66 или CoolReader 3.0.15 и выше будет выглядеть так (геометрически):

1:1		
2:1	2:2	2:3
3:1		
4:1		
5:1	5:2	

т.е. ячейка 5:2 не растянута до конца строки. Эта «некрасивость» будет, надеюсь, исправлена (после изучения SDK по OOo)

Таблицы, где *ячейка* может содержать *разно число строк* – игнорируются (из-за сложности разработки алгоритма). Т.е. такая таблица будет проигнорирована:

1:1		
2.1.1	2.1.2	2:2

4.1.10. Экспорт Графики (Картинки, Формулы и Диаграммы) из ячеек Таблицы

Согласно схеме fb2 картинки в таблице fb2-файла могут быть только как in-line картинки внутри тэга <th> или <td>. Например: <td align="left"><image l:href="#img_2.png"/></td>. Конвертер распознает привязку Графики **только** как «К Абзацу» «К символу» и «Как символ». Другие – *игнорируются* и Графика *не будет экспортирован!* Будьте внимательны!

4.2. Работа с Текстовыми 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

Экспорт для Текстовых Врезок сделан полностью:

1. Учитывается опция «Удаление пустых абзацев»,
2. Экспорт нумерованных списков из Текстовых Врезок.
3. Экспорт маркированных списков из Текстовых Врезок, в том числе и маркеро-картинок, как картинок.

4. Экспортируются символьные стили для слов Врезок (гиперссылка, переход на другое место документа, курсив, верхний регистр ...). Работает выбор режима экспорта символьных стилей.

5. Экспорт Графики (Картинки, Формулы и Диаграммы) из Врезок. Распознается привязка Графики только как «К Абзацу», «К символу» и «Как символ». Другие – игнорируются и тэг Графика не будет экспортирован!

6. Экспорт Таблиц из Врезок.

Вложенные Врезки во Врезки игнорируются!

Возможны 4 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).

4. Как Сложную структуру текста, «обрамленного» Абзацевыми Стилями (Поэма, Цитата...).

Первые три режима позволяют конвертировать текст врезки соответственно в виде Таблицы, Текста и Цитаты. При этом Абзацевые стили (Поэма, Цитата...) игнорируются. Для режима «Таблица» при экспорте учитывается *выравнивание абзацев Врезки по ширине*. 4-й режим – для конвертации Врезки в виде Структуры. Работают все стили Абзацев! Т.е. текст после соответствующего «обрамления» стилями можно экспортировать, как Цитаты, Сложные Стихи, Уровни (Заголовки), Подзаголовки...

Совет: В 4-м режиме будьте очень внимательны! Если врезка – не в начале документа, и текст оформлен, как Аннотация или Эпиграф – книга будет **невалидной**! Поэтому, этот режим лучше все использовать для экспорта Сложных Стихов. Цитат, Подзаголовков, Заголовков (Уровней)... Хотя для этого удобнее вообще не пользоваться Врезками...

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

4.3. Работа с Картинками и другой графикой

4.3.1. Допустимые виды графики (Картинки, Формулы, Диаграммы)

fb2-читалки, обычно, «понимают» только png и jpeg форматы, закодированные в base64. Редкие – еще и gif... А в экспортируемом документе могут «содержаться» и bmp, tif, gif и другие картинки, растровых и векторных форматов. Часто встречаются и формулы и диаграммы (их тоже хотелось бы «видеть» в fb2-файле).

Конвертер позволяет конвертировать *все типы* картинок, которые могут быть загружены в ООо документ, а также **формулы** и **диаграммы**. При этом, картинки «понимаемого» читалками форматов (png и jpeg) кодируются в base64 как png и jpeg соответственно. Для всех остальных типов картинок, а так же для формул и диаграмм предусмотрен следующий алгоритм: сначала идет преобразование их в выбранный вами в Настройках формат (png, jpg или jpeg), и только потом они кодируются в base64. Выбор формата для преобразования позволяет «играть» размерами fb2-файла (разные картинки – цветные или черно-белые имеют разный размер в png и jpeg). Все «не родные» для fb2-читалок форматы условно разбиты на следующие:

- gif (очень часто встречается в html-документах);
- tif, tiff, bmp («тяжеловесные растры»);
- другие растровые и векторные форматы;
- формулы и диаграммы.

Для каждой из этих групп можно выбрать формат преобразования: png, jpg или jpeg. Описание см. в разделе данной справки, посвященной [Настройкам](#) Конвертирования.

4.3.2. Привязка картинок (формул, диаграмм) к тексту

Всего существует 4 вида картинок:

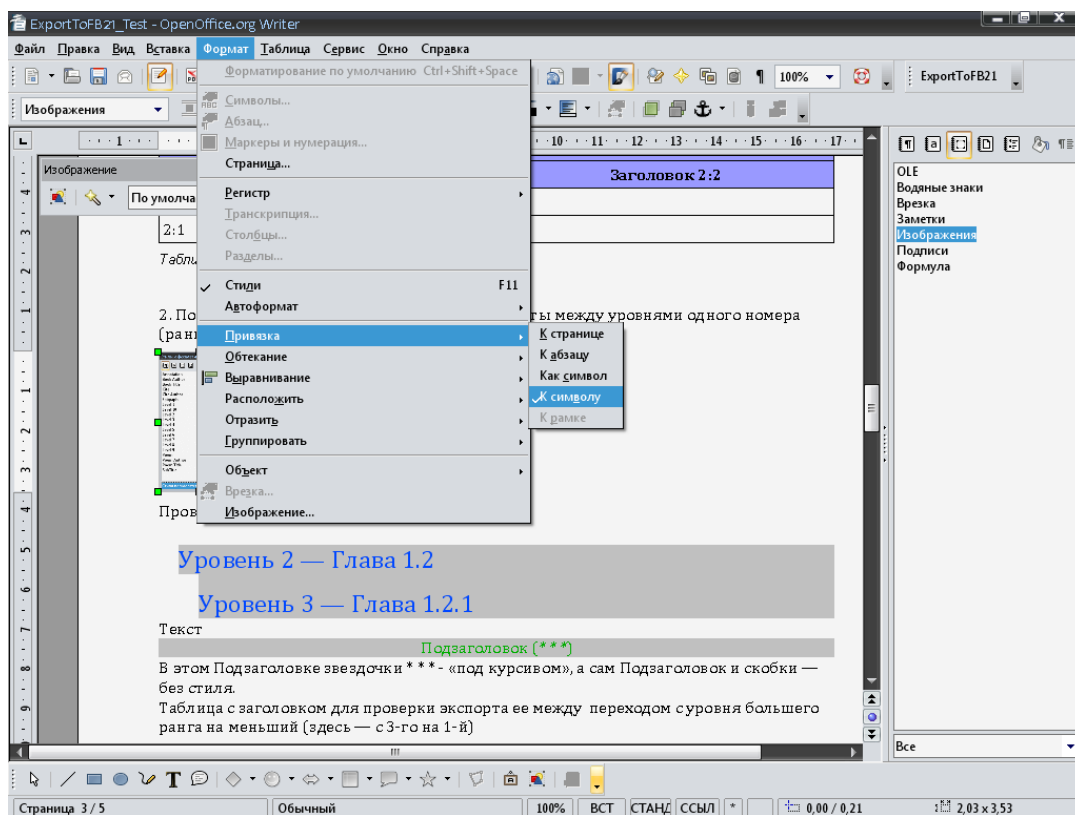
1. С привязкой к странице.
2. С привязкой к абзацу.
3. С привязкой к символу.
4. С привязкой как символ.

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).

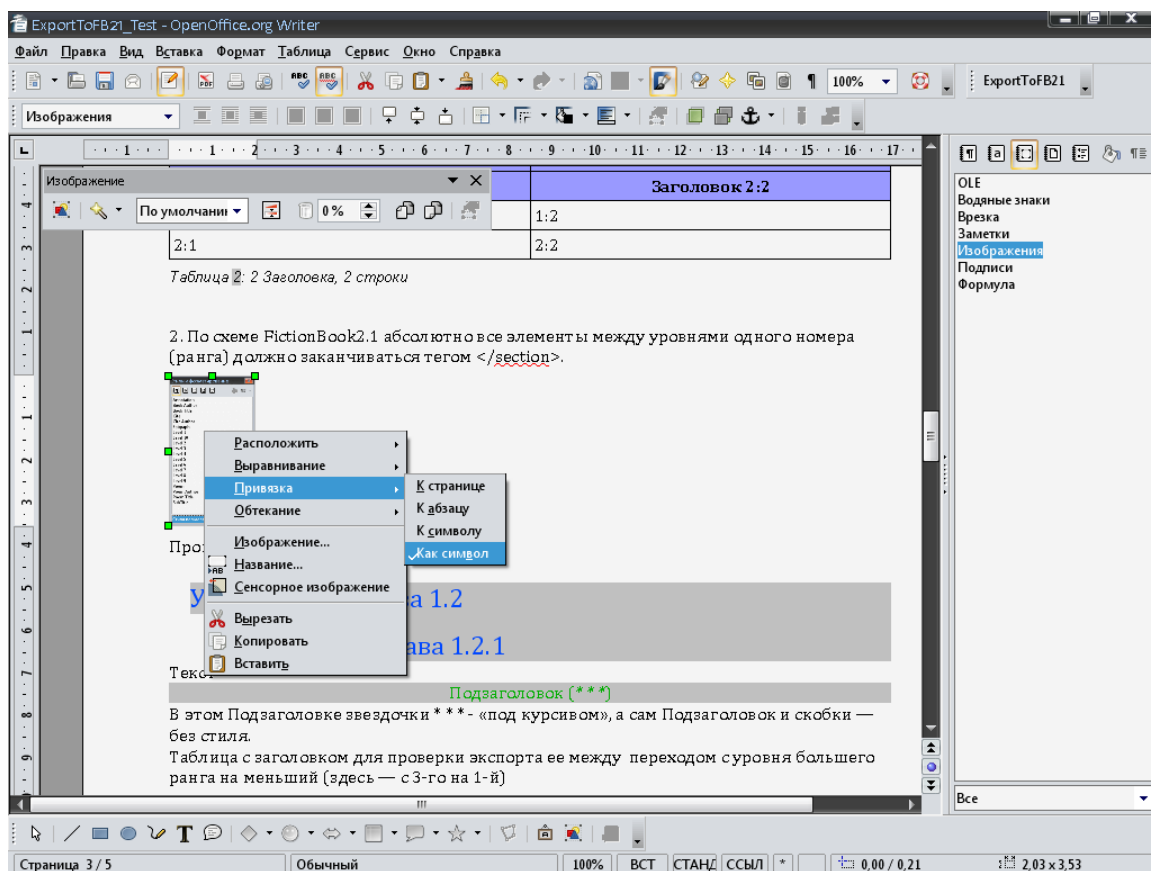
1-й вариант – картинка с привязкой к странице «не отлавливается» конвертором, и картинка будет проигнорирована. Картинки с названиями во врезках тоже пока «не отлавливаются», и экспортировано будет только их название из врезок в виде текста. Поэтому не используйте эти два вида картинок! Если надо задать 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.

Нужную привязку картинки можно осуществить несколькими способами.

4.3.2.1. Через меню «Формат» → «Привязка»

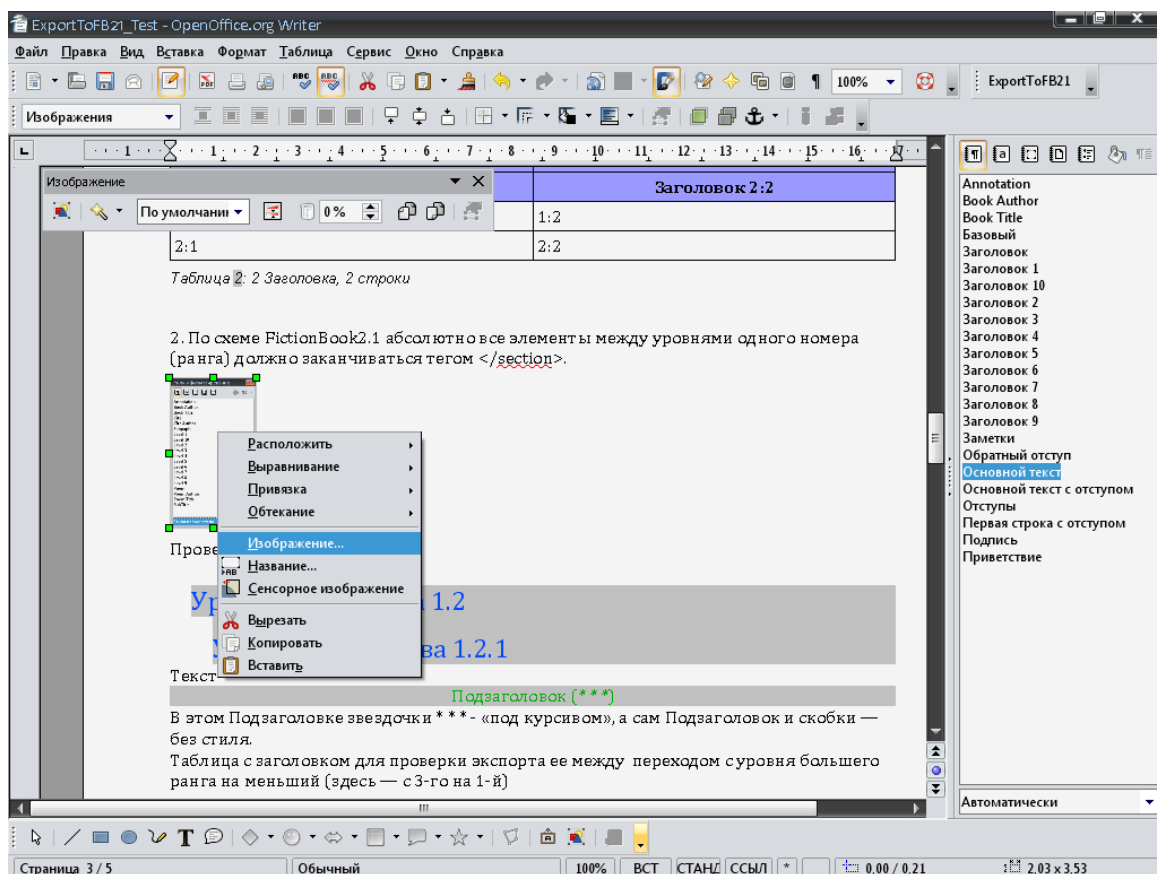


4.3.2.2. Через контекстное меню «Привязка»

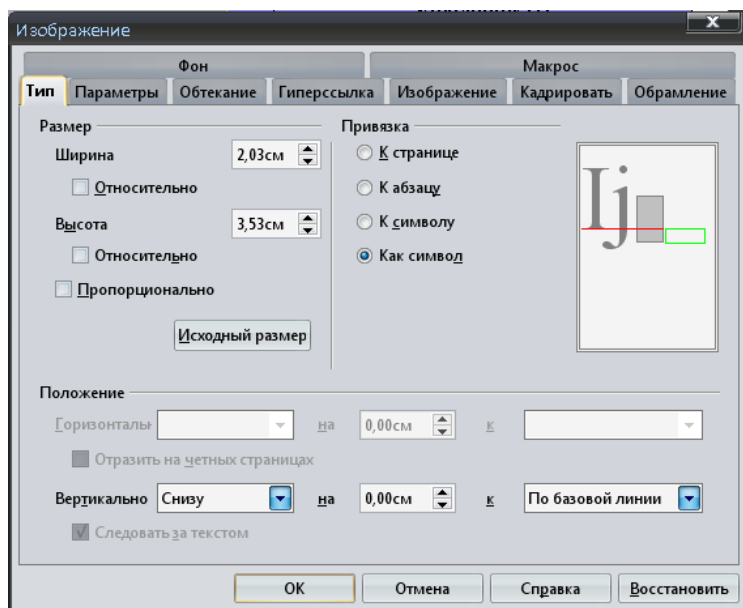


4.3.2.3. Через диалог «Изображение»

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

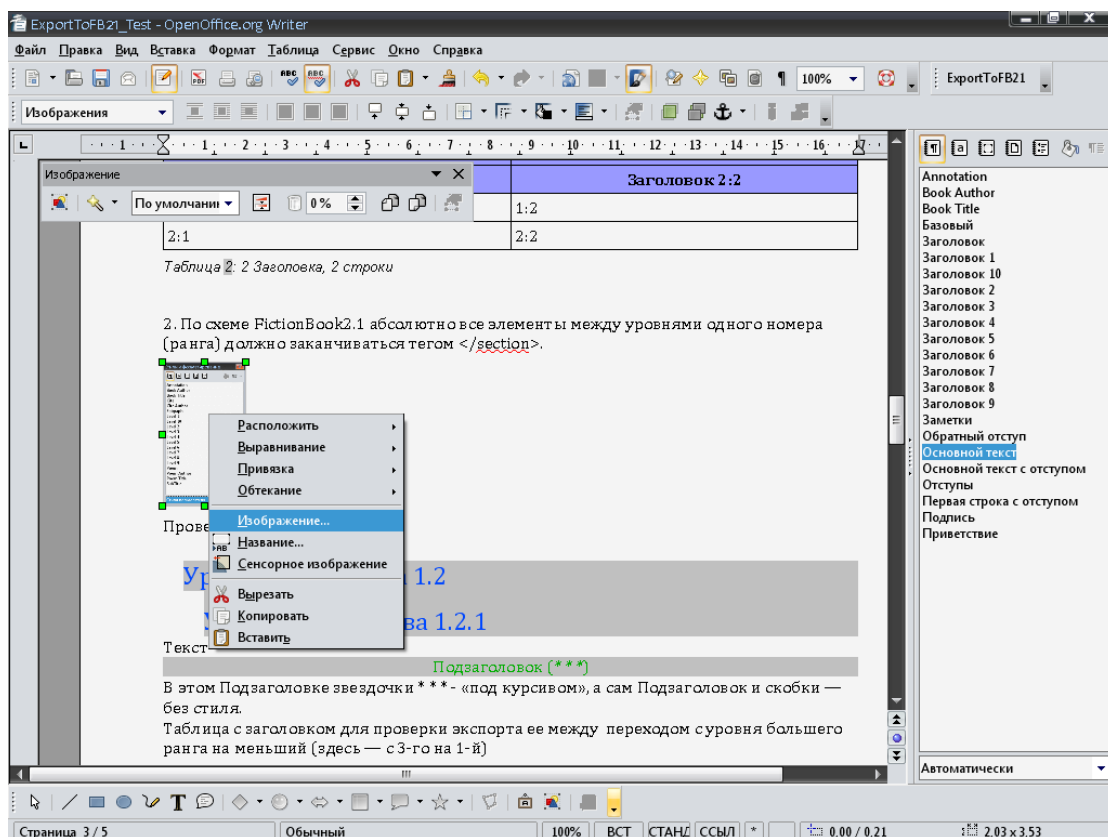


В появившемся диалоге на вкладке «Тип» выберите Привязку картинки:

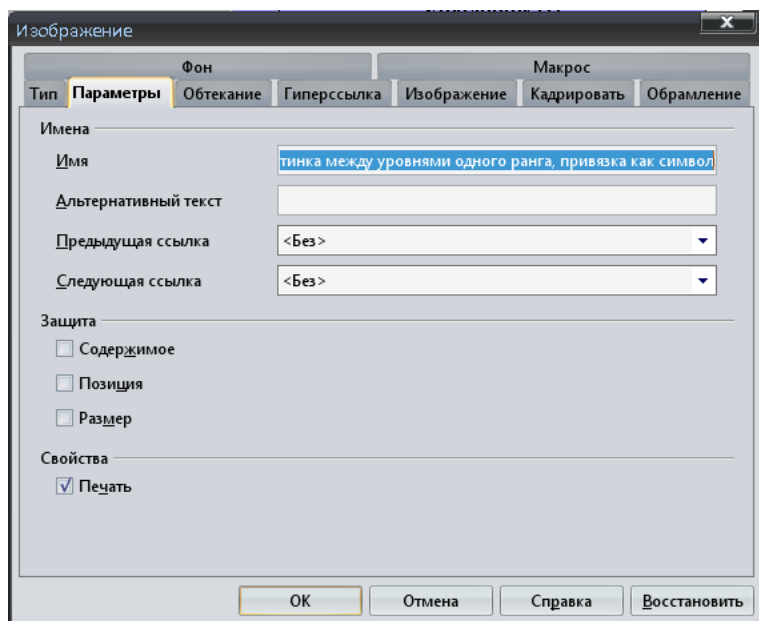


4.3.3. Задание имени картинке

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.

4.3.5. Обложки Книги

Конвертор конвертирует и обложки книги. Главное условие: все картинки, формулы и диаграммы, находящиеся в документе **выше** Названия Книги – абзаца со стилем *Book Title* (по-умолчанию) воспринимаются конвертором, как обложки.

4.3.6. Особенности работы с Графикой для Unix-подобных систем

1. Для кодирования картинок используется утилита *base64*. Пакет с ней обычно сразу устанавливается при установке дистрибутива Linux. С *Ubuntu* – это так. Если же на вашем дистрибутиве Linux почему-то не установлен этот пакет, то его необходимо доустановить. Чтобы проверить, установлен ли пакет *base64* или нет, выполните в терминале команду: *base64 -?* Если пакет установлен, то вы увидите краткую справку, если нет – то терминал выдаст соответствующее сообщение.

4.3.7. Условия успешного экспорта Графики

Т.к. Конвертор позволяет экспортировать самые различные типы графики в base64 в fb2-файл, то для этого он сначала преобразует (это не относится к png и jpeg) графику в формат, «допустимых» для читалок (см. [Настройки](#)), а только потом кодирует в base64.

К сожалению, в силу каких-то внутренних особенностей ООо это преобразование в файл может либо не осуществиться, либо вместо полноценной картинки получится «обрубок» из нескольких байт. Такую картинку не откроет ни один графический выювер – файл «битый». ООо определяет тип любой Графики неверно (в основном, как *x-vclgraphic*), и сохраняются «обрубки» картинок, а они в свою очередь кодируются в base64 так: **iVBORw0KGgo=**. Всего несколько байт, вместо «полноценного» кода. Книга экспортируется, а вот картинка в читалке не отобразится – закодированный «обрубок» картинки не воспринимается (как и должно быть).

Есть несколько случаев, когда это происходит:

1. Если работа **любого** макроса в ООо (ООоFBTools или любого другого вами установленного) была аварийно прервана или сознательно вами остановлена, то экспорт Графики уже не будет корректно осуществлен.

2. Если вы вставили Графику в текст (из файла, как Графический объект и т.д.), не закрывали редактор, а сразу же экспортировали книгу в fb2.

Поэтому, чтобы избежать всех этих проблем с «обрубленными» картинками и получить

корректный результат экспорта, необходимо жестко придерживаться следующих правил:

1. Если **любой** макрос был остановлен либо аварийно завершил свою работу – **перезагрузите** редактор, **и только потом** экспортируйте книгу в fb2.

2. Если вы правите текст Книги, устраняете грамматические ошибки, «отбиваете» стилями Заголовки, Цитаты и т.д., вставляете в нее картинки, то сохраните книгу, **перезагрузите** редактор, **и только потом** экспортируйте книгу в fb2.

При соблюдении этих условий у меня прекрасно экспортировались картинки с dpi 300 и размерами 5000 x 5360 (просто чисто для эксперимента)!!! Хотя, чем меньше dpi и размеры картинки (до разумного предела) – тем лучше.

4.4. Работа с ссылками внутри документа

4.4.1. Создание ссылок

Очень полезная функция конвертора!

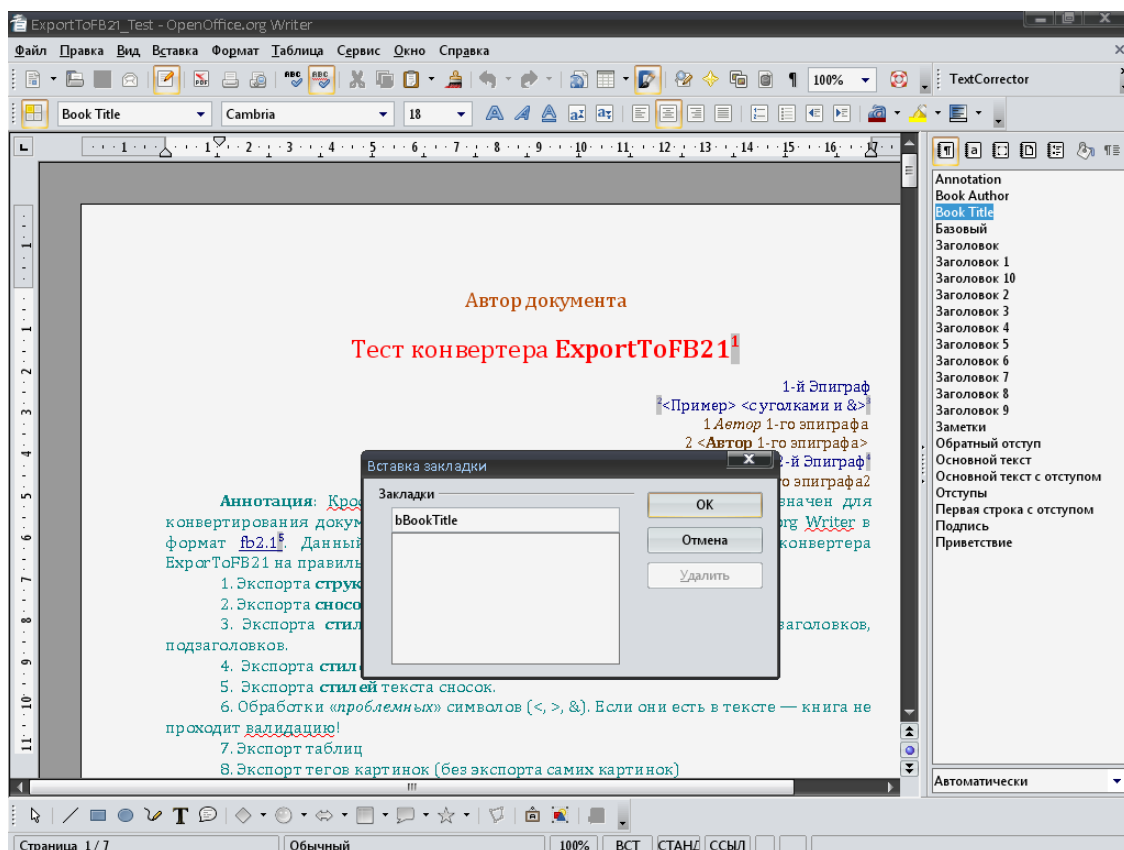
Например, надо из текста сноски по клику на какое-то слово перейти на Заголовок Поэмы. Или из Названия Книги – на Автора Цитаты. Или – из Автора Эпиграфа – на какой-нибудь раздел книги. Это легко теперь можно сделать. Как?

Прежде всего для простоты определимся с терминологией. Абзац, куда ссылается ссылка назовем «цель», а слово(а), которое(ые) ссылается(ются) на «приемник» – «источник».

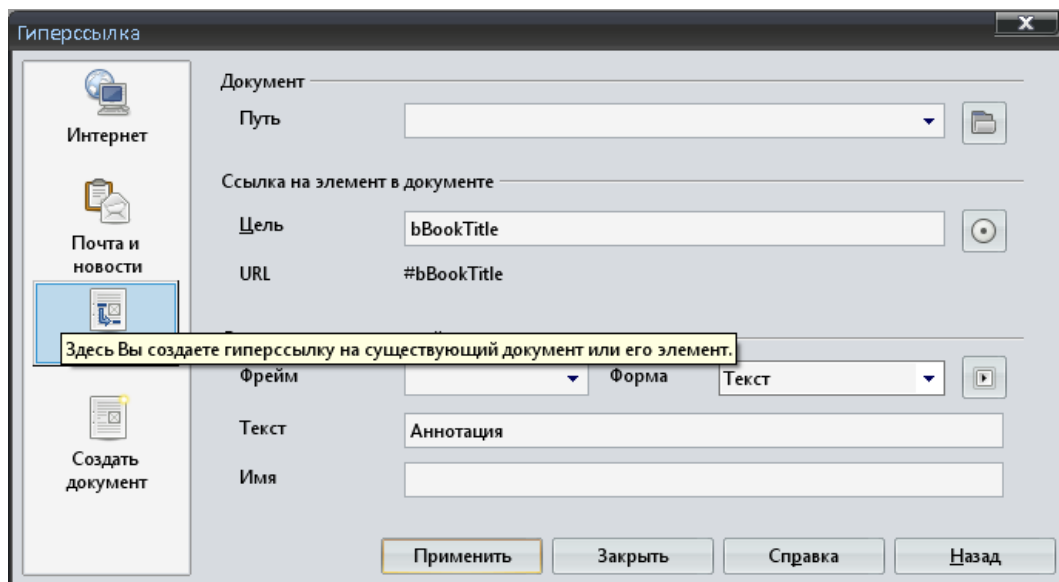
1. Поставьте курсор в любое место абзаца – «цели» (обычно сначала). Выберите меню «Вставка» → «Закладка». В появившемся диалоге введите имя закладки (желательно информативно, чтобы самому потом не запутаться, какая закладка к чему относится).

Важно: имена закладок – по-английски, *и ни в коем случае* впереди имени закладки *не ставьте* символ «#»! Работать ничего не будет!

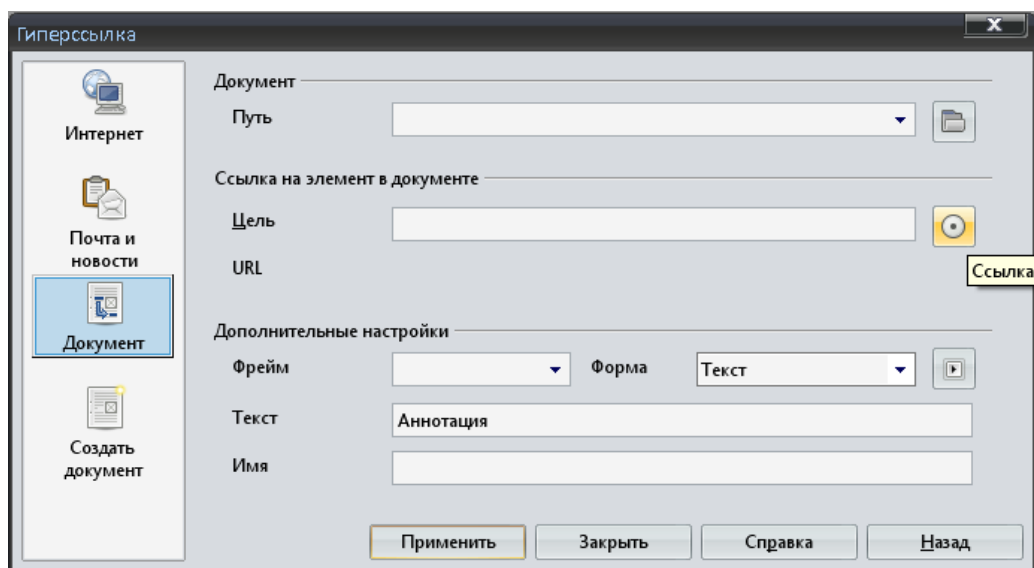
Пример ввода закладки для Названия Книги (сюда планируем отправить ссылку из сноски):



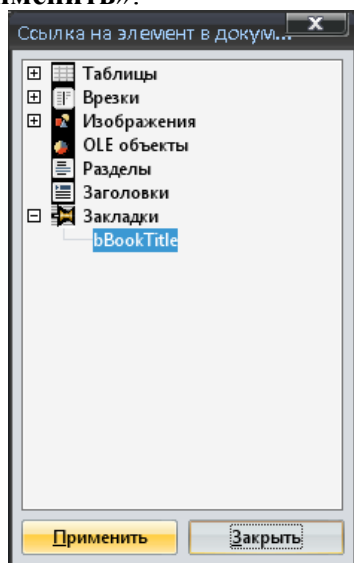
2. Выделите «источник». Через меню «Вставка» выберите «Гиперссылка», и в появившемся диалоге выделите иконку «Документ». В поле «Цель» введите имя нужной закладки, нажмите кнопку «Принять». Пример:



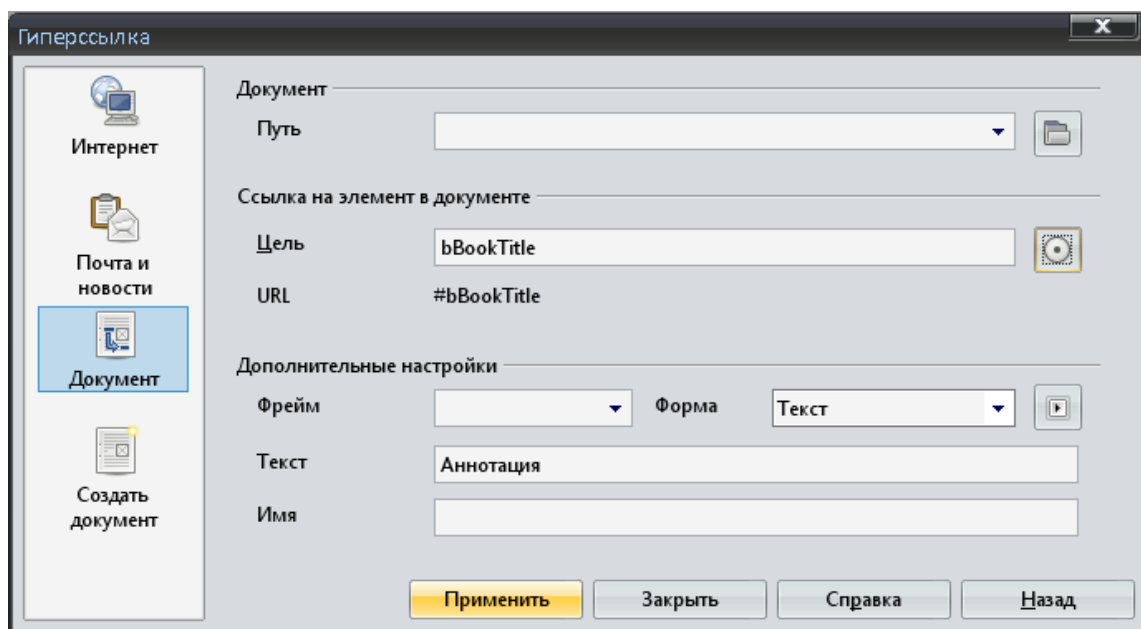
Можно вводить имя закладки не вручную (чтобы случайно не ошибиться), а с помощью диалога «Ссылка на элемент в документе». Эта кнопка – справа от поля «Цель»:



В дереве элементов выбираем «Закладки», а там – нужную нам закладку, нажимаем «Применить»:



Теперь жмем «Закреть» для возврата в диалог «Гиперссылка», и в нем нажимаем кнопку «Применить»:



Теперь закрываем это диалог – и т.д.

Сконвертируйте тестовый документ **ExportToFB21_Test.odt**, и посмотри на самые различные (и немислимые) варианты ссылок. Их работоспособность в fb2 файле можно проверить или в fb2–редакторе (кнопки перехода по ссылкам) или в читалках, просто нажав на ссылки мышкой.

4.4.2. Что важно помнить:

1. Ссылки на таблицы и (или) на тестовые врезки НЕ РАБОТАЮТ – нужно переделывать весь их алгоритм, а оно вряд ли того стоит...

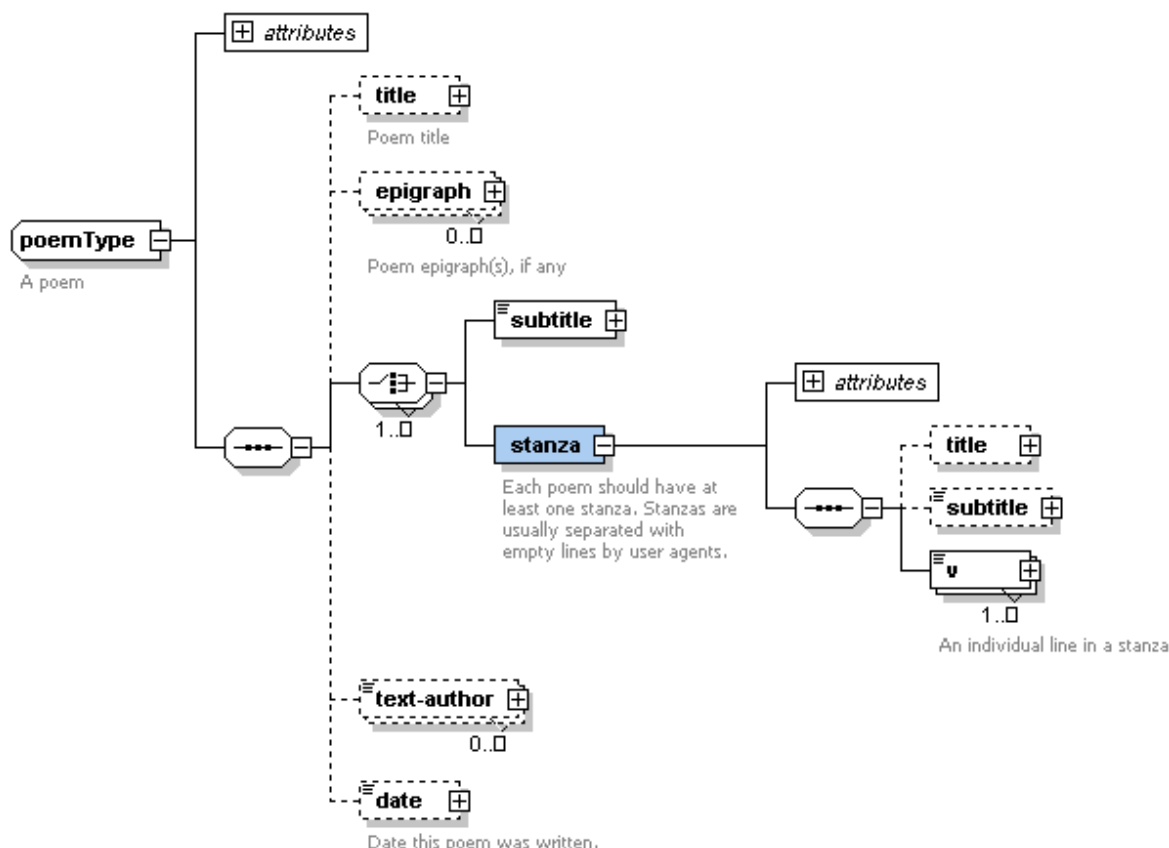
2. Если включить опцию сборки нескольких подзаголовков (SubTitle) в одну строку (тоже самое – и для заголовков Поэмы), то в результирующем Subtitle (или заголовке Поэмы) атрибут id будет взят из последнего Subtitle (или заголовке Поэмы) с закладкой! Это значит, что другие “ссылки” на переставшие существовать абзацы просто будут “указывать в никуда”! Учитывайте это!

3. Как уже было сказано выше, *ни в коем случае* впереди имени закладки *не ставьте* символ «#»! Работать ничего не будет! Это «служебный» символ для обозначения ссылки для атрибута «href» тега <a> fb2 – файла (да и в xml, тоже). Например: <a l:href="#bTitle" type="note"> . Если впереди имени закладки поставить символ «#» (<a l:href="##bTitle" type="note">), то ссылка просто работать не будет, т.к. в абзаце, куда она ссылается, в атрибуте “id” будет не «#bTitle», а «bTitle», которое конвертор берет из имени закладки – Bookmark.Name. Писать дополнительный код проверки – это обречь конвертор на еще медленное конвертирование (Open BASIC – язык интерпретатор), а это не есть хорошо. Поэтому – будьте внимательны, когда вводите имя закладки!

4.5. Работа со стихами

4.5.1. Структура стиха

Согласно схеме FictionBook стих (роем) имеет следующую структуру:



Т.е. он состоит из необязательных элементов (порядок очередности важен!) **Заголовок**, **Эпиграф**. Потом – бесконечное число чередований обязательных элементов **Подзаголовок** или **Четверостишие** (одно или бесконечное число). Потом – опять необязательные элементы **Автор стиха** (один или ни одного) и **Дата написания**. Все это полностью реализовано в конвертере.

В свою очередь элемент Stanza (Четверостишие) сам имеет два необязательных элемента **Заголовок** и **Подзаголовок**, и один обязательный – **v**. Элемент **v** реализован в конвертере, а два необязательных элемента **Заголовок** и **Подзаголовок** для Stanza – нет. Это связано с тем, что пользователю будет тяжело не запутаться с «отбивкой» стилями.

Если вы отформатируете стилями текст стиха (поэмы) в соответствии с этой схемой на картинке, то конвертер экспортирует текст корректно. Если вы переставите местами какой-либо элемент структуры Стиха – то в fb2 файле Поэма будет «разбита» на несколько, но валидация сработает, либо книга не будет валидной. Все зависит от того, что вы перепутаете. Так что будьте внимательны и сверяйтесь с этой схемой Поэмы!

4.5.1.1. Важная особенность работы со Стихами

Т.к. Стих может иметь свой Эпиграф, и Эпиграф может иметь свой стих, то, для того, чтобы конвертер смог отличить где Эпиграф относится к Стиху, а где – Стих принадлежит Эпиграфу, надо сделать следующее. Если Эпиграф Стиха имеет Автора, то лучше вставить его со стилем *Epigraph Author*. Если Автора Эпиграфа нет, то надо после Эпиграфа Стиха поставить пустой абзац стиля *Epigraph Author*.

Если этого не сделать, то конвертер (он не человек – не знает логику Стихов), «посчитает» Стих после Эпиграфа, не как основной Стих, а как Стих этого Эпиграфа.

Будьте внимательны!

4.5.2. Виды стихов

Конвертор позволяет экспортировать 3 вида стихов:

1. Несколько поэм, идущих одна за другой. Это достигается вставки *между ними пустого абзаца без форматирования* и отключением в Диалоге Настроек опции «Удалять пустые строки».

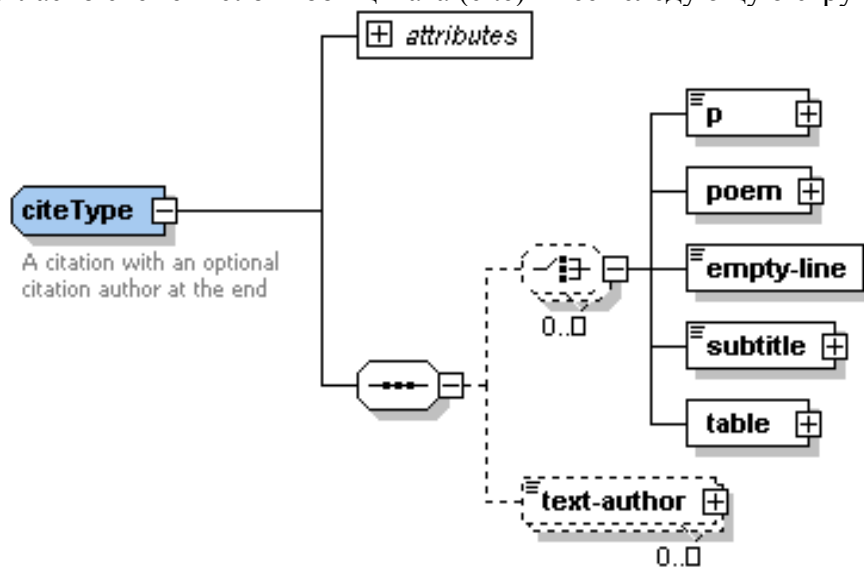
2. Поэма, где все четверостишия собраны под одной stanza. Для этого между абзацами со стихами не должно быть ни одной пустой строки.

3. Поэма, где каждое четверостишие (или набор абзацев) взято в отдельное stanza. Достигается это вставкой между каждым четверостишием пустой строки со стилем Поэм. Работает и с включенной опцией «Удалять пустые абзацы»!

4.6. Работа с цитатами

4.6.1. Структура цитаты

Согласно схеме FictionBook цитата (cite) имеет следующую структуру:



Т.е. он состоит из бесконечного числа чередований обязательных элементов (порядок очередности неважен) **Текст, Поэма, Пустая строка, Подзаголовок, Таблица**. Потом – необязательный элемент **Автор цитаты** (один или ни одного). Все это полностью реализовано в конверторе.

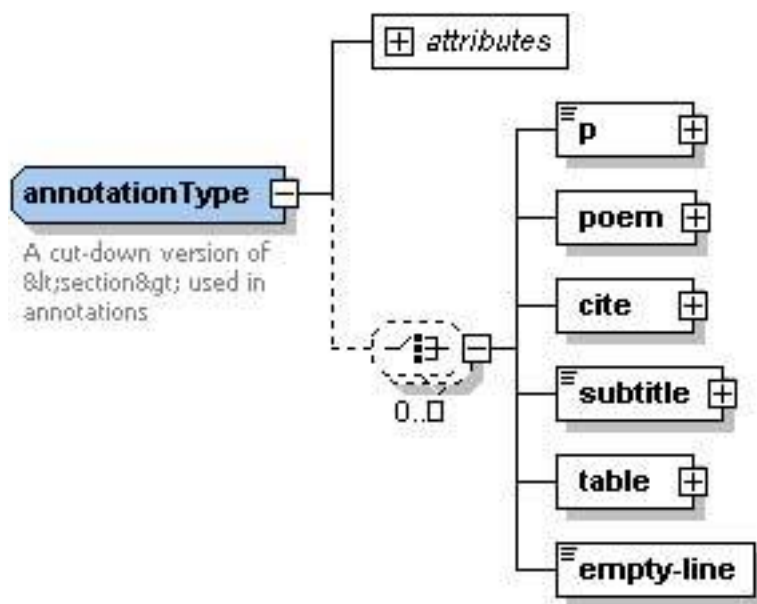
Если вы отформатируете стилями текст цитаты в соответствии с этой схемой на картинке, то конвертор экспортирует текст корректно. Если вы переставите местами **Автора Цитаты** и какой-либо элемент структуры Цитаты – то в fb2 файле книга не будет валидной. Так что будьте внимательны и сверяйтесь с этой схемой Цитаты!

Для форматирования Поэмы - элемента Цитаты, пользуйтесь теми же стилями, что и для Поэмы.

4.7. Работа с аннотациями

4.7.1. Структура аннотации

Согласно схеме FictionBook аннотация (annotation) имеет следующую структуру:

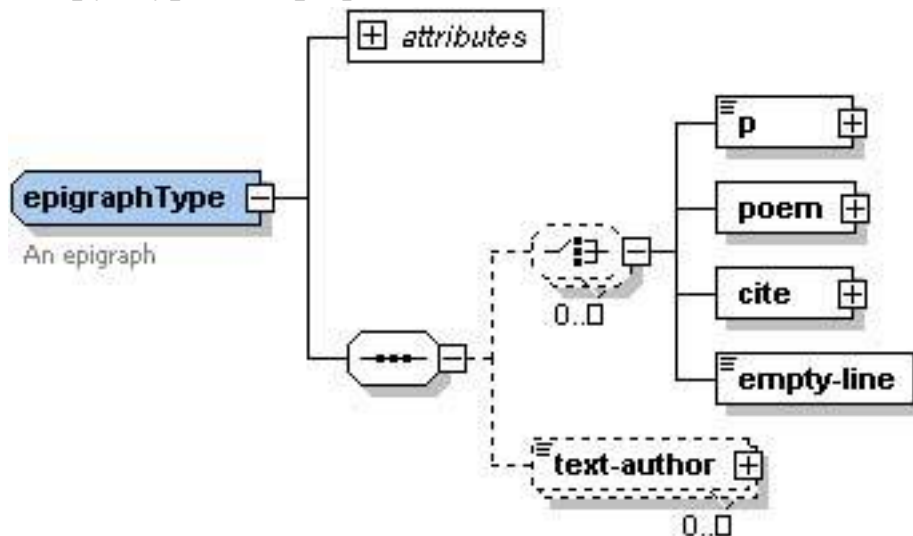


Т.е. он состоит из бесконечного числа чередований обязательных элементов (порядок очередности неважен) **Текст, Поэма, Цитата, Подзаголовок, Таблица, Пустая строка**. Все это полностью реализовано в конверторе.

Для форматирования Поэмы - элемента Аннотации, пользуйтесь теми же стилями, что и для Поэмы.

4.8. Работа с эпиграфами

4.8.1. Структура эпиграфа



Т.е. он состоит из бесконечного числа чередований обязательных элементов (порядок очередности неважен) **Текст, Поэма, Цитата, Пустая строка**. Потом – необязательный элемент **Автор цитаты** (один или ни одного). Все это полностью реализовано в конверторе.

Если вы отформатируете стилями текст эпиграфа в соответствии с этой схемой на картинке, то конвертор экспортирует текст корректно. Если вы переставите местами **Автора Эпиграфа** и какой-либо элемент структуры Цитаты – то в fb2 файле книга не будет валидной. Так что будьте внимательны и сверяйтесь с этой схемой Эпиграфа!

Для форматирования Поэмы - элемента Эпиграфа, пользуйтесь теми же стилями, что и

для Поэмы.

4.8.2. Отделение одного элемента эпиграфа от другого

Как правило, конвертор сам распознает один элемент от другого внутри Эпиграфа по характерным признакам (Автор или Дата Стиха, Автор Цитаты и т.д.). Но в одном случае, когда после Цитаты Эпиграфа сразу идет Поэма конвертор «вложит» ее во внутрь Цитаты (т.к. Цитата тоже может содержать Поэму). Если же эта Поэма принадлежит не Цитате, а непосредственно Эпиграфу, то между Цитатой и этой Поэмой *вставьте пустую строку стиля Текста Эпиграфа (Epigraph)*. Тогда Конвертор распознает и Цитату, и Поэму, как два самостоятельных элемента Эпиграфа.

Пример этого см. в тестовой файле.

Поэкспериментируйте.

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 предопределенных 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 - жанры схемы fb2.1

Книга (Title Info)

Жанр: 0 ... Язык: ru Оригинал:

Авторы

Имя: Отчество: Фамилия: 0

Ник: Email: Сайт:

Название: 1Тест конвертера ExportToFB212 Часть 1

Keywords: 1 2 3 4 5 6 7 8 9 0

Date text: 2008 Date value: 2008-03-31

Переводчики

Имя: Отчество: Фамилия: 0

Ник: Email: Сайт:

Серия: Номер: 0

FB2 Документ (Document Info)

Версия: 1.0 ID: OOO-ExportToFB21-2008730152359

Авторы

Имя: Вадим Отчество: Фамилия: Кузнецов 0

Ник: DikBSD Email: Сайт:

Programs used: ExportToFB21

Date Text: 30.07.2008 Date Value: 2008-07-30

Source OCR:

Source URLs:

Бумажная Книга (Publish Info)

Название: Название Бумажной Книги ISBN:

Издатель: Издательство Город: Город Год: 2000

Серия: Номер: 0

Создать fb2-файл

Отмена

Настройки

Жанры: ☒ fb2.1 ☐ fb2.2

Сохранить как... D:_Projects\OOoFBTools\doc\ ...

Кодировка: utf-8

Custom Info

Тип:

Value:

История fb2 файла

1.0 Создание fb2-файлов

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 FictionBook. 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.

Список **Жанр** служит для выбора жанра конвертируемой книги. Их может быть несколько. Счетчик справа от него – **match** – служит для задания степени соответствия книги выбранному Жанру в %. Атрибут **match** тэга **genre** сделан *по умолчанию необязательным*, как и задано в схеме FictionBook. Т.е. в контроле **match** по умолчанию стоит 0. В этом случае этот атрибут не заносится в тэг **genre**. Если же значение счетчика **match** изменится, и станет больше 0, то это значение будет добавлено, как значение атрибута **match** тэга **genre**.

Два переключателя **fb2.1** и **fb2.2** служат для выбора схемы жанров: FictionBook 2.1 или FictionBook 2.2. В зависимости от выбора схемы жанров меняется и выпадающий список

жанров. Это полезно, когда вы готовите книгу для Библиотеки, которая работает со схемой FictionBook 2.1.

Один момент требует особого разъяснения. Около некоторых полей для ввода данных есть кнопка «+» и счетчик числа добавленных элементов. Если в эти поля вы ввели данные, но не нажимали кнопку «+», то в сконвертированном fb2-файле 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 «+». Значение счетчика увеличится на 1, поля очистятся. Можно вводить новые значения, если это требуется.

При работе с несколькими Авторами, Переводчиками и т.д. Если в списке уже есть данные (например, для 2-х Переводчиков), и вы ввели в поля данные на 3-го, то можно: ЛИБО нажать кнопку «+», счетчик отобразит «3», поля очистятся, ЛИБО не нажимать кнопку «+», а данные так и оставить в полях. При экспорте они благополучно считаются.

Если вам надо загрузить свой профиль (данные об Авторе fb2 документа и Source OCR), 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.

Для того, чтобы ввести данные описания книги вы можете использовать любой из 2-х нижеописанных вариантов:

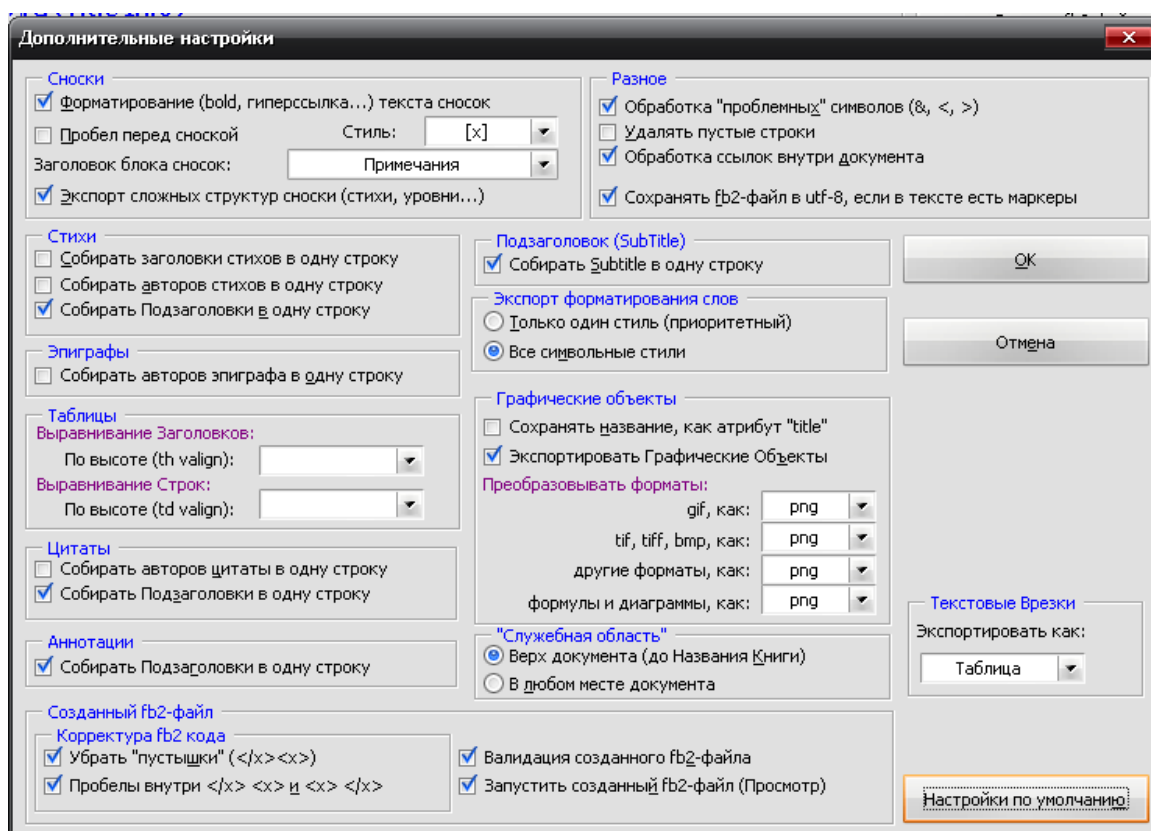
1. Вообще не использовать «Служебные стили», а вручную заносить данные в соответствующие поля формы конвертора. **Плюс:** Нет «лишней» информации в тексте книги, которую можно забыть отформатировать «Служебными стилями», и эти данные экспортируются в текст книги fb2 формата. **Минус:** При повторном конвертировании уже обработанной и ранее сохраненной книги надо будет опять вводить вручную все эти данные описания книги.

2. Использовать «Служебные стили». **Плюс:** Все данные описания книги хранятся в тексте самой книги и не переносятся в текст книги fb2 формата, а только в тэги <Description>. При повторном конвертировании уже обработанной и ранее сохраненной книги не надо будет опять обрабатывать все эти данные описания книги. **Минус:** Можно по невнимательности каким-либо данным придать не тот «Служебный стиль», и эти данные не будут экспортированы в тэги раздела описания книги в fb2 формате! Поэтому будьте внимательны!

При использовании 2-го способа все данные заносятся в текстовые массивы на стадии инициализации конвертора, и их уже «случайно испортить» вводя другую информацию форме конвертора нельзя. При этом всегда можно в поля диалога вводить дополнительные данные – они тоже будут экспортированы в тэги описания книги.

Совет: прочтите [«Служебные стили» конвертора](#), [«“Служебные стили” «по умолчанию»](#)» и [«Работа со «служебными стилями» в документе»](#) !!!

6.2. Дополнительные настройки



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

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 }, (x) или (x). По умолчанию будет использоваться [x].

3. Опция «Форматирование (bold, italic, гиперссылка...) текста сносок» позволяет конвертировать курсивы, болды гиперссылки и т.д. в тексте сноски.

Включение этой опции увеличивает время экспортирования документа.

4. Опция «Экспорт сложных структур сноски (стихи, уровни...)» позволяет конвертировать сложные стихи, сложные цитаты и т.д., а также многоуровневый текст (как в книге) в тексте сноски.

Включение этой опции увеличивает время экспортирования документа.

5. Список «Заголовок сносок» позволяет выбрать одно из predefined или же ввести свое название для секции текста сносок в результирующем fb2-файле (Примечания, Сноски, Комментарии или какое-нибудь свое).

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

7. 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).

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

9. «Обработка "проблемных" символов» позволяет корректно экспортировать символы &, < и >, как & , < и >

Включение этой опции также увеличивает время экспортирования документа.

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

11. Опция "сохранить fb2-файл в utf-8, если в тексте есть маркеры" предназначена для сохранения fb2-файла жестко в utf-8 (не зависимо от выбранной пользователем кодировки), если в тексте есть хоть один маркированный список. В этом случае маркеры экспортируются правильно (многие маркеры в windows-1251 не отображаются). Опция по-умолчанию включена.

12. 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.

Опция «Экспортировать Графические Объекты» позволяет либо включить экспорт картинок, формул и диаграмм, либо отключить их экспорт, а оставить экспорт только их тэгов. Это сделано для того, что не на всех операционных системах ООо или производные от него редакторы (как на Mac Os X) может вести себя одинаково при преобразовании графики. И, если вдруг на вашей системе конвертор будет «слетать» при экспорте графики – просто пока отключите эту опцию – картинок в fb2-файле не будет, зато все остальные «изыски» конвертора будут работать.

Опции подгруппы *Преобразовать форматы* предназначены для преобразования разных графических форматов, внедренных в документ к форматам, «родным» для fb2-читалок. Картинки png, jpeg перед кодированием не преобразуются – это основные форматы для читалок. Но, если в документе находятся форматы gif, tif, и т.д., а так же формулы и диаграммы, то все они сначала преобразуются в выбранный в этих настройках формат (понимаемый читалками), а потом уже кодируются в base64.

Можно поэкспериментировать с этими настройками, чтобы добиться наименьшего размера сгенерированного fb2-файла.

13. Для Таблиц предусмотрено выравнивание и Заголовков, и Строк в Ячейках по высоте (по ширине выравнивание считывается автоматически). При выборе «пустого» вида выравнивания для Заголовков и Строк Таблиц соответствующие атрибуты для тэгов не будут заноситься в код fb2 файлы.

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.

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

15. Опция «Обработка ссылок внутри документа» позволяет вводить в соответствующие тэги результирующего fb2 файла атрибуты id=" " для перехода на них из различных мест текста. Подробности см. в [п. 4.4](#).

16. Опции группы «Экспорт форматирования слов» дают возможность выбрать режим форматирования: либо «Только один стиль (приоритетный)», либо «Все символьные стили».

Режим «Только один стиль (приоритетный)»

При этом режиме слово, отформатированное сразу несколькими символьными стилями (гиперссылкой, полужирным и курсивом и т.д.) экспортируется **только с одним** стилем, причем стиль выбирается согласно списка приоритета стилей: 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``) нельзя. Этот режим работает быстрее.

Например, есть такой текст:

Пример `^{экспортируемого}` текста.

После экспорта fb2-код будет для этого режима таким:

```
<p><strong>Пример </strong><a  
l:href="http://www.google.ru/">экспортируемого</a><strong> текста.</strong></p>
```

Режим «Все символьные стили»

При этом режиме слово, отформатированное сразу несколькими символьными стилями (гиперссылкой, полужирным и курсивом и т.д.) так и экспортируется – со всеми стилями.

Например, все для того же текста:

Пример `^{экспортируемого}` текста.

Результат будет уже другой:

```
<p><strong>Пример </strong><sup><emphasis><strong><a  
l:href="http://www.google.ru/">экспортируемого</a></strong></emphasis></sup><strong>  
текста.</strong></p>
```

Для сравнения – код, приведенный ниже, сформирован в редакторе FictionBook Editor:

```
<p><strong>Пример </strong><a  
l:href="http://www.google.ru/"><strong><emphasis><sup>экспортируемого</sup></emphasis></  
strong></a><strong> текста.</strong></p>
```

Этот режим включен по-умолчанию.

17. Группа опций «Корректурa fb2 кода»: «Убрать “пустышки” `</x><x>`» и «Пробелы внутри `</x><x>` и `<x></x>`». Опции позволяют в процессе конвертирования документа существенно уменьшить размер результирующего fb2 файла. Группы тэгов ``, `</sub>_{` и т.д. удаляются («пустышки»), а ``, `}<sub>`, ``, `` и т.д. заменяются на пробелы.

18. Опции группы «Служебная область» служат для того, чтобы «подсказать» конвертору, где быстрее найти информацию-описание книги, отформатированную «Служебными стилями» (данные об Авторах, Переводчиках, Издательстве и т.д.). Кому-то удобнее размещать эти данные в самом верху текста документа, кому-то – в самом низу, а кому-то – где угодно. Здесь важно помнить следующее:

- Если вам удобно размещать информацию о книге в самом *низу* текста, то включите опцию **Диалога Настроек** опции «В любом месте документа», но при этом, если ваш документ очень большой, на считывание этих данных затратится несколько секунд (перебор всего текста). При маленьких объемах книги время на инициализацию конвертора практически ничтожно, как и в случае опции «Верх Документа (До Названия Книги)».
- Если вам важна скорость инициализации конвертора, то размещайте информацию о книге в самом *верху* текста, при этом включите опцию **Диалога Настроек** опции «Верх Документа (До Названия Книги)» (задана по умолчанию). На больших документах инициализация конвертора практически мгновенная — перебор только до стиля Названия Документа.

19. Группа опций «Созданный fb2-файл»:

Опция «Запустить созданный fb2-файл (Просмотр)» позволяет запустить созданный fb2-файл в приложении, которое ассоциировано с расширением fb2. Если fb2 файлы запускает по-умолчанию читалка – то она и запустится, если – FB2 редактор – то он и запустится, если

xml-редактор, то он и запустится. Если с fb2 не ассоциирована никакая из этих программ, то запустится текстовый редактор, либо браузер, в зависимости от настроек системы. Эта опция удобна тем, что можно сразу после экспорта текста посмотреть, что же у нас получилось. Опция по-умолчанию включена.

Опция «Валидация созданного fb2-файла» позволяет после создания fb2-файла проверить его на соответствие схеме FictionBook.xsd. Опция по-умолчанию включена. Пока валидация работает только для системы Windows.

При нажатии кнопки ОК все контролы диалога становятся недоступны (защита от случайного нажатия), запускается процесс инициализации конвертора (из-за Области «Служебных стилей»). Ход работа иллюстрируется прогрессбаром на OOo Writer.

Все настройки формы «Дополнительные настройки» сохраняется в файл *settings.txt*, который помещается в жестко заданную папку *config* пакета *OpenOffice.org*, откуда и загружается. Для Unix-подобных систем – это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows – это папка *C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools*, где XXX – ваш логин (имя пользователя).

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 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.

3. 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.

4. 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. Работа с профилями конвертора

Каждый раз вводить одинаковые данные в поля конвертора довольно утомительно, тем более, если создаешь очень много книг. Поэтому предусмотрена возможность создания и загрузки в нужные поля данных разных профилей.

8.1. Профиль Автора Книги

Очень удобно создать для разных авторов книг свои профили и загружать их, по мере оцифровки их книг. Например, вы оцифровываете все произведения Федора Михайловича Достоевского. Удобнее для каждой его книги просто загрузить его профиль, чем набивать вручную его данные.

На форме конвертора, в группе «Книга» (Title Info)» есть три кнопки для работы с

профилем – «Загрузить профиль...», «Сохранить профиль...» и «Очистить поля ввода»:

Для удобства работы создайте на жестком диске папку «Авторы книг» (или как вам нравится). Здесь будут храниться файлы профиля Авторов книг.

Например, для Достоевского создаем «его» профиль:

В результате все его данные и «его» жанр сохранены в файле. При оцифровке другой его книги просто загружаем его профиль.

Бывает, что один автор работает в разных жанрах. А наш профиль сохраняет один

выбранный жанр. Тогда при загрузке профиля этого автора надо просто выбрать нужный жанр.

Кнопка «**Очистить поля ввода**» нужна для очистки полей Автора книги и его жанра. Иногда бывает полезно...

Важно: При вводе названия файла профиля Автора Книги **не ставьте на конце точку!** Конвертор вылетает – OpenBasic OOo Writer'a не считывает в этом случае название файла.

8.2. Профиль Автора fb2–документа

Профиль Автора fb2-файла может быть 2-х видов: 1. По-умолчанию (нет возможности указать, куда сохранять); и 2. Произвольный (можно сохранять в любое место)

Профиль состоит из данных Автора fb2–документа (имя, фамилия, отчество, ник, email и сайт), Source OCR, Source URLs и истории создания документа.

На форме конвертора, в группе «FB2 Документ (Document Info)» есть 4 кнопки для работы с профилем: «> >» и «< <» – для загрузки и сохранения профиля **по-умолчанию**, и «А <<» и «А >>» – для загрузки и сохранения профиля в произвольный txt-файл.

Профиль Автора fb2-файла по-умолчанию

Кнопка «> >» служит для сохранения данных профиля, а «< <» – для загрузки данных профиля в поля конвертора. В обоих случаях появится предупреждение и перезаписи данных в текстовом файле профиля или в полях конвертора (для сохранения и загрузки профиля соответственно). Это – защита от случайного нажатия кнопок и потери информации.

Поскольку эти данные практически не меняются, то для загрузки и сохранения профиля не предусмотрен диалог выбора файла. Файл профиля *profile.txt* помещается в жестко заданную папку *config* пакета *OpenOffice.org*, откуда и загружается. Для Unix–подобных систем – это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows – это папка *C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools*, где XXX – ваш логин (имя пользователя).

Произвольный профиль Автора fb2-файла

Кнопка «> > А» служит для сохранения данных профиля в произвольный txt-файл, а «А <<» – для загрузки данных профиля в поля конвертора из ранее сохраненного файла.

При старте диалога конвертера автоматически загружается профиль Автора fb2-файла ПО-УМОЛЧАНИЮ. Его можно изменить и сохранить. Если же надо загрузить (и (или) сохранить) профиль другого Автора fb2-файла, то используйте кнопки "А <<" и ">> А".

8.3. Профиль Издательства

Удобно также сделать свои профили и для издательств, а потом их подгружать.

На форме конвертора, в группе «Бумажная Книга» (Publish Info)» есть три кнопки для работы с профилем – «**Загрузить профиль...**», «**Сохранить профиль...**» и «**Очистить поля ввода**»:

Бумажная Книга (Publish Info)

Название:

Издатель: Город:

Серия:

ISBN:

Номер:

Кнопки: Загрузить профиль..., Загрузить данные об Издательстве из отдельного профиля., Очистить поля ввода

Так же, как и для профиля Автора Книги создайте на жестком диске папку «Издательства» (или как вам нравится). Здесь будут храниться файлы профиля Издательств.

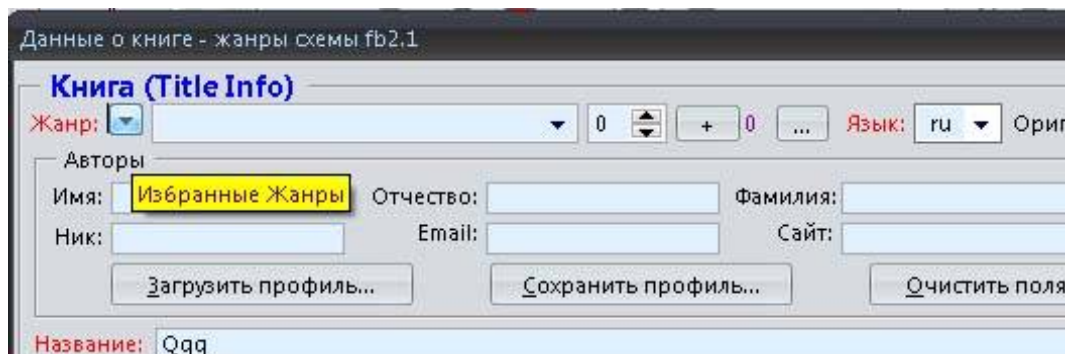
Кнопка «**Очистить поля ввода**» нужна для очистки полей Бумажной книги. Это так же иногда бывает полезно...

Важно: При вводе названия файла профиля Издательства **не ставьте на конце точку!** Конвертор вылетает – OpenBasic OOo Writer`а не считывает в этом случае название файла.

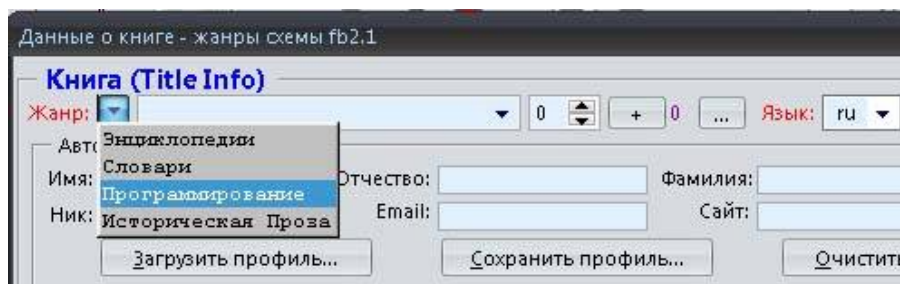
9. Избранные жанры

Как правило, создатели fb2-книги используют очень небольшое число жанров. Каждый раз выбирать жанр для книги из огромного списка – утомительно. Поэтому сделана возможность создания Избранных Жанров и работа со списком Избранных Жанров – сортировка, перемещение, добавление и удаление.

При запуске конвертора список Избранных Жанров пуст (он расположен между меткой «Жанр:» и основным списком Жанров):



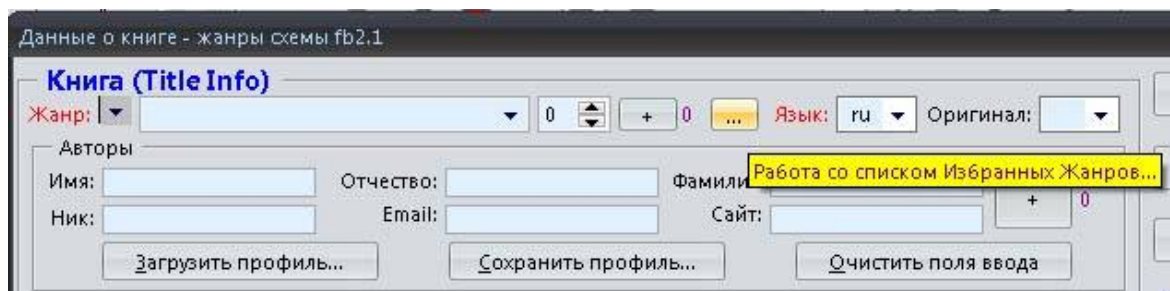
После добавления «любимых» жанров в Избранное, получаем что-то вроде этого:



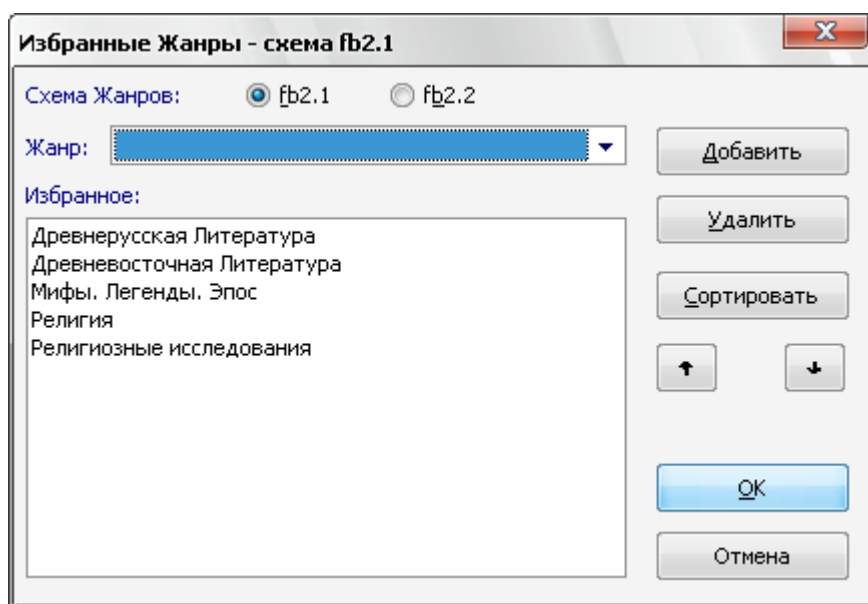
При выборе жанра из Избранного, этот жанр отображается в основном списке жанров. Дальше – работаем, как обычно.

9.1. Создание списка Избранных Жанров

Для создания Избранного Жанров нажмите кнопку «...» (расположена между счетчиком жанров и языком книги):



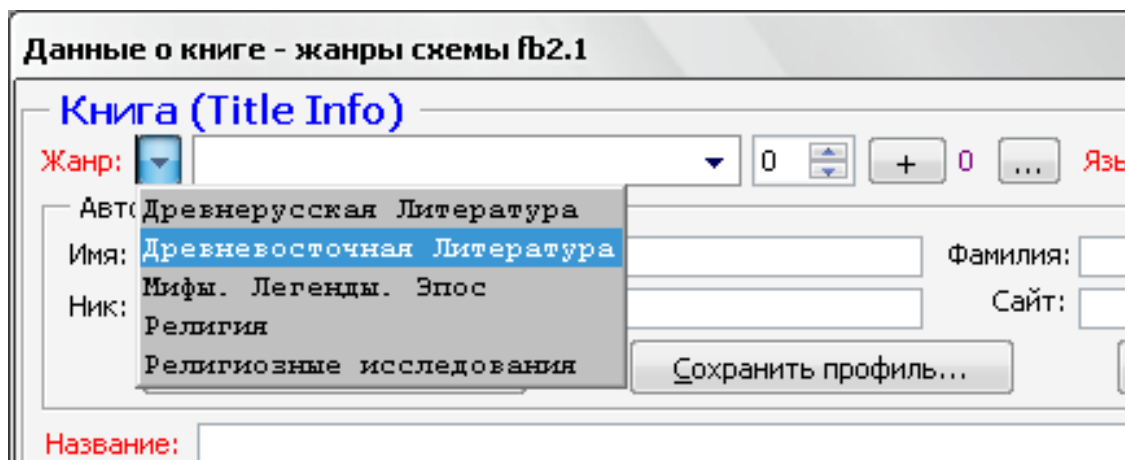
Появится диалог, работа с которым особых комментариев не требует. На картинке уже имеется некоторый список:



Для удобства можно отсортировать список жанров или переместить нужный жанр выше или ниже в списке...

После создания списка нажатия кнопки ОК список сохраняется в соответствующий файл. Файл Избранных Жанров сохраняются в файл *mygenres.txt*, который помещается в жестко заданную папку *config* пакета *OpenOffice.org*, откуда и загружается. Для Unix-подобных систем – это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows – это папка *C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools*, где XXX – ваш логин (имя пользователя).

При нажатия на выпадающий список Избранных Жанров на Главной Форме можно выбрать нужный Жанр,



и он будет отображен в основном выпадающем списке жанров:

Данные о книге - жанры схемы fb2.1

Книга (Title Info)

Жанр: ▼ Древневосточная Литература ▼ 0 ▲ ▼ + 0 ...

Авторы

Имя: Отчество: Фамилия:

Ник: Email: Сайт:

Название:

После этого этот Жанр можно будет либо использовать для экспорта книги, либо добавить в список жанров экспортируемой книги (кнопка «+»).

II. Text Corrector - Корректор текста документа

Для работы с корректором текста надо вызвать его диалог, который выглядит так:

Корректор текста документа ✕

Разорванные дефисы и переносы символом ¶

☐ Удалить разрывы для дефисов и переносов слов

☐ Обрабатывать только разорванные дефисы

☒ Дефисы и удаление из слов только символа ¶ без переноса (-)

☐ Дефисы и удаление из слов разрыва: и переноса (-) и абзаца (¶)

Абзацы и разрывы

☐ Собрать разорванные символом ¶ строки абзаца в один абзац

☐ По пробелам в начале абзаца

☐ По .!?..."'; в конце абзаца

☒ По "пустому" абзацу

☐ Удалить пустые абзацы (замена множественных ¶ одним)

☐ Разрыв строки:

☒ Удалить ☐ Заменить на абзац (¶) ☐ Заменить на 1 пробел

Пробелы и Табуляции

☒ Удалить все лишние пробелы и табуляции

Кавычки, Дефисы, Разное

☐ Заменить двойной дефис (--) на тире

☒ Вставка пробела между "слипшимися" символом и минусом, тире или дефисом

☒ Заменить дефис, используемый вместо тире на тире

☒ Вместо минуса - дефис в "кое-что", "куда-то" и т.д.

☒ Заменить кавычки " " и " " на « »

☒ Упорядочить многоточие

☐ Заменить (c) на ©, (r) на ® и (TM) на ™

☒ Удалить мягкие переносы

☐ Удалить разрывы страниц

☐ Изменить привязку всех картинок

Картинки

Привязать: К абзацу ▼

Все выбранные опции автоматически запоминаются, и при следующем вызове Корректора они будут автоматически установлены.

II.1. Описание некоторых опций

Многие опции Корректора Текста интуитивно понятны. Некоторые, требующие

пояснения, опишем подробнее.

II.1.1. Удалить разрывы для дефисов и переносов слов

Очень важная опция! Очень часто требуется исправить текст, который содержит слова с дефисами и переносами, разорванными символом абзаца. В ручную – долго и утомительно.

Данная опция имеет 3 режима работы: 1. Обрабатывать только разорванные дефисы. 2. Дефисы и удаление из слов только символа ¶ без переноса (-). 3. Дефисы и удаление из слов разрыва: и переноса (-) и абзаца (¶).

Т.е. разорванные дефисы обрабатываются во всех трех режимах – удаляется разрыв, а в двух последних режима еще и обрабатываются слова, не обязательно содержащие дефисы, но разорванные на переносе.

Поиск и корректура осуществляется везде – в тексте, в таблицах, в текстовых врезках и в сносках.

Рассмотрим каждый режим.

1. Обрабатывать только разорванные дефисы

Есть два вида слов с дефисом: предлоги, местоимения и т.д. (все-таки, во-первых, ты-то, из-за...) и другие слова (белым-бело, сестра-хозяйка, по-хорошему, иди-ка...). *Таких* слов сотни, тысячи... Если составить полный список *всех этих* слов, и «загрузить» его в Корректор для сравнения всех найденных в тексте дефисных слов с разрывами, то на сравнение только одного слова из книги понадобится несколько часов! Поэтому, в Корректоре механизм распознавания *таких дефисных* слов сделан следующим образом:

1. Правая или левая часть найденного в тексте слова с разорванным дефисом (или переносом) сравнивается со словами-шаблонами, которые в большой вероятности принадлежат словам с дефисами: «-то», «-ка», «-таки», «-ли», «по-», «кое-», «кой-», «из-», «во-», «в-». Если есть соответствие, то Корректор удаляет разрыв – символ ¶, оставляя при этом дефис, или символ «минус» – смотря, что стояло.

Конечно, есть вероятность того, что в тексте встретится не дефисное слово, соответствующее этому алгоритму, а просто слово с разорванным переносом (например: «по-¶шли», «доло-¶то»), и, вместо «-¶» удалится только «¶», но такая вероятность мала. Во всяком случае, потом текст можно «прогнать» [ручным инструментом поиска слов с дефисом](#), и удалить символ "-" там, где он остался. Но – лучше такой алгоритм, чем...

2. И левая, и правая части найденного слова с разорванным дефисом (или переносом) сравниваются с полным шаблоном: «как-никак», «какой-никакой». Если есть соответствие, то Корректор удаляет разрыв – символ ¶, оставляя при этом дефис, или символ «минус» – смотря, что стояло. Список этих полных шаблонов по мере развития Корректора может расти.

3. Составлен список слов, правая дефисная часть которых – либо слово «нибудь», либо – «либо». Вот этот список (он со временем может пополниться): "где", "откуда", "зачем", "почему", "отчего", "кто", "как", "когда", "какой", "какого", "какому", "каком", "какая", "какую", "какое", "какого", "какие", "каких", "каким", "кем", "кого", "кому", "ком", "куда", "что", "чем", "чём", "чему", "чей", "чья", "чьего", "чьему", "чьей", "чьи", "чьих", "чьими", "чье", "чьё", "чего", "сколько", "скольких", "скольким", "сколькими".

Когда Корректор нашел в тексте разорванный абзацем дефис, то он определяет правую часть, если это – «нибудь» или «либо», то далее следует проверка по вышеназванному списку. Если есть соответствие, то Корректор удаляет разрыв – символ ¶, оставляя при этом дефис, или символ «минус» – смотря, что стояло.

Если же этого соответствия не было, то найденное в тексте слово «не считается»

дефисным, а просто с разорванным переносом (например, смот-¶ри), и в этом режиме обработки игнорируется.

2. Дефисы и удаление из слов только символа ¶ без переноса (-)

Все, что было написано для режима «Обрабатывать только разорванные дефисы» работает и в этом режиме. Только в дополнение к этому, удаляется не только разрыв дефисных слов, но и слова с разорванными переносами не игнорируются, и из них удаляется только символ «¶». Символ «-» не трогается. Этот режим удобен для исправления разрывов не только дефисов, но и слов, типа «давным-давно», «сестра-хозяйка», «белый-белый» и т. п. Например, слово «давным-¶давно» станет «давным-давно», а «смот-¶ри» – «смот-ри». В первом случае получили то, что и надо, а во втором – нет. Текст после коррекции можно «прогнать» [ручным инструментом поиска слов с дефисом](#), и удалить «-» там, где он остался. Универсальный алгоритм предусматривал бы сравнение со списком из тысяч слов в разных склонениях и временных формах, и требовал бы часы, на отработку. Поэтому, этот алгоритм проще, хотя и не все корректно обрабатывает.

3. Дефисы и удаление из слов разрыва: и переноса (-) и абзаца (¶)

Все, что было написано для режима «Обрабатывать только разорванные дефисы» работает и в этом режиме.

Этот режим похож на описанный выше, с одним лишь отличием. Удаляется не только разрыв дефисных слов, но и оба символа разрыва «-¶» в слова с разорванными переносами. Например, слово «давным-¶давно» станет «давнымдавно», а «смот-¶ри» – «смотри». В первом случае получили не то, что надо, а во втором – именно то, что надо.

Таблица сравнения результата работы 3-х режимов коррекции

Для сравнения режимов возьмем такой произвольный текст (отражены и непечатные символы). **Синим** цветом показаны «дефисные» слова с разорванным дефисом, **красным** - слова с разорванным переносом, **зеленым** – «дефисные» слова без разрывов (просто для того, чтобы убедиться. Что с ними ничего не случится):

1. Кое-что; сколько-¶
нибудь.¶
2. По-русски; по-¶
хорошему.¶
3. Сам-то; далеко-¶
то; ты-¶
то¶
4. Кое-что; кое-¶
куда...¶
5. Этот текст показывает: **немного-¶**
немало; а работу 3-х режимов обработки слов с
разорван-¶
ным переносом или дефисами.¶
Во-первых; он хоть как-¶
то; но показывает работу Корректора Текста.¶
Во-¶
вторых - можно сравнить результат и **все-таки**
определиться для себя, какой режим лучше
использовать.¶

1. Обрабатывать <u>только</u> разорванные дефисы	2. Дефисы и удаление из слов <u>только</u> символа ¶ без переноса (-)	3. Дефисы и удаление из слов разрыва: <u>и</u> переноса (-) <u>и</u> абзаца (¶)
---	---	---

1. Кое-что, сколько-нибудь.¶ 2. По-русски, по-хорошему.¶ 3. Сам-то, далеко-то, ты-то¶ 4. Кое-что, кое-куда...¶ 5. Этот текст показывает немного-¶ немало , а работу 3-х режимов обработки слов с разорван-¶ ным переносом или дефисами.¶ Во-первых , он хоть как-то , но показывает работу Корректора Текста.¶ Во-вторых -- можно сравнить результат, и все-таки определимся для себя, какой режим лучше использовать.¶	1. Кое-что, сколько-нибудь.¶ 2. По-русски, по-хорошему.¶ 3. Сам-то, далеко-то, ты-то¶ 4. Кое-что, кое-куда...¶ 5. Этот текст показывает немного-немало , а работу 3-х режимов обработки слов с разорван-ным переносом или дефисами.¶ Во-первых , он хоть как-то , но показывает работу Корректора Текста.¶ Во-вторых -- можно сравнить результат, и все-таки определимся для себя, какой режим лучше использовать.¶	1. Кое-что, сколько-нибудь.¶ 2. По-русски, по-хорошему.¶ 3. Сам-то, далеко-то, ты-то¶ 4. Кое-что, кое-куда...¶ 5. Этот текст показывает немногонемало , а работу 3-х режимов обработки слов с разорванным переносом или дефисами.¶ Во-первых , он хоть как-то , но показывает работу Корректора Текста.¶ Во-вторых -- можно сравнить результат, и все-таки определимся для себя, какой режим лучше использовать.¶
Пояснения		
1. Корректно исправляются самые распространенные дефисные слова. 2. Менее распространенные дефисные слова не исправляются (здесь – « немного-¶ » немало »)). 3. Другие слова с разрывами не исправляются (здесь – « разорван- ¶ным »))	1. Корректно исправляются самые распространенные дефисные слова. 2. Корректно исправляются менее распространенные дефисные слова (здесь – « немного-немало »)). 3. Другие слова с разрывами не исправляются (здесь – « разорван- ¶ным »))	1. Корректно исправляются самые распространенные дефисные слова. 2. Менее распространенные дефисные слова не исправляются, части его «слипнутся» (здесь – « немногонемало »)). 3. Корректно исправляются другие слова с разрывами (здесь – « разорванным »))

Как видим, наиболее «универсальный» – это 2-й режим (он стоит по-умолчанию). Хотя иногда после него надо «ручная доводка».

II.1.2. Собрать разорванные символом ¶ строки абзаца в один абзац

При выборе этой опции сначала отработывают алгоритмы удаления из текста «мягких переносов» и «Удаление разрывов для дефисов и переносов слов», даже если эти опции и не были включены! Это необходимо для корректной склейки разорванных строк одного абзаца. При этом, будет задействован помеченный режим опции «Удаление разрывов для дефисов и переносов слов».

Тоже очень полезная опция. Она имеет 2 варианта работы (алгоритма): «По пробелам в начале абзаца» и «По .!?!...»”“”“; в конце абзаца», т.е. завершение абзаца определяется по выше названным символам, а все абзацы между – воспринимаются, как разорванные строки одного и того же абзаца, и склеиваются.

Режим «По пробелам в начале абзаца»:

Часто «работает» с книгами, скаченными из сети, которые после OCR сохранились, как простой текст. Красная строка «отбивалась», как несколько пробелов, а перевод строки сохранялся как окончание абзаца. Т.е. все строки каждого абзаца сохранились, как отдельный абзац. Править руками такую книгу удовольствия мало... Корректор позволяет собрать такие строки в «нормальные» абзацы. При этом идентификатором «нового» абзаца является наличие в начале строки 1 или нескольких пробелов.

Пример. Есть такой текст (для наглядности показаны и непечатаемые символы):

```

файл·из·библиотеки·фензина·http://www.fenzin.ru¶
Любишь·фантасстику?·Давай·на·фензин!¶
¶
.....Сергей·ЛУКЪЯНЕНКО¶
¶
.....ПРИСТАНЬ·ЖЕЛТЫХ·КОРАБЛЕЙ¶
¶
¶
¶
.....ЧАСТЬ·ПЕРВАЯ··ПУТЬ¶
¶
.....Шлюпка·все-таки·не·выдержала·..До·поверхности·оставалось·..всего¶
несколько·километров,·когда·по·пульту·пробежала·волна·алых·аварийных¶
огней,·а·за·стеной·взвыла·..сирена·..Индикатор·перегрева·реактора¶
стремительно·прыгнул·вверх,·и·Кирилл·понял,·что·пора·уходить·Он·задержал¶
дыхание,·вжался·в·кресло·и·рванул·рычаг·катапульты·..Над·головой·хлопнул¶
раскрывшийся·люк,·и·его·выбросило·в·поток·спрессованного·..воздуха.¶
Разноцветные·ленты·стриммер-парашюта·выхлестнулись·из·ранца,·одновременно¶
лопнули·ремни,·и·он·отделился·от·кресла·..А·шлюпка,·с·виду·совершенно¶
невредимая,·неслась·вверх,·стремительно·исчезая·в·голубоватой·дымке¶
облаков.¶
.....Кирилл·проводил·взглядом·ее·трехгранный·корпус·и·взялся·за·упругие¶
парашютные·фалы·Земля·надвигалась·слишком·уж·быстро,·видимо,·здесь·была¶
небольшая·разница·в·давлении·или·силе·тяжести·Он·упал·удачно,·на·бок,¶
прокатился·по·низкой,·выгоревшей·траве,·пытаясь·сбить·рвущийся·на·ветру¶
стриммер,·а·камень,·на·который·его·тащило,·заметил·слишком·поздно...¶
¶
¶

```

После обработки Корректором со всеми опциями получаем следующее:

```

|файл·из·библиотеки·фензина·http://www.fenzin.ru¶
Любишь·фантасстику?·Давай·на·фензин!¶
Сергей·ЛУКЪЯНЕНКО¶
ПРИСТАНЬ·ЖЕЛТЫХ·КОРАБЛЕЙ¶
¶
ЧАСТЬ·ПЕРВАЯ··ПУТЬ¶
Шлюпка·все-таки·не·выдержала·До·поверхности·оставалось·всего·несколько
километров,·когда·по·пульту·пробежала·волна·алых·аварийных·огней,·а·за·стеной
взвыла·сирена·Индикатор·перегрева·реактора·стремительно·прыгнул·вверх,·и·Кирилл
понял,·что·пора·уходить·Он·задержал·дыхание,·вжался·в·кресло·и·рванул·рычаг
катапульты·Над·головой·хлопнул·раскрывшийся·люк,·и·его·выбросило·в·поток
спрессованного·воздуха·Разноцветные·ленты·стриммер-парашюта·выхлестнулись·из
ранца,·одновременно·лопнули·ремни,·и·он·отделился·от·кресла·А·шлюпка,·с·виду
совершенно·невредимая,·неслась·вверх,·стремительно·исчезая·в·голубоватой·дымке
облаков.¶
Кирилл·проводил·взглядом·ее·трехгранный·корпус·и·взялся·за·упругие·парашютные
фалы·Земля·надвигалась·слишком·уж·быстро,·видимо,·здесь·была·небольшая·разница
в·давлении·или·силе·тяжести·Он·упал·удачно,·на·бок,·прокатился·по·низкой,
выгоревшей·траве,·пытаясь·сбить·рвущийся·на·ветру·стриммер,·а·камень,·на·который
его·тащило,·заметил·слишком·поздно...¶
¶

```

Такому тексту для наглядности абзацев осталось только задать отступ 1-й строки, и все!

Режим «По .!?...»”\””;; в конце абзаца»:

Не менее часто встречаются книги, скаченные из сети, у которых нет Красной строки, а перевод строки сохранялся как окончание абзаца. Т.е. все строки каждого абзаца сохранились, как отдельный абзац. Этот вариант еще труднее править руками. Корректор позволяет собрать такие строки в «нормальные» абзацы. При этом идентификатором

завершения абзаца является наличие в его конце вышеназванные символы.

Пример. Есть такой текст (для наглядности показаны и непечатаемые символы):

¶
С Ревкиным случилось вот что. Он эту ученую статью кандидата Ушастого воспринял ¶ как выпад против себя лично. Он перед зеркалом даже вертелся тайком от Аглаи, ¶ вершиками обмеривал свой собственный череп от затылка до подбородка и от уха до уха, и измерения эти оказались удручающими для него. А удлинить свою голову при ¶ помощи спецпарика он, конечно, не мог, ибо народ его знал таким, каков он есть, ¶ и терять свой авторитет он не желал. Однако он понял, что власть его уходит из ¶ рук, как песок, и потому решился на самый отчаянный шаг. Он определенно заявил, ¶ что подобного безобразия терпеть не будет, и послал в обком статью Ушастого и ¶ свою докладную записку, в которой называл статью псевдонаучной и шарлатанской, ¶ утверждал, что она инспирирована, конечно, новым начальником Тех. Кому Надо, ¶ который с самого начала ведет себя вызывающе, игнорирует партийные органы и тем ¶ самым противопоставляет свое Учреждение партии. Опровергая основные положения ¶ статьи, Ревкин пошел на весьма рискованное, если не сказать безумное, ¶ возражение, написав, что если бы марксистское мировоззрение действительно влияло ¶ на строение черепа, то самый вытянутый череп был бы у товарища Сталина, ибо ¶ именно он обладает самым последовательным марксистским мировоззрением. "Между ¶ тем, -- утверждал Ревкин, -- стоит посмотреть на любую ¶ фотографию товарища Сталина, чтобы убедиться в абсурдности доводов К. Ушастого". ¶ Ревкин предлагал Привлечь Фигурина к строгой партийной ответственности. Записку ¶ он эту направил в обком, но копию ее еще выше, в ЦК. Не дожидаясь ответа, он ¶ решил проявить свою власть на месте. Он написал короткую записку: "Т. Фигурин, ¶ вам необходимо срочно зайти в РК для выяснения некоторых обстоятельств". Записку ¶ отправил с шофером Мотей и стал ждать ответа. Он явно нервничал и ни на чем не ¶ мог сосредоточить внимание. Мотя вернулась через сорок минут. ¶ -- Почему так долго? -- напустился на нее Ревкин. Ответить она не успела. Вслед ¶ за ней вошли два ¶ рослых молодых человека в штатском, и один из них, улыбнувшись, спросил: ¶ -- Где у вас оружие, хозяин? ¶ И самое удивительное, Ревкин не спросил у них никаких документов, сразу показал ¶ на ящик стола, в котором лежал револьвер. ¶ В сопровождении молодых людей Ревкин вышел в приемную. ¶
¶

После обработки Корректором со всеми опциями и задания отступа 1-й строки абзаца получаем следующее:

С Ревкиным случилось вот что. Он эту ученую статью кандидата Ушастого воспринял как выпад против себя лично. Он перед зеркалом даже вертелся тайком от Аглаи, вершками обмеривал свой собственный череп от затылка до подбородка и от уха до уха, и измерения эти оказались удручающими для него. А удлинить свою голову при помощи спецпарика он, конечно, не мог, ибо народ его знал таким, каков он есть, и терять свой авторитет он не желал. Однако он понял, что власть его уходит из рук, как песок, и потому решился на самый отчаянный шаг. Он определенно заявил, что подобного безобразия терпеть не будет, и послал в обком статью Ушастого и свою докладную записку, в которой называл статью псевдонаучной и шарлатанской, утверждал, что она инспирирована, конечно, новым начальником Тех Кому Надо, который с самого начала ведет себя вызывающе, игнорирует партийные органы и тем самым противопоставляет свое Учреждение партии. Опровергая основные положения статьи, Ревкин пошел на весьма рискованное, если не сказать безумное, возражение, написав, что если бы марксистское мировоззрение действительно влияло на строение черепа, то самый вытянутый череп был бы у товарища Сталина, ибо именно он обладает самым последовательным марксистским мировоззрением. «Между тем, — утверждал Ревкин, — стоит посмотреть на любую фотографию товарища Сталина, чтобы убедиться в абсурдности доводов К. Ушастого». Ревкин предлагал Привлечь фигурина к строгой партийной ответственности. Записку он эту направил в обком, но копию ее еще выше, в ЦК. Не дожидаясь ответа, он решил проявить свою власть на месте. Он написал короткую записку: «Т. фигурин, вам необходимо срочно зайти в РК для выяснения некоторых обстоятельств». Записку отправил с шофером Мотей и стал ждать ответа. Он явно нервничал и ни на чем не мог сосредоточить внимание. Мотя вернулась через сорок минут. ¶

— Почему так долго? — напустился на нее Ревкин. Ответить она не успела. Вслед за ней вошли два рослых молодых человека в штатском, и один из них, улыбнувшись, спросил: ¶

— Где у вас оружие, хозяин? ¶

И самое удивительное, Ревкин не спросил у них никаких документов, сразу показал на ящик стола, в котором лежал револьвер. ¶

В сопровождении молодых людей Ревкин вышел в приемную. ¶

¶

Если же в данном примере в самом верху удалить пустой абзац, то результат будет чуть-чуть иной для верхней строки (т.к. обработка идет от 1-го найденного шаблона):

С Ревкиным случилось вот что. Он эту ученую статью кандидата Ушастого воспринял ¶

как выпад против себя лично. Он перед зеркалом даже вертелся тайком от Аглаи, вершками обмеривал свой собственный череп от затылка до подбородка и от уха до уха, и измерения эти оказались удручающими для него. А удлинить свою голову при помощи спецпарика он, конечно, не мог, ибо народ его знал таким, каков он есть, и терять свой авторитет он не желал. Однако он понял, что власть его уходит из рук, как песок, и потому решился на самый отчаянный шаг. Он определенно заявил, что подобного безобразия терпеть не будет, и послал в обком статью Ушастого и свою докладную записку, в которой называл статью псевдонаучной и шарлатанской, утверждал, что она инспирирована, конечно, новым начальником Тех Кому Надо, который с самого начала ведет себя вызывающе, игнорирует партийные органы и тем самым противопоставляет свое Учреждение партии. Опровергая основные положения статьи, Ревкин пошел на весьма рискованное, если не сказать безумное, возражение, написав, что если бы марксистское мировоззрение действительно влияло на строение черепа, то самый вытянутый череп был бы у товарища Сталина, ибо именно он обладает самым последовательным марксистским мировоззрением. «Между тем, — утверждал Ревкин, — стоит посмотреть на любую фотографию товарища Сталина, чтобы убедиться в абсурдности доводов К. Ушастого». Ревкин предлагал Привлечь фигурина к строгой партийной ответственности. Записку он эту направил в обком, но копию ее еще выше, в ЦК. Не дожидаясь ответа, он решил проявить свою власть на месте. Он написал короткую записку: «Т. фигурин, вам необходимо срочно зайти в РК для выяснения некоторых обстоятельств». Записку отправил с шофером Мотей и стал ждать ответа. Он явно нервничал и ни на чем не мог сосредоточить внимание. Мотя вернулась через сорок минут. ¶

— Почему так долго? — напустился на нее Ревкин. Ответить она не успела. Вслед за ней вошли два рослых молодых человека в штатском, и один из них, улыбнувшись, спросил: ¶

— Где у вас оружие, хозяин? ¶

И самое удивительное, Ревкин не спросил у них никаких документов, сразу показал на ящик стола, в котором лежал револьвер. ¶

В сопровождении молодых людей Ревкин вышел в приемную. ¶

¶

Не во всех случаях этот 2-й режим все корректно определит (Названия заголовков, как правило, не имеют точек в конце абзаца). Но, тем не менее, это лучше, чем ничего.

Режим «По “пустому” абзацу»:

Еще один вариант распространенных в сети текстов – строки одного параграфа разорваны абзацем, а сами параграфы отделены друг от друга одним или несколькими «пустыми» абзацами. Такие тексты получаются при копировании текста из многих pdf-журналов. Режим «По “пустому” абзацу» - программа ищет блок текста (как правило разорванные строки одного абзаца) по идентификаторам – «пустой» абзац. Найденное склеивается в один абзац.

Например, есть текст:

Фрагментация и файловые системы Unix¶

¶
Те, кто уже сталкивался с существенным снижением производительности¶
при работе с дисками в Windows, из-за их сильной фрагментированности, часто¶
задаются вопросом – нужно ли им проводить некоторый периодический анализ и¶
дефрагментацию их Unix-систем. В большинстве случаев ответ – нет. Большинство¶
Unix-систем не подвержены фрагментации в той же степени, что и FAT или¶
NTFS. Однако, прежде чем давать быстрый ответ “да” или “нет”, давайте¶
взглянем на то, что такое фрагментация и по какой причине Unix¶
системы не так ей подвержены, в отличие от Windows-систем и что¶
делать, если Unix-система фрагментирована.¶

¶ ¶ Что вызывает фрагментацию?¶

¶
Фрагментация происходит, когда система не может расположить целый¶
файл на одном непрерывном участке диска. Вместо этого система разбивает¶
файл на несколько фрагментов, которые записываются в разные места на диске и¶
система затем должна поддерживать некую структуру файловой¶
системы для отслеживания мест, где расположены все фрагменты файла.¶

После обработки Корректором текст выглядит уже так:

Фрагментация и файловые системы Unix¶

¶ Те, кто уже сталкивался с существенным снижением производительности при работе с дисками в Windows, из-за их сильной фрагментированности, часто задаются вопросом – нужно ли им проводить некоторый периодический анализ и дефрагментацию их Unix-систем. В большинстве случаев ответ — нет. Большинство Unix-систем не подвержены фрагментации в той же степени, что и FAT или NTFS. Однако, прежде чем давать быстрый ответ «да» или «нет», давайте взглянем на то, что такое фрагментация и по какой причине Unix-системы не так ей подвержены, в отличие от Windows-систем и что делать, если Unix-система фрагментирована. ¶

¶

¶

Что вызывает фрагментацию?¶

¶

Фрагментация происходит, когда система не может расположить целый файл на одном непрерывном участке диска. Вместо этого система разбивает файл на несколько фрагментов, которые записываются в разные места на диске и система затем должна поддерживать некую структуру файловой системы для отслеживания мест, где расположены все фрагменты файла. ¶

Иногда некоторые абзацы не отделены от других пустым абзацем, и являются самостоятельными абзацами, не разорванными символом абзаца. К сожалению, они склеются с вышестоящими абзацами. Чтобы этого не произошло, надо просмотреть весь текст, и перед «целыми», неразорванными абзацами вставить пустой абзац, если его нет. Тогда Корректор корректно обработает текст и склеит нужные абзацы в один. Так, что лучше немного потратить время на просмотр текста и вставку пустых абзацев в нужные места, если их там нет, нежели часами вручную склеивать строки в абзацы!

Совет: Без надобности не включайте опцию Корректора «Собрать разорванные символом ¶ строки абзаца в один абзац»! Иначе Корректор может склеить и «нормальные» абзацы воедино! Это особенно касается 1-го режима работы этой опции.

II.2. Настройки Корректора Текста

Настройка формы Корректора сохраняется в файл *corrector.txt*, который помещается в жестко заданную папку *config* пакета *OpenOffice.org*, откуда и загружается. Для Unix-подобных систем — это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows — это папка *C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools*, где XXX — ваш логин (имя пользователя).

II.3. Как запускать Корректор?

Создайте на уже созданной ранее панели инструментов (в моем случае FictionBook) или на любой удобной для вас кнопку для вызова Корректора Текста из макросов по следующему пути: Макрос **OpenOffice.org / Мои макросы / OOoFBTools / TextCorrector**. Нужный здесь макрос — **TextCorrector**. Как вытаскивать макросы на кнопки подробно описано в этой справке в [п.2.2](#). Сделайте по аналогии.

А можно просто воспользоваться соответствующей кнопкой на панели инструментов или пунктом меню OOoFBTools → Text Corrector... И кнопка, и пункт меню автоматически создаются при установке расширения.

II.4. Это важно!

При работе с Корректором Текста важно помнить следующее: Обработку текста лучше проводить **ДО** того, как вы «отформатировали» стилими документ для конвертации его в формат fb2! Т.к., если в Корректоре выбрана опция «Удалить пустые абзацы...», и в «отформатированном» документе они есть, то после удаления этих пустых абзацев, абзацы, следующие за ними, «теряют» свои стили. Если это абзацы-уровни (Заголовки, Части книги), которые должны быть взяты в секции <section>, то корректность уровней нарушается, что в результирующем файле fb2 проявится, как некорректность расстановки секций. И дело тут не в конверторе – просто после удаления пустых абзацев Корректором и утраты следующим абзацем своего стиля, документ получается «отформатированным» стилими не корректно. Вводить в Корректор логику, отлавливающую стили последующих абзацев и восстанавливающую их – значит сделать то, что Корректор перестанет быть универсальным.

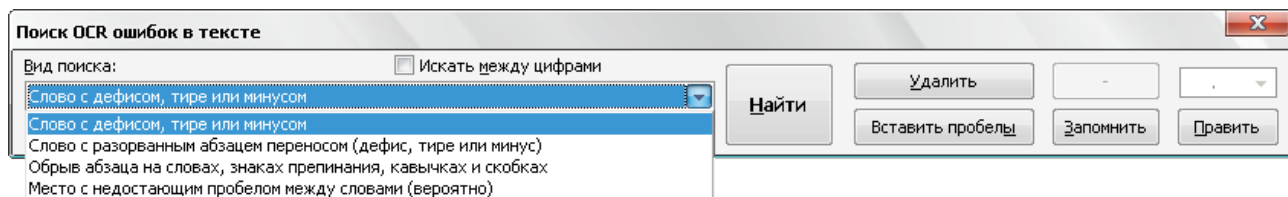
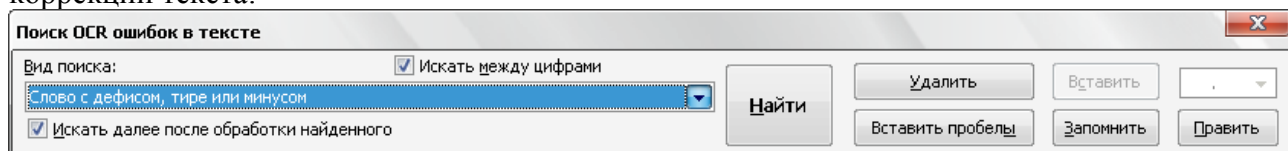
Конечно, можно подправить документ нужными стилими, и экспорт пройдет «на ура», но зачем двойная работа.

Поэтому, оговорюсь еще раз: Обработку текста лучше проводить **ДО** того, как вы «отформатировали» стилими документ для конвертации его в формат fb2!

III. Инструменты для «ручного» исправления OCR-ошибок

III.1. Поиск OCR ошибок в тексте

Эти инструменты предназначены для ручного поиска и самостоятельной корректировки слов. Они часто бывают полезны после OCR. Многие ошибки в тексте можно исправить с помощью автоматического Корректора Текста. Но он не все корректирует «до конца». Например, при корректировании разрыва абзацем дефиса и переноса слов, в зависимости от выбора режима корректировки, часто остаются не обработанными либо слова с дефисами, либо слова с переносами, т.к. есть случаи, которые невозможно программно отследить, и только человек способен правильно решить, что это – слово с дефисом или же слово с разорванным переносом. Для такого случая и подобных и созданы ручные инструменты коррекции текста:



Диалог «Поиск OCR ошибок в тексте» выполнен, как *немодальный*, т.е. с него легко «перейти» на найденного им слово для ручной коррекции, не закрывая при этом этот диалог. Но эта возможность может понадобиться крайне редко, т.к. предусмотрена кнопка «Удалить» («Вставить», «Обработать» - название и назначение меняется в зависимости от вида поиска), нажатие на которой вносит необходимые изменения в найденный текст.

Диалог «запоминает» свое местоположение на экране, и при следующем его вызове он отобразится в том месте, где вы его закрыли в прошлый раз. Это удобно. Его можно расположить над меню и панелями инструментов ООо (сделано по умолчанию), и он не будет

«мешать» видеть текст во время поиска.

Поиск ручным корректором и внесение изменений осуществляется везде – в тексте, в таблицах, в текстовых врезках и в сносках.

Поиск начинается от позиции курсора в документе и идет «по-кругу».

Корректировка найденного текста по нажатию кнопки (справа от кнопки «Найти») работает «интеллектуально», т.к., если после того, как вы нашли что-то (этот текст выделяется), а потом перешли курсором в другое место, или выделили случайно другое место в тексте, и нажали на корректирующую кнопку, то, если нет выделения – ничего не будет откорректировано. А если выделение есть, то в нем будет проведен поиск OCR-ошибки, согласно выбранного вида поиска, и в случае обнаружения, произойдет коррекция текста.

Конечно, все случаи «защиты от глупости» не предусмотреть...

Опция «Искать далее после обработки найденного»: Если она выключена, то после того, как инструмент нашел ошибку OCR, и вы ее обработали с помощью кнопок обработки, текстовый курсор остается на найденном, и чтобы найти новую ошибку OCR, надо нажать кнопку «Найти». И так – до конца книги. Если же опция «Искать далее после обработки найденного» **включена**, то после обработки найденной ошибки **автоматически** происходит **переход на новую ошибку** в тексте! Это очень удобно!

Поиск слов с дефисами, тире или минусом

Инструмент позволяем искать в тексте слова с дефисами, например: «кое-что», «серый-серый», «по-своему», «ты-ли», «ангел-хранитель» и т.д.

Польза этого инструмента – после работы Корректора Текста в автоматическом режиме можно найти слова, где из разорванных переносов был удален только знак абзаца, а дефис (или «минус») – оставлен. Если и он должен быть удален, то теперь это можно сделать вручную. Например, нашли: «замеча-тельно». Здесь перенос – явно лишний. Удаляем...

С другой стороны, можно найти другие слова с тире, дефисом или минусом и заменить эти символы «на ваш вкус». Обрабатывается и неразрывный дефис.

Опция «**Искать между цифрами**» нужна для возможности включить/отключить поиск минуса, тире и дефиса между цифрами. Очень часто в книгах встречаются такие конструкции: 12-34, 1972-2010 и т.д. (В христианской литературе – это ссылка на стихи Библии, например, Мф. 5:1-8. В исторических книгах – это указание на отрезок времени. И т.д.). В таких книгах, как правило, минус, тире или дефис между цифрами – это не ошибка OCR, а норма. При вычитке инструмент часто находит такие цифры с дефисами, что утомляет. Чтобы пропустить эти «правильные» цифры и не тратить на них время нужно отключить опцию «**Искать между цифрами**».

Для поиска используется шаблон регулярных выражений:

При включенной опции «**Искать между цифрами**»:

`<[:alnum:]+[-" & chr(8209) & chr(8211) & chr(8212) & "][:alnum:]+`

При выключенной опции «**Искать между цифрами**»:

`<[:alpha:]+[-" & chr(8209) & chr(8211) & chr(8212) & "][:alpha:]+`

Кнопка **Вставить пробелы** активна только для этого режима. При нажатии на ней, если найдено слово, содержащие дефис, минус или тире, то слева и справа от этого минуса, дефиса или тире ставятся пробелы. Это полезно для «слипшихся» слов и тире... Например, найдено слово «Ты–нет!». После нажатия кнопки **Вставить пробелы**, получаем «Ты – нет!», как и должно было быть.

В этом режиме поиска предусмотрена возможность Запоминать найденное «хорошее» дефисное слово (Кнопка **Запомнить**), и Редактировать файл списка этих дефисных слов (Кнопка **Править**). Слова «запоминаются» в .txt-файл *words_defis_list.txt*, который создается

в папке config пакета OOo. Для Unix-подобных систем – это папка /home/XXX/.openoffice.org3/user/config/OOoFBTools. Для Windows – это папка C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools, где XXX – ваш логин (имя пользователя).

Как это работает? Когда найдено очередное слово с минусом, неразрывным дефисом, дефисом или тире, то, чтобы инструмент не останавливался в следующий раз на этом же слове, его можно «Запомнить». И теперь оно будет «благополучно» пропускаться при поиске.

Предусмотрены защиты от «глупости»: слово не будет «запоминаться», если:

- нет выделения в тексте;
- выделено слово, не содержащее ни минуса, ни тире, ни дефиса, ни неразрывного дефиса;

- случайно пользователем выделено несколько абзацев.

Во всех этих случаях будет выдано соответствующее предупреждение.

При Правке файла списка дефисных слов (нажатие кнопки **Править**) файл `words_defis_list.txt` будет запущен в ассоциированном с расширением .txt в вашей системе текстовом редакторе. Лучше ассоциировать .txt с редактором, который поддерживает Unicode и корректно «видит» перевод строк (абзацы). В Windows **notepad.exe** для этого не подходит. В Windows я пользуюсь **Notepad++.exe**.

Поиск слов, в которых абзац разорвал перенос (дефис, тире или минус)

Этот инструмент позволяет искать в документе разрыва дефисов и слов. Например: «давным-¶», «из-¶», «пересмотре-¶». Если после знака переноса стоит один или пробелов, то такие слова тоже находятся. Например: «давным- ¶», «из- ¶», «пересмотре- ¶»

Обрабатывается и неразрывный дефис.

Для поиска используется шаблон регулярных выражений:

```
"\<[:alnum:]+\[-" & chr(8209) & chr(8211) & chr(8212) & "]"$ \<[:alnum:]+\[-" & chr(8209) & chr(8211) & chr(8212) & "]"[:space:]$"
```

Поиск обрыва абзаца на словах, знаках препинания, кавычках и скобках

Например, есть текст:

Это пример того, как можно «найти»¶
разрыв предложения на кавычках,¶
запятой, (тексте в скобках) ¶
что часто бывает нужно после OCR.

Еще пример:

Здесь – разрыв после пробела и тире –¶

А здесь – разрыв после 2-х пробелов, минуса и пробела - ¶

Для поиска используется шаблон регулярных выражений:

```
"\<[:alnum:]+$ \<[:alnum:]+\{ \},»"'"';:]$ \<[:alnum:]+\{ \},»"'"';:] + \{ \},»"'"';:]$ [:space:]+  
[-" & chr(8209) & chr(8211) & chr(8212) & "]"$ & "[:space:]+\[-" & chr(8209) & chr(8211) & chr(8212) & "]"[:space:]+$"
```

Т.е. отлавливаются не только разрывы на запятых, скобках и словах, но и на минуса, тире и дефисах, слева от которых стоит 1 или более пробелов, а справа – ни один или множество пробелов (частая ошибка OCR).

Кнопка «Обработать» позволяет удалить разрыв предложения и вставить пробел.

В этом варианте поиска активируется выпадающий список знака пунктуации, который можно выбрать, и при нажатии кнопки «Вставить» он будет вставлен «на свое место».

Кнопка «Склеить» позволяет просто удалить разрыв, не вставляя пробел. Это полезно, т.к. часто при OCR (особенно в FR10) слово оказывается разорванным не как, например «по-

«шел», а «пошел». Поэтому «Склейка» склеивает разорванное слово.

Поиск вероятного недостающего пробела между словами

Ищет слова с вероятным отсутствием пробела между ними по знакам препинания, закрывающим скобкам.

Опция «**Искать между цифрами**» нужна для возможности включить/отключить поиск недостающего пробела между цифрами. Очень часто в книгах встречаются такие конструкции: 12,34, 1972:2010 и т.д. (В христианской литературе – это ссылка на стихи Библии, например, Ин. 3:16. В математических книгах и книгах по программированию – это указание на числовые диапазоны, просто дробные числа. И т.д.). В таких книгах, как правило, отсутствие пробела между цифрами – это не ошибка OCR, а норма. При вычитке инструмент часто находит такие цифры "без пробела", что утомляет. Чтобы пропустить эти «правильные» цифры и не тратить на них время нужно отключить опцию «**Искать между цифрами**».

Для поиска используется шаблон регулярных выражений:

При включенной опции «Искать между цифрами»:

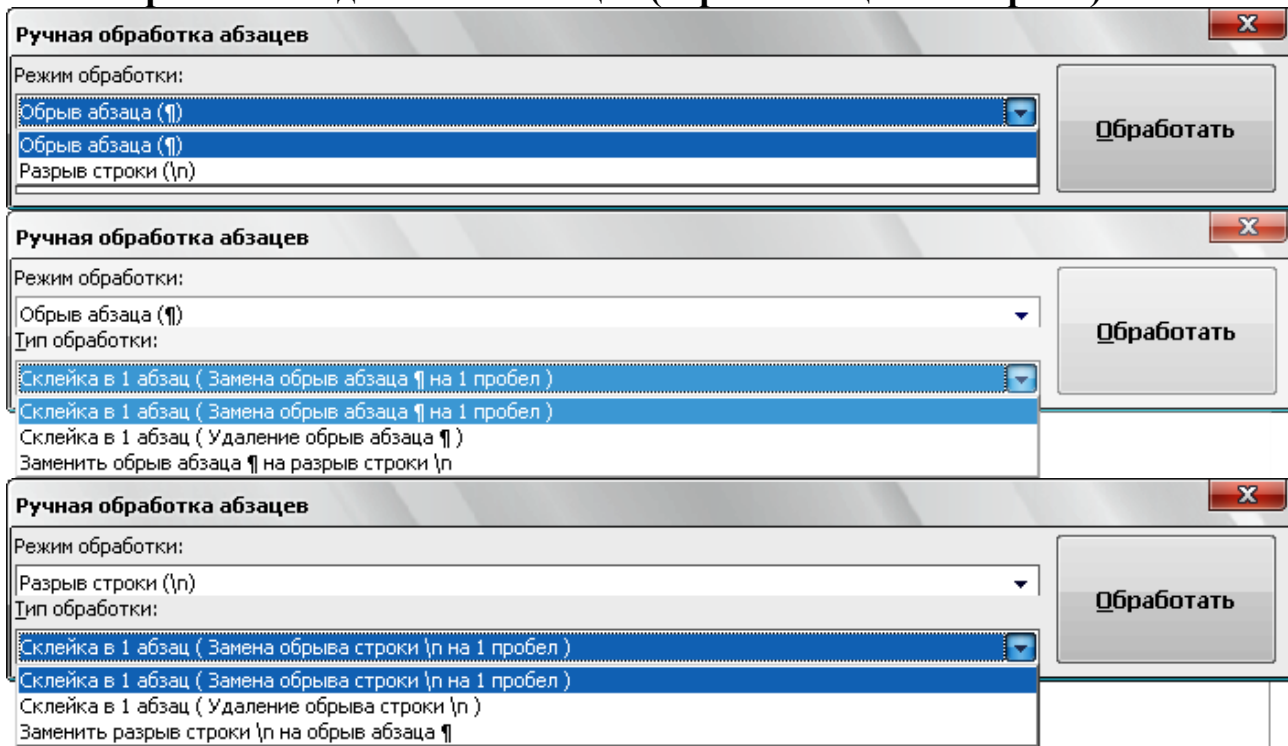
```
\\<[:alnum:]+[)]\\(\\{\\[,!?.>”\\'";«“”[:alpha:]+\\<[:alpha:]+[:digit:]+\\<[:digit:]+[:alpha:]+
```

При выключенной опции «Искать между цифрами»:

```
\\<[:alpha:]+()\\)(\\{[,!?!...»”\\'";«“”[:alpha:]+\\<[:alpha:]+[:digit:]+\\<[:digit:]+[:alpha:]+
```

При корректировании в этом режиме поиска программа сама определяет, где поставить пробел – слева или справа от найденного знака препинания, скобки или кавычек.

III.2. Обработка выделенных абзацев (обрыв абзаца или строки)



Часто требуется не автоматическая обработка текста, а ручная некоторых фрагментов.

Данный набор инструментов может работать как с одним выделенным фрагментом текста, так и с несколькими выделенными областями текста. Если нет ни одного выделенного фрагмента, то инструменты могут обработать весь документ (появляется соответствующее окно с вопросом).

Выделять можно только Текст. Инструменты не работают с Врезками и Таблицами.

Например, есть текст с обрывом абзацев (показаны скрытые символы):

```
Death·in·the·Afternoon¶
¶
To¶
Pauline¶
¶
Chapter·One¶
¶
At·the·first·bullfight·I·ever·went·to·I·expected·to·be·horrified,
and·perhaps·sickened·by·what·I·had·been·told·would·happen,
to·the·horses·.·Every·thing·I·had·read·about·the·bull·ring·in-
sisted·on·that·point;·most·people·who·wrote·of·it·condemned,
bullfighting·outright·as·a·stupid·brutal·business,·but·even·those
that·spoke·well·of·it·as·an·exhibition·of·skill·and·as·a·spectacle,
deplored·the·use·of·the·horses·and·were·apologetic·about·the
whole·thing·.·The·killing·of·the·horses·in·the·ring·was·consid-
ered·indefensible·.·I·suppose,·from·a·modern·moral·point·of
view,·that·is,·a·Christian·point·of·view,·the·whole·bullfight·is
indefensible;·there·is·certainly·much·cruelty,·there·is·always
danger,·either·sought·or·unlooked·for,·and·there·is·always
death,·and·I·should·not·try·to·defend·it·now,·only·to·tell·honest-
ly·the·things·I·have·found·true·about·it·.·To·do·this·I·must
be·altogether·frank,·or·try·to·be,·and·if·those·who·read·this
decide·with·disgust·that·it·is·written·by·some·one·who·lacks
their·the·readers',·fineness·of·feeling·I·can·only·plead·that·this
may·be·true·.·But·whoever·reads·this·can·only·truly·make·such
a·judgment·when·he,·or·she,·has·seen·the·things·that·are·spoken
of·and·knows·truly·what·their·reactions·to·them·would·be·.¶
¶
Once·I·remember·Gertrude·Stein·talking·of·bullfights·spoke
of·her·admiration·for·Joselito·and·showed·me·some·pictures·of
him·in·the·ring·and·of·herself·and·Alice·Toklas·sitting·in·the
first·row·of·the·wooden·barreras·at·the·bull·ring·at·Valencia¶
¶
1¶
¶
DEATH·IN·THE·AFTERNOON¶
¶
with·Joselito·and·his·brother·Gallo·below,·and·I·had·just·come
from·the·Near·East,·where·the·Greeks·broke·the·legs·of·their
baggage·and·transport·animals·and·drove·and·shoved·them
off·the·quay·into·the·shallow·water·when·they·abandoned·the
city·of·Smyrna,·and·I·remember·saying·that·I·did·not·like·the
bullfights·because·of·the·poor·horses·.·I·was·trying·to·write·then
and·I·found·the·greatest·difficulty,·aside·from·knowing·truly
what·you·really·felt,·rather·than·what·you·were·supposed·to
```

В этом тексте есть и обрыв абзаца, и обрыв строки (тэг
 в html-файлах).

Делаем следующее:

1. Выделяем те строки, где есть разрыв строки:

Death in the Afternoon¶

¶

To¶

Pauline¶

¶

Chapter One¶

¶

At the first bullfight I ever went to I expected to be horrified, and perhaps sickened by what I had been told would happen to the horses. Every tiling I had read about the bull ring insisted on that point; most people who wrote of it condemned bullfighting outright as a stupid brutal business, but even those that spoke well of it as an exhibition of skill and as a spectacle deplored the use of the horses and were apologetic about the whole thing. The killing of the horses in the ring was considered indefensible. I suppose, from a modern moral point of view, that is, a Christian point of view, the whole bullfight is indefensible; there is certainly much cruelty, there is always danger, either sought or unlooked for, and there is always death, and I should not try to defend it now, only to tell honestly the things I have found true about it. To do this I must be altogether frank, or try to be, and if those who read this decide with disgust that it is written by some one who lacks their, the readers', fineness of feeling I can only plead that this may be true. But whoever reads this can only truly make such a judgment when he, or she, has seen the things that are spoken of and knows truly what their reactions to them would be.

¶

Запускаем инструменты **Ручная обработка абзацев**, выбираем Режим обработки: **Разрыв строки (\n)** и Тип обработки: **Склейка в 1 абзац (Замена обрыва строки \n на 1 пробел)**. После нажатия **Обработать**, получили следующее:

Death in the Afternoon¶

¶

To¶

Pauline¶

¶

Chapter One¶

¶

At the first bullfight I ever went to I expected to be horrified and perhaps sickened by what I had been told would happen to the horses. Every tiling I had read about the bull ring insisted on that point; most people who wrote of it condemned bullfighting outright as a stupid brutal business, but even those that spoke well of it as an exhibition of skill and as a spectacle deplored the use of the horses and were apologetic about the whole thing. The killing of the horses in the ring was consid-ered indefensible. I suppose, from a modern moral point of¶

view, that is, a Christian point of view, the whole bullfight is¶

indefensible; there is certainly much cruelty, there is always¶

danger, either sought or unlooked for, and there is always¶

death, and I should not try to defend it now, only to tell hon-estly the things I have found true about it. To do this I must be altogether frank, or try to be, and if those who read this decide with disgust that it is written by some one who lacks their, the readers', fineness of feeling I can only plead that this may be true. But whoever reads this can only truly make such a judgment when he, or she, has seen the things that are spoken of and knows truly what their reactions to them would be.¶

¶

Once I remember Gertrude Stein talking of bullfights spoke¶

of her admiration for Joselito and showed me some pictures of¶

him in the ring and of herself and Alice Toklas sitting in the¶

first row of the wooden barreras at the bull ring at Valencia¶

¶

Теперь выделяем все строки, разорванные абзацев, причем в каждое выделение должны попасть только те строки, которые должны составить один целый абзац:

Death in the Afternoon¶

¶

To¶

Pauline¶

¶

Chapter One¶

¶

At the first bullfight I ever went to I expected to be horrified and perhaps sickened by what I had been told would happen to the horses. Every tiling I had read about the bull ring in sisted on that point; most people who wrote of it condemned bullfighting outright as a stupid brutal business, but even those that spoke well of it as an exhibition of skill and as a spectacle deplored the use of the horses and were apologetic about the whole thing. The killing of the horses in the ring was consid-ered indefensible. I suppose, from a modern moral point of¶

view, that is, a Christian point of view, the whole bullfight is¶

indefensible; there is certainly much cruelty, there is always¶

danger, either sought or unlooked for, and there is always¶

death, and I should not try to defend it now, only to tell hon-estly the things

I have found true about it. To do this I must be altogether frank, or try to be,

and if those who read this decide with disgust that it is written by some one

who lacks their, the readers', fineness of feeling I can only plead that this

may be true. But whoever reads this can only truly make such a judgment when he,

or she, has seen the things that are spoken of and knows truly what their

reactions to them would be.¶

¶

Once I remember Gertrude Stein talking of bullfights spoke¶

of her admiration for Jogelito and showed me some pictures of¶

him in the ring and of herself and Alice Toklas sitting in the¶

first row of the wooden barreras at the bull ring at Valencia¶

¶

Для инструмента **Ручная обработка абзацев**, выбираем Режим обработки: **Обрыв абзаца (¶)** и Тип обработки: **Склейка в 1 абзац (Замена обрыва строки \n на 1 пробел)**. После нажатия **Обработать**, получили следующее:

Death in the Afternoon¶

¶

To¶

Pauline¶

¶

Chapter One¶

¶

At the first bullfight I ever went to I expected to be horrified and perhaps sickened by what I had been told would happen to the horses. Every tiling I had read about the bull ring insisted on that point; most people who wrote of it condemned bullfighting outright as a stupid brutal business, but even those that spoke well of it as an exhibition of skill and as a spectacle deplored the use of the horses and were apologetic about the whole thing. The killing of the horses in the ring was considered indefensible. I suppose, from a modern moral point of view, that is, a Christian point of view, the whole bullfight is indefensible; there is certainly much cruelty, there is always danger, either sought or unlooked for, and there is always death, and I should not try to defend it now, only to tell honestly the things I have found true about it. To do this I must be altogether frank, or try to be, and if those who read this decide with disgust that it is written by some one who lacks their, the readers', fineness of feeling I can only plead that this may be true. But whoever reads this can only truly make such a judgment when he, or she, has seen the things that are spoken of and knows truly what their reactions to them would be.¶

¶

Once I remember Gertrude Stein talking of bullfights spoke of her admiration for Joselito and showed me some pictures of him in the ring and of herself and Alice Toklas sitting in the first row of the wooden barreras at the bull ring at Valencia¶

¶

Что и требовалось. Разорванные строки 2-х абзаце мы склеили, каждые в свой абзац. Разобраться с остальными Типами обработки этих инструментов не сложно.

III.3. Инструменты работы с примечаниями и сносками

Пожалуй, самым нудным и утомительным в вычитке текста является обработка и создание сносок из примечаний. Это – постоянные «прыжки» по тексту – с места текста примечания, вырезая его текст в буфер – на «его» номер в главе, удаление этого номера, вставки сноска, потом вставка из буфера примечания... И так – «до умопомрачения».

Набор инструментов для работы со сносками созданы как раз, чтобы облегчить этот процесс в полуавтоматическом режиме. Доступ к ним – либо через меню **OooFBTools**, либо – через панель инструментов **Генерация сноска** (иконки со временем могут измениться):



1. Генерация сноска.
2. Вставка закладки для списка примечаний.
3. Вставка закладки для главы с № примечаний.
4. Преобразование № примечаний в верхнем индексе по шаблону.

5. Удаление всех служебных закладок.
6. Интерактивное преобразование номеров примечаний в верхнем индексе по шаблону
7. Нумерация выделенных абзацев.

И инструмент **Генерация сносок** и инструмент **Преобразование № примечаний в верхнем индексе по шаблону** активно используют работу с буфером обмена (Cut, Paste) и реальным положением видимого курсора экрана! Поэтому, пока программа не завершит работу, ничего не делайте ни с мышкой, ни с клавиатурой!

Инструмент **Преобразование № примечаний в верхнем индексе по шаблону** работает и с текстом, и с таблицами, т.е. цифра в верхнем индексе может преобразовываться в шаблонный вид и из ячеек таблиц. Врезки – игнорируются.

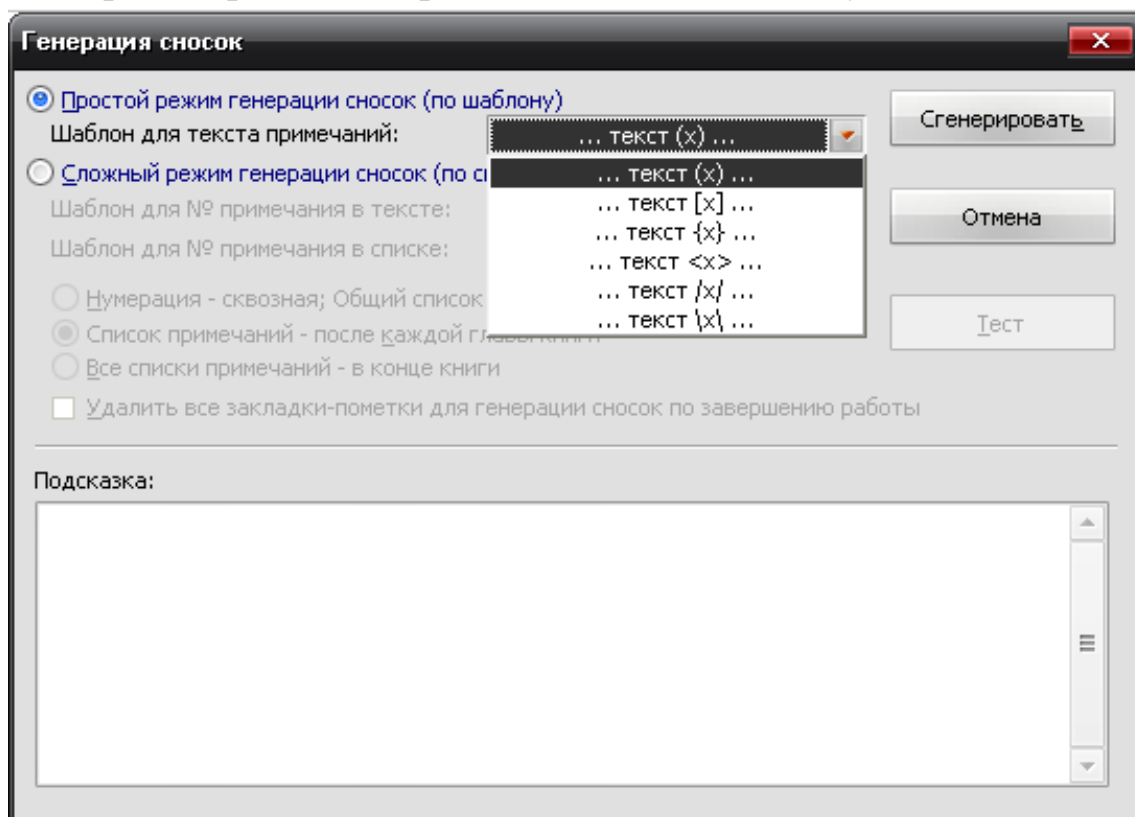
Инструменты **Генерация сносок** и **Нумерация выделенных абзацев** работают только с текстом документа. Текстовые Врезки – игнорируются, т.к. OOoWrither не позволяет в них вставлять сноски. Таблицы тоже игнорируются (из-за сложности проверки выхода курсора за пределы Таблицы и некорректности получаемого результата).

Теперь – подробнее о каждом инструменте.

1. Генерация сносок

Есть 2 основных режима генерации сносок: Простой и Сложный.

1.1. Простой режим генерация сносок (по шаблону)



Генерация сносок в этом режиме возможно в выделенных фрагментах текста (их может быть сколько угодно), либо во всем документе.

При запуске генерации сносок, если нет ни одного выделения в тексте, появится окно с вопросом, обрабатывать ли весь текст? Если нажать кнопку ОК, то будет обработан весь

документ. Если – Отмена, то работа остановится, и вы можете выделить нужные области текста. Если выделения есть, то программа ничего не запрашивает, а производит обработку всех этих выделений.

В каких случаях используется простой режим генерации сносок?

Очень часто источник цитаты или мысли, на которую ссылается автор книги, находится в скобках (круглых, квадратных и т.д.). А это как раз и есть «кандидат» на сноски.

Например, есть текст книги (Рыбаков, Язычество древней Руси):

К предметам, связанным с бытом волхвов, следует отнести находки неолитических кремневых орудий и стрел. Такая находка была сделана В. А. Городцовым в вятическом кургане у с. Барыбина. Исследователь сопоставил её с обычаем хоронить колдунов с "громовыми стрелами". (Городцов В. А. Археология. Каменный период. М., 1923, с.77-78.) Обычно под громовыми стрелами подразумевают фульгуриты, но вполне вероятно, что к ним причисляли и изделия каменного века. Подобные находки есть и в других местах. Особый интерес представляет кремневый наконечник копья, найденный в Новгороде в слоях рубежа XIII - XIV вв. Кремень оправлен в серебро с чернью. М. В. Седова определяет дом, в котором найден этот талисман, как дом волхва, так как в его фундаменте зарыты 4 детских черепа. (Седова М. В. Амулет из древнего Новгорода. - Сов. археология, 1957, № 4, рис. 1. с. 166-167.) Наличие православного восьмиконечного креста на серебряной оправе говорит о любопытном синкретизме представлений этого колдуна-знахаря.

Запустив диалог **Генерации сносок**, выбираем **«Простой режим генерации сносок (по шаблону)»**. А в нем – из выпадающего списка – шаблон для текста сносок (в нашем случае – это текст внутри круглых скобках). Нажав кнопку **Сгенерировать** мы получаем текст, где вместо круглых скобок стоит очередная сноска, а в тексте сноски – текст из круглых скобок:

К предметам, связанным с бытом волхвов, следует отнести находки неолитических кремневых орудий и стрел. Такая находка была сделана В. А. Городцовым в вятическом кургане у с. Барыбина. Исследователь сопоставил её с обычаем хоронить колдунов с "громовыми стрелами". ¹ Обычно под громовыми стрелами подразумевают фульгуриты, но вполне вероятно, что к ним причисляли и изделия каменного века. Подобные находки есть и в других местах. Особый интерес представляет кремневый наконечник копья, найденный в Новгороде в слоях рубежа XIII - XIV вв. Кремень оправлен в серебро с чернью. М. В. Седова определяет дом, в котором найден этот талисман, как дом волхва, так как в его фундаменте зарыты 4 детских черепа. ² Наличие православного восьмиконечного креста на серебряной оправе говорит о любопытном синкретизме представлений этого колдуна-знахаря.

¹Городцов В. А. Археология. Каменный период. М., 1923, с.77-78.

²Седова М. В. Амулет из древнего Новгорода. - Сов. археология, 1957, № 4, рис. 1. с. 166-167.

Этот текст содержит в скобках только ссылки на источники – книги. Но в реальных книгах после OCR встречается ситуация, когда в скобках (или других идентификаторах из шаблона) находятся не только ссылки на авторов книг, но и уточнения, пояснения... Если

запустить Простой режим генерации сносок для всего документа, то и эти уточнения и пояснения тоже обрабатываются, как кандидаты на сноски. А это не то, что мы ожидаем. Для решения этой проблемы есть 2 пути. 1). Круглые скобки вокруг авторов книг заменить на фигурные, или другие идентификаторы из списка шаблона. 2). Выделять только те фрагменты книги, где в скобках находятся действительно кандидаты на сноски, и не выделять те, которые являются уточнениями и т.д., хотя и находятся тоже в таких же скобках. Тогда они и не будут преобразованы в сноски.

Вот пример текста (в нем, для иллюстрации, желтым маркером выделены кандидаты на сноски, а оранжевым – уточнения, которые не должны быть обработаны, как сноски):

Длинные курганы кривичей, насчитывающие иногда свыше двух десятков захоронений, очевидно, являются родовыми усыпальницами, пополнявшимися на протяжении нескольких поколений. Древнейшие длинные курганы появились на Псковщине в VI - VII вв. н. э., а затем они распространились и по Смоленской земле и по Полоцкой (VII - VIII вв.). "Массовое захоронение в валообразных насыпях прекращается в IX в. ..." (Седов В. В. Длинные курганы кривичей, с. 35.)

Не менее интересными представляются своеобразные курганы ильменских словен, называемые сопками - от старого глагола "съпати" - сыпать. (Седов В. В. Новгородские сопки. М., 1970.)

Значит, нам надо курсором выделить только кандидатов на сноски (на картинке – желтый маркер), используя клавишу **Ctrl**:

Длинные курганы кривичей, насчитывающие иногда свыше двух десятков захоронений, очевидно, являются родовыми усыпальницами, пополнявшимися на протяжении нескольких поколений. Древнейшие длинные курганы появились на Псковщине в VI - VII вв. н. э., а затем они распространились и по Смоленской земле и по Полоцкой (VII - VIII вв.). "Массовое захоронение в валообразных насыпях прекращается в IX в. ..." (Седов В. В. Длинные курганы кривичей, с. 35.)

Не менее интересными представляются своеобразные курганы ильменских словен, называемые сопками - от старого глагола "съпати" - сыпать. (Седов В. В. Новгородские сопки. М., 1970.)

Точность выделения не имеет значения – главное – чтобы наши идентификаторы (в данном случае круглые скобки) попали в это выделение.

Теперь, после генерации сносок в Простом режиме генерирования получаем следующее:

Длинные курганы кривичей, насчитывающие иногда свыше двух десятков захоронений, очевидно, являются родовыми усыпальницами, пополнявшимися на протяжении нескольких поколений. Древнейшие длинные курганы появились на Псковщине в VI - VII вв. н. э., а затем они распространились и по Смоленской земле и по Полоцкой (VII - VIII вв.). "Массовое захоронение в валообразных насыпях прекращается в IX в. ..."¹

Не менее интересными представляются своеобразные курганы ильменских словен, называемые сопками - от старого глагола "съпати" - сыпать.²

1 Седов В. В. Длинные курганы кривичей, с. 35.

2 Седов В. В. Новгородские сопки. М., 1970.

Как видим, все получилось, как надо.

Использование выделений в тексте довольно удобно, когда не надо обрабатывать весь документ. Идеально было бы, если бы все идентификаторы кандидатов на сноски отличались бы от круглых или квадратных скобок. Тогда можно просто обработать весь документ, не думая о том, чтобы под обработку не попали и уточнения в круглых скобках, или страницы оригинала в квадратных скобках.

1.2. Сложный режим генерация сносок (по спискам примечаний)

Сложный режим генерации сносок делится на 3 вида (по виду книг с примечаниями):

- 1). Нумерация примечаний в Тексте – сквозная; Общий Список примечаний – в конце книги.
- 2). Список примечаний – после каждой главы книги.
- 3). Все списки примечаний – в конце книги.

Генерация сносок

☐ Простой режим генерации сносок (по шаблону)
 Шаблон для текста примечаний: ... текст (x) ...

☒ Сложный режим генерации сносок (по спискам примечаний)
 Шаблон для № примечания в тексте: ... текст {x} ...
 Шаблон для № примечания в списке: x. Текст ...

☐ Нумерация - сквозная; Общий список примечаний - в конце книги
☒ Список примечаний - после каждой главы книги
☐ Все списки примечаний - в конце книги

☐ Удалить все закладки-пометки для генерации сносок по завершению работы

Подсказка:

Документ не имеет Таблиц и Текстовых Врезок.
 Это - важное условие для корректной работы Теста и Генератора сносок!

Сгенерировать Отмена Тест

Для каждого из этих вариантов расположения блоков текста примечаний можно выбрать Шаблон для номера примечания в тексте:

Генерация сносок

☐ Простой режим генерации сносок (по шаблону)
 Шаблон для текста примечаний: ... текст (x) ...

☒ Сложный режим генерации сносок (по спискам примечаний)
 Шаблон для № примечания в тексте: ... текст {x} ...
 Шаблон для № примечания в списке: x. Текст ...

☐ Нумерация - сквозная; Общий список примечаний - в конце книги
☒ Список примечаний - после каждой главы книги
☐ Все списки примечаний - в конце книги

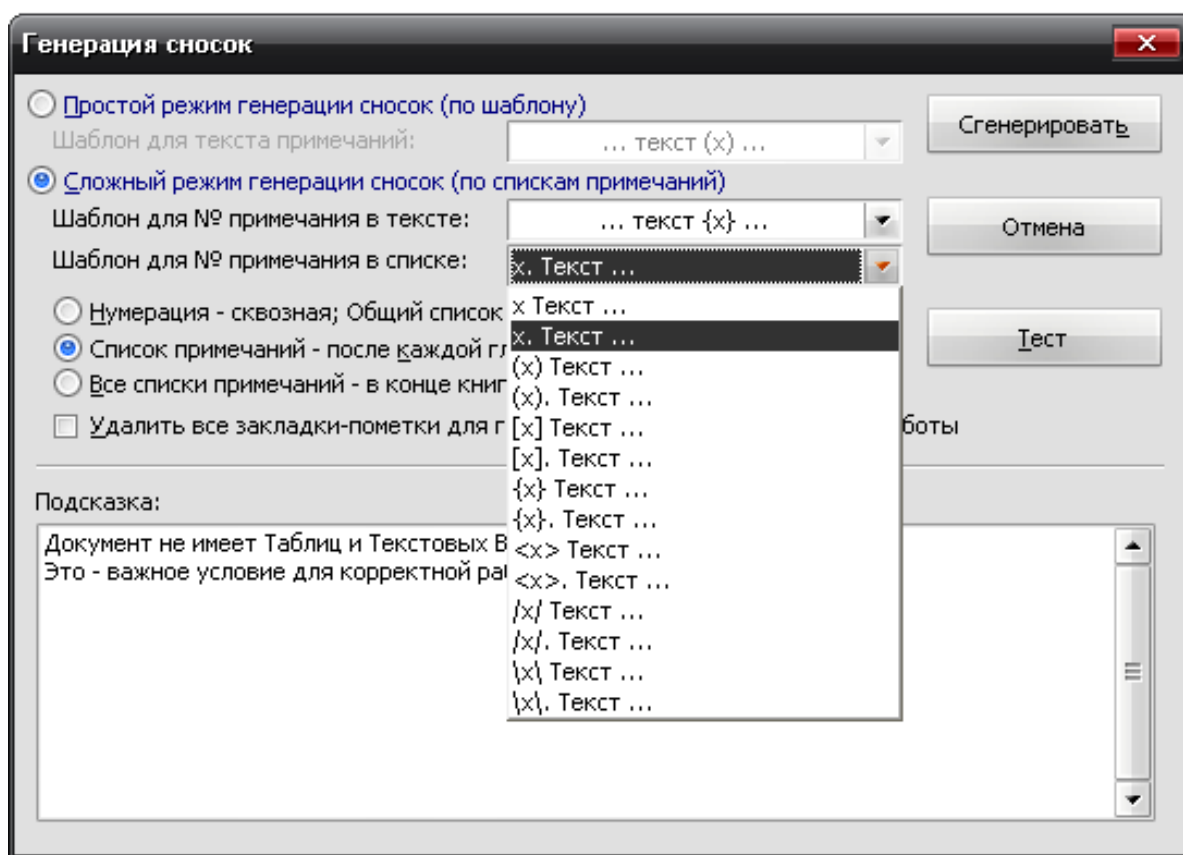
☐ Удалить все закладки-пометки для генерации сносок по завершению работы

Подсказка:

Документ не имеет Таблиц и Текстовых Врезок.
 Это - важное условие для корректной работы Теста и Генератора сносок!

Сгенерировать Отмена Тест

И Шаблон для номера примечания в списке примечаний:



В Сложном режиме генерации сносок обрабатывается весь документ. Обработка выделенных фрагментов текста не предусмотрена, т.к. она избыточна. Нужные зоны поиска задаются закладками (об этих инструментах см. [ниже](#)).

При генерации сносок ищется соответствие номеров примечаний в главах с номерами этих примечаний в списке примечаний по каждой главе. Если такое соответствие не найдено, то в списке примечаний такое примечание *остается*, не удаляется.

1.2.0. Очень ВАЖНО: корректность документа для сложной генерации сносок и/или Теста!

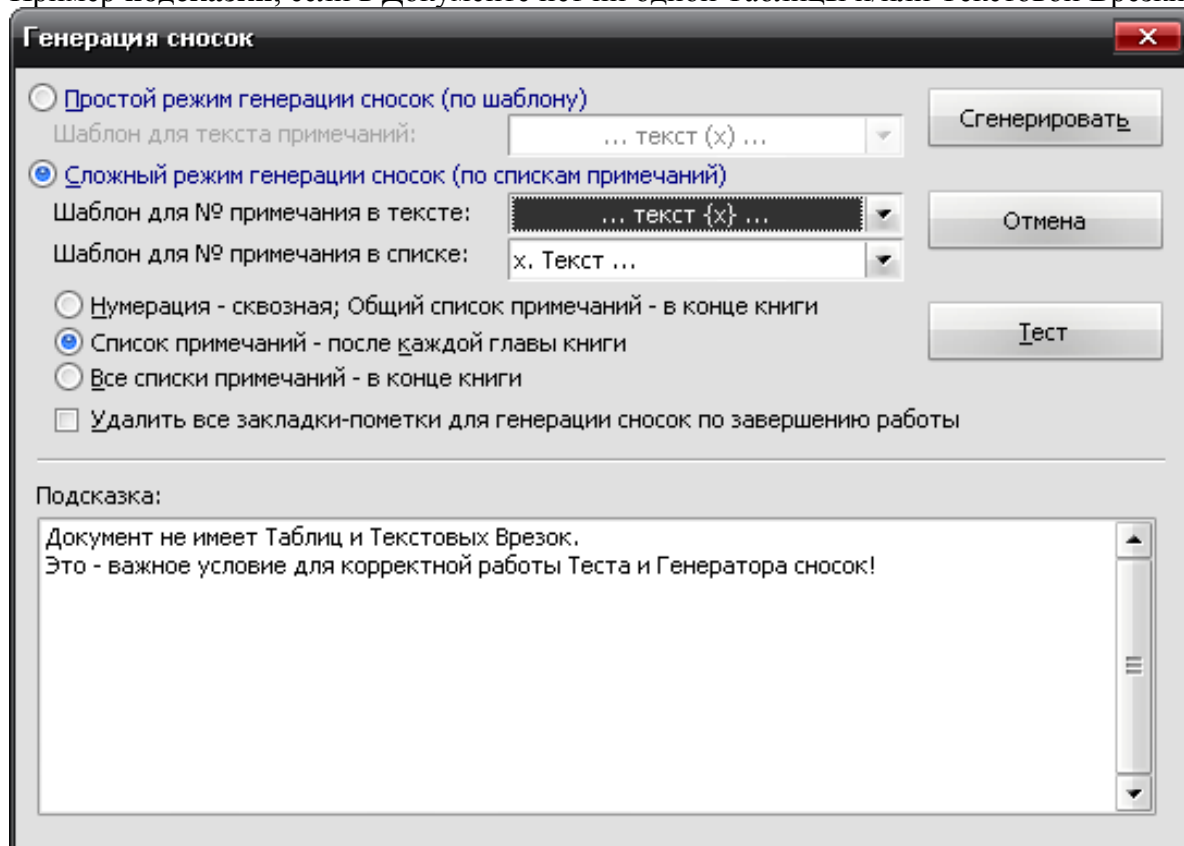
Корректность работы Генератора сносок и/или Теста в режиме сложной генерации зависит от наличия (вернее – отсутствия) Таблиц и/или Текстовых Врезок в Списках примечаний (эти области документа отмечены закладками `_ooo_ocr_tc_foot_note_list_zone_dikbsd_ ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_1 ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_N`):

1. OOo Writer **не может** вставлять Таблицы в сноски.
2. Таблицы и Врезки «путают» Тест и Генератор сносок – результат может быть не верным или вообще – программа «вылетает».

Поэтому, если в Документе есть хоть одна Врезка или Таблица, то выдается **сообщение-подсказка** в текстовом поле формы Генератора сносок, что нужно сделать. Дальнейшая работа - на ваш страх и риск. Лучшее - просмотреть Документ, в частности **все Списки примечаний** (они помечены закладками `_ooo_ocr_tc_foot_note_list_zone_dikbsd_ ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_1 ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_N` - легко найти в Навигаторе OOo - по F5), и откорректировать Списки. Для этого надо в Списках примечаний текст Врезки вынести из нее в текст Списка примечаний, Врезку удалить. А Таблицы можно отскриншотить в виде картинки, саму таблицу удалить, а после генерации

сносок вставить эту картинку в нужное место нужной сноски.

Пример подсказки, если в Документе нет ни одной Таблицы и/или Текстовой Врезки:



Пример подсказки, если в Документе есть хоть одна Таблица и/или Текстовая Врезка:

Генерация сносок

☐ Простой режим генерации сносок (по шаблону)
 Шаблон для текста примечаний: ... текст (x) ...

☒ Сложный режим генерации сносок (по спискам примечаний)
 Шаблон для № примечания в тексте: ... текст {x} ...
 Шаблон для № примечания в списке: x. Текст ...

☐ Нумерация - сквозная; Общий список примечаний - в конце книги
☒ Список примечаний - после каждой главы книги
☐ Все списки примечаний - в конце книги

☐ Удалить все закладки-пометки для генерации сносок по завершению работы

Сгенерировать

Отмена


Тест

Подсказка:

В Документе имеются Таблицы и/или Текстовые Врезки.
 OOo Writer не может размещать Таблицы в сносках.
 Генератор сносок и Тест не могут корректно работать, если в Списках примечаний есть Таблицы и/или Текстовые Врезки.
 Поэтому, посмотрите, есть ли в Списках Примечаний Таблиц и Врезок, и сделайте для них следующее:
 1. Вынесите текст из Врезки в соответствующий ей текст Списка примечаний.
 2. Таблицу уберите из Списка примечаний, а после генерации сносок вставьте в нужное место текста сноски картинку этой Таблицы.

В этом случае, при нажатии на кнопку **Тест** или **Сгенерировать**, выдается запрос на дальнейшую работу:

Генерация сносок

 В Документе имеются Таблицы и/или Текстовые Врезки.
 OOo Writer не может размещать Таблицы в сносках.
 Генератор сносок и Тест не могут корректно работать, если в Списках примечаний есть Таблицы и/или Текстовые Врезки.
 Поэтому, посмотрите, есть ли в Списках Примечаний Таблиц и Врезок, и сделайте для них следующее:
 1. Вынесите текст из Врезки в соответствующий ей текст Списка примечаний.
 2. Таблицу уберите из Списка примечаний, а после генерации сносок вставьте в нужное место текста сноски картинку этой Таблицы.

Продолжить выполнение задания?

OK

Отмена

Если в Списках примечаний Таблиц и/или Врезок нет, но они есть в других частях Документа (Главы книги и т.д.), то не обращайте внимание на это предупреждение, жмите кнопку OK. И все буде ОК ☺. Но, если в Списках примечаний есть Таблица и/или Врезка, то все же лучше от них «избавиться», как об этом написано выше, и только потом производить Тест и/или Генерацию сносок!

В будущем планируется упростить такую проверку на наличие Таблиц и Врезок, чтобы их наличие проверялось только в Списках примечаний (закладка _ooo_ocr_tc_foot_note_list_zone_dikbsd_), а не во всем документе. Тогда лишние подсказки сообщения не будут выдаваться – будет удобнее. А пока – не получается проверить наличие Таблиц и Врезок именно в самой закладке (они есть там, то объект закладки показывает, что их там нет! ☺). Но это – скорее для программистов. Кстати, если кто подскажет, как в

закладке определить наличие Таблицы и Врезки (код) – буду благодарен!

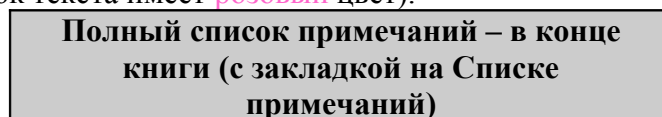
1.2.1. Режим 1: Нумерация примечаний – сквозная; Общий Список примечаний – в конце книги

К этому виду книг относятся книги, в которых Номера примечаний находятся либо во всех, либо в некоторых главах, а полный сквозной список примечаний для этих номеров находится в конце книги.

Схематично это выглядит так:



Для того, чтобы генератор сносок «знал», откуда ему брать текст примечаний для каждого номера-ссылки из текста на *это* примечание, нужно указать список примечаний в конце книги. Сделать это можно, выделив этот список мышкой и поставив на него закладку для списка примечаний (кнопка «Вставка закладки для списка примечаний») – (в таблице ниже этот отрывок текста имеет **розовый** цвет).



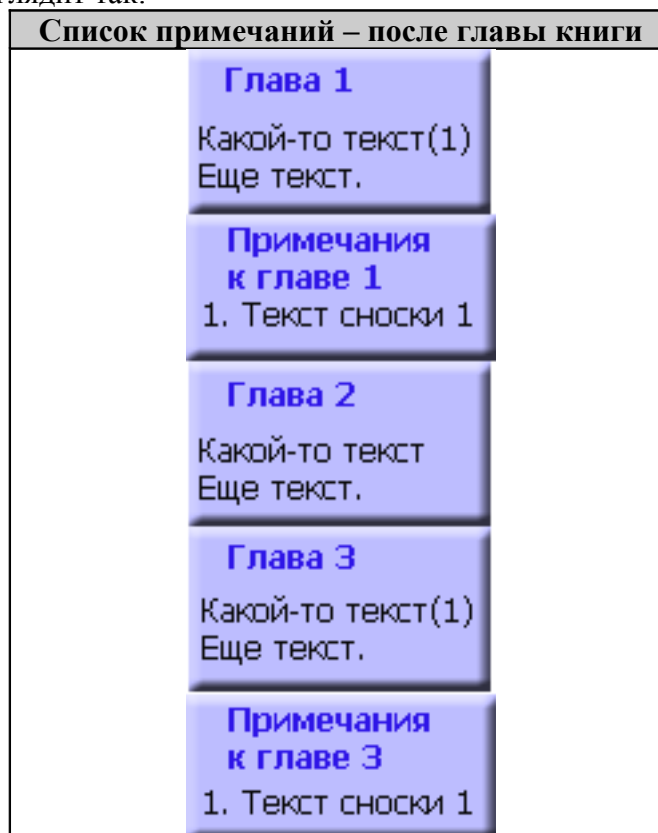


Теперь можно запустить генератор.

1.2.2. Режим 2: Список примечаний – после каждой главы книги

К этому виду книг относятся книги, в которых Номера примечаний находятся либо во всех, либо в некоторых главах, а список примечаний для этих номеров располагается после той главы, где есть «их» номера примечаний.

Схематично это выглядит так:



Расстановка закладок: Инструмент «Вставка закладки для списка примечаний»

Для указания программе списков примечаний необходимо соответствующий фрагмент текста выделить мышкой, и нажать на кнопку **«Вставка закладки для списка примечаний»** (в таблице ниже этот отрывок текста имеет **розовый** цвет). Этот инструмент интеллектуального типа. Он проверяет, нет ли на этом месте уже такой же закладки. Или, может быть, эта часть текста уже входит в такую же закладку. Или этот выделенный фрагмент текста частично перекрывает уже установленную закладку для списка текста примечаний. В этих случаях выдается соответствующее предупреждение, и закладка не ставится.

Это сделано для того, чтобы не было нарушено ВАЖНОЕ соотношение – **число глав с номерами примечаний ((1)... (N) и т.д.) не должно отличаться от числа закладок-пометок списков этих примечаний!** Если это будет нарушено (часто по невнимательности) – тоже при запуске генерации сносок произойдет следующее: текст примечания главы, например 5, оказался бы применен к главе, например, 3. А это не есть хорошо! Поэтому, следите за вышеназванным соотношением числа глав с № примечаний и числом закладок для их текста!

Поставить эту закладку легко. Выделите мышкой фрагмент текста книги, где находится список текста примечаний для конкретной главы. Очень важно, чтобы в эту зону выделения попал **ВСЬ** список, т.е. номер 1-го примечания в этом списке, и последние символы последнего примечания из этого списка. Иначе то примечание, номер которого лишь частично попал в закладку или вовсе не попал – не будет обработано, а. значит, из него не будет сгенерирована и сноска.

Недостающие примечания, ошибки OCR и методы их коррекции

Часто приходится вычитывать и конвертировать текст после OCR. Который делал кто-то другой. И в этом тексте масса опечаток. То номера сноски в главе нет, то в списке текста сносок имеются не все сноски... Часто бывает и так: есть текст, в некоторых главах есть номера примечаний, а списка текста этих примечаний для этой главы – нет. В идеале – лучше всего найти бумажную книгу и исправить, дополнить текст нужными сносками. Но, если оригинала книги нет, то можно поступить следующим образом.

Например, у нас есть такая книг, которая имеет главы с номерами примечаний, а самих примечаний для некоторых глав – нет. Если поставить закладки на имеющиеся списки примечаний и запустить генерацию сносок – то примечания, относящиеся к одной главе, могут быть сгенерированы, как сноски для другой. В таблице дан пример такой книги и метод, как обойти это проблему:

Текст, как есть	Добавление «пустых» закладок
-----------------	------------------------------

<div>Глава 1</div> <div>Какой-то текст(1) Еще текст.</div> <div>Примечания к главе 1</div> <div>1. Текст сноски 1</div> <div>Глава 2</div> <div>Какой-то текст(1) Еще текст(2).</div> <div>Глава 3</div> <div>Какой-то текст(1) Еще текст.</div> <div>Примечания к главе 3</div> <div>1. Текст сноски 1</div>	<div>Глава 1</div> <div>Какой-то текст(1) Еще текст.</div> <div>Примечания к главе 1</div> <div>1. Текст сноски 1</div> <div>Глава 2</div> <div>Какой-то текст(1) Еще текст(2).</div> <div></div> <div>Глава 3</div> <div>Какой-то текст(1) Еще текст.</div> <div>Примечания к главе 3</div> <div>1. Текст сноски 1</div>
---	---

Как видим, в этой книге во 2-й главе есть номера примечаний, а самого списка с текстом этих примечаний в конце главы – нет. Чтобы примечания к главе 3 не сгенерировались, как сноски для 2-й главы, нам надо поставить в конце главы 2 «пустую» закладку для «фиктивного» блока списка текста примечаний. Тогда будет соответствие между главами, содержащими номера примечаний и закладками на списках этих примечаний в конце этих глав.

Для добавления «пустой» закладки можно, либо поставить закладку с помощью инструмента **«Вставка закладки для списка примечаний»** в конец текста главы 2, либо после главы 2 вставить пустой абзац, а на него поставить закладку этим же инструментом.

Теперь при генерации сносок каждые примечания будут «приписаны» строго к «своей» главе. Для главы 2 (в нашем случае) вместо номеров примечаний (1) и (2) будут стоять сноски, текст которых будет отсутствовать – «пустые сноски».

1.2.3. Режим 3: Все списки примечаний глав – в конце книги

Эта опция выбирается тогда, когда все списки примечаний глав находятся в конце книги.

Схематично это выглядит так:

Все списки примечаний глав – в конце книги

Глава 1 Какой-то текст(1) Еще текст.
Глава 2 Какой-то текст(1) Еще текст(2).
Конец Книги
Примечания к главе 1 1. Текст сноски 1
Примечания к главе 2 1. Текст сноски 1 2. Текст сноски 2

Расстановка закладок: Инструменты «Вставка закладки для списка примечаний» и «Вставка закладки для главы с номером примечания»

Работа с закладками в этом режиме Генерации сносок немного отличается от предыдущего. Пометка списков примечаний закладкой с помощью инструмента «Вставка закладки для списка примечаний» ничем не отличается от того, что было написано выше. Отличие состоит только в том, для этого режима генерации сносок требуется использовать еще один инструмент: «Вставка закладки для главы с номером примечания».

Для чего нужен этот инструмент? Т.к. в результате ошибок OCR в одной главе, например, 1-й, могут быть номера примечаний с 1 по 10, потом сразу 12 и по 15 (т.е. нет номера 11). А в следующей главе 2 номера примечаний все – с 1-го по 20. Как в этом случае программе «понять», что номера примечаний из 2-й главы: 11, 16-20 не относятся к главе 1?

Делать анализ всех списков примечаний для выявления количества номеров примечаний по каждой главе тоже бессмысленно – в главе 1, например 15 примечаний, но из-за ошибок OCR номер 11 примечания отсутствует. А в списке примечаний для этой 1-й главы – всего 3 примечания – 1, 2 и 3-е (тоже из-за ошибок OCR). Как видим, установить соответствие списка примечаний с конкретной главой по числу номеров примечаний невозможно.

Поэтому, наиболее лучший способ – это просто «подсказать» программе, где начинается и заканчивается зона поиска номеров примечаний для конкретного списка примечаний. И делается это при помощи инструмента «Вставка закладки для главы с номером примечания».

Схематично, документ с 2-мя видами закладок выглядит так:

Все списки примечаний глав – в конце книги



На рисунке **зеленым** цветом показаны проставленные закладки для главы с *номером* примечания в начале тех глав, где есть номера примечаний. *Эта закладка ставится в самое начало главы (до номера 1-го примечания).*

В каких случаях ставится Закладка для главы с номером примечания

Эта закладка ставится в *начало главы только* в том случае, если *в ней есть номера примечаний, для которых есть список этих примечаний в конце книги.*

Иными словами, должно быть соблюдено ВАЖНОЕ соотношение – **число глав с номерами примечаний, а, значит, и число закладок для этих глав, не должно отличаться от числа закладок–пометок списков этих примечаний!** Если это будет нарушено (часто по невнимательности) – то при попытке сгенерировать сноски будет выдано соответствующее предупреждение и программа не запустится. Иначе произошло бы следующее: текст примечаний одной главы оказался бы применен к другой. А это не правильно! Поэтому, следите за вышеназванным соотношением!

Примеры неверного проставления закладки для главы с номером примечаний

- 1). Когда ставят эту закладку в начале главы, где нет ни одного номера примечаний.
- 2). Когда ставят эту закладку в начале главы, где есть номера примечаний, но нет в конце книги списка примечаний для этой главы (из-за ошибок OCR).
- 3). Когда ставят эту закладку в начале главы, где нет номера примечаний (из-за ошибок OCR), а в конце книги для нее есть список примечаний.

Недостающие примечания, ошибки OCR и методы их коррекции

Самый идеальный вариант – это все-таки найти бумажную книгу и восстановить недостающие примечания... Если же это невозможно, то поступить и по-другому. Ниже в таблицах будет схематично показаны некоторые ошибки OCR и пути их коррекции для корректной генерации сносок, когда нет бумажного оригинала книги.

Для главы 2 нет списка примечаний	
Текст, как есть	Добавление закладок
<div>Глава 1</div> <div>Какой-то текст(1) Еще текст.</div> <div>Глава 2</div> <div>Какой-то текст(5) Еще текст.</div> <div>Глава 3</div> <div>Какой-то текст(1) Еще текст(2).</div> <div>— Конец Книги —</div> <div>Примечания к главе 1</div> <div>1. Текст сноски 1</div> <div>Примечания к главе 3</div> <div>1. Текст сноски 1 2. Текст сноски 2</div>	<div>Глава 1</div> <div>Какой-то текст(1) Еще текст.</div> <div>Глава 2</div> <div>Какой-то текст(5) Еще текст.</div> <div>Глава 3</div> <div>Какой-то текст(1) Еще текст(2).</div> <div>— Конец Книги —</div> <div>Примечания к главе 1</div> <div>1. Текст сноски 1</div> <div></div> <div>Примечания к главе 3</div> <div>1. Текст сноски 1 2. Текст сноски 2</div>

В этом примере в книге всего 3 главы. Во всех главах есть номера примечаний. Но списки примечаний в конце книги есть только для глав 1 и 3. 1). На начало каждой главы, где есть номера примечаний ставим «закладку для главы с номером примечаний» (зеленый цвет на картинке). 2). На списки примечаний ставим «закладку для списков примечаний». 3). Теперь, чтобы при генерации сносок примечания главы 3 не сгенерировались, как сноски главы 2 ставим «пустую» «закладку для списков примечаний» между списком для главы 1 и списком для главы 3 (вставляем пустой абзац, а на него – эту закладку). Теперь – все в порядке.

В главе 2 нет номеров сносок, а список примечаний - есть	
Текст, как есть	Добавление закладок

<p>Глава 1 Какой-то текст(1) Еще текст.</p> <p>Глава 2 Какой-то текст. Еще текст.</p> <p>Глава 3 Какой-то текст(1) Еще текст(2).</p> <p>Конец Книги</p> <p>Примечания к главе 1 1. Текст сноски 1</p> <p>Примечания к главе 2 1. Текст сноски 1 5. Текст сноски 5</p> <p>Примечания к главе 3 1. Текст сноски 1 2. Текст сноски 2</p>	<p>Глава 1 Какой-то текст(1) Еще текст.</p> <p>Глава 2 Какой-то текст. Еще текст.</p> <p>Глава 3 Какой-то текст(1) Еще текст(2).</p> <p>Конец Книги</p> <p>Примечания к главе 1 1. Текст сноски 1</p> <p>Примечания к главе 2 1. Текст сноски 1 5. Текст сноски 5</p> <p>Примечания к главе 3 1. Текст сноски 1 2. Текст сноски 2</p>
--	--

В этом примере в книге тоже всего 3 главы. Номера примечаний есть только в главах 1 и 3. Списки примечаний в конце книги есть для всех глав. 1). На начало главы, где есть номера примечаний ставим «закладку для главы с номером примечаний» (зеленый цвет на картинке). 2). Ставим «закладку для списков примечаний» только на те списки примечаний, в главах которых есть их номера. Теперь – все в порядке. Конечно, примечания для 2-й главы не будут сгенерированы, как сноски, но это лучше, чем ничего, или. Когда примечания от одной главы сгенерируются, как сноски для другой.

1.2.4. Опция «Удалить все закладки-пометки для генерации сносок по завершению работы»

В диалог Генерации сносок есть опция «Удалить все закладки-пометки для генерации сносок по завершению работы». Если она включена, то все закладки, поставленные на главы и списки примечаний автоматически удаляются по окончанию работы. Если эта опция не выключена, то эти закладки остаются. Это удобно, чтобы после генерации сносок посмотреть, а что осталось в списках примечаний? Для этого с помощью средств OOoWriter можно переходить от закладки к закладке. Такой «осмотр» полезен: возможно в главе есть номер необработанного примечаний, только он не соответствует полностью шаблону поиска, и поэтому был проигнорирован. Тогда, после просмотра всех списков примечаний и исправления *в заранее сохраненной копии* документа всех опечаток, можно снова сгенерировать сноски.

1.3. Инструмент «Удаление всех служебных закладок»

Этот инструмент просто удаляет из документа все оставшиеся, если они есть, закладки-пометки, нужные для генерации сносок. Зачем в обработанном документе уже ненужные закладки? Тем более, что при экспорте в fb2-формат закладки рассматриваются, как id для тэга <r> для ссылки-перехода на этот абзац. А это будет явная «пустышка», т.к. на этот абзац перехода не будет (после генерации сносок все уже «ненужные» остатки списков примечаний нужно удалить)...

1.4. Важные замечания

Генератор сносок игнорирует Таблицы и Текстовые Врезки.

Часто после OCR из-за неправильного экспорта распознанного текста, часть этого текста может оказаться во Врезке. И, если в ней есть номера ссылок на примечания, то этот текст просто необходимо перенести из Врезки в основной текст книги, а саму Врезку – удалить. Тогда и эти номера примечаний будут обработаны генератором.

В реальных книгах не так часто встречаются номера примечаний в Таблицах. Если же это есть, то их можно будет после генерации обработать вручную.

2. Тесты к Генератору сносок

На диалоге Генератора Сносок есть кнопка **Тест**. Она активна только для сложной генерации сносок. В зависимости от режима генерации выполняются разные тесты.

Для чего это нужно?

Очень часто книга вычитывается и делается после OCR, когда в тексте может быть масса ошибок, которые могут распространяться и на номера примечаний. Сгенерировав сноски в такой книге можно заметить, что не все примечания обработаны, чего-то не хватает, а какие-то сноски содержат «чужой текст».

Чтобы быть уверенным, что сноски сгенерируются корректно, сначала нужно выполнить тест, немного подождать, потом проанализировать результат (он будет автоматически показан в браузере, стоящему по-умолчанию в системе) в виде html-файла.

2.1. Примеры ошибок OCR и «ложного срабатывания» генератора сносок

Рассмотренные ниже примеры – результат анализа работы Теста. Запустите Тест, просмотрите результат – повторяющиеся номера, отсутствующие номер и т.д. По этим данным можно многое понять об ошибках в тексте и найти нужные номера примечаний в книге – они в тесте показаны в виде списка.

«Ложное» срабатывание генератора – это его абсолютно верная реакция на ошибки или некорректности в тексте книги.

1. Отсутствующие номера примечаний и опечатки

Например, для обозначения номеров примечаний в Тексте используются круглые скобки и цифры, а в списке – в начале каждого абзаца номер с точкой после него (самый распространенный вариант книг):

Корректный текст книги	Нехватка номера-указателя в Тексте	Нехватка номера примечания в Списке примечаний
Какой-нибудь текст (1). Еще текст (2).	Какой-нибудь текст (1). Еще текст (2-).	Какой-нибудь текст (1). Еще текст (2).

Еще текст (3). Примечания: 1. Текст 1-го примечания. 2. Текст 2-го примечания. 3. Текст 3-го примечания.	Еще текст (3}. Примечания: 1. Текст 1-го примечания. 2. Текст 2-го примечания. 3. Текст 3-го примечания.	Еще текст (3). Примечания: 1. Текст 1-го примечания. 2. Текст 2-го примечания. 3, Текст 3-го примечания.
---	---	---

В 1-й колонке – пример корректного текста книги. Есть четкое соответствие между номерами примечаний в Тексте и соответствующими им примечаниями в Списке примечаний (номера выделены синим цветом). Сгенерируется 3 сноски.

Во 2-й колонке таблицы – ситуация, когда в Тексте книги либо нет номера примечания, либо рядом с цифрой – «посторонний» символ, либо закрывающая скобка «распознала» не как круглая, а как фигурная (эти «дефекты» показаны красным цветом). А генератор «настроен» именно на круглые скобки и цифры внутри них. В этом случае сгенерируется только 1-я сноска. Все остальное останется нетронутым.

В 3-й колонке таблицы – пример плохого экспорта после OCR Списка примечаний – 1-е и 2-е примечания «слиплись» в один абзац, а у 3-го после номера вместо точки стоит запятая. Будет сгенерировано только 1 сноска из 1-го примечания, причем ее текст будет содержать и 1-е примечание, и 2-е, т.к. это один абзац. 3-е же примечание будет вообще не найдено.

После исправлений всех этих опечаток все примечания будут корректно сгенерированы в сноски.

2. Повторяющиеся номера примечаний

Часто в текстах можно встретить ситуацию, когда номера примечаний повторяются либо по замыслу автора книги, либо из-за OCR-ошибок, либо гол, весь и другая информация в скобках цифрах воспринимается генератором, как номер сноски.

Одинаковые номера-указатели в Тексте – замысел автора книги

Исходный текст книги	Откорректированный текст книги
Какой-нибудь текст (50). Еще текст (51). Еще текст – (50). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания.	Какой-нибудь текст (50). Еще текст (51). Еще текст – (5000). Примечания: 50. Текст 50-го примечания. 5000. Текст 50-го примечания. 51. Текст 51-го примечания.

Здесь автор книги дважды ссылается на одно и то же примечание – (50). Это вполне допустимо, но при генерации это 2-е указание на примечание не будет обработано. Чтобы и оно сгенерировалось в сноску, его надо изменить, придав ему заведомо уникальное значение, а в Списке примечаний добавить еще один абзац с текстом от примечаний 50. а номер – изменить на **уникальный**, как это показано во 2-й колонке таблицы примера. Теперь все будет сгенерировано в сноски корректно.

Одинаковые номера-указатели в Тексте – какие-то цифры воспринимаются, как номера примечаний

Исходный текст книги	Откорректированный текст книги
Какой-нибудь текст – в скобках – число лет (50). Еще текст (49). Еще текст (50).	Какой-нибудь текст – в скобках – число лет {50}. Еще текст (49). Еще текст (50).

Примечания: 49. Текст 49-го примечания. 50. Текст 50-го примечания.	Примечания: 49. Текст 49-го примечания. 50. Текст 50-го примечания.
--	--

В этом примере рассмотрена ситуация, когда число лет человека, взятое в круглые скобки, воспринимается, как дублирование примечания (50). При генерации сносок именно это число лет (50) будет воспринято, как указатель на примечание из Списка: «50. Текст 2-го примечания.». А это – не то, что должно быть. Надо запустить тест, посмотреть, есть ли повторяющиеся номера (будет показан их список, если они есть), и те числа, которые не являются указателями на примечание – как-то выделить особо, чтобы после генерации сносок найти их восстановить их прежний вид. Например, число лет в скобках (50) можно «защитить», изменив круглые скобки на фигурные: {50}.

Одинаковые номера-указатели в Тексте – при распознавании (OCR) какой-то номер примечаний распознанся неправильно

Исходный текст книги	Откорректированный текст книги
Какой-нибудь текст (50). Еще текст (51). ... Еще текст (59). Еще текст (50). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 59. Текст 59-го примечания. 60. Текст 60-го примечания.	Какой-нибудь текст (50). Еще текст (51). ... Еще текст (59). Еще текст (60). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 59. Текст 59-го примечания. 60. Текст 60-го примечания.

Одинаковые номера-указатели в Списке примечаний – при распознавании (OCR) какой-то номер примечаний распознанся неправильно

Исходный текст книги	Откорректированный текст книги
Какой-нибудь текст (50). Еще текст (51). ... Еще текст (59). Еще текст (60). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 59. Текст 59-го примечания. 50. Текст 60-го примечания.	Какой-нибудь текст (50). Еще текст (51). ... Еще текст (59). Еще текст (60). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 59. Текст 59-го примечания. 60. Текст 60-го примечания.

3. Не соответствие числа номеров примечаний из Текста числу примечаний в Списке примечаний и наоборот

В корректном тексте книги с примечаниями число номеров-ссылок на соответствующие им примечания в Списке примечаний и число примечаний в этом Списке должны совпадать. Иногда Тест может показать, что это не так. Примеры:

В Тексте есть отсутствующие номера из Списка примечаний

Исходный текст книги	Откорректированный текст книги
Какой-нибудь текст (50). Еще текст (51). ... Еще текст (58). Еще текст (59). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 68. Текст 58-го примечания. 59. Текст 59-го примечания.	Какой-нибудь текст (50). Еще текст (51). ... Еще текст (58). Еще текст (59). Примечания: 50. Текст 50-го примечания. 51. Текст 51-го примечания. ... 58. Текст 58-го примечания. 59. Текст 59-го примечания.

Тест показал, что номер (58) из Текста отсутствует в Списке примечаний, а номер из Списка примечаний (68) отсутствует в Тексте. В этом примере в Списке примечаний ошибка в номере 68. Вместо этого номера должен стоять номер 58. – это – OCR-ошибка.

2.2. Книга с номерами примечаний без ошибок

В хорошо распознанном тексте должны иметься следующее:

1. Число номеров-ссылок на соответствующие им примечания в Списке примечаний и число примечаний в этом Списке должны совпадать. Если это соотношение не соблюдается – генерация сносок не будет корректной.

2. **Все** номера примечаний в Тексте должны быть «обрамлены» в **одни и те же** символы, например круглые скобки.

3. **Все** примечания из Списка должны идти с начала абзаца и соответствовать шаблону – х. или {х} или (х) или <х> и т.д.

4. Число обработанных номеров примечаний должно быть равно числу Уникальных, т.е. не повторяющихся номеров.

5. Должны быть выполнены следующие условия:

Число Повторяющихся номеров = 0;

Число Отсутствующих номеров = 0;

Число Отсутствующих номеров из Текста в Списке примечаний = 0;

Число Отсутствующих номеров из Списка примечаний в Тексте = 0

2.3. Использование данных анализа Теста

В генерируемом отчете Теста для каждого из Режимов содержится масса таблиц, помогающих понять – что в книге не в порядке с примечаниями (если таковое есть). Анализ делается в Зонах поиска, помеченной 2-мя закладками – Глава Книги и Список «ее» примечаний. По каждой такой Зоне выдается своя информация нескольким критериям: Сколько обработано номеров примечаний, сколько есть неповторяющихся номеров, сколько повторяющихся, сколько отсутствующих, и в Главе, и в Списке «ее» примечаний. А также – сколько и какие номера из Главы отсутствуют в Списке примечаний и наоборот.

По всем эти данным выдаются таблицы с соответствующими номерами.

В конце отчета – **важные** таблицы – какие Зоны (Главы и «их» Списки примечаний) не требуют корректуры (с ними все в порядке), а какие – требуют исправления ошибок для успешной генерации сносок. В этих таблицах отображаются названия закладок, по которым с помощью Навигатора в OpenOffice.org можно легко перейти в нужное место и исправить ошибки в нумерации сносок и т.д.

Пользуйтесь этой информацией!!!

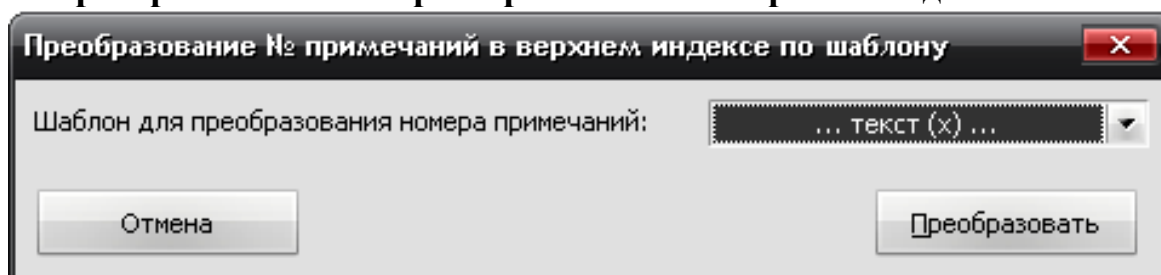
2.4. Важное замечание о Тесте

Часто пользователь может по невнимательности ошибочно выбрать не тот режим сложной генерации сносок, сделать Тест, а потом удивляться, почему тест выдал «странные» данные – то нет вообще ни одного номера примечаний (а в Документе они есть), то число уникальных (неповторяющихся) номеров примечаний в Главе или Списке примечаний вообще нет! И т.д. Чтобы этого не было четко следуйте следующему:

1. Если Книга у вас имеет вид: Примечания есть в каждой главе, а Списки примечаний – в конце книги, то в форме Генератора сносок выберите именно этот режим, а не Список примечаний – после каждой главы! Т.е. не ошибитесь с выбором режима генерации сносок!

2. Если в Главах Книги номера примечаний отмечены у вас фигурными скобками, например, {1}... {51}, номера примечаний в Списке примечаний – цифрой в круглых скобках без точки после закрывающей скобки, например: (1) Текст примечаний, то в форме диалога в шаблонах нужно выбрать именно такой шаблон! Не перепутайте! Иначе и Тест, и результат генерации сносок вас «приятно разочарует» ☺. Будьте внимательны!

3. Преобразование номеров примечаний в верхнем индексе по шаблону



Инструмент «запоминает» свое положение на экране и последний выбранный шаблон из списка.

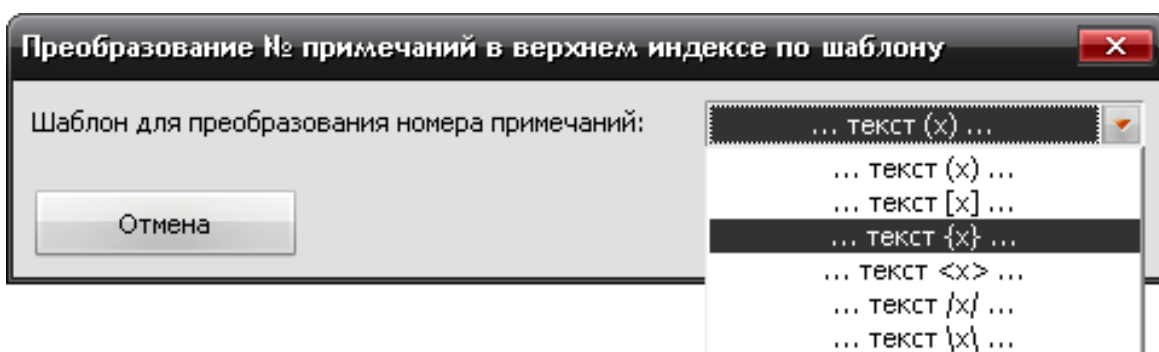
Для чего нужен этот инструмент?

Очень часто встречаются книги, сохраненные после OCR в формат, поддерживающий верхний индекс текста (rtf, doc...). В таких книгах *номера примечаний* часто представлены *цифрами в верхнем индексе*. Инструмент Генерации сносок «не умеет» работать с такими номерами примечаний по ряду причин: 1). Часто в документах встречаются в верхнем индексе и степени цифр из математических формул и т.д. И, если эту степень посчитать за номер примечания и сгенерировать из нее сноску в формуле – это нарушит логику текста. А инструмент Генерации сносок должен быть как можно более нейтральным. Поэтому, логичнее, сначала преобразовать цифры – номера примечаний в верхнем индексе в текст по определенному шаблону (цифра в круглых или фигурных скобках и т.д.), а потом уже сгенерировать по ним сноски.

С этой целью и создан Инструмент преобразования номеров примечаний в верхнем индексе по шаблону. Конечно, он тоже «не понимает», где – номер примечания, а где – степень числа или переменной из формулы. Поэтому применять его нужно осторожно. Или – запомнить, где в тексте встречаются формулы, а потом их исправить.

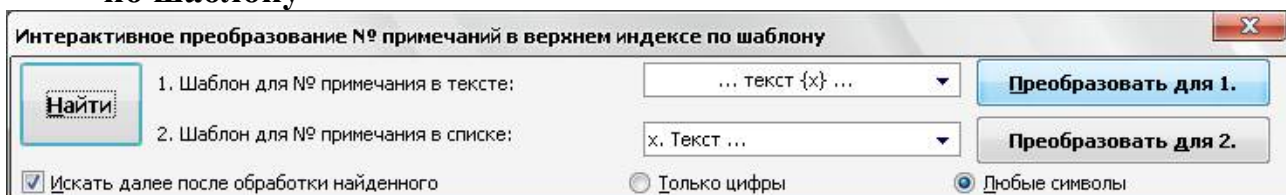
Алгоритм этого инструмента представляет собой анализатор всего документа по абзацам. Поэтому он не предусматривает (пока) поиск в выделенных зонах текста.

Лучше всего выбрать шаблон для преобразования номеров примечаний в верхнем индексе, отличный от круглых (x) или квадратных [x] скобок, т.к. в тексте могут встречаться пояснения в виде цифр в этих скобках, и при генерации сносок они могут обрабатываться, как сноски! Лучше избрать шаблон в виде фигурных скобок {x}:



Работа инструмента проста. Просто нажмите кнопку **Преобразовать**. После работы он выдаст сообщение о числе преобразованных номеров примечаний.

4. Интерактивное преобразование номеров примечаний в верхнем индексе по шаблону



Этот инструмент полезен в тех случаях, где нельзя использовать инструмент автоматического преобразования номеров примечаний в верхнем индексе (см. выше). Автоматическое преобразование нельзя использовать в таких книгах, как Библия, Коран, математические книги и т.д., в которых цифры в верхнем индексе обозначают номера стихов (Библия, Коран...) или степень числа в формулах. При использовании автоматического преобразования все эти цифры тоже будут преобразованы по шаблону (например, {1}), а это – неверно.

Поэтому для таких книг и разработан инструмент интерактивного преобразования номеров примечаний для дальнейшей автоматической генерации сносок в книге (см. соответствующий инструмент выше).

Как он работает?

Поиск осуществляется от местоположения видимого курсора, и движется «по кругу». Нажатие кнопки «Найти» *выделяет* найденные цифры в верхнем индексе. Если эти цифры действительно имеют отношения к номерам примечаний, то для их преобразования есть два выпадающих списка шаблонов и две соответствующие этим спискам кнопки преобразования.

Если включена опция «Искать далее после обработки найденного», то после нажатия одной из кнопок преобразования найденное будет преобразовано по шаблону и будет автоматически найдены другие цифры в верхнем индексе.

После OCR цифрами в верхнем индексе в книге могут обозначаться не только номера примечаний, но и номера сносок в списках текста сносок. Например:

Простой текст¹.
Еще текст².

¹ Текст сноски 1.

² Текст сноски 2.

Здесь – и номера примечаний в тексте (сноски), и номера примечаний в списке примечаний – в верхнем индексе. Инструмент найдет все такие номера. И, для каждого из этих 2-х видов цифр в верхнем индексе предусмотрен свой шаблон преобразования и соответствующая ему кнопка преобразования.

Например, для этого текста на картинке, после ручной обработке с шаблонами, показанными на картинке выше результат будет таким:

Простой текст{1}.

Еще текст{2}.

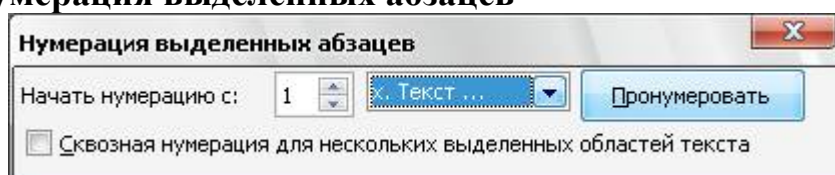
1. Текст сноски 1.

2. Текст сноски 2.

Теперь такой текст легко может быть использован для работы автоматического генератора сносок (см. выше).

В инструменте предусмотрен выбор поиска либо цифр в верхнем индексе, либо любых символов в верхнем индексе. Последнее часто бывает очень полезно, т.к. после OCR многие цифры примечаний «распознаются, как случайные символы верхнего индекса. Инструмент найдет и их...

5 Нумерация выделенных абзацев



Инструмент написан по следующим причинам:

1. Разработчики OOo Writer очень часто меняют свойства и методы многих объектов и, в частности, нумерованных списков, из-за чего пакет OOoFBTools перестает нормально обрабатывать нумерованные списки.

2. Инструмент [автогенерации сносок](#) не может работать с нумерованными списками.

Поэтому, нумерованные списки можно очень просто заменить на абзацы с нумерацией.

Данный инструмент позволяет пронумеровывать либо весь документ, либо 1 или несколько выделенных фрагментов текста. Он игнорирует Таблицы, Текстовые Врезки и Сноски (внизу страницы). Можно задать начальный номер для самого первого абзаца. Можно сделать Сквозную нумерацию для нескольких выделенных областей текста, если включить соответствующую опцию.

Если обрабатываемый абзац - нумерованный список, то автонумерация отключается. Если же этот абзац - простой абзац, то он пронумеровывается без автонумерации. Т.е. в любом случае для обрабатываемых абзацев автонумерация отключается!

Можно задать вид нумерации (полезно для последующей автогенерации сносок).

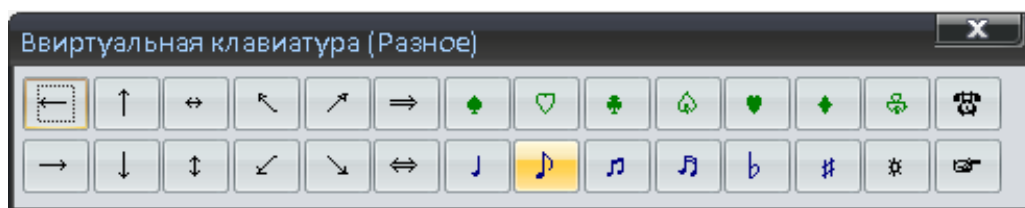
Скажу от себя – мне этот инструмент экономит массу времени!

IV. Виртуальные клавиатуры

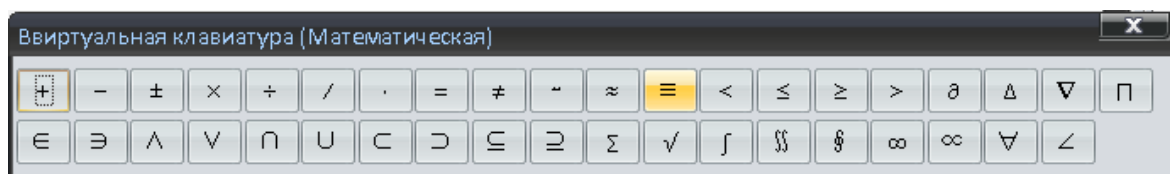
IV.1. Виды виртуальных клавиатур

Иногда требуется в текст документа (корректурa, правка после OCR) ввести Unicode символы. Удобнее для этого использовать соответствующие виртуальные клавиатуры:

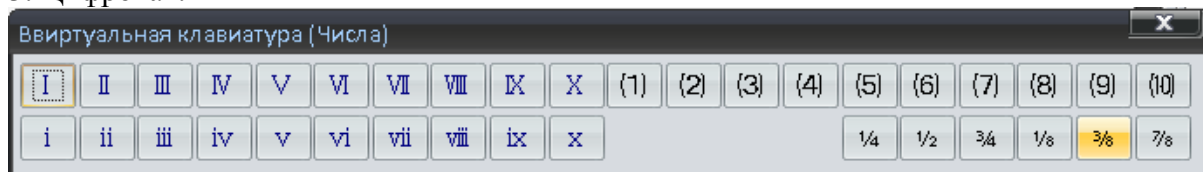
1. Различные символы (стрелки, ноты...):



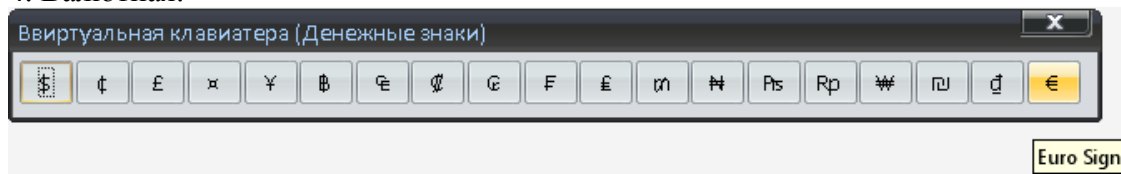
2. Математическая:



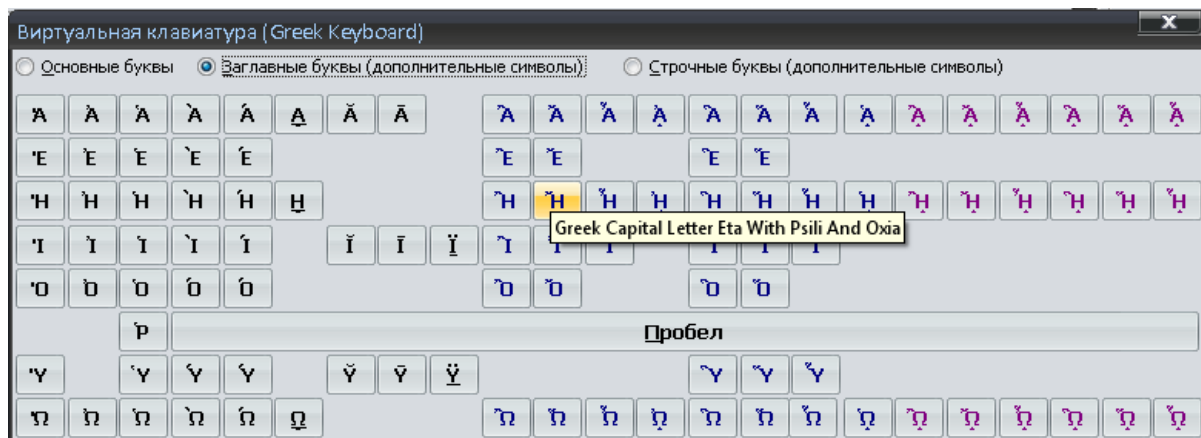
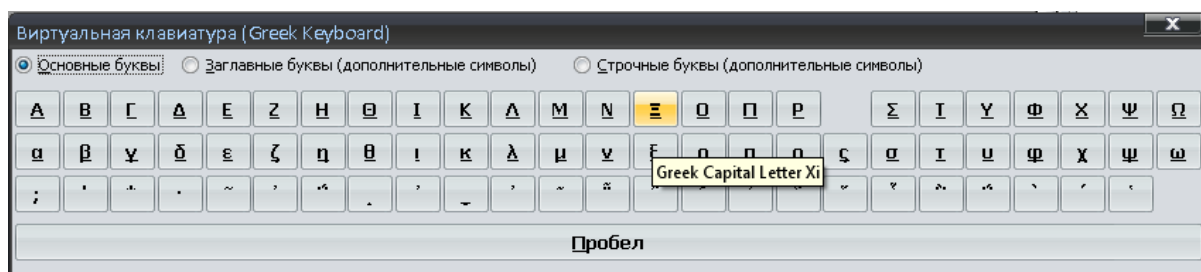
3. Цифровая:



4. Валютная:



5. Греческая виртуальная клавиатура:



6. Еврейская виртуальная клавиатура:



Для каждой кнопки виртуальных клавиатур (валютная, греческая и еврейская) заданы соответствующие всплывающие подсказки. Пользуйтесь ими, чтобы узнать название символов и (или) букв.

Для клавиатуры с разными символами цветами выделены символы одной логической группы.

Для греческой виртуальной клавиатуры цветами выделены три вида букв – без дополнительных знаков, с 2-мя дополнительными знаками и с 3-мя дополнительными знаками. Так легче искать то, что нужно.

IV.2. Как запускать виртуальные клавиатуры?

Вынесите соответствующие виртуальным клавиатурам макросы на нужную вам панель (где уже располагаются кнопки конвертора, или на отдельную панель, что гораздо удобнее).

Для «Различные символы (стрелки, ноты...)»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / MiscKeyboard**.

Для «Математической виртуальной клавиатуры»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / MatchKeyboard**.

Для «Цифровой виртуальной клавиатуры»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / NumbersKeyboard**.

Для «Валютной виртуальной клавиатуры»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / CurrencyKeyboard**.

Для «Греческой виртуальной клавиатуры»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / GreekKeyboard**.

Для «Еврейской виртуальной клавиатуры»: макрос **OpenOffice.org / Мои макросы / OOoFBTools / VirtualKeyboards / HebrewKeyboard**.

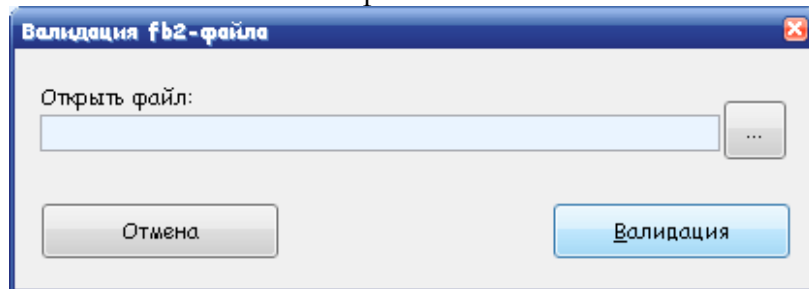
Как вытаскивать макросы на кнопки подробно описано в этой справке в [п.2.2](#). Сделайте по аналогии.

IV.2. Это важно!

Если вы при корректуре документа использовали эти виртуальные клавиатуры, и хотите экспортировать его в fb2 формат, то для экспорта в конвертере выбирайте кодировку **utf-8**! Почему? Потому что символы виртуальных клавиатур – Unicode символы. И сохранение Unicode символов в другой кодировке (в нашем случае в windows-1251) все испортит – они просто не экспортируются нормально. Поэтому, в этом случае при конвертировании используйте **utf-8**!

V. Валидаторы

Валидатор предназначен для того, чтобы проверить, соответствует ли fb2-файл схеме FictionBook. Т.е. – есть ли в файле ошибки.



FB2 Валидатор позволяет выбрать нужный fb2-файл и проверить его. В случае его валидности выдается соответствующее сообщение. Если же файл не соответствует хоть в чем-то схеме – то выдается сообщение с указанием номера строки и символа в файле, где найдена ошибка, а также краткое описание, что за ошибка.

Запускается Валидатор либо через кнопку fb2Validator на панели инструментов OOoFBTools, либо через меню OOoFBTools -> fb2Validator... (и панель инструментов, и меню создаются автоматически при установке пакета расширения OOoFBTools).

Пока Валидатор работает только для системы Windows.

Планы по развитию пакета OOoFBTools (TODO)

А. Конвертор ExportToFB21

1. Сделать то, что не реализовано (см. [п.1.3.](#)), по мере возможности и способности.
2. Оптимизация скорости парсинга.
3. Минивалидатор форматирования стилями текста – выдает сообщение о соответствии/не соответствии помеченного стилями документа схеме FictionBook2.2.
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. Что-нибудь еще...

Б. «Ёфикация»

1. «Ёфикация», т.е. автоматический и ручной поиск и замена в соответствующих словах буквы «е» на «ё».

В. Импорт fb2-файлов

1. Как xml-код.
2. Как текст книги, с сохранением всей структуры выделением соответствующими стилями (стихи, таблицы...).

Вопросы – Ответы

Установка пакета OOoFBTools

1. После установке пакета не видно, чтобы он обновился до более нового

1. Для OOo Writer версии ниже 3.0:

После обновления пакета OOoFBTools или установки новой версии пакета *обязательно* нужно перезапустить редактор OOo Writer. Это правило распространяется на *любые* расширения, которые Вы хотите установить на OOo.

1. Для OOo Writer версии 3.0 и выше:

После обновления пакета OOoFBTools или после его удаления и установки более новой версии *обязательно* нужно либо перезагрузить компьютер, либо перелогиниться (этого будет достаточно без перезагрузки компьютера). Одной перезагрузки редактора будет недостаточно. Это правило распространяется на *любые* расширения, которые Вы хотите установить на OOo.

2. Где находятся все служебные файлы пакета OOoFBTools – файлы профилей, настроек...

Все эти файлы хранятся в папке *config* пакета *OpenOffice.org*.

Для OOo версии 2:

Для Unix-подобных систем – это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows – это папка *C:\Documents and Settings\XXX\Application Data\Openoffice.org3\user\config\OOoFBTools*, где XXX – ваш логин (имя пользователя).

Для OOo версии 3:

Для Unix-подобных систем – это папка */home/XXX/.openoffice.org3/user/config/OOoFBTools*. Для Windows – это папка *C:\Documents and Settings\XXX\Application Data\OpenOffice.org3\user\config\OOoFBTools*, где XXX – ваш логин (имя пользователя).

Стили

1. Не могу найти в редакторе стили, которыми надо форматировать разные структуры текста

Скорее всего, Вы не загрузили файл стилей конвертора *fb21_styles.ott*. Как это сделать – см. подробно [здесь](#).

2. Мне неудобно пользоваться названиями стилей. Более привычно использовать не Level N, а Заголовок N (и т.д.).

Каждый пользователь может легко настроить названия стилей «под себя» просто переименовав их в диалоге «Стили и форматирование» (F11), а потом переопределив их с помощью инструмента «Пользовательские стили». Как это сделать подробно описано [здесь](#) и [здесь](#).

Стили в диалоге «Стили и форматирование» (F11) нужны непосредственно для создания структуры документа. В диалоге переопределения стилей (OOoFBTools -> Пользовательские стили) задаются стили, с которыми будет работать конвертер. И эти стили должны по названию полностью совпадать с названиями стилей в диалоге «Стили и форматирование» (F11)! Иначе структура книги не будет экспортирована.

Например, в диалоге «Стили и форматирование» (F11) вы переименовали стиль Level 1 в Заголовок1, а в диалоге переопределения стилей он так и остался Level 1. Это неверно. Надо и его переименовать в Заголовок1 и нажать кнопку «Принять».

3. Я переопределил стили (переименовал и в диалоге «Стили и форматирование» (F11) и в диалоге переопределения стилей OOoFBTools -> Пользовательские стили). Задал структуру книги этими новыми стилями. Все прекрасно экспортировалось! Но, когда я загрузил новую книгу для экспорта, загрузил в нее стили (OOoFBTools ->Загрузка шаблона стилей в документ) то оказалось, что все мои новые стили исчезли!

Дело в том, что вы переименовали название стилей, **загруженных в Ваш документ**. Стили же в шаблоне остались неперименованными.

Можно сделать следующее:

1. После загрузки стилей в каждую новую книгу – всякий раз в диалоге «Стили и форматирование» (F11) переименовывать их на нужные. Это утомительно.

2. Создать отдельный шаблон стилей *fb21_styles.ott*, содержащий Ваши стили, загрузить его и работать теперь с ним:

Для этого откройте в редакторе чистый лист.

Загрузите в него стили, как обычно (OOoFBTools ->Загрузка шаблона стилей в документ).

Переименуйте нужные Вам стили.

Сохраните Ваши стили в рабочий шаблон: Файл -> Сохранить Как... ->Тип файла (Шаблон текстового документа ODF (.ott)). Название Вашего шаблона лучше дать осмысленное, например ЗаголовкиN.ott – если вы переименовывали Level N в ЗаголовокN.

Зайдите в диалог управления Шаблонами редактора: Файл -> Шаблоны -> Управление. Два раза кликните мышкой на Мои шаблоны, Выделите шаблон *fb21_styles.ott* и удалите его либо через контекстное меню, либо через кнопку Команды. Теперь, стоя здесь же на Мои шаблоны сделайте импорт Вашего нового шаблона либо через контекстное меню, либо через кнопку Команды.

Все! Теперь для каждой новой Книги при загрузке в нее стилей (OOoFBTools ->Загрузка шаблона стилей в документ) будут загружаться именно Ваши переименованные стили.

Таких пользовательских шаблонов стилей на базе файла *fb21_styles.ott* можно наделать сколько угодно под разные нужды.

Только **не забывайте** про то, что названия стилей из диалога «Стили и форматирование» (F11), которыми Вы форматируете документ должно **полностью совпадать** с названиями стилей в диалоге переопределения пользовательских стилей (OOoFBTools -> Пользовательские стили)! Иначе – результат Вас разочарует.

4. Мне никак не удается в диалоге «Стили и форматирование» (F11) переименовать Level 1 в Заголовок 1.

Стиль Заголовок 1 – жестко «зашит» в редакторе для русской локализации, как основной стиль Заголовков. Редактор и не даст его переименовать.

Здесь есть 3 пути:

1. Вы можете Level 1 переименовать в Заголовок1 или _Заголовок 1.
2. Просто выберите или введите Заголовок 1 на стандартной панели форматирования редактора в списке стилей.
3. В диалоге «Стили и форматирование» (F11) в выпадающем списке (внизу диалога) выберите «Автоматически». Появятся все стандартные для редактора стили, в том числе и Заголовок 1 ... Заголовок 10.

5. Я хочу экспортировать в fb2-формат файл Microsoft Word. Но в этом файле заголовки УЖЕ отформатированы стилями Заголовок 1 ... Заголовок 10. А в OOoFBTools используются для заголовков стили Level 1 ... Level 10. Что делать?

Для таких случаев есть два пути, после загрузки такого файла в OOo Writer.

Первый путь: Переименовываем названия одних стилей (Заголовок 1 и т.д.) на название других (Level 1 и т.д.). Для этого загружаем в документ стили (OOoFBTools ->Загрузка шаблона стилей в документ). Нажимаем **Ctrl+F** для вызова диалога поиска. В нем нажимаем кнопку **Больше параметров** и ставим галочку на **Искать по стилям**. В выпадающих списках **Найти** и **Заменить на** появятся все стили. В **Найти** выберите **Заголовок 1**, в **Заменить на** – **Level 1**. Нажимаем кнопку **Заменить все**. И т.д., для всех стилей, которые нужно заменить.

Второй путь: Запустите OOoFBTools диалог переопределения стилей (OOoFBTools -> Пользовательские стили). В нем либо загрузите профиль из файла со стилями, где стили Level «подменяются» на Заголовок, либо переименуйте Level 1 ... Level N в Заголовок 1 ... Заголовок N. Если хотите, сохраните на будущее этот профиль стилей. Нажмите кнопку **Принять**. Все, теперь можно делать экспорт. **Не забудьте:** Теперь конвертер будет воспринимать стили заголовков ВСЕХ книг, как Заголовок N !!! Если надо опять вернуться к Level N, в диалоге переопределения стилей нажмите кнопку **По умолчанию**, потом – **Принять**.

Картинки

1. Какую графику конвертер может экспортировать в fb2-файл?

Конвертер экспортирует в fb2 только те графические объекты (картинки, формулы и диаграммы), которые удовлетворяют следующим требованиям:

1. Привязка графики только 3-х видов: К Абзацу, К Символу или Как символ. Другая привязка игнорируется
2. Графика, вложенная в сложные Врезки или сложные Таблицы – игнорируется. Сложная Врезка – это Врезка, вложенная в другую Врезку. Сложная Таблица – это таблица, вложенная в ячейку другой Таблицы (странички форумом). Такие Врезки и Таблицы нужно привести к более простому виду.
3. Картинка есть либо в документе, либо на жестком диске, а в документе есть на нее ссылка. Если же в документе есть ссылка на какой-нибудь сайт, где «лежит» эта картинка, то

она экспортироваться не будет, т.к. до нее «не добраться».

2. Никак не экспортируются картинки

Здесь может быть несколько вариантов:

1. Нет нужных пакетов в системе:

Если у Вас система – Linux, то должен быть установлен пакет *base64*. Если Ваша система – Mac Os X – то нужен *OpenSSL*. Эти пакеты обычно установлены по-умолчанию. Проверьте, есть ли они у Вас в системе (в терминале наберите название пакета...).

2. Неверно задана привязка картинок в тексте:

Конвертор «понимает» для картинок только 3 привязки – «К абзацу», «К символу» и «Как символ». Последняя используется для экспорта «буквиц» – картинок–символов (внутри предложения). Другие привязки игнорируются, и картинки не экспортируются. Вообще-то эти другие привязки и не нужны.

3. Никак не экспортируются обложки книги. В fb2-файле в разделе <description> нет тэга <coverpage>

Это может быть по нескольким причинам:

1. Те же, что и в вопросе «Никак не экспортируются картинки».

2. Конвертер воспринимает, как обложки **только те картинки**, которые **расположены выше** Названия Книги (стиль **Book Title**). Картинки, расположенные ниже конвертируются, как иллюстрации Книги!

3. Конвертер (пока) не понимает картинки-обложки, размещенные в Текстовых Врезках. Все остальные картинки (не обложки), расположенные во Врезках конвертер понимает и экспортирует. Поэтому, если Ваша обложка помещена во Врезку – вынесите картинку из нее!

4. Никак не экспортируются картинки в ячейках таблицы

Посмотрите, экспортируются ли картинки из текста книги (не таблиц). Если нет, то проблема не в таблицах. Решение – см. [выше](#). Если экспортируются, то посмотрите *привязку* картинок в ячейках таблицы.

Картинки в ячейках таблиц *всегда* должны иметь привязку «К Абзацу», «К символу» или «Как символ»! Иначе они «не видятся» конвертором, а, следовательно, и не экспортируются. Обычно, при вставке картинки или графики в ячейку таблицы, по-умолчанию ставится «К абзацу» – все нормально. Если Вы изменили привязку на другую, отличную от 3-х вышеназванных, то картинки экспортироваться не будут.

5. Теги картинок экспортируются, а сами картинки – нет. Base64-код какой-то «битый», «обгрызанный», всего несколько символов

Это происходит из-за того, что Вы экспортируете книгу *сразу* после редактирования! По невыясненным пока причинам ООо Writer не корректно экспортирует картинку в файл на диске, если был изменен хоть один символ текста.

Выход: Экспорт книги нужно *всегда* делать сразу после открытия книги в ООо. Если Вы хоть что-то исправляли в тексте – *перезапустите* редактор ООо Writer.

Подробнее – см. [здесь](#).

6. Не экспортируется картинка, вложенная во Врезку

Тут может быть несколько причин:

1. Если вы изменили хоть один символ в тексте, и, не перезагружая редактор пытаетесь делать экспорт – картинки не экспортируются. Это – особенность ОпенОфиса. Чтобы картинки экспортировались, надо делать экспорт **на только что открытом** файле.

2. Все картинки во Врезках экспортируются. Проверьте, не лежит ли эта ваша Врезка в другой Врезке (бывает часто на открытых документах MS Word)? Конвертор не умеет экспортировать данные **вложенных** Врезок в другие Врезки! Просто вытащите картинку из этих Врезок, перезагрузите файл и делайте экспорт. Если же вам обязательно надо, чтобы

картинка находилась во Врезке, то вытащите внутреннюю Врезку с картинкой из внешней, и удалите пустую внешнюю Врезку. Перезагрузите файл и экспортируйте.

7. Как определить, все ли картинки экспортировались в fb2-файл?

После окончания экспорта выдается сообщение о завершении конвертации. Если в документе есть графические объекты, то в этом сообщении отображается информация, **сколько** графических объектов **из скольких** было экспортировано. Если эти цифры одинаковые, то – все в порядке. Если же, например, «5 из 6», то какая-то картинка не соответствует условиям успешного экспорта. Ее надо найти и исправить либо привязку, либо вытащить из сложной Врезки или Сложной Таблицы. Об условиях успешного экспорта картинок см в этом FAQе вопрос «[Какую графику конвертер может экспортировать в fb2-файл?](#)».

С другой стороны, если после экспорта выдано сообщение, что экспортировано, например, 5 картинок из 6 найденных, то это может быть еще и по следующей причине: В документе есть ссылка на какой-нибудь сайт, где «лежит» не экспортированная картинка – она не экспортирована, т.к. до нее «не добраться».

8. Не экспортируются Картинки и Графика из текста Сносок (Примечаний)

Конвертер пока не умеет делать такой экспорт. Существующий алгоритм анализатора документа не позволяет «вычленить» Графику в тексте Сносок. Переделывать алгоритм не хочется (очень большой и сложный) – можно наделать ошибок, и все, что уже работает, перестанет функционировать. В будущем буду думать, КАК внести изменения в существующий алгоритм анализатора.

Если есть необходимость в картинках в тексте Сносок, то вставьте их в fb2-редакторе.

Таблицы

1. Не экспортируются картинки из таблиц

Ответ на этот вопрос дан [выше](#).

2. Некоторые таблицы либо не экспортируются, либо экспортируются не полностью – часть ячеек теряется

Таблицы существуют 3-х видов: 1). Простые матрицы; 2). Полусложные – разные строки имеют разное число колонок; и 3). Сложные – в отдельные ячейки таблицы вложены другие таблицы.

Конвертер отлично экспортирует 1-й вид таблиц. 2-й вид таблиц после экспорта в fb2 не всегда имеет первоначальное расположение ячеек в одной строке по отношению к ячейкам в другой строке (в будущем планируется доработать – есть определенные сложности). Но число ячеек в каждой строке экспортируется четко. 3-й вид таблиц либо просто игнорируется, либо, если они «не очень сложны» для конвертора – то частично экспортируются, игнорируя при этом все вложенные таблицы.

3-й вид таблиц чаще всего встречается в html-файлах – с форумов и т.д.

Совет – по возможности, старайтесь преобразовывать таблицы в более простой вид – к 1-му или 2-му типу.

Более подробно о работе с таблицами см. [здесь](#).

Разметка структур стилями

1. В некоторых частях текста есть «подпись» его авторов, но это не цитаты (например, Предисловие, а в его конце – ФИО его автора). Как мне экспортировать это ФИО не просто как текст (тэг <p>), а как Автора (тэг <text-author>)?

Это сделать довольно просто.

Например, в книге есть Предисловие, размеченное стилем Level N:

Предисловие

Что-то там написано...
Еще что-то написано...

Иванов Иван Иванович

После экспорта получаем:

```
<section>
<title>
<p><strong>Предисловие</strong></p>
</title>
<p>Что-то там написано...</p>
<p>Еще что-то написано...</p>
<p>Иванов Иван Иванович</p>
</section>
```

Т.е. Иванов Иван Иванович экспортировался просто, как текст.

Чтобы «его» экспортировать, как Автора Предисловия, не экспортируя при этом само Предисловие, как Цитату, нужно перед Автором вставить пустой абзац, придав ему стиль **Cite**, а Автору – стиль **Cite Author**. Тогда после экспорта получим:

```
<section>
<title>
<p><strong>Предисловие</strong></p>
</title>
<p>Что-то там написано...</p>
<p>Еще что-то написано...</p>
<cite>
<empty-line/>
<text-author>Иванов Иван Иванович</text-author>
</cite>
</section>
```

Что и требовалось.

Если не вводить пустой абзац стиля **Cite**, то конвертер не воспримет стиль **Cite Author** Автора Предисловия, как **Cite Author**, и экспортирует его просто как текст (и правильно сделает, иначе fb2-файл получился бы невалидным).

Этот прием можно применять в любых случаях, когда надо экспортировать Автора текста, как Автора, но сам текст – просто, как текст.

Валидация

1. После экспорта fb2-файл получается не валидным – не соответствует схеме. Программы-библиотекари и некоторые Internet-библиотеки его не принимают из-за ошибок в файле

Тут может быть несколько вариантов:

1. Элементы книги (стихи, эпиграфы, заголовки) должны быть отформатированы определенными стилями. При этом их расположение в книге должно *четко* соответствовать

схеме *FictionBook.xsd*. Если это так, то результирующий fb2-файл получается валидным.

В zip-архиве пакета OOoFBTools есть папка *scheme*, которая содержит документацию по схеме FictionBook2.1. Она очень наглядна и показывает, какой элемент структуры где должен располагаться для того, чтобы fb2-файл был без ошибок, валидным. Пользуйтесь этой документацией, когда форматируете стилями текст книги!

Совет: внимательно изучите весь [этот](#) и [этот](#) пункт справки.

2. Часто в тексте встречаются символы <, >, и &, которые в таком виде в fb2-файле «считаются» ошибкой. В Настройках Конвертора предусмотрена опция «Обработка проблемных символов» (включена по-умолчанию). Проверьте, не выключили ли Вы ее?

3. Какая-то скрытая ошибка в конвертере, о какой пока еще не знает разработчик. В этом случае, пожалуйста, вышлите мне Ваш **размеченный стилями** текстовый файл, который Вы пытаетесь экспортировать в fb2-формат мне на e-mail dik-bsd@mail.ru

Вылет ошибки при конвертировании

1. При экспорте некоторых документов у меня вылетает сообщение об ошибке и экспорт прерывается

А. Чаще всего это происходит на документах, весь текст которых или большая его часть отформатированы одним стилем. Например, сборник стихов – весь текст отформатирован, как Поэма (Роем).

Дело в том, что Анализатор извлекает данные из документа по определенному стилю и аккумулирует их в виде строк. Это сделано для эффективности работы парсера и последующего этапа – Создание файла. Длина строки не должна превышать 64К. В документах, где весь текст или большая его часть отформатированы одним стилем, как правило, размер этого текста превышает 64К. В результате происходит переполнение и выдается ошибка.

Я не стал разбивать считывание строк на блоки меньше 64К, т.к. это не только усложнило бы написание Анализатора, но и очень существенно увеличило бы время экспорта (2 часа вместо 5 минут.)

Поэтому, чтобы избежать такой проблемы нужно делать следующее:

1). Понимать, что делаете. Не нужно форматировать весь текст одним стилем. Если вы создаете сборник стихов, то там, как правило есть авторы отдельных стихов. Выделите их стилем Роем Author. Это значительно уменьшит размер блоков текста под структуру (в данном случае – под Роем).

2). Если все же действительно нужно, чтобы весь текст был одного стиля (например – цитата (Cite)), то хотя бы в нескольких местах поставьте пустой абзац без какого-либо стиля.

Б. Сообщения об ошибке могут появляться и по другим причинам (все не предусмотреть). В таком случае, пожалуйста, вышлите мне ваш .odt-файл по email (он указан в О программе пакета).

Вычитка