

Вадим Кузнецов (DikBSD)
**Описание работы OooFBTools –
межплатформенного пакета для
обработки текста и конвертации в fb2
формат**

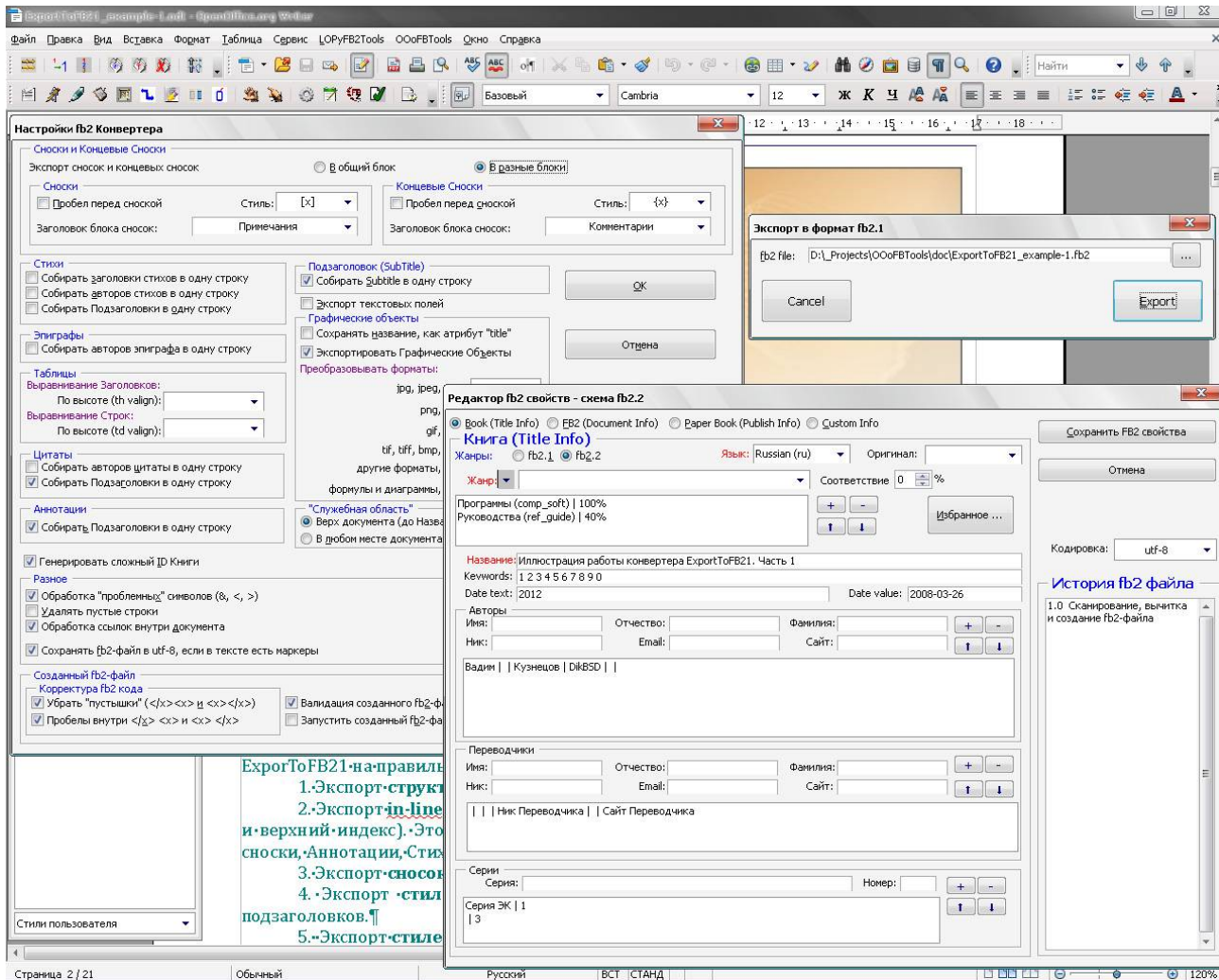
Данный пакет инструментов, как следует из названия этого мануала. Предназначен для типографики (обработка и корректировка текста, генерация сносок, исправление ошибок форматирования после OCR и т. д.), а также для конвертации текста в популярный в Рунете формат электронных книг — fb2.1.

Пакет состоит из большого числа инструментов. В данном руководстве в иллюстрациях подробно описывается каждый инструмент: принципы его работы и результат его применения.

Конвертер ExportToFB21

I. Описание и установка

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

Конвертер 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 Поэм (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 “Заголовок 1”, “ Заголовок 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 конвертера 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 &#x26; и < 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 (опция 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 **Маркеров-Картинок**) из **Текста** документа и **Текстовых Врезок**.

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

11. Экспорт **ссылок внутри документа** (переход из любого места документа в любое место документа, переход из любого места документа на текст любой сноски, переход из текста одной сноски на текст другой сноски, переход с текста сноски на любое место 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 конвертера «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 **Маркеров-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 также 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).

1.4. Лицензионное соглашение

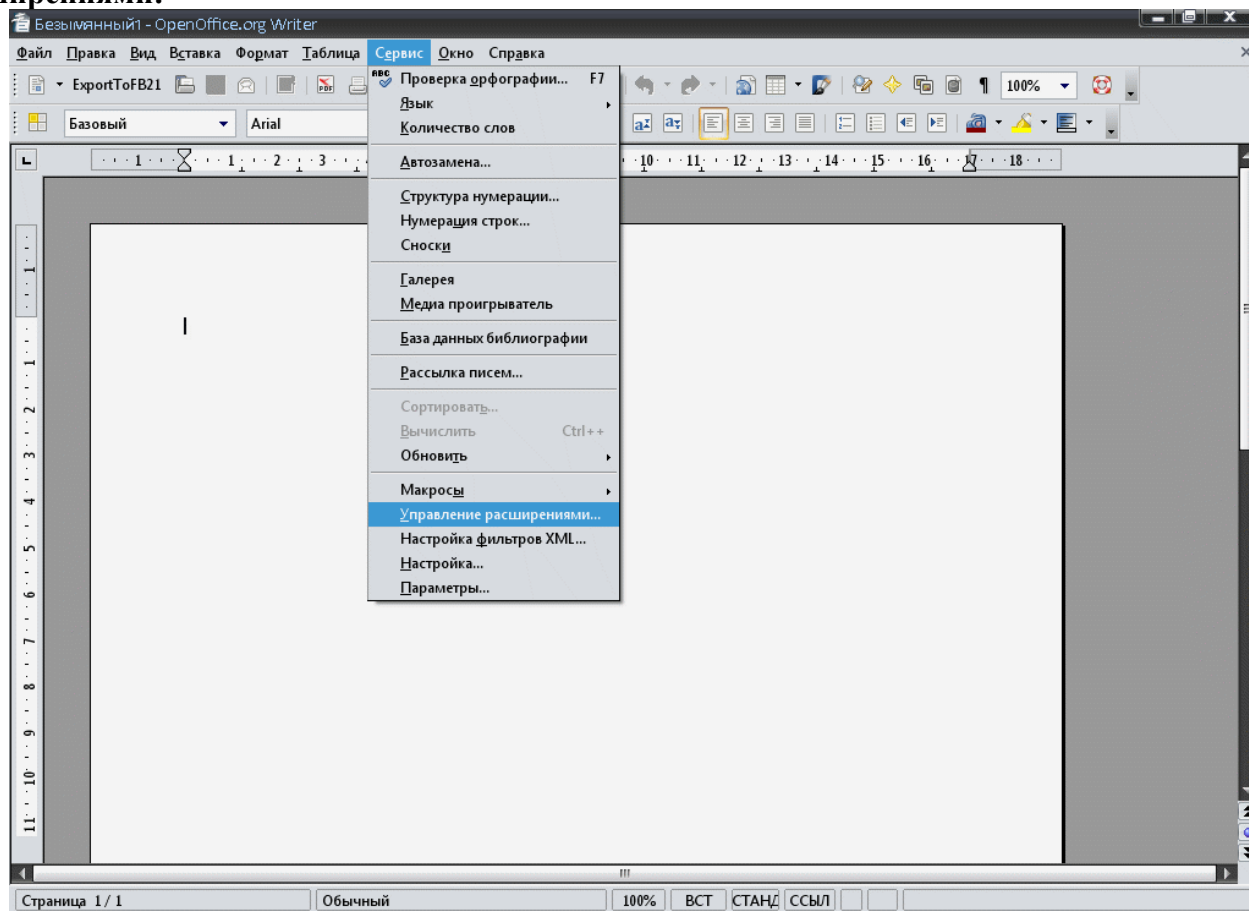
Конвертер создается и распространяется под лицензией GPL. Т.е. его можно свободно использовать, распространять, на базе его кода создавать свои проекты, тоже под лицензией GPLv3.

2. Установка конвертера (пакет OOoFBTools)

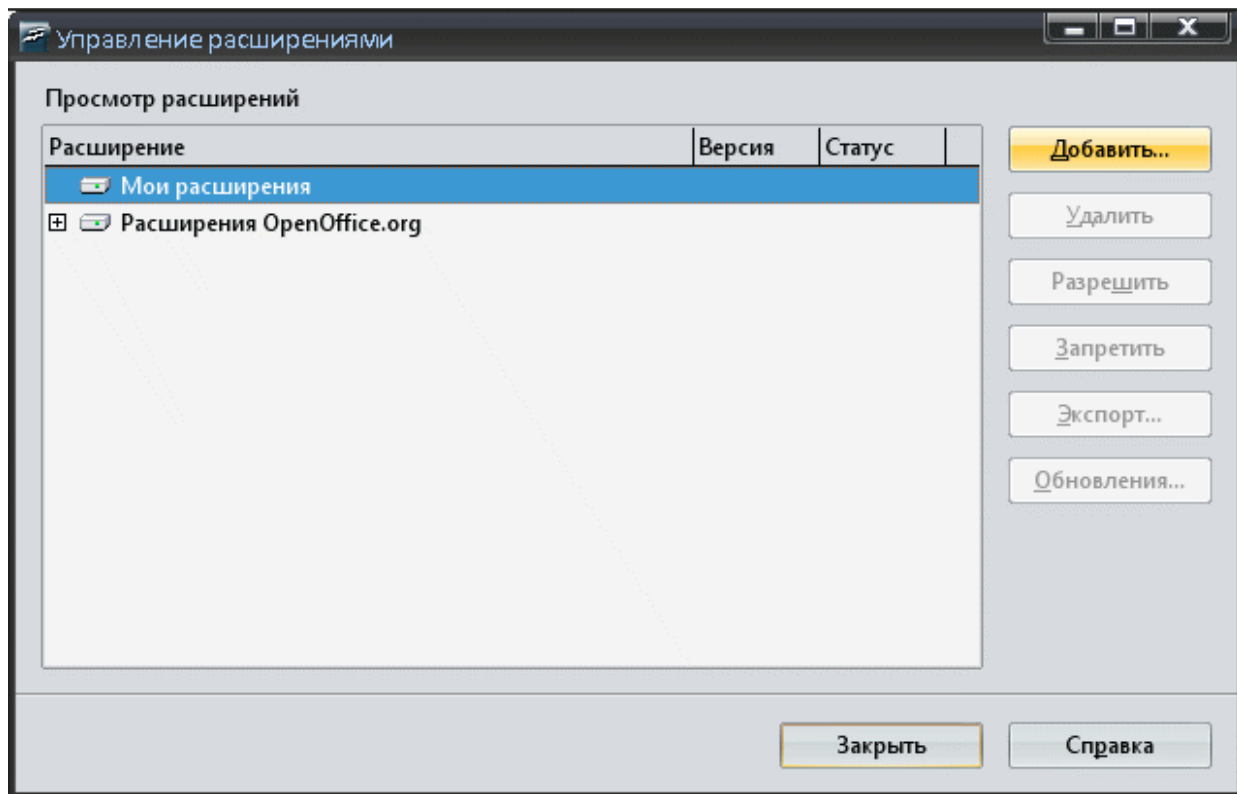
2.1. Установка расширения OOoFBTools.oxt с конвертером

2.1.1. Первый способ (для всех операционных систем)

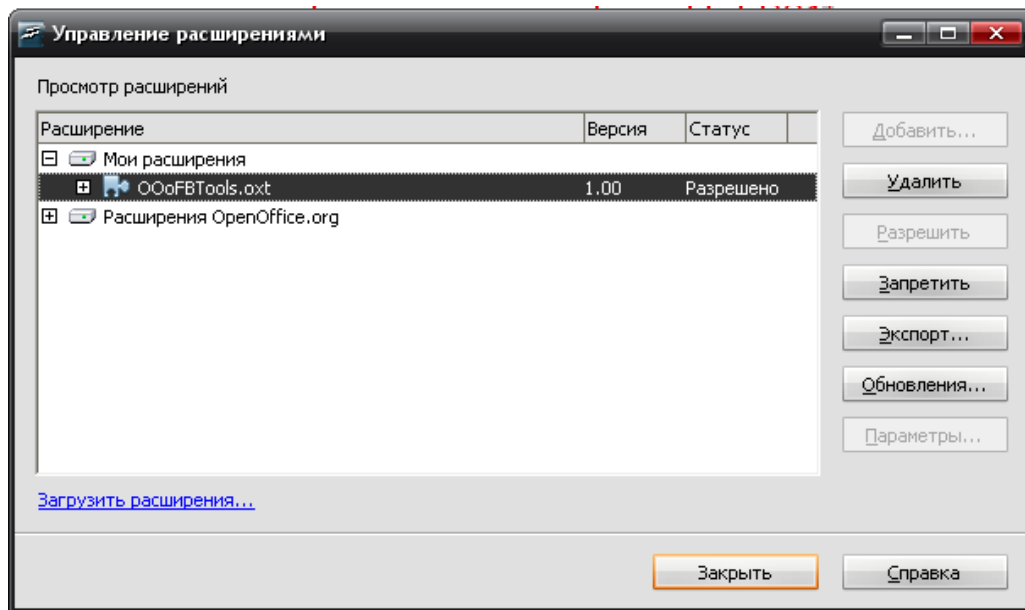
Вызываем диалог «Управление расширениями» через меню Сервис -> Управление расширениями:



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



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

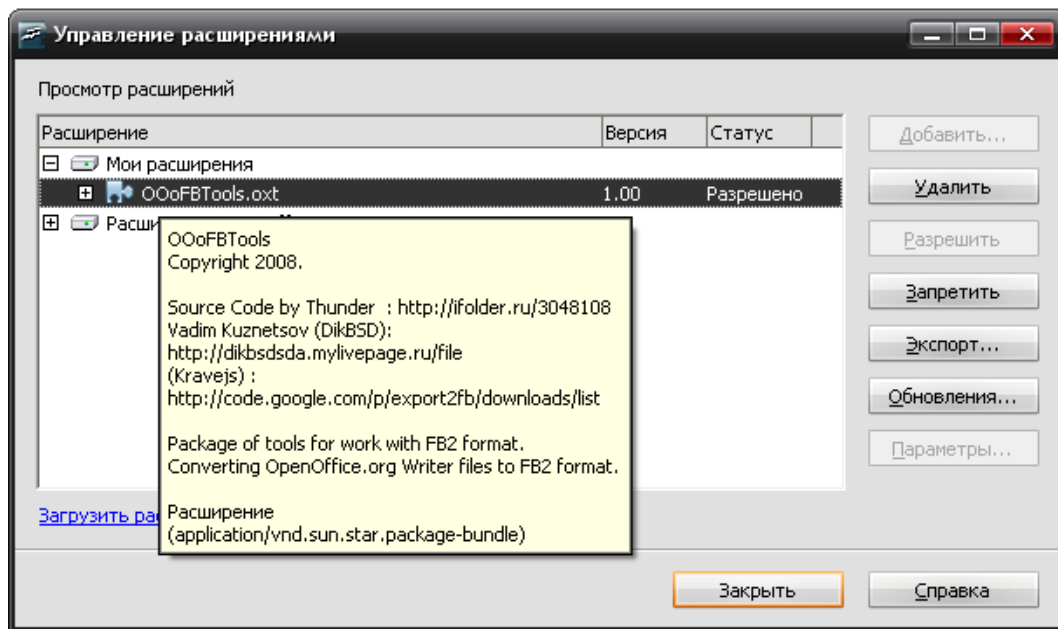


2.1.2. Второй способ (только для Microsoft Windows)

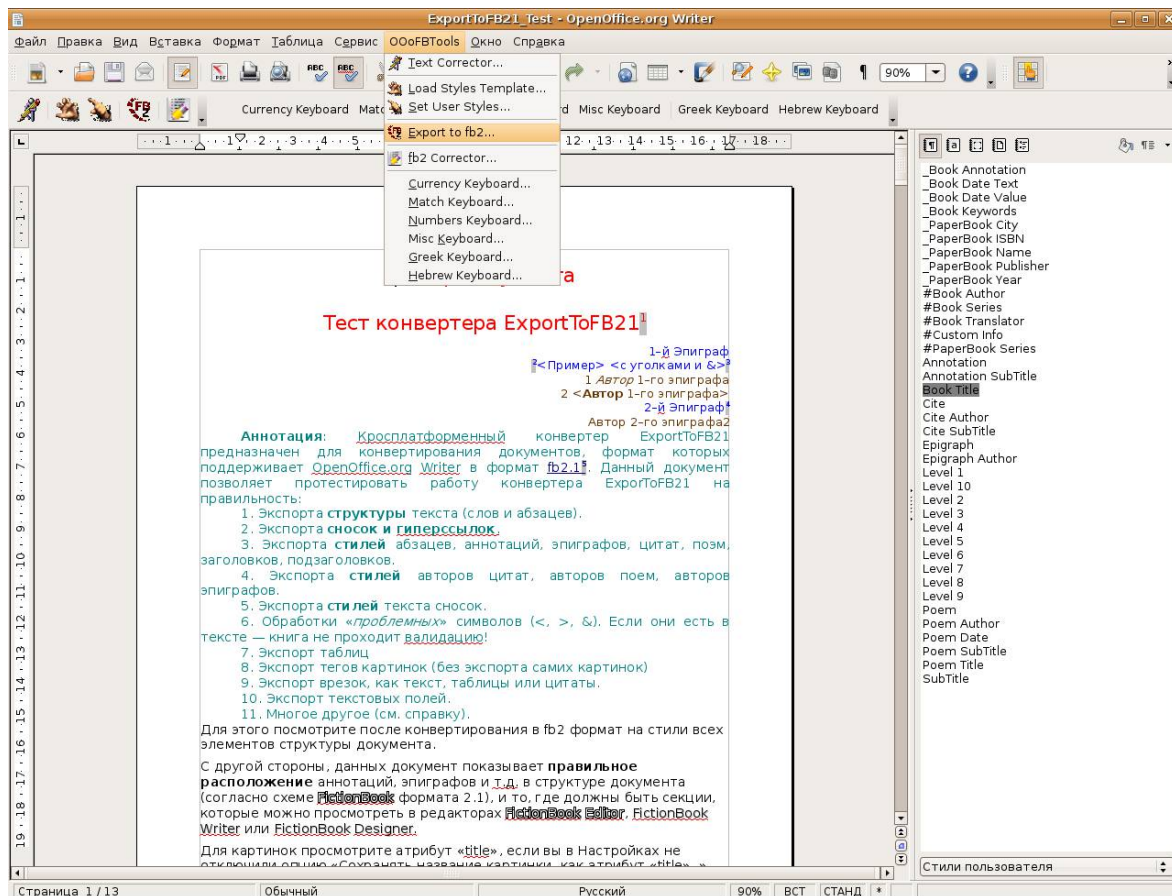
Второй способ применим *только* к установке любого расширения **.oxt** - просто из проводника или файлового менеджера запустите это расширения (по Enter или двойному клику мышки).

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. Установка дополнительных платформозависящих пакетов и приложений

2.2.1. 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 -?* или *base64 --version* (в зависимости от системы). Если пакет установлен, то вы увидите краткую справку, если нет – то терминал выдаст соответствующее сообщение.

Для валидации fb2 файла на Unix системах должна быть установлена утилита *xmllin* (обычно она включена в пакет *libxml2-utils*).

Для генерации GUID для ID книги нужна утилита *uuidgen*. Обычно она установлена во многих дистрибутивах по-умолчанию. Проверить это можно, выполнив в терминале команду: *uuidgen -t*. Если *uuidgen* установлена, то отобразится GUID. Если ее нет в системе, то нужно установить пакет *uuid-runtime* или *libuuid1*.

2.2.2. Для системы Mac Os X

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

Для валидации fb2 файла на Unix системах должна быть установлена утилита *xmllin* (обычно она включена в пакет *libxml2-utils*).

Для генерации GUID для ID книги нужна утилита *uuidgen*. Обычно она установлена по-умолчанию. Проверить это можно, выполнив в терминале команду: *uuidgen*. Если *uuidgen* установлена, то отобразится GUID.

2.2.3. Для системы Windows

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

Все же остальные утилиты для генерации GUID для ID fb2-файла и кодирования картинок находятся в самом расширении OOoFBTools.

II. Экспорт документов в fb2 формат

Процесс экспорта в fb2 формат состоит из 4-х этапов:

А. Подготовка текста (разметка стилями).

Б. Задание требуемых настроек экспорта (как правило, делается один раз, и по мере необходимости вносятся корректуры)

В. Задание описания Книги в Редакторе fb2 свойств.

Г. Экспорт в fb2 формат.

Рассмотрим подробно каждый из этапов

А. Подготовка документа к экспорту

Для того, чтобы экспортировать документ в fb2 формат, его надо предварительно подготовить – разметить Заголовки, Эпиграфы, Цитаты и т.д. соответствующими стилями. Конвертер по стилям автоматически определяет, в какую «структуру» экспортировать текст.

1. Работа со стилями структуры документа

1.1. Виды стилей

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

1. 6 символьных Inline стилей (применяются как для отдельных символов, слов, так и для нескольких абзацев – **полужирный**, *курсив*, ~~зачеркнутый~~, ^{верхний индекс}, _{нижний индекс}, code) и

2. Стили абзаца – их действие распространяется на весь абзац.

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

1.1.1. Символьные (Inline) стили

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

Соответствие стилей с тегами формата fb2.1

Назначение (fb2-тэг)	Стиль
Полужирный 	Bold
Курсив <emphasis>	<i>Italic</i>
Зачеркивание <strikethrough>	Зачеркивание (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Зачеркивание ")
Код программ – <code>	Символьный стиль «code», а также Тень или Контур (клик правой клавишей мышки на слове, выбрать "Стиль" -> "Тень" или "Курсив").
Верхний индекс – <sup>	^{"Верхний индекс"}
Нижний индекс – <sub>	_{"Нижний индекс"}

Назначение стилей вполне понятно из их названий. Пару слов о **code**. Этот стиль применяется для разметки (выделения) **кода программ**, а также **примеров** и **иллюстраций** в книгах. Такие слова и/или абзацы в тексте выделяются моноширинным шрифтом (в нашем случае – Courier New).

Для использования этих стилей достаточно выделить слово (фразу) и нажать соответствующую кнопку **Панели fb2 стилей** (см. ниже) или выбрать соответствующий пункт меню или контекстного меню. Здесь все просто.

1.1.2. Стили абзаца

Основные стили применяются для «выделения» в тексте документа (книги) таких структур, как

аннотация, цитата, поэма, заголовок, и т.д. Эти структуры могут включать в себя один и несколько абзацев целиком.

Для некоторых языков Названия стилей для удобства работы локализованы. Много стилей имеют такие же названия, как они называются и в OpenOffice (LibreOffice, StarOffice). Так привычнее работать с текстом.

Языки, для которых стили локализованы:

**Русский,
Украинский,
Белорусский,
Английский,
Немецкий,
Французский,
Испанский,
Польский,
Итальянский.**

Для всех остальных языков загружаются АНГЛИЙСКИЕ стили. Если надо еще локализовать стили – пишите.

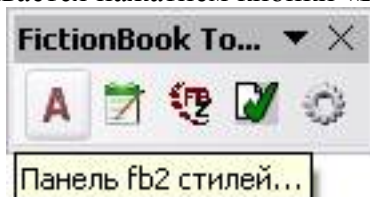
Ниже приведена таблица соответствия объектов текста и локализованных стилей для 2-х языков – Английский и Русский. Для своего родного языка вы можете посмотреть названия стилей. Запустив **Панель fb2 стилей** (см. ниже).

Таблица локализованных стилей

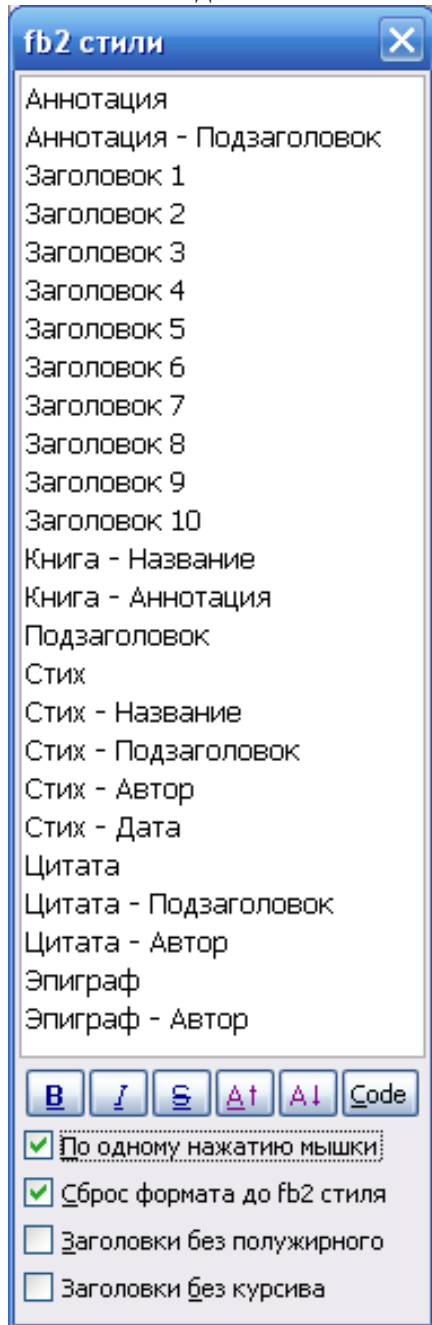
Название объекта	Стиль на Английском	Стиль на Русском
Название Книги	Book Title	Книга - Название
Аннотация на Книгу	Book Annotation	Книга - Аннотация
Аннотация (Главы, Раздела...)	Annotation	Аннотация
Подзаголовок Аннотации	Annotation Subtitle	Аннотация - Подзаголовок
Заголовок 1 ... Заголовок 10	Heading 1 ... Heading 10	Заголовок 1 ... Заголовок 10
Подзаголовок	Subtitle	Подзаголовок
Стих	Poem	Стих
Название Стиха	Poem Title	Стих – Название
Подзаголовок Стиха	Poem Subtitle	Стих – Подзаголовок
Автор Стиха	Poem Author	Стих – Автор
Дата написания Стиха	Poem Date	Стих – Дата
Цитата	Quotations	Цитата
Подзаголовок Цитаты	Quotations Subtitle	Цитаты – Подзаголовок
Автор Цитаты	Quotations Author	Цитата – Автор
Эпиграф	Epigraph	Эпиграф
Автор Эпиграфа	Epigraph Author	Эпиграф – Автор

1.2. Разметка структуры текста с помощью Панели fb2 стилей

Для удобной разметки текста создан специальный Инструмент – «**Панель fb2 стилей**». Он вызывается нажатием кнопки «**Панель fb2 стилей**» на панели инструментов FictionBook Tools.



Панель выглядит так:



При первом нажатии кнопки, fb2 стили автоматически загружаются в документ, и появляется **Панель fb2 стилей**. При всех последующих ее вызовах стили не загружаются в документ – они уже загружены. И Инструмент это автоматически определяет. Таким образом, экономится время на открытие самой Панели.

Стили Абзаца находятся в Списке стилей. Символьные стили задаются нажатием одной из 6-ти кнопок.

Для форматирования определенного абзаца(ев) надо просто дважды щелкнуть левой клавишей мышки на нужный стиль.

Переключатель «**По одному нажатию мышки**» дает возможность форматировать стилем не по 2-м щелчкам мышки, а по одному.

Переключатель «**Сброс формата до fb2 стиля**» позволяет перед форматированием нужным стилем очистить все стили выбранного абзаца. Т.е. стиль абзаца приводится к «**Базовый**» (убираются все курсивы и т.д., размер шрифта всех слов становится одинаковым). И только потом этому абзацу

задается выбранный fb2 стиль. Такая работа может потребоваться, если отдельные слова абзаца не отформатировались нужным стилем из-за различных значений атрибутов. Все это никак не влияет на экспорт в fb2. Просто текст начинает выглядеть более эстетично.

Переключатель «**Заголовки без полужирного**» дает возможность убирать полужирное начертание одного или нескольких абзацев, которые размечаются стилем Заголовок 1 ююю Заголовок 10 с панели fb2 стилей.

Переключатель «**Заголовки без курсива**» дает возможность убирать курсив одного или нескольких абзацев, которые размечаются стилем Заголовок 1 ююю Заголовок 10 с панели fb2 стилей.

Переключатели «**Заголовки без полужирного**» и «**Заголовки без курсива**» полезны, когда вы не желаете, чтобы в экспортируемом fb2 файле Заголовки книги, моделируемые тэгами <section> содержали теги полужирного начертания и/или курсива <emphasis>. Обычно читалка сама придает стиль Заголовкам книги. А избыточные тэги и <emphasis> в fb2 файле приведут к тому, что Заголовки книги всегда будут с полужирным начертанием и/или курсивом.

1.2.1. Использование переключателя «Сброс формата до fb2 стиля»

Работу этого переключателя лучше всего рассмотреть на следующем примере.

Допустим, у нас есть такой текст:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);  
result = pow(result, 5);
```

Здесь абзац «**Вопрос:** Как написать код на языке C для вычисления $\sin^5 8$?» -это наш будущий Заголовок. А все абзацы под ним – это будущая Цитата.

Если мы, запустив **Панель fb2 стилей**, включим опцию «Сброс формата до fb2 стиля» и нажмем на стиль **Заголовок 1**, то получим следующее:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);  
result = pow(result, 5);
```

Видно, что курсив слов **Вопрос** и **sin** остался. Размер шрифта слова **sin** – тоже не изменился. Главное – степень синуса – цифра **5** так и осталась в верхнем индексе.

Если же мы включим опцию «Сброс формата до fb2 стиля» и нажмем на стиль **Заголовок 1**, то результат будет уже другим – все символы будут приведены к одному виду:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);
result = pow(result, 5);
```

Как видим – «пропал» и курсив, и степень числа 5. Если курсив обязательно нужен в заголовке – его придется устанавливать вручную. Степень синуса – тоже.

Теперь разметим нужные абзацы стилем **Цитата**. Можно не выделять абзацы целиком – достаточно захватить хоть один символ. Опцию «Сброс формата до fb2 стиля» выключим, и нажмем стиль Цитата:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);
result = pow(result, 5);
```

Сохранилось все: и курсив, и полужирное начертание слова **Ответ**. Абзацы Кода программы сохранили свой шрифт Courier New.

Теперь включим опцию «Сброс формата до fb2 стиля» и нажмем стиль **Цитата**. Получим следующее:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);
result = pow(result, 5);
```

«Пропало» все – и курсив, и полужирное начертание, и шрифт Courier New. В данном случае можно выделить (теперь уже целиком!!!) абзацы с кодом программы и нажать на кнопку **Code** **Панели fb2** стилей:

Вопрос: Как написать код на языке C для вычисления $\sin^5 8$?

Ответ: Делаем возведение функции в степень, т.е. Вычисляем значение функции.

Потом полученное значение возводим в 5 степень:

```
double result = sin(8);
result = pow(result, 5);
```

1.2.2. Общие рекомендации по использованию переключателя «Сброс формата до fb2 стиля»

Опцию можно включать при разметке абзацев fb2 стилями в том случае, если:

1. В абзацах нет слов с курсивами, полужирным начертанием, верхнего или нижнего индекса и т.д.

2. После OCR часто в абзацах есть «невидимые» атрибуты, и после разметки fb2 стилем без включенной опции «Сброс формата до fb2 стиля» отдельные слова приобретают разный размер шрифта и т.д. В таком случае данную опцию лучше включить.

3. Для нас абсолютно не важен курсив и т.д. слов в размечаемом абзаце (художественная литература).

Опцию лучше не включать:

1. Когда важно исходное форматирование текста - курсив, полужирное начертание, степени математических формул и т.д.

1.3. Использование «быстрых клавиш» при работе с fb2 стилями

Для упрощения и ускорения разметки документа, fb2 стили можно «подвесить» на «быстрые клавиши».

В коде Панели fb2 стилей есть ряд функций, предназначенных для назначения им «быстрых клавиш». Все функции находятся в модуле **FB2Styles**. Такие функции написаны только для стилей абзаца. Символьные стили (полужирный, курсив и т.д. редко используются, или их легко использовать с помощью либо стандартных быстрых клавиш, либо – кнопок **Панели fb2 стилей**.)

В следующей таблице для стилей Русского и Английского языков показано соответствие fb2 стиля и функции, на которую можно назначить «быструю клавишу». Эта таблица применима и к остальным языкам, для которых fb2 стили локализованы.

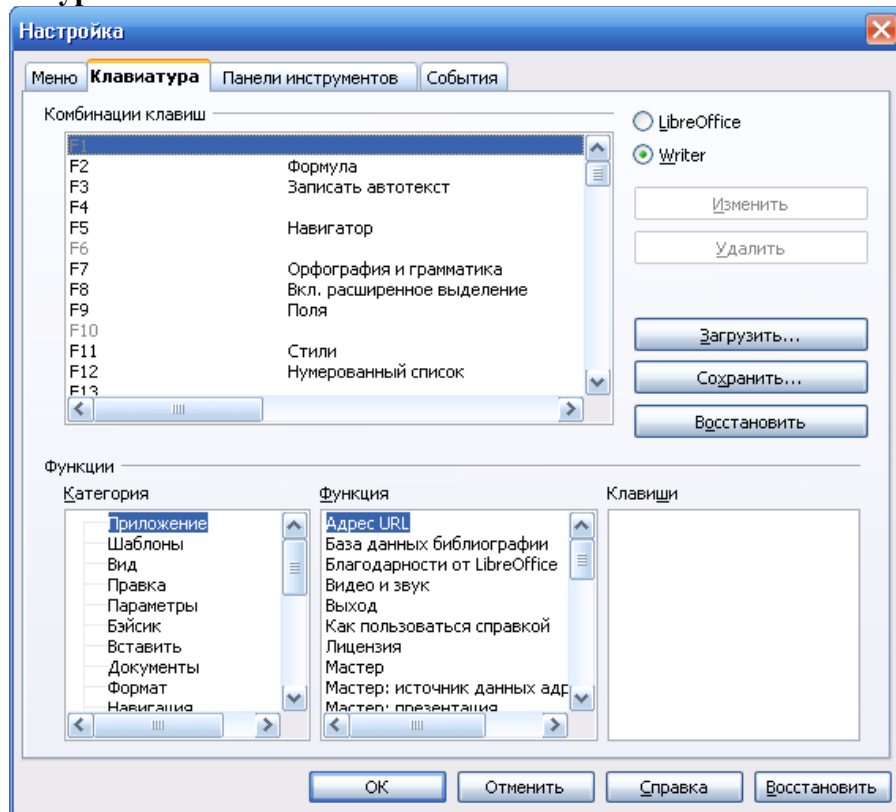
Стиль на Английском	Стиль на Русском	Функция модуля FB2Styles
Book Title	Книга - Название	ShortcutsForBookTitle
Book Annotation	Книга - Аннотация	ShortcutsForBookAnnotation
Annotation	Аннотация	ShortcutsForAnnotation
Annotation Subtitle	Аннотация - Подзаголовок	ShortcutsForAnnotationSubTitle
Heading 1 ... Heading 10	Заголовок 1 ... Заголовок 10	ShortcutsForHeading1 ... ShortcutsForHeading10
Subtitle	Подзаголовок	ShortcutsForSubTitle
Poem	Стих	ShortcutsForPoem
Poem Title	Стих – Название	ShortcutsForPoemTitle
Poem Subtitle	Стих – Подзаголовок	ShortcutsForPoemSubTitle
Poem Author	Стих – Автор	ShortcutsForPoemAuthor
Poem Date	Стих – Дата	ShortcutsForPoemDate
Quotations	Цитата	ShortcutsForCite
Quotations Subtitle	Цитаты – Подзаголовок	ShortcutsForCiteSubTitle
Quotations Author	Цитата – Автор	ShortcutsForCiteAuthor
Epigraph	Эпиграф	ShortcutsForEpigraph
Epigraph Author	Эпиграф – Автор	ShortcutsForEpigraphAuthor

Мы покажем, как задать быструю клавишу стилю **Заголовок 1**. Назначение клавиш для остальных стилей – аналогично.

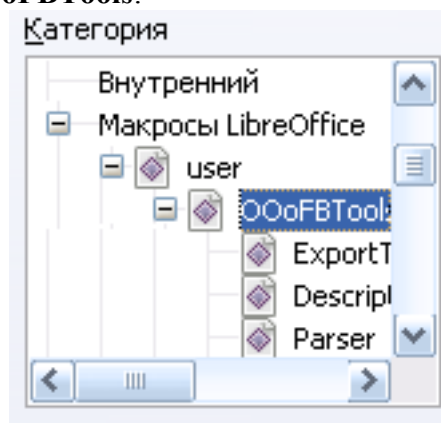
В нашем примере для стиля Заголовок 1 назначим комбинацию клавиш Ctrl+1. Это довольно удобно – левой рукой держишь Ctrl, а правой – нажимаешь цифру 1 на цифровой клавиатуре. Вы можете задействовать те клавиши, работать с которыми вам более комфортно.

Итак, запускаем диалог **Настройка** (меню **Сервис->Настройка**). Переходим на вкладку

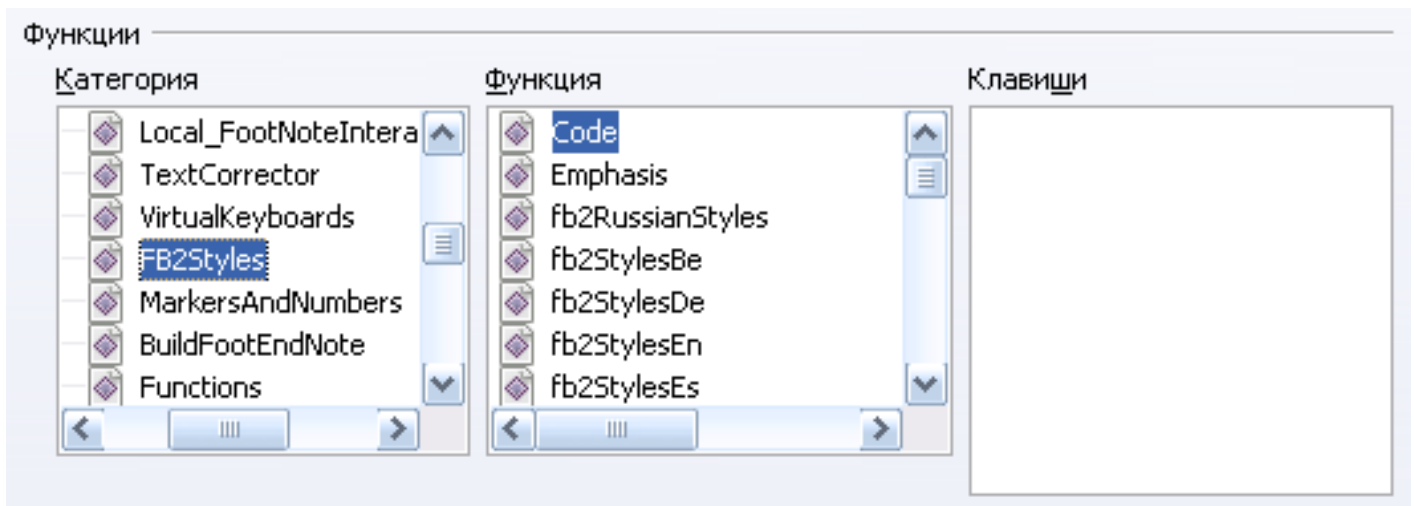
Клавиатура.



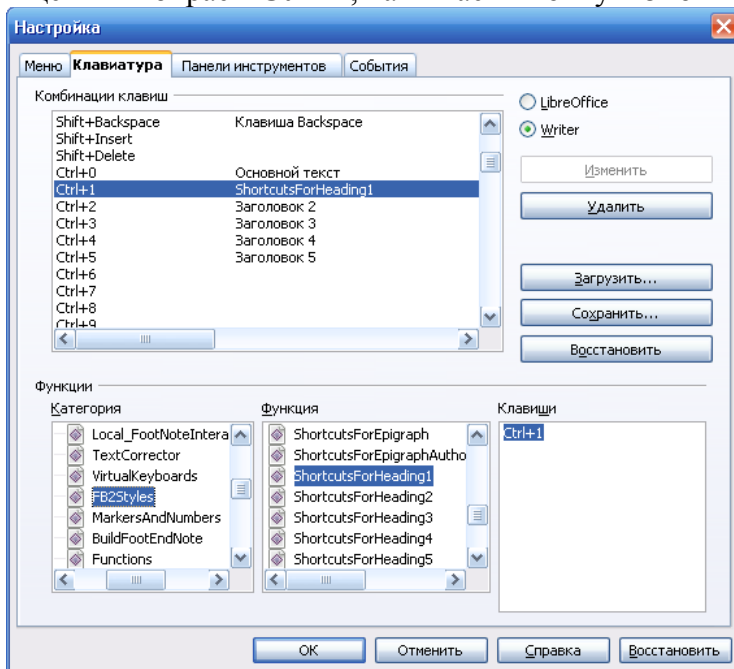
В древовидном списке Категория находим **Макросы LibreOffice** (для OpenOffice будет **Макросы OpenOffice**, для StarOffice – **Макросы StarOffice**). Раскрываем вложенные ветви **user- >OOoFBTools**:



Функции для работы с fb2 стилями находятся в модуле **FB2Styles**. Выделяем его. В окне **Функции** отобразятся все функции:



Теперь ищем функцию «**ShortcutsForHeading1**». Выделяем ее. В окне **Комбинации клавиш** ищем и выбираем **Ctrl+1**, нажимаем кнопку **Изменить**:



В окне **Комбинация клавиш** видно, что нужной нам комбинации **Ctrl+1** уже задан стиль **Заголовок 1**. По-идеи, можно было бы так и оставить – ничего не назначать. Но в этом случае, при нажатии **Ctrl+1** абзац будет размечен стилем **Заголовок 1**, но дополнительные обработки текста произведены не будут. Поэтому в любом случае меняем его на функцию «**ShortcutsForHeading1**».

Аналогично назначаем быстрые клавиши для тех fb2 стилей, с которыми мы работаем наиболее часто. Функции всех стилей приведены в таблице выше.

Нажимаем кнопку **Ок**. Все, теперь для того, чтобы быстро и удобно было работать со стилями.

Примечание: К сожалению, не на все клавиши можно подключать выполнение пользовательских функций. Много клавиш жестко предопределены в офисе. Такие клавиши и комбинации клавиш выделены серым цветом в окне **Комбинация клавиш** диалога **Настройки** офиса. Все попытки назначить им команды ни к чему не приведут. Поэтому, придется довольствоваться теми комбинациями, которые «разрешены» - они все выделены черным цветом.

Примечание: Разметка текста fb2 стилями при помощи назначенных быстрых клавиш учитывает и то, нужно ли сбрасывать форматирование абзаца до применения fb2 стиля. Вы можете

включить/отключить эту возможность на **Панели fb2 стилей** с помощью переключателя Сброс формата до fb2 стиля. Теперь можно закрыть Панель стилей и работать при помощи быстрых клавиш.

2. Особенности работы с текстом

2.1. Работа с таблицами

2.1.1. 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.

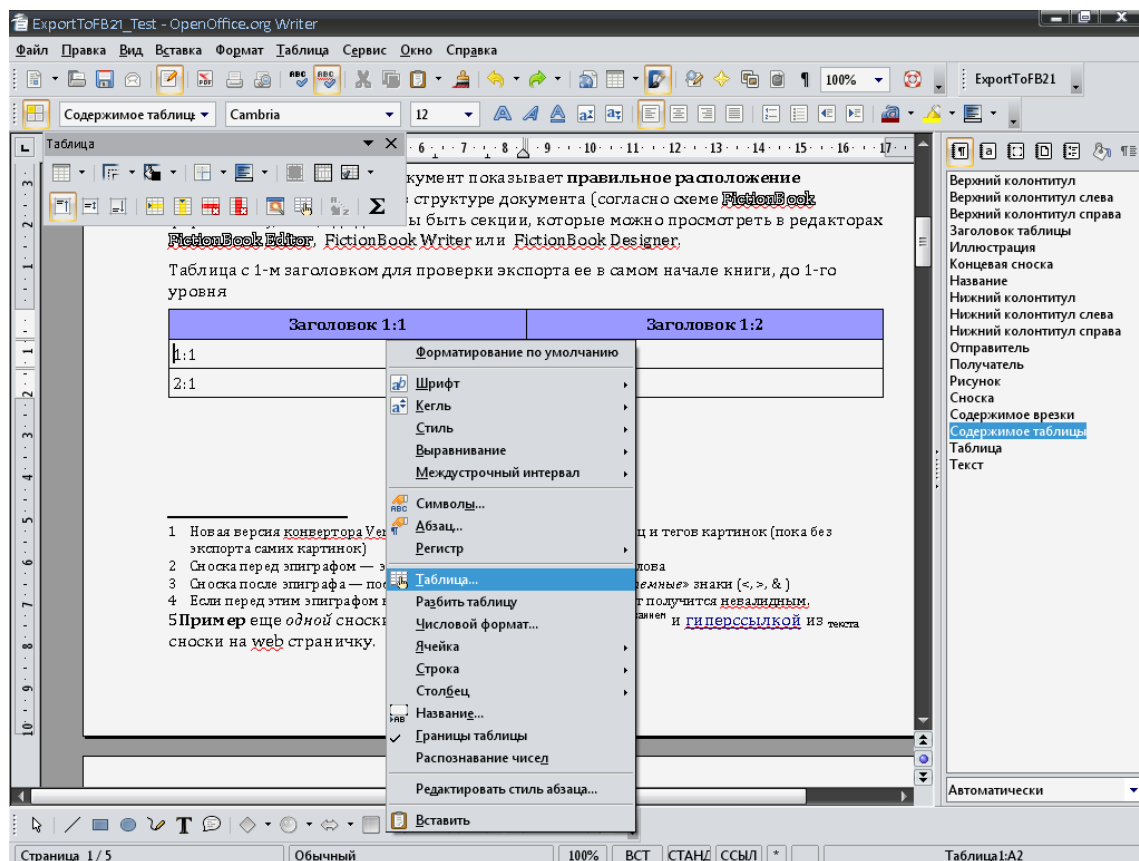
Конвертер корректно экспортирует **только простые** таблицы. Экспортируются и сложные таблицы, но не все сложные таблицы экспортируются согласно их «геометрии». Сложные таблицы, где в ячейку вложена другая таблица – игнорируются (ввиду чрезвычайной сложности создания алгоритма...).

2.1.2. Название таблицы

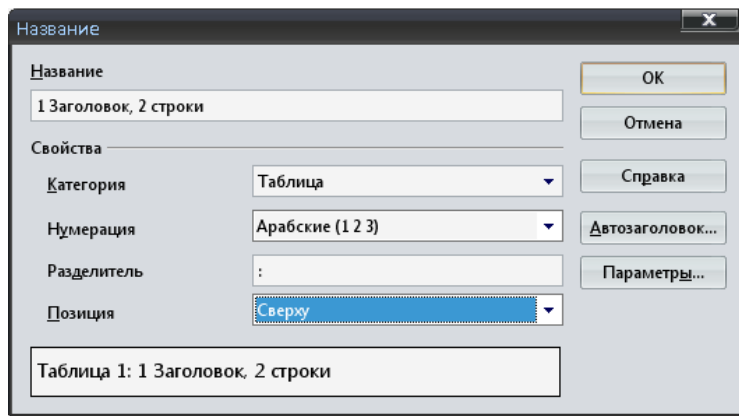
При желании можно дать таблице название (сверху или снизу таблицы). Это название также будет экспортироваться, с учетом форматирования.

Делается это так.

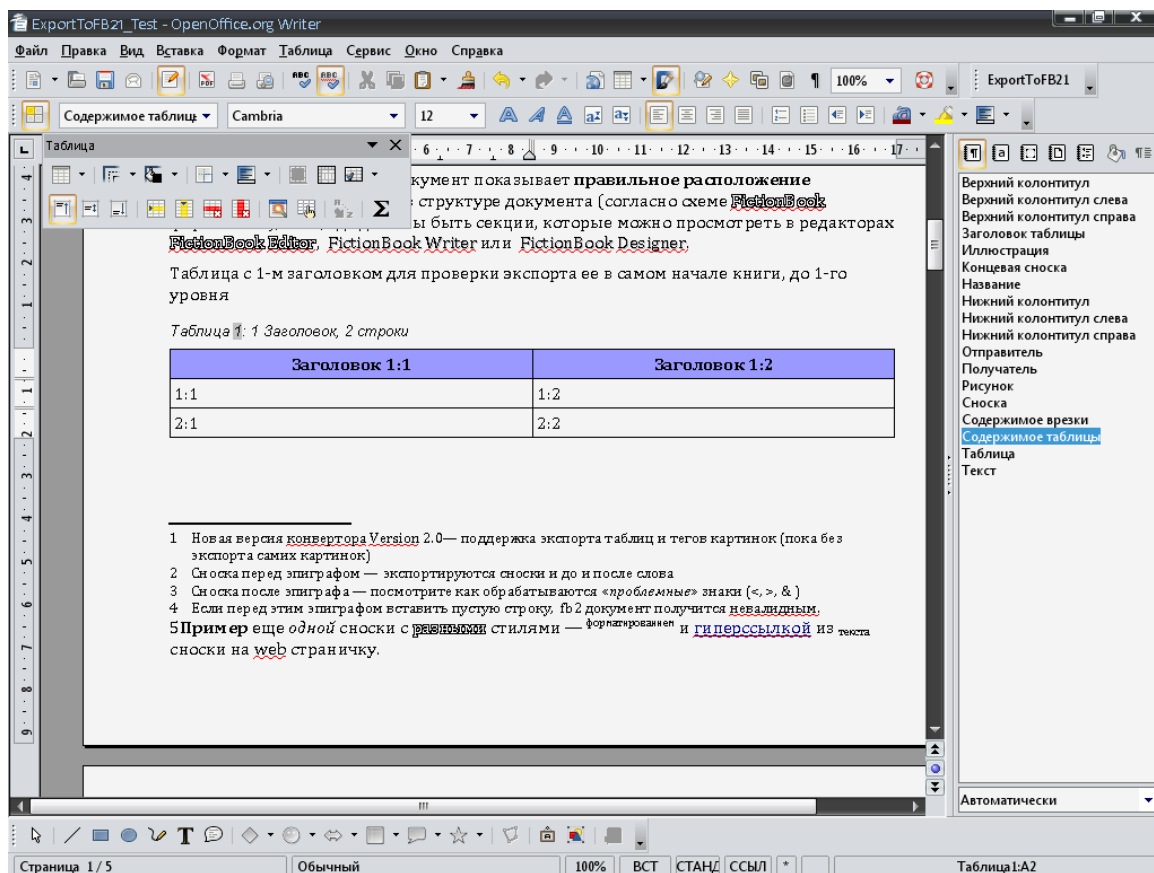
На нужной таблице кликните правой клавишей мышки и выберите «Название».



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



В результате получим следующее:

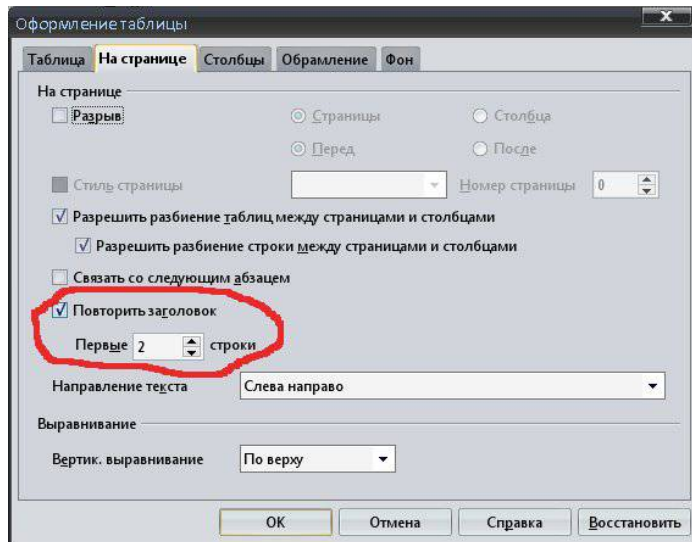


В названии имеется текстовое поле со сквозной нумерацией таблицы — оно тоже экспортируется.

2.1.3. Задание числа заголовков

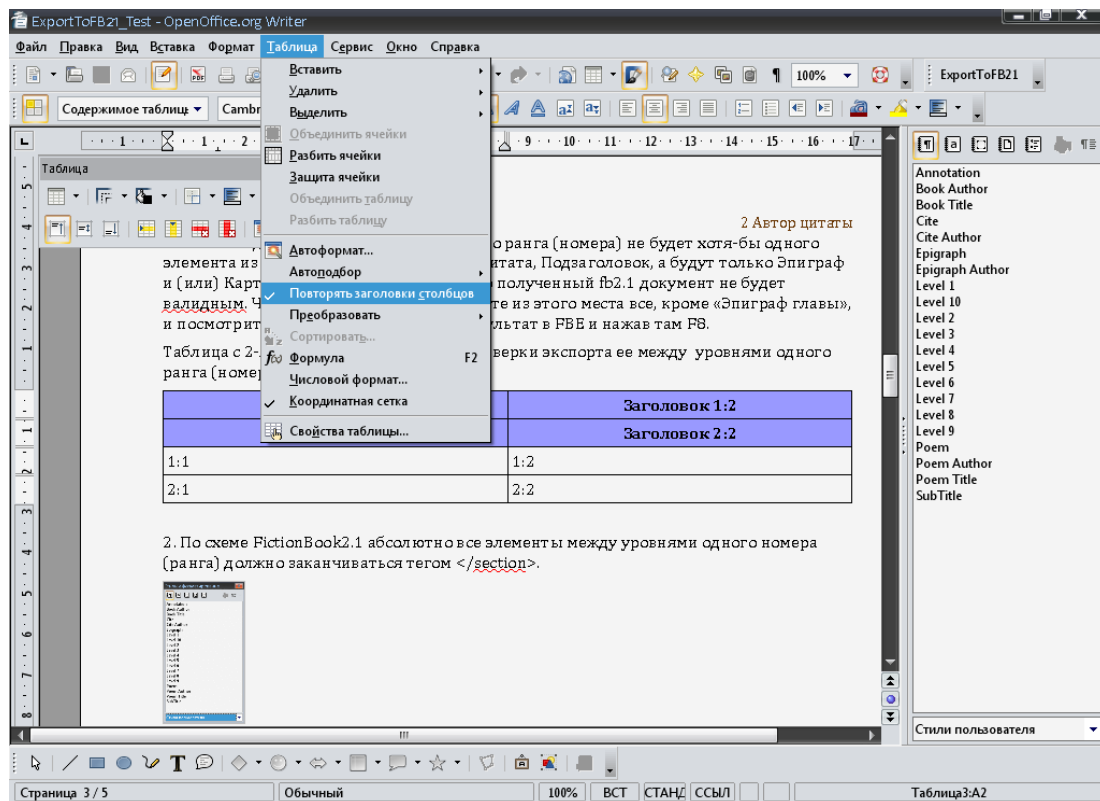
Если наша таблица содержит заголовок, или несколько, то надо задать число строк, из которых он будет состоять, и включить опцию, сигнализирующую, что заголовок есть.

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.



2.1.4. Важен порядок вставки таблиц в документ

В связи с особенностью алгоритма конвертера 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*!

2.1.5. Выравнивание строк ячеек таблицы

Можно выравнивать строки в каждой ячейке **по ширине**, как вам угодно – конвертер 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 конвертера 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.

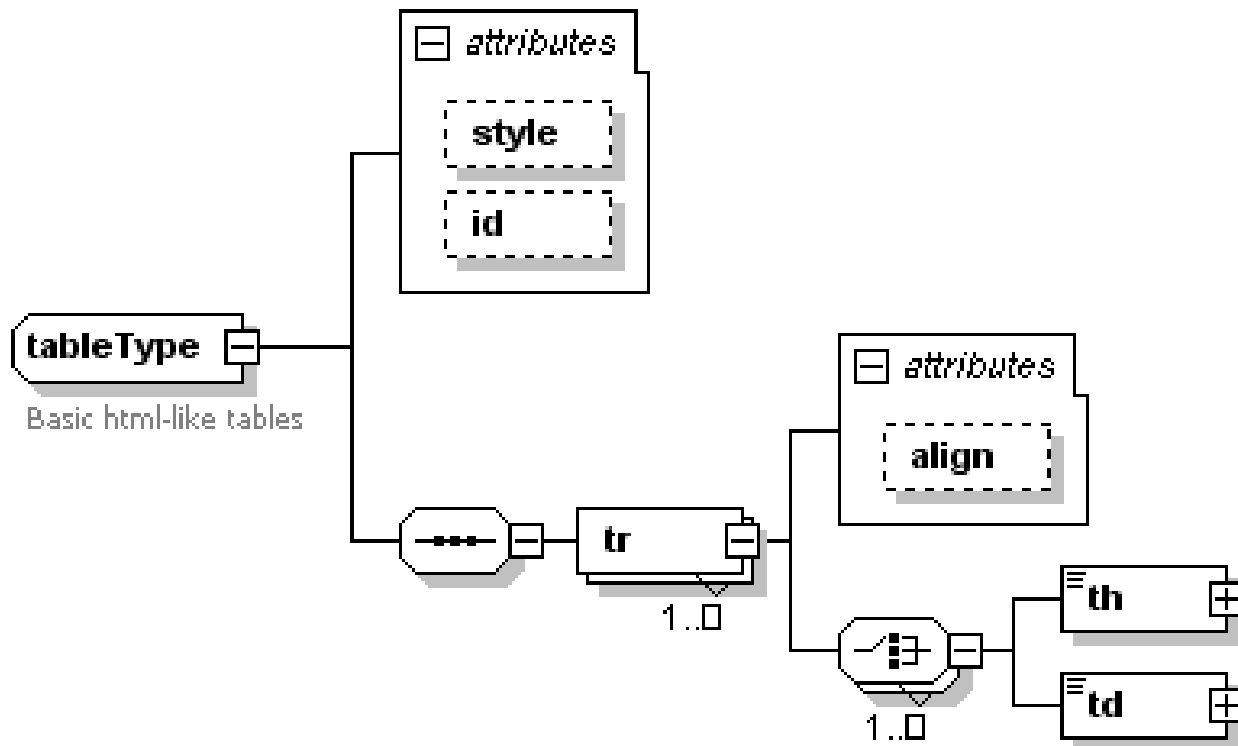
2.1.6. «Пустые» ячейки таблицы

Иногда требуется, чтобы в таблице были ячейки с «пустым» содержанием. Конвертер 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).

2.1.7. Структура элементов таблицы

Таблица формата fb2 имеет следующую структуру:

2.1.7.1. 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>

```

1.1.7.2. 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:

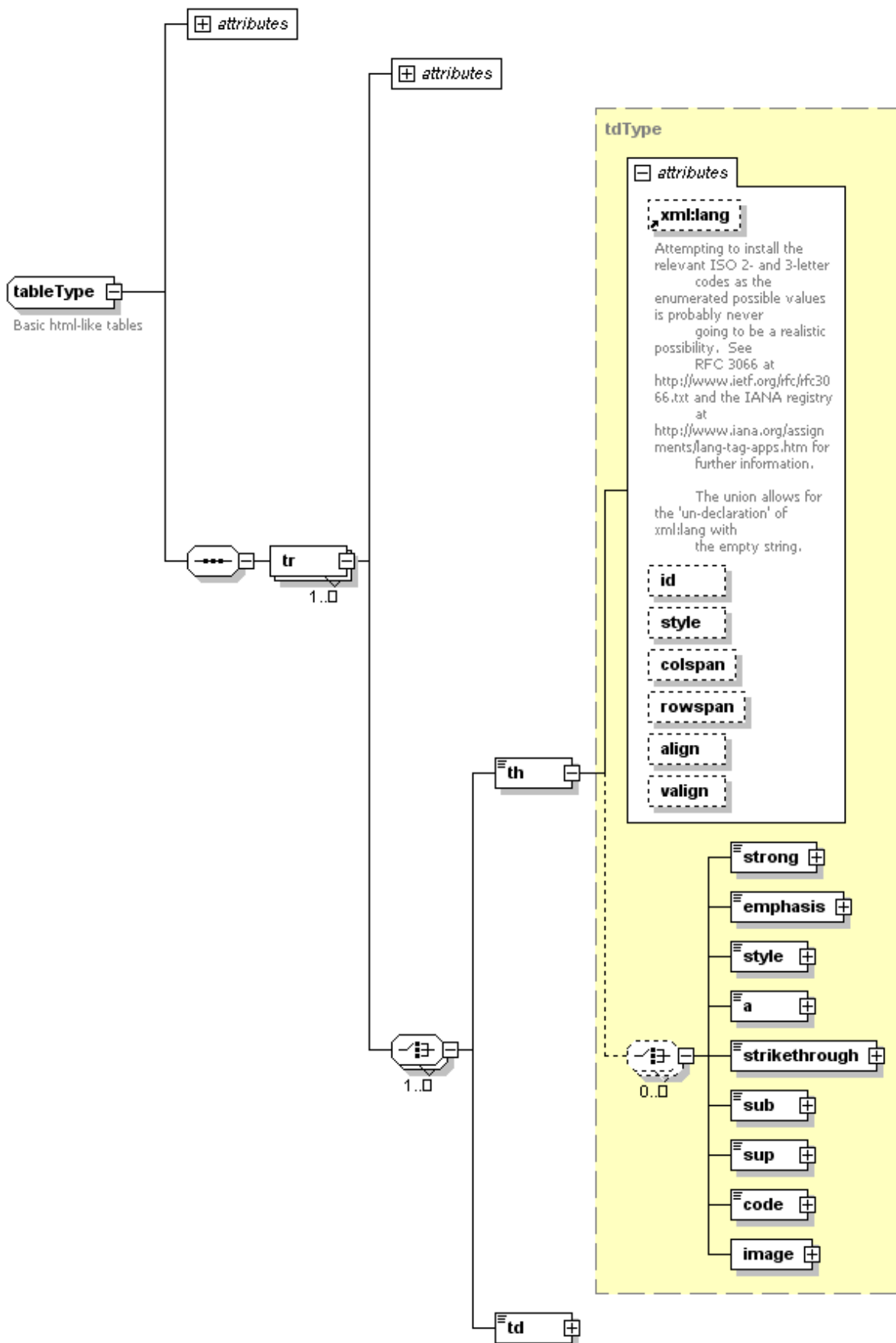
```

<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:



2.1.7.3. 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:

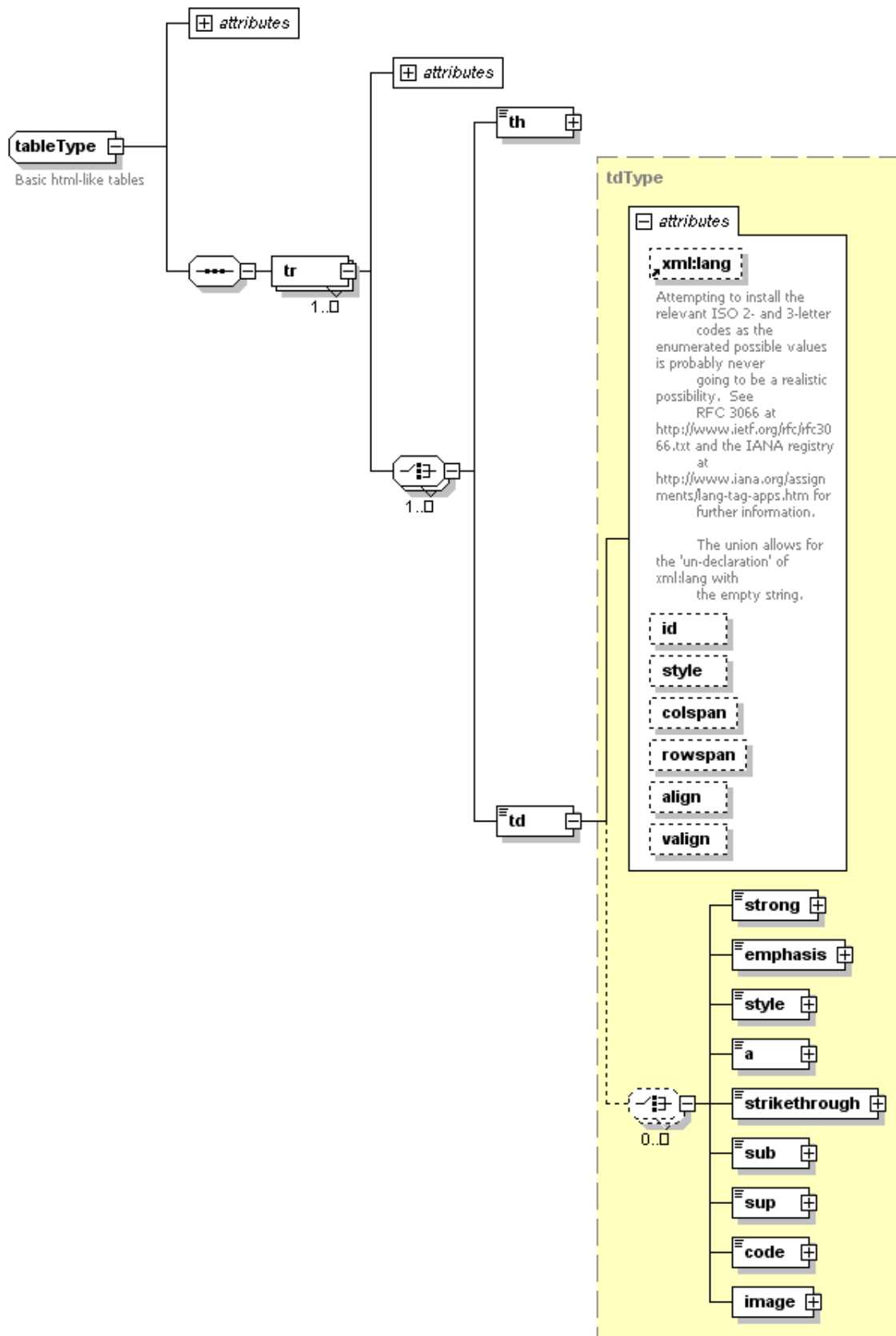
```
<table>
  <tr>
    <td/>
    <td> </td>
  </tr>
</table>
```

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:



2.1.8. 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)	1:2
1:1(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?

Конвертер **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 1:2
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)

Конвертер 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 границы ячеек.

2.1.8.1. Ячейки: экспорт абзацев как строк и экспорт абзацев как одну строку

Конвертер по умолчанию экспортирует многоабзацевые ячейки как абзацы. Но, иногда надо, чтобы абзацы определенной ячейки сконвертировались, как одна строка (слить их в один, разделив пробелом).

Например, есть таблица:

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...

В 1-й ячейке два абзаца по сути дела выражают одно и тоже. 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 конвертере 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...

2.1.9. Экспорт сложных таблиц

Пока конвертер 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 *разное число строк* – игнорируются (из-за сложности разработки алгоритма). Т.е. такая таблица будет проигнорирована:

1:1		
2.1.1	2.1.2	2:2

2.1.10. 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!

2.2. Работа с Текстовыми Врезками

2.2.1. 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.

Конвертер «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.

2.2.2. 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 маркеров-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!

Конвертер поддерживает 4 варианта экспорта Текстовых 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».

2.3. Работа с Картинками и другой графикой

2.3.1. Допустимые виды графики (Картинки, Формулы, Диаграммы, все виды векторных стрелок, схем)

fb2-читалки, обычно, «понимают» только png и jpeg форматы, закодированные в base64. Некоторые – 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, диаграммы, векторные стрелки, схемы** и т.д.. 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.

2.3.2. Привязка картинок (формул, диаграмм) к тексту

Всего существует 4 вида картинок:

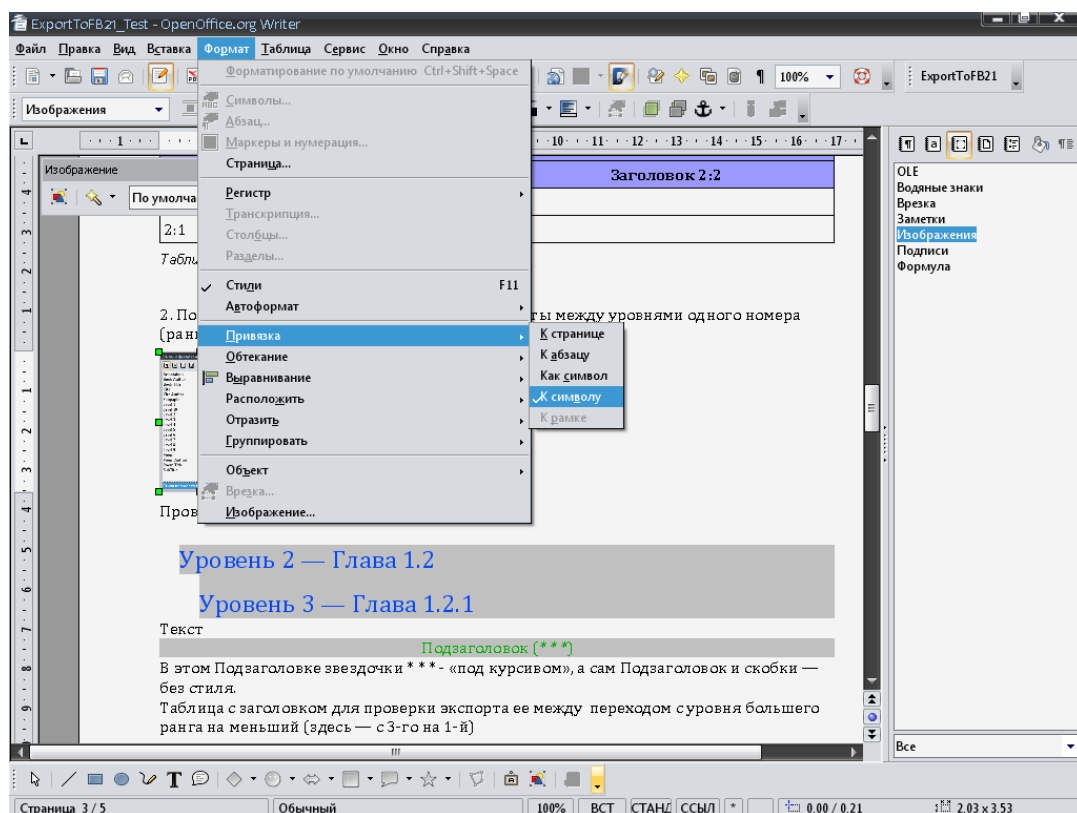
1. С привязкой к странице.
2. С привязкой к абзацу.
3. С привязкой к символу.
4. С привязкой как символ.

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).

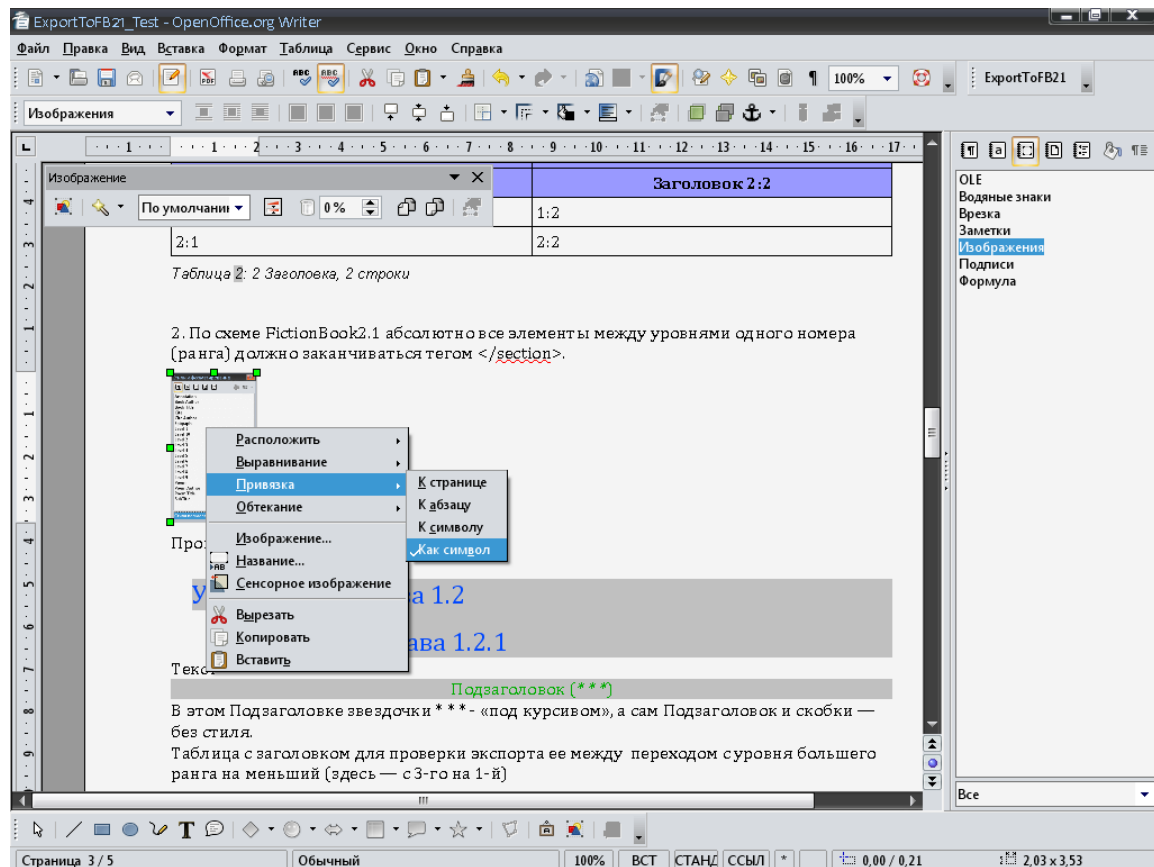
1-й вариант – картинка с привязкой к странице «не отлавливается» конвертером, и картинка будет проигнорирована. Картинки с названиями во врезках тоже пока «не отлавливаются», и экспортировано будет только их название из врезок в виде текста. Поэтому не используйте эти два вида картинок! Если надо задать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

2.3.2.1. Через меню «Формат» → «Привязка»

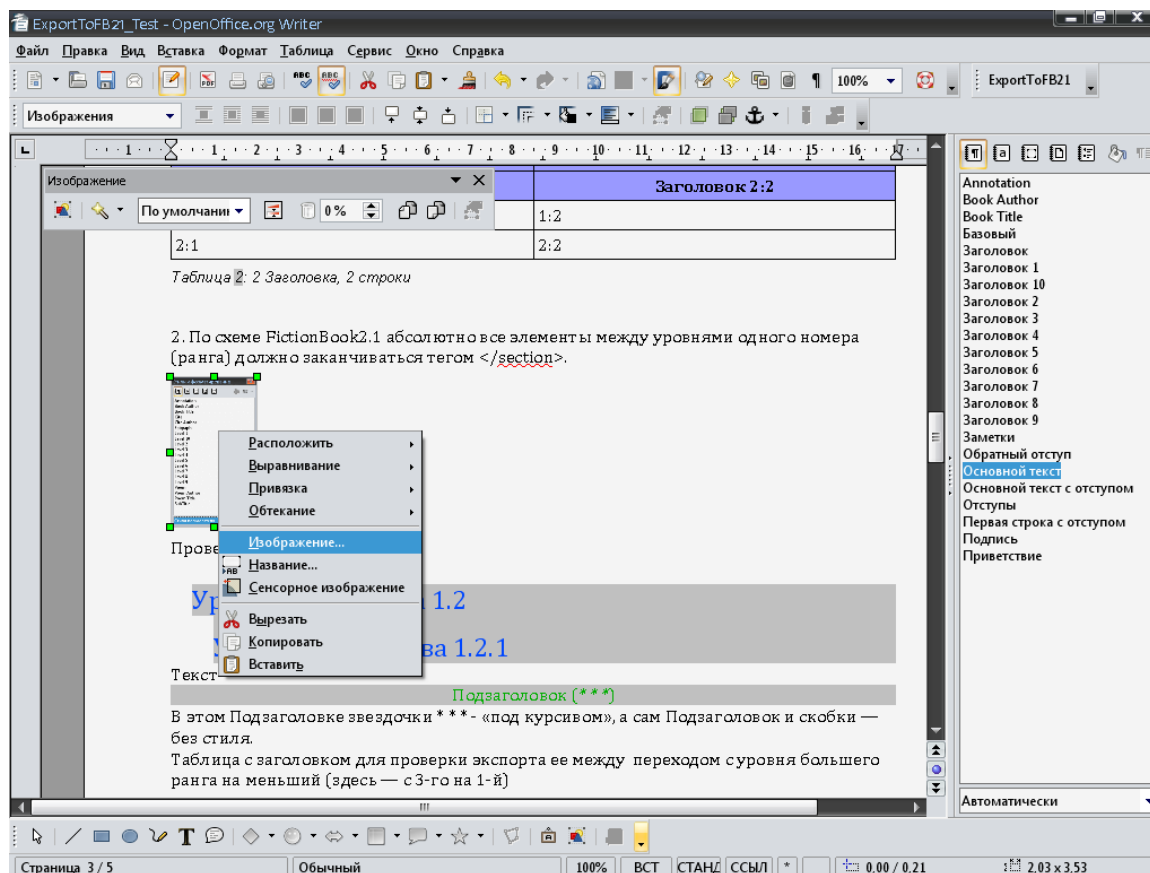


2.3.2.2. Через контекстное меню «Привязка»

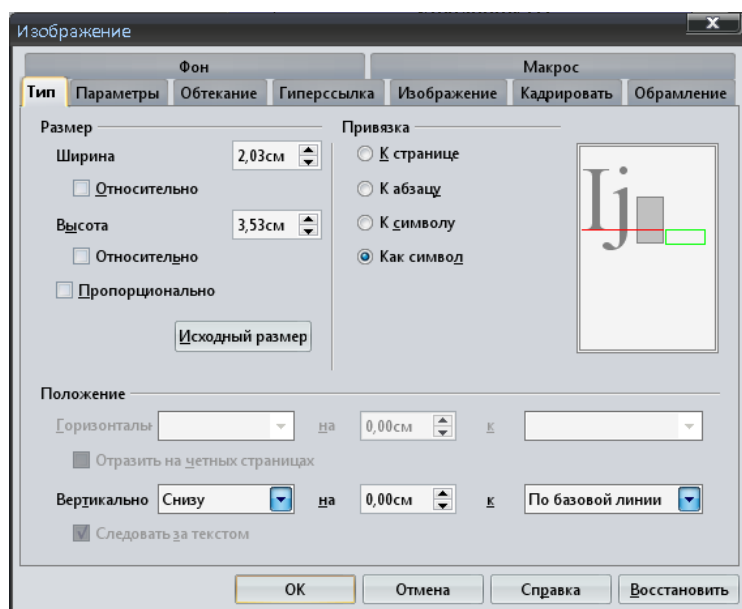


2.3.2.3. Через диалог «Изображение»

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

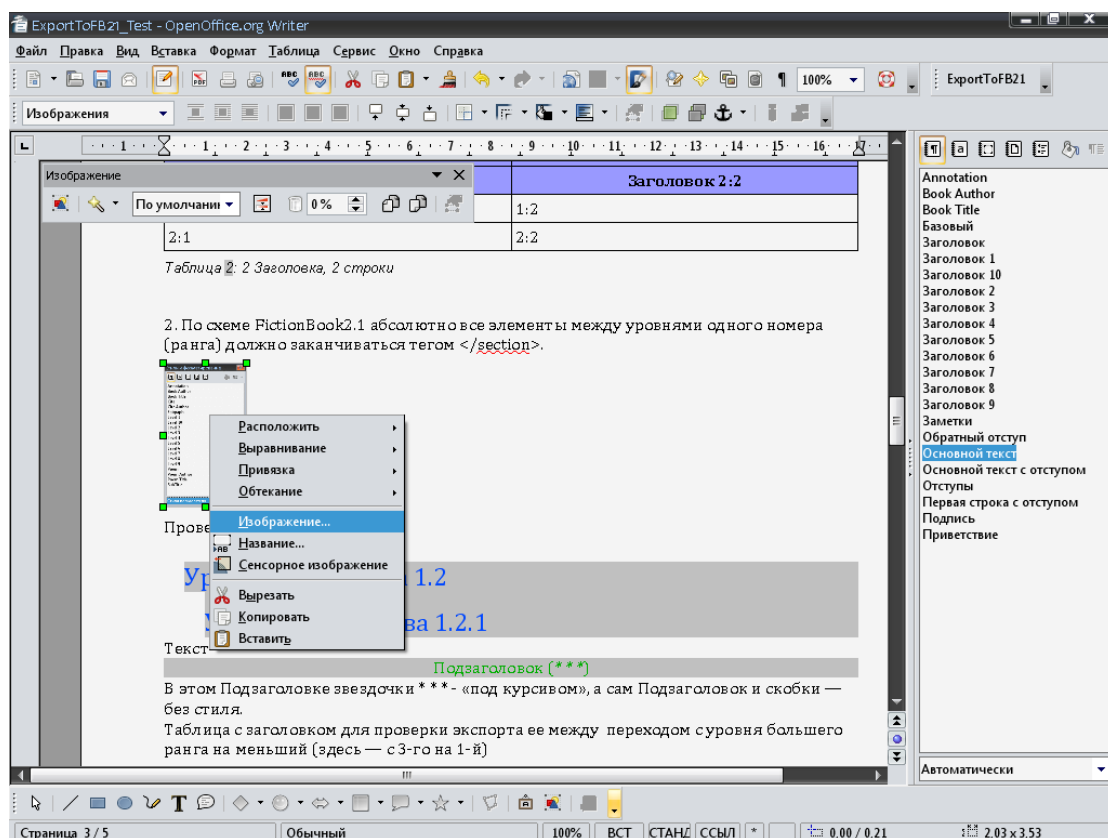


В появившемся диалоге на вкладке «Тип» выберите Привязку картинки:

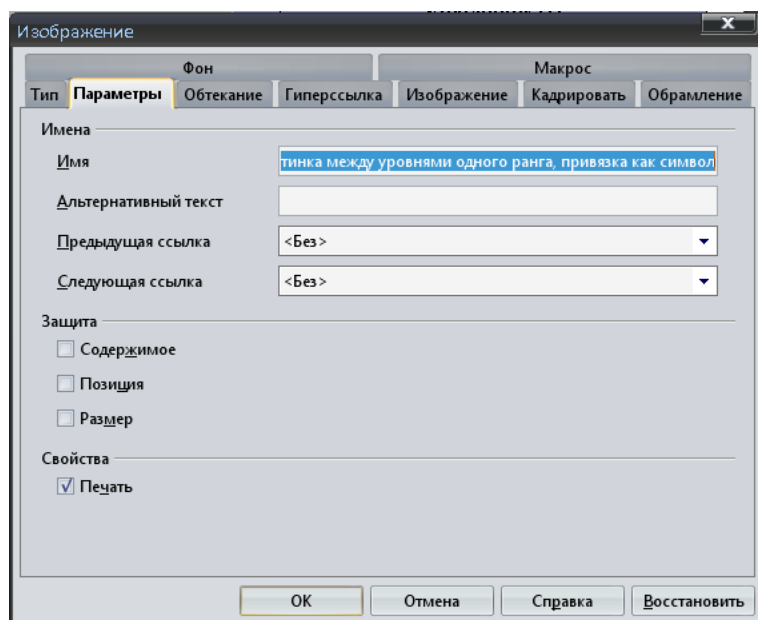


2.3.3. Задание имени картинке

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



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



2.3.4. 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](#)).

Конвертер 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, конвертер позволяет экспортировать такую структуру.

2.3.5. Обложки Книги

Конвертер 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 конвертером, как обложки.

2.3.6. Особенности работы с Графикой для 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.

2.3.7. Условия успешного экспорта Графики

Этот раздел более не актуален для OoFBTools, начиная с версии 2.28. То есть, все ниженаписанное в этом разделе про перезагрузку редактора можно не делать — конвертер сам автоматически перезагружает документ перед экспортом. Но на всякий случай текст этого раздела сохранен...

Т.к. Конвертер 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.

2.4. Работа с ссылками внутри документа

2.4.1. Создание ссылок

Очень полезная функция конвертера!

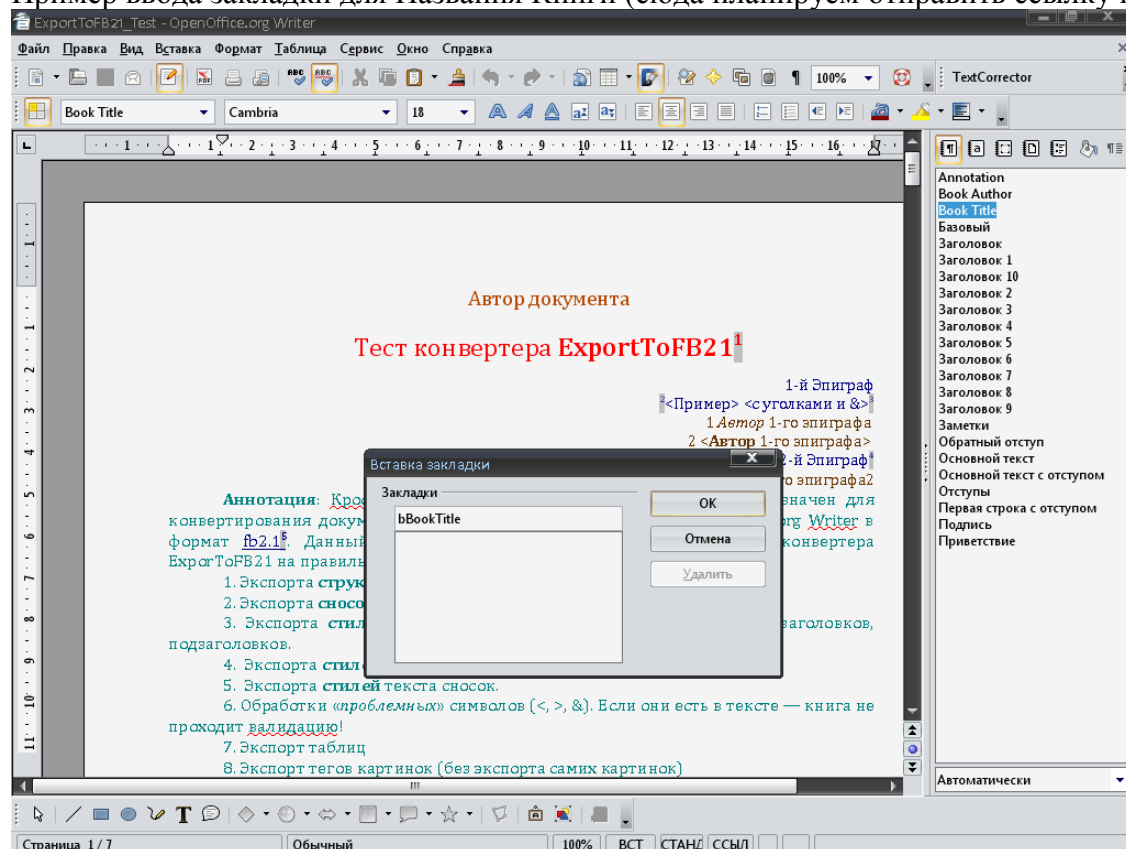
Например, надо из текста сноски по клику на какое-то слово перейти на Заголовок Поэмы. Или из Названия Книги – на Автора Цитаты. Или – из Автора Эпиграфа – на какой-нибудь раздел книги. Это легко теперь можно сделать. Как?

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

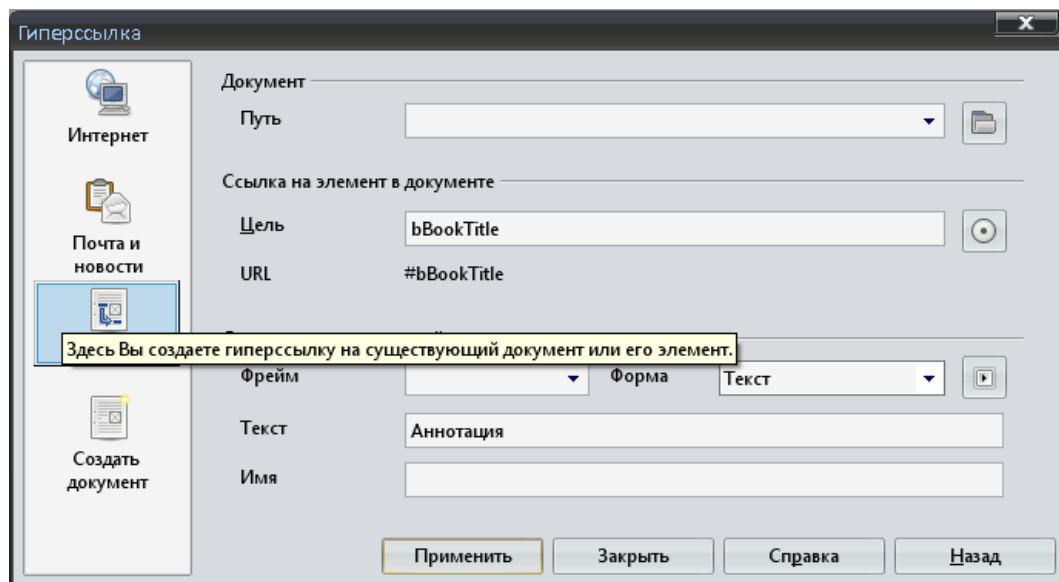
1. Поставьте курсор в любое место абзаца – «цели» (обычно в начало). Выберите меню «Вставка» → «Закладка». В появившемся диалоге введите имя закладки (желательно информативно, чтобы самому потом не запутаться, какая закладка к чему относится).

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

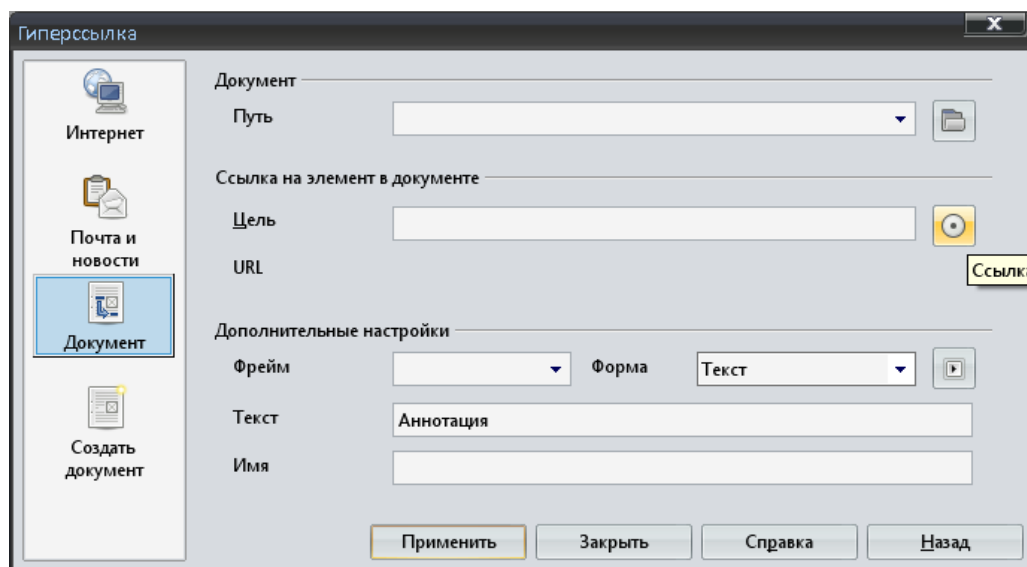
Пример ввода закладки для Названия Книги (сюда планируем отправить ссылку из сноски):



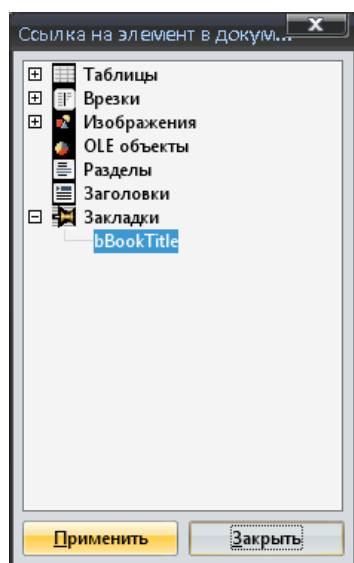
2. Выделите «источник». Через меню «Вставка» выберите «Гиперссылка», и в появившемся диалоге выделите иконку «Документ». В поле «Цель» введите имя нужной закладки, нажмите кнопку «Принять». Пример:



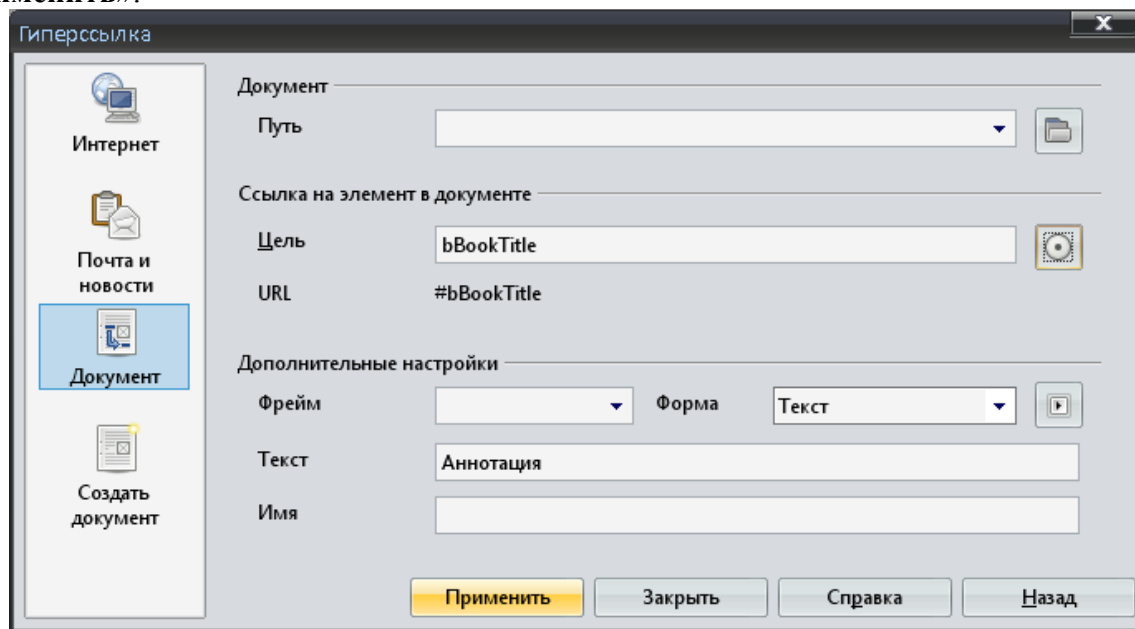
Можно вводить имя закладки не вручную (чтобы случайно не ошибиться), а с помощью диалога «Ссылка на элемент в документе». Эта кнопка – справа от поля «Цель»:



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



Теперь жмем «Заккрыть» 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 посмотрите 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.

2.4.2. Что важно помнить:

1. Ссылки на таблицы и (или) на текстовые 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!

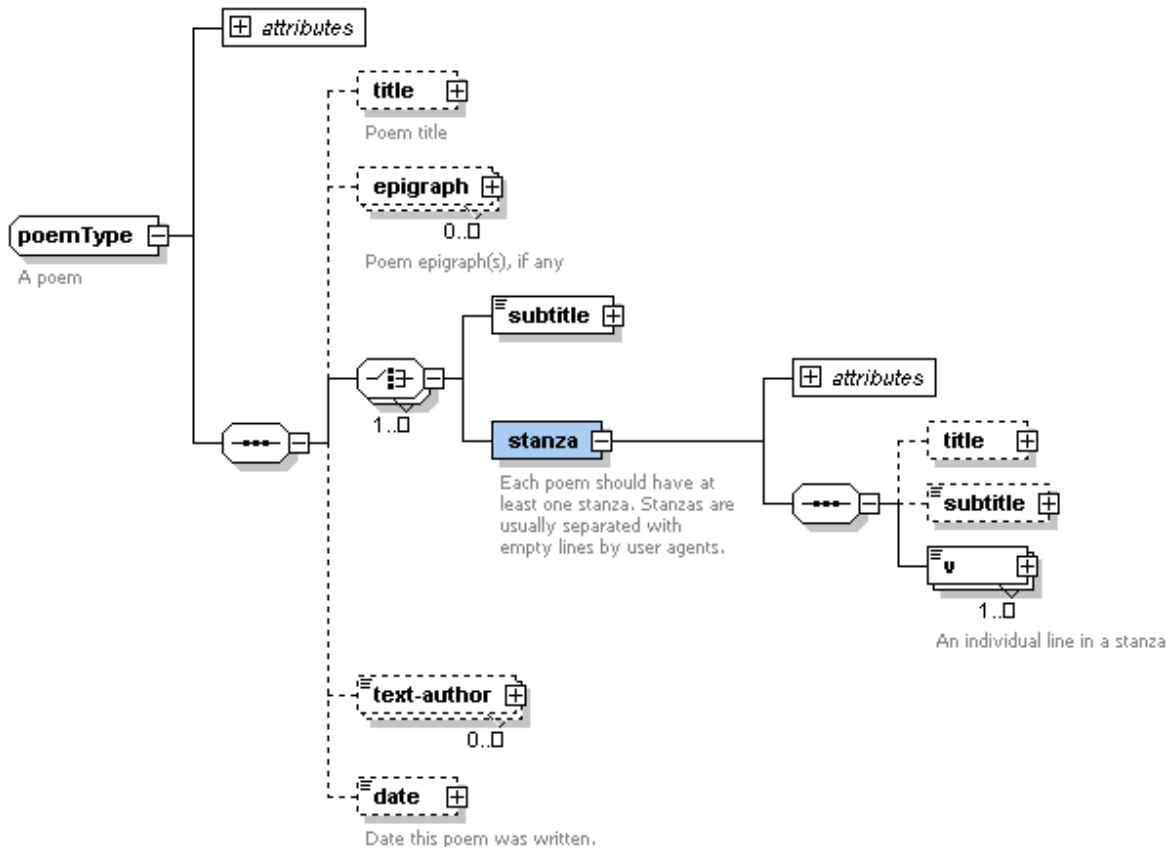
3. Как уже было сказано выше, *ни в коем случае* впереди имени закладки *не ставьте* символ «#»! Работать ничего не будет! Это «служебный» символ для обозначения ссылки для атрибута «href»

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 конвертер берет из имени закладки – Bookmark.Name. Писать дополнительный код проверки – это обречь конвертер на еще медленное конвертирование (Open BASIC – язык интерпретатор), а это не есть хорошо. Поэтому – будьте внимательны, когда вводите имя закладки!

2.5. Работа со стихами

2.5.1. Структура стиха

Согласно схеме FictionBook стих (поем) 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 работает, либо книга не будет валидной. Все зависит от того, что вы перепутаете. Так что будьте внимательны и сверяйтесь с этой схемой Поэмы!

2.5.1.1. 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!

2.5.2. Виды стихов

Конвертер 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».

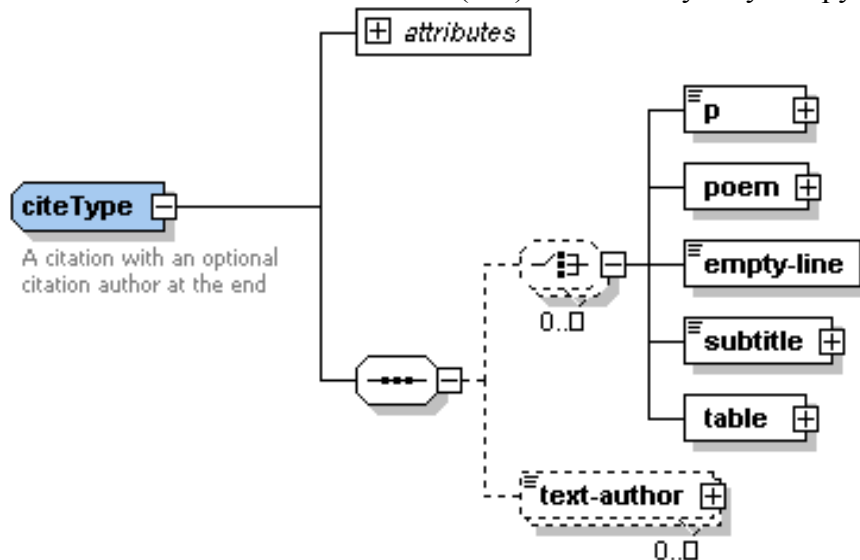
2. Поэма, где все четверостишия собраны под одной stanza. Для этого между абзацами со стихами не должно быть ни одной пустой строки.

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»!

2.6. Работа с цитатами

2.6.1. 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 конвертере.

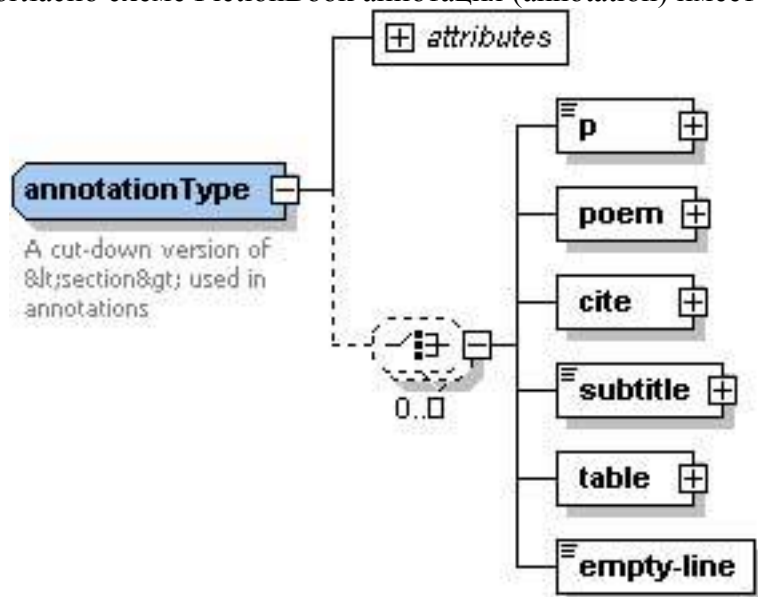
Если вы отформатируете стилями текст цитаты в соответствии с этой схемой на картинке, то конвертер экспортирует текст корректно. Если вы переставите местами **Автора Цитаты** и какой-либо элемент структуры Цитаты – то в fb2 файле книга не будет валидной. Так что будьте внимательны и сверяйтесь с этой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

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

2.7. Работа с аннотациями

2.7.1. Структура аннотации

Согласно схеме FictionBook аннотация (annotation) имеет следующую структуру:

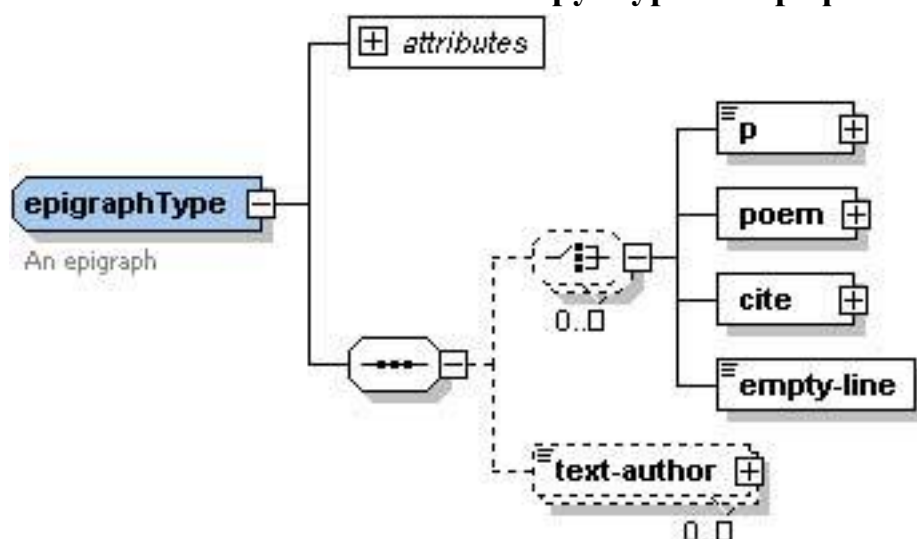


Т.е. он состоит из бесконечного числа чередований обязательных элементов (порядок очередности неважен) **Текст**, **Поэма**, **Цитата**, **Подзаголовок**, **Таблица**, **Пустая строка**. Все это полностью реализовано в конвертере.

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

2.8. Работа с эпитафиями

2.8.1. Структура эпитафия



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

Если вы отформатируете стилями текст эпитафия в соответствии с этой схемой на картинке, то конвертер экспортирует текст корректно. Если вы переставите местами **Автора Эпитафия** 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.

2.8.2. Отделение одного элемента эпиграфа от другого

Как правило, конвертер 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 конвертер «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 Конвертер распознает и Цитату, и Поэму, как *два самостоятельных* элемента Эпиграфа.

Пример этого см. в файле-примере ExportToFB21_example-1.odt.

Поэкспериментируйте.

2.9. Работа с внутренними ссылками (гиперссылки)

Гиперссылки используются для перехода из одного места документа в другое.

Конвертер позволяет экспортировать следующие внутренние ссылки:

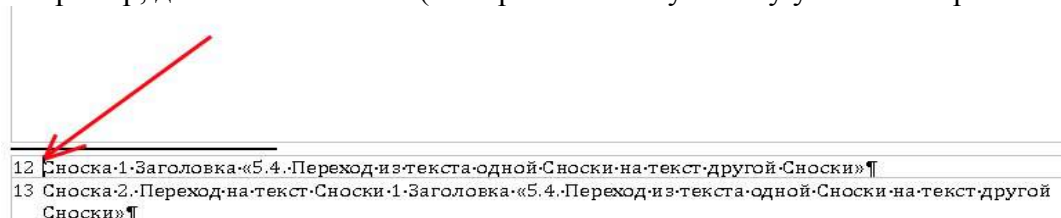
1. Переход из любого места Документа на любое другое место Документа (текст, Цитаты, Стихи, Заголовки и т.д.).
2. Переход из любого места Документа на любое место в тексте Сноски.
3. Переход с текста любой Сноски на любое место Документа.
4. Переход из текста одной Сноски на текст другой Сноски.

Конвертер пока не может экспортировать внутренние ссылки (переходы), связанные с Таблицами и Текстовыми Врезками.

Для задания внутренних ссылок (переходов) необходимо сделать следующее:

1. На абзац, **КУДА** нужно перейти, ставим закладку.

Например, для Текста Сноски (на картинке на эту сноску указывает красная стрелка):



Часть-6.-Текстовые-Врезки,-Поля-и-другое¶

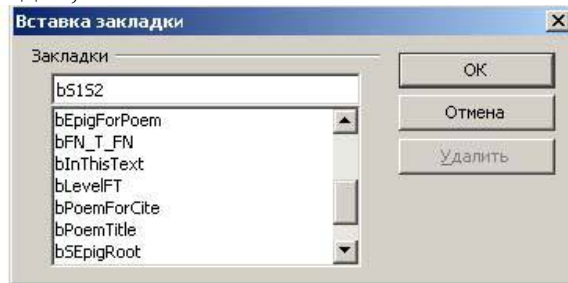
6.1.-Текстовые-врезки¶

Врезки могут экспортироваться либо в виде *простого текста*, либо в виде *цитаты*, либо в виде *таблицы*. Выбор-варианта-осуществляется в диалоге **Настройки** Конвертера.¶

Для-врезки-существует-4-вида-привязки:¶

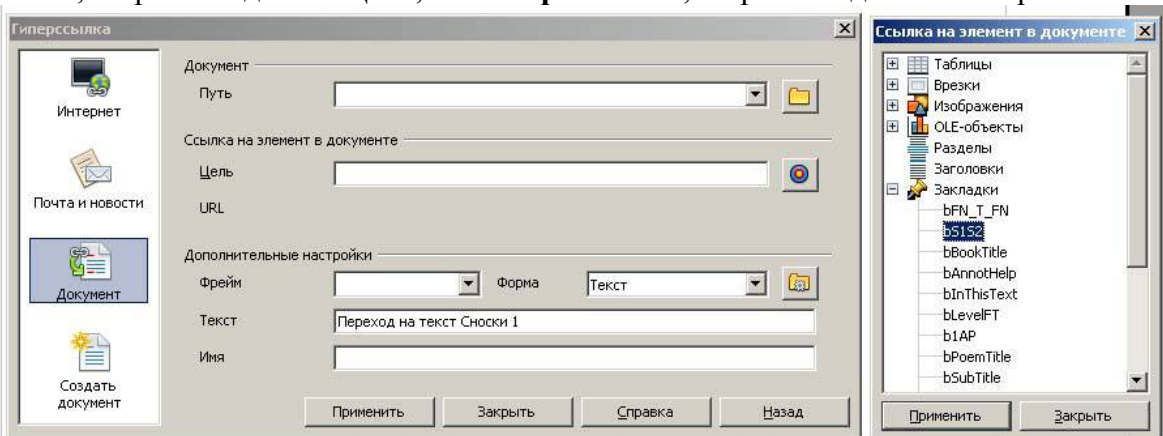
- 1.-К-странице.¶
- 2.-К-абзацу.¶
- 3.-К-символу.¶

Ставим курсор в начало абзаца, вызываем диалог **Закладка** (меню **Вставка->Закладка...**).
Задаем там нужное имя закладки, жмем **Ок**:



12 Сноска-1-Заголовка-«5.4-Переход-из-текста-одной-Сноски-на-текст-другой-Сноски»¶
13 Сноска-2-Переход-на-текст-Сноски-1-Заголовка-«5.4-Переход-из-текста-одной-Сноски-на-текст-другой-Сноски»¶

2. В абзаце, **ОТКУДА** нужно перейти, выделяем слово или словосочетание-переход, и форматируем его, как гиперссылку (меню **Вставка->Гиперссылка**). В открывшемся диалоге для **Документа** выбираем **Цель** — **Закладки**, а там — нужную закладку для абзаца **КУДА**. Жмем **Применить**, закрываем диалог **Цели**, жмем **Применить**, закрываем диалог **Гиперссылки**.



12 Сноска-1-Заголовка-«5.4-Переход-из-текста-одной-Сноски-на-текст-другой-Сноски»¶
13 Сноска-2-Переход-на-текст-Сноски-1-Заголовка-«5.4-Переход-из-текста-одной-Сноски-на-текст-другой-Сноски»¶

Все!

3. 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 исключением 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.

3.1. 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**] – (Quotations)
 [**Подзаголовок**] – (Subtitle)
 [**Стих**] – (Poem)

3.2. 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 Заголовок 1 ... Заголовок 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

3.2.1. 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] – (Quotations)
 [Подзаголовок] – (Subtitle)
 [Стих] – (Poem)

3.2.2. 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] } – (Quotations)
 { [Подзаголовок] } – (Subtitle)
 { [Стих] } – (Poem)

3.2.3. 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 Заголовок 2 и Заголовок 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.

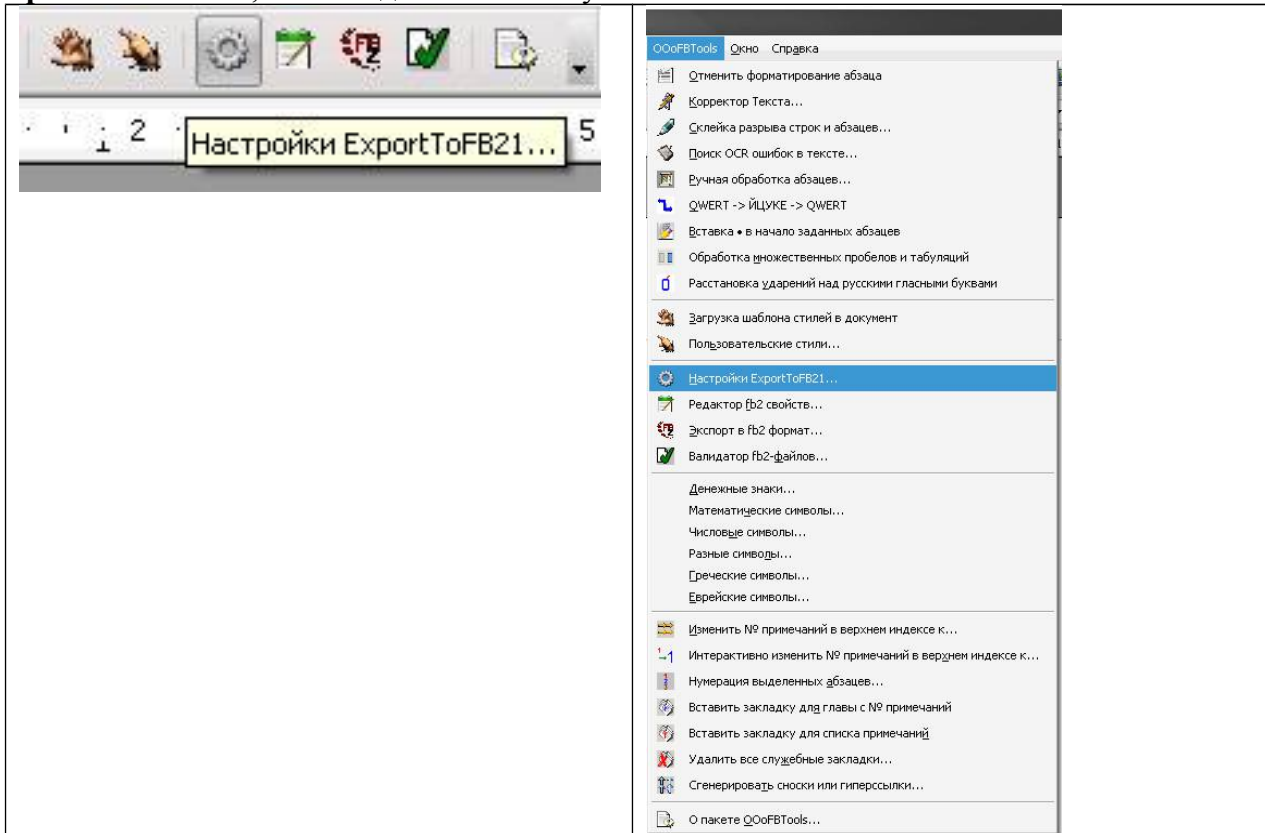
то полученный fb2 документ не будет валидным! Номер следующего (смежного) Уровня по возрастанию не может отличаться от предыдущего более чем на 1. По убыванию – как угодно. Поэтому, при разметке структуры будьте очень внимательны!

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

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

Б. Задание требуемых настроек экспорта

Диалог «Настройки fb2 Конвертера» вызывается нажатием либо кнопки «Настройки ExportToFB21 ...», либо – одноименным пунктом меню OoFBTools:



Диалог обеспечивает ряд «удобств» при работе с конвертером.

Настройки fb2 Конвертера

Сноски и Концевые Сноски
 Экспорт сносок и концевых сносок: ☐ В общий блок ☒ В разные блоки

Сноски
☐ Пробел перед сноской Стил: [x] Заголовок блока сносок: Примечания

Концевые Сноски
☐ Пробел перед сноской Стил: {x} Заголовок блока сносок: Комментарии

Стихи
☐ Собирать Заголовки Стихов в одну строку
☐ Собирать Авторов Стихов в одну строку
☒ Собирать Подзаголовки Стихов в одну строку

Эпиграфы
☐ Собирать Авторов Эпиграфа в одну строку

Таблицы
 Выравнивание Заголовков: По высоте (th valign):
 Выравнивание Строк: По высоте (td valign):

Цитаты
☐ Собирать Авторов Цитаты в одну строку
☒ Собирать Подзаголовки Цитаты в одну строку

Аннотации
☒ Собирать Подзаголовки Аннотации в одну строку

Подзаголовок (SubTitle)
☒ Собирать Подзаголовки в одну строку

Графические объекты
☐ Сохранять название, как атрибут "title"
☒ Экспортировать Графические Объекты

Преобразовывать форматы:
 jpg, jpeg, как: jpeg
 png, как: png
 gif, как: png
 tif, tiff, bmp, как: png
 другие форматы, как: png
 формулы и диаграммы, как: png

Mode: ☐ Original (As in Text) ☒ Logical Size

☒ Экспорт текстовых полей

Экспорт качества jpeg
 Цветовой режим: ☒ Цвет ☐ Серый
 Качество: 100 %

Текстовые Врезки
 Экспортировать как: Структура

Вложенные списки
 Число пробелов для отступа: 3

Разное
☒ Обработка "проблемных" символов (&, <, >)
☐ Удалять пустые строки
☒ Обработка ссылок внутри документа
☒ Сохранять fb2-файл в utf-8, если в тексте есть маркеры
☒ Шрифт 'Courier New', как определитель листингов программ

Созданный fb2-файл
☒ Обработать "пустые" тэги (</x><x>, <x></x>, </x> <x> и <x> </x>)
☒ Валидация созданного fb2-файла

☐ Запустить созданный fb2-файл (Просмотр)

OK
Отмена

Настройки по умолчанию

1. Экспорт сносок

Есть возможность выбора – экспортировать сноски и концевые сноски в один блок или же в разные блоки. Для каждого вида сносок можно выбрать свои настройки:

а. Вставлять или нет пробел между словом и сноской (многим удобнее читать, когда после слова и перед номером сноски стоит пробел).

б. Стил сноски: x, [x], [x], {x}, { x }, (x) или (x). По умолчанию для сносок будет использоваться [x], а для концевых сносок {x}.

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

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

3. 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).

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

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

Включение этой опции также увеличивает время экспортирования документа.

6. Опция «Удаление пустых строк» говорит сама за себя – удаляются все пустые строки из

документа в результирующем fb2 файле.

7. 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.

8. 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 конвертер будет «слетать» при экспорте графики – просто пока отключите эту опцию – картинок в fb2-файле не будет, зато все остальные «изыски» конвертера 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.

Для jpeg есть возможность выбора, в каком цветовом режиме экспортировать картинки: в градациях серого или в цвете. Преобразование цветовой схемы делается в момент экспорта. Также можно выбирать процент качества jpeg при экспорте.

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

9. 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.

10. 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.

11. 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](#).

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

13. Опция «Экспорт текстовых полей» служит для экспорта любых текстовых полей (времени, номеров страниц и т.д.)

14. Опция «Генерировать сложный ID Книги» служит для генерации ID Книги, исходя из времени создания fb2-файла. Используется генерация случайных чисел.

Такой тип генерации ID используется автоматически, если по какой-либо причине не генерируется полноценный GUID для ID для fb2-файла (чисто теоретически, так как проблем с генерацией GUID не должно быть).

15. Группа опций «Созданный fb2 файл»:

Опция «Обработать “пустые” тэги `</x><x>`, `<x></x>`, `</x> <x>` и `<x> </x>`» позволяет 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.

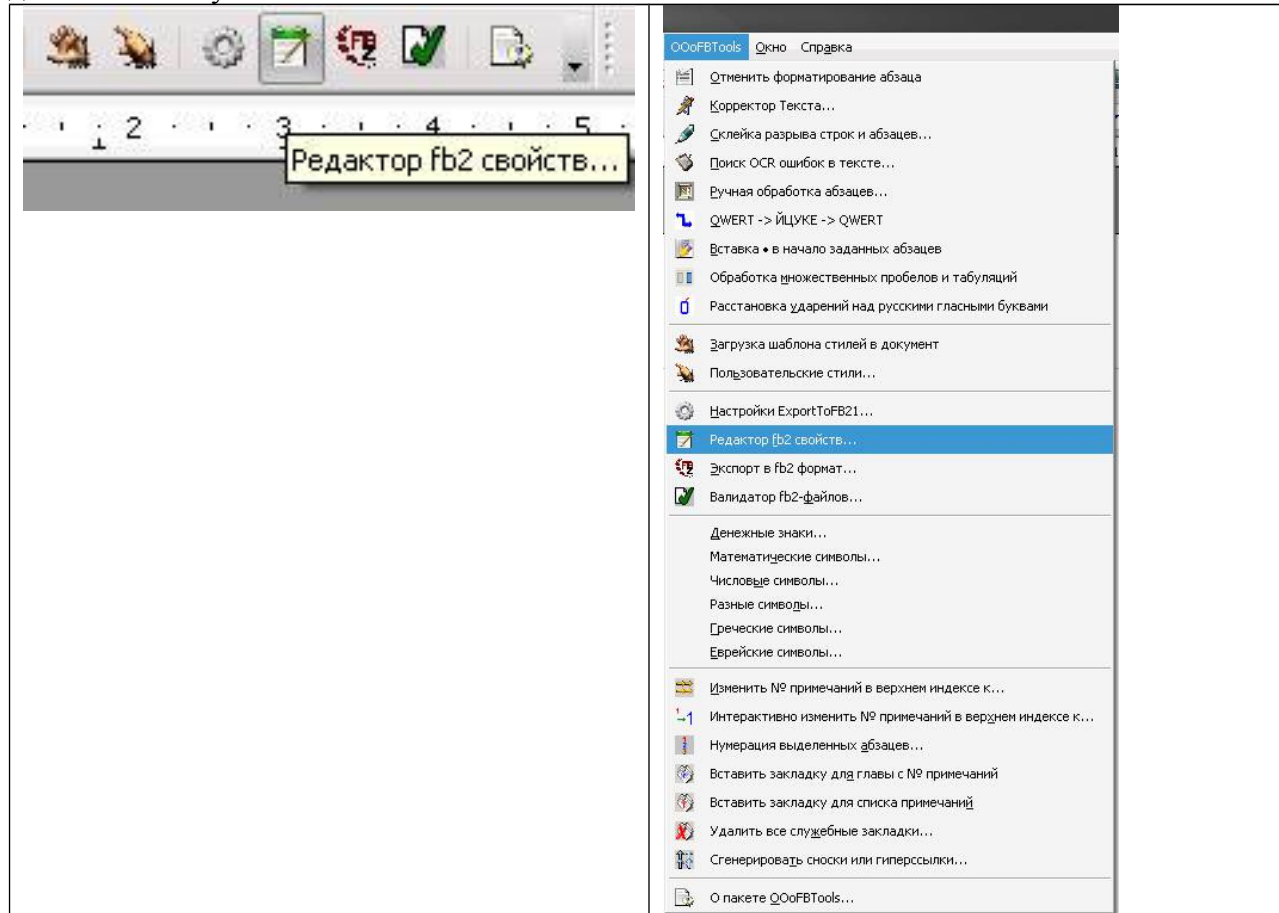
16. Группа опций «Созданный fb2-файл»:

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 просмотреть, 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.

В. Задание описания Книги в Редакторе fb2 свойств

Редактор fb2 свойств вызывается нажатием либо кнопки «Редактор fb2 свойств ...», либо – одноименным пунктом меню OooFBTools:



Редактор fb2 свойств предназначен для сохранения и редактирования описания книги в свойствах документа. В дальнейшем, при конвертировании в fb2-формат, конвертер считывает эти свойства из документа и на их основании формирует раздел <description> fb2-файла. В свойства документа сохраняется так же и выбранная кодировка (windows-1251 или utf-8), в которой конвертер должен создать fb2-файл. Это удобно.

Редактор имеет несколько «вкладок», по которым сгруппированы данные описания книги.

1. Общие правила при работе с Редактором fb2 свойств

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.

Для Жанров Книги, Авторы Книги, Переводчиков Книги, Авторы fb2-файла, Серий Книги, Серий Бумажной Книги и Пользовательской информации (Custom Info) можно формировать списки. Для этого в соответствующие поля вводим необходимую информацию и добавляем ее в список нажатием кнопки «+». При необходимости удалить какой-то элемент списка (например, неверно введенный Жанр) необходимо выделить его мышкой и нажать кнопку «-».

Например, для Авторы Книги:

Добавление всех Авторы в Список производится нажатием кнопки "+". Но данные последнего Автора можно просто оставить в полях ввода, не добавляя их в Список нажатием кнопки "+".

Если Автор Книги всего 1, то можно его данные оставить в полях ввода, не добавляя в Список нажатием кнопки "+". Т.е., при сохранении fb2-данных в свойства, вся информация в полях будет автоматически добавлена в Соответствующие Списки.

Кнопки «↑» и «↓» служат для перемещения выбранного элемента списка вверх и вниз соответственно.

Для Жанров Книги, Авторы Книги, Переводчиков Книги, Авторы fb2-файла, Серий Книги, Серий Бумажной Книги и Пользовательской информации (Custom Info) есть свои кнопки добавления в список «+», удаления из списка «-», перемещения выбранного элемента вверх «↑» и вниз «↓» в списке.

2. Вкладка Book (Title Info)

Редактор fb2 свойств - схема fb2.2

☒ Book (Title Info) ☐ FB2 (Document Info) ☐ Paper Book (Publish Info) ☐ Custom Info

Книга (Title Info)

Язык: Russian (ru) Оригинал:

Жанр: Соответствие: 0 %

Название:

Авторы

Имя: Отчество: Фамилия:

Ник: Email: Сайт:

Серии

Название: Номер:

Переводчики

Имя: Отчество: Фамилия:

Ник: Email: Сайт:

Ключевые слова:

Дата (текст): Дата (значение):

Кодировка: utf-8

Жанры:

☐ fb2.1

☒ fb2.2

☐ fb2 LibRuSec

История fb2 файла

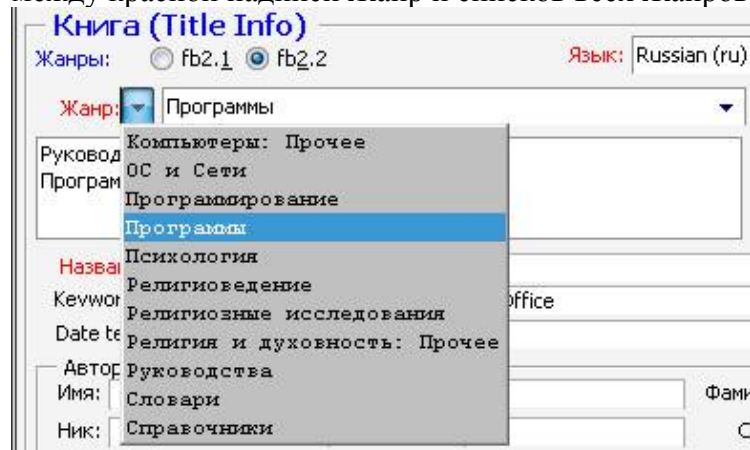
1.0 Распознавание, вычитка и создание fb2-файла

Здесь вводятся основные данные по описанию Книги.

Список **Жанр** служит для выбора жанра конвертируемой книги. Их может быть несколько. Счетчик справа от него – **Соответствие** – 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 **Соответствие** по умолчанию стоит 0. В этом случае этот атрибут не заносится в тэг **genre**. Если же значение счетчика **Соответствие** изменится, и станет больше 0, то это значение будет добавлено, как значение атрибута **match** тэга **genre**.

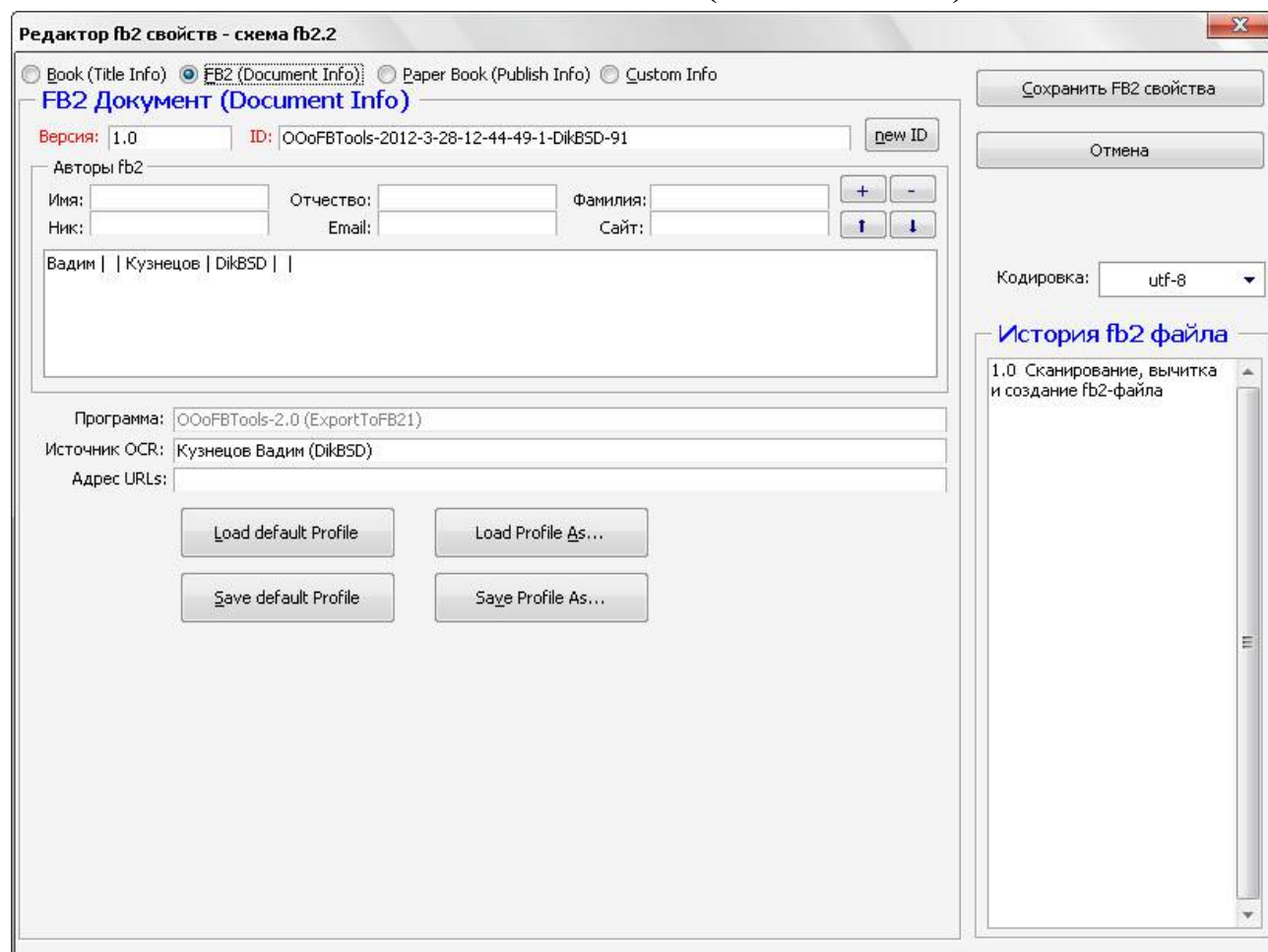
Три переключателя **fb2.1**, **fb2.2** и **fb2 LibRuSec** служат для выбора схемы жанров: FictionBook 2.1, FictionBook 2.2 или FictionBook 2 LibRuSec. В зависимости от выбора схемы жанров меняется и выпадающий список жанров. Это полезно, когда вы готовите книгу для Библиотеки, которая работает со схемой FictionBook 2.1.

Кнопка «**Избранное ...**» служит для вызова диалога создания списка наиболее часто используемых вами Жанров. Этот список отображается в выпадающем списке, который находится между красной надписи Жанр и списков всех Жанров:



Подробности работы с Избранными Жанрами см [здесь](#).

3. Вкладка FB2 (Document Info)



Здесь задаются данные о создателе fb2-файла, версии и ID fb2-файла, и другое.

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

Кнопки **Load default Profile**, **Save default Profile**, **Load Profile As ...**, **Save Profile As ...** служат для загрузки и сохранения готового профиля Автора fb2-файла. Подробнее см. [здесь](#).

4. Вкладка Paper Book (Publish Info)

Редактор fb2 свойств - схема fb2.2

☐ Book (Title Info)
 ☐ FB2 (Document Info)
 ☒ Paper Book (Publish Info)
 ☐ Custom Info

Бумажная Книга (Publish Info)

Название:
 Издатель: Город:
 ISBN: Год:

Серии
 Серия: Номер:

Кодировка:

История fb2 файла

1.0 Сканирование, вычитка и создание fb2-файла

Здесь вводятся данные для Бумажной Книги: ее Название, Издательство, выпустившее эту книгу, Город, где книга была издана. Год издания, ISBN Книги, а также одну или несколько Серий Бумажной Книги (обычно она указывается на обложке ☺).

Кнопки «**Загрузить профиль ...**» и «**Сохранить профиль ...**» служат для создания готового профиля Издательства, сохранения его в текстовый файл и, при необходимости, загрузки этих данных в соответствующие поля **Редактора fb2 свойств**. Это очень удобно – не надо каждый раз вручную вводить одни и те же данные на Издательство. Подробнее см. [здесь](#).

5. Вкладка Custom Info

Здесь можно задавать дополнительную информацию о книге. Например, о редакторе книги, иллюстраторе, авторских правах и т.д. Все, что может быть полезным для описания Книги.

6. Сохранение описания книги в свойства документа

Введенное описание Книги сохраняется в свойства документа. Для этого нажмите кнопку «**Сохранить FB2 свойства**».

Если какое-то из **обязательных** полей (названия выделены **красным** цветом) *не заполнено*, то появится соответствующее предупреждение.

7. Работа с профилями Редактора fb2 свойств

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

7.1. 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 конвертера, в группе «FB2 Документ (Document Info)» есть 4 кнопки для работы с профилем: «**Load default profile**» и «**Save default profile**» – для загрузки и сохранения профиля **по-умолчанию**, и «**Load profile As ...**» и «**Save profile As ...**» – для загрузки и сохранения профиля в произвольный txt-файл.

Профиль Автора fb2-файла по-умолчанию

При загрузке или сохранении профиля автора fb2-файла по умолчанию появится предупреждение и перезаписи данных в текстовом файле профиля или в списке Редактора (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 папку .OOoFBTools, которая находится в [домашней папке пользователя](#).

Произвольный профиль Автора fb2-файла

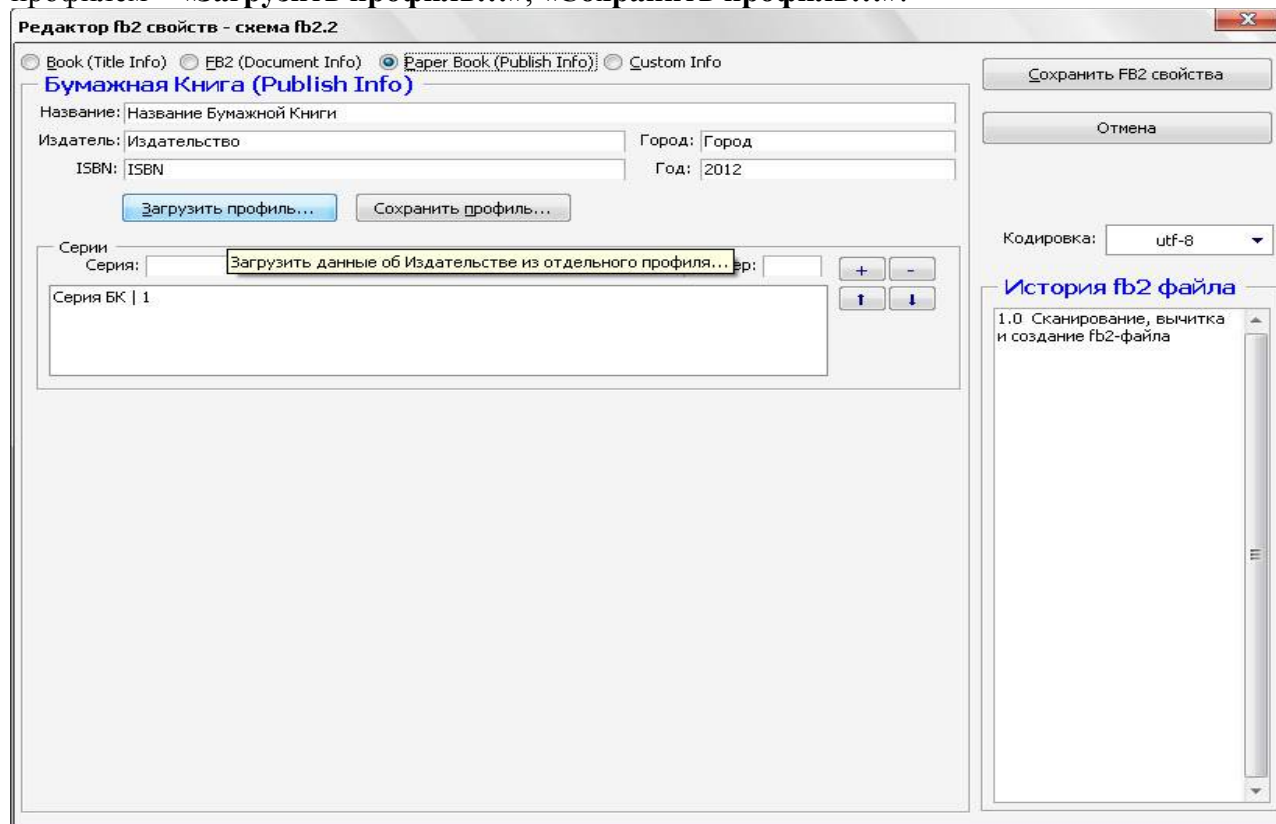
При старте диалога Редактора автоматически загружаются данные автора fb2-файла из свойств документа, если они есть. Если же эти данные не были сохранены с помощью Редактора fb2 свойств, то загружается профиль Автора fb2-файла ПО-УМОЛЧАНИЮ. Его можно изменить и сохранить. Если же надо загрузить (и (или) сохранить) профиль другого Автора fb2-файла, то используйте кнопки " **Load profile As ...**" и " **Save profile As ...**".

7.2. Профиль Издательства

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

На форме Редактора, в группе «Бумажная Книга» (Publish Info)» есть две кнопки для работы с

профилем – «Загрузить профиль...», «Сохранить профиль...»:



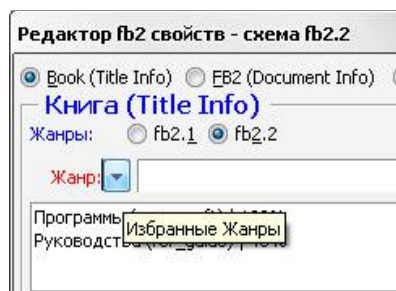
Так же, как и для профиля Автора Книги создайте на жестком диске папку «Издательства» (или как вам нравится). Здесь будут храниться файлы профиля Издательства.

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

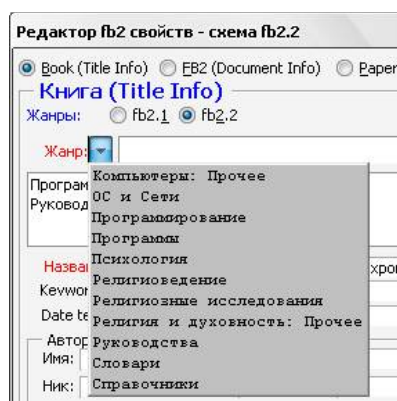
8. Избранные жанры

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



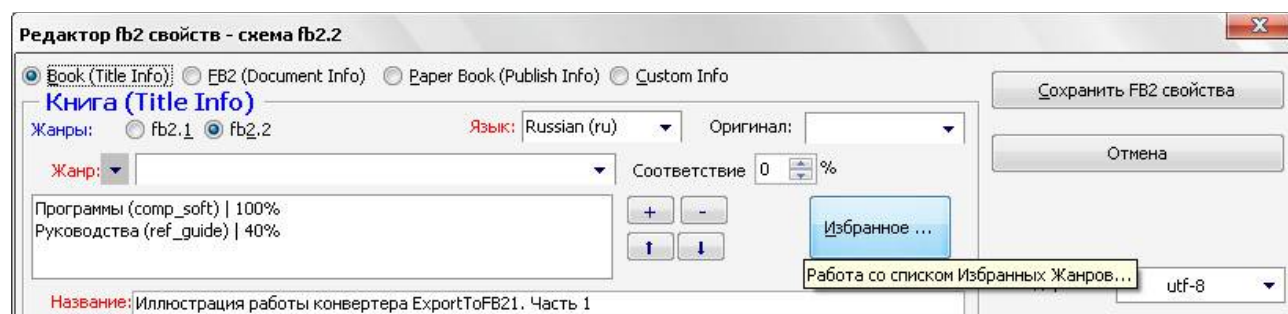
После добавления часто используемых жанров в Избранное, получаем что-то вроде этого:



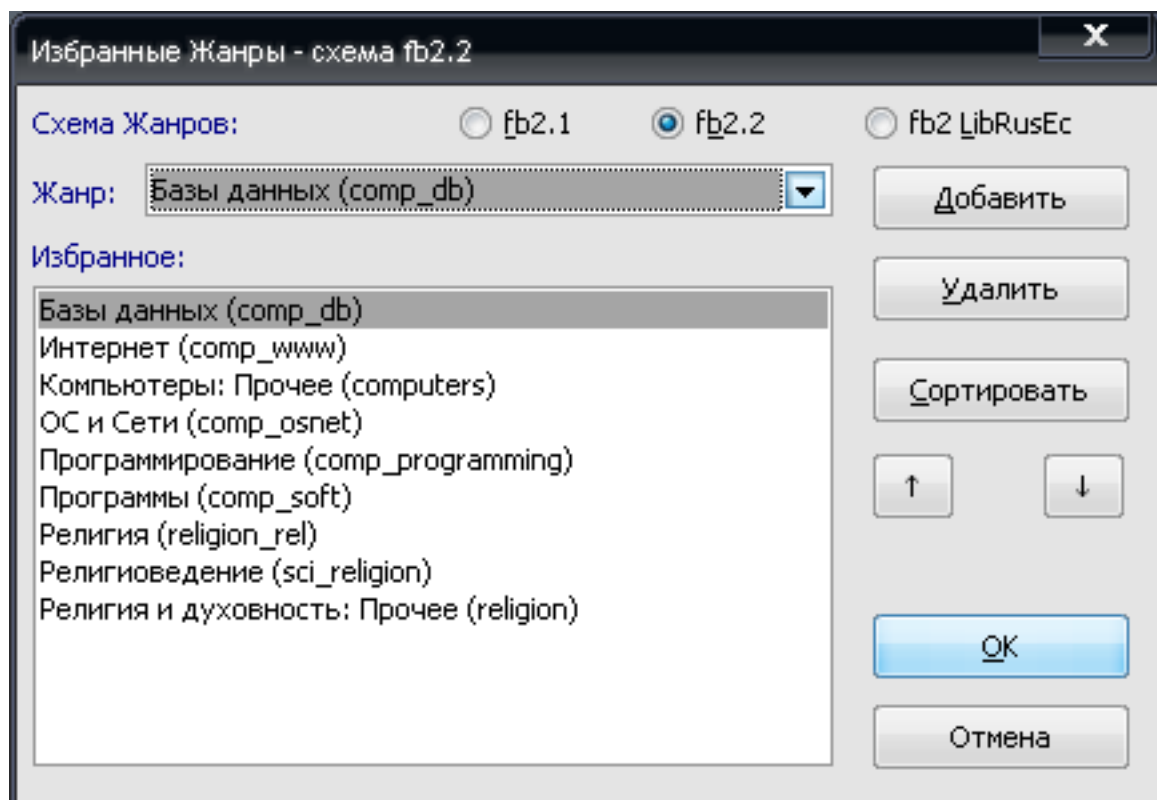
При выборе жанра из Избранного, этот жанр отображается в основном списке жанров. Дальше – работаем, как обычно.

9.1. Создание списка Избранных Жанров

Для создания Избранного Жанров нажмите кнопку «Избранное ...»



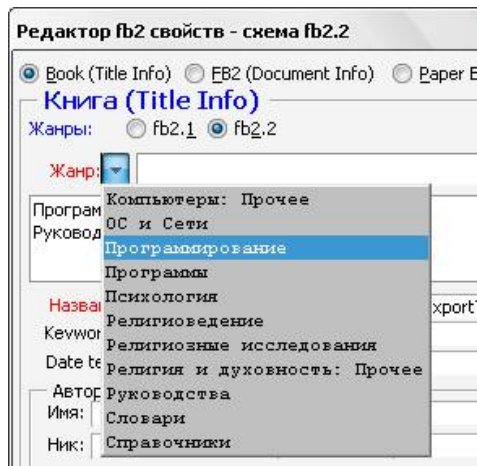
Появится диалог, работа с которым особых комментариев не требует. На картинке уже имеется некоторый список:



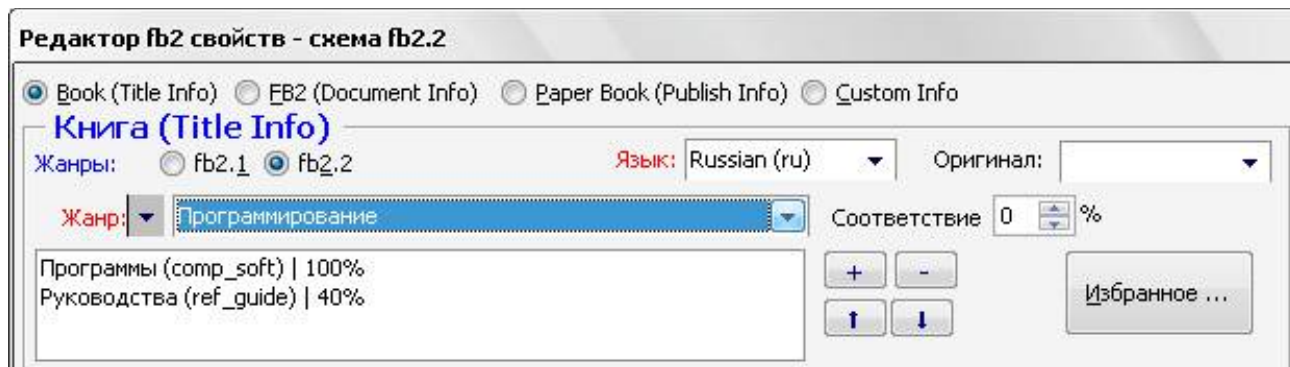
Для удобства можно отсортировать список жанров или переместить нужный жанр выше или ниже в списке...

После создания списка нажатия кнопки ОК список сохраняется в соответствующий файл. Файл Избранных Жанров сохраняются в файл *mygenres.txt*, который помещается в папку .OOOFBTools, которая находится в [домашней папке пользователя](#).

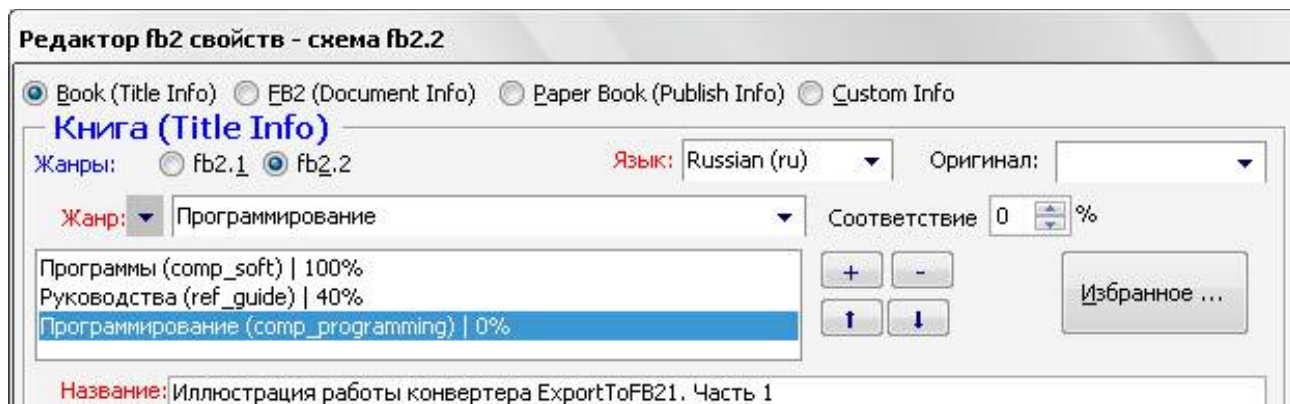
При нажатия на выпадающий список Избранных Жанров на Главной Форме можно выбрать нужный Жанр,



и он будет отображен в основном выпадающем списке жанров:

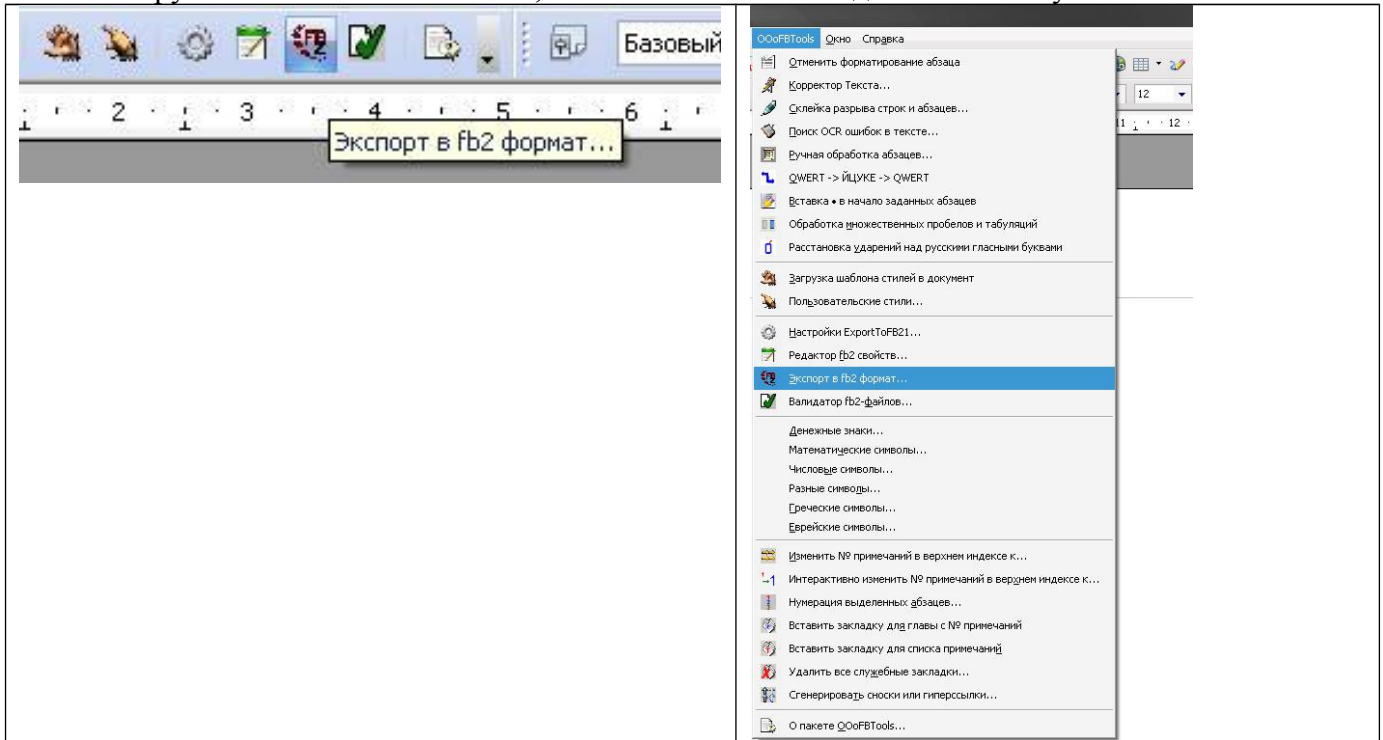


После этого этот Жанр можно будет либо добавить в список жанров экспортируемый книги (кнопка «+»).



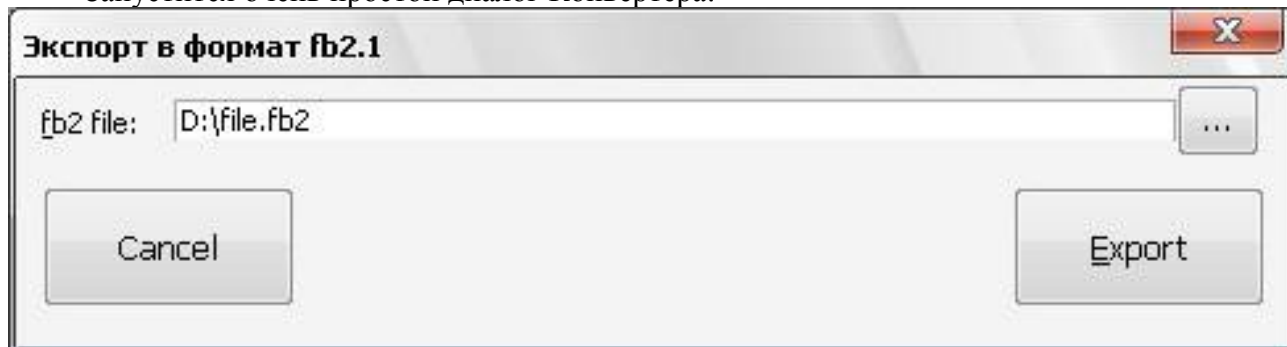
Г. Экспорт в fb2 формат

Для вызова диалога Конвертора можно нажать либо кнопку «Экспорт в fb2 формат...» на панели инструментов **FictionBook Tools**, либо – использовать одноименный пункт меню **OooFBTools**:



Если документ не был ни разу сохранен, то появится соответствующее сообщение с приглашением сохранить документ. Если документ был ранее сохранен, но в нем были сделаны изменения, которые не были сохранены, то появится сообщение с приглашением сохранить документ и перезапустить редактор для корректного экспорта картинок. Впрочем, от этого можно отказаться, нажав кнопку **Ок**.

Запустится очень простой диалог Конвертера:



Кнопка «. . . » служит для изменения местоположения и/или создаваемого fb2-файла. По умолчанию fb2 –файл называется так же, как и открытый экспортируемый документ.

Для конвертации текста достаточно просто нажать кнопку «**Экспорт**» и подождать некоторое время. Процесс экспортирования наглядно отображается в строке состояния редактора.

Когда требуется редактирование результата в fb2–редакторах

1. Когда вы хотите внести в полученный fb2 документ тэги и возможности, пока отсутствующие в конвертере (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 конвертере 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).

Инструменты коррективы текста

I. Автоматическая коррективировка текста

Для коррективировки текста можно использовать либо интерактивные инструменты (см. ниже), либо – автоматические. Рассмотрим их подробнее.

A. Text Corrector - Корректор текста документа

Для удобства работы и уменьшения размеров диалога инструмента все опции Корректора разбиты на несколько вкладок-групп:

Корректор текста документа

☒ **Пробелы, абзацы и разрывы** ☐ Дефисы, тире и другое

Удаление пробелов и/или табуляций

- ☒ Удалить все простые пробелы в начале абзаца
- ☒ Удалить все неразрывные пробелы в начале абзаца
- ☒ Удалить все табуляции в начале абзаца
- ☒ Удалить пробелы внутри скобок, например: (слова) -> (слова)
- ☒ Удалить пробелы внутри кавычек, например: « слова » -> «слова»
- ☒ Удалить пробел перед знаком препинания
- ☒ Заменить множество пробелов между словами на: ☒ Пробел ☐ Табуляцию
- ☒ Обращать и неразрывные пробелы

Вставка пробелов

- ☒ Вставить пробел после знака препинания
- ☒ Вставить пробел между кавычкой и минусом, тире или дефисом
- ☒ Вставить пробел между знаком препинания и минусом, тире или дефисом
- ☒ Вставить неразрывные пробелы между Фамилией и инициалами и наоборот
- ☒ Вставить неразрывный пробел в сокращениях (т.д., т.п., т.к., т.н., т.ч. т.е., н.э., до Р.Х. ...)
- ☒ Вставка неразрывного пробела в датах (19 января ...)

Абзацы и разрывы

- ☐ Удалить разрывы страниц
- ☒ Удалить мягкие переносы
- ☐ Удалить все пустые абзацы ¶
- ☒ Разрыв строки:
 - ☐ Удалить
 - ☒ Заменить на абзац (¶)
 - ☐ Заменить на 1 пробел

Дефисы и переносы, разорванные символом абзаца(¶) и строки

- ☒ Удалить разрывы для дефисов и переносов слов
 - ☐ Обращать только разорванные дефисы
 - ☒ Удаление только разрыва абзаца(¶) и строки(BR); перенос (-) не удаляется
 - ☐ Удаление и разрыва переноса(-) и абзаца(¶) и строки(BR)

Готовые настройки

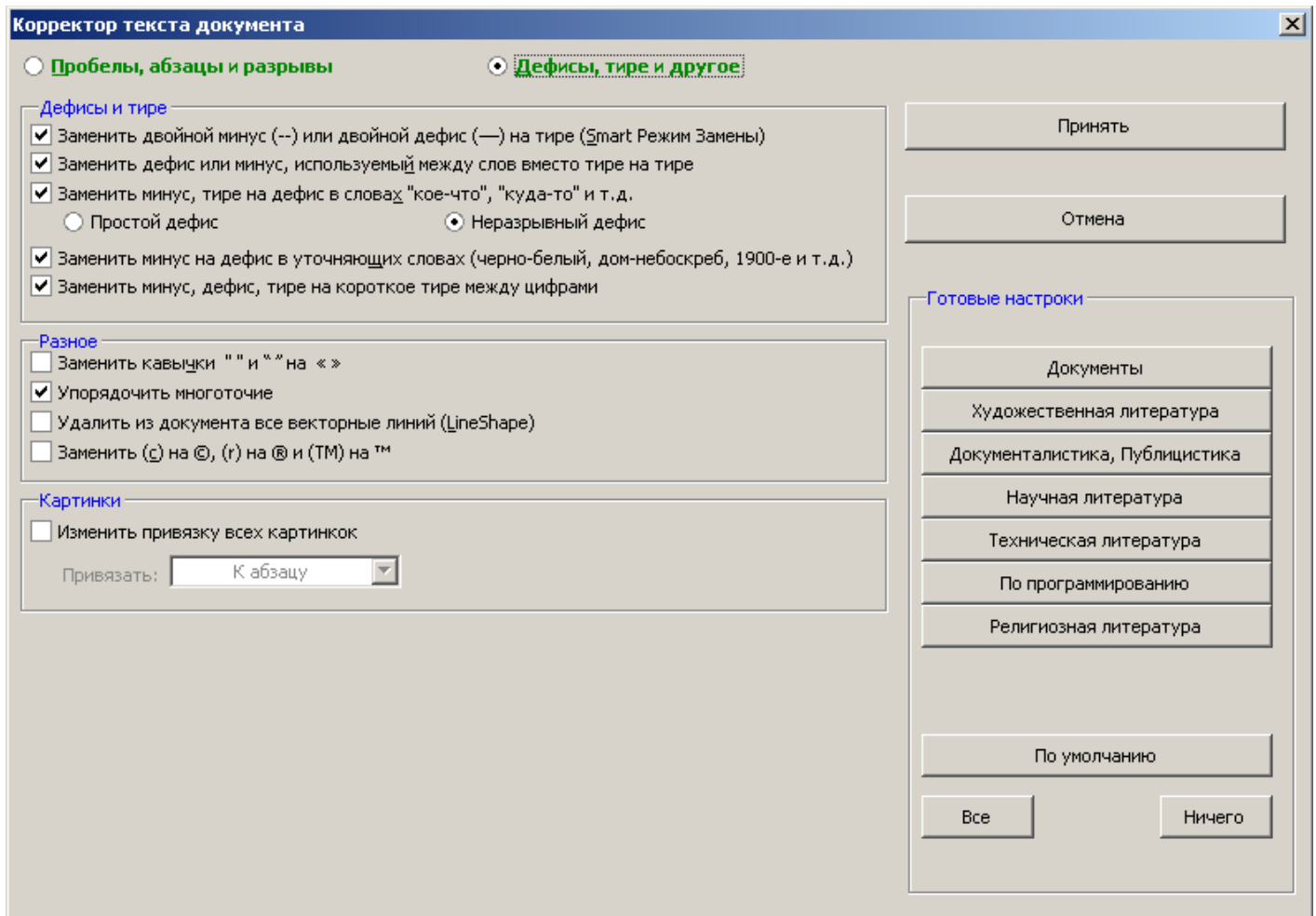
- Документы
- Художественная литература
- Документалистика, Публицистика
- Научная литература
- Техническая литература
- По программированию
- Религиозная литература

Принять

Отмена

По умолчанию

Все Ничего



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

1. Готовые настройки для книг разного вида

В Корректоре предусмотрены наборы готовых настроек для книг разного вида: Документы, Художественные, Религиозные, Научные и т.д. По нажатию соответствующих кнопок одни опции включаются, другие – отключаются. Это позволяет создавать наиболее оптимальную обработку для книг любого вида. Если какая-то опция для определенного вида книг включена или отключена, на ваш взгляд, неверно, то ее легко можно включить/отключить.

2. Описание некоторых опций

Многие опции Корректора Текста интуитивно понятны. Некоторые, требующие пояснения, опишем подробнее.

2.1. 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 режимах 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 вероятностью 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 только разорванные дефисы	2. Дефисы и удаление из слов только символа ¶ без переноса (-)	3. Дефисы и удаление из слов разрыва: и переноса (-) и 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 определиться 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 определиться 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 определиться 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 нужна «ручная доводка».

2.2. Заменить двойной минус (--) или двойной дефис (--) на тире (Smart Режим Замены)

Назначение этой опции понятно из ее названия. Пояснение требует только Smart режим. Обычно двойной минус встречается в книгах, которые были набраны еще в эпоху, когда тире имитировалось двойным минусом. Алгоритм обработки текста по этой опции создан таким образом, что «интеллектуально» подходит к двойным минусам в разных видах книг. Например, почти во всех книгах, кроме книг по программированию двойной минус следует заменять на тире и в прямой речи (в начале абзаца), и в середине предложений. Но в книгах по программированию двойным минусом обозначается либо декремент (например, --i), либо опция командной строки (например, --file). Поэтому рассматриваемая опция Корректора и создана, как Smart – т.е. Корректор отличает с определенной степенью погрешности, где следует заменять двойной минус на тире, а где – нет. Поэтому эту опцию можно использовать и для корректуры книг по программированию – декременты и опции сохраняют «свои» минусы, а в тексте – заменяются на тире.

3. 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 конвертере – 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!

Б. Склейка разорванных строк абзаца

Склейка разорванных строк абзаца

Идентификатор нового абзаца:

- ☒ Пробелы в начале абзаца
- ☐ Отступ в начале абз "Определитель" каждого нового абзаца - 1 или несколько стартовых пробелов
- ☐ Символы .!?... "»"; в конце абзаца
- ☐ "Пустой" абзац или одиночный разрыв строки
- ☐ Разрыв предложений в абзаце на строчных буквах, цифрах, знаках препинания, кавычках и скобках

Обработка разрывов дефисных слов и переносов:

- ☐ Удаление только разрыва абзаца(¶) И строки(BR); перенос (-) не удаляется
- ☒ Удаление И разрыва переноса(-) И разрыва абзаца(¶) И разрыва строки(BR)

Пост-обработка:

- ☒ Удалить все простые пробелы в начале абзаца
- ☒ Удалить все неразрывные пробелы в начале абзаца
- ☒ Удалить все табуляции в начале абзаца
- ☒ Заменить множество пробелов между словами на 1 пробел

Склеить

Отмена

Данный инструмент полезен для обработки многих текстов из Интернет-Библиотек, например, Мошкова, а также – некоторых html-файлов, где есть масса разрывов строк
.

Инструмент ищет признак (идентификатор) нового абзаца по определенным признакам (режимы работы), и все отдельные строки (абзацы), которые оказываются между этими абзацами-маркерами воспринимаются, как части одного разорванного абзаца и склеиваются воедино.

Не зависимо от режима Склейки, Инструмент автоматически удаляет все мягкие переносы и пробелы в конце абзацев. Иначе результат склейки будет неверным.

Рассмотрим режимы работы инструмента – идентификаторы нового абзаца, маркеры.

1. Режимы склейки (по идентификатору нового абзаца)

Пробелы в начале абзаца

Часто «работаешь» с книгами, загруженными из сети, которые после OCR сохранялись, 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... Инструмент позволяет собрать такие строки в «нормальные» абзацы. При этом идентификатором «нового» абзаца является наличие в начале строки одного или нескольких пробелов.

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

```

файл·из·библиотеки·фензина·http://www.fenzin.ru¶
Любишь·фантастику?·Давай·на·фензин!¶
¶
¶
.....Сергей·ЛУКЬЯНЕНКО¶
¶
.....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 Инструментом получаем следующее:

```

|файл·из·библиотеки·фензина·http://www.fenzin.ru¶
|Любишь·фантастику?·Давай·на·фензин!¶
|Сергей·ЛУКЬЯНЕНКО¶
|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 каждого абзаца, и все!

Отступ в начале абзаца (Красная строка)

Часто встречается текст, в котором строки абзаца разорваны символом абзаца, но каждый новый абзац определен отступом (Красная строка). Такой текст можно получить из некоторых pdf файлов.

Например, есть такой текст (из интернета):

```

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 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¶

```

После обработки получаем 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 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 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¶

Символы «!?...»”:; в конце абзаца»

Не менее часто встречаются книги, скаченные из сети, у которых нет Красной строки, обозначенной пробелами или отступом, 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, ¶ 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 сцепарика 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 Инструментом Склейки 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 сцепарика 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.

“Пустой” абзац или одиночный разрыв строки

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. При выборе данного режима обработки, программа ищет блок текста (как правило разорванные строки одного абзаца) по идентификаторам – «пустой» абзац (одиночный разрыв строки
). 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 Инструментом Склейки получаем следующее:

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 Инструмент Склейки 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!

2. Обработка разрывов дефисных слов и переносов

В тексте могут встречаться разрывы на дефисных словах (кое-кто, кого-то и т.д.), а так же на переносах слов. Инструмент Склейки абзацев обрабатывает такие случаи разрывов, исходя из 2-х вариантов:

1. Удаление только разрыва абзаца (¶) и строки (BR); перенос (-) не удаляется
2. Удаление и разрыва переноса (-) и абзаца (¶) и строки (BR)

По сути, это та же обработка разрывов, что и в Корректоре Текста.

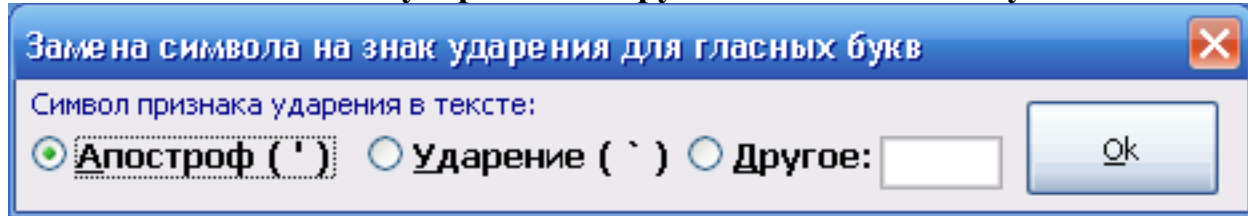
Инструмент будет обрабатывать такие разрывы в любом случае, иначе он не сможет корректно склеить все строки, принадлежащие к одному абзацу.

3. Пост-обработка

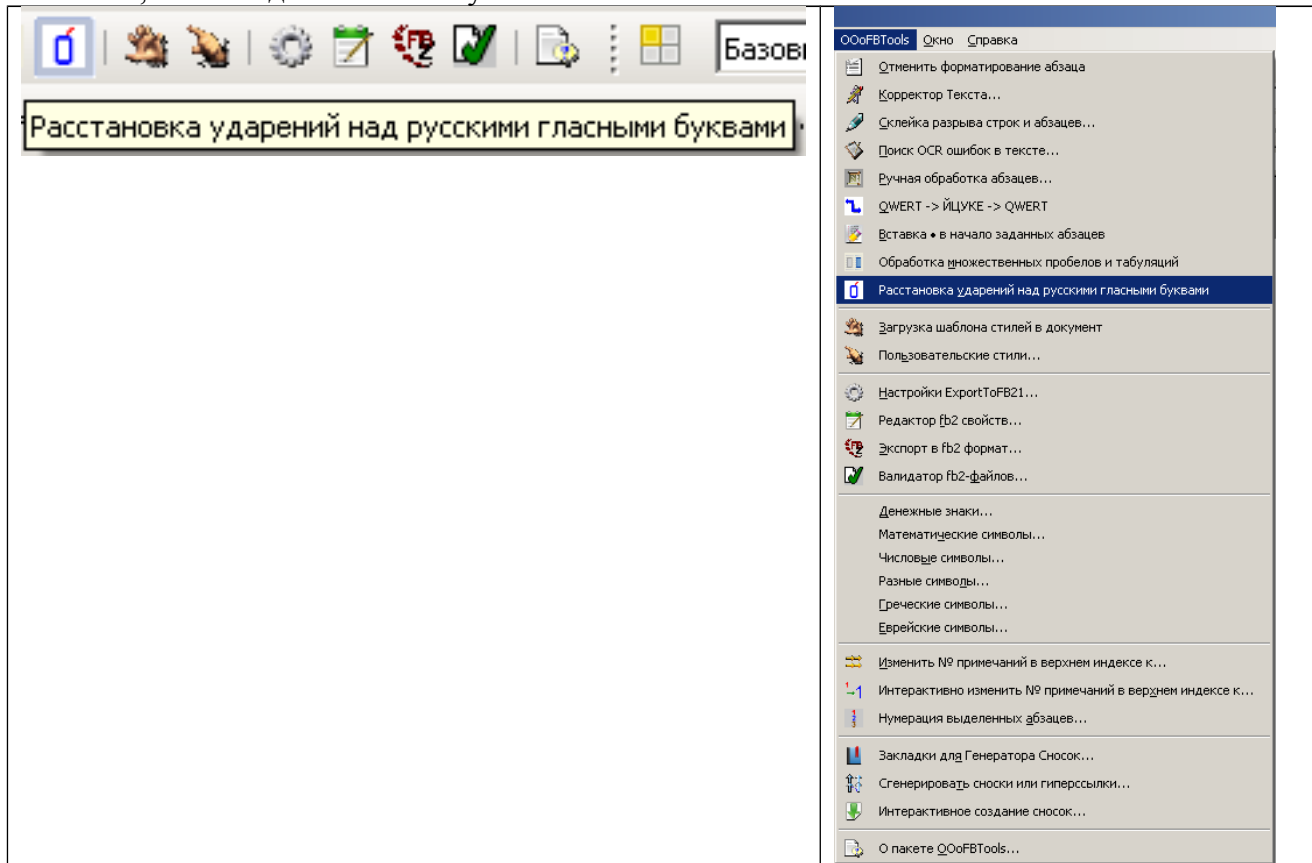
После склейки разорванных абзацев в тексте могут остаться множественные пробелы и табуляции, а также пробелы в начале абзаца. При желании вы можете включить опции для их удаления.

Удалении пробелов в конце абзаца, если они есть, происходит автоматически.

В. Расстановка ударений над русскими гласными буквами



Для запуска диалога Расстановки ударений над русскими гласными буквами нужно нажать либо кнопку «**Расстановка ударений над русскими гласными буквами**» панели инструментов OooFBTools, либо – одноименный пункт меню OooFBTools:



Назначение этого инструмента понятно из его названия.

В диалоге Инструмента есть выбор символа – признака ударения в тексте. Это позволяет более гибко задавать признак ударения в документе, когда встречаются такие слова, как "О'Нил".

Варианты выбора символа – признака ударения:

1. Апостроф (')
2. Ударение (`) и
3. Другое.

В 3-м случае в поле можно ввести символ(ы), но важно помнить, что он (они) не должны совпадать с символами регулярных выражений: \, |, и т.д. В 3-м случае допускается вводить несколько символов.

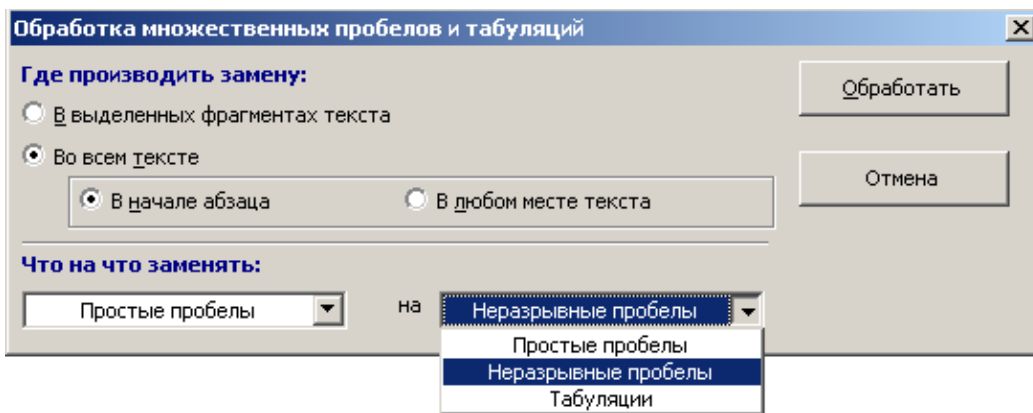
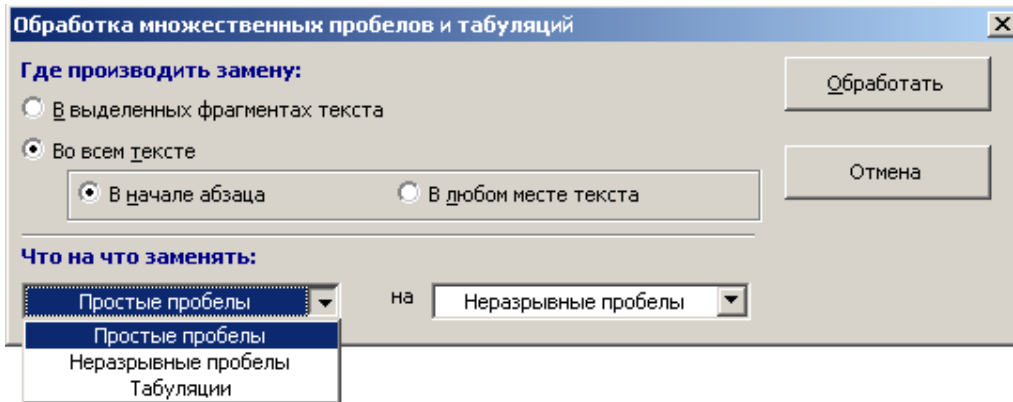
Как происходит обработка текста? Если вам надо поставить над гласной буквой русского языка ударение, то справа от этой гласной поставьте специальный символ-признак ударения, например, апостроф ('). Инструмент ищет такое сочетание – гласная и рядом – апостроф, и проставляет ударение.

Например: Надо проставить ударение над буквой «а» в слове «написать». Ставим апостроф справа от «а»: «написа'ть». Запускаем инструмент – получаем: «написа́ть».

Важно: Будьте внимательны – если в редакторе включена или введена автозамена для апострофов, то вместо апострофа (') можно получить, например, символ кавычки-лапки (‘). В 1-м Режиме работы «Апостроф (')» Инструмент распознает *только* апостроф!

Важно: Часто в тексте апостроф служит для создания слов: «Д’Арк», «Дэвид О’Нил» и т.д. Тогда лучше всего для задания символа признака ударения использовать не апостроф. А знак ударения (`), а в диалоге Инструмента – режим 2: «Ударение (`)». Можно еще использовать какой-нибудь иной символ, или несколько символов, а в Инструменте – Режим 3 «Другое», задав в поле этот (эти) символ (символы).

Г. Обработка множественных пробелов и табуляций



Этот инструмент служит для поиска либо множественных простых пробелов, либо множественных неразрывных пробелов, либо множественных табуляций в различных местах текста и замены их на равное число либо простых пробелов, либо неразрывных пробелов, либо табуляций.

Такая обработка текста очень полезна для книг и документов, где для смыслового и/или функционального восприятия требуется сохранить структуру некоторых абзацев. В форматах doc, odt и ряде других этого можно добиться выставляя отступы абзацев с помощью горизонтальной линейки текста. Но в таких форматах, как fb2 это не сработает, т.к. этот формат вообще не предусматривает сохранение структур текста. Поэтому данный инструмент здесь очень пригодится.

Инструментом можно обрабатывать научные, научно-религиозные книги, книги по программированию...

Инструмент работает в 2-х режимах: 1. В Выделенных фрагментах текста. 2. Во всем тексте. Обработка не производится во Врезках, Таблицах и Сносках!

Рассмотрим оба режима.

1. Обработка текста в выделенных фрагментах

Диалог инструмента – немодальный, поэтому можно легко переходить из диалога в текст для выделения фрагментов и обратно.

Зона может захватывать и слова, и единичные пробелы между слов, и целые абзацы. Главное – чтобы в выделении оказались множественные табуляции пробелы, которые необходимо заменить на неразрывные.

Важное условие: выделение должно либо захватывать с правой стороны хоть один символ слова, либо хотя бы «прикасаться» к ближайшему слову. Т.е. между выделением и ближайшим словом не должно быть ни одного пробела и/или табуляции! Иначе инструмент ничего не заменит.

Если надо заменить, например, множественные пробелы на такое же число неразрывных в некоторых фрагментах абзацев, то проще всего выделить эти фрагменты и запустить обработку в режиме «В выделенных фрагментах текста». Например, есть такой текст книги про FreeBSD:

В итоге к рубежу 90-х годов выяснилось, что исходного (проприетарного) UNIX-кода в составе берклианских UNIX осталось не так уж и много. И родилась идея создания полностью открытой операционной системы, распространяемой свободно и в исходных текстах. К этому же времени прекратилось (не иначе как вследствие распада мировой системы социализма) финансирование группы CSRG, и она столь же благополучно распалась — все в жизни имеет свою оборотную сторону, даже крах коммунистической идеологии. Но дело ее не пропало. Его подхватили, расширили, укрепили и закалили в боях многие из бывших членов CSRG.¶

Поскольку здесь начинается уже собственно история FreeBSD, пора вспомнить и о персоналиях. Именно созданием общедоступной UNIX-системы, причем — на общедоступной же платформе, сиречь 1x86, озаботились Вильям и Линна Джолитц. Базируясь на одном из побегов ветви 3BSD — BSD Net/2, они дописали недостающие компоненты и создали 386BSD — первую BSD-систему, претендовавшую на звание открытой в собственном смысле слова. И, насколько я понимаю, первый берклианский побег, портированный на PC (IBM-совместимые компьютеры, как их тогда еще задумчиво называли).¶

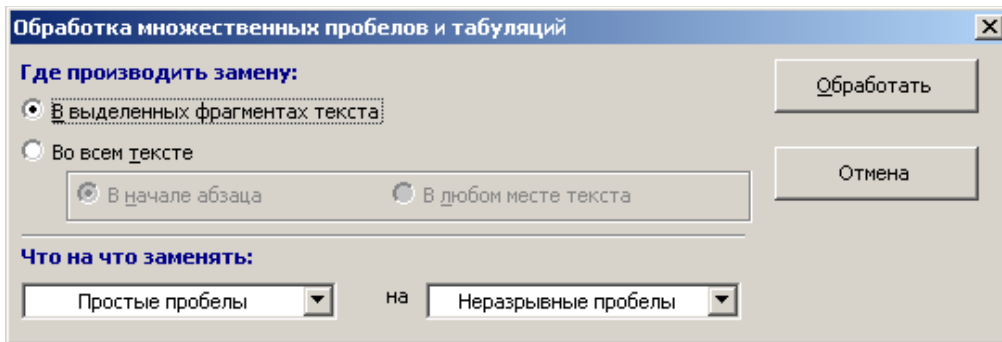
Нам необходимо заменить на неразрывные пробелы в 1-м абзаце 5 пробелов до слово UNIX и 5 пробелов после, и во 2-м абзаце – 5 пробелов до слова FreeBSD и 5 – после. Для этого выделяем эти места мышкой, удерживая клавишу Ctrl:

В итоге к рубежу 90-х годов выяснилось, что исходного (проприетарного) UNIX-кода в составе берклианских UNIX осталось не так уж и много. И родилась идея создания полностью открытой операционной системы, распространяемой свободно и в исходных текстах. К этому же времени прекратилось (не иначе как вследствие распада мировой системы социализма) финансирование группы CSRG, и она столь же благополучно распалась — все в жизни имеет свою оборотную сторону, даже крах коммунистической идеологии. Но дело ее не пропало. Его подхватили, расширили, укрепили и закалили в боях многие из бывших членов CSRG.¶

Поскольку здесь начинается уже собственно история FreeBSD, пора вспомнить и о персоналиях. Именно созданием общедоступной UNIX-системы, причем — на общедоступной же платформе, сиречь 1x86, озаботились Вильям и Линна Джолитц. Базируясь на одном из побегов ветви 3BSD — BSD Net/2, они дописали недостающие компоненты и создали 386BSD — первую BSD-систему, претендовавшую на звание открытой в собственном смысле слова. И, насколько я понимаю, первый берклианский побег, портированный на PC (IBM-совместимые компьютеры, как их тогда еще задумчиво называли).¶

На картинке выделение захватывает несколько букв слева и справа от множественных пробелов. Можно было выделить и весь абзац.

Теперь запускаем инструмент обработки пробелов и табуляций, и задаем Режим – В выделенных фрагментах текста, Искать – простые пробелы, Заменить – на неразрывные пробелы:



Нажимаем кнопку Обработать, получаем следующее:

В итоге к рубежу 90-х годов выяснилось, что исходного (проприетарного) UNIX-кода в составе берклианских UNIX осталось не так уж и много. И родилась идея создания полностью открытой операционной системы, распространяемой свободно и в исходных текстах. К этому же времени прекратилось (не иначе как вследствие распада мировой системы социализма) финансирование группы CSRG, и она столь же благополучно распалась — все в жизни имеет свою оборотную сторону, даже крах коммунистической идеологии. Но дело ее не пропало. Его подхватили, расширили, укрепили и закалили в боях многие из бывших членов CSRG.¶

Поскольку здесь начинается уже собственно история FreeBSD, пора вспомнить и о персоналиях. Именно созданием общедоступной UNIX-системы, причем — на общедоступной же платформе, сиречь 1x86, озаботились Вильям и Линна Джолитц. Базируясь на одном из побегов ветви 3BSD — BSD Net/2, они дописали недостающие компоненты и создали 386BSD — первую BSD-систему, претендовавшую на звание открытой в собственном смысле слова. И, насколько я понимаю, первый берклианский побег, портированный на PC (IBM-совместимые компьютеры, как их тогда еще задумчиво называли).¶

Что и требовалось...

2. Обработка во всем тексте

В этом режиме инструмент может обрабатывать текст либо в начале абзацев, либо в любом месте текста.

Работу инструмента в этом режиме лучше всего проиллюстрировать на тексте с программным листингом (включено отображение непечатных символов, чтобы увидеть символы табуляции):

Фрагмент листинга программы:

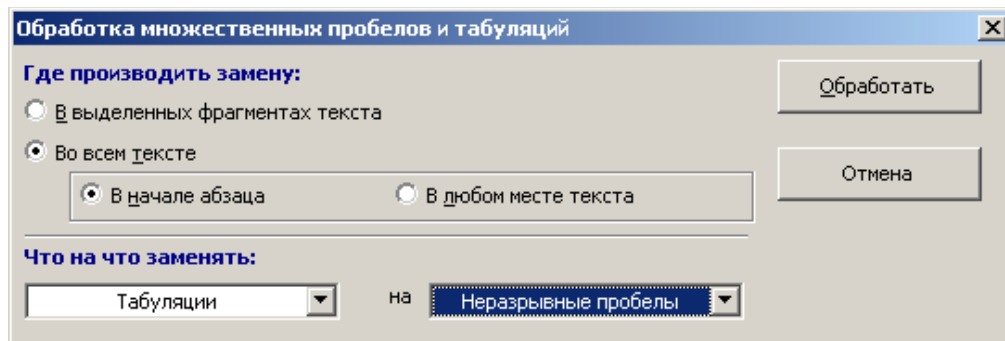
```

public static List<string> MakeFileListFromDir(string sFromDir, bool bSort,
bool bFB2Only) {
    List<string> lFilesList = null;
    if (Directory.Exists(sFromDir)) {
        string[] files = Directory.GetFiles(sFromDir);
        if (files.Length != 0) {
            lFilesList = new List<string>();
            foreach (string sFile in files) {
                if (bFB2Only) {
                    if (Path.GetExtension(Path.GetFileName(sFile)).ToLower() == ".fb2") {
                        lFilesList.Add(sFile);
                    }
                } else {
                    lFilesList.Add(sFile);
                }
            }
            if (bSort) {
                lFilesList.Sort();
            }
        }
    }
    return lFilesList;
}

```

Надо заменить все стартовые табуляции в листинге на такое же число неразрывных пробелов. Выделять вручную фрагменты с табуляциями в каждом абзаце – нерационально. Проще использовать 2-й режим обработки инструмента.

Запускаем инструмент в режиме «Во всем тексте». Выбираем искать Табуляции и Заменить их на Неразрывные пробелы. Нажимаем кнопку Обработать:



Получаем, что требуется:

Фрагмент листинга программы:

```

public static List<string> MakeFileListFromDir(·string·sFromDir,·bool·bSort,
bool·bFB2Only)·{
    List<string>·lFilesList·==·null;
    if(·Directory.Exists(·sFromDir·)·)·{
        string[]·files·==·Directory.GetFiles(·sFromDir·);
        if(·files.Length·!=·0·)·{
            lFilesList·==·new·List<string>();
            foreach(·string·sFile·in·files·)·{
                if(·bFB2Only·)·{
                    if(·Path.GetExtension(·Path.GetFileName(·sFile·)·).ToLower()·==·".fb2"·)·{
                        lFilesList.Add(·sFile·);
                    }
                }·else·{
                    lFilesList.Add(·sFile·);
                }
            }
        }
        if(·bSort·)·{
            lFilesList.Sort();
        }
    }
    return·lFilesList;
}

```

Если же выбрать Обработку не в «Начале абзаца», а «В любом месте текста», то результат будет несколько иной, т.к. при обработке «Начале абзаца» обрабатываются **все** стартовые пробелы или табуляции, в режиме «В любом месте текста» - только множественные:

Фрагмент листинга программы:

```

public static List<string> MakeFileListFromDir(·string·sFromDir,·bool·bSort,
bool·bFB2Only)·{
    List<string>·lFilesList·==·null;
    if(·Directory.Exists(·sFromDir·)·)·{
        string[]·files·==·Directory.GetFiles(·sFromDir·);
        if(·files.Length·!=·0·)·{
            lFilesList·==·new·List<string>();
            foreach(·string·sFile·in·files·)·{
                if(·bFB2Only·)·{
                    if(·Path.GetExtension(·Path.GetFileName(·sFile·)·).ToLower()·==·".fb2"·)·{
                        lFilesList.Add(·sFile·);
                    }
                }·else·{
                    lFilesList.Add(·sFile·);
                }
            }
        }
        if(·bSort·)·{
            lFilesList.Sort();
        }
    }
    return·lFilesList;
}

```

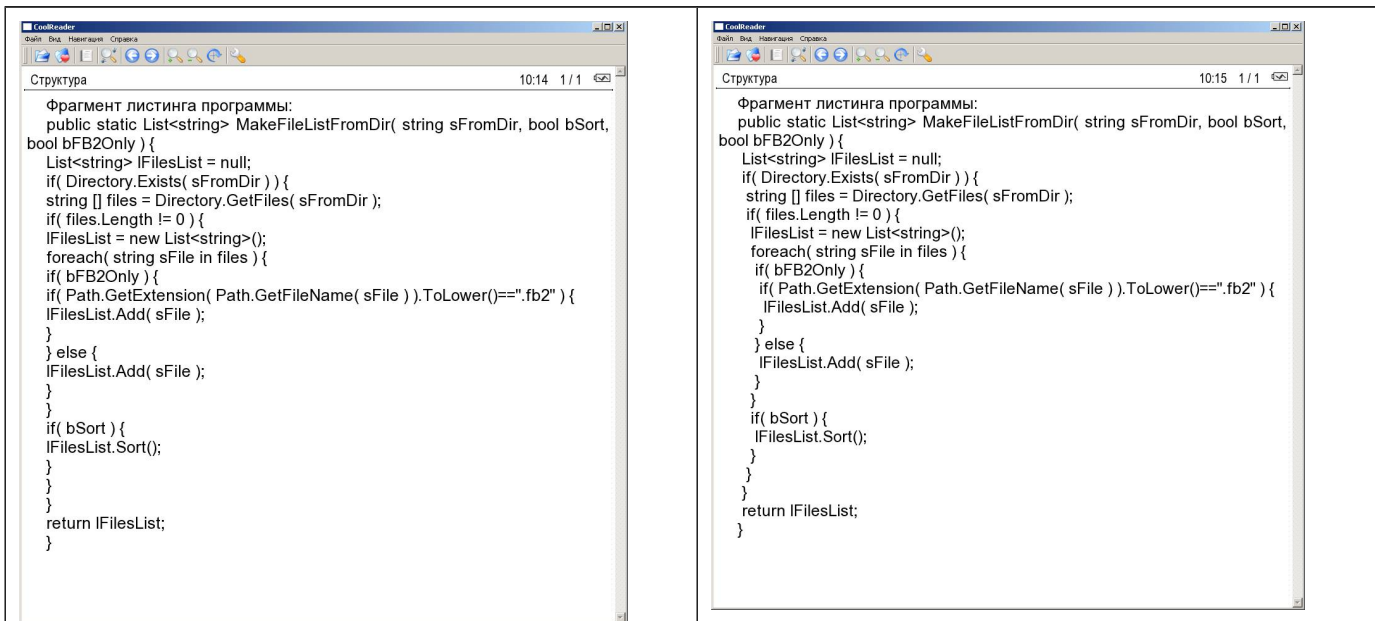
Т.е., в 4, 5 строках сверху и в 2,3 строках снизу табуляции остались не замененными на неразрывные пробелы, и, следовательно, в экспортируемом fb2-файле они «нарушат» структуру листинга, т.к. читалки игнорируют стартовые табуляции и пробелы, кроме неразрывных.

Давайте посмотрим на то, как отображают читалки fb2-файл этого листинга до замены табуляций на неразрывные пробелы и после.

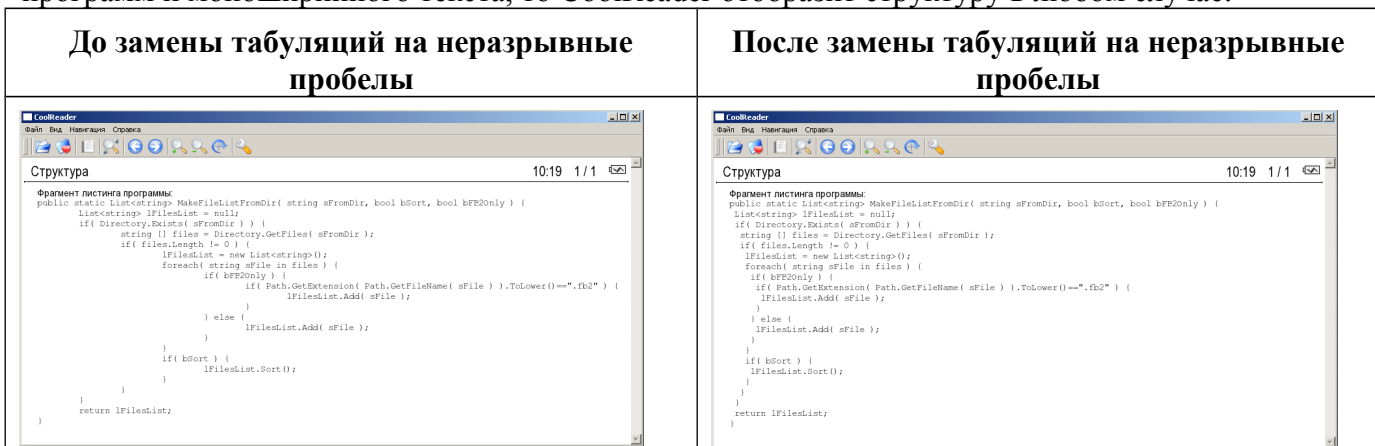
В CoolReader 3

До замены табуляций на неразрывные пробелы

После замены табуляций на неразрывные пробелы

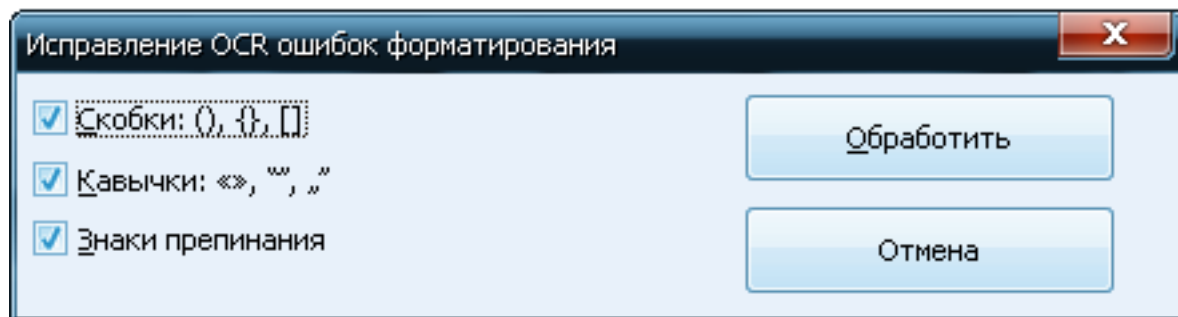


Но, если отформатировать листинг программы OOoFBTools-стилем code (см. подробнее о стилях в соответствующем разделе справки), который используется для оформления текстов программ и моноширинного текста, то CoolReader отобразит структуру в любом случае:



Это – особенность CoolReader. Другие читалки могут работать по-иному, например вообще не отображать множественные простые пробелы или табуляции в начале абзацев, независимо, оформлен ли текст структуры fb-2 тэгом <code> или нет. Поэтому заменять множественные табуляции и/или пробелы на неразрывные полезно в любом случае. Тем более, что в религиозных книгах структуры не форматируют OOoFBTools-стилем code.

Д. Обработка OCR ошибок форматирования

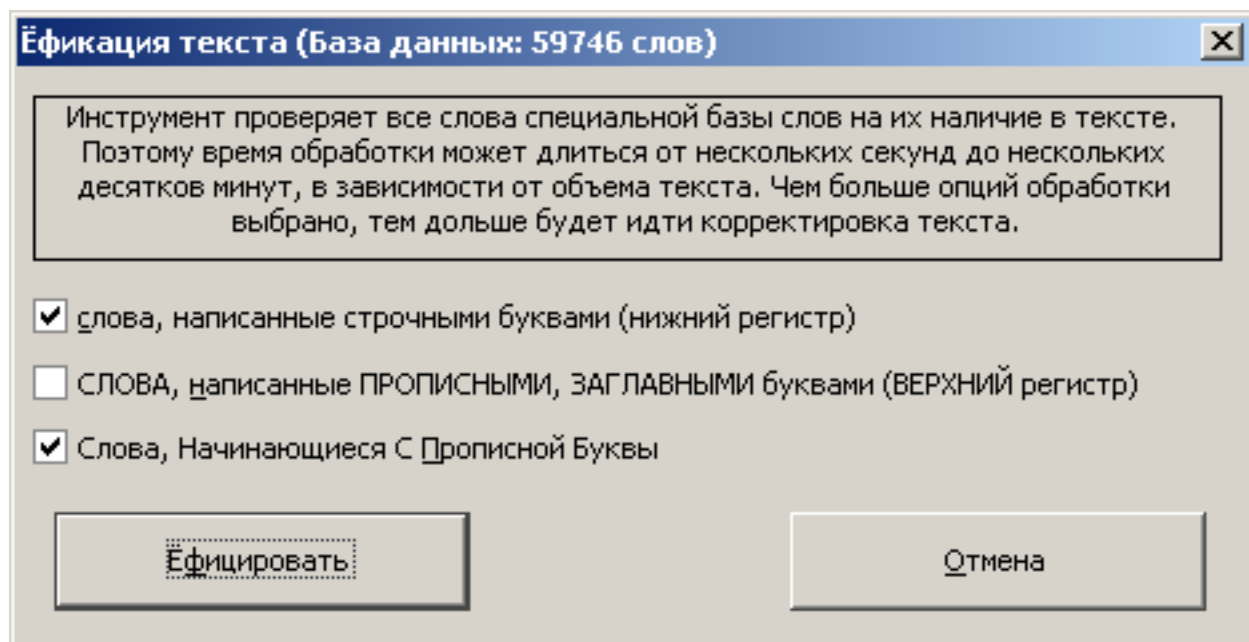


Часто после OCR распознавания текста, если в кавычках и/или скобках есть текст с курсивом, то сами кавычки/скобки распознаются тоже с курсивом, хотя в бумажной книге они – без курсива. Также при распознавании очень часто неверно форматируются курсивом знаки препинания, стоящие после слов/кавычек/скобок с курсивом.

Данный инструмент ищет по тексту документа (сноски, таблицы и текстовые врезки не обрабатываются) вышеописанные OCR ошибки форматирования и снимает курсив со скобок, кавычек и/или знаков препинания. Используется интеллектуальный алгоритм определения, где курсив должен остаться, а где – это явная OCR ошибка.

Инструмент позволяет опционально выбирать объекты исправления.

Е. Ёфикация русского текста



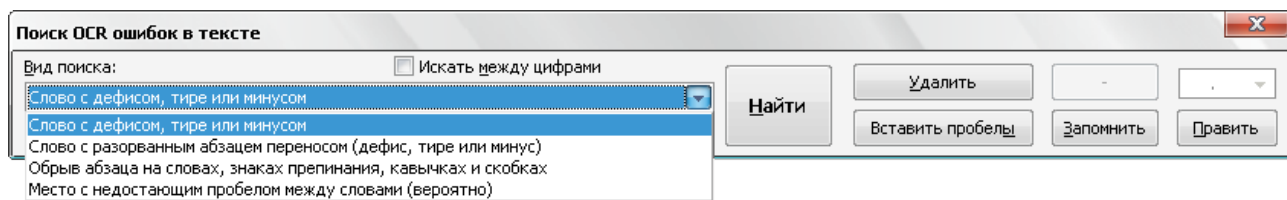
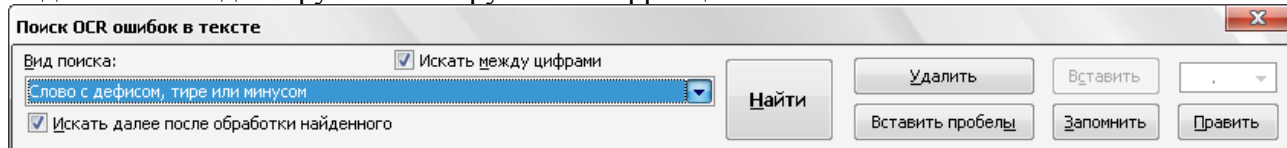
Инструмент предназначен для поиска и замены в автоматическом режиме слов, в которых вместо буквы «ё» («Ё») стоит буква «е» (Е). Обрабатывается только русский текст (оно и понятно). Используется довольно большая база слов, поэтому на обработку текста может потребоваться много времени.

Важное замечание: Чем больше опция включено, тем дольше будет идти обработка текста.

II. Интерактивные инструменты обработки текста

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 папку .OOOFBTools, которая находится в [домашней папке 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 минусе, тире и дефисах, слева от которых стоит 1 или более пробелов, а справа – ни один или множество пробелов (частая ошибка OCR).

Кнопка «Обработать» позволяем 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 «по¶шел». 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.

2. Обработка выделенных абзацев (обрыв абзаца или строки)

The image displays three sequential screenshots of a software dialog box titled "Ручная обработка абзацев" (Manual paragraph processing). Each window has a close button (X) in the top right corner and an "Обработать" (Process) button on the right.

- First screenshot:** The "Режим обработки:" (Processing mode) dropdown is set to "Обрыв абзаца (¶)" (Paragraph break). The "Тип обработки:" (Processing type) dropdown is also set to "Обрыв абзаца (¶)".
- Second screenshot:** The "Режим обработки:" dropdown is set to "Обрыв абзаца (¶)". The "Тип обработки:" dropdown is set to "Склейка в 1 абзац (Замена обрыв абзаца ¶ на 1 пробел)" (Join into 1 paragraph (Replace paragraph break ¶ with 1 space)).
- Third screenshot:** The "Режим обработки:" dropdown is set to "Разрыв строки (↵)" (Line break). The "Тип обработки:" dropdown is set to "Склейка в 1 абзац (Замена обрыва строки ↵ на 1 пробел)" (Join into 1 paragraph (Replace line break ↵ with 1 space)).

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 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.¶
¶
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 (¶)** и Тип обработки: **Склейка в 1 абзац (Замена обрыва строки ¶ 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 ¶

¶

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 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¶

¶

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. ¶

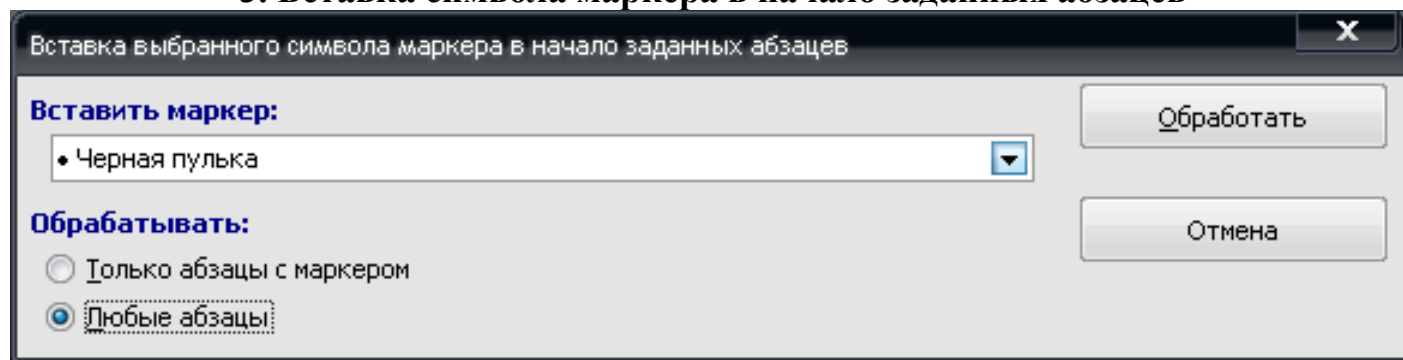
¶

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.

3. Вставка символа маркера в начало заданных абзацев



Данный инструмент позволяет делать следующее:

1. Вставлять в заданные абзацы текста выбранного из обширного выпадающего списка символ маркера.
2. Обрабатывать либо выделенные абзацы, либо весь документ.
3. Обрабатывать либо только маркированные абзацы, либо любые не маркированные абзацы, либо – и те и другие. Когда инструмент находит маркированный абзац, он удаляет из него маркер и его признак.

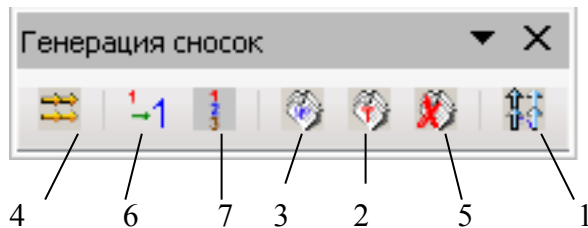
Очень часто после OCR, некоторые виды маркеров в тексте экспортируются в fb2-файл не корректно, что проявляется их «кривым» отображением в читалках. Данный инструмент решает эту проблему.

При вставке Юникодных символов маркера (со 2-го и далее в выпадающем списке) следует учитывать следующее: Экспорт текста с маркерами требует кодировки UTF-8, что влечет за собой увеличение размера результирующего fb2-файла. Используя данный инструмент можно заменить все маркеры на символ маркера, что позволит делать экспорт текста в кодировке Windows-1251 для уменьшения размера fb2-файла. Конечно, при условии, что в тексте отсутствуют Юникодные символы.

4. Инструменты работы с примечаниями, сносками и гиперссылками

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 сноски, 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 сносок или гиперссылок** (иконки со временем могут измениться):



1. Генерация Сносок или Гиперссылок на примечания.
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 или гиперссылок** 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 или гиперссылок** и **Нумерация выделенных абзацев** работают только с текстом документа. Текстовые Врезки – игнорируются, т.к. OoWrither не позволяет в них вставлять сноски. Таблицы тоже игнорируются (из-за сложности проверки выхода курсора за пределы Таблицы и некорректности получаемого результата).

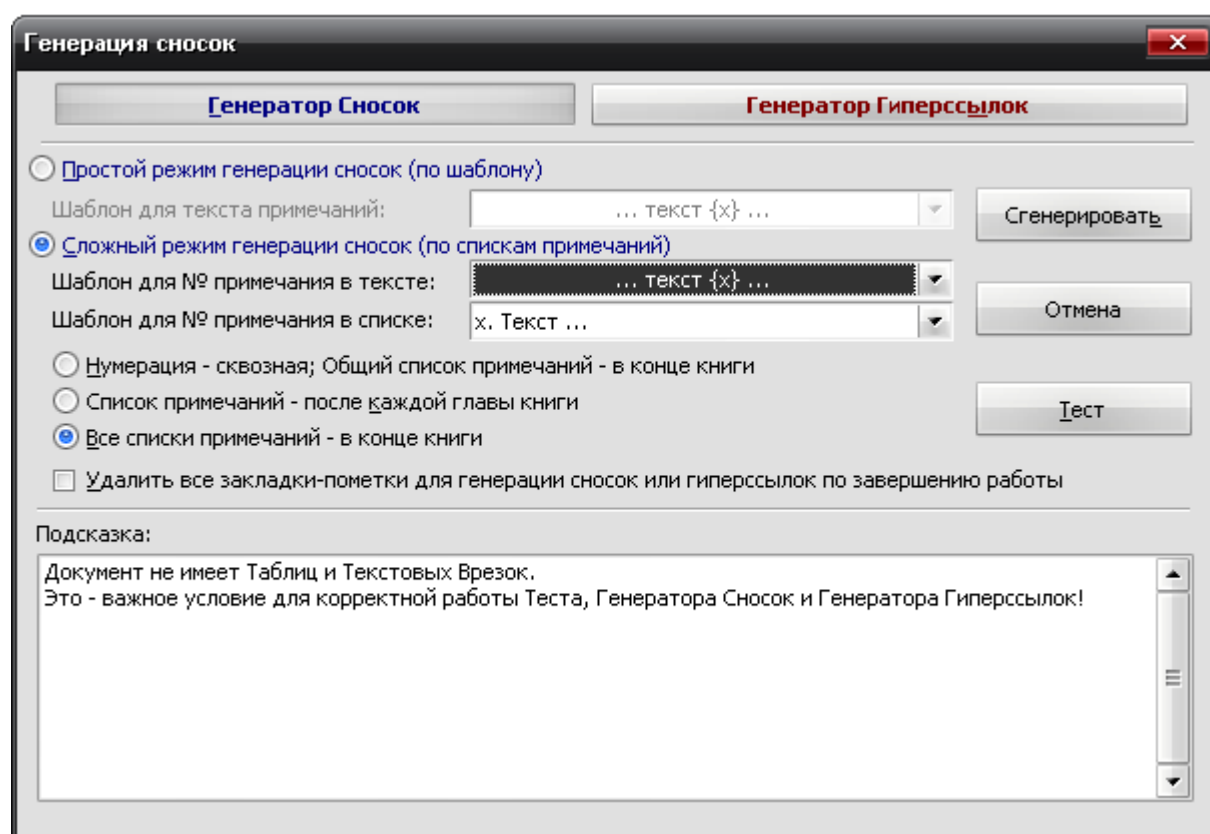
Теперь – подробнее о каждом инструменте.

4.1. Генерация сносок или гиперссылок

Механизм работы Генератора Сносок и Генератора Гиперссылок одинаков. Различие состоит в том, что Генератор сносок *перемещает* текст примечания в сгенерированную сноску, а Генератор гиперссылок формирует из соответствующего места книги гиперссылку на нужное примечание, ничего не делая с самим текстом примечания.

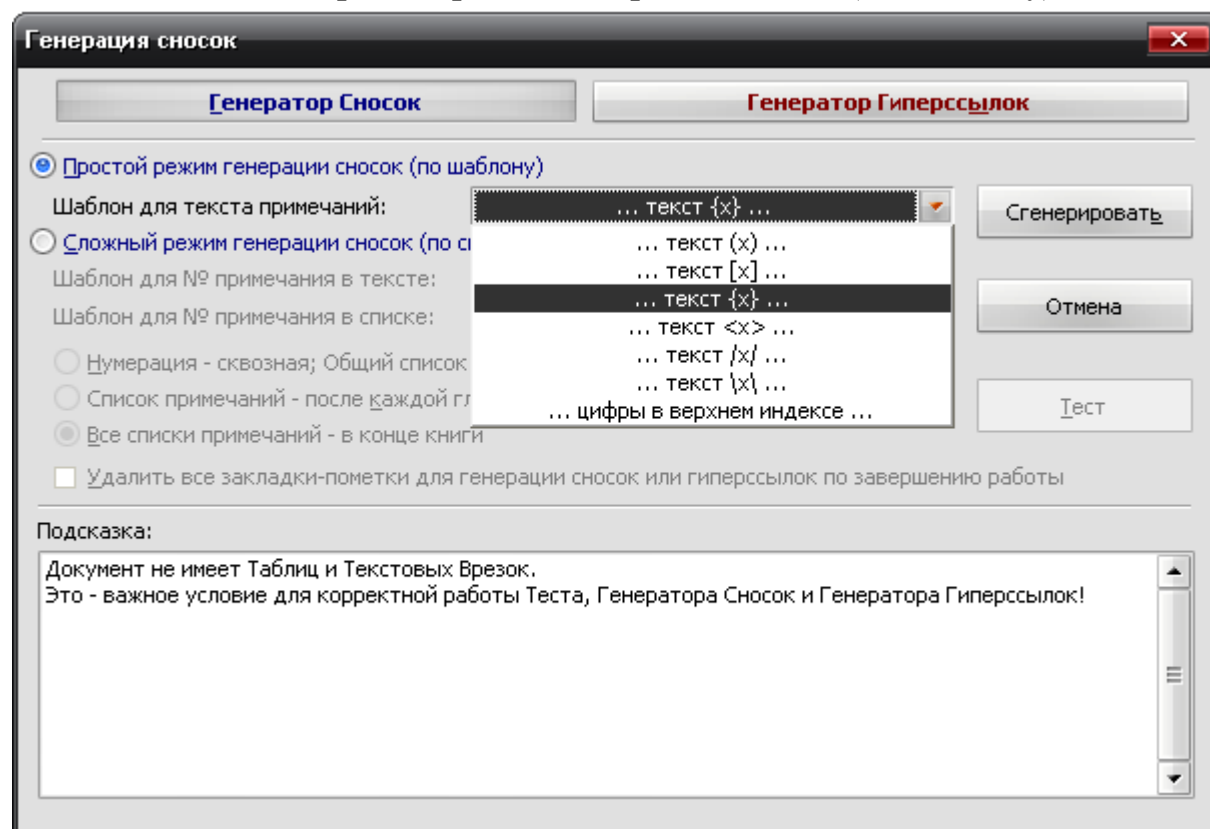
Переключение между ними осуществляется с помощью «залипающих» кнопок.

А. Генератор сносок



Есть 2 основных режима генерации сносок: Простой и Сложный.

4.1.1. 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.

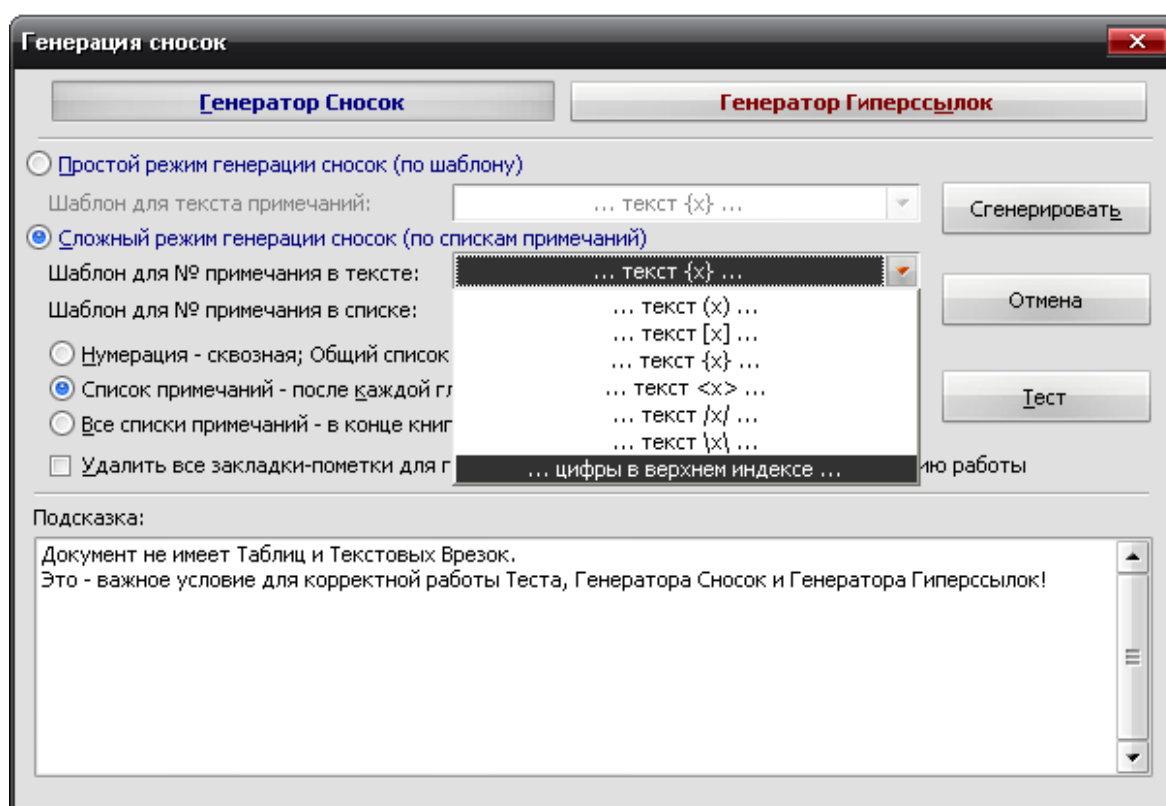
Использование выделений в тексте довольно удобно, когда не надо обрабатывать весь документ. Идеально было бы, если бы все идентификаторы кандидатов на сноски отличались бы от круглых или квадратных скобок. Тогда можно просто обработать весь документ, не думая о том, чтобы под обработку не попали и уточнения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

4.1.2. Сложный режим генерация сносок (по спискам примечаний)

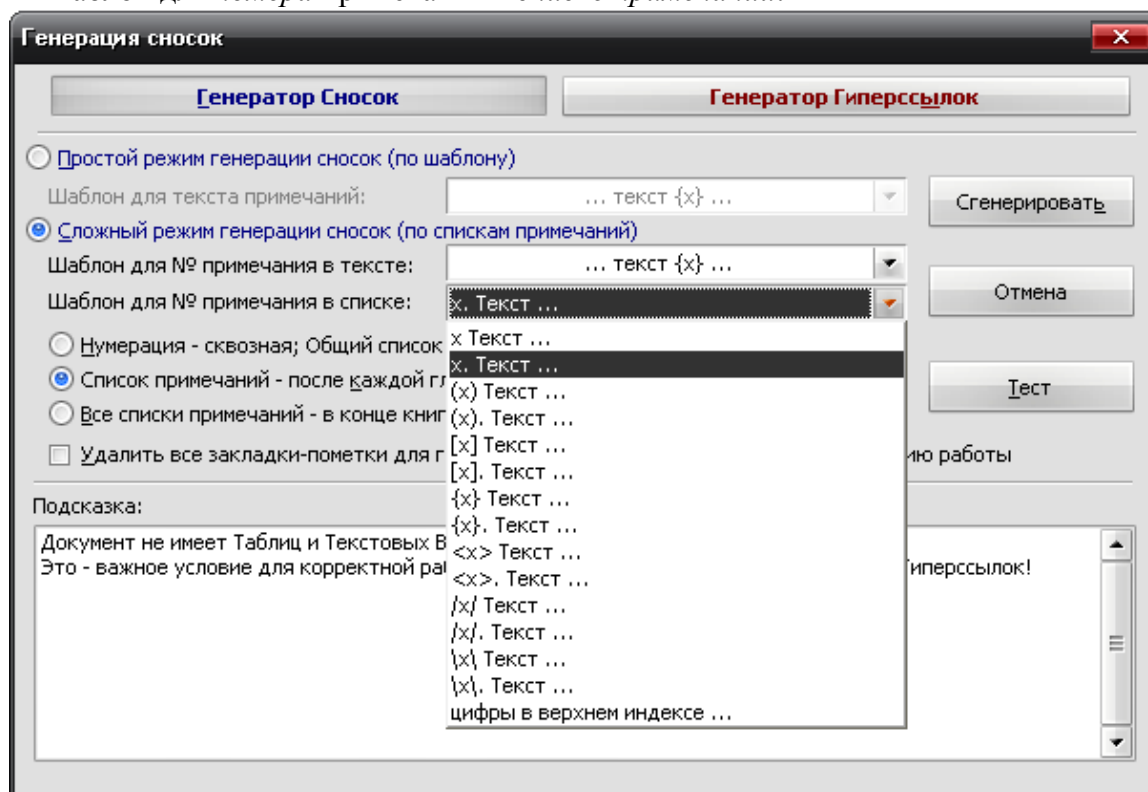
Сложный режим генерации сносок делится на 3 вида (по виду книг с примечаниями):

- 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.

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



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.

4.1.2.0. Очень ВАЖНО: корректность документа для сложной генерации сносок!

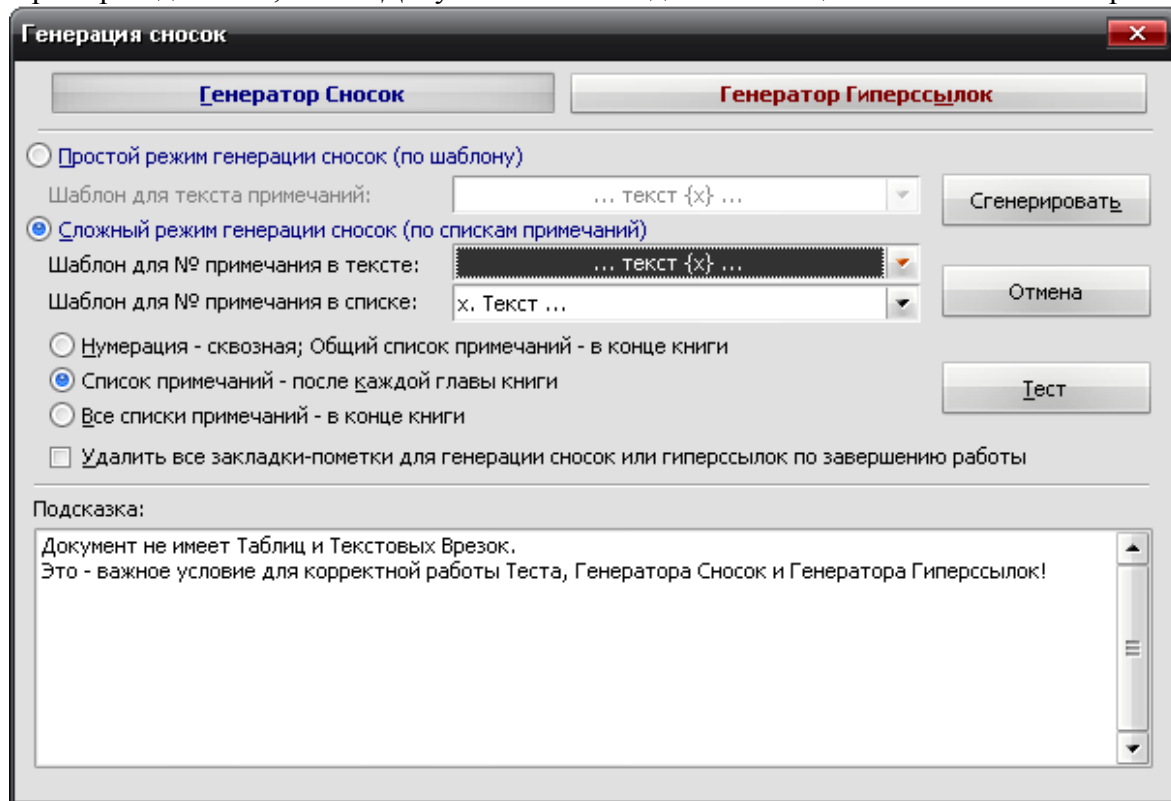
Корректность работы Генератора сносок и/или Теста в режиме сложной генерации зависит от наличия (вернее – отсутствия) Таблиц и/или Текстовых Врезок в Списках примечаний (эти области документа отмечены закладками _ooo_ocr_tc_foot_note_list_zone_dikbsd_ ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_1 ... _ooo_ocr_tc_foot_note_list_zone_dikbsd_N)¹:

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».

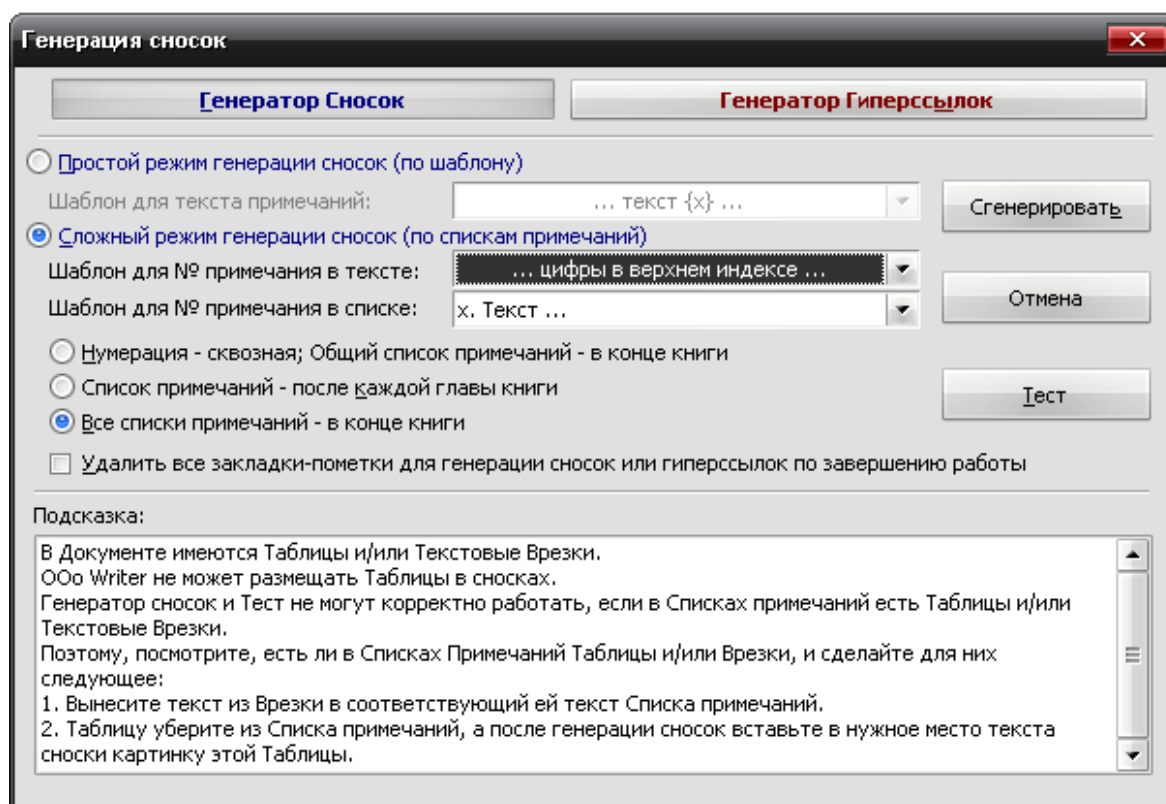
Поэтому, если в Документе есть хоть одна Врезка или Таблица, то выдается **сообщение-подсказка** в текстовом поле формы Генератора сносок, что нужно сделать. Дальнейшая работа - на ваш страх и риск. Лучшее - просмотреть Документ, в частности **все Списки примечаний** (они помечены закладками _ooo_ocr_tc_foot_note_list_zone_dikbsd_ ... _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:

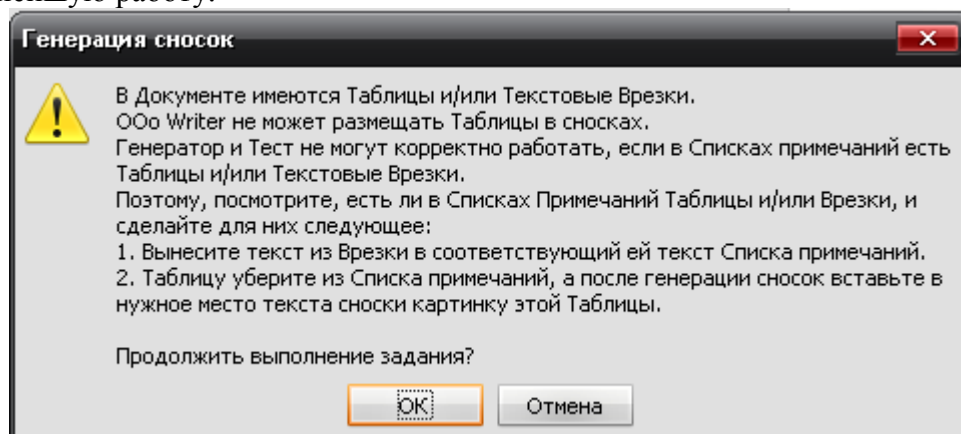


Пример подсказки, если в Документе есть хоть одна Таблица и/или Текстовая Врезка:

¹ В названия служебных закладок входит Ник автора пакета OoFBTools. Это сделано не из-за тщеславия автора ☺, а для того, чтобы максимально исключить вероятность наличия в обрабатываемом документе одноименных закладок.



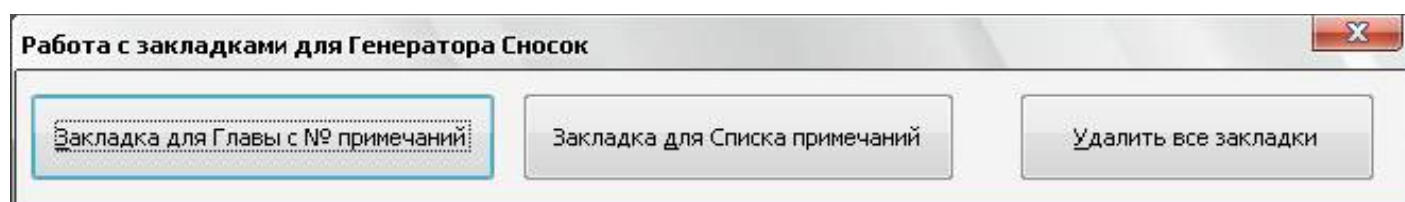
В этом случае, при нажатии на кнопку **Тест** или **Сгенерировать**, выдается запрос на дальнейшую работу:



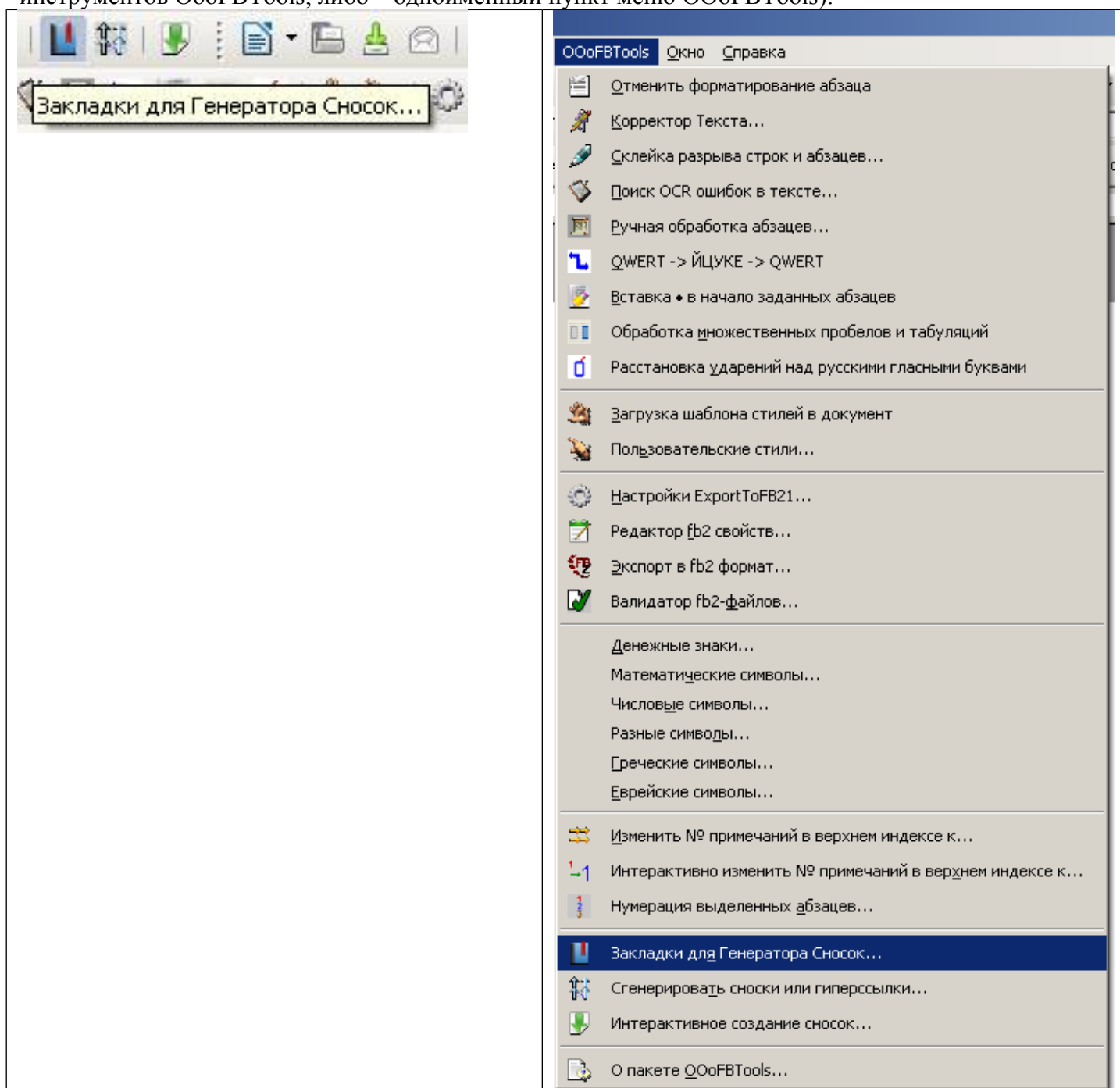
Если в Списках примечаний Таблиц и/или Врезок нет, но они есть в других частях Документа (Главы книги и т.д.), то не обращайте внимание на это предупреждение, жмите кнопку ОК, и все будет ОК ☺. Но, если в Списках примечаний есть Таблица и/или Врезка, то все же лучше от них «избавиться», как об этом написано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

4.1.2.1. Диалог Работы с закладками для Генератора Сносок



Генератор сносок в своем поиске номеров сносок и их текста «ориентируется» на специальные закладки для Главы книги и для соответствующего этой Главе Списка текста сносок (об этом – см. ниже). Пользователь проставляет эти закладки при помощи Инструмента **Работа с закладками Генератора Сносок** (Нажимаем либо кнопку **Закладки для Генератора Сносок...** на панели инструментов OoFBTools, либо – одноименный пункт меню OoFBTools):



Далее в описании работы Генератора сносок будут использоваться кнопки этого Инструмента.

Например, есть такой текст:

Глава 1
Текст¹ с несколькими² сносками³.

Примечания к Главе 1

- ¹ Текст 1-й сноски
- ² Текст 2-й сноски
- ³ Текст 3-й сноски

Кнопка «**Закладка для Главы с № примечаний**» используется для вставки рабочей закладки в начале Главы Книги, в тексте которой (Главы) есть номера примечаний. Для приведенного примера на картинке– ставим курсор мышки слева от «Глава 1»..

Кнопка «**Закладка для Списка примечаний**» используется для вставки рабочей закладки на выделенную область Списка примечаний. Для приведенного примера:

Глава 1
Текст¹ с несколькими² сносками³.

Примечания к Главе 1

- ¹ Текст 1-й сноски
- ² Текст 2-й сноски
- ³ Текст 3-й сноски

Кнопка «**Удалить все закладки**» используется для удаления всех закладок **после** Генерации Сносок. В самом Генераторе есть опция, позволяющая удалять эти закладки после Генерации сносок. но ее лучше отключить, чтобы после Генерации по закладкам можно было легко найти нужное место – пустая область от бывших Списков примечаний и т.д. и удалить эти «пустые области».

Далее будут описаны Режимы работы Генератора Сносок. Но, помните - все закладки должны быть проставлены **до** его вызова (впрочем, он сам уведомит вас, если вы забыли это сделать, или не все закладки проставлены ☺)

4.1.2.2. 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:

Полный список примечаний – в конце

книги	
	Глава 1 Какой-то текст(1) Еще текст.
	Глава 2 Какой-то текст Еще текст.
	Глава 3 Какой-то текст(2) Еще текст.
	— Конец Книги — Примечания 1. Текст сноски 1 2. Текст сноски 2

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» Инструмента **Работа с закладками для Генератора Сносок**) – (в таблице ниже этот отрывок текста имеет **розовый** цвет).

Полный список примечаний – в конце книги (с закладкой на Списке примечаний)	
	Глава 1 Какой-то текст(1) Еще текст.
	Глава 2 Какой-то текст Еще текст.
	Глава 3 Какой-то текст(2) Еще текст.
	— Конец Книги — Примечания 1. Текст сноски 1 2. Текст сноски 2

Теперь можно запустить генератор.

4.1.2.3. 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:

Список примечаний – после главы книги	
	Глава 1 Какой-то текст(1) Еще текст.
	Примечания к главе 1 1. Текст сноски 1
	Глава 2 Какой-то текст Еще текст.
	Глава 3 Какой-то текст(1) Еще текст.
	Примечания к главе 3 1. Текст сноски 1

(1) Расстановка закладок: Инструмент «Вставка закладки для списка примечаний»

Для указания программе списков примечаний необходимо соответствующий фрагмент текста выделить мышкой, и нажать на кнопку «Закладка для Списка примечаний» Инструмента **Работа с закладками для Генератора Сносок** (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.

(2) 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 книга, 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 data-bbox="267 575 573 737"> <p>Глава 1</p> <p>Какой-то текст(1) Еще текст.</p> </div> <div data-bbox="267 737 573 884"> <p>Примечания к главе 1</p> <p>1. Текст сноски 1</p> </div> <div data-bbox="267 884 573 1045"> <p>Глава 2</p> <p>Какой-то текст(1) Еще текст(2).</p> </div> <div data-bbox="267 1045 573 1207"> <p>Глава 3</p> <p>Какой-то текст(1) Еще текст.</p> </div> <div data-bbox="267 1207 573 1360"> <p>Примечания к главе 3</p> <p>1. Текст сноски 1</p> </div>	<div data-bbox="927 575 1232 737"> <p>Глава 1</p> <p>Какой-то текст(1) Еще текст.</p> </div> <div data-bbox="927 737 1232 884"> <p>Примечания к главе 1</p> <p>1. Текст сноски 1</p> </div> <div data-bbox="927 884 1232 1045"> <p>Глава 2</p> <p>Какой-то текст(1) Еще текст(2).</p> </div> <div data-bbox="927 1045 1232 1207"> <p> </p> </div> <div data-bbox="927 1207 1232 1360"> <p>Глава 3</p> <p>Какой-то текст(1) Еще текст.</p> </div> <div data-bbox="927 1360 1232 1514"> <p>Примечания к главе 3</p> <p>1. Текст сноски 1</p> </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 кнопки **«Закладка для Списка примечаний»** **Инструмента Работа с закладками для Генератора Сносок** 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».

4.1.2.4. 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:



(1) 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 кнопки «**Закладка для Списка примечаний**» Инструмента **Работа с закладками для Генератора Сносок** 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 кнопки «**Закладка для Главы с номером примечания**» Инструмента **Работа с закладками для Генератора Сносок**.

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





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).

(2) 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!

(3) 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.

(4) 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>

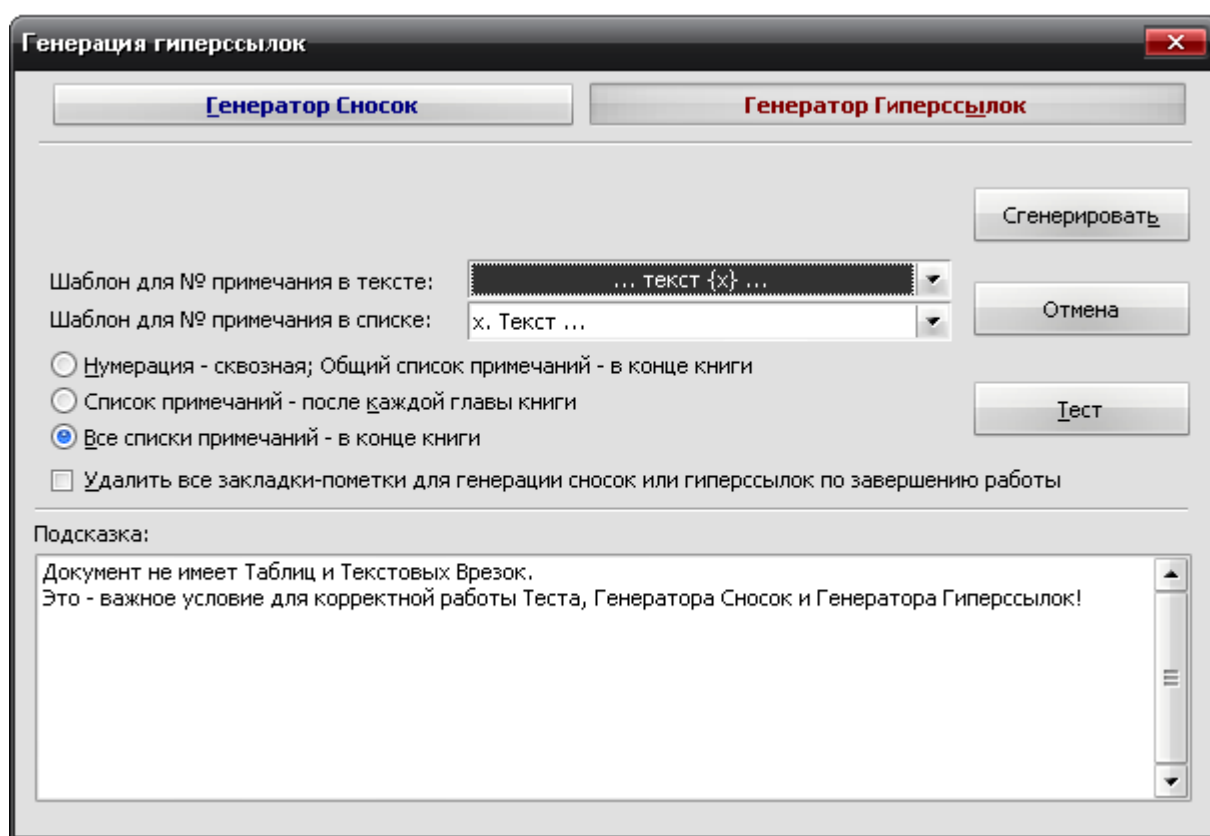
В этом примере в книге всего 3 главы. Во всех главах есть номера примечаний. Но списки примечаний в конце книги есть только для глав 1 и 3. 1). На начало каждой главы, где есть номера примечаний ставим «Закладку для главы с номером примечаний» (зеленый цвет на картинке). 2). На списки примечаний ставим «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 — 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 — 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**» с помощью Инструмента **Работа с закладками для Генератора Сносок** (зеленый цвет на картинке). 2). Ставим «**Закладку для Списков примечаний**» с помощью Инструмента **Работа с закладками для Генератора Сносок** 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.

Б. Генератор Гиперссылок



Генератор гиперссылок применяется в том случае, когда по какой-либо причине желательно сделать не сноски из примечаний, а просто переходы из соответствующих абзацев книги на примечания. В этом случае Списки примечаний остаются в тексте, и электронная книга становится максимально похожа на бумажную книгу (кому как нравится).

Расстановка закладок и работа режимов генерации – точно такие же, как и для [Генератора Сносок в Сложном режиме генерации](#). Поэтому, не будем повторяться.

В. Опция «Удалить все закладки-пометки для генерации сносок или гиперссылок по завершению работы»

В диалог Генерации сносок есть опция «Удалить все закладки-пометки для генерации сносок или гиперссылок 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 (хотя, использовать [Тест](#) удобнее)? Для этого с помощью средств OOoWriter (навигатор вызывается по F5) 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.

Г. 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 <р> 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)...

Д. Важные замечания

Генераторы сносок и гиперссылок игнорируют 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.

Е. Тесты к Генераторам сносок и гиперссылок

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 или гиперссылки сгенерируются корректно, сначала нужно выполнить тест, немного подождать, потом проанализировать результат (он будет автоматически показан в браузере, стоящем по-умолчанию в системе) в виде html-файла.

1. Примеры ошибок OCR и «ложного срабатывания» генераторов

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

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 или гиперссылки.

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 (гиперссылка). 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 (гиперссылки).

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

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 распознался 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 распознался 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) Несоответствие 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. 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

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 (вызывается по F5) можно легко перейти в нужное место и исправить ошибки в нумерации примечаний и т.д.

Пользуйтесь этой информацией!!!

4. Важное замечание о Тесте

Часто пользователь может по невнимательности ошибочно выбрать не тот режим сложной генерации сносок (гиперссылок), 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 (гиперссылок)!

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 (гиперссылок) вас «приятно разочарует» ☺. Будьте внимательны!

4.2. Интерактивное преобразование номеров примечаний в верхнем индексе по шаблону

Интерактивное преобразование № примечаний в верхнем индексе по шаблону

Найти

1. Шаблон для № примечания в тексте: ... текст [x] ... Преобразовать для 1.

2. Шаблон для № примечания в списке: x. ... Преобразовать для 2.

☒ Искать далее после обработки найденного ☐ Только цифры ☒ Любые символы

Режим поиска символов

☒ Только в верхнем индексе ☐ Только в нижнем индексе ☐ И в верхнем и в нижнем индексе

Этот инструмент полезен в тех случаях, где нельзя использовать инструмент автоматического преобразования номеров примечаний в верхнем индексе (см. выше) – в документе есть математические формулы со степенями (цифры в верхнем индексе) и т.д. 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 сносков в книге (см. соответствующий инструмент выше).

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 сносков они могут преобразоваться по шаблону! Лучше избрать шаблон в виде фигурных скобок {x}:

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 сносков в списках текста сносков. 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.

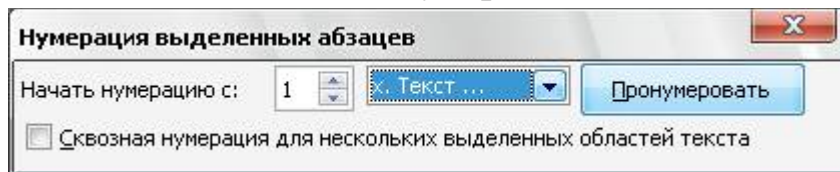
2. Текст сноски 2.

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...

P.S. Этот инструмент просто незаменим для поиска символов или цифр в любом индексе. Это полезно в ряде случаев – для исправления неверно распознанных символов, для поиска номеров сносок для последующей ручной вставки сноски в это место и т.д.

4.3. 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 или гиперссылок](#) **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!

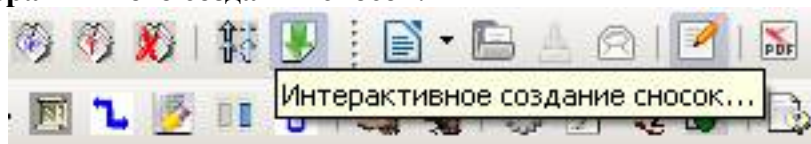
Можно задать вид нумерации (полезно для последующей автогенерации сносок или гиперссылок).

Мне этот инструмент экономит массу времени!

4.4. Интерактивное создание сносок

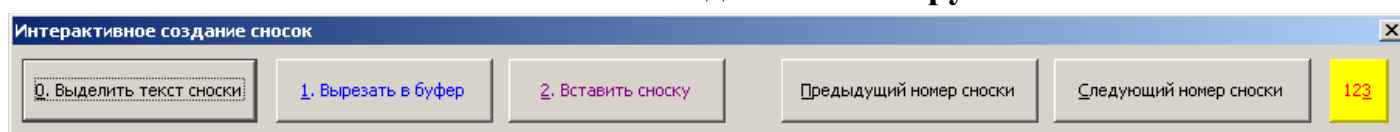
Есть множество книг, где списки сносок расположены в конце каждой, или почти каждой страницы. Генератор сносок не умеет обрабатывать такие сноски.

Чтобы облегчить обработку этого вида сносок был разработан Инструмент для **Интерактивного создания сносок**.



Пользоваться Инструментом чрезвычайно просто – поочередное нажатие 2-х – 3-х кнопок.

4.4.1. Использование диалога инструмента



Кнопка "Выделение цветом номеров сносок для наглядности" (желтая кнопка с текстом "123"), используется для того, чтобы выделить номера сносок в тексте **красным** цветом и **желтым** маркером, а номера сносок в списках сносок - **красным** цветом и **голубым** маркером. Гораздо удобнее ориентироваться при работе с инструментами, когда сноски "подсвечены".

Методика интерактивного создания сносок

Рассмотрим пример работы с небольшим фрагментом текста. После OCR текст страниц и списки сносок идут последовательно один за другим:

Текст книги¹ с *несколькими*² сносками³.

В конце страницы расположен список текста сносок

¹ Текст 1-й сноски.

² Текст 2-й сноски.

³ Текст 3-й сноски.

*Текст*⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

⁴ Текст 4-й сноски.

⁵ Текст 5-й сноски.

⁶ Текст 6-й сноски.

В данном примере отдельные слова и номера сносок в верхнем индексе имеют свои символьные стили – **полужирный**, *курсив* и подчёркивание. Инструмент интерактивного создания сносок убирает эти символьные стили у номеров сносок. Иначе – наклонный и/или курсивный номер сноски в тексте

будет некрасиво выглядеть.

Запускаем Инструмент Интерактивного создания сносок.

1. «Подсвечивание» цветом номеров сносок

Этот этап не является обязательным. Он просто обеспечивает удобство ориентации в тексте: всегда отличать искать глазами нужный текст (в данном случае – номера сносок) среди цветного текста, нежели в черно-белом.

Нажимаем на желтую кнопку «123».

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок

¹ Текст 1-й сноски.

² Текст 2-й сноски.

³ Текст 3-й сноски.

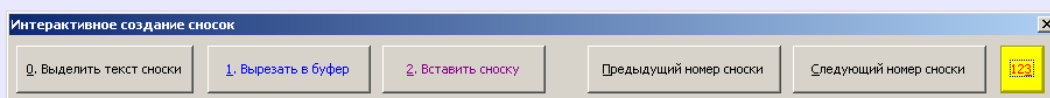
Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

⁴ Текст 4-й сноски.

⁵ Текст 5-й сноски.

⁶ Текст 6-й сноски.



Как видим, стало значительно удобнее и легче видеть, где в книге расположен номер сноски в верхнем индексе, а где – текст этой сноски в списке сносок.

2. Выделение текста 1-й сноски в списке сносок

Ставим курсор в *любое* место текста 1-й сноски и нажимаем кнопку «0. Выделить текст сноски». В дальнейшем, вплоть до конца списка сносок, эта кнопка не потребуется, так как все последующие сноски *до конца списка* будут выделяться автоматически.

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок

¹ Текст 1-й сноски.

² Текст 2-й сноски.

³ Текст 3-й сноски.

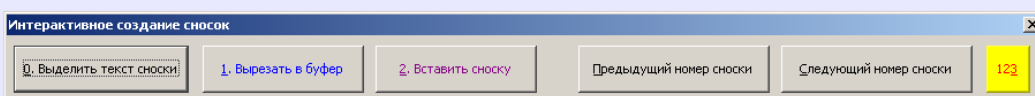
Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

⁴ Текст 4-й сноски.

⁵ Текст 5-й сноски.

⁶ Текст 6-й сноски.



Важно помнить, что автоматически выделяется только один абзац текста сноски! Если сноска состоит из более, чем одного абзаца, то все последующие абзацы нужно будет выделить самостоятельно.

3. Вырезание текста сноски в буфер обмена

Нажимаем кнопку «1. Вырезать в буфер»:

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок



Текст 2-й сноски.

Текст 3-й сноски.

Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

Текст 4-й сноски.

Текст 5-й сноски.

Текст 6-й сноски.

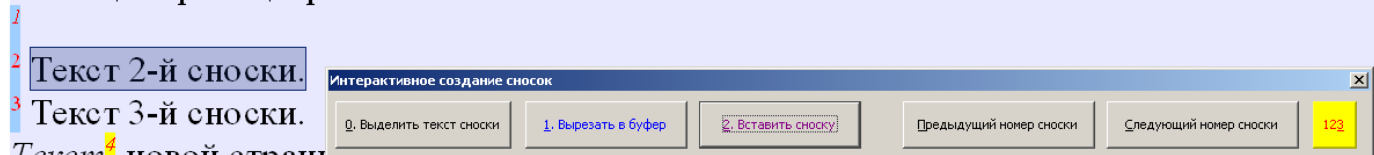
В списке сносок текст 1-й сноски вырезался в буфер, и автоматически был осуществлён переход на первый в книге номер в верхнем индексе (он выделяется). В данном примере этот номер совпал с номером 1-й сноски. В различных технических и научных статьях и книгах цифры в верхнем индексе могут представлять собой не только номер сноски, но и степень в формулах... Если инструмент автоматически перешел не на номер сноски, а на другое число в верхнем индексе, для поиска номера сноски нужно нажать кнопку «Следующий номер сноски» до тех пор, пока не будет найден номер нужной нам сноски. Обычно в большинстве книг такое действие по поиску номера сноски не требуется.

4. Генерация сноски

Теперь, нажав кнопку «2. Вставить сноску» мы генерируем сноску на месте найденного ее номера. В самом низу страницы редактора – текст сгенерированной сноски:

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок



Текст 2-й сноски.

Текст 3-й сноски.

Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

Текст 4-й сноски.

Текст 5-й сноски.

Текст 6-й сноски.

1 Текст 1-й сноски.

Когда сноска сгенерировалась, курсор автоматически переходит на список текста сносок, откуда был вырезан текст вставленной сноски. В данном случае – на текст 2-й сноски. Это очень удобно, т.к. не надо всякий раз просматривать текст в поисках списка сносок. Текст этой сноски автоматически выделяется, игнорируя номер сноски и пробелы между ним и текстом, если они есть.

5. Цикл создания сносок

Цикл генерации сносок теперь сводится просто к поочередному нажатию 2-х кнопок и беглому просмотру – та ли сноска обрабатывается. Нажимаем кнопку «1. Вырезать в буфер», потом – кнопку

«2. Найти номер сноски» – и переходим на следующий номер сноски от только что сгенерированной. Переставлять вручную курсор никуда не надо – инструмент автоматически запоминает положение последней созданной сноски и с этого места начинает поиск следующей. Выделять текст новой сноски тоже не надо – инструмент после вставки текста предыдущей сноски автоматически выделяет текст новой.

Когда мы сгенерировали последнюю сноску в списке сносок данной страницы (в нашем примере – это сноска «Текст 3-й сноски»), то инструмент автоматически выделяет следующий абзац, который не является текстом сноски:

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок



Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

⁴ Текст 4-й сноски.

⁵ Текст 5-й сноски.

⁶ Текст 6-й сноски.

1 Текст 1-й сноски.

2 Текст 2-й сноски.

3 Текст 3-й сноски.

6. Удаление «отработанной» области списка текста сносок

В тексте осталась область с номерами от обработанного списка сносок. Ее надо выделить и удалить вручную:

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок

Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок

⁴ Текст 4-й сноски.

⁵ Текст 5-й сноски.

⁶ Текст 6-й сноски.



1 Текст 1-й сноски.

2 Текст 2-й сноски.

3 Текст 3-й сноски.

Дальнейшая работа сводится к повторению шагов 2-6. Обычно все сводится к последовательному нажатию 2-х кнопок (быстрых клавиш – см. ниже).

Вот результат работы:

Текст книги¹ с несколькими² сносками³.

В конце страницы расположен список текста сносок

Текст⁴ новой страницы⁵ со сносками⁶.

В конце страницы расположен следующий список текста сносок



1Текст 1-й сноски.

2Текст 2-й сноски.

3Текст 3-й сноски.

4Текст 4-й сноски.

5Текст 5-й сноски.

6Текст 6-й сноски.

Цвет уже сгенерированных номеров сносок можно легко вернуть к черному, выделив весь текст документа и применив к нему нужный цвет шрифта.

4.4.2.Необходимость в хотя бы беглом контроле работы

Если номера сносок в документе распознались в OCR-программе без ошибок (цифры в верхнем индексе, а не буквы), то работу по вышеописанному циклу создания сносок можно вести практически механически – нажатием последовательно 2-х – 3-х кнопок. К сожалению, на практике из-за плохого качества бумаги и сканирования, номера сносок могут быть распознаны неверно: либо буква вместо цифры, причем – не в верхнем индексе; либо номер сноски в тексте после OCR может отсутствовать и т.д. Поэтому между нажатием 2-х кнопок инструмента нужно хотя бы бегло просматривать – та ли сноска?

На всякий случай в Инструменте предусмотрены 2 кнопки для перехода к предыдущему номеру и к следующему номеру сноски в тексте (цифра в верхнем индексе). Иногда это бывает необходимо.

В идеале очень желательно перед генерацией сносок в интерактивном режиме хотя бы бегло просмотреть текст документа и исправить неверно распознанные номера сносок.

4.4.3.Интеллектуальность инструмента

Инструмент при нахождении текста сносок может различать «полезную» информацию — он пропускает пробелы, табуляции и неразрывные пробелы до текста сносок. Это предотвращает занесение в сноски ненужных символов.

4.4.4.Использование быстрых клавиш

Можно значительно увеличить скорость создания сносок, если назначить некоторым кнопкам клавиатуры выполнение функций кнопок Инструмента. Тогда можно будет, вызвав диалог Инструмента, и сдвинув его куда-нибудь в сторону (в район панелей инструментов), чтобы он не закрывал рабочую зону текста, просто нажимать поочередно пару кнопок на клавиатуре. Не надо будет постоянно двигать курсор мышки то к одной кнопке, то к другой.

Функции инструмента находятся в модуле **FootNoteInteractiveMaker**.

Таблица функций кнопок Инструмента

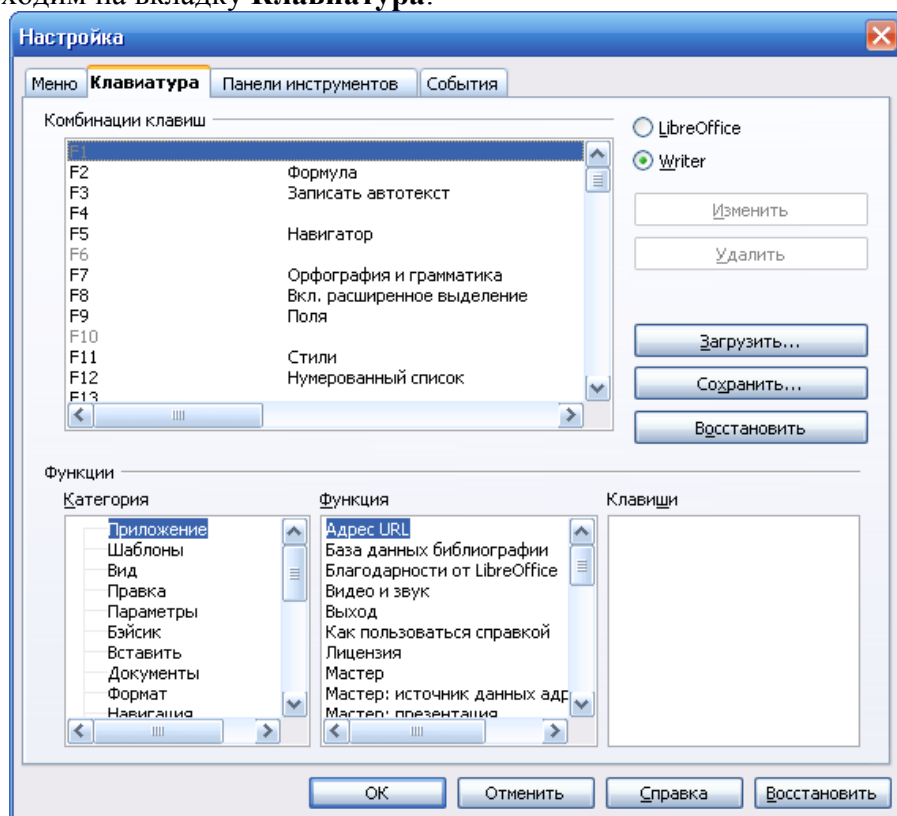
Кнопка Инструмента	Функция
0. Выделить текст сноски	SelectFNTextPara
1. Вырезать в буфер	CutSelectedTextInBuferAndGoToFootNoteUpIndex

2. Вставить сноску	InsertNoteText
Предыдущий номер сноски	FindPreviousFootNoteUpIndex
Следующий номер сноски	FindNextFootNoteUpIndex

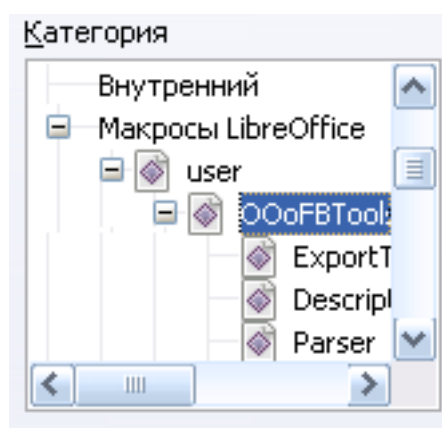
Функции можно назначать любым кнопкам клавиатуры, как вам будет удобно. Мне комфортнее работать с таким назначением кнопок:

Инструмента	Функция	Кнопка клавиатуры
0. Выделить текст сноски	SelectFNTextPara	Ctrl+Up
1. Вырезать в буфер	CutSelectedTextInBuferAndGoToFootNoteUpIndex	Ctrl+- (минус цифровых кнопок)
2. Вставить сноску	InsertNoteText	Ctrl++ (плюс цифровых кнопок)
Предыдущий номер сноски	FindPreviousFootNoteUpIndex	Ctrl+Left
Следующий номер сноски	FindNextFootNoteUpIndex	Ctrl+Right

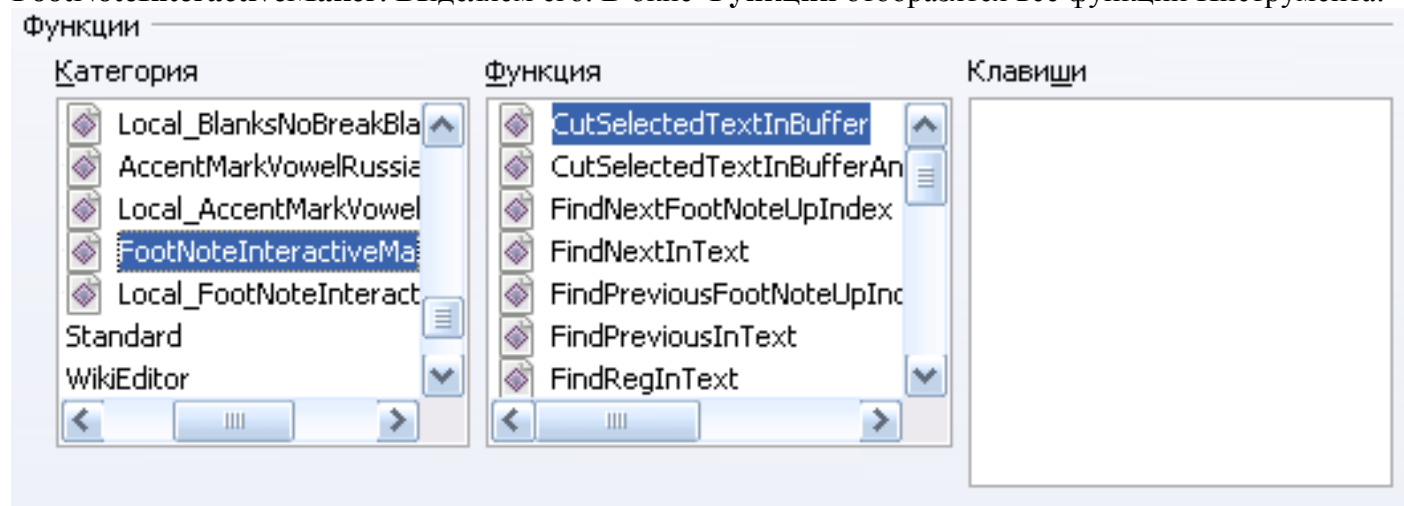
Делаем следующее. Из меню **Сервис->Настройка** вызываем диалог Настроек Офиса. Переходим на вкладку **Клавиатура**:



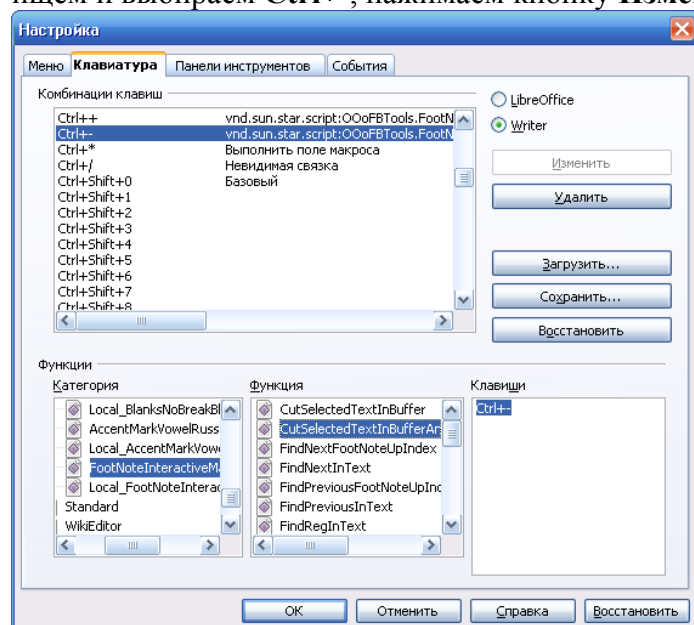
В древовидном списке Категория находим **Макросы LibreOffice** (для OpenOffice будет **Макросы OpenOffice**, для StarOffice – **Макросы StarOffice**). Раскрываем вложенные ветви **user->OOoFTools**:



Функции Инструмента Интерактивного создания сносок находятся в модуле **FootNoteInteractiveMaker**. Выделяем его. В окне **Функции** отобразятся все функции Инструмента:



Теперь ищем функцию кнопки «1. Вырезать в буфер». Это – **CutSelectedTextInBuferAndGoToFootNoteUpIndex**. Выделяем ее. В окне **Комбинации клавиш** ищем и выбираем **Ctrl+-**, нажимаем кнопку **Изменить**:



Аналогично назначаем для **Ctrl++** функцию **InsertNoteText** (работа кнопки «2. Вставить

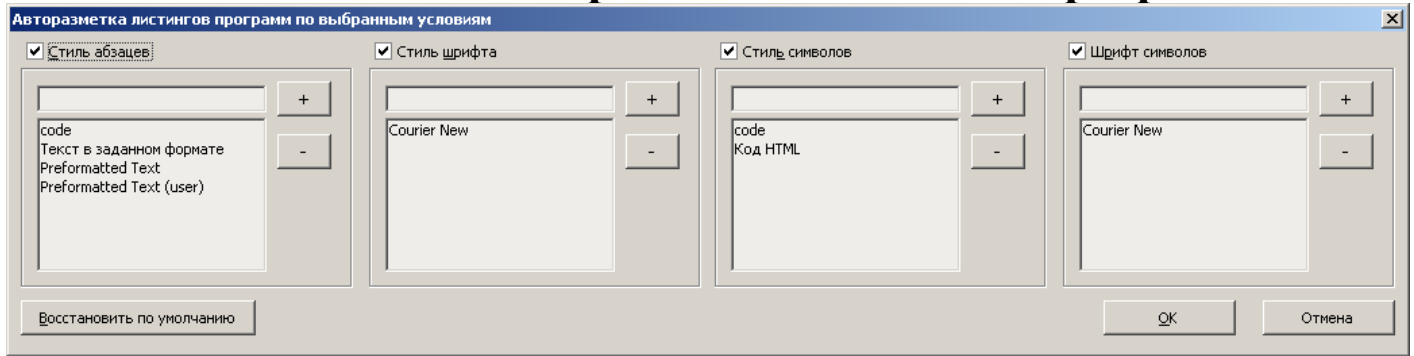
сноску»). Для **Ctrl+Left** – функцию **FindPreviousFootNoteUpIndex** (работа кнопки «**Предыдущий номер сноски**»). Для **Ctrl+Right** – функцию **FindNextFootNoteUpIndex** (работа кнопки «**Следующий номер сноски**»)

Нажимаем кнопку **Ok**. Все, теперь для того, чтобы быстро и удобно было работать со сносками, достаточно вызвать Инструмент, сдвинуть его в сторону (на панель инструментов) и работать только быстрыми кнопками.

Важно: При работе с быстрыми кнопками Инструмент должен быть запущен **обязательно!** Для того, чтобы панель инструмента не отвлекала и не мешала, ее можно сдвинуть как можно ближе к границе экрана.

С помощью этого инструмента можно создать за 1 час от 280 до 400 сносок.

5. Автоматическая разметка листингов программ



Предназначен для поиска в документе листингов программ, слов, операторов и т.д., и их разметки стилем **code** и шрифтом **Courier New**.

Используются следующие критерия идентификации текста, как листинга программы:

1. Шрифт для абзацев,
2. Стиль для абзацев,
3. Шрифт для символов
4. Стиль для символов.

Можно задействовать как один критерий, так и все 4.

При выборе критериев для Символов, инструмент отформатирует отдельные слова стилем **code** и шрифтом **Courier New**.

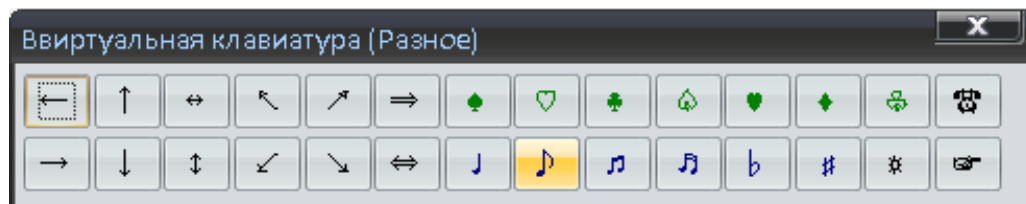
При выборе критериев для Абзацев, инструмент отформатирует найденные абзацы стилем **code** и шрифтом **Courier New**, а также заменит в этих абзацах множественные простые пробелы на неразрывные.

III. Виртуальные клавиатуры

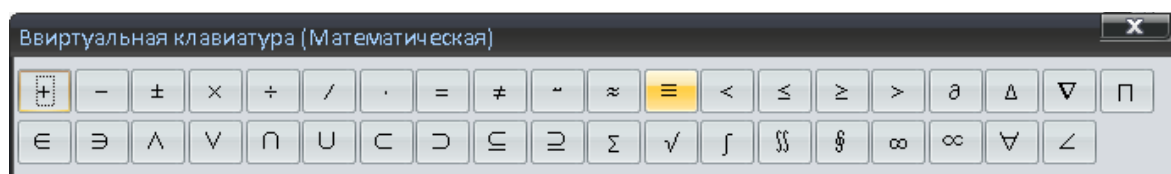
1. Виды виртуальных клавиатур

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

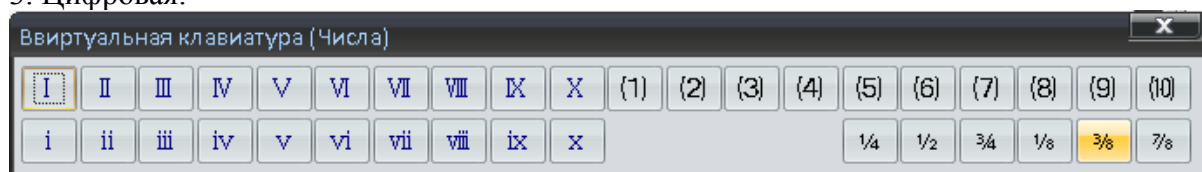
1. Различные символы (стрелки, ноты...):



2. Математическая:



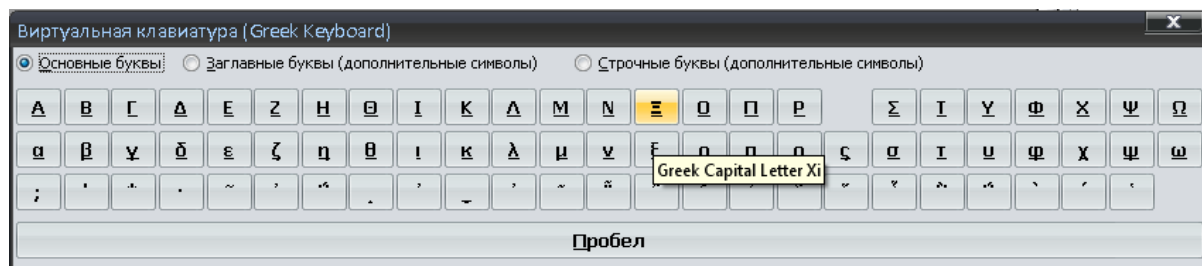
3. Цифровая:

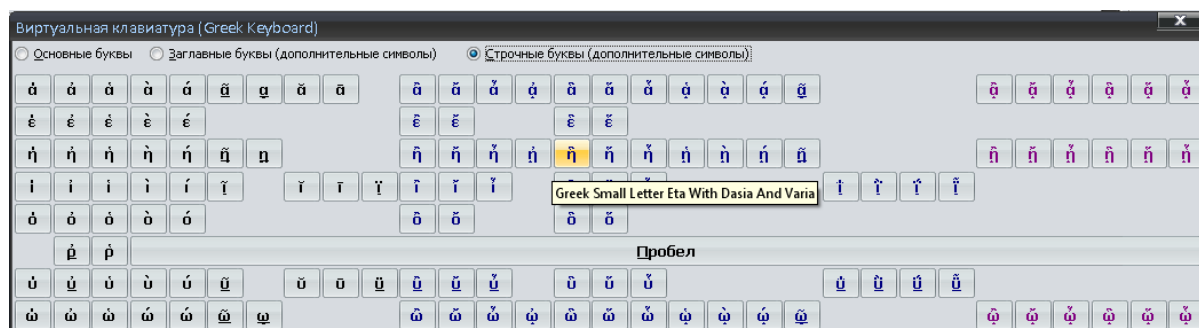
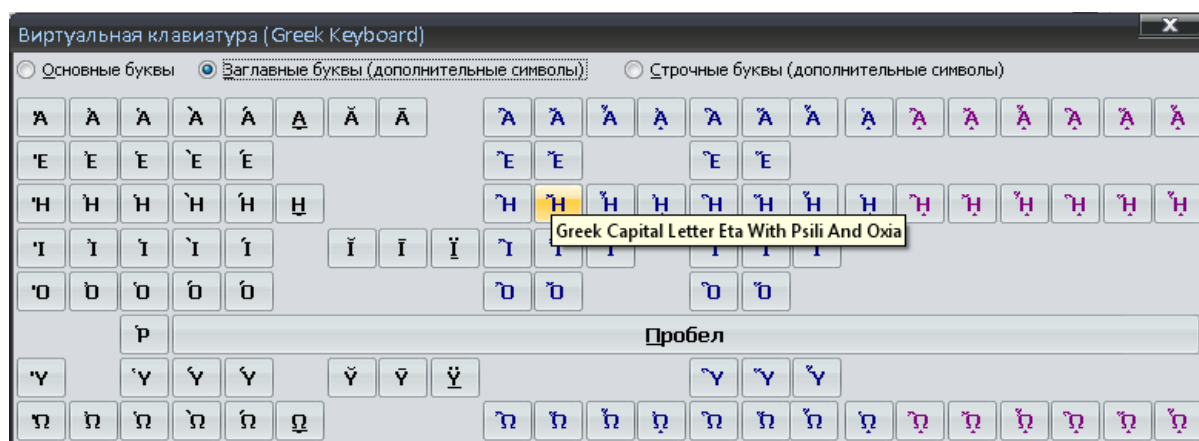


4. Валютная:



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.

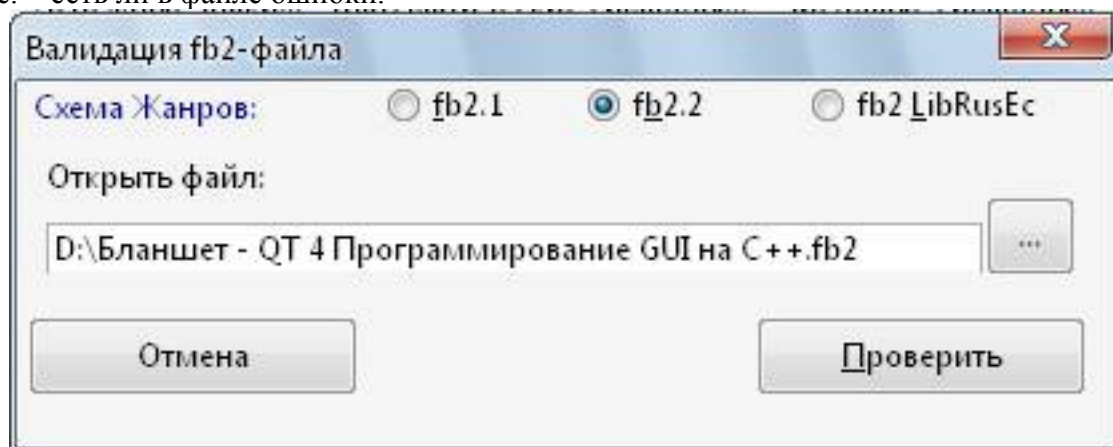
Для греческой виртуальной клавиатуры цветами выделены три вида букв – без дополнительных знаков, с 2-мя дополнительными знаками и с 3-мя дополнительными знаками. Так легче искать то, что нужно.

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**!

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.

В Валидаторе предусмотрен выбор FictionBook схемы. Схема «fb2 LibRusEc» отличается от fb2.1 и fb2.2 только гораздо большим списком разрешенных жанров.

При валидации будьте внимательны при выборе схемы Жанров. Если книга сделана с поддержкой списка жанров Либрусек, и в книге есть Жанры, присущие только этой схеме, а вы в Валидаторе выбрали схему «fb2.1» или «fb2.2», то Валидатор сообщит о невалидности проверяемого fb2-файла (и будет абсолютно прав).

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.

Папка хранения всех служебных файлов (профили и настройки) всех инструментов OOoFBTools

Все файлы профилей и настроек хранятся в домашней папке пользователя. Для Unix-подобных систем – это папка *в домашней папке в папке .OOoFBTools*, т.е. в ***D:\Мои документы\OOoFBTools*** для Windows и ***/home/dikbsd/.OOoFBTools*** для Linux. Для MacOS X – аналогично. В примерах показана моя учетная запись (*dikbsd*), у вас в системе будет ваша.

Это сделано для удобства переноса файлов при смене системы и/или версии OOo или LO.

Планы по развитию пакета OOoFBTools (TODO)

А. Конвертер 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. 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 профилей и настроек пакета OOoFBTools

Для Unix-подобных систем – это папка в домашней папке в папке *.OOoFBTools*, т.е. в *D:\Мои документы\OOoFBTools* для Windows и */home/dikbsd/.OOoFBTools* для Linux. Для MacOS X – аналогично. В примерах показана моя учетная запись (dikbsd), у вас в системе будет ваша.

3. При установке расширения пакета OOoFBTools постоянно появляются сообщения об ошибке и пакет не устанавливается

Такое часто случается и на других расширениях при частой их переустановке.

Сами расширения в этом «не виноваты».

Простейший способ устранить эту ошибку установки пакета – удалить всю папку OpenOffice.org, которая располагается в папке настроек текущего пользователя.

Для Windows XP это - C:\Documents and Settings\VadimK\Application Data\.

Т.е. надо удалить папку C:\Documents and Settings\VadimK\Application Data\ OpenOffice.org,

Удаление этой папки часто помогает решить проблему и тогда, когда установленное и работающее расширение вдруг перестает работать.

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:

Конвертер «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 **Книга - Название**). 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» конвертером, 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

Этот раздел больше не актуален, т. к. конвертер автоматически перед экспортом перезагружает документ. Раздел оставлен «на всякий случай»...

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)? Конвертер не умеет экспортировать данные **вложенных** Врезок в другие Врезки! Просто вытащите картинку из этих Врезок, перезагрузите файл и сделайте 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».

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 конвертера – 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 Заголовок 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.

Вариант 1 (метод уже устарел, но работает)

Чтобы «его» экспортировать, как Автора Предисловия, не экспортируя при этом само Предисловие, как Цитату, нужно перед Автором вставить пустой абзац, придав ему стиль **Цитата**, а Автору – стиль **Цитата - Автор**. 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 **Цитата**, то конвертер не воспримет стиль **Цитата - Автор** Автора Предисловия, как **Цитата - Автор**, 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.

Вариант 2 (доработка конвертера)

Конвертер доработан так, чтобы экспортировать Автора Цитаты после текста без введения пустого абзаца стиля **Цитата - Автор** между самым текстом и Автором Цитаты. Просто отформатируйте Автора Цитаты после текста стилем **Цитата - Автор**. После экспорта получим примерно следующее:

```

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

```

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 OoFBTools 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 Конвертера 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 (он указан в **О программе пакета**).

Сообщение об ошибке при конвертировании

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

А. Чаше всего это происходит на очень больших документах. Из-за ограничения OoO и Lo средствами StarBasic не удается экспортировать очень большой текст, храня его в памяти. Это пока неразрешимая проблема конвертера.

Выход из этой ситуации предоставляет **LOPyFB2Tools** – замечательный форк OoFBTools. Он написан на языке Python, и поэтому не имеет перечисленных выше ограничений StarBasic, и прекрасно экспортирует очень большие документы.

Скачать **LOPyFB2Tools** можно по адресу: <http://ru.fbtools.org/home/lopyfb2tools>

Исходный код: <https://gitorious.org/lopyfb2tools>

Б. Еще одна распространенная причина «падения» конвертера – когда весь текст документа 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 Анализатора 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 **Стих - Автор**. Это значительно уменьшит размер блоков текста под структуру (в данном случае – под **Стих**).

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

В. 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