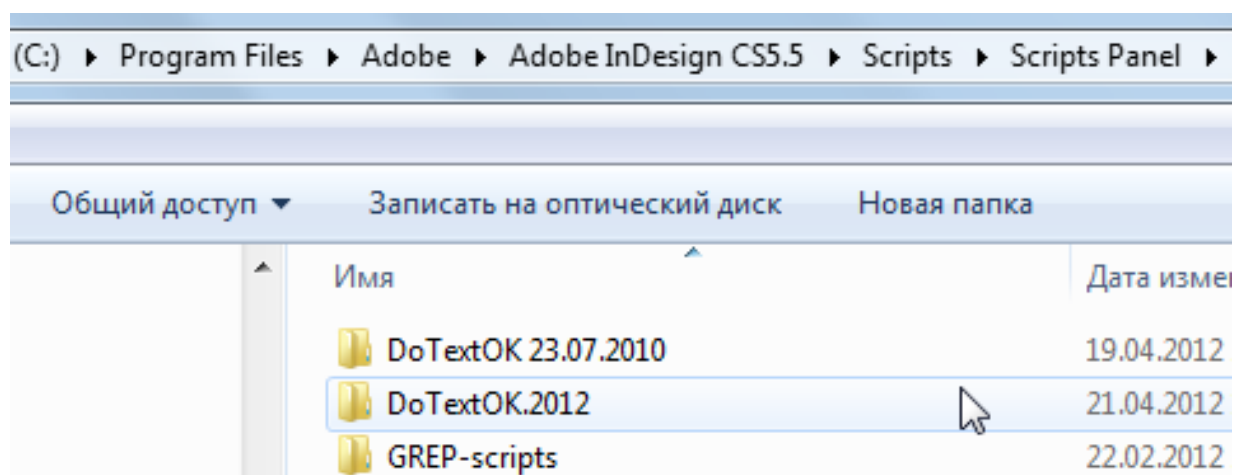


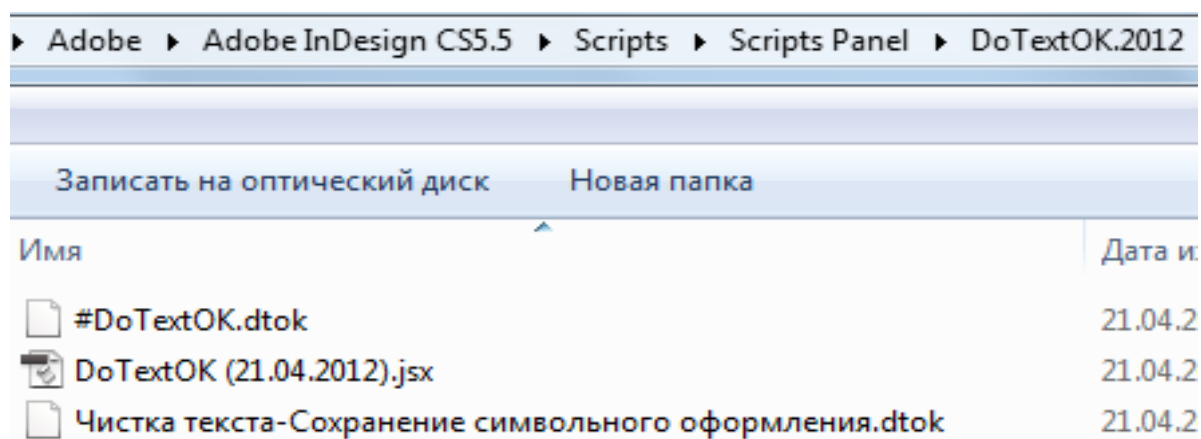
DoTextOK-2012

1. Размещение скрипта

Скрипты помещаются в каталоге Program Files\Adobe\Adobe InDesign CS5.5\Scripts\Scripts Panel. Но лучшее решение — сделать отдельную папку для этого скрипта. На этом рисунке она называется DoTextOK.2012.



Вот как может выглядеть содержимое этой папки:



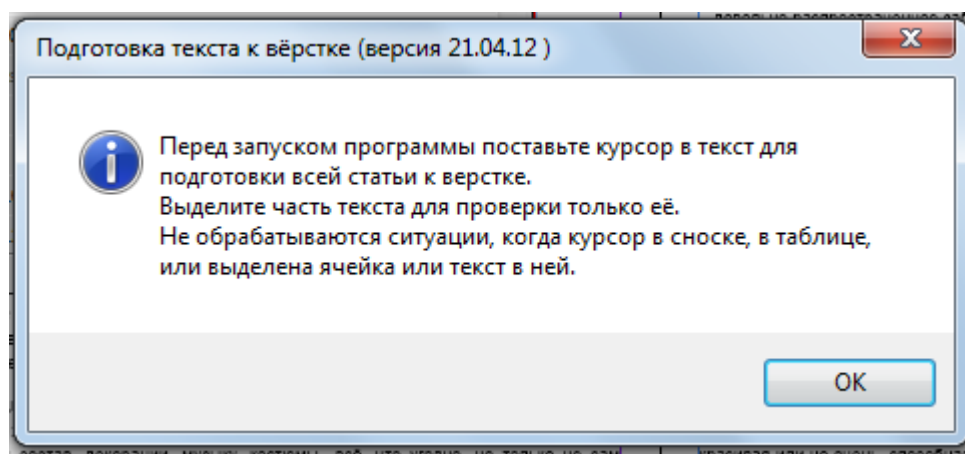
Первая строка – это имеющийся всегда шаблон настроек. Он создается при первом запуске скрипта, в нём сохраняются установки последнего запуска скрипта.

Вторая строка – это сам скрипт.

Третья – созданный шаблон установок обработки текста. Таких шаблонов на разные случаи можно сделать сколько угодно.

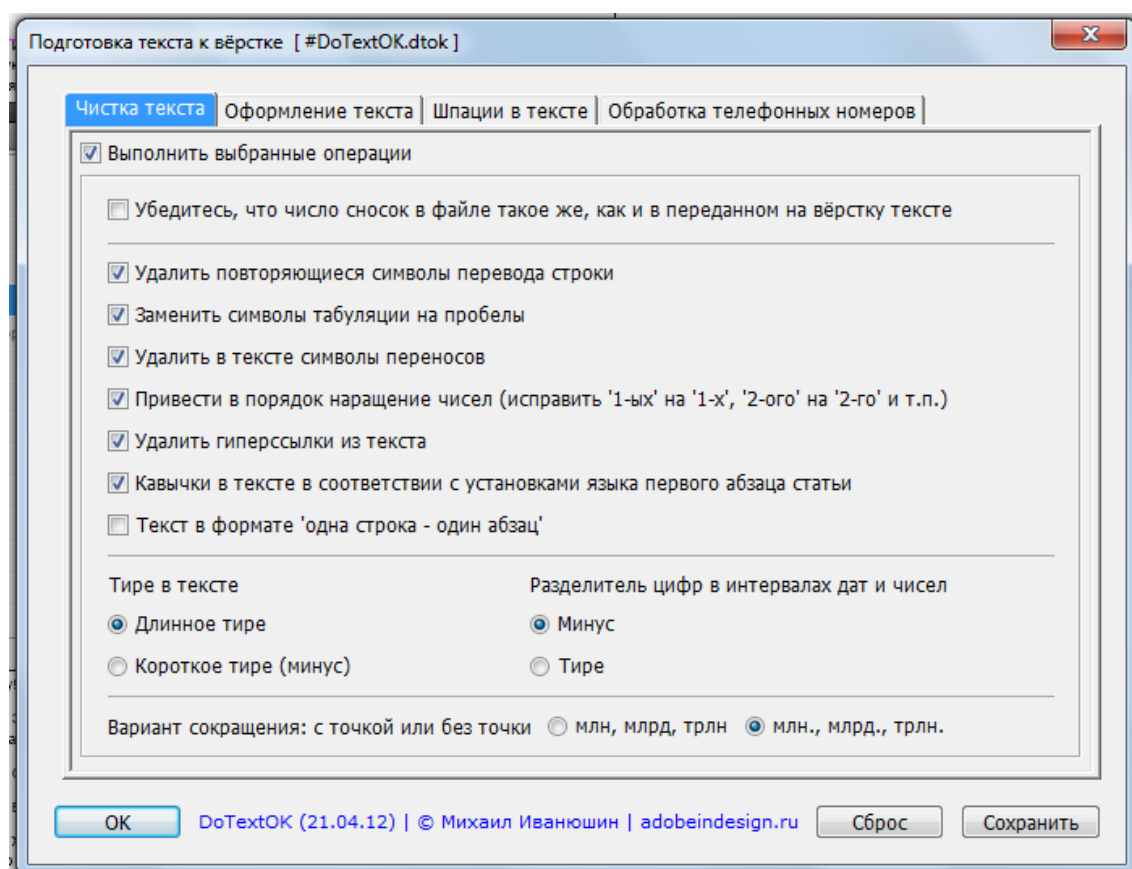
2. Запуск программы

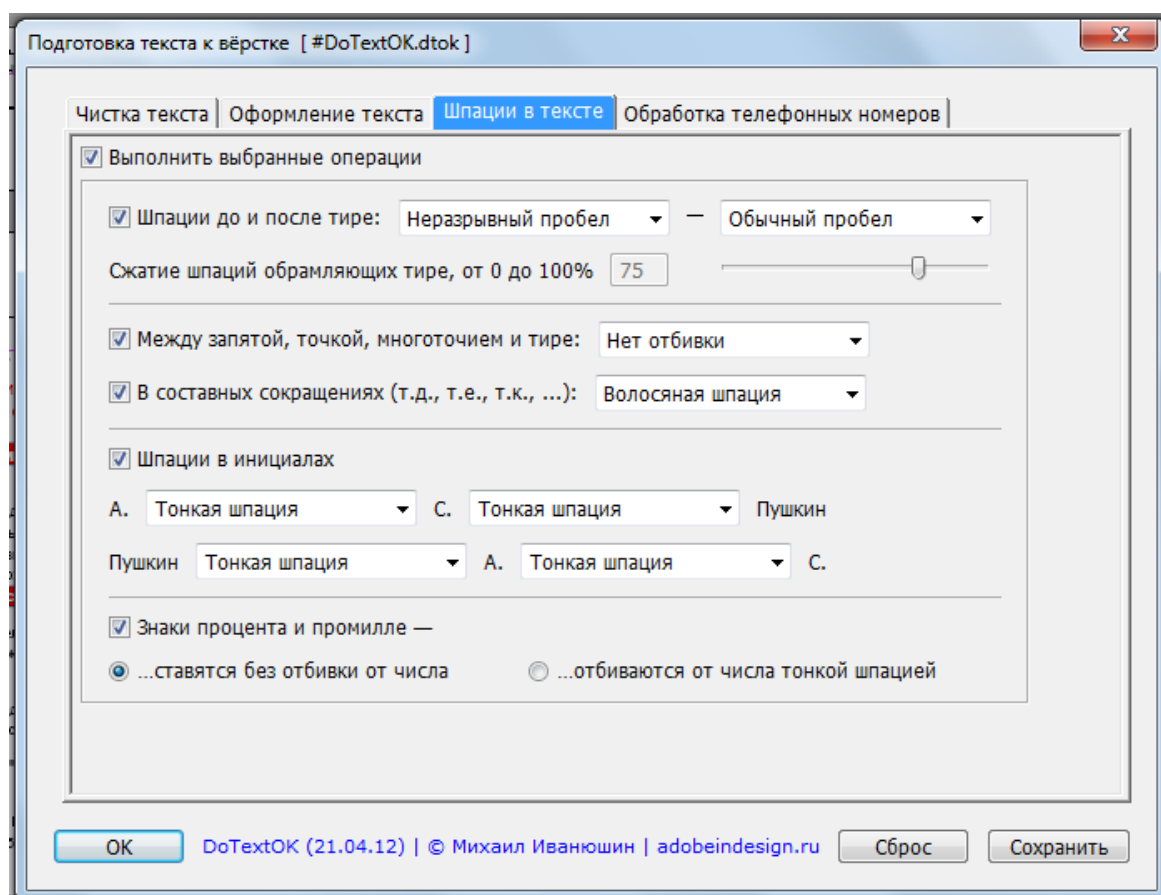
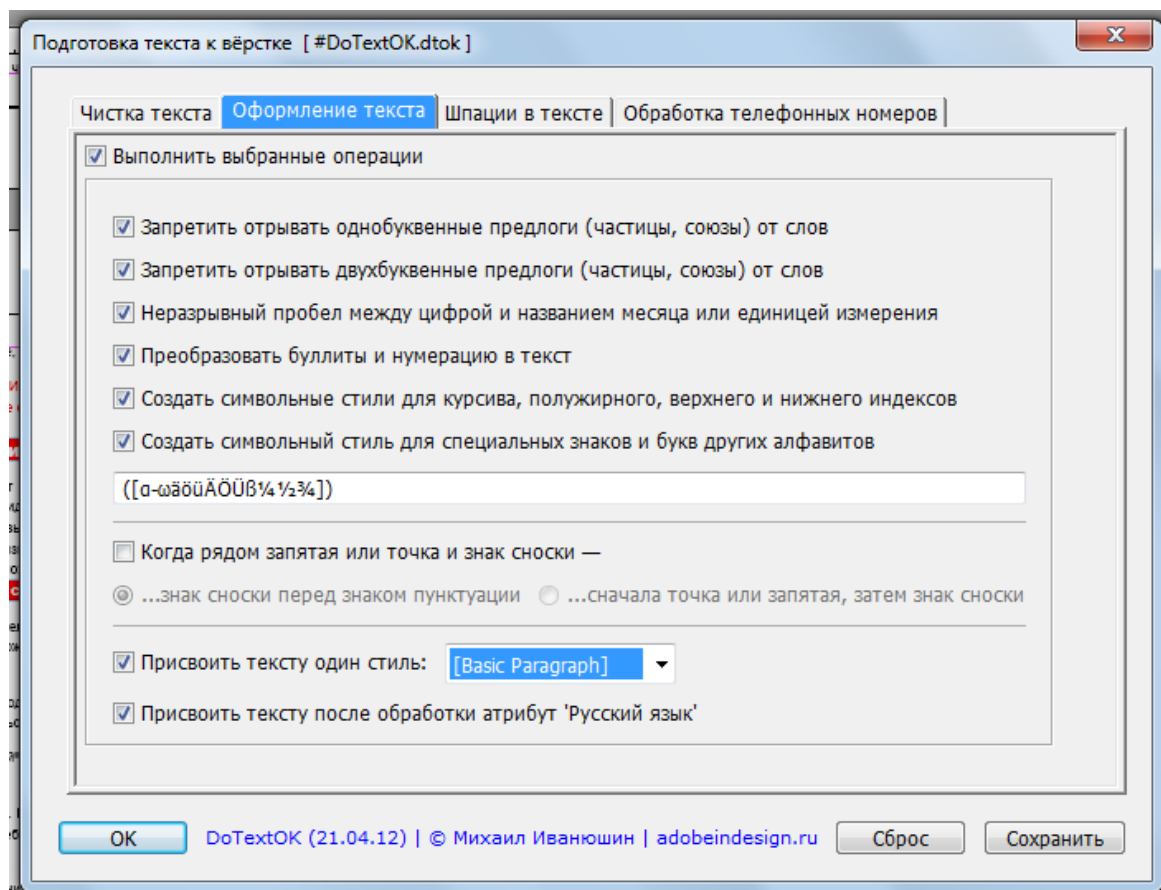
Хотя кроме текста статьи обрабатываются и сноски, и таблицы, перед исполнением скрипта курсор должен быть в именно тексте статьи. Если это условие выполнено не будет, появится сообщение:

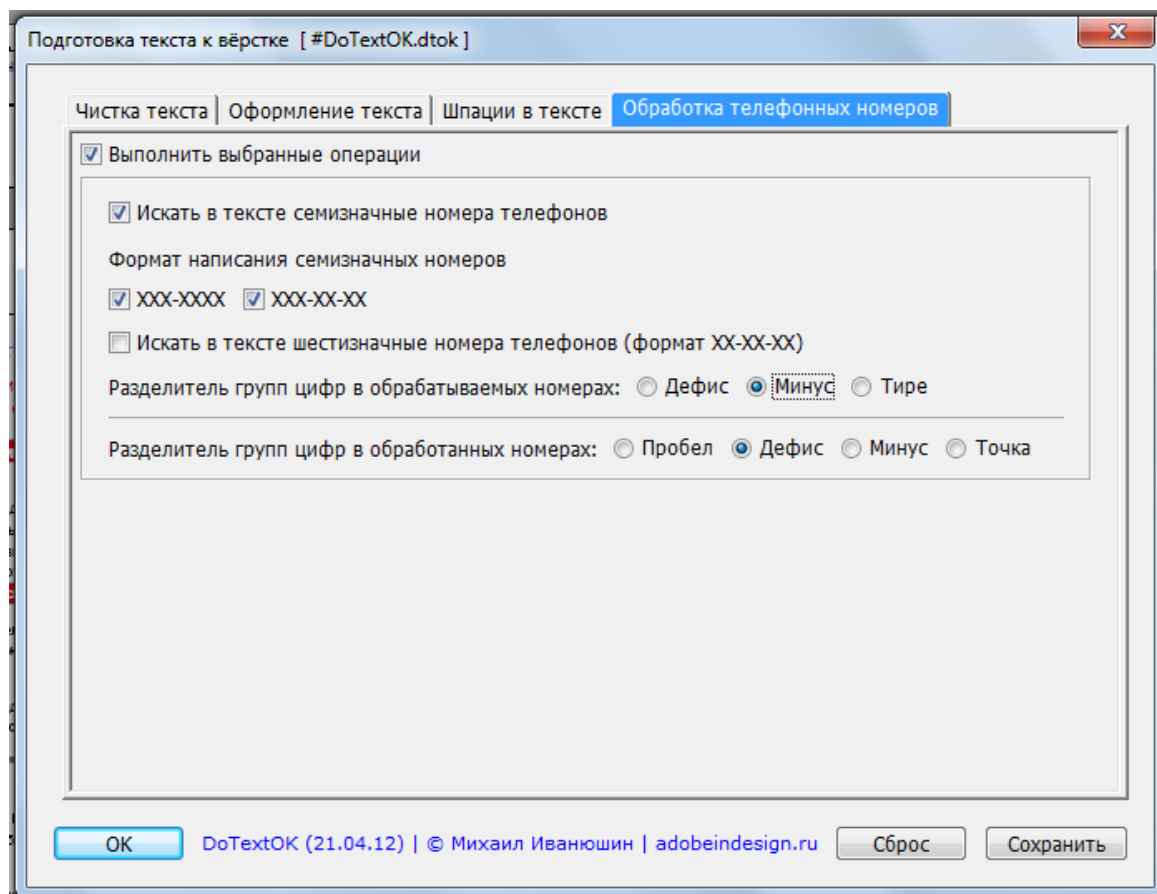


Если надо обработать только таблицу, выделите не только её, но и несколько символов перед ней.

При первом запуске программы, и при последующих, если не создавались шаблоны установок обработки, на экране сразу появится окно программы. Вот все её вкладки:





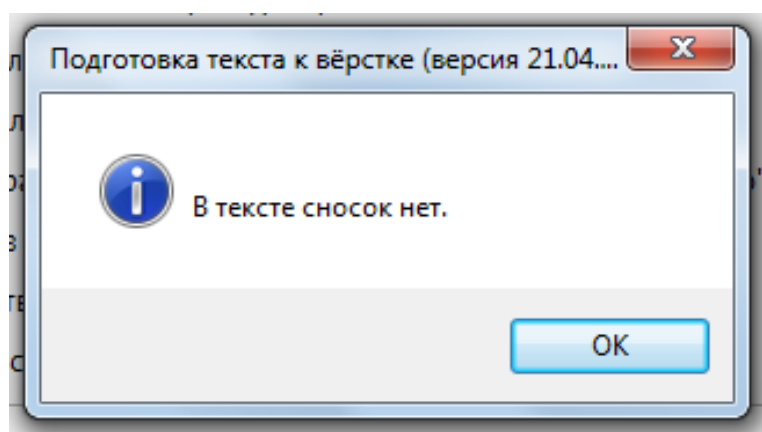


В титуле окна в квадратных скобках выводится название выбранного шаблона установок.

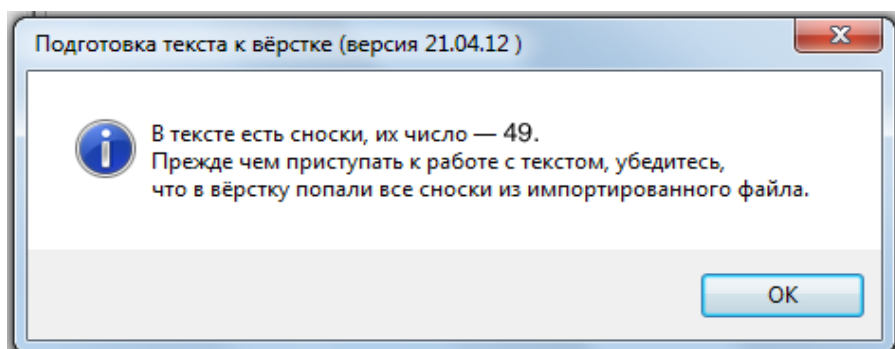
Флажок «Выполнить выбранные операции» есть на всех вкладках, он разрешает/запрещает выборку и исполнение всех операций текущей вкладки.



☐ Убедиться, что число сносок в файле такое же, как и в переданном на вёрстку тексте

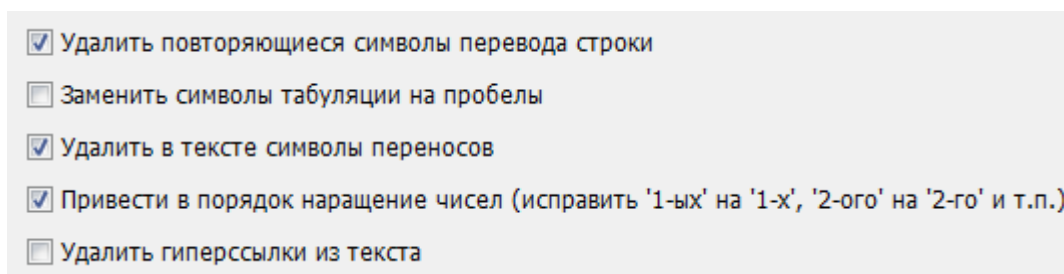


Или

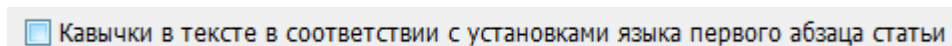


Самое время открыть текстовый файл и убедиться, что там число сносок такое же.

Если не совпадает, то пробуйте менять формат импортируемого файла. Был .doc, импортируйте .rtf, был .rtf, импортируйте .doc. Обновите версию программы.

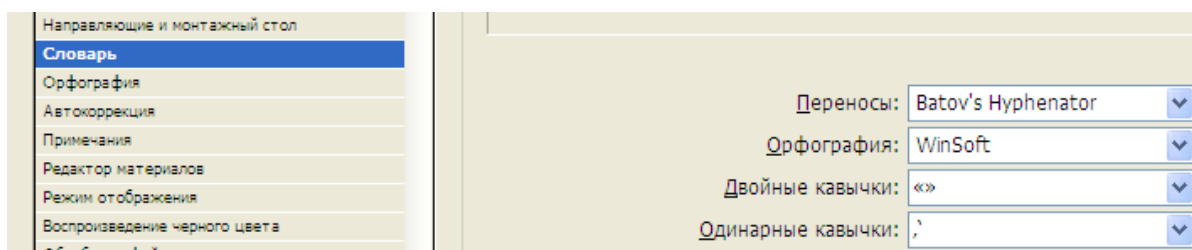


Эти опции понятны из их названий.



Если важно, чтобы кавычки были правильными, а не абы какими, то назначьте нужный язык первому символу статьи.

Какие будут кавычки, определяется настройками программы InDesign для двойных кавычек, это окно выводится на экран нажатием клавиш Ctrl+K:



Но если в работе есть вложенные кавычки, то для наведения порядка с ними воспользуйтесь скриптом DoQuotesOK.

<http://adobeindesign.ru/2009/03/12/programma-doquotesok-kontrol-kavyчек-v-tekste/>

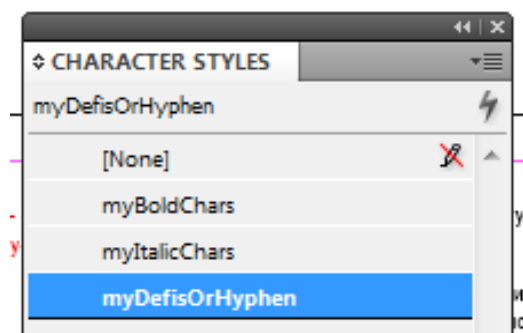
☒ Текст в формате 'одна строка - один абзац'

Таких файлов много в библиотеке Максима Мошкова <http://lib.ru/>

Вот как может выглядеть текст после импорта:

```
¶
. . . . . Когда Ethernet-кадр попадает в драйвер сетевого интерфейса Ethernet, ¶
он может быть направлен либо в модуль ARP (Address Resolution Protocol - ¶
адресный протокол), либо в модуль IP (Internet Protocol - межсетевой про- ¶
токол). . . . На то, куда должен быть направлен Ethernet-кадр, указывает зна- ¶
чение поля типа в заголовке кадра. ¶
¶
```

К сожалению, в таких txt-файлах дефисы, знаки переноса и тире отображаются одним символом. Поэтому в процессе преобразования этих однострочных абзацев в нормальный текст всем найденным в конце строки дефисам назначается символьный стиль myDefisOrHyphen, чтобы потом просмотреть эти отмеченные этим стилем символы и принять решение по каждому из них.



Назначение этих опций очевидно.

Тире в тексте	Разделитель цифр в интервалах дат и чисел
<input checked="" type="radio"/> Длинное тире	<input checked="" type="radio"/> Минус
<input type="radio"/> Короткое тире (минус)	<input type="radio"/> Тире
Вариант сокращения: с точкой или без точки <input type="radio"/> млн, млрд, трлн <input checked="" type="radio"/> млн., млрд., трлн.	

Оформление текста

☒ Запретить отрывать однобуквенные предлоги (частицы, союзы) от слов

Если этот флажок установлен, вот эти однобуквенные частицы будут привязаны к слову рядом.

```
////
// SpaceLetterNonbreackingSpace - однобуквенные частицы, которые не будут оторваны от следующего слова
var SpaceLetterNonbreackingSpace = "(a|a,|в|и|и,|к|о|с|у|с.т.п.ч.|\"O|\"K)"; // Две последние прописные буквы.
// NonbreackingSpaceLetterSpace - однобуквенные частицы, которые не будут оторваны от предыдущего слова
var NonbreackingSpaceLetterSpace = "(б|ж)";
```

☒ Запретить отрывать двухбуквенные предлоги (частицы, союзы) от слов

Если этот флажок установлен, вот эти двухбуквенные слова будут привязаны к слову рядом.

```
////  
// SpaceTwoLetterNonbreackingSpace - двухбуквенные частицы, которые не будут оторваны от следующего слова  
var SpaceTwoLetterNonbreackingSpace = "(не|ни|на|ну|ну,|от|об|из|за|да|да,|но|но,|по|до|во|со|ко|та|ту|то|те|см.|им." +  
"|He|Hi|Ha|Hy|Hy,|От|Об|Из|За|Да|Да,|Но|Но,|По|До|Во|Со|Ко|Та|Ту|То|Те|См.|Им.|\\\"Об)";  
// NonbreackingSpaceTwoLetterSpace - двухбуквенные частицы, которые не будут оторваны от предыдущего слова  
var NonbreackingSpaceTwoLetterSpace = "(ли|же|ль|бы|бы,|же,)";
```

☒ Неразрывный пробел между цифрой и названием месяца или единицей измерения

Если флажок установлен, единицы измерения и названия месяцев будут привязаны к числам.

```
////  
// привязываемые к числу единицы измерений и названия месяцев  
myUnits = "(\\d)( )(?|кг|мм|дм|см|м|км|л|В|А|Вт|W|°C|января|февраля|марта|апреля|мая|июня|июля|августа|сентября|октября|ноября|декабря)";  
////
```

Все эти данные размещены в начале скрипта, и их можно редактировать.

☒ Преобразовать буллиты и нумерацию в текст

Назначение этого флажка очевидно.

☒ Создать символьные стили для курсива, полужирного, верхнего и нижнего индексов

Тоже понятный флажок.

☒ Создать символьный стиль для специальных знаков и букв других алфавитов

([α-ωäöüÄÖÜß¼½¾])

Этот флажок в процессе работы изменил своё название. Поначалу он назывался «Создать символьный стиль для греческих букв». С его помощью искали символы в диапазоне α-ω. Потом добавился поиск любого знака гарнитуры Symbol и присвоение найденным символам этого стиля.

Проблема сохранения в вёрстке греческих букв была решена, но терялись ещё и немецкие буквы с умляутиками. И были файлы с дробными. Это и натолкнуло собрать все такие специальные знаки в один массив. В результате изменилось название опции и добавлена возможность пользователям самим добавлять свои символы, которые важно не потерять. Эта строка редактируемая, главное, чтобы искомые символы были внутри круглых и квадратных скобок.

☒ Когда рядом запятая или точка и знак сноски —
☒ ...знак сноски перед знаком пунктуации ☐ ...сначала точка или запятая, затем знак сноски

Этот флажок доступен, если в работе есть сноски. Вопрос в том, как всё же должны стоять рядом точка/запятая и знак сноски. С одной стороны, все руководства по правилам вёрстки предписывают ставить знаки препинания после знака сноски. И никого не волнует, что между последней буквой слова и точкой, под знаком сноски появляется дыра, соизмеримая с шириной пробела.

И вот в 2006 году выходит на русском языке книга Р. Брингхёрста «Основы стиля в типографике». Под редакцией Владимира Ефимова.

И как же там изящно смотрятся сноски! Вот скан части страницы 160:

Рисунок шрифтов для ли
основными факторами. Во-пе
ментов знаков невозможно бе
риц (поэтому у курсивной *f*
чахлая головка и хвост). Во-е
лится только на 18 единиц, чт
ций знаков.³³ И в-третьих, к ←
обычно спарены. Поэтому в б
каждой курсивной буквы долж
мого аналога.³⁴ ←

Многие шрифты, разрабо
успех, несмотря на эти ограни

Представьте, как убого смотрелись бы дыры под сносками 33 и 34, если эти знаки сносок сделать по правилам. В общем, каждый решает сам. Надо знать правила, чтобы иногда сознательно отходить от них.

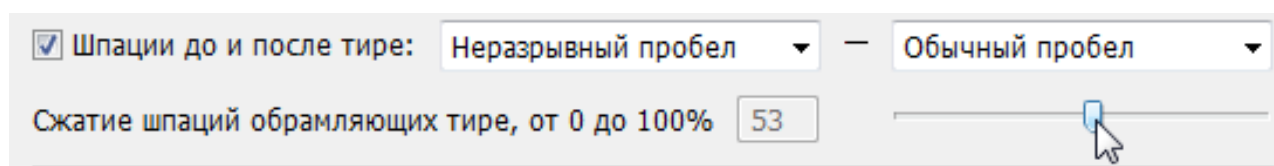
☒ Присвоить тексту один стиль: [Basic Paragraph] ▼
☒ Присвоить тексту после обработки атрибут 'Русский язык'

Эти флажки понятны без объяснений.

Шпации в тексте

☒ Шпации до и после тире: Неразрывный пробел — Обычный пробел

Сжатие шпаций обрамляющих тире, от 0 до 100% 53



Если хочется сделать окружающие тире отбивки менее широкими, то этот инструмент поможет. Имеющиеся в индизайне различные шпации все неразрывные. Поэтому тире, обрамлённое такими шпациями, никогда не будет в конце строки. Так что лучшее решение — использовать не шпации, а пробелы, и ползунком (на него на картинке указывает курсор) установить нужный процент сжатия.

Эти установки понятны без объяснений.

☒ Между запятой, точкой, многоточием и тире: Волосяная шпация

☒ В составных сокращениях (т.д., т.е., т.к., ...): Волосяная шпация

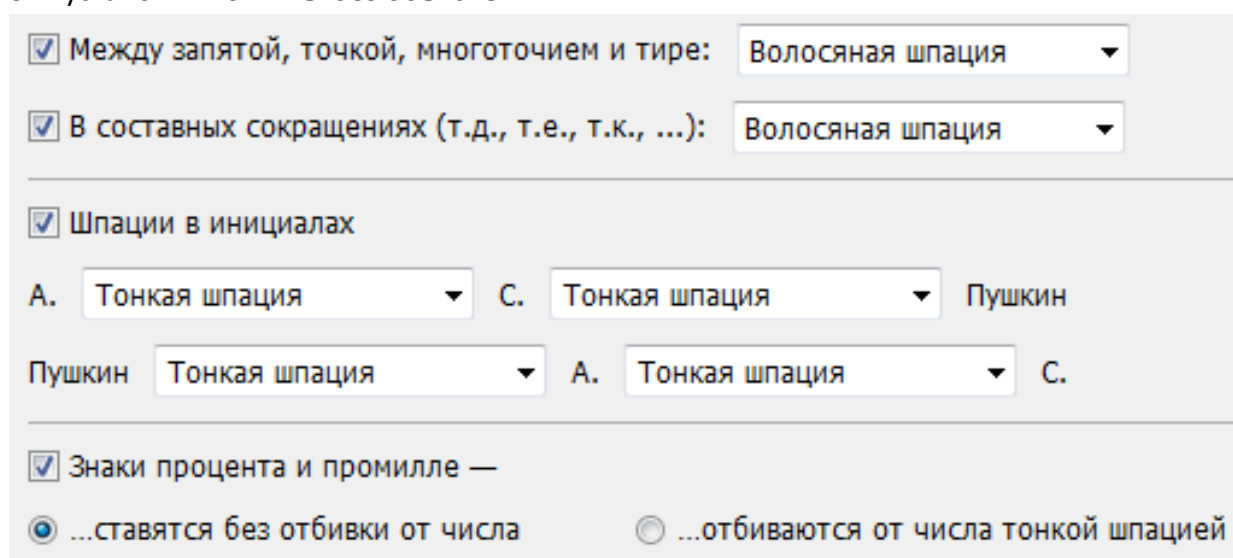
☒ Шпации в инициалах

A. Тонкая шпация C. Тонкая шпация Пушкин

Пушкин Тонкая шпация A. Тонкая шпация C.

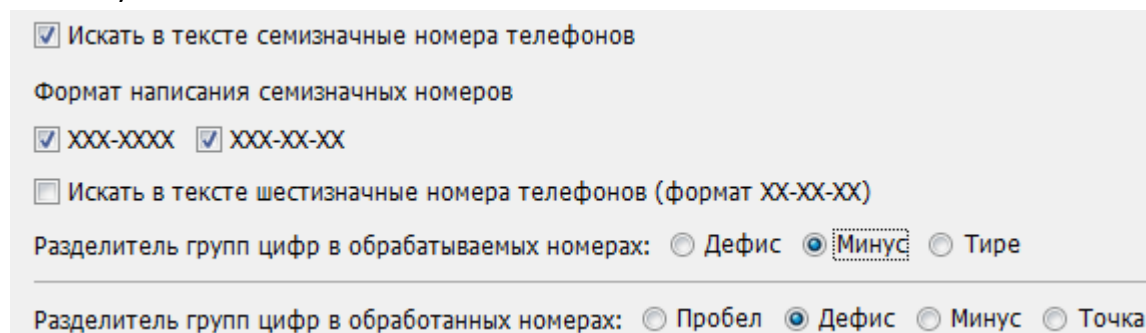
☒ Знаки процента и промилле —

☒ ...ставятся без отбивки от числа ☐ ...отбиваются от числа тонкой шпацией



Обработка телефонных номеров

Все эти установки понятны из названий.

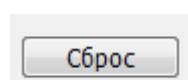
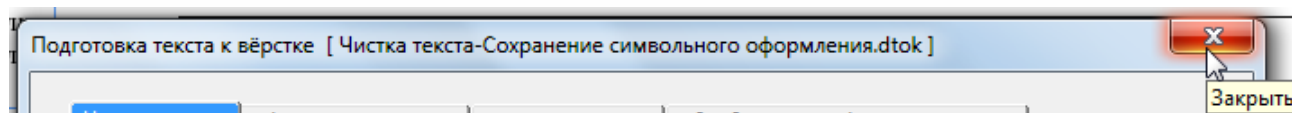


Скриншот окна настроек «Обработка телефонных номеров». В нем есть следующие элементы:

- Чекбокс ☒ Искать в тексте семизначные номера телефонов
- Группа чекбоксов «Формат написания семизначных номеров»:
 - ☒ XXX-XXXX
 - ☒ XXX-XX-XX
- Чекбокс ☐ Искать в тексте шестизначные номера телефонов (формат XX-XX-XX)
- Группа радиокнопок «Разделитель групп цифр в обрабатываемых номерах»:
 - ☐ Дефис
 - ☒ Минус
 - ☐ Тире
- Группа радиокнопок «Разделитель групп цифр в обработанных номерах»:
 - ☐ Пробел
 - ☒ Дефис
 - ☐ Минус
 - ☐ Точка

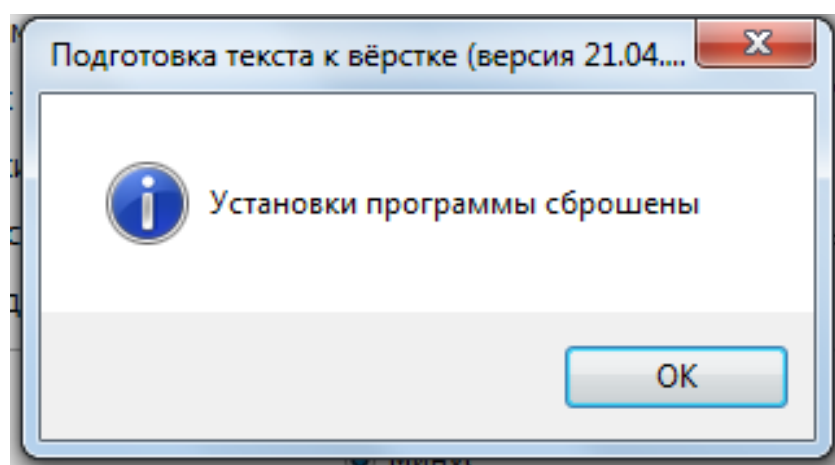
3. Управляющие кнопки

Красный крестик в верхнем правом углу используется как кнопка отмены ('Cancel')



при нажатии на эту кнопку установки будут сброшены. Обновляются данные всех вкладок, кроме «Обработки телефонных номеров».

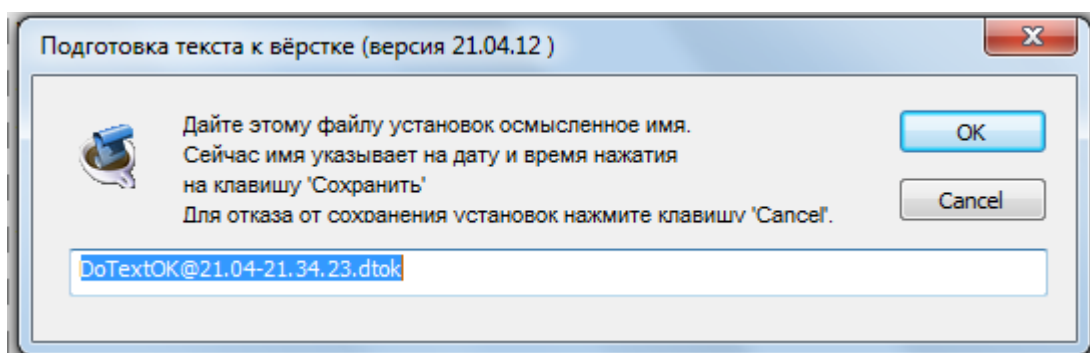
На экран будет выведено такое сообщение:



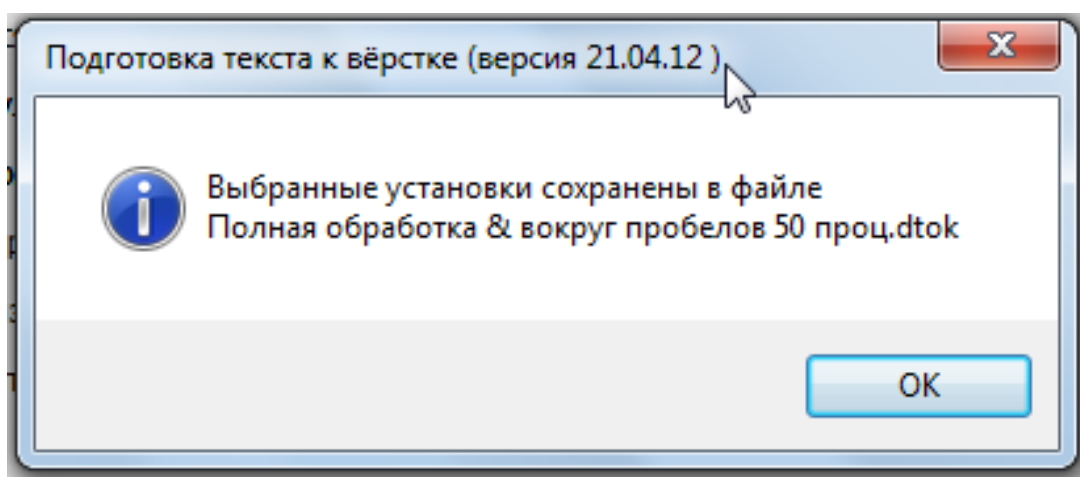
Сохранить

Текущие установки сохраняются в отдельном файле.

По умолчанию в имени файла указываются дата и время создания файла.



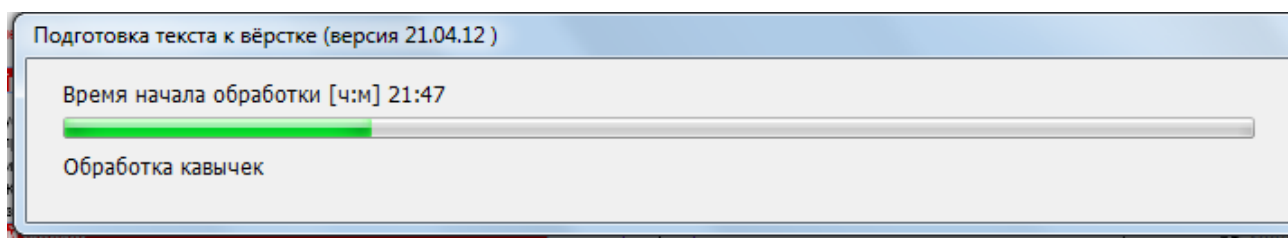
После ввода имени и нажатия на кнопку OK на экране будет сообщение:



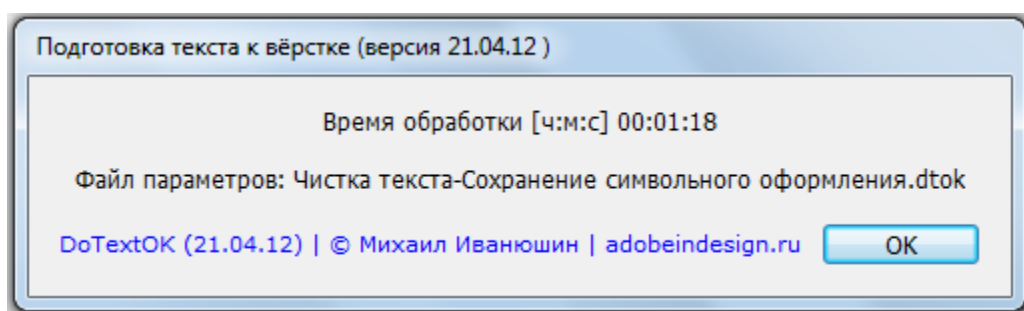
OK

Эта кнопка для запуска процедуры обработки текста. Текущие установки сохраняются в файле #DoTextOK.dtok.

После нажатия на эту кнопку появляется окно, в котором есть информация о времени начала обработки, и под прогрессбаром отображается, что выполняется сейчас.

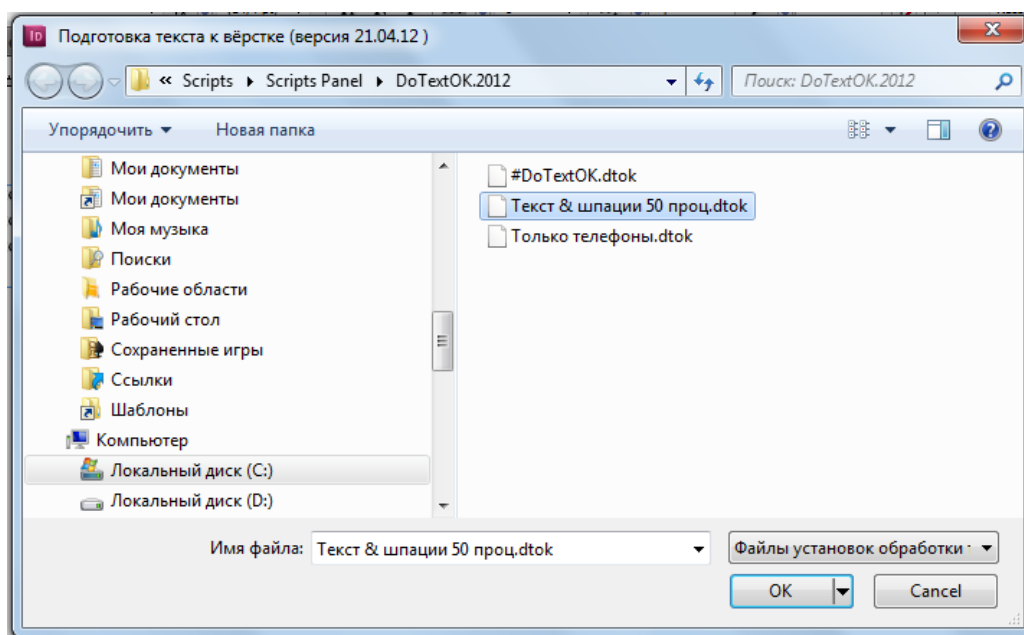


После завершения обработки выводится окно с информацией о времени обработки и использованном шаблоне установок.



4. Выбор шаблона настроек

При первом запуске или при запусках без создания новых шаблонов файл установок один, и выбор не нужен, поэтому программа сразу выводит окно установки параметров. Если шаблонов несколько, то появляется окно выбора данных. Эти шаблоны могут быть полезны фрилансерам, вынужденным приспосабливаться к несовпадению понимания оформления текста разными заказчиками.



В окне можно переименовывать и удалять файлы. Название выбранного файла будет отображаться в титуле окна программы, и в сообщении о времени выполнения задания.

После обработки последние установки будут сохранены в файле #DoTextOK.dtok.

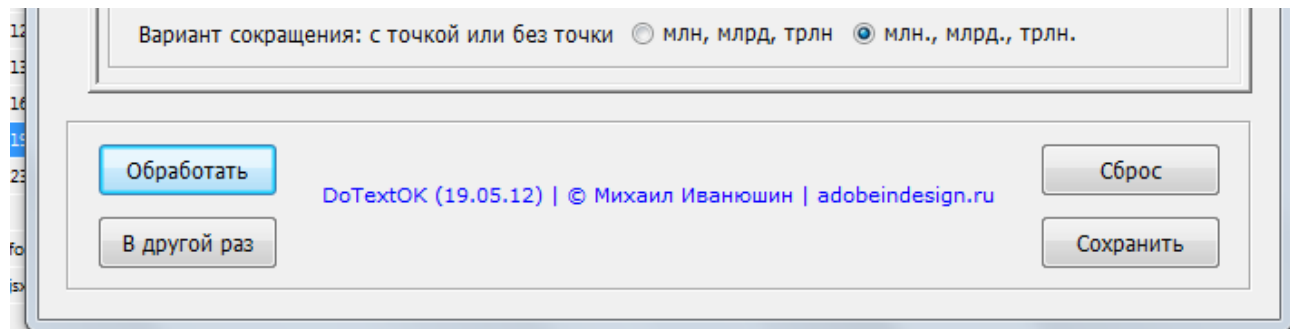
Щелчок на клавише Esc или кнопке Cancel эквивалентен выбору файла #DoTextOK.dtok с установками последней обработки.

Михаил Иванюшин
ivanyushin#yandex.ru

История изменений

К 19.05.2012 было сделано следующее:

= Изменился вид и число кнопок основного окна:



= Массив одно- и двухбуквенных частиц выделен в отдельный файл WordsAndUnits.jsx.

Идея в том, что пока этого не было, после любого обновления скрипта приходилось бы переносить в скрипт все свои наработки – частицы и единицы измерения, которые не надо отрывать. Файл формата jsx, чтобы его можно было править в непосредственно отладчике.

= Скрипт доработан в плане удаления переводов строки и пробелов в начале и конце содержимого всех ячеек таблиц.

= Если в тексте есть сноски, им назначается символьный стиль myFootnotes. В этом стиле определено только то, что символы будут в верхнем регистре. Это гарантирует, что сноски визуально не потеряются при переопределении стиля абзаца текста. Это стилевое оформление прикладывается как к номерам сносок в тексте, так и к номерам в блоках самих сносок.

= В первых релизах гиперссылки удалялись во всем документе, независимо от того, выбран весь документ или выделена его часть. В результате, если работать с текстом частями – обработать часть, поставить где нужно гиперссылки, при переходе к другому фрагменту ранее созданные гиперссылки терялись. Теперь гиперссылки удаляются только из выделенного фрагмента.

= В результате переписки с пользователями решено сделать начало диалогов с длинного тире, независимо от того, какое тире выбрано в программе. После этого тире идёт полукруглая шпация.

= Если в тексте есть единицы измерения с дробной чертой: ц/га, руб./долл. и пр., этой дробной черте назначается атрибут No Break. Благодаря этому эти единицы измерения не разрываются.

= при обработке текста первыми релизами замечено неожиданное появление неразрывного пробела между числом и словом 'городов': 10^городов.

Причиной этому было то, что программа первую букву этого слова принимала за 'г' из массива myUnits. Изменение этого массива позволило избавиться от этой ошибки.

```
myUnits = "(\\d)( )(руб|ц/га|г[.,]|кг|мм|дм|см|м[.,]|км|л[.,]|В|А|Вт|
```

Очевидно, что 'м' и 'л' тоже могут быть так же обработаны. Поэтому эти запись этих единиц измерения тоже изменена.

Конечно, желательно хоть немного понимать инструмент грер, чтобы понимать, что исправляется.

```
В частности, добавлены падежи слова 'квартал': |декабря|квартал\\.);
```

```
'\.' – это любой знак.
```

= Исправлена неверная обработка апострофов в английских текстах. До этого I'm / you can't превращалось в I»m / you can»t. Осталась одна «заноза» английской грамматики: там нельзя логически разделить, одиночная кавычка после s — это апостроф или все-таки кавычка.

Это встречается редко, вот примеры: Girls' Generation / Boys' Brigade / Boys' or Girls' Snow Flurry Boots.

Но раз такое теоретически возможно, эту ситуацию надо отлавливать. Поэтому после s ставится принудительно апостроф и он красится цветом myInfoColor.
