**Как подготовить шаблон Word документа для работы**

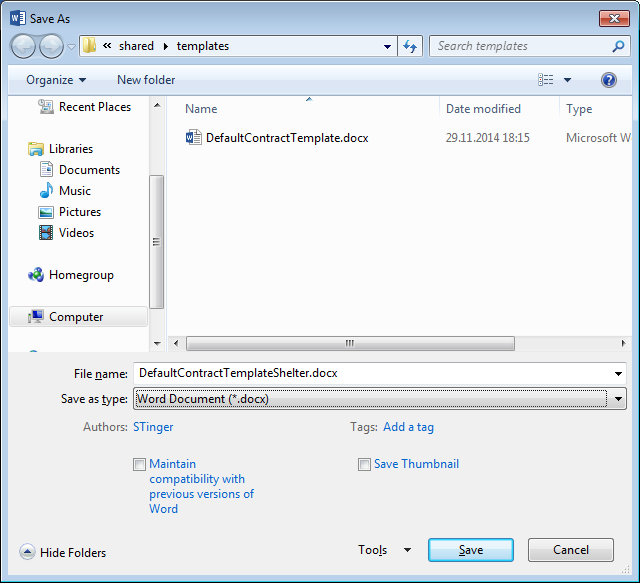
1. Все шаблоны должны быть в формате .docx, т.е. MS Word 2007 и выше
2. MS Word вставляет тонны левых тегов в созданные .docx файлы, поэтому некоторые длинные метатеги, которые мы будем заменять, бьются на несколько частей и в реальности замена происходить не будет. Чтобы это полечить, каждый созданный шаблон необходимо «исправить» вручную. Это однократная операция в том случае, если наши метатеги не будут меняться в будущем (к ним не будут применяться операции форматирования текста), в противном случае, операцию придется повторить после подобных изменений.

Теперь о том, как это выглядит:

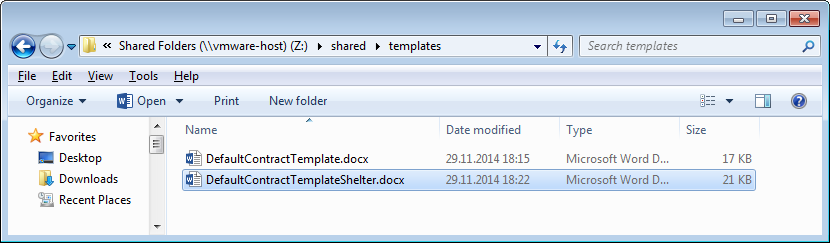
Нам потребуется заранее установленные архиватор WinRar (крякать не нужно) и текстовый редактор Noptead++.

Напоминаю, что шаблон должен быть сохранен в формате .docx!!!

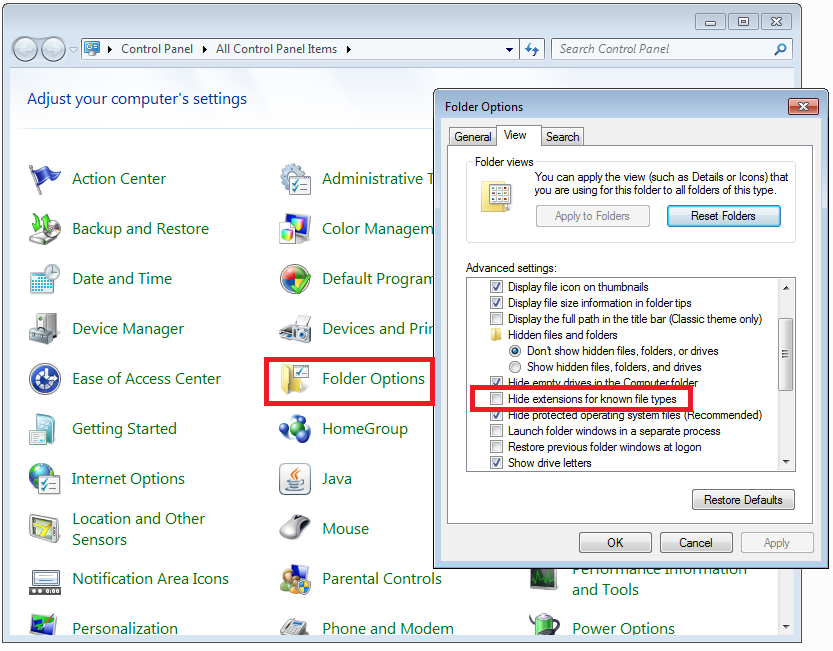
Лучше всего производить изменения во временной директории, скопировав туда исходный документ и после модификаций, забрав оттуда новый.



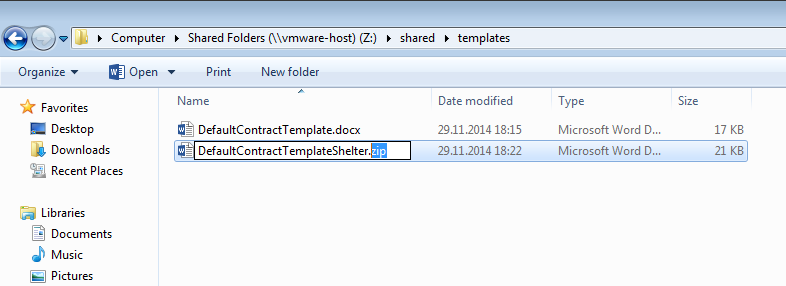
Закрываем шаблон, если он открыт и идем в ту папку, где он лежит.



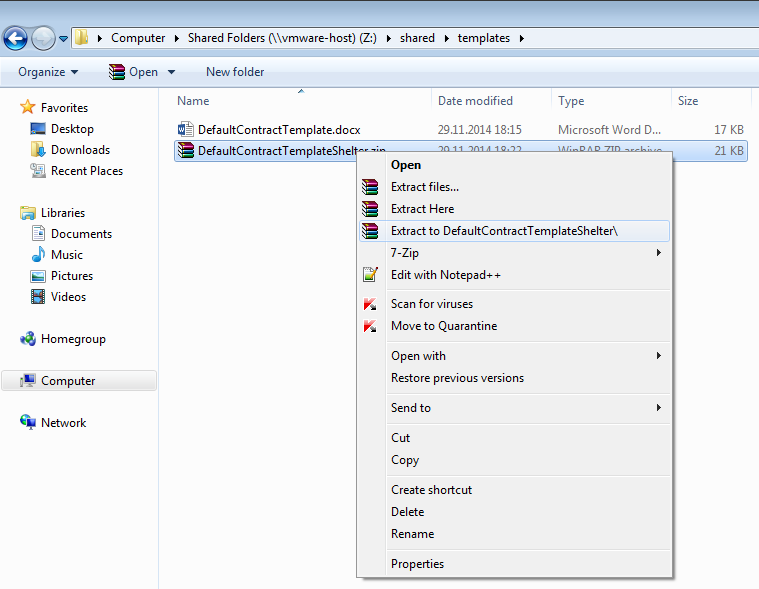
Здесь важное отступление: на компьютере нужно включить отображение расширений для известных типов файлов, иначе хвост «.docx» отображаться не будет. Это можно сделать в Панели Управления -> Настройка папок (или как-то так), снять там соответствующий чекбокс и сказать ОК. Для англоязычной винды это выглядит так:



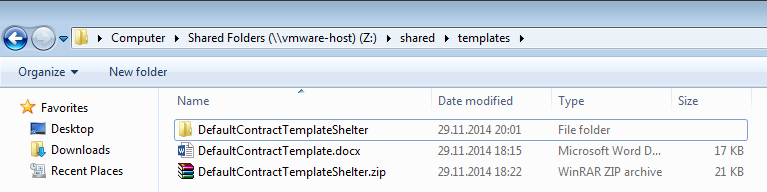
Затем мы выделяем нужный нам документы и нажимаем <F2> в Проводнике для переименования (или правый клик + Переименовать…) и переименовываем хвост файла в .zip. На предупреждение забиваем.



Здесь мы его и распаковываем. DOCX – это упакованный в ZIP архив документ DOC, поэтому, его можно распаковать обычным архиватором. Щелкаем правой кнопкой по нему, выбираем распаковку в поддиректорию с этим же именем:

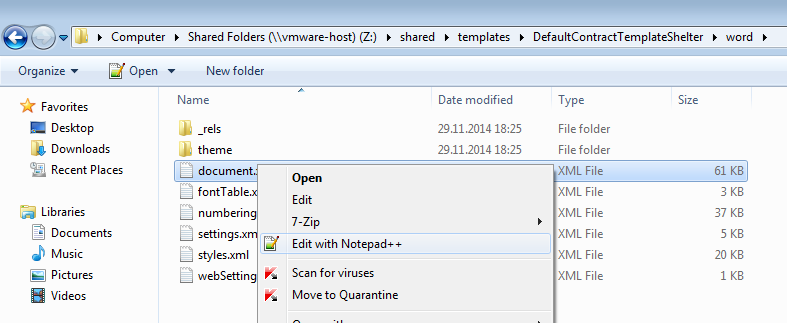


Видим

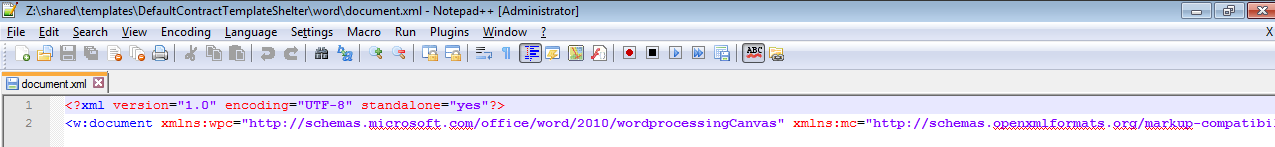


Заходим в распакованный шаблон, в подпапку word. Там находим файл document.xml, который содержит основную разметку документа и его текст с XML формате.

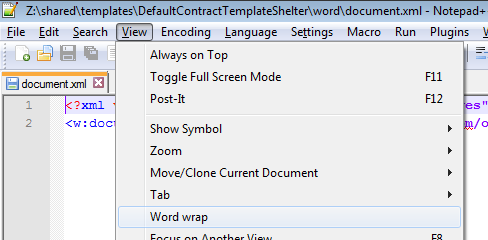
Открываем его с помощью установленного Notepad++:



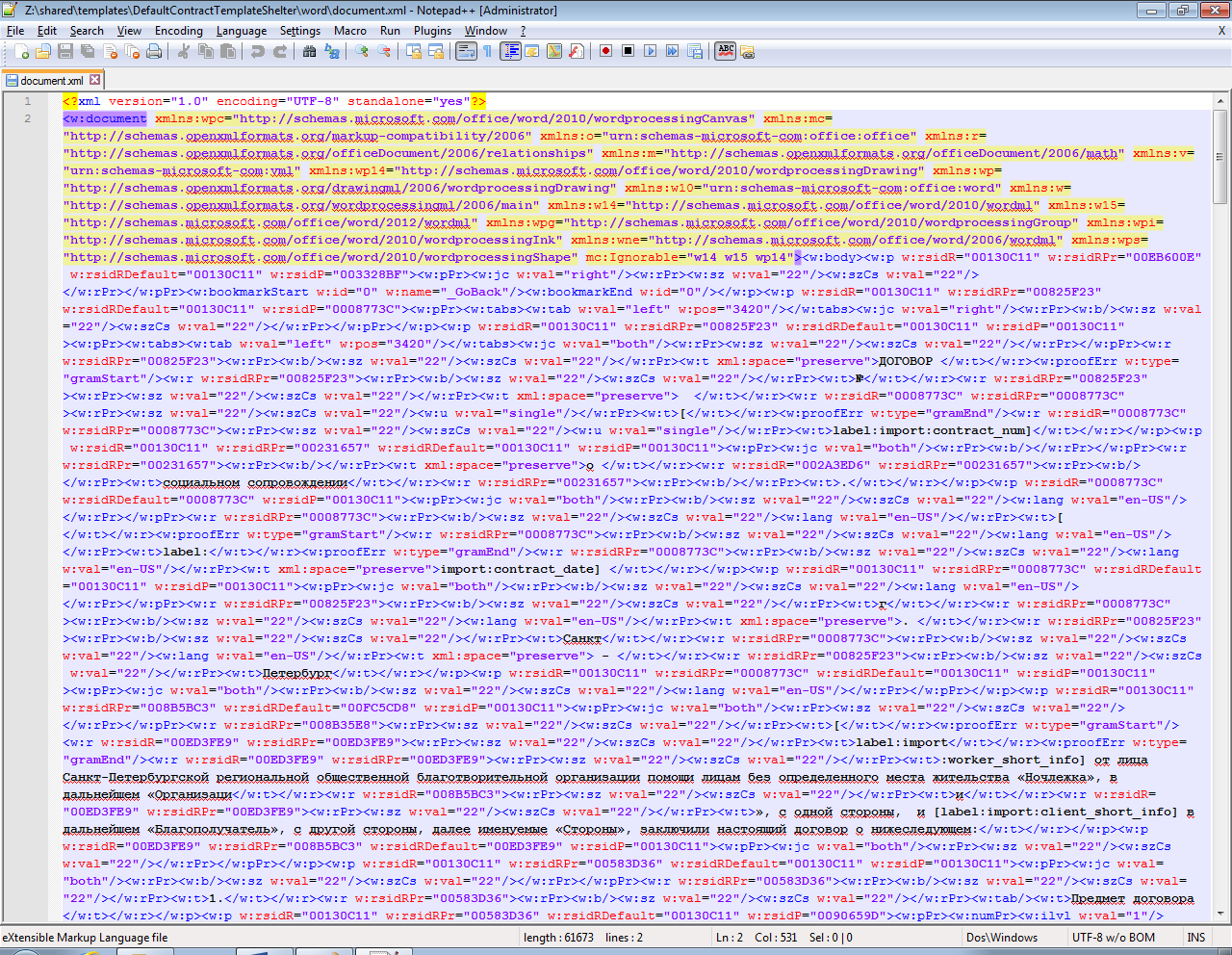
И видим:



С этим работать неудобно, поэтому включаем перенос по словам (Word Wrap) в меню View (Вид):



И видим на экране весь XML текст документа:



Прямо на этой картинке (на этом примере) мы видим то, как MS Word воспринимает текст с метатегами, например, [label:import:contract\_num]: [ ***мусор*** label:import:contract\_num ***мусор*** ]

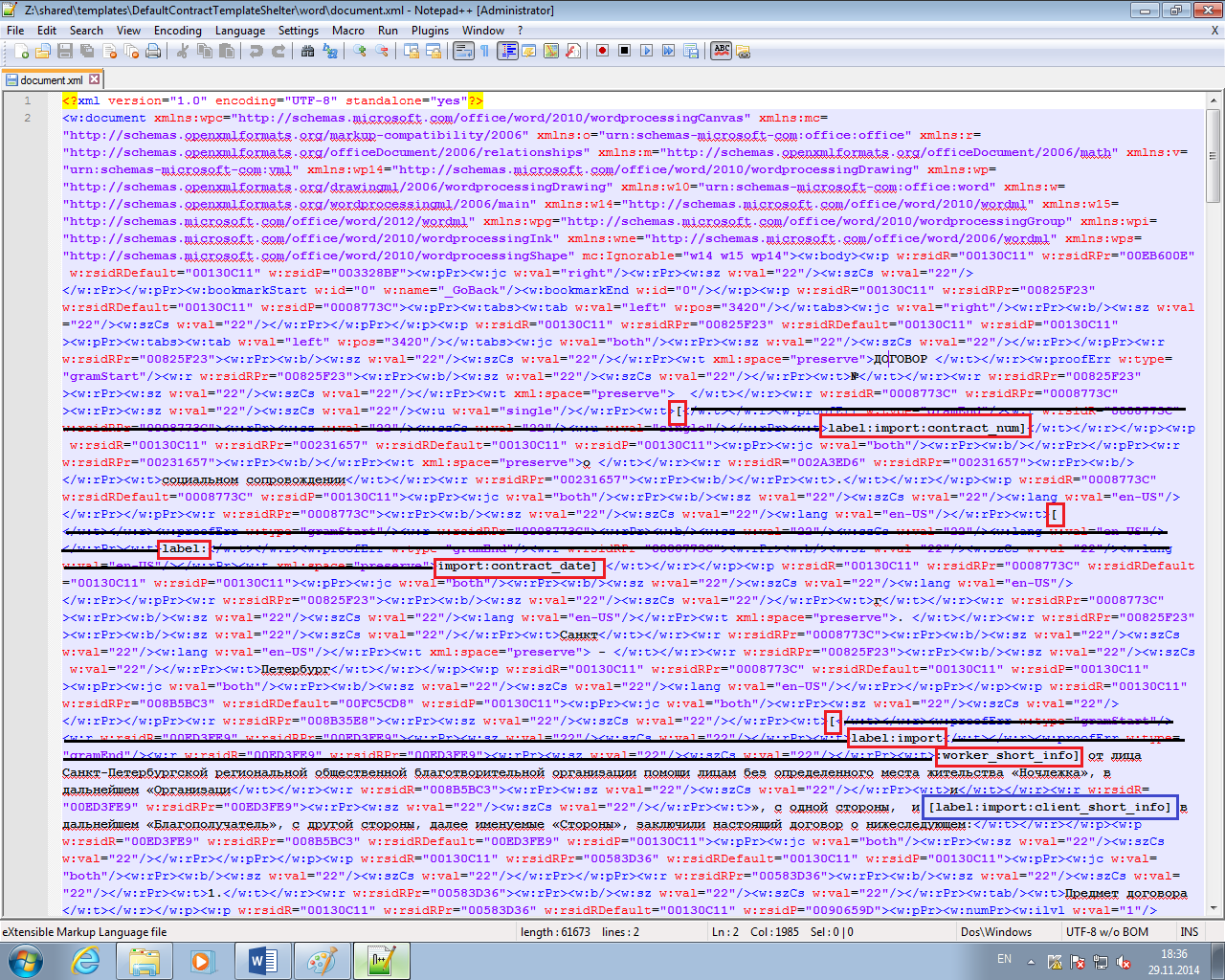
Движок формирования документов, к сожалению, не во всех случаях может правильно распознать эти метатеги, т.к. оперирует отдельными словами и, если к нем разрозненно применены разные стили, то разбивка на слова не работает правильно. Если приведенный пример будет заменен в 60% случаях, то [ ***мусор*** label: ***мусор*** import:contract\_date] – вообще никогда. Увы, это ограничение движков Apache POI и docx4j и повлиять на это мы пока никак не можем, они еще развиваются.

Собственно, к чему речь: проходим **по всему** документу и склеиваем все [метатег], нещадно вырубая все то, что встречается между ними. После чего, сохраняем. См. пример ниже (синим прямоугольником отмечено то, что уже и так хорошо, не трогаем).

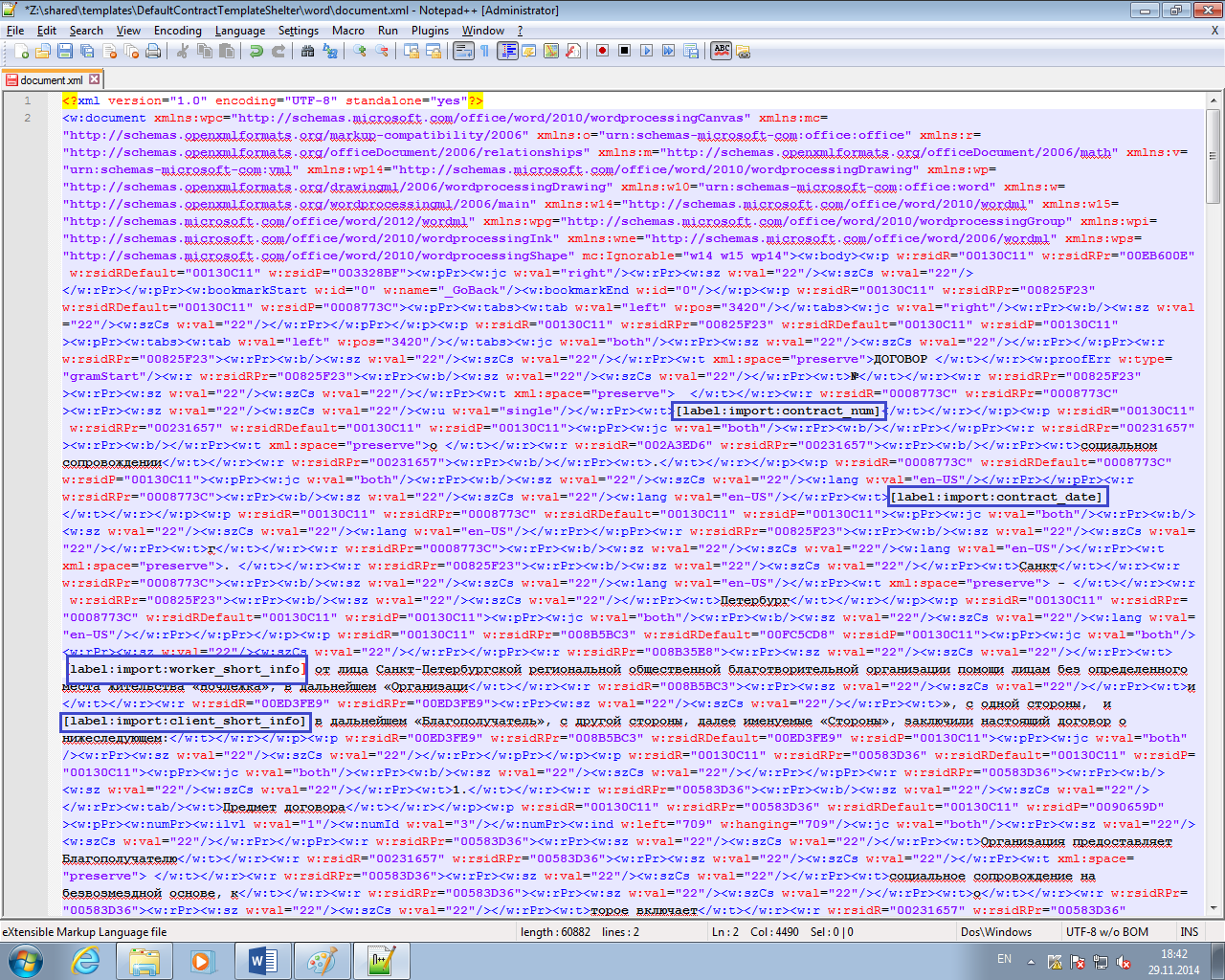
Собственно, Notepad++ используем потому что он замечательно подсвечивает весь XML и очень легко отделить обычный текст (черный) от тегов (синие), атрибутов (красные) и их значений (фиолетовые).

Зачеркнутый текст – это тот текст, который надо удалить.

**Правим только метатеги, остальной текст не трогать!!!**

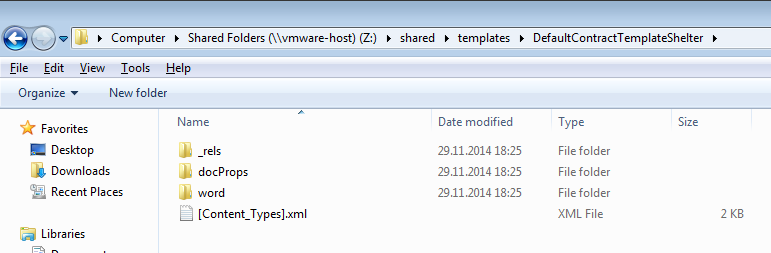


Должно получиться как на следующей картинке:



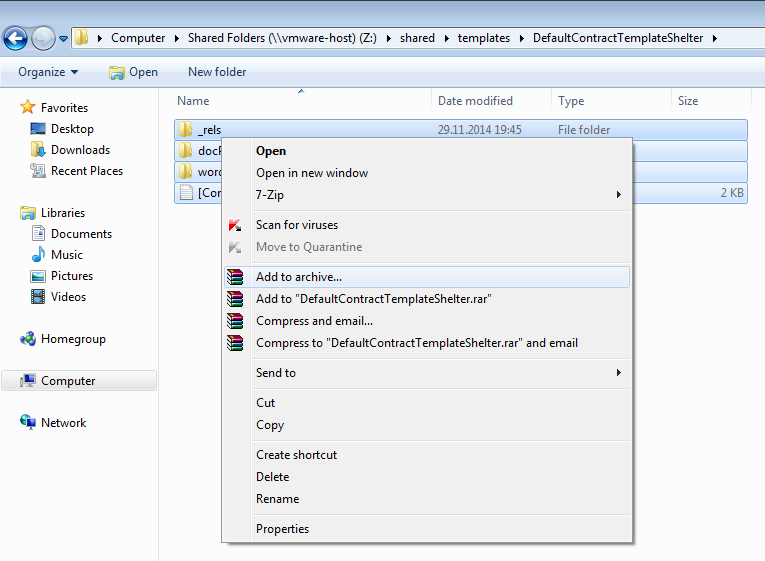
Отлично! Мы прошли по всему документу до самого конца и поправили все несоответствия. Теперь все метатаги выглядят как цельные слова. Закрываем document.xml с сохранением.

Теперь это надо запаковать назад. Выходим чуть выше, в распакованные кишки документа, чтобы это выглядело так:

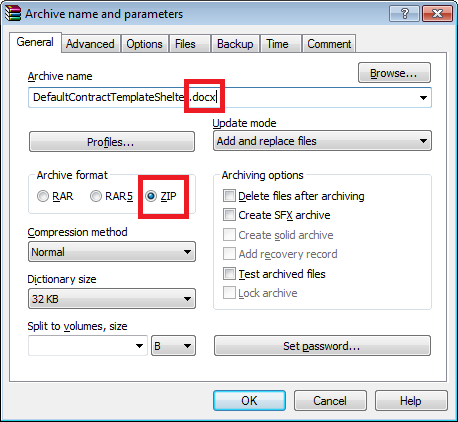


Затем выделяем все файлы протягиванием мышки (или нажатием кнопок Ctrl+A на клавиатуре, или через Shift/Ctrl и мышкой, короче, любым известным способом) и кликаем по выделенному правой кнопкой:

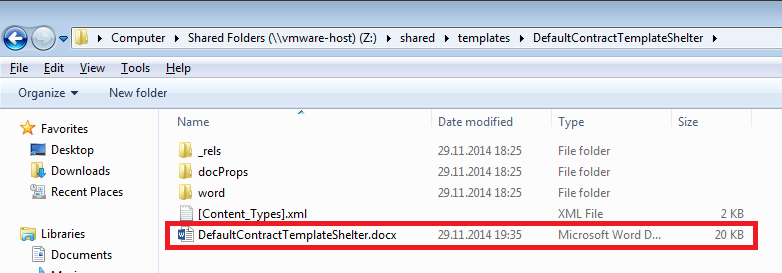
Нас интересует добавление в новый архив выделенных файлов и директорий с помощью WinRar:



Меняем тип архива с rar на zip и правим имя файла так, чтобы оно было имя\_файла.docx:



Ну и жмем ОК. Архив создается, видим рядом новый «документ MS Word». Открываем, проверяем что он нормально открывается.



Вырубаем новый документ куда нужно (например, на место старого документа), а все, что было распаковано ранее и переименованный старый документ удаляем, так как уже не нужны.