**ТЗ на то, что нужно запрограммировать за ближайший месяц**

**(в широком понимании слова «месяц» =) )**

Статусы:  
Не готово  
Сделано **(комментарий)**  
В процессе **(комментарий)**  
Не применимо (выполняется с другим пунктом / происходит само) **(комментарий)**

1. BackEnd
   1. Корректное открывание/закрывание тегов в клонах и контексте при изготовлении (выделении) информационных элементов
   2. Проверка на пересечение с уже выделенными клонами
   3. Убирание фильтрации самопересечений
   4. Проверка на мусорные клоны «is a», «there is» и т.д.  
      **(появился задел для использования средств обработки естественных языков;  
      словарь составляется)**
   5. Проверка всего процесса на соответствие статье (может, еще что-то забыли)
   6. Модуль рефакторинга -- генерации самих предложенных информационных элементов, которые выбрал пользователь, и замещения их вхождение в текст ссылками на них
      1. Для информационных элементов без точек расширения
      2. Для информационных элементов с очками расширения
   7. Поддержка выделения группы клонов в словарь (если есть – проверить, что работает)
2. FrontEnd
   1. Достаточно удобный браузер. Скорее всего похожий на ту раскрашенную HTMLку, которую я тебе присылал (по крайней мере мне она нравится), с возможностью отметить информационные элементы для утверждения.
   2. Поддержка средств манипуляций с одиночными группами и с группами-парами **(уточнить что это за функциональность)**
   3. Поддержать просмотр клонов в тексте документа (из браузера)  
      **(можно считать подпунктом 2.1)**
3. Интеграция
   1. Вызов из DocLine в виде мастера.  
      **(в процессе интеграция)**
   2. Публикация в репозитории. Там сейчас есть одна из предыдущих версий, надо будет обновить и дополнить.  
      **(заведён отдельный репозиторий, установлена связь между репозиториями для автоматического обновления)**
4. Веб-страничка
   1. Отдельно написать по-русски и по-английски про новую функциональность с поиском клонов (несколько предложений, найти подходящее место на сайте)
   2. Выложить новую версию DocLine вместе с нашим тулом на сайт  
      **(автоматически обновляется в репозитории по мере разработки в результате 3.2)**
5. Эксперименты
   1. Провести полноценные эксперименты на Java-документах  
      **(после окончания «своих» тестов займётся Екатерина Клочкова)**

1. Перспектива
   1. Поддержка многофайловых документов
   2. Ускорить (если получится, распараллелить) реализацию
   3. искать комбинации из N групп (с N-1 extension points).
   4. Переработать алгоритм – как?
   5. Курс на собственный искатель клонов (заодно разобраться в нижнеуровневых алгоритмах).  
      **(занимается Александра Михайлова)**