Текст для презы

Добрый вечер! Мы - люди до этого незнакомые - объединились в один коллектив, чтобы представить проект “Отечественные записки: головная боль для цензора, «чума» для ретроградов”. Прежде, чем мы перейдем к сути презентации позвольте полить немного воды, чтобы уложиться во временной промежуток.

Журнал «Отечественные записки» - колыбель русской демократической мысли, настольное чтиво отцов русской демократии и особ приближенных к императору, колыбель русского литературоцентризма. Вырвал звание самого толстого журнала 19 века у ньюскульного современника.

Почему звание толстый журнал это престижно:

1. Журнал заменял газету, рассказывая о значимых событиях в политической, экономической и социальной жизни страны.

2. научную энциклопедию, сообщая читателям знания, малодоступные рядовому гражданину.

3. художественную книгу, т.к. большая часть литературных произведений публиковалась прежде в журналах, и лишь позже, в случае успеха, они выходили как самостоятельные издания.

Ну и суммируя все вышеперечисленные достоинства, было куда выгоднее купить Том Отечественных записок, чем покупать три других более дорогих книги.

Теперь что касается, ТЗ

Главная идея, на которой завязан финальный продукт - это создать цифровое издание самого знакового для русской культуры журнала. Что нам для этого понадобится:

* сканы страниц в старой орфографии
* Fine Reader
* кофе
* сканы с распознанным текстовым слоем в старой орфографии с Google Books
* снова кофе
* текст со слоем нормализованной орфографии

Как минимум мы планируем:

- создать таблицу с информацией по каждому выпуску журнала (где находится, в каком виде.

-сделать корпус с нормализованной орфографией для 10 выпусков

И чтобы вы понимали, таблицу мы уже создали. Как бы 80 % работы уже сделано.

Далее, если смотреть в завтрашний день трезвым взглядом, то наш ожидаемый результат (Expected product (EP)) корпус с нормализованной орфографией для 100 выпусков + мы хотим сделать классный сайт, где и разместим результат наших трудов.

Мы достаточно оптимистично настроены, поэтому наш идеальный результат совпадает с ожидаемым.

Как мы планируем работать:

* сначала мы создаем таблицу. В ней у нас указаны года, в соответствии с ними номера томов, и ссылки на сам текст. Часть ссылок мы нашли на Викитеке, в Google Books, а все остальные - в электронном каталоге Российской Национальной Библиотеки.
* Когда таблица создана, мы смотрим, какие тома насколько распознаны. Поскольку этот этап мы уже частично выполнили, то в наших планах прогнать все тексты через Fine Reader, выполнив при этом серию экспериментов с языковыми регистрами. (Русский, русский+ французский, просто русский).
* Далее следует, пожалуй, большая часть нашего проекта, когда мы проверяем насколько хорошо распознан текст, исправляя при необходимости опечатки и неточности. Для начала планируется 200 страниц на человека в неделю, далее пойдут целые тома.
* В зависимости от скорости вычитки мы будем корректировать наш план. В наших интересах откорректировать как можно больше томов, чтобы потом мы могли прогнать их через программу, которая переделает русский язык 19 века в классический русский язык 21 века. Таким образом у нас уже получается огромный массив вычитанных текстов Отечественных Записок на современном русском языке.
* Останется сделать сайт. Ну, а поскольку веб-девелопмент у нас будет на втором курсе, мы заложим интригу до следующего года.

Итак, наш план действий вы знаете, источники и инструменты названы. Настрой на победу у нас тоже есть. Спасибо за ваше внимание, мы готовы ответить на вопросы.