**Проектное предложение**

|  |  |
| --- | --- |
| Тип проекта | *Прикладной* |
| Название проекта | Отечественные записки 3.0: создаем цифровой архив |
| Подразделение инициатор проекта | Центр цифровых гуманитарных исследований ФГН НИУ ВШЭ |
| Руководитель проекта | Заковоротная Е. М. |
| Заказчик проекта / востребованность проекта | Проект выиграл грант АНО «Инфокультура» по цифровой архивации ([подтверждение](https://www.hse.ru/ma/dh/news/348090631.html)). В гранте участвовали 24 заявки, в т.ч. от других сильных проектов из НИУ ВШЭ. Таким образом, востребованность проекта подтверждена мнением экспертного сообщества, представители которого составляли жюри гранта. |
| Основная проектная идея / описание решаемой проблемы | «Отечественные записки» ⎯ один из ключевых российских журналов XIX века. На страницах журнала с 1818 по 1884 гг. публиковали свои произведения В. А. Жуковский, В. Ф. Одоевский, В. Г. Белинский, М. А. Бакунин,  А. И. Герцен, Н. А. Некрасов, М. Ф. Салтыков-Щедрин, А. С. Островский, Ф. М. Достоевский, Л. Н. Толстой. Несмотря на значимость журнала для понимания русской культурной и общественной жизни XIX века, его выпуски нельзя найти онлайн в адекватной машиночитаемой форме. В сети доступны либо нераспознанные изображения, либо разрозненные PDF-документы с текстовым слоем неприемлемого качества: доля ошибочно распознанных символов там достигает 50%, символы дореформенной орфографии утрачены.  Из-за этого невозможны ни адекватный полнотекстовый поиск, ни автоматическая обработка текстов. Проведение исторических, лингвистических, филологических исследований с использованием «Отечественных записок» как междисциплинарного текстового корпуса (ср. многочисленные количественные исследования на основе Национального корпуса русского языка, выходящие далеко за пределы собственно лингвистики) требует разработки открытого общедоступного цифрового издания, отвечающего требованиям работы с данными в 2020 году.  Для создания полного корпуса для всех выпусков журнала «Отечественные записки» требуется привлечение исполнителей для первичной вычитки распознанного текста — необходимо вычитать 253 выпуска журнала по 1000 – 1200 страниц каждый, исправив ошибки OCR, неизбежные для материала такой давности. |
| Цель проекта | Открытое общедоступное цифровое издание журнала «Отечественные записки». |
| Планируемые результаты проекта, специальные или функциональные требования к результату | 1. создание электронной базы с детальными метаданными по каждому выпуску журнала; 2. создание и публикация под свободной лицензией корпуса с нормализованной орфографией для всех выпусков в международном формате текстовой разметки культурного наследия TEI/XML (Text Encoding Initiative); 3. создание web-оболочки корпуса. |
| Требования к участникам с указанием ролей в проектной команде при групповых проектах | От участников требуется работа во вычитке и корректировке результатов OCR. Менеджерами и менторами выступит постоянная команда проекта — сотрудники Центра цифровых гуманитарных исследований НИУ ВШЭ. |
| Количество вакантных мест на проекте | 40 (такое количество обусловлено большим количеством материала, к концу проекта необходимо вычитать порядка 40 новых томов по 500 -1000 страниц) |
| Проектное задание | Получив том издания в формате .docx (в дореформенной орфографии), вычитать его на предмет ошибок распознавания и исправить их. Сдать вычитанный .docx-документ |
| Критерии отбора студентов |  |
| Сроки и график реализации проекта | Вычитка томов продлится с ноября 2020 по 20е марта 2021. |
| Трудоемкость (часы в неделю) на одного участника | *3* |
| Количество кредитов | *3* |
| Форма итогового контроля | *Экзамен для проектов от 3 кредитов* |
| Формат представления результатов, который подлежит оцениванию | *Вычитанные файлы .docx* |
| Образовательные результаты проекта | Погружение в литературную, общественную и культурную жизнь русского образованного общества 40-х-50-х годов XIX века Знакомство с журнальным языком и медийными практиками эпохи  Глубокое понимание генезисоа специфического русского феномена «толстого литературного журнала» |
| Критерии оценивания результатов проекта с указанием всех требований и параметров | Оценка выставляется за вычитанный том по формуле «10 \* (процент исправленных ошибок + 10%)*».* Таким образом, за том, в котором исправлены 90% ошибок и не исправлены 10%, выставляется 10. Оценка производится на 1 % страниц тома выбранных автоматически методом случайного сэмплирования. |
| Возможность пересдач при получении неудовлетворительной оценки | *Да* |
| Рекомендуемые образовательные программы | *Филология, Журналистика, История и др. гуманитарные специальности* |
| Территория | *Старая Басманная 21/4 (Факультет гуманитарных наук)* |