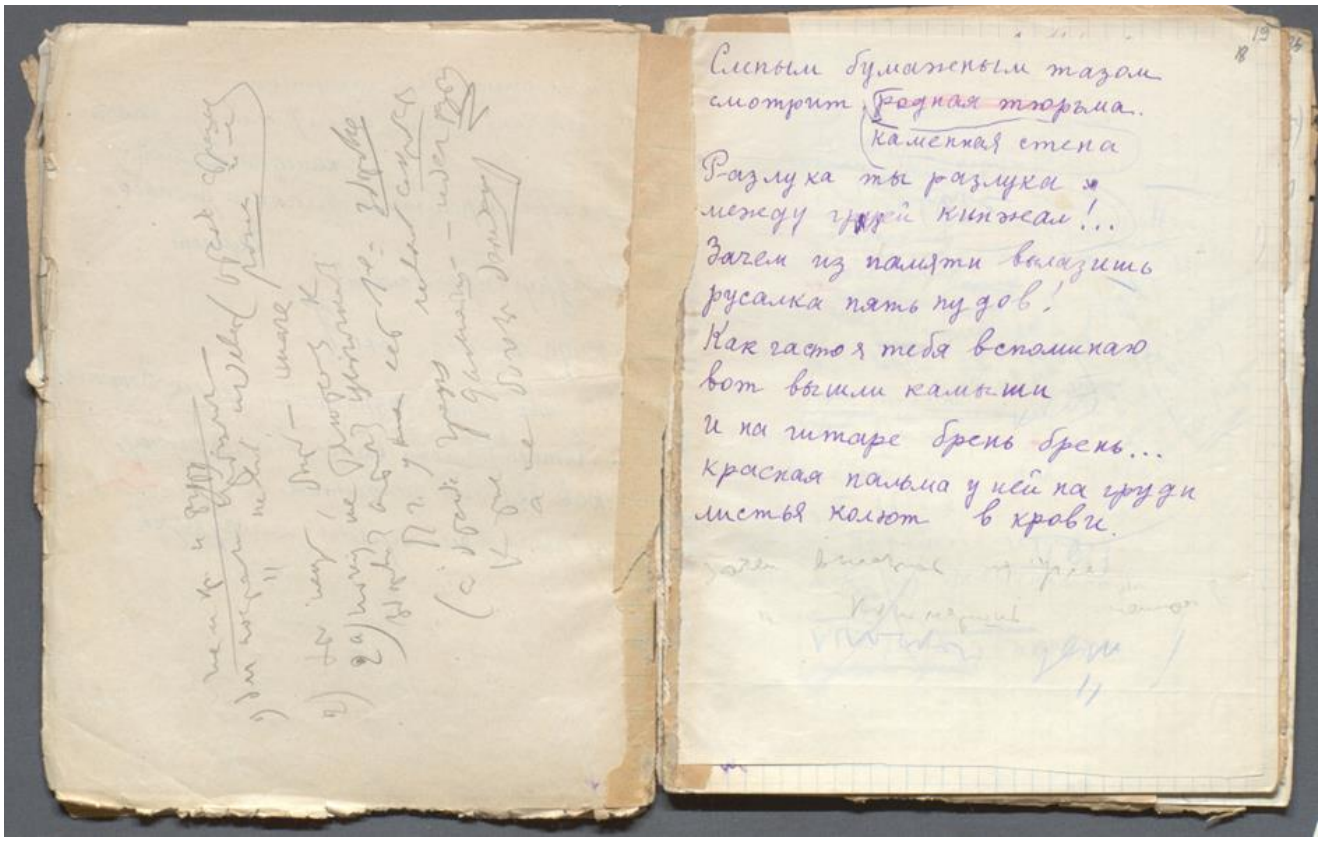
Ольга Ермакова

Записные книжки
и рабочие тетради писателей в цифровой среде

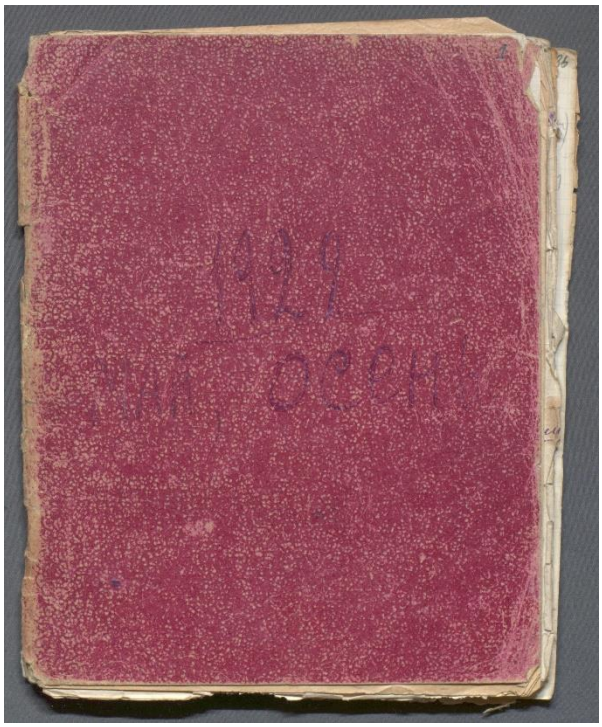
Цифровой поворот и резкий скачок в развитии компьютерных технологий, который мы наблюдали в последние два десятилетия, не только дал толчок развитию точных наук, но и предоставил ученым возможность по-новому взглянуть на объекты, относящиеся к области гуманитарного знания. В частности, этими объектами стали проблемы филологии и текстологии, областей, традиционно ассоциирующихся с качественными, а не количественными методами исследования и потому, на первый взгляд, стоящих в стороне от новых трендов в обработке больших данных и новых возможностей программирования. Произошел подобных сдвиг потому, что тексты стали перемещаться в диджитал пространство: и касается это не только опубликованных произведений отечественных писателей и поэтов, но и материалов, которые довольно долгое время оставались в архивах, музеях и библиотеках: это рукописи, черновики, рабочие тетради, дневники и альбомы авторов.



Здесь и далее публикуется рабочая тетрадь Алексея Крученых
РГАЛИ. Ф. 1334. Оп. 1. Ед. хр. 48. Л. 18 – 19.

Шаг первый: выход в цифровое пространство

Традиционно в качестве объекта для академических исследований выступал критически установленный текст, то есть тот вариант художественного произведения, который был зафиксирован учеными как окончательный. Более того, даже к настоящему времени, более 90 % процентов рукописных документов не опубликованы и фактически не изучены. Среди них – рабочие тетради, дневники, переписки, которые рассматривались лишь как вспомогательный материал для исследования основного текста художественного произведения. На рубеже XX – XXI вв. ситуация изменилась. В фокус внимания исследователей попали не только на рукописи известных авторов, но и на официальные документы, обуславливающие и определяющие жизнь писательского сообщества: стенограммы, протоколы архивных заседаний, деловая переписка.

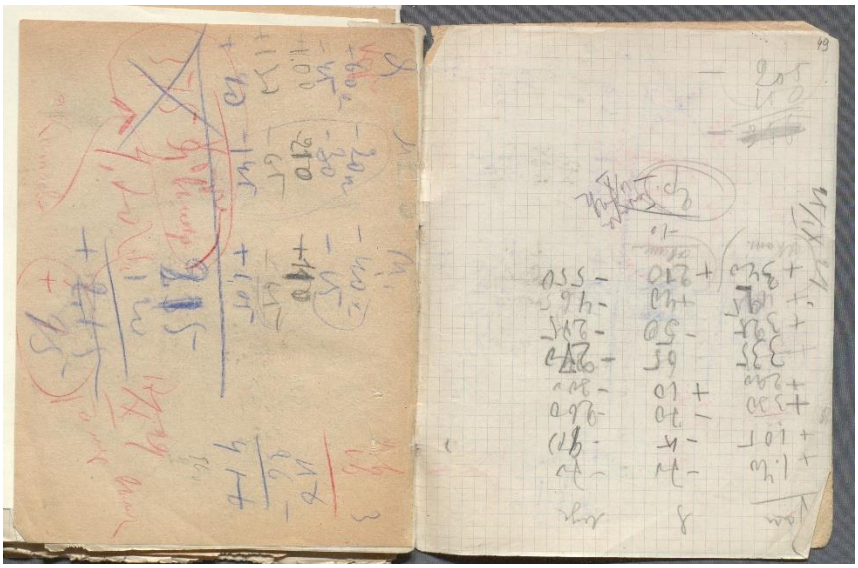


Рабочая тетрадь Алексея крученых. Обложка

В связи с ростом интереса к огромному пласту документов, которые ранее практически не были исследованы, филологам стал необходим практичный инструментарий для работы с рукописями – наравне с подлинником текстологи начали использовать электронные копии документов. Работая с цифровой рукописью, исследователь может масштабировать текст или увеличить яркость строк, написанных выцветшими чернилами.

Цифровые архивы, подобно офлайн-хранилищам, собирают и систематизируют рукописное наследие, среди которого рабочие тетради с набросками произведений, письма, записные книжки, неизвестные редакции широко тиражированных текстов. Важно, что электронное собрание позволяет увидеть источник в среде его бытования. Например, можно рассмотреть и исследовать набросок рассказа вместе с письмом, в которое он был вложен, первые варианты стихотворений вместе с посвященными им дневниковыми записями.

Исследователь продвигается от нескольких строчек в записной книжке или подчеркнутой карандашом фразы к рабочим тетрадям, письмам, рисункам. Затем следуют черновые автографы, еще очень далекие от окончательного варианта, после – письма и дневники, и наконец, первый беловик.



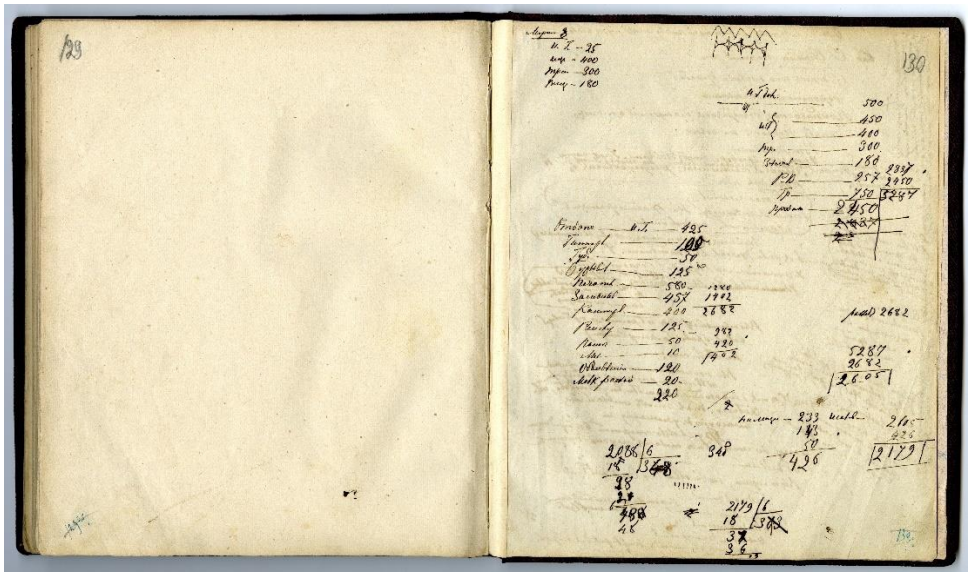
Расчеты Алексея Крученых. Форзац рабочей тетради

Подобный переход рукописных текстов в цифровое пространство стал первым шагом к тому, чтобы рукописные тексты сблизились с большими данными.

Шаг второй: большие данные цифрового архива

Цифровые архивы, хранящие в себе десятки тысяч отсканированных документов, технологически представляют собой нераспознанный корпус растровых изображений, вполне соответствуют категориям, по которым определяют большие данные. Среди них объём (volume), скорость приращения (velocity) и многообразие (variety). И накопленный объем цифровых изображений, и темпы роста виртуальных архивов соответствуют этим критериям. Однако последний критерий – вариативность – ставит перед исследователями важную задачу: необходимо не просто наблюдать за многообразием цифровых документов, но и находить взаимосвязи между ними, а также унифицировать растровые изображения для того, чтобы обработать их с помощью технологий data mining – интеллектуального анализа изображений. Здесь могут помочь стандарты описания архивных документов и гиперссылки, предоставляющие исследователю возможность самостоятельно моделировать состав и размер массива данных. Среди примеров подобных

архивов «Объединенный цифровой архив рукописей Ф.М. Достоевского» (ИРЛИ РАН) и «”Стенограмма”: Политика и литература. Цифровой архив литературных организаций 1920-х – 1930-х годов» (ИМЛИ РАН).



Федор Достоевский. Расчеты в рабочей тетради
РГАЛИ. Ф. 212. Оп. 1. Ед. хр. 9. Л. 129 – 130.

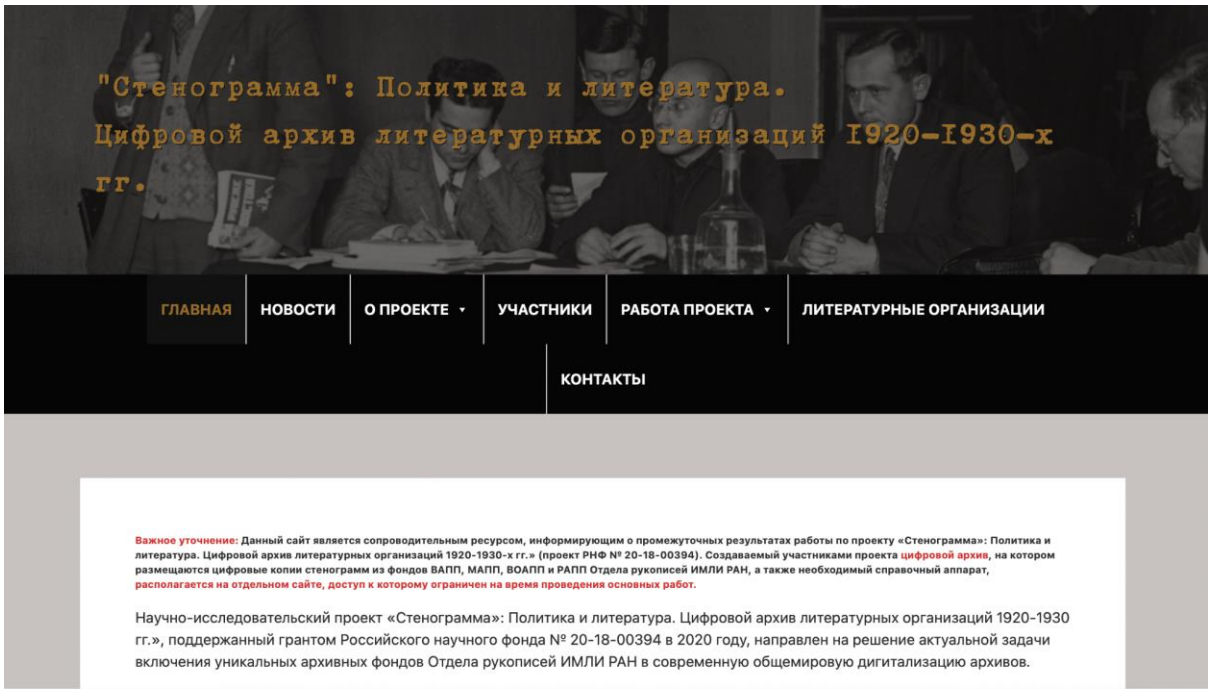
С некоторой оговоркой к группе подобных архивов можно отнести и проект ИРЛИ РАН «А.С. Пушкин: Электронное академическое издание. Тексты произведений». Но отличается от остальных этот проект тем, что Pushkin Digital в первую очередь является уникальным цифровым научным изданием в гипертекстовом формате, и возможность использовать черновики Пушкина как растровый архив – результат вторичный.

Сайт «Объединенного цифрового архива рукописей Ф.М. Достоевского». ИРЛИ РАН

Наиболее близко к управлению «большой цифрой» из названных архивов “подошел” проект «”Стенограмма”: Политика и литература». В проекте задействован нераспознанный массив официальных документов –

машинописных отчетов о съездах, конференциях, заседаниях литературных институций, протоколов и анкет их участников, учетных листков и финансовых ведомостей – в совокупности насчитывающих десятки тысяч страниц.

.



Сайт проекта «“Стенограмма”: политика и литература»

Информационный массив выстроен по модели интерактивного древа – ветвящейся сети источников, которую исследователь формирует сам. От конкретной стенограммы он может перейти к персоналии и всему комплексу связанных с ней документов. Исследователь также может осуществлять поиск по определенным категориям. Например, организациям и сосредоточить внимание на источниках, посвященных институциональной политике. С помощью гибкого управления большими данными комплекс из десятков тысяч растровых страниц редуцируется до нескольких десятков отдельных страниц, непосредственно затрагивающих проблематику исследования. Остается один шаг до интеллектуального анализа данных: если сделать единый аппарат ключевых слов и унифицировать алгоритмы поиска, то сжатие информационного массива от десятков тысяч возможно-нужных страниц до нескольких десятков совершенно необходимых можно будет выполнить автоматически. Технологии искусственного интеллекта могут быть задействованы для решения исследовательских задач и при полнотекстовом поиске. В этом случае с помощью программного модуля будет сформирован растровый паттерн искомой лексемы (графическая имитация машинописи), а затем автоматизировано сопоставлен с множеством словоформ полного

машинописного корпуса стенограмм и протоколов. На помощь ученому приходят технологии сбора, обработки и анализа больших данных, которые позволяет ему, с одной стороны, быстро найти необходимую информацию среди десятков тысяч текстов, с другой – по-новому взглянуть на привычный объект исследования.

Цифровые рукописи прошли немало шагов на пути к сближению с большими данными. За последние 10 лет эго-документы «вышли» из хранилищ и библиотек в цифровое пространство, затем, усилиями исследователей, организовались в упорядоченные и общедоступные цифровые архивы, часть которых, благодаря особенностям устройства и каталогизации открыла возможность для интеллектуального анализа данных. Благодаря цифровизации традиционное гуманитарное знание сблизилось с методами работы, которые обычно относятся к точным или компьютерным наукам. И это «притяжение противоположного» предлагает ученым возможность изучения привычных объектов под новым углом цифрового видения.