

9. Егор Перетц [Электронный ресурс] // Проект «Прожито». – Режим доступа: <https://prozhito.org/person/764>. – Дата доступа: 3.03.2021.
10. Ефремов, С. О. Щоденники (1923–1929) [Электронный ресурс] / С. О. Ефремов // Проект «Прожито». – Режим доступа: <https://prozhito.org/notes?date=%221924-01-01%22&diaries=%5B410%5D>. – Дата доступа: 3.03.2021.
11. Жихарев, С. П. Записки современника [Электронный ресурс] / С. П. Жихарев // Проект «Прожито». – Режим доступа: <https://prozhito.org/notes?date=%221808-01-01%22&diaries=%5B515%5D>. – Дата доступа: 3.03.2021.
12. Историк в эпоху перемен: Н. И. Кареев в 1914–1931 гг.: исследование и материалы [Электронный ресурс] // Проект «Прожито». – Режим доступа: <https://prozhito.org/notes?date=%221926-01-01%22&diaries=%5B172%5D>. – Дата доступа: 3.03.2021.
13. Орешников, А. В. Дневник. 1915–1933 [Электронный ресурс] // Проект «Прожито». – Режим доступа: <https://prozhito.org/notes?date=%221930-01-01%22&diaries=%5B172%5D>. – Дата доступа: 3.03.2021.
14. Чикаленко, Е. Х. Щоденник (1907–1917) [Электронный ресурс] / Е. Х. Чикаленко // Проект «Прожито». – Режим доступа: <https://prozhito.org/notes?date=%221909-01-01%22&diaries=%5B421%5D>. – Дата доступа: 3.03.2021.

**Базарова Т. А., Проскуракова М. Е.**

### **«АВТОГРАФЫ ПЕТРА ВЕЛИКОГО: ЧТЕНИЕ ТЕХНОЛОГИЯМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»: ЭТАПЫ РЕАЛИЗАЦИИ ПРОЕКТА**

В докладе представлены результаты научно-исследовательского проекта «Автографы Петра Великого: Чтение технологиями искусственного интеллекта». Задачей проекта стало создание алгоритма, который способен автоматически распознавать и транскрибировать автографы Петра Великого.

Одной из особенностей русского делового письма до начала XVIII века является стандартизации почерка писцов и широкое использование типичных каллиграфических приемов. При составлении документов служащие канцелярий XVI–XVII вв. использовали устоявшиеся словосочетания, языковые формулы, повторяя их из документа в документ. Благодаря этому оказывается возможным чтение даже угасшего текста и восстановление целых фраз, которые не сохранились, но почти наверняка содержались в тексте.

В начале XVIII в. получает распространение новое явление. Сановники эпохи Петра Великого начинают вести переписку друг с другом по деловым и личным вопросам. Нередко они пишут письма своей рукой, не прибегая к помощи секретарей. Эти написанные беглым малоразборчивым почерком тексты чрезвычайно сложны для чтения и транслитерации.

Петр был первым русским царем, кто лично стал подписывать документы и собственноручно писать письма (до него грамоты заверялись печатью). Поэтому до нашего времени сохранился огромный массив (более десяти тысяч документов), написанных или подписанных Петром Великим.

В автографах петровских сановников также, как и в автографах самого Петра, можно увидеть продолжение скорописной традиции в использовании выносных (надстрочных) букв и сокращений, связности написания двух графем, одна из которых написана в строке, а другая над строкой (лигатура), отсутствии пробелов между словами. Однако в случае с автографами Петра Великого мы сталкиваемся также с яркой индивидуальной манерой. Перед нами письмо «спешащего» человека: торопящегося поделиться новостями, отдать распоряжение, сделать выговор провинившемуся чиновнику. Эта особенность автографов Петра Великого затрудняет их прочтение. Другая сложность заключается в богатстве лексики и языка петровских писем. Петр часто использовал иностранные слова, топонимы, узкоспециальные термины, метафоры. Поэтому чтение автографов царя требует не только навыков чтения документов XVII в. и начала XVIII в., но и глубокого знания исторического контекста.

Перспективы применения новых методов при изучении автографов первого российского императора открылись благодаря научно-исследовательскому проекту «Автографы Петра Великого: Чтение технологиями искусственного интеллекта», инициированному в июне 2020 г. Российским историческим обществом и ПАО «Сбербанк». В Санкт-Петербургском институте истории РАН (далее – СПбИИ РАН) была сформирована рабочая группа, состоящая из научных сотрудников – специалистов по истории Петровской эпохи, а также палеографии и археографии.

Реализация данного проекта находится в русле проблем, стоящих перед мировой наукой. В настоящее время активные поиски в области компьютерного транскрибирования рукописного текста ведут крупные европейские научные центры. В рамках международного проекта «The Recognition and Enrichment of Archival Documents (READ)» (<https://eadh.org/projects/read>) была создана платформа Transkribus [4], предназначенная для машинного чтения древних рукописей (<http://transkribus.eu>) [5]. Платформа включает более 50 моделей (58) для распознавания текстов и позволяет обрабатывать документы, написанные ранними формами кириллического письма – уставом и полууставом. Модель для транслитерации документов, написанных полууставом, построена на 38374 строках; комбинированная модель для чтения уставного и полууставного письма построена на 75422 строках [3].

Однако тестовое транскрибирование нескольких образцов полууставного письма выявило многочисленные неточности и ошибки прочтения. Прежде всего, следует отметить низкое качество распознавания надстрочных знаков, которые в отдельных случаях «читаются» алгоритмом как самостоятельные строки и неверно идентифицируются как предлоги или окончания слов верхней строки. Кроме того, основной массив русской делопроизводственной документации XV – начала XVIII в. написан скорописью, для чтения которой в Transkribus не создано модели.

Участниками российского проекта были выявлены и отобраны документы на основе материалов 9–13 томов издания «Письма и бумаги императора Петра Великого», вышедших из печати в 1950–2003 гг. [2]. Эти тома включают в себя документы с 1709 по 1713 г. Решение взять за основу не первые, а последние тома продолжающегося издания было связано с тем, что легенды опубликованных в них документов содержали наиболее близкие к современным архивные шифры Российского государственного архива древних актов (далее – РГАДА). Источниковую базу проекта также дополнили документы из собрания СПБНИИ РАН. Основным критерием отбора документа являлось наличие в нем текста, написанного рукой Петра (от двух-трех слов до нескольких страниц).

Изучение значительного массива документов позволило выявить характерные черты почерка Петра Великого. Особенностью его автографов является малая вариативность в написании графем: как правило, буква имеет одну, реже две формы. В то же время начертание ряда выносных знаков близко друг к другу, а порой даже идентично. Так, например, схожим надстрочным знаком передаются буквы «н», «с», «т» [1, л. 66]. Прочтение и транскрибирование текста также затрудняет отсутствие пробелов между словами и частые сокращения слов, в том числе использование одного и того же знака при сокращении разных слов. Петр употреблял два вида сокращений: сокращения до начальной буквы слова или начальных букв нескольких слов, например, «г ф» – «г (енерал)-ф (ельдмаршал)» и сокращения слогов в отдельных словах, например, «де» – «де (нь)» [1, л. 66]. Применяя сиглы (сокращения до первой буквы), Петр буквой «г» передавал сразу несколько слов «господин», «генерал», «Головин». Отсутствие вариативности в сокращении разных слов, начинающихся с одной буквы, встречается и в других случаях.

В почерке Петра Великого наблюдается непоследовательность в написании заглавных и строчных букв. Как правило, заглавная буква проставлена только в первой строке, где указывалась должность (или чин) и имя адресата, в начале основного текста послания, а также в подписи – имени государя. Кроме того, в автографах Петра существует вариативность в написании букв «ер» (ъ) и «ерь» (ь), что является характерной чертой в письме XVII – начала XVIII в. В изученном комплексе автографов нет случаев использования буквы «кси» (ѣ) и «пси» (ѵ) [1, л. 66]. В собственноручных текстах Петра Великого также не удалось обнаружить букву «ферт» (ф), вместо которой употреблялась «фита» (Ѳ). Примечательно, что эту же букву царь использовал в качестве выносной в конце слов. В таких случаях она служила «заменой» букве «в»: «Ушакоф», «охотникоф», «футоф».

В ходе реализации проекта петербургские палеографы совместно со специалистами по анализу данных Сбербанка Д. В. Димитровым и М. С. Потаниным выработали методику передачи текста петровских автографов для последующей компьютерной обработки. Основным принципом стала максимально точная передача знаков (букв), которые использовал Петр Великий. Помимо букв современного нам алфавита при компьютерном наборе использовались «ять» (ѣ) и «и десятеричная» (і), надстрочные знаки вносились в строку, мягкий и твердый знаки при их отсутствии не восполнялись, сокращения не раскрывались, заглавные буквы, знаки препинания, знаки переноса строки не расставлялись [1, л. 67]. Компьютерному набору подлежал только собственноручный текст государя в документе (включая зачеркнутые буквы, слова и предложения, а также текст, вписанный на полях). Удалось выявить случаи, когда текст Петра Великого был подвергнут издателями «Писем и бумаг» цензуре. Фразы, содержащие вульгарные просторечные слова, не вошли в публикацию. В ходе реализации описываемого проекта такого рода лексика не опускалась и петровский автограф воспроизводился дословно. Набранный текст несколько раз сверялся с цифровой копией подлинника, все спорные случаи обсуждались. Каждой строке (в цифровой копии и компьютерном наборе) присваивался номер. Если документ состоял из нескольких страниц, то текст на каждой из них имел свою нумерацию. Таким образом было выделено 9656 строк.

Следующий этап работы начался после загрузки цифровых копий в web-приложение Computer Vision Annotation Tool (далее – CVAT). Это было связано с необходимостью построчной разметки цифровых копий документов (в соответствии с компьютерным набором). На данном этапе к выполнению проекта подключились аспиранты и магистранты Научно-исследовательского университета «Высшая школа экономики» (Москва и Санкт-Петербург), имевшие опыт работы со скорописью XVII–XVIII вв. Таким образом, весь оцифрованный материал был размечен и подготовлен к последующей обработке сотрудниками ПАО «Сбербанк».

Созданный для машинного чтения автографов Петра Великого алгоритм безошибочно распознает свыше 97% текста [1, с. 68–71]. Разработанная модель способна быстро обучаться на новом почерке, не требуя большого количества транскрибированных и размеченных в CVAT данных (достаточно всего несколько страниц с новым почерком). Как представляется, дальнейшее совершенствование алгоритма пойдет по пути обучения его распознаванию отдельных слов (вне контекста), заглавных букв (в том числе и буквиц), а также сокращений, вставок на полях и помет из документов XVI–XVIII вв.

1. Распознать и транскрибировать: Автографы Петра Великого и технологии искусственного интеллекта / Т. А. Базарова [и др.] // Воронцово поле. – 2020. – № 4. – С. 36–41.
2. Письма и бумаги императора Петра Великого. – М.; Л.: Изд-во Академии наук СССР, 1950–1962. – Т. 9. – Вып. 1, 2; Т. 10; Т. 11. – Вып. 1; М.: Наука, 1964–1975. – Т. 11. – Вып. 2; Т. 12; М.: Древлехранилище, 2003. – Т. 13.
3. Rabus, A. Recognizing handwritten text in Slavic manuscripts: A neural-network approach using Transkribus [Electronic resource] / A. Rabus. – Mode of access: [https://www.academia.edu/38835297/Recognizing\\_handwritten\\_text\\_in\\_Slavic\\_manuscripts\\_A\\_neural\\_network\\_approach\\_using\\_Transkribus\\_1\\_Achim\\_Rabus](https://www.academia.edu/38835297/Recognizing_handwritten_text_in_Slavic_manuscripts_A_neural_network_approach_using_Transkribus_1_Achim_Rabus). – Date of access: 21.03.2021.
4. The Recognition and Enrichment of Archival Documents (READ) project [Electronic resource]. – Mode of access: <https://eadh.org/projects/read>. – Date of access: 21.03.2021.
5. Transkribus. Transcribe. Collaborate. Share [Electronic resource]. – Mode of access: <https://readcoop.eu/transkribus/>. – Date of access: 21.03.2021.

**Ткаченко Д. С.**

### **РОССИЙСКИЕ ИСТОРИКО-КУЛЬТУРНЫЕ ПАМЯТНИКИ XIX – НАЧАЛА XX В. НА КАВКАЗЕ: ОПЫТ СОЗДАНИЯ ВИРТУАЛЬНЫХ РЕКОНСТРУКЦИЙ**

Одним из динамично развивающихся направлений информационных технологий, имеющих потенциал использования в ходе исторических исследований, становится моделирование трехмерного пространства, или «виртуальной реальности». Кроме дублирования реальных историко-культурных объектов посредством 3D-технологий можно осуществить реконструкцию утраченных объектов – воссоздать их образ, используя набор данных, полученных через описание или изображение объекта, составленных до его уничтожения.

Одним из шагов в данном направлении стала реализация проекта РГНФ № 16-01-12012 «Российские историко-культурные памятники XIX – начала XX в. на Северном Кавказе: научно-исследовательский ресурс» [7], разработанный автором в 2016–2017 гг. с целью провести виртуальную реконструкцию ныне утраченных имперских памятников на Кавказе.

По своему архитектурному облику, реконструированные в виде трехмерных моделей памятники могли быть разных видов. Прежде всего, следует отметить наличие в российской империи типового архитектурного проекта, очень близкого тому стилю, который использовался европейскими странами. Это была пирамидальная стела, помещенная на квадратное основание. Одним из первых маркеров такого типа стал памятник майору Монтезору, погибшему вместе со своим отрядом в июле 1804 г. под селением Серали. Помимо описания [4, л. 2–2об.], сохранился рисунок этого памятника, а также фотография торжеств, проведенных в 1904 г. во время освящения монумента после реставрации. Изображения дают все необходимые детали для создания трехмерной модели памятника.

Создание реконструкции проводится по технологии лофтинговой фигуры, когда квадратное основание и несколько уменьшающихся в диаметре срезов памятника проецируются на линию, очерчивающую высоту монумента. В результате создается точная копия того пирамидального обелиска, геометрия которого легла в основу оригинального памятника. Текстурой служит изображение каменных блоков из черного гранита – материала, которым по описанию современников был облицован оригинальный монумент. Образ таблички так же достаточно четко представлен на рисунке и фотографии памятника, что делает возможным ее точное воспроизведение в виде трехмерной модели.

Аналогичный по своему архитектурному решению памятник был возведен на Гунибе солдатами Апшеронского полка для того, чтобы увековечить роль своей части в штурме последней твердыни Шамиля. Еще одним памятником, изготовленным по вариации типового архитектурного проекта, стал памятник князю Цицианову, установленному в г. Баку на одноименном сквере в 1846 г. Его официальное описание содержится в отчете губернатора перед МВД [5, л. 3]. Сохранилось так же большое количество фотографий этого памятника, снятых с различных ракурсов, которые позволяют создать не только детальную реконструкцию стелы, но также и элементы историко-культурного ландшафта. Техничко-строительный комитет МВД хранил и архитектурный проект всего Цициановского сквера с точным указанием на нем мест расположения основных объектов. Это делает возможным в совокупности с информацией об образах окружающих сквер зданий, воссоздать не только памятник, но и часть городского ландшафта старого Баку.

Кроме типового проекта в имперских памятниках на Кавказе использовались и более сложные архитектурные решения. Из них наибольший интерес представляет реконструкция первого