

## **ЦИФРОВАЯ ПАЛЕОГРАФИЯ И АНАЛИЗ ПОЧЕРКА: ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ К ДРЕВНИМ РУКОПИСЯМ**

Нарыжный Д.И. (ИТМО)

Научный руководитель – доцент Попков Р.А. (ИТМО)

Введение. Палеография – историко-филологическая дисциплина, исследующая эволюцию систем письма и графических форм букв в их материальном и культурном контексте. Она позволяет датировать, локализовать и атрибутировать рукописи, раскрывая социальные аспекты письменной традиции [1]. С развитием вычислительных технологий возникло направление цифровой палеографии, ориентированное на извлечение признаков из оцифрованных документов и автоматическую классификацию почерков. Исследователи подчёркивают, что цифровая палеография не только переносит методы машинного обучения в гуманитарную область, но и заставляет переосмыслить, как цифровые представления влияют на понимание самого объекта [3]. Мотивацией для таких исследований является масштаб и скорость обработки данных, объективность метрик, возможность верифицировать гипотезы специалистов и междисциплинарность этого направления.

Основная часть. Задачи цифровой палеографии включают идентификацию писцов и стилей, классификацию рукописей по языку и системе письма, датировку и атрибуцию текстов, сегментацию и распознавание рукописного текста, а также применение языковых моделей для восстановления лакун и нормализации исторической орфографии. Эти задачи требуют обработки изображений рукописей, выделения текстовых областей, сегментации строк и символов и анализа графических особенностей почерка [2]. Исследования показывают, что классические методы, основанные на ручном конструировании признаков и статистических моделях, обладают интерпретируемостью и устойчивостью на малых выборках, но требуют значительных усилий по инженерии признаков и плохо масштабируются. Современные свёрточные нейронные сети достигают высокой точности, однако их переносимость ограничена и они склонны к переобучению на фон или особенности конкретной рукописи, что приводит к резкому падению качества на «слепых» наборах данных. Дефекты рукописей – выцветание чернил, утраты носителя, нестандартные сокращения – а также высокая стоимость разметки усугубляют проблему обработки данных.

В работе предлагается методика классификации почерков внутри крупных рукописей с использованием сочетания классических и глубоких моделей. Для получения устойчивых оценок предлагается вычислять метрики на уровне рукописей и разбивать выборку по рукописям, что позволяет оценить способность моделей обобщать на новые документы. Для разметки используется концепция «soft-labeling», при которой вместо одной жёсткой метки класса используется вектор степеней принадлежности, отражающий переходные стили письма; такая разметка ближе к реальности и помогает избежать переобучения. Для проверки того, на какие элементы рукописи обращает внимание модель, используются карты внимания (Grad-CAM), позволяющие отсеивать модели, ориентирующиеся на фон или дефекты вместо букв.

Применение глубоких сетей для классификации еврейских рукописей показывает, что нейросети могут автоматически извлекать сложные признаки и достигать точности, сопоставимой с экспертом [3]. Авторы подчёркивают, что создание такой модели требует предварительной подготовки датасета, включающего основные типы письма и различные режимы письма; используется как жёсткая, так и мягкая маркировка, что позволяет отражать неопределённость при отнесении к стилю.

**Выводы.** Описанная методика демонстрирует, что цифровая палеография является междисциплинарным полем, где методы машинного обучения и компьютерного зрения решают традиционные задачи историков и филологов. Реализация предложенных подходов позволит ускорить атрибуцию и датировку рукописей, стандартизировать обработку оцифрованных коллекций и облегчить работу специалистов. Разработанные модели могут быть применены для анализа рукописей различных традиций, включая греческие и еврейские манускрипты, и послужат основой для создания инструментов, интегрированных в электронные библиотеки. Перспективными направлениями являются расширение экспериментов, применение доменной адаптации и развитие методов совместного обучения для многокорпусных данных.

#### **Список использованных источников:**

[1] Paleography, study of ancient and medieval handwriting // Encyclopaedia Britannica. URL: <https://www.britannica.com/topic/paleography> (дата обращения: 17.02.2026).

[2] Dhali M.A., He S., Popović M. и др. A Digital Palaeographic Approach towards Writer Identification in the Dead Sea Scrolls // Proc. of ICPRAM 2017. P. 693–702.

[3] Droby A., Rabaev I., Shapira D. и др. Digital Hebrew Paleography: Script Types and Modes // Journal of Imaging. 2022. Vol.8, No.5:143.