

DOI: 10.26794/3030-7097-2026-2-3-6-15
УДК 621.382.323(045)

Формирование объединенного алфавита древнетюркских диалектов для аннотации изображений

А.Д. Бородина, М.В. Кызыл-оол, Р.А. Кочкаров

Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

В статье представлена разработка объединенного алфавита древнетюркских диалектов для решения задач автоматизированного распознавания орхон-енисейских рунических надписей. Актуальность исследования обусловлена разрозненностью существующих классификаций рун и отсутствием единой системы их сопоставления, что затрудняет формирование совместимых наборов данных и снижает эффективность моделей компьютерного зрения. **Цель исследования** заключается в создании унифицированного алфавита, обеспечивающего корректную аннотацию изображений и возможность обучения многоалфавитных нейросетевых моделей для распознавания и сравнительного анализа рунических символов. **Методология** исследования базируется на комплексном анализе визуальных, фонетических и историко-филологических характеристик рун четырех источников: памятника Кюль-Тегину, рукописи Ирк Битиг, орхонских и енисейских надписей. В ходе работы выделены структурно устойчивые символы, стилиевые вариации и уникальные маркеры диалектов, на основе чего сформирован консолидированный набор классов рун. **Полученные результаты** создают основу для стандартизации разметки изображений, повышения устойчивости распознавания и дальнейшего развития многоалфавитных систем компьютерного зрения в исследованиях древнетюркской письменности.

Ключевые слова: орхон-енисейская письменность; руническая письменность; древнетюркский язык; систематизация алфавитов; распознавание символов; компьютерное зрение; искусственный интеллект; многоалфавитная модель

Для цитирования: Бородина А.Д., Кызыл-оол М.В., Кочкаров Р.А. Формирование объединенного алфавита древнетюркских диалектов для аннотации изображений. *Цифровые решения и технологии искусственного интеллекта*. 2026;2(3):6-15. DOI: 10.26794/3030-7097-2026-2-3-6-15

ORIGINAL PAPER

Formation of a Unified Alphabet of Ancient Turkic Dialects for Image Annotation

A.D. Borodina, M.V. Kyzyl-ool, R.A. Kochkarov

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

ABSTRACT

This paper is devoted to the development of a unified alphabet of Ancient Turkic dialects for solving the problems of automated recognition of Orkhon–Yenisei runic inscriptions. The relevance of the study is determined by the fragmentation of existing rune classifications and the absence of a unified system for correlating symbols across different sources and regions, which complicates the formation of compatible datasets and reduces the efficiency of computer vision models. **The research** is to create a unified alphabet that ensures correct image annotation and enables the training of multi-alphabet neural network models for the recognition and comparative analysis of runic symbols. **The methodology** is based on a comprehensive analysis of visual, phonetic, and historical-philological characteristics of runes from four sources: the Kül Tegin monument, the Irk Bitig manuscript, Orkhon inscriptions, and Yenisei inscriptions. During the study, structurally stable symbols, stylistic variations, and unique dialect markers were identified, on the basis of which a consolidated set of rune classes was formed. **The obtained results** provide a foundation for the standardization of image annotation, improvement of recognition robustness, and further development of multi-alphabet computer vision systems in the study of Ancient Turkic writing.

Keywords: Orkhon-Yenisei script; runic writing; Ancient Turkic language; alphabet systematization; symbol recognition; computer vision; artificial intelligence; multi-alphabet model

For citation: Borodina A.D., Kyzyl-ool M.V., Kochkarov R.A. Formation of a unified alphabet of ancient Turkic Dialects for Image Annotation. *Digital Solutions and Artificial Intelligence Technologies*. 2026;2(3):6-15. DOI: 10.26794/3030-7097-2026-2-3-6-15



ВВЕДЕНИЕ

Рунические надписи Орхоно-Енисейского региона представляют собой уникальное культурное наследие древнетюркских народов. Большинство из них высечены на камнях, а тексты посвящены памятным событиям или важным людям. Также сохранилась и рукописная «Гадательная книга» (Ирк Битиг), содержащая поэтические миниатюры и выводы к ним. Эти памятники, возникшие в VI–X вв., являются важнейшими источниками для изучения истории, языка и культуры древних тюрков¹.

История изучения орхоно-енисейских рунических письменностей берет свое начало с работы датского лингвиста Вильгельма Томсена, впервые расшифровавшего орхонские надписи в 1893 г. Это открытие стало основой для дальнейших исследований и положило начало систематическому изучению этого вида письменности. В XX в. значительный вклад в эту область внес С.Е. Малов, чьи труды способствовали глубокому пониманию структуры и значения текстов, написанных этими письменностями [1]. Современные исследования продолжают развивать тему, применяя технологии машинного обучения². Однако, несмотря на значимость темы, исследование рунических надписей сталкивается с рядом проблем. Одной из ключевых является разрозненность существующих классификаций и отсутствие единой схемы систематизации [2, 3].

В последние годы интерес к орхоно-енисейским письменностям усилился в рамках цифровых гуманитарных исследований, где предпринимаются попытки автоматизации анализа памятников с использованием методов компьютерного зрения и машинного обучения³ [4, 5]. Однако большинство существующих работ опираются на локальные классификации символов, что ограничивает возможность объединения данных и переноса моделей между различными корпусами [6].

Современные методы компьютерного зрения и глубокого обучения позволяют решать задачи автоматического распознавания древних письменностей с высокой точностью, однако их эффективность напрямую зависит от качества и согласованности исходных данных⁴ [7]. В случае орхоно-енисейских рунических памятников основной проблемой остается отсутствие единой системы классификации символов. Такая разрозненность приводит к ряду последствий:

- во-первых, невозможно объединить существующие наборы данных для обучения единой модели, так как один и тот же символ в разных классификациях может иметь разные обозначения;
- во-вторых, модели компьютерного зрения, обученные на одном наборе, теряют обобщающую способность при применении к надписям из других регионов или источников;
- в-третьих, возникает риск ошибочной интерпретации редких или региональных вариантов начертания, что снижает точность распознавания.

Ранее нами была предложена система классификации на основе текста с памятника Кюль-Тегину⁵ [8]. Однако данный подход оказывается неприменимым в случаях, когда в надписях присутствуют руны, отсутствующие в исходной классификации. Для решения данной проблемы все символы, не вошедшие в нашу классификацию, отмечались одним классом. В результате в модель подавались как один класс руны с абсолютно разным начертанием. Это негативно сказывалось на процессе обучения и приводило к снижению валидационной точности модели [9].

Научная новизна настоящего исследования заключается в формировании объединенного алфавита, ориентированного на прикладные задачи аннотации изображений и обучения многоалфавитных моделей компьютерного зрения, а не на традиционную лингвистическую классификацию. Объединение разрозненных классификаций в единый алфавит позволяет решить эти проблемы. Единая система символов создает универсальную основу для аннотации изображений, что делает возможным формирование крупного, сбалансированного корпуса данных. Это, в свою очередь, обеспечивает стабильность и воспроизводимость обучения моделей, упрощает перенос знаний между различными источниками и позволяет строить многоалфавит-

¹ Кормушин И.В. Древние тюркские языки. Учебное пособие. М.: Восточная литература; 2006. 30 с. URL: <https://www.elibrary.ru/qrijsof>

² См., например: Пойкалайнен А.М., Сердюк М.С., Таратин А.А., Кочкаров Р.А. Распознавание тюркских рунических надписей на примере Ирк Битиг. 2024. URL: <https://t.me/Digitalisimus>; Исследование по цифровой обработке гуманитарных данных. URL: <https://core.ac.uk/download/pdf/74335732.pdf>; Войников А. Цифровые методы анализа письменных источников. URL: <http://biblio.darial-online.ru/text/Voinikov/06.pdf>; Куприянова Н. Применение цифровых методов в гуманитарных исследованиях. URL: <https://elib.sfu-kras.ru/bitstream/handle/2311/156100/kupriyanova.pdf?sequence=1&isAllowed=y>

³ Кочкаров Р.А., Бородина А.Д., Виноградова В.В. и др. Классификация и разметка символов рунических надписей Евразии методами машинного обучения. Отчет о НИР/НИОКР (итоговый). М.; 2025. URL: <https://www.elibrary.ru/upauqb>

⁴ Распознавание орхоно-енисейских рунических надписей методами машинного обучения. URL: <https://habr.com/ru/companies/fa/articles/907950/?ysclid=ms64e72r1540015961>

⁵ См. также: Задача озвучивания орхоно-енисейских рунических надписей. URL: <https://habr.com/ru/companies/fa/articles/1028000/?ysclid=ms648hbj357696357>

ные нейросети, способные учитывать региональные и временные особенности письменности [10].

В данном контексте целью настоящего исследования является создание объединенного алфавита четырех наречий древнетюркского языка для обучения многоалфавитной модели компьютерного зрения. Достижение этой цели позволит не только разметить изображения рун, но и в дальнейшем обучить модель компьютерного зрения для распознавания символов всех включенных диалектов.

ОБЗОР СУЩЕСТВУЮЩИХ КЛАССИФИКАЦИЙ

Существующие классификации орхоно-енисейских письменностей представляют собой множество систем, разработанных различными исследователями в течение XX в. Их количество превышает двадцать, и они часто противоречат друг другу, что создает значительные трудности для унификации данных и их дальнейшего анализа. Разрозненность классификаций приводит к путанице, усложняет процесс сопоставления данных и затрудняет междисциплинарное сотрудничество [2, 3]. Для решения этой проблемы в своих исследованиях мы решили сформировать рунические алфавиты из источников, имеющих точную историческую и географическую привязку: памятник Кюль-Тегину, Ирк Битиг («Гадательная книга»), а также орхонские и енисейские памятники. Разделив надписи с памятников на символы, мы создали наборы уникальных символов для каждого источника отдельно. Таким образом получилось четыре алфавита, которые далее будут называться Кюль-Тегин, Ирк Битиг, орхонский и енисейский [11–14]. Они отличаются стилем написания, но можно заметить идентичные или похожие руны (рис. 1).

Ирк Битиг, как и орхоно-енисейская письменность, представляет собой важный этап в развитии тюркской письменности. Книга была создана в VII–VIII вв., что делает ее современницей памятника Кюль-Тегину, который был установлен в 732 г. Тем не менее начертание рун в Ирк Битиг значительно отличается от остальных трех алфавитов. Отчасти причина этого может быть в разных способах написания — символы Кюль-Тегина выбиты на камне, когда Ирк Битиг был написан на бумаге. Отличия особо заметны на одинаковых рунах. На рис. 1 классы 006, 106, 206 и 306 объективно обозначают одну руну. При этом самые сильные отличия в начертании наблюдаются у Ирк Битиг.

Рунический состав алфавитов тоже различается. В каждом алфавите есть уникальные символы, не имеющие визуальных аналогов. Енисейская классификация самая многочисленная и состоит из 41 руны. Оставшиеся три алфавита содержат 38 символов.

Ирк Битиг	Енисейский	Орхонский	Кюль-Тегин
001 	101 	201 	301 
002 			302 
003 	103 	203 	303 
004 	104 	204 	304 
005 	105 	205 	305 
006 	106 	206 	306 
007 	107 	207 	307 

Рис. 1 / Fig. 1. Фрагмент таблицы письменностей диалектов / Fragment of the Table of Dialects' Scripts

Источник / Source: составлено авторами / Compiled by the authors.

Все вышеперечисленное затрудняет объединение этих диалектов в общую систему классификации. Одна и та же руна может обозначаться по-разному, а некоторые символы отсутствуют в отдельных системах. Без единого алфавита символов невозможно сформировать крупные обучающие выборки и проводить сравнение результатов между исследованиями⁶ [15, 16].

МЕТОДИКА ОБЪЕДИНЕНИЯ АЛФАВИТОВ

Цель создания объединенного алфавита — разметить наибольшее возможное количество символов на изображениях с руническими надписями. Следовательно, это будет прикладной инструмент машинного зрения, а не лингвистическая классификация. Алфавит должен включить не только все уникальные руны из четырех диалектов, но

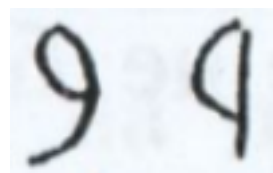


Рис. 2 / Fig. 2. Руны орхонского и енисейского наречия / Runes of the Orkhon and Yenisei Dialects

Источник / Source: составлено авторами / Compiled by the authors.

⁶ Воронцов К.В. Разметка данных для обучения нейросетевых моделей языка как способ формализации гуманитарных знаний. Доклад. XVIII научная конференция межрегиональной ассоциации «История и компьютер» «Историческая информатика как Historical Data Science». М.; 2022.

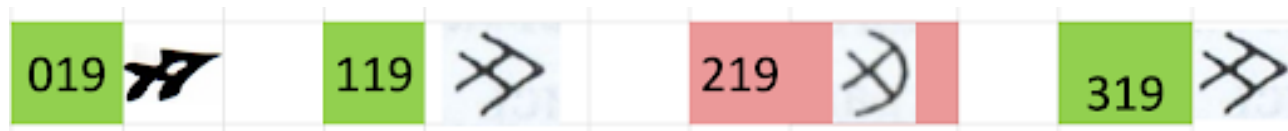


Рис. 3 / Fig. 3. Сравнение четырех наречий: Ирк Битиг, Енисейский, Орхонский, Кюль-Тегин / Comparison of Four Dialects: Irk Bitig, Yenisei, Orkhon, and Kül Tegin

Источник / Source: составлено авторами / Compiled by the authors.

и учесть все варианты начертания. Это значит, что если один и тот же звук или руна имеют два существенно отличающихся начертания, то такие варианты будут включены в систематизацию как отдельные классы. Такой подход позволит модели машинного обучения формировать устойчивые векторные представления для каждого символа и надежно определять их даже при визуальных вариациях начертания.

При создании общего алфавита также было важно не перегрузить его вариациями начертания. Если облик и форма символов, обозначающих один звук, имеет незначительные отличия (рис. 2), то определяли их в один класс.

После формирования объединенного алфавита нужно было выделить уникальные символы, которые могли бы указывать на определенный диалект. Самый простой вариант, когда в других алфавитах нет даже похожих символов. Например, класс 002 алфавита Ирк Битиг на рис. 1. В таком случае эта руна входит в общий алфавит отдельным классом, а также становится «маркером» диалекта Ирк Битиг. Если она встречается в надписи, то это может указывать на ее принадлежность именно к этому наречию. Также уникальными символами могут стать и те руны, аналоги которых есть во всех алфавитах. Например, класс 219 Орхонского алфавита хоть и похож на руны из других алфавитов, имеет заметное отличие. Выделение его в отдельный класс не только улучшит его точность распознавания, но также поможет определить надпись с ним к Орхонскому наречию (рис. 3).

В результате сопоставления (рис. 4) был сформирован новый объединенный алфавит, содержащий руны всех четырех наречий.

СОЗДАНИЕ НОВОГО ОБЪЕДИНЕННОГО АЛФАВИТА

Процесс формирования нового объединенного алфавита является центральным этапом исследования, поскольку именно он создает основу для дальнейшей разметки изображений. На данном этапе четыре независимых набора: Кюль-Тегин, Ирк Битиг, орхонский и енисейский — будут интегрированы в единую систему классов, отображающую как общие, так и различающиеся руны древ-

034	↓	134	↑ ↓	234	↑ ↓	334	↓
035	⦿	135	⦿ ⦿	235	⦿ ⦿	335	⦿
036	⦿				⦿	336	
037	≡	137	:	237	:	337	:
038	⦿	Знак числа - идет обычно между страницами					338
		139	◁ ▷	239	◁ ▷	339	◁
		140	М	240	М	340	М

Рис. 4 / Fig. 4. Визуальные различия рун четырех наречий: Ирк Битиг, Енисейского, Орхонского и Кюль-Тегин / Visual Differences Between Runes of Four Dialects: Irk Bitig, Yenisei, Orkhon, and Kül Tegin

Источник / Source: составлено авторами / Compiled by the authors.

нетюркской письменности. Ранее был собран набор классов для начертаний рун 4 диалектов (рис. 1), который мы собираемся объединить в один общий.

Объединение проводилось поэтапно. Сначала были сопоставлены все руны, встречающиеся в каждом из четырех источников. На основе визуального сходства, фонетических интерпретаций и историко-филологических описаний были выделены группы идентичных или функционально эквивалентных символов. Если руна из разных диалектов совпадала по структуре, но различалась стилистическими особенностями (толщина линий, углы штрихов, орнаментальные элементы), она включалась в общий класс как вариант начертания. Например, классы 105 и 204 на рис. 1. В случаях, когда символы демонстрировали устойчивые различия (структурные расхождения, различное число элементов, уникальные добавочные штрихи), они фиксировались как отдельные классы, которые будут упомянуты в следующем разделе. Следовательно, объединенный алфавит формировался не как простое пересечение четырех наборов, а как их интегративная схема, отражающая внутреннюю вариативность и устраняющая многократное дублирование классов.

Рассмотрим структуру нового алфавита и количество классов. В результате сопоставления был получен консолидированный набор рун, включающий:

- базовые руны, встречающиеся во всех четырех наборах;
- вариативные руны, имеющие отличающиеся начертания при сохранении общей функции;
- уникальные символы, характерные только для одного или двух источников и отсутствующие в остальных.

301	↘ ↗
302	✂
303	└
304	>
305	4 4
306	5
307	8 8
308	3 3
309	X
310	1 1
311	7 7
312	D
313	9 9
314	H
315	Y Y
316	R
317	✓
318	Y
319	→

Рис. 5 / Fig. 5. Объединенный набор рун /
Combined Set of Runes

Источник / Source: составлено авторами / Compiled by the authors.

Каждому элементу присвоен идентификатор класса, не зависящий от конкретного наречия. Такая структура позволяет использовать набор и для аннотации данных, и для написания программных алгоритмов, работающих с конкретными символами.

Итоговое количество классов определяется числом структурно различных рун. Для практического применения в задаче классификации каждый класс соответствует одному элементу пространства меток. Также мы сохранили вспомогательный класс неизвестной руны, которым можно пометить символы, не вошедшие в нашу классификацию.

Весь набор представлен в виде таблицы соответствий, в которой для каждого объединенного класса указаны:

- идентификатор руны, общей для четырех наборов;
- визуальные варианты начертания.

Стандартизация меток обеспечила совместимость разметки, корректный экспорт в форматы YOLO/CNN и единообразие аннотаций при работе с графическим корпусом. Фрагмент итогового общего алфавита представлен на рис. 5.

Итоги формирования общего алфавита: выделена 51 руна и один вспомогательный класс. В новом объединенном алфавите каждый символ формально принадлежит одному классу, даже если встречается в нескольких источниках. Теперь разметка текста становится однозначной. Каждый символ имеет непротиворечивую и стабильную метку.

ВЫДЕЛЕНИЕ УНИКАЛЬНЫХ ПРИЗНАКОВ НАРЕЧИЙ

После формирования объединенного алфавита стала возможна систематическая оценка распределения рун по четырем диалектам. Систематизация знаков позволила сопоставлять символы напрямую, не учитывая различия в стилевых вариациях, что послужило основой для выявления уникальных рун, характерных для определенных диалектов.

Для отнесения руны к категории уникальных использовались следующие критерии:

1. Структурная автономность: символ должен сохранять отличимую форму, не являющуюся стилевым вариантом другой руны в объединенном алфавите.

2. Локализованность: руна не присутствует хотя бы в одном из наборов.

При анализе алфавитов четырех диалектов была составлена таблица сопоставления. В ней указано, к каким наречиям относятся руны (рис. 6).

Там же отмечены уникальные руны для конкретного наречия (рис. 7) и выбывающие руны с другой вариацией начертания (рис. 8).

1		Енисейский, Орхонский (второй вариант)
2		Енисейский, Орхонский, Кюль-Тегин
3		Ирк Битиг
4		Енисейский, Орхонский, Кюль-Тегин
5		Енисейский, Орхонский (первый вариант)
6		Енисейский, охронский, Ирк Битиг

Рис. 6 / Fig. 6. Общие руны, с указанием к каким наречиям они относятся / General Runes, with an Indication of which Dialects they Belong to

Источник / Source: составлено авторами / Compiled by the authors.

7		Кюль-тегин
8		Енисейский

Рис. 7 / Fig. 7. Уникальные руны для одного наречия / Unique Runes for One Dialect

Источник / Source: составлено авторами / Compiled by the authors.

54		Енисейский — это просто перевернутый символ 4
----	---	---

Рис. 8 / Fig. 8. Руна с другой вариацией начертания / A Rune with an Alternative Form of Inscription

Источник / Source: составлено авторами / Compiled by the authors.

Выделение уникальных рун имеет ключевое значение для статистических и аналитических исследований в этой области. Они позволяют повысить точность определения происхождения надписей. Кроме того, наличие таких наборов дает возможность применять алгоритм, где сначала определяется сам текст, а затем по уникальным символам — предположительное наречие надписи.

Таким образом, выявление уникальных признаков наречий стало возможным благодаря созданию общего алфавита. Полученные результаты демонстрируют, что даже в пределах близкородственных письменностей сохраняется выраженная диалектная специфика.

ИТОГ СОСТАВЛЕНИЯ АЛФАВИТА

Проведенный анализ отдельных диалектов, составление общего алфавита и поиск уникальных символов показали, что уникальные признаки описывают отличительные особенности различных наречий:

- *Кюль-Тегин* характеризуется наличием нескольких сложных по структуре и хорошо стандартизированных рун, что связано с монументальным характером надписи и строгой эпиграфической нормой.
- *Ирк Битиг* демонстрирует набор упрощенных символов и специфических вариантов начертания, обусловленных рукописной традицией и большей свободой писца.
- *Орхонские памятники* содержат руны, устойчиво встречающиеся в надписях на камне, незначительно отличающиеся от енисейских вариантов формой штрихов и структурой внутренних элементов.
- *Енисейская письменность* выделяется присутствием нескольких редких, но строго локальных рун, отсутствующих в других наборах, что делает их ключевыми для атрибуции фрагментов.

В совокупности эти различия позволяют говорить о наличии выраженной региональной стратификации внутри рунической традиции.

ПРИМЕНЕНИЕ В ЗАДАЧАХ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Алфавитом, который получился в результате этого исследования, будет размечен корпус изображений рунических надписей. Это изображения реальных памятников, а также перенесенные на бумажные и цифровые носители тексты. Так как корпус включает изображения памятников из разных регионов и созданных в различные эпохи, то и надписи на них могут относиться к разным диалектам. Использование алфавита со всеми диалектами позволит разметить эти снимки наиболее полно. При разметке изображений каждая руна получает уникальную метку из объединенной системы, что делает данные совместимыми с современными архитектурами нейросетей, такими как YOLO и сверточные нейронные сети (CNN).

Модель YOLO используется на этапе детекции — она определяет координаты символов на изобра-

жении и присваивает им классы из единого набора [17]. Сверточные сети применяются на этапе классификации для уточнения принадлежности выделенных фрагментов к конкретным рунам. Унификация алфавита позволяет обучать обе модели на согласованных данных и устраняет проблему несовпадений меток между разными корпусами.

Описанный в работе подход помогает модели формировать устойчивые векторные представления и уверенно различать даже схожие по форме символы. Это повышает точность и обобщающую способность системы, делая ее пригодной для работы с надписями из разных регионов, эпох и носителей.

В исследовании распознавания рун алфавита Кюль-Тегин возникла техническая проблема из-за символов, не вошедших в набор. Они подходили для обучения модели обнаружения, но модель классификации, наоборот, сбивали. В итоге модель ошибочно относила часть известных рун к этому классу. Использование объединенного алфавита должно решить эту проблему. В процессе обучения классификатора предполагается отказ от использования класса неизвестной руны. Для предотвращения искажения процесса обучения будут закрашены. Ожидается, что использование только известных рун обеспечит достаточный объем данных для обучения, а ансамбль моделей детекции и классификации продемонстрирует более высокие показатели по сравнению с предыдущими экспериментами.

ВЫВОДЫ

Проделанная работа показала, что объединение четырех классификаций древнетюркских рун в единый алфавит действительно помогает упорядочить разрозненный материал и устранить противоречия, накопившиеся в разных исследовательских традициях. Полученные результаты согласуются с выводами современных исследований в области

цифровой эпиграфики и компьютерного зрения, в которых подчеркивается ключевая роль стандартизации данных для повышения устойчивости и воспроизводимости моделей машинного обучения. Единый набор символов делает разметку более предсказуемой и сокращает количество ошибок, которые обычно возникают при работе с несколькими несогласованными системами. Это важно для задач машинного обучения, где качество данных напрямую влияет на итоговую точность моделей.

Получившийся алфавит охватывает весь спектр рун четырех диалектов орхоно-енисейской письменности и одновременно фиксирует различия между вариантами начертания. Такой формат позволяет корректно применять систематизацию при обучении моделей и избежать ситуации, когда непохожие по форме руны оказываются в одном классе, а визуально близкие — в разных.

Отдельной задачей стало выделение рун, характерных только для одного наречия. Этот шаг помогает не только распознавать символы, но и учитывать особенности происхождения надписей. Такие маркеры могут служить инструментом атрибуции и обогатить аналитическую часть будущих исследований.

Подготовленный алфавит создает прочную основу для разработки многоалфавитной системы компьютерного зрения. Он упрощает обучение нейросетей на разнородных изображениях, повышает стабильность распознавания и помогает моделям уверенно работать с текстами, созданными в разных регионах и эпохах.

Следующим шагом станет расширение корпуса изображений и проверка созданной структуры в реальных задачах автоматического распознавания. Новый алфавит уже используется для аннотации изображений, которые станут основой для дальнейших исследований тюркской письменности с помощью технологий искусственного интеллекта.

СПИСОК ИСТОЧНИКОВ

1. Убрытова Е.И. Сергей Ефимович Малов: (К 75-летию со дня рождения). В сб.: Тюркологический сборник. М.: Издательство Академии наук СССР; 1951. URL: https://www.orientalstudies.ru/rus/images/pdf/journals/p_ts_1_1951_02_ubryatova.pdf
2. Кляшторный С.Г. Древнетюркские рунические памятники как источник по истории Средней Азии. М.; 1964. 215 с. URL: <https://www.elibrary.ru/qaujxn>
3. Кызласов И.Л. Рунические письменности евразийских степей. М.: Наука; 1994. 318 с. URL: <https://www.elibrary.ru/wjrref>
4. Пенская Е.Н. Машинное обучение и представление информации: новые возможности цифровых архивов. *Имагология и компаративистика*. 2025;23:380-389. URL: <https://doi.org/10.17223/24099554/23/18>
5. Примова Ф.М. Искусственный интеллект в филологии. *Международная конференция академических наук*. 2025;4(1)63-65. URL: <https://doi.org/10.5281/zenodo.14682779>



6. Рудакова С.В. Особенности взаимодействия филологии с искусственным интеллектом. В сб.: Общество. Наука. Инновации (НПК-2024). Киров: Вятский государственный университет; 2024:70-73. URL: <https://www.elibrary.ru/eeekohu>
7. Матвеева Т.В. Категориально-текстовый анализ речевого произведения: обоснование и применение метода. *Quaestio Rossica*. 2024;12(3):901-920. URL: <https://doi.org/10.15826/qr.2024.3.915>
8. Кочкаров Р.А., Бородин А.Д. Распознавание орхоно-енисейских рунических надписей методами машинного обучения. *Нелинейный мир*. 2025;23(3):96-106. URL: <https://doi.org/10.18127/j20700970-202503-12>
9. Кормушин И.В. О взглядах Э.Р. Тенишева на историю древних литературных тюркских языков. *Вестник Института языкознания РАН*. 2021;1-2 (30-31):45-53. URL: <https://www.elibrary.ru/mbprias>
10. Кочкаров У.Ю., Беляева В.Н., Кочкаров Р.А., Кочкаров А.А. Подготовка набора визуальных данных для машинного распознавания рунической письменности. В сб.: Проблематика и методология современных тюркологических исследований. Коллективная монография. М.: Издательство МБА; 2025:162-173. URL: <https://www.elibrary.ru/chxcmz>
11. Лебедев Ю.С., Попов П.В. Погребение VIII-IX вв. из Астраханской области и горшок с рунической надписью. *Российская археология*. 2023;1:178-186. URL: <https://doi.org/10.31857/S0869606323010130>
12. Мунчаев Р.М., Корневский С.Н., ред. Проблемы археологии Кавказа. Вып. 1. М.: ТAYC; 2012. 248 с. URL: <https://www.elibrary.ru/qpxfjv>
13. Кормушин И.В. Тюркские енисейские эпитафии. Монография. М.: Наука; 1997. 338 с. URL: <https://www.elibrary.ru/sfhmdj>
14. Томсен В.Л.П. Дешифровка орхонских и енисейских надписей. Ада Чир-Суу — Отечество: краеведческий альманах. 2018;4:129-134. URL: <https://www.elibrary.ru/ntspvz>
15. Володин А.Ю. Цифровые гуманитарные исследования в контексте датаизма. *Современная наука: Актуальные проблемы теории и практики. Серия: Гуманитарные науки*. 2024;7(2):6-9. URL: <https://doi.org/10.37882/2223-2982.2024.7-2.05>
16. Сухан Д.Д., Плюснина Е.А. Метаразметка и визуализация данных в корпусе текстов по корпусной лингвистике. В сб.: Компьютерная лингвистика и вычислительные онтологии. Вып. 8. СПб.: Университет ИТМО; 2024:45-60. URL: <https://doi.org/10.17586/2541-9781-2024-8-45-60>
17. Кочкаров Р.А., Бородин А.Д. Систематизация орхоно-енисейских рунических письменностей и перспективы их автоматизированного анализа методами машинного обучения. *Динамика сложных систем — XXI век*. 2025;19(5):97-103. URL: <https://doi.org/10.18127/j19997493-202505-11>

REFERENCES

1. Ubryatova E.I. Sergey Efimovich Malov: (On the 75th Anniversary of His Birth). In: Turkological Collection. Moscow: Publishing House of the USSR Academy of Sciences; 1951. URL: https://www.orientalstudies.ru/rus/images/pdf/journals/p_ts_1_1951_02_ubryatova.pdf (In Russ.).
2. Klyashtorny S.G. Ancient Turkic Runic Monuments as a Source on the History of Central Asia. Moscow; 1964. 215 p. URL: <https://www.elibrary.ru/qaujxn> (In Russ.).
3. Kyzlasov I.L. Runic Scripts of the Eurasian Steppes. Moscow: Nauka; 1994. 318 p. URL: <https://www.elibrary.ru/wjrref> (In Russ.).
4. Penskaya E.N. Machine Learning and Information Presentation: New Opportunities for Digital Archives. *Imagology and Comparative Studies*. 2025;23:380-389. (In Russ.). URL: <https://doi.org/10.17223/24099554/23/18>
5. Primova F.M. Artificial Intelligence in Philology. *International Conference of Academic Sciences*. 2025;4(1):63-65. (In Russ.). URL: <https://doi.org/10.5281/zenodo.14682779>
6. Rudakova S.V. Features of the Interaction of Philology with Artificial Intelligence. In: Society. Science. Innovations (NPC 2024). Kirov: Vyatka State University; 2024:70-73. URL: <https://www.elibrary.ru/eeekohu> (In Russ.).
7. Matveeva T.V. Categorical-textual analysis of a speech work: Justification and application of the method. *Quaestio Rossica*. 2024;12(3):901-920. (In Russ.). URL: <https://doi.org/10.15826/qr.2024.3.915>
8. Kochkarov R.A., Borodina A.D. Recognition of Orkhon-Yenisei Runic Inscriptions Using Machine Learning Methods. *Non-Linear World*. 2025;23(3):96-106. (In Russ.). URL: <https://doi.org/10.18127/j20700970-202503-12>
9. Kormushin I.V. On E.R. Tenishev's Views on the History of Ancient Literary Turkic Languages. *Bulletin of the Institute of Linguistics of the Russian Academy of Sciences*. 2021;1-2(30-31):45-53. URL: <https://www.elibrary.ru/mbprias> (In Russ.).

10. Kochkarov U.Yu., Belyaeva V.N., Kochkarov R.A., Kochkarov A.A. Preparation of a set of visual data for machine recognition of runic writing. In: Problems and Methodology of Modern Turkological Research. Monograph. Moscow: MBA Publishing House; 2025. URL: <https://www.elibrary.ru/chxcmz> (In Russ.).
11. Lebedev Yu.S., Popov P.V. Burial of the 8th-9th Centuries from the Astrakhan Region and a Pot with a Runic Inscription. *Russian Archaeology*. 2023;1:178-186. (In Russ.). URL: <https://doi.org/10.31857/S0869606323010130>
12. Munchaev R.M., Korenevsky S.N., ed. Problems of the Caucasus Archaeology. Vol. 1. Moscow: TAUS; 2012. 248 p. URL: <https://www.elibrary.ru/qpxfjv> (In Russ.).
13. Kormushin I.V. Turkic Yenisei epitaphs. Monograph. Moscow: Nauka; 1997. URL: <https://www.elibrary.ru/sfhmdj> (In Russ.).
14. Tomsen V.L.P. Deciphering the Orkhon and Yenisei Inscriptions. *Ada Chir Suu — Fatherland: Local History Almanac*. 2018;4:129-134. URL: <https://www.elibrary.ru/ntspvz> (In Russ.).
15. Volodin A. Yu. Digital Humanities Research in the Context of Dataism. *Modern Science: Current Issues in Theory and Practice. Series: Humanities Sciences*. 2024;7(2):6-9. (In Russ.). URL: <https://doi.org/10.37882/2223-2982.2024.7-2.05>
16. Suhan D.D., Plyusnina E.A. Meta-markup and data visualization in the corpus of texts on corpus linguistics. In: Computer linguistics and computational ontologies. Issue 8. St. Petersburg: ITMO University; 2024:45-60. (In Russ.). URL: <https://doi.org/10.17586/2541-9781-2024-8-45-60>
17. Kochkarov R.A., Borodina A.D. Systematization of the Orkhon-Yenisei Runic Scripts and Prospects for Their Automated Analysis Using Machine Learning Methods. *Dynamics of Complex Systems — 21st Century*. 2025;19(5):97-103. (In Russ.). URL: <https://doi.org/10.18127/j19997493-202505-11>

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Анастасия Дмитриевна Бородин — студентка бакалавриата, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Anastasia D. Borodina — Bachelor's degree student, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<https://orcid.org/0009-0002-5126-9854>
225034@edu.fa.ru

Монгун-Ай Вячеславовна Кызыл-оол — студентка бакалавриата, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Mongun-Ai V. Kyzyl-ool — Bachelor's degree student, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<https://orcid.org/0009-0006-5385-9603>

Автор для корреспонденции / Corresponding author:
221361@edu.fa.ru

Расул Ахматович Кочкаров — доктор технических наук, заместитель декана по научной работе, доцент кафедры искусственного интеллекта факультета информационных технологий и анализа больших данных, Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

Rasul A. Kochkarov — Dr. Sci. (Tech.), Deputy Dean for Research, Associate Professor of the Department of Artificial Intelligence, Faculty of Information Technology and Big Data Analysis, Financial University under the Government of the Russian Federation, Moscow, Russian Federation
<https://orcid.org/0000-0003-3186-3901>
rkochkarov@fa.ru

Заявленный вклад авторов:

А.Д. Бородин — участие в систематизации алфавитов, анализе визуальных различий рун и подготовке экспериментальной части исследования.

М.В. Кызыл-оол — участие в анализе источников, формировании объединенного алфавита, подготовка иллюстративных материалов и текста статьи.

Р.А. Кочкаров — разработка общей концепции исследования, научное руководство, анализ результатов и подготовка финальной версии текста.

Authors' declared contributions:

A. D. Borodina — participation in alphabet systematization, analysis of visual differences in runes, and preparation of the experimental part of the study.

M. V. Kyzyl-ool — participation in source analysis, development of the unified alphabet, preparation of visual materials and manuscript drafting.

R. A. Kochkarov — development of the general research concept, scientific supervision, analysis of results, and preparation of the final version of the manuscript.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила 22.04.2026; после рецензирования 10.07.2026; принята к публикации 01.08.2026.

Авторы прочитали и одобрили окончательный вариант рукописи.

The article was submitted on 22.04.2026; revised on 10.07.2026; accepted for publication on 01.08.2026.

The authors read and approved the final version of the manuscript.