

Научная статья

УДК 004.412

<https://doi.org/10.35266/1999-7604-2024-3-3>



Сравнительный анализ метрик векторного расстояния растровых изображений

*Александр Александрович Детков*¹✉, *Валерия Алексеевна Воронина*²,
*Юлия Владимировна Гарифуллина*³, *Андрей Максимович Корепанов*⁴,
*Алина Юрьевна Вишнякова*⁵

^{1, 2, 3, 4, 5}Уральский федеральный университет имени первого Президента России Б. Н. Ельцина,
Екатеринбург, Россия

¹samik1982@mail.ru✉

³u.garifullina@yandex.ru

⁵Alina.vishniakova@urfu.ru

Аннотация. В статье описаны метрики и алгоритмы, позволяющие оценить различие или сходство следующих исследуемых объектов: расстояние Хэмминга, расстояние Левенштейна, сходство Джаро – Винклера, коэффициент Серенсена, SIFT, перцептивный хэш-алгоритм. С целью выявления алгоритма, оптимально подходящего для практического применения, а именно для оценки уникальности логотипов путем сравнения их друг с другом, метрики и алгоритмы были реализованы в виде программ на языке программирования Python и протестированы на предмет скорости и качества их работы. Наиболее быстро работающий алгоритм сравнения логотипов основан на расчете перцептивной хэш-суммы с последующим вычислением расстояния Хэмминга, наиболее точные результаты сравнения логотипов были получены в процессе работы программы, основанной на расчете расстояния Хэмминга. В качестве оптимального алгоритма, объединяющего в себе высокую скорость работы и качество полученных результатов, был выбран программно-реализованный алгоритм, основанный на расчете расстояния Хэмминга.

Ключевые слова: метрика, расстояние, изображение, сходство, различие, сравнение изображений, алгоритм, Python

Для цитирования: Детков А. А., Воронина В. А., Гарифуллина Ю. В., Корепанов А. М., Вишнякова А. Ю. Сравнительный анализ метрик векторного расстояния растровых изображений // Вестник кибернетики. 2024. Т. 23, № 3. С. 22–30. <https://doi.org/10.35266/1999-7604-2024-3-3>.

Original article

Comparative analysis of vector distance metrics for raster images

*Aleksandr A. Detkov*¹✉, *Valeriya A. Voronina*², *Yuliya V. Garifullina*³,
*Andrey M. Korepanov*⁴, *Alina Yu. Vishnyakova*⁵

^{1, 2, 3, 4, 5}Ural Federal University named after the First President of Russia B. N. Yeltsin,
Yekaterinburg, Russia

¹samik1982@mail.ru✉

³u.garifullina@yandex.ru

⁵Alina.vishniakova@urfu.ru

Abstract. The article describes metrics and algorithms that allow evaluating the difference or similarity of the objects under study: Hamming distance, Levenshtein distance, Jaro-Winkler similarity, Dice-Sørensen coefficient, SIFT, Perceptual Hash algorithm. To identify an algorithm that is optimally suitable for evaluating the uniqueness of logos by comparing them with each other, we implemented metrics and algorithms as programs in the Python

programming language. Then, we tested their speed and quality of work. The fastest working algorithm for comparing logos is based on the calculation of a perceptual hash sum, followed by the calculation of the Hamming distance. The most accurate results of comparing logos were obtained during the operation of the program based on the calculation of the Hamming distance. We chose a software-implemented algorithm based on Hamming distance calculation as the optimal algorithm that combines high speed of operation and the quality of the obtained results.

Keywords: metric, distance, image, similarity, difference, image comparison, algorithm, Python

For citation: Detkov A. A., Voronina V. A., Garifullina Yu. V., Korepanov A. M., Vishnyakova A. Yu. Comparative analysis of vector distance metrics for raster images. *Proceedings in Cybernetics*. 2024;23(3):22–30. <https://doi.org/10.35266/1999-7604-2024-3-3>.

ВВЕДЕНИЕ

Метрика – это функция, определяющая расстояние в метрическом пространстве; мера, определяющая расстояние между элементами множества [1]. Метрика определена свойствами: больше или равна нулю; равна нулю при условии, что элементы множества совпадают; симметричная; подчиняется закону треугольника. Введение метрики на множестве определяет метрическое пространство [2]. Метрическое пространство – это множество, в котором определено расстояние между любой парой элементов [3].

Понятие расстояния введено для количественного измерения сходства и различия двух последовательностей. Мера расстояния – функция, присваивающая численное значение паре последовательностей: чем больше расстояние, тем меньше подобие [4].

В настоящее время известно множество метрик, позволяющих оценить различие или сходство исследуемых объектов, например, расстояние Хэмминга, расстояние Левенштейна, сходство Джаро – Винклера, коэффициент Серенсена, каскад Хаар и др.

Метрики сходства или различия широко используются и применяются для решения различных задач – распознавания образов, классификации документов, исправления ошибок, обнаружения вирусов, обнаружения дубликатов и т. д.

Так, метрики сходства или различия применимы для такой предметной области как «интеллектуальная собственность», в узком смысле «правовая охрана средств индивидуализации». Средства индивидуализации – обозначения, при помощи которых коммерческие фирмы и их продукция, а именно

товары и услуги, идентифицируются среди множества других подобных, приобретают уникальность. Правовая охрана интеллектуальной собственности, согласно Статье 1232 ГК РФ, предоставляется на основе государственной регистрации, которую осуществляет федеральный орган исполнительной власти Роспатент [5]. По статистике, около 30 % поданных заявок на регистрацию обозначения получают отказ, на основании наличия тождественных или схожих до степени смешения товарных знаков, зарегистрированных ранее. Проведение предварительной проверки обозначения на предмет наличия схожих или тождественных товарных знаков, зарегистрированных ранее, позволит снизить риски отказа в его государственной регистрации. Использование метрик позволит автоматизировать процесс проверки изобразительного обозначения: уменьшить сроки, повысить качество проверки и минимизировать человеческое влияние на результат.

Целью данного исследования является разработка и оценка эффективности и скорости работы программ, позволяющих выявить различие или сходство двух изобразительных обозначений между собой, путем расчета метрик измерения расстояния между двумя последовательностями.

Задачи исследования:

1. Поиск и анализ научных статей на тему расчета метрик измерения расстояний между последовательностями, их применения для решения различных задач и возможности их программной реализации.

2. Поиск и изучение существующих метрик, измеряющих расстояния между последовательностями.

3. Анализ метрик с точки зрения их применимости для сравнения изобразительных обозначений.

4. Программная реализация алгоритма расчета метрик различия или схожести изобразительных обозначений.

5. Проведение испытаний и оценка эффективности и скорости работы программ при сравнении изобразительных обозначений, основанных на расчете метрик различия или схожести.

МАТЕРИАЛЫ И МЕТОДЫ

Источниками информации в области товарных знаков и их государственной регистрации выступили Руководство по осуществлению административных процедур и действий в рамках предоставления государственной услуги по регистрации товарного знака [6] и четвертая часть Гражданского кодекса Российской Федерации [5]. В качестве источника информации о метриках измерения расстояния между последовательностями были использованы научные труды российских и зарубежных ученых. Для поиска и проведения литературного обзора научных статей были использованы ресурсы поисковой системы eLIBRARY, поиск проведен по вышеуказанным ключевым словам. Для поиска и анализа программных кодов, реализующих расчет метрик расстояния, были использованы веб-сервисы, позволяющие просматривать исходные коды программ [7]. Для оценки применимости и эффективности использования различных метрик расстояния между последовательностями были рассчитаны следующие величины: расстояние Хэмминга, расстояние Левенштейна, расстояние Джаро – Винклера, мера Серенсена.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

В ходе проведения исследования была найдена и рассмотрена статья: «Актуальные задачи анализа изображений», подтверждающая актуальность выбранной темы. В статье авторы (А. И. Либерман и Т. Э. Шульга) представляют обзор современного состояния интеллектуального анализа изобра-

жений, рассматривают сферы применения анализа изображений и анализируют методы искусственного интеллекта для решения различных практических задач. Авторы приходят к выводу о возможности комбинации методов анализа изображений и онтологического моделирования с целью дальнейшего развития области анализа изображений [8].

Е. Ю. Коломийцева в статье «Анализ изображений в интернете: возможности современных сервисов» рассматривает анализ изображений с помощью искусственного интеллекта с точки зрения применимости и ценности в области журналистской работы. Автор приводит примеры параметров изображений, которые в настоящее время можно анализировать с использованием современных сервисов, например, такие параметры как определение шрифта, определение того, что изображено на фото, распознавание эмоций человека, определение личности человека и другое [9].

В статье «Анализ изображений. Определение дубликатов и степени размытости изображений», автором которой является Д. В. Шестоперов, было предложено решение проблемы затрудненного поиска фотографий в большом файловом хранилище, не имеющем фильтрации – анализ изображений и удаление неподходящих фотографий с использованием алгоритма разностного хеширования и сравнения хэшей при помощи расстояния Хэмминга для поиска дубликатов изображений [10].

Л. К. Ситниковой, М. О. Еланцевым и Р. О. Султановым в статье «Двухэтапный метод поиска изображений в большой базе изображений» были зафиксированы результаты анализа способов поиска в базе данных большого размера. Авторами был сделан вывод о том, что совместное применение метода сравнения по перцептивному хэшу и метода классификации при помощи нейронной сети является наиболее эффективным [11].

После рассмотрения научных статей и публикаций были сделаны выводы об актуальности исследуемой темы и определены

наиболее часто используемые для анализа изображений метрики измерения расстояний между последовательностями.

Метрики и алгоритмы, позволяющие оценить различие или сходство исследуемых объектов:

1. Расстояние Хэмминга – метрика различия объектов одинаковой размерности; метрика измерения расстояния между двумя последовательностями. Метрикой для определения меры различия между двоичными векторами в векторном пространстве кодовых последовательностей – расстоянием Хэмминга $d(x; y)$ между двумя двоичными последовательностями – векторами: x и y , длины n – называется число позиций, в которых они различны. В общем виде расстояние Хэмминга для объектов и размерностей задается функцией [12]:

$$d_{ij} = \sum_{k=1}^p |x_{ik} - x_{jk}|. \quad (1)$$

2. Расстояние Левенштейна – метрика сходства между двумя последовательностями; метрика измеряющая по модулю разность между двумя последовательностями символов. Расстояние Левенштейна определяется как минимальное количество односимвольных операций – вставки, удаления, замены, необходимых для превращения одной последовательности элементов в другую. Пусть S_1 и S_2 – две строки некоторого алфавита, длины которых M и N соответственно, тогда расстояние Левенштейна $d(S_1, S_2)$, можно рассчитать по формуле:

$$D(i, j) = \begin{cases} d(S_1, S_2) = D(M, N), \text{ где} \\ 0, & i = 0, j = 0 \\ i, & j = 0, i > 0 \\ \min\{ \\ D(i, j - 1) + 1, \\ D(i - 1, j) + 1, & j = 0, i > 0 \\ D(i - 1, j - 1) + m(S_1[i], S_2[j]) \\ \} \end{cases} \quad (2)$$

где $m(a, b)$ равна нулю, если $a = b$ и единице в противном случае $\min\{a, b, c\}$ возвращает наименьшее из аргументов. Шаг по i символизирует удаление из первой строки, по

j – вставку в первую строку, а шаг по обоим индексам символизирует замену символа или отсутствие изменений [13].

3. Сходство Джаро – Винклера – мера схожести для измерения расстояния между двумя последовательностями символов.

Расстояние Джаро – Винклера – это минимальное число односимвольных преобразований, которое необходимо для того, чтобы изменить одно слово в другое. Чем меньше расстояние Джаро – Винклера d_w для двух последовательностей, тем больше сходства имеют эти последовательности друг с другом. Результат нормируется так, что $d_w = 0$ – отсутствие сходства, $d_w = 1$ – точное сходство. Сходство Джаро – Винклера равно $1 - d_w$.

При расчете расстояния Джаро – Винклера используется коэффициент масштабирования p , что позволяет присвоить более благоприятные рейтинги строкам, которые совпадают друг с другом от начала до определенной длины l , которая называется префиксом.

Пусть S_1 и S_2 – две строки, расстояние Джаро – Винклера d_w можно рассчитать по формуле:

$$d_w = d_j + (l_p (1 - d_j)), \quad (3)$$

где d_j – расстояние Джаро для строк S_1 и S_2 ;

l – длина общего префикса от начала строки до максимума четырех символов;

p – постоянный коэффициент масштабирования, использующийся для того, чтобы скорректировать оценку в сторону повышения для выявления наличия общих префиксов: $p = const = 0,1$ [14].

4. Коэффициент Серенсена – бинарная мера сходства. Коэффициент Серенсена создан для определения схожести любых множеств. Коэффициент Серенсена K_s рассчитывается по формуле:

$$K_s = \frac{2c}{a + b}, \quad (4)$$

где a – количество видов первой экосистемы;

b – количество видов второй экосистемы;

c – количество общих для 1-й и 2-й экосистемы видов [15].

5. Алгоритм SIFT (масштабно-инвариантная трансформация признаков) – алгоритм выявления признаков в компьютерном зрении для выявления и описания локальных признаков в изображениях.

Первостепенно в алгоритме SIFT извлекаются ключевые точки объектов из набора изображений и запоминаются в базе данных. Заданный объект в новом изображении распознается путем сравнения признаков из нового изображения с признаками из базы данных и нахождения признаков-кандидатов на основе евклидова расстояния между векторами признаков. Ключевые точки отбираются из набора соответствий в новом изображении, на основании наиболее близкого сходства с объектом по его местоположению, масштабу и ориентации. Подходящие блоки признаков определяются с помощью реализации хэш-таблицы обобщенного преобразования Хафа, каждый согласующийся с объектом блок из трех или более признаков подлежит дальнейшей подробной проверке. Резко отклоняющиеся блоки признаков отбрасываются. Определяется набор признаков, который с высокой вероятностью подтвердит наличие искомого объекта на рассматриваемом изображении [16].

6. Перцептивный хэш-алгоритм. Хэш-алгоритмы основаны на генерации сравнимых хэшей. Данные, используемые для генерации хэша, выполняют роль источника случайных чисел – одинаковые данные дадут одинаковый результат, разные данные – разный результат. Характеристики изображений используются для генерации индивидуальных отпечатков, которые можно сравнивать друг с другом. Выводы, которые можно сделать на основании сравнения хэшей: хэши отличаются – данные разные; хэши совпадают – данные одинаковые. Изменение размера, соотношения сторон и цветовых характеристик (яркость, контраст и т. д.) сравниваемых изображений не повлияет на результат вычисления перцептивного хэша.

Этапы получения перцептивного хэша:

– уменьшение размера рассматриваемого изображения;

– перевод изображения в градации серого;
– вычисление среднего значения яркости получившегося изображения;

– бинаризация изображения, выборка пикселей, значение которых больше среднего;

– построение хэша, перевод полученных отдельных значений в одно 64-битное значение [17].

На основании рассмотренных выше методов были разработаны алгоритмы для сравнения логотипов с целью проверки их уникальности. Основой для алгоритмов выступили шесть методов: расстояние Хэмминга, расстояние Левенштейна, сходство Джаро – Винклера, коэффициент Серенсена, SIFT и метод перцептивной хэш-суммы с последующим вычислением расстояния Хэмминга для сравнения. Методы реализованы на языке программирования Python с использованием библиотек: OpenCV, NumPy, Editdistance, Strsimpy и ImageHash.

Для реализации алгоритма, основанного на расчете расстояния Хэмминга, использована формула расчета количества несовпадающих битов между двумя бинарными строками с применением библиотек: OpenCV, NumPy.

Расстояние Левенштейна было реализовано с помощью алгоритма Дамерау – Левенштейна, который позволяет находить минимальное количество операций редактирования, необходимых для превращения одной строки в другую, реализован с применением библиотек: Editdistance, NumPy.

Для сравнения строк на основе сходства Джаро – Винклера использована формула, которая учитывает совпадение символов в строках и их расстояние друг от друга, реализован с помощью библиотек: Strsimpy, NumPy.

Алгоритм, основанный на расчете коэффициента Серенсена, реализован как отношение числа общих символов в строках к общему числу символов, с применением библиотек: Strsimpy, NumPy.

Метод SIFT реализован с помощью построения гауссовых пирамид и поиска экстремумов на разных уровнях пирамиды, а также определения градиентов и ориентаций угло-

вых точек на изображении с использованием библиотеки OpenCV.

Перцептивный хэш-алгоритм был реализован путем вычисления хэш-суммы на основе яркости каждого пикселя и дальнейшего вычисления расстояния Хэмминга между хэш-суммами двух изображений, с использованием библиотеки ImageHash.

Разработанные алгоритмы протестированы путем сравнения заранее подготовленных тестовых логотипов между собой. Проверка работы алгоритмов была нацелена на получение результатов вывода программ, а также скорости их обработки.

Разработанными программами были обработаны четыре пары логотипов, размером 1024×1024 пикселя:

- 1) полностью черный/полностью белый (далее – черно-белые);
- 2) существующий логотип/его инвертированная версия (далее – инвертированные);
- 3) существующий логотип/точно такой же существующий логотип (далее – идентичные);
- 4) существующий логотип/другой существующий логотип (далее – разные).

Результаты тестирования представлены в табл. 1. Сходство логотипов изменяет-

ся в процентах, где 100 % – полное сходство, 0 % – отсутствие сходства.

Результаты работы программ подтверждают корректную работу алгоритмов: черно-белые логотипы – 0 % сходства, идентичные алгоритмы – 100 % сходства. Результаты сравнения инвестированной пары логотипов: расстояние Хэмминга – 0 % сходства, SIFT – 4 %, расстояние Левенштейна – 17 %, коэффициент Серенсена – 23 %, сходство Джаро – Винклера – 33 %, перцептивный хэш-алгоритм – 38 %. Результаты сравнения разных логотипов существенно отличаются: расстояние Хэмминга, расстояние Левенштейна и хэш-сумма – 60–70 % сходства, сходство Джаро – Винклера – 87 %, коэффициент Серенсена – 13 %, SIFT – 1 % сходства.

Результаты тестирования программ на предмет скорости их работы представлены в табл. 2.

Из результатов тестирования программ на скорость их работы видно, что в основе самого быстроработающего алгоритма лежит перцептивный хэш-алгоритм с последующим вычислением расстояния Хэмминга, далее по скорости работы идут метод SIFT и коэффициент Серенсена, после – расстояние Хэм-

Таблица 1

Результаты тестирования

Типы изображений	Метрики векторного расстояния					
	Расстояние Хэмминга	Расстояние Левенштейна	Сходство Джаро – Винклера	Коэффициент Серенсена	SIFT	Хэш-сумма
Черный/белый	0 %	0 %	0 %	0 %	0 %	0 %
Инвертированный	0 %	17 %	33 %	23 %	4 %	38 %
Идентичный	100 %	100 %	100 %	100 %	100 %	100 %
Разный	60 %	62 %	87 %	13 %	1 %	70 %

Примечание: составлено авторами.

Таблица 2

Результаты тестирования

Время выполнения метрик	Метрики векторного расстояния					
	Расстояние Хэмминга	Расстояние Левенштейна	Сходство Джаро – Винклера	Коэффициент Серенсена	SIFT	Хэш-сумма
Скорость обработки данных, с	1,3	1 281	45 108	1,08	0,249	0,212

Примечание: составлено авторами.

минга, расстояние Левенштейна и сходство Джаро – Винклера.

На основании проделанной работы и полученных результатов можно сделать вывод о том, что алгоритм, основанный на расчете расстояния Хэмминга, является оптимальным для сравнения логотипов с целью оценки их уникальности: высокая скорость обработки данных и точность результата сравнения. Неприменимой на практике является программа, основанная на расчете сходства Джаро – Винклера, скорость работы которой низкая, а качество результата работ невысокое.

ЗАКЛЮЧЕНИЕ

В ходе проведения исследования были проанализированы научные статьи, тематика которых связана с математическими метриками, позволяющими измерить расстояние между последовательностями, с применением данных метрик для решения различных задач, а также с возможностью программной реализации алгоритмов расчета расстояния между последовательностями. Обзор научных статей подтвердил актуальность темы исследования, позволил определить и изучить часто используемые в ходе анализа математиче-

ские метрики: расстояние Хэмминга, расстояние Левенштейна, сходство Джаро – Винклера, коэффициент Серенсена, алгоритм SIFT и перцептивный хэш-алгоритм.

На основании изученного материала был реализован программный алгоритм расчета метрик различия или сходства изобразительных обозначений. Для каждой метрики были применены собственные алгоритмы и формулы в целях их реализации.

Проведены испытания и оценка эффективности и скорости работы программ при сравнении изобразительных обозначений, основанных на расчете метрик различия или схожести путем сравнения заранее подготовленных тестовых логотипов между собой. В ходе проверки были получены результаты, проанализирована скорость их обработки, на основе чего была разработана программа, позволяющая выявить различие или сходство двух изобразительных обозначений между собой путем расчета метрик измерения расстояния между двумя последовательностями с помощью алгоритма, основанного на расчете расстояния Хэмминга, позволившего достичь целевого результата соотношения скорости и сходства логотипов.

Список источников

1. Большой энциклопедический словарь. 2-е изд., перераб. и доп. М. ; СПб. : Большая российская энциклопедия, Норинт, 2000. 1434 с.
2. Определение метрики и ее свойства. Шкалы: номинальная, порядковая, интервальная, относительная. URL: <https://studfile.net/preview/8985938/page:7/> (дата обращения: 04.05.2024).
3. Битюцков В. И., Войцеховский М. И., Иванов А. Б. и др. Математическая энциклопедия. М. : Советская Энциклопедия, 1984. 1246 с.
4. Метрика (математика). URL: [https://ru.wikibrief.org/wiki/Metric_\(mathematics\)](https://ru.wikibrief.org/wiki/Metric_(mathematics)) (дата обращения: 04.05.2024).
5. Гражданский кодекс Российской Федерации (часть четвертая) от 18.12.2006 № 230-ФЗ. Доступ из СПС «КонсультантПлюс».
6. Об утверждении Руководства по осуществлению административных процедур и действий в рамках предоставления государственной услуги по государственной регистрации товарного знака, знака обслуживания, коллективного знака и выдаче свидетельств на товарный знак, знак обслуживания,

References

1. Bolshoy entsiklopedicheskiy slovar. 2nd ed., revised. Moscow; Saint Petersburg: Bolshaya rossiyskaya entsiklopediya, Norint, 2000. 1434 p. (In Russ.).
2. Opredelenie metriki i ee svoystva. Shkaly: nominalnaya, poryadkovaya, intervalnaya, otnositelnaya. URL: <https://studfile.net/preview/8985938/page:7/> (accessed: 04.05.2024). (In Russ.).
3. Bityutskov V. I., Voytsekhovskiy M. I., Ivanov A. B. et al. Matematicheskaya entsiklopediya. Moscow: Sovetskaya Entsiklopediya, 1984. 1246 p. (In Russ.).
4. Metric (mathematics). URL: [https://ru.wikibrief.org/wiki/Metric_\(mathematics\)](https://ru.wikibrief.org/wiki/Metric_(mathematics)) (accessed: 04.05.2024). (In Russ.).
5. Civil Code of the Russian Federation (Pt. 4) of 18.12.2006 No. 230 Federal Law. Accessed through Law assistance system “Consultant Plus”. (In Russ.).
6. Ob utverzhdenii Rukovodstva po osushchestvleniyu administrativnykh protsedur i deystviy v ramkakh predostavleniya gosudarstvennoy uslugi po gosudarstvennoy registratsii tovarnogo znaka, znaka obsluzhivaniya, kollektivnogo znaka i vydache svidetelstv na tovarnyy znak, znak obsluzhivaniya, kollektivnyy znak,

- коллективный знак, их дубликатов : приказ ФГБУ ФИПС 20.01.2020 № 12 (ред. от 25.03.2022). URL: <https://legalacts.ru/doc/prikaz-fgbu-fips-ot-20012020-n-12-ob-utverzhdanii/> (дата обращения: 04.05.2024).
7. GitHub. URL: <https://github.com/?ysclid=lhjgyylis6v237864613> (дата обращения: 04.05.2024).
 8. Либерман А. И., Шульга Т. Э. Актуальные задачи анализа изображений // Проблемы управления в социально-экономических и технических системах : сб. науч. ст. XVIII Междунар. науч.-практич. конф., 14–15 апреля 2022 г., г. Саратов. Саратов : Наука, 2022. С. 41–46.
 9. Коломийцева Е. Ю. Анализ изображений в Интернете: возможности современных сервисов // Медиа в современном мире. 61-е Петербургские чтения : ст. участников ежегодного апрельского научного форума, 21–22 апреля 2022 г., Санкт-Петербург. Т. 2. СПб. : Медиапир, 2022. С. 66–67.
 10. Шестоперов Д. В. Анализ изображений. Определение дубликатов и степени размытости изображений // Вопросы устойчивого развития общества. 2022. № 5. С. 1089–1096.
 11. Ситникова Л. К., Еланцев М. О., Султанов Р. О. Двухэтапный метод поиска изображений в большой базе изображений // Выставка Инноваций – 2022 (весенняя сессия) : сб. материалов XXXIII Республиканской выставки-сессии студенческих инновационных проектов, 29 апреля 2022 г., г. Ижевск. Ижевск : Ижевский государственный технический университет имени М. Т. Калашникова, 2022. С. 304–309.
 12. Питерсон У. У., Уэлдон Э. Коды, исправляющие ошибки / пер. с англ., под ред. Р. Л. Добрушина, С. И. Самойленко. М. : Мир, 1976. 594 с.
 13. Метод Левенштейна. URL: https://spravochnick.ru/informatika/metod_levenshteyna/?ysclid=li96moyh99224030240 (дата обращения: 10.05.2024).
 14. Brinardi L., Seng H. Text documents plagiarism detection using Rabin Karp and Jaro-Winkler Distance algorithms // Indonesian Journal of Electrical Engineering and Computer Science. 2017. Vol. 5, no. 2. P. 462–471.
 15. Трасс Х. Х. Геоботаника: история и современные тенденции развития. Л. : Наука, 1976. 252 с.
 16. Построение SIFT дескрипторов и задача сопоставления изображений. URL: <https://habr.com/ru/articles/106302/> (дата обращения: 10.05.2024).
 17. Как бороться с репостами или пара слов о перцептивных хешах. Перцептивные хэш-алгоритмы. URL: <https://habr.com/ru/articles/237307/> (дата обращения: 10.05.2024).
 - ikh dublikatov: prikaz FGBU FIPS 20.01.2020 No. 12 (ed. 25.03.2022). URL: <https://legalacts.ru/doc/prikaz-fgbu-fips-ot-20012020-n-12-ob-utverzhdanii/> (accessed: 04.05.2024). (In Russ.).
 7. GitHub. URL: <https://github.com/?ysclid=lhjgyylis6v237864613> (accessed: 04.05.2024).
 8. Liberman A. I., Shulga T. E. Aktualnye zadachi analiza izobrazheniy. In: *Collection of scientific Articles of the XVIII International scientific-practical conference “Problemy upravleniya v sotsialno-ekonomicheskikh i tekhnicheskikh sistemakh”*, April 14–15, 2022, Saratov. Saratov: Nauka, 2022. P. 41–46. (In Russ.).
 9. Kolomiytseva E. Yu. Analiz izobrazheniy v Internetе: vozmozhnosti sovremennykh servisov. In: *Articles of the participants of the annual April scientific forum “Media v sovremennom mire. 61-e Peterburgskie chteniya”*, April 21–22, 2022, Saint Petersburg. Vol. 2. Saint Petersburg: Mediapir, 2022. P. 66–67. (In Russ.).
 10. Shestoperov D. V. Image analysis. Identification of blur and duplicates of images. *Voprosi ustoychivogo razvitiya obshchestva*. 2022;(5):1089–1096. (In Russ.).
 11. Sitnikova L. K., Elantcev M. O., Sultanov R. O. A two-step method of searching for similar images in a large image database. In: *materials of the XXXIII Republican exhibition-session of student innovation projects “Vystavka Innovatsiy – 2022 (vesenniyaya sessiya)”*, April 29, 2022, Izhevsk. Izhevsk: Izhevskiy gosudarstvennyy tekhnicheskii universitet imeni M. T. Kalashnikova, 2022. P. 304–309. (In Russ.).
 12. Piterson U. U., Ueldon E. Kody, ispravlyayushchie oshibki. Trans. Eds. R. L. Dobrushin, S. I. Samoylenko. Moscow: Mir, 1976. 594 p. (In Russ.).
 13. Metod Levenshteyna. URL: https://spravochnick.ru/informatika/metod_levenshteyna/?ysclid=li96moyh99224030240 (accessed: 10.05.2024). (In Russ.).
 14. Brinardi L., Seng H. Text documents plagiarism detection using Rabin Karp and Jaro-Winkler Distance algorithms. *Indonesian Journal of Electrical Engineering and Computer Science*. 2017;5(2):462–471.
 15. Trass H. Vegetation science. History and contemporary trends of development. Leningrad: Nauka, 1976. 252 p. (In Russ.).
 16. Postroenie SIFT deskriptorov i zadacha sopostavleniya izobrazheniy. URL: <https://habr.com/ru/articles/106302/> (accessed: 10.05.2024). (In Russ.).
 17. Kak borotsya s repostami ili para slov o pertseptivnykh khashakh. Pertseptivnye khash-algoritmy. URL: <https://habr.com/ru/articles/237307/> (accessed: 10.05.2024). (In Russ.).

Информация об авторах

А. А. Детков – кандидат экономических наук, доцент.

В. А. Воронина – студент.

Ю. В. Гарифуллина – студент.

А. М. Корепанов – студент.

А. Ю. Вишнякова – старший преподаватель.

About the authors

A. A. Detkov – Candidate of Sciences (Economics), Associate Professor.

V. A. Voronina – Student.

Yu. V. Garifullina – Student.

A. M. Korepanov – Student.

A. Yu. Vishnyakova – Senior Lecturer.