

ПОИСК НАИЛУЧШЕГО АЛГОРИТМА ВИЗУАЛИЗАЦИИ ГРАФА (НА ПРИМЕРЕ РЕЗУЛЬТАТА РАБОТЫ АЛГОРИТМА ПОИСКА НАИБОЛЬШЕГО ПУТИ)

СЕКАЧЕВ Герман Дмитриевич

ФГБОУ ВО «Казанский национальный исследовательский технический
университет им. А.Н. Туполева – КАИ»
г. Казань, Россия

В данном исследовании рассматривается процесс выбора наиболее эффективного алгоритма визуализации результата обработки направленного циклического графа, на примере задачи поиска наибольшего простого пути.

Ключевые слова: компьютерная графика; визуализация данных; VI; наибольший путь в графе; теория графов.

Компьютерная графика охватывает все виды и формы представления изображений, доступных для восприятия человеком. Она находит свое применение не только в компьютерном мире, но и в различных сферах человеческой деятельности: научных исследованиях (визуализация строения вещества, векторных полей и т. д.), медицине (компьютерная томография), опытно-конструкторских разработках и т. п. [4]

Компьютерная графика – необходимый инструмент в кино, рекламе, искусстве, при создании архитектурных презентаций, компьютерных игр и программ обучения [5].

Визуализация данных с использованием графики – это наиболее наглядная и эффективная форма представления информации. А при моделировании и обработке изображений и вовсе единственная из возможных.

Примером данных, с которыми имея визуализацию работать эффективнее являются графы. Граф – в теории графов это структура, представляющая множество объектов (вершин), в которой некоторые пары объектов в некотором смысле «связаны» ребрами.

Наибольший путь в направленном графе есть наибольшая последовательность вершин, в которой каждая вершина соединена со следующей ребром.

Проблемой данного исследования стала эффективность отображения результата вы-

полнения алгоритма нахождения наибольшего пути в графе.

Целью исследования является нахождение наилучшего алгоритма укладки графа на плоскости для максимально эффективного отображения результата выполнения алгоритма нахождения наибольшего пути.

Задачи:

1. Изучение процесса визуализации данных.

2. Получение результата (наибольшего пути в графе) обработанных данных для дальнейшей визуализации.

3. Выбор лучшего алгоритма для наиболее эффективного отображения результата.

Человеческий мозг способен обрабатывать изображения в 60 000 раз быстрее, чем текст. 90% информации, передаваемой в мозг – это зрительные данные. А это значит, что визуализация данных представляет собой использование врожденных навыков человека для повышения эффективности обработки данных.

Визуальное представление информации является прекрасным инструментом при проведении научных исследований, технических расчетов, наглядным аргументом в научных статьях, не говоря уже о том, что текст с картинками воспринимается лучше, чем «серый» текст, а картинки с текстом воспринимаются еще лучше.

Из обучающего процесса мы знаем, что гораздо лучше запоминается материал тех учебных занятий, которые сопровождаются слайдами и презентациями. Визуализация данных позволяет преобразовать текст в понятные и логичные графики, схемы, диаграммы, сделать инфографику.

Говоря максимально просто, визуализации данных – это отображение пространства данных в графическом пространстве [8].

Грамотно представленная информация в процессе ее переработки в визуальный вид лучше усваивается.

В 2011 г. ученые Э. Соьер и Р. Хогарт провели среди экономистов небольшой эксперимент, доказавший, что часто визуализация данных позволяет воспринимать изложенную информацию лучше, чем текст или цифры. Участники исследования были разделены на 3 группы, в зависимости от изучаемой формы представленной информации. В окончании эксперимента выяснилось, что: в 1 группе, где данные были представлены в виде классической статистической отчетности, 72% участников дали неправильный ответ; во 2 группе, где данные были в виде графиков и классической статистической отчетности, неверный ответ дали 61%; в 3 группе, где данные были представлены только в виде графической информации, ошиблись только 3% [2].

Процесс реализации визуализации выглядит так: сначала данные обрабатываются и отфильтровываются, затем преобразовываются в визуально выражаемую форму (Visual Form), чтобы в конечном итоге отобразить их в виде для пользователя виде (View) [8].

Специалистам по визуализации данных необходимо освоить следующие технологические стеки:

- основная математика: тригонометрические функции, линейная алгебра, геометрические алгоритмы;
- графика: холст, SVG, WebGL, вычислительная графика, теория графов;
- инженерные алгоритмы: базовые алгоритмы, статистические алгоритмы, общие алгоритмы компоновки;
- анализ данных: очистка данных, статистика, моделирование данных;

– эстетика дизайна: принципы дизайна, эстетическое суждение, цвет, взаимодействие, познание;

– визуальные основы: визуальное кодирование, визуальный анализ, графическое взаимодействие;

– решение для визуализации: правильное использование графиков, общие сценарии бизнес-визуализации [8].

Итак, поскольку визуализация данных позволяет проще и лучше усваивать информацию, спрос на технологии, позволяющие улучшить качество визуализации растет с каждым годом.

Сложные анимационные и динамические изображения, созданные современными графическими системами, помогают визуализировать процессы и явления, происходящие в реальности, и вне её. Наиболее известный пример такой системы – симулятор полетов. Он используется в обучении пилотов для моделирования процесса и обстановки полета.

Компьютерная графика применяется при визуализации результатов, научного, технического и медицинского направлений [7]. Научная графика помогает отобразить результаты, которые нельзя увидеть в реальном мире. Деловая графика – предлагает визуальное представление при оформлении отчетов. Конструкторская графика используется архитекторами и инженерами. Художественная – создает произвольные изображения, используемые в видеоматериалах, рекламных роликах, компьютерных играх, и даже мультфильмах [1]. Когнитивная – позволяет получить новые знания, раскрывая сущность явлений или изучать процессы интеллектуального мира используя какие-то созданные изображения [3].

Когда данные визуализированы – их можно обработать быстрее. Визуализация данных объединяет всю информацию, которая может быть упущена при традиционном подходе, в котором используются текст или таблицы. Использование инструментов для визуализации данных повышает скорость и эффективность представления информации.

Задача о самом длинном пути – это задача из теории графов и теории алгоритмов, которая заключается в нахождении максимального пути в заданном графе. Определенные связанные

данные могут быть представлены в виде графов. Множество крупных компаний, такие как Facebook, Amazon, Microsoft, Google используют их, так как они показывают природу, глубину и взаимозависимость отношений, созданные их бизнес-решениями. Граф один из наиболее удобных и понятных способов визуализации информации. Это помогает при решении задач связанных с прокладыванием путей (например, для транспорта или путей сети интернет). Они являются неотъемлемой частью в представлении и обработки BigData у крупных компаний [9].

Если речь заходит о задаче поиска максимального пути для маленьких графов, то есть для тех, количество вершин и ребер которых невелико, она не является трудной. Для небольших графов обычно применяются алгоритмы перебора вершин, сложности $O(n)$, в которых время выполнение алгоритма напрямую зависит от параметра n , характеризующего количество входящей информации алгоритма, но на практике такие графы встречаются редко. Поэтому возникает проблема, когда количество ребер и вершин резко возрастает. Обычный перебор значений крайне неэффективен и необходимо прибегать к другим методам обработки.

Рассмотренный в статье [6] алгоритм является эффективным способом обработки очень больших графов, он позволяет сократить время выполнения по сравнению с простым перебором с использованием GPU. При исходном графе в 1840944 вершин и 7831764 ребер, время сократилось в ≈ 20 раз, что при некоторых графах позволяет сокращать целые дни при выполнении, не теряя при этом точность полученных результатов.

В процессе проведения исследования, для визуализации графа и результата будет использоваться инструмент – Gephi.

Gephi – это пакет программного обеспечения для сетевого анализа и визуализации,

что делает его идеально подходящим для работы с графами.

На сегодняшний день довольно трудно найти инструмент, который справится с графом с большим количеством вершин, поэтому нам необходимо сделать выборку (часть генеральной совокупности элементов, которые будут охвачены в данном эксперименте) из имеющихся данных.

Выборка должна удовлетворять следующим условиям:

- состоять примерно из 100 тысяч строк исходной таблицы, чтобы избежать ошибки нехватки памяти;

- содержать минимум 1 из итоговых полученных в результате работе алгоритма наидлиннейших путей.

Возьмем первый путь из полученных в результате работы алгоритма в статье [6] и будем делать выборку на его основе. Следовательно, итоговая выборка должна содержать все имеющиеся у нас IP адреса. По итогу мы получаем граф из: 38675 вершин, 122281 ребер.

После того как мы загрузили выборку в Gephi отметим вершины в нашем выбранном пути. Изменим их размер с 10 на 25, и сменим их цвет с черного на красный.

Воспользуемся некоторыми из доступных алгоритмов укладки, предварительно приведя все вершины в начало координат. Выбраны будут самые известные и часто используемые алгоритмы, а также те, которые подходят для работы с графом исходных размеров.

1. ForceAtlas2 – алгоритм, который вызывает разбиение данных на классы/группы, с помощью силы отталкивания (насколько каждая вершина будет отталкивать другие вершины) и силы притяжения (насколько связанные под одним кластером вершины будут близки друг к другу). Веса ребер учитываются, если они присутствуют в исходных данных. Для обработки исходного графа потребовалось около 25 секунд (рисунок 1).

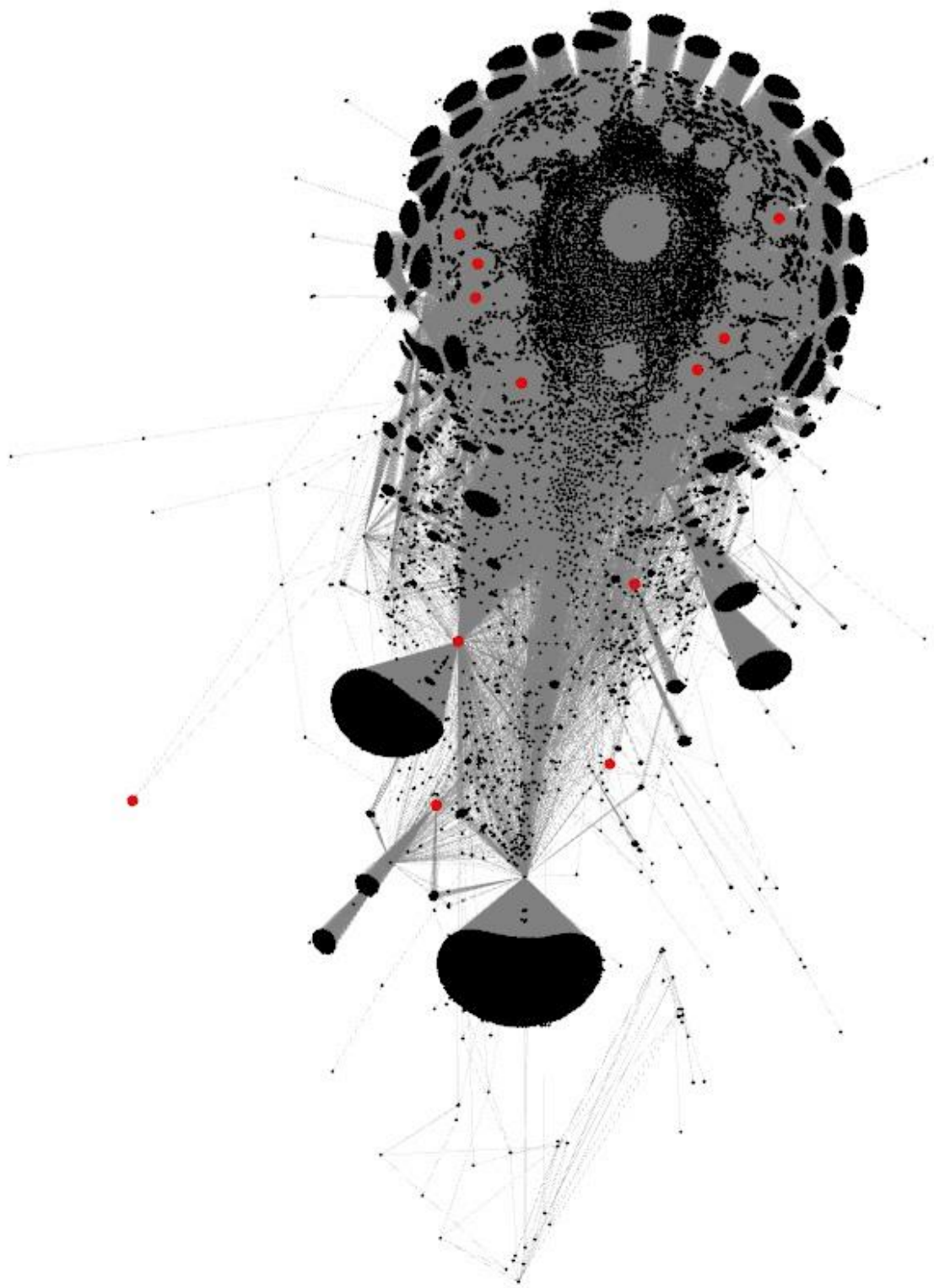


Рисунок 1. ForceAtlas2

2. Yifan Hu – алгоритм, который тоже используется для выявления классов/групп, отлично подходит для графов с небольшим количеством ребер ограничение составляет 100 – 100000 вершин, также как и ForceAtlas2 ис-

пользует силы отталкивания и притяжения при укладке, но не учитывает вес ребер. Для обработки исходного графа потребовалось около 34 секунд (рисунок 2).

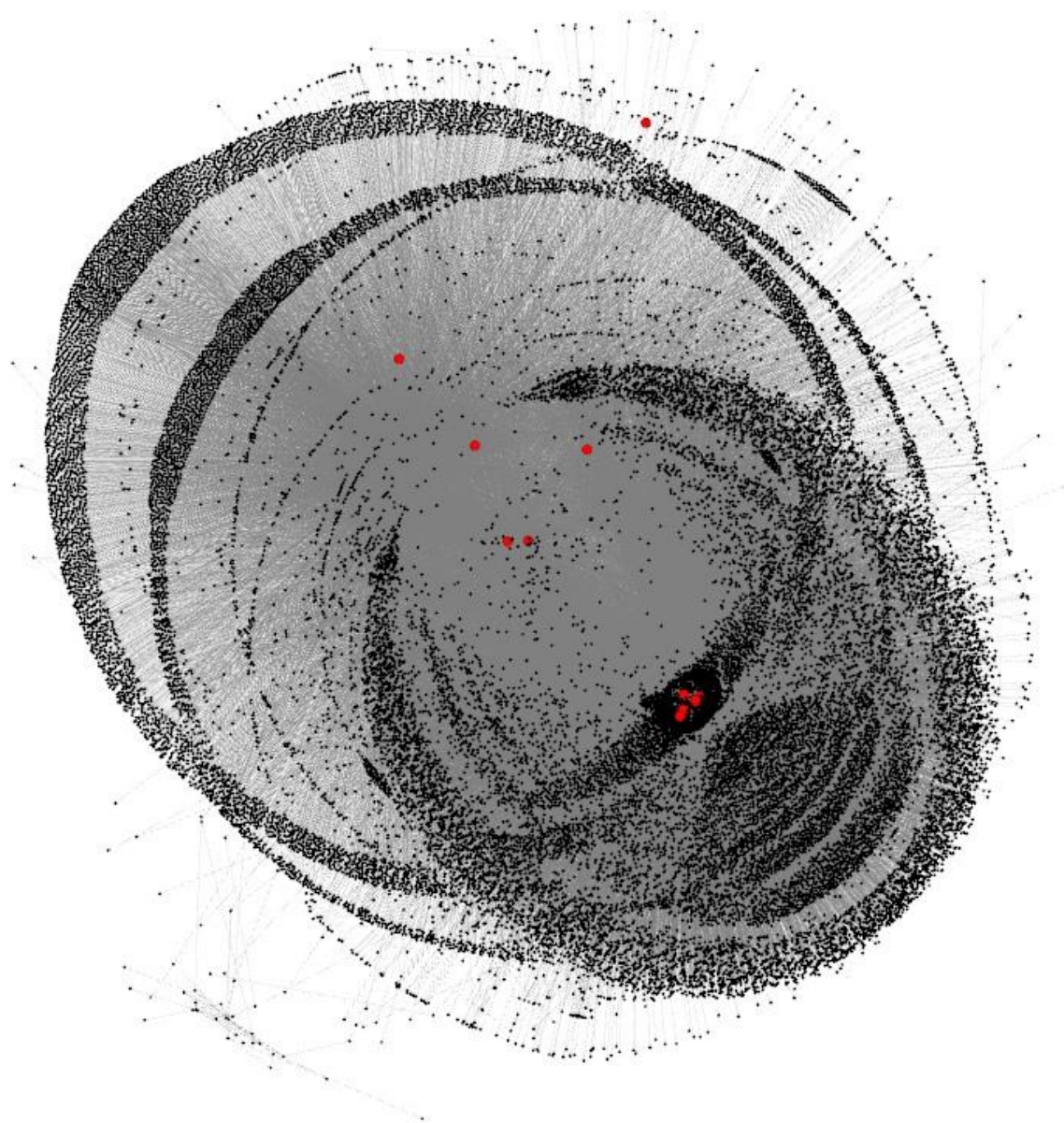


Рисунок 2. Yifan Hu

3. OpenOrd – алгоритм укладки предназначенный для очень больших графов. Скорость работы довольно высокая, но точность является средней. Принцип работы сильно отличается от 2-х вышестоящих алгоритмов.

Сначала все вершины помещаются в начало координат, а потом укладываются с помощью алгоритма имитации отжига. Для обработки исходного графа потребовалось около 17 секунд (рисунок 3).

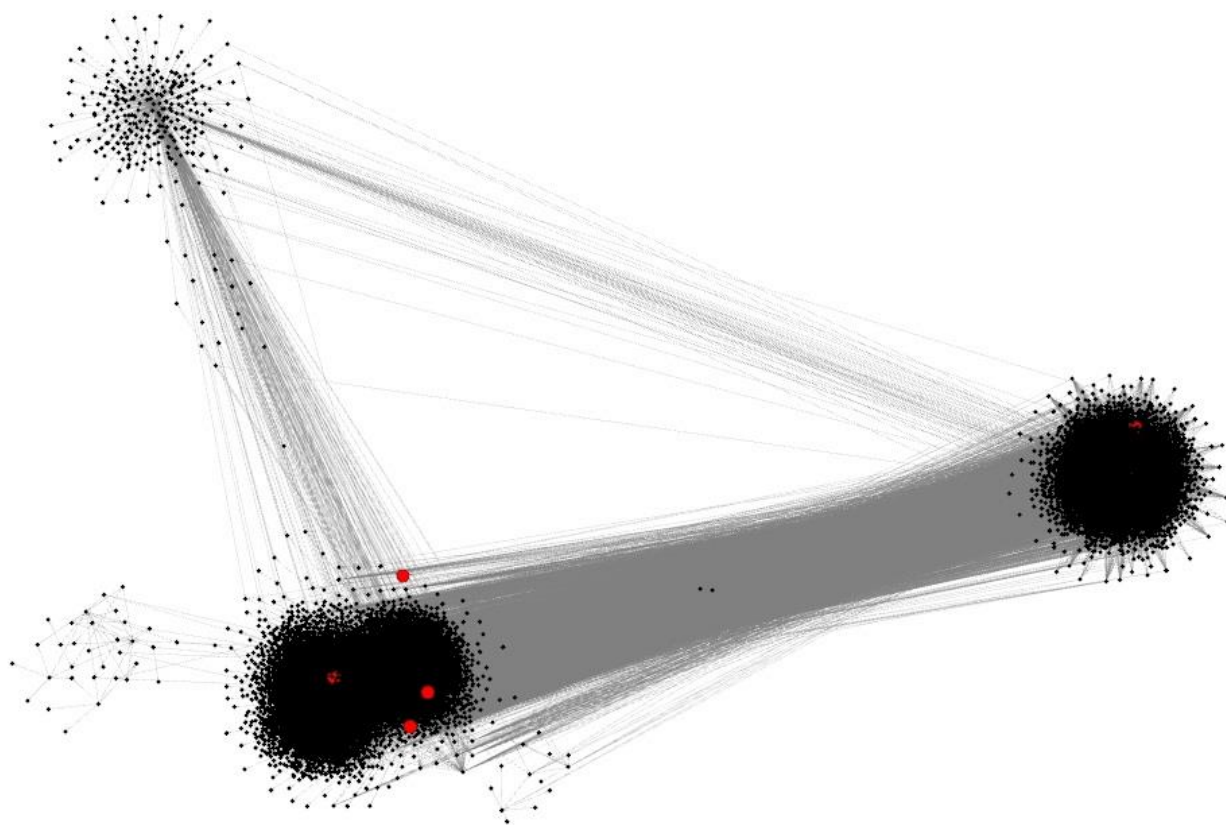


Рисунок 3. OpenOrd

Как можно заметить OpenOrd является самым быстрым алгоритмом укладки, но из-за небольшого размера графа точность укладки слишком мала, что делает его крайне малоэффективным для исходного графа. В свою очередь Yifan Hu, довольно медленный в исполнении и не очень хорошо отображает деление на классы. ForceAtlas2 показывает отличный результат, скорость его выполнения выше алгоритма Yifan Hu, а итоговая укладка отлично передает деление данных на группы.

В результате проведенного исследования был сделан вывод, что лучшим алгоритмом

укладки является ForceAtlas2, так как он, в совокупности с высокой скоростью обработки, отлично передает деление данных на классы/группы, что позволяет лучше отобразить результат алгоритма получения наибольшего пути.

Автор статьи выражает благодарность кандидату физико-математических наук, доценту кафедры прикладной математики и информатики Казанского национального исследовательского технического университета им. А.Н. Туполева – КАИ Арутюновой Наталье Константиновне за ценные отзывы, которые помогли улучшить статью.

СПИСОК ЛИТЕРАТУРЫ

1. Вишневская Л.А. Компьютерная графика для школьников. – М.: Новое знание, 2007. – 160 с.
2. Зачем и как использовать визуализацию данных? Методы визуализации. – URL:<https://leal-ly.ru/internet/zachem-i-kak-ispolzovat-vizualizaciyu-dannyh-metody-vizualizacii/> (дата обращения: 21.06.2024).
3. Зенкин А.А. Когнитивная компьютерная графика. – М.: Наука, 1997. – 187 с.
4. Коджаспирова Г.М., Петров К.В. Технические средства обучения и методика их использования. – 2 изд. – М.: Академия, 2001. – 256 с.

5. Коджастирова Г.М., Петров К.В. Технические средства обучения и методика их использования: учеб. пособие для студентов высш. учеб. заведений. 2-е изд., перераб. и доп. – М.: Академия, 2005. – 350 с.
6. Секачев Г.Д. Finding the longest path in directed cyclic graph // Английский язык в сфере профессиональной коммуникации: VII Всероссийская молодежная научная конференция. – Казань: ИП Сагиева, 2021. – С. 275.
7. Ходжаева Д.Ф., Омонов А.А., Курбанова Ш.М. Компьютерная графика в образовании // Наука, техника и образование. – 2020. – № 4(68). – С. 95-97.
8. Что такое визуализация данных // Русские Блоги. – URL:<https://russianblogs.com/article/3500165683/> (дата обращения: 22.06.24).
9. This map shows how the most innovative companies are using knowledge graphs // The Business of Business. – URL:<https://www.businessofbusiness.com/articles/biggest-corporate-knowledge-graph-users/> (дата обращения: 20.06.24).

FINDING THE BEST GRAPH VISUALIZATION ALGORITHM (ON THE EXAMPLE OF THE RESULT OF THE WORK OF THE ALGORITHM FOR FINDING THE LONGEST PATH)

SEKACHEV German Dmitrievich

Kazan National Research Technical University named after A.N. Tupolev – KAI
Kazan, Russia

This research considers the process of selecting the most efficient algorithm for visualizing the result of directed cyclic graph processing, using the problem of finding the largest simple path as an example.

Keywords: computer graphics, data visualization, BI, longest path in graph, graph theory.
