

УДК 004.67

ОПТИМИЗАЦИЯ РАБОТЫ С МАССИВАМИ ПРИ ПОМОЩИ КРИВОЙ ГИЛЬБЕРТА НА ПРИМЕРЕ ГЕОЛОКАЦИИ

Баринов В.Р.

Московский политехнический университет, Факультет информационных технологий.

Научный руководитель: к.т.н., директор научно-образовательного центра инфокогнитивных технологий Филиппович Ю.Н.

Ключевые слова: массивы, геолокация, кривая Гильберта, многомерные массивы, оптимизация, программирование, библиотеки.

Аннотация: Массивы – одна из основных структур любого языка программирования. Массивы часто используются для решения сложных задач. В реальной жизни используются чаще всего многомерные массивы. Однако, работа с многомерными массивами представляет определённые трудности, связанные со сложностью их структуры. При этом программный код становится громоздким. В данной статье рассматривается возможность использования многомерных массивов при геолокации. Геолокация используется большим количеством приложений для определения местоположения пользователя. Таким образом, задача определения координат становится очень актуальной. Решением данной задачи занимались сотрудники компаний и в XX веке, но на современном этапе эта задача не потеряла своей актуальности. При геолокации каждая географическая точка описывается, как минимум, двумя координатами. Большинство приложений, использующих геолокацию, определяют положение объекта по этим данным. Однако, для некоторых программ созданные СУБД не подходят. Для решения подобных задач корпорация Google создала библиотеку S2, работающую с координатами по принципу кривой Гильберта. В статье приводится метод, которым задаются клетки в системе. Далее рассматривается принцип библиотеки S2, применяемой при разбиении поверхности Земного шара на клетки для работы с координатами. На основании анализа работы библиотеки S2 автор сделал вывод о том, что сделать вывод о том, что двумерные и трёхмерные массивы можно преобразовывать в одномерные, в том случае, если не требуется сохранять вид матрицы. На основании вышеизложенного, автор статьи предлагает применять подобный способ для оптимизации работы с некоторыми двумерными и трёхмерными массивами и компилирования программ.

OPTIMIZATION OF WORK WITH MASSIVES WITH THE HILBERT CURVE ON THE EXAMPLE OF GELOCATION

Barinov V. Moscow Polytechnic University, Faculty of Information Technology.

Scientific adviser: Ph.D., director of the scientific and educational center of infocognitive technologies Filippovich Y.

Keywords: arrays, geolocation, Hilbert curve, multidimensional arrays, optimization, programming, libraries.

Abstract: Arrays are one of the main structures of any programming language. Arrays are often used to solve complex problems. In real life, multidimensional arrays are most often used. However, working with multidimensional arrays presents certain difficulties associated with the complexity of their structure. In this case, the program code becomes cumbersome. This article discusses the possibility of using multidimensional arrays for geolocation. Geolocation is used by a large number of applications to determine the location of the user. Thus, the problem of determining the coordinates becomes very urgent. The decision of this task involved

employees of companies in the XX century, but at the present stage this task has not lost its relevance. When geolocation, each geographic point is described, at least, by two coordinates. Most applications using geolocation determine the position of the object from these data. However, for some programs, DBMSs are not suitable. To solve such problems, Google has created a library S2, working with coordinates on the principle of the Hilbert curve. The article describes the method by which the cells in the system are given. Next, the principle of the library S2, applied when the surface of the globe is divided into cells for working with coordinates, is considered. Based on the analysis of the library S2, the author concluded that to conclude that two-dimensional and three-dimensional arrays can be converted into one-dimensional arrays, in the event that it is not required to preserve the form of the matrix. Based on the foregoing, the author of the article suggests using this method to optimize work with some two-dimensional and three-dimensional arrays and compiling programs.

Массивы – одна из основных структур любого языка программирования. Данное утверждение основывается на том, что программирование по своей сути – это решение большого количества задач. Массивы данных используются постоянно, они позволяют связывать различные элементы, структурировать и сохранять информацию.

В то же время массивы в языке программирования имеют сложную структуру с достаточно большим количеством ограничений. При работе с ними код может становиться довольно громоздким достаточно быстро при том, что производится довольно банальное сравнение значений «координат», при этом структура кода имеет уже четырёхуровневый цикл.

При решении реальных задач очень часто приходится прибегать к использованию многомерных массивов, так как в нашем мире практически нет вещей, обладающих лишь одной характеристикой. Рассмотрим, к примеру, геолокацию – достаточно популярный в наши дни источник информации для разработчиков.

Как известно, геолокация (geolocation)- это определение реального географического местоположения какого-либо электронного устройства (мобильного телефона, компьютера и т.д.), подключённого к сети Интернет.

В современном мире практически каждый житель планеты пользуется мобильным телефоном и мобильным Интернетом. Таким образом, координаты людей, использующих электронные устройства, становятся легкодоступным ресурсом. Поэтому с каждым годом возрастает количество приложений, использующих подобную информацию. Среди них можно отметить популярные игры, связанные с походом в реальные места, сервисы для заказа различных услуг и товаров, персональные приложения компаний, сервисы для случайных знакомств, и многое другое.

Все эти приложения объединяет важный фактор - необходимость работы с географическими координатами, т.е. тремя измерениями – долготой, шириной и высотой. Отметим, что высоту, как элемент, в данном случае можно игнорировать. При этом мы всё равно сталкиваемся, как минимум, с двумерным массивом, отвечающим за фактически одно значение.

Современные геолокационные сервисы всё больше ориентированы на точное определение местоположения их пользователей. В условиях постоянно изменяющихся требований к таким сервисам, возникает необходимость предоставлять более качественные и быстрые алгоритмы обработки данных с возможностью фильтрации результатов выборки, автоматического масштабирования и расширения базы данных и сбора статистики.

Как указывают исследователи Цветков-Омеличев А.Г. и Белая Т.И. в своих работах, «существующие подходы к определению конкретных объектов на настоящий момент слабо оптимизированы и представляют собой привязку объектов к так называемым «местам». При этом для конкретного объекта создаётся специальная точка на карте с именем и координатами (широта и долгота). Другие геолокационные объекты могут быть привязаны к этой точке, которая может определяться по имени или системой предлагаются другие похожие точки, выбранные на основе определённого радиуса отдаления от исходной точки. При таком подходе поиск объекта осуществляется на основании имени точки на карте, а не его местоположения. Данный подход, на наш взгляд, имеет существенный недостаток. А именно, в случае существования геолокационных точек с любым расстоянием между ними, но с похожими названиями, при поисковом запросе такая система выполнит выборку всех таких вариантов, что делает такой поиск абсолютно нерелевантным. [4]

Использование такого подхода для определения местоположения объекта совершенно не подходит для транспортных компаний. Поэтому для оптимизации работы некоторые компании создают дополнительные программные продукты, в том числе и мобильные. Например, некоторые из них используют мобильное приложение, созданное с использованием системы ANTOR LogisticsMaster™ компании ANTOR, которое позволяет не только прокладывать маршрут с учётом пробок, но и решать многие другие задачи. При её создании разработчики устранили недостатки системы «TopLogistic» компании «TopPlan». В частности, данный продукт позволял разрабатывать маршруты только на территории России с использованием карт «TopPlan». [2]

Кроме того, объёмы данных при использовании геолокации достаточно велики и могут стать серьёзной статьёй расходов для компаний. Стоит отметить, что часть навигационных данных являются ошибочными и избыточными, что связано с различными факторами.

Существуют 3 способа фильтрации данных геолокации, сравнительный анализ которых приведён в статье Алёшина О.Г.:

- Метод быстрой фильтрации потока данных, содержащих глобальную позицию наблюдаемого объекта;
- Адаптивный алгоритм обработки потока навигационных данных на основе метода диагностической фильтрации;
- Блочный-временной алгоритм фильтрации геолокационных данных. Обоснованное применение данных способов фильтрации данных позволяет значительно снизить нагрузку на канал передачи данных. [1]

Стоит отметить, что работа с координатами – это задача, отнюдь, не новая. Актуальность её обусловлена жизнедеятельностью людей. [6], [7]. Уже в XX веке многие ИТ-специалисты занимались решением данной задачи и поэтому, при разработке приложения работающего с координатами, стоит взглянуть на готовые решения, это значительно упростит задачу. [5]

Существуют несколько способов определения текущего местоположения: с помощью широко известных API Google Maps и Яндекс Карты, а также расчётов на серверах с помощью программных средств, при этом требуется написание кода. Существуют известные программные методы, например, `getFromlocationName()`, которые позволяют трансформировать предоставленную информацию (почтовый адрес) в координаты (широту, долготу) и обратно. [3]

Стоит отметить, что информационные массивы в данном случае имеют большой объём, а при создании мобильных приложений данный фактор является одним из ключевых, так как он может влиять на быстродействие. При этом стоит учитывать и тот факт, что большой вес программ может препятствовать его установке на устройства с низкой производительностью, а в дальнейшем иметь проблемы с установкой обновлений. [8]

Исходя из постановки задачи, можно понять, что координаты – это огромное количество точек, находящихся на нашем земном шаре. Однако, система ширины и долготы достаточно универсальна и количество координат не бесконечно (можно значительно увеличить их количество, но большого выигрыша по точности это не принесёт). Таким образом, мы видим, что работа с координатами по большому счёту – довольно закрытая и постоянная, без изменяющихся переменных.

В связи с этим, большинство СУБД предоставляют свои сервисы по работе с координатами, в качестве примера может служить расширение PostGIS к популярной СУБД PostgreSQL. Но СУБД не всегда может подходить к вашему проекту, это может быть связано как с трудностями реализации, так и с необходимостью спецификации.

Отличной альтернативой, решающей многие из вышеуказанных проблем может стать библиотека S2, от компании Google, работающая с координатами по принципу кривой Гильберта. [9]

Отметим, что координаты – это частный случай двумерного массива. Таким образом, изучив работу библиотеки, мы сможем частично упростить работу с некоторыми массивами. Библиотека берёт на себя большинство действий с математическими понятиями, однако, стоит отметить, что библиотека представляет Землю как единичную сферу.

Библиотека S2 подразумевает разбиение планеты на клетки, с последующим разбиением клеток на более мелкие клетки. Таким образом, мы получаем иерархическую структуру, где клетка нулевого уровня занимает наибольшую площадь, а клетка максимального, тридцатого уровня имеет площадь меньше квадратного сантиметра. В итоге получается система, которая покрывает практически любые потребности, зависящие от поставленной задачи. Одним из наиболее показательных примеров работы библиотеки, является карта Гавайских островов (рис.1):

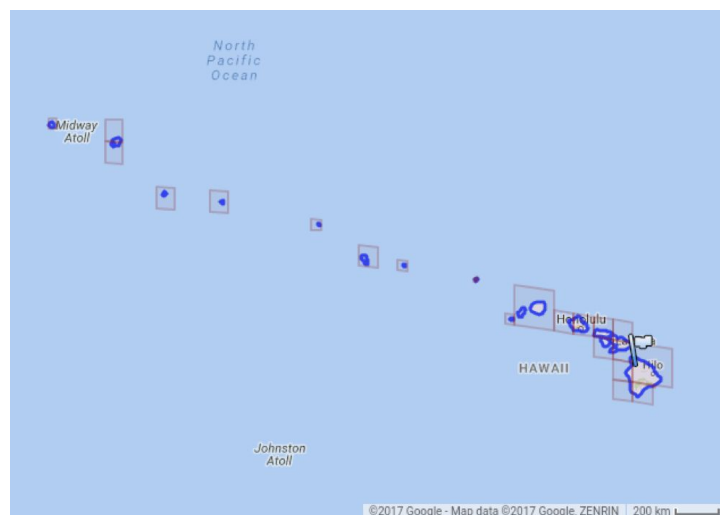


Рис. 1: Гавайские острова, разделённые по координатам посредством библиотеки S2

Важно отметить, что «собирать» определённые географические объекты (такие как штат Гавайи) можно даже из клеток разных иерархических уровней.

Рассмотрим метод, которым задаются клетки в системе. Для начала стоит перенести сферическую поверхность на плоскость: для этого окружаем сферу кубом и проецируем изначальные клетки нулевого уровня на поверхность. Сделать развёртку куба, очевидно, легко. Далее, мы разделяем каждую из граней куба на четыре части и получаем первый уровень иерархии. Аналогично делим каждый из квадратов первого уровня и получаем второй. Принцип получения последующих 29 уровней аналогичен.

Как уже было отмечено выше, сделать развёртку куба нетрудно, но мы всё ещё остаёмся ограничены рамками двумерного массива. Превратить его в одномерный возможно с помощью кривой Гильберта, также известной, как заполняющей пространство кривой Гильберта, непрерывно фрактально заполняющей пространство кривой, описанной немецким математиком Давидом Гильбертом в 1891 году. Функция фрактальна, поэтому позволяет описывать любой из указанных нами уровней. Каждый из поворотов кривой Гильберта указывает на центр точки, указанной клетки, расстояние между точками фиксировано, поэтому мы можем «распрямить» кривую и получить одномерный объект (линию), описывающую плоскость (рис.1). Также, стоит отметить, что кривая Гильберта применима и для трёхмерного пространства.

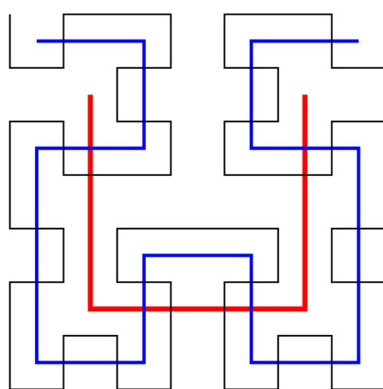


Рис. 2. Первые 3 шага кривой Гильберта

Ещё одной интересной особенностью является довольно удобная кодировка любой клетки, для этого нам будет достаточно одного 64-битного числа. Изначально у нас есть 6 граней куба, для их кодировки достаточно 3 бит. Затем мы каждую из 6 граней разбиваем на 4 равные части 30 раз. Чтобы задать одну из 4 частей, на которые мы разбиваем клетку на каждом из уровней, нам хватит двух бит. Таким образом, мы получаем 63 бита, последний отводим под определение уровня. [4]

В итоге мы получаем следующие преимущества, связанные с особенностями работы библиотеки S2:

1. Мы можем любую точку на земном шаре закодировать одним числом. В зависимости от уровня мы получаем разную точность.
2. Свойство Гильбертовой кривой, заключающееся в том, что точки, которые находятся рядом на ней, находятся рядом и в пространстве, и тот факт, что у нас есть всего лишь одно число, позволяют нам для поиска пользоваться обычным деревом типа B-дерева.

3. Разбиение земного шара на уровни даёт нам простой framework для дробления нашей системы. К примеру, на нулевом уровне мы можем разделить наш сервис на шесть клеток, на первом уровне — на 24 и т. д.

Итак, мы выяснили, что для геолокации, которая фактически является большим массивом двумерных координат, можно использовать одномерную функцию. Как было указано ранее, кривая Гильберта работает и для трёхмерных пространств. Таким образом, можно сделать вывод о том, что двумерные и трёхмерные массивы можно преобразовывать в одномерные. Однако, метод не является универсальным: кривая Гильберта в данном случае служит неким объединителем точек, задающих координату задаваемого значения. Если в результате важно сохранить вид матрицы, или выделить одну из двух (или трёх) задаваемых координат, то развёртка кривой нам не подойдёт.

Список использованной литературы:

1. Алёшин О.Г. Обзор современных методов фильтрации данных геолокации. *European Science*. 2017. №6(28). с.27-31.
2. Попов Б.Н., Сазонов Д.С. Геоинформационная система транспортной логистики. Информационные технологии и системы: управление, экономика, транспорт, право. 2014. №3. с.176-184
3. Соломатин А.А. Сравнение методов по нахождению расстояния от пользователя до конкретного объекта на примере приложения под ANDROID. *Электронный журнал: наука, техника и образование*. 2015. №3(3). С. 78-86.
4. Цветков-Омеличев А.Г., Белая Т. Автоматизированное формирование базы данных нахождения геообъектов на основе математической модели города и страны. *Инженерный вестник Дона*. 2015 .Т 37. № 3. С.42.
5. Бьерн Страуструп. Программирование. Принципы и практика использования С++, изд. «ЛитРес», 2017 – 1248 стр.
6. Генри С. Уоррен. (мл.) Алгоритмические трюки для программистов, 2-е издание, 512 стр., ИД «ВИЛЬЯМС», 2014.
7. Anthony T. Holdener III. HTML5 Geolocation. – Sebastopol: O'Reilly Media. – 2011. – 112 p.
8. Alasdair Allan. Geolocation in iOS. Mobile Positioning and Mapping on iPhone and iPad. - Sebastopol: O'Reilly Media. – 2012. – 116 p.

9. Google для мобильных устройств. - URL blog.greensmm.ru/?p=571.