

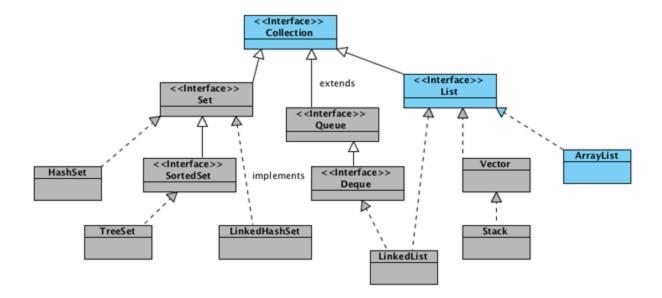


Структуры данных в картинках. ArrayList

Java*

Приветствую вас, хабралюди!

Взбрело мне в голову написать несколько статей, о том как реализованы некоторые структуры данных в Java. Надеюсь, статьи будут полезны визуалам (картинки наше всё), начинающим javaвизуалам а также тем кто уже умеет писать new ArrayList(), но слабо представляет что же происходит внутри.



Сегодня поговорим о ArrayList-ax

ArrayList — реализует интерфейс List. Как известно, в Java массивы имеют фиксированную длину, и после того как массив создан, он не может расти или уменьшаться. ArrayList может менять свой размер во время исполнения программы, при этом не обязательно указывать размерность при создании объекта. Элементы ArrayList могут быть абсолютно любых типов в том числе и null.

Создание объекта

```
ArrayList<String> list = new ArrayList<String>();
```

Только что созданный объект list, содержит свойства elementData и size.

Хранилище значений **elementData** есть ни что иное как массив определенного типа (указанного в generic), в нашем случае **String[]**. Если вызывается конструктор без параметров, то по умолчанию будет создан массив из 10-ти элементов типа Object (с приведением к типу, разумеется).

```
elementData = (E[]) new Object[10];
 0
        1
               2
                      3
                                    5
                                          6
                                                 7
                                                               9
nuli
       null
              null
                     null
                            null
                                  null
                                          null
                                                null
                                                       null
                                                              null
```

Вы можете использовать конструктор **ArrayList(capacity)** и указать свою начальную емкость списка.

Добавление элементов

Внутри метода add(value) происходят следующие вещи:

1) проверяется, достаточно ли места в массиве для вставки нового элемента;

```
ensureCapacity(size + 1);
```

2) добавляется элемент в конец (согласно значению size) массива.

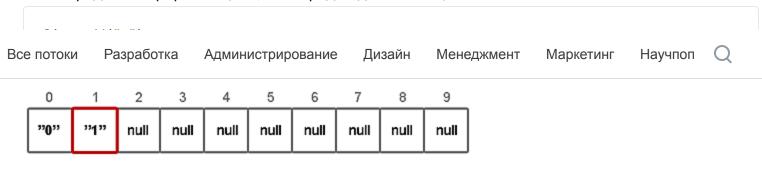
```
elementData[size++] = element;
```

Весь метод ensureCapacity(minCapacity) рассматривать не будем, остановимся только на паре интересных мест. Если места в массиве не достаточно, новая емкость рассчитывается по формуле (oldCapacity * 3) / 2 + 1. Второй момент это копирование элементов. Оно осуществляется с помощью native метода System.arraycopy(), который написан не на Java.

```
// newCapacity - новое значение емкости
elementData = (E[])new Object[newCapacity];

// oldData - временное хранилище текущего массива с данными
System.arraycopy(oldData, 0, elementData, 0, size);
```

Ниже продемонстрирован цикл, поочередно добавляющий 15 элементов:



...

```
list.add("9");

0 1 2 3 4 5 6 7 8 9

"0" "1" "2" "3" "4" "5" "6" "7" "8" "9"
```

```
list.add("10");
```

При добавлении 11-го элемента, проверка показывает что места в массиве нет. Соответственно создается новый массив и вызывается **System.arraycopy()**.



После этого добавление элементов продолжается



...

```
list.add("14");
                    3
                                5
                                       6
                                             7
                                                                10
                                                                      11
                                                                            12
                                                                                  13
                                                                                                15
"0"
      "1"
            "2"
                   "3"
                               "5"
                                      "6"
                                            "7"
                                                   "8"
                                                               "10"
                                                                                  "13"
                                                                           "12"
                                                                                               null
```

Добавление в «середину» списка

```
list.add(5, "100");
```

Добавление элемента на позицию с определенным индексом происходит в три этапа:

1) проверяется, достаточно ли места в массиве для вставки нового элемента;

```
ensureCapacity(size+1);
```

2) подготавливается место для нового элемента с помощью **System.arraycopy()**;

```
System.arraycopy(elementData, index, elementData, index + 1, size - index);
 +75
         322K
                      1454
                              5
                                         7
0
       1
                  3
                                                          10
                                                                11
                                                                      12
                                                                           13
                                                                                 14
                                                                                       15
                                                    "8"
"0"
     "1"
           "2"
                 "3"
                       "4"
                             "5"
                                  "5"
                                        "6"
                                              "7"
                                                          "9"
                                                               "10"
                                                                     "11"
                                                                           "12"
                                                                                 "13"
```

3) перезаписывается значение у элемента с указанным индексом.

```
elementData[index] = element;
size++;
0
                               5
                                           7
                                                              10
                                                                          12
                                                                                13
                   3
                                      6
                                                                    11
                                                                                             15
                  "3"
                        "4"
                             "100"
                                           "6"
                                                                   "10"
                                                                                "12"
                                                                                      "13"
```

Как можно догадаться, в случаях, когда происходит вставка элемента по индексу и при этом в вашем массиве нет свободных мест, то вызов **System.arraycopy()** случится дважды: первый в **ensureCapacity()**, второй в самом методе **add(index, value)**, что явно скажется на скорости всей операции добавления.

В случаях, когда в исходный список необходимо добавить другую коллекцию, да еще и в «середину», стоит использовать метод **addAll(index, Collection)**. И хотя, данный метод скорее всего вызовет **System.arraycopy()** три раза, в итоге это будет гораздо быстрее поэлементного добавления.

Удаление элементов

Удалять элементы можно двумя способами:

— по индексу remove(index)

— по значению remove(value)

С удалением элемента по индексу всё достаточно просто

```
list.remove(5);
```

Сначала определяется какое количество элементов надо скопировать

```
int numMoved = size - index - 1;
```

затем копируем элементы используя System.arraycopy()

```
System.arraycopy(elementData, index + 1, elementData, index, numMoved);
```

уменьшаем размер массива и забываем про последний элемент

```
elementData[--size] = null; // Let gc do its work
```

При удалении по значению, в цикле просматриваются все элементы списка, до тех пор пока не будет найдено соответствие. Удален будет лишь первый найденный элемент.

Дополнение 1: Как верно заметил @MikeMirzayanov, при удалении элементов текущая величина сарасіtу не уменьшается, что может привести к своеобразным утечкам памяти. Поэтому не стоит пренебрегать методом **trimToSize()**.

Итоги

- Быстрый доступ к элементам по индексу за время О(1);
- Доступ к элементам по значению за линейное время O(n);
- Медленный, когда вставляются и удаляются элементы из «середины» списка;
- Позволяет хранить любые значения в том числе и null;
- Не синхронизирован.

Ссылки

Исходник ArrayList
Исходник ArrayList из JDK7
Исходники JDK OpenJDK & trade 6 Source Release — Build b23

Пишите в комментариях пожелания/замечания и есть ли смысл продолжать.

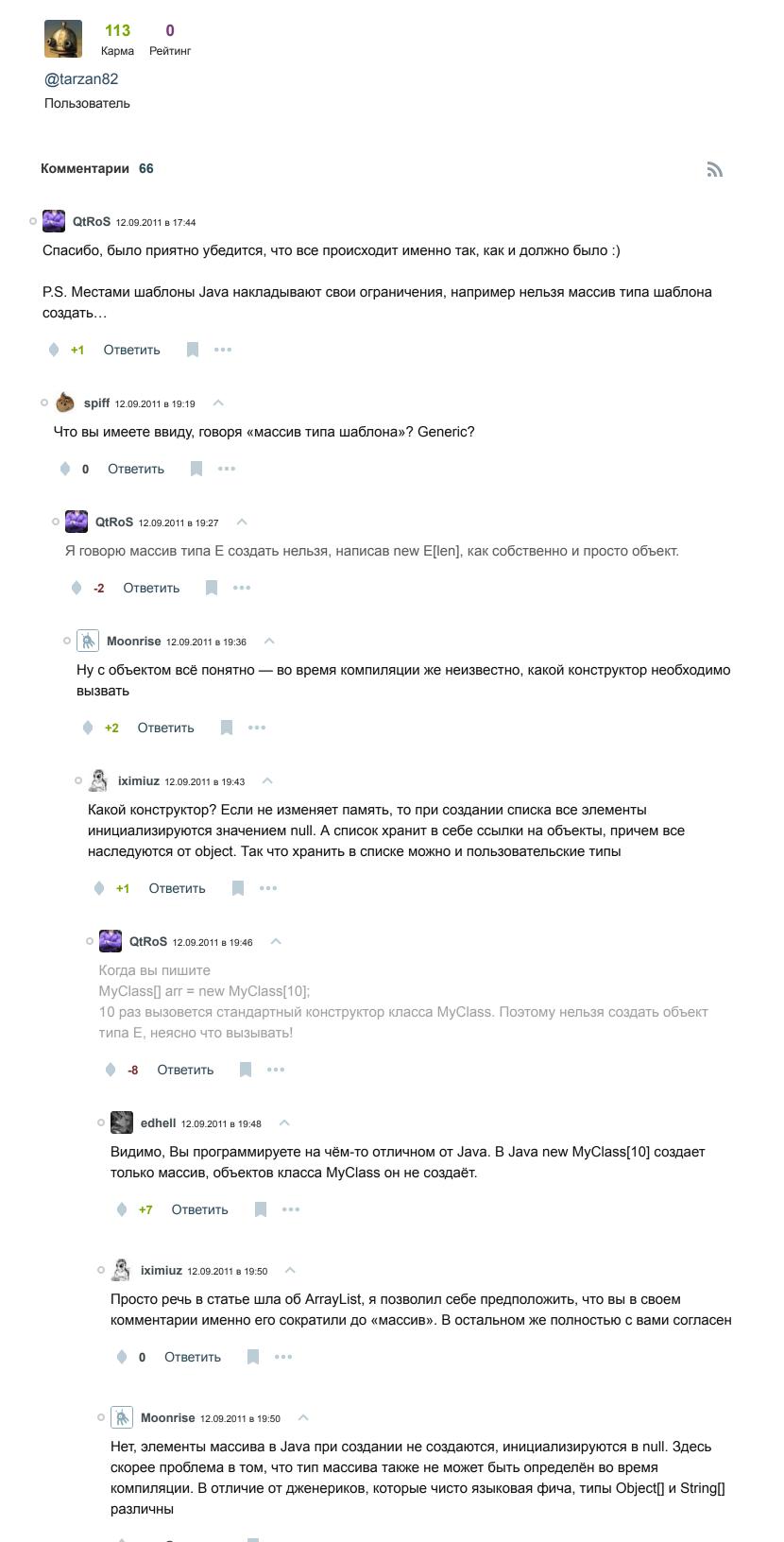
Теги: java, структуры данных, ArrayList

Хабы: Java

Редакторский дайджест

Присылаем лучшие статьи раз в месяц

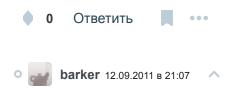
Электропочта





Очевидно, что с появлением generic'ов они переписывали этот код. Соответственно, либо

Очевидно, что с появлением generic'ов они переписывали этот код. Соответственно, либо приведение каждый раз при создании массива, либо каждый раз при обращении. По-моему первый путь экономнее

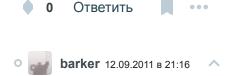


Соответственно, либо приведение каждый раз при создании массива, либо каждый раз при обращении. По-моему первый путь экономнее

Какое ещё приведение? Ни в первом ни во втором случае нет никакого приведения. И вообще никакого приведения не бывает с дженериками. В реальности array останется массивом Object и в том числе в байткоде. Вы как-то превратно понимаете что такое java-дженерики.



Я все прекрасно понимаю. Есть такая вещь — простота редактирования кода. Я имею ввиду «виртуальное» приведение — то, которое нужно компилятору



А, ок. Да, в случае исходников тут всё понятно — вроде как можно и правда переписать Object[] на Т[], но резона, видимо, не было) Да и разницы было бы не много — в некоторых геттерах убралось бы «виртуальное приведение» (Т[]). С точки зрения внешнего интерфейса и производительности всё останется идентичным.





Наверное потому, что в Sun старались писать хороший код? По-моему, код, в котором без особой на то нужды игнорируются предупреждения компилятора о типобезопасности generic-ов, не слишком хорош.



A return (T)elementData[index]; по вашему его не генерирует? Невозможно на Java написать генерик коллекцию на основе массива без type-safe warning'а к сожалению



Да, сглупил. Но всё же я остаюсь при своём мнении о «хорошести» кода. Альтернативный аргумент: операция (T) elementData[index] осталась бы корректна, если бы вместо Т был реальный тип, а операция (T[]) new Object[capacity] возможна **только** для generic-типов. Посути, когда мы пишем приводение к T[], мы на самом деле рассчитываем на то что эта операция не сработает, что несколько странно. В чуть-более сложной ситуации типа class MyCollection<T extends Number> это допущение сразу перестаёт работать:

```
class A<T extends Number>:
    (T[]) new Object[1];
    --> java.lang.ClassCastException: [Ljava.lang.Object; cannot be cast to
[Ljava.lang.Number;
```

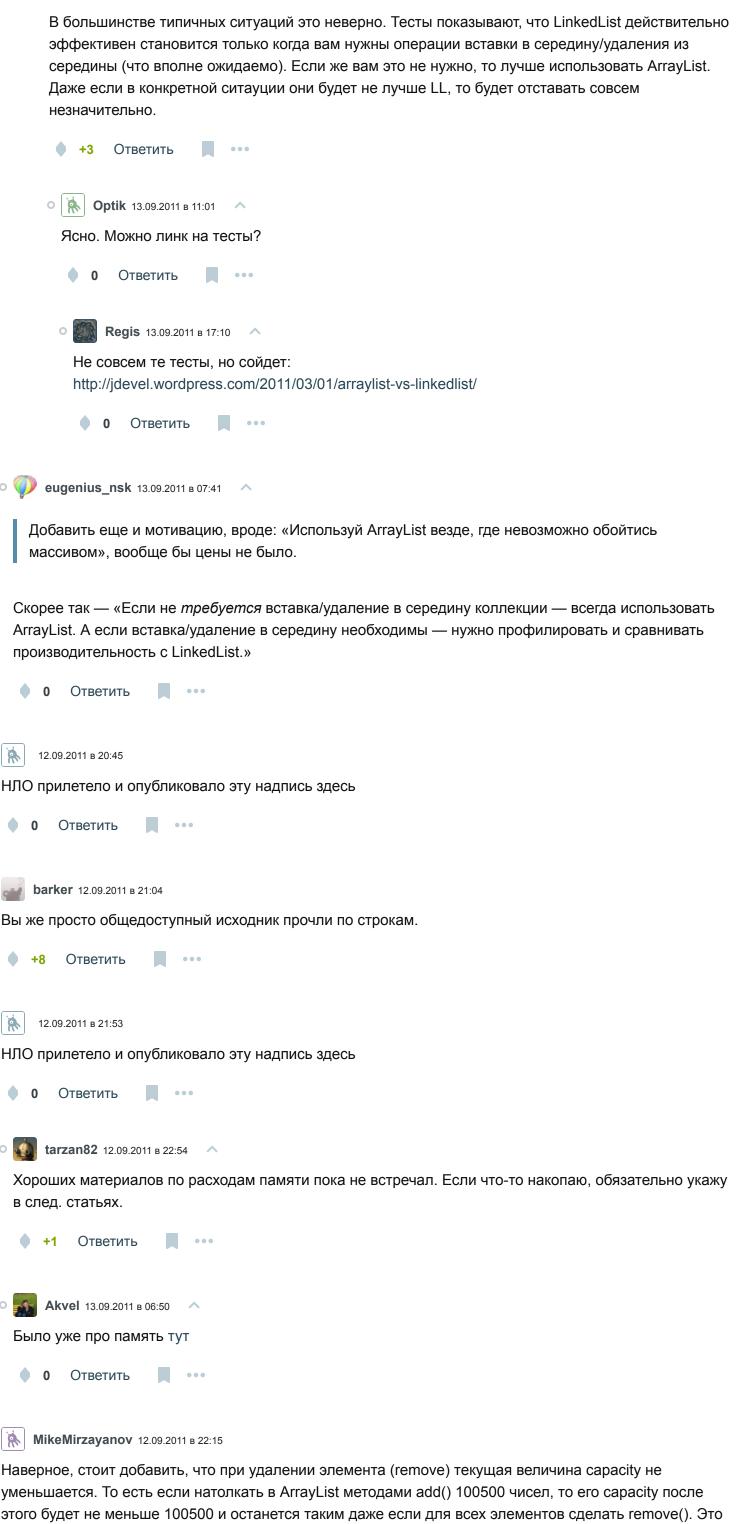
Кстати, написать коллекцию без warning-ов можно, достаточно передать в конструктор Class<T>. Только это никому не нужно.





Проще говоря, ArrayList ведёт себя как обычный массив. Единственно стоит учитывать, что при переполнении массив удваивается в размере.

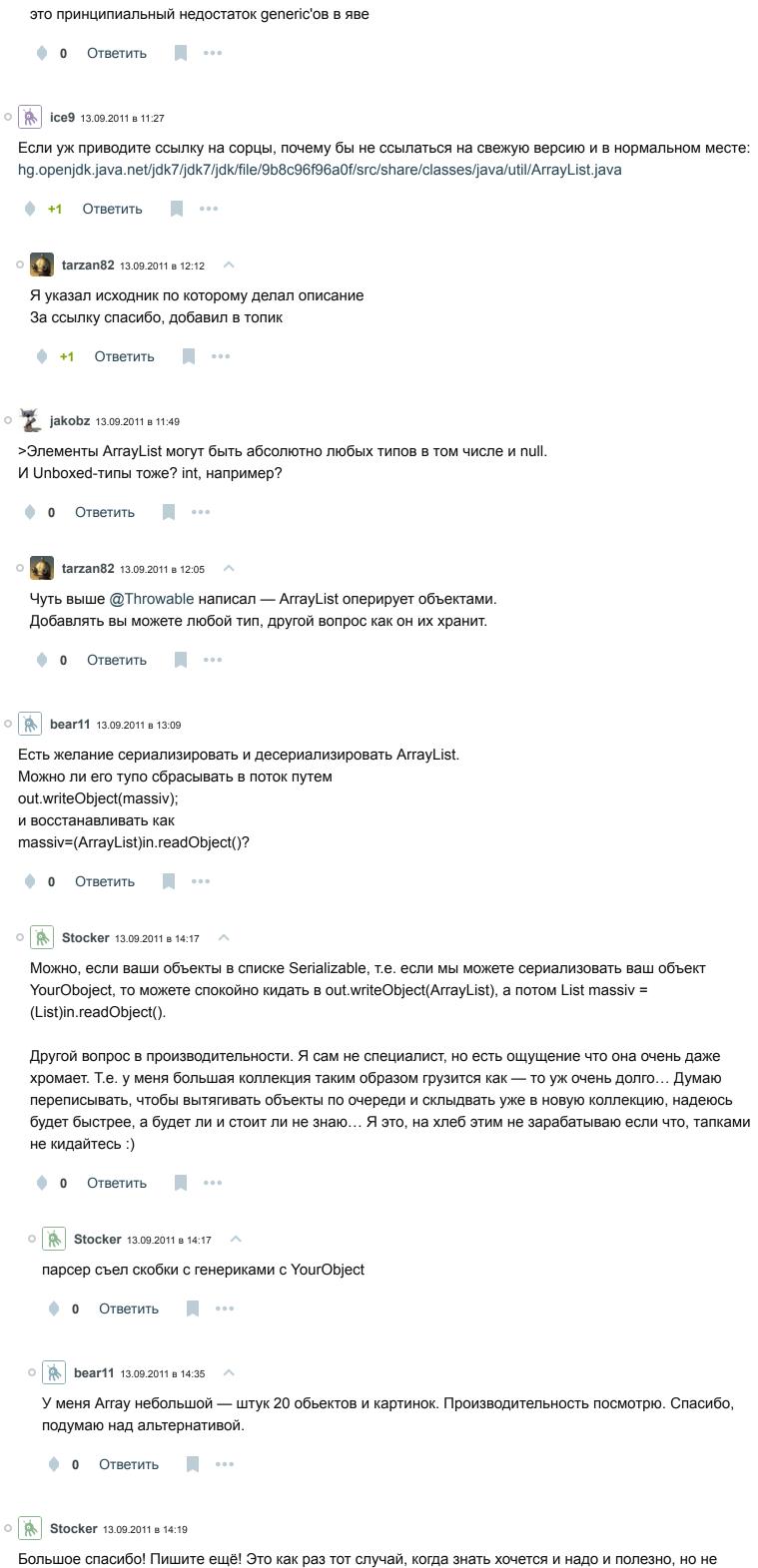
Чуть интереснее с TreeMap/TreeSet (красно-чёрные деревья) и HashMap/HashSet (хэш-таблица в виде массива списков). Ответить edhell 12.09.2011 в 19:42 поправка: не в два раза, а в полтора: int newCapacity = (oldCapacity * 3)/2 + 1 Ответить **spiff** 12.09.2011 в 20:29 Дада. Поэтому, если вы заранее знаете сколько будете хранить в контейнере — лучше сразу указать его размер в конструкторе, тем самым избежать копирование (arraycopy), которое как известно хоть и вылизано до последней строчки, но всеравно жутко дорогое. Ответить TimID 12.09.2011 B 20:19 Вам бы книги писать по программированию «для чайников» Очень все красиво и наглядно. Ответить tarzan82 12.09.2011 в 21:22 Спасибо, рад что понравилось Ответить **spiff** 12.09.2011 B 20:37 Мне кажется надо продолжать. Особенно мне понравились итоги — краткие выводы: стоимость операций. Добавить еще и мотивацию, вроде: «Используй ArrayList везде, где невозможно обойтись массивом», вообще бы цены не было. Ответить о **tarzan82** 12.09.2011 в 21:27 🔥 Спасибо, постараюсь учесть в будущих статьях Ответить Optik 12.09.2011 в 22:48 «Используй ArrayList везде, где невозможно обойтись массивом» По-моему, это не совсем верно. Поправьте, если ошибаюсь, но часто бывают ситуации, когда предпочтительней к примеру LinkedList. И еще, мне кажется, говоря о коллекциях, стоит упомянуть итераторы. +2 Ответить **Akvel** 13.09.2011 в 06:49 Судя по недавнему обзору Сравнение потребления памяти у разных структур хранения данных LinkedList уж больно прожорлив по памяти 0 Ответить Optik 13.09.2011 в 07:45 ^ У LinkedList есть преимущество по скорости при наполнении элементами. При этом проход итератором по списку, подозреваю, будет одинаков. Не ожидал, что LinkedList так много жрет памяти. Но при относительно больших по размеру объектах разница нивелируется, а учитывая, что ArrayList может выделить под новый элемент в полтора раза больше памяти, тогда он и «пожирней» может оказаться. Ответить **Regis** 13.09.2011 в 09:04



У LinkedList есть преимущество по скорости при наполнении элементами.

уменьшается. То есть если натолкать в ArrayList методами add() 100500 чисел, то его capacity после этого будет не меньше 100500 и останется таким даже если для всех элементов сделать remove(). Это может привести к своеобразным утечкам памяти. Иногда может оказаться полезным метод trimToSize() или запись вида «а = new ArrayList<T>()» для освобождения старого буфера для ссылок в «а».

Еще можно отметить, что если вы создаете ArrayList и знаете, что положите в него не менее n элементов, то немного эффективнее будет создать его как a = new ArraysList<Integer>(n), чтобы буфер не перераспределялся log n раз. Ответить о **tarzan82** 13.09.2011 в 00:13 Спасибо, добавил ваше замечание в топик Ответить о rauch 12.09.2011 в 22:33 A как же Navigable*(Set|Map) в иерархии классов на первой картинке? Ответить Optik 12.09.2011 в 23:05 Разве карты относятся к коллекциям? Ответить eugenius_nsk 13.09.2011 в 07:45 NavigableSet — да, упущен. А вот NavigableMap (как и SortedMap и, собственно, Map) — не являются Collection-ами. Ответить **barker** 13.09.2011 B 09:10 Не являются коллекциями — значит их реализации в стандартной библиотеке Java не имплементируют Collection<E>? Ну тогда да, но это странное определение. Конечно же, все Мар являются коллекциями. Просто их ветка на этом рисунке располагалась бы рядом. Просто почему-то их проигнорировали :(Ответить Optik 13.09.2011 в 11:07 Note also that the hierarchy consists of two distinct trees — a Map is not a true Collection. download.oracle.com/javase/tutorial/collections/interfaces/index.html Или я чего то не понимаю? Вы объясните. Я java изучаю недавно, так что без эмоций прошу. Ответить 12.09.2011 в 23:27 НЛО прилетело и опубликовало эту надпись здесь • 0 Ответить О **Lof** 13.09.2011 в 10:53 Наследование коллекций в Java, то что я всегда не понимал... Ответить **Throwable** 13.09.2011 B 11:22 Недостаток ArrayList в том, что он оперирует объектами, и для примитивов его использовать очень неэффективно, хотя в реальной жизни это часто требуется. Вот пара библиотек, которые оперируют коллекциями из примитивов: trove.starlight-systems.com/ pcj.sourceforge.net/ 0 Ответить



corristo 16.09.2011 в 10:20

Большое спасибо! Пишите ещё! Это как раз тот случай, когда знать хочется и надо и полезно, но не настолько чтобы самому копаться в исходниках, а Ваша статья как раз то что нужно и по объёму и по



уровню детализации :)

+2 © 8.9K 59 41 +41 18 июля 2021 в 16:55 Две открытые библиотеки для обучения байесовских сетей и идентификации структуры данных +6 ② 2.3K 48 1 +1 минуточку внимания Разместить КАРЬЕРНАЯ Web Back Событие Турбо Турбо Карьерная неделя для Тетрис на стероидах: тестируем Как проходит интервью QAразработчиков 1С в телеграме War Robots на Steam Deck инженеров в Тинькофф ВАКАНСИИ Java разработчик от 130 000 · Докзилла · Можно удаленно Администратор баз данных (Hadoop Администратор) от 300 000 до 350 000 · АО «ГНИВЦ» · Можно удаленно Senior Software Developer, Database Engine (Java or C++) от 400 000 · Querify Labs · Можно удаленно Lead Software Architect, Database Engine (Java or C++) от 550 000 · Querify Labs · Можно удаленно Ведущий инженер-программист в центр роботехники Сбера от 280 000 до 300 000 · Сбер · Москва Больше вакансий на Хабр Карьере ЛУЧШИЕ ПУБЛИКАЦИИ ЗА СУТКИ вчера в 20:49 Что не так с ДЭГ Москвы на этот раз? **©** 26K +261 33 98 +98 вчера в 23:33 Смерть Mozilla — это смерть открытого Интернета +88 **3K** 36 213 +213 вчера в 16:03 Бифуркация (фантастический рассказ) +22 **3K** 12 **15 +15** сегодня в 07:06 ЦБ хочет компенсировать страдания российских инвесторов за счет «недружественных» нерезидентов +21 3.9K 5 3 +3

Как снизить зависимость кода от структуры данных?

Гайд по межсетевому экранированию (nftables) +17 **4.5K** 100 2 +2 Хотим узнать, что вы думаете о своём месте работы (анонимно) Опрос читают сейчас Смерть Mozilla — это смерть открытого Интернета **3K** 213 +213 Что не так с ДЭГ Москвы на этот раз? 98 +98 **36K** Активность найма на IT-рынке в августе 2022 **19K** 15 +15 Крякнул софт? Суши сухари **37K** 132 +132 Американские компании начали убирать кнопки Facebook** для авторизации со своих сайтов **◎** 5.7K **◎** 7 +7 Финтех и геймдев: где зеленей трава, а нагрузка больше Подкаст 🖸 РАБОТА Java разработчик 425 вакансий Все вакансии Ваш аккаунт Разделы Информация Услуги Войти Публикации Устройство сайта Корпоративный блог Регистрация Новости Для авторов Медийная реклама Для компаний Хабы Нативные проекты Компании Документы Образовательные Авторы Соглашение программы Песочница Конфиденциальность Стартапам Мегапроекты













Настройка языка

Вернуться на старую версию

© 2006–2022, Habr