

# 一. 初识算法

---

## 1.1 什么是算法？

---

### 定义

在数学和计算机科学领域，算法是一系列有限的严谨指令，通常用于解决一类特定问题或执行计算

In mathematics and computer science, an **algorithm** (/ˈælgərɪðəm/) is a finite sequence of rigorous instructions, typically used to solve a class of specific problems or to perform a computation. <sup>1</sup>

### Introduction to Algorithm <sup>2</sup>

不正式的说，算法就是任何定义优良的计算过程：接收一些值作为输入，在有限的时间内，产生一些值作为输出。

Informally, an algorithm is any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output in a finite amount of time.

## 1.2 什么是数据结构？

---

### 定义

在计算机科学领域，数据结构是一种数据组织、管理和存储格式，通常被选择用来高效访问数据

In computer science, a **data structure** is a data organization, management, and storage format that is usually chosen for efficient access to data

### Introduction to Algorithm <sup>2</sup>

数据结构是一种存储和组织数据的方式，旨在便于访问和修改

A data structure is a way to store and organize data in order to facilitate access and modifications

接下来我们通过对一个非常著名的二分查找算法的讲解来认识一下算法

## 1.3 二分查找 <sup>3</sup>

---

二分查找算法也称折半查找，是一种非常高效的工作于有序数组的查找算法。后续的课程中还会学习更多的查找算法，但在此之前，不妨用它作为入门。

## 二分查找基础版

需求：在有序数组  $A$  内，查找值  $target$

- 如果找到返回索引
- 如果找不到返回  $-1$

算法描述	
前提	给定一个内含 $n$ 个元素的有序数组 $A$ ，满足 $A_0 \leq A_1 \leq A_2 \leq \dots \leq A_{n-1}$ ，一个待查值 $target$
1	设置 $i = 0, j = n - 1$
2	如果 $i > j$ ，结束查找，没找到
3	设置 $m = floor(\frac{i+j}{2})$ ， $m$ 为中间索引， $floor$ 是向下取整 ( $\leq \frac{i+j}{2}$ 的最小整数)
4	如果 $target < A_m$ 设置 $j = m - 1$ ，跳到第2步
5	如果 $A_m < target$ 设置 $i = m + 1$ ，跳到第2步
6	如果 $A_m = target$ ，结束查找，找到了

java 实现

```
1 public static int binarySearch(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     while (i <= j) {
4         int m = (i + j) >> 1;
5         if (target < a[m]) {           // 在左边
6             j = m - 1;
7         } else if (a[m] < target) {    // 在右边
8             i = m + 1;
9         } else {
10            return m;
11        }
12    }
13    return -1;
14 }
```

- $i, j$  对应着搜索区间  $[0, a.length - 1]$  (注意是闭合的区间)， $i <= j$  意味着搜索区间内还有未比较的元素， $i, j$  指向的元素也可能是比较的目标
  - 思考：如果不加  $i == j$  行不行？
  - 回答：不行，因为这意味着  $i, j$  指向的元素会漏过比较
- $m$  对应着中间位置，中间位置左边和右边的元素可能不相等 (差一个)，不会影响结果
- 如果某次未找到，那么缩小后的区间内不包含  $m$

## 二分查找改变版

另一种写法

```
1 public static int binarySearch(int[] a, int target) {
2     int i = 0, j = a.length;
3     while (i < j) {
4         int m = (i + j) >> 1;
5         if (target < a[m]) {           // 在左边
6             j = m;
7         } else if (a[m] < target) {    // 在右边
8             i = m + 1;
9         } else {
10            return m;
11        }
12    }
13    return -1;
14 }
```

- $i, j$  对应着搜索区间  $[0, a.length)$  (注意是左闭右开的区间),  $i < j$  意味着搜索区间内还有未比较的元素,  $j$  指向的**一定不是**查找目标
  - 思考: 为啥这次不加  $i == j$  的条件了?
  - 回答: 这回  $j$  指向的不是查找目标, 如果还加  $i == j$  条件, 就意味着  $j$  指向的还会再次比较, 找不到时, 会死循环
- 如果某次要缩小右边界, 那么  $j = m$ , 因为此时的  $m$  已经**不是**查找目标了

## 衡量算法好坏

### 时间复杂度

下面的查找算法也能得出与之前二分查找一样的结果, 那你能说出它差在哪里吗?

```
1 public static int search(int[] a, int k) {
2     for (
3         int i = 0;
4         i < a.length;
5         i++
6     ) {
7         if (a[i] == k) {
8             return i;
9         }
10    }
11    return -1;
12 }
```

考虑最坏情况下 (没找到) 例如 `[1,2,3,4]` 查找 5

- `int i = 0` 只执行一次
- `i < a.length` 受数组元素个数  $n$  的影响, 比较  $n + 1$  次
- `i++` 受数组元素个数  $n$  的影响, 自增  $n$  次
- `a[i] == k` 受元素个数  $n$  的影响, 比较  $n$  次

- `return -1`, 执行一次

粗略认为每行代码执行时间是  $t$ , 假设  $n = 4$  那么

- 总执行时间是  $(1 + 4 + 1 + 4 + 4 + 1) * t = 15t$
- 可以推导出更一般地公式为,  $T = (3 * n + 3)t$

如果套用二分查找算法, 还是 `[1,2,3,4]` 查找 5

```

1  public static int binarySearch(int[] a, int target) {
2      int i = 0, j = a.length - 1;
3      while (i <= j) {
4          int m = (i + j) >>> 1;
5          if (target < a[m]) {           // 在左边
6              j = m - 1;
7          } else if (a[m] < target) {    // 在右边
8              i = m + 1;
9          } else {
10             return m;
11         }
12     }
13     return -1;
14 }

```

- `int i = 0, j = a.length - 1` 各执行 1 次
- `i <= j` 比较  $\text{floor}(\log_2(n) + 1)$  再加 1 次
- `(i + j) >>> 1` 计算  $\text{floor}(\log_2(n) + 1)$  次
- 接下来 `if()` `else if()` `else` 会执行  $3 * \text{floor}(\log_2(n) + 1)$  次, 分别为
  - `if` 比较
  - `else if` 比较
  - `else if` 比较成立后的赋值语句
- `return -1`, 执行一次

结果:

- 总执行时间为  $(2 + (1 + 3) + 3 + 3 * 3 + 1) * t = 19t$
- 更一般地公式为  $(4 + 5 * \text{floor}(\log_2(n) + 1)) * t$

**注意:**

左侧未找到和右侧未找到结果不一样, 这里不做分析

两个算法比较, 可以看到  $n$  在较小的时候, 二者花费的次数差不多

但随着  $n$  越来越大, 比如说  $n = 1000$  时, 用二分查找算法 (红色) 也就是  $54t$ , 而蓝色算法则需要  $3003t$

画图采用的是 [Desmos](#) | [图形计算器](#)

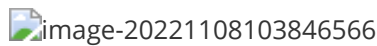
计算机科学中，**时间复杂度**是用来衡量：一个算法的执行，随数据规模增大，而增长的时间成本

- 不依赖于环境因素

如何表示时间复杂度呢？

- 假设算法要处理的数据规模是  $n$ ，代码总的执行行数用函数  $f(n)$  来表示，例如：
  - 线性查找算法的函数  $f(n) = 3 * n + 3$
  - 二分查找算法的函数  $f(n) = (\text{floor}(\log_2(n)) + 1) * 5 + 4$
- 为了对  $f(n)$  进行化简，应当抓住主要矛盾，找到一个变化趋势与之相近的表示法

## 大 O 表示法<sup>4</sup>



其中

- $c, c_1, c_2$  都为常数
- $f(n)$  是实际执行代码行数与  $n$  的函数
- $g(n)$  是经过化简，变化趋势与  $f(n)$  一致的  $n$  的函数

## 渐进上界

渐进上界 (asymptotic upper bound)：从某个常数  $n_0$  开始， $c * g(n)$  总是位于  $f(n)$  上方，那么记作  $O(g(n))$

- 代表算法执行的最差情况

例1

- $f(n) = 3 * n + 3$
- $g(n) = n$
- 取  $c = 4$ ，在  $n_0 = 3$  之后， $g(n)$  可以作为  $f(n)$  的渐进上界，因此表示法写作  $O(n)$

例2


- $f(n) = 5 * \text{floor}(\log_2(n)) + 9$
- $g(n) = \log_2(n)$
- $O(\log_2(n))$

已知  $f(n)$  来说，求  $g(n)$

- 表达式中相乘的常量，可以省略，如
  - $f(n) = 100 * n^2$  中的 100
- 多项式中数量规模更小（低次项）的表达式，如
  - $f(n) = n^2 + n$  中的  $n$
  - $f(n) = n^3 + n^2$  中的  $n^2$
- 不同底数的对数，渐进上界可以用一个对数函数  $\log n$  表示

- 例如:  $\log_2(n)$  可以替换为  $\log_{10}(n)$ , 因为  $\log_2(n) = \frac{\log_{10}(n)}{\log_{10}(2)}$ , 相乘的常量  $\frac{1}{\log_{10}(2)}$  可以省略
- 类似的, 对数的常数次幂可省略
  - 如:  $\log(n^c) = c * \log(n)$

## 常见大 $O$ 表示法

 image-20221108114915524

按时间复杂度从低到高

- 黑色横线  $O(1)$ , 常量时间, 意味着算法时间并不随数据规模而变化
- 绿色  $O(\log(n))$ , 对数时间
- 蓝色  $O(n)$ , 线性时间, 算法时间与数据规模成正比
- 橙色  $O(n * \log(n))$ , 拟线性时间
- 红色  $O(n^2)$  平方时间
- 黑色朝上  $O(2^n)$  指数时间
- 没画出来的  $O(n!)$

## 渐进下界

渐进下界 (asymptotic lower bound): 从某个常数  $n_0$  开始,  $c * g(n)$  总是位于  $f(n)$  下方, 那么记作  $\Omega(g(n))$

## 渐进紧界

渐进紧界 (asymptotic tight bounds): 从某个常数  $n_0$  开始,  $f(n)$  总是在  $c_1 * g(n)$  和  $c_2 * g(n)$  之间, 那么记作  $\Theta(g(n))$

## 空间复杂度

与时间复杂度类似, 一般也使用大  $O$  表示法来衡量: 一个算法执行随数据规模增大, 而增长的**额外**空间成本

```

1  public static int binarySearchBasic(int[] a, int target) {
2      int i = 0, j = a.length - 1;    // 设置指针和初值
3      while (i <= j) {                // i~j 范围内有东西
4          int m = (i + j) >>> 1;
5          if(target < a[m]) {          // 目标在左边
6              j = m - 1;
7          } else if (a[m] < target) { // 目标在右边
8              i = m + 1;
9          } else {                    // 找到了
10             return m;
11         }
12     }
13     return -1;
14 }
```

## 二分查找性能

下面分析二分查找算法的性能

时间复杂度

- 最坏情况:  $O(\log n)$
- 最好情况: 如果待查找元素恰好在数组中央, 只需要循环一次  $O(1)$

空间复杂度

- 需要常数个指针  $i, j, m$ , 因此额外占用的空间是  $O(1)$

## 二分查找平衡版

```
1 public static int binarySearchBalance(int[] a, int target) {
2     int i = 0, j = a.length;
3     while (1 < j - i) {
4         int m = (i + j) >>> 1;
5         if (target < a[m]) {
6             j = m;
7         } else {
8             i = m;
9         }
10    }
11    return (a[i] == target) ? i : -1;
12 }
```

思想:

1. 左闭右开的区间,  $i$  指向的可能是目标, 而  $j$  指向的不是目标
2. 不奢望循环内通过  $m$  找出目标, 缩小区间直至剩 1 个, 剩下的这个可能就是要找的 (通过  $i$ )
  - $j - i > 1$  的含义是, 在范围内待比较的元素个数  $> 1$
3. 改变  $i$  边界时, 它指向的可能是目标, 因此不能  $m + 1$
4. 循环内的平均比较次数减少了
5. 时间复杂度  $\Theta(\log(n))$

## 二分查找 Java 版

```
1 private static int binarySearch0(long[] a, int fromIndex, int toIndex,
2                                   long key) {
3     int low = fromIndex;
4     int high = toIndex - 1;
5
6     while (low <= high) {
7         int mid = (low + high) >>> 1;
8         long midVal = a[mid];
9
10        if (midVal < key)
```

```

11         low = mid + 1;
12     else if (midval > key)
13         high = mid - 1;
14     else
15         return mid; // key found
16 }
17 return -(low + 1); // key not found.
18 }

```

- 例如 [1, 3, 5, 6] 要插入 2 那么就是找到一个位置，这个位置左侧元素都比它小
  - 等循环结束，若没找到，low 左侧元素肯定都比 target 小，因此 low 即插入点
- 插入点取负是为了与找到情况区分
- -1 是为了把索引 0 位置的插入点与找到的情况进行区分

## Leftmost 与 Rightmost

有时我们希望返回的是最左侧的重复元素，如果用 Basic 二分查找

- 对于数组 [1, 2, 3, 4, 4, 5, 6, 7]，查找元素4，结果是索引3
- 对于数组 [1, 2, 4, 4, 4, 4, 5, 6, 7]，查找元素4，结果也是索引3，并不是最左侧的元素

```

1 public static int binarySearchLeftmost1(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     int candidate = -1;
4     while (i <= j) {
5         int m = (i + j) >>> 1;
6         if (target < a[m]) {
7             j = m - 1;
8         } else if (a[m] < target) {
9             i = m + 1;
10        } else {
11            candidate = m; // 记录候选位置
12            j = m - 1;     // 继续向左
13        }
14    }
15    return candidate;
16 }

```

如果希望返回的是最右侧元素

```

1 public static int binarySearchRightmost1(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     int candidate = -1;
4     while (i <= j) {
5         int m = (i + j) >>> 1;
6         if (target < a[m]) {
7             j = m - 1;
8         } else if (a[m] < target) {
9             i = m + 1;
10        } else {
11            candidate = m; // 记录候选位置
12            i = m + 1;     // 继续向右

```



```

13     }
14 }
15 return candidate;
16 }

```

## 应用

对于 Leftmost 与 Rightmost, 可以返回一个比 -1 更有用的值

Leftmost 改为

```

1 public static int binarySearchLeftmost(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     while (i <= j) {
4         int m = (i + j) >>> 1;
5         if (target <= a[m]) {
6             j = m - 1;
7         } else {
8             i = m + 1;
9         }
10    }
11    return i;
12 }

```

- leftmost 返回值的另一层含义:  $< target$  的元素个数
- 小于等于中间值, 都要向左找

Rightmost 改为

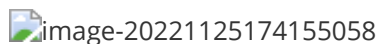
```

1 public static int binarySearchRightmost(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     while (i <= j) {
4         int m = (i + j) >>> 1;
5         if (target < a[m]) {
6             j = m - 1;
7         } else {
8             i = m + 1;
9         }
10    }
11    return i - 1;
12 }

```

- 大于等于中间值, 都要向右找

几个名词



范围查询:

- 查询  $x < 4$ ,  $0..leftmost(4) - 1$
- 查询  $x \leq 4$ ,  $0..rightmost(4)$
- 查询  $4 < x$ ,  $rightmost(4) + 1..\infty$
- 查询  $4 \leq x$ ,  $leftmost(4)..\infty$

- 查询  $4 \leq x \leq 7$ ,  $leftmost(4) \dots rightmost(7)$
- 查询  $4 < x < 7$ ,  $rightmost(4) + 1 \dots leftmost(7) - 1$

**求排名:**  $leftmost(target) + 1$

- $target$  可以不存在, 如:  $leftmost(5) + 1 = 6$
- $target$  也可以存在, 如:  $leftmost(4) + 1 = 3$

**求前任 (predecessor) :**  $leftmost(target) - 1$

- $leftmost(3) - 1 = 1$ , 前任  $a_1 = 2$
- $leftmost(4) - 1 = 1$ , 前任  $a_1 = 2$

**求后任 (successor) :**  $rightmost(target) + 1$

- $rightmost(5) + 1 = 5$ , 后任  $a_5 = 7$
- $rightmost(4) + 1 = 5$ , 后任  $a_5 = 7$

**求最近邻居:**

- 前任和后任距离更近者

## 二. 基础数据结构

### 2.1 数组

#### 概述

##### 定义

在计算机科学中, 数组是由一组元素 (值或变量) 组成的数据结构, 每个元素有至少一个索引或键来标识

In computer science, an **array** is a data structure consisting of a collection of *elements* (values or variables), each identified by at least one *array index* or *key*

因为数组内的元素是**连续存储**的, 所以数组中元素的地址, 可以通过其索引计算出来, 例如:

```
1 | int[] array = {1,2,3,4,5}
```

知道了数组的**数据**起始地址  $BaseAddress$ , 就可以由公式  $BaseAddress + i * size$  计算出索引  $i$  元素的地址

- $i$  即索引, 在 Java、C 等语言都是从 0 开始
- $size$  是每个元素占用字节, 例如  $int$  占 4,  $double$  占 8

#### 小测试

```
1 | byte[] array = {1,2,3,4,5}
```

已知 array 的数据的起始地址是 0x7138f94c8，那么元素 3 的地址是什么？

答：0x7138f94c8 + 2 \* 1 = 0x7138f94ca

## 空间占用

Java 中数组结构为

- 8 字节 markword
- 4 字节 class 指针（压缩 class 指针的情况）
- 4 字节 数组大小（决定了数组最大容量是  $2^{32}$ ）
- 数组元素 + 对齐字节（java 中所有对象大小都是 8 字节的整数倍<sup>5</sup>，不足的要对齐字节补足）

例如

```
1 | int[] array = {1, 2, 3, 4, 5};
```

的大小为 40 个字节，组成如下

```
1 | 8 + 4 + 4 + 5*4 + 4(alignment)
```

## 随机访问性能

即根据索引查找元素，时间复杂度是  $O(1)$

## 动态数组

java 版本

```
1 | public class DynamicArray implements Iterable<Integer> {
2 |     private int size = 0; // 逻辑大小
3 |     private int capacity = 8; // 容量
4 |     private int[] array = {};
5 |
6 |
7 |     /**
8 |      * 向最后位置 [size] 添加元素
9 |      *
10 |     * @param element 待添加元素
11 |     */
12 |     public void addLast(int element) {
13 |         add(size, element);
14 |     }
15 |
16 |     /**
17 |      * 向 [0 .. size] 位置添加元素
18 |      *
19 |     * @param index 索引位置
```

```

20     * @param element 待添加元素
21     */
22     public void add(int index, int element) {
23         checkAndGrow();
24
25         // 添加逻辑
26         if (index >= 0 && index < size) {
27             // 向后挪动，空出待插入位置
28             System.arraycopy(array, index,
29                             array, index + 1, size - index);
30         }
31         array[index] = element;
32         size++;
33     }
34
35     private void checkAndGrow() {
36         // 容量检查
37         if (size == 0) {
38             array = new int[capacity];
39         } else if (size == capacity) {
40             // 进行扩容，1.5 1.618 2
41             capacity += capacity >> 1;
42             int[] newArray = new int[capacity];
43             System.arraycopy(array, 0,
44                             newArray, 0, size);
45             array = newArray;
46         }
47     }
48
49     /**
50     * 从 [0 .. size) 范围删除元素
51     *
52     * @param index 索引位置
53     * @return 被删除元素
54     */
55     public int remove(int index) { // [0..size)
56         int removed = array[index];
57         if (index < size - 1) {
58             // 向前挪动
59             System.arraycopy(array, index + 1,
60                             array, index, size - index - 1);
61         }
62         size--;
63         return removed;
64     }
65
66
67     /**
68     * 查询元素
69     *
70     * @param index 索引位置，在 [0..size) 区间内
71     * @return 该索引位置的元素
72     */
73     public int get(int index) {
74         return array[index];

```

```

75     }
76
77     /**
78      * 遍历方法1
79      *
80      * @param consumer 遍历要执行的操作，入参：每个元素
81      */
82     public void foreach(Consumer<Integer> consumer) {
83         for (int i = 0; i < size; i++) {
84             // 提供 array[i]
85             // 返回 void
86             consumer.accept(array[i]);
87         }
88     }
89
90     /**
91      * 遍历方法2 - 迭代器遍历
92      */
93     @Override
94     public Iterator<Integer> iterator() {
95         return new Iterator<Integer>() {
96             int i = 0;
97
98             @Override
99             public boolean hasNext() { // 有没有下一个元素
100                 return i < size;
101             }
102
103             @Override
104             public Integer next() { // 返回当前元素,并移动到下一个元素
105                 return array[i++];
106             }
107         };
108     }
109
110     /**
111      * 遍历方法3 - stream 遍历
112      *
113      * @return stream 流
114      */
115     public IntStream stream() {
116         return IntStream.of(Arrays.copyOfRange(array, 0, size));
117     }
118 }

```

- 这些方法实现，都简化了 index 的有效性判断，假设输入的 index 都是合法的

### 插入或删除性能

头部位置，时间复杂度是  $O(n)$

中间位置，时间复杂度是  $O(n)$

尾部位置，时间复杂度是  $O(1)$  (均摊来说)

## 二维数组

```
1 int[][] array = {
2     {11, 12, 13, 14, 15},
3     {21, 22, 23, 24, 25},
4     {31, 32, 33, 34, 35},
5 };
```

内存图如下

- 二维数组占 32 个字节，其中 array[0], array[1], array[2] 三个元素分别保存了指向三个一维数组的引用
- 三个一维数组各占 40 个字节
- 它们在内层布局上是**连续**的

更一般的，对一个二维数组  $Array[m][n]$

- $m$  是外层数组的长度，可以看作 row 行
- $n$  是内层数组的长度，可以看作 column 列
- 当访问  $Array[i][j]$ ,  $0 \leq i < m, 0 \leq j < n$  时，就相当于
  - 先找到第  $i$  个内层数组（行）
  - 再找到此内层数组中第  $j$  个元素（列）

### 小测试

Java 环境下（不考虑类指针和引用压缩，此为默认情况），有下面的二维数组

```
1 byte[][] array = {
2     {11, 12, 13, 14, 15},
3     {21, 22, 23, 24, 25},
4     {31, 32, 33, 34, 35},
5 };
```

已知 array **对象** 起始地址是 0x1000，那么 23 这个元素的地址是什么？

答：

- 起始地址 0x1000
- 外层数组大小：16字节对象头 + 3元素 \* 每个引用4字节 + 4 对齐字节 = 32 = 0x20
- 第一个内层数组大小：16字节对象头 + 5元素 \* 每个byte1字节 + 3 对齐字节 = 24 = 0x18
- 第二个内层数组，16字节对象头 = 0x10，待查找元素索引为 2
- 最后结果 = 0x1000 + 0x20 + 0x18 + 0x10 + 2\*1 = 0x104a

# 局部性原理

这里只讨论空间局部性

- cpu 读取内存（速度慢）数据后，会将其放入高速缓存（速度快）当中，如果后来的计算再用到此数据，在缓存中能读到的话，就不必读内存了
- 缓存的最小存储单位是缓存行（cache line），一般是 64 bytes，一次读的数据少了不划算啊，因此最少读 64 bytes 填满一个缓存行，因此读入某个数据时也会读取其**临近的数据**，这就是所谓**空间局部性**

## 对效率的影响

比较下面 ij 和 ji 两个方法的执行效率

```
1  int rows = 1000000;
2  int columns = 14;
3  int[][] a = new int[rows][columns];
4
5  Stopwatch sw = new Stopwatch();
6  sw.start("ij");
7  ij(a, rows, columns);
8  sw.stop();
9  sw.start("ji");
10 ji(a, rows, columns);
11 sw.stop();
12 System.out.println(sw.prettyPrint());
```

ij 方法

```
1  public static void ij(int[][] a, int rows, int columns) {
2      long sum = 0L;
3      for (int i = 0; i < rows; i++) {
4          for (int j = 0; j < columns; j++) {
5              sum += a[i][j];
6          }
7      }
8      System.out.println(sum);
9  }
```

ji 方法

```
1  public static void ji(int[][] a, int rows, int columns) {
2      long sum = 0L;
3      for (int j = 0; j < columns; j++) {
4          for (int i = 0; i < rows; i++) {
5              sum += a[i][j];
6          }
7      }
8      System.out.println(sum);
9  }
```

执行结果

```

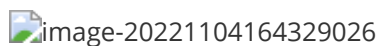
1 0
2 0
3 Stopwatch '': running time = 96283300 ns
4 -----
5 ns          %      Task name
6 -----
7 016196200  017%   ij
8 080087100  083%   ji

```

可以看到 ij 的效率比 ji 快很多，为什么呢？

- 缓存是有限的，当新数据来了后，一些旧的缓存行数据就会被覆盖
- 如果不能充分利用缓存的数据，就会造成效率低下

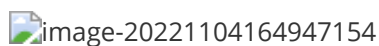
以 ji 执行为例，第一次内循环要读入  $[0, 0]$  这条数据，由于局部性原理，读入  $[0, 0]$  的同时也读入了  $[0, 1] \dots [0, 13]$ ，如图所示



但很遗憾，第二次内循环要的是  $[1, 0]$  这条数据，缓存中没有，于是再读入了下图的数据



这显然是一种浪费，因为  $[0, 1] \dots [0, 13]$  包括  $[1, 1] \dots [1, 13]$  这些数据虽然读入了缓存，却没有及时用上，而缓存的大小是有限的，等执行到第九次内循环时



缓存的第一行数据已经被新的数据  $[8, 0] \dots [8, 13]$  覆盖掉了，以后如果再想读，比如  $[0, 1]$ ，又得到内存去读了

同理可以分析 ij 函数则能充分利用局部性原理加载到的缓存数据

### 举一反三

1. I/O 读写时同样可以体现局部性原理
2. 数组可以充分利用局部性原理，那么链表呢？

答：链表不行，因为链表的元素并非相邻存储

## 越界检查

java 中对数组元素的读写都有越界检查，类似于下面的代码

```

1 bool is_within_bounds(int index) const
2 {
3     return 0 <= index && index < length();
4 }

```

- 代码位置：`openjdk\src\hotspot\share\oops\arrayOop.hpp`

只不过此检查代码，不需要由程序员自己来调用，JVM 会帮我们调用

## 2.2 链表



# 概述

## 定义

在计算机科学中，链表是数据元素的线性集合，其每个元素都指向下一个元素，元素存储上并不连续


In computer science, a **linked list** is a linear collection of data elements whose order is not given by their physical placement in memory. Instead, each element points to the next.

可以分为<sup>6</sup>


- 单向链表，每个元素只知道其下一个元素是谁

 image-20221110083407176

- 双向链表，每个元素知道其上一个元素和下一个元素

 image-20221110083427372

- 循环链表，通常的链表尾节点 tail 指向的都是 null，而循环链表的 tail 指向的是头节点 head

 image-20221110083538273

链表内还有一种特殊的节点称为哨兵（Sentinel）节点，也叫做哑元（Dummy）节点，它不存储数据，通常用作头尾，用来简化边界判断，如下图所示

 image-20221110084611550

## 随机访问性能

根据 index 查找，时间复杂度  $O(n)$

## 插入或删除性能

- 起始位置： $O(1)$
- 结束位置：如果已知 tail 尾节点是  $O(1)$ ，不知道 tail 尾节点是  $O(n)$
- 中间位置：根据 index 查找时间 +  $O(1)$

## 单向链表

根据单向链表的定义，首先定义一个存储 value 和 next 指针的类 Node，和一个描述头部节点的引用

```
1 public class SinglyLinkedList {
2
3     private Node head; // 头部节点
4
5     private static class Node { // 节点类
6         int value;
7         Node next;
8
9         public Node(int value, Node next) {
10             this.value = value;
11             this.next = next;
```

```

12     }
13 }
14 }

```

- Node 定义为内部类，是为了对外**隐藏**实现细节，没必要让类的使用者关心 Node 结构
- 定义为 static 内部类，是因为 Node **不需要**与 SinglyLinkedList 实例相关，多个 SinglyLinkedList 实例能共用 Node 类定义

## 头部添加

```

1 public class SinglyLinkedList {
2     // ...
3     public void addFirst(int value) {
4         this.head = new Node(value, this.head);
5     }
6 }

```

- 如果 this.head == null，新增节点指向 null，并作为新的 this.head
- 如果 this.head != null，新增节点指向原来的 this.head，并作为新的 this.head
  - 注意赋值操作执行顺序是从右到左

## while 遍历

```

1 public class SinglyLinkedList {
2     // ...
3     public void loop() {
4         Node curr = this.head;
5         while (curr != null) {
6             // 做一些事
7             curr = curr.next;
8         }
9     }
10 }

```

## for 遍历

```

1 public class SinglyLinkedList {
2     // ...
3     public void loop() {
4         for (Node curr = this.head; curr != null; curr = curr.next) {
5             // 做一些事
6         }
7     }
8 }

```

- 以上两种遍历都可以把**要做的事**以 Consumer 函数的方式传递进来

- Consumer 的规则是一个参数，无返回值，因此像 System.out::println 方法等都是 Consumer
- 调用 Consumer 时，将当前节点 curr.value 作为参数传递给它

## 迭代器遍历

```
1 public class SinglyLinkedList implements Iterable<Integer> {
2     // ...
3     private class NodeIterator implements Iterator<Integer> {
4         Node curr = head;
5
6         public boolean hasNext() {
7             return curr != null;
8         }
9
10        public Integer next() {
11            int value = curr.value;
12            curr = curr.next;
13            return value;
14        }
15    }
16
17    public Iterator<Integer> iterator() {
18        return new NodeIterator();
19    }
20 }
```

- hasNext 用来判断是否还有必要调用 next
- next 做两件事
  - 返回当前节点的 value
  - 指向下一个节点
- NodeIterator 要定义为非 static 内部类，是因为它与 SinglyLinkedList 实例相关，是对某个 SinglyLinkedList 实例的迭代

## 递归遍历

```
1 public class SinglyLinkedList implements Iterable<Integer> {
2     // ...
3     public void loop() {
4         recursion(this.head);
5     }
6
7     private void recursion(Node curr) {
8         if (curr == null) {
9             return;
10        }
11        // 前面做事
12        recursion(curr.next);
13        // 后面做事
14    }
```

## 尾部添加

```

1  public class SinglyLinkedList {
2      // ...
3      private Node findLast() {
4          if (this.head == null) {
5              return null;
6          }
7          Node curr;
8          for (curr = this.head; curr.next != null; ) {
9              curr = curr.next;
10         }
11         return curr;
12     }
13
14     public void addLast(int value) {
15         Node last = findLast();
16         if (last == null) {
17             addFirst(value);
18             return;
19         }
20         last.next = new Node(value, null);
21     }
22 }

```

- 注意，找最后一个节点，终止条件是 `curr.next == null`
- 分成两个方法是为了代码清晰，而且 `findLast()` 之后还能复用

## 尾部添加多个

```

1  public class SinglyLinkedList {
2      // ...
3      public void addLast(int first, int... rest) {
4
5          Node sublist = new Node(first, null);
6          Node curr = sublist;
7          for (int value : rest) {
8              curr.next = new Node(value, null);
9              curr = curr.next;
10         }
11
12         Node last = findLast();
13         if (last == null) {
14             this.head = sublist;
15             return;
16         }
17         last.next = sublist;
18     }
19 }

```

- 先串成一串 sublist
- 再作为一个整体添加

## 根据索引获取

```
1 public class SinglyLinkedList {
2     // ...
3     private Node findNode(int index) {
4         int i = 0;
5         for (Node curr = this.head; curr != null; curr = curr.next, i++) {
6             if (index == i) {
7                 return curr;
8             }
9         }
10        return null;
11    }
12
13    private IllegalArgumentException illegalIndex(int index) {
14        return new IllegalArgumentException(String.format("index [%d] 不合法%n", index));
15    }
16
17    public int get(int index) {
18        Node node = findNode(index);
19        if (node != null) {
20            return node.value;
21        }
22        throw illegalIndex(index);
23    }
24 }
```

- 同样，分方法可以实现复用

## 插入

```
1 public class SinglyLinkedList {
2     // ...
3     public void insert(int index, int value) {
4         if (index == 0) {
5             addFirst(value);
6             return;
7         }
8         Node prev = findNode(index - 1); // 找到上一个节点
9         if (prev == null) { // 找不到
10            throw illegalIndex(index);
11        }
12        prev.next = new Node(value, prev.next);
13    }
14 }
```

- 插入包括下面的删除，都必须找到上一个节点

## 删除

```
1 public class SinglyLinkedList {
2     // ...
3     public void remove(int index) {
4         if (index == 0) {
5             if (this.head != null) {
6                 this.head = this.head.next;
7                 return;
8             } else {
9                 throw illegalIndex(index);
10            }
11        }
12        Node prev = findNode(index - 1);
13        Node curr;
14        if (prev != null && (curr = prev.next) != null) {
15            prev.next = curr.next;
16        } else {
17            throw illegalIndex(index);
18        }
19    }
20 }
```

- 第一个 if 块对应着 removeFirst 情况
- 最后一个 if 块对应着至少得两个节点的情况
  - 不仅仅判断上一个节点非空，还要保证当前节点非空

## 单向链表（带哨兵）

观察之前单向链表的实现，发现每个方法内几乎都有判断是不是 head 这样的代码，能不能简化呢？

用一个不参与数据存储的特殊 Node 作为哨兵，它一般被称为哨兵或哑元，拥有哨兵节点的链表称为带头链表

```
1 public class SinglyLinkedListSentinel {
2     // ...
3     private Node head = new Node(Integer.MIN_VALUE, null);
4 }
```

- 具体存什么值无所谓，因为不会用到它的值

加入哨兵节点后，代码会变得比较简单，先看几个工具方法

```
1 public class SinglyLinkedListSentinel {
2     // ...
3
4     // 根据索引获取节点
5     private Node findNode(int index) {
6         int i = -1;
7         for (Node curr = this.head; curr != null; curr = curr.next, i++) {
```

```

8         if (i == index) {
9             return curr;
10        }
11    }
12    return null;
13 }
14
15 // 获取最后一个节点
16 private Node findLast() {
17     Node curr;
18     for (curr = this.head; curr.next != null; ) {
19         curr = curr.next;
20     }
21     return curr;
22 }
23 }

```

- findNode 与之前类似，只是 i 初始值设置为 -1 对应哨兵，实际传入的 index 也是  $[-1, \infty)$
- findLast 绝不会返回 null 了，就算没有其它节点，也会返回哨兵作为最后一个节点

这样，代码简化为

```

1 public class SinglyLinkedListSentinel {
2     // ...
3
4     public void addLast(int value) {
5         Node last = findLast();
6         /*
7         改动前
8         if (last == null) {
9             this.head = new Node(value, null);
10            return;
11        }
12        */
13        last.next = new Node(value, null);
14    }
15
16    public void insert(int index, int value) {
17        /*
18        改动前
19        if (index == 0) {
20            this.head = new Node(value, this.head);
21            return;
22        }
23        */
24        // index 传入 0 时，返回的是哨兵
25        Node prev = findNode(index - 1);
26        if (prev != null) {
27            prev.next = new Node(value, prev.next);
28        } else {
29            throw illegalIndex(index);
30        }
31    }

```

```

32
33     public void remove(int index) {
34         /*
35         改动前
36         if (index == 0) {
37             if (this.head != null) {
38                 this.head = this.head.next;
39                 return;
40             } else {
41                 throw illegalIndex(index);
42             }
43         }
44         */
45         // index 传入 0 时, 返回的是哨兵
46         Node prev = findNode(index - 1);
47         Node curr;
48         if (prev != null && (curr = prev.next) != null) {
49             prev.next = curr.next;
50         } else {
51             throw illegalIndex(index);
52         }
53     }
54
55     public void addFirst(int value) {
56         /*
57         改动前
58         this.head = new Node(value, this.head);
59         */
60         this.head.next = new Node(value, this.head.next);
61         // 也可以视为 insert 的特例, 即 insert(0, value);
62     }
63 }

```

- 对于删除, 前面说了【最后一个 if 块对应着至少得两个节点的情况】, 现在有了哨兵, 就凑足了两个节点

## 双向链表 (带哨兵)

```

1     public class DoublyLinkedListSentinel implements Iterable<Integer> {
2
3         private final Node head;
4         private final Node tail;
5
6         public DoublyLinkedListSentinel() {
7             head = new Node(null, 666, null);
8             tail = new Node(null, 888, null);
9             head.next = tail;
10            tail.prev = head;
11        }
12
13        private Node findNode(int index) {
14            int i = -1;
15            for (Node p = head; p != tail; p = p.next, i++) {

```



```
16         if (i == index) {
17             return p;
18         }
19     }
20     return null;
21 }
22
23 public void addFirst(int value) {
24     insert(0, value);
25 }
26
27 public void removeFirst() {
28     remove(0);
29 }
30
31 public void addLast(int value) {
32     Node prev = tail.prev;
33     Node added = new Node(prev, value, tail);
34     prev.next = added;
35     tail.prev = added;
36 }
37
38 public void removeLast() {
39     Node removed = tail.prev;
40     if (removed == head) {
41         throw illegalIndex(0);
42     }
43     Node prev = removed.prev;
44     prev.next = tail;
45     tail.prev = prev;
46 }
47
48 public void insert(int index, int value) {
49     Node prev = findNode(index - 1);
50     if (prev == null) {
51         throw illegalIndex(index);
52     }
53     Node next = prev.next;
54     Node inserted = new Node(prev, value, next);
55     prev.next = inserted;
56     next.prev = inserted;
57 }
58
59 public void remove(int index) {
60     Node prev = findNode(index - 1);
61     if (prev == null) {
62         throw illegalIndex(index);
63     }
64     Node removed = prev.next;
65     if (removed == tail) {
66         throw illegalIndex(index);
67     }
68     Node next = removed.next;
69     prev.next = next;
70     next.prev = prev;
```

```

71     }
72
73     private IllegalArgumentException illegalIndex(int index) {
74         return new IllegalArgumentException(
75             String.format("index [%d] 不合法%n", index));
76     }
77
78     @Override
79     public Iterator<Integer> iterator() {
80         return new Iterator<Integer>() {
81             Node p = head.next;
82
83             @Override
84             public boolean hasNext() {
85                 return p != tail;
86             }
87
88             @Override
89             public Integer next() {
90                 int value = p.value;
91                 p = p.next;
92                 return value;
93             }
94         };
95     }
96
97     static class Node {
98         Node prev;
99         int value;
100        Node next;
101
102        public Node(Node prev, int value, Node next) {
103            this.prev = prev;
104            this.value = value;
105            this.next = next;
106        }
107    }
108 }

```


## 环形链表（带哨兵）

双向环形链表带哨兵，这时哨兵既作为头，也作为尾

 image-20221229144232651

 image-20221229143756065

 image-20221229153338425

 image-20221229154248800

参考实现

```

1 public class DoublyLinkedListSentinel implements Iterable<Integer> {
2

```

```

3      @Override
4      public Iterator<Integer> iterator() {
5          return new Iterator<>() {
6              Node p = sentinel.next;
7
8              @Override
9              public boolean hasNext() {
10                 return p != sentinel;
11             }
12
13             @Override
14             public Integer next() {
15                 int value = p.value;
16                 p = p.next;
17                 return value;
18             }
19         };
20     }
21
22     static class Node {
23         Node prev;
24         int value;
25         Node next;
26
27         public Node(Node prev, int value, Node next) {
28             this.prev = prev;
29             this.value = value;
30             this.next = next;
31         }
32     }
33
34     private final Node sentinel = new Node(null, -1, null); // 哨兵
35
36     public DoublyLinkedListSentinel() {
37         sentinel.next = sentinel;
38         sentinel.prev = sentinel;
39     }
40
41     /**
42      * 添加到第一个
43      * @param value 待添加值
44      */
45     public void addFirst(int value) {
46         Node next = sentinel.next;
47         Node prev = sentinel;
48         Node added = new Node(prev, value, next);
49         prev.next = added;
50         next.prev = added;
51     }
52
53     /**
54      * 添加到最后一个
55      * @param value 待添加值
56      */
57     public void addLast(int value) {

```

```

58     Node prev = sentinel.prev;
59     Node next = sentinel;
60     Node added = new Node(prev, value, next);
61     prev.next = added;
62     next.prev = added;
63 }
64
65 /**
66  * 删除第一个
67  */
68 public void removeFirst() {
69     Node removed = sentinel.next;
70     if (removed == sentinel) {
71         throw new IllegalArgumentException("非法");
72     }
73     Node a = sentinel;
74     Node b = removed.next;
75     a.next = b;
76     b.prev = a;
77 }
78
79 /**
80  * 删除最后一个
81  */
82 public void removeLast() {
83     Node removed = sentinel.prev;
84     if (removed == sentinel) {
85         throw new IllegalArgumentException("非法");
86     }
87     Node a = removed.prev;
88     Node b = sentinel;
89     a.next = b;
90     b.prev = a;
91 }
92
93 /**
94  * 根据值删除节点
95  * <p>假定 value 在链表中作为 key，有唯一性</p>
96  * @param value 待删除值
97  */
98 public void removeByValue(int value) {
99     Node removed = findNodeByValue(value);
100     if (removed != null) {
101         Node prev = removed.prev;
102         Node next = removed.next;
103         prev.next = next;
104         next.prev = prev;
105     }
106 }
107
108 private Node findNodeByValue(int value) {
109     Node p = sentinel.next;
110     while (p != sentinel) {
111         if (p.value == value) {
112             return p;

```

```

113         }
114         p = p.next;
115     }
116     return null;
117 }
118
119 }

```

## 2.3 递归

### 概述

#### 定义

计算机科学中，递归是一种解决计算问题的方法，其中解决方案取决于同一类问题的更小子集

In computer science, **recursion** is a method of solving a computational problem where the solution depends on solutions to smaller instances of the same problem.

比如单链表递归遍历的例子：

```

1 void f(Node node) {
2     if(node == null) {
3         return;
4     }
5     println("before:" + node.value)
6     f(node.next);
7     println("after:" + node.value)
8 }

```

说明：

1. 自己调用自己，如果说每个函数对应着一种解决方案，自己调用自己意味着解决方案是一样的（有规律的）
2. 每次调用，函数处理的数据会较上次缩减（子集），而且最后会缩减至无需继续递归
3. 内层函数调用（子集处理）完成，外层函数才能算调用完成

#### 原理

假设链表中有 3 个节点，value 分别为 1，2，3，以上代码的执行流程就类似于下面的伪码

```

1 // 1 -> 2 -> 3 -> null f(1)
2
3 void f(Node node = 1) {
4     println("before:" + node.value) // 1
5     void f(Node node = 2) {
6         println("before:" + node.value) // 2
7         void f(Node node = 3) {
8             println("before:" + node.value) // 3
9             void f(Node node = null) {
10                 if(node == null) {
11                     return;

```

```

12         }
13     }
14     println("after:" + node.value) // 3
15 }
16     println("after:" + node.value) // 2
17 }
18     println("after:" + node.value) // 1
19 }

```

## 思路

1. 确定能否使用递归求解
2. 推导出递推关系，即父问题与子问题的关系，以及递归的结束条件

例如之前遍历链表的递推关系为

$$f(n) = \begin{cases} \text{停止} & n = null \\ f(n.next) & n \neq null \end{cases}$$

- 深入到最里层叫做**递**
- 从最里层出来叫做**归**
- 在**递**的过程中，外层函数内的局部变量（以及方法参数）并未消失，**归**的时候还可以用到

## 单路递归 Single Recursion

### E01. 阶乘

用递归方法求阶乘

- 阶乘的定义  $n! = 1 \cdot 2 \cdot 3 \cdots (n-2) \cdot (n-1) \cdot n$ ，其中  $n$  为自然数，当然  $0! = 1$
- 递推关系

$$f(n) = \begin{cases} 1 & n = 1 \\ n * f(n-1) & n > 1 \end{cases}$$

代码

```

1 private static int f(int n) {
2     if (n == 1) {
3         return 1;
4     }
5     return n * f(n - 1);
6 }

```

拆解**伪码**如下，假设  $n$  初始值为 3

```

1  f(int n = 3) { // 解决不了,递
2      return 3 * f(int n = 2) { // 解决不了,继续递
3          return 2 * f(int n = 1) {
4              if (n == 1) { // 可以解决,开始归
5                  return 1;
6              }
7          }
8      }
9  }

```

## E02. 反向打印字符串

用递归反向打印字符串,  $n$  为字符在整个字符串  $str$  中的索引位置

- **递**:  $n$  从 0 开始, 每次  $n + 1$ , 一直递到  $n == str.length() - 1$
- **归**: 从  $n == str.length()$  开始归, 从归打印, 自然是逆序的

递推关系

$$f(n) = \begin{cases} \text{停止} & n = str.length() \\ f(n + 1) & 0 \leq n \leq str.length() - 1 \end{cases}$$

代码为

```

1  public static void reversePrint(String str, int index) {
2      if (index == str.length()) {
3          return;
4      }
5      reversePrint(str, index + 1);
6      System.out.println(str.charAt(index));
7  }

```

拆解伪码如下, 假设字符串为 "abc"

```

1  void reversePrint(String str, int index = 0) {
2      void reversePrint(String str, int index = 1) {
3          void reversePrint(String str, int index = 2) {
4              void reversePrint(String str, int index = 3) {
5                  if (index == str.length()) {
6                      return; // 开始归
7                  }
8              }
9              System.out.println(str.charAt(index)); // 打印 c
10         }
11         System.out.println(str.charAt(index)); // 打印 b
12     }
13     System.out.println(str.charAt(index)); // 打印 a
14 }

```

# 多路递归 Multi Recursion

## E01. 斐波那契数列

- 之前的例子是每个递归函数只包含一个自身的调用，这称之为 single recursion
- 如果每个递归函数例包含多个自身调用，称之为 multi recursion

### 递推关系

$$f(n) = \begin{cases} 0 & n = 0 \\ 1 & n = 1 \\ f(n-1) + f(n-2) & n > 1 \end{cases}$$

下面的表格列出了数列的前几项

F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
0	1	1	2	3	5	8	13	21	34	55	89	144	233

### 实现

```
1 public static int f(int n) {
2     if (n == 0) {
3         return 0;
4     }
5     if (n == 1) {
6         return 1;
7     }
8     return f(n - 1) + f(n - 2);
9 }
```

### 执行流程



- 绿色代表正在执行（对应递），灰色代表执行结束（对应归）
- 递不到头，不能归，对应着深度优先搜索

### 时间复杂度

- 递归的次数也符合斐波那契规律， $2 * f(n+1) - 1$
- 时间复杂度推导过程
  - 斐波那契通项公式  $f(n) = \frac{1}{\sqrt{5}} * (\frac{1+\sqrt{5}}{2}^n - \frac{1-\sqrt{5}}{2}^n)$
  - 简化为:  $f(n) = \frac{1}{2.236} * (1.618^n - (-0.618)^n)$
  - 带入递归次数公式  $2 * \frac{1}{2.236} * (1.618^{n+1} - (-0.618)^{n+1}) - 1$
  - 时间复杂度为  $\Theta(1.618^n)$

1. 更多 Fibonacci 参考 [7](#) [8](#) [9](#)

2. 以上时间复杂度分析，未考虑大数相加的因素

### 变体1 - 兔子问题 [7](#)



image-20221110155655827



- 第一个月，有一对未成熟的兔子（黑色，注意图中个头较小）
- 第二个月，它们成熟
- 第三个月，它们能产下一对新的小兔子（蓝色）
- 所有兔子遵循相同规律，求第  $n$  个月的兔子数

### 分析

兔子问题如何与斐波那契联系起来呢？设第  $n$  个月兔子数为  $f(n)$

- $f(n)$  = 上个月兔子数 + 新生的小兔子数
- 而【新生的小兔子数】实际就是【上个月成熟的兔子数】
- 因为需要一个月兔子就成熟，所以【上个月成熟的兔子数】也就是【上上个月的兔子数】
- 上个月兔子数，即  $f(n - 1)$
- 上上个月的兔子数，即  $f(n - 2)$

因此本质还是斐波那契数列，只是从其第一项开始

### 变体2 - 青蛙爬楼梯

- 楼梯有  $n$  阶
- 青蛙要爬到楼顶，可以一次跳一阶，也可以一次跳两阶
- 只能向上跳，问有多少种跳法

### 分析

n	跳法	规律
1	(1)	暂时看不出
2	(1,1) (2)	暂时看不出
3	(1,1,1) (1,2) (2,1)	暂时看不出
4	(1,1,1,1) (1,2,1) (2,1,1) (1,1,2) (2,2)	最后一跳，跳一个台阶的，基于 $f(3)$ 最后一跳，跳两个台阶的，基于 $f(2)$
5	...	...

- 因此本质上还是斐波那契数列，只是从其第二项开始
- 对应 leetcode 题目 [70. 爬楼梯 - 力扣 \(LeetCode\)](#)

## 递归优化-记忆法

上述代码存在很多重复的计算，例如求  $f(5)$  递归分解过程

 image-20221207092417933

可以看到（颜色相同的是重复的）：

- $f(3)$  重复了 2 次
- $f(2)$  重复了 3 次
- $f(1)$  重复了 5 次
- $f(0)$  重复了 3 次

随着  $n$  的增大，重复次数非常可观，如何优化呢？

**Memoization** 记忆法（也称备忘录）是一种优化技术，通过存储函数调用结果（通常比较昂贵），当再次出现相同的输入（子问题）时，就能实现加速效果，改进后的代码

```
1 public static void main(String[] args) {
2     int n = 13;
3     int[] cache = new int[n + 1];
4     Arrays.fill(cache, -1);
5     cache[0] = 0;
6     cache[1] = 1;
7     System.out.println(f(cache, n));
8 }
9
10 public static int f(int[] cache, int n) {
11     if (cache[n] != -1) {
12         return cache[n];
13     }
14
15     cache[n] = f(cache, n - 1) + f(cache, n - 2);
16     return cache[n];
17 }
```

优化后的图示，只要结果被缓存，就不会执行其子问题



- 改进后的时间复杂度为  $O(n)$
- 请自行验证改进后的效果
- 请自行分析改进后的空间复杂度

#### 注意

1. 记忆法是动态规划的一种情况，强调的是自顶向下的解决
2. 记忆法的本质是空间换时间

## 递归优化-尾递归

### 爆栈

用递归做  $n + (n - 1) + (n - 2) \dots + 1$

```
1 public static long sum(long n) {
2     if (n == 1) {
3         return 1;
4     }
5     return n + sum(n - 1);
6 }
```

在我的机器上  $n = 12000$  时，爆栈了

```
1 | Exception in thread "main" java.lang.StackOverflowError
2 |     at Test.sum(Test.java:10)
3 |     at Test.sum(Test.java:10)
4 |     at Test.sum(Test.java:10)
5 |     at Test.sum(Test.java:10)
6 |     at Test.sum(Test.java:10)
7 |     ...
```

为什么呢？

- 每次方法调用是需要消耗一定的栈内存的，这些内存用来存储方法参数、方法内局部变量、返回地址等等
- 方法调用占用的内存需要等到**方法结束时**才会释放
- 而递归调用我们之前讲过，不到最深不会回头，最内层方法没完成之前，外层方法都结束不了
  - 例如， $sum(3)$  这个方法内有个需要执行  $3 + sum(2)$ ， $sum(2)$  没返回前，加号前面的 3 不能释放
  - 看下面伪码

```
1 | long sum(long n = 3) {
2 |     return 3 + long sum(long n = 2) {
3 |         return 2 + long sum(long n = 1) {
4 |             return 1;
5 |         }
6 |     }
7 | }
```

## 尾调用

如果函数的最后一步是调用一个函数，那么称为尾调用，例如

```
1 | function a() {
2 |     return b()
3 | }
```

下面三段代码**不能**叫做尾调用

```
1 | function a() {
2 |     const c = b()
3 |     return c
4 | }
```

- 因为最后一步并非调用函数

```
1 | function a() {
2 |     return b() + 1
3 | }
```

- 最后一步执行的是加法

```

1 function a(x) {
2     return b() + x
3 }

```

- 最后一步执行的是加法

一些语言<sup>10</sup> 的编译器能够对尾调用做优化，例如

```

1 function a() {
2     // 做前面的事
3     return b()
4 }
5
6 function b() {
7     // 做前面的事
8     return c()
9 }
10
11 function c() {
12     return 1000
13 }
14
15 a()

```

没优化之前的伪码

```

1 function a() {
2     return function b() {
3         return function c() {
4             return 1000
5         }
6     }
7 }

```

优化后伪码如下

```

1 a()
2 b()
3 c()

```

为何尾递归才能优化？

调用 a 时

- a 返回时发现：没什么可留给 b 的，将来返回的结果 b 提供就可以了，用不着我 a 了，我的内存就可以释放

调用 b 时

- b 返回时发现：没什么可留给 c 的，将来返回的结果 c 提供就可以了，用不着我 b 了，我的内存就可以释放

如果调用 a 时

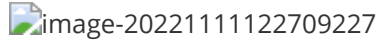
- 不是尾调用，例如 `return b() + 1`，那么 `a` 就不能提前结束，因为它还得利用 `b` 的结果做加法

## 尾递归

尾递归是尾调用的一种特例，也就是最后一步执行的是同一个函数

## 尾递归避免爆栈

安装 Scala



Scala 入门

```
1 object Main {
2     def main(args: Array[String]): Unit = {
3         println("Hello scala")
4     }
5 }
```

- Scala 是 java 的近亲，java 中的类都可以拿来重用
- 类型是放在变量后面的
- Unit 表示无返回值，类似于 void
- 不需要以分号作为结尾，当然加上也对

还是先写一个会爆栈的函数

```
1 def sum(n: Long): Long = {
2     if (n == 1) {
3         return 1
4     }
5     return n + sum(n - 1)
6 }
```

- Scala 最后一行代码若作为返回值，可以省略 return

不出所料，在  $n = 11000$  时，还是出了异常

```
1 println(sum(11000))
2
3 Exception in thread "main" java.lang.StackOverflowError
4     at Main$.sum(Main.scala:25)
5     at Main$.sum(Main.scala:25)
6     at Main$.sum(Main.scala:25)
7     at Main$.sum(Main.scala:25)
8     ...
```

这是因为以上代码，还不是尾调用，要想成为尾调用，那么：

1. 最后一行代码，必须是一次函数调用

2. 内层函数必须**摆脱**与外层函数的关系，内层函数**执行后**不依赖于外层的变量或常量

```
1 def sum(n: Long): Long = {
2     if (n == 1) {
3         return 1
4     }
5     return n + sum(n - 1) // 依赖于外层函数的 n 变量
6 }
```

如何让它执行后就摆脱对 n 的依赖呢？

- 不能等递归回来再做加法，那样就必须保留外层的 n
- 把 n 当做内层函数的一个参数传进去，这时 n 就属于内层函数了
- 传参时就完成累加，不必等回来时累加

```
1 sum(n - 1, n + 累加器)
```

改写后代码如下

```
1 @tailrec
2 def sum(n: Long, accumulator: Long): Long = {
3     if (n == 1) {
4         return 1 + accumulator
5     }
6     return sum(n - 1, n + accumulator)
7 }
```

- accumulator 作为累加器
- @tailrec 注解是 scala 提供的，用来检查方法是否符合尾递归
- 这回 sum(10000000, 0) 也没有问题，打印 50000005000000

执行流程如下，以**伪码**表示 *sum*(4, 0)

```
1 // 首次调用
2 def sum(n = 4, accumulator = 0): Long = {
3     return sum(4 - 1, 4 + accumulator)
4 }
5
6 // 接下来调用内层 sum，传参时就完成了累加，不必等回来时累加，当内层 sum 调用后，外层 sum
  空间没必要保留
7 def sum(n = 3, accumulator = 4): Long = {
8     return sum(3 - 1, 3 + accumulator)
9 }
10
11 // 继续调用内层 sum
12 def sum(n = 2, accumulator = 7): Long = {
13     return sum(2 - 1, 2 + accumulator)
14 }
15
```

```

16 // 继续调用内层 sum，这是最后的 sum 调用完就返回最后结果 10，前面所有其它 sum 的空间早已释放
17 def sum(n = 1, accumulator = 9): Long = {
18     if (1 == 1) {
19         return 1 + accumulator
20     }
21 }

```

本质上，尾递归优化是将函数的递归调用，变成了函数的循环调用

### 改循环避免爆栈

```

1 public static void main(String[] args) {
2     long n = 1000000000;
3     long sum = 0;
4     for (long i = n; i >= 1; i--) {
5         sum += i;
6     }
7     System.out.println(sum);
8 }

```

## 递归时间复杂度-Master theorem 11

若有递归式

$$T(n) = aT\left(\frac{n}{b}\right) + f(n)$$

其中

- $T(n)$  是问题的运行时间， $n$  是数据规模
- $a$  是子问题个数
- $T\left(\frac{n}{b}\right)$  是子问题运行时间，每个子问题被拆成原问题数据规模的  $\frac{n}{b}$
- $f(n)$  是除递归外执行的计算

令  $x = \log_b a$ ，即  $x = \log_{\text{子问题缩小倍数}} \text{子问题个数}$

那么

$$T(n) = \begin{cases} \Theta(n^x) & f(n) = O(n^c) \text{ 并且 } c < x \\ \Theta(n^x \log n) & f(n) = \Theta(n^x) \\ \Theta(n^c) & f(n) = \Omega(n^c) \text{ 并且 } c > x \end{cases}$$

### 例1

$$T(n) = 2T\left(\frac{n}{2}\right) + n^4$$

- 此时  $x = 1 < 4$ ，由后者决定整个时间复杂度  $\Theta(n^4)$
- 如果觉得对数不好算，可以换为求【 $b$  的几次方能等于  $a$ 】

### 例2

$$T(n) = T\left(\frac{7n}{10}\right) + n$$

- $a = 1, b = \frac{10}{7}, x = 0, c = 1$
- 此时  $x = 0 < 1$ , 由后者决定整个时间复杂度  $\Theta(n)$

### 例3

$$T(n) = 16T\left(\frac{n}{4}\right) + n^2$$

- $a = 16, b = 4, x = 2, c = 2$
- 此时  $x = 2 = c$ , 时间复杂度  $\Theta(n^2 \log n)$

### 例4

$$T(n) = 7T\left(\frac{n}{3}\right) + n^2$$

- $a = 7, b = 3, x = 1.?, c = 2$
- 此时  $x = \log_3 7 < 2$ , 由后者决定整个时间复杂度  $\Theta(n^2)$

### 例5

$$T(n) = 7T\left(\frac{n}{2}\right) + n^2$$

- $a = 7, b = 2, x = 2.?, c = 2$
- 此时  $x = \log_2 7 > 2$ , 由前者决定整个时间复杂度  $\Theta(n^{\log_2 7})$

### 例6

$$T(n) = 2T\left(\frac{n}{4}\right) + \sqrt{n}$$

- $a = 2, b = 4, x = 0.5, c = 0.5$
- 此时  $x = 0.5 = c$ , 时间复杂度  $\Theta(\sqrt{n} \log n)$

### 例7. 二分查找递归

```

1  int f(int[] a, int target, int i, int j) {
2      if (i > j) {
3          return -1;
4      }
5      int m = (i + j) >>> 1;
6      if (target < a[m]) {
7          return f(a, target, i, m - 1);
8      } else if (a[m] < target) {
9          return f(a, target, m + 1, j);
10     } else {
11         return m;
12     }
13 }
```

- 子问题个数  $a = 1$
- 子问题数据规模缩小倍数  $b = 2$
- 除递归外执行的计算是常数级  $c = 0$



$$T(n) = T\left(\frac{n}{2}\right) + n^0$$

- 此时  $x = 0 = c$ , 时间复杂度  $\Theta(\log n)$

#### 例8. 归并排序递归

```

1 void split(B[], i, j, A[])
2 {
3     if (j - i <= 1)
4         return;
5     m = (i + j) / 2;
6
7     // 递归
8     split(A, i, m, B);
9     split(A, m, j, B);
10
11    // 合并
12    merge(B, i, m, j, A);
13 }
```

- 子问题个数  $a = 2$
- 子问题数据规模缩小倍数  $b = 2$
- 除递归外, 主要时间花在合并上, 它可以用  $f(n) = n$  表示

$$T(n) = 2T\left(\frac{n}{2}\right) + n$$

- 此时  $x = 1 = c$ , 时间复杂度  $\Theta(n \log n)$

#### 例9. 快速排序递归

```

1 algorithm quicksort(A, lo, hi) is
2     if lo >= hi || lo < 0 then
3         return
4
5     // 分区
6     p := partition(A, lo, hi)
7
8     // 递归
9     quicksort(A, lo, p - 1)
10    quicksort(A, p + 1, hi)
```

- 子问题个数  $a = 2$
- 子问题数据规模缩小倍数
  - 如果分区分的好,  $b = 2$
  - 如果分区没分好, 例如分区1的数据是0, 分区2的数据是  $n - 1$
- 除递归外, 主要时间花在分区上, 它可以用  $f(n) = n$  表示

情况1 - 分区分的好

$$T(n) = 2T\left(\frac{n}{2}\right) + n$$

- 此时  $x = 1 = c$ , 时间复杂度  $\Theta(n \log n)$

情况2 - 分区没分好

$$T(n) = T(n-1) + T(1) + n$$

- 此时不能用主定理求解

## 递归时间复杂度-展开求解

像下面的递归式，都不能用主定理求解

### 例1 - 递归求和

```
1 long sum(long n) {
2     if (n == 1) {
3         return 1;
4     }
5     return n + sum(n - 1);
6 }
```

$$T(n) = T(n-1) + c, \quad T(1) = c$$

下面为展开过程

$$T(n) = T(n-2) + c + c$$

$$T(n) = T(n-3) + c + c + c$$

...

$$T(n) = T(n - (n-1)) + (n-1)c$$

- 其中  $T(n - (n-1))$  即  $T(1)$
- 带入求得  $T(n) = c + (n-1)c = nc$

时间复杂度为  $O(n)$

### 例2 - 递归冒泡排序

```
1 void bubble(int[] a, int high) {
2     if(0 == high) {
3         return;
4     }
5     for (int i = 0; i < high; i++) {
6         if (a[i] > a[i + 1]) {
7             swap(a, i, i + 1);
8         }
9     }
10    bubble(a, high - 1);
11 }
```

$$T(n) = T(n-1) + n, \quad T(1) = c$$

下面为展开过程

$$T(n) = T(n-2) + (n-1) + n$$

$$T(n) = T(n-3) + (n-2) + (n-1) + n$$

...

$$T(n) = T(1) + 2 + \dots + n = T(1) + (n-1) \frac{2+n}{2} = c + \frac{n^2}{2} + \frac{n}{2} - 1$$

时间复杂度  $O(n^2)$

注：

- 等差数列求和为 个数 \*  $\frac{|\text{首项}-\text{末项}|}{2}$

### 例3 - 递归快排

快速排序分区没分好的极端情况

$$T(n) = T(n-1) + T(1) + n, \quad T(1) = c$$

$$T(n) = T(n-1) + c + n$$

下面为展开过程

$$T(n) = T(n-2) + c + (n-1) + c + n$$

$$T(n) = T(n-3) + c + (n-2) + c + (n-1) + c + n$$

...

$$T(n) = T(n - (n-1)) + (n-1)c + 2 + \dots + n = \frac{n^2}{2} + \frac{2cn+n}{2} - 1$$

时间复杂度  $O(n^2)$

不会推导的同学可以进入 <https://www.wolframalpha.com/>

- 例1 输入  $f(n) = f(n-1) + c, f(1) = c$
- 例2 输入  $f(n) = f(n-1) + n, f(1) = c$
- 例3 输入  $f(n) = f(n-1) + n + c, f(1) = c$

## 2.4 队列

### 概述

计算机科学中，queue 是以顺序的方式维护的一组数据集合，在一端添加数据，从另一端移除数据。习惯来说，添加的一端称为**尾**，移除的一端称为**头**，就如同生活中的排队买商品

In computer science, a **queue** is a collection of entities that are maintained in a sequence and can be modified by the addition of entities at one end of the sequence and the removal of entities from the other end of the sequence

先定义一个简化的队列接口

```
1 public interface Queue<E> {
```

```

2
3    /**
4     * 向队列尾插入值
5     * @param value 待插入值
6     * @return 插入成功返回 true, 插入失败返回 false
7     */
8    boolean offer(E value);
9
10   /**
11    * 从对列头获取值, 并移除
12    * @return 如果队列非空返回对头值, 否则返回 null
13    */
14    E poll();
15
16   /**
17    * 从对列头获取值, 不移除
18    * @return 如果队列非空返回对头值, 否则返回 null
19    */
20    E peek();
21
22   /**
23    * 检查队列是否为空
24    * @return 空返回 true, 否则返回 false
25    */
26    boolean isEmpty();
27
28   /**
29    * 检查队列是否已满
30    * @return 满返回 true, 否则返回 false
31    */
32    boolean isFull();
33 }

```

## 链表实现

下面以**单向环形带哨兵**链表方式来实现队列

 image-20221230150105089

 image-20221230150141318

 image-20221230150153271

代码

```

1 public class LinkedListQueue<E>
2     implements Queue<E>, Iterable<E> {
3
4     private static class Node<E> {
5         E value;
6         Node<E> next;
7
8         public Node(E value, Node<E> next) {
9             this.value = value;

```

```

10         this.next = next;
11     }
12 }
13
14 private Node<E> head = new Node<>(null, null);
15 private Node<E> tail = head;
16 private int size = 0;
17 private int capacity = Integer.MAX_VALUE;
18
19 {
20     tail.next = head;
21 }
22
23 public LinkedListQueue() {
24 }
25
26 public LinkedListQueue(int capacity) {
27     this.capacity = capacity;
28 }
29
30 @Override
31 public boolean offer(E value) {
32     if (isFull()) {
33         return false;
34     }
35     Node<E> added = new Node<>(value, head);
36     tail.next = added;
37     tail = added;
38     size++;
39     return true;
40 }
41
42 @Override
43 public E poll() {
44     if (isEmpty()) {
45         return null;
46     }
47     Node<E> first = head.next;
48     head.next = first.next;
49     if (first == tail) {
50         tail = head;
51     }
52     size--;
53     return first.value;
54 }
55
56 @Override
57 public E peek() {
58     if (isEmpty()) {
59         return null;
60     }
61     return head.next.value;
62 }
63
64 @Override

```

```

65     public boolean isEmpty() {
66         return head == tail;
67     }
68
69     @Override
70     public boolean isFull() {
71         return size == capacity;
72     }
73
74     @Override
75     public Iterator<E> iterator() {
76         return new Iterator<E>() {
77             Node<E> p = head.next;
78             @Override
79             public boolean hasNext() {
80                 return p != head;
81             }
82             @Override
83             public E next() {
84                 E value = p.value;
85                 p = p.next;
86                 return value;
87             }
88         };
89     }
90 }

```

## 环形数组实现

### 好处

1. 对比普通数组，起点和终点更为自由，不用考虑数据移动
2. “环”意味着不会存在【越界】问题
3. 数组性能更佳
4. 环形数组比较适合实现有界队列、RingBuffer 等

 image-20221228175413998

### 下标计算

例如，数组长度是 5，当前位置是 3，向前走 2 步，此时下标为  $(3 + 2) \% 5 = 0$

 image-20221228180357257

$$(cur + step) \% length$$

- cur 当前指针位置
- step 前进步数
- length 数组长度

注意：

- 如果 step = 1，也就是一次走一步，可以在  $\geq length$  时重置为 0 即可

## 判断空



## 判断满



满之后的策略可以根据业务需求决定

- 例如我们要实现的环形队列，满之后就拒绝入队

## 代码

```
1 public class ArrayQueue<E> implements Queue<E>, Iterable<E>{
2
3     private int head = 0;
4     private int tail = 0;
5     private final E[] array;
6     private final int length;
7
8     @SuppressWarnings("all")
9     public ArrayQueue(int capacity) {
10         length = capacity + 1;
11         array = (E[]) new Object[length];
12     }
13
14     @Override
15     public boolean offer(E value) {
16         if (isFull()) {
17             return false;
18         }
19         array[tail] = value;
20         tail = (tail + 1) % length;
21         return true;
22     }
23
24     @Override
25     public E poll() {
26         if (isEmpty()) {
27             return null;
28         }
29         E value = array[head];
30         head = (head + 1) % length;
31         return value;
32     }
33
34     @Override
35     public E peek() {
36         if (isEmpty()) {
37             return null;
38         }
39         return array[head];
40     }
41
42     @Override
43     public boolean isEmpty() {
```

```

44         return tail == head;
45     }
46
47     @Override
48     public boolean isFull() {
49         return (tail + 1) % length == head;
50     }
51
52     @Override
53     public Iterator<E> iterator() {
54         return new Iterator<E>() {
55             int p = head;
56             @Override
57             public boolean hasNext() {
58                 return p != tail;
59             }
60
61             @Override
62             public E next() {
63                 E value = array[p];
64                 p = (p + 1) % array.length;
65                 return value;
66             }
67         };
68     }
69 }

```

## 判断空、满方法2

引入 size

```

1  public class ArrayQueue2<E> implements Queue<E>, Iterable<E> {
2
3      private int head = 0;
4      private int tail = 0;
5      private final E[] array;
6      private final int capacity;
7      private int size = 0;
8
9      @SuppressWarnings("all")
10     public ArrayQueue2(int capacity) {
11         this.capacity = capacity;
12         array = (E[]) new Object[capacity];
13     }
14
15     @Override
16     public boolean offer(E value) {
17         if (isFull()) {
18             return false;
19         }
20         array[tail] = value;
21         tail = (tail + 1) % capacity;
22         size++;
23         return true;

```



```

24     }
25
26     @Override
27     public E poll() {
28         if (isEmpty()) {
29             return null;
30         }
31         E value = array[head];
32         head = (head + 1) % capacity;
33         size--;
34         return value;
35     }
36
37     @Override
38     public E peek() {
39         if (isEmpty()) {
40             return null;
41         }
42         return array[head];
43     }
44
45     @Override
46     public boolean isEmpty() {
47         return size == 0;
48     }
49
50     @Override
51     public boolean isFull() {
52         return size == capacity;
53     }
54
55     @Override
56     public Iterator<E> iterator() {
57         return new Iterator<E>() {
58             int p = head;
59
60             @Override
61             public boolean hasNext() {
62                 return p != tail;
63             }
64
65             @Override
66             public E next() {
67                 E value = array[p];
68                 p = (p + 1) % capacity;
69                 return value;
70             }
71         };
72     }
73 }

```

### 判断空、满方法3

- head 和 tail 不断递增，用到索引时，再用它们进行计算，两个问题

- 如何保证 head 和 tail 自增超过正整数最大值的正确性
- 如何让取模运算性能更高
- 答案: 让 capacity 为 2 的幂

```
1 public class ArrayQueue3<E> implements Queue<E>, Iterable<E> {
2
3     private int head = 0;
4     private int tail = 0;
5     private final E[] array;
6     private final int capacity;
7
8     @SuppressWarnings("all")
9     public ArrayQueue3(int capacity) {
10         if ((capacity & capacity - 1) != 0) {
11             throw new IllegalArgumentException("capacity 必须为 2 的幂");
12         }
13         this.capacity = capacity;
14         array = (E[]) new Object[this.capacity];
15     }
16
17     @Override
18     public boolean offer(E value) {
19         if (isFull()) {
20             return false;
21         }
22         array[tail & capacity - 1] = value;
23         tail++;
24         return true;
25     }
26
27     @Override
28     public E poll() {
29         if (isEmpty()) {
30             return null;
31         }
32         E value = array[head & capacity - 1];
33         head++;
34         return value;
35     }
36
37     @Override
38     public E peek() {
39         if (isEmpty()) {
40             return null;
41         }
42         return array[head & capacity - 1];
43     }
44
45     @Override
46     public boolean isEmpty() {
47         return tail - head == 0;
48     }
49
50     @Override
51     public boolean isFull() {
```

```

52         return tail - head == capacity;
53     }
54
55     @Override
56     public Iterator<E> iterator() {
57         return new Iterator<E>() {
58             int p = head;
59
60             @Override
61             public boolean hasNext() {
62                 return p != tail;
63             }
64
65             @Override
66             public E next() {
67                 E value = array[p & capacity - 1];
68                 p++;
69                 return value;
70             }
71         };
72     }
73 }

```

## 2.5 栈

### 概述

计算机科学中，**stack** 是一种线性的数据结构，只能在其一端添加数据和移除数据。习惯来说，这一端称之为**栈顶**，另一端不能操作数据的称之为**栈底**，就如同生活中的一摞书

先提供一个栈接口

```

1  public interface Stack<E> {
2      /**
3       * 向栈顶压入元素
4       * @param value 待压入值
5       * @return 压入成功返回 true，否则返回 false
6       */
7      boolean push(E value);
8
9      /**
10     * 从栈顶弹出元素
11     * @return 栈非空返回栈顶元素，栈为空返回 null
12     */
13     E pop();
14
15     /**
16     * 返回栈顶元素，不弹出
17     * @return 栈非空返回栈顶元素，栈为空返回 null
18     */
19     E peek();

```

```

20
21     /**
22      * 判断栈是否为空
23      * @return 空返回 true, 否则返回 false
24      */
25     boolean isEmpty();
26
27     /**
28      * 判断栈是否已满
29      * @return 满返回 true, 否则返回 false
30      */
31     boolean isFull();
32 }

```

## 链表实现

```

1  public class LinkedListStack<E> implements Stack<E>, Iterable<E> {
2
3      private final int capacity;
4      private int size;
5      private final Node<E> head = new Node<>(null, null);
6
7      public LinkedListStack(int capacity) {
8          this.capacity = capacity;
9      }
10
11     @Override
12     public boolean push(E value) {
13         if (isFull()) {
14             return false;
15         }
16         head.next = new Node<>(value, head.next);
17         size++;
18         return true;
19     }
20
21     @Override
22     public E pop() {
23         if (isEmpty()) {
24             return null;
25         }
26         Node<E> first = head.next;
27         head.next = first.next;
28         size--;
29         return first.value;
30     }
31
32     @Override
33     public E peek() {
34         if (isEmpty()) {
35             return null;
36         }
37         return head.next.value;

```

```

38     }
39
40     @Override
41     public boolean isEmpty() {
42         return head.next == null;
43     }
44
45     @Override
46     public boolean isFull() {
47         return size == capacity;
48     }
49
50     @Override
51     public Iterator<E> iterator() {
52         return new Iterator<E>() {
53             Node<E> p = head.next;
54             @Override
55             public boolean hasNext() {
56                 return p != null;
57             }
58
59             @Override
60             public E next() {
61                 E value = p.value;
62                 p = p.next;
63                 return value;
64             }
65         };
66     }
67
68     static class Node<E> {
69         E value;
70         Node<E> next;
71
72         public Node(E value, Node<E> next) {
73             this.value = value;
74             this.next = next;
75         }
76     }
77 }

```

## 数组实现

```

1  public class ArrayStack<E> implements Stack<E>, Iterable<E>{
2      private final E[] array;
3      private int top = 0;
4
5      @SuppressWarnings("all")
6      public ArrayStack(int capacity) {
7          this.array = (E[]) new Object[capacity];
8      }
9
10     @Override

```

```
11     public boolean push(E value) {
12         if (isFull()) {
13             return false;
14         }
15         array[top++] = value;
16         return true;
17     }
18
19     @Override
20     public E pop() {
21         if (isEmpty()) {
22             return null;
23         }
24         return array[--top];
25     }
26
27     @Override
28     public E peek() {
29         if (isEmpty()) {
30             return null;
31         }
32         return array[top-1];
33     }
34
35     @Override
36     public boolean isEmpty() {
37         return top == 0;
38     }
39
40     @Override
41     public boolean isFull() {
42         return top == array.length;
43     }
44
45     @Override
46     public Iterator<E> iterator() {
47         return new Iterator<E>() {
48             int p = top;
49             @Override
50             public boolean hasNext() {
51                 return p > 0;
52             }
53
54             @Override
55             public E next() {
56                 return array[--p];
57             }
58         };
59     }
60 }
```

## 应用

模拟如下方法调用

```
1 public static void main(String[] args) {
2     System.out.println("main1");
3     System.out.println("main2");
4     method1();
5     method2();
6     System.out.println("main3");
7 }
8
9 public static void method1() {
10    System.out.println("method1");
11    method3();
12 }
13
14 public static void method2() {
15    System.out.println("method2");
16 }
17
18 public static void method3() {
19    System.out.println("method3");
20 }
```

模拟代码

```
1 public class CPU {
2     static class Frame {
3         int exit;
4
5         public Frame(int exit) {
6             this.exit = exit;
7         }
8     }
9     static int pc = 1; // 模拟程序计数器 Program counter
10    static ArrayStack<Frame> stack = new ArrayStack<>(100); // 模拟方法调用栈
11
12    public static void main(String[] args) {
13        stack.push(new Frame(-1));
14        while (!stack.isEmpty()) {
15            switch (pc) {
16                case 1 -> {
17                    System.out.println("main1");
18                    pc++;
19                }
20                case 2 -> {
21                    System.out.println("main2");
22                    pc++;
23                }
24                case 3 -> {
25                    stack.push(new Frame(pc + 1));
26                    pc = 100;
27                }
28            }
29        }
30    }
31 }
```

```
28         case 4 -> {
29             stack.push(new Frame(pc + 1));
30             pc = 200;
31         }
32         case 5 -> {
33             System.out.println("main3");
34             pc = stack.pop().exit;
35         }
36         case 100 -> {
37             System.out.println("method1");
38             stack.push(new Frame(pc + 1));
39             pc = 300;
40         }
41         case 101 -> {
42             pc = stack.pop().exit;
43         }
44         case 200 -> {
45             System.out.println("method2");
46             pc = stack.pop().exit;
47         }
48         case 300 -> {
49             System.out.println("method3");
50             pc = stack.pop().exit;
51         }
52     }
53 }
54 }
55 }
```

## 2.6 双端队列

### 概述

双端队列、队列、栈对比

	定义	特点
队列	一端删除（头） 另一端添加（尾）	First In First Out
栈	一端删除和添加（顶）	Last In First Out
双端队列	两端都可以删除、添加	
优先级队列		优先级高者先出队
延时队列		根据延时时间确定优先级
并发非阻塞队列	队列空或满时不阻塞	
并发阻塞队列	队列空时删除阻塞、队列满时添加阻塞	

注1:



- Java 中 LinkedList 即为典型双端队列实现，不过它同时实现了 Queue 接口，也提供了栈的 push pop 等方法

注2:

- 不同语言，操作双端队列的方法命名有所不同，参见下表

操作	Java	JavaScript	C++	leetCode 641
尾部插入	offerLast	push	push_back	insertLast
头部插入	offerFirst	unshift	push_front	insertFront
尾部移除	pollLast	pop	pop_back	deleteLast
头部移除	pollFirst	shift	pop_front	deleteFront
尾部获取	peekLast	at(-1)	back	getRear
头部获取	peekFirst	at(0)	front	getFront

- 吐槽一下 leetCode 命名比较 low
- 常见的单词还有 enqueue 入队、dequeue 出队

## 接口定义

```

1 public interface Deque<E> {
2
3     boolean offerFirst(E e);
4
5     boolean offerLast(E e);
6
7     E pollFirst();
8
9     E pollLast();
10
11    E peekFirst();
12
13    E peekLast();
14
15    boolean isEmpty();
16
17    boolean isFull();
18 }

```

## 链表实现

```

1 /**
2  * 基于环形链表的双端队列
3  * @param <E> 元素类型
4  */
5 public class LinkedListDeque<E> implements Deque<E>, Iterable<E> {

```

```
6
7  @Override
8  public boolean offerFirst(E e) {
9      if (isFull()) {
10         return false;
11     }
12     size++;
13     Node<E> a = sentinel;
14     Node<E> b = sentinel.next;
15     Node<E> offered = new Node<>(a, e, b);
16     a.next = offered;
17     b.prev = offered;
18     return true;
19 }
20
21 @Override
22 public boolean offerLast(E e) {
23     if (isFull()) {
24         return false;
25     }
26     size++;
27     Node<E> a = sentinel.prev;
28     Node<E> b = sentinel;
29     Node<E> offered = new Node<>(a, e, b);
30     a.next = offered;
31     b.prev = offered;
32     return true;
33 }
34
35 @Override
36 public E pollFirst() {
37     if (isEmpty()) {
38         return null;
39     }
40     Node<E> a = sentinel;
41     Node<E> polled = sentinel.next;
42     Node<E> b = polled.next;
43     a.next = b;
44     b.prev = a;
45     size--;
46     return polled.value;
47 }
48
49 @Override
50 public E pollLast() {
51     if (isEmpty()) {
52         return null;
53     }
54     Node<E> polled = sentinel.prev;
55     Node<E> a = polled.prev;
56     Node<E> b = sentinel;
57     a.next = b;
58     b.prev = a;
59     size--;
60     return polled.value;
```

```

61     }
62
63     @Override
64     public E peekFirst() {
65         if (isEmpty()) {
66             return null;
67         }
68         return sentinel.next.value;
69     }
70
71     @Override
72     public E peekLast() {
73         if (isEmpty()) {
74             return null;
75         }
76         return sentinel.prev.value;
77     }
78
79     @Override
80     public boolean isEmpty() {
81         return size == 0;
82     }
83
84     @Override
85     public boolean isFull() {
86         return size == capacity;
87     }
88
89     @Override
90     public Iterator<E> iterator() {
91         return new Iterator<E>() {
92             Node<E> p = sentinel.next;
93             @Override
94             public boolean hasNext() {
95                 return p != sentinel;
96             }
97
98             @Override
99             public E next() {
100                 E value = p.value;
101                 p = p.next;
102                 return value;
103             }
104         };
105     }
106
107     static class Node<E> {
108         Node<E> prev;
109         E value;
110         Node<E> next;
111
112         public Node(Node<E> prev, E value, Node<E> next) {
113             this.prev = prev;
114             this.value = value;
115             this.next = next;

```

```

116     }
117 }
118
119 Node<E> sentinel = new Node<>(null, null, null);
120 int capacity;
121 int size;
122
123 public LinkedListDeque(int capacity) {
124     sentinel.next = sentinel;
125     sentinel.prev = sentinel;
126     this.capacity = capacity;
127 }
128 }

```

## 数组实现

```

1  /**
2   * 基于循环数组实现，特点
3   * <ul>
4   *     <li>tail 停下来的位置不存储，会浪费一个位置</li>
5   * </ul>
6   * @param <E>
7   */
8  public class ArrayDeque<E> implements Deque<E>, Iterable<E> {
9
10     /**
11         h
12         t
13         0  1  2  3
14         b      a
15     */
16     @Override
17     public boolean offerFirst(E e) {
18         if (isFull()) {
19             return false;
20         }
21         head = dec(head, array.length);
22         array[head] = e;
23         return true;
24     }
25
26     @Override
27     public boolean offerLast(E e) {
28         if (isFull()) {
29             return false;
30         }
31         array[tail] = e;
32         tail = inc(tail, array.length);
33         return true;
34     }
35
36     @Override
37     public E pollFirst() {

```

```
38         if (isEmpty()) {
39             return null;
40         }
41         E e = array[head];
42         array[head] = null;
43         head = inc(head, array.length);
44         return e;
45     }
46
47     @Override
48     public E pollLast() {
49         if (isEmpty()) {
50             return null;
51         }
52         tail = dec(tail, array.length);
53         E e = array[tail];
54         array[tail] = null;
55         return e;
56     }
57
58     @Override
59     public E peekFirst() {
60         if (isEmpty()) {
61             return null;
62         }
63         return array[head];
64     }
65
66     @Override
67     public E peekLast() {
68         if (isEmpty()) {
69             return null;
70         }
71         return array[dec(tail, array.length)];
72     }
73
74     @Override
75     public boolean isEmpty() {
76         return head == tail;
77     }
78
79     @Override
80     public boolean isFull() {
81         if (tail > head) {
82             return tail - head == array.length - 1;
83         } else if (tail < head) {
84             return head - tail == 1;
85         } else {
86             return false;
87         }
88     }
89
90     @Override
91     public Iterator<E> iterator() {
92         return new Iterator<E>() {
```

```


93         int p = head;
94         @Override
95         public boolean hasNext() {
96             return p != tail;
97         }
98
99         @Override
100        public E next() {
101            E e = array[p];
102            p = inc(p, array.length);
103            return e;
104        }
105    };
106
107
108    E[] array;
109    int head;
110    int tail;
111
112    @SuppressWarnings("unchecked")
113    public ArrayDequeue(int capacity) {
114        array = (E[]) new Object[capacity + 1];
115    }
116
117    static int inc(int i, int length) {
118        if (i + 1 >= length) {
119            return 0;
120        }
121        return i + 1;
122    }
123
124    static int dec(int i, int length) {
125        if (i - 1 < 0) {
126            return length - 1;
127        }
128        return i - 1;
129    }
130 }

```

数组实现中，如果存储的是基本类型，那么无需考虑内存释放，例如

 image-20230110084245095

但如果存储的是引用类型，应当设置该位置的引用为 null，以便内存及时释放

 image-20230110084632543

## 2.7 优先级队列

# 无序数组实现

## 要点

1. 入队保持顺序
2. 出队前找到优先级最高的出队，相当于一次选择排序

```
1 public class PriorityQueue1<E extends Priority> implements Queue<E> {
2
3     Priority[] array;
4     int size;
5
6     public PriorityQueue1(int capacity) {
7         array = new Priority[capacity];
8     }
9
10    @Override // O(1)
11    public boolean offer(E e) {
12        if (isFull()) {
13            return false;
14        }
15        array[size++] = e;
16        return true;
17    }
18
19    // 返回优先级最高的索引值
20    private int selectMax() {
21        int max = 0;
22        for (int i = 1; i < size; i++) {
23            if (array[i].priority() > array[max].priority()) {
24                max = i;
25            }
26        }
27        return max;
28    }
29
30    @Override // O(n)
31    public E poll() {
32        if (isEmpty()) {
33            return null;
34        }
35        int max = selectMax();
36        E e = (E) array[max];
37        remove(max);
38        return e;
39    }
40
41    private void remove(int index) {
42        if (index < size - 1) {
43            System.arraycopy(array, index + 1,
44                             array, index, size - 1 - index);
45        }
46        array[--size] = null; // help GC
47    }
48 }
```

```

49     @Override
50     public E peek() {
51         if (isEmpty()) {
52             return null;
53         }
54         int max = selectMax();
55         return (E) array[max];
56     }
57
58     @Override
59     public boolean isEmpty() {
60         return size == 0;
61     }
62
63     @Override
64     public boolean isFull() {
65         return size == array.length;
66     }
67 }

```

- 视频中忘记了 help GC，注意一下

## 有序数组实现

### 要点

1. 入队后排好序，优先级最高的排列在尾部
2. 出队只需删除尾部元素即可

```

1  public class PriorityQueue2<E extends Priority> implements Queue<E> {
2
3      Priority[] array;
4      int size;
5
6      public PriorityQueue2(int capacity) {
7          array = new Priority[capacity];
8      }
9
10     // o(n)
11     @Override
12     public boolean offer(E e) {
13         if (isFull()) {
14             return false;
15         }
16         insert(e);
17         size++;
18         return true;
19     }
20
21     // 一轮插入排序
22     private void insert(E e) {
23         int i = size - 1;
24         while (i >= 0 && array[i].priority() > e.priority()) {

```



```

25         array[i + 1] = array[i];
26         i--;
27     }
28     array[i + 1] = e;
29 }
30
31 // o(1)
32 @Override
33 public E poll() {
34     if (isEmpty()) {
35         return null;
36     }
37     E e = (E) array[size - 1];
38     array[--size] = null; // help GC
39     return e;
40 }
41
42 @Override
43 public E peek() {
44     if (isEmpty()) {
45         return null;
46     }
47     return (E) array[size - 1];
48 }
49
50 @Override
51 public boolean isEmpty() {
52     return size == 0;
53 }
54
55 @Override
56 public boolean isFull() {
57     return size == array.length;
58 }
59 }

```

## 堆实现

计算机科学中，堆是一种基于树的数据结构，通常用**完全二叉树**实现。堆的特性如下

- 在大顶堆中，任意节点 C 与它的父节点 P 符合  $P.value \geq C.value$
- 而小顶堆中，任意节点 C 与它的父节点 P 符合  $P.value \leq C.value$
- 最顶层的节点（没有父亲）称之为 root 根节点

In computer science, a **heap** is a specialized tree-based data structure which is essentially an almost complete tree that satisfies the **heap property**: in a *max heap*, for any given node C, if P is a parent node of C, then the *key* (the *value*) of P is greater than or equal to the key of C. In a *min heap*, the key of P is less than or equal to the key of C. The node at the "top" of the heap (with no parents) is called the **root** node

例1 - 满二叉树 (Full Binary Tree) 特点：每一层都是填满的



例2 - 完全二叉树 (Complete Binary Tree) 特点: 最后一层可能未填满, 靠左对齐



例3 - 大顶堆



例4 - 小顶堆



完全二叉树可以使用数组来表示



特征

- 如果从索引 0 开始存储节点数据
  - 节点  $i$  的父节点为  $\text{floor}((i - 1)/2)$ , 当  $i > 0$  时
  - 节点  $i$  的左子节点为  $2i + 1$ , 右子节点为  $2i + 2$ , 当然它们得  $< \text{size}$
- 如果从索引 1 开始存储节点数据
  - 节点  $i$  的父节点为  $\text{floor}(i/2)$ , 当  $i > 1$  时
  - 节点  $i$  的左子节点为  $2i$ , 右子节点为  $2i + 1$ , 同样得  $< \text{size}$

代码

```
1 public class PriorityQueue4<E extends Priority> implements Queue<E> {
2
3     Priority[] array;
4     int size;
5
6     public PriorityQueue4(int capacity) {
7         array = new Priority[capacity];
8     }
9
10    @Override
11    public boolean offer(E offered) {
12        if (isFull()) {
13            return false;
14        }
15        int child = size++;
16        int parent = (child - 1) / 2;
17        while (child > 0 && offered.priority() > array[parent].priority()) {
18            array[child] = array[parent];
19            child = parent;
20            parent = (child - 1) / 2;
21        }
22        array[child] = offered;
23        return true;
24    }
25 }
```

```

26
27     private void swap(int i, int j) {
28         Priority t = array[i];
29         array[i] = array[j];
30         array[j] = t;
31     }
32
33     @Override
34     public E poll() {
35         if (isEmpty()) {
36             return null;
37         }
38         swap(0, size - 1);
39         size--;
40         Priority e = array[size];
41         array[size] = null;
42
43         shiftDown(0);
44         return (E) e;
45     }
46
47     void shiftDown(int parent) {
48         int left = 2 * parent + 1;
49         int right = left + 1;
50         int max = parent;
51         if (left < size && array[left].priority() > array[max].priority()) {
52             max = left;
53         }
54         if (right < size && array[right].priority() > array[max].priority())
55     {
56         max = right;
57     }
58     if (max != parent) {
59         swap(max, parent);
60         shiftDown(max);
61     }
62
63     @Override
64     public E peek() {
65         if (isEmpty()) {
66             return null;
67         }
68         return (E) array[0];
69     }
70
71     @Override
72     public boolean isEmpty() {
73         return size == 0;
74     }
75
76     @Override
77     public boolean isFull() {
78         return size == array.length;
79     }

```

## 2.8 阻塞队列

之前的队列在很多场景下都不能很好地工作，例如

1. 大部分场景要求分离向队列放入（生产者）、从队列拿出（消费者）两个角色、它们得由不同的线程来担当，而之前的实现根本没有考虑线程安全问题
2. 队列为空，那么在之前的实现里会返回 null，如果就是硬要拿到一个元素呢？只能不断循环尝试
3. 队列为满，那么再之前的实现里会返回 false，如果就是硬要塞入一个元素呢？只能不断循环尝试

因此我们需要解决的问题有

1. 用锁保证线程安全
2. 用条件变量让**等待非空线程**与**等待不满线程**进入**等待状态**，而不是不断循环尝试，让 CPU 空转

有同学对线程安全还没有足够的认识，下面举一个反例，两个线程都要执行入队操作（几乎在同一时刻）

```
1 public class TestThreadUnsafe {
2     private final String[] array = new String[10];
3     private int tail = 0;
4
5     public void offer(String e) {
6         array[tail] = e;
7         tail++;
8     }
9
10    @Override
11    public String toString() {
12        return Arrays.toString(array);
13    }
14
15    public static void main(String[] args) {
16        TestThreadUnsafe queue = new TestThreadUnsafe();
17        new Thread(() -> queue.offer("e1"), "t1").start();
18        new Thread(() -> queue.offer("e2"), "t2").start();
19    }
20 }
```

执行的时间序列如下，假设初始状态 tail = 0，在执行过程中由于 CPU 在两个线程之间切换，造成了指令交错

线程1	线程2	说明
array[tail]=e1		线程1 向 tail 位置加入 e1 这个元素，但还没来得及执行 tail++
	array[tail]=e2	线程2 向 tail 位置加入 e2 这个元素，覆盖掉了 e1
	tail++	tail 自增为1
tail++		tail 自增为2

线程1	线程2	说明
		最后状态 tail 为 2，数组为 [e2, null, null ...]

糟糕的是，由于指令交错的顺序不同，得到的结果不止以上一种，宏观上造成混乱的效果

## 单锁实现

Java 中要防止代码段交错执行，需要使用锁，有两种选择

- synchronized 代码块，属于关键字级别提供锁保护，功能少
- ReentrantLock 类，功能丰富

以 ReentrantLock 为例

```
1 ReentrantLock lock = new ReentrantLock();
2
3 public void offer(String e) {
4     lock.lockInterruptibly();
5     try {
6         array[tail] = e;
7         tail++;
8     } finally {
9         lock.unlock();
10    }
11 }
```

只要两个线程执行上段代码时，锁对象是同一个，就能保证 try 块内的代码的执行不会出现指令交错现象，即执行顺序只可能是下面两种情况之一

线程1	线程2	说明
lock.lockInterruptibly()		t1对锁对象上锁
array[tail]=e1		
	lock.lockInterruptibly()	即使 CPU 切换到线程2，但由于t1已经对该对象上锁，因此线程2卡在这儿进不去
tail++		切换回线程1 执行后续代码
lock.unlock()		线程1 解锁
	array[tail]=e2	线程2 此时才能获得锁，执行它的代码
	tail++	

- 另一种情况是线程2 先获得锁，线程1 被挡在外面
- 要明白**保护的本质**，本例中是保护的是 tail 位置读写的安全

事情还没有完，上面的例子是队列还没有放满的情况，考虑下面的代码（这回锁同时保护了 tail 和 size 的读写安全）

```

1 ReentrantLock lock = new ReentrantLock();
2 int size = 0;
3
4 public void offer(String e) {
5     lock.lockInterruptibly();
6     try {
7         if(isFull()) {
8             // 满了怎么办?
9         }
10        array[tail] = e;
11        tail++;
12
13        size++;
14    } finally {
15        lock.unlock();
16    }
17 }
18
19 private boolean isFull() {
20     return size == array.length;
21 }

```

之前是返回 false 表示添加失败，前面分析过想达到这么一种效果：

- 在队列满时，不是立刻返回，而是当前线程进入等待
- 什么时候队列不满了，再唤醒这个等待的线程，从上次的代码处继续向下运行

ReentrantLock 可以配合条件变量来实现，代码进化为

```

1 ReentrantLock lock = new ReentrantLock();
2 Condition tailWaits = lock.newCondition(); // 条件变量
3 int size = 0;
4
5 public void offer(String e) {
6     lock.lockInterruptibly();
7     try {
8         while (isFull()) {
9             tailWaits.await(); // 当队列满时，当前线程进入 tailWaits 等待
10        }
11        array[tail] = e;
12        tail++;
13
14        size++;
15    } finally {
16        lock.unlock();
17    }
18 }
19
20 private boolean isFull() {
21     return size == array.length;
22 }

```

- 条件变量底层也是个队列，用来存储这些需要等待的线程，当队列满了，就会将 offer 线程加入条件队列，并暂时释放锁

- 将来我们的队列如果不满了（由 poll 线程那边得知）可以调用 tailWaits.signal() 来唤醒 tailWaits 中首个等待的线程，被唤醒的线程会再次抢到锁，从上次 await 处继续向下运行

思考为何要用 while 而不是 if，设队列容量是 3

操作前	offer(4)	offer(5)	poll()	操作后
[1 2 3]	队列满，进入 tailWaits 等待			[1 2 3]
[1 2 3]			取走 1，队列不满，唤醒线程	[2 3]
[2 3]		抢先获得锁，发现不满，放入 5		[2 3 5]
[2 3 5]	从上次等待处直接向下执行			[2 3 5 ?]

关键点：

- 从 tailWaits 中唤醒的线程，会与新来的 offer 的线程争抢锁，谁能抢到是不一定的，如果后者先抢到，就会导致条件又发生变化
- 这种情况称之为**虚假唤醒**，唤醒后应该重新检查条件，看是不是得重新进入等待

最后的实现代码

```

1  /**
2   * 单锁实现
3   * @param <E> 元素类型
4   */
5  public class BlockingQueue1<E> implements BlockingQueue<E> {
6      private final E[] array;
7      private int head = 0;
8      private int tail = 0;
9      private int size = 0; // 元素个数
10
11     @SuppressWarnings("all")
12     public BlockingQueue1(int capacity) {
13         array = (E[]) new Object[capacity];
14     }
15
16     ReentrantLock lock = new ReentrantLock();
17     Condition tailWaits = lock.newCondition();
18     Condition headWaits = lock.newCondition();
19
20     @Override
21     public void offer(E e) throws InterruptedException {
22         lock.lockInterruptibly();
23         try {
24             while (isFull()) {
25                 tailWaits.await();
26             }

```

```

27         array[tail] = e;
28         if (++tail == array.length) {
29             tail = 0;
30         }
31         size++;
32         headwaits.signal();
33     } finally {
34         lock.unlock();
35     }
36 }
37
38 @Override
39 public void offer(E e, long timeout) throws InterruptedException {
40     lock.lockInterruptibly();
41     try {
42         long t = TimeUnit.MILLISECONDS.toNanos(timeout);
43         while (isFull()) {
44             if (t <= 0) {
45                 return;
46             }
47             t = tailwaits.awaitNanos(t);
48         }
49         array[tail] = e;
50         if (++tail == array.length) {
51             tail = 0;
52         }
53         size++;
54         headwaits.signal();
55     } finally {
56         lock.unlock();
57     }
58 }
59
60 @Override
61 public E poll() throws InterruptedException {
62     lock.lockInterruptibly();
63     try {
64         while (isEmpty()) {
65             headwaits.await();
66         }
67         E e = array[head];
68         array[head] = null; // help GC
69         if (++head == array.length) {
70             head = 0;
71         }
72         size--;
73         tailwaits.signal();
74         return e;
75     } finally {
76         lock.unlock();
77     }
78 }
79
80 private boolean isEmpty() {
81     return size == 0;

```



```

82     }
83
84     private boolean isFull() {
85         return size == array.length;
86     }
87 }

```

- `public void offer(E e, long timeout) throws InterruptedException` 是带超时的版本，可以只等待一段时间，而不是永久等下去，类似的 `poll` 也可以做带超时的版本，这个留给大家了

### 注意

- JDK 中 `BlockingQueue` 接口的方法命名与我的示例有些差异
  - 方法 `offer(E e)` 是非阻塞的实现，阻塞实现方法为 `put(E e)`
  - 方法 `poll()` 是非阻塞的实现，阻塞实现方法为 `take()`

## 双锁实现

单锁的缺点在于：

- 生产和消费几乎是不冲突的，唯一冲突的是生产者和消费者它们有可能同时修改 `size`
- 冲突的主要是生产者之间：多个 `offer` 线程修改 `tail`
- 冲突的还有消费者之间：多个 `poll` 线程修改 `head`

如果希望进一步提高性能，可以用两把锁

- 一把锁保护 `tail`
- 另一把锁保护 `head`

```

1  ReentrantLock headLock = new ReentrantLock(); // 保护 head 的锁
2  Condition headWaits = headLock.newCondition(); // 队列空时，需要等待的线程集合
3
4  ReentrantLock tailLock = new ReentrantLock(); // 保护 tail 的锁
5  Condition tailWaits = tailLock.newCondition(); // 队列满时，需要等待的线程集合

```

先看看 `offer` 方法的初步实现

```

1  @Override
2  public void offer(E e) throws InterruptedException {
3      tailLock.lockInterruptibly();
4      try {
5          // 队列满等待
6          while (isFull()) {
7              tailWaits.await();
8          }
9
10         // 不满则入队
11         array[tail] = e;
12         if (++tail == array.length) {
13             tail = 0;
14         }
15     }

```

```

16         // 修改 size (有问题)
17         size++;
18
19     } finally {
20         tailLock.unlock();
21     }
22 }

```

上面代码的缺点是 size 并不受 tailLock 保护，tailLock 与 headLock 是两把不同的锁，并不能实现互斥的效果。因此，size 需要用下面的代码保证原子性

```

1  AtomicInteger size = new AtomicInteger(0);    // 保护 size 的原子变量
2
3  size.getAndIncrement(); // 自增
4  size.getAndDecrement(); // 自减

```

代码修改为

```

1  @Override
2  public void offer(E e) throws InterruptedException {
3      tailLock.lockInterruptibly();
4      try {
5          // 队列满等待
6          while (isFull()) {
7              tailWaits.await();
8          }
9
10         // 不满则入队
11         array[tail] = e;
12         if (++tail == array.length) {
13             tail = 0;
14         }
15
16         // 修改 size
17         size.getAndIncrement();
18
19     } finally {
20         tailLock.unlock();
21     }
22 }

```

对称地，可以写出 poll 方法

```

1  @Override
2  public E poll() throws InterruptedException {
3      E e;
4      headLock.lockInterruptibly();
5      try {
6          // 队列空等待
7          while (isEmpty()) {
8              headWaits.await();
9          }
10
11         // 不空则出队

```

```

12     e = array[head];
13     if (++head == array.length) {
14         head = 0;
15     }
16
17     // 修改 size
18     size.getAndDecrement();
19
20 } finally {
21     headLock.unlock();
22 }
23 return e;
24 }

```

下面来看一个难题，就是如何通知 headWaits 和 tailWaits 中等待的线程，比如 poll 方法拿走一个元素，通知 tailWaits：我拿走一个，不满了噢，你们可以放了，因此代码改为

```

1  @Override
2  public E poll() throws InterruptedException {
3      E e;
4      headLock.lockInterruptibly();
5      try {
6          // 队列空等待
7          while (isEmpty()) {
8              headWaits.await();
9          }
10
11         // 不空则出队
12         e = array[head];
13         if (++head == array.length) {
14             head = 0;
15         }
16
17         // 修改 size
18         size.getAndDecrement();
19
20         // 通知 tailWaits 不满（有问题）
21         tailWaits.signal();
22
23     } finally {
24         headLock.unlock();
25     }
26     return e;
27 }

```


问题在于要使用这些条件变量的 await(), signal() 等方法需要先获得与之关联的锁，上面的代码若直接运行会出现以下错误

```

1 java.lang.IllegalMonitorStateException

```

那有同学说，加上锁不就行了吗，于是写出了下面的代码

 image-20230208160343493

发现什么问题了？两把锁这么嵌套使用，非常容易出现死锁，如下所示



因此得避免嵌套，两段加锁的代码变成了下面平级的样子



性能还可以进一步提升

1. 代码调整后 offer 并没有同时获取 tailLock 和 headLock 两把锁，因此两次加锁之间会有空隙，这个空隙内可能有其它的 offer 线程添加了更多的元素，那么这些线程都要执行 signal()，通知 poll 线程队列非空吗？
  - 每次调用 signal() 都需要这些 offer 线程先获得 headLock 锁，成本较高，要想法减少 offer 线程获得 headLock 锁的次数
  - 可以加一个条件：当 offer 增加前队列为空，即从 0 变化到不空，才由此 offer 线程来通知 headWaits，其它情况不归它管
2. 队列从 0 变化到不空，会唤醒一个等待的 poll 线程，这个线程被唤醒后，肯定能拿到 headLock 锁，因此它具备了唤醒 headWaits 上其它 poll 线程的先决条件。如果检查出此时有其它 offer 线程新增了元素（不空，但不是从 0 变化而来），那么不妨由此 poll 线程来唤醒其它 poll 线程

这个技巧被称之为级联通知（cascading notifies），类似的原因

3. 在 poll 时队列从满变化到不满，才由此 poll 线程来唤醒一个等待的 offer 线程，目的也是为了减少 poll 线程对 tailLock 上锁次数，剩下等待的 offer 线程由这个 offer 线程间接唤醒

最终的代码为

```
1 public class BlockingQueue2<E> implements BlockingQueue<E> {
2
3     private final E[] array;
4     private int head = 0;
5     private int tail = 0;
6     private final AtomicInteger size = new AtomicInteger(0);
7     ReentrantLock headLock = new ReentrantLock();
8     Condition headWaits = headLock.newCondition();
9     ReentrantLock tailLock = new ReentrantLock();
10    Condition tailWaits = tailLock.newCondition();
11
12    public BlockingQueue2(int capacity) {
13        this.array = (E[]) new Object[capacity];
14    }
15
16    @Override
17    public void offer(E e) throws InterruptedException {
18        int c;
19        tailLock.lockInterruptibly();
20        try {
21            while (isFull()) {
22                tailWaits.await();
23            }
24            array[tail] = e;
25            if (++tail == array.length) {
26                tail = 0;
27            }
28            c = size.getAndIncrement();
```

```

29         // a. 队列不满，但不是从满->不满，由此offer线程唤醒其它offer线程
30         if (c + 1 < array.length) {
31             tailWaits.signal();
32         }
33     } finally {
34         tailLock.unlock();
35     }
36     // b. 从0->不空，由此offer线程唤醒等待的poll线程
37     if (c == 0) {
38         headLock.lock();
39         try {
40             headWaits.signal();
41         } finally {
42             headLock.unlock();
43         }
44     }
45 }
46
47 @Override
48 public E poll() throws InterruptedException {
49     E e;
50     int c;
51     headLock.lockInterruptibly();
52     try {
53         while (isEmpty()) {
54             headWaits.await();
55         }
56         e = array[head];
57         if (++head == array.length) {
58             head = 0;
59         }
60         c = size.getAndDecrement();
61         // b. 队列不空，但不是从0变化到不空，由此poll线程通知其它poll线程
62         if (c > 1) {
63             headWaits.signal();
64         }
65     } finally {
66         headLock.unlock();
67     }
68     // a. 从满->不满，由此poll线程唤醒等待的offer线程
69     if (c == array.length) {
70         tailLock.lock();
71         try {
72             tailWaits.signal();
73         } finally {
74             tailLock.unlock();
75         }
76     }
77     return e;
78 }
79
80 private boolean isEmpty() {
81     return size.get() == 0;
82 }
83

```

```

84     private boolean isFull() {
85         return size.get() == array.length;
86     }
87
88 }

```

双锁实现的非常精巧，据说作者 Doug Lea 花了一年的时间才完善了此段代码

## 2.9 堆

以大顶堆为例，相对于之前的优先级队列，增加了堆化等方法

```

1  public class MaxHeap {
2      int[] array;
3      int size;
4
5      public MaxHeap(int capacity) {
6          this.array = new int[capacity];
7      }
8
9      /**
10     * 获取堆顶元素
11     *
12     * @return 堆顶元素
13     */
14     public int peek() {
15         return array[0];
16     }
17
18     /**
19     * 删除堆顶元素
20     *
21     * @return 堆顶元素
22     */
23     public int poll() {
24         int top = array[0];
25         swap(0, size - 1);
26         size--;
27         down(0);
28         return top;
29     }
30
31     /**
32     * 删除指定索引处元素
33     *
34     * @param index 索引
35     * @return 被删除元素
36     */
37     public int poll(int index) {
38         int deleted = array[index];
39         swap(index, size - 1);
40         size--;
41         down(index);

```

```

42         return deleted;
43     }
44
45     /**
46      * 替换堆顶元素
47      * @param replaced 新元素
48      */
49     public void replace(int replaced) {
50         array[0] = replaced;
51         down(0);
52     }
53
54     /**
55      * 堆的尾部添加元素
56      *
57      * @param offered 新元素
58      * @return 是否添加成功
59      */
60     public boolean offer(int offered) {
61         if (size == array.length) {
62             return false;
63         }
64         up(offered);
65         size++;
66         return true;
67     }
68
69     // 将 offered 元素上浮：直至 offered 小于父元素或到堆顶
70     private void up(int offered) {
71         int child = size;
72         while (child > 0) {
73             int parent = (child - 1) / 2;
74             if (offered > array[parent]) {
75                 array[child] = array[parent];
76             } else {
77                 break;
78             }
79             child = parent;
80         }
81         array[child] = offered;
82     }
83
84     public MaxHeap(int[] array) {
85         this.array = array;
86         this.size = array.length;
87         heapify();
88     }
89
90     // 建堆
91     private void heapify() {
92         // 如何找到最后这个非叶子节点 size / 2 - 1
93         for (int i = size / 2 - 1; i >= 0; i--) {
94             down(i);
95         }
96     }

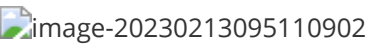
```

```

97
98 // 将 parent 索引处的元素下潜：与两个孩子较大者交换，直至没孩子或孩子没它大
99 private void down(int parent) {
100     int left = parent * 2 + 1;
101     int right = left + 1;
102     int max = parent;
103     if (left < size && array[left] > array[max]) {
104         max = left;
105     }
106     if (right < size && array[right] > array[max]) {
107         max = right;
108     }
109     if (max != parent) { // 找到了更大的孩子
110         swap(max, parent);
111         down(max);
112     }
113 }
114
115 // 交换两个索引处的元素
116 private void swap(int i, int j) {
117     int t = array[i];
118     array[i] = array[j];
119     array[j] = t;
120 }
121
122 public static void main(String[] args) {
123
124     int[] array = {1, 2, 3, 4, 5, 6, 7};
125     MaxHeap maxHeap = new MaxHeap(array);
126     System.out.println(Arrays.toString(maxHeap.array));
127 }
128 }
```

建堆

Floyd 建堆算法作者（也是之前龟兔赛跑判环作者）：



- 1. 找到最后一个非叶子节点
- 2. 从后向前，对每个节点执行下潜

一些规律

- 一棵满二叉树节点个数为  $2^h - 1$ ，如下例中高度  $h = 3$  节点数是  $2^3 - 1 = 7$
- 非叶子节点范围为  $[0, size/2 - 1]$

算法时间复杂度分析



下面看交换次数的推导：设节点高度为 3

	本层节点数	高度	下潜最多交换次数（高度-1）
--	-------	----	----------------



	本层节点数	高度	下潜最多交换次数 (高度-1)
4567 这层	4	1	0
23这层	2	2	1
1这层	1	3	2

每一层的交换次数为：节点个数 \* 此节点交换次数，总的交换次数为

$$\begin{aligned}
 &4 * 0 + 2 * 1 + 1 * 2 \\
 &\frac{8}{2} * 0 + \frac{8}{4} * 1 + \frac{8}{8} * 2 \\
 &\frac{8}{2^1} * 0 + \frac{8}{2^2} * 1 + \frac{8}{2^3} * 2
 \end{aligned}$$

即

$$\sum_{i=1}^h \left( \frac{2^h}{2^i} * (i - 1) \right)$$

在 <https://www.wolframalpha.com/> 输入

```
1 | Sum[(40)Divide[Power[2,x],Power[2,i]]*(40)i-1(41)\(41),{i,1,x}]
```

推导出

$$2^h - h - 1$$

其中  $2^h \approx n$ ,  $h \approx \log_2 n$ , 因此有时间复杂度  $O(n)$

## 2.10 二叉树

二叉树是这么一种树状结构：每个节点最多有两个孩子，左孩子和右孩子

重要的二叉树结构

- 完全二叉树 (complete binary tree) 是一种二叉树结构，除最后一层以外，每一层都必须填满，填充时要遵从先左后右
- 平衡二叉树 (balance binary tree) 是一种二叉树结构，其中每个节点的左右子树高度相差不超过 1

## 存储

存储方式分为两种

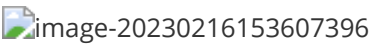
1. 定义树节点与左、右孩子引用 (TreeNode)
2. 使用数组，前面讲堆时用过，若以 0 作为树的根，索引可以通过如下方式计算
  - 父 = floor((子 - 1) / 2)
  - 左孩子 = 父 \* 2 + 1
  - 右孩子 = 父 \* 2 + 2

# 遍历

遍历也分为两种

- 1. 广度优先遍历 (Breadth-first order)：尽可能先访问距离根最近的节点，也称为层序遍历
- 2. 深度优先遍历 (Depth-first order)：对于二叉树，可以进一步分成三种（要深入到叶子节点）
  - 1. pre-order 前序遍历，对于每一棵子树，先访问该节点，然后是左子树，最后是右子树
  - 2. in-order 中序遍历，对于每一棵子树，先访问左子树，然后是该节点，最后是右子树
  - 3. post-order 后序遍历，对于每一棵子树，先访问左子树，然后是右子树，最后是该节点

## 广度优先



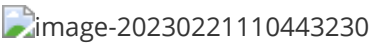
本轮开始时队列	本轮访问节点
[1]	1
[2, 3]	2
[3, 4]	3
[4, 5, 6]	4
[5, 6]	5
[6, 7, 8]	6
[7, 8]	7
[8]	8
[]	

- 1. 初始化，将根节点加入队列
- 2. 循环处理队列中每个节点，直至队列为空
- 3. 每次循环内处理节点后，将它的孩子节点（即下一层的节点）加入队列























注意

- 以上用队列来层序遍历是针对 `TreeNode` 这种方式表示的二叉树
- 对于数组表现的二叉树，则直接遍历数组即可，自然为层序遍历的顺序

## 深度优先



栈暂存	已处理	前序遍历	中序遍历
[1]	1  左  右	1	

栈暂存	已处理	前序遍历	中序遍历
[1, 2]	2 ✓ 左  右  1 ✓ 左  右 	2	
[1, 2, 4]	4 ✓ 左 ✓ 右 ✓ 2 ✓ 左  右  1 ✓ 左  右 	4	4
[1, 2]	2 ✓ 左 ✓ 右 ✓ 1 ✓ 左  右 		2
[1]	1 ✓ 左 ✓ 右 		1
[1, 3]	3 ✓ 左  右  1 ✓ 左 ✓ 右 	3	
[1, 3, 5]	5 ✓ 左 ✓ 右 ✓ 3 ✓ 左  右  1 ✓ 左 ✓ 右 	5	5
[1, 3]	3 ✓ 左 ✓ 右  1 ✓ 左 ✓ 右 		3
[1, 3, 6]	6 ✓ 左 ✓ 右 ✓ 3 ✓ 左 ✓ 右  1 ✓ 左 ✓ 右 	6	6
[1, 3]	3 ✓ 左 ✓ 右 ✓ 1 ✓ 左 ✓ 右 		
[1]	1 ✓ 左 ✓ 右 ✓		
[]			

## 递归实现

```

1  /**
2   * <h3>前序遍历</h3>
3   * @param node 节点
4   */
5  static void preOrder(TreeNode node) {
6      if (node == null) {
7          return;
8      }
9      System.out.print(node.val + "\t"); // 值
10     preOrder(node.left); // 左
11     preOrder(node.right); // 右
12 }
13
14 /**
15 * <h3>中序遍历</h3>
16 * @param node 节点
17 */
18 static void inOrder(TreeNode node) {

```

```

19     if (node == null) {
20         return;
21     }
22     inOrder(node.left); // 左
23     System.out.print(node.val + "\t"); // 值
24     inOrder(node.right); // 右
25 }
26
27 /**
28  * <h3>后序遍历</h3>
29  * @param node 节点
30  */
31 static void postOrder(TreeNode node) {
32     if (node == null) {
33         return;
34     }
35     postOrder(node.left); // 左
36     postOrder(node.right); // 右
37     System.out.print(node.val + "\t"); // 值
38 }

```

## 非递归实现

### 前序遍历

```

1  LinkedListStack<TreeNode> stack = new LinkedListStack<>();
2  TreeNode curr = root;
3
4  while (!stack.isEmpty() || curr != null) {
5      if (curr != null) {
6          stack.push(curr);
7          System.out.println(curr);
8          curr = curr.left;
9      } else {
10         TreeNode pop = stack.pop();
11         curr = pop.right;
12     }
13
14 }

```

### 中序遍历

```

1  LinkedListStack<TreeNode> stack = new LinkedListStack<>();
2  TreeNode curr = root;
3
4  while (!stack.isEmpty() || curr != null) {
5      if (curr != null) {
6          stack.push(curr);
7          curr = curr.left;
8      } else {
9          TreeNode pop = stack.pop();
10         System.out.println(pop);
11         curr = pop.right;
12     }
13 }

```

## 后序遍历

```

1  LinkedListStack<TreeNode> stack = new LinkedListStack<>();
2  TreeNode curr = root;
3  TreeNode pop = null;
4
5  while (!stack.isEmpty() || curr != null) {
6      if (curr != null) {
7          stack.push(curr);
8          curr = curr.left;
9      } else {
10         TreeNode peek = stack.peek();
11         if (peek.right == null || peek.right == pop) {
12             pop = stack.pop();
13             System.out.println(pop);
14         } else {
15             curr = peek.right;
16         }
17     }
18 }

```

对于后序遍历，向回走时，需要处理完右子树才能 pop 出栈。如何知道右子树处理完成呢？

- 如果栈顶元素的 *right*  $\equiv$  *null* 表示没啥可处理的，可以出栈
- 如果栈顶元素的 *right*  $\neq$  *null*,
  - 那么使用 *lastPop* 记录最近出栈的节点，即表示从这个节点向回走
  - 如果栈顶元素的 *right*  $\equiv$  *lastPop* 此时应当出栈

对于前、中两种遍历，实际以上代码从右子树向回走时，并未走完全程（stack 提前出栈了）后序遍历以上代码是走完全程了

## 统一写法

下面是一种统一的写法，依据后序遍历修改

```

1  LinkedList<TreeNode> stack = new LinkedList<>();
2
3  TreeNode curr = root; // 代表当前节点
4  TreeNode pop = null; // 最近一次弹栈的元素
5  while (curr != null || !stack.isEmpty()) {

```

```

6      if (curr != null) {
7          colorPrintln("前: " + curr.val, 31);
8          stack.push(curr); // 压入栈, 为了记住回来的路
9          curr = curr.left;
10     } else {
11         TreeNode peek = stack.peek();
12         // 右子树可以不处理, 对中序来说, 要在右子树处理之前打印
13         if (peek.right == null) {
14             colorPrintln("中: " + peek.val, 36);
15             pop = stack.pop();
16             colorPrintln("后: " + pop.val, 34);
17         }
18         // 右子树处理完成, 对中序来说, 无需打印
19         else if (peek.right == pop) {
20             pop = stack.pop();
21             colorPrintln("后: " + pop.val, 34);
22         }
23         // 右子树待处理, 对中序来说, 要在右子树处理之前打印
24         else {
25             colorPrintln("中: " + peek.val, 36);
26             curr = peek.right;
27         }
28     }
29 }
30
31 public static void colorPrintln(String origin, int color) {
32     System.out.printf("\033[%dm%s\033[0m\n", color, origin);
33 }

```

一张图演示三种遍历

 Sorted\_binary\_tree\_ALL\_RGB.svg

- 红色: 前序遍历顺序
- 绿色: 中序遍历顺序
- 蓝色: 后续遍历顺序

## 三. 练习

### 3.1 时间复杂度

用函数  $f(n)$  表示算法效率与数据规模的关系, 假设每次解决问题需要 1 微秒 ( $10^{-6}$  秒), 进行估算:

1. 如果  $f(n) = n^2$  那么 1 秒能解决多少次问题? 1 天呢?
2. 如果  $f(n) = \log_2(n)$  那么 1 秒能解决多少次问题? 1 天呢?
3. 如果  $f(n) = n!$  那么 1 秒能解决多少次问题? 1 天呢?

参考解答

1. 1秒  $\sqrt{10^6} = 1000$  次, 1天  $\sqrt{10^6 * 3600 * 24} \approx 293938$  次
2. 1秒  $2^{1,000,000}$  次, 一天  $2^{86,400,000,000}$
3. 推算如下

- $10! = 3,628,800$  1秒能解决 1,000,000 次, 因此次数为 9 次
- $14! = 87,178,291,200$ , 一天能解决 86,400,000,000 次, 因此次数为 13 次

## 3.2 二分查找

- [69. x 的平方根 - 力扣 \(LeetCode\)](#)

### E01. 二分查找-力扣 704 题

**要点:** 减而治之, 可以用递归或非递归实现

给定一个  $n$  个元素有序的 (升序) 整型数组 `nums` 和一个目标值 `target`, 写一个函数搜索 `nums` 中的 `target`, 如果目标值存在返回下标, 否则返回 -1

例如

```
1 输入: nums = [-1,0,3,5,9,12], target = 9
2 输出: 4
3 解释: 9 出现在 nums 中并且下标为 4
4
5 输入: nums = [-1,0,3,5,9,12], target = 2
6 输出: -1
7 解释: 2 不存在 nums 中因此返回 -1
```

**参考答案:** 略, 可以用讲过的任意一种二分求解

### E02. 搜索插入位置-力扣 35 题

**要点:** 理解谁代表插入位置

给定一个排序数组和一个目标值

- 在数组中找到目标值, 并返回其索引
- 如果目标值不存在于数组中, 返回它将会被按顺序插入的位置

例如

```
1 输入: nums = [1,3,5,6], target = 5
2 输出: 2
3
4 输入: nums = [1,3,5,6], target = 2
5 输出: 1
6
7 输入: nums = [1,3,5,6], target = 7
8 输出: 4
```

**参考答案1:** 用二分查找基础版代码改写, 基础版中, 找到返回 `m`, 没找到 `i` 代表插入点, 因此有

```
1 public int searchInsert(int[] a, int target) {
```

```

2     int i = 0, j = a.length - 1;
3     while (i <= j) {
4         int m = (i + j) >>> 1;
5         if (target < a[m]) {
6             j = m - 1;
7         } else if (a[m] < target) {
8             i = m + 1;
9         } else {
10            return m;
11        }
12    }
13    return i; // 原始 return -1
14 }

```

**参考答案2：**用二分查找平衡版改写，平衡版中

- 如果  $\text{target} == a[i]$  返回  $i$  表示找到
- 如果  $\text{target} < a[i]$ ，例如  $\text{target} = 2$ ， $a[i] = 3$ ，这时就应该在  $i$  位置插入 2
- 如果  $a[i] < \text{target}$ ，例如  $a[i] = 3$ ， $\text{target} = 4$ ，这时就应该在  $i+1$  位置插入 4

```

1 public static int searchInsert(int[] a, int target) {
2     int i = 0, j = a.length;
3     while (1 < j - i) {
4         int m = (i + j) >>> 1;
5         if (target < a[m]) {
6             j = m;
7         } else {
8             i = m;
9         }
10    }
11    return (target <= a[i]) ? i : i + 1;
12    // 原始 (target == a[i]) ? i : -1;
13 }

```

**参考答案3：**用 leftmost 版本解，返回值即为插入位置（并能处理元素重复的情况）

```

1 public int searchInsert(int[] a, int target) {
2     int i = 0, j = a.length - 1;
3     while(i <= j) {
4         int m = (i + j) >>> 1;
5         if(target <= a[m]) {
6             j = m - 1;
7         } else {
8             i = m + 1;
9         }
10    }
11    return i;
12 }

```



## E03. 搜索开始结束位置-力扣 34 题

给你一个按照非递减顺序排列的整数数组 `nums`，和一个目标值 `target`。请你找出给定目标值在数组中的开始位置和结束位置。

如果数组中不存在目标值 `target`，返回 `[-1, -1]`。

你必须设计并实现时间复杂度为  $O(\log n)$  的算法解决此问题

例如

```
1  输入: nums = [5,7,7,8,8,10], target = 8
2  输出: [3,4]
3
4  输入: nums = [5,7,7,8,8,10], target = 6
5  输出: [-1,-1]
6
7  输入: nums = [], target = 0
8  输出: [-1,-1]
```

### 参考答案

```
1  public static int left(int[] a, int target) {
2      int i = 0, j = a.length - 1;
3      int candidate = -1;
4      while (i <= j) {
5          int m = (i + j) >>> 1;
6          if (target < a[m]) {
7              j = m - 1;
8          } else if (a[m] < target) {
9              i = m + 1;
10         } else {
11             candidate = m;
12             j = m - 1;
13         }
14     }
15     return candidate;
16 }
17
18 public static int right(int[] a, int target) {
19     int i = 0, j = a.length - 1;
20     int candidate = -1;
21     while (i <= j) {
22         int m = (i + j) >>> 1;
23         if (target < a[m]) {
24             j = m - 1;
25         } else if (a[m] < target) {
26             i = m + 1;
27         } else {
28             candidate = m;
29             i = m + 1;
30         }
31     }
```

```

31     }
32     return candidate;
33 }
34
35 public static int[] searchRange(int[] nums, int target) {
36     int x = left(nums, target);
37     if(x == -1) {
38         return new int[] {-1, -1};
39     } else {
40         return new int[] {x, right(nums, target)};
41     }
42 }

```

## 3.3 递归 - single recursion

### E03. 二分查找

```

1 public static int binarySearch(int[] a, int target) {
2     return recursion(a, target, 0, a.length - 1);
3 }
4
5 public static int recursion(int[] a, int target, int i, int j) {
6     if (i > j) {
7         return -1;
8     }
9     int m = (i + j) >> 1;
10    if (target < a[m]) {
11        return recursion(a, target, i, m - 1);
12    } else if (a[m] < target) {
13        return recursion(a, target, m + 1, j);
14    } else {
15        return m;
16    }
17 }

```

### E04. 冒泡排序

```

1 public static void main(String[] args) {
2     int[] a = {3, 2, 6, 1, 5, 4, 7};
3     bubble(a, 0, a.length - 1);
4     System.out.println(Arrays.toString(a));
5 }
6
7 private static void bubble(int[] a, int low, int high) {
8     if(low == high) {
9         return;
10    }
11    int j = low;
12    for (int i = low; i < high; i++) {
13        if (a[i] > a[i + 1]) {

```

```

14         swap(a, i, i + 1);
15         j = i;
16     }
17 }
18 bubble(a, low, j);
19 }
20
21 private static void swap(int[] a, int i, int j) {
22     int t = a[i];
23     a[i] = a[j];
24     a[j] = t;
25 }

```

- low 与 high 为未排序范围
- j 表示的是未排序的边界，下一次递归时的 high
  - 发生交换，意味着有无序情况
  - 最后一次交换（以后没有无序）时，左侧 i 仍是无序，右侧 i+1 已然有序
- 视频中讲解的是只考虑 high 边界的情况，参考以上代码，理解在 low .. high 范围内的处理方法

## E05. 插入排序

```

1 public static void main(String[] args) {
2     int[] a = {3, 2, 6, 1, 5, 7, 4};
3     insertion(a, 1, a.length - 1);
4     System.out.println(Arrays.toString(a));
5 }
6
7 private static void insertion(int[] a, int low, int high) {
8     if (low > high) {
9         return;
10    }
11    int i = low - 1;
12    int t = a[low];
13    while (i >= 0 && a[i] > t) {
14        a[i + 1] = a[i];
15        i--;
16    }
17    if (i + 1 != low) {
18        a[i + 1] = t;
19    }
20    insertion(a, low + 1, high);
21 }

```

- 已排序区域：[0 .. i .. low-1]
- 未排序区域：[low .. high]
- 视频中讲解的是只考虑 low 边界的情况，参考以上代码，理解 low-1 .. high 范围内的处理方法
- 扩展：利用二分查找 leftmost 版本，改进寻找插入位置的代码

## E06. 约瑟夫问题<sup>12</sup>

$n$  个人排成圆圈，从头开始报数，每次数到第  $m$  个人（ $m$  从 1 开始）杀之，继续从下一个人重复以上过程，求最后活下来的人是谁？

### 方法1

根据最后的存活者  $a$  倒推出它在上一轮的索引号

$f(n,m)$	本轮索引	为了让 $a$ 是这个索引，上一轮应当这样排	规律
$f(1,3)$	0	x x x a	$(0 + 3) \% 2$
$f(2,3)$	1	x x x 0 a	$(1 + 3) \% 3$
$f(3,3)$	1	x x x 0 a	$(1 + 3) \% 4$
$f(4,3)$	0	x x x a	$(0 + 3) \% 5$
$f(5,3)$	3	x x x 0 1 2 a	$(3 + 3) \% 6$
$f(6,3)$	0	x x x a	

### 方法2

设  $n$  为总人数， $m$  为报数次数，解返回的是这些人的索引，从0开始

$f(n, m)$	解	规律
$f(1, 3)$	0	
$f(2, 3)$	0 1 $\Rightarrow$ 1	$3\%2=1$
$f(3, 3)$	0 1 2 $\Rightarrow$ 0 1	$3\%3=0$
$f(4, 3)$	0 1 2 3 $\Rightarrow$ 3 0 1	$3\%4=3$
$f(5, 3)$	0 1 2 3 4 $\Rightarrow$ 3 4 0 1	$3\%5=3$
$f(6, 3)$	0 1 2 3 4 5 $\Rightarrow$ 3 4 5 0 1	$3\%6=3$

### 一. 找出等价函数

规律：下次报数的起点为  $k = m \% n$

- 首次出列人的序号是  $k - 1$ ，剩下的  $n - 1$  个人重新组成约瑟夫环
- 下次从  $k$  开始数，序号如下
  - $k, k + 1, \dots, 0, 1, k - 2$ ，如上例中 3 4 5 0 1

这个函数称之为  $g(n - 1, m)$ ，它的最终结果与  $f(n, m)$  是相同的。

### 二. 找到映射函数

现在想办法找到  $g(n-1, m)$  与  $f(n-1, m)$  的对应关系, 即

$$3 \rightarrow 0$$

$$4 \rightarrow 1$$

$$5 \rightarrow 2$$

$$0 \rightarrow 3$$

$$1 \rightarrow 4$$

映射函数为

$$mapping(x) = \begin{cases} x - k & x = [k..n-1] \\ x + n - k & x = [0..k-2] \end{cases}$$

等价于下面函数

$$mapping(x) = (x + n - k) \% n$$

代入测试一下

$$3 \rightarrow (3 + 6 - 3) \% 6 \rightarrow 0$$

$$4 \rightarrow (4 + 6 - 3) \% 6 \rightarrow 1$$

$$5 \rightarrow (5 + 6 - 3) \% 6 \rightarrow 2$$

$$0 \rightarrow (0 + 6 - 3) \% 6 \rightarrow 3$$

$$1 \rightarrow (1 + 6 - 3) \% 6 \rightarrow 4$$

综上有

$$f(n-1, m) = mapping(g(n-1, m))$$

### 三. 求逆映射函数

映射函数是根据  $x$  计算  $y$ , 逆映射函数即根据  $y$  得到  $x$

$$mapping^{-1}(x) = (x + k) \% n$$

代入测试一下

$$0 \rightarrow (0 + 3) \% 6 \rightarrow 3$$

$$1 \rightarrow (1 + 3) \% 6 \rightarrow 4$$

$$2 \rightarrow (2 + 3) \% 6 \rightarrow 5$$

$$3 \rightarrow (3 + 3) \% 6 \rightarrow 0$$

$$4 \rightarrow (4 + 3) \% 6 \rightarrow 1$$

因此可以求得

$$g(n-1, m) = mapping^{-1}(f(n-1, m))$$

### 四. 递推式

代入推导

$$\begin{aligned} f(n, m) &= g(n-1, m) \\ &= mapping^{-1}(f(n-1, m)) \\ &= (f(n-1, m) + k) \% n \\ &= (f(n-1, m) + m \% n) \% n \\ &= (f(n-1, m) + m) \% n \end{aligned}$$

最后一步化简是利用了模运算法则

$(a + b) \% n = (a \% n + b \% n) \% n$  例如

- $(6 + 6) \% 5 = 2 = (6 + 6 \% 5) \% 5$
- $(6 + 5) \% 5 = 1 = (6 + 5 \% 5) \% 5$
- $(6 + 4) \% 5 = 0 = (6 + 4 \% 5) \% 5$

最终递推式

$$f(n, m) = \begin{cases} (f(n-1, m) + m) \% n & n > 1 \\ 0 & n = 1 \end{cases}$$

## 3.4 递归 - multi recursion

### E02. 汉诺塔 <sup>13</sup>

Tower of Hanoi, 是一个源于印度古老传说: 大梵天创建世界时做了三根金刚石柱, 在一根柱子从下往上按大小顺序摞着 64 片黄金圆盘, 大梵天命令婆罗门把圆盘重新摆放在另一根柱子上, 并且规定

- 一次只能移动一个圆盘
- 小圆盘上不能放大圆盘

下面的动图演示了4片圆盘的移动方法




使用程序代码模拟圆盘的移动过程, 并估算出时间复杂度

#### 思路

- 假设每根柱子标号  $a, b, c$ , 每个圆盘用  $1, 2, 3 \dots$  表示其大小, 圆盘初始在  $a$ , 要移动到的目标是  $c$
- 如果只有一个圆盘, 此时是最小问题, 可以直接求解
  - 移动圆盘1  $a \mapsto c$
- 如果有两个圆盘, 那么
  - 圆盘1  $a \mapsto b$
  - 圆盘2  $a \mapsto c$
  - 圆盘1  $b \mapsto c$
- 如果有三个圆盘, 那么
  - 圆盘12  $a \mapsto b$
  - 圆盘3  $a \mapsto c$
  - 圆盘12  $b \mapsto c$
- 如果有四个圆盘, 那么
  - 圆盘123  $a \mapsto b$
  - 圆盘4  $a \mapsto c$
  - 圆盘123  $b \mapsto c$

```
1 public class E02HanoiTower {
2
3
4     /*
5         源 借 目
6         h(4, a, b, c) -> h(3, a, c, b)
7             a -> c
8             h(3, b, a, c)
9     */
10    static LinkedList<Integer> a = new LinkedList<>();
11    static LinkedList<Integer> b = new LinkedList<>();
12    static LinkedList<Integer> c = new LinkedList<>();
13
14    static void init(int n) {
15        for (int i = n; i >= 1; i--) {
16            a.add(i);
17        }
18    }
19
20    static void h(int n, LinkedList<Integer> a,
21        LinkedList<Integer> b,
22        LinkedList<Integer> c) {
23        if (n == 0) {
24            return;
25        }
26        h(n - 1, a, c, b);
27        c.addLast(a.removeLast());
28        print();
29        h(n - 1, b, a, c);
30    }
31
32    private static void print() {
33        System.out.println("-----");
34        System.out.println(a);
35        System.out.println(b);
36        System.out.println(c);
37    }
38
39    public static void main(String[] args) {
40        init(3);
41        print();
42        h(3, a, b, c);
43    }
44 }
```

## E03. 杨辉三角 <sup>14</sup>

 image-20221219172514410

### 分析

把它斜着看

```
1      1
2     1  1
3    1  2  1
4   1  3  3  1
5  1  4  6  4  1
```

- 行  $i$ , 列  $j$ , 那么  $[i][j]$  的取值应为  $[i-1][j-1] + [i-1][j]$
- 当  $j = 0$  或  $i = j$  时,  $[i][j]$  取值为 1

### 题解

```
1 public static void print(int n) {
2     for (int i = 0; i < n; i++) {
3         if (i < n - 1) {
4             System.out.printf("%" + 2 * (n - 1 - i) + "s", " ");
5         }
6
7         for (int j = 0; j < i + 1; j++) {
8             System.out.printf("%-4d", element(i, j));
9         }
10        System.out.println();
11    }
12 }
13
14 public static int element(int i, int j) {
15     if (j == 0 || i == j) {
16         return 1;
17     }
18     return element(i - 1, j - 1) + element(i - 1, j);
19 }
```

### 优化1

是 multiple recursion, 因此很多递归调用是重复的, 例如

- recursion(3, 1) 分解为
  - recursion(2, 0) + recursion(2, 1)
- 而 recursion(3, 2) 分解为
  - recursion(2, 1) + recursion(2, 2)

这里 recursion(2, 1) 就重复调用了, 事实上它会重复很多次, 可以用 `static AtomicInteger counter = new AtomicInteger(0)` 来查看递归函数的调用总次数

事实上, 可以用 **memoization** 来进行优化:



```

1 public static void print1(int n) {
2     int[][] triangle = new int[n][];
3     for (int i = 0; i < n; i++) {
4         // 打印空格
5         triangle[i] = new int[i + 1];
6         for (int j = 0; j <= i; j++) {
7             System.out.printf("%-4d", element1(triangle, i, j));
8         }
9         System.out.println();
10    }
11 }
12
13 public static int element1(int[][] triangle, int i, int j) {
14     if (triangle[i][j] > 0) {
15         return triangle[i][j];
16     }
17
18     if (j == 0 || i == j) {
19         triangle[i][j] = 1;
20         return triangle[i][j];
21     }
22     triangle[i][j] = element1(triangle, i - 1, j - 1) + element1(triangle, i
- 1, j);
23     return triangle[i][j];
24 }

```

- 将数组作为递归函数内可以访问的遍历，如果  $triangle[i][j]$  已经有值，说明该元素已经被之前的递归函数计算过，就不必重复计算了

## 优化2

```

1 public static void print2(int n) {
2     int[] row = new int[n];
3     for (int i = 0; i < n; i++) {
4         // 打印空格
5         createRow(row, i);
6         for (int j = 0; j <= i; j++) {
7             System.out.printf("%-4d", row[j]);
8         }
9         System.out.println();
10    }
11 }
12
13 private static void createRow(int[] row, int i) {
14     if (i == 0) {
15         row[0] = 1;
16         return;
17     }
18     for (int j = i; j > 0; j--) {
19         row[j] = row[j - 1] + row[j];
20     }
21 }

```

注意：还可以通过每一行的前一项计算出下一项，不必借助上一行，这与杨辉三角的另一个特性有关，暂不展开了

力扣对应题目，但递归不适合在力扣刷高分，因此只列出相关题目，不做刷题讲解了

- [118. 杨辉三角 - 力扣 \(LeetCode\)](#)
- [119. 杨辉三角 II - 力扣 \(LeetCode\)](#)

## 3.5 链表

### E01. 反转单向链表-力扣 206 题

对应力扣题目 [206. 反转链表 - 力扣 \(LeetCode\)](#)

```
1  输入: head = [1,2,3,4,5]
2  输出: [5,4,3,2,1]
3
4  输入: [1,2]
5  输出: [2,1]
6
7  输入: []
8  输出: []
```

#### 方法1

构造一个新链表，从**旧链表**依次拿到每个节点，创建新节点添加至**新链表头部**，完成后新链表即是倒序的

```
1  public ListNode reverseList(ListNode o1) {
2      ListNode n1 = null;
3      ListNode p = o1;
4      while (p != null) {
5          n1 = new ListNode(p.val, n1);
6          p = p.next;
7      }
8      return n1;
9  }
```

评价：简单直白，就是得新建节点对象

#### 方法2

与方法1 类似，构造一个新链表，从**旧链表头部**移除节点，添加到**新链表头部**，完成后新链表即是倒序的，区别在于原题目未提供节点外层的容器类，这里提供一个，另外一个区别是并不去构造新节点

```
1  static class List {
2      ListNode head;
3  }
```

```

4     public List(ListNode head) {
5         this.head = head;
6     }
7
8     public ListNode removeFirst(){
9         ListNode first = head;
10        if (first != null) {
11            head = first.next;
12        }
13        return first;
14    }
15
16    public void addFirst(ListNode first) {
17        first.next = head;
18        head = first;
19    }
20 }

```

代码

```

1 public ListNode reverseList(ListNode head) {
2     List list1 = new List(head);
3     List list2 = new List(null);
4     ListNode first;
5     while ((first = list1.removeFirst()) != null) {
6         list2.addFirst(first);
7     }
8     return list2.head;
9 }

```

评价：更加面向对象，如果实际写代码而非刷题，更多会这么做

### 方法3

递归，在归时让  $5 \rightarrow 4$ ,  $4 \rightarrow 3 \dots$

首先，写一个递归方法，返回值用来拿到最后一个节点

```

1 public ListNode reverseList(ListNode p) {
2     if (p == null || p.next == null) { // 不足两个节点
3         return p; // 最后一个节点
4     }
5     ListNode last = reverseList(p.next);
6     return last;
7 }

```

- 注意1：递归终止条件是 `curr.next == null`，目的是到最后一个节点就结束递归，与之前递归遍历不一样
- 注意2：需要考虑空链表即 `p == null` 的情况

可以先测试一下

```

1 | ListNode o5 = new ListNode(5, null);
2 | ListNode o4 = new ListNode(4, o5);
3 | ListNode o3 = new ListNode(3, o4);
4 | ListNode o2 = new ListNode(2, o3);
5 | ListNode o1 = new ListNode(1, o2);
6 | ListNode n1 = new E01Leetcode206().reverseList(o1);
7 | System.out.println(n1);

```

会打印

```
1 | [5]
```

下面为伪码调用过程，假设节点分别是  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$ ，先忽略返回值

```

1 | reverseList(ListNode p = 1) {
2 |     reverseList(ListNode p = 2) {
3 |         reverseList(ListNode p = 3) {
4 |             reverseList(ListNode p = 4) {
5 |                 reverseList(ListNode p = 5) {
6 |                     if (p == null || p.next == null) {
7 |                         return p; // 返回5
8 |                     }
9 |                 }
10 |                // 此时p是4, p.next是5
11 |            }
12 |            // 此时p是3, p.next是4
13 |        }
14 |        // 此时p是2, p.next是3
15 |    }
16 |    // 此时p是1, p.next是2
17 | }

```

接下来，从  $p=4$  开始，要让  $5 \rightarrow 4$ ,  $4 \rightarrow 3 \dots$

```

1 | reverseList(ListNode p = 1) {
2 |     reverseList(ListNode p = 2) {
3 |         reverseList(ListNode p = 3) {
4 |             reverseList(ListNode p = 4) {
5 |                 reverseList(ListNode p = 5) {
6 |                     if (p == null || p.next == null) {
7 |                         return p; // 返回5
8 |                     }
9 |                 }
10 |                // 此时p是4, p.next是5, 要让5指向4, 代码写成 p.next.next=p
11 |                // 还要注意4要指向 null, 否则就死链了
12 |            }
13 |            // 此时p是3, p.next是4
14 |        }
15 |        // 此时p是2, p.next是3
16 |    }
17 |    // 此时p是1, p.next是2
18 | }

```

最终代码为：

```
1 public ListNode reverseList(ListNode p) {
2     if (p == null || p.next == null) { // 不足两个节点
3         return p; // 最后一个节点
4     }
5     ListNode last = reverseList(p.next);
6     p.next.next = p;
7     p.next = null;
8     return last;
9 }
```

Q：为啥不能在递的过程中倒序？

A：比如

- $1 \rightarrow 2 \rightarrow 3$  如果递的过程中让  $2 \rightarrow 1$  那么此时  $2 \rightarrow 3$  就被覆盖，不知道接下来递给谁
- 而归的时候让  $3 \rightarrow 2$  不会影响上一层的  $1 \rightarrow 2$

评价：单向链表没有 prev 指针，但利用递归的特性【记住了】链表每次调用时相邻两个节点是谁

#### 方法4

从链表每次拿到第二个节点，将其从链表断开，插入头部，直至它为 null 结束

1. 设置指针 o1(旧头)、n1(新头)、o2(旧老二)，分别指向第一，第一，第二节点

$\frac{n1}{1} \frac{o1}{1} \rightarrow \frac{o2}{2} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

2. 将 o2 节点从链表断开，即 o1 节点指向第三节点

$\frac{n1}{1} \frac{o1}{1} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null, \frac{o2}{2}$

3. o2 节点链入链表头部，即

$\frac{o2}{2} \rightarrow \frac{n1}{1} \frac{o1}{1} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

4. n1 指向 o2

$\frac{n1}{2} \frac{o2}{2} \rightarrow \frac{o1}{1} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

5. o2 指向 o1 的下一个节点，即

$\frac{n1}{2} \rightarrow \frac{o1}{1} \rightarrow \frac{o2}{3} \rightarrow 4 \rightarrow 5 \rightarrow null$

6. 重复以上 2 ~ 5 步，直到 o2 指向 null

7. 还应当考虑边界条件，即链表中不满两个元素时，无需走以上逻辑

参考答案

```
1 public ListNode reverseList(ListNode o1) {
2     if (o1 == null || o1.next == null) { // 不足两个节点
3         return o1;
4     }
5     ListNode o2 = o1.next;
6     ListNode n1 = o1;
7     while (o2 != null) {
```

```

8         o1.next = o2.next;
9         o2.next = n1;
10        n1 = o2;
11        o2 = o1.next;
12    }
13    return n1;
14 }

```

## 方法5

要点：把链表分成两部分，思路就是不断从链表2的头，往链表1的头搬移

1. n1 指向 null，代表**新链表**一开始没有元素，o1 指向**原链表**的首节点

$\frac{n1}{null}, \frac{o1}{1} \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

2. 开始循环，o2 指向**原链表**次节点

$\frac{n1}{null}, \frac{o1}{1} \rightarrow \frac{o2}{2} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

3. 搬移

$\frac{o1}{1} \rightarrow \frac{n1}{null}, \frac{o2}{2} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

4. 指针复位

$\frac{n1}{1} \rightarrow null, \frac{o1\ o2}{2} \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow null$

5. 重复 2 ~ 4 步

6. 当 o1 = null 时退出循环

## 参考答案

```

1 public ListNode reverseList(ListNode o1) {
2     if (o1 == null || o1.next == null) {
3         return o1;
4     }
5     ListNode n1 = null;
6     while (o1 != null) {
7         ListNode o2 = o1.next;
8         o1.next = n1;
9         n1 = o1;
10        o1 = o2;
11    }
12    return n1;
13 }

```

评价：本质上与方法2 相同，只是方法2更为面向对象

## E02. 根据值删除节点-力扣 203 题

例如

```
1  输入: head = [1,2,6,3,6], val = 6
2  输出: [1,2,3]
3
4  输入: head = [], val = 1
5  输出: []
6
7  输入: head = [7,7,7,7], val = 7
8  输出: []
```

### 方法1

图中 s 代表 sentinel 哨兵 (如果不加哨兵, 则删除第一个节点要特殊处理), 例如要删除 6

```
1  p1   p2
2  s -> 1 -> 2 -> 6 -> 3 -> 6 -> null
```

- 如果 p2 不等于目标, 则 p1, p2 不断后移

```
1      p1   p2
2  s -> 1 -> 2 -> 6 -> 3 -> 6 -> null
3
4      p1   p2
5  s -> 1 -> 2 -> 6 -> 3 -> 6 -> null
```

- p2 == 6, 删除它, 注意 p1 此时保持不变, p2 后移

```
1      p1   p2
2  s -> 1 -> 2 -> 3 -> 6 -> null
```

- p2 不等于目标, 则 p1, p2 不断后移

```
1      p1   p2
2  s -> 1 -> 2 -> 3 -> 6 -> null
```

- p2 == 6, 删除它, 注意 p1 此时保持不变, p2 后移

```
1      p1   p2
2  s -> 1 -> 2 -> 3 -> null
```

- p2 == null 退出循环

最后代码

```

1 public ListNode removeElements(ListNode head, int val) {
2     ListNode sentinel = new ListNode(-1, head);
3     ListNode p1 = sentinel;
4     ListNode p2;
5     while ((p2 = p1.next) != null) {
6         if (p2.val == val) {
7             p1.next = p2.next;
8         } else {
9             p1 = p1.next;
10        }
11    }
12    return sentinel.next;
13 }

```

## 方法2

思路，递归函数负责返回：从当前节点（我）开始，完成删除的子链表

1. 若我与 v 相等，应该返回下一个节点递归结果
2. 若我与 v 不等，应该返回我，但我的 next 应该更新（让我能带上后续删过的子链表）

```

1 removeElements(ListNode p=1, int v=6){
2     1.next=removeElements(ListNode p=2, int v=6){
3         2.next=removeElements(ListNode p=6, int v=6){
4             removeElements(ListNode p=3, int v=6){
5                 3.next=removeElements(ListNode p=6, int v=6){
6                     removeElements(ListNode p=null, int v=6){
7                         // 没有节点,返回
8                         return null
9                     }
10                }
11                return 3
12            }
13            }
14            return 2
15        }
16        return 1
17    }

```

## 代码

```

1 public ListNode removeElements(ListNode head, int val) {
2     if (head == null) {
3         return null;
4     }
5     if (head.val == val) {
6         return removeElements(head.next, val);
7     } else {
8         head.next = removeElements(head.next, val);
9         return head;
10    }
11 }

```



## E03. 删除倒数节点-力扣 19 题

例如

```
1  输入: head = [1,2,3,4,5], n = 2
2  输出: [1,2,3,5]
3
4  输入: head = [1], n = 1
5  输出: []
6
7  输入: head = [1,2], n = 1
8  输出: [1]
```

另外题目提示

- 链表至少一个节点
- n 只在合理范围

### 方法1

思路，写一个递归函数，用来返回下一个节点的倒数序号

```
1  recursion(ListNode p=1, int n=2) {
2      recursion(ListNode p=2, int n=2) {
3          recursion(ListNode p=3, int n=2) {
4              recursion(ListNode p=4, int n=2) {
5                  recursion(ListNode p=5, int n=2) {
6                      recursion(ListNode p=null, int n=2) {
7                          return 0; // 最内层序号0
8                      }
9                      return 1; // 上一次返回值+1
10                 }
11                 return 2;
12             }
13             if(返回值 == n == 2) {
14                 // 删除 next
15             }
16             return 3;
17         }
18         return 4;
19     }
20     return 5;
21 }
```

但上述代码有一个问题，就是若删除的是第一个节点，它没有上一个节点，因此可以加一个哨兵来解决

代码

```
1  public ListNode removeNthFromEnd(ListNode head, int n) {
2      ListNode sentinel = new ListNode(-1, head);
3      recursion(sentinel, n);
```

```

4     return sentinel.next;
5 }
6
7 public int recursion(ListNode p, int n) {
8     if (p == null) {
9         return 0;
10    }
11    int nth = recursion(p.next, n);
12    if (nth == n) {
13        p.next = p.next.next;
14    }
15    return nth + 1;
16 }

```

Q: p.next.next 不怕空指针吗?

A:

- p 是待删除节点的上一个节点，如果能递归回到 p，那么 p.next 肯定有值，不会是 null
- 且题目说明了  $n \geq 1$ ，不会因为  $\text{nth} == 0$  而让 p.next 指向最后的 null

## 方法2

快慢指针，p1 指向待删节点的上一个，p2 先走  $n + 1$  步

```

1  i=0
2  p2
3  s -> 1 -> 2 -> 3 -> 4 -> 5 -> null
4
5      i=1
6      p2
7  s -> 1 -> 2 -> 3 -> 4 -> 5 -> null
8
9          i=2
10         p2
11  s -> 1 -> 2 -> 3 -> 4 -> 5 -> null
12
13          i=3 从此开始 p1 p2 依次向右平移，直到 p2 移动到末尾
14  p1          p2
15  s -> 1 -> 2 -> 3 -> 4 -> 5 -> null
16
17          p1          p2
18  s -> 1 -> 2 -> 3 -> 4 -> 5 -> null

```

代码

```

1  public ListNode removeNthFromEnd(ListNode head, int n) {
2      ListNode s = new ListNode(-1, head);
3      ListNode p1 = s;
4      ListNode p2 = s;
5      for (int i = 0; i < n + 1; i++) {
6          p2 = p2.next;
7      }
8      while (p2 != null) {

```

```

9         p1 = p1.next;
10        p2 = p2.next;
11    }
12    p1.next = p1.next.next;
13    return s.next;
14 }

```

### 方法3

```

1  public ListNode removeNthFromEnd(ListNode head, int n) {
2      Composite c = recursion(head, n);
3      return c.node;
4  }
5
6  static class Composite {
7      ListNode node;
8      int nth;
9
10     public Composite(ListNode node, int nth) {
11         this.node = node;
12         this.nth = nth;
13     }
14 }
15
16 public Composite recursion(ListNode p, int n) {
17     if (p == null) {
18         return new Composite(null, 1);
19     }
20     Composite c = recursion(p.next, n);
21     if (c.nth != n) {
22         p.next = c.node;
23         c.node = p;
24     }
25     c.nth += 1;
26     return c;
27 }

```

## E04. 有序链表去重-力扣 83 题

例如

```

1  输入: head = [1,1,2]
2  输出: [1,2]
3
4  输入: head = [1,1,2,3,3]
5  输出: [1,2,3]

```

注意: 重复元素保留一个

### 方法1

```

1 | p1    p2
2 | 1 -> 1 -> 2 -> 3 -> 3 -> null

```

- $p1.val == p2.val$  那么删除  $p2$ , 注意  $p1$  此时保持不变

```

1 | p1    p2
2 | 1 -> 2 -> 3 -> 3 -> null

```

- $p1.val != p2.val$  那么  $p1, p2$  向后移动

```

1 |          p1    p2
2 | 1 -> 2 -> 3 -> 3 -> null
3 |
4 |          p1    p2
5 | 1 -> 2 -> 3 -> 3 -> null

```

- $p1.val == p2.val$  那么删除  $p2$

```

1 |          p1    p2
2 | 1 -> 2 -> 3 -> null

```

- 当  $p2 == null$  退出循环

代码

```

1 | public ListNode deleteDuplicates(ListNode head) {
2 |     // 链表节点 < 2
3 |     if (head == null || head.next == null) {
4 |         return head;
5 |     }
6 |     // 链表节点 >= 2
7 |     ListNode p1 = head;
8 |     ListNode p2;
9 |     while ((p2 = p1.next) != null) {
10 |         if (p1.val == p2.val) {
11 |             p1.next = p2.next;
12 |         } else {
13 |             p1 = p1.next;
14 |         }
15 |     }
16 |     return head;
17 | }

```

## 方法2

递归函数负责返回：从当前节点（我）开始，完成去重的链表

1. 若我与 next 重复，返回 next
2. 若我与 next 不重复，返回我，但 next 应当更新

```

1 | deleteDuplicates(ListNode p=1) {

```

```

2   deleteDuplicates(ListNode p=1) {
3       1.next=deleteDuplicates(ListNode p=2) {
4           2.next=deleteDuplicates(ListNode p=3) {
5               deleteDuplicates(ListNode p=3) {
6                   // 只剩一个节点，返回
7                   return 3
8               }
9           }
10          return 2
11      }
12      return 1
13  }
14  }

```

代码

```

1   public ListNode deleteDuplicates(ListNode p) {
2       if (p == null || p.next == null) {
3           return p;
4       }
5       if(p.val == p.next.val) {
6           return deleteDuplicates(p.next);
7       } else {
8           p.next = deleteDuplicates(p.next);
9           return p;
10      }
11  }

```

## E05. 有序链表去重-力扣 82 题

例如

```

1   输入: head = [1,2,3,3,4,4,5]
2   输出: [1,2,5]
3
4   输入: head = [1,1,1,2,3]
5   输出: [2,3]

```

注意: **重复元素一个不留**

### 方法1

递归函数负责返回: 从当前节点 (我) 开始, 完成去重的链表

1. 若我与 next 重复, 一直找到下一个不重复的节点, 以它的返回结果为准
2. 若我与 next 不重复, 返回我, 同时更新 next

```

1   deleteDuplicates(ListNode p = 1) {
2       // 找下个不重复的
3       deleteDuplicates(ListNode p = 1) {
4           deleteDuplicates(ListNode p = 1) {
5               deleteDuplicates(ListNode p = 2) {
6                   2.next=deleteDuplicates(ListNode p = 3) {

```

```

7          // 只剩一个节点，返回
8          return 3;
9      }
10     return 2;
11 }
12 }
13 }
14 }

```

代码

```

1 public ListNode deleteDuplicates(ListNode p) {
2     if (p == null || p.next == null) {
3         return p;
4     }
5     if (p.val == p.next.val) {
6         ListNode x = p.next.next;
7         while (x != null && x.val == p.val) {
8             x = x.next;
9         }
10        return deleteDuplicates(x);
11    } else {
12        p.next = deleteDuplicates(p.next);
13        return p;
14    }
15 }

```

## 方法2

p1 是待删除的上一个节点，每次循环对比 p2、p3 的值

- 如果 p2 与 p3 的值重复，那么 p3 继续后移，直到找到与 p2 不重复的节点，p1 指向 p3 完成删除
- 如果 p2 与 p3 的值不重复，p1, p2, p3 向后平移一位，继续上面的操作
- p2 或 p3 为 null 退出循环
  - p2 为 null 的情况，比如链表为 1 1 1 null

```

1 p1 p2 p3
2 s, 1, 1, 1, 2, 3, null
3
4 p1 p2 p3
5 s, 1, 1, 1, 2, 3, null
6
7 p1 p2 p3
8 s, 1, 1, 1, 2, 3, null
9
10 p1 p3
11 s, 2, 3, null
12
13 p1 p2 p3
14 s, 2, 3, null
15
16 p1 p2 p3

```

```
17 | s, 2, 3, null
```

代码

```
1 public ListNode deleteDuplicates(ListNode head) {
2     if (head == null || head.next == null) {
3         return head;
4     }
5
6     ListNode s = new ListNode(-1, head);
7     ListNode p1 = s;
8     ListNode p2;
9     ListNode p3;
10    while ((p2 = p1.next) != null && (p3 = p2.next) != null) {
11        if (p2.val == p3.val) {
12            while ((p3 = p3.next) != null
13                && p3.val == p2.val) {
14            }
15            p1.next = p3;
16        } else {
17            p1 = p1.next;
18        }
19    }
20    return s.next;
21 }
```

## E06. 合并有序链表-力扣 21 题

例

```
1 输入: l1 = [1,2,4], l2 = [1,3,4]
2 输出: [1,1,2,3,4,4]
3
4 输入: l1 = [], l2 = []
5 输出: []
6
7 输入: l1 = [], l2 = [0]
8 输出: [0]
```

### 方法1

- 谁小，把谁链给 p，p 和小的都向后平移一位
- 当 p1、p2 有一个为 null，退出循环，把不为 null 的链给 p

```

1 | p1
2 | 1  3  8  9  null
3 |
4 | p2
5 | 2  4  null
6 |
7 | p
8 | s  null

```

代码

```

1 | public ListNode mergeTwoLists(ListNode p1, ListNode p2) {
2 |     ListNode s = new ListNode(-1, null);
3 |     ListNode p = s;
4 |     while (p1 != null && p2 != null) {
5 |         if (p1.val < p2.val) {
6 |             p.next = p1;
7 |             p1 = p1.next;
8 |         } else {
9 |             p.next = p2;
10 |            p2 = p2.next;
11 |        }
12 |        p = p.next;
13 |    }
14 |    if (p1 != null) {
15 |        p.next = p1;
16 |    }
17 |    if (p2 != null) {
18 |        p.next = p2;
19 |    }
20 |    return s.next;
21 | }

```

- 可以自行验证例中后两种情况

## 方法2

递归函数应该返回

- 更小的那个链表节点，并把它剩余节点与另一个链表再次递归
- 返回之前，更新此节点的 next

```

1 | mergeTwoLists(p1=[1,3,8,9], p2=[2,4]) {
2 |     1.next=mergeTwoLists(p1=[3,8,9], p2=[2,4]) {
3 |         2.next=mergeTwoLists(p1=[3,8,9], p2=[4]) {
4 |             3.next=mergeTwoLists(p1=[8,9], p2=[4]) {
5 |                 4.next=mergeTwoLists(p1=[8,9], p2=null) {
6 |                     return [8,9]
7 |                 }
8 |                 return 4
9 |             }
10 |            return 3
11 |        }

```



```

12         return 2
13     }
14     return 1
15 }

```

## E07. 合并多个有序链表-力扣 23 题

例

```

1  输入: lists = [[1,4,5],[1,3,4],[2,6]]
2  输出: [1,1,2,3,4,4,5,6]
3  解释: 链表数组如下:
4  [
5      1->4->5,
6      1->3->4,
7      2->6
8  ]
9  将它们合并到一个有序链表中得到。
10 1->1->2->3->4->4->5->6

```

### 方法1

递归

```

1  public ListNode mergeKLists(ListNode[] lists) {
2      if (lists.length == 0) {
3          return null;
4      }
5      return merge(lists, 0, lists.length - 1);
6  }
7
8  public ListNode split(ListNode[] lists, int i, int j) {
9      System.out.println(i + " " + j);
10     if (j == i) {
11         return lists[i];
12     }
13     int m = (i + j) >>> 1;
14     return mergeTwoLists(
15         split(lists, i, m),
16         split(lists, m + 1, j)
17     );
18 }

```

还可以用优先级队列求解，这个放在后面讲

## E08. 查找链表中间节点-力扣 876 题

例如

```
1  输入: [1,2,3,4,5]
2  输出: 此列表中的中间节点 3 (序列化形式: [3,4,5])
3
4  输入: [1,2,3,4,5,6]
5  输出: 此列表中的中间节点 4 (序列化形式: [4,5,6])
```

- 偶数节点时，中间点是靠右的那个

解法：快慢指针，快指针一次走两步，慢指针一次走一步，当快指针到链表结尾时，慢指针恰好走到链表的一半

```
1  public ListNode middleNode(ListNode head) {
2      ListNode p1 = head; // 慢指针，中间点
3      ListNode p2 = head; // 快指针
4      while (p2 != null && p2.next != null) {
5          p1 = p1.next;
6          p2 = p2.next;
7          p2 = p2.next;
8      }
9      return p1;
10 }
```

## E09. 回文链表-力扣 234 题

所谓回文指正着读、反着读，结果一样，例如

```
1  [1,2,2,1]
2  [1,2,3,2,1]
```

它们都是回文链表，不是回文的例子

```
1  [1,2,3,1]  --反过来-->  [1,3,2,1]
```

解法

```
1  /*
2      步骤1. 找中间点
3      步骤2. 中间点后半个链表反转
4      步骤3. 反转后链表与原链表逐一比较
5  */
6  public boolean isPalindrome(ListNode head) {
7      ListNode middle = middle(head);
8      ListNode newHead = reverse(middle);
9      while (newHead != null) {
10         if (newHead.val != head.val) {
```

```

11         return false;
12     }
13     newHead = newHead.next;
14     head = head.next;
15 }
16 return true;
17 }
18
19 private ListNode reverse(ListNode o1) {
20     ListNode n1 = null;
21     while (o1 != null) {
22         ListNode o2 = o1.next;
23         o1.next = n1;
24         n1 = o1;
25         o1 = o2;
26     }
27     return n1;
28 }
29
30 private ListNode middle(ListNode head) {
31     ListNode p1 = head; // 慢
32     ListNode p2 = head; // 快
33     while (p2 != null && p2.next != null) {
34         p1 = p1.next;
35         p2 = p2.next.next;
36     }
37     return p1;
38 }

```

## 优化后解法

```

1 public boolean isPalindrome(ListNode h1) {
2     if (h1 == null || h1.next == null) {
3         return true;
4     }
5     ListNode p1 = h1; // 慢指针，中间点
6     ListNode p2 = h1; // 快指针
7     ListNode n1 = null; // 新头
8     ListNode o1 = h1; // 旧头
9     // 快慢指针找中间点
10    while (p2 != null && p2.next != null) {
11        p1 = p1.next;
12        p2 = p2.next.next;
13
14        // 反转前半部分
15        o1.next = n1;
16        n1 = o1;
17        o1 = p1;
18    }
19    if (p2 != null) { // 节点数为奇数
20        p1 = p1.next;
21    }
22    // 同步比较新头和后半部分

```

```

23     while (n1 != null) {
24         if (n1.val != p1.val) {
25             return false;
26         }
27         p1 = p1.next;
28         n1 = n1.next;
29     }
30     return true;
31 }

```

## E10. 环形链表-力扣 141 题

本题以及下题，实际是 Floyd's Tortoise and Hare Algorithm （Floyd 龟兔赛跑算法） 15

除了 Floyd 判环算法外，还有其它的判环算法，详见 [https://en.wikipedia.org/wiki/Cycle\\_detection](https://en.wikipedia.org/wiki/Cycle_detection)

如果链表上存在环，那么在环上以不同速度前进的两个指针必定会在某个时刻相遇。算法分为两个阶段

阶段1

- 龟一次走一步，兔子一次走两步
- 当兔子能走到终点时，不存在环
- 当兔子能追上龟时，可以判断存在环

阶段2

- 从它们第一次相遇开始，龟回到起点，兔子保持原位不变
- 龟和兔子一次都走一步
- 当再次相遇时，地点就是环的入口

为什么呢？

- 设起点到入口走  $a$  步（本例是 7），绕环一圈长度为  $b$ （本例是 5），
- 那么**从起点开始，走  $a + \text{绕环 } n \text{ 圈}$ ，都能找到环入口**
- 第一次相遇时
  - 兔走了  $a + \text{绕环 } n \text{ 圈}$ （本例 2 圈）+  $k$ ， $k$  是它们相遇距环入口位置（本例 3，不重要）
  - 龟走了  $a + \text{绕环 } n \text{ 圈}$ （本例 0 圈）+  $k$ ，当然它绕的圈数比兔少
  - 兔走的距离是龟的两倍，所以**龟走的 = 兔走的 - 龟走的 = 绕环  $n$  圈**
- 而前面分析过，如果走  $a + \text{绕环 } n \text{ 圈}$ ，都能找到环入口，因此从相遇点开始，再走  $a$  步，就是环入口

阶段1 参考代码（判断是否有环）

```

1 public boolean hasCycle(ListNode head) {
2     ListNode h = head; // 兔
3     ListNode t = head; // 龟
4     while (h != null && h.next != null) {
5         t = t.next;
6         h = h.next.next;
7         if(h == t){
8             return true;
9         }
10    }
11    return false;
12 }

```

## E11. 环形链表-力扣 142 题

阶段2 参考代码（找到环入口）

```

1 public ListNode detectCycle(ListNode head) {
2     ListNode t = head; // 龟
3     ListNode h = head; // 兔
4     while (h != null && h.next != null) {
5         t = t.next;
6         h = h.next.next;
7         if (h == t) {
8             t = head;
9             while (true) {
10                if (h == t) {
11                    return h;
12                }
13                h = h.next;
14                t = t.next;
15            }
16        }
17    }
18    return null;
19 }

```

- 还有一道扩展题目，也可以用判环算法思想来解：就是 287 题，寻找重复数

## Ex1. 删除节点-力扣 237 题

这道题目比较简单，留给大家自己练习

例如

```
1 输入: head = [4,5,1,9], node = 5
2 输出: [4,1,9]
3
4
5 输入: head = [4,5,1,9], node = 1
6 输出: [4,5,9]
```

注意: 被删除的节点**不是**末尾节点

### 参考答案

```
1 public class Ex1Leetcode237 {
2     /**
3      *
4      * @param node 待删除节点, 题目已说明肯定不是最后一个节点
5      */
6     public void deleteNode(ListNode node) {
7         node.val = node.next.val;    // 下一个节点值赋值给待"删除"节点
8         node.next = node.next.next;  // 把下一个节点删除
9     }
10
11     public static void main(String[] args) {
12         ListNode o5 = new ListNode(5, null);
13         ListNode o4 = new ListNode(4, o5);
14         ListNode o3 = new ListNode(3, o4);
15         ListNode o2 = new ListNode(2, o3);
16         ListNode o1 = new ListNode(1, o2);
17         System.out.println(o1);
18         new Ex1Leetcode237().deleteNode(o3);
19         System.out.println(o1);
20     }
21 }
```

输出

```
1 [1,2,3,4,5]
2 [1,2,4,5]
```

## Ex2. 共尾链表-力扣 160 题

原题叫做**相交**链表, 个人觉得用**共尾**链表更形象些, 此题更像是一道脑筋急转弯, 留给大家练习

例如, 下图的两个链表 [1, 2, 4, 5] 与 [3, 4, 5] 它们中 [4, 5] 是相同的, 此时应返回节点 4



非共尾的情况, 如下图所示, 此时返回 null



思路, 称两个链表为 a=[1, 2, 4, 5], b=[3, 4, 5], 图中用 N 代表 null

1. 遍历 a, 遇到 null 时改道遍历 b
2. 与此同时, 遍历 b, 遇到 null 时改道遍历 a

- 在此过程中，如果遇到相同的节点，即为找寻目标，返回即可，如下图中的第二次出现的 4
- 相同节点应该比较其引用值，图中数字只是为了便于区分

1	1	2	4	5	N	3	4	5	N
2	3	4	5	N	1	2	4	5	N

如果两个链表长度相同，则可以更早找到目标，例如 a=[1, 4, 5], b=[3, 4, 5]，第一次出现 4 时，即可返回

1	1	4	5	N	3	4	5	N
2	3	4	5	N	1	4	5	N

如果是非共尾的情况，如 a=[1, 2, 4], b=[3, 5]，可以看到，唯一相等的情况，是遍历到最后那个 N 此时退出循环

1	1	2	4	N	3	5	N
2	3	5	N	1	2	4	N

代码

```
1 public ListNode getIntersectionNode(ListNode a, ListNode b) {
2     ListNode p1 = a;
3     ListNode p2 = b;
4     while (true) {
5         if (p1 == p2) {
6             return p1;
7         }
8         if (p1 == null) {
9             p1 = b;
10        } else {
11            p1 = p1.next;
12        }
13        if (p2 == null) {
14            p2 = a;
15        } else {
16            p2 = p2.next;
17        }
18    }
19 }
```

## 3.6 数组

### E01. 合并有序数组

将数组内两个区间内的有序元素合并

例

1	[1, 5, 6, 2, 4, 10, 11]
---	-------------------------

可以视作两个有序区间

```
1 | [1, 5, 6] 和 [2, 4, 10, 11]
```

合并后，结果仍存储于原有空间

```
1 | [1, 2, 4, 5, 6, 10, 11]
```

## 方法1

递归

- 每次递归把更小的元素复制到结果数组

```
1 merge(left=[1,5,6],right=[2,4,10,11],a2=[]){
2     merge(left=[5,6],right=[2,4,10,11],a2=[1]){
3         merge(left=[5,6],right=[4,10,11],a2=[1,2]){
4             merge(left=[5,6],right=[10,11],a2=[1,2,4]){
5                 merge(left=[6],right=[10,11],a2=[1,2,4,5]){
6                     merge(left=[],right=[10,11],a2=[1,2,4,5,6]){
7                         // 拷贝10, 11
8                     }
9                 }
10            }
11        }
12    }
13 }
```

代码

```
1 public static void merge(int[] a1, int i, int iEnd, int j, int jEnd,
2                           int[] a2, int k) {
3     if (i > iEnd) {
4         System.arraycopy(a1, j, a2, k, jEnd - j + 1);
5         return;
6     }
7     if (j > jEnd) {
8         System.arraycopy(a1, i, a2, k, iEnd - i + 1);
9         return;
10    }
11    if (a1[i] < a1[j]) {
12        a2[k] = a1[i];
13        merge(a1, i + 1, iEnd, j, jEnd, a2, k + 1);
14    } else {
15        a2[k] = a1[j];
16        merge(a1, i, iEnd, j + 1, jEnd, a2, k + 1);
17    }
18 }
```

测试



```

1  int[] a1 = {1, 5, 6, 2, 4, 10, 11};
2  int[] a2 = new int[a1.length];
3  merge(a1, 0, 2, 3, 6, a2, 0);

```

## 方法2

代码

```

1  public static void merge(int[] a1, int i, int iEnd,
2                          int j, int jEnd,
3                          int[] a2) {
4      int k = i;
5      while (i <= iEnd && j <= jEnd) {
6          if (a1[i] < a1[j]) {
7              a2[k] = a1[i];
8              i++;
9          } else {
10             a2[k] = a1[j];
11             j++;
12         }
13         k++;
14     }
15     if (i > iEnd) {
16         System.arraycopy(a1, j, a2, k, jEnd - j + 1);
17     }
18     if (j > jEnd) {
19         System.arraycopy(a1, i, a2, k, iEnd - i + 1);
20     }
21 }

```

测试

```

1  int[] a1 = {1, 5, 6, 2, 4, 10, 11};
2  int[] a2 = new int[a1.length];
3  merge(a1, 0, 2, 3, 6, a2);

```

## 3.7 队列

### E01. 二叉树层序遍历-力扣 102 题

```

1  class Solution {
2      public List<List<Integer>> levelOrder(TreeNode root) {
3          List<List<Integer>> result = new ArrayList<>();
4          if (root == null) {
5              return result;
6          }
7          LinkedListQueue<TreeNode> queue = new LinkedListQueue<>();
8          queue.offer(root);
9          int c1 = 1;      // 本层节点个数
10         while (!queue.isEmpty()) {

```

```

11         int c2 = 0;        // 下层节点个数
12         List<Integer> level = new ArrayList<>();
13         for (int i = 0; i < c1; i++) {
14             TreeNode node = queue.poll();
15             level.add(node.val);
16             if (node.left != null) {
17                 queue.offer(node.left);
18                 c2++;
19             }
20             if (node.right != null) {
21                 queue.offer(node.right);
22                 c2++;
23             }
24         }
25         c1 = c2;
26         result.add(level);
27     }
28     return result;
29 }
30
31 // 自定义队列
32 static class LinkedListQueue<E> {
33
34     private static class Node<E> {
35         E value;
36         Node<E> next;
37
38         public Node(E value, Node<E> next) {
39             this.value = value;
40             this.next = next;
41         }
42     }
43
44     private final Node<E> head = new Node<>(null, null);
45     private Node<E> tail = head;
46     int size = 0;
47     private int capacity = Integer.MAX_VALUE;
48
49     {
50         tail.next = head;
51     }
52
53     public LinkedListQueue() {
54     }
55
56     public LinkedListQueue(int capacity) {
57         this.capacity = capacity;
58     }
59
60     public boolean offer(E value) {
61         if (isFull()) {
62             return false;
63         }
64         Node<E> added = new Node<>(value, head);
65         tail.next = added;

```

```

66         tail = added;
67         size++;
68         return true;
69     }
70
71     public E poll() {
72         if (isEmpty()) {
73             return null;
74         }
75         Node<E> first = head.next;
76         head.next = first.next;
77         if (first == tail) {
78             tail = head;
79         }
80         size--;
81         return first.value;
82     }
83
84     public E peek() {
85         if (isEmpty()) {
86             return null;
87         }
88         return head.next.value;
89     }
90
91     public boolean isEmpty() {
92         return head == tail;
93     }
94
95     public boolean isFull() {
96         return size == capacity;
97     }
98 }
99 }

```

## Ex1. 设计队列-力扣 622 题

由于与课堂例题差别不大，这里只给出参考解答

基于链表的实现

```

1  public class Ex1Leetcode622 {
2
3      private static class Node {
4          int value;
5          Node next;
6          Node(int value, Node next) {
7              this.value = value;
8              this.next = next;
9          }
10     }
11     private final Node head = new Node(-1, null);
12     private Node tail = head;

```

```
13     private int size = 0;
14     private int capacity = 0;
15
16     {
17         tail.next = head;
18     }
19
20     public Ex1Leetcode622(int capacity) {
21         this.capacity = capacity;
22     }
23
24     public boolean enqueue(int value) {
25         if(isFull()) {
26             return false;
27         }
28         Node added = new Node(value, head);
29         tail.next = added;
30         tail = added;
31         size++;
32         return true;
33     }
34
35     public boolean dequeue() {
36         if(isEmpty()) {
37             return false;
38         }
39         Node first = head.next;
40         head.next = first.next;
41         if (first == tail) {
42             tail = head;
43         }
44         size--;
45         return true;
46     }
47
48     public int Front() {
49         if(isEmpty()) {
50             return -1;
51         }
52         return head.next.value;
53     }
54
55     public int Rear() {
56         if(isEmpty()) {
57             return -1;
58         }
59         return tail.value;
60     }
61
62     public boolean isEmpty() {
63         return head == tail;
64     }
65
66     public boolean isFull() {
67         return size == capacity;
```

```
68     }  
69 }
```

注意:

- Leetcode 的实现里 deQueue (出队) 返回值是布尔值, 并不会返回队头元素
- 它期望用法是先用 Front 返回对头元素, 再 deQueue 出队

## 3.8 栈

### E01. 有效的括号-力扣 20 题

一个字符串中可能出现 `[]` `()` 和 `{}` 三种括号, 判断该括号是否有效

有效的例子

```
1  ()[]{}  
2  
3  ([{}])  
4  
5  ()
```

无效的例子

```
1  []  
2  
3  ([)]  
4  
5  ([
```

思路

- 遇到左括号, 把要配对的右括号放入栈顶
- 遇到右括号, 若此时栈为空, 返回 false, 否则把它与栈顶元素对比
  - 若相等, 栈顶元素弹出, 继续对比下一组
  - 若不等, 无效括号直接返回 false
- 循环结束
  - 若栈为空, 表示所有括号都配上对, 返回 true
  - 若栈不为空, 表示右没配对的括号, 应返回 false

答案 (用到了课堂案例中的 ArrayStack 类)

```
1  public boolean isValid(String s) {  
2      ArrayStack<Character> stack = new ArrayStack<>(s.length() / 2 + 1);  
3      for (int i = 0; i < s.length(); i++) {  
4          char c = s.charAt(i);  
5          if (c == '(') {  
6              stack.push(')');  
7          } else if (c == '[') {  
8              stack.push(']');  
9          } else if (c == '{') {
```

```

10         stack.push('}');
11     } else {
12         if (!stack.isEmpty() && stack.peek() == c) {
13             stack.pop();
14         } else {
15             return false;
16         }
17     }
18 }
19 return stack.isEmpty();
20 }

```

## E02. 后缀表达式求值-力扣 120 题

后缀表达式也称为逆波兰表达式，即运算符写在后面

- 从左向右进行计算
- 不必考虑运算符优先级，即不用包含括号

示例

```

1  输入: tokens = ["2","1","+","3","*"]
2  输出: 9
3  即: (2 + 1) * 3
4
5  输入: tokens = ["4","13","5","/","+"]
6  输出: 6
7  即: 4 + (13 / 5)

```

题目假设

- 数字都视为整数
- 数字和运算符个数给定正确，不会有除零发生

代码

```

1  public int evalRPN(String[] tokens) {
2      LinkedList<Integer> numbers = new LinkedList<>();
3      for (String t : tokens) {
4          switch (t) {
5              case "+" -> {
6                  Integer b = numbers.pop();
7                  Integer a = numbers.pop();
8                  numbers.push(a + b);
9              }
10             case "-" -> {
11                 Integer b = numbers.pop();
12                 Integer a = numbers.pop();
13                 numbers.push(a - b);
14             }
15             case "*" -> {
16                 Integer b = numbers.pop();
17                 Integer a = numbers.pop();

```

```

18         numbers.push(a * b);
19     }
20     case "/" -> {
21         Integer b = numbers.pop();
22         Integer a = numbers.pop();
23         numbers.push(a / b);
24     }
25     default -> numbers.push(Integer.parseInt(t));
26 }
27 }
28 return numbers.pop();
29 }

```

## E03. 中缀表达式转后缀

```

1 public class E03InfixToSuffix {
2     /*
3         思路
4         1. 遇到数字，拼串
5         2. 遇到 + - * /
6             - 优先级高于栈顶运算符 入栈
7             - 否则将栈中高级或平级运算符出栈拼串，本运算符入栈
8         3. 遍历完成，栈中剩余运算符出栈拼串
9             - 先出栈，意味着优先运算
10        4. 带 ()
11            - 左括号直接入栈
12            - 右括号要将栈中直至左括号为止的运算符出栈拼串
13
14        |   |
15        |   |
16        |   |
17        -----
18
19        a+b
20        a+b-c
21        a+b*c
22        a*b+c
23        (a+b)*c
24
25        */
26     public static void main(String[] args) {
27         System.out.println(infixToSuffix("a+b"));
28         System.out.println(infixToSuffix("a+b-c"));
29         System.out.println(infixToSuffix("a+b*c"));
30         System.out.println(infixToSuffix("a*b-c"));
31         System.out.println(infixToSuffix("(a+b)*c"));
32         System.out.println(infixToSuffix("a+b*c+(d*e+f)*g"));
33     }
34
35     static String infixToSuffix(String exp) {
36         LinkedList<Character> stack = new LinkedList<>();
37         StringBuilder sb = new StringBuilder(exp.length());
38         for (int i = 0; i < exp.length(); i++) {

```

```

39         char c = exp.charAt(i);
40         switch (c) {
41             case '+', '-', '*', '/' -> {
42                 if (stack.isEmpty()) {
43                     stack.push(c);
44                 } else {
45                     if (priority(c) > priority(stack.peek())) {
46                         stack.push(c);
47                     } else {
48                         while (!stack.isEmpty()
49                             && priority(stack.peek()) >= priority(c))
50                             sb.append(stack.pop());
51                     }
52                     stack.push(c);
53                 }
54             }
55         }
56         case '(' -> {
57             stack.push(c);
58         }
59         case ')' -> {
60             while (!stack.isEmpty() && stack.peek() != '(') {
61                 sb.append(stack.pop());
62             }
63             stack.pop();
64         }
65         default -> {
66             sb.append(c);
67         }
68     }
69 }
70 while (!stack.isEmpty()) {
71     sb.append(stack.pop());
72 }
73 return sb.toString();
74 }
75
76 static int priority(char c) {
77     return switch (c) {
78         case '(' -> 0;
79         case '*', '/' -> 2;
80         case '+', '-' -> 1;
81         default -> throw new IllegalArgumentException("不合法字符:" + c);
82     };
83 }
84 }

```



## E04. 双栈模拟队列-力扣 232 题

给力扣题目用的**自实现栈**，可以定义为静态内部类

```
1  class ArrayStack<E> {
2
3      private E[] array;
4      private int top; // 栈顶指针
5
6      @SuppressWarnings("all")
7      public ArrayStack(int capacity) {
8          this.array = (E[]) new Object[capacity];
9      }
10
11     public boolean push(E value) {
12         if (isFull()) {
13             return false;
14         }
15         array[top++] = value;
16         return true;
17     }
18
19     public E pop() {
20         if (isEmpty()) {
21             return null;
22         }
23         return array[--top];
24     }
25
26     public E peek() {
27         if (isEmpty()) {
28             return null;
29         }
30         return array[top - 1];
31     }
32
33     public boolean isEmpty() {
34         return top == 0;
35     }
36
37     public boolean isFull() {
38         return top == array.length;
39     }
40 }
```

参考解答，注意：题目已说明

- 调用 push、pop 等方法的次数最多 100

```
1  public class E04Leetcode232 {
2
3      /*
4          队列头        队列尾
5          s1            s2
6          顶    底    底    顶
```

```

7         abc
8
9         push(a)
10        push(b)
11        push(c)
12        pop()
13    */
14    ArrayStack<Integer> s1 = new ArrayStack<>(100);
15    ArrayStack<Integer> s2 = new ArrayStack<>(100);
16
17    public void push(int x) {
18        s2.push(x);
19    }
20
21    public int pop() {
22        if (s1.isEmpty()) {
23            while (!s2.isEmpty()) {
24                s1.push(s2.pop());
25            }
26        }
27        return s1.pop();
28    }
29
30    public int peek() {
31        if (s1.isEmpty()) {
32            while (!s2.isEmpty()) {
33                s1.push(s2.pop());
34            }
35        }
36        return s1.peek();
37    }
38
39    public boolean empty() {
40        return s1.isEmpty() && s2.isEmpty();
41    }
42
43 }

```

## E05. 单队列模拟栈-力扣 225 题

给力扣题目用的**自实现**队列，可以定义为静态内部类

```

1    public class ArrayQueue3<E> {
2
3        private final E[] array;
4        int head = 0;
5        int tail = 0;
6
7        @SuppressWarnings("all")
8        public ArrayQueue3(int c) {
9            c -= 1;
10           c |= c >> 1;
11           c |= c >> 2;

```

```

12         c |= c >> 4;
13         c |= c >> 8;
14         c |= c >> 16;
15         c += 1;
16         array = (E[]) new Object[c];
17     }
18
19     public boolean offer(E value) {
20         if (isFull()) {
21             return false;
22         }
23         array[tail & (array.length - 1)] = value;
24         tail++;
25         return true;
26     }
27
28     public E poll() {
29         if (isEmpty()) {
30             return null;
31         }
32         E value = array[head & (array.length - 1)];
33         head++;
34         return value;
35     }
36
37     public E peek() {
38         if (isEmpty()) {
39             return null;
40         }
41         return array[head & (array.length - 1)];
42     }
43
44     public boolean isEmpty() {
45         return head == tail;
46     }
47
48     public boolean isFull() {
49         return tail - head == array.length;
50     }
51 }

```

参考解答，注意：题目已说明

- 调用 push、pop 等方法的次数最多 100
- 每次调用 pop 和 top 都能保证栈不为空

```

1 public class E05Leetcode225 {
2     /*
3         队列头      队列尾
4         cba
5         顶          底
6
7         queue.offer(a)
8         queue.offer(b)
9         queue.offer(c)

```

```

10     */
11     ArrayQueue3<Integer> queue = new ArrayQueue3<>(100);
12     int size = 0;
13     public void push(int x) {
14         queue.offer(x);
15         for (int i = 0; i < size; i++) {
16             queue.offer(queue.poll());
17         }
18         size++;
19     }
20
21     public int pop() {
22         size--;
23         return queue.poll();
24     }
25
26     public int top() {
27         return queue.peek();
28     }
29
30     public boolean empty() {
31         return queue.isEmpty();
32     }
33 }

```

## 3.9 双端队列

### E01. 二叉树 Z 字层序遍历-力扣 103 题

```

1 public class E01Leetcode103 {
2     public List<List<Integer>> zigzagLevelOrder(TreeNode root) {
3         List<List<Integer>> result = new ArrayList<>();
4         if (root == null) {
5             return result;
6         }
7         LinkedList<TreeNode> queue = new LinkedList<>();
8         queue.offer(root);
9         boolean leftToRight = true;
10        int c1 = 1;
11        while (!queue.isEmpty()) {
12            int c2 = 0;
13            LinkedList<Integer> deque = new LinkedList<>();
14            for (int i = 0; i < c1; i++) {
15                TreeNode n = queue.poll();
16                if (leftToRight) {
17                    deque.offerLast(n.val);
18                } else {
19                    deque.offerFirst(n.val);
20                }
21                if (n.left != null) {
22                    queue.offer(n.left);
23                }

```

```

24         }
25         if (n.right != null) {
26             queue.offer(n.right);
27             c2++;
28         }
29     }
30     c1 = c2;
31     leftToRight = !leftToRight;
32     result.add(deque);
33 }
34
35 return result;
36 }
37
38 public static void main(String[] args) {
39     TreeNode root = new TreeNode(
40         new TreeNode(
41             new TreeNode(4),
42             2,
43             new TreeNode(5)
44         ),
45         1,
46         new TreeNode(
47             new TreeNode(6),
48             3,
49             new TreeNode(7)
50         )
51     );
52     List<List<Integer>> lists = new
E01Leetcode103().zigzagLevelOrder(root);
53     for (List<Integer> list : lists) {
54         System.out.println(list);
55     }
56 }
57 }

```

## Ex1. 设计双端队列-力扣 641 题

与课堂例题也是差别不大，请参考

## 3.10 优先级队列

### E01. 合并多个有序链表-力扣 23 题

这道题目之前解答过，现在用刚学的优先级队列来实现一下

题目中要从小到大排列，因此选择用小顶堆来实现，自定义小顶堆如下

```

1 public class MinHeap {
2
3     ListNode[] array;
4     int size;
5

```

```

6     public MinHeap(int capacity) {
7         array = new ListNode[capacity];
8     }
9
10    public void offer(ListNode offered) {
11        int child = size++;
12        int parent = (child - 1) / 2;
13        while (child > 0 && offered.val < array[parent].val) {
14            array[child] = array[parent];
15            child = parent;
16            parent = (child - 1) / 2;
17        }
18        array[child] = offered;
19    }
20
21    public ListNode poll() {
22        if (isEmpty()) {
23            return null;
24        }
25        swap(0, size - 1);
26        size--;
27        ListNode e = array[size];
28        array[size] = null; // help GC
29
30        down(0);
31
32        return e;
33    }
34
35    private void down(int parent) {
36        int left = 2 * parent + 1;
37        int right = left + 1;
38        int min = parent;
39        if (left < size && array[left].val < array[min].val) {
40            min = left;
41        }
42        if (right < size && array[right].val < array[min].val) {
43            min = right;
44        }
45        if (min != parent) {
46            swap(min, parent);
47            down(min);
48        }
49    }
50
51    private void swap(int i, int j) {
52        ListNode t = array[i];
53        array[i] = array[j];
54        array[j] = t;
55    }
56
57    public boolean isEmpty() {
58        return size == 0;
59    }
60 }

```

```

1 public class E01Leetcode23 {
2     public ListNode mergeKLists(ListNode[] lists) {
3         // 1. 使用 jdk 的优先级队列实现
4         // PriorityQueue<ListNode> queue = new PriorityQueue<>
4         (Comparator.comparingInt(a -> a.val));
5         // 2. 使用自定义小顶堆实现
6         MinHeap queue = new MinHeap(lists.length);
7         for (ListNode head : lists) {
8             if (head != null) {
9                 queue.offer(head);
10            }
11        }
12        ListNode s = new ListNode(-1, null);
13        ListNode p = s;
14        while (!queue.isEmpty()) {
15            ListNode node = queue.poll();
16            p.next = node;
17            p = node;
18            if (node.next != null) {
19                queue.offer(node.next);
20            }
21        }
22        return s.next;
23    }
24 }

```

提问：

- 能否将每个链表的所有元素全部加入堆，再一个个从堆顶移除？

回答：

- 可以是可以，但对空间占用就高了，堆的一个优点就是用有限的空间做事情

## 3.11 堆

### E01. 堆排序

算法描述

1. heapify 建立大顶堆
2. 将堆顶与堆底交换（最大元素被交换到堆底），缩小并下潜调整堆
3. 重复第二步直至堆里剩一个元素

可以使用之前课堂例题的大顶堆来实现

```

1  int[] array = {1, 2, 3, 4, 5, 6, 7};
2  MaxHeap maxHeap = new MaxHeap(array);
3  System.out.println(Arrays.toString(maxHeap.array));
4
5  while (maxHeap.size > 1) {
6      maxHeap.swap(0, maxHeap.size - 1);
7      maxHeap.size--;
8      maxHeap.down(0);
9  }
10 System.out.println(Arrays.toString(maxHeap.array));

```

## E02. 数组中第K大元素-力扣 215 题

小顶堆（可删去用不到代码）

```

1  class MinHeap {
2      int[] array;
3      int size;
4
5      public MinHeap(int capacity) {
6          array = new int[capacity];
7      }
8
9      private void heapify() {
10         for (int i = (size >> 1) - 1; i >= 0; i--) {
11             down(i);
12         }
13     }
14
15     public int poll() {
16         swap(0, size - 1);
17         size--;
18         down(0);
19         return array[size];
20     }
21
22     public int poll(int index) {
23         swap(index, size - 1);
24         size--;
25         down(index);
26         return array[size];
27     }
28
29     public int peek() {
30         return array[0];
31     }
32
33     public boolean offer(int offered) {
34         if (size == array.length) {
35             return false;
36         }
37         up(offered);

```



```

38         size++;
39         return true;
40     }
41
42     public void replace(int replaced) {
43         array[0] = replaced;
44         down(0);
45     }
46
47     private void up(int offered) {
48         int child = size;
49         while (child > 0) {
50             int parent = (child - 1) >> 1;
51             if (offered < array[parent]) {
52                 array[child] = array[parent];
53             } else {
54                 break;
55             }
56             child = parent;
57         }
58         array[child] = offered;
59     }
60
61     private void down(int parent) {
62         int left = (parent << 1) + 1;
63         int right = left + 1;
64         int min = parent;
65         if (left < size && array[left] < array[min]) {
66             min = left;
67         }
68         if (right < size && array[right] < array[min]) {
69             min = right;
70         }
71         if (min != parent) {
72             swap(min, parent);
73             down(min);
74         }
75     }
76
77     // 交换两个索引处的元素
78     private void swap(int i, int j) {
79         int t = array[i];
80         array[i] = array[j];
81         array[j] = t;
82     }
83 }

```

题解

```

1 public int findKthLargest(int[] numbers, int k) {
2     MinHeap heap = new MinHeap(k);
3     for (int i = 0; i < k; i++) {
4         heap.offer(numbers[i]);
5     }
6     for (int i = k; i < numbers.length; i++) {
7         if(numbers[i] > heap.peek()){
8             heap.replace(numbers[i]);
9         }
10    }
11    return heap.peek();
12 }

```

求数组中的第 K 大元素，使用堆并不是最佳选择，可以采用快速选择算法

## E03. 数据流中第K大元素-力扣 703 题

上题的小顶堆加一个方法

```

1 class MinHeap {
2     // ...
3     public boolean isFull() {
4         return size == array.length;
5     }
6 }

```

题解

```

1 class KthLargest {
2
3     private MinHeap heap;
4
5     public KthLargest(int k, int[] nums) {
6         heap = new MinHeap(k);
7         for(int i = 0; i < nums.length; i++) {
8             add(nums[i]);
9         }
10    }
11
12    public int add(int val) {
13        if(!heap.isFull()){
14            heap.offer(val);
15        } else if(val > heap.peek()){
16            heap.replace(val);
17        }
18        return heap.peek();
19    }
20
21 }

```

求数据流中的第 K 大元素，使用堆最合适不过

## E04. 数据流的中位数-力扣 295 题

可以扩容的 heap, max 用于指定是大顶堆还是小顶堆

```
1 public class Heap {
2     int[] array;
3     int size;
4     boolean max;
5
6     public int size() {
7         return size;
8     }
9
10    public Heap(int capacity, boolean max) {
11        this.array = new int[capacity];
12        this.max = max;
13    }
14
15    /**
16     * 获取堆顶元素
17     *
18     * @return 堆顶元素
19     */
20    public int peek() {
21        return array[0];
22    }
23
24    /**
25     * 删除堆顶元素
26     *
27     * @return 堆顶元素
28     */
29    public int poll() {
30        int top = array[0];
31        swap(0, size - 1);
32        size--;
33        down(0);
34        return top;
35    }
36
37    /**
38     * 删除指定索引处元素
39     *
40     * @param index 索引
41     * @return 被删除元素
42     */
43    public int poll(int index) {
44        int deleted = array[index];
45        swap(index, size - 1);
46        size--;
47        down(index);
48        return deleted;
49    }
```

```

50
51  /**
52   * 替换堆顶元素
53   *
54   * @param replaced 新元素
55   */
56  public void replace(int replaced) {
57      array[0] = replaced;
58      down(0);
59  }
60
61  /**
62   * 堆的尾部添加元素
63   *
64   * @param offered 新元素
65   */
66  public void offer(int offered) {
67      if (size == array.length) {
68          grow();
69      }
70      up(offered);
71      size++;
72  }
73
74  private void grow() {
75      int capacity = size + (size >> 1);
76      int[] newArray = new int[capacity];
77      System.arraycopy(array, 0,
78                      newArray, 0, size);
79      array = newArray;
80  }
81
82  // 将 offered 元素上浮：直至 offered 小于父元素或到堆顶
83  private void up(int offered) {
84      int child = size;
85      while (child > 0) {
86          int parent = (child - 1) / 2;
87          boolean cmp = max ? offered > array[parent] : offered <
array[parent];
88          if (cmp) {
89              array[child] = array[parent];
90          } else {
91              break;
92          }
93          child = parent;
94      }
95      array[child] = offered;
96  }
97
98  public Heap(int[] array, boolean max) {
99      this.array = array;
100     this.size = array.length;
101     this.max = max;
102     heapify();
103 }

```

```

104
105 // 建堆
106 private void heapify() {
107     // 如何找到最后这个非叶子节点 size / 2 - 1
108     for (int i = size / 2 - 1; i >= 0; i--) {
109         down(i);
110     }
111 }
112
113 // 将 parent 索引处的元素下潜：与两个孩子较大者交换，直至没孩子或孩子没它大
114 private void down(int parent) {
115     int left = parent * 2 + 1;
116     int right = left + 1;
117     int min = parent;
118     if (left < size && (max ? array[left] > array[min] : array[left] <
array[min])) {
119         min = left;
120     }
121     if (right < size && (max ? array[right] > array[min] : array[right]
< array[min])) {
122         min = right;
123     }
124     if (min != parent) { // 找到了更大的孩子
125         swap(min, parent);
126         down(min);
127     }
128 }
129
130 // 交换两个索引处的元素
131 private void swap(int i, int j) {
132     int t = array[i];
133     array[i] = array[j];
134     array[j] = t;
135 }
136 }

```

## 题解

```

1 private Heap left = new Heap(10, false);
2 private Heap right = new Heap(10, true);
3
4 /**
5  为了保证两边数据量的平衡
6  <ul>
7      <li>两边数据一样时,加入左边</li>
8      <li>两边数据不一样时,加入右边</li>
9  </ul>
10 但是，随便一个数能直接加入吗？
11  <ul>
12      <li>加入左边前，应该挑右边最小的加入</li>
13      <li>加入右边前，应该挑左边最大的加入</li>
14  </ul>
15  */
16 public void addNum(int num) {
17     if (left.size() == right.size()) {

```

```

18         right.offer(num);
19         left.offer(right.poll());
20     } else {
21         left.offer(num);
22         right.offer(left.poll());
23     }
24 }
25
26 /**
27  * <ul>
28  *     <li>两边数据一致，左右各取堆顶元素求平均</li>
29  *     <li>左边多一个，取左边元素</li>
30  * </ul>
31  */
32 public double findMedian() {
33     if (left.size() == right.size()) {
34         return (left.peek() + right.peek()) / 2.0;
35     } else {
36         return left.peek();
37     }
38 }

```

本题还可以使用平衡二叉搜索树求解，不过代码比两个堆复杂

## 3.12 二叉树

### E04. 对称二叉树-力扣 101 题

```

1 public boolean isSymmetric(TreeNode root) {
2     return check(root.left, root.right);
3 }
4
5 public boolean check(TreeNode left, TreeNode right) {
6     // 若同时为 null
7     if (left == null && right == null) {
8         return true;
9     }
10    // 若有一个为 null (有上一轮筛选，另一个肯定不为 null)
11    if (left == null || right == null) {
12        return false;
13    }
14    if (left.val != right.val) {
15        return false;
16    }
17    return check(left.left, right.right) && check(left.right, right.left);
18 }

```

类似题目：Leetcode 100 题 - 相同的树

## E05. 二叉树最大深度-力扣 104 题

### 后序遍历求解

```
1  /*
2      思路：
3      1. 得到左子树深度，得到右子树深度，二者最大者加一，就是本节点深度
4      2. 因为需要先得到左右子树深度，很显然是后序遍历典型应用
5      3. 关于深度的定义：从根出发，离根最远的节点总边数，
6          注意：力扣里的深度定义要多一
7
8      深度2      深度3      深度1
9      1          1          1
10     / \       / \       / \
11    2   3     2   3     2   3
12               \       \
13              4         4
14  */
15 public int maxDepth(TreeNode node) {
16     if (node == null) {
17         return 0; // 非力扣题目改为返回 -1
18     }
19     int d1 = maxDepth(node.left);
20     int d2 = maxDepth(node.right);
21     return Integer.max(d1, d2) + 1;
22 }
```

### 后序遍历求解-非递归

```
1  /*
2      思路：
3      1. 使用非递归后序遍历，栈的最大高度即为最大深度
4      */
5  public int maxDepth(TreeNode root) {
6      TreeNode curr = root;
7      LinkedList<TreeNode> stack = new LinkedList<>();
8      int max = 0;
9      TreeNode pop = null;
10     while (curr != null || !stack.isEmpty()) {
11         if (curr != null) {
12             stack.push(curr);
13             int size = stack.size();
14             if (size > max) {
15                 max = size;
16             }
17             curr = curr.left;
18         } else {
19             TreeNode peek = stack.peek();
20             if(peek.right == null || peek.right == pop) {
21                 pop = stack.pop();
22             } else {
23                 curr = peek.right;
24             }
25         }
26     }
27 }
```

```

25     }
26 }
27 return max;
28 }

```

## 层序遍历求解

```

1  /*
2     思路:
3     1. 使用层序遍历, 层数即最大深度
4  */
5  public int maxDepth(TreeNode root) {
6      if(root == null) {
7          return 0;
8      }
9      Queue<TreeNode> queue = new LinkedList<>();
10     queue.offer(root);
11     int level = 0;
12     while (!queue.isEmpty()) {
13         level++;
14         int size = queue.size();
15         for (int i = 0; i < size; i++) {
16             TreeNode node = queue.poll();
17             if (node.left != null) {
18                 queue.offer(node.left);
19             }
20             if (node.right != null) {
21                 queue.offer(node.right);
22             }
23         }
24     }
25     return level;
26 }

```

## E06. 二叉树最小深度-力扣 111 题

### 后序遍历求解

```

1  public int minDepth(TreeNode node) {
2      if (node == null) {
3          return 0;
4      }
5      int d1 = minDepth(node.left);
6      int d2 = minDepth(node.right);
7      if (d1 == 0 || d2 == 0) {
8          return d1 + d2 + 1;
9      }
10     return Integer.min(d1, d2) + 1;
11 }

```



相较于求最大深度，应当考虑：

- 当右子树为 null，应当返回左子树深度加一
- 当左子树为 null，应当返回右子树深度加一

上面两种情况满足时，不应该再把为 null 子树的深度 0 参与最小值比较，例如这样

```
1 |      1
2 |     /
3 |    2
```

- 正确深度为 2，若把为 null 的右子树的深度 0 考虑进来，会得到错误结果 1

```
1 |      1
2 |     \
3 |      3
4 |     \
5 |      4
```

- 正确深度为 3，若把为 null 的左子树的深度 0 考虑进来，会得到错误结果 1

## 层序遍历求解

遇到的第一个叶子节点所在层就是最小深度

例如，下面的树遇到的第一个叶子节点 3 所在的层就是最小深度，其他 4，7 等叶子节点深度更深，也更晚遇到

```
1 |      1
2 |     / \
3 |    2   3
4 |   / \
5 |  4   5
6 |   /
7 |  7
```

## 代码

```
1 public int minDepth(TreeNode root) {
2     if(root == null) {
3         return 0;
4     }
5     Queue<TreeNode> queue = new LinkedList<>();
6     queue.offer(root);
7     int level = 0;
8     while (!queue.isEmpty()) {
9         level++;
10        int size = queue.size();
11        for (int i = 0; i < size; i++) {
12            TreeNode node = queue.poll();
13            if (node.left == null && node.right == null) {
14                return level;
15            }
16        }
17    }
18 }
```

```

16         if (node.left != null) {
17             queue.offer(node.left);
18         }
19         if (node.right != null) {
20             queue.offer(node.right);
21         }
22     }
23 }
24 return level;
25 }

```

效率会高于之前后序遍历解法，因为找到第一个叶子节点后，就无需后续的层序遍历了

## E07. 翻转二叉树-力扣 226 题

```

1 public TreeNode invertTree(TreeNode root) {
2     fn(root);
3     return root;
4 }
5
6 private void fn(TreeNode node){
7     if (node == null) {
8         return;
9     }
10    TreeNode t = node.left;
11    node.left = node.right;
12    node.right = t;
13    fn(node.left);
14    fn(node.right);
15 }

```

先交换、再递归或是先递归、再交换都可以

## E08. 后缀表达式转二叉树

```

1 static class TreeNode {
2     public String val;
3     public TreeNode left;
4     public TreeNode right;
5
6     public TreeNode(String val) {
7         this.val = val;
8     }
9
10    public TreeNode(TreeNode left, String val, TreeNode right) {
11        this.left = left;
12        this.val = val;
13        this.right = right;
14    }
15
16    @Override

```

```

17     public String toString() {
18         return this.val;
19     }
20 }
21
22  /*
23     中缀表达式          (2-1)*3
24     后缀（逆波兰）表达式  21-3*
25
26     1.遇到数字入栈
27     2.遇到运算符，出栈两次，与当前节点建立父子关系，当前节点入栈
28
29     栈
30     |   |
31     |   |
32     |   |
33     -----
34
35     表达式树
36         *
37       /  \
38      -   3
39     /  \
40    2   1
41
42     21-3*
43  */
44  public TreeNode constructExpressionTree(String[] tokens) {
45      LinkedList<TreeNode> stack = new LinkedList<>();
46      for (String t : tokens) {
47          switch (t) {
48              case "+", "-", "*", "/" -> { // 运算符
49                  TreeNode right = stack.pop();
50                  TreeNode left = stack.pop();
51                  TreeNode parent = new TreeNode(t);
52                  parent.left = left;
53                  parent.right = right;
54                  stack.push(parent);
55              }
56              default -> { // 数字
57                  stack.push(new TreeNode(t));
58              }
59          }
60      }
61      return stack.peek();
62  }

```

## E09. 根据前序与中序遍历结果构造二叉树-力扣 105 题

- 先通过前序遍历结果定位根节点
- 再结合中序遍历结果切分左右子树

```

1  public class E09Leetcode105 {

```

```

2
3      /*
4          preOrder = {1,2,4,3,6,7}
5          inOrder = {4,2,1,6,3,7}
6
7          根 1
8              pre      in
9          左  2,4      4,2
10         右  3,6,7    6,3,7
11
12
13         根 2
14         左 4
15
16         根 3
17         左 6
18         右 7
19     */
20
21     public TreeNode buildTree(int[] preOrder, int[] inOrder) {
22         if (preOrder.length == 0) {
23             return null;
24         }
25         // 创建根节点
26         int rootValue = preOrder[0];
27         TreeNode root = new TreeNode(rootValue);
28         // 区分左右子树
29         for (int i = 0; i < inOrder.length; i++) {
30             if (inOrder[i] == rootValue) {
31                 // 0 ~ i-1 左子树
32                 // i+1 ~ inOrder.length -1 右子树
33                 int[] inLeft = Arrays.copyOfRange(inOrder, 0, i); // [4,2]
34                 int[] inRight = Arrays.copyOfRange(inOrder, i + 1,
35 inOrder.length); // [6,3,7]
36
37                 int[] preLeft = Arrays.copyOfRange(preOrder, 1, i + 1); //
[2,4]
38                 int[] preRight = Arrays.copyOfRange(preOrder, i + 1,
39 inOrder.length); // [3,6,7]
40
41                 root.left = buildTree(preLeft, inLeft); // 2
42                 root.right = buildTree(preRight, inRight); // 3
43                 break;
44             }
45         }
46         return root;
47     }

```

- 代码可以进一步优化，涉及新数据结构，以后实现

## E10. 根据中序与后序遍历结果构造二叉树-力扣 106 题

- 先通过后序遍历结果定位根节点
- 再结合中序遍历结果切分左右子树

```
1 public TreeNode buildTree(int[] inOrder, int[] postOrder) {
2     if (inOrder.length == 0) {
3         return null;
4     }
5     // 根
6     int rootValue = postOrder[postOrder.length - 1];
7     TreeNode root = new TreeNode(rootValue);
8     // 切分左右子树
9     for (int i = 0; i < inOrder.length; i++) {
10        if (inOrder[i] == rootValue) {
11            int[] inLeft = Arrays.copyOfRange(inOrder, 0, i);
12            int[] inRight = Arrays.copyOfRange(inOrder, i + 1,
13inOrder.length);
14            int[] postLeft = Arrays.copyOfRange(postOrder, 0, i);
15            int[] postRight = Arrays.copyOfRange(postOrder, i,
16postOrder.length - 1);
17            root.left = buildTree(inLeft, postLeft);
18            root.right = buildTree(inRight, postRight);
19            break;
20        }
21    }
22    return root;
23 }
```

- 代码可以进一步优化，涉及新数据结构，以后实现

## 附录

### 参考文章

### 落选题目

### 反转字符数组

```
1 public static void main(String[] args) {
2     char[] array = "abcde".toCharArray();
3     recursion(array, 0, array.length - 1);
4     System.out.println(Arrays.toString(array));
5 }
6
7 public static void recursion(char[] array, int i, int j) {
8     if (i >= j) {
9         return;
```

```

10     }
11     swap(array, i, j);
12     recursion(array, ++i, --j);
13 }
14
15 public static void swap(char[] array, int i, int j) {
16     char c = array[i];
17     array[i] = array[j];
18     array[j] = c;
19 }






```

- 第一次交换的是 array[0] 和 array[4]
- 第二次交换的是 array[1] 和 array[3]
- 第三次 i = j = 2, 开始返回
- 如果 array.length 是偶数, 则会在 i > j 时返回

## 力扣高评价题目列表

引用自 [面试最常考的 100 道算法题分类整理! - 知乎\(zhihu.com\)](#)

带  是本课程讲解过的

- [1. Two Sum \(两数之和\)](#), Easy, 11757 likes 
- [2. Add Two Numbers \(两数相加\)](#), Medium, 6524 likes
- [3. Longest Substring Without Repeating Characters \(无重复字符的最长子串\)](#), Medium, 5845 likes 
- [4. Median of Two Sorted Arrays \(寻找两个正序数组的中位数\)](#), Hard, 4303 likes
- [5. Longest Palindromic Substring \(最长回文子串\)](#), Medium, 3896 likes
- [15. 3Sum \(三数之和\)](#), Medium, 3582 likes
- [53. Maximum Subarray \(最大子序和\)](#), Easy, 3533 likes
- [7. Reverse Integer \(整数反转\)](#), Easy, 2970 likes
- [11. Container With Most Water \(盛最多水的容器\)](#), Medium, 2659 likes
- [42. Trapping Rain Water \(接雨水\)](#), Hard, 2552 likes
- [20. Valid Parentheses \(有效的括号\)](#), Easy, 2544 likes 
- [10. Regular Expression Matching \(正则表达式匹配\)](#), Hard, 2273 likes
- [26. Remove Duplicates from Sorted Array \(删除有序数组中的重复项\)](#), Easy, 2146 likes 
- [136. Single Number \(只出现一次的数字\)](#), Easy, 1958 likes 
- [22. Generate Parentheses \(括号生成\)](#), Medium, 1946 likes
- [206. Reverse Linked List \(反转链表\)](#), Easy, 1886 likes 
- [21. Merge Two Sorted Lists \(合并两个有序链表\)](#), Easy, 1832 likes 
- [70. Climbing Stairs \(爬楼梯\)](#), Easy, 1791 likes 
- [300. Longest Increasing Subsequence \(最长递增子序列\)](#), Medium, 1773 likes
- [121. Best Time to Buy and Sell Stock \(买卖股票的最佳时机\)](#), Easy, 1766 likes
- [72. Edit Distance \(编辑距离\)](#), Hard, 1743 likes
- [14. Longest Common Prefix \(最长公共前缀\)](#), Easy, 1707 likes
- [198. House Robber \(打家劫舍\)](#), Medium, 1585 likes
- [9. Palindrome Number \(回文数\)](#), Easy, 1568 likes
- [146. LRU Cache \(LRU 缓存机制\)](#), Medium, 1544 likes
- [19. Remove Nth Node From End of List \(删除链表的倒数第 N 个结点\)](#), Medium, 1494 likes 
- [33. Search in Rotated Sorted Array \(搜索旋转排序数组\)](#), Medium, 1493 likes

- [46. Permutations \(全排列\)](#), Medium, 1484 likes
- [101. Symmetric Tree \(对称二叉树\)](#), Easy, 1483 likes ✓
- [84. Largest Rectangle in Histogram \(柱状图中最大的矩形\)](#), Hard, 1472 likes
- [39. Combination Sum \(组合总和\)](#), Medium, 1466 likes
- [13. Roman to Integer \(罗马数字转整数\)](#), Easy, 1436 likes
- [23. Merge k Sorted Lists \(合并K个升序链表\)](#), Hard, 1436 likes ✓
- [17. Letter Combinations of a Phone Number \(电话号码的字母组合\)](#), Medium, 1436 likes
- [322. Coin Change \(零钱兑换\)](#), Medium, 1414 likes
- [32. Longest Valid Parentheses \(最长有效括号\)](#), Hard, 1400 likes
- [287. Find the Duplicate Number \(寻找重复数\)](#), Medium, 1325 likes
- [122. Best Time to Buy and Sell Stock II \(买卖股票的最佳时机 II\)](#), Easy, 1306 likes
- [160. Intersection of Two Linked Lists \(相交链表\)](#), Easy, 1302 likes ✓
- [55. Jump Game \(跳跃游戏\)](#), Medium, 1292 likes
- [76. Minimum Window Substring \(最小覆盖子串\)](#), Hard, 1280 likes
- [200. Number of Islands \(岛屿数量\)](#), Medium, 1270 likes
- [78. Subsets \(子集\)](#), Medium, 1269 likes
- [31. Next Permutation \(下一个排列\)](#), Medium, 1260 likes
- [96. Unique Binary Search Trees \(不同的二叉搜索树\)](#), Medium, 1257 likes
- [148. Sort List \(排序链表\)](#), Medium, 1248 likes
- [236. Lowest Common Ancestor of a Binary Tree \(二叉树的最近公共祖先\)](#), Medium, 1238 likes ✓
- [25. Reverse Nodes in k-Group \(K 个一组翻转链表\)](#), Hard, 1230 likes
- [6. ZigZag Conversion \(Z 字形变换\)](#), Medium, 1226 likes
- [152. Maximum Product Subarray \(乘积最大子数组\)](#), Medium, 1223 likes
- [215. Kth Largest Element in an Array \(数组中的第K个最大元素\)](#), Medium, 1211 likes ✓
- [8. String to Integer \(atoi\) \(字符串转换整数 \(atoi\)\)](#), Medium, 1168 likes
- [41. First Missing Positive \(缺失的第一个正数\)](#), Hard, 1163 likes
- [283. Move Zeroes \(移动零\)](#), Easy, 1162 likes
- [141. Linked List Cycle \(环形链表\)](#), Easy, 1161 likes ✓
- [98. Validate Binary Search Tree \(验证二叉搜索树\)](#), Medium, 1156 likes ✓
- [124. Binary Tree Maximum Path Sum \(二叉树中的最大路径和\)](#), Hard, 1152 likes
- [105. Construct Binary Tree from Preorder and Inorder Traversal \(从前序与中序遍历序列构造二叉树\)](#), Medium, 1149 likes ✓
- [34. Find First and Last Position of Element in Sorted Array \(在排序数组中查找元素的第一个和最后一个位置\)](#), Medium, 1137 likes ✓
- [239. Sliding Window Maximum \(滑动窗口最大值\)](#), Hard, 1114 likes
- [142. Linked List Cycle II \(环形链表 II\)](#), Medium, 1097 likes ✓
- [139. Word Break \(单词拆分\)](#), Medium, 1097 likes
- [45. Jump Game II \(跳跃游戏 II\)](#), Medium, 1094 likes
- [169. Majority Element \(多数元素\)](#), Easy, 1089 likes
- [234. Palindrome Linked List \(回文链表\)](#), Easy, 1072 likes ✓
- [62. Unique Paths \(不同路径\)](#), Medium, 1072 likes
- [189. Rotate Array \(旋转数组\)](#), Medium, 1057 likes
- [94. Binary Tree Inorder Traversal \(二叉树的中序遍历\)](#), Easy, 1052 likes ✓
- [56. Merge Intervals \(合并区间\)](#), Medium, 1051 likes
- [88. Merge Sorted Array \(合并两个有序数组\)](#), Easy, 1041 likes ✓
- [560. Subarray Sum Equals K \(和为K的子数组\)](#), Medium, 1036 likes
- [279. Perfect Squares \(完全平方数\)](#), Medium, 1035 likes
- [35. Search Insert Position \(搜索插入位置\)](#), Easy, 1005 likes
- [24. Swap Nodes in Pairs \(两两交换链表中的节点\)](#), Medium, 996 likes

- [85. Maximal Rectangle \(最大矩形\)](#), Hard, 983 likes
- [28. Implement strStr\(\) \(实现 strStr\(\)\)](#), Easy, 982 likes
- [92. Reverse Linked List II \(反转链表 II\)](#), Medium, 980 likes
- [155. Min Stack \(最小栈\)](#), Easy, 979 likes
- [79. Word Search \(单词搜索\)](#), Medium, 979 likes
- [27. Remove Element \(移除元素\)](#), Easy, 967 likes
- [51. N-Queens \(N 皇后\)](#), Hard, 965 likes
- [75. Sort Colors \(颜色分类\)](#), Medium, 961 likes
- [102. Binary Tree Level Order Traversal \(二叉树的层序遍历\)](#), Medium, 960 likes ✓
- [48. Rotate Image \(旋转图像\)](#), Medium, 960 likes
- [95. Unique Binary Search Trees II \(不同的二叉搜索树 II\)](#), Medium, 955 likes
- [64. Minimum Path Sum \(最小路径和\)](#), Medium, 954 likes
- [406. Queue Reconstruction by Height \(根据身高重建队列\)](#), Medium, 947 likes
- [226. Invert Binary Tree \(翻转二叉树\)](#), Easy, 941 likes ✓
- [437. Path Sum III \(路径总和 III\)](#), Medium, 937 likes
- [104. Maximum Depth of Binary Tree \(二叉树的最大深度\)](#), Easy, 937 likes ✓
- [237. Delete Node in a Linked List \(删除链表中的节点\)](#), Easy, 936 likes ✓
- [337. House Robber III \(打家劫舍 III\)](#), Medium, 929 likes
- [18. 4Sum \(四数之和\)](#), Medium, 918 likes
- [91. Decode Ways \(解码方法\)](#), Medium, 904 likes
- [207. Course Schedule \(课程表\)](#), Medium, 897 likes
- [37. Sudoku Solver \(解数独\)](#), Hard, 897 likes
- [175. Combine Two Tables \(组合两个表\)](#), Easy, 891 likes
- [416. Partition Equal Subset Sum \(分割等和子集\)](#), Medium, 886 likes
- [238. Product of Array Except Self \(除自身以外数组的乘积\)](#), Medium, 885 likes
- [114. Flatten Binary Tree to Linked List \(二叉树展开为链表\)](#), Medium, 877 likes ✓

---

1. "Definition of ALGORITHM". *Merriam-Webster Online Dictionary*. [Archived](#) from the original on February 14, 2020. Retrieved November 14, 2019. [↗](#)

2. Introduction to Algorithm 中文译作《算法导论》 [↗](#) [↗](#)

3. 主要参考文档 [https://en.wikipedia.org/wiki/Binary\\_search\\_algorithm](https://en.wikipedia.org/wiki/Binary_search_algorithm) [↗](#)

4. 图片及概念均摘自 Introduction to Algorithm 4th, 3.1节, 3.2节 [↗](#)

5. jdk 版本有关, 64 位 jdk, 按 8 字节对齐 [↗](#)

6. 图片引用自 wikipedia linkedlist 条目, [https://en.wikipedia.org/wiki/Linked\\_list](https://en.wikipedia.org/wiki/Linked_list) [↗](#)

7. Fibonacci 介绍: [https://en.wikipedia.org/wiki/Fibonacci\\_number](https://en.wikipedia.org/wiki/Fibonacci_number) [↗](#) [↗](#)

8. 几种计算Fibonacci数列算法的时间复杂度比较 - 知乎 (zhihu.com) [↗](#)

9. 几种斐波那契数列算法比较 [Fast Fibonacci algorithms \(nayuki.io\)](https://nayuki.io/Fast_Fibonacci_algorithms) [↗](#)

10. 我知道的有 C++, Scala [↗](#)

11. 与主定理类似的还有 Akra-Bazzi method, [https://en.wikipedia.org/wiki/Akra%E2%80%93Bazzi\\_method](https://en.wikipedia.org/wiki/Akra%E2%80%93Bazzi_method) [↗](#)

12. Josephus problem 主要参考 [https://en.wikipedia.org/wiki/Josephus\\_problem](https://en.wikipedia.org/wiki/Josephus_problem) [↗](#)

13. 汉诺塔图片资料均来自 [https://en.wikipedia.org/wiki/Tower\\_of\\_Hanoi](https://en.wikipedia.org/wiki/Tower_of_Hanoi) [↗](#)

14. 也称为 Pascal's triangle [https://en.wikipedia.org/wiki/Pascal%27s\\_triangle](https://en.wikipedia.org/wiki/Pascal%27s_triangle) [↗](#)

15. 龟兔赛跑动画来自于 [Floyd's Hare and Tortoise Algorithm Demo - One Step! Code \(onestepcode.com\)](https://onestepcode.com/Floyd's_Hare_and_Tortoise_Algorithm_Demo) [↗](#)