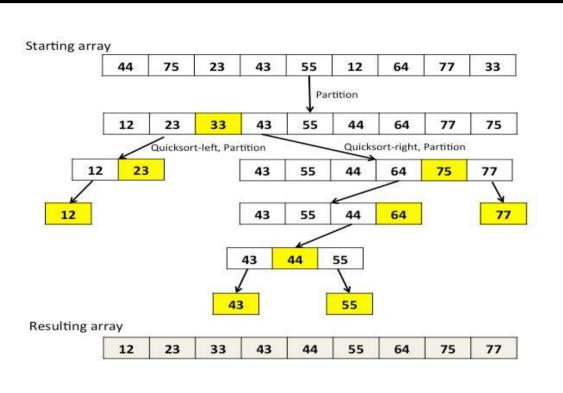
2017

PISL 05

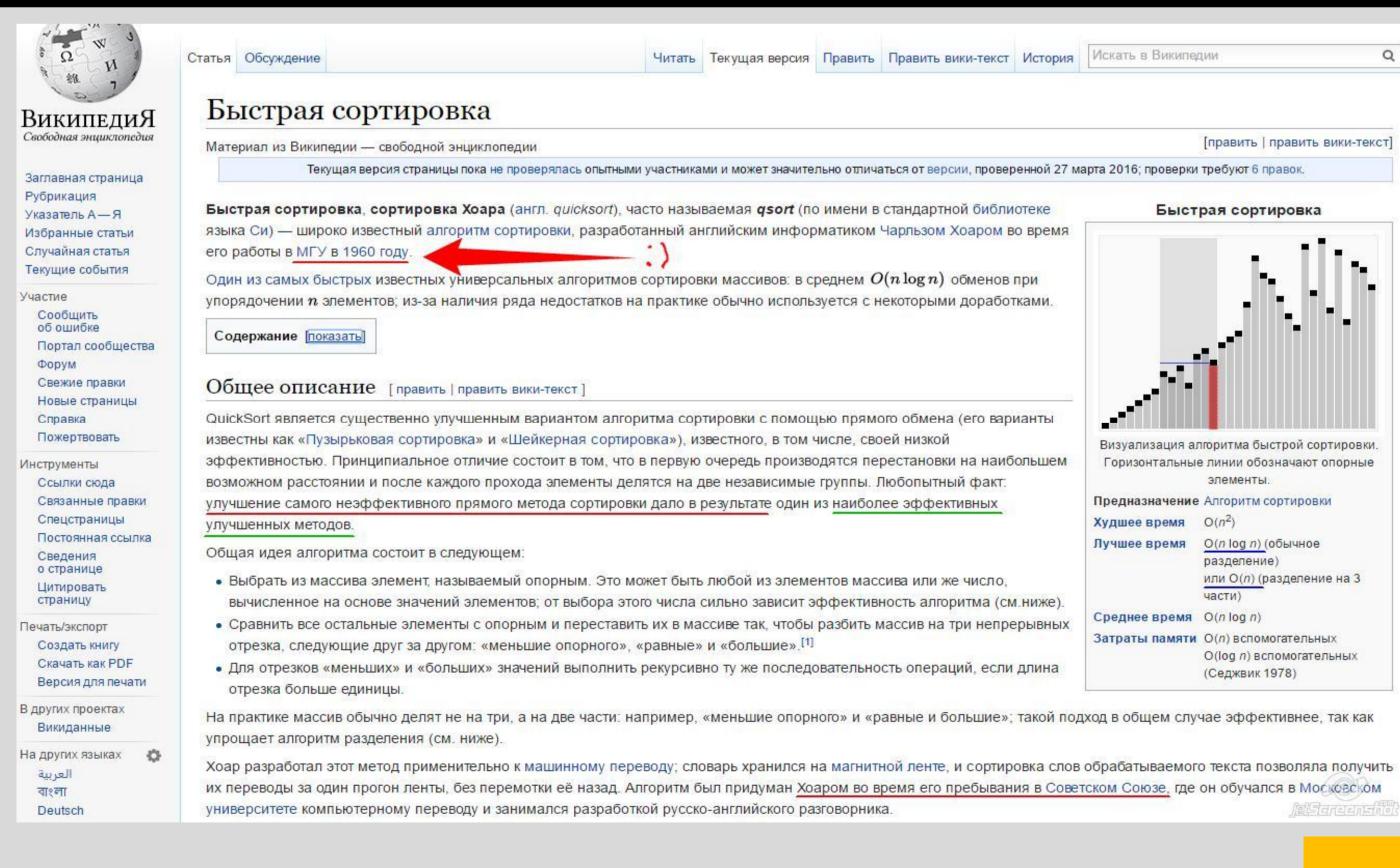
PISL05. Quick Sort, Heap Sort, Count Sort etc.





Кафедра экономической информатики Бгуир, 201

PISL04. Quick sort



PISL05. Quick Sort, Heap Sort, Count Sort etc.

æ Quick Sort.

- § Принцип сортировки
- § Доказательство эффективности
- § 0-оценкі
- § Оптимизация Quick Sort

e Selection sort

- Принцип сортировки
- § О-оценки

Heap sort

- § Принцип сортировки
- § 0-оценки
- **20** Сортировки, быстрее чем O(n log n)
- **2** Задачи (A) (B) (C)

æ Материалы: http://tinyurl.com/ei-pisl

æ Github: https://github.com/Khmelov/PISL2017-01-26

Кафедра экономической информатики Бгуир 2017

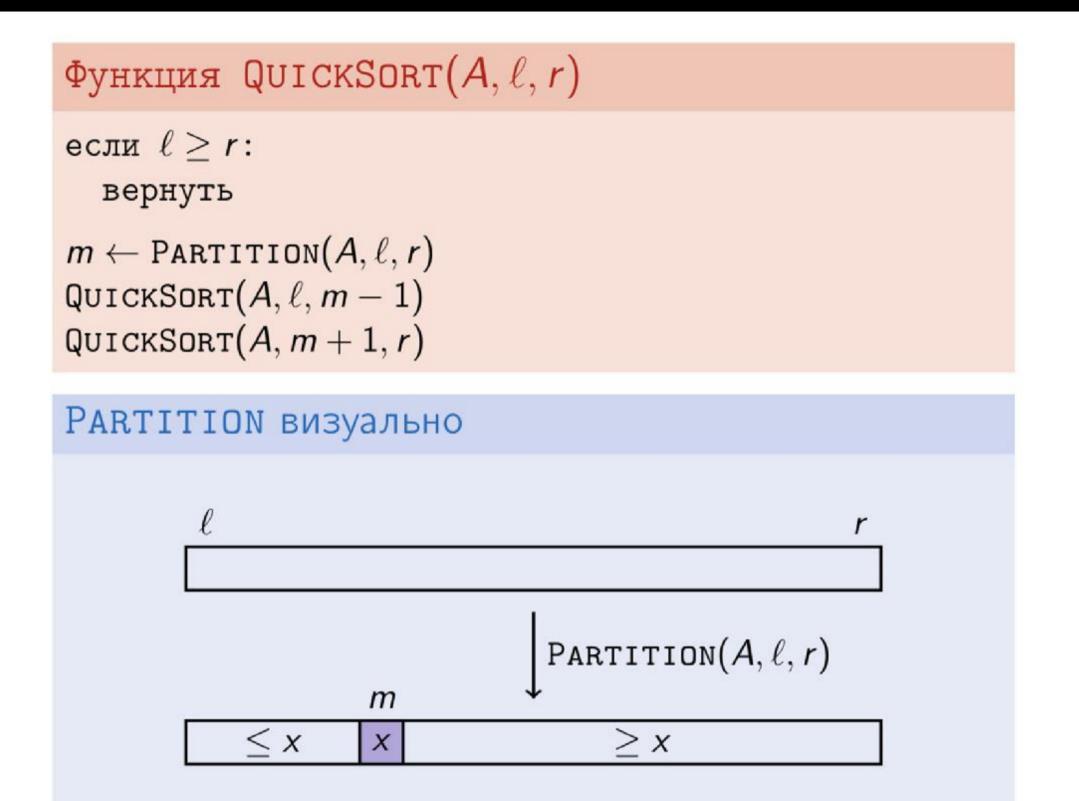
PISL05. Quick Sort, Heap Sort, Count Sort etc.

```
если \ell \geq r: вернуть
```

Функция QUICKSORT (A, ℓ, r)

 $m \leftarrow \text{PARTITION}(A, \ell, r)$ QUICKSORT $(A, \ell, m-1)$

QUICKSORT(A, m+1, r)



Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

```
Функция PARTITION(A, \ell, r)

x \leftarrow A[\ell]

j \leftarrow \ell

для i от \ell + 1 до r:

если A[i] \leq x:

j \leftarrow j + 1

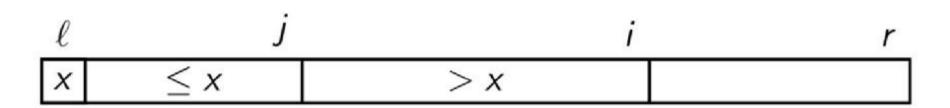
обменять A[j] и A[i]

обменять A[\ell] и A[j]

вернуть j
```

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Разбиение: основные идеи



- $\mathbf{x} = A[\ell]$ опорный элемент
- lacktriangleright двигаем i от $\ell+1$ до r, поддерживая следующий инвариант:
 - $lacksymbol{A}[k] \leq x$ для всех $\ell+1 \leq k \leq j$
 - lacksquare A[k] > x для всех $j+1 \le k \le i$
- пусть i только что увеличился; если A[i] > x, не делаем ничего; в противном случае меняем A[i] с A[j+1] и увеличиваем j

Кафедра экономической информатики Бгуир 2017

6

PISL05. Quick Sort, Heap Sort, Count Sort etc.

```
Функция PARTITION(A, \ell, r)

x \leftarrow A[\ell]

j \leftarrow \ell

для i от \ell + 1 до r:

если A[i] \leq x:

j \leftarrow j + 1

обменять A[j] и A[i]

обменять A[\ell] и A[j]

вернуть j
```

Функция РАRTITION(A, ℓ, r) $x \leftarrow A[\ell]$ $j \leftarrow \ell$ для i от $\ell + 1$ до r: если $A[i] \leq x$: $j \leftarrow j + 1$ обменять A[j] и A[i]обменять $A[\ell]$ и A[j]вернуть j

Время работы: O(n).

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Плохие и хорошие разделители

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

$$T(n) = n + (n-1) + (n-2) + \ldots = \Theta(n^2)$$

$$T(n) = T(n-5) + T(4) + n:$$

$$T(n) \ge n + (n-5) + (n-10) + \ldots = \Theta(n^2)$$

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Плохие и хорошие разделители

■
$$T(n) = T(n-1) + n$$
:
 $T(n) = n + (n-1) + (n-2) + ... = \Theta(n^2)$

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Плохие и хорошие разделители

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

$$T(n) = n + (n-1) + (n-2) + \ldots = \Theta(n^2)$$

$$T(n) = T(n-5) + T(4) + n:$$

$$T(n) \ge n + (n-5) + (n-10) + \ldots = \Theta(n^2)$$

$$T(n) = 2T(n/2) + n$$
: $T(n) = O(n \log n)$

Плохие и хорошие разделители

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

$$T(n) = n + (n-1) + (n-2) + \dots = \Theta(n^2)$$

■
$$T(n) = T(n-5) + T(4) + n$$
:
 $T(n) \ge n + (n-5) + (n-10) + \ldots = \Theta(n^2)$

$$T(n) = 2T(n/2) + n$$
: $T(n) = O(n \log n)$

$$T(n) = T(n/10) + T(9n/10) + n$$
:

$$T(n) = O(n \log n)$$

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

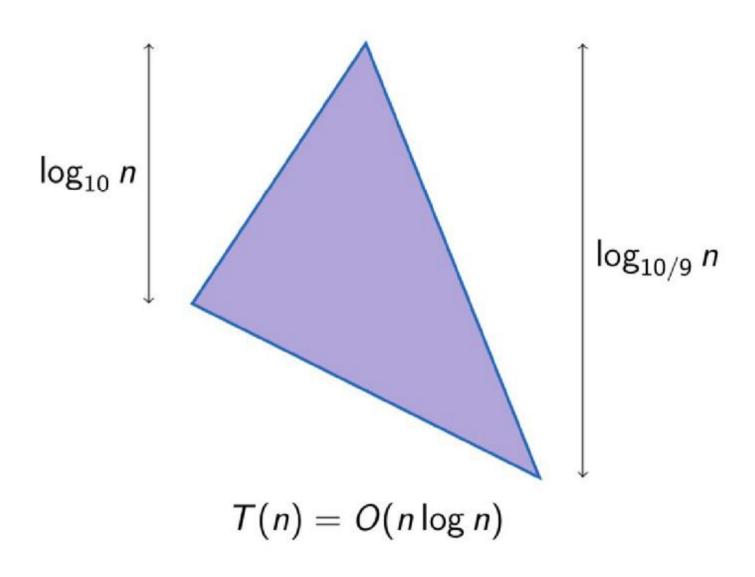
Случайный разделитель

• чтобы разбить A относительно случайного разделителя, обменяем $A[\ell]$ со случайным элементом и вызовем Рактітіон (A,ℓ,r)

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сбалансированные разбиения

$$T(n) = T(n/10) + T(9n/10) + O(n)$$

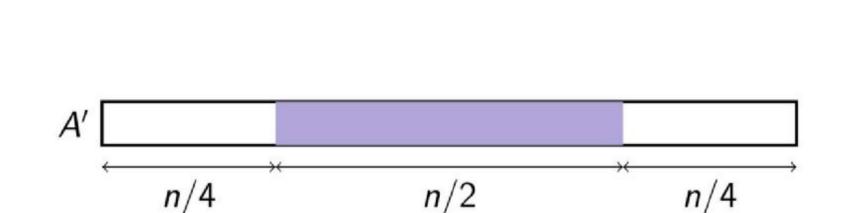


Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Случайный разделитель

- чтобы разбить A относительно случайного разделителя, обменяем $A[\ell]$ со случайным элементом и вызовем Рактітіон (A,ℓ,r)
- \blacksquare важное наблюдение: половина элементов A дают сбалансированное разбиение:



Время работы

Теорема

Допустим, что все элементы массива $A[1 \dots n]$ различны и что разделитель всегда выбирается равномерно случайным образом. Тогда среднее время работы алгоритма QuickSort(A) есть $O(n \log n)$, в то время как время работы в худшем случае есть $O(n^2)$.

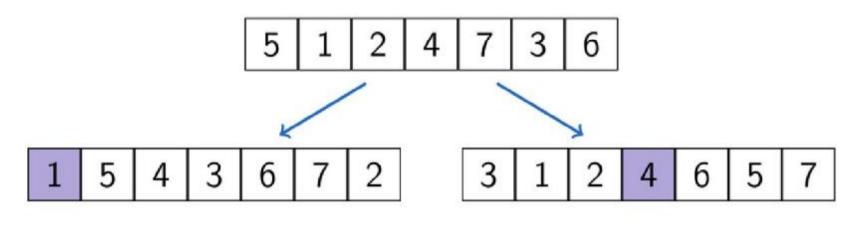
Кафедра экономической информатики Бгуир 2017

17

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Идея доказательства: сравнения

- время работы пропорционально количеству сравнений
- сбалансированные разбиения лучше, потому что они лучше уменьшают количество необходимых сравнений:



узнали только, что 1 — минимум

3,1,2 уже никогда не сравнятся с 6,5,7

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Время работы

Теорема

Допустим, что все элементы массива $A[1 \dots n]$ различны и что разделитель всегда выбирается равномерно случайным образом. Тогда среднее время работы алгоритма QuickSort(A) есть $O(n \log n)$, в то время как время работы в худшем случае есть $O(n^2)$.

Замечание

Усреднение берётся по случайным числам алгоритма, а не по входам.

Кафедра экономической информатики Бгуир 2017

18

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Идеи доказательства: вероятность

$$\mathsf{Prob}\,(1$$
 и 9 сравнятся $)=rac{2}{9}$

 $\mathsf{Prob}\left(\mathsf{3}\;\mathsf{u}\;\mathsf{4}\;\mathsf{сравнятся}\right)=1$

Доказательство

 \blacksquare для i < j положим

$$\chi_{ij} = egin{cases} 1 & A'[i] \ u \ A'[j] \ \text{сравнились} \ 0 & \text{в противном случае} \end{cases}$$

- **при временты** A'[i] и A'[j] либо сравниваются ровно один раз, либо не сравниваются вообще (мы всегда сравниваем с разделителем)
- lacktriangle это, в частности, означает, что время работы в худшем случае не больше $O(n^2)$

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Равные элементы

■ визуализация: www.sorting-algorithms.com/quick-sort

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Доказательство

- важное замечание: $\chi_{ij} = 1$, если и только если первый разделитель, выбранный из $A'[i \dots j]$, это A'[i] или A'[j]
- $lacksymbol{\blacksquare}$ тогда $\mathsf{Prob}(\chi_{ij}) = rac{2}{j-i+1}$ и $\mathsf{E}(\chi_{ij}) = rac{2}{j-i+1}$
- тогда среднее время работы есть

$$E \sum_{i=1}^{n} \sum_{j=i+1}^{n} \chi_{ij} = \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(\chi_{ij})$$

$$= \sum_{i < j} \frac{2}{j-i+1}$$

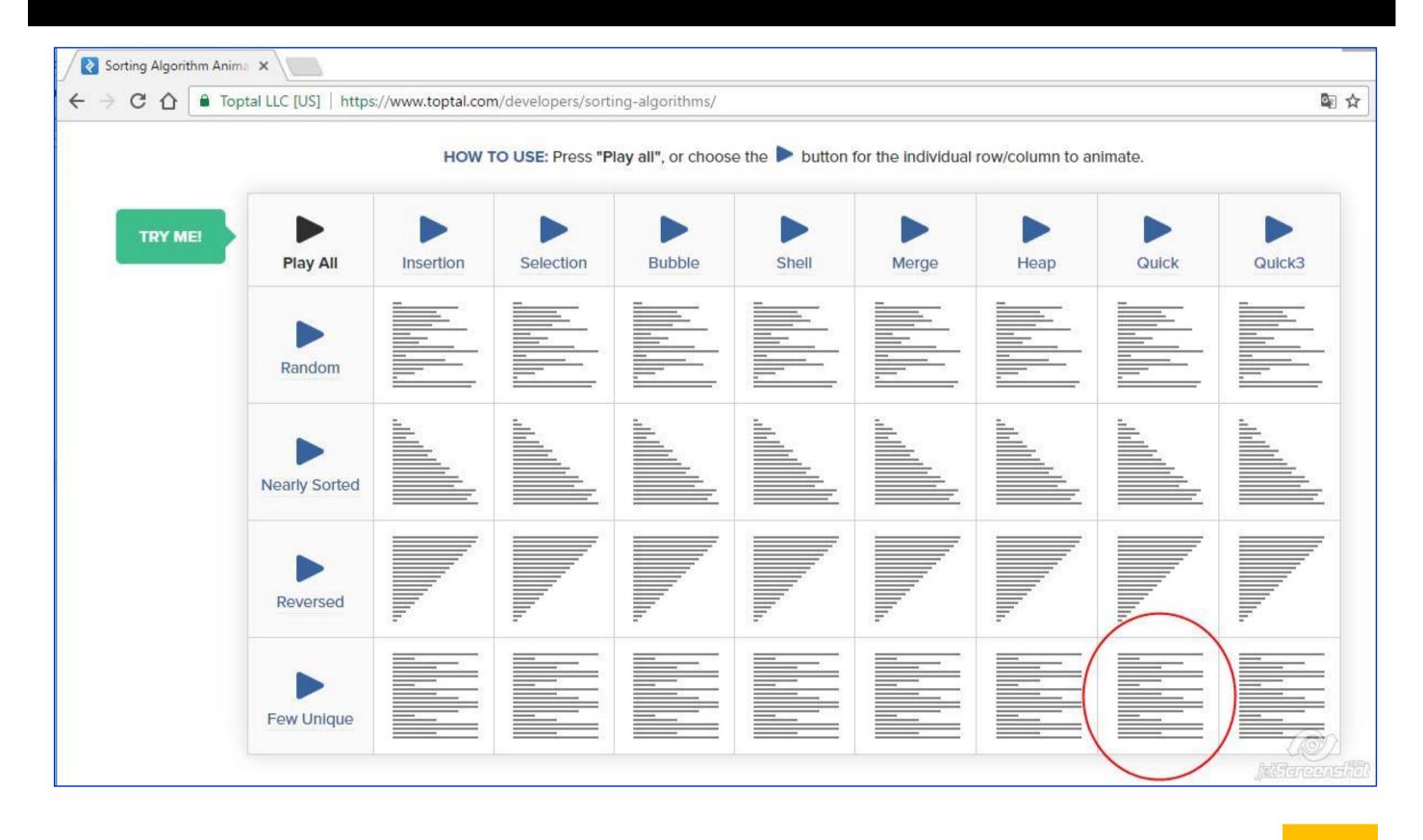
$$\leq 2n \cdot \left(\frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}\right)$$

$$= \Theta(n \log n) \square$$

Кафедра экономической информатики Бгуир 2017

22

PISL05. Quick Sort, Heap Sort, Count Sort etc.



Равные элементы

- визуализация: www.sorting-algorithms.com/quick-sort
- если все элементы массива равны между собой, то рассмотренная реализация алгоритма быстрой сортировки будет работать квадратичное время

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Элиминация хвостовой рекурсии

```
Процедура QUICKSORT(A, \ell, r)
```

```
пока \ell < r:
   m \leftarrow \text{PARTITION}(A, \ell, r)
   QUICKSORT(A, \ell, m-1)
   \ell \leftarrow m+1
```

Элиминируя рекурсивный вызов для более длинного массива, мы гарантируем, что глубина рекурсии (а значит, и дополнительная память) будет в худшем случае не более $O(\log n)$.

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Равные элементы

- визуализация: www.sorting-algorithms.com/quick-sort
- если все элементы массива равны между собой, то рассмотренная реализация алгоритма быстрой сортировки будет работать квадратичное время
- чтобы обойти это препятствие, массив можно разбивать на три части вместо двух: < x, = x и > x(3-разбиение)

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Интроспективная сортировка: $O(n \log n)$ в худшем случае

- запускает быструю сортировку с простой эвристикой выбора разделителя (например, медиана из первого, среднего и последнего элементов)
- \blacksquare если глубина рекурсии превышает порог $c \log n$, быстрая сортировка прерывается и запускается алгоритм с гарантированным временем $O(n \log n)$ в худшем случае (например, сортировка кучей)

Заключение

- Быстрая сортировка работает за время $O(n \log n)$ в среднем случае и за $O(n^2)$ в худшем случае.
- Усреднение берётся по случайным числам алгоритма, но не по входам.
- Очень эффективен на практике.
- Если в массиве могут быть равные числа, стоит использовать 3-разбиение вместо 2-разбиения.
- \blacksquare Элиминация хвостовой рекурсии позволяет сделать так, чтобы алгоритм быстрой сортировки использовал не более $O(\log n)$ дополнительной памяти.

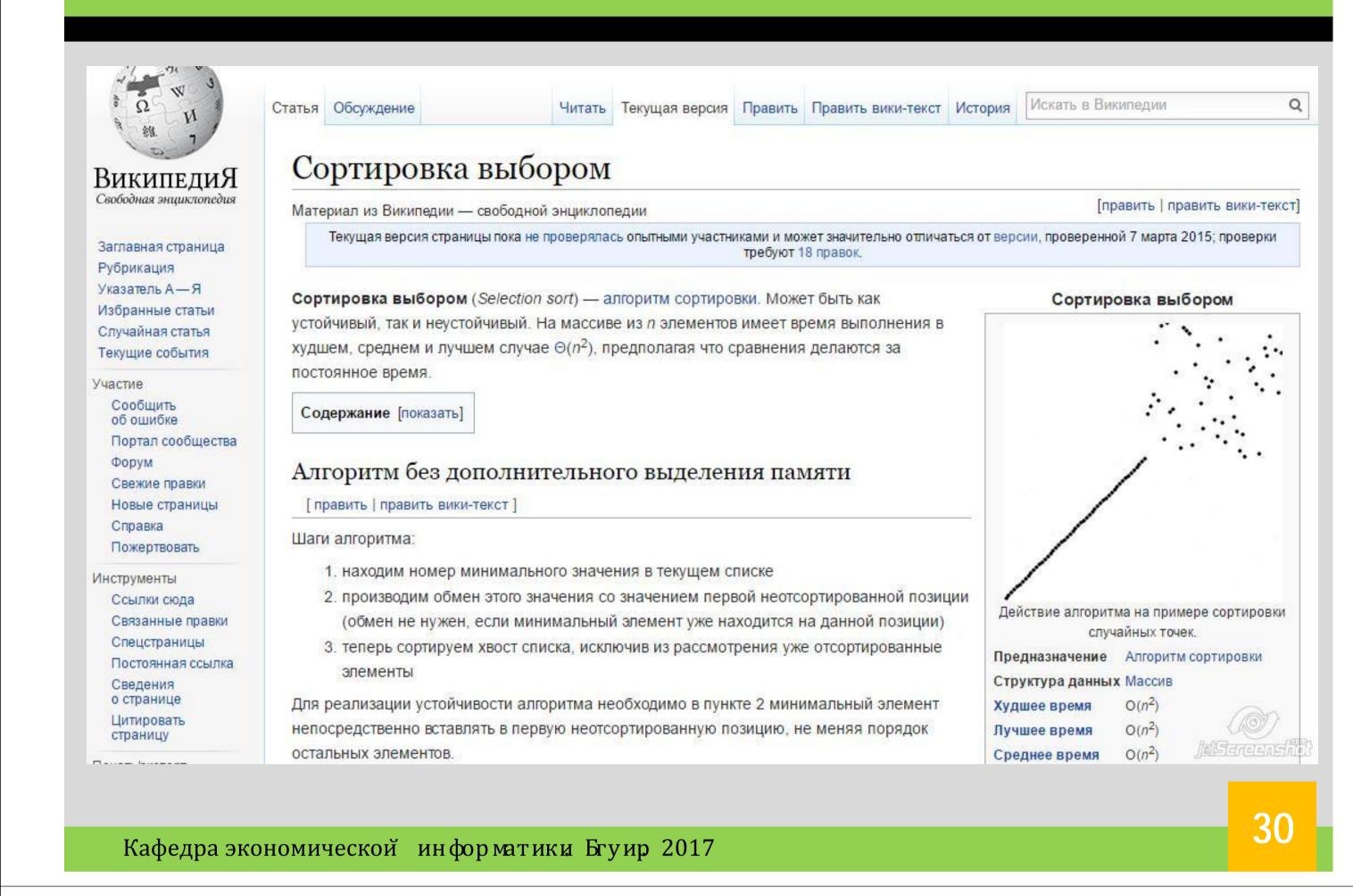
Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка выбором

```
Процедура SELECTIONSORT(A[1 \dots n])
для i от 1 до n:
k \leftarrow i
для j от i+1 до n:
если A[j] < A[k]:
k \leftarrow j
обменять A[i] и A[k]
```

PISL04. Selection sort



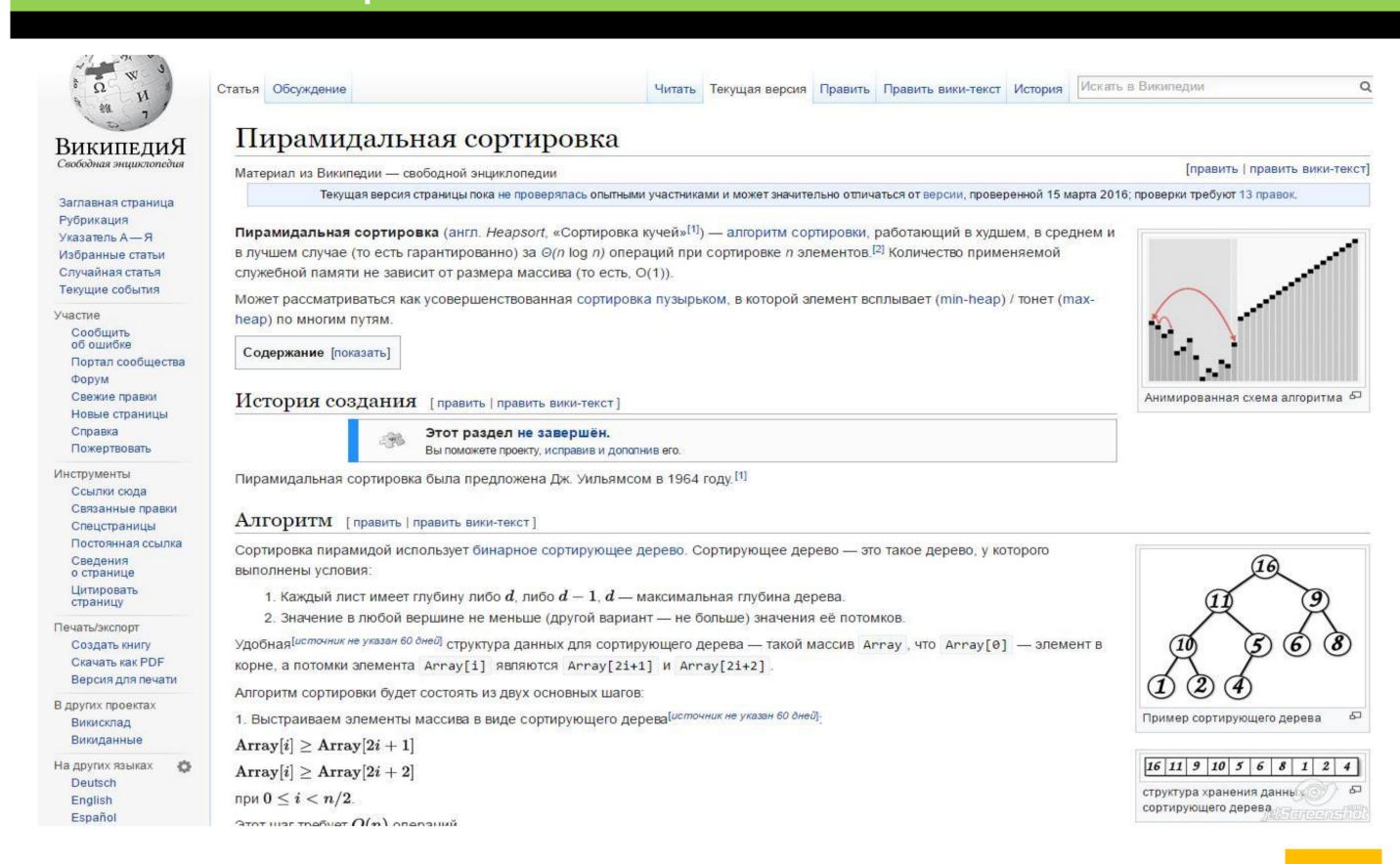
PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка выбором

```
Процедура SELECTIONSORT(A[1 \dots n])
для i от 1 до n:
k \leftarrow i
для j от i+1 до n:
если A[j] < A[k]:
k \leftarrow j
обменять A[i] и A[k]
```

Время работы: $O(n^2)$.

PISL04. Heap sort



Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка кучей

```
Процедура HEAPSORT(A[1...n])

H \leftarrow \{\} {мин-куча} для i от 1 до n:
   INSERT(H, A[i])

для i от 1 до n:
   A'[i] \leftarrow \text{ExtractMin}(H)
```

Время работы: $O(n \log n)$.

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка кучей

```
Процедура HEAPSORT(A[1...n])

H \leftarrow \{\} {мин-куча}
для i от 1 до n:
   INSERT(H, A[i])
для i от 1 до n:
   A'[i] \leftarrow \text{ExtractMin}(H)
```

Кафедра экономической информатики Бгуир 2017

3

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка кучей на месте

```
Процедура HEAPSORT(A[1 \dots n])

Виіldмахнеар(A)

size \leftarrow n

для i от n до 2:

обменять A[size] и A[1]

size \leftarrow size - 1

SiftDown(A, 1)
```

Замечание

Переменная size используется внутри SIFTDOWN: куча располагается в подмассиве $A[1 \dots size]$.

Построение кучи

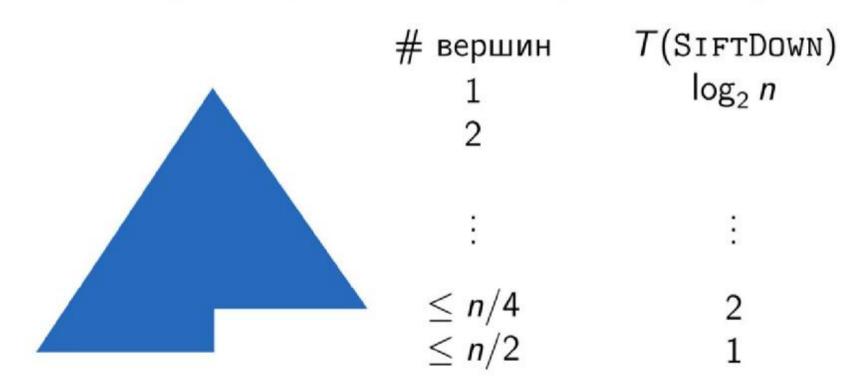
Процедура BUILDMAXHEAP(A[1...n])

для i от $\lfloor n/2 \rfloor$ до 1: SIFTDOWN(A, i)

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Время работы построения кучи

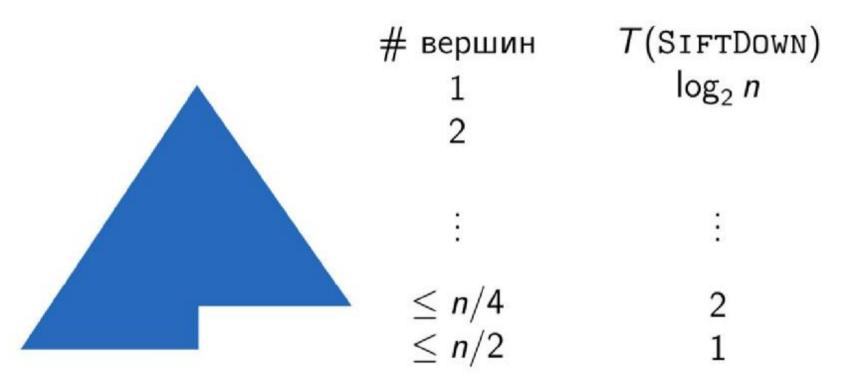


$$T(BUILDHEAP) \le \frac{n}{2} \cdot 1 + \frac{n}{4} \cdot 2 + \frac{n}{8} \cdot 3 + \cdots$$

 $\le n \cdot \sum_{i=1}^{\infty} \frac{i}{2^i} = 2n$

PISL05. Quick Sort, Heap Sort, Count Sort etc.

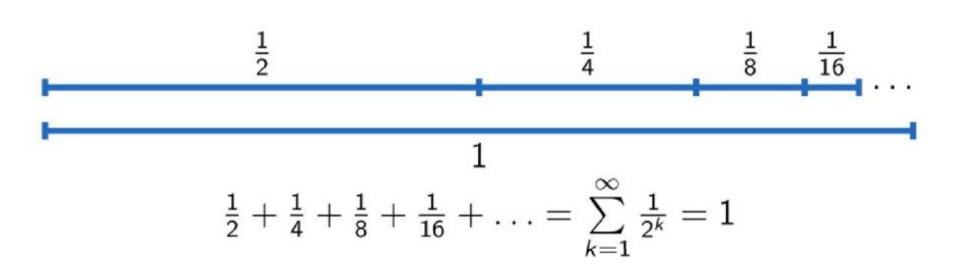
Время работы построения кучи



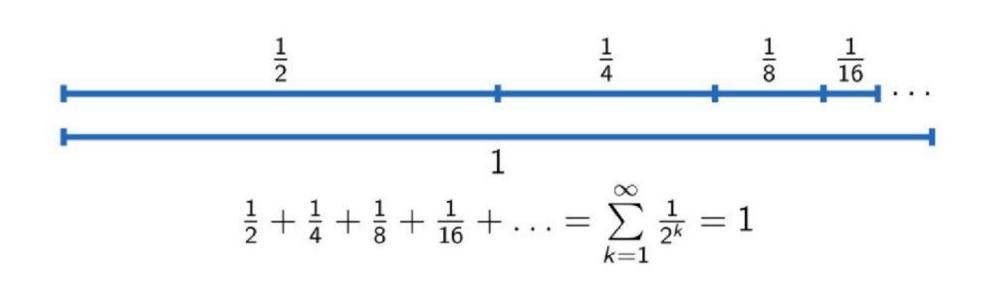
Кафедра экономической информатики Бгуир 2017

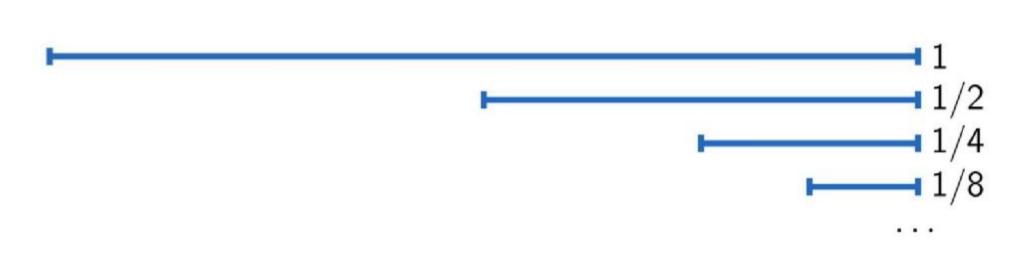
PISL05. Quick Sort, Heap Sort, Count Sort etc.

Оценка суммы



Оценка суммы





Кафедра экономической информатики Бгуир 2017

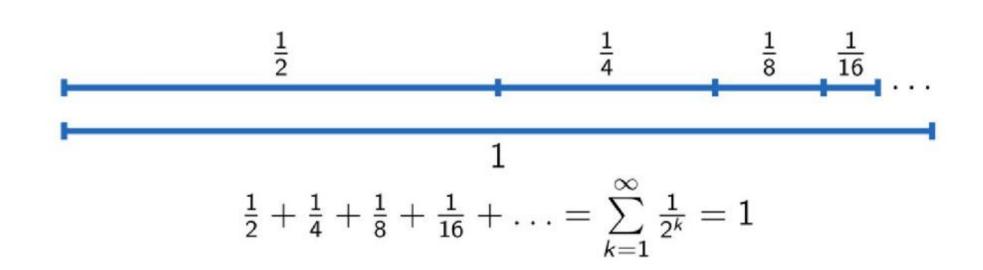
PISL05. Quick Sort, Heap Sort, Count Sort etc.

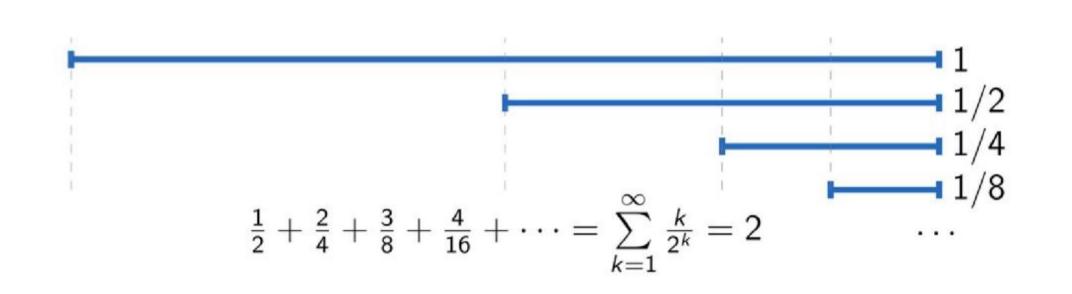
Заключение

■ Время работы сортировки кучей: $O(n \log n)$.

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Оценка суммы





Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Заключение

- Время работы сортировки кучей: $O(n \log n)$.
- Время работы построения кучи: O(n).

Заключение

- Время работы сортировки кучей: $O(n \log n)$.
- Время работы построения кучи: O(n).
- Сортирует на месте.

Кафедра экономической информатики Бгуир 2017

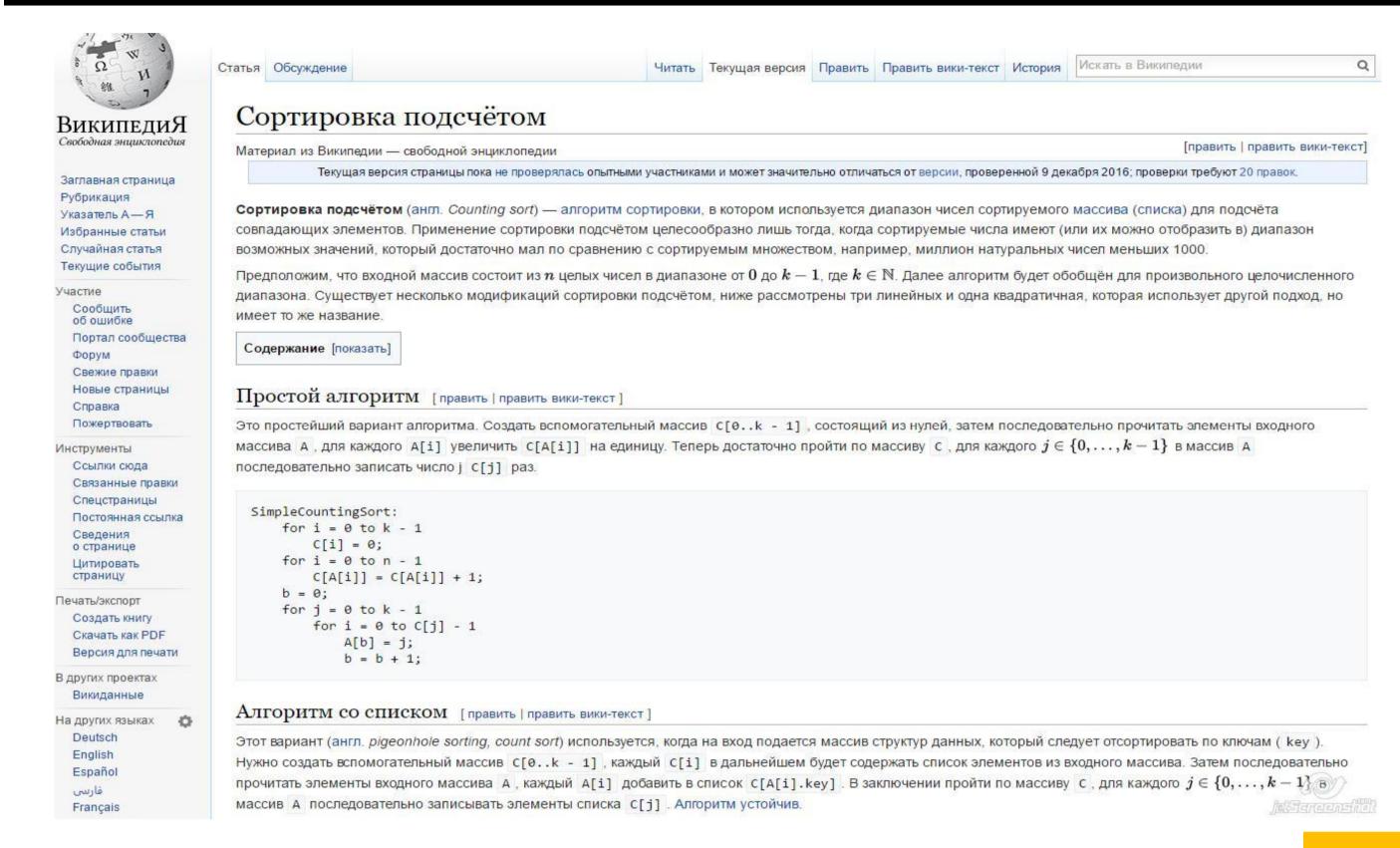
PISL05. Quick Sort, Heap Sort, Count Sort etc.

Сортировка подсчётом

Сортировка небольших по величине целых чисел:

2 1 1 1 3 2 2 2 3 2 3 2 1 1

PISL04. Count sort



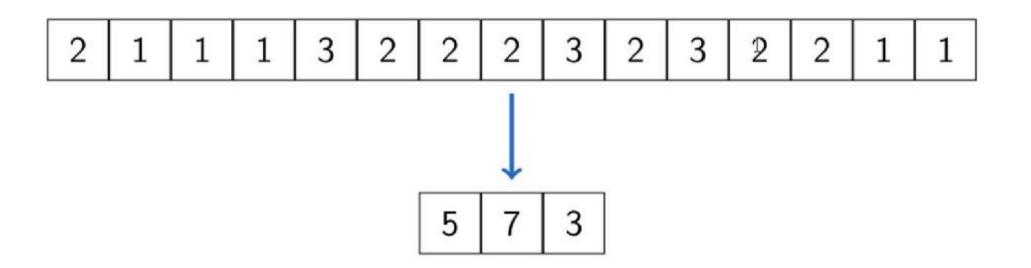
Кафедра экономической информатики Бгуир 2017

46

PISL05. Quick Sort, Heap Sort, Count Sort etc.

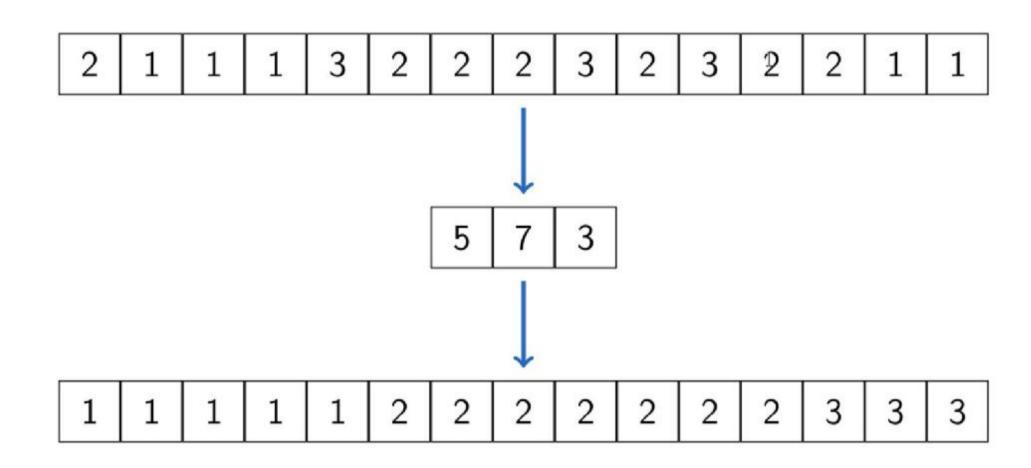
Сортировка подсчётом

Сортировка небольших по величине целых чисел:



Сортировка подсчётом

Сортировка небольших по величине целых чисел:



Кафедра экономической информатики Бгуир 2017

49

PISL05. Quick Sort, Heap Sort, Count Sort etc.

```
public static int[] sort(int[] array) {
     int min, max = min = array[0];
     // тупо находим максимальное и минимальное значение
     for (int i = 1; i < array.length; i++) {</pre>
                                                              static int[] sort(int[] array, int min, int max) {
          if (array[i] < min) {</pre>
                                                                  // счетчик это такой массив в котором мы будем считать, как часто встречаются
                   min = array[i];
                                                                  // числа в сортируемом массиве.
                                                                  // donycmum\ maccub\ pabeh = \{0,2,1,5,1\},\ min = 0,\ max = 5
                                                                  // счетчик = count[0]....count[5]
          if (array[i] > max) {
                                                                  int[] count = new int[max - min + 1];
               max = array[i];
                                                                  for (int i = 0; i < array.length; i++) {</pre>
                                                                      // подсчитываем сколько раз встречается число,
                                                                      // встретилось +1 к счетчику
      // понеслась
                                                                      count[array[i] - min]++;
     return sort(array, min, max);
                                                                   // например. count[0]=1, count[1]=2, count[3]=0, count[4]=0, count[5]=1
                                                                  int idx = 0;
                                                                  // теперь все готово
                                                                  // пробегаем по всему счетчику (от 0 до 5)
                                                                  // count[i] - показывает сколько раз встречается то или иное число
                                                                  for (int i = 0; i < count.length; i++) {</pre>
                                                                      // count[0]=1, значит array[0]=0;
public static void main(String[] args) {
                                                                      // count[1]=2, значит вставляем два раза array[1]=array[2]=1;
   Random rnd = new Random();
                                                                      // count[2]=1, опять только один раз. array[3]=2;
   int arr1[] = new int[1024*1024];
                                                                      // count[3]=0, значит ничего не вставляем и т.д.
   for (int i = 0; i < arr1.length; ++i) {</pre>
                                                                      for (int j = 0; j < count[i]; j++) {
      arr1[i] = rnd.nextInt(1000) + 123;
                                                                           array[idx++] = i + min;
   int arr2[] = new int[arr1.length];
   System.arraycopy(arr1, 0, arr2, 0, arr1.length);
   long t1 = System.currentTimeMillis();
                                                                  // ну собственно и всё
   sort(arr1);
                                                                   return array;
   long t2 = System.currentTimeMillis();
   Arrays.sort(arr2);
   long t3 = System.currentTimeMillis();
   System.out.printf("counting sort: %d merge sort: %d \r\n", (t2-t1), (t3-t2))
                                                                         http://programador.ru/sorting-positive-int-linear-time/
   System.out.println(Arrays.equals(arr1, arr2));
```

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Стабильная сортировка подсчётом

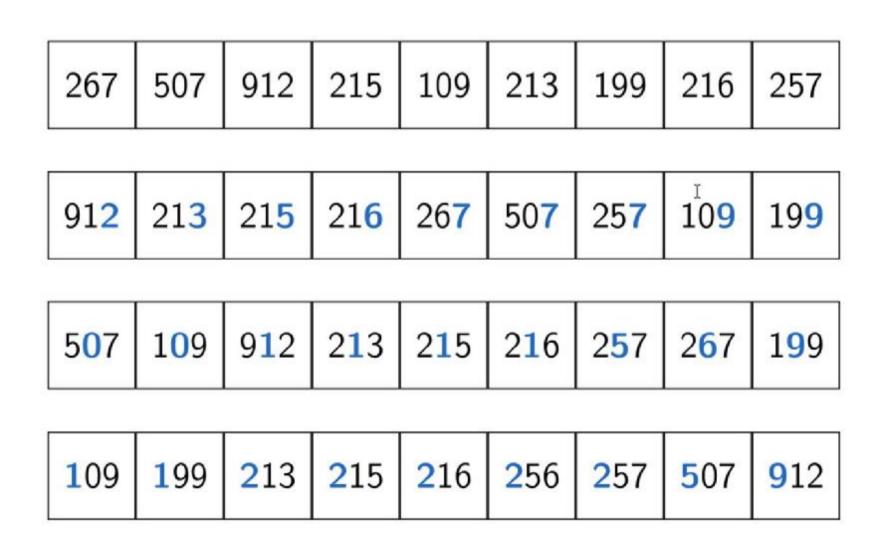
```
Процедура COUNTSORT(A[1...n])
\{массив A содержит целые числа от 1 до M\}
создать массив B[1...M] \leftarrow [0,0,...,0]
для j от 1 до n:
  B[A[j]] \leftarrow B[A[j]] + 1
для i от 2 до M:
  B[i] \leftarrow B[i] + B[i-1]
для j от n до 1:
  A'[B[A[j]]] \leftarrow A[j]
  B[A[j]] \leftarrow B[A[j]] - 1
```

Время работы: O(n+M).

Кафедра экономической информатики Бгуир 2017

PISL05. Quick Sort, Heap Sort, Count Sort etc.

Цифровая сортировка



Время работы: O(nd), где d — число разрядов во входных числах.

Заключение

- Алгоритм сортировки подсчётом сортирует массив размера n, содержащий целые числа от 1 до M, за время O(n+M).
- Нижняя оценка $\Omega(n \log n)$ не применима \hat{k} алгоритму сортировки подсчётом, поскольку он основан не на сравнениях.
- Алгоритм сортировки подсчётом можно реализовать так, чтобы он был стабильным: равные элементы будут идти в том же порядке, в каком они были в исходном массиве.
- Алгоритмом цифровой сортировки можно отсортировать числа с d разрядами за время O(nd).

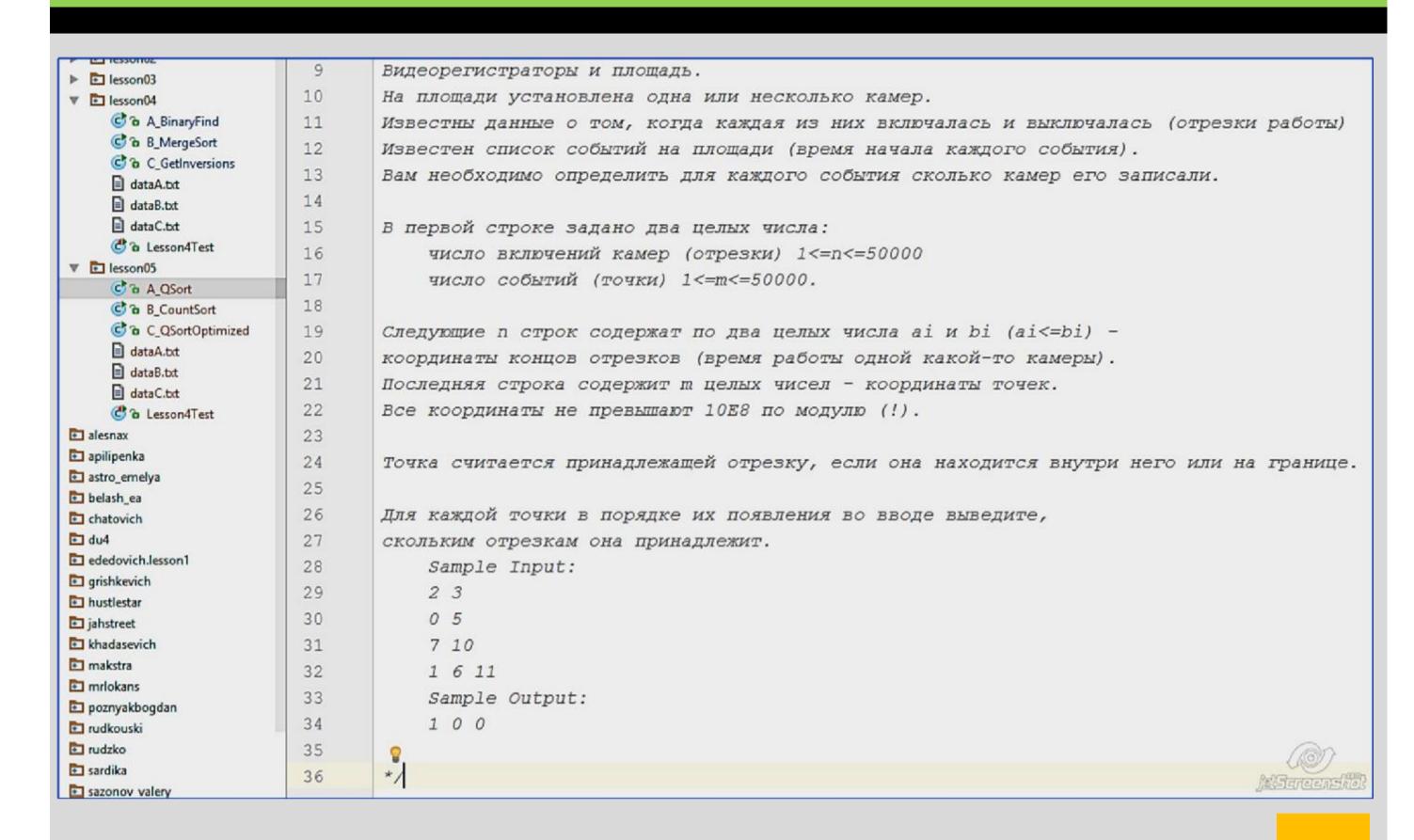
Кафедра экономической информатики Бгуир 2017

53

Задание Б.

```
lesson02
                             +import ...
 lesson03
 lesson04
                             1*
    C & A_BinaryFind
    C & B MergeSort
                              Первая строка содержит число 1<=n<=10000, вторая - n натуральных чисел, не превышающих 10.
    C & C_GetInversions
                     11
                              Выведите упорядоченную по неубыванию последовательность этих чисел.
    dataA.txt
                     12
    dataB.txt
                     13
    dataC.txt
                              При сортировке реализуйте метод со сложностью O(n)
    & b Lesson4Test
  lesson05
                     15
                              Пример: https://karussell.wordpress.com/2010/03/01/fast-integer-sorting-algorithm-on/
    C & A QSort
                             Вольный перевод: http://programador.ru/sorting-positive-int-linear-time/
    C & B_CountSort
    C & C_QSortOptimized
    dataA.txt
    dataB.txt
                              public class B CountSort {
    dataC.txt
    & Lesson4Test
alesnax
                      21
apilipenka
                     22
                                  int[] countSort(InputStream stream) throws FileNotFoundException {
astro_emelya
                                       //подготовка к чтению данных
belash_ea
                                       Scanner scanner = new Scanner(stream);
chatovich
du4
                                                                                               11111111111111111111111111111111
                                       //!!!!!!!!!!!!!!!!!!!!!!!!!!!!
                                                                           НАЧАЛО ЗАЛАЧИ
                      26
                                       //размер массива
grishkevich
                                      int n = scanner.nextInt();
hustlestar
                                       int[] points=new int[n];
iahstreet
khadasevich
makstra makstra
                                       //читаем точки
                                       for (int i = 0; i < n; i++) {
poznyakbogdan
                                           points[i]=scanner.nextInt();
rudkousk
rudzko 🗀
                      33
sardika
                      34
                                       //тут реализуйте логику задачи с применением сортировки подсчетом
sazonov_valery
```

Задание А.



Кафедра экономической информатики Бгуир 2017

54

Задание С.

```
package by.it.a khmelov.lesson05;
▼ 🛅 a_khmelov
  lesson01
  lesson02
                               +import ...
  lesson03
  ▼ lesson04
      C & A_BinaryFind
      € B_MergeSort
                                Видеорегистраторы и площадь 2.
      C & C_GetInversions
                                Условие то же что и в задаче А.
      dataA.txt
      dataB.txt
      dataC.txt
                                         По сравнению с задачей А доработайте алгоритм так, чтобы
      C & Lesson4Test
                                         1) он оптимально использовал время и память:
  ▼ lesson05
                                              - за стек отвечает элиминация хвостовой рекурсии,
      C & A_QSort
                                              - за сам массив отрезков - сортировка на месте
      C & B_CountSort
      C_QSortOptimized
                                              - рекурсионные вызовы должны проводится на основе 3-разбиения
      dataA.txt
      dataB.txt
                                        2) при поиске подходящих отрезков для точки реализуйте метод бинарного поиска,
      dataC.txt
                                         помните при реализации, что поиск множественный
      C & Lesson4Test
 alesnax
                                          (т.е. отрезков, подходящих для точки, может быть много)
 apilipenka
 astro_emelya
                                    Sample Input:
 belash_ea
                                    2 3
 chatovich
 du4
                                    0 5
 ededovich.lesson1
                                    7 10
 grishkevich
                                    1 6 11
 hustlestar
                                    Sample Output:
 iahstreet
 khadasevich
                                    1 0 0
 makstra
 mrlokans
 poznyakbogdan
 rudkouski
rudzko
```