

АЛГОРИТМЫ СОРТИРОВКИ ЧАСТЬ 2

Лекция 4

Иванов Г.В.

Сортировки 2

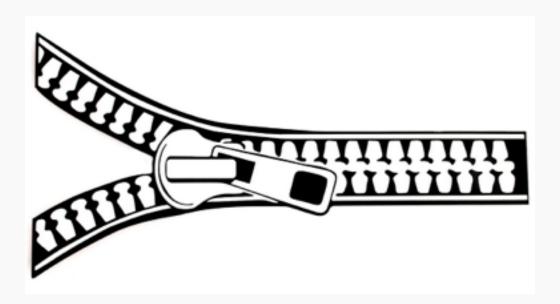


- 1. Сортировка слиянием
- 2. Быстрая сортировка
- 3. Порядковые статистики
- 4. Сортировки без сравнений

Слияние 2х упорядоченных массивов «



- Выберем массив, крайний элемент которого меньше
- Извлечём этот элемент в массив-результат
- Продолжаем, пока один из массивов не опустеет
- ❖ Копируем остаток второго массива в конец массива-результата





```
void merge(int *a, int a len, int *b, int b len, int *c) {
  int i=0; int j=0;
  for (;i < a len and j < b len;) {</pre>
    if (a[i] < b[j]) {
      c[i+j] = a[i];
     ++i;
    } else {
      c[i+j] = b[j];
     ++j;
  if (i==a len) {
    for (; j < b len; ++ j) { c[i+j] = b[j]; }</pre>
  } else {
    for (;i < a len;++i) { c[i+j] = a[i]; }</pre>
```




Сложность *О(n+m)* Количство сравнений

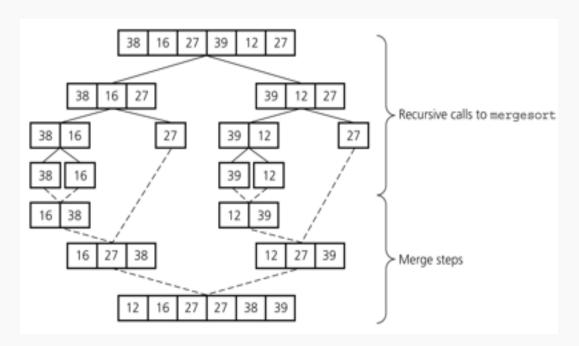
- Лучшее: min(n,m)
- Худшее n+m

Нисходящая сортировка слиянием



- Разбить массив на 2 части
- Отсортировать каждую часть рекурсивно

♦ Объединить 1 и 2 части



Нисходящая сортировка слиянием



```
void merge_sort(int *data, int size, int *buffer) {
  if (size < 2) return;
  merge_sort(data, size / 2, buffer);
  merge_sort(&data[size / 2], size - size / 2, buffer);

  merge(&data[0], size / 2, &data[size/2], size - size / 2, buffer);

  for (size_t pos = 0; pos < size; ++ pos) {
    data[pos] = buffer[pos];
  }
}</pre>
```

Восходящая сортировка слиянием



- ◆ Разбить массив на 2^k частей размером не больше m.
- ♦ Отсортировать каждую часть другим алгоритмом
- ❖ Объединить 1 и 2, 3 и 4, ... n-1 и n части.
- ♦ Повторить шаг 3, пока не останется одна часть

1	X	В	R	S	T	U	Α	C	N	P	D
2	В	X	R	S	T	U	A	C	N	P	D
4	В	R	S	Χ	Α	С	Т	U	D	N	Р
8	Α	В	С	R	S	Т	U	Χ	D	N	Р
16	A	В	С	D	N	Р	R	S	Т	U	Χ

Восходящая сортировка слиянием



```
void merge sort(int *data, size t size, int *buffer) {
  for(size t chunk size = 1; chunk size < size; chunk size *= 2) {</pre>
    size t offset = 0;
    for (; offset + chunk size < size; offset += 2 * chunk size) {</pre>
      size t right size = chunk size;
      if (offset + chunk size + right size > size) {
        right size = size - offset - chunk size;
      merge(
        &data[offset], chunk size,
        &data[offset + chunk size], right size,
        &buffer[offset]);
    for(size t pos = 0; pos < size; ++pos) {</pre>
      data[pos] = buffer[pos];
```

Восходящая сортировка слиянием



```
void merge sort fast(int *data, size t size, int *buffer) {
  bool is swapped = false;
  for(size t chunk size = 1; chunk size < size; chunk size *= 2, is swapped = !is swapped) {</pre>
    size t offset = 0;
    for (; offset + chunk size < size; offset += 2 * chunk size) {</pre>
      size t right size = chunk size;
      if (offset + chunk size + right size > size) {
        right size = size - offset - chunk size;
      merge(
        &data[offset], chunk size,
        &data[offset + chunk size], right size,
        &buffer[offset]);
    for (size t pos = offset; pos < size; ++pos) {</pre>
      buffer[pos] = data[pos];
    std::swap(data, buffer);
  if (is swapped) {
    std::swap(data, buffer);
    for(size t pos = 0; pos < size; ++pos) {</pre>
      data[pos] = buffer[pos];
```

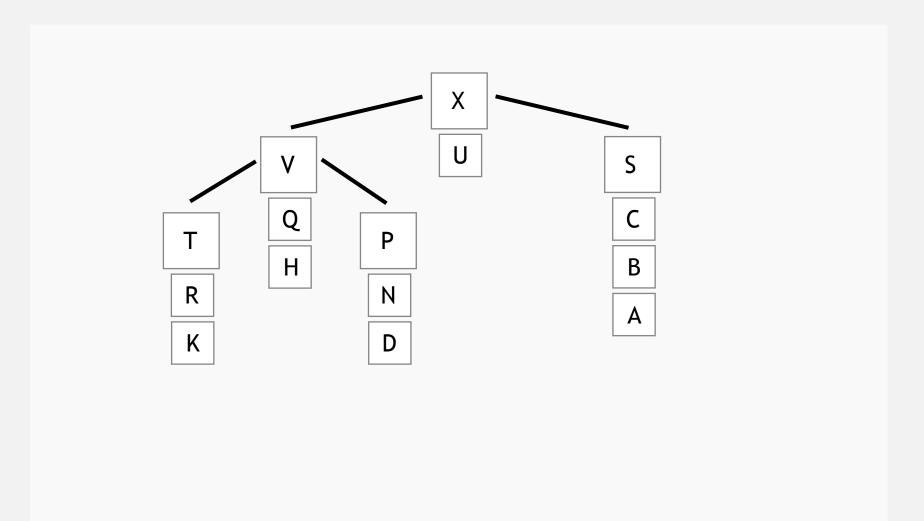


Дано k упорядоченных массивов суммарным размером n: A1, A2, .. Ak

- \bullet Построить кучу из k массивов O(k)
- ❖ Перенести первый элемент из вершины кучи в результат O(1)
- Если массив на вершине кучи пуст извлечь элемент из кучи
- ❖ Восстановить порядок массивов в куче O(log(k))
- Повторить пока куча не пуста

Суммарная сложность: O(k + n*log(k))







- ❖ Построим бинарное дерево с массивами А1..Ак в листьях
- В узловых вершинах будем хранить указатель на минимальный элемент поддерева
- Изъятие элемента из узловой вершины изъятие из минимального из дочерних узлов
- При изъятии элемента из списка, его размер уменьшается на 1
- Если список пуст его сосед перемещается на место родительской узловой вершины
- При изъятии происходит одно сравнение на каждом уровне дерева

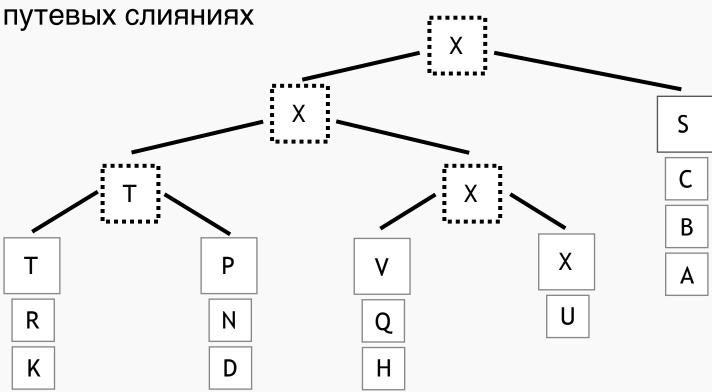
Высота дерева *log(k)*

Итого O(n*log(k))



Экономия на операциях копирования: п

Количество сравнений такое же как и при log(k) 2x



Quick sort



QuickSort: Split



- ❖ Установим 2 указателя: i в − начало массива, j − в конец
- ❖ Будем помнить под каким указателем лежит «пивот»
- ❖ Если a[j] > a[i], поменяем элементы массива под i, j
- ❖ Сместим на 1 указатель, не указывающий на «пивот»
- ❖ Продолжим пока *i* != *j*

QuickSort: Split



i									j					
Ε	F	С	В	S	Н	Z	D	Α	G					
Е	F	С	В	S	Н	Z	D	Α	G					
Α	F	С	В	S	Н	Z	D	Ε	G					
Α	Е	С	В	S	Н	Z	D	F	G					
Α	D	С	В	S	Н	Z	Ε	F	G					
Α	D	С	В	S	Н	Z	Ε	F	G					
Α	D	С	В	S	Н	Z	Е	F	G					
Α	D	С	В	Ε	Н	Z	S	F	G					
Α	D	С	В	Ε	Н	Z	S	F	G					
Α	D	С	В	Ε	Н	Z	S	F	G					

Сортировка Хоара



```
void quick sort(int *a, int n) {
  int i = 0;
  int j = n - 1;
  bool side = 0;
  while (i != j) {
    if (a[i] > a[j]) {
      swap(a[i], a[j]);
      side = !side;
    if (side) {
      ++i;
    } else{
      --i;
  if (i > 1) quick sort(a, i);
  if (n > i+1) quick sort(a + (i+1), n - (i+1));
```

 $ext{vocuse}$

Quicksort: анализ



Предположим, что split делит массив в соотношении 1:1

■
$$T(n) \le c_1 + c_2 n + T\left(\frac{n}{2}\right) + T\left(\frac{n}{2}\right) =$$

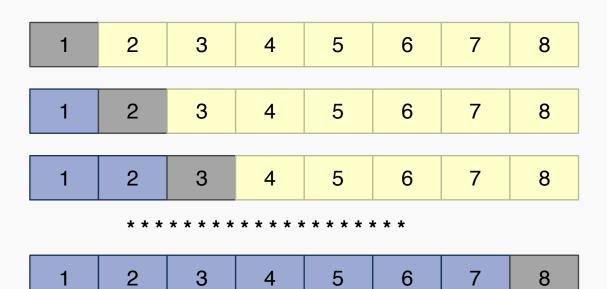
$$= \sum_{k=0}^{\log_2 n} \{c_1 2^k + c_2 n\} = c_1 n + c_2 n \log(n)$$

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

Quicksort: анализ



- Упорядоченный массив делится в соотношении 1:n-1
- $T(n) = c_1 + c_2 n + T(n-1) = \frac{1}{2}c_2 n^2 + c_1 n = O(n^2)$



Quicksort: выбор пивота



1й элемент

Серединный элемент

Медиана трёх

Случайный элемент

Медиана

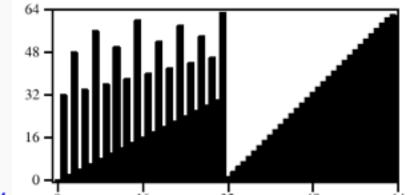
Медиана по трём случайным

. . .

Quicksort-killer



- Последовательность, приводящая к времени: $T(n) = O(n^2)$
- [1,2,3,...п] ⇔ первый элемент
- Для любого предопределённого порядка выбора пивота, существует killer-последовательность



http://www.cs.ca.i6...oaci32.caa, 48.oag, i64.dmspe.pdf

Рандомизованная медиана



Медиана⇔ *RandSelect(A[1, N], k)*; *k = N/2*

Случайно выберем элемент из А: А[j]

Разобьём **А** на 2 части: меньше/больше **А[j]**

Пусть позиция **А[j]** в разделённом массиве: **k**

k < j?

- Найдём: *RandSelect(A[1, j], k)*
- Иначе: RandSelect(A[j+1, N], k-j)

Quicksort: медиана за линейное время «



Медиана *⇔ SELECT(A[1,N], k)*; *k = N/2*

Разобьём массив на пятёрки

Отсортируем каждую пятёрку

Найдём медиану середин пятёрок

Разобьём массив на 2 группы: меньше/больше медианы пятёрок

Пусть индекс медианы в массиве ј

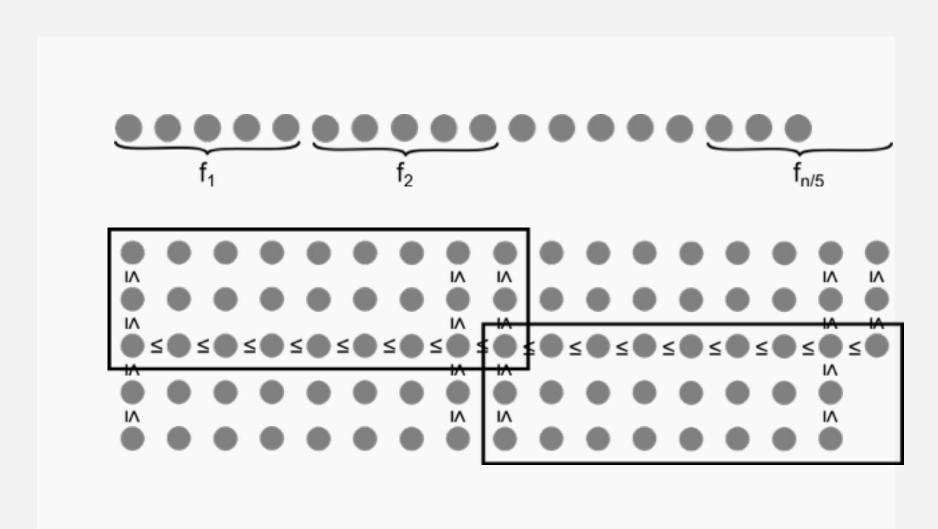
$$j > k$$
?

■ найдём: *SECLECT(A[1, j], k)*

■ иначе: *SELECT(A[j+1, N], k-j)*

Медиана за линейное время





Медиана за линейное время: анализ «



- Разобьём массив на пятёрки
- Отсортируем каждую пятёрку $c_1 \mathit{N}$
- Найдём медиану середин пятёрок T(N/5)
- Разобьём массив на 2 группы <> медианы медиан: $c_2 N$
 - найдём: SECLECT(A[1, j], k) => T(j)
 - иначе: SELECT(A[j+1, N], k-j) => T(N-j); $0.3N \le j \le 0.7N$;
- $T(N) \le T(\frac{N}{5}) + cN + T(0.7N); \Rightarrow T(N) = O(N);$

Сравнение сортировок



Name ◆	Best ◆	Average •	Worst ¢	Memory ¢	Stable •				
Quicksort	$n \log n$	$n \log n$	n^2	$\log n$ on average, worst case is n	typical in-place sort is not stable; stable versions exist				
Merge sort	$n \log n$	$n \log n$	$n \log n$	Depends further explanation needed, worst case is $n \\$	Yes				
In-place merge sort	-	-	$n (\log n)^2$	1	Yes				
Heapsort	$n \log n$	$n \log n$	$n \log n$	1	No				
Insertion sort	n	n^2	n^2	1	Yes				
Introsort	$n \log n$	$n \log n$	$n \log n$	$\log n$	No				
Selection sort	n^2	n^2	n^2	1	No				
Timsort	n	$n \log n$	$n \log n$	n	Yes				
Shell sort	n	$n(\log n)^2$ or $n^{3/2}$	Depends on gap sequence; best known is $n(\log n)^2$	1	No				
Bubble sort	n	n^2	n^2	1	Yes				
Binary tree sort	n	$n \log n$	$n \log n$	n	Yes				
Cycle sort	-	n^2	n^2	1	No				
Library sort	_	$n \log n$	n^2	n	Yes				
Patience sorting	_	-	$n \log n$	n	No				
Smoothsort	n	$n \log n$	$n \log n$	1	No				
Strand sort	n	n^2	n^2	n	Yes				
Tournament sort	_	$n \log n$	$n \log n$	$n^{[4]}$					
Cocktail sort	n	n^2	n^2	1	Yes				
Comb sort	n	$n \log n$	n^2	1	No				
Gnome sort	n	n^2	n^2	1	Yes				
Franceschini's method ^[5]	_	$n \log n$	$n \log n$	1	Yes				

Сортировки без сравнений



- 1. Сортировка подсчетом
- 2. Поразрядная сортировка от младших к старшим
- 3. Поразрядная сортировка от старших к младшим
- 4. Поразрядная быстрая сортировка



```
Algorithm 3.10: CountingSort(R)

Input: (Multi)set R = \{k_1, k_2, \dots k_n\} of integers from the range [0..\sigma).

Output: R in nondecreasing order in array J[0..n).

(1) for i \leftarrow 0 to \sigma - 1 do C[i] \leftarrow 0

(2) for i \leftarrow 1 to n do C[k_i] \leftarrow C[k_i] + 1

(3) sum \leftarrow 0

(4) for i \leftarrow 0 to \sigma - 1 do // cumulative sums

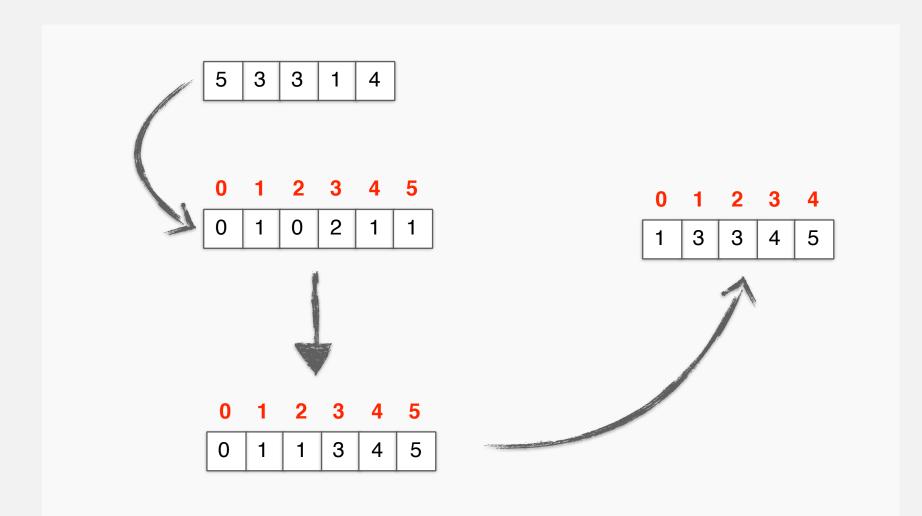
(5) tmp \leftarrow C[i]; C[i] \leftarrow sum; sum \leftarrow sum + tmp

(6) for i \leftarrow 1 to n do // distribute

(7) J[C[k_i]] \leftarrow k_i; C[k_i] \leftarrow C[k_i] + 1

(8) return J
```





 $ext{vocuse}$



```
void count sort(int *data, int size, int range) {
  int *count = new int[range];
  int *aux = new int[size];
  for (int i = 1; i < range; ++i) {</pre>
    count[i] = 0;
  for (int i = 0; i < size; ++i) {</pre>
    ++count[data[i] + 1];
  for (int i = 1; i < range; ++i) {</pre>
    count[i] += count[i - 1];
  for (int i = 0; i < size; ++i) {</pre>
    aux[count[data[i]]++] = data[i];
  for (int i = 0; i < size; i++) {</pre>
    data[i] = aux[i];
  delete [] aux;
  delete [] count;
```



- + O(n) linear time
- + stable
- + нет сравнений
- требует O(n) памяти
- O(n) перемещений



Algorithm 3.11: LSDRadixSort(\mathcal{R})

Input: Set $\mathcal{R} = \{S_1, S_2, \dots, S_n\}$ of strings of length m over the alphabet $[0..\sigma)$. Output: \mathcal{R} in increasing lexicographical order.

- (1) for $\ell \leftarrow m-1$ to 0 do CountingSort(\mathcal{R}, ℓ)
- (2) return \mathcal{R}



Α	00001	R	10010	Т	10100\ X	11000	Р	10000	Α	00001
S	10011	Т	10100	×	11000 P	10000	Α	00001	Α	00001
0	01111	Ν	01110	P	100001/A	00001	Α	00001	Ε	00101
R	10010	Х	11000	L	011001/1	01001	R	10010	Е	00101
Т	10100	Ρ	10000	Α	000011//A	00001	S	10011	G	00111
1	01001	L	01100	- 1	01001/WR	10010	Т	10100	- 1	01001
N	01110	Α	00001	E	00101VWS		Ε	00101	L	01100
G	00111	S	10011	Α	00001/\/\T	10100	Ε	00101	M	01101
E	00101	0	01111	M	01101\/\)\'L	0 1 1 0 0	G	0 0 1 1 1	N	01110
X	11000	1	01001	E	00101XX,E	00101	X	1 1 0 0 0	0	01111
Α	00001	G	00111	R	10010'\ M	0 1 1 0 1	- 1	0 1 0 0 1	Р	10000
M	01101	Ε	00101	N	01110 E	00101	L	0 1 1 0 0	R	10010
Р	10000	Α	00001	s	10011 N	0 1 1 1 0	M	01101	S	10011
L	01100	Μ	01101	0	01111-0	0 1 1 1 1	N	01110	Т	10100
E	00101	Ε	00101	G	00111 - G	00111	0	01111	X	11000
			_		_					



```
4 void jsw radix pass ( int a[], int aux[], int n, int radix )
 5 {
    int i;
 6
    int count[RANGE] = {0};
 8
 9
    for (i = 0; i < n; i++)
10
      ++count[digit ( a[i], radix ) + 1];
11
12
    for ( i = 1; i < RANGE; i++ )
13
      count[i] += count[i - 1];
14
15
    for (i = 0; i < n; i++)
      aux[count[digit ( a[i], radix )]++] = a[i];
16
17
18
    for (i = 0; i < n; i++)
     a[i] = aux[i];
19
20 }
21
```



- + O(n*key len) время работы
- + O(n) время работы
- + stable
- can't use unstable sort for buckets

MSD raddix sort



```
Algorithm 3.15: MSDRadixSort(\mathcal{R}, \ell)
```

Input: Set $\mathcal{R} = \{S_1, S_2, \dots, S_n\}$ of strings over the alphabet $[0..\sigma)$ and the length ℓ of their common prefix.

Output: R in increasing lexicographical order.

- (1) if $|\mathcal{R}| < \sigma$ then return StringQuicksort(\mathcal{R}, ℓ)
- (2) $\mathcal{R}_{\perp} \leftarrow \{S \in \mathcal{R} \mid |S| = \ell\}; \ \mathcal{R} \leftarrow \mathcal{R} \setminus \mathcal{R}_{\perp}$
- (3) $(\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_{\sigma-1}) \leftarrow \text{CountingSort}(\mathcal{R}, \ell)$
- (4) for $i \leftarrow 0$ to $\sigma 1$ do $\mathcal{R}_i \leftarrow \mathsf{MSDRadixSort}(\mathcal{R}_i, \ell + 1)$
- (5) return $\mathcal{R}_{\perp} \cdot \mathcal{R}_0 \cdot \mathcal{R}_1 \cdots \mathcal{R}_{\sigma-1}$

MSD raddix sort



01000	0 1 0 0 0	0 0 1 0 1	00001	00001	00001
10000	0 0 0 0 1	0 0 0 0 1	0 0 1 0 1	0 0 1 0 1	00101
01100	0 1 1 0 0	0 0 1 1 1	0 0 1 1 1	0 0 1 1 1	00111
0 0 1 1 1	0 0 1 1 1	01100	0 1 0 0 0	01000	01000
0 1 1 1 0	0 1 1 1 0	0 1 1 1 0	0 1 1 1 0	0 1 1 0 0	0 1 1 0 0
10101	0 1 1 0 1	0 1 1 0 1	0 1 1 0 1	0 1 1 0 1	01101
10010	0 1 1 1 0	0 1 1 1 0	0 1 1 1 0	0 1 1 1 0	0 1 1 1 0
10000	0 0 1 0 1	0 1 0 0 0	0 1 1 0 0	0 1 1 1 0	0 1 1 1 0
00101	10000	10000	10000	10000	10000
0 1 1 1 0	10010	10010	10010	1 0 0 0 0	10000
1 1 0 1 1	1 1 0 1 1	1 0 1 0 1	10000	10010	10010
				No. of	
1 1 1 0 1	1 1 1 0 1	1 0 0 0 0	1 0 1 0 1	1 0 1 0 1	10101
0 1 1 0 1	1 0 1 0 1	1 0 1 0 1	1 0 1 0 1	1 0 1 0 1	10101
10111	1 0 1 1 1	10111	10111	10111	10111
00001	10000	1 1 1 0 1	1 1 0 1 1	11011	11011
10101	10101	1 1 0 1 1	1 1 1 0 1	11101	11101

MSD raddix sort



- + O(n*key_len) время работы*
- + O(n) памяти
- + нет сортировки частей размером 1

- + could be unstable
- рекурсивная реализация

MSD vs LSD



362	291	207	2 07	237	2 37	216	211
436	36 <mark>2</mark>	4 <mark>3</mark> 6	2 53	318	2 16	211	216
291	25 <mark>3</mark>	2 <mark>5</mark> 3	2 91	216	211	2 <mark>3</mark> 7	237
487	436	3 <mark>6</mark> 2	<mark>3</mark> 62	462	2 68	2 <mark>6</mark> 8	268
207	487	487	<mark>3</mark> 97	211	3 18	318	318
253	207	291	4 36	268	462	4 <mark>6</mark> 2	46 <mark>0</mark>
397	397	3 <mark>9</mark> 7	<mark>4</mark> 87	460	4 60	4 <mark>6</mark> 0	462

 $ext{vocuse}$

LSD Radix Sorting:

Sort by the last digit, then

by the middle and the first one

Sort by the first digit, then sort each of the groups by the next digit

Как сортировать ключи разной длинны?



- * расширить алфавит пустым символом
- сложность LSD ~ максимальной длине
- + MSD сортирует только непустуе суффиксы

Как сортировать длинные строки?



дорогое сравнение если строки отличаются только на конце много итераций LSD на длинных ключах

MSD sort vs QuickSort



первый бит == 1 как опорный элемент

 $ext{vocuse}$

Binary QuickSort



разделим массив на две части

- с ведущим разрядом 0
- о с ведущим разрядом 1ы

рекурсивно отсортируем обе части

 $ext{vocuse}$

Binary quicksort



```
quicksortB(int a[], int l, int r, int w)
\{ int i = 1, j = r; \}
  if (r <= 1 || w > bitsword) return;
  while (j != i)
     while (digit(a[i], w) == 0 && (i < j)) i++;
     while (digit(a[j], w) == 1 && (j > i)) j--;
     exch(a[i], a[j]);
  if (digit(a[r], w) == 0) j++;
  quicksortB(a, l, j-1, w+1);
  quicksortB(a, j, r, w+1);
```

3-way split



actinian doenobite actinian jeffrey donelrad bracteal coenobite actinian doenobite bracteal donelrad conelrad secureness dumin secureness dilatedly chariness cumin dentesimal inkblot chariness bracteal effrey dankerous displease displease dircumflex millwright millwright millwright repertoire repertoire repertoire dourness dourness dourness southeast centesimal southeast fondler fondler fondler interval interval interval reversionary reversionary reversionary dilatedly dumin secureness inkblot dhariness dilatedly dentesimal southeast inkblot dankerous jeffrey cankerous circumflex dircumflex displease

3-way raddix quick sort



```
#define ch(A) digit(A, D)
void quicksortX(Item a[], int 1, int r, int D)
  int i, j, k, p, q; int v;
  if (r-1 <= M) { insertion(a, 1, r); return; }
  v = ch(a[r]); i = 1-1; j = r; p = 1-1; q = r;
  while (i < j)
     while (ch(a[++i]) < v);
     while (v < ch(a[--j])) if (j == 1) break;
     if (i > j) break;
     exch(a[i], a[j]);
     if (ch(a[i]) == v) { p++; exch(a[p], a[i]); }
     if (v==ch(a[j])) { q--; exch(a[j], a[q]); }
  if (p == q)
     { if (v != '\0') quicksortX(a, 1, r, D+1);
       return; }
  if (ch(a[i]) < v) i++;
  for (k = 1; k <= p; k++, j--) exch(a[k], a[j]);
  for (k = r; k \ge q; k--, i++) exch(a[k], a[i]);
  quicksortX(a, 1, j,(D));
  if ((i == r) && (ch(a[i]) == v)) i++;
  if (v != '\0') quicksortX(a, j+1, i-1, D+1)
  quicksortX(a, i, r,(D))
```

3-way raddix quick sort



```
for
      for
             bet
                    bet
                           a | ce
tip
ilk
             dug
     dug
                    and
                           a | nd
     ilk
             cab
                    ace
                           b et
dim
     dim
             dim
                    clab
tag
     ago
             ago
                    claw
jot
      and
             and
                    c|ue
sob
      fee
             egg
                    egg
nob
     cue
             cue
                    dug
sky
      caw
             caw
                    dim
             flee
hut
      hut
ace
             for
      ace
bet
     bet
             f|ew
men
      cab
             ilk
             gig
hut
egg
      egg
few
     few
      jlay
jot
jay
owi
             jajy
      jjoy
             joly
joy
      jjam
             jojt
rap
gig
      owl
             owl
                    m | en
wee
      wee
             now
                    owl
was
      was
             nob
                    nob
cab
     men
             men
                    now
wad
     wad
             rap
caw
      sky
                    sky
             sky
                    tip
cue
     nob
             was
                           sob
                           t|ip
fee
      sob
             sob
                    sob
                                  talr
                           tap
tap
     tap
             tap
                    tap
                                  talp
                           t|ag
ago
     tag
                    tag
             tag
                                  tag
tar
     tar
                           t ar
                                  tip
             tar
                    tar
dug
      tip
             tip
                    w as
and
                    w | ee
      now
             wee
jam
             wad
                    w|ad
     rap
```

Спасибо за внимание!

