Sıralama Algoritmaları

Sıralama Algoritmaları Analiz Özet

Adı ¢	Ortalama ♦	En Kötü ◆	Bellek 	Kararlı mı? ♦	Yöntem ♦
Kabarcık Sıralaması	_	O(n²)	O(1)	Evet	Değiştirme
Kokteyl Sıralaması	_	O(n²)	O(1)	Evet	Değiştirme
Tarak Sıralaması	O(n log n)	O(n log n)	O(1)	Hayır	Değiştirme
Cüce Sıralaması	_	O(n²)	O(1)	Evet	Değiştirme
Seçmeli Sıralama	O(n²)	O(n²)	O(1)	Hayır	Seçme
Eklemeli Sıralama	O(n + d)	O(n²)	O(1)	Evet	Ekleme
Kabuk Sıralaması	_	O(n log² n)	O(1)	Hayır	Ekleme
Ağaç Sıralaması	O(n log n)	O(n log n)	O(n)	Evet	Ekleme
Kütüphane Sıralaması	O(n log n)	O(n²)	O(n)	Evet	Ekleme
Birleştirmeli Sıralama	O(n log n)	O(n log n)	O(n)	Evet	Birleştirme
Yerinde Birleştirmeli Sıralama	O(n log n)	O(n log n)	O(1)	Evet	Birleştirme
Yığın Sıralaması	O(n log n)	O(n log n)	O(1)	Hayır	Seçme
Rahat Sıralama	_	O(n log n)	O(1)	Hayır	Seçme
Hızlı Sıralama	O(n log n)	O(n²)	O(log n)	Hayır	Bölümlendirme
İçgözlemle Sıralama	O(n log n)	O(n log n)	O(log n)	Hayır	Melez
Sabir Siralamasi	_	O(n²)	O(n)	Hayır	Ekleme
Íplik Sıralaması	O(n log n)	O(n²)	O(n)	Evet	Seçme

Alt Sınırları Sıralama Doğrusal-Zaman (linear time) Sıralaması

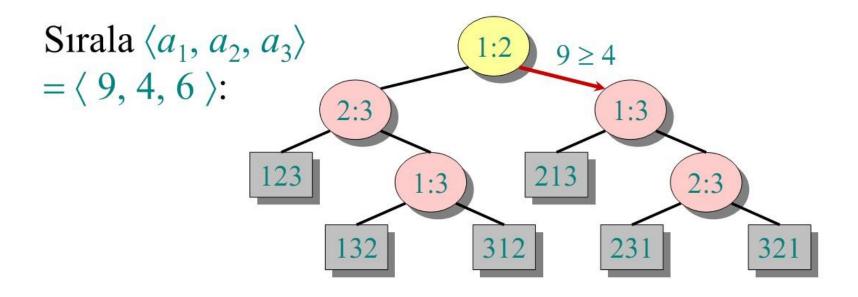
Ne kadar hızlı sıralayabiliriz?

- ulletn adet sayı $O(n \ lgn)$ zamanda sıralayan çeşitli algoritmalar vardır.
 - Birleştirmeli ve bellek yığın (heap) sıralama bu <u>üst sınıra</u> en kötü durumda ulaşır.
 - Hızlı (quick) sıralama <u>ise ortalama zamanda</u> ulaşır.
- Bu algoritmaların her biri için $\Omega(n \, lgn)$ sürede çalışmasına sebep olan n girdi <u>sayıda bir sıra üretebiliriz</u>.
- Bu algoritmalarda *belirledikleri sıralı düzen* sadece girdi elemanları **arasındaki karşılaştırmaya dayanır.** Bu algoritmalar <u>karşılaştırmalı sıralama algoritmalardır.</u>

Sıralama İçin Alt Sınırlar

- n sayıda elemanı karşılaştırarak sıralamak için en kötü durumda $\Omega(n \lg n)$ karşılaştırma yapılmasını kanıtlamalıyız.
- Karşılaştırma sıralamasında $\langle a_1, a_2, \dots, a_n \rangle$ girdi dizisi hakkında düzenli bir bilgi edinmek için karşılaştırmalar kullanılır.
- Verilen a_i ve a_j elemanları göreceli bir düzene karar vermek için $a_i < a_j, a_i \le a_j, a_i = a_j, a_i \ge a_j$ veya $a_i > a_j$ testlerinden birinden geçirilir. Burada = olma veya > , < olma durumları yerine $a_i \le a_j, a_i \ge a_j, a_i >$ değerlendirmemiz yeterli olacaktır.
- n tane sayı n! düzende karşımıza gelebilir.

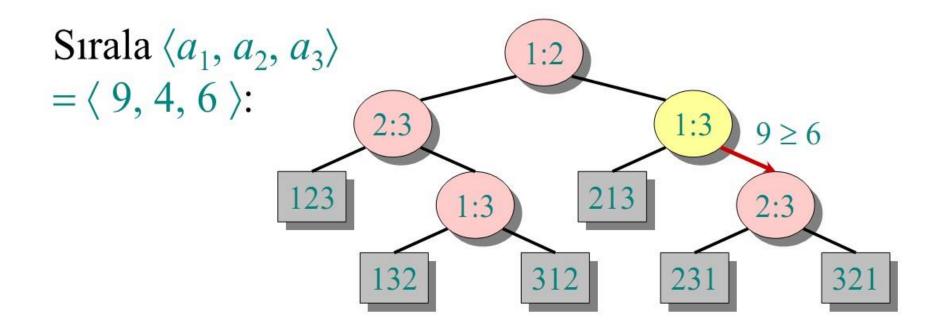
Alt sınır belirlememizde karar ağaçlarından faydalanabiliriz.



* Araya yerleştirme algoritmasına uyan bir karar ağacı örneği

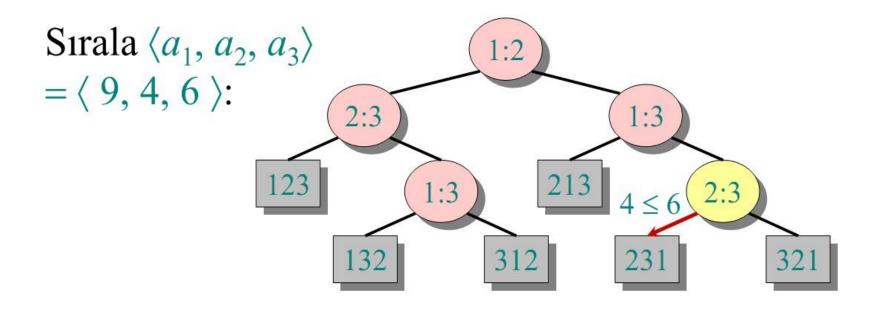
Her iç boğumun etiketlenmesi i:j; $i, j \in \{1, 2, ..., n\}$ için.

- Sol alt-ağaç $a_i \le a_j$ ise, ardarda karşılaştırmaları gösterir.
- Sağ alt-ağaç $a_i \ge a_j$ ise, ardarda karşılaştırmaları gösterir.



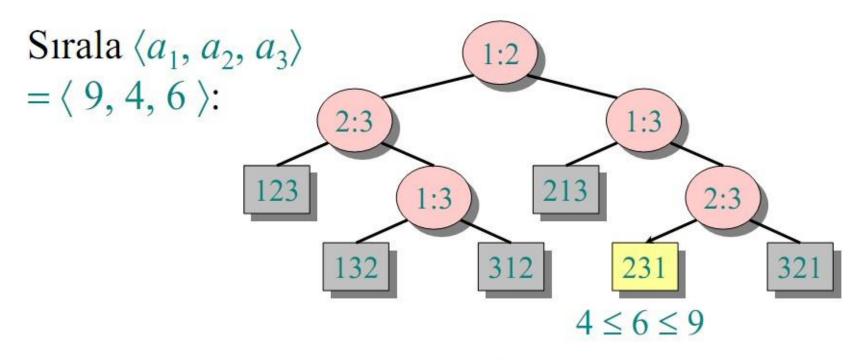
Her iç boğumun etiketlenmesi i:j; $i,j \in \{1, 2, ..., n\}$ için:

- Sol alt-ağaç $a_i \le a_j$ ise, ardarda karşılaştırmaları gösterir.
- Sağ alt-ağaç $a_i \ge a_j$ ise, ardarda karşılaştırmaları gösterir.



Her iç boğumun etiketlenmesi i:j; $i,j \in \{1, 2, ..., n\}$ için.

- Sol alt-ağaç $a_i \le a_j$ ise, ardarda karşılaştırmaları gösterir.
- Sağ alt-ağaç $a_i \ge a_j$ ise, ardarda karşılaştırmaları gösterir.



Her yaprakta $\langle \pi(1), \pi(2), ..., \pi(n) \rangle$ permütasyonu vardır bu $a_{\pi(1)} \le a_{\pi(2)} \le \cdots \le a_{\pi(n)}$ sıralamasının tamamlanmış olduğunu gösterir.

Karar-ağacı modeli

- Bir karar ağacı her karşılaştırma sıralaması uygulanmasını modelleyebilir:
 - Her n giriş boyutu için bir ağaç.
 - Algoritmayı **iki elemanı karşılaştırdığında** bölünüyormuş gibi görün.
 - Ağaç tüm olası komut izlerindeki karşılaştırmalar içerir.
 - Algoritmanın çalışma zamanı = takip edilen yolun uzunluğu.
 - En kötü-durum çalışma zamanı = ağacın boyu.

Karar-ağacı sıralamasında alt sınır

Teorem. n elemanı sıralayabilen bir karar-ağacının yüksekliği (boyu) $\Omega(n \lg n)$ olmalıdır.

Kanıtlama. Ağacın ≥ n! yaprağı olmalıdır, çünkü ortada n! olası permütasyon vardır. Boyu h olan bir ikili ağacın ≤ 2^h yaprağı olur. Böylece, $n! \le 2^h$.

```
∴ h \ge \lg(n!) (lg monoton artışlı)

\ge \lg ((n/e)^n) (Stirling'in formülü)

= n \lg n - n \lg e

= \Omega(n \lg n).
```

Karar-ağacı sıralamasında alt sınır

Doğal sonuç:

Yığın sıralaması ve birleştirme sıralaması asimptotik olarak en iyi karşılaştırma sıralaması algoritmalarıdır.

Doğrusal zamanda sıralama

❖ Sayma sıralaması (Counting Sort):

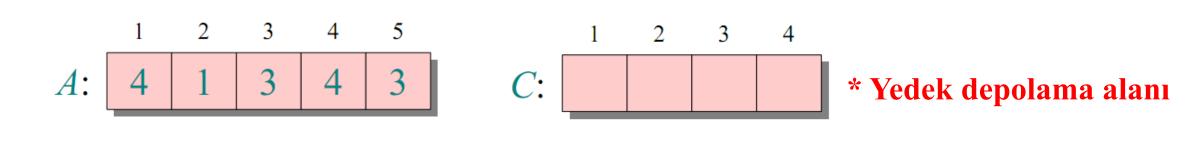
Elemanlar arası karşılaştırma yok.

Giriş: A[1..n], burada A[j] \in {1, 2, ..., k}.

- **k**, *küçük* ise <u>iyi bir algoritma olur.</u>
- k, büyük ise *çok kötü bir algoritma olur*.

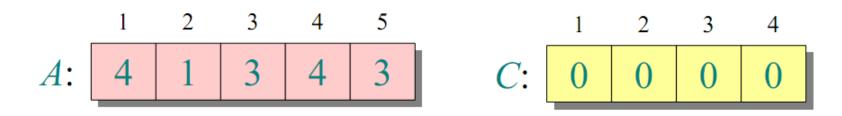
Çıkış: B[1..n], sıralı.

Yedek depolama: C[1..k].



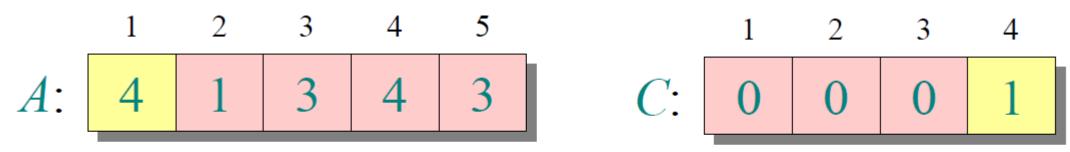
B:

o Dizi girişi 1 ile 4 arasındadır. O zaman **k=4** olur. Bir başka ifade ile dizinin en büyük elemanı bulunur.



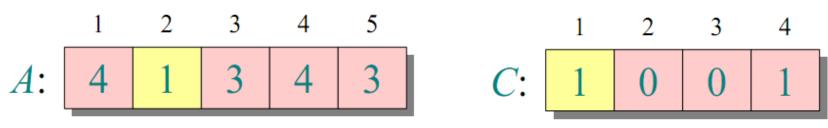
Dizinin elemanları ilk etapta sıfırlanır.

Döngü 1 for
$$i \leftarrow 1$$
 to k do $C[i] \leftarrow 0$

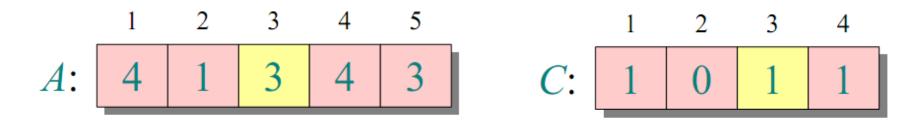


İlk elemandan başlanarak her bir elemandan kaç tane var sayılır.

for *j* ← 1 to *n*
do
$$C[A[j]]$$
 ← $C[A[j]] + 1$ $\triangleright C[i] = |\{\text{key} = i\}|$



for
$$j \leftarrow 1$$
 to n
do $C[A[j]] \leftarrow C[A[j]] + 1 \quad \triangleright C[i] = |\{\text{key} = i\}|$ Döngü 2



for
$$j \leftarrow 1$$
 to n
do $C[A[j]] \leftarrow C[A[j]] + 1 \quad \triangleright C[i] = |\{\text{key} = i\}| \quad \textbf{D\"{o}ng\"{u}} \ \textbf{2}$



for
$$j \leftarrow 1$$
 to n
do $C[A[j]] \leftarrow C[A[j]] + 1 \quad \triangleright C[i] = |\{\text{key} = i\}|$ Döngü 2

$$\begin{array}{ll} \textbf{D\"{o}ng\"{u}} \ \textbf{2} & \textbf{for} \ j \leftarrow 1 \ \textbf{to} \ n \\ \textbf{do} \ C[A[j]] \leftarrow C[A[j]] + 1 \quad \triangleright C[i] = |\{\text{key} = i\}| \end{array}$$

for
$$i \leftarrow 2$$
 to k
do $C[i] \leftarrow C[i] + C[i-1]$ $\triangleright C[i] = |\{\text{key } \le i\}|$ Döngü 3

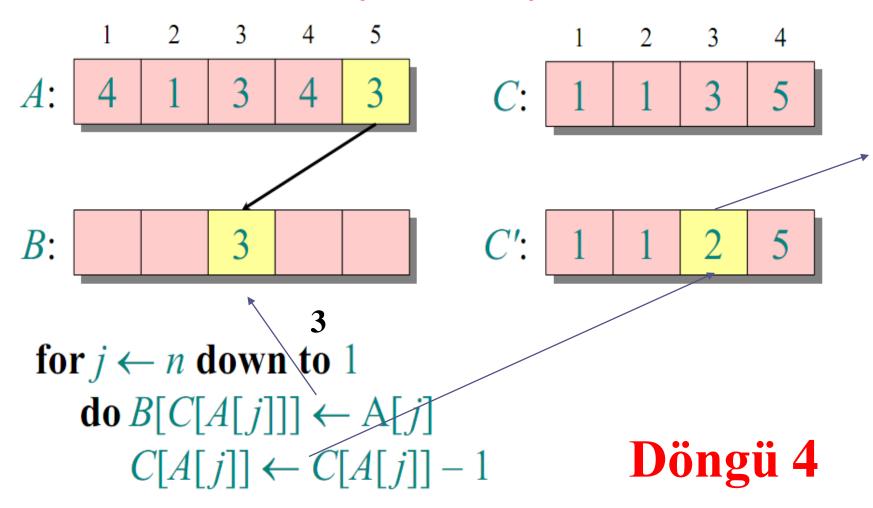
- Bir elemanın değeri bir önceki elemanın değeriyle toplanır ve elemana yazılır.
- Bir başka ifadeyle *C[i]*, i den küçük veya eşit olan elamanların sayısını içerir.

for
$$i \leftarrow 2$$
 to k
do $C[i] \leftarrow C[i] + C[i-1]$ $\triangleright C[i] = |\{\text{key } \le i\}|$ **Döngü 3**

- Bir elemanın değeri bir önceki elemanın değeriyle toplanır ve elemana yazılır.
- Bir başka ifadeyle *C[i]*, i den küçük veya eşit olan elamanların sayısını içerir.

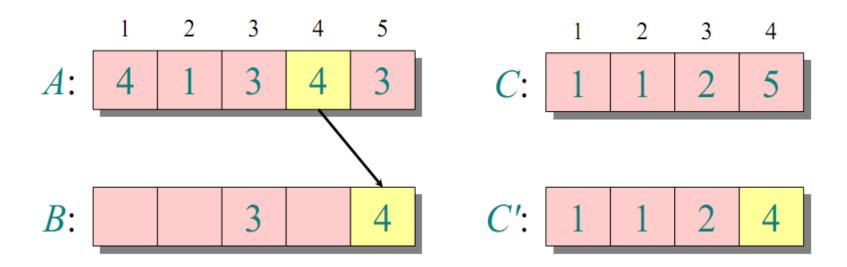
for
$$i \leftarrow 2$$
 to k
do $C[i] \leftarrow C[i] + C[i-1]$ $\triangleright C[i] = |\{\text{key } \le i\}|$ Döngü 3

- Bir elemanın değeri bir önceki elemanın değeriyle toplanır ve elemana yazılır.
- Bir başka ifadeyle *C[i]*, i den küçük veya eşit olan elamanların sayısını içerir.



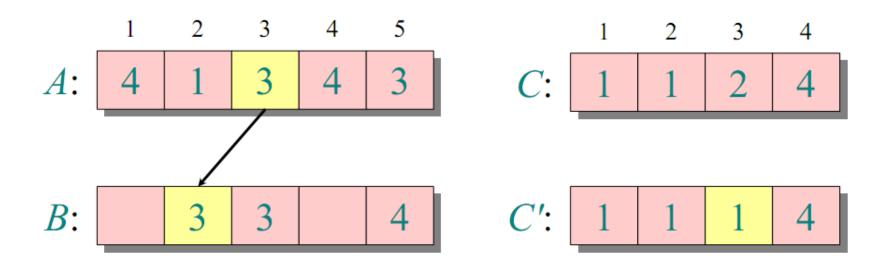
* 3'ten küçük ve eşit 2 tane eleman var. Dolayısıyla bu sayı 3. indekste yer alacaktır.

- A dizisinin **sondan** itibaren elemanı seçilir.
- A[j] elemanını çıktı dizisi B'de doğru pozisyona yerleştirir.



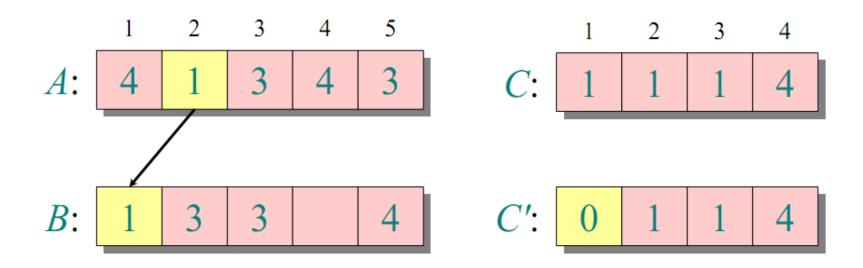
for
$$j \leftarrow n$$
 down to 1
do $B[C[A[j]]] \leftarrow A[j]$
 $C[A[j]] \leftarrow C[A[j]] - 1$

- A dizisinin **sondan** bir azaltılarak elemanı seçilir.
- A[j] elemanını çıktı dizisi B'de doğru pozisyona yerleştirir.



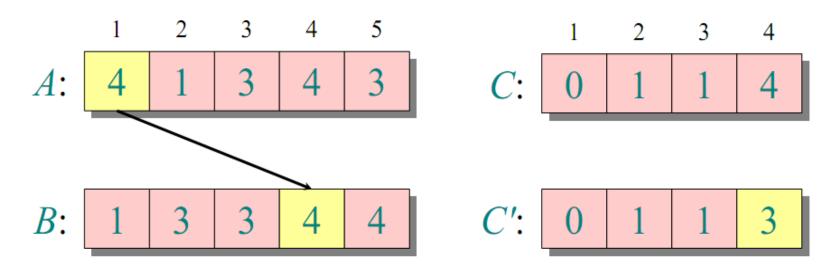
for
$$j \leftarrow n$$
 down to 1
do $B[C[A[j]]] \leftarrow A[j]$
 $C[A[j]] \leftarrow C[A[j]] - 1$

- A dizisinin sondan birer birer azaltılarak elemanı seçilir.
- A[j] elemanını çıktı dizisi B'de doğru pozisyona yerleştirir.



for
$$j \leftarrow n$$
 down to 1
do $B[C[A[j]]] \leftarrow A[j]$
 $C[A[j]] \leftarrow C[A[j]] - 1$

- A dizisinin sondan birer birer azaltılarak elemanı seçilir.
- A[j] elemanını çıktı dizisi B'de doğru pozisyona yerleştirir.



for
$$j \leftarrow n$$
 down to 1
do $B[C[A[j]]] \leftarrow A[j]$
 $C[A[j]] \leftarrow C[A[j]] - 1$

- A dizisinin sondan birer birer azaltılarak elemanı seçilir.
- A[j] elemanını çıktı dizisi B'de doğru pozisyona yerleştirir.

Çözümleme

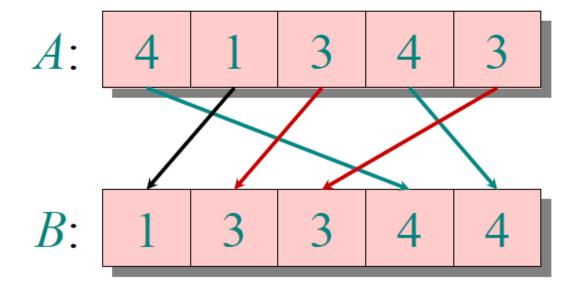
```
\Theta(k) \begin{cases} \mathbf{for} \ i \leftarrow 1 \ \mathbf{to} \ k \\ \mathbf{do} \ C[i] \leftarrow 0 \end{cases}
      \Theta(n) \begin{cases} \mathbf{for} \ j \leftarrow 1 \ \mathbf{to} \ n \\ \mathbf{do} \ C[A[j]] \leftarrow C[A[j]] + 1 \end{cases}
      \Theta(k) \begin{cases} \text{for } i \leftarrow 2 \text{ to } k \\ \text{do } C[i] \leftarrow C[i] + C[i-1] \end{cases}
      \Theta(n) \begin{cases} \mathbf{for} \ j \leftarrow n \ \mathbf{down} \ \mathbf{to} \ 1 \\ \mathbf{do} \ B[C[A[j]]] \leftarrow A[j] \\ C[A[j]] \leftarrow C[A[j]] - 1 \end{cases}
\Theta(n+k)
```

Not: Bu döngüler iç içe değildir.

Çalışma Zamanı

- o k = O(n) ise, sayma sıralaması Θ(n) süresi alır. Eğer k=n² veya k=2n çok kötü bir algoritma olur.
- o k tamsayı olmalı.
- \circ Ama sıralamalar $\Omega(n \lg n)$ süresi alıyordu! (karar ağacı)
- Hata nerede?
- O Yanıt:
- \circ Karşılaştırma sıralaması $\Omega(n \lg n)$ süre alır.
- O Sayma sıralaması bir karşılaştırma sıralaması değildir.
- Aslında elemanlar arasında bir tane bile karşılaştırma yapılmaz!

Sayma sıralaması *kararlı* bir sıralamadır: eşit eşit elemanlar arasındaki düzeni korur.



Sayma/Sayarak Sıralamanın Avantaj ve Dezavantajları

- **□** Avantajları:
 - ❖ n ve k da doğrusaldır (lineer).
 - Kararlı yapıdadır.
 - Kolay uygulanır.
- **□** Dezavantajları:
 - Yerinde sıralama yapmaz. Ekstra depolama alanına ihtiyaç duyar.
 - Sayıların küçük tam sayı olduğu varsayılır.
 - O Byte ise ek dizinin boyutu en fazla 28= 256 olur fakat sayılar int ise yani 32 bit lik sayılar ise 232= 4.2 milyar sayı eder oda yaklaşık **16 Gb yer tutar.**

Taban (Radix) sıralaması

- ☐En az öneme sahip basamaktan *başlayarak* sıralama yapar.
- □Birden fazla anahtara göre sıralama gerektiğinde kolaylıkla kullanılabilir.
- DÖrneğin tarihe göre sıralamada yıl, ay, gün'e göre sıralama yapılır. Tarih sıralamasında önce gün, sonra ay ve yıl' a göre kolayca sıralanır.

Taban (Radix) sıralaması

- O Basamak basamak sıralama.
- Kötü fikir: sıralamaya önceli en önemli basamaktan başlamak.
- O İyi fikir: Sıralamaya en önemsiz basamaktan başlamak ve ek kararlı sıralama uygulamak.

Taban (Radix) sıralaması -örnek

3 2 9	7	20	7	20	3 2 9
457	3	5 5	3	29	3 5 5
657	4	3 6	4	3 6	4 3 6
839	4	5 7	8	3 9	4 5 7
4 3 6	6	5 7	3	5 5	6 5 7
720	3	29	4	5 7	720
3 5 5	8	3 9	6	5 7	8 3 9
1		又	7	Z /	7

Taban (Radix) sıralaması – örnek 2

Basamak konumunda tümevarım

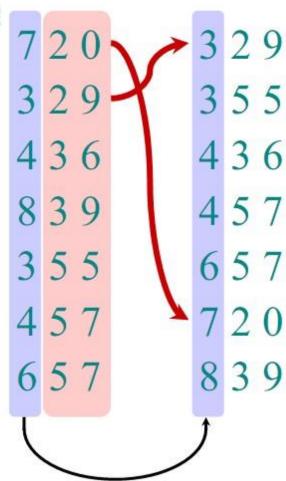
- Sayıların düşük düzeyli
 t 1 basamaklarına göre
 sıralandığını varsayın.
- t basamağında sıralama yapın.

7	20	3	2	9
3	29	3	5	5
4	3 6	4	3	6
8	39	4	5	7
3	5 5	6	5	7
4	5 7	7	2	0
6	57	8	3	9
1)		
(j	200	

Taban (Radix) sıralaması – örnek 2

Basamak konumunda tümevarım

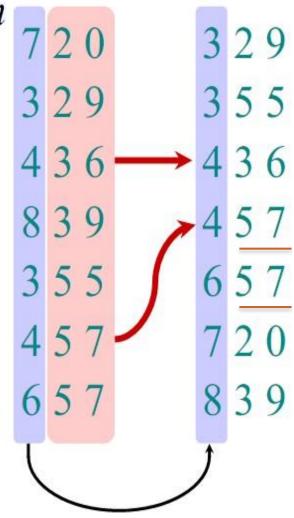
- Sayıların düşük düzeyli
 t –1 basamaklarına göre
 sıralandığını varsayın.
- *t* basamağında sıralama yapın.
 - t basamağında farklı olan iki sayı doğru sıralanmış.



Taban (Radix) sıralaması – örnek 2

Basamak konumunda tümevarım

- •Sayıların düşük düzeyli t −1 basamaklarına göre sıralandığını varsayın.
- t basamağında sıralama yapın.
 - t basamağında farklı olan iki sayı doğru sıralanmış.
 - t basamağındaki iki eşit sayının girişteki sıraları muhafaza edilmiş ⇒ doğru sıra.



Taban sıralamasının çözümlemesi

- Sayma sıralamasını ek kararlı sıralama varsayın.
- Herbiri *b* bit olan *n* bilgiişlem sözcüğünü sıralayın.
- Her sözcüğün basamak yapısı *b/r* taban-2^r olarak görülebilir.

Örnek: 32-bit sözcük

 $r = 8 \Rightarrow b/r = 4$ ise, taban-28 basamak durumunda sıralama 4 geçiş yapar; veya $r = 16 \Rightarrow b/r = 2$ ise, taban-216 basamakta 2 geçiş yapar.

Kaç geçiş yapmalıyız?

Taban sıralamasının çözümlemesi

Hatırla: Sayma sıralaması $\Theta(n + k)$ süresini alır; (0 ile k - 1 aralığında n sayıyı sıralamak için). Her b-bitlik sözcük r-bitlik parçalara ayrılırsa, sayma sıralamasının her geçişi $\Theta(n + 2^r)$ süre alır. Bu durumda b/r geçiş olduğundan, elimizde:

$$T(n,b) = \Theta\left(\frac{b}{r}(n+2r)\right)$$
 olur.

r' yi, T(n, b)' yi en aza düşürecek gibi seçin:
r'yi arttırmak daha az geçiş demektir, ama
r >> lg n olduğundan, süre üstel olarak artar.

r' yi seçmek

T(n,b)'yi türevini alıp 0' a eşitleyerek en aza düşürün.

Veyahut da, istemediğimiz değer $2^r >> n$ olduğundan, bu sınırlamaya bağlı kalarak r'yi olabildiğince büyük seçmenin asimptotik bir sakıncası olmadığını gözleyin.

 $r = \lg n$ seçimi $T(n, b) = \Theta(bn/\lg n)$ anlamına gelir.

• 0 ile $n^d - 1$ aralığındaki sayılarla $b = d \lg n$ 'yi elde ederiz. \Rightarrow taban sıralaması $\Theta(dn)$ süresini alır.

Not: Değer aralığı 0 ... 2^b -1= 0 ... n^d kabul edip her iki tarasın logaritması alınır. **b=dlogn** olur. Burada **d basamak sayısıdır.**

Sonuçlar

Pratikte taban sıralaması büyük girişler için hızlıdır; aynı zamanda kod yazması ve bakımı kolaydır.

Örnek (32-bitlik sayılar için):

- En çok 3 geçiş (≥ 2000 sayının sıralanmasında).
- Birleştirme sıralaması /çabuk sıralama [1g 2000] en az 11 geçiş yaparlar.

Dezavantajı: Çabuk sıralamanın aksine, taban sıralamasının yer referansları zayıftır ve bu nedenle ince ayarlı bir çabuk sıralama, dik bellek sıradüzeni olan günümüz işlemcilerinde daha iyi çalışır.

Pratikte radix sort yani taban sıralaması, <u>sayılarınız gerçekten küçük</u> <u>değilse</u> çok hızlı bir algoritma <u>değildir</u>.

Kova Sıralama (BucketSort)

- Kova Sıralaması (ya da sepet sıralaması), sıralanacak bir diziyi parçalara <u>ayırarak</u> sınırlı <u>sayıdaki kovalara</u> (ya da sepetlere) atan bir sıralama algoritmasıdır.
- Ayrışma işleminin ardından her kova kendi içinde ya farklı bir algoritma kullanılarak ya da kova sıralamasını özyinelemeli olarak çağırarak sıralanır.

Kova Sıralama (BucketSort)

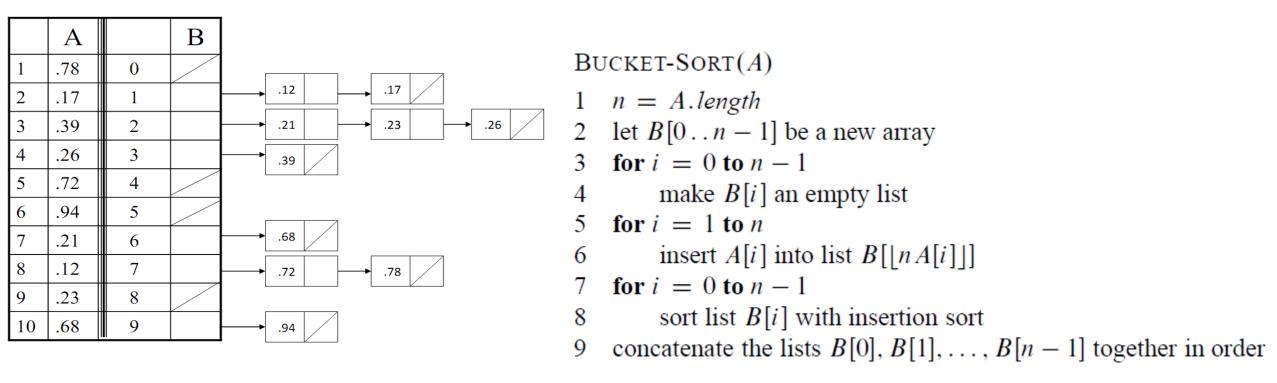
Kova sıralaması aşağıdaki biçimde çalışır:

- Başlangıçta boş olan bir "kovalar" dizisi oluştur.
- Asıl dizinin üzerinden geçerek her öğeyi ilgili aralığa denk gelen kovaya at.
- Boş olmayan bütün kovaları sırala.
- Boş olmayan kovalardaki bütün öğeleri yeniden diziye al.

Kova Sıralama (BucketSort)

- Kova sıralaması doğrusal zamanda çalışır.
- Girişin düzgün dağılımlı olduğu kabul edilir.
- Random olarak [0,1) aralığında oluşturulmuş giriş bilgileri olduğu kabul edilir.
- Temel olarak [0, 1) aralığını n eşit alt aralığa böler ve girişi bu aralıklara dağıtır.
- Aralıklardaki değerleri insert sort (araya sokma) ile sıralar.
- Aralıkları bir biri ardına <u>ekleyerek sıralanmış diziyi elde</u>
 <u>eder.</u>

Kova/Sepet Sıralama (Bucket Sort)



Yukarıdaki şekilde n = 10 için BUCKET-SIRALAMA işlemi.

(a) Giriş dizisi A[1..10] (b) Algoritmanın 8. satırından sonra sıralanmış listelerin (kovalar) B[0..9] dizisi. Kova i, yarı açık aralıktaki [i/10, i+1/10) değerleri tutar. Sıralanan çıktı, B[0],B[1], ...B[9] listelerinin sırasına göre bir birleştirmeden oluşur.

Kova/Sepet Sıralama (Bucket Sort)

```
BUCKET-SORT(A)
                                                                                      O(n^2)
                                                En kötü durum performansı
O(n)
       1 \quad n = A.length
                                                                                     O(n+k)
                                                Ortalama durum performansı
O(n) 2 let B[0..n-1] be a new array
O(n) 3 for i = 0 to n - 1
O(n)
       4 make B[i] an empty list
O(n)
       5 for i = 1 to n
O(n)
      6 insert A[i] into list \mathcal{B}[\lfloor nA[i] \rfloor]
         for i = 0 to n \rightarrow 1
O(n)
O(n_i^2) 8
               sort list B[i] with insertion sort
         concatenate the lists B[0], B[1], \ldots, B[n-1] together in order
O(n)
```

Her i kovasının aynı $E(n_i^2)$ değerine sahip olması şaşırtıcı değildir.

A girdi dizisindeki her bir değerin herhangi bir kovaya düşme olasılığı eşit derecededir (1/n).

Kaynakça

- ► Algoritmalar : Prof. Dr. Vasif NABİYEV, Seçkin Yayıncılık
- ▶ Algoritmalara Giriş: Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, Palme YAYINCILIK
- ► Algoritmalar : Robert Sedgewick , Kevin Wayne, Nobel Akademik Yayıncılık
- M.Ali Akcayol, Gazi Üniversitesi, Algoritma Analizi Ders Notları
- Doç. Dr. Erkan TANYILDIZI, Fırat Üniversitesi, Algoritma Analizi Ders Notları
- http://www.bilgisayarkavramlari.com