

CENG 213 Veri Yapıları 8: Özetleme Fonksiyonları(Hash)

Öğr.Gör. Şevket Umut ÇAKIR

Pamukkale Üniversitesi

Hafta 8

1 Bağlı Listesiz Özetleme

- Doğrusal Sondalama
- Karesel Sondalama
- Çift Özetleme
- Yeniden Özetleme

2 $O(1)$ Erişime Sahip Özetleme

- Cuckoo Özetleme
- Hopscotch Özetleme

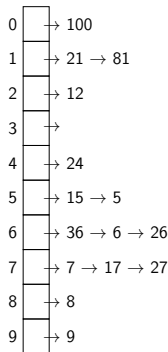
Tanım

Özetleme(hashing) işlemi bir veriyi sabit boyutlu bir alana/tabloya izdüşürme işlemidir.

- Örneğin $f(x) = x \bmod 13$ fonksiyonu doğal sayıları 13 uzunluklu bir tabloya izdüşürür
- Özetleme işleminin amacı arama ve ekleme işlemlerinin karmaşıklığını sabit zamana indirmektir

Ayrık Zincirleme/Seperate Chaining

- Tablodaki her bir konum bir bağlı liste içerir
- 21, 7, 36, 8, 9, 17, 6, 24, 26, 15, 27, 81, 12, 5, 100 değerleri eklensin.



- Bağlı listelerde yeni elemanlar için yer ayırma işlemi zaman alıcı olabildiği için ayrık zincirleme dezavantajlı olabilir
- Sırasıyla $h_0(x), h_1(x), h_2(x), \dots$ konumlarındaki ilk boş kısma yerleştirilir. $h_i(x) = (\text{hash}(x) + f(i)) \bmod \text{TabloBoyutu}$
- Özet fonksiyonu bir konum üretir
- Üretilen konum boş ise değer yerleştirilir
- Üretilen konum dolu ise boş bir konum bulunana kadar birer birer sonraki konumlara bakılır
- Doğrusal sondalama için $f(i) = i$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		89

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası
0			
1			
2			
3			
4			
5			
6			
7			
8			18
9		89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası
0				49
1				
2				
3				
4				
5				
6				
7				
8			18	18
9		89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası
0				49	49
1					58
2					
3					
4					
5					
6					
7					
8			18	18	18
9		89	89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0				49	49	49
1					58	58
2						69
3						
4						
5						
6						
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Doğrusal Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0				49	49	49
1					58	58
2						69
3						
4						
5						
6						
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $h(x) = x \bmod 10$

Karesel Sondalama(Deneme)/Quadratic Probing

- Sırasıyla $h_0(x), h_1(x), h_2(x), \dots$ konumlarındaki ilk boş kısma yerleştirilir. $h_i(x) = (\text{hash}(x) + f(i)) \bmod \text{TabloBoyutu}$
- Özet fonksiyonu bir konum üretir
- Üretilen konum boş ise değer yerleştirilir
- Üretilen konum dolu ise deneme sayısının karesi kadar ileri bakılır
- Karesel sondalama için $f(i) = i^2$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		89

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası
0			
1			
2			
3			
4			
5			
6			
7			
8			18
9		89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası
0				49
1				
2				
3				
4				
5				
6				
7				
8			18	18
9		89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası
0				49	49
1					
2					58
3					
4					
5					
6					
7					
8			18	18	18
9		89	89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0				49	49	49
1						
2					58	58
3						69
4						
5						
6						
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $h(x) = x \bmod 10$

Tablo: Karesel Sondalama Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0				49	49	49
1						
2					58	58
3						69
4						
5						
6						
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $h(x) = x \bmod 10$

- Sırasıyla $h_0(x), h_1(x), h_2(x), \dots$ konumlarındaki ilk boş kısma yerleştirilir. $h_i(x) = (\text{hash}(x) + f(i)) \bmod \text{TabloBoyutu}$
- Çakışma durumunda ikinci bir özet fonksiyonu kullanılır
- Çift özetleme için $f(i) = i \cdot \text{hash}_2(x)$
- $\text{hash}_2(x) = R - (x \bmod R)$
- R tablo boyutundan küçük bir asal sayıdır

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Ekleme Örneği

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası
0			
1			
2			
3			
4			
5			
6			
7			
8			18
9		89	89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Ekleme Örneği

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası
0				
1				
2				
3				
4				
5				
6				49
7				
8			18	18
9		89	89	89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Ekleme Örneği

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası
0					
1					
2					
3					58
4					
5					
6				49	49
7					
8			18	18	18
9		89	89	89	89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0						69
1						
2						
3					58	58
4						
5						
6				49	49	49
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

Tablo: Çift Özetleme Ekleme Örneği
89, 18, 49, 58, 69 ekleme

	Boş tablo	89 sonrası	18 sonrası	49 sonrası	58 sonrası	69 sonrası
0						69
1						
2						
3					58	58
4						
5						
6				49	49	49
7						
8			18	18	18	18
9		89	89	89	89	89

Tablo: $hash_1(x) = x \bmod 10$
 $hash_2(x) = 7 - (x \bmod 7)$

- Düğüm yapısında **bool** isActive alanı tutulur. Bu alan false ise değer silinmiştir

```
1 public class HashEntry<T> {  
2     T value;  
3     boolean isActive;  
4     public HashEntry(T value, boolean isActive) {  
5         this.value = value;  
6         this.isActive = isActive;  
7     }  
8     public HashEntry(T value) {  
9         this(value, true);  
10    }  
11 }
```

- Eğer tablo çok dolarsa ekleme işlemi daha çok zaman alacaktır ve karesel sondalamada uygun konum bulmak imkansız hale gelebilir
- Bu durumda mevcut tablonun en az iki katı büyüklüğünde yeni bir özet tablosu ve özetleme fonksiyonu oluşturulur ve tüm değerler(silinenler hariç) yeni tabloya eklenir

Yeniden Özetleme Örnek

Tablo: 13, 15, 24, 6, 23 değerleri eklendikten sonra yeni özet tablosu

0	6	0	
1	15	1	
2	23	2	
3	24	3	
4		4	
5		5	
6	13	6	6
		7	23
		8	24
		9	
		10	
		11	
		12	
		13	13
		14	
		15	15
		16	

- Yarısından çoğu boş olan iki adet özet tablosu bulunur
- Her tablo için ayrı bir özetleme fonksiyonu bulunur
- Değerler bu iki tablodan birinde bulunur

Cuckoo Özetleme Örneği

Tablo: Cuckoo Özetleme Örneği

Tablo 1	
0	B
1	C
2	
3	E
4	

Tablo 2	
0	D
1	
2	A
3	
4	F

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

- Eleman ilk tabloya eklenir
- Elemanın eklendiği konum(ilk tablodaki) doluysa, üzerine yerleştirdiğimiz elemanı ikinci tabloya taşırız
- Taşıma sırasında ikinci tablodaki konum da doluysa, o konumdaki elemanı birinci tabloya taşırız

Tablo: A'nın eklenmesi

Tablo 1	
0	A
1	
2	
3	
4	

Tablo 2	
0	
1	
2	
3	
4	

A: 0,2

Tablo: B'nin eklenmesi

Tablo 1	
0	B
1	
2	
3	
4	

Tablo 2	
0	
1	
2	A
3	
4	

A: 0,2

B: 0,0

Tablo: C'nin eklenmesi

Tablo 1	
0	B
1	C
2	
3	
4	

Tablo 2	
0	
1	
2	A
3	
4	

A: 0,2

B: 0,0

C: 1,4

Tablo: D'nin eklenmesi

Tablo 1	
0	B
1	D
2	
3	
4	

Tablo 2	
0	
1	
2	A
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

Tablo: E'nin eklenmesi

Tablo 1	
0	B
1	D
2	
3	E
4	

Tablo 2	
0	
1	
2	A
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

Tablo: F'nin eklenmesi, $F \rightarrow E$

Tablo 1	
0	B
1	D
2	
3	F
4	

Tablo 2	
0	
1	
2	A
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

Tablo: F'nin eklenmesi, $E \rightarrow A$

Tablo 1	
0	B
1	D
2	
3	F
4	

Tablo 2	
0	
1	
2	E
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

Tablo: F'nin eklenmesi, $A \rightarrow B$

Tablo 1	
0	A
1	D
2	
3	F
4	

Tablo 2	
0	
1	
2	E
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

Tablo: F'nin eklenmesi, $B \rightarrow \text{Tablo2}[0]$

Tablo 1	
0	A
1	D
2	
3	F
4	

Tablo 2	
0	B
1	
2	E
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

Tablo: G'nin eklenmesi

Tablo 1	
0	A
1	D
2	
3	F
4	

Tablo 2	
0	B
1	
2	E
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

G: 1,2

Tablo: G'nin eklenmesi

Tablo 1	
0	A
1	D
2	
3	F
4	

Tablo 2	
0	B
1	
2	E
3	
4	C

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

G: 1,2

Tablo: $G \rightarrow D, D \rightarrow B, B \rightarrow A, A \rightarrow E, E \rightarrow F, F \rightarrow C, C \rightarrow G$

Tablo: G'nin eklenmesi

Tablo 1	
0	B
1	C
2	
3	E
4	

Tablo 2	
0	D
1	
2	A
3	
4	F

A: 0,2

B: 0,0

C: 1,4

D: 1,0

E: 3,2

F: 3,4

G: 1,2

Tablo: Çeşitli varyasyonlara göre en fazla doluluk oranları

	Her hücrede 1 eleman	Her hücrede 2 eleman	Her hücrede 4 eleman
2 Özetleme fonksiyonu	0.49	0.86	0.93
3 Özetleme fonksiyonu	0.91	0.97	0.98
4 Özetleme fonksiyonu	0.97	0.99	0.999

Seksek(Hopscotch) Özetleme

- Klasik doğrusal sondalamayı geliştirmek için ortaya atılmıştır
- Amaç sondalama uzunluğunu kısıtlamaktır
- Bir eleman özgün konumundan en fazla belirli bir uzaklıkta yer alabilir(MAX_DIST)
- Eğer eleman özgün konumundan çok uzaksa geriye doğru giderek uygun elemanlar yer değiştirilir
- Hop(sıçrama) bilgisi bitlerden oluşur ve hangi uzaklıktaki elemanların mevcut özet değerine sahip olduğunu gösterir
- Örneğin MAX_DIST=4 için 5. konumdaki 1011 değeri 5,7 ve 8. konumdaki elemanların özet değerinin 5 olduğunu gösterir

Tablo: Örnek Hopscotch Tablosu

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9	B	1000
10	E	0000
11	G	1000
12	F	1000
13		0000
14		0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

- Eklenecek konum boş ise eklenir($hash(x)$)
- Eklenecek konum boş değilse doğrusal sondalama ile ilk boş konum bulunur:
 - Boş konum özgün konumdan MAX_DIST uzaklığından az ise eklenir
 - Aksi takdirde boş konum ile MAX_DIST-1 arasında yer değiştirilebilecek eleman aranır ve istenilen uzaklık değeri elde edilene kadar bu işlem tekrarlanır
 - Yer değiştirecek eleman ararken Hop değerine bakmak yeterlidir, uzaklığı dikkate alarak değeri 1 olan hop değerlerinin gösterdiği elemanlarla yer değişimi yapılır

Tablo: H:9 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9	B	1000
10	E	0000
11	G	1000
12	F	1000
13		0000
14		0000
...		

A: 7
B: 9
C: 6
D: 7
E: 8
F: 12
G: 11
H:9

Tablo: H:9 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9	B	1000
10	E	0000
11		0010
12	F	1000
13	G	0000
14		0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

H:9

Tablo: H:9 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9	B	1010
10	E	0000
11	H	0010
12	F	1000
13	G	0000
14		0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

H:9

Tablo: I:6 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9	B	1010
10	E	0000
11	H	0010
12	F	1000
13	G	0000
14		0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

H: 9

I: 6

Tablo: l:6 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop	
...			
6	C	1000	A: 7
7	A	1100	B: 9
8	D	0010	C: 6
9	B	1010	D: 7
10	E	0000	E: 8
11	H	0001	F: 12
12	F	1000	G: 11
13		0000	H:9
14	G	0000	I:6
...			

Tablo: l:6 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop	
...			
6	C	1000	A: 7
7	A	1100	B: 9
8	D	0010	C: 6
9	B	1010	D: 7
10	E	0000	E: 8
11	H	0001	F: 12
12		0100	G: 11
13	F	0000	H:9
14	G	0000	I:6
...			

Tablo: I:6 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1000
7	A	1100
8	D	0010
9		0011
10	E	0000
11	H	0001
12	B	0100
13	F	0000
14	G	0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

H: 9

I: 6

Tablo: I:6 değerinin eklenmesi(MAX_DIST=4)

	Eleman	Hop
...		
6	C	1001
7	A	1100
8	D	0010
9	I	0011
10	E	0000
11	H	0001
12	B	0100
13	F	0000
14	G	0000
...		

A: 7

B: 9

C: 6

D: 7

E: 8

F: 12

G: 11

H: 9

I: 6

17, 2, 8, 4, 3, 15 değerlerini $h(x) = x \bmod 7$ özet fonksiyonunu kullanarak aşağıdaki sek sek(hopscotch) özet tablosuna ekleyin(MAX_DIST=3).

	Değer	Hop
0		
1		
2		
3		
4		
5		
6		

17, 2, 8, 4, 3, 15 değerlerini $h(x) = x \bmod 7$ özet fonksiyonunu kullanarak aşağıdaki sek sek(hopscotch) özet tablosuna ekleyin(MAX_DIST=3).

	Değer	Hop
0		
1	8	100
2	2	100
3	17	101
4	4	100
5	3	000
6	15	

17, 2, 8, 4, 3, 15 değerlerini $h(x) = x \bmod 7$ özet fonksiyonunu kullanarak aşağıdaki sek sek(hopscotch) özet tablosuna ekleyin(MAX_DIST=3).

	Değer	Hop
0		
1	8	100
2	2	100
3	17	101
4	4	100
5	3	000
6	15	

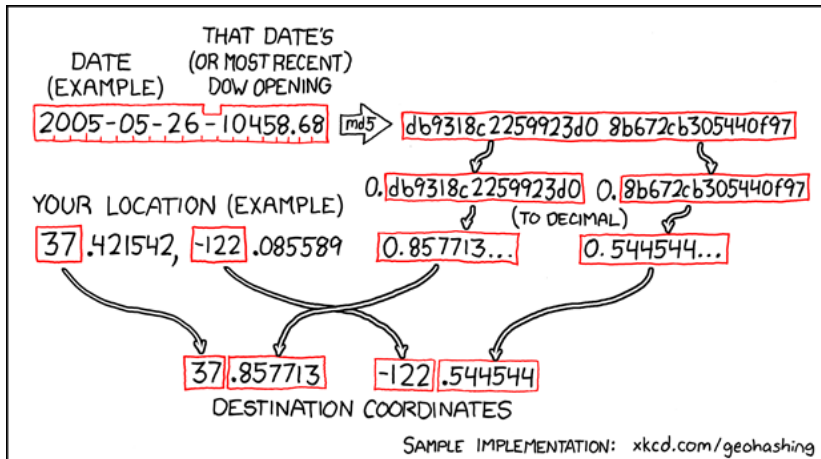
	Değer	Hop
0		
1	8	100
2	2	100
3	17	101
4	15	001
5	3	000
6	4	000

17, 2, 8, 4, 3, 15 değerlerini $h(x) = x \bmod 7$ özet fonksiyonunu kullanarak aşağıdaki sek sek(hopscotch) özet tablosuna ekleyin(MAX_DIST=3).

	Değer	Hop
0		
1	8	100
2	2	100
3	17	101
4	4	100
5	3	000
6	15	

	Değer	Hop
0		
1	8	100
2	2	100
3	17	101
4	15	001
5	3	000
6	4	000

	Değer	Hop
0		
1	8	110
2	15	001
3	17	101
4	2	001
5	3	000
6	4	000



1