ARAMA (SEARCHING) (2)

Kıyım Fonksiyonu (Hashing II)

- Eğer kötü bir seçim, **bazı sabit bir kıyım fonksiyonu** ile **çırpılanacak anahtarları seçerse;**
- **Bu kötü seçim *Θ (n)*** zamanda ortalama alımı destekleyecek aynı hücreye, çırpılanan ***n*** anahtarı seçebilir.
- Herhangi bir sabit kıyım fonksiyonu, **en kötü duruma maruz bırakabilir.**

Kıyım Fonksiyonu (Hashing II)

- Durumu iyileştirmenin yolu **sabit bir kıyım fonksiyonu seçmek yerine** **<u>rastgele bir kıyım fonksiyonu seçmektir.</u>**
- **Rastgele bir kıyım fonksiyonunun seçilmesi, <u>evrensel kıyım olarak adlandırılır.</u>**

Evrensel Kıyım

- Evrensel kıyımlamada uygulamanın başlangıcında dikkatlice tasarlanmış **sınıf fonksiyonları arasından rassal bir çırpı fonksiyonu seçeriz.**
- **Hızlı sıralamada olduğu** gibi **randomizasyon hiçbir girdinin** **en kötü durum davranışına yol açmayacağını garanti eder.**
- **Bu algoritma aynı girdi de dahil her bir uygulamada farklı davranış gösterecektir.**

Evrensel Kıyım

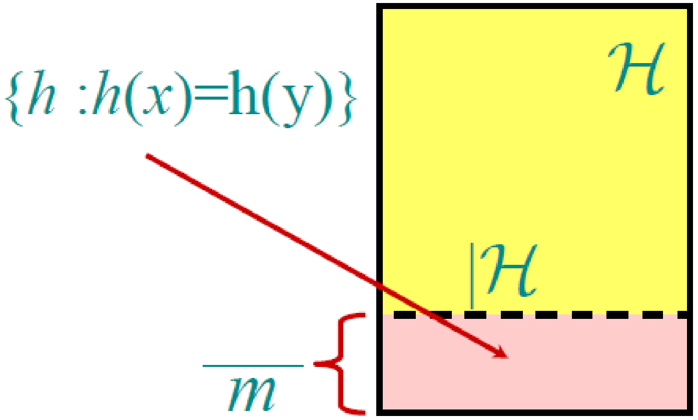
- Çırpı fonksiyonunu rastgele seçtiğimiz için **algoritma herhangi bir girdi** için <u>**ortalama performansı garanti eder.**</u>
- Bu algoritma **aynı girdi de dahil** <u>her bir uygulamada farklı davranış gösterecektir.</u>

Evrensel kıyım fonksiyonu

Tanım. U bir anahtarlar evreni ve $\mathcal{H}$ de sınırlı sayıdaki kıyım fonksiyonlarının kümesi olsun; herbiri U' yu $\{0, 1, \ldots, m-1\}$' e eşlemlesin.

$\mathcal{H}$ ' nin ***evrensel*** olması için: $x, y \in U$ ve $x \neq y$, ile $|\{h \in \mathcal{H} : h(x) = h(y)\}| = |\mathcal{H}|/m$ olması gerekir.

Yani, $\boldsymbol{x}$ ile $\boldsymbol{y}$ arasında bir çarpışma olasılığı : $1/m$ 'dir; koşul: h' nin $\mathcal{H}$' den rastgele seçimi.



Evrensellik iyidir

Teorem. h (tekbiçimli olarak) rastgele seçilmiş bir kıyım fonksiyonu olsun; seçim evrensel bir $\mathcal{H}$ kıyım fonksiyonları setinden yapılmış olsun. h' nin n rastgele anahtarı T tablosundaki m yuvaya kıyımladığını farzedin.
Bu durumda verilen bir x anahtarı için:

$$E[x \text{ ile çarpışma sayısı}] < n/m.$$

Teoremin kanıtı

Kanıt. C_x, T' nin içindeki anahtarlarla x' in toplam çarpışma sayısını gösteren rastgele değişken olsun; ve

$$c_{xy} = \begin{cases} 1 \text{ eğer } h(x) = h(y), \\ 0 \text{ (diğer durumlarda)} \end{cases} \quad \text{olsun.}$$

Not: $E[c_{xy}] = 1/m$ ve $C_x = \sum_{y \in T - \{x\}} c_{xy}$.

Teoremin kanıtı

$$E[C_x] = E\left[\sum_{y \in T-\{x\}} c_{xy}\right]$$

- Her iki tarafın da beklenenini bulun.

$$= \sum_{y \in T-\{x\}} E[c_{xy}]$$

- Beklenenin doğrusallığı (expectation).

$$= \sum_{y \in T-\{x\}} 1/m$$

- $E[c_{xy}] = 1/m$.

$$= \frac{n-1}{m} . \quad \square$$

- Cebir.

Bir evrensel kıyım fonksiyonları setini yapılandırmak

m asal sayı olsun. k anahtarını $r + 1$ basamağa ayrıştırın; herbirinin set içinde değeri $\{0, 1, \ldots, m-1\}$ olsun. Yani, $k = \langle k_0, k_1, \ldots, k_r \rangle$ ve $0 \le k_i < m$ olsun.

Rastgele yapma stratejisi:

$a = \langle a_0, a_1, \ldots, a_r \rangle$ olsun; burada a_i, $\{0, 1, \ldots, m-1\}$ arasından rastgele seçilmiştir.

Tanım: $h_a(k) = \sum_{i=0}^{r} a_i k_i \bmod m$. ***Nokta çarpım, mod m*** (ölçke)

$\mathcal{H} = \{h_a\}$ ne büyüklükte? $|\mathcal{H}| = m^{r+1}$.

* a'nın ilk değeri m tane olabilir. İkinci değeri de m tane.

Bir evrensel kıyım fonksiyonları setini yapılandırmak

ÖRNEK : a string uzunluğu **3** ve Tablo boyutu **131** olsun.

$$h_a(k) = \sum_{i=0}^{r} a_i k_i \bmod m$$

let a = <35, 100, 21>

$$h_a(\text{“xyz”}) = (35*120 + 100*121 + 21*122) \;\%\; 131$$
$$= 129$$

Dezavantajı: k$_i$ değeri tablo boyutundan büyük olabilir.
Bu yüzden **tablo boyutunu, k**$_i$ değerinden büyük seçilmeli

Nokta - çarpım kıyım fonksiyonların evrenselliği

Teorem. $\mathbb{H} = \{h_a\}$ seti evrenseldir.

Kanıt. $x = \langle x_0, x_1, \ldots, x_r \rangle$ olduğunu varsayın ve $y = \langle y_0, y_1, \ldots, y_r \rangle$ farklı anahtarlar olsun. Yani, en az bir basamakta farklı olsunlar ve log pozisyonu 0 olsun. Kaç $h_a \in \mathbb{H}$ için x ve y çarpışırlar?

$h_a(x) = h_a(y)$ olması gerekir ve bunun anlamı:

$$\sum_{i=0}^{r} a_i x_i \equiv \sum_{i=0}^{r} a_i y_i \pmod{m}$$

Kanıt (devamı)

Benzer yaklaşımla, elimizde

$$\sum_{i=0}^{r} a_i(x_i - y_i) \equiv 0 \pmod{m}$$

veya

$$a_0(x_0 - y_0) + \sum_{i=1}^{r} a_i(x_i - y_i) \equiv 0 \pmod{m},$$

olur ve bu da şu anlama gelir:

$$\underline{a_0}(x_0 - y_0) \equiv -\sum_{i=1}^{r} a_i(x_i - y_i) \pmod{m}.$$

Sayı teorisinin gerçeği

Teorem. m asal sayı olsun. Herhangi bir $z \in \mathbb{Z}_m$ ve $z \neq 0$ için, özgün bir $z^{-1} \in \mathbb{Z}_m$ vardır ve bu durumda:

$$z \cdot z^{-1} \equiv 1 \pmod{m}$$. (ölçke m) olur.

Örnek: $m = 7$.

z	1	2	3	4	5	6
z^{-1}	1	4	5	2	3	6

Kanıt (devamı)

Elimizde

$$a_0(x_0 - y_0) \equiv -\sum_{i=1}^{r} a_i(x_i - y_i) \pmod{m} \text{ var,}$$

ve $x_0 \neq y_0$, olduğundan tersi de $(x_0 - y_0)^{-1}$ olmalıdır, ve bu da şu anlama gelir:

$$a_0 \equiv \left(-\sum_{i=1}^{r} a_i(x_i - y_i)\right) \cdot (x_0 - y_0)^{-1} \pmod{m}.$$

Yani, herhangi bir $a_1, a_2, \ldots, a_r$, seçiminde tek a_0 seçimi, x ile y 'nin çarpışmasına neden olur.

Kanıt (tamamlanması)

S. Kaç tane h_a , x ile y' nin çarpışmasına neden olur?

C. Her $a_1, a_2, \ldots, a_r$ için m seçenek vardır ama, bunlar bir kez seçildiğinde, sadece bir tane a_0 x ile y' yi çarpıştırabilir, yani

$$a_0 = \left(\left(-\sum_{i=1}^{r} a_i (x_i - y_i) \right) \cdot (x_0 - y_0)^{-1} \right) \bmod m .$$

Böylece, çarpışmaya neden olabilecek h' lerin sayısı $= m^r \cdot 1 = m^{r-1} = |\mathcal{H}|/m.$

$a_1, a_2, \ldots a_r$; *Her biri m tane değer alır.*

a_0 değeri m olasılıktan 1 tane olacağı (çarpışmayı sağlayacak) için buraya 1 konuldu.

Mükemmel Kıyım

Şu ana kadar kıyımlar **beklenen zamanda başarımla ilgiliydi.**

Kıyım , beklenen süre bağlamında iyi bir uygulama.

Mükemmel kıyım ise şu sorulara ilgilenir :

- **Farz edin ki size bir anahtar kümesi verilsin.**
- **Statik bir tablo oluşturmanız istensin.**
- Böylece **en kötü zamanda** tabloda **anahtarı arayabileyim.**

Mükemmel Kıyım

- Elimizde sabit bir anahtar kümesi, **programlama dilindeki saklı kelimeler** kümesi ya da CD-ROM daki dosya isim kümeleri olabilir.

- Bir başka örnek vermek gerekirse ***İngilizcedeki en sık kullanılan 100*** *veya* ***1000*** *sözcük gibi bir şey.*

Mükemmel Kıyım

- Sözcüğün İngilizcede sık kullanılıp kullanılmadığına **tabloya bakarak hızlı bir şekilde anlamalıyız.**
- Bu işi ***beklenen başarımla değil*** de **garantilenmiş en kötü durum zamanında yapabilmeliyiz.**

Mükemmel Kıyım

Problem: Verilen ***n adet anahtar*** için **statik bir kıyım tablosu yaratmak.**

Diğer bir deyişle , *yeni girdi* veya silme yapılmayacak.

Sadece elemanları oraya koyacağız.

Büyüklüğü ise , ***m = O(n)***.

m = O(n) boyutunda **bir tablo** ve **en kötü durumda arama O(1) zamanı alacak.**

Mükemmel Kıyım

- Buradaki fikir, **iki aşamalı** bir veri tanımlaması yapmaktır.

  Fikir , **bir kıyım tablomuz olacak** , yuvalara kıyım yapacağız.
- Ancak zincirleme işlemini kullanmak yerine **ikinci bir kıyım tablosu daha olacak.**
- **İkinci tabloya ikinci bir kıyım daha yapacağız.**
- İkinci düzeyde hiç çarpışma olmadan kıyım yapmak.

Mükemmel Kıyım

- Birinci düzeyde çarpışma olabilir.
- Birinci tabloda çarpışan her şeyi, *ikinci düzeydeki tabloya koyacağız.*

Ama bu tabloda çarpışma olmayacak.

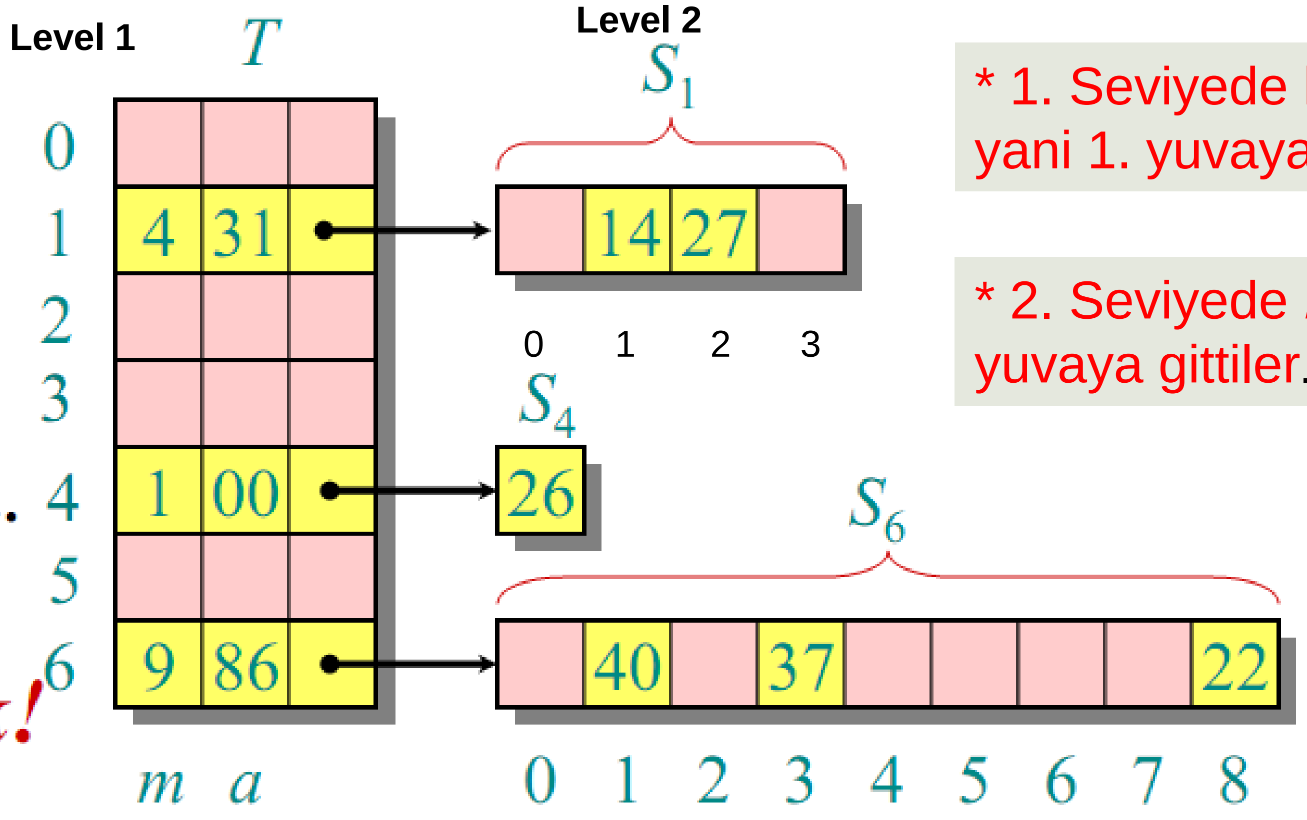
- Dolayısıyla evrensel bir kıyım fonksiyonu bulalım.
  - Rastgele bir fonksiyon seçiyoruz.
  - Yapacağımız bu düzeye yani ilk düzeye kıyım yapmak.

Mükemmel Kıyım Fonksiyonu

n anahtarlı bir set verilirse, bir statik kıyım tablosunu boyutu $m = O(n)$ olacak şekilde yapılandırın ve ARAMA (SEARCH) *en kötü durumda* $\Theta(1)$ süre alsın.

FİKİR: Her iki düzeyde de evrensel kıyım ile 2 düzeyli veri tanımlama.

2. düzey çarpışması yok!



* 1. Seviyede **h(14)=h(27)=1** yani 1. yuvaya gittiler.

* 2. Seviyede h_{31}**(14)=1** h_{31}**(27)=2** yuvaya gittiler.

Mükemmel Kıyım Fonksiyonu

TEOREM $\mathcal{H}$, evrensel kıyım fonksiyonu türlerinden biri olsun ve boyutu da $m = n^2$ olsun. Bu durumda, eğer bir rastgele $h \in \mathcal{H}$ ' yi n anahtarı tabloya kıyımlamakta kullanırsak, beklenen çarpışma sayısı en çok $1/2$ olur.

Kanıt. Evrenselliğin tanımı gereği, tablodaki belirli 2 anahtarın h altında çarpışma olasılığı $1/m = 1/n^2$ olur. Çarpışma olasılıklı $\binom{n}{2}$ çift anahtar olduğundan, çarpışmaların beklenen sayısı:

2 anahtar çifti seçilir.

$$\binom{n}{2} \cdot \frac{1}{n^2} = \frac{n(n-1)}{2} \cdot \frac{1}{n^2} < \frac{1}{2}.$$

Mükemmel Kıyım -Örnek

- Örnek : K={10,22,37,40,52,60,70,72,75} 9 elamanlı bir anahtar kümesi mod yani m=n=9 olur. Hash fonksiyonumuz: **h(k)=((a*k+b) mod p) mod m**
- a=3, b=42, p =101, m=9 (a ve b değerleri 0-101 arasında rastgele üretilen sayılar) Öncelikle ilk kıyım tablomuzda çakışmaların sayısını bulalım
- h(10)=0
- h(60)=2
- h(72)=2
- h(75)=2
- h(70)=5
- h(22)=7
- h(37)=7
- h(40)=7
- h(52)=7

0. ve 5. indiste 1 çakışma

2. indiste 3 çakışma

7. indiste 4 çakışma

	n_i
0	1
1	
2	3
3	
4	
5	1
6	
7	4
8	

Mükemmel Kıyım -Örnek

- Çakışmaları bulduktan sonra tek çakışmaya sahip değerler için a_i ve b_i değerlerini 0, diğerleri için ise 0-p arasında random seçelim, 2.kıyım (S_i)tablosunun büyüklüğü ise $m_i=n_i^2$ olacak
- 2.kıyım fonksiyonunda çakışma olmayacak şekilde yapılandıralım
- **$h_i(k)=((a_i*k+b_i) \bmod p) \bmod m_i$**

	n_i	m_i	a_i	b_i
0	1	**1**	0	0
1				
2	3	**9**	10	18
3				
4				
5	1	**1**	0	0
6				
7	4	**16**	23	88
8				

S_0

10
0

S_2

			60	**72**			**75**	
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

S_5

70
0

S_7

			40					**52**	**22**					**37**	
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

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