

## BİL3000 Mühendislik Projesi Dersi

### Ödev 1–Hash Tablosunda Kelime Arama

Liste ve dizi gibi veri yapıları içerisinde, belli bir değerin daha etkin aranabilmesi için Hash Tablosu kullanılır. Bu ödevde metin dosyasında bulunan kelimelerin Hash Tablosu mantığı kullanılarak bir diziye kopyalanması ve bu dizide arama işlemi yapılması istenmektedir.

**Metin dosyasında** her satırda bir kelime olacak şekilde en az 100 adet kelime bulunmalıdır. Kelimeler Türkçe karakter içermemelidir. Sadece küçük harflerden oluşmalıdır, bunu garantilemek için diziye yerleştirmek için ele alınan kelimenin veya aranacak kelimenin karakterleri küçük harfe dönüştürülmedir. Kelimelerin tümü Hash Tablosu mantığı ile bir diziye aktarılmalıdır ve arama işlemi bu dizide gerçekleştirilmelidir.

#### Aranacak bir kelimenin,

- dizide bulunması durumunda *“Girdiğiniz ..... kelimesi metin dosyasında bulunmaktadır”* gibi bir mesajla arama sonucu bildirilmelidir.

Aranan kelime dizide bulunmuyorsa,

- aranan kelime her seferinde sadece bir karakteri silinerek yeniden arama yapılmalıdır. Örneğin ilk aramada bulunamayan *“kedi”* kelimesi için *“edi”*, *“kdi”*, *“kei”* ve *“ked”* için de arama yapılmalı ve *“Girdiğiniz kedi kelimesi metin dosyasında edi ve kei olarak bulunmuştur”* benzeri bir mesajla benzerlerinin bulunduğu bildirilmelidir. ( $m$  karaktere sahip bir kelime için  $m$  olasılığın tümü için arama yapılmalıdır)
- aranan kelimenin her komşu iki karakteri yer değiştirilerek yeniden arama yapılmalıdır. Önceki örneğimiz olan *“kedi”* kelimesi için *“ekdi”*, *“kdei”* ve *“keid”* için de arama yapılmalıdır ve durumu özetleyen benzer bir mesajla arama sonucu bildirilmelidir. ( $m$  karaktere sahip bir kelime için  $m-1$  olasılığın tümü için arama yapılmalıdır)

#### Hash Tablosu:

Metin dosyasında bulunan kelimelerin diziye Hash Tablosu mantığıyla yerleştirilmesi için bir Hash fonksiyonu kullanılır. Bu mantığı anlamak için, öncelikle daha anlaşılır olmasından dolayı tam sayıların diziye yerleştirilmesi örneğini ele alalım. Dizimiz 10 elemanlı bir dizi olsun. Bu diziye sırasıyla 11, 28, 39, 38 ve 78 sayılarını yerleştirmek istiyoruz. Aşağıdaki şekilde yerleştirme mantığı açıklanmıştır.

11, 28 ve 39 eklenmesi	38 eklenmesi	78 eklenmesi																																																												
<table><tr><td>0</td><td></td></tr><tr><td>1</td><td>11</td></tr><tr><td>2</td><td></td></tr><tr><td>3</td><td></td></tr><tr><td>4</td><td></td></tr><tr><td>5</td><td></td></tr><tr><td>6</td><td></td></tr><tr><td>7</td><td></td></tr><tr><td>8</td><td>28</td></tr><tr><td>9</td><td>39</td></tr></table> <div>H=hash(key,N) =key%N 11%10=1, 28%10=8, 39%10=9</div>	0		1	11	2		3		4		5		6		7		8	28	9	39	<table><tr><td></td><td></td></tr><tr><td>11</td><td></td></tr><tr><td>38</td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td>28</td><td></td></tr><tr><td>39</td><td></td></tr></table> <div>H=hash(38,10)=8 i=0, (8+0)%10=8, dolu i=1, (8+1)%10=9, dolu i=2, (8+4)%10=2, boş</div>			11		38												28		39		<table><tr><td></td><td></td></tr><tr><td>11</td><td></td></tr><tr><td>38</td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td></td><td></td></tr><tr><td>78</td><td></td></tr><tr><td>28</td><td></td></tr><tr><td>39</td><td></td></tr></table> <div>H=hash(78,10)=8 i=0, (8+0)%10=8, dolu i=1, (8+1)%10=9, dolu i=2, (8+4)%10=2, dolu i=3, (8+9)%10=7, boş</div>			11		38										78		28		39	
0																																																														
1	11																																																													
2																																																														
3																																																														
4																																																														
5																																																														
6																																																														
7																																																														
8	28																																																													
9	39																																																													
11																																																														
38																																																														
28																																																														
39																																																														
11																																																														
38																																																														
78																																																														
28																																																														
39																																																														

**Şekil:** Hash Tablosu yerleşimi ve çakışmanın (collision) ikinci dereceden tarama (*quadratic probing*) ile çözülmesi

Burada kullanılan  $hash(key, N)$  bir hash fonksiyonudur ve diziye yerleştirilecek tam sayının ( $key$ ) dizi boyutuna ( $N$ ) göre modunu almaktadır. Hash fonksiyonu tam sayının dizide yerleşeceği konumun indisini üretmektedir. 38 ve 78 sayılarının yerleştirileceği konum bulunurken çakışma (*collision*) ortaya çıkmıştır ve bu sorun ikinci dereceden tarama (*quadratic probing*) yöntemiyle çözülmüştür.

**Kelimelerin tabloya yerleştirilmesinde** ise farklı hash fonksiyonları kullanılabilir. Bunlardan biri, kelimenin içerdiği karakterlerin ASCII karşılıklarının kelime içerisinde bulundukları konum ile ağırlıklandırılarak toplanması ve elde edilen değer kelimenin yerleştirileceği dizinin boyutuna göre modunun alınmasıdır. Örneğin “*kedi*” kelimesi için,

$$\text{TOPLAM} = 1 \times 107(k) + 2 \times 101(e) + 3 \times 100(d) + 4 \times 105(i) = 1029$$
$$H = \text{hash}(1029, 100) = 29$$

Bu örneğe göre 100 elemanlı bir dizide  $dizi[29]$  değeri “*kedi*” olmalıdır veya bir şekilde “*kedi*” kelimesi ile ilişkilendirilmelidir.

### Program Yapısı:

Kısaca tanımlamak gerekirse, programınız ana fonksiyon içerisinde çağrılan fonksiyonlardan oluşmalıdır. Mutlaka ayrı birer fonksiyon yazılması gereken işler şunlardır:

- metindeki kelimelerin diziye Hash Tablosu mantığı ile yerleştirilmesi ve yerleşim sonucunda oluşan dizinin ekranda gösterilmesi,
- verilen kelimenin dizide aranması,
- bulunamayan kelimenin karakterlerinin eksiltiyle dizide aranması (a),
- bulunamayan kelimenin komşu karakterlerinin yer değiştirilmesiyle dizide aranması (b)

Fonksiyonların döndürdüğü değerlere bağlı olarak ekranda gerekli mesajlar ana fonksiyonda verilmelidir. Fonksiyon ve programda kullanılan tüm değişken, sabit v.s. adları üstlendikleri görev veya kullanıldıkları amaca uygun olmalı ve gerekli kısaltmaları içermelidir. Programınızda gerekli yerlerde gerekli açıklama satırları bulunmalıdır.

**Henüz hatalardan arındırılmayan (çalışmayan) programlar değerlendirme dışı bırakılacaklardır!!!**

**Tablo: Puanlama**

Madde	Açıklama	Yüzde (%)
Program Yapısı (Ana fonksiyon ve diğer fonksiyonlar) (%100)		10
İşlevsellik	<ul style="list-style-type: none"><li>• Dosyanın okunması ve kelimelerin ayrıştırılması (Reading and parsing file) (%10)</li><li>• Kelimelerin Hash Tablosu mantığında diziye yerleştirilmesi (%20)</li><li>• Verilen kelimenin aranması (%20)</li><li>• Karakter eksiltiyle aranması (a) (%25)</li><li>• Karakterlerin yer değiştirilerek aranması (b) (%25)</li></ul>	60
Ödev Raporu		10
Hocanın ödevi anlattığı derse katılım		10
Sunum (video)		10

**Önemli Uyarı:** Hocanın ödevi anlattığı derste yoklama alınacaktır. Katılmayan öğrencilerin ödevin tekrar anlatılması talepleri tuhaf karşılanacak ve karşılık bulmayacaktır!!!