

## Konu:Hashing

**Problem:** Bu ödevde harf yazım kontrolü (spell checker) işlemi yapan bir sistem tasarlanacaktır. Sistem, dışarıdan verilen bir kelimenin sözlükte olup olmadığını kontrol edecek, kelimeyi sözlükte bulursa ‘kelime doğru yazılmıştır’ mesajı verecek, kelimeyi sözlükte bulamazsa sözlükte bulunan kelimelerden doğru kelime önerilerini verecektir. Bu ödevde hash tablosu oluşturulacak sözlük için ekte verilen **smalldictionary.txt** dosyası kullanılacaktır. Ödev üç ana bölümden oluşacaktır :

1. **Hash Tablosu Oluşturma** : Doğru yazılmış kelimelerin saklandığı sözlük dosyasındaki (ekteki smalldictionary) kelimeler hash tablosuna yerleştirilir. Bu işlem bir defa yapılacaktır.
2. **Verilen Kelimeyi Hash Tablosunda Arama**: Yazım kontrolünden geçirmek için dışarıdan okunan kelime hash tablosunda aranır. Eğer kelime sözlükte bulunamazsa, yanlış kelime yerine sözlükte bulunan benzer kelimeler önerilir.
3. **Yanlış Kelime Yerine Kelime Önerme** : Yanlış kelime yerine kelime önerme işlemi için her seferinde verilen kelimedeki tek harfi değiştiren basit bir yöntem kullanılacaktır. Bu yöntemde kelimedeki her harf sırası ile ‘a’ dan ‘z’ ye harflerle değiştirilerek elde edilen yeni kelimenin hash tablosunda olup olmadığına bakılacaktır. Eğer elde edilen yeni kelime hash tablosunda varsa yanlış kelime yerine önerilen kelimelerden biri olarak ekrana yazdırılacaktır. Örneğin yanlış kelime **sate** olsun. Bunun yerine sistemin önerebileceği kelimeler **late, same, save, ...** olabilir.

**Hash Tablosu aşağıdaki şartları sağlayacak şekilde oluşturulmalıdır:**

1. Tabloya yerleştirilecek büyük harflerden oluşan bir kelimenin sayı karşılığını aşağıdaki gibi hesaplayınız.

```
for(i=key=0;i < length(str); i++)  
    key += (str[i] - 'a')*26*i ;
```

2. Hash tablosunu oluştururken **OpenAddress, Çakışma(collusion)** problemini çözmek için **Quadratik Probing** yöntemini kullanınız. Hash fonksiyonu olarak bölme (division) yöntemini kullanınız. Hash fonksiyonlarını aşağıdaki bağıntıları kullanarak belirleyiniz:  $c_1=2$ ,  $c_2=1$  alınız

```
h(key, i) = (h1(key) + i^2) % m  
h1(key) = h'(key) + c1*i + c2*i^2)
```

3. **Tablo uzunluğunu gösteren m değerinin belirlenmesi için** aşağıdaki bağıntıyı kullanınız :

$m = \text{EnküçükAsalSayı} \geq \text{TablodakiElemanSayısı} / \text{LoadFactor}$

R değerini m’den küçük ilk asal sayı olarak alınız.

4. LoadFactor’u 0.60 alınız. Tablo uzunluğunu belirleme işlemi elle yapınız, bunu için program yazmayınız. Program çalışmaya başladığında hash tablosu oluşturulmasının istenip istenmediğini kullanıcıya sormalı, eğer oluşturması isteniyorsa hash tablosunu oluşturmalı, istenmiyorsa mevcut hash tablosunun saklı olduğu dosya ismini sormalı ve dosya var ise bu dosyadan hash tablosunu okumalıdır.

1. Yaptığınız çalışmayı yöntem ve uygulama bölümlerinden oluşan bir raporda anlatınız.

2. Yöntem bölümünde problemi ve gerçekleştirdiğiniz çözümü kısaca anlatıp,

algoritmanıza ait pseudo code'unu yazınız.

3.Uygulama bölümünde aşağıdaki 20 kelime için kelimelerin hash tablosunda aranma adım sayısını hesaplayınız. Ortalama aranma sayısının ikili arama yönteminden daha hızlı olup olmadığını değerlendiriniz.

FACE FACT GOLE GOLF HOLE HOLK BISK BITE DAME DAMN  
NIMS NINE PAIN PAIR LIVE LOAD YARD YARE ZERO ZEST

### **Ödev Teslimi ile ilgili önemli bilgiler:**

Aşağıda verilen bütün bilgileri içeren tek bir belge hazırlayarak

**23.04.2021 saat 17:00'a** kadar [online.yildiz.edu.tr](http://online.yildiz.edu.tr) sistemi üzerinden

**HW2\_OgrenciNumarasi.rar** dosyasını yükleyiniz.

### **Teslim Edilecekler:**

1. Yaptığınız çalışmayı yöntem ve uygulama bölümlerinden oluşan bir raporda anlatınız.

**Yöntem** bölümünde problemi ve çözüm için önerdiğiniz yöntemi kısaca anlatınız

2. **Uygulama** bölümünde test verileri ile elde edeceğiniz sonuçların ekran görüntülerini açıklayarak gösteriniz

3. Algoritmanızın **C** dilinde programını hazırlayarak dokümana da ekleyiniz.

#### **4. Teslim Edilecekler**

a. HW#\_OgrenciNumarasi.rar (Örn: HW2\_15011001.rar)

i. OgrenciNumarasi.pdf (Örn: 15011001.pdf)

ii. OgrenciNumarasi.c (Örn: 15011001.c)

### **Değerlendirme**

#### ***Algoritma Tasarımı ve Programın Çalışması: (%80)***

1. Ödev, istenilen işlerin tamamını yerine getirmelidir.
2. Gereksiz kontrollerden ve işlemlerden arınmış bir tasarım yapılmalıdır.
3. Program hatasız çalışmalıdır.
4. Programın çalışması sırasında, konuyu bilmeyen kişilerin rahatlıkla anlayabilmesi için, giriş ve çıkışlarda mesajlarla bilgi verilmelidir.

#### ***Rapor Dokümantasyonu: (%20)***

1. Raporun kapak sayfasında, dersin adı, öğrencinin ad, soyad ve numarası, ödev konusu bilgileri yer almalıdır.
2. Kaynak kodda değişken deklarasyonu yapılırken her değişken tek satırda tanımlanmalı, tanımın yanına değişkenin ne için kullanılacağı açıklama olarak yazılmalıdır.
3. Değişken isimleri anlamlı olmalıdır.
4. Her fonksiyonun yaptığı iş, parametreleri ve dönüş değeri açıklanmalıdır.
5. Gerekli yerlerde açıklama satırları ile kodda yapılan işlemler açıklanmalıdır.
6. Gereksiz kod tekrarı olmamalıdır.
7. Kaynak kodun formatı düzgün, okunabilir ve takip edilebilir olmalıdır.