

2021-2022 Güz Yarıyılı Algoritma Analizi 3. Ödevi

Ders Yürütücüleri
Prof. Dr. M. Elif KARSLIGİL
Doç. Dr. M. Amaç GÜVENSAN

Konu: Hashing

Problem: Bu ödevde sorgulanan anahtar kelimelerin hangi web sayfasında olduğunu bulan basit bir *arama motoru* algoritması tasarlamamız ve kodlamamız istenmektedir. Problemde hash tablosunu hangi aşamada ve nasıl kullanacağınızı sizin belirlemeniz beklenmektedir.

Açıklama: Web sayfasının linki ve bu sayfada bulunan anahtar kelimeler bir text dosyada satırlar halinde aşağıdaki gibi verilmektedir. Arama için bir veya birden fazla kelime aralarında **ve,veya** bağlaçları kullanılarak verildiğinde şartı sağlayan web sayfası linkleri verilmelidir.

Örnek:

Bilgileri içeren dosya örneği:

<http://www.mevsimler.com>

sonbahar yaz ilkbahar kış kar güneş yağmur

<http://www.meteoroloji.org>

sis kar yağmur

<http://www.arac.com.tr>

motorsiklet kamyon tır bisiklet

Sorgu Örnekleri:

- kar ve yağmur
<http://www.mevsimler.com>
<http://www.meteoroloji.org>
- kamyon
<http://www.arac.com.tr>
- sis ve kar
<http://www.meteoroloji.org>
- motorsiklet veya yaz
<http://www.mevsimler.com>
<http://www.arac.com.tr>
- kardanadam
Bu kelimenin geçtiği doküman bulunmamaktadır.
- motorsiklet ve yaz
Bu iki kelimenin birlikte geçtiği doküman bulunmamaktadır.

Hash Tablosu:

Hash tablosunu aşağıdaki şekilde tasarlayınız.

1. Tabloya yerleştirirken indis belirlemek için kullanılacak anahtarların string olduğunu varsayarsak bu string'in sayı karşılığını Horner kuralını kullanarak hesaplayınız.
2. Hash tablosunu oluştururken **open-addressing**, çakışma(collision) problemini çözmek için **linear probing** yöntemini kullanınız.
3. Hash tablosunun uzunluğunu, giriş dosyası verildiğinde hash tablosuna yerleşecek eleman sayısını hesaplayarak, Load-Factor değerine göre olabilecek en yakın asal sayı olarak belirleyiniz. Load-Factor değerini dışardan okuyunuz. Hash tablosunun uzunluğunu belirlemek için, bir sayının asal sayı olup olmadığını bulma işlemini de içeren bir fonksiyon yazınız.

Ödev Süresi:

1. Ödevinizi **20.11.2022 Pazar saat 23:59'a** kadar online.yildiz.edu.tr üzerinden **HW3_OgrenciNumarasi.rar** dosyasını yükleyiniz.
2. Ödev hakkında sorularınızı 18.11.22 Cuma Günü Saat 17.00'a kadar Classroom grubundan sorabilirsiniz.

Ödevinizi hazırlarken Classroom Sayfasında paylaşılan Ödev Teslim Kuralları başlığındaki kurallara uymanız gerekmektedir.

Teslim Edilecekler:

1. Ödevinizin çalışmasını ekteki **arama.txt** dosyası için deneyiniz.
2. Yaptığınız çalışmayı yöntem, uygulama ve sonuç bölümlerinden oluşan bir raporda anlatınız.
 - a. **Yöntem** bölümünde problemi ve çözüm için önerdiğiniz yöntemi adım adım anlatınız. Hash tablonuzu ve tablo için tasarladığınız struct'ın bileşenlerini kısaca açıklayınız.
 - b. **Uygulama** bölümünde **arama.txt** dosyasını giriş olarak alıp, aşağıdaki durumlar için elde ettiğiniz sonuçlara ait program çıktılarınızın ekran görüntülerini ekleyerek gösteriniz. Programınız kullanıcının isteğine göre normal ve detay modu olmak üzere iki durum için çıktı vermelidir:
 - i. **Normal modda çalıştırma:** Programınız *normal modda* çalıştırıldığında verilen kelimeler için bu kelimelerin bulunduğu web sayfalarının linkini veriniz.
Örnek:
 > sis veya yaz
<http://www.mevsimler.com>
<http://www.meteoroloji.org>
 > sonbahar ve kar
<http://www.mevsimler.com>
 - ii. **Detay modda çalıştırma:** Programınız *detay modda* çalıştırıldığında verilen giriş dosyası için aşağıdaki bilgileri ekrana yazdırınız:
 1. Hash tablosunun uzunluğu
 2. Hash tablosunu oluşturan her kayıt ve tabloya kaçınıcı denemede yerleştiği
 3. Tablo tamamlandığında hash tablosunun içeriği (her indiste hangi bilgi olduğu)

4. Normal moddaki sorgular ve çıktıları

c. **Sonuç** bölümünde

- i. Programınızı load factor'un 0.1, 0.5 ve 0.9 değerleri için çalıştırarak çakışma durumunun nasıl değiştiğini gözlemleyiniz ve bununla ilgili yorumlarınızı yazınız.
- ii. Hash tablosunun uzunluğunu belirleyen fonksiyonunuzun karmaşıklığını hesaplayınız.
- iii. Arama motoru algoritmanızın tamamının karmaşıklığını (hash tablosunun uzunluğunu belirleyen fonksiyonunuzu hesaba katmadan) aşağıdaki iki durum için hesaplayarak sonucu 1-2 cümle ile yorumlayınız.
 1. hashing yerine başka hangi yöntemi kullanabilirdiniz? Bu yöntemi kullandığınızda en kötü durumdaki karmaşıklık ne olur?
 2. hashing yöntemi kullandığınızda en kötü durumdaki karmaşıklık ne olur?

3. Kısa bir **video** (5-10 dk.) hazırlayınız. Video içeriğinde problemi, geliştirdiğiniz çözümü ve kodunuzu anlatınız. Ardından farklı girdiler ile programınızın çalışmasını gösteriniz. Video linkini raporunuza ekleyiniz. Video linkini public paylaşmayınız, kopyaya sebep olmaktadır.

Teslim Edilecek Dokümanlar:

- a. HW#_OgrenciNumarasi.rar (Örn: HW3_15011001.rar)
 - i. OgrenciNumarasi.pdf (Örn: 15011001.pdf)
 - ii. OgrenciNumarasi.c (Örn: 15011001.c)
 - iii. Uygulama video linki