

2020-2021 Güz Yarıyılı

Algoritma Analizi

Ödev – 2

Ders Yürütücüleri

Doç. Dr. M. Elif KARSLIGİL

Dr. Öğr. Üyesi M. Amaç GÜVENSAN

Konu : Hashing Algoritmasının Kullanımı

Problem: Bu ödevde, bir kelimenin geçtiği dokümanları listeleyen bir sistem tasarlanacaktır. Bir kelimenin hangi dokümanlarda geçtiğini bulmak için her seferinde bütün dokümanlara bakmak çok zaman alıcıdır. Bunun yerine bu ödevde, yeni gelen bir dokümandaki kelimeler hashing ile bir sözlüğe yerleştirilecek ve bir kelime arandığında yine hashing yöntemi ile sözlükte aranarak içinde yer aldığı dokümanlar bulunacaktır.

Sistem 3 alt bölümden oluşacaktır :

- 1. Yeni gelen dokümanı sözlüğe ekleme :** Bir doküman geldiğinde dokümanda bulunan kelimeleri hash fonksiyonundan geçirerek eğer hash tablosunda yoksa hem kelimeyi hem de dokümanın adını tabloya ekleyin, kelime tabloda varsa sadece dokümanın adını bu kelimenin bulunduğu adrese ekleyin.
Hash tablosunu bir dosyada saklayın. Program yeniden çalıştırıldığında mevcut dosyayı kullanarak, yeni bilgileri ekleyin.
- 2. Sorgulanan kelimeyi arama :** Bir kelime sorgulandığında hash fonksiyonundan geçirerek kelimeyi tabloda arayın. Eğer tabloda varsa geçtiği dokümanların isimlerini ekrana yazdırın. Tabloda yoksa, olmadığını mesajla belirtin. Arama işleminin kaç adımda tamamlandığını ekrana yazdırınız.
- 3. Hash tablosunu oluşturma :** Hash tablosu aşağıdaki kurallara göre oluşturulacaktır :
 - Hash tablosunu oluştururken *openaddress*, çakışma problemini çözmek için *double hashing* yöntemleri kullanılacaktır. Buna göre:
$$h(\text{key}, i) = [h1(\text{key}) + i * h2(\text{key})] \bmod M$$
$$h1(\text{key}) = \text{key} \bmod M$$
$$h2(\text{key}) = 1 + (\text{key} \bmod MM)$$
 - Dosyaların içinde sadece kelimeler olduğu, özel karakterler olmadığı varsayılmıştır. Kelimeler küçük-büyük harf karışık olabilir. Örneğin “Araba” ve “araba” aynı kelimelerdir. Kelimeleri sayıya çevirirken Horner Metodu kullanılacaktır.
 - Tablo uzunluğunu gösteren **M** değerini 1000 değerine en yakın ASAL sayı olarak alınız.
 - h2 fonksiyonunda $MM = M - 1$ alınız.
 - Loadfactor’ü her yeni eklenen kelime için ekrana yazdırıp LoadFactor 0.8’i geçtiğinde kullanıcıya UYARI MESAJI gösteriniz. LoadFactor 1.0 olduğunda EKLEME işlemini DURDURUNUZ ve dosyada EKLENEMEYEN kelimeleri EKRANA yazdırınız.
 - İPUCU:** LoadFactor’ün son değerini dosyada saklayınız. Yeni eklenen kelimeler için bu değeri GÜNCELLEYİNİZ.

Ödev Teslimi ile ilgili önemli bilgiler:

Aşağıda verilen bütün bilgileri içeren tek bir doküman hazırlayarak **02.12.2020 saat 23:59'a** kadar online.yildiz.edu.tr adresi üzerinden **OgrenciNumarasi.rar** dosyasını yükleyiniz.

Teslim Edilecekler:

1. Algoritmanızın **C** dilinde programını hazırlayarak dokümana da ekleyiniz.
2. Her algoritma için farklı durumlarını gösteren ekran çıktılarını dokümanda paylaşınız.
3. İlk soruda her iki şık için karmaşıklık analizini yapınız.
4. **Teslim Edilecekler**
 - a. OgrenciNumarasi.rar (Örn: 15011001.rar)
 - i. OgrenciNumarasi.pdf (Örn: 15011001.pdf)
 - ii. OgrenciNumarasi.c (Örn: 15011001.c)
 - iii. OgrenciNumarasi.txt (Hash Tablosunun saklandığı dosyanın son hali-Örn:15011001.txt)

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.
8. **ÖRNEK ÇIKTI EKLANLARINI mutlaka rapora ekleyiniz.**