

Yapay Zeka Temelleri 2. Ödevi

Bu ödevde, bir dizi string'i (metinsel ifade) sayılara dönüştürerek her bir string'in bir sayı karşılığı olması sağlanacaktır. String'ler sayıya Horner kuralı kullanılarak ve aşağıdaki açıklanan hash fonksiyonları kullanılarak dönüştürülmelidir.

$$\text{key} = s_0 \times \text{base}^{n-1} + s_1 \times \text{base}^{n-2} + \dots + s_{n-1} \times \text{base}^0$$

S_0, S_1, \dots, S_{n-1} sırasıyla string'deki karakterlerin ASCII kod değeri karşılıklarıdır. n string'in uzunluğudur (örnek: "ABCD" için $n = 4$). "base" bir sayı olup (örneğin 31, 37, 101 vb.) aşağıdaki formülde kullanılacaktır.

Örnek: Diyelim ki "AB" metnini (2 karakter) $\text{base} = 31$ ile sayıya dönüştürelim:

- A karakterinin ASCII değeri 65 (örnek olarak alıyoruz).
- B karakterinin ASCII değeri 66.
- $n = 2$.

$$\begin{aligned} \text{key}(\text{"AB"}) &= 65 \times 31^{2-1} + 66 \times 31^{2-2} \\ &= 65 \times 31 + 66 \times 1 \\ &= 2015 + 66 = 2081 \end{aligned}$$

Artık "AB" için 2081 sayısını elde etmiş oluyoruz.

Daha sonra elde ettiğimiz bu key değerini aşağıdaki formülde kullanarak string'in sayısal karşılığını hesaplıyoruz.

$$h(\text{key}, i) = (h_1(\text{key}) + i \times h_2(\text{key})) \mod m$$

- **key:** String'den sayıya dönüştürülmüş değerdir. (Örnekte "AB \rightarrow 2081).
- $h_1(\text{key})$: Birinci hash fonksiyonu sonucu.
- $h_2(\text{key})$: İkinci hash fonksiyonu sonucu.
- i : Çakışma sayısı veya deneme sayısı (i 'den başlar, her çalışmada 1 artar).
- m : Hash tablosunun boyutu.

Burada:

$$h_1(\text{key}) = (\text{key}) \bmod m$$
$$h_2(\text{key}) = 1 + ((\text{key}) \bmod m2)$$

ve

$$m2 = m - 3$$

Örnek: Elimizde 4 adet (string1, string2, string3, string4) string olsun. $m = 11$, $m2 = 8$ olarak belirlenmiş olsun. string1 için Horner kuralı ile key'i hesaplandıktan sonra $i = 0$ için $h(\text{key}, i)$ hesaplanır. Bu işlem sonucunda oluşan sayı daha önce bu string için oluşturulmadıysa string1'in karşılığı belirlenmiş olur. Daha sonra aynı işlem string2 için gerçekleştirilir ve $i = 0$ için oluşan sayı daha önce oluşmuş ise (çakışma olmuş demektir) string2 için $i = 1$ kullanılarak yeni bir sayı hesaplanır ve bu şekilde çakışma olmayana kadar i sayısı 1 artırılır. string2 için çakışmayan sayı bulunduktan sonra string3 ve string4 için aynı işlemler ile sayısal karşılıklar ve çakışma sayıları hesaplanır. Sonuç olarak 4 string için 4 farklı sayı oluşmuş olacaktır.

Görev: Hash hesaplama formülündeki base değerini asal bir sayı seçmenin çakışma sayısını azaltıp azaltmadığını Monte Carlo simülasyonu kullanarak incelemenizdir. Bunun için rastgele stringler oluşturularak base değerini bir asal ve bir asal olmayan (ve bir yüksek ve bir düşük) sayı seçerek çakışma oranlarını hesaplayınız.

Tüm denemelerde m değerini tabloya ekleyeceğiniz string sayısının iki katından büyük ilk asal sayı olacak şekilde belirleyiniz.

Teslim etmeniz gerekenler:

- .py formatında kodlarınız.
- .pdf formatında hazırlanmış rapor.

Rapor İçeriği:

- Seçtiğiniz taban (base) değerleri.
- Elde ettiğiniz çakışma sayıları.
- Asal / asal olmayan taban farkı.
- Gözlemlerinizi ve sonuçlarımız.

Son teslim tarihi: Ödevinizi numara.zip şeklinde isimlendirerek online.yildiz.edu.tr adresi üzerinden 10 Ocak 2025 Cuma 23.59 tarihine kadar teslim ediniz.