**Ayşe Melis Akçakaya**

**2024481022**

**Ünite sonu soruları**

**10. hangi veri türü, trie veya bloom filtres, alan açısından daha verimlidir?**

**#**Trie ve Bloom filtresi, farkli kullanım senaryolarına ve gereksinimlere sahiptir, bu nedenle hangisinin daha verimli olduğu konusunda kesin bir yanıt vermek zordur. Ancak, genel olarak konuşursak:

Alan Açısından Verimlilik:

Trie yapısı, anahtarlarını saklanması için her bir karakterin ayrı bir düğümde temsil edilmesi gerektiğinden, genellikle bellek alanı açısından daha fazla yer kaplar. Özellikle anahtarlar çok uzunsa veya çok sayıda farklı karakter içeriyorsa, trie yapısı daha fazla bellek kullanabilir.

Bloom filtresi ise genellikle sabit boyutlu bir bit dizisi kullanır. Bu nedende, saklanan öge saysına bağlı olarak sabit bir bellek kullanımı vardır. Ancak, bloom filtresi yanlış pozitif sonuçlar üretebilir, bu da yanlış pozitiflerin sayısını azaltmak için filtre boyutunu artırmayı gerektirebilir.

Genel olarak, alan açısından bakıldığında, bloom filtresi genellikle daha az bellek kullanır.

Hız Açısından Verimlilik

Trie yapısı, anahtarların aranması için tipik olarak daha hızlıdır çünkü her bir karakterin ayı bir düğümde temsil edilmesi, arama işlemlerini çok hızlı hale getirir. Ancak, bu hız, bellek kullanımının artmasıyla birlikte gelir.

Bloom filtresi ise tipik olarak çok hızlıdır çünkü bir ögenin kümede olup olmadığını belirlemek için sabit bir işlem yapar. Ancak, yanlış pozitif sonuçlar olabilir ve bu yanlış pozitifleri azaltmak için ek işlemler gerekebilir, bu da hızı etkileyebilir.

Genel olarak, alan ve hız açısından bakıldığında, bloom filtresi genellikle daha hızlıdır.

Sonuç olarak, alan ve hız açısından hangi veri yapısının daha verimli olduğu, kullanım senaryosuna ve gereksinimlere bağlıdır.

**Programlama problemleri**

**1.metin web sitesine gidin ve kelime sözlüğünü indirin. Bu kelime listesi için bir bloom filtresi oluşturun ve bağımsızlık bildirgesini yazım denetimi yapmak için kullanın, yanlış yazılan tüm kelimeleri ekrana yazdırın.**

**#**

import requests

import re

from bloom\_filter2 import BloomFilter

# 1. Adım: Kelime sözlüğünü indir (örnek URL)

dictionary\_url = "https://raw.githubusercontent.com/csariyildiz/turkish-wordlist/master/words.txt"

try:

response = requests.get(dictionary\_url)

response.raise\_for\_status()

dictionary\_words = response.text.splitlines()

print(f"Sözlükte {len(dictionary\_words)} kelime bulundu.")

except Exception as e:

print("Kelime sözlüğü indirilemedi:", e)

dictionary\_words = []

# 2. Adım: Bloom filtresi oluşturun

# Burada, sözlükteki kelime sayısını maksimum eleman sayısı olarak belirliyoruz

bloom = BloomFilter(max\_elements=len(dictionary\_words) or 100000, error\_rate=0.001)

for word in dictionary\_words:

# Tüm kelimeleri küçük harfe çeviriyoruz

bloom.add(word.lower())

# 3. Adım: Bağımsızlık Bildirgesi metnini indir (örnek URL)

declaration\_url = "https://raw.githubusercontent.com/abovetheclouds/declaration\_of\_independence/master/declaration.txt"

try:

resp\_decl = requests.get(declaration\_url)

resp\_decl.raise\_for\_status()

declaration\_text = resp\_decl.text

except Exception as e:

print("Bağımsızlık Bildirgesi metni indirilemedi:", e)

declaration\_text = ""

# 4. Adım: Metni kelimelere ayırın ve yazım denetimi yapın

# Burada basit bir regex ile kelimeleri çekiyoruz

words\_in\_text = re.findall(r'\b\w+\b', declaration\_text.lower())

# Yazım hatası olarak kabul edilen kelimeleri tutmak için bir küme oluşturuyoruz

misspelled\_words = set()

for word in words\_in\_text:

if word not in bloom:

misspelled\_words.add(word)

# Sonuçları ekrana yazdırın

print("\nYanlış yazılan (sözlükte bulunmayan) kelimeler:")

for word in sorted(misspelled\_words):

print(word)

Kelime Sözlüğü İndirme:

requests.get yöntem ile belirtilen URL'den kelime listesi çekiyoruz ve satır satır ayrıştırıyoruz.

Bloom Filtresi Oluşturma:

İndirilen kelimelerin sayısına göre filtre kapasitesi ayarlanıyor. Hata payı %0,1 (0,001) olarak belirlendi.

Metnin İşlenmesi:

Bağımsızlık Bildirgesi metni indirildikten sonra, regex yoluyla tüm kelimeler elde edilir.

Her kelime küçük harflere dönüştürülüyor ki sözlükle uyumlu olsun.

Yazım Denetimi:

Eğer kelime Bloom filtresinde yoksa, bu kelime “yanlış” veya sözlükte bulunmuyorsa kabul ediliyor ve sonuç ayarlarına ekleniyor.