ÜYELİK YAPILARI 8

Berfin İSSİ  
2024481023

8.3.3 TRİE ÜYELİK

Bir Trie’deki üyeliğin kontrol edilebilmesi özyinelemeli olarak da gerçekleştirilebilir. Adımlar, ilk başta tamamen sezgisel olmayabilecek bir temel durum içerir. Boş anahtar herhangi bir Trie’nin üyesi olarak rapor edilir çünkü üyelik kontrol edilirken çalışır. Trie’ye eklenen sentinel birimi ile, boş bir anahtar için True döndürmek tamamen güvenlidir çünkü herhangi bir gerçek anahtar en azından sentinel karakterlerden oluşacaktır. Burada özetlenen algoritmada, sentinelin anahtara zaten eklenmiş olduğu varsıyılır. Üyelik testi için adımlar aşağıdaki gibidir:  
1. Anahtarın uzunluğu 0 ise, True döndürerek başarıyı raporlayın.  
2. Eğer baktığımız düğüm None ise False döndürerek başarısızlığı bildirir.  
3. Anahtarın ilk birimi geçerli düğümdeki birimle eşleşiyorsa, aşağıdaki   
düğümden başlayarak anahtarın üyeliğini kontrol edin.  
 4. Aksi taktirde, Trie’deki bir sonraki düğümden başlayarak anahtarın üyeliğini kontrol edin.  
Yine bu kod, bir while döngüsü ile iteratif olarak uygulanabilir. Geçerli düğümün ve anahtarın geri kalanının izini sürer. Öz yenilemeli veya yenilemeli uygulaması eşit derecede iyi çalışacaktır.

8.3.4 TRİE DENEMELERİ VE BLOOM FİLTELERİNİN KARŞILAŞTIRILMASI

Bloom filtreler, üyeliği test etmek için Trie’den çok daha hızlıdır. Bununla birlikte Trie kabul edilebilir derecede iyi çalışır. En uzun Bloom filtresi arama süresi basit bir deneyde 160 µs, en uzun Trie araması ise 217 µs olmuştur. Tabii ki Trie daha fazla yer kaplar, ancak ortak önekler bir Trie’deki düğümleri paylaşarak depolamaya göre biraz yer tasarrufu sağlar. Her bir kelime, bir hash kümesinde olduğu gibi, bir veri yapısında ayrı ayrı yer alır. Yazım denetimi amacıyla bir Trie’nin belirgin avantajları vardır, çünkü yazım alternatifleri kolayca bulunabilir. Yaygın yazım hataları dört katagoriden birine girer:   
Teh gibi karakterlerin the yerine transpoze edilmesi  
Their yerine thei gibi karakterler bırakılması  
Thre gibi ekstra karakterler yerine the   
There yerine thare gibi yanlış karakterler  
Bir Trie içinde arama yapıldığında bir kelime bulunmazsa, alternatifler yazımların bir seçimini bulmak için bu alternatifler de aranabilir. Dahası , bu alternatif yazımlar, alternatiflerin bir listesini hızlı bir şekilde bir araya getirmek için bir Trie’de paralel olarak aranabilir. Alternatif yazımları bulmak için bir Bloom filtersi kullanılmaz, çünkü bu bilgi filtreye girildikten sonra kaybolur. Elbette bir Trie , Bloom filtresinde mümkün olduğu gibi asla yanlış pozitif raporlanmayacaktır.

8.4 BÖLÜM ÖZETİ

Trie Denemeleri ve Bloom filterleri, üyeliği test etmek için kullanılan iki veri yapısıdır. Bloom filtreleri nispeten küçüktür ve zamanın belirli bir yüzdesinde yanlış pozitifler üretecektir. Trie Denemeleri daha büyüktür, yanlış pozitifler üretmez ve aranan anahtara yakın alternatif anahtar değerleri bulmak için kullanılabilir. Her iki veri yapısı çalışacak olsa da yazım denemtimi için, yazım düzeltilmesi bir Trie tarafından desteklenirken, bir Bloom filteresi yardımcı olmaz. Verimlilik açısından bakıldığında , Bloom filtereleri küme üyeliğini daha verimli bir şekilde test eder. Bazen ortaya çıkan yanlış pozitif sonuçlara maruz kalabilir . Bununla birlikte, bir Trie aynı zamanda bir Bloom filtresinden daha fazla yer kaplar. Hem Bloom filtresi hem de Trie , mikrosaniyeler içinde sözlükteki kelimelerin üyeliğini tespit eder. Bloom filtresinin en kötü süresi 160 µs iken,Trie’nin en kötü süresi her ikisinde de gerçekleştirilen gayri resmi test için 217 µs’dir .  
Boyut gereksinimleri de elbette bir endişe kaynağıdır. Bu bölümde hem Bloom filtresinin hem de Trie’nin geliştirmesinde kullanılan örnek sözlük 109.583 kelime içeriyordu. Bu sözcük sözlüğü için Bloom filtresi yaklaşık 128 KB boyutundaydı. Sonraki ve takip eden işaretçilerin her birinin 4 byte ve anahtar birimlerinin ( yani kelime karakterlerinin ) her birinin 1 byte aldığı varsayıldığında , Trie’nin boyutu kabaca 3,1 MB’tır. Bloom filtresi Trie’den çok daha küçük olsa da , her ikisi de bilgisayaların depolama kapatisesinin sınırları içindedir.