

Algoritma Analizi-3

Basit Arama Motoru

Versiyon 1.0

Prepared by Onur Demir

Grup 2-Ödev 3

onur.demir2@std.yildiz.edu.tr

18.11.2022

Video Linki: https://youtu.be/P42f0Pn5gfc

İçerik Tablosu

İç	erik	Tablosu	ii
		yon Tarihi	
1.	Υö	intem	.1
	1.1	Problem	1
	1.2	Problem Tanımı	
	1.3	Hash tablosunun problemdeki rolü	
	1.4	Struct Bilesenlerinin tanımı	
	1.5	Struct Bileşenlerinin tanımı	
2.	Uv	gulama1-	-4
	2.1	y gulama1 - Normal moddaki bazı bileşenlerin çıktıları	. 1
	2.2	Detay modda calısan bilesenlerin çıktıları	
	2.3	Detay modda çalışan bileşenlerin çıktıları	
		J 3 3	
2	Sa	nuç	4
Э.	2.1	Πυς	,4
	3.1	Load Factörün analizi (bucket dağılımı analizi)	••
	4.2	Sieves of Erothenes algoritmasının karmaşıklık hesabı	••
	4.3	Hashing kullanılması ve kullanılmaması durumuna göre algoritmanın analizi	

Revizyon Tarihi

İsim	Tarih	Değişim Sebebi	Versiyon
Onur Demir	18.11.2022	Tüm rapor yazıldı sunuldu.	1.0

Video Linki : https://youtu.be/P42f0Pn5gfc

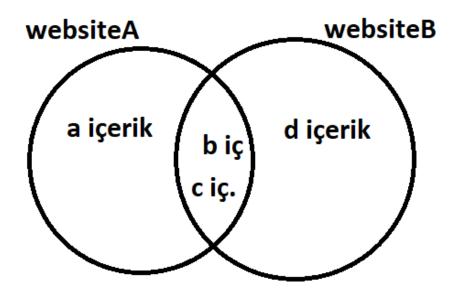
1. Yöntem

1.1 Problem

Bu ödevde sorgularıan anahtar kelimelerin hangi web sayfasında olduğunu bulan basit bir arama motoru algoritması tasarlamanız ve kodlamanız istenmektedir. Problemde hash tablosunu hangi aşamada ve nasıl kullanacağınızı sizin belirlemeniz beklenmektedir

1.2 Problem Tanımı

Problem eleman kümeleri içerisindeki keşişen ve birleşen kümeleri bulmaktır.



1.3 Hash Tablosunun Problemdeki Rolü

Bu problemde hash tablosu hızlı arama yapmak için kullanılabilir. İçeriklerin ait olduğu kümeler hash tablosunda oluşacaktır bu kümelere bucket denir. Burada web sayfa içeriği ve web sayfasının adı, anahtar değer çifti olarak kullanılabilir.Bu şekilde hızlı open addressing ve linear probing kullanarak arama yapılacaktır.

1.4 Struct Bileşenlerinin Tanımı

```
/*
64 bitlik key ile o keye ait site değeri hash tablosunda kullanılır.
*/
typedef struct KeyValuePair {
   unsigned long long int key;
   char *value;
   int orderOfInsert;
} KeyValuePair;
```

1.3 konuda belirtildiği gibi bir anahtar değer çifti bulundurmalıyız.

```
/*
Key Value değerlerini tutacak hash tablosudur.
örneğin
1 - (15,"a.com")
2 - (15,"b.com")
... şeklinde
*/
typedef struct HashTable {
    KeyValuePair **table; // bu TABLO DİZİSİ bu ilk durum kontrolü için bu şekilde aslında
    int size;
} HashTable;
```

Hash tablosunun yapısal şeması tablonun boyutu ve değer çiflerini bulunduran bir dizi tutar. String

```
/*
geleneksel string liste hali kullanımı
*/
typedef struct String {
   char a;
   struct String * nextC;
} String;
```

Okuma ve parse işlemi sırasında kullanılan String yapısı çünkü dinamik olarak input okunmalı.

```
Reading ilk okuma sırasında veriyi listeleyerek tüm terimlerin sayısını tutar.

Bundan sonra okuma bu veri uyapısı üzerinden yapılır
diğer bir faydası işe ilk başta bulunması istenen ayrık tterim sayısını elde etmek için kullanılacaktır.
toplam terim sayısı set veri yapısının kullancağı hash tablosunun boyutu olacaktır.
*/
typedef struct Term {
   char * termName;
   struct Term *next;
} Term;

typedef struct ReadStruct {
   char * websiteName ;
   Term * terms;
   struct ReadStruct * next;
} ReadStruct;

typedef struct Reading {
   ReadStruct * website;
   int totalTermCount; // bu tutulmayabilir ama set kullanılacaksa faydallı olabilir.
} Reading;
```

İlk başta website adı ve içeriğini tutacak yapılar

Reading yapısının detaylı anlatırsak tüm websiteleri ve bu websitelerin toplam sayısını tutar.

ReadStruct websitesinin içeriğini tutacak olan yapıdır.Bunun içerisinde websitesinin terimleri websitesinin adı ve sonraki websitesine olan pointer kullanılır.Burada linkli liste yapısı vardır.

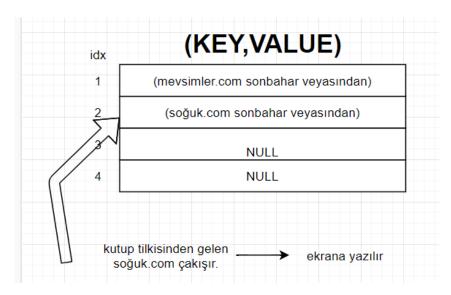
Term ise her websitesinden erişilebilen websitesinin terimlerini(içeriğini) ifade eder kendisinden sonraki eleman websitesindeki diğer içeriği verir ve içeriğin adını tutar.

1.5 Tabloya yerleşimin nasıl yapıldığı örneği

Tabloya yerleştirme yaparken Lineer Probing ile yapılır ve aşağıdaki gibi bir tablo oluşur. Değerler eklenirken aynı degree sahip olanlar kümelenecektir bu ümelenme kullanılarak problem çözülebilir.

	(KEY,VALUE)
1	(sonbahar, mevsimler.com)
2	(sonbahar,soğuk.com)
3	NULL
4	(kedi,hayvanlar.com)
5	(kedi,petshop.com)
6	(kutup tilkisi, soğuk.com)
7	NULL
8	defter,kırtasiye.com
9	kalem,kırtasiye.com

Sonbahar veya Kedi için sonbahar kümesi ve kedi kümesindekiler birleştirilir. Bu sayede "mevsimler.com, soğuk.com, hayvanlar.com, petshop.com" veya ile bu tablodan elde edilir. Kesişim için ise veya kümesinden elde edilenler value değerleri hash edilerek bulunur.Aşağıda bu örnek görülmektedir.Örneğin kutup tilkisi ve sonbahar için soğuk.com elde edilir ama tüm veya bilgisideki değerler set olarak yazılır. Çakışma olduğunda program durur ve bu değeri yazdırır.



2. Uygulama

2.1 Normal Moddaki bazı bileşenlerin Çıktıları

Normal modda ekrana çıktı yazdırılırması sonuçları aşağıdaki gibidir.

```
D:\yeni\Projeler\cpp_algo\sea X
University Computers IT AI R&D Education https://ce.yildiz.edu.tr
News https://edition.cnn.com
Entertainment SocialNetwork https://www.instagram.com
AI Competition Dataset Cloud https://www.kaggle.com
Computers Coding Tutorials https://www.tutorialspoint.com
E-Trade https://www.amazon.com
Entertainment Reviews Movies Series https://www.rottentomatoes.com
Computers IT Physics Education https://www.udemy.com
Entertainment Education News https://www.youtube.com
E-Trade https://www.hepsiburada.com
Entertainment SocialNetwork https://www.reddit.com
News https://weather.com
News SocialNetwork https://twitter.com
Entertainment Blogs Education https://medium.com
SocialNetwork Business https://www.linkedin.com
Education IT AI https://www.coursera.org
RealEstate Cars Motorcycles https://www.sahibinden.com
Cars https://www.motors.co.uk
Entertainment Reviews Movies Series https://www.imdb.com
Entertainment Movies Series https://www.netflix.com
Computers IT Competition https://leetcode.com
Normal mod mu Detay mod mu? (normal/detay)
normal
SocialNetwork
https://www.instagram.com https://www.reddit.com https://twitter.com https://www.linkedin.com
Normal mod mu Detay mod mu? (normal/detay)
normal
Normal mod mu Detay mod mu? (normal/detay)
SocialNetwork ve Entertainment
https://www.instagram.com https://www.reddit.com
Normal mod mu Detay mod mu? (normal/detay)
normal
News veya Cars
https://edition.cnm.com https://www.youtube.com https://weather.com https://twitter.com https://www.sahibinden.com https://www.motors.co.uk
Normal mod mu Detay mod mu? (normal/detay)
normal
0nur
Normal mod mu Detay mod mu? (normal/detay)
AI ve SocialNetwork
Normal mod mu Detay mod mu? (normal/detay)
```

er bir iterasyonda en büyük local maksimum değer bulunur ve global maksimum ile karşılaştırılır.

2.2 Detay Modda Çalışan Bileşenlerin Çıktıları

Detay modda ekran çıktısı sonuçları aşağıdaki gibidir.

```
D:\yeni\Projeler\cpp_algo\sea X
Normal mod mu Detay mod mu? (normal/detay)
Hash Tablosunun uzunlugu : 109
HASH TABLE
0 || NULL
  || NULL
2 | key: 1992673583 value: https://medium.com order:0
 || key: 1990183173528 value: https://www.instagram.com order:0
  || key: 208013049512 value: https://www.kaggle.com order:1
 || key: 1990183173528 value: https://www.rottentomatoes.com order:2
 || key: 1990183173528 value: https://www.youtube.com order:3
7 || key: 1990183173528 value: https://www.reddit.com order:4
8 || key: 1990183173528 value: https://medium.com order:5
9 || key: 1990183173528 value: https://www.imdb.com order:6
10 || key: 1990183173528 value: https://www.netflix.com order:7
11 || key: 71627823357 value: https://www.rottentomatoes.com order:0
12 || key: 71627823357 value: https://www.imdb.com order:1
13 || key: 71627823357 value: https://www.netflix.com order:2
14 || NULL
15 || NULL
16 || NULL
17 || NULL
18
  || NULL
19 || NULL
20
  || NULL
21 || NULL
22 || NULL
23 || NULL
24 || NULL
25 || NULL
26 || NULL
  || NULL
27
  || NULL
29 || NULL
30 || NULL
31 || NULL
32 || NULL
33 || NULL
34 || NULL
35 || NULL
36 || NULL
37 || NULL
38 || NULL
39 || NULL
40 || key: 236026655589 value: https://www.rottentomatoes.com order:0
41 | key: 236026655589 value: https://www.imdb.com order:1
42 || key: 843536928360 value: https://ce.yildiz.edu.tr order:0
43 || key: 843536928360 value: https://www.tutorialspoint.com order:1
44 || key: 843536928360 value: https://www.udemy.com order:2
45 || key: 1112933639830 value: https://www.sahibinden.com order:0
46 || key: 843536928360 value: https://leetcode.com order:4
47 || NULL
48 || key: 76662859849 value: https://www.rottentomatoes.com order:0
49 || key: 76662859849 value: https://www.imdb.com order:1
```

```
D:\yeni\Projeler\cpp_algo\sea × + ~
50 || key: 76662859849 value: https://www.netflix.com order:2
51 || NULL
52 || NULL
53 || NULL
54 || key: 72757 value: https://ce.yildiz.edu.tr order:0
55 || key: 162704697796 value: https://www.amazon.com order:0
56 | key: 72757 value: https://www.udemy.com order:2
57 || key: 162704697796 value: https://www.hepsiburada.com order:2
58 | key: 473258350624 value: https://www.linkedin.com order:4
59 || key: 72757 value: https://www.coursera.org order:5
60 | key: 72757 value: https://leetcode.com order:6
61 || NULL
62 || NULL
63 || NULL
64 || key: 1928978200127 value: https://www.instagram.com order:0
65 || key: 1928978200127 value: https://www.reddit.com order:1
66 || key: 1928978200127 value: https://twitter.com order:2
67 || key: 75161453 value: https://edition.cnn.com order:0
68 || key: 826203662712 value: https://ce.yildiz.edu.tr order:0
69 || key: 826203662712 value: https://www.udemy.com order:1
70 || key: 826203662712 value: https://www.youtube.com order:2
71 | key: 75161453 value: https://www.youtube.com order:4
72 || key: 75161453 value: https://weather.com order:5
73 || key: 75161453 value: https://twitter.com order:6
74 | key: 826203662712 value: https://medium.com order:6
75 || key: 1928978200127 value: https://www.linkedin.com order:11
76 || key: 826203662712 value: https://www.coursera.org order:8
77 || NULL
78 || NULL
79 || NULL
80 || NULL
81 || NULL
82 || NULL
83 || NULL
84 || NULL
85 || key: 1293351909953 value: https://www.kaggle.com order:0
86 | key: 1293351909953 value: https://leetcode.com order:1
87 || key: 235298768569 value: https://www.udemy.com order:0
88 || key: 939351717355 value: https://www.tutorialspoint.com order:0
89 || NULL
90 || NULL
91 || key: 64728 value: https://ce.yildiz.edu.tr order:0
92 || key: 1223012781842 value: https://ce.yildiz.edu.tr order:0
93 || key: 64728 value: https://www.kaggle.com order:2
94 || key: 64728 value: https://www.coursera.org order:3
95 || NULL
96 || NULL
97 || NULL
98 || NULL
99 || NULL
100 || key: 64878753 value: https://www.sahibinden.com order:0
101 || key: 62736171446 value: https://www.tutorialspoint.com order:0
102 || key: 1328855819102 value: https://www.sahibinden.com order:1
103 || key: 2481488 value: https://ce.yildiz.edu.tr order:0
```

```
D:\yeni\Projeler\cpp_algo\sea X
99 || NULL
100 || key: 64878753 value: https://www.sahibinden.com order:0
101 | key: 62736171446 value: https://www.tutorialspoint.com order:0
102 | key: 1328855819102 value: https://www.sahibinden.com order:1
103 | key: 2481488 value: https://ce.yildiz.edu.tr order:0
104 | key: 2021315723 value: https://www.kaggle.com order:1
105 | key: 64878753 value: https://www.motors.co.uk order:5
106 || NULL
107 || NULL
108 || NULL
a ve b
Bu iki kelimenin birlikte gectigi dokuman bulunmamaktadir.
Normal mod mu Detay mod mu? (normal/detay)
normal
News ve Cars
Normal mod mu Detay mod mu? (normal/detay)
normal
Computers ve News
Normal mod mu Detay mod mu? (normal/detay)
normal
Computers ve IT
https://ce.yildiz.edu.tr https://www.udemy.com https://leetcode.com
Normal mod mu Detay mod mu? (normal/detay)
normalS
```

2.3 Detay modda çalıştırılan örneklerin incelenmesi

Yukarıdaki örnekte gördüğümüz gibi öbeklenmeler bulunakta ve bu öbeklenmeler birbirine yakın olan değerler null değerleri arasında olacak şekilde dağılmıştır.Örneğin 64 66 arasındaki değerler ve yine bu key değerine ait olan diğer değer ise 75 dedir.

3. Uygulama

3.1 Load Factörün Analizi (bucket dağılımı analizi)

Load factorü 0.5 için detay çıktısı yukarıda verilmiştir bu bölümde 0.1 ve 0.9 için hash tablosu incelenecektir.0.9 için

```
D:\yeni\Projeler\cpp_algo\sea X
Normal mod mu Detay mod mu? (normal/detay)
normal
Entertainment ve Reviews
https://www.rottentomatoes.com https://www.imdb.com
Normal mod mu Detay mod mu? (normal/detay)
Hash Tablosunun uzunlugu : 61
HASH TABLE
0 || key: 208013049512 value: https://www.kaggle.com order:0
  || key: 75161453 value: https://www.youtube.com order:3
  || key: 75161453 value: https://weather.com order:4
  || key: 75161453 value: https://twitter.com order:5
  || key: 1992673583 value: https://medium.com order:1
  || NULL
  || key: 1293351909953 value: https://www.kaggle.com order:0
  || key: 64728 value: https://ce.yildiz.edu.tr order:0
  || key: 2481488 value: https://ce.yildiz.edu.tr order:0
 || key: 64728 value: https://www.kaggle.com order:2
10 || key: 64728 value: https://www.coursera.org order:3
11 || key: 64878753 value: https://www.sahibinden.com order:4
12
  || key: 64878753 value: https://www.motors.co.uk order:5
   || key: 1293351909953 value: https://leetcode.com order:7
   || key: 939351717355 value: https://www.tutorialspoint.com order:0
15
  || key: 76662859849 value: https://www.rottentomatoes.com order:0
  | key: 1223012781842 value: https://ce.yildiz.edu.tr order:0
   || key: 76662859849 value: https://www.imdb.com order:2
  || key: 76662859849 value: https://www.netflix.com order:3
19 || NULL
20 || key: 2021315723 value: https://www.kaggle.com order:0
21 || NULL
22 || NULL
23
     NULL
24
     key: 1990183173528 value: https://www.instagram.com order:0
25
     key: 843536928360 value: https://ce.yildiz.edu.tr order:0
  || key: 843536928360 value: https://www.tutorialspoint.com order:1
  || key: 1990183173528 value: https://www.rottentomatoes.com order:3
  key: 71627823357 value: https://www.rottentomatoes.com order:3
  || key: 843536928360 value: https://www.udemy.com order:4
  || key: 1990183173528 value: https://www.youtube.com order:6
31 || key: 1990183173528 value: https://www.reddit.com order:7
32
  || key: 1990183173528 value: https://medium.com order:8
33
  || key: 826203662712 value: https://ce.yildiz.edu.tr order:0
     key: 826203662712 value: https://www.udemy.com order:1
     key: 826203662712 value: https://www.youtube.com order:2
     key: 826203662712 value: https://medium.com order:3
36
37
   || key: 826203662712 value: https://www.coursera.org order:4
  || key: 1328855819102 value: https://www.sahibinden.com order:0
  || key: 1990183173528 value: https://www.imdb.com order:15
40 || key: 71627823357 value: https://www.imdb.com order:15
41 || key: 1990183173528 value: https://www.netflix.com order:17
     key: 71627823357 value: https://www.netflix.com order:17
43 || key: 1928978200127 value: https://www.instagram.com order:0
```

Burada 24.değerden bir keye ait bir değer 4.değerde de bulunmakta key değeri neredeyse tüm tabloyu taramıştır.

```
D:\yeni\Projeler\cpp_algo\sea X
15 || key: 76662859849 value: https://www.rottentomatoes.com order:0
16 || key: 1223012781842 value: https://ce.yildiz.edu.tr order:0
17 || key: 76662859849 value: https://www.imdb.com order:2
18 || key: 76662859849 value: https://www.netflix.com order:3
20 || key: 2021315723 value: https://www.kaggle.com order:0
21 || NULL
22 || NULL
23 ||
     NULL
24 ||
     key: 1990183173528 value: https://www.instagram.com order:0
     key: 843536928360 value: https://ce.yildiz.edu.tr order:0
     key: 843536928360 value: https://www.tutorialspoint.com order:1
27
   || key: 1990183173528 value: https://www.rottentomatoes.com order:3
  || key: 71627823357 value: https://www.rottentomatoes.com order:3
29 || key: 843536928360 value: https://www.udemy.com order:4
30 || key: 1990183173528 value: https://www.youtube.com order:6
31 || key: 1990183173528 value: https://www.reddit.com order:7
32 || key: 1990183173528 value: https://medium.com order:8
33 || key: 826203662712 value: https://ce.yildiz.edu.tr order:0
34 | key: 826203662712 value: https://www.udemy.com order:1
35 || key: 826203662712 value: https://www.youtube.com order:2
36 || key: 826203662712 value: https://medium.com order:3
37 || key: 826203662712 value: https://www.coursera.org order:4
38 || key: 1328855819102 value: https://www.sahibinden.com order:0
     key: 1990183173528 value: https://www.imdb.com order:15
     key: 71627823357 value: https://www.imdb.com order:15
     key: 1990183173528 value: https://www.netflix.com order:17
42 | key: 71627823357 value: https://www.netflix.com order:17
43 || key: 1928978200127 value: https://www.instagram.com order:0
44 || key: 1928978200127 value: https://www.reddit.com order:1
45 || key: 72757 value: https://ce.yildiz.edu.tr order:0
46 || key: 72757 value: https://www.udemy.com order:1
47 || key: 1928978200127 value: https://twitter.com order:4
48 || key: 235298768569 value: https://www.udemy.com order:0
49 | key: 162704697796 value: https://www.amazon.com order:0
50 | key: 162704697796 value: https://www.hepsiburada.com order:1
51 || key: 1928978200127 value: https://www.linkedin.com order:8
52 | key: 473258350624 value: https://www.linkedin.com order:4
53 || key: 72757 value: https://www.coursera.org order:8
     key: 236026655589 value: https://www.rottentomatoes.com order:0
     key: 1112933639830 value: https://www.sahibinden.com order:12
     key: 236026655589 value: https://www.imdb.com order:2
     key: 843536928360 value: https://leetcode.com order:32
   || key: 72757 value: https://leetcode.com order:13
  || key: 75161453 value: https://edition.cnn.com order:0
60 || key: 62736171446 value: https://www.tutorialspoint.com order:0
```

0.1 Load faktör için ise 557 boyutlu bir gash tablosu oluşmaktadır ve bucket dizilimi çoğunlukla ardışık olarak ilerlemektedir.

```
D:\yeni\Projeler\cpp_algo\sea
    II NULL
406
    Ш
      NULL
    || NULL
408
    || NULL
      key: 1328855819102 value: https://www.sahibinden.com order:0
   | NULL
411
    || NULL
    || NULL
413
   | NULL
414
    || NULL
415
    II NULL
416
    Ш
      NULL
418
    || NULL
    || NULL
    || NULL
420
   | NULL
421
422
    || NULL
   || NULL
423
424
   | NULL
   | key: 843536928360 value: https://ce.yildiz.edu.tr order:0
425
   || key: 843536928360 value: https://www.tutorialspoint.com order:1
      key: 843536928360 value: https://www.udemy.com order:2
428
    | key: 843536928360 value: https://leetcode.com order:3
    | NULL
429
      key: 75161453 value: https://edition.cnn.com order:0
    || key: 75161453 value: https://www.youtube.com order:1
    | key: 75161453 value: https://weather.com order:2
    || key: 75161453 value: https://twitter.com order:3
    II NULL
434
435
      NULL
```

3.2 Sieves of Eratosthenes algoritmasının karmaşıklık hesabı

Bu algoritmada en yakın asalı bulmak için önce bir elek oluşturur. Bu elek 2 3 5 7 diye bu katları elimine eder. Bu elek oluşturulurken toplan elenen eleman sayısını 2 den elenenleri n/2 3 den elenenleri n/3 alırsak;

$$\frac{n}{2} + \frac{n}{3} + \frac{n}{5} + \frac{n}{7} + \dots p$$

$$\frac{n * (1/2 + 1/3 + 1/5 + 1/7 \dots \infty)}{n * (1/2 + 1/3 + 1/5 + 1/7 \dots \infty)}$$

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \dots = log(log(n))$$

Harmonic seri açılımından bulunur.(İspata Wolfram alphadan bakılabilir.)

$$O(n * (log(log(n))))$$
Karmaşıklığı bulunur.

Her seferinde eleğin oluşturulması gerektiği için en iyi ve en kötü durum $\theta(N*Log(Log(N)))$ durumuna yakınsar.

Space Complexity O(n) dir

3.3 Hashing kullanılması ve kullanılmaması durumuna göre algoritmanın analizi

Hashing kullanıldığında en kötü durum için eğer load faktörü düşük ise tüm tabloyu tarayacaktır. Bu durumda O(N) complexity de tablo taranır.

En iyi durumda load factor aşırı küçükse ulaşılır ama bu sefer space complexity artacağından algoritma yavaş çalışacaktır bunun için load factoru ortalama bir değer alınıp hash fonksiyonu çakışmayacak şekilde yapılandırılmalıdır.Örneğin bazı durumlarda load factor düşük olmasına rağmen bir keye ait olmasını beklediğimiz satırlar başka key ile collision olmaktaydı. Eğer tam doğru collision engellenmiş bir hash ve iyi bir load değeri ile algoritma O(K) ile tek seferde tüm verileri çekecektir. Burada K veya ile çekilecek eleman sayısıdır.

Ve durumunda ise En kötü durumda ilk tabloda N ile arnır ikinci tabloda ise N ile aranır O(N+N) den O(N) en kötü durum karmaşıklığı olur.En iyi durumda ise ilk tabloda O(K) ikinci tabloda 2->O(1) ile arama yapar Bu durumda en iyi durumda O(K) ile bu işlemi gerçekleştirecektir.

Hashing kullanılmasaydı veya işlemi için tüm veri üzerinden geçilecekti.Bu durumda karmaşıklık O(N) olacaktı. N burada toplam terim sayısı olacaktır. Ve için ise tüm tablo taranacaktır ve tarananlar arasındaki kesişenler bulunacaktır.Bu durumda karmaşıklık O(N+K^2) olacaktır.

N: Toplam terim sayısı

K: küme üzerindeki veri sayısı