

# Programlama Laboratuvarı 1. Dönem 3. Proje Ödevi

Mustafa Efe Tamer  
Bilgisayar Mühendisliği 2.Öğretim  
Kocaeli Üniversitesi  
Kocaeli, Türkiye  
Okul No : 220202084  
[efetamer8@gmail.com](mailto:efetamer8@gmail.com)

İnan Yiğit Oğuz  
Bilgisayar Mühendisliği 2. Öğretim  
Kocaeli Üniversitesi  
Kocaeli, Türkiye  
Okul No : 220202014  
[yigitoguz14@outlook.com](mailto:yigitoguz14@outlook.com)

## 1 – Özet

Bu projede 50.000 kullanıcı var ve bunların her birinin kullanıcı adı, adı, takip ettiklerinin sayısı ve takip ettikleri, takipçilerinin sayısı ve takipçileri, dili, bölgesi ve tweetleri bulunmakta. Bütün bu bilgiler bir Json dosyasında tutulmaktadır ve biz de işlemlerimizi bu Json dosyası üzerinden yapmaktayız. Örnek vermek gerekirse kullanıcı ve bilgilerini hash tablolarını atmak, kullanıcıları düğüm haline getirip kenarlar ile her birini takip ettikleri ve takipçilerine bağlamak, kullanıcıların ilgi alanlarını bulup hash tablosuna atıp ortak ilgi alanlarına sahip kullanıcıları listelemek vb.

## 2 – Giriş

İlk olarak yüksek hacimli bir Json dosyasını terminal yardımı ile indirdik ardından IntelliJ Idea idemizden belli bir metot yardımı ile Json dosyamızı okuduk. Daha sonra içindeki her bir kullanıcının bilgilerini farklı bir sınıfta hash Tablolarına attık. Ek olarak hash tablolarına attığımız takipçi, takip edilen bilgilerimiz ışığında her bir kullanıcıyı, takipçiyi ve takip edileni teker teker düğümlere (nodes) attık sonrasında kenarlar (edges) yardımı ile bu düğümleri birbirine bağladık. İlgi alanları

için Türkçe doğal dil işleme araştırdık ve zembereği kurduk ‘jar’ dosyalarını indirerek akabinde ise bu dosyaları idemizden bağımlılık olarak ekledik fakat zemberek yetersiz kaldı çünkü bizim için cümleleri anlamlandırıp ilgi alanını belirleyebilmesi gerekiyordu ama zemberek sadece tweetler’deki kelimeleri seçip eklerine köklerine göre ayırıyor ve kelimelerin türlerini belirliyordu. Mesela : “ben bugün maç yaptım” cümlesinden ilgi alanını ‘futbol’ olarak belirlemesi gerekirken o ‘ben’ ‘bugün’ ‘maç’ gibi ayırıp tür ve ek, köklerine ayırdı. Biz de bunun üstüne NLP kullanmak yerine mecburi olarak kelime frekanslarına göre gittik. Yani bir kullanıcının attığı tweetlerdeki en fazla geçen iki kelimeyi ve bunların frekanslarını buluyor böylece o iki kelime bizim ilgi alanlarımız oluyor fakat burada şöyle bir sıkıntı çıkıyor gereksiz kelimeleri ilgi alanı olarak belirleyebiliyor mesela bağlaçları edatları vb. Biz de bunun sonucunda gereksiz kelimeleri çıkarmak için edatlar, bağlaçlar, ekler gibi bir hazır liste aldık ve frekansı en fazla olan iki kelime bunlardan biri ise onu atladık ve diğer en fazla geçene baktık. Programımız daha sonrasında ise bizden istediğimiz kullanıcılarımızı girmemizi istiyor ve bu kullanıcının ilgi alanlarını ve frekanslarını buluyor ardından ilgi alanları ile örtüşen ilgi alanlarına sahip kullanıcıları listeliyor ve ortak ilgi alanının ne olduğunu gösteriyor.

### 3 – Yöntem

Json dosyasını ‘objectMapper.readTree’ ile okuyup JsonNode türünde bir değişkene attık. Sonrasında her bir kullanıcının bilgilerini ‘Kullanıcı’ sınıfına gönderdik ve oradan da HashTablosu sınıfına gönderip ‘put’ methodu ile hash tablolarına ekledik. Graf sınıfına da yine JsonNode türündeki değişkenimizi atıp her kullanıcıyı düğüm olarak tuttuk ve ardından her bir düğüme (kullanıcıya) takip ettiklerini ve takipçilerini kenarlar ile bağladık (hash tablosundaki get methodunun yardımı ile), altına da toplam düğüm bağlama sayısını yazdırdık. En fazla geçen kelime1 ve 2 olarak iki değişken oluşturup başlangıç olarak ‘null’ a eşitledik ve frekanslarını da ‘0’ a eşitledik. Sonrasında tweetleri ayırdık ve boşluklara göre parçaladık ardından sayaç yardımı ile kelime frekansını bulduk tabikide gereksiz kelime listesinde olup olmadığına göre ayıklayarak. Sonrasında bunları hash tablolarında tuttuk. Ek olarak program hangi kullanıcıyı girmek istediğimizi bizden istiyor ve girilen kullanıcı numarasına göre kullanıcıyı bulup ilgi alanlarının ne olduğunu hash tablosundan çekiyor ve iki ilgi alanını getiriyor ardından da bu ortak ilgi alanlarına sahip kullanıcıyı buluyor ve adıyla birlikte ortak olan ilgi alanını gösteriyor.

### 4 – Deneysel Sonuçlar

```
"murphy.augustine"-->"hickle.braden"
"murphy.augustine"-->"turner.pfeffer"
"murphy.augustine"-->"mohamed80"
"murphy.augustine"-->"snitzsche"
"murphy.augustine"-->"cstroman"
"murphy.augustine"-->"osinski.marisa"
"murphy.augustine"-->"shyanne19"
"antonina.simonis"-->"ahammes"
"antonina.simonis"-->"tressa77"
"antonina.simonis"-->"miller.vicenta"
"antonina.simonis"-->"baumbach.taylor"
"antonina.simonis"-->"rowland11"
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "kutch.trevor"
Benzer ilgi alanları : 'müzik'
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "schaefter.therese"
Benzer ilgi alanları : 'müzik'
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "schuyler.schmeler"
Benzer ilgi alanları : 'tarihinde'
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "easton.padberg"
Benzer ilgi alanları : 'müzik'
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "zhirthe"
Benzer ilgi alanları : 'müzik'
```

```
Bu ikisinin ilgi alanı benzerdir : "mccullough.gabe" - "kauer"
Benzer ilgi alanları : 'müzik'
```

```
Kullanıcı adı : "satterfield.hosea"
Birinci ilgi alanı : 'olur.', Frekans: 3
İkinci ilgi alanı : 'tarihi', Frekans: 3
```

```
Lutfen aramak istediğiniz kullanıcı sırasını giriniz - (çıkışı için ise '-1' giriniz) : 67
```

```
1. ilgi alanı : eğitim
```

```
2. ilgi alanı : Galatasaray
```

```
Bu kullanıcı ile benzer ilgi alanını sahip kullanıcı : "amina.batz"
```

```
"medhurst.jennifer"-->"amina.batz"
```

```
Benzer olan ilgi alanı : eğitim
```

```
Bu kullanıcı ile benzer ilgi alanını sahip kullanıcı : "cecilia.casper"
```

```
"medhurst.jennifer"-->"cecilia.casper"
```

```
Benzer olan ilgi alanı : eğitim
```

```
Bu kullanıcı ile benzer ilgi alanını sahip kullanıcı : "heaney.frank"
```

```
"medhurst.jennifer"-->"heaney.frank"
```

```
Benzer olan ilgi alanı : eğitim
```

### 5 – Katkıları

Genel olarak kodlarımızı hep ortaklaşa yaptık bir methodun bir kısmını birimiz aynı methodun eksik kalan yerlerini veya güncellemeleri diğerimiz yaptık. Başlı başına birimizin yaptığı bir method yoktur.

## 6 – Sonuç

Bu proje sayesinde Json dosyasının ne olduğunu ve Json dosyası üzerinde işlemler yapmayı öğrendik. Json dosyasındaki yazım şeklini ve Json dosyasının içindeki verileri nasıl hash tablolarına atacağımızı öğrendik. Bu arada hash tabloları ile de ilk defa burada tanıştık ve onların neden, nasıl kullanıldığını ve bize nasıl fayda sağladığını görmüş olduk. Tweeter, Facebook vb. gibi sosyal mecraların arkasında dönen işlemlere en temelden bile olmuş olsa tanık olma fırsatımız oldu ve bilinç kazanmış olduk. Çevrimsel, yönlü, yönsüz grafları araştırdık ve işimize yarayanını kullandık ve böylece graflar ile çalışmayı öğrendik. Düğümler oluşturmayı onları birbirlerine bağlamayı öğrendik ve bunları en ilkel şekilde hiçbir kütüphanenin hiçbir hazır metodunu kullanmadan yaptık. Kısacası bu proje bize hem yeni bir şeyler öğretti hem de bildiklerimiz üzerinde pratikler yapmayı ve farklı amaçlar için bildiklerimizi şekillendirmeyi öğretti. Artık sosyal medyaya girdiğimiz zaman önümüze ilgi alanlarına göre ‘kullanıcıların’, ‘postların’ çıkmasındaki arka planı biraz da olsa anlayabilmek güzel oldu.

## 7 – Kaynakça

<http://www.flowgorithm.org/download/>

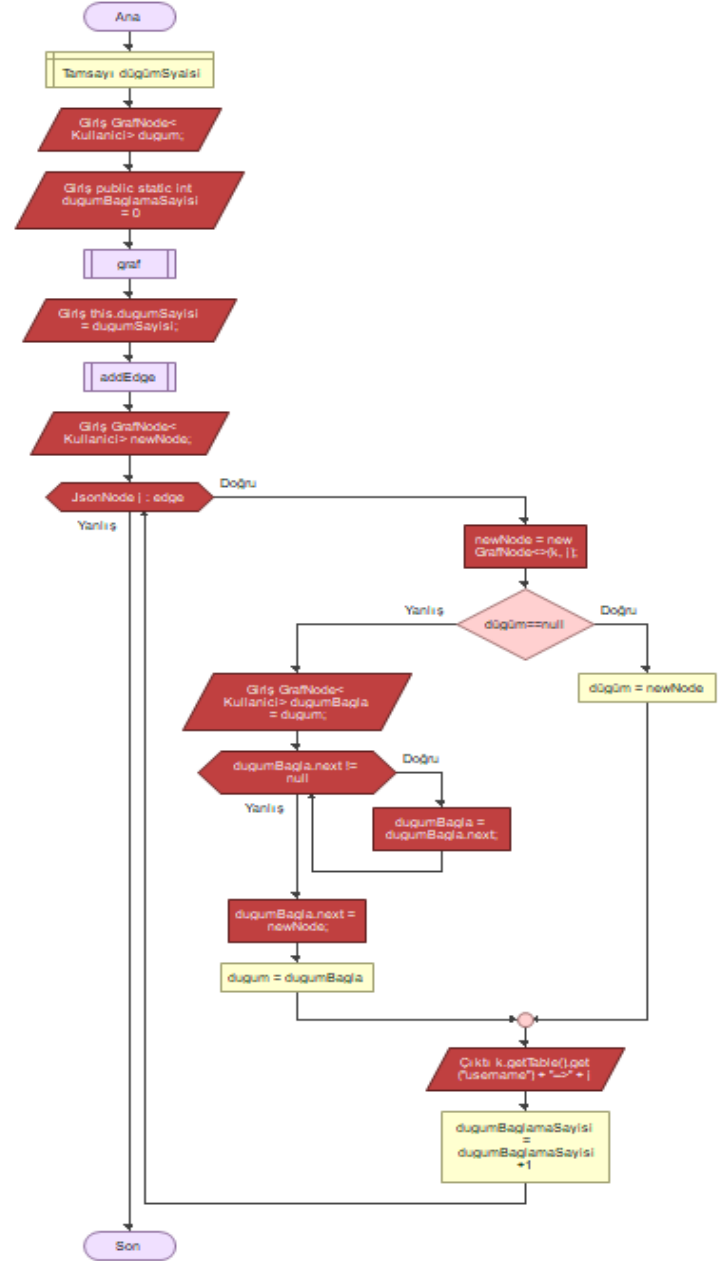
<https://sourceforge.net/projects/jung/>

<https://www.turkceogretimi.com/tavsiyeler/en-cok-kullanilan-1000-turkce-kelime?start=1>

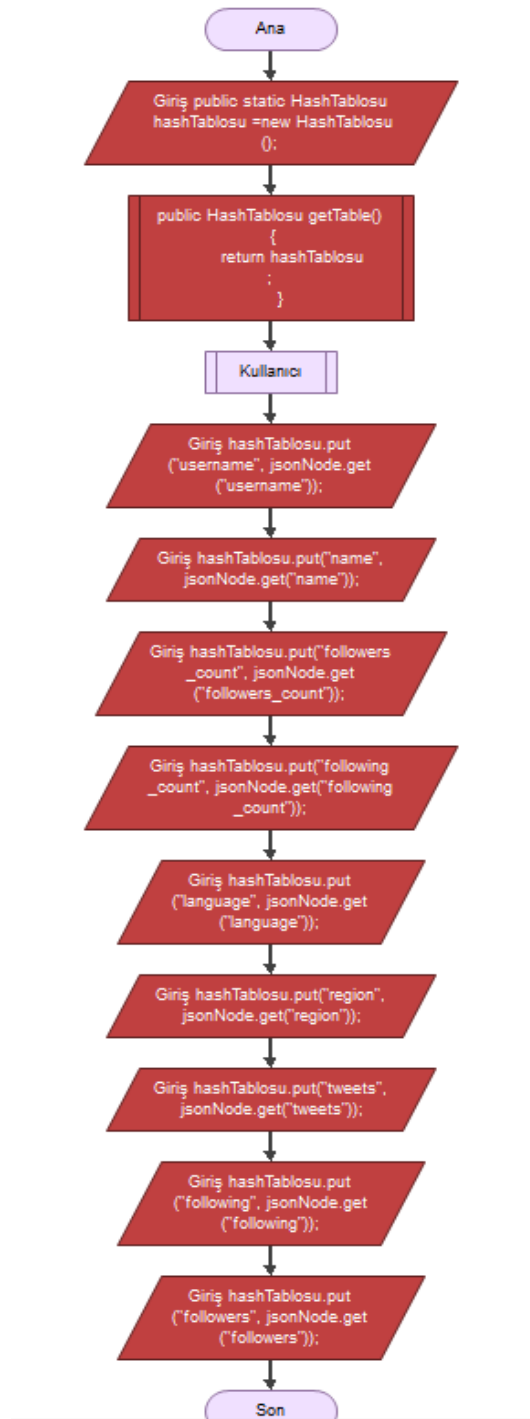
<https://github.com/ahmetaa/zembere-k-nlp>

<https://sourceforge.net/projects/json-lib/>

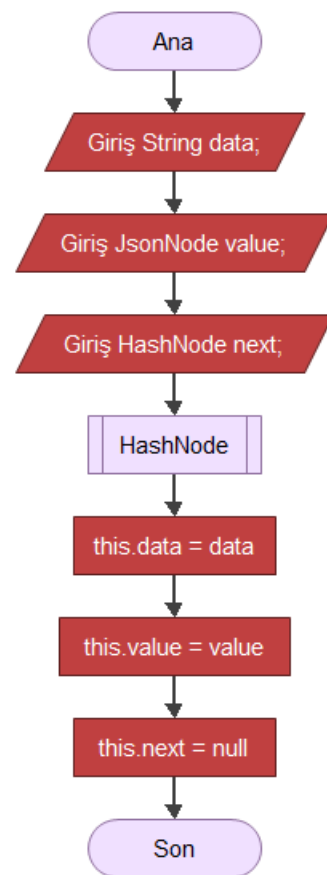
### Graf Akış Diyagramı :



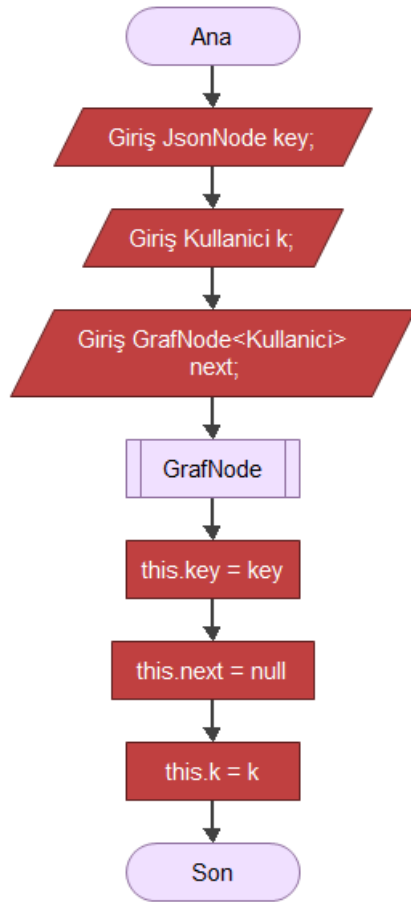
## Kullanıcı Akış Diyagramı :



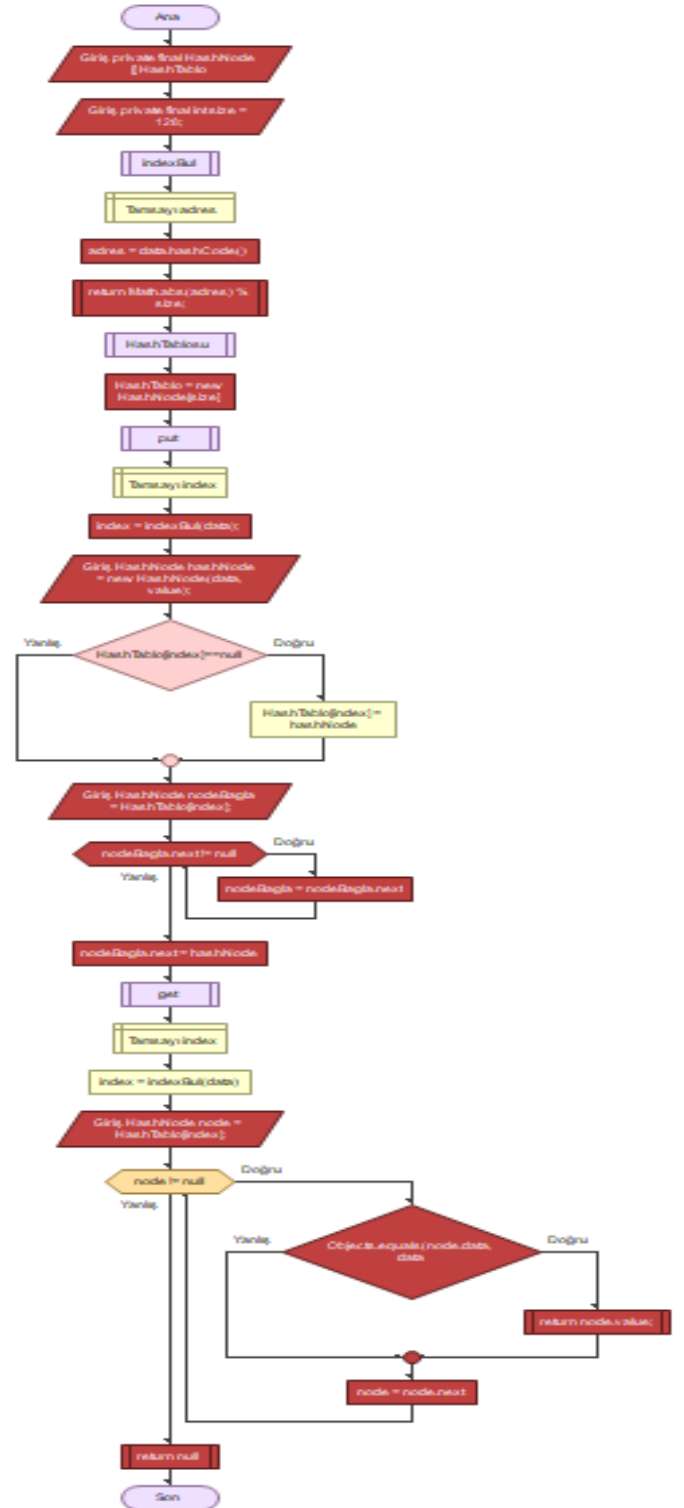
## HashNode Akış Diyagramı :



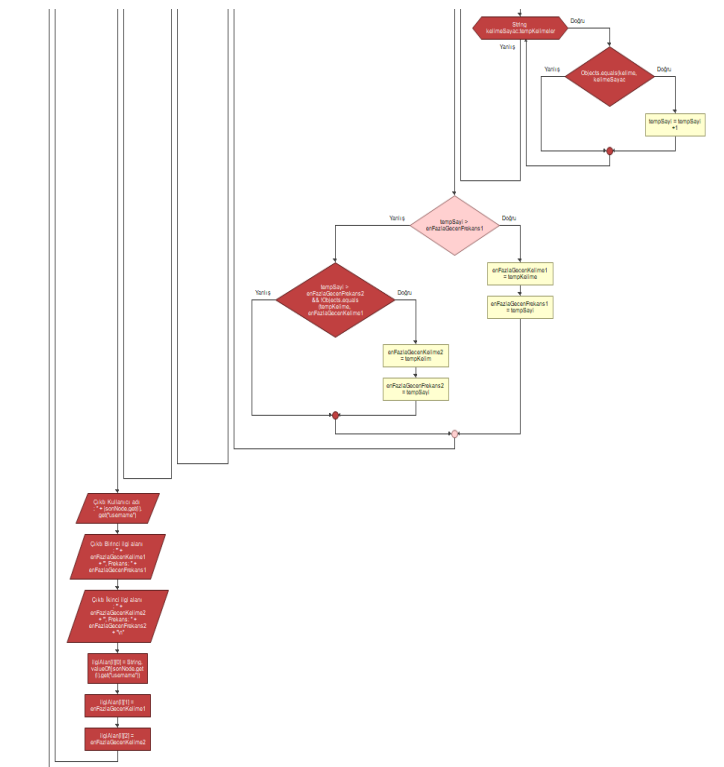
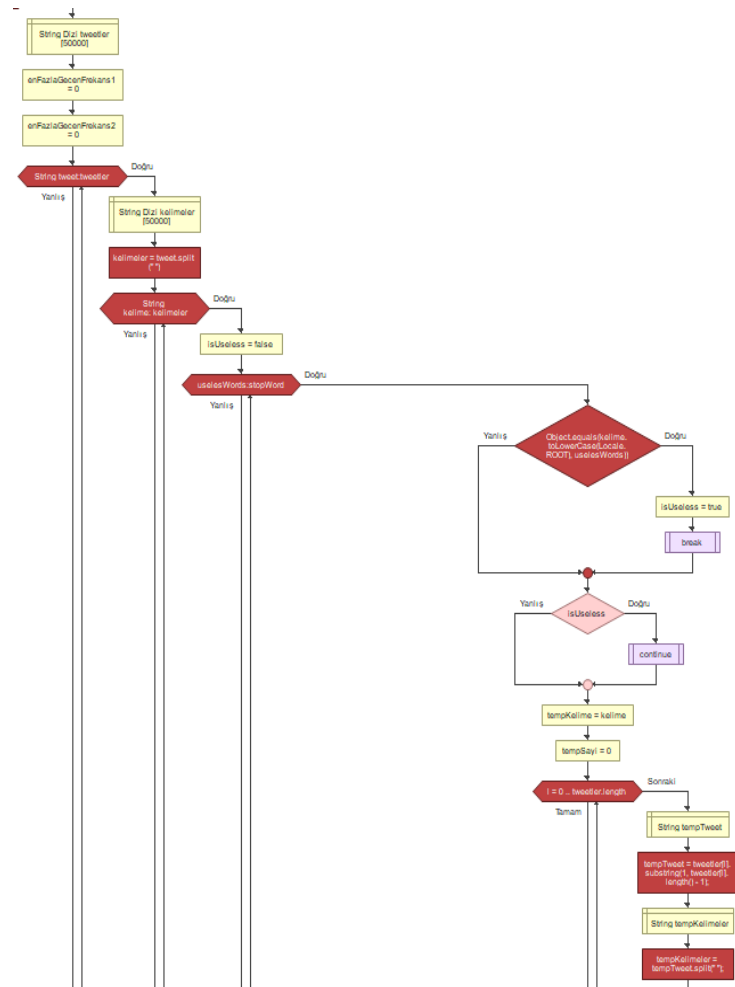
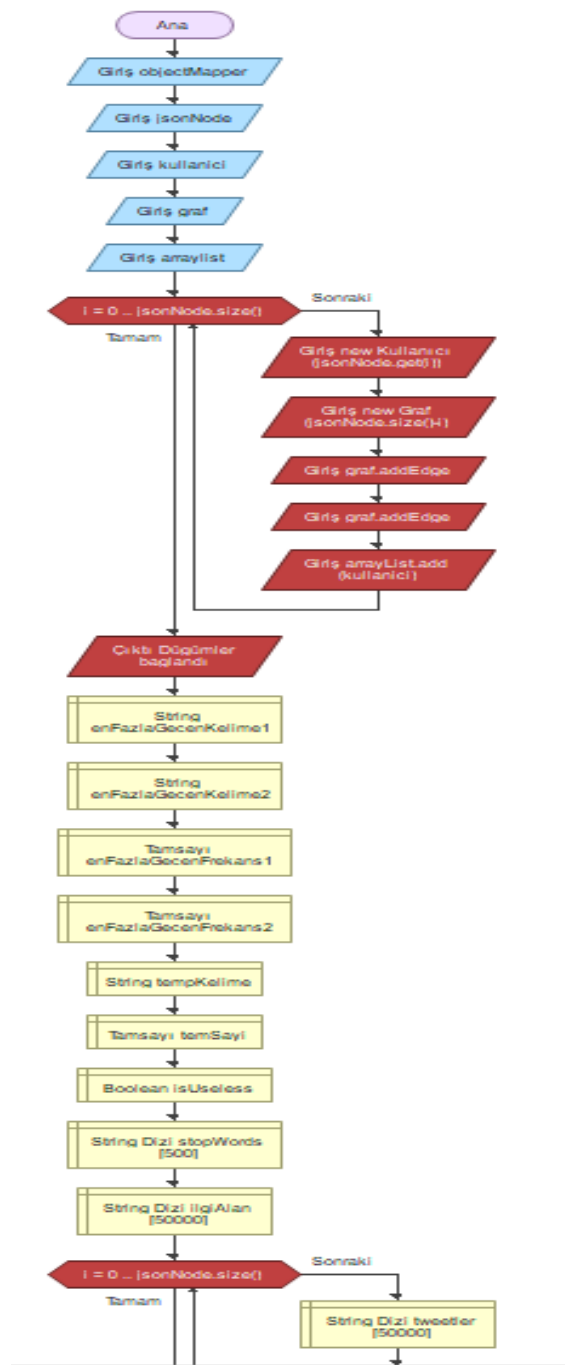
GrafNode Akış Diyagramı :



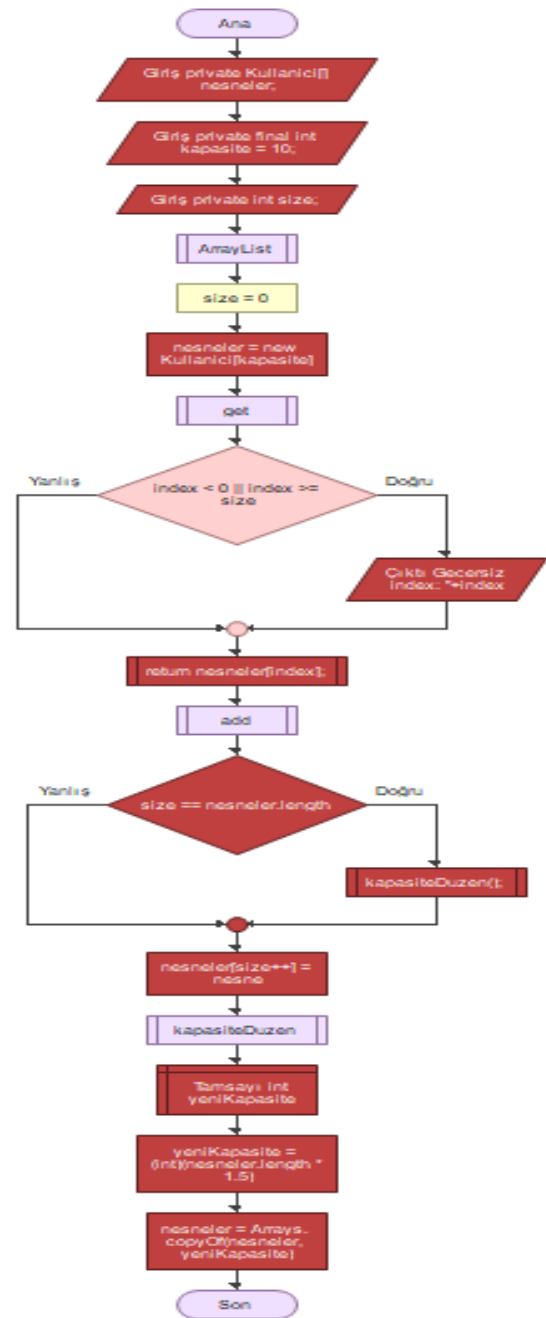
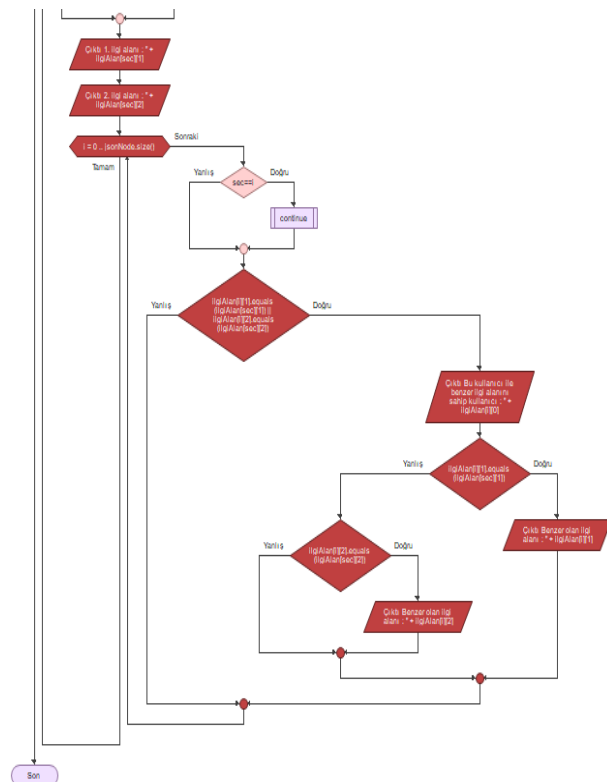
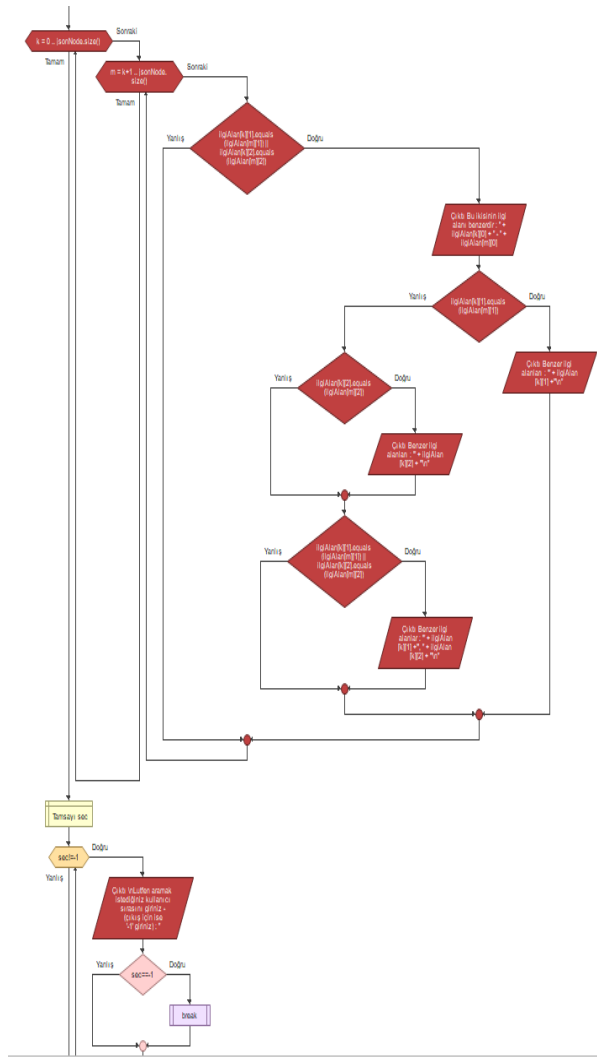
HashTablosu Akış Diyagramı :



Main Akış Diyagramı :



### ArrayList Akış Diyagramı :



UML Diyagramı :

