



## **BLM 4580 Doğal Dil İşleme Giriş Dersi**

### **1.Ödev**

**Konu :** Adresler için Regex hazırlanması

**Hazırlayan:** 16011002 Aykut AKDENİZ

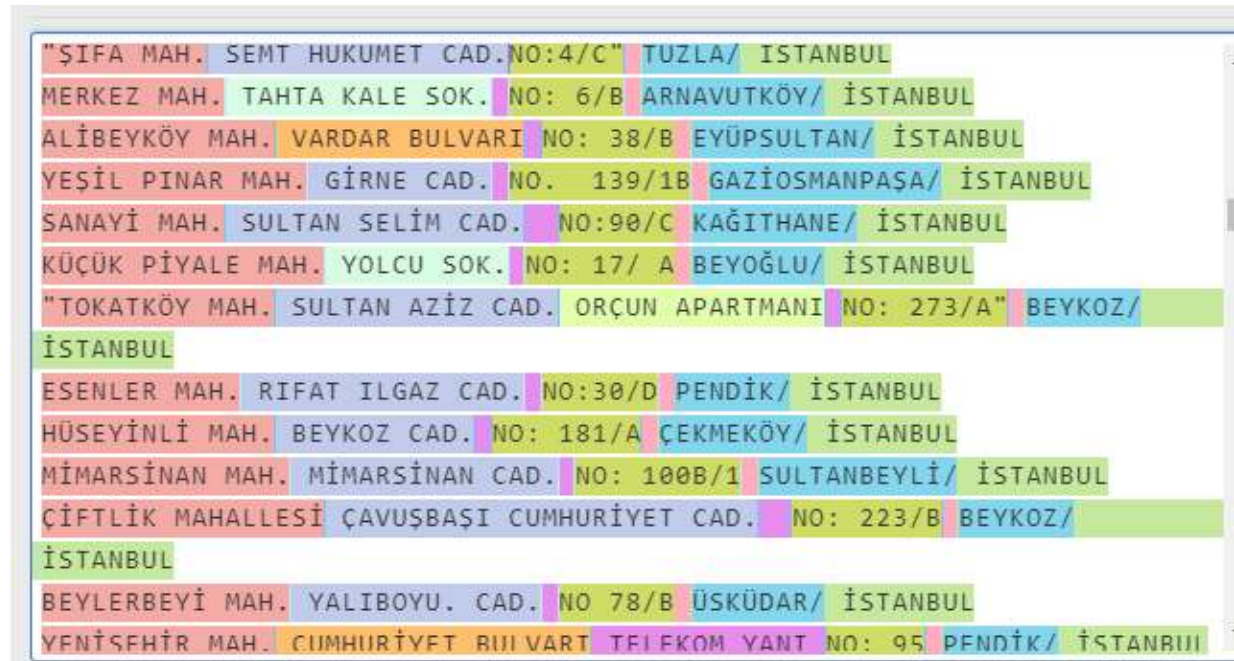
Adreslerin formatında bazı örneklerde değişse de genel olarak sırayla mahalle-bulvar-cadde-sokak-apartman-no-ilçe-il olarak gittiği için buna göre regex hazırlanmıştır.(<https://regex101.com/> adresinde PCRE(PHP) ile yapılmıştır.)(İl olarak [A-ZĞÜŞİÖÇ]{3-14} çalışma süresini çok fazla arttırdığı için il yerine İstanbul yazılmıştır.)

```
^(?<Adres>(?(?<Mahalle>.[^V.^n]*M\.*MA?H\s?[\.|\s]).*MAHALLES?[İİ]?)(?(?<Bulvar>.*BULVARI|. *B\.*U?L\.*V\.*BUL\.)|(?(?<Cadde>.*C\.*[s\.]CA?D\s?[s\.].*CADDES?[İİ]?s)|(?(?<Sokak>.*S\.*[s\.]SO?K\s?[s\.].*SOKA[KĞG]I?)(?(?<Apartman>.*sAPT\.*APARTMANI?)(?(?<Bilinmeyen>.*)(?(?<No>[s\.]?NO\s?[s\.]*s*[0-9A-Z]+[V-]?[A-Z0-9]*[sV-\'"]?[A-Z0-9]*[s\.'"]?)(?(?<Tanim_cumlesi>.*s)(?(?<Ilce>\p{L}+V)(?(?<Il>sİSTANBUL))+$
```

Regex adres dosyasındaki tüm adresleri grup içerisine alabiliyor ancak istasyon, bayii, camii, durak, büfe önü, bayii vb. örneklerin grup olarak aynı grup içerisine alıp tam olarak ismini tanımlayamıyor(Tanım cümlesi başlığında toplanmışlardır). Aşağıda verilen başlıkta bulunan durumlarda yanlış gruplandırırsa da adres bilgilerinde bulunan 6831 adresi de kapsamaktadır.

**Regex ile başarılı bulunan adres sayısı: 6788**

**Regex'in başarı oranı: %99.37**



### Yanlış Gruplandırılan Örnekler

Genel olarak durağı, camii, iskele gibi kesin olmayan tarif içeren adreslerde belirli bir gruplama yapılamamıştır ve bu kısımlar Tanım\_cumlesi grubuna alınmıştır. Sitesi, Apartmanı, iş merkezi gibi bilgiler için de ifade uzunluğu artacağı için "Bilinmeyen" olarak

gruplandırılmıştır. Başına No eklenmemiş rastgele numara bilgileri de belirsizlikten dolayı eklenememiştir.

ÖRNEK MAH. DOĞ. ARS. BLV FİKRİ SÖN CAD. GİRİŞİ AGENA E NO. 215 9/2  
ESENİYURT/ İSTANBUL

Yukarıdaki örnekte “/2” kısmını no grubuna dahil edememiştir.

KUYUMCU KENT ATÖLYE DURAĞI, 29 EKİM CAD. LADİN SOK. K:1, SK:9, NO:6  
BAHÇELİEVLER/ İSTANBUL

Yukarıdaki örnekte durak bilgisini tanımadığı için cadde bilgisine eklemiştir.

SOĞUKSU MH. M.AKİF ERSOY CD. NO:141/C BEYKOZ/ İSTANBUL

Yukarıdaki örnekte m. Kısmından dolayı mahalle bilgisini ön kısmından ayıramamıştır.

MİMAR SİNAN MAH. HÂKİMİYETİ MİLLİYE CAD. MİHRİMAH SULTAN CAMİİ ÖNÜ  
ÜSKÜDAR/ İSTANBUL

Yukarıdaki örnekte “Mihrimah” kısmındaki mah kısmından dolayı mahalle grubuna dahil edilmiştir.

MEHMET AKİF ERSOY CAD. BOSNA BULVARI NO. 71/2 ÜSKÜDAR/ İSTANBUL

Yukarıdaki örnekte ise diğer adres bilgilerinde bulvar caddeden önce gelmesine rağmen burada sırası değiştiği için bulvar grubunu caddeyi de içerecek şekilde bulmuştur.

SOKULLU CAD. GÜZELTEPE. MAH EYÜPSULTAN/ İSTANBUL

Yukarıdaki örnekte mahalle, caddeden sonra geldiği için mahalle kısmını olduğundan uzun bulmuştur.

SAHİP MOLLA CAD. NO:37 PAŞABAHÇE BEYKOZ/ İSTANBUL

Uzun ve karmaşık yapıda “No” bilgilerinin genelini kapsayacak bir ifade yazıldığı için bu örnekteki gibi yanlış harfleri de dahil edebilmektedir.