



Doğal Dil İşleme Dersi 1. Ödev

Alperen Özer

17011069

Regex: ^(([\w,ş,ı,ğ,ü,ç,ö,\.,]*)(MA?H\.?))?(([\w,ş,ı,ğ,ü,ç,ö,\.,]*)(B\.?L\.?V\.?))?(([\w,ş,ı,ğ,ü,ç,ö,\.,]*)(CA?D\.?))?(([\w,ş,ı,ğ,ü,ç,ö,\.,]*)(SOK\.?))?([\w,ş,ı,ğ,ü,ç,ö,\.,]*)?(NO \?:?\.\.? ?[\w,-,V,-]+)?([\w,ş,ı,ğ,ü,ç,ö,\.,]*)*([\w,ş,ı,ğ,ü,ç,ö]*)\/? ?([\w,ş,ı,ğ,ü,ç,ö]*)\$

Örnek link: <https://regex101.com/r/rAC4xJ/2>

Denemede 6832 örnek bulunmaktadır. Verilen regex ile bunların 6004'ü bulunmaktadır. Bu, verilen örneklerin %88'inin bulunduğu anlamına gelmektedir.

Regex satır başından itibaren önce mahalle ismini ve olası mahalle kısaltmalarını, sonra bulvar ismini ve olası bulvar kısaltmalarını, sonra cadde ismini ve olası cadde kısaltmalarını, sonra sokak ismini ve olası sokak kısaltmalarını, sonra numarayı, sonra da ilçeyi ve ili bulmaya çalışmaktadır.

Bulunan bazı örnekler aşağıdaki gibidir:

```
VELİBABA MAH. ANKARA CAD. NO:118 PENDİK/ İSTANBUL
CAFERAĞA MAH. MODA CAD. NO: 21/B KADIKÖY/ İSTANBUL
İNÖNÜ MAH. ULUSU CAD. NO:123/C ATAŞEHİR/ İSTANBUL
OSMANGAZİ MAH. HİHAL CAD. NO:131-133/A SANCAKTEPE/ İSTANBUL
KAYIŞDAĞI MAH. USLU CAD. NO:113 ATAŞEHİR/ İSTANBUL
ŞİFA MAH. HÜRRİYET CAD. NO:49/A TUZLA/ İSTANBUL
MERVE MAH. ATATÜRK CAD. NO.344/C SANCAKTEPE/ İSTANBUL
ŞİFA MAH. SEKERPINAR CAD. DİNÇÖZ SOK. NO:26 TUZLA/ İSTANBUL
MİMAR SİNAN MAH. MİMAR SİNAN CAD. NO:41 ÇEKMEKÖY/ İSTANBUL
ALİBEYKÖY MAH. VARDAR BULVARI NO:1/C EYÜPSULTAN/ İSTANBUL
BULGURLU CAD. TOYGAR SİTESİ ALP APT. NO:118/1 ÜSKÜDAR/ İSTANBUL
GÜZELTEPE MAH. GÜZELTEPE CAD. NO:51/B/2 ÜSKÜDAR/ İSTANBUL
YUKARI DUDULLU MAH. ACISU CAD. NO:63/A ÜMRANİYE/ İSTANBUL
ATALAR MAH. KÖROĞLU CAD. NO:16/A KARTAL/ İSTANBUL
MAREŞAL ÇAKMAK MAH. NECİP FAZIL KISAKÜREK CAD. NO:133/A GÜNGÖREN/ İSTANBUL
ŞEHİT MUHTAR MAH. MİS SOK. NO:4 BEYOĞLU/ İSTANBUL
İSTİKLAL MAH. ATATÜRK CAD. NO.61/A ESENYURT/ İSTANBUL
MOLLA GÜRANİ MAH. AŞAĞI GUREBA CAD. NO:22 FATİH/ İSTANBUL
MAHMUT ŞEVKET PAŞA MAH. KURUCAY CAD. NO:5 ŞİŞLİ/ İSTANBUL
SİNAN PAŞA MAH. BARBAROS BULVARI KÖY İÇİ CAD. GİRİŞİ NO:1/7 BEŞİKTAŞ/ İSTANBUL
BAĞDAT CAD. BAĞDAT APT. NO:268/C KADIKÖY/ İSTANBUL
CUMHURİYET MAH. ŞEHİT TURAN ÇAMUR CAD. NO 1/A ÇEKMEKÖY/ İSTANBUL
```

Bulunamayan bazı örnekler ise açıklamasıyla beraber aşağıdaki gibidir:

```
5. BÖLGE G-5 CAD. BÜFE 2 BAŞAKŞEHİR/ İSTANBUL
KARADENİZ MAH. 1176 SOK. C YAN CEPHE 2/3 GAZİOSMANPAŞA/ İSTANBUL
MARMARA MAH. HÜRRİYET BULVARI 233-C BEYLİKDÜZÜ/ İSTANBUL
```

Hatanın sebebi “-“ ve “/” gibi işaretler kullanılmasıdır, bunlar sadece NO:’da kontrol edilmektedir. Bu karakterler eklendiği zaman başarı oranı fazla yükselmemekte ancak işlem yükü artmaktadır. Bu sebeple işlem yükünü az tutmak için bu karakterler eklenmemiştir.

Bulunamayan bazı örneklerde de NO: dışında bir yerde, daire numarası veya kat eklemek için, “:” kullanılmıştır. Yukarıdakiyle aynı sebepten eklenmemiştir.