



BLM4580 - Doğal Dil İşleme

Ödev Raporu - 1

Enes Kılıçarslan - 17011011

Kullanılan Regex:

((?<mahalle>.*MA?H(\.)?)(?<cadde>.*CA?D(DEŞİ)?(\.)?)(?<sokak>.*SOK(\.)?(AK)?)(?<duzensiz>.+?(?=NO)))(?<numara>(\s)?NO(\s)*?([:.\s)*[0-9]{0,3}(\s)?(-)?(\s)*?[A-Z0-9]{0,3})?)|(((\s)*?(?<semt>[^\s]*)\s)?(?<sehir>İSTANBUL)))

Mahalle => .*MA?H(\.)?

Cadde => .*CA?D(DEŞİ)?(\.)?

Sokak => .*SOK(\.)?(AK)?

Diğer => .+?(?=NO)

Numara => (\s)?NO(\s)*?([:.\s)*[0-9]{0,3}(\s)?(-)?(\s)*?[A-Z0-9]{0,3})?)

Semt => [^\s]*

Şehir => \s?İSTANBUL

Eksiksiz Bulunan Adres Sayısı: 5613

Tam anlamıyla eşleme işlemi içerisinde düzensiz ifadeler bulunmayan adresler için sağlanmıştır. Bu adreslerin formatı => MAH|CAD|SOK|DİĞER|NUMARA|SEMT|ŞEHİR şeklindedir.

Aramayı gerçekleştirirken regex ifadesi sağdan sola, aranan cümle soldan sağa tarandığı için numaradan sonra gelen düzensiz ifadeler bulunamamıştır. Numaradan önce gelen ve mahalle, cadde, sokak gibi regex içerisinde belirtilen ifadelerden sonra gelen düzensiz ifadelerin tespiti de zaten "Numaraya kadar olan kısımları al" şeklinde ifade ettiğimiz +?(?=NO) komutu ile gerçekleşmiştir.

Eksik Adres Sayısı: 1218

Yukarıdaki kurala eksik uyan adreslerden bazıları aşağıda verilmiştir.

YENİBOSNA METRO İSTASYONU BAKIRKÖY/ İSTANBUL

KENNEDY CAD. SİRKECI ARABALI VAPUR İSKELESİ FATİH/ İSTANBUL

YILDIZ POSTA CAD. TÜRK TELEKOM ÖNÜ GAZETE BAYİİ BEŞİKTAŞ/ İSTANBUL

ARMAĞAN EVLER MAH. ALEMDAĞ CAD. SİTE OTOBÜS DURAĞI YANI ÜMRANİYE/ İSTANBUL

BABA HASAN ALEMİ MAH. ATATÜRK BULVARI FATİH/ İSTANBUL

TOPKAPI TİCARET MERKEZİ KARŞISI ÜSTGEÇİT ALTI ZEYTİNBURNU/ İSTANBUL

MİLLET CAD. YUSUF PAŞA ZİRAAT BANKASI ÖNÜ FATİH/ İSTANBUL

E5 KARAYOLU ÜZERİ GÖZTEPE OTOBÜS DURAGI YANI KADIKÖY/ İSTANBUL

ÇAĞLAYAN MAH. ÇAĞLAYAN BENZİN İSTASYONU YANI KAĞITHANE/ İSTANBUL

Bulunamayanlar Hakkında:

İçerisinde yukarıdaki regex komutuna göre en az bir “veya” işlemi için dahi olumlu sonuç veren tüm adresler bulundu/eksik bulundu kategorisine dahil olacağından bulunamayan adres mevcut değildir. En azından her adres bilgisinde İSTANBUL içeriği mevcuttur.

Örnek Eşleşme ve Gruplar:



Online Deneme: <https://regex101.com/r/0ZJRAI/1>