

DOĞAL DİL İŞLEME 1. ÖDEV RAPORU

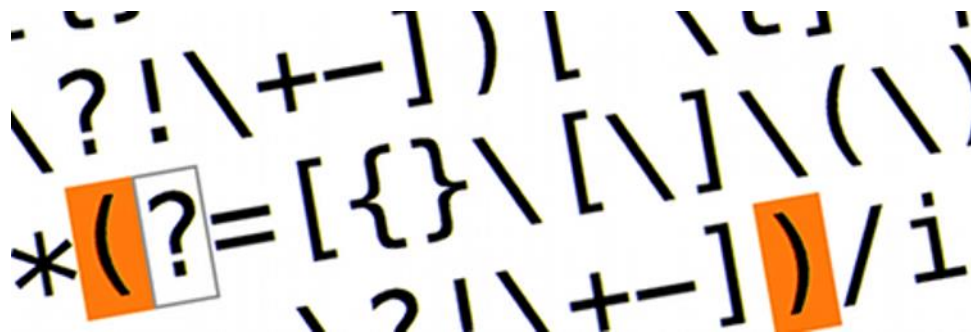
AD : Ahmet

SOYAD: AYDIN

NUMARA : 16011003

DERSİN YÜRÜTÜCÜSÜ: Prof. Dr. Banu DİRİ

ÖDEV KONUSU : Verilen Adresleri Uygun Şekilde Ayıran RegEx Komutunu Oluşturma.



Adreslemeler ana hatlarıyla mahalle,cadde,sokak,no,ilçe/il şeklinde yazılmış.

Eğer mahalle adı girilmediyse ilk olarak cadde adı girilmiştir. Cadde adı girilmediyse sokak adı vb. girilmiştir.Bu yüzden her durum OR komutuyla birbirine bağlanmıştır.

YAZILAN REGEX KOMUTU

```
((.* MA?HA?L?L?E?S?İ?\.?)(.* CA?DD?E?S?İ?\.?)(.* SO?KA?K?\.?)*(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))|((.* CA?DD?E?S?İ?\.?)(.* SO?KA?K?\.?)*(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))|((.* SO?KA?K?\.?)(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))|((.* NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))|((.*\s)(\p{L}*V.*))
```

İLK DURUM RENK YEŞİL

```
((.* MA?HA?L?L?E?S?İ?\.?)(.* CA?DD?E?S?İ?\.?)(.* SO?KA?K?\.?)*(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))
```

*Mahalle adı girilmiş olan adresler bu regex komutu ile bulunur.

İKİNCİ DURUM RENK TURKUAZ

```
((.* CA?DD?E?S?İ?\.?)(.* SO?KA?K?\.?)*(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))
```

*Mahalle adı girilmeyen adresler bu regex komutu ile bulunur.

ÜÇÜNCÜ DURUM RENK KIRMIZI

```
((.* SO?KA?K?\.?)(.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))
```

*Mahalle ve cadde adı girilmeyen adresler bu regex komutu ile bulunur.

DÖRDÜNCÜ DURUM RENK SARI

```
((.*( NO[^0-9]*[\\-\\s\:] [0-9A-z\\-]*)))(.*\s)(\p{L}*V.*))
```

*Mahalle , cadde ve sokak adı girilmeyen adresler bu regex komutu ile bulunur.

BEŞİNCİ DURUM RENK GRİ

`(.*\s)(\p{L}*\V.*)`

*Mahalle , cadde , sokak ve no gibi hiçbir belirleyici bilgi verilmeyen adresler bu regex komutu ile bulunur.

HER BİR REGEX KOMUTUNUN NİTELİĞİNİN AÇIKLANMASI

`(.* MA?HA?L?L?E?S?İ?\.)` = Mahalle adları MH.,MAH. MAHALLESİ veya MAH kısaltması ile verilmiştir.

`(.* CA?DD?E?S?İ?\.)` = Cadde adları CD. , CAD. veya CADDESİ şeklinde verilmiştir.

`(.* SO?KA?K?\.)` = Sokak adları SK. , SOK. veya SOKAK şeklinde verilmiştir.

`.*(NO[^0-9]*[\\-\\s\\:][0-9A-z\\-]*)` = Numaralar aşağıdaki gibi değişik şekillerde girildiği için karışık bir regex komutu yazılmıştır.

NO:2 NO:15/2 NO:4/B NO:7A NO :11 NO:131-133/A NO:58-A

`(.*\s)(\p{L}*\V.*)` = İlçe ve il isimlerini alır.

Regex Komutu Çıktısı

REGULAR EXPRESSION

6823 matches, 2313484 steps (~1.75s)

```

/ gm
((.*MA?HA?L?L?E?S?i?i?.?)*(.*CA?DD?E?S?i?i?.?)*(.*SOKA?K??.?)*(.*(NO[^\0-9]*
[\\/\-\\s\.:][\0-9A-z\\/\-]*)*(.*\s)(\p{L}*\\/.*))|((.*CA?DD?E?S?i?i?.?)
(.*SOKA?K??.?)*(.*(NO[^\0-9]*[\\/\-\\s\.:][\0-9A-z\\/\-]*)*(.*\s)
(\p{L}*\\/.*))|((.*SOKA?K??.?)*(.*(NO[^\0-9]*[\\/\-\\s\.:][\0-9A-z\\/\-]*)*
(.*\s)(\p{L}*\\/.*))|((.*(NO[^\0-9]*[\\/\-\\s\.:][\0-9A-z\\/\-]*)*(.*\s)
(\p{L}*\\/.*))|((.*\s)(\p{L}*\\/.*))

```

TEST STRING

YENİBOSNA METRO İSTASYONU BAKIRKÖY/ İSTANBUL
KENNEDY CAD. SİRKEÇİ ARABALI VAPUR İSKELESİ FATİH/ İSTANBUL
YAVUZTÜRK MAH. KARADENİZ CAD. NO:2 ÜSKÜDAR/ İSTANBUL
HAMİDİYE MAH. ALPEREN SOK. NO:15/2 ÇEKMEKÖY/ İSTANBUL
UĞUR MUMCU MAH. YUNUS EMRE CAD. NO:25 KARTAL/ İSTANBUL
BAĞLARBAŞI MAH İNÖNÜ CAD NO:3 MALTEPE/ İSTANBUL
HASANPAŞA MAH. FAHRETTİN KERİM GÖKAY CAD. KADIKÖY/ İSTANBUL
P.T.T. EVLERİ BAHÇEKÖY CAD. NO: 53 SARIYER/ İSTANBUL
KARAKÖY YER ALTI GEÇİDİ NO:24 BEYOĞLU/ İSTANBUL
ÖRNEK MAH. DOĞ. ARS. BLV FİKRİ SÖN CAD. GİRİŞİ AGENA E NO. 215 9/2 ESENİYURT/ İSTANBUL
GÜRSEL MAH.28 NİSAN CAD.NO:4/B KAĞITHANE/ İSTANBUL
ATATÜRK MAH. ALEMDAĞ CAD. NO:61 ÜMRANİYE/ İSTANBUL
YILDIZ POSTA CAD. TÜRK TELEKOM ÖNÜ GAZETE BAYI BEŞİKTAŞ/ İSTANBUL
ARMAĞAN EVLER MAH. ALEMDAĞ CAD. SİTE OTOBÜS DURAĞI YANI ÜMRANİYE/ İSTANBUL
FETİHTEPE MAH. FATİH SULTAN CAD. NO:37/B BEYOĞLU/ İSTANBUL
BOZKURT MAH. KURTULUŞ CAD. NO:135/A ŞİŞLİ/ İSTANBUL
PAŞABAĞCI MAH. BARBAROS CAD. NO:4/A BEYOĞLU/ İSTANBUL
YEŞİL PINAR MAH. ŞÜKRAN SOK. NO:36/B EYÜPSULTAN/ İSTANBUL
MUSTAFA KEMAL PAŞA CAD. AZİMKAR SOK.14/1 FATİH/ İSTANBUL
YENİ DOĞAN MAH. KİŞLA CAD. NUR SANAYİ SİTESİ NO: 69 BAYRAMPAŞA/ İSTANBUL
DENİZ KÖŞKLER MAH. ESKİ LONDRA ASFALTI NO: 22/3 AVCILAR/ İSTANBUL
ORTAÇEŞME SONDURAK BEYOĞLU/ İSTANBUL
MEHMET AKİF ERSOY MAH. NATO YOLU BOSNA BULVARI NO: 115 ÜSKÜDAR/ İSTANBUL
DERBENT MAH. DEREİÇİ OTOBÜS SON DURAK SARIYER/ İSTANBUL
LEVAZIM MAH. KORU SOK NO:7A BEŞİKTAŞ/ İSTANBUL
ÇENGELKÖY MAH. ÇENGELKÖY CAD. NO:35/A ÜSKÜDAR/ İSTANBUL
BÜYÜKDERE CAD. NİMET ABLA CAMİ YANI ŞİŞLİ/ İSTANBUL
REŞİTPAŞA MAH. POSTAYOLU CAD. NO :85 SARIYER/ İSTANBUL
RUMELİ HİSAR ÜSTÜ BEBEK YOLU SOK SARIYER/ İSTANBUL
HASANPAŞA MAH. KURBAĞALIDERE CAD. NO:23/A KADIKÖY/ İSTANBUL
SEYRANTEPE MAH. İ. KARAOĞLANOĞLU CAD. NO: 153/B KAĞITHANE/ İSTANBUL
TAHTAKALE MAH. GAFFAR OKAN CAD. NO:10 AVCILAR/ İSTANBUL
BABA HASAN ALEMİ MAH. ATATÜRK BULVARI FATİH/ İSTANBUL
TOPKAPI TİCARET MERKEZİ KARŞISI ÜSTGEÇİT ALTI ZEYİNBURNU/ İSTANBUL
MİLLET CAD. YUSUF PAŞA ZİRAAT BANKASI ÖNÜ FATİH/ İSTANBUL
SÖĞÜTLÜ ÇEŞME MAH. FEVZİ ÇAKMAK CAD. NO:42 KÜÇÜKÇEKMECE/ İSTANBUL
MALTEPE MAH. MATBAACILAR SİTESİ MALTEPE METROBÜS ÇIKIŞI NO:133 BAYRAMPAŞA/ İSTANBUL
ATATÜRK BULVARI 32 C FATİH/ İSTANBUL

Hatalar

Hata 1:

ORHANGAZİ MAH. NECMETTİN ERBAKAN CAD. NO:127 C PENDİK/ İSTANBUL

Adres NO formatı yukarıda da belirtildiği üzere zaten çok kapsamlı daha da karışık olmaması adına maalesef bu istisnai hata görmezden geliniyor.

Hata 2:

ATATÜRK MAH. ATAŞEHİR BULVARI CAD. 28/A KARDELEN BLOKLARI ATAŞEHİR/ İSTANBUL

Adres NO'su verilirken NO etiketi ile belirtilmediği için NO bilgisi BLOK bilgisi ile iç içe geçmiş vaziyette. Benzer örnek;

MERKEZ MAH. EYÜP CAD. DÜNYA İŞ MERKEZİ 10/A GAZİOSMANPAŞA/ İSTANBUL

Hata 3:

ÇİFTLİK MAHALLESİ ÇAVUŞBAŞI CUMHURİYET CAD. NO: 223/B BEYKOZ/ İSTANBUL

Mahallesi ismi girilirken MAHALLESİ kelimesini de bulmasına yönelik RegEx komutu girilmesine rağmen (BKNZ : (. *MA?HA?L?L?E?S?İ?\.?)) maalesef atlıyor. O yüzden cadde isminin bir kısmı mahalle ismine dahil oluyor.

Hata 4:

SEYİT NİZAM MAH. AKŞEMSETTİN TRAMVAY DURAĞI ÖĞRETMENLER PARKI NO:1 ZEYTİNBURNU/ İSTANBUL

RegEx komutu yazılırken DURAK ve PARK nitelikleri kontrol edilmediği için birlikte tek nitelik olarak sayılıyorlar. Kolay bir şekilde kontrolü sağlanabilir. Fakat nadir denk geldiği için optimizasyonu sağlamak adına bu nitelikler RegEx komutuna dahil edilmemiştir.

Hata 5:

MUHSİNE HATUN MAH TEVAŞİ ÇEŞMESİSOK NO :9 A FATİH/ İSTANBUL

SOKAK kısmını bütün halinde doğru bir şekilde almış fakat SOK niteliği sokak adıyla bitişik yazıldığı için sokak ismi doğru nitelik kategorisinde yer almamış. NO kısmı da ilk hatada denildiği gibi çok farklı formatla yazıldığı için işlem karmaşıklığını artırmamak adına nadirde olsa yanlış yazılmıştır.

Hata 6:

ESKİ EDİRNE ASFALTI NO:168 GAZİOSMANPAŞA/ İSTANBUL

Mahalle, cadde, sokak gibi nitelikler girilmediği zaman maalesef yukarıda da örneği olduğu gibi ESK((.*SO?KA?K?\.?)) kelimesini SOKAK niteliği olarak sayıyor. Benzer örnek;

YENİ VAPUR İSKELESİ KARŞISI KADIKÖY/ İSTANBUL

Bu hatanın çözümü rapor yazılırken fark edilmiş olup düzeltilmiştir. Çözüm olarak nitelik adlarından önce 'boşluk' karakteri gelmesi şartı konmuştur. Çözümünden sonraki çıktılar;

ESKİ EDİRNE ASFALTI NO:168 GAZİOSMANPAŞA/ İSTANBUL

YENİ VAPUR İSKELESİ KARŞISI KADIKÖY/ İSTANBUL

Bu şekilde Hata 3 de çözülmüştür. 😊

ÇİFTLİK MAHALLESİ ÇAVUŞBAŞI CUMHURİYET CAD. NO: 223/B BEYKOZ/ İSTANBUL

SONUÇ

Regex komut çıktısı görselinde de görüldüğü üzere tüm adresler bulunuyor. Genel olarak sorunumuz adres bilgisini doğru seçmede değil adres bilgisini verimli bir şekilde yorumlama. Fakat yukarıda da görüldüğü üzere yorumlarken oluşan hataların hepsinin çözümü kolaylıkla sağlanabilir ama performansın ciddi bir şekilde düşmemesi adına nadir de olsa oluşan hatalara göz yumulmuştur.

Başarılı bulunan adres sayısı = 6778

Başarı oranı = %99.22