

YILDIZ TEKNİK ÜNİVERSİTESİ
ELEKTRİK-ELEKTRONİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



DOĞAL DİL İŞLEMeye GİRİŞ

ÖDEV: 1

Regex Komutları İle Adres Tespiti

16011073 – İpek KOÇ

2020

Regex kodu:

```
^((.*)\sMAHALLES?[İİ]?|(.*) MAH?[\\.\s])?((.*)\sBULVARI |  
(.*)\sBU?LV[\\.\s])?((.*)CADDES?[İİ]?|(.*)\sCA?D[\\.\s])?  
((.*)\sSOKAK[\\s]|(.*)\sSO?K[\\.\s])?((.*)S[İİ]TES[İİ]|  
(.*)\sS[İİ]?T[\\.\s])?( \sNO.*)?(.*)?(.*)?(.*)\s/)(.*)$
```

Açıklama: Ödev kapsamında regex kodu yazılarak adres tespiti istenmiştir. Burada adres tespiti yapılacak olan dosya içerisinde 6831 adet adres bulunmaktadır.

Çözüm: Dosya içerisinde yer alan adresler incelenerek bulunması gereken yapılar tespit edilmiştir. Bunlar: mahalle, bulvar, cadde, sokak, site, numara, ilçe ve ildir. Ayrıca yalnızca ilçe ve il bilgisi ile lokasyon tanımı olan adresler de göz önünde bulundurulmuştur.

Başarı: 6831 adres yer alan dosyadan adres bulma olarak değerlendirildiğinde 6831 adres bulabilmektedir. Fakat aşağıda belirtileceği gibi bazı olumsuz gruplamalar yer almaktadır.

Doğru bulunan adres sayısı = 6678

Hatalı gruplanan adres sayısı = 153

Başarı oranı = %97.76

Örnek Durumlar:

1. Başarılı durum: 6831 adrese ulaşıldı ve gruplara bölündü.

REGULAR EXPRESSION 6831 matches, 7277438 steps (~8.74s)

```

^((.*)\sMAHALLES?([İİ])?|(.*) MAH?H?[\.\s])?((.*)\sBULVARI |
(.*)\sBU?LV[\.\s])?((.*)CADDES?([İİ])?|(.*)\sCA?D[\.\s])?
((.*)\sSOKAK[\s]|(.*)\sSO?K[\.\s])?((.*)S([İİ])TES([İİ])|
(.*)\sS([İİ])?T[\.\s])?( \sNO.*)?(.*)?(.*)?(\\|)(.*)$

```

TEST STRING

ORTAÇEŞME SONDURAK BEYKÖZ/ İSTANBUL
MEHMET AKİF ERSOY MAH. NATO YOLU BOSNA BULVARI NO: 115 ÜSKÜDAR/
İSTANBUL
DERBENT MAH. DEREİCİ OTOBÜS SON DURAK SARIYER/ İSTANBUL
LEVAZİM 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

2. Tarifler ile ilgili hatalı gruplama ve doğru gruplama farkı: “Rumeli Hisar Üstü” adres içerisinde bulunduğu konum dolayısı ile sokak bilgisinden ayrılmamıştır.

RUMELİ HİSAR ÜSTÜ BEBEK YOLU SOK SARIYER/ İSTANBUL
TOPKAPI TİCARET MERKEZİ KARŞISI ÜSTGEÇİT ALTI ZEYTİNBURNU/ İSTANBUL
MİLLET CAD. YUSUF PAŞA ZİRAAT BANKASI ÖNÜ FATİH/ İSTANBUL

3. Sokak bilgisi hatası: Belirleyici kelimelerden önce boşluk gelmelidir. Bu sebeple “SOKAK” ya da “SOK.” kelimeleri öncesinde bulunan stringe birleşirse bütün olarak ele alınır. “S.” kontrolü yapılmamıştır. Çünkü adres tanımları içerisinde bulunan kısaltmalar ile karışabilmektedir.

BARBAROS MAH. 171.SOKAK NO:2/A BAĞCILAR/ İSTANBUL
K.TEPE MAH FİLİZ S. NO:16 BAKIRKÖY/ İSTANBUL
SÜMER MAH. 18.SOK.NO:17/A ZEYTİNBURNU/ İSTANBUL
ESENEVLER MAH. 3001.CAD. 28/A ÜMRANİYE/ İSTANBUL

4. Numara bilgisinin adres tanımlama içerisinde verilmesi durumu:

AYDINLI MAH. DERSADET CAD. 26/1 103 NO' LU HALK EKMEK BÜFESİ TUZLA/
İSTANBUL

5. Cadde hatası: 3. örnekte sokak için yaşanan benzer durum geçerlidir. “C” kısaltma olarak alınabileceğinden cadde tanımı içerisinde yer almamaktadır.

ESENTEPE M. TOROS C. 5 B KARTAL/ İSTANBUL
"EMEK PINAR C RESİM SOKAK" SANCAKTEPE/ İSTANBUL

6. Apartman bilgisinin numara dahil olması: Apartman bilgisi tarif olarak kabul edildiğinden numara bilgisinden sonra yer alması gerekmektedir. Aksi halde durum 3’te yer aldığı gibi numara bilgisine dahil edilir.

CUMHURİYET MH. AŞIK VEYSEL CD. NAMLI APT. NO:57 KÜÇÜKÇEKMECE/
İSTANBUL

ÖZET: Yukarıda verilen örnekler ve bunlara benzeyen durumlarda gruplama hatası yaşanmaktadır. Fakat tüm adresler tespit edilebilmektedir.