



Yıldız Teknik Üniversitesi
Bilgisayar Mühendisliği Bölümü
Doğal Dil İşleme Dersi Ödev-1

Serkan Kaan Bahşı

17011055

KULLANILAN REGEX:

([\\s]+[Mm][Aa][Hh]([Aa][Ll][Ll][Ee])*[\\s]+)|([\\s]+[Cc][Aa][Dd]([Dd][Ee])*[\\s]+)|([\\s]+[Ss][Oo][Kk]([Aa][Kk])*[\\s]+)|([\\s]+[Aa][Pp][Tt][\\s]+)|([\\s]*[Nn][Oo][\\s]*([:])*[\\s]*([0-9]+[\\s]*)*[A-Z]*)([\\s]*[V][\\s]*)

Regex:

- 1- Mahalle, cadde, sokak, no, apt kısımlarını ayırt edebiliyor.
- 2- No'dan sonra gelen sayılar ve semboller ayırt edilebiliyor.
- 3- İl, ilçelerin ayrımı yapılıyor.
- 4- Blok ve site ayrımı yapılamıyor.

Yukarıda verilen regex'i regex101 sitesinde denedim. Sonuçlar şu şekilde:

Doğru Sonuçlar:

YEŞİL PINAR MAH. GİRNE CAD. MUHTAR CAMİİ SOK. NO:1 EYÜPSULTAN/ İSTANBUL
F. ÇAKMAK MAH. OSMANGAZİ CAD. NO: 72/B BAĞCILAR/ İSTANBUL
YAVUZ SELİM MAH. ANTALYA CAD. NO:16/B SULTANBEYLİ/ İSTANBUL
ŞEHREMİNİ MAH. MİLLET CAD. NO:121-B FATİH/ İSTANBUL
ZÜMRÜTEVLER MAH. NİL CAD. NO:37 MALTEPE/ İSTANBUL
TALAT PAŞA MAH. KESKİNGİL CAD. NO :8 KAĞITHANE/ İSTANBUL
YEŞİLKÖY MAH. İSTASYON CAD. NO: 10/2A D:1 BAKIRKÖY/ İSTANBUL
YAMANEVLER MAH. KÜÇÜKSU CAD. NO:26 ÜMRANİYE/ İSTANBUL
SOĞUKSU MH. M.AKİF ERSOY CD. NO:141/C BEYKOZ/ İSTANBUL
19 MAYIS MAH. İNÖNÜ CAD. NO:48/B KADIKÖY/ İSTANBUL
DUMLUPINAR MAH. ANADOLU CAD. NO:53/B ÜMRANİYE/ İSTANBUL
İSTİKLAL MAH. HACI RÜSTEMOĞLU SOK. NO:3/A ÜMRANİYE/ İSTANBUL
İHASANİYE MAH. HAREM İSKELE CAD. NO.6/A ÜSKÜDAR/ İSTANBUL

Mahalle, cadde, sokak ve No düzgün bir şekilde verildiğinde adres bilgilerini parçalara ayırabiliyor.

Eksik Sonuçlar:

KORDONBOYU MAH. ANKARA CAD. TOGAY APT. ANT SİTESİ A BLOK 175/1 KARTAL/ İSTANBUL
SARAÇ İŞHAK MAH ÇİFTE GELİNLER İSTANBUL OTELCİLER ODASI 40 1 FATİH/ İSTANBUL
TELSİZ MAH. 85.CADDESİ NO:6-A ZEYTİNBURNU/ İSTANBUL

Fazladan bilgi veya bir bilginin farklı bir şekilde verilmesi durumlarında. Bilgileri eksik alıyor. Regex'de site numarası, blok numarası için çözümler yok.

Bulunamayanlar:

YENİBOSNA METRO İSTASYONU BAKIRKÖY/ İSTANBUL

EVREN MAHALLESİ İSTİKLAL CADDESİ HÜNKAR SOK. NO: 117/C BAĞCILAR/ İSTANBUL

Bu tarz çok bilgi içermeyen örneklerde ise. En fazla il/ilçe ayrımı belli oluyor. İkinci örnekte mahallesi kelimesini tespit edemedi.

Sonuçlar:

- 1- Regex'in doğru bir şekilde ayrabilmesi için belirli kurallar lazım.
- 2- Adres yapısı belirginleştikçe başarı artmakta.
- 3- Düzensiz yazımlar ve beklenmedik durumlar karşılaşılan en büyük zorluk.