



Doğal Dil İşleme

Ödev 1

Ödev Konusu:
RegEx

16011089

ENES TEZCAN

YÖNTEM

Her alan için ayrı ayrı RegEx komutları:

Adresler Mahalle, Cadde, Sokak, Diğer(Yanı,Merkezi, Durağı vb.), No Ve Şehir olarak ayrılmıştır

RegEx Komutu:

```
(?<MAH>\s?[A-z,\.,ç,Ç,ğ,Ğ,ı,İ,ö,Ö,ş,Ş,ü,Ü,0-9,\-  
," ,\s]+MA?HA?L?L?E?S?İ?\.\?[\s\.]*)*|(?<CAD>[A-  
z,\.,ç,Ç,ğ,Ğ,ı,İ,ö,Ö,ş,Ş,ü,Ü,0-9,\-  
," ,\s]+CA?DD?E?S?İ?\.\?)*[\s\.]*|(?<SOK>[A-  
z,\.,ç,Ç,ğ,Ğ,ı,İ,ö,Ö,ş,Ş,ü,Ü,0-9,\-  
," ,\s]+(BULVA?R?I?|SO?KA?K?Ğ?I?\.\?))*[\s\.]*|(?<DIGER>[  
A-z,\.,ç,Ç,ğ,Ğ,ı,İ,ö,Ö,ş,Ş,ü,Ü,0-9,\-  
," ,\s]+(. *APT\.\?|. *DURAK?Ğ?I?\.\?|. *İSTASYONU?|. *YANI|. *EKM  
EK|. *PLAZA|. *AVM|. *MERKEZİ?I?|KARŞISI))*[\s\.]*|(?<NO>NO?\s?  
.\?:\?:\?[\s\?[\0-9\\\/A-Z?])*|(?<sehir>[A-  
Z,Ç,Ğ,İ,ö,Ö,Ş,Ü,\s]+\?\?\/?\.\?\sİSTANBUL$)
```

Eksiksiz bulduğu adres sayısı: 4013

Hiç bulamadığı adres sayısı: 0

Eksik bulduğu adres sayısı: 0

- Mahalle, Cadde, No olarak belirmediği adresler sıkıntılı bulunmuştur.
- Beklenmedik boşluklar ve Cad. Mah. gibi kısaltmalardan sonra boşluk gelmemesi durumlarında hatalar meydana gelmiştir.

UYGULAMA

REGULAR EXPRESSION v4

63283 matches, 47215705 steps (~40.62s)

```
// (?<MAH>\s?[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+MA?HA?L?L?E?S?İ?\.?  
[\s\.]*)*(?<CAD>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+CA?DD?E?S?İ?\.?)?  
[\s\.]*(?<SOK>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+(BULVA?R?I?|SO?KA?K?  
Ğ?I?\.?)*)[\s\.]*(?<DİĞER>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+(.?APT\.  
|.?.*DURAK?Ğ?I?\.?.|.?.*İSTASYONU?|.?.*YANI|.?.*EKMEK|.?.*PLAZA|.?.*AVM|.?.*MERKEZİ?I?|KARŞISI)))*  
[\s\.]*(?<NO>NO?\s?\.\.?\:?\s?[\0-9\\\/A-Z?]*)?(?<şehir>[A-Z,Ç,Ğ,İ,Ö,Ş,Ü,\s]+\|?  
\/?\.\.?\sİSTANBUL$)
```

TEST STRING

HASEKİ KÜÇÜK MÜHENDİS SOK. NO:7 FATİH/ İSTANBUL
BAHÇE ŞEHİR İSPARTA KULE MEVKİİ PARK 1 ILGAZ VİLLALARI YANI BAŞAKŞEHİR/ İSTANBUL
ÖRNEK TEPE MAH. ETİBANK CAD. NO: 105 BEYOĞLU/ İSTANBUL
ÇAPA MİLLET CAD.-TATLI PINAR CAD. KESİŞİMİ FATİH/ İSTANBUL
BİRLİK MAH. K. KARABEKİR CAD. NO:92 ESENLER/ İSTANBUL
CAMİİ KEBİR MAH. DERE BOYU CAD. NO :43/A BEYOĞLU/ İSTANBUL
FATİH MAH. CENGİZ TOPEL CAD. SARDUNYA SOK. 84 B BÜYÜKÇEKMECE/ İSTANBUL
5. GAZETECİLER SİTESİ EBRU SOK. OTOBÜS DURAĞI BEŞİKTAŞ/ İSTANBUL
MİTHAT PAŞA CAD. NO:2 BAHÇELİEVLER/ İSTANBUL
KEMANKEŞ MAH. KEMANKEŞ CAD. DENİZ İST. GENEL. KARŞISI BEYOĞLU/ İSTANBUL
YUNUS EMRE MAH. BARBAROS CAD. NO:3 SANCaktepe/ İSTANBUL
İSTİKLAL MAH. ALEMDAĞ CAD. NO:272/C ÜMRANİYE/ İSTANBUL
MİMAR SİNAN MAH. ATLAS SAK. NO:18/B ATAŞEHİR/ İSTANBUL
ZÜMRÜTEVLER MAH.SERHAT SOK. NO:1 MALTEPE/ İSTANBUL
AYDINLI MAH. ATATÜRK CAD. NO:65/2 TUZLA/ İSTANBUL

REGULAR EXPRESSION v4

63283 matches, 47215705 steps (~40.62s)

```
// (?<MAH>\s?[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+MA?HA?L?L?E?S?İ?\.?  
[\s\.]*)*(?<CAD>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+CA?DD?E?S?İ?\.?)?  
[\s\.]*(?<SOK>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+(BULVA?R?I?|SO?KA?K?  
Ğ?I?\.?)*)[\s\.]*(?<DİĞER>[A-z,\.,ç,ç,ğ,ğ,ı,ı,ö,ö,ş,ş,ü,ü,0-9,\-,"'\s]+(.?APT\.  
|.?.*DURAK?Ğ?I?\.?.|.?.*İSTASYONU?|.?.*YANI|.?.*EKMEK|.?.*PLAZA|.?.*AVM|.?.*MERKEZİ?I?|KARŞISI)))*  
[\s\.]*(?<NO>NO?\s?\.\.?\:?\s?[\0-9\\\/A-Z?]*)?(?<şehir>[A-Z,Ç,Ğ,İ,Ö,Ş,Ü,\s]+\|?  
\/?\.\.?\sİSTANBUL$)
```

TEST STRING

OSMAN GAZİ MAH. LÜKMAN KERİM CAD. NO: 93 DARICA/ İSTANBUL
OSMAN GAZİ MAH. AŞIROĞLU CAD. NO: 122 GEBZE/ İSTANBUL
Y.DUDULLU MAH. ŞENEL CAD. NO: 78 ÜMRANİYE/ İSTANBUL
MERKEZ MAH. SADABAT CAD. NO: 26 KAĞITHANE/ İSTANBUL
ÜNİVERSİTE MAH. BAHAR SOK. ELİF SİT. A BLOK D: 4 AVCILAR/ İSTANBUL
MUSTAFA KEMAL PAŞA CAD. NO:1 BAHÇELİEVLER/ İSTANBUL
İSKENDER PAŞA MAH. SİMİTÇİ ŞAKİR SOK. NO: 32/B FATİH/ İSTANBUL
İNCİRKÖY MAH. SOĞUKSU CAD. NO: 1/A BEYKOZ/ İSTANBUL
SAHRAYICEDİT MAH. F. KERİM GÖKAY CAD. NO: 276/A KADIKÖY/ İSTANBUL
POSTANE MAH. FEVZİ ÇAKMAK CAD. NO: 118/B TUZLA/ İSTANBUL
MAHMUTBEY HACI BOSTAN CAD. NO:24 BAĞCILAR/ İSTANBUL
GÖZTEPE MAH. ZAMBAK SOK. NO:42 BAĞCILAR/ İSTANBUL
OSMANİYE MAH. ÜMRANİYE SOK. NO:7 BAKIRKÖY/ İSTANBUL
GÜMÜŞSUYU CAD. LİTTROS YOLU KOŞE HAN. NO:14-5-6 ZEYTİNBURNU/ İSTANBUL
YAYLA MAH. CİNDERESİ SOK NO 29/A ŞİŞLİ/ İSTANBUL
ÇAĞLAYAN MAH. VATAN CAD. NO:39/B KAĞITHANE/ İSTANBUL