

DOĞAL DİL İŞLEMEYE GİRİŞ

1. ÖDEV

Emre Nitim

17011079

Konu: Verilen txt dosyası içerisindeki adreslerin yazılacak RegEx ile tespit edilmesi.

1. RegEx Komut Satırı Analizi

Yöntem: RegEx ifadesini yazmadan önce bir adreste bulunması gereken bilgileri belirledim ve bunu mahalle, cadde, sokak, no, daire, kat, ilçe ve il olarak parçalara böldüm. Daha sonra bu parçaları bulan RegEx ifadeleri yazıp bunları | sembolü ile birleştirdim ve tüm bu ifadeye + sembolü ekleyerek adresleri tespit eden RegEx'i tamamladım.

a) Tüm Komut Satırı

```
((.+ (MAHALLESİ|MAH|MH) [,\\. ]*)|(.+ (CADDESİ|CAD|CD) [,\\. ]*)|(.+ (SOKAK|SOKAĞI|SOK|SK) [,\\. ]*)|(NO *[:]? *[0-9]+( *[/\\-\\]? *([0-9]* *[a-zA-Z]|[0-9]+))?)|((DAİRE|D) *[:]? *[0-9]+ +)|((KAT|K) *[:]? *[0-9]+[, ]+)|([a-zA-ZıİğĞüÜşŞöÖçÇ]+ *[/\\-\\]? *[a-zA-ZıİğĞüÜşŞöÖçÇ]+))+
```

b) Mahalleyi Bulan Komut Parçası

```
((.+ (MAHALLESİ|MAH|MH) [,\\. ]*)
```

c) Caddeyi Bulan Komut Parçası

```
(.+ (CADDESİ|CAD|CD) [,\\. ]*)
```

d) Sokağı Bulan Komut Parçası

```
(.+ (SOKAK|SOKAĞI|SOK|SK) [,\\. ]*)
```

e) No'yu Bulan Komut Parçası

```
(NO *[:]? *[0-9]+( *[/\\-\\]? *([0-9]* *[a-zA-Z]|[0-9]+))?)? +)
```

f) Daireyi Bulan Komut Parçası

```
((DAİRE|D) *[:]? *[0-9]+ +)
```

g) Katı Bulan Komut Parçası

```
((KAT|K) *[:]? *[0-9]+[, ]+)
```

h) İlçe ve Şehri Bulan Komut Parçası

```
([a-zA-ZıİğĞüÜşŞöÖçÇ]+ *[/\\-\\]? *[a-zA-ZıİğĞüÜşŞöÖçÇ]+)
```

2. Tespit Edilen Adresler

Yazdığım RegEx'i verilen metinde regex101.com üzerinden çalıştırdım ve aşağıdaki sonuçları elde ettim.

a) Tamamı Doğru Bulunan Adres Miktarı

Kullandığım RegEx 6831 farklı adresten 4583 tanesini eksiksiz buldu.

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

b) Eksik Bulunan Adres Miktarı

Kullandığım RegEx 6831 farklı adresten 2248 tanesini eksik olarak buldu.

ÖMERLİ MAH. TERME SOK. A.V.M ÇARŞI NO:13 ARNAVUTKÖY/ İSTANBUL
CERRAH PAŞA MAH. KOCAMUSTAFA PAŞA CAD. NO.59A FATİH/ İSTANBUL
ORDU CAD. MARMARA ÇARŞISI ÖNÜ GAZETE BAYİ 35/1 FATİH/ İSTANBUL

c) Tamamı Eksik Olan Adres Miktarı

Kullandığım RegEx 6831 farklı adresin her birinden en az 1 kısmı doğru bir şekilde buldu.

BAŞAK ŞEHİR 4.ETAP1. KISIM BAŞAKŞEHİR/ İSTANBUL