

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
ELEKTRİK ELEKTRONİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

DERSİN ADI

BLM4580 Doğal Dil İşlemeye Giriş

Regex İle Adres Tespiti

DERSİ VEREN ÖĞRETİM ÜYESİ

Prof. Dr. Banu DİRİ

ÖDEVİ YAPAN ÖĞRENCİ

Elif Özkan

17011024

ÖDEV TESLİM TARİHİ

11.11.2020

Adresler.txt içerisindeki adresleri en başarılı şekilde tespit edilebilmesi için aşağıdaki RegEx komutu yazılmıştır.

```
([0-9a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Mm][Aa]?[Hh][Aa]?[Ll]*[Ee]?[Ss]?[İi]?[\.\.]?[\\.\.])|([0-9a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Bb][Uu]?[Ll][Vv]?[Aa]?[Rr]?[İi]?[\\.\.]?[\\.\.])|([0-9a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Cc][Aa]?[Dd]+[Ee]?[Ss]?[İi]?[\\.\.]?[\\.\.])|([0-9a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Ss][Oo]?[Kk][Aa]?[Kk]?[\\.\.]?[\\.\.])|([a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Aa][Pp][Aa]?[Rr]?[Tt][Mm]?[Aa]?[Nn]?[İi]?[\\.\.]?[\\.\.])|([a-zA-ZÂÏÖÜŞÇĞıöüğşç\-\\" \.]*[Ss][İi][Tt][Ee]?[Ss]?[İi]?[\\.\.]?[\\.\.])|([Nn][Oo]?[Uu]?[\\.\.]?[\\.\.])|([0-9VA-Za-z\-\"]*)([a-zA-ZÂÏÖÜŞÇĞıöüğşç]*V ?)|([a-zA-ZÂÏÖÜŞÇĞıöüğşç]*$)
```

Bu komut ile;

- MAH., MH., MAHALLESİ vb. tanımlanan mahalleler,
- CAD., CD., CADDESİ vb. tanımlanan caddeler,
- BUL., BL., BULVARI şeklinde tanımlanan bulvarlar,
- SK., SOK., SOKAK şeklinde tanımlanan sokaklar,
- SİT., SİTESİ, APT., APARTMANI şeklinde tanımlanan binalar,
- NO:, NO., NU:, NU. Şeklinde tanımlanan kapı numaraları,
- İlçe / İl şeklinde tanımlanan konum bilgileri elde edilebilmektedir.

Kullanılan regex’te tam olarak 9 grup vardır. İlk grup mahalle, ikinci grup bulvar, 3.grup cadde, 4.grup sokak, 5.grup apartman, 6.grup site, 7.grup no, 8.grup ilçe, 9.grup il bilgisi aramaktadır.

Tüm gruplar için yapılan aramalar sonucunda tam olarak 39.344 eşleşme bulunmuştur. Ancak satırdaki bütün elemanların eksiksiz bir şekilde gruplandığı, yani adresin tüm bilgilerinin ayrıştırılabildiği satır sayısı 4550’dir. Bu durum göz önüne alındığında başarı oranı %66.6’dır.

Aşağıda birkaç örnek bulunmaktadır.

Çıktıda görüldüğü üzere, kapı numarasının “D:” şeklinde tanımlandığı satırda kapı numarası bilgisi elde edilememiştir. Bunun için regex’e ekleme yapılmalıdır.

KÜPLÜCE MAH.	MEHMET AKİF ERSOY CAD.	NO:1 ÜSKÜDAR/	İSTANBUL
TARLABAŞI BULVARI	NO:190/B BEYOĞLU/	İSTANBUL	
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
VALİDEATİK MAH.	NUHKUYUSU CAD.	NO:69/A ÜSKÜDAR/	İSTANBUL
GÜZELTEPE MAH.	KÜÇÜK NAMAZGAH CAD.	NO:43/A ÜSKÜDAR/	İSTANBUL

ATATÜRK MAH. ALEMDAĞ CAD. NO:61 ÜMRANİYE/ İSTANBUL
YILDIZ POSTA CAD. TÜRK TELEKOM ÖNÜ GAZETE BAYII 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
PAŞABAHÇE MAH. BARBAROS CAD. NO:4/A BEYKOZ/ İSTANBUL
YEŞİL PINAR MAH. ŞÜKRAN SOK. NO:36/B EYÜPSULTAN/ İSTANBUL
MUSTAFA KEMAL PAŞA CAD. AZİMKAR SOK.14/1 FATİH/ İSTANBUL
YENİ DOĞAN MAH. KIŞLA CAD. NUR SANAYİ SİTESİ NO: 69 BAYRAMPAŞA/ İSTANBUL
DENİZ KÖŞKLER MAH. ESKİ LONDRA ASFALTI NO: 22/3 AVCILAR/ İSTANBUL
ORTAÇEŞME SONDURAK BEYKOZ/ İSTANBUL
MEHMET AKİF ERSOY MAH. NATO YOLU BOSNA BULVARI NO: 115 ÜSKÜDAR/ İSTANBUL
DERBENT MAH. DEREİCİ 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

Yukarıda gösterilen çıktıda ise bulunamayan adres kısımlarının hiçbirinde mahalle, cadde, sokak bilgilerinin olmadığı görülmektedir. Aramak için kullanılan regex'te durak, cami, asfalt gibi kelimeler aranmadığından ötürü bu bilgiler elde edilememektedir.

GULBAHAR MH GULBAG CD NO:2/D ŞİŞLİ/ İSTANBUL
AYAZAGA MH SEHİT İLHAN YURT SK. N77 ŞİŞLİ/ İSTANBUL
SANAYİ MAHALLESİ SULTAN SELİM CADDESİ NO:66 KAĞITHANE/ İSTANBUL
ŞEHİT ERTUĞRUL KABATAŞ CAD. NO:1/1 ŞİŞLİ/ İSTANBUL
ŞAİR NEDİM CAD. NO:65/A BEŞİKTAŞ/ İSTANBUL
SANAYİ MAH SULTAN SELİM CAD NO:39 KAĞITHANE/ İSTANBUL
PROFILO ALISVERIS MRK. NO:137 KAT:1 ŞİŞLİ/ İSTANBUL
INONU CAD.41/A CELİKTEPE KAĞITHANE/ İSTANBUL
SANAYİ MAH.S.SELİM CAD.NO:34/A KAĞITHANE/ İSTANBUL
SINANPASA MAH.MUMCU BAKKAL SOK.NO:2 BEŞİKTAŞ/ İSTANBUL
ORTABAYIR MAH. TALATPAŞA CAD. NO:67/A KAĞITHANE/ İSTANBUL
İSMET İNÖNÜ CADDESİ NO:4-C1 KAĞITHANE/ İSTANBUL
SINANPASA MH.ORTABAHCE CD.NO:17 BEŞİKTAŞ/ İSTANBUL
SANAYİ MAH. SULTAN SELİM CAD. NO:4 KAĞITHANE/ İSTANBUL
SANAYİ MAH. SULTAN SELİM CAD. 1 C KAĞITHANE/ İSTANBUL
TALATPAŞA CD.NO:43 KAĞITHANE/ İSTANBUL
CİHANNÜMA MAH. BEŞİKTAŞ/ İSTANBUL

Yukarıdaki örnekte de boşluk beklenirken yazıma devam edilmesi, ‘.’ kullanılmaması gibi istisnai durumlardan dolayı eşleşme bulunamamıştır. Bu gibi durumlar da regex yazılırken göz önünde bulundurulmalıdır.

Tüm örneklerde görüldüğü üzere regex'in son kısmında bulunan “([a-zA-ZÂİÖÜŞÇĞİöüşç]*V ?)|([a-zA-ZÂİÖÜŞÇĞİöüşç]*\$)” kısmı sayesinde satır sonlarındaki ilçe ve il bilgisi her satır için eksiksiz bulunmuştur.