



**YILDIZ TEKNİK ÜNİVERSİTESİ
ELEKTRİK-ELEKTRONİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
BLM 4580 DOĞAL DİL İŞLEMeye GİRİŞ**

RegEx ile Adres Ayırıştırma

ÖDEV 1

Muhammet Ali Şen – 20011701

01.11.2022

Giriş

Problem

Text biçiminde verilen veri içerisindeki adresleri ayırtmak amacına uygun bir Regular Expression (düzenli ifade) oluşturmak. Bu adreslerin mahalle, cadde, bulvar, sokak, blok, apartman, no, daire, posta kodu, ilçe (semt), şehir bilgilerini ayırtmaktan ibarettir.

Veri

Elimizdeki veri text (metin) formatındadır ve 600 adet adres içermektedir. Bu satırların herbiri farklı bir adresi barındırmaktadır. Bu adreslerin büyük bir kısmı, Türkiye adres yazım alışkanlıklarına uyum u olacak şekilde Mahalle, Cadde, Bular, Sokak, Blok, Apartman, Numara, Daire, Posta Kodu, İlçe, Şehir formatında verilmiştir ancak oluşturulacak RegEx ile bu format a uymayan adres bilgilerinin de başarıyla okunup ayrıştırılması hedeflenmektedir.

Yaklaşım

Ele alınan problem bir metnin regular expression ile taranıp ayrıştırılmasını gerektirdiğinden, çalışma ortamının yazılan ifadenin söz dizilimini kontrol edebilen, başarısını ölçebilen ve verilen veri üzerindeki taramasını ayrıntılarıyla bize gösterebilen bir ortam olması gerekliliği ortaya çıkmıştır. Bu gereklilik doğrultusunda www.regex101.com sitesi çalışma ortamı olarak seçilmiştir.

Türkiye'deki ev ve ofis adreslerinin yazım formatı başka ülkelere kıyasla daha karmaşık bir formattadır. Bu format doğrultusunda adresi verilecek mekânın konum bilgisinin bütün detayları adrese dahil edilir ve genellikle bu bilgiler büyük yerleşim yerinden daralarak yazılır. Bu ödev kapsamında ele alınmış olan adreslerin geneli de bu şekilde yazılmıştır. Bazı istisnalar dışında, adreslerin genel yazılış tarzlarının şu ilerleyişe sadık kaldığı görülmektedir:

Mahalle, Cadde, Bulvar, Sokak, Blok, Apartman, Numara, Daire, Posta Kodu, İlçe, Şehir

Bu formatın yanı sıra adres içerisinde site, kompleks, iş hanı, şehirler arası yola göre tarif gibi detaylar da eklenmektedir. Bahsedilen bu detayların adreste yerleştirilebileceği her yerin ayrı ayrı taranması ve her ek bilginin neyi ifade ettiğinin çözümlenmeye çalışılması ifadeyi oldukça kalabalıklaştıracığından, adresler deki genel yazılış biçimi incelenmiş ve bu ek bilgileri en yüksek doğrulukta yakalamak ve kodu da basitlikten uzaklaştırmamak hedefiyle bir RegEx formatı belirlenmiştir. Buna göre yazılan RegEx in formatı aşağıdaki gibi planlanmıştır.

Mahalle (Varsa), Cadde (Varsa), Bulvar (Varsa), Sokak (Varsa), Blok (Varsa), Apartman (Varsa), Numara (Varsa), Daire (Varsa), İlçe (Zorunlu), Şehir (Zorunlu)

Eldeki adres bilgilerinin tamamı İstanbul içi adresler olmasına rağmen oluşturulan RegEx te bu durum göz ardı edilmiştir.

Adresler arası ele alınması gereken başka bir problem de mahalle, sokak gibi kelimelerin yazılışlarının veya kısaltma biçimlerinin yazan kişinin tercihinine bağlı olarak değişiyor olmasıdır. Yazılan RegEx ile muhtemel bütün kısaltmaların bulunması sağlanmıştır. Buna göre muhtemel olduğu varsayılan yazılışlar ve kısaltmaların bir kısmı şu şekilde ifade edilmiştir.

MH, MAH, MAHALLE, MAHALLESİ, MAHALLESİ: **Mahalle**
CD, CAD, CADDE, CADDESİ, CADDESİ: **Cadde**
SK, SOK, SOKAK, SOKAĞI, SOKAĞI: **Sokak**
N., N:, NO., NO:, NR., NR.: **Numara**

Bu ve bunun gibi adres parçalarının içerebileceği karakterler ve aralarına konulacak boşluk ve noktalama işaretleri düşünülerek her senaryoda başarılı olunması hedeflenmiştir.

Oluşturulan RegEx ve Başarı Oranı

Regular Expression

Yaklaşım bölümünde anlatılmış olan faktörler gözönünde bulundurularak aşağıdaki RegEx oluşturulmuştur;

```
REGULAR EXPRESSION v2 596 matches (449 007 steps, 38.0ms)
: / ^(\\"?(([\n]+)(MAHALLE|MAHALLES|I|I|MA?H)[\.,\s]+)?([A-z0-9ŞÜÂĞİÖşçğİöÇÜ"\. - ])* / gm
BULVARI)?(([\n]+)(CADDES|I|I|CADDE|CA?D)[\.,\s]+)?(([\n]+)(\s|\.|.)+
(SOKA|GĞ|I|SOKAK|SO?K)[\.,\s]+)?([\w]sĞÜŞİÖÇÂ)\. - ", +[\.,\s]+)?(P<N00>(NO|NR|N)?
(\s*)(\.:)?(\s*)?([\w]sĞÜŞİÖÇ\.\s"-:]{1,7})[\.,\s]+)?([A-zÜÂçğİö-9-öçĞİÖşşü]*
APT\.? )?([\n]+[\.,\s]+)?((?P<Semt>[\n\s]+\/)(\s*)(?P<Sehir>[\d\s]+))$
```

Başarı Durumu

Oluşturulan RegEx verilen adres bilgilerinin neredeyse tamamını ayrı ayrı olacak şekilde algılamayı başarmıştır. RegEx ile veri üzerinde yapılan tarama sonucunda, il ve ilçe bilgileri dahil olmak 596 adet adres tespit edilmiştir. Oransal olarak (596/600) 99,33 olarak hesaplanabilir.

Kullanılan ortamın şartları sebebiyle, yapılan tarama sonucu bulunan mahalle, cadde, bulvar, sokak, blok, apartman, no, daire, posta kodu, ilçe, şehir ve ek bilgiler ayrı ayrı olacak şekilde bir veri yapısında toplanamamıştır.

	MAH.	CAD.	BLV.	SOK.	BLOK.	APT.	NO.	DAİRE.	PK.	SEMT	ŞEHİR
Adet	520	432	23	188	4	6	495	331	5	598	598
Bulunan	506	421	20	167	3	4	477	320	3	596	596
Oran	97,30	97,45	86,95	88,82	75	66	96,36	96,67	60	99,66	99,66

Doğru Bulunamayan veya Eksik Bulunan Adreslerden Bazıları;

KAPTAN PAŞA MAH. YAY GEÇİDİ SOK. NO:6/C BEYOĞLU

Fenerbahçe, Iğrıp Sk. No:13, 34726 İstanbul

YENİ VAPUR İSKELESİ KARŞISI KADIKÖY/ İSTANBUL

ESKİ EDİRNE ASFALTI NO:168 GAZİOSMANPAŞA/ İSTANBUL

Ziya Gökalp, Mahallesi, Eski Turgut Özal Cd. No:29 D:Kat.2, 34000 İkitelli
Osb/Başakşehir/İstanbul

GAYRETTEPE MAH. YILDIZ POSTA CAD. SİNAN PASAJI GİRİŞİ NO:38 BEŞİKTAŞ/
İSTANBUL

RUMELİ HİSAR ÜSTÜ NİSPETİYE CAD. NO: 41/B SARIYER/ İSTANBUL

Ekran Görüntüleri

The screenshot shows a regular expression search interface. The top bar indicates 'REGULAR EXPRESSION' and 'v2' with a dropdown arrow. On the right, it shows '596 matches (449 007 steps, 38.0ms)'. The main area contains a complex regular expression: `^(\"?(([\n]+)(MAHALLE|MAHALLES|Iİ|MA?H)[\.,\s]+)?([A-z0-9ŞÜĞİÖŞÇĞİöçü\".\-])* / gm`. Below the expression, there is a 'TEST STRING' section containing a list of addresses, each followed by a small icon. The addresses are: GÜNEŞ-TEPE-MAH.-NECİP-FAZIL-KISAKÜREK-CAD.-NO:155-GÜNGÖREN-/İSTANBUL, EYÜP-SULTAN-MAH.-KZILAY-CAD.-NO:2-SANCAKTEPE-/İSTANBUL, RUMELİ-HİSAR-ÜSTÜ-NİSPETİYE-CAD.-NO:41/B-SARIYER-/İSTANBUL, DİKİLİTAŞ-MAH.-EMİRHAN-CAD.-İŞ-BANKASI-KARŞISI-BEŞİKTAŞ-/İSTANBUL, MERKEZ-EFENDİ-MAH.-G.43-SOK.-NO:4/A-ZEYTİNBURNU-/İSTANBUL, NİŞANCA-MAH.-BALCI-YOKUŞU-SOK.-NO:2/A-EYÜPSULTAN-/İSTANBUL, İCADIYE-MAH.-CUMHURİYET-CAD.-NO:173-ÜSKÜDAR-/İSTANBUL, GARAJ-SOK.-NO:9/1-ŞİŞLİ-/İSTANBUL, BATIKÖY-MAH.-ÇARMIKLI-CAD.-NO:1-6-BÜYÜKÇEKMECE-/İSTANBUL, ÇIRPICI-MAH.-TRAMVAY-SON-DURAĞI-ZEYTİNBURNU-/İSTANBUL, ZEYREK-MAH.-GÜL-BAHÇE-SOK.-01-11-FATİH-/İSTANBUL, GÜLSUYU-MAH.-KAZIMKARABEKİR-CAD.-NO:50-MALTEPE-/İSTANBUL, TEVFİK-BEY-MAH.-MEKTEP-SOK.-NO:18-20-KÜÇÜKÇEKMECE-/İSTANBUL, YENİ-VAPUR-İSKELESİ-KARŞISI-KADIKÖY-/İSTANBUL, KAYIŞDAĞI-MAH.-USLU-CAD.-NO:48/A-ATAŞEHİR-/İSTANBUL, BAĞLARBAŞI-MAH.-ATATÜRK-CAD.-NO:46-MALTEPE-/İSTANBUL, KAVAKPINAR-MAH.-ABDİ-İPEKÇİ-CAD.-NO:70/B-PENDİK-/İSTANBUL, BOSTANCI-MAH.-ÇETİN-EMEÇ-BULVARI-NO:1-10/A-KADIKÖY-/İSTANBUL, YILDIRIM-MAH.-KINALI-ADA-SOK.-NO:14/A-BAYRAMPAŞA-/İSTANBUL, BÜYÜK-SHİR-MAH.-GONCA-SOK.-NO:18/A-BEYLİKDÜZÜ-/İSTANBUL, YENİBOSNA-KULELİ-İETT-OTOBÜS-DURAĞI-D-100-KARAYOLU-BAHÇELİEVLER-/İSTANBUL, GÜLTEPE-MAH.-ALAGEYİK-SOK.-NO:11/A-KÜÇÜKÇEKMECE-/İSTANBUL.

MERKEZ MAH. ÖMERLİ CAD. NO:46/A ÇEKMEKÖY / İSTANBUL
KEMAL TÜRKLER MAH. UYSAL CAD. NO:46 SANCAKTEPE / İSTANBUL
İÇMELER MAH. ANKARA CAD. NO:20/A TUZLA / İSTANBUL
BULGURLU MAH. BULGURLU CAD. NO:104/A ÜSKÜDAR / İSTANBUL
MEHMET AKİF MAH. ULUBATLI HASAN CAD. NO:17/B ÇEKMEKÖY / İSTANBUL
MERKEZ MAH. KEMERBURGAZ CAD. NO:56 KAĞITHANE / İSTANBUL
MİTHAT YILMAZ CAD. İETT OTOBÜS DURAĞI SARIYER / İSTANBUL
PİYALE PAŞA MAH. HALICIOĞLU CAD. NO:1/A BEYOĞLU / İSTANBUL
YUNUS EMRE MAH. ADEM YAVUZ CAD. NO:4 SULTANGAZİ / İSTANBUL
BALAT MAH. DARÜŞŞAFAKA CAD. NO:54/1 FATİH / İSTANBUL
CUMHURİYET MAH. YÜZYIL CAD. NO:41/D KARTAL / İSTANBUL
MEHMET AKİF MAH. ŞAHİNBEY CAD. NO:7/A ÇEKMEKÖY / İSTANBUL
BULGURLU MAH. BULGURLU CAD. NO:6/A ÜSKÜDAR / İSTANBUL
KEMAL TÜRKLER MAH. ATATÜRK CAD. NO:150 SANCAKTEPE / İSTANBUL
GÜMÜŞSUYU MAH. ÇEŞMİ BÜLBÜL SOK. NO:34/1A BEYKOZ / İSTANBUL
GAYRETTEPE MAH. YILDIZ POSTA CAD. SİNAN PASAJI GİRİŞİ NO:38 BEŞİKTAŞ / İSTANBUL
YILDIZ TABYA CAD. NO:111 GAZİOSMANPAŞA / İSTANBUL
YENİ MAHALLE 10 TEMMUZ CAD. NO:15/C BAKIRKÖY / İSTANBUL
SAADET DERE MAH. FEVZİ ÇAKMAK CAD. NO:33/B ESENYURT / İSTANBUL
CİBALİ MAH. ÜSKÜPLÜ CAD. NO:12/A FATİH / İSTANBUL
ÜNİVERSİTE MAH. ZAMBAK SOK. NO:7/B AVCILAR / İSTANBUL
75.YIL MAH. FEVZİ ÇAKMAK CAD. NO:41-2 SULTANGAZİ / İSTANBUL

ŞEHLİ MAH. YUNUS EMRE CAD. NO:32/A PENDİK PENDİK / İSTANBUL
FENERYOLU FAHRETTİN KERİM GÖKAY CAD. ARALIK SOK. NO:1/1 KADIKÖY / İSTANBUL
ELMALI KENT MAH. ADEM YAVUZ CAD. NO:79/B ÜMRANİYE / İSTANBUL
MERKEZ MAH. ÇAVUŞBAŞI CAD. NO:42/C ÇEKMEKÖY / İSTANBUL
İNÖNÜ MH. KAR. SK. KÖŞEBAŞI ATAŞEHİR / İSTANBUL
MECİDİYE MH. FATİH BULVARI NO:470 SULTANBEYLİ / İSTANBUL
YENİSAHRA MAH. KARAMAN CAD. NO:3/1A ATAŞEHİR / İSTANBUL
MİMAR SİNAN MAH. SELAMİ ALİ EFENDİ CAD. NO:9/1 ÜSKÜDAR / İSTANBUL
KAYIŞDAĞ CAD. NO:221/A KADIKÖY / İSTANBUL
ÇAKMAK MAH. TAVUKÇUYOLU CAD. NO:30/3 ÜMRANİYE / İSTANBUL
ÇINARDERE MAH. AYAZMA CAD. NO:4 PENDİK / İSTANBUL
AKPINAR KÖYÜ NO:3/A EYÜPSULTAN / İSTANBUL
ZEYTİNLİK MAH. PTT ÖNÜ GAZETE BAYİİ BAKIRKÖY / İSTANBUL
MERKEZ MAH. ŞEHİT MİTHAT YILMAZ CAD. KEKİK SOK. NO:1 SARIYER / İSTANBUL
ABDİ İPEKÇİ MH. İSTASYON CD. 1549. SK. NO:18/C DARICA / İSTANBUL
AYDINTEPE MAH. İBRAHİM KORAY SOK. NO:3/A TUZLA / İSTANBUL
DENİZ CD. KÜÇÜKSO. SK. NO:15 PENDİK PENDİK / İSTANBUL
CAFER AĞA MAH. ASIM GÜNDÜZ CAD. NO:84 KADIKÖY / İSTANBUL
HÜRRİYET MAH. ESKİ EDİRNE ASFALTI 184 GAZİOSMANPAŞA / İSTANBUL
ZÜHTÜPAŞA MAH. BAĞDAT CAD. NO:39 KADIKÖY / İSTANBUL
RASİMPAŞA MAH. HALİTAĞA CAD. NO:39/2/1A KADIKÖY / İSTANBUL
FINDIKLI MAH. ÇINAR CAD. NO:19 MALTEPE / İSTANBUL

PETROL İŞ MAH. RAHMANLAR CAD. NO: 25 KARTAL / İSTANBUL
EVLİYA ÇELABİ MAH. ŞİNASİ DURAL CAD. NO:105 TUZLA / İSTANBUL
AMBARLI MAH. FEVZİ ÇAKMAK CAD. NO:48/3 AVCILAR / İSTANBUL
BATTALGAZİ MAHALLESİ SELÇUKHAN CADDESİ NO:92 SULTANBEYLİ / İSTANBUL
MEVLANA MAH. SANCAKLAR CAD. NO:112B SANCAKTEPE / İSTANBUL
Y.DUDULLU MAH. NATOYOLU CAD. NO: 136 ÜMRANİYE / İSTANBUL
SELAMİ ALİ MAH. ŞEHİT ASTEĞMEN HALİL ÖGEL SOK. NO: 4 ÜSKÜDAR / İSTANBUL
BATTALGAZİ MAH. KARADENİZ CAD. KILIÇHAN. SOK. NO:3 SULTANBEYLİ / İSTANBUL
Mehmet Nesih Özmen, Yüksel Sk. NO: 19/B, 34173 CARGO/İstanbul
CEVİZLİ MAH. S. YOLU, KENNEDY CAD. ÜSTGEÇİT ALTI BAKIRKÖY / İSTANBUL
AKSARAY MAH. ÇAKIR AĞA CAMİ SOK. FATİH / İSTANBUL
HASANPAŞA MAH. FATİH BUL. NO:16 SULTANBEYLİ / İSTANBUL
GÖKTÜRK MERKEZ MAH. KIVRIM SOK. NO: 1/1 A-A EYÜPSULTAN / İSTANBUL
SAHRAVICEDİT MAH. GUZİDE SOK. NO:1/A /1 KADIKÖY / İSTANBUL
Fenerbahçe, Iğrıp Sk. No:13, 34726 İstanbul
HÜRRİYET MAH. MİTHATPAŞA CAD. NO: 33/A KARTAL / İSTANBUL
TANTAVİ MAHALLESİ SUIŞ CADDESİ NO:3-A ÜMRANİYE / İSTANBUL
İÇERENKÖY MAH. RAKKAS SOK. NO:12 ATAŞEHİR / İSTANBUL
AŞIK VEYSEL MAH. 3001 CAD. NO:88/A ATAŞEHİR / İSTANBUL
Yenimahalle Mahallesi, Atatürk Blv. 39/1, 07980 İlkadım/Antalya
KAPTAN PAŞA MAH. YAY GEÇİDİ SOK. NO:6/C BEYOĞLU

HÜSEYİNLİ MAH. BEYKOZ CAD. NO: 181/A ÇEKMEKÖY / İSTANBUL
MİMARŞİNAN MAH. MİMARŞİNAN CAD. NO: 100B/1 SULTANBEYLİ / İSTANBUL
ÇİFTLİK MAHALLESİ ÇAVUŞBAŞI CUMHURİYET CAD. NO: 223/B BEYKOZ / İSTANBUL
BEYLERBEYİ MAH. YALIBOYU CAD. NO 78/B ÜSKÜDAR / İSTANBUL
YENİŞEHİR MAH. CUMHURİYET BULVARI TELEKOM YANI NO: 95 PENDİK / İSTANBUL
YENİŞEHİR MAH. MİLLET CAD. MUSTAFA AKYOL BELDE İŞ HANI NO:25 PENDİK / İSTANBUL
DUA TEPE MAH. BARUTHANE CAD. NO: 78/B ŞİŞLİ / İSTANBUL
MALKOÇOĞLU MAH. ESKİ EDİRNE ASFALTI NO: 793/A SULTANGAZİ / İSTANBUL
KARADENİZ MAH. MEHMET AKİF CAD. NO:74/A GAZİOSMANPAŞA / İSTANBUL
ESENTEPE MAH. SİLAHTAR AĞA CAD. NO: 263/A/1 EYÜPSULTAN / İSTANBUL
GENÇOSMAN MAH. MERVE SOK. NO:1/1 GÜNGÖREN / İSTANBUL
MEHMET AKİF MAH. RECEP AYAN CAD. NO:1 ÇEKMEKÖY / İSTANBUL
CUMHURİYET MAH. MEŞECİK SOK. 31/A ÜSKÜDAR / İSTANBUL
İNÖNÜ MAH. ULUSU CAD. B BLOK 43/B 12 ATAŞEHİR / İSTANBUL
İHLAMUR KUYU MAH. ALEMDAĞ CAD. NO:231 ÜMRANİYE / İSTANBUL
BULGURLU MAH. ATALAR SOK. NO:1/A ÜSKÜDAR / İSTANBUL
MEVLANA MAH. VEYSEL KARANİ CAD. NO:207/A SANCAKTEPE / İSTANBUL
YENİDOĞAN MAH. ORDU CAD. NO:136/A1 GAZİOSMANPAŞA / İSTANBUL
KOCASİNAN MERKEZ MAH. MAHMUTBEY CAD. NO:278/A BAHÇELİEVLER / İSTANBUL
İNCİRTEPE MAH. DOĞAN ARASLI BULVARI NO:70 ESEN YURT / İSTANBUL
KAZIMKARABEKİR MAH. DR. GÜRKAN CEYLAN CAD. NO:58/B GAZİOSMANPAŞA / İSTANBUL
KOCAMUSTAFA PAŞA MAH. ORG. ABDURRAHMAN NAZİF GÜRMAN CAD. NO:62 FATİH / İSTANBUL

Yararlanılan Site
www.regex101.com