Öncelikle bütün ifadeleri gruplayıp ortak özelliklerini aradığımda bir çoğunun /İstanbul ile bittiğini gördüm.

Bu ifadeleri **(.\*[/]\s\*[İ](stanbul|STANBUL))** regexi altında topladım. Kalan uç case olan **(T.\*ÜMRANİYE)** ifadesini de ayrı bir case olarak yazdım. Başlangıç ifadesi aşağıdaki gibidir. **(.\*[/]\s\*[İ](stanbul|STANBUL))|(T.\*ÜMRANİYE)**

Daha sonra sonuçları incelerken birinci çözümümdeki regex'in yeterli olmadığını gördüm.

**"Ahmet Bey ve ekibi tamirat ve tadilat için Hürriyet Mah. Abdi İpekçi Cad. No:38/C KARTAL/İSTANBUL"** cümlesinde ifadenin tamamı adres tanımına giriyordu. Bu yüzden özelleştirme yapmaya karar verdim. Bu ve diğer cümleleri inceleyince Adreslerin büyük harfle başlaması gerektiği kararını verdim. İfadenin başındaki **.\*** regex sorunlu olduğu için değiştirmek için incelemeye başladım.

Büyük harfle başlayan kelimeleri(tek harfliler de dahil) alan regex’i **[A-Z0-9]\w\*** şeklinde yazdım. Bunların kendini tekrar etmesi için sonunda boşluk veya tab karakterleri de olabilir diye düşünerek regex’i genişlettim.

**([A-Z0-9]\w[\t ]\*)+**

Sonrasında adreslerin içinde **.**, **:**, **-**, **/** karakterleri de olabileceği için regex’in ikinci kısmı olan \w kısmını **[\w\.:\-/]** şeklinde genişlettim. Sorgu bu haliyle aşağıdaki haline ulaştı.

**([A-Z0-9][\w\.:\-/]\*[\t ]\*)+([/]\s\*[İ](stanbul|STANBUL))|(T.\*ÜMRANİYE)**

Sonrasında örneklerde inceleme yaparken aslında kelime bazında alma işlemi yaptığımız için **önü** kelimesinin ve “HIRKA-I SERİF MAH MAH. LÜLECİYEKTA CAD. LÜLECİ YEKTA SOK. NO **:** 1 DAİRE **:** 1 KAT **:** 1 FATİH/ İSTANBUL” örneğinki iki nokta işaretlerinin ayrı birer cümle olarak görüldüğünü fark ettim. Bundan dolayı sorgunun büyük harfle başlamalı dediğimiz kısmına(**[A-Z0-9]**) önü ve iki nokta ifadelerini de ekleyerek **([A-Z0-9]|önü|:)** haline getirdim. Sorgu son haliyle aşağıdaki hali aldı.

**(([A-Z0-9]|önü|:)[\w\.:\-/]\*[\t ]\*)+([/]\s\*[İ](stanbul|STANBUL))|(T.\*ÜMRANİYE)**

Bu adımdan sonra aslında ilk kelimesinde ö ve : ifadelerinin olmaması gerektiği düşündüm. Bundan dolayı bir defalığa mahsus **([A-Z0-9][\w\.:\-/]\*[\t ]\*)** ifadesinin ilk başta kullanılması gerektiğini düşünerek son ifadeyi oluşturdum.

**([A-Z0-9][\w\.:\-/]\*[\t ]\*)(([A-Z0-9]|önü|:)[\w\.:\-/]\*[\t ]\*)+([/]\s\*[İ](stanbul|STANBUL))|(T.\*ÜMRANİYE)**

Çoğu programlama dilinde ve programda regex kontrolünde türkçe karakterler sorun verecektir. BUnlar için de bir düzeltme olarak

* [A-Z0-9] = [A-ZÇĞİÖŞÜ0-9]
* [\w] = [a-zA-ZÇĞİÖŞÜçğıöşü0-9]

Olacak şekilde güncelledim. İfade son haliyle aşağıdaki gibidir.

Bu renktekiler baştaki birincil kontrolü mavi, ana sorgu kısmını pembe, sorgu sonundaki kısmı gri ve uç case’i(Ümraniye ile biten) sarı ile boyanmış şekilde ayarladım.

**Sorgunun Son Hali:**

([A-ZÇĞİÖŞÜ0-9][a-zA-ZÇĞİÖŞÜçğıöşü0-9\.:\-/]\*[\t ]\*)(([A-ZÇĞİÖŞÜ0-9]|önü|:)[a-zA-ZÇĞİÖŞÜçğıöşü0-9\.:\-/]\*[\t ]\*)+([/]\s\*[İ](stanbul|STANBUL))|(T.\*ÜMRANİYE)

<https://regex101.com/> üzerinde ecma script(js) ile çalıştırılan sorgunun sonucu aşağıdaki gibidir. Ayrıca Python, golang ve java 8 ile de çalıştırarak aynı sonuçları aldım.

