**Doğal Dil İşlemeye Kavramsal Bakış**

Ödev 1

Teslim Tarihi : 7 Mart 2016 Ders Saati

Teslim Şekli : Rapor halinde elden (hard copy)

1. Zipf Kanunu araştırılacak ve kısa bir bilgi verilecek.
2. Toplamda (1 A4 sayfada 500 kelime olduğu düşünülecek olursa 50 sayfada toplam 25000 kelime) 25 bin kelimeden oluşan hem Türkçe hem de İngilizce doküman oluşturulacak.

Hem Türkçe için hem İngilizce için ayrı ayrı karakter sıklıkları çıkarılacak ve her iki dilin karakter kullanım sıklığının Zipf kanuna uyup uymadığı gösterilecektir (Grafik olarak göstermeniz tercih edilir).

1. Sadece Türkçe için her kelimenin Zemberek veya başka bir araç yardımıyla kök ve ekleri ayrılacak olup, sadece eklerin kullanım frekansları çıkarılarak yine Zipf kanununa uyup uymadığı gösterilecektir.

Teslim edilecek raporda, Zipf kanunu hakkında bilgi, karakter ve eklerin dillerdeki dağılımlarının Zipf Kanununa uyup uymadığının grafik olarak gösterim ile yorumlanması.

Ödev 2

Teslim Tarihi : 14 Mart 2016 Ders Saati

Teslim Şekli : Rapor halinde elden (hard copy)

1. Farklı kaynaklardan farklı formatlarda toplanmış olan Türkçe yazılmış adresleri bir doküman içerisinde toplayıp (En az 20 farklı adres olmalıdır) yazacağınız RegEx ile sırasıyla mahalle, cadde, sokak, apartman, no, daire no, semt, il ve ülke bilgisini çıkaran bir uygulama yazınız.
2. Her adres bilgisi içerisinde yukarıdaki tüm alanlar bulunmayabilir. Bulunmayan bilgiler boş bırakılacaktır.
3. Grup üzerinden Alper arkadaşınız bir kargo firmasının veri tabanında yer alan adres bilgilerinden belli sayıda sizinle paylaşacak.
4. Yazdığınız RegEx kodunu çalıştırdığınızda aldığınız sonucun başarı olarak değerlendirmesi de yapılacak.

Teslim edilecek raporda, öncelikli olarak sadece yazdığınız RegEx ifadesini paylaşıp, arkasından mevcut adresi ve regex çalıştırıldıktan sonra mahalle, cadde, sokak, apartman, no, daire no, semt, il ve ülke bilgilerinden hangilerinin bulunduğu bilgisi verilecek.

Raporun sonunda her alan için başarı değerlendirmesi yapılacaktır.