**Türkçe :**

Argüman olarak girilen directory'de bulunan yazıların 1-gram, 2-gram, 3-gram, 4-gram ve 5-gram' ları hesaplanıp .txt uzantılı dosyalara yazılır. Bu işlem çok büyük directory' lerde özellikle four ve five gram hesaplarken uzun sürecektir. O yüzden programın main fonksiyonunda dosyaları oluşturma satırları yoruma alınmıştır. Ben bunları oluşturmuş olduğum için Chain Rule ile Markov Assumption kullanarak olasılık hesapladım. Olasılığını hesaplayacağınız string' in boyutu girdiğiniz gram değerinden küçük ise program -1.0 döndürecektir. Programda %5' lik kısım için ayrı bir hesaplama yapmam gerekiyordu o yüzden isRemain isimli parametre kullandım ve %100' lük kısım için hesaplama yapmak istenirse bu parametre 0 olarak gönderilir. Benim kullandığım directory şu adresteki : "http://www.kemik.yildiz.edu.tr/data/File/1150haber.rar" raws\_texts' dir. Ayrıca oluşturduğum N-gram' ları içeren dosyaları şu linkte paylaşıyorum : "http://s6.dosya.tc/server8/twtyxg/n-grams.rar.html"

**English :**

1-gram, 2-gram, 3-gram, 4-gram and 5-grams of the letters in the directory entered as an argument are calculated and files with .txt extension are written. This will take a long time, especially in very large directories, when calculating four and five grams. That's why the lines in the main function of the program have been interpreted. I calculated the probability using the Markov Assumption with the Chain Rule because I created them. If the size of the string is less than the gram value you entered, the program will return -1.0. I had to do a separate calculation for the 5% part in the program, so I used the isRemain named parameter, and if I want to do calculation for 100% part, this parameter is sent as 0. My directory is in the following directory: "http://www.kemik.yildiz.edu.tr/data/File/1150haber.rar" is raws\_texts. I also share files containing N-grams that I created on the link "http://s6.dosya.tc/server8/twtyxg/n-grams.rar.html"