

# 2020-2021 Dönemi Algoritma Ve Programlama Dersi Proje Raporu :

## Projenin Konusu:

Kullanıcı tarafından girilen bir metnin İngilizce dilinde mi yoksa Almanca dilinde mi yazıldığıнын otomatik olarak belirlenmesi.

## Proje Teslim Tarihi:

30.12.2020

## Hazırlayanlar:

Adı , Soyadı: Hazar Karademir

Öğrenci No: 20360859084

Adı , Soyadı: Mehmet Can Uslu

Öğrenci No: 20360859084

## Proje Görevi:

Kullanıcıdan bir metin alınacaktır. Kullanıcıdan alınan metin üzerinde önceden belirlenmiş olan bigram ve trigramların frekansları hesaplanacaktır (frekans hesaplaması: her bir bigramın ve her bir trigramın bu metinde yüzde kaç kez geçtiğinin bulunması). Bu hesaplamalar sonucunda elde edilen frekans matrisi ile taslak kod içerisinde verilen frekans matrisi (bu matriste İngilizce ve Almanca dillerinin frekans değerleri yer almaktadır) karşılaştırılacaktır. İngilizce ve Almanca dillerinin her birinin frekansları ile elde edilen frekans matrisi arasında ayrı ayrı uzaklık hesabı yapılacaktır . Hangi dil ile daha küçük uzaklık değeri elde edildiyse, metnin yazıldığı dil olarak o dil tahmin edilecektir.

## Uygulanan Adımlar:

**Adım 1:** Verilmiş C dosyasını incelendi. Bu dosyada verilen bigramlar , trigramlar ve bu bigram ve trigramların İngilizce ve Almanca dillerindeki frekans değerleri incelendi. C dosyasındaki fonksiyonların fonksiyon bildirimleri incelendi.

**Adım 2:** Taslak kod içerisinde prototipleri verilmiş olan fonksiyonları gerekli işlemleri yerine getirecek şekilde yazıldı. İhtiyaç olarak taslak koda ek fonksiyonlar tanımlandı.

**a) Kullanıcıdan alınan metnin temizlenmesi:** Kullanıcıdan alınan metinde yer alan yabancı karakterler temizlendi.

Bu işlem için void filter\_str(char str[]); fonksiyonu kullanıldı.

Bu fonksiyon parametre olarak aldığı char dizisindeki tüm karakterleri küçük harfle değiştirdi.

Bu fonksiyon parametre olarak aldığı char dizisindeki tüm karakterleri tek tek kontrol etti. a-z karakterleri (küçük harf) veya boşluk karakteri(ASCII: 32) dışındaki diğer tüm karakterleri boşluk karakteriyle değiştirildi.

**b) Taslak kodda verilmiş olan tüm bigramların, kullanıcıdan alınan metin üzerindeki frekanslarının hesaplanması:** Verilen tüm bigramların, kullanıcıdan alınan metin üzerindeki frekansları hesaplanarak bir frekans matrisi oluşturuldu. Bu bigram frekans matrisi ekrana yazdırıldı.

Bu işlem için void calculate\_frequencies\_bi(char str[]); fonksiyonu kullanıldı.

Global "matrix\_bigram\_strings" dizisindeki bigramların her biri, kullanıcıdan alınan "str" karakter dizisinde sırasıyla arandı ve frekansları global "calculated\_frequencies" dizisine sırasıyla yerleştirildi.

**c) Taslak kodda verilmiş olan tüm trigramların, kullanıcıdan alınan metin üzerindeki frekanslarının hesaplanması:** Verilen tüm trigramların, kullanıcıdan alınan metin üzerindeki frekansları hesaplanarak bir frekans matrisi oluşturuldu. Bu trigram frekans matrisi ekrana yazdırıldı.

Bu işlem için void calculate\_frequencies\_tri(char str[]); fonksiyonu kullanıldı.

Global “matrix\_trigram\_strings” dizisindeki trigramların her biri, kullanıcıdan alınan “str” karakter dizisinde sırasıyla arandı ve frekansları global “calculated\_frequencies” dizisine sırasıyla yerleştirildi.

**d) Taslak kodda verilmiş olan frekans matrisleri (frequencies\_eng ve frequencies\_germ) ile elde edilen frekans matrisi (calculated\_frequencies) arasında uzaklık hesabının yapılması:**

Bu işlem için void calculate\_distances(); fonksiyonu kullanıldı.

Bu fonksiyon, taslak kodda verilmiş olan frekans matrisleri ile, kullanıcı tarafından girilen metinden elde ettiğimiz frekans matrisi olan calculated\_frequencies arasındaki uzaklığını hesaplamaktadır. İlk olarak frequencies\_eng dizisi ile, yani İngilizce dili için verilmiş olan frekans matrisi ile calculated\_frequencies dizisi arasında uzaklık hesabı yapıldı. İkinci olarak, frequencies\_germ dizisi ile, yani Almanca dili için verilmiş olan frekans matrisi ile calculated\_frequencies dizisi arasında uzaklık hesabı yapıldı. Elde edilen değerler sırasıyla global “distances” dizisinde ilgili indise atandı.

**e) Metnin yazıldığı dilin tahmin edilmesi:**

Bu işlem için void detect\_lang(); fonksiyonu kullanıldı.

Global scope’ta bulunan “distances” dizisindeki değerler dikkate alınarak metnin yazıldığı dil tahmin edildi . Tahmin edilen dil ekrana yazdırıldı.