

BURSA TEKNİK ÜNİVERSİTESİ

BİLGİSAYAR MÜHENDİSLİĞİ

BLM0111 ALGORİTMALAR VE PROGRAMLAMA DERSİ

PROJE ÖDEVİ

- Ödevin süresi 3 haftadır.
- Projeler 2'şerli gruplar ile yapılmalıdır.
- Yazılan kodlarda yorum satırları olmalı ve yapılan işlemler açıklanmalıdır.
- Projenin bitiminde en az 3 sayfalık bir proje raporu hazırlanmalıdır. Proje raporunun ilk sayfasında öğrencilerin isim, soyisim ve öğrenci numarası bilgileri yer almalıdır.
- Klasörün adı sadece her iki öğrencinin numarasının arada tire karakteri (-) olacak şekilde birleşiminden oluşmalıdır.
- Her öğrencinin proje dosyalarını kendi Github hesabından paylaşması gerekmektedir.
- Ödev tesliminde proje raporunuzun, C kodlarınızın bulunduğu Codeblocks proje klasörünüzün ve "github.txt" içerisinde bireysel Github reponuzun bağlantısının yer aldığı dosyalar bir klasörde toplanarak zip haline getirilmeli ve o şekilde ekampus.btu.edu.tr adresinden teslim edilmelidir.
- Zamanında teslim edilmeyen ödevler için 1 hafta geç teslim süresi bulunmaktadır ancak geç teslim edilen ödevlerden 30 puan kırılacaktır.
- Kopya çekildiğinin tespit edilmesi halinde o ödevlere puan verilmeyecektir.

Ödev ile ilgili sorularınız için Arş. Gör. Esmâ İBİŞ veya Arş. Gör. Muhammed Ömer Faruk SELVİ'ye ulaşabilirsiniz.

Başarılar dileriz...

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

Amaç: Fonksiyon yapılarının öğrenilmesi ve öğrenilen bilgilerin uygulanabilmesi, döngülerin öğrenilmesi ve öğrenilen bilgilerin uygulanabilmesi, char dizileri üzerinde çeşitli işlemlerin gerçekleştiriminin öğrenilmesi ve öğrenilen bilgilerin uygulanabilmesi.

Açıklama: N-gram: n adet elemandan oluşan ardışık dizidir. N=2 olduğu durumda **bigram** olarak isimlendirilmekte, N=3 olduğu durumda **trigram** olarak isimlendirilmektedir. Doğal dil işleme, hesaplamalı dilbilim gibi alanlarda sıklıkla kullanılmaktadır.

Görev: 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 ettiğiniz frekans matrisi ile size 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 sizin elde ettiğiniz frekans matrisi arasında ayrı ayrı uzaklık hesabı yapılacaktır (öklid uzaklığını kullanabilirsiniz). Hangi dil ile daha küçük uzaklık değeri elde edildiyse, metnin yazıldığı dil olarak o dil tahmin edilecektir.

Yapılacak İşlemler:

Adım 1: Size verilen C dosyasını inceleyiniz. Bu dosyada bigramlar ve trigramlar ayrı ayrı diziler olarak size verilmiştir. Ayrıca bu bigram ve trigramların İngilizce ve Almanca dillerindeki frekans değerleri yine bir dizi olarak size verilmiştir. Kullanmanız gereken fonksiyonların fonksiyon bildirimleri yazılmıştır.

Adım 2: Taslak kod içerisinde prototipleri verilmiş olan fonksiyonları gerekli işlemleri yerine getirecek şekilde yazınız. Fonksiyonların imzasını değiştirmemelisiniz. İhtiyaç duymanız halinde bunlara ek olarak başka fonksiyonlar da tanımlayıp kullanabilirsiniz

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

Bu işlem için `void filter_str(char str[]);` fonksiyonu kullanılmalıdır.

Bu fonksiyon parametre olarak aldığı char dizisindeki tüm karakterleri tek tek kontrol etmelidir. A-Z, karakterleri (büyük harf), a-z karakterleri (küçük harf) veya boşluk karakteri(ASCII: 32) dışındaki diğer tüm karakterleri boşluk karakteriyle değiştirmelidir.

Örnek girdi:

DRESSAIT SA LOQUACES ET POSSEDER IL QU. ECHOS NUITS VIN OEUFs NET ART
SABRE QUE. BOUTS VOTRE DAIMS SI JE COUPS MONTE RESTE. LABOUR AU MA
ECLATE ROCHES ME PIECES DURANT. NEZ ROUGEATRES PRIMEVERES
LUMINEUSES GOURMETTES JEU MOUSTACHUS BOULEVARDS. PHRASES
EXTREME THEATRE PEU SUR. FIN VOICI ADORE SENTI COURs UNE SON.
SACHANT ENSUITE EST PRAIRIE TOI CAVITES DRAPENT.

Örnek Çıktı:

DRESSAIT SA LOQUACES ET POSSEDER IL QU ECHOS NUIITS VIN OEUFS NET ART
SABRE QUE BOUTS VOTRE DAIMS SI JE COUPS MONTE RESTE LABOUR AU MA
ECLATE ROCHES ME PIECES DURANT NEZ ROUGEATRES PRIMEVERES
LUMINEUSES GOURMETTES JEU MOUSTACHUS BOULEVARDS PHRASES
EXTREME THEATRE PEU SUR FIN VOICI ADORE SENTI COURS UNE SON
SACHANT ENSUITE EST PRAIRIE TOI CAVITES DRAPENT

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 oluşturulmalıdır. Bu bigram frekans matrisi ekrana yazdırılmalıdır.

Bu işlem için `void calculate_frequencies_bi(char str[]);` fonksiyonu kullanılmalıdır.

Global “matrix_bigram_strings” dizisindeki bigramların her biri, kullanıcıdan alınan “str” karakter dizisinde sırasıyla aranmalı ve frekansları global “calculated_frequencies” dizisine sırasıyla yerleştirilmelidir.

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 oluşturulmalıdır. Bu trigram frekans matrisi ekrana yazdırılmalıdır.

Bu işlem için `void calculate_frequencies_tri(char str[]);` fonksiyonu kullanılmalıdır.

Global “matrix_trigram_strings” dizisindeki trigramların her biri, kullanıcıdan alınan “str” karakter dizisinde sırasıyla aranmalı ve frekansları global “calculated_frequencies” dizisine sırasıyla yerleştirilmelidir.

d) Taslak kodda verilmiş olan frekans matrisleri (frequencies_eng ve frequencies_germ) ile sizin elde ettiğimiz frekans matrisi (calculated_frequencies) arasında uzaklık hesabının yapılması:

Bu işlem için `void calculate_distances();` fonksiyonu kullanılmalıdır.

Bu fonksiyon, taslak kodda verilmiş olan frekans matrisleri ile, kullanıcı tarafından girilen metinden elde ettiğiniz frekans matrisi olan calculated_frequencies arasındaki öklit uzaklığını hesaplamalıdır. İlk olarak frequencies_eng dizisi ile, yani İngilizce dili için verilmiş olan frekans matrisi ile calculated_frequencies dizisi arasında uzaklık hesabı yapılmalıdır. İkinci olarak, frequencies_germ dizisi ile, yani Almanca dili için verilmiş olan frekans matrisi ile calculated_frequencies dizisi arasında uzaklık hesabı yapılmalıdır. Elde edilen değerler sırasıyla global “distances” dizisinde ilgili indise atamalıdır.

e) Metnin yazıldığı dilin tahmin edilmesi:

Bu işlem için `void detect_lang();` fonksiyonu kullanılmalıdır.

Global scope’ta bulunan “distances” dizisindeki değerler dikkate alınarak metnin yazıldığı dil tahminlenmelidir. Tahmin edilen dil ekrana yazdırılmalıdır.