Bilecik Şeyh Edebali Üniversitesi ;

NEfyaz Enes Özmert

Dönem Projesi(35272606506)

Amaç: Web sitelerinde ki bilgileri kullanarak(Google Api ile) ve bunları Nlp , HtmlAgilityPack ,regex gibi kütüphanelerle işleyerek sorulara cevap veren kütüphane oluşturmak.

Platform: Net.Core .dll

Kütüphane nerelerde kullanılır : .net core destekleyen platformlar için uygundur.

Dil: C#

Çalışma mantığı: Elimizde ki soru cümlesini Google api ile aratıp linkler elde ediyoruz sayfa başına 10 link bu linklerin kaynak kodlarını elde edip html ve js gibi web kodlarından arındırıyoruz arındırılan sitedeki metinleri cümleler halinde gramer yapısı çözülen kelimelerin cümlenin içinde bulunup bulunmamasına göre listelere ekliyoruz. Bu listeleri daha sonra aynı web sitesine sahip olan cümleler(paragraf) şeklinde yeni bir listeye ekleyip en çok kelime kökü çıkan paragrafı sonuç olarak kabul ediyoruz.

Temel olarak yapılan işlem bir filtreleme işlemdir.

Birçok listeyi ve bilgiyi süzgeçten geçirip soru cümlesine ait olan kısmı kullanıcıya sunuyoruz.

Kullanılan Harici Kütüphaneler:

-Zemberek Nlp -HtmlAgilityPack - System.Text.RegularExpressions; -Newtonsoft.Json

Hangi kütüphane ne amaçla kullanılıyor?

Zemberek Nlp büyük ve gelişmiş bir doğal dil işleme kütüphanedir. Projede Normalizasyon , Tokenization , MorphologyAnalizer(kelime kök isim ve diğer gramerler için) kullanılmaktadır.

HtmlAgilityPack projede string olarak indirilip okunan web sitelerinde html js vs. ayıklamak için kullanılmaktadır.

System.Text.RegularExpressions projede html js vs. için regex adlı metodu ile <> , {} arasında olan yerleri temizlemek ve ayıklamak için kullanılır.

Newtonsoft.Json projede Google apiden aldığımız linkler json formatında gelir bunları listeye aktarmak için newtonsoft’un json’u kullunılır.

Projede bulunan sınıflar:

1) ConvertNew 2) GQuestionAnswer 3) GSearch 4) GWebClient 5) HtmlTagFilter 6) MorphologyAnalizer 7) Normalization 8)NMain

Projede bulunan sınıflar ne için kullanıldı:

ConvertNew=UFT8’i UFT16 ya convert için kullanıldı.

GQuestionAnswer=Soruların cevaplarını hazırlamak için kullanıldı.

GSearch=Arama motorundan web sitelinklerini elde etmek için kullanıldı.

GWebClient=web sitelerini string indirip verileri ede etmek için kullanıldı.

HtmlTagFilter=Web sitelerini ayıklmak için kullanıldı.

MorphologyAnalizer=Soru cümlelerini analiz etmek için kullanıldı.

Normalization=Soru cümlelerini normalize etmek için kullanıldı.,

NMain=Diğer Class’ları birleştirmek kodu okunabilir hale getirmek için kullanıldı.

Projede bulunan Liste sınıflar:

1)Galaxy 2)SolarSystem 3)Word

Projede tasarım nasıl oluşturuldu?

Proje filtreleme mantığı kullanır uzay yapısını örnek alır. En siyah dış bizim için uzaydır Galaxy bu alanda webde bulunan linkler yer alır yani bilgilerin tamamı (reklamlar kodlar her şey dahil) onun içinde ki üçgen kısım ise bizim için solar system yani güneş sistemidir burada linklerin sitring olarak okunması ve ayıklanmasını ifade eder en ortada ki mavi yer ise dünya Word ‘u ifade eder eşleşen cümlelerin birleştiği ve rank yani derecelendirdiği sorunun cevaplandığı kısımdır.

Projenin İşlem basamaklar:

Normalizasyon , Analiz Gramer yapısı(İsim kök bulma) ,Google api ile cümleyi webde arama link olarak listeye eklemek , elde edilen linkleri webclient httpresponse ile kaynak kodu elde etmek, kaynak kodlarını filtrelemek ,filtre edilen kaynak kodlarını cümlelere bölmek(tokenazation ) bu cümlelerede kökleri ve isimleri aramak ve listelemek , listeleri aynı web sitelerine göre birleştirip rank verip yeniden listelemek.

Etkin olan arama sonuçları :

En az üç kelime olmak şartı ile doğru sonuçlar bulunabiliyor.

Kuantum fiziği nedir? , Renk körlüğü nedir ? , Silikon vadisi nerdedir? … gibi soruların cevaplarını 100% alabiliyoruz.

Kullanımı: Projenize bir adet metin belgesi (Html5ElementList) ekleyiniz. Projenize bu klasörü (Resources) ana dizine ekleyiniz. Projenize bu kütüphaneyi(Nefyaz.dll) ekleyiniz.

Kütüphane basit bir kod ile çalışır.

Kod:



var Reader = Console.ReadLine();//ekrandaki soru cümlesini oku

var cWorld = NMain.worlds(Reader);//NMainde bütün işlemleri yap ve liste olarak cWorld'a aktar.

foreach (var item in cWorld)//Listele

{

Console.WriteLine("id: / " + item.Id + "/\n" + "WebSite: /" + item.WebSite + "/\n" + "AnswerToTheQuestion: /" + item.AnswerToTheQuestion + "/\n" + "KeyWord: /" + item.KeyWord + "/\n" + "Rank: /" + item.Rank + "/\n" + "\n");

}

Console.Read();

İşte bu kadar isterseniz arama sonuçlarını çoğaltmak için i yerine en az 1 olmak üzere değer verebilirsiniz. NMain.worlds(Reader,i);

Kurulması gereken kütüphaneler:

Nuget:

Zemberek Nlp All

HtmlAgilityPack

NewtonSoft Json

Normal Dll:

NEfyaz