

WEB INDEXLEME UYGULAMASI

Sefa Mert GÜNGÖR

Bilgisayar Mühendisliği Bölümü

Kocaeli Üniversitesi

180201086

Miraç Onur SAKA

Bilgisayar Mühendisliği Bölümü

Kocaeli Üniversitesi

180201081

ÖZET- Bu projede temel amaç verilen bir urldeki sayfa içeriğine göre birden fazla web sayfasını benzerlik yönünden indexleyip sıralayan web tabanlı bir uygulama geliştirmektir. Bu proje sayesinde bazı python kütüphanelerini, web indexleme yöntemlerini ve Flask Framework'ü kullanarak web tabanlı uygulama yazma becerisi geliştirilmiş oldu.

Anahtar kelimeler-

Python, Flask, Web İndexleme, Veri Çekmek, HTML, CSS, Bootstrap

install packages) 3. Parti python paketlerini kurulması için kullanılan bir scripttir. PyPI paketlerini kurmamızı, güncellememizi ya da kaldırmamıza yarar. Gerekli kütüphaneler import edildikten sonra Flask Framework'ü ile projenin algoritmaları gerçekleştirilmeye başlanmıştır. Proje 3 aşamadan oluşmaktadır. Proje isterlerinden olan 2. ve 3. ister 2. Aşama, 4. ve 5. ister 3. aşama olarak uygulamanın içinde yer almaktadır. Projenin her bir isterini gerçekleştirmek için farklı algoritmalar ve formüller geliştirilmiştir.

II.HAZIRLIKLAR VE BİLGİLER

I.GİRİŞ

Öncelikle bootstrap kullanarak hazırlanacak web uygulamamızın görsel tasarımı oluşturulmuştur. Daha sonra pythonda bulunan PIP scriptini kullanarak gerekli paketler install edilmiştir. Pip(PIP

Bu proje Python programlama diliyle çalışan bir Framework olan Flask'ta geliştirilmiş olup, geliştirme ortamı olarak "Visual Studio Code" kullanılmıştır. İlk önce bizden istenen durumları değerlendirip tartışarak ne yapmamız gerektiği hakkında karara varıp projeyi hangi adımlara göre takip edeceğimize karar verdik. Daha sonrasında projeye başladık. Projeye başladıktan sonra ilk olarak projeyi gerçekleştirmek için gerekli

olan ortamların kurulumunu yaptık. Daha sonra Python dilinin web programlama açısından daha uygun olacağını düşünerek Python dilini araştırmaya ve Flask Framework'ünün kullanımını öğrenmeye başladık. Bize gerekli olabilecek algoritmalar ve kütüphaneleri belirleyerek bunların kullanımının üzerine yoğunlaştık.

III.YÖNTEM

Uygulamada ilk olarak karşımıza yönlendirme sayfası çıkmaktadır. Bu sayfada yapılacak seçime göre istenilen aşama gerçekleştirilebilmektedir. 1. aşamada verilen bir URL için içeriğinde bulunan her bir kelimenin kaçar defa yer aldığını bulan bir algoritma yer almaktadır. 2. Aşamada verilen 2 URL'in anahtar kelimelerini bulan ve hem bunlarla hem de sayfa içeriğindeki başlıklarla birlikte hesaplanan 2 sayfanın birbirine benzerlik oranını hesaplayan bir algoritma bulunur. 3. aşamada ise 1. girilen URL'deki içeriklerle diğer URL'lerin (istenildiği kadar URL girilebilir sadece bu aşamadan önce belirtilen yerde ne kadar URL girilmesinin istendiği belirtilmelidir) maksimum 3 alt katman olacak şekilde içine girip çıkardığı tüm URL'ler ile benzerlik oranını karşılaştırıp onları sıraya koyan ve daha sonra bir ağaç yapısı şeklinde katmanlarına ayırıp yazdıran bir algoritma bulunmaktadır.

IV.YALANCI KOD

- 1) Web uygulaması çalıştırıldı.
- 2) Yönlendirme sayfası karşımıza çıktı. **ek1*
- 3) 1. aşamaya tıklandı
- 4) 1. aşamada bir url girilmesi istendi.
- 5) Web sayfası içeriğinde bulunan bütün kelimelerin frekansları çıkarıldı. **ek2*
- 6) 2. aşamaya tıklandı
- 7) 2. aşamada benzerlik oranının karşılaştırılması istenen 2 url girildi.
- 8) Girilen her bir url in anahtar kelimelerini buldu ve benzerlik oranlarını çıkardı. **ek3*
- 9) 3. aşamaya tıklandı.
- 10) 3. aşamadan önce kaç tane url girileceği seçildi. **ek4*
- 11) 3. Aşamada 1 tane ana url ve karşılaştırılması istenen urler girildi. **ek5*
- 12) Maksimum 3 alt katman olacak şekilde içine girip çıkardığı tüm urler ile benzerlik oranını karşılaştırıp onları sıraya koyup daha sonra bir ağaç yapısı şeklinde katmanlarına ayırıp yazdırdı. **ek6*

V.EKRAN GÖRÜNTÜLERİ

*ek1

[Ana Sayfa](#) [Arama 1](#) [Arama 23](#) [Arama 45](#)

Web Indexleme Uygulamasına Hoşgeldiniz

Asama 1

Sayfada Geçen Kelimelerin Frekanslarını Hesaplayınız

g

Asama 2/3

Anahtar Kelime Çıkartma ve Benzerlik Çıkarma

g

Asama 4/5

Sitelere Indexleme Sıralama, Semantik Analiz

g

*ek3

Indexleme işlemi başarılı bir şekilde gerçekleşmiştir!!

Lutfen İlk Url'i Giriniz

Url

Lutfen İkinci Url'i Giriniz

Url

Gönder

Kelime	Tekrar
turizm	16
samsun	14
kültür	9
genel	6
bilgiler	5

Kelime	Tekrar
gözet	20
samsun	16
tl	11
turları	10
keşfet	8

15.65

*ek4

Lutfen Url Sayisi Girin

Git

*ek2

Indexleme işlemi başarılı bir şekilde gerçekleşmiştir!!

Lutfen Frekans Hesaplamak için Url Giriniz

Url	Gönder
Kelime	Tekrar
haberler.com	4
haberler	7
son	8
dakika	12
ve	16
güncel	3
haber	9
türkiye'nin	2
en	3
kapsamlı	2
portali	2
haberleri	3
dakika	1
magazin	1
spor	5
ekonomi	1
indirim	1

*ek5

Lutfen İlk Url'i Giriniz

Url

Lutfen Diğer Url'leri Giriniz

Url

Gönder

*ek6

- 100.0 -- https://www.haberler.com/ -- {'süresi': 109, 'dakika': 12, 'haber': 9, 'son': 8, 'haberler': 7} -- ['an', 'zaman', 'geri']
 - 56.92 -- https://www.haberler.com/ekonomi/ -- {'süresi': 24, 'ekonomi': 12, 'haberler': 7, 'haberleri': 6, 'son': 6} -- []
 - 52.6 -- https://www.haberler.com/son-dakika/ -- {'son': 56, 'dakika': 40, 'kırmızı': 27, 'içişleri': 17, 'ilk': 16} -- ['an', 'zaman']
 - 8.33 -- https://www.haberler.com/pudra-sekeri-cekim-diyen-kursat-ayvatoglu-14029855-haber/ -- {'kokain': 18, 'ifadesinde': 13, 'kürşat': 12, 'ayvatoğlu': 11, 'pudra': 10} -- []

37.47

- 37.47

VII.REFERANSLAR

- 1- <https://getbootstrap.com/>
- 2- <https://www.w3resource.com/>
- 3- <https://stackoverflow.com/>
- 4- <https://es-anlamli.gen.sx/>
- 5- <https://medium.com/>
- 6- <https://gelecegiyazanlar.turkcell.com.tr/>
- 7- <https://www.udemy.com/>
- 8- <https://www.geeksforgeeks.org/>
- 9- <https://www.youtube.com/channel/UCZNZj3mkdCGJfCoKyl4bSYQ>

VI.SONUÇ

Bu proje sayesinde web indexleme yöntemleri hakkında bilgi edinilmesini, yeni python paketlerinin öğrenilmesi, Flask Framework'ü hakkında bilgi edinilmesi gerçekleştirilmiştir. Web tabanlı uygulama yazma becerisi edinilmiştir. Bootstrap teknolojisi bileşenleri daha kapsamlı bir şekilde incelenmiştir. HTML ve CSS hakkındaki bilgiler geliştirilmiş olup web sitesi tasarlama konusunda tecrübelenilmiştir.