# KOCAELİ ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ YAZILIM LABORATUVARI

Tuğba ÇEVİKER 210202056 Samethan KAZANBAŞ 210202043

### I. ÖZET

Bu döküman Yazılım Laboratuvarının dördüncü projesi olan Web Scraping Akademi Uygulamasının çözümünü açıklamaya yönelik oluşturulmuştur. Dökümanda özet , giriş , deneysel sonuçlar , sonuç , yöntem ve kaynakça kısımlarına yer verilmiştir.

# II. GİRİŞ

Google Akademik gibi akademik arama motorları üzerinden web scraping yöntemiyle aratılan akademik yayınlara ait bilgilerin kaydedildiği bir veritabanıyla birlikte bu bilgilerin webden aratılması, görüntülenmesi ve istenilen özelliklere göre sorguların yapılmasını sağlayan bir web arayüzü geliştirilmelidir. Proje sayesinde web scraping ile bir web sayfasından bilgiye erişim sağlama; MongoDB veritabanı ile Elasticsearch sorgu yapılarını kullanma ve web kodlama becerilerinin geliştirilmesi amaçlanmaktadır.

## 1)WEB SCRAPING

En az bir akademik arama motorundan web scraping kullanılarak girilen anahtar kelimelere göre listelenmiş en az ilk 10 akademik yayının bilgileri,oluşturduğunuz web arayüzünde görüntülenmelidir.Kullanıcının arama yapmak için kullanacağı anahtar kelimeler oluşturacağınız kendi web sayfanızdaki bir text alanı üzerinden girilecektir. İstenen yayına ilişkin bilgiler doğrudan akademik arama motorunun sayfasından çekilebileceği gibi arama motoru sayfasındaki link üzerinden yönlendirilecek diğer bir web sayfasından da elde edilebilir. İstenen her yayın için pdf dosyası mutlaka indirilmelidir. Daha sonra tercihe göre yayın bilgileri ya web sayfası üzerinden ya da indirilmiş pdf dosyasının içeriğinden elde edilebilir.

#### 2)VERİTABANI

Web scraping ile elde edilen veriler MongoDb kullanılarak veritabanına kaydedilecektir. Veritabanında tutulması beklenen bilgiler: Yayın id , yayın adı , yazarların isimleri , yayın türü , yayımlanma tarihi , yayıncı adı , anahtar kelimler(aranan), anahtar kelimler (makale) , özet , referanslar , alıntı sayısı , doi numarası (varsa) , url adresi.

## 3)WEB SAYFASI

Erişilen yayınların bilgilerinin kullanıcıya gösterilmesi için bir web sayfası oluşturmanız beklenmektedir. Web sayfasında aratılacak yayınlar için bir text alanı oluşturulmalı ve bu text alanı girilecek anahtar kelimeler üzerinden ilgili arama motorunun yayınları aratıp bilgilerini web sayfasına getirmesi sağlanmalıdır. Web sayfası ilk açıldığında veri tabanında tutulan bilgiler ana sayfada gösterilmelidir.

## III. YÖNTEM

# 1)WEB SCRAPING

Öncelikle projeyi hangi dilde yapcağımıza karar verdik. Projeyi C dilinde ve .Net Core 6.0 projesi olarak yapmaya karar verdik. Yaptığımız araştırmalarda .Net Core 6.0 da Web Scraping yapabilmemiz için kullanacağımız belli başalı kütüphaneler vardı. Bunlar Selenium ve HtmlAgiliyPazk kütüphaneleriydi. Biz bu kütüphaneler arasından HtmlAgilityPack kütüphanesini kullanmaya karar verdik. Ardından web scraping uygulayacağımız siteye karar vermemiz gerekiyordu. Burda da "dergipark.com" sitesini kullanmaya karar verdik. HtmlAgilityPack kullanarak dergipark sitesinini herhangi bir uzantısına ulaşmamızı sağlayan kod parçası.

```
String url = "https://dergipask.org.tr/tr/search?gs" + search;
url = url + "Searcinomarticles";
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search?gs" + search;
var https://dergipask.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/tr/search.org.tr/t
```

WebScraping işlemiyle ilgili siteinin html kodundan istediğimiz yeri istediğimiz şekilde çekebiliyoruz. Başlangıç controllerımızda yazdığımız algoritma sayesinde bir makalenin özelliklerini Proje dökümanında da belirtildiği gibi böldük. Oluşturduğumuz Article nesnesine de bu özellikleri mongoDb veritabanında da kullanılacak bir şekilde düzenleyip atamasını yapabildik.

### 2)Veritabanı

Veritabanı olarak dökümanda da belirtilen mongoDB veritabanını kullandık. HtmlAgilityPack kütüphanesi sayesinde böldüğümüz her veriyi atadığımız article nesnelerini veritabanına ArticleService sınıfında yazdığımız Create fonksiyonu sayesinde ekleyebildik.

```
_id: ObjectId('05f7afdc5d6f62b7edb995db')
yayin_adi: "OKUL KAVRAMINA ILI$KIN METAFORLAR"
yazar_adi: "PınarARSLAN"
yayin_turu: "Makaleler"
yayinlamma_tarihi: 2020-04-05T21:00:00.000+00:00
konular: ""
anahtarkelime_makale: "Metafor, Okul, Okul Metaforları"
ozet: "Metaforlar, olayların oluşumu, işleyişi ile ilgili olarak bireyin düşü..."
referanslar: "Akkaya, N. (2009). Öğretmen adaylarının öğretmenlik mesleğine yönelik ..."
alinti_sayisi: 1
doi_numarasi: ""
url_adres: "https://dergipark.org.tr/tr/pub/jils/issue/53618/668509"
download: "https://dergipark.org.tr/tr/download/article-file/1038100"
```

Projemizin veritabanıyla bağlanabilmesi içinde .Net Core 6.0 projemize Connection String eklememiz gerekiyordu.

Bunu da projemizde yer alan app.settings dosyasına aşağıdaki gib ekledik.

```
{
    "WebScrapingDatabaseSettings": {
        "WebScrapingCollectionName": "Article",
        "ConnectionString": "mongodb://localhost:27017",
        "DatabaseName": "WebScraping"
    },
```

## 3)Web Sayfası

Proje çalıştırıldığında ekrana Arama yapabileceğiniz arama motorumuz geliyor.



Ardından kullanıcı istediği kelimeyi aratıyor ve ekrana arattığı kelimeyle alakalı en fazla 24 sonuç geliyor. Sonuç ekranı da aşağıdaki gibidir.

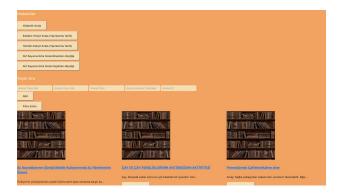


Kullanıcı ilgili makalenin adına basıp direkt dergipark üzerindeki linke ulaşabilir veya makaleye git butonuylailgili sitedeki bilgilerin gösterildiği yeni bir sayfaya geçer.

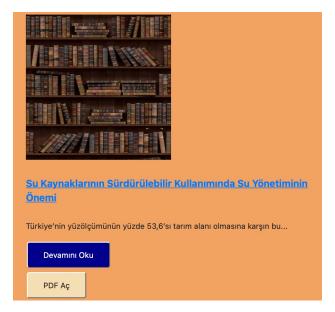
Kullanıcının makaleye git butonuna basıp gittiği nakaleler arka planda çalışan algoritma ile mongodb veritabanına kaydedilir.

Giriş ekranın altında yer alan kaydettiğim makaleler butonuna basılırsa ekrana kullanıcının makaleleri gelir.

Kullanıcı bu ekrandan makalenin özelliklerine göre dinamik bir filtreleme ve sıralama işlemi yapabilir.



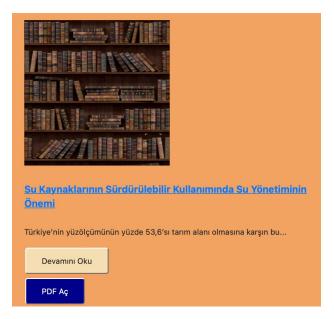
Kaydedilen her bir makalenin altında yer alan Devamını oku butonuyla kullanıcı makalenin bilgi sayfasına ulaşır.





Kaydedilen her bir makalenin altında yer alan Pdf AÇ butonuyla kullanıcı makalenin Pdfini ekranda görüntüleyebilir.

Önamli Notlar !( Eğer bir makalenin pdfi yoksa veya isim bilgisine ulaşılamıyorsa bu makale veritabanına kaydedilmez. Dinamik filtreleme ve sıralama işlemleri ElasticSearch yapısıyla indexleme yöntemi kullanılarak yapılmıştır.)





# IV. DENEYSEL SONUÇLAR

Arama motoruna istenilen kelime aratıldığında ekrana gelen ilk 24 sonucu listeliyoruz.Aşağıdaki fotoğraflarda görebilirsiniz.Daha sonrasında da makaleye git butonuna basıldığında makaleye ait bilgiler ekrana gelmektedir .Özet, yazar , yayımlandığı yıl vs .Bunu da aşağıdaki fotoğraftan görebilirisiniz.

Alttaki resimde görülen bir filtreleme örneğidir. Çay kelimesi aranmak istenen kelime labelına yazıldığında ve ara butonuna tıklandığında aşağıdaki gibi bir görüntü oluşur.

Article Card İçerikleri:	
Na Karmikiaram Similiridadir Kalimennda ka Yinatminia Önuni Pad ÖlKAN Japai 1970'S, Haras HUMA JAian 18745 (2013, 7)	rt Blind Driende Erreis, Sept. 2003, 100-103,
Follow papiles alederden alsons trito vertica (data yfficels elementarignen) ymyticidet en kapendian wenne delglift. Di mobilet in hymeddaren i kvernenne v pridordellet i bellandriane toen samendelglift. Di mobilet på prive som entrallet kolo sa kynellan kollomaliste verylenerinen ymentalet i koloni mellet mellet på prijatendare, in hymeddaren i konstrukt, alebediktig ver geve far kynelden Sa ytenden.	
	Makaleye Git
Sa Serimonia Sa Hidda es Sa ESA Corressindo Esdechnolódico i Banda Eddello - 2003, atamos Consenhos brinas es la collectiva e trans-	the three CRT Sep 2, MIS, O - M.
obseased in the two selected it. I giolescendar to topositionine recycles point metabolic tre scirifical characteristics, on the confirmation of the confirmation process and the responsabilities and the confirmation process and the responsabilities are the resource basing pointing in the two sense forestends for learning and the confirmation of the confirmation of the historical are historical	
	Makainye Git
Sa to Manteria, Ampiema Madest Fred VEMAX ; 5000; <u>Main Spiralate or Core</u> , CM 1. No.	
2000-2000 liter / ginditel Biyoki yezir sa kirdigi yinaemshabidi. Biyoki biyiki nirid odintir ologumaha Samumahdabi-Dix vi ngipalari nahatan odoki merkeboloki Rigoki ne dokesi problemler. Bu gendelmler sa politikasa gerelikini. Goydale venndi ne kollinemu onopurah yapmahla da Dikhideria sa kirdi kotora Bi wikilima hdabi.	
	Makalaye Git
Sa Gircell ( National Right SARDERMETOGLU ), 1989. Engine Entertain Visation Entertain Depth (Nr. Nov. I. 1987, 141 - 146, <u>1970) Nov.</u>	90.2501
deligned or redistriye git profes kined, bilged te yestl dejekt benil in servicinan nenya Quantum olot fording bu farm' by yelening profesigi. Fording redistrict receives destructing da dejekt from senson elekt fat delignik in kito kapelani celangar. Eleya silaman çek kala say endekte dispini, kitolika entirazen pisterin calmansın seley elektrikat, çekse, içlektir. İstanlır saktar Bir nektrici ve kitolikat in konsulet ve in kayasıtılarını kalanını yelenindi. Çekse, içlektir.	
	Makaleye Git
Trans Mater Olivab Ser. Se. Arab. St. Sanat Serve Sanat Ser Transis Angerica Material Material Material Sci. (ICC), Assessive Sanates Control Services Co.	38 S. Say J. 2021, 279 - 466,
Be çalışma, doğal kaşınık ve kama hizmeti olanık sanımlarını ne kaşınağının erişin, poşluşmı ve doğum semilarını tişdiri qarkın çabilmilde deli pşasa neklasilminasının keriği güncü kevinde rolan Su- riyak bil, Sanal iye ve Sanal Sir Tulanist kerinde kerindelakti Semiların kerindelin beşiyik kerindi pi pşasa	

• B
Editybaseen 13ds
Exercision to
Iblin. 105. Kraiktósz
31.365.5 Yik 2013 Saye: 1,190 - 195,01.02 2003
Dol DZCO Block AND N Home HCRMA Block ANTAS
The Confidence of the Confiden
Salvation
Sathalahirik
Santanias.

Yayın Ara				
Arama Yayın Adı	Arama Yazar Adı	Arama Özet	Arama Anahtar Kelimeler	Arama ID
ARA				
Filtre Sıfırla				
CAY VE CAY EENOLİKI	LERININ ANTIOKSIDAN	AVTIVITESI		
	ın çok tüketilen bir içecektir. So			
	ar you tollean oil lyecekill. St			
Devamini Oku				
PDF Aç				

## V. SONUÇ

İlk defa web scraping içeren bir projede bulunduk , edindiğimiz bilgiler ve deneyimler bizi çok geliştirdi. Htm-lagiltypack kullanımı hakkında bilgi sahibi olduk.MONGODB ve Elasticsearch yapılarının kullandık ve bir daha bu tarz bir proje içinde bulunursak kullanmamız ve ekip arkadaşlarımıza yardımcı olmamız daha kolay olacak.

#### VI. KAYNAKÇA

https://dergipark.org.tr/tr/search?q=nemsection=articles https://www.youtube.com/watch?v=ullxVcavOSY https://www.youtube.com/watch?v=m9zFq6KS94Y https://www.youtube.com/watch?v=mh48aGK85LQ https://www.youtube.com/watch?v=-bt\_u4Loof g

VII. AKIŞ ŞEMASI SONRAKI SAYFALARDADIR