

Kocaeli Üniversitesi

Bilgisayar Mühendisliği Bölümü

Yazılım Laboratuvarı II

Web Scraping Akademi Uygulaması

Oğuzhan Çelik-190202105

Can Güneri-190202094

GİRİŞ

Projenin Özeti

Yazılım Laboratuvarı II Projesi olarak bizden “Web Scraping Akademi Uygulaması” adındaki web scraping ile bir web sayfasından bilgiye erişim sağlama; MongoDB veritabanı ile Elasticsearch sorgu yapılarını kullanma ve web kodlama becerilerinin geliştirilmesi amaçlanan bir proje geliştirmemiz istenmiştir.

Web Scraping Akademi Uygulaması, kullanıcıların Google Scholar üzerinde belirledikleri bir konu hakkında arama yaparak verileri çekmelerini ve kaydetmelerini sağlayan bir uygulamadır. Python dilinde geliştirilen bu uygulama, kullanıcı dostu bir arayüzle donatılmıştır. Kullanıcılar, uygulamayı başlattıklarında bir arama yapmak istedikleri makale konusunu girer ve kaç sayfa veri çekmek istediklerini belirlerler. Ardından, verilerin kaydedileceği Excel dosyasının klasör konumunu seçerler ve arama işlemini başlatırlar.

Uygulama, kullanıcıların veri çekme sürecini izlemelerini ve takip etmelerini sağlar. Veriler, hem Excel dosyasına hem de MongoDB veritabanına kaydedilir, böylece kullanıcılar istedikleri zaman verileri geri çağırabilir ve analiz edebilirler. Bu sayede, kullanıcılar Google Scholar'dan hızlı ve etkili bir şekilde veri çekme işlemlerini gerçekleştirebilirler. Uygulama ayrıca kullanıcıları, verilerin başarıyla kaydedildiği bilgisiyle bilgilendirir, böylece kullanıcılar işlemin tamamlandığını doğrulayabilirler.

Web Scraping Akademi Uygulaması, kullanıcıların Google Scholar üzerinde arama yaparak bilimsel makaleler hakkında veri çekmelerini ve bu verileri kaydetmelerini sağlayan bir uygulamadır. Bu uygulama, Python programlama dili kullanılarak geliştirilmiştir ve kullanıcı dostu bir arayüzle donatılmıştır.

Uygulamanın temel amacı, araştırmacıların ve öğrencilerin belirledikleri konularla ilgili en güncel bilimsel verilere kolayca erişmelerini sağlamaktır. Kullanıcılar, uygulamayı başlattıklarında aramak istedikleri makale konusunu girerler ve kaç sayfa veri çekmek istediklerini belirlerler. Uygulama, kullanıcının veri çekme sürecini izlemesini ve kayıt altına almasını sağlar.

Ayrıca, uygulama verileri hem Excel dosyasına hem de MongoDB veritabanına kaydederek kullanıcıların verilere istedikleri zaman erişmelerini sağlar. Bu sayede, araştırmacılar ve öğrenciler Google Scholar'dan kolayca veri çekme işlemlerini gerçekleştirebilir ve bu verileri analiz ederek çalışmalarına katkı sağlayabilirler. Uygulama, kullanıcıları verilerin başarıyla kaydedildiği bilgisiyle de bilgilendirir, böylece kullanıcılar işlemin tamamlandığını doğrulayabilirler.

1. TEMEL BİLGİLER

Projenin temelinde Python programlama dili ve Visual Studio Code geliştirme ortamı kullanılmıştır. Kodlar, scraping.py başlığı altında organize edilmiş ve geliştirme sürecinde gerekli olan kütüphaneler (requests, BeautifulSoup, pandas, tkinter, pymongo) kullanılmıştır. Bu kütüphaneler, web scraping işlemleri için gereken HTTP istekleri, HTML içeriğinin analizi, veri manipülasyonu, kullanıcı arayüzü oluşturma ve MongoDB ile etkileşim gibi işlevleri sağlar.

Kodlar, Google Scholar gibi bir web sitesinden bilimsel makale verilerini çekme işlemlerini gerçekleştirir. requests kütüphanesi ile web sitesine HTTP istekleri gönderilir ve BeautifulSoup kütüphanesi ile sayfa içeriği analiz edilir. Bu analiz sonucunda, makale başlıkları, yazarlar, yayın yılları, URL'ler gibi veriler çekilir ve kullanıcı arayüzünde gösterilir. Veriler, kullanıcının belirlediği Excel dosyasına kaydedilir ve aynı zamanda MongoDB veritabanına da aktarılır.

Python'un esnekliği ve kütüphanelerinin sağladığı fonksiyonellikler sayesinde, Web Scraping Akademi Uygulaması projenin ana yapısını oluşturur ve kullanıcıların web scraping işlemlerini kolayca gerçekleştirmesini sağlar. Bu temel bileşenler, projenin işlevselliği ve başarılı çalışması için önemlidir.

2. YÖNTEM

Web Scraping Akademi Uygulaması, Python programlama dilinde geliştirilmiş bir web scraping aracıdır. Bu uygulama, kullanıcıların Google Scholar üzerinde belirli bir konu veya anahtar kelime ile arama yaparak makaleleri çekmesini ve bu verileri Excel dosyasına ve MongoDB veritabanına kaydetmesini sağlar.

Makaleleri Tara Metodu

Projenin en temel yöntemlerinden biri olan makaleleri_tara(sorgu, sayfa_sayisi) metodu, Google Scholar üzerinde yapılan aramalardan makaleleri çekmek için kullanılır. Bu metodun işleyişi şu adımlar üzerinde gerçekleşir:

1. URL Oluşturma: Kullanıcı tarafından belirtilen sorgu ve sayfa sayısı bilgileri temel alınarak Google Scholar üzerindeki arama URL'si oluşturulur. Bu URL, requests kütüphanesi aracılığıyla taranacak sayfanın adresini belirler.
2. Sayfa İçeriğini Alma: Oluşturulan URL üzerinden yapılan istek sonucunda elde edilen HTML içeriği, BeautifulSoup kütüphanesi ile analiz edilir. Bu sayede arama sonuçlarındaki makalelerin HTML yapısı incelenerek gerekli bilgiler çekilir.

3. Makale Bilgilerini Alma: Sayfa içeriği taranarak her bir makalenin başlık, yazarlar, yayın yılı ve URL bilgileri çekilir. Bu bilgiler, veri_listesi adlı bir liste içinde sözlük formatında saklanır.
4. Veri Listesini Döndürme: Tüm sayfa tarandıktan sonra elde edilen veri_listesi, metodun dışında kullanılmak üzere döndürülür.

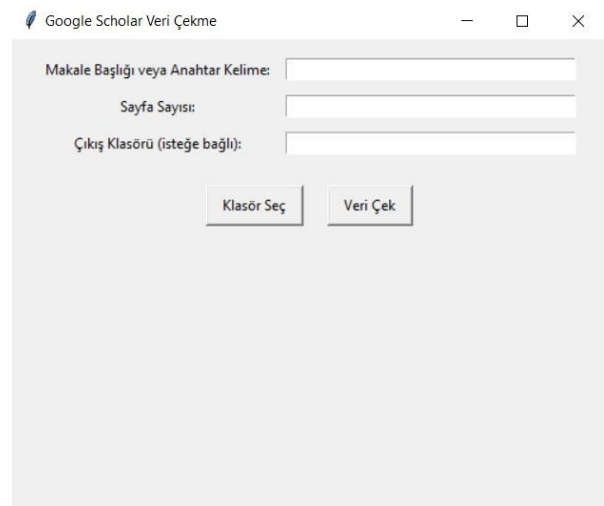
Excel Aktar Metodu

Uygulamanın bir diğer önemli metodu olan excel_aktar(veri_listesi, dosya_adi) metodu, çekilen verileri Excel dosyasına aktarmak için kullanılır. Bu metodun işleyişi şu şekildedir:

1. Veri Dönüştürme: Çekilen veri_listesi, pandas kütüphanesi kullanılarak DataFrame'e dönüştürülür. Bu sayede veriler tablo formatında düzenlenir.
2. Excel'e Yazma: DataFrame içindeki veriler, to_excel metodu aracılığıyla belirtilen dosya adıyla birlikte Excel dosyasına yazılır. Bu dosya, kullanıcının belirlediği klasör konumuna kaydedilir.

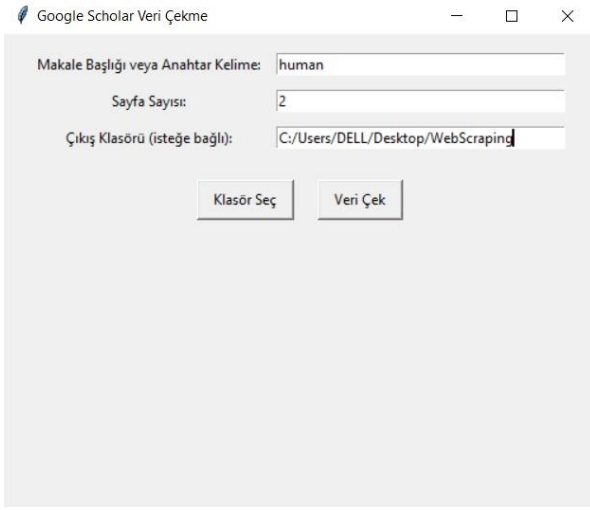
Klasör Seç ve Veri Çek Metotları

Uygulamanın kullanıcı arayüzü tarafında yer alan klasor_sec() ve makaleleri_tara_aksiyon() metodları da önemlidir. klasor_sec() metodunun görevi, kullanıcıya klasör seçme penceresi açarak çıkış klasörünü belirlemesine olanak tanımaktır. makaleleri_tara_aksiyon() metod ise "Veri Çek" butonuna basıldığında çalışacak işlemleri gerçekleştirir.



Ana Pencere ve Kullanıcı Arayüzü

Uygulamanın ana penceresi, tkinter kütüphanesi kullanılarak oluşturulur. Giriş alanları, etiketler, klasör seçme ve veri çekme butonları, durum etiketi gibi kullanıcı arayüzü bileşenleri eklenir. Bu sayede kullanıcılar, kolayca arama yapabilir, verileri çekebilir ve Excel dosyasına kaydedebilirler.



Web Scraping Akademi Uygulaması'nın temel yöntemleri ve kullanıcı arayüzü, web scraping işlemlerini kullanıcılar için daha kolay ve erişilebilir hale getirir. Bu sayede kullanıcılar, Google Scholar üzerindeki verilere daha hızlı ve etkili bir şekilde erişebilir ve bu verileri analiz edebilirler.

3. SONUÇ

Web Scraping Akademi Uygulaması, modern araştırma ve veri çekme tekniklerinin güçlü bir örneği olarak karşımıza çıkmaktadır. Bu uygulama, akademik dünyada bilimsel makalelerin araştırılması ve verilerin toplanması gibi önemli bir ihtiyaca yanıt vermektedir. Kullanıcı dostu arayüzü ve etkileyici yetenekleri ile araştırmacıların işlerini kolaylaştırmak için tasarlanmıştır.

Öncelikle, uygulama kullanıcılarına Google Scholar üzerinde istedikleri konuda arama yapma ve çıkan sonuçları detaylı bir şekilde inceleme imkanı sunar. Bu sayede araştırmacılar, ilgili makaleleri hızlı bir şekilde bulabilir ve gereken bilgilere erişebilirler. Sayfa sayısı belirtme özelliği, kullanıcıların geniş kapsamlı aramalar yapmasına ve daha fazla veriye ulaşmasına olanak tanır.

Web Scraping Akademi Uygulaması'nın en dikkat çeken özelliklerinden biri, çekilen verileri Excel dosyasına aktarma ve MongoDB veritabanına kaydetme yeteneğidir. Bu sayede kullanıcılar, çekilen verileri kolayca depolayabilir ve istedikleri zaman erişebilirler. Excel dosyasına aktarılan verilerin düzenli bir şekilde sunulması, araştırmacıların verileri analiz etme ve raporlama süreçlerini kolaylaştırır.

Ayrıca, uygulamanın klasör seçme özelliği kullanıcıların verileri istedikleri konumda saklamalarına olanak sağlar. Bu da verilerin düzenli bir şekilde yönetilmesini ve depolanmasını sağlayarak kullanıcı deneyimini artırır.

Web Scraping Akademi Uygulaması'nın sunduğu bu özellikler, akademik araştırmaların verimliliğini artırmak ve araştırmacılara zaman kazandırmak için son derece önemlidir. Uygulamanın kullanıcı dostu arayüzü, güçlü veri çekme yetenekleri ve verileri düzenli bir şekilde depolama özellikleri, araştırmacıların bilimsel çalışmalarını daha etkili bir şekilde yürütmelerine yardımcı olur.

Sonuç olarak, Web Scraping Akademi Uygulaması'nın geliştirilmesiyle araştırmacılar, bilimsel verilere hızlı ve güvenilir bir şekilde erişebilirler. Bu uygulama, akademik dünyada bilgiye erişimi kolaylaştırarak araştırmacıların çalışmalarını daha verimli hale getirir ve bilimsel ilerlemenin önünü açar.

