NO. 1

```
import bs4
import requests
from urllib.request import urlopen as uReq
from bs4 import BeautifulSoup

response = requests.get('https://unair.ac.id/news/')
rawhtml = response.text
soup = BeautifulSoup(rawhtml, 'html.parser')

for i in soup.find_all('h2'):
    print(i.get_text())
```

```
                    Websvaganza 2023 Hadirkan Bazar Kosmetik untuk
Dorong Kepercayaan Diri dan Hilangkan Insecure


                    Mahasiswa UNAIR Sabet Juara 1 Kategori Lomba
Infografis KOMINFO Jawa Timur


                    Pentingnya Mengenal Potensi Diri Melalui
Pemahaman Emosional


                    Dukung Pendidikan Merata UNAIR Berikan Beasiswa
Siswa Sekolah Dasar


                    Apa Sih UKM Wanala ?


                    Ancam Boikot SpaceX, Dosen UNAIR Sebut Israel
Tidak Ingin Aksi Genosida Diketahui Dunia Luar


                    Dr Andriyanto, Alumnus UNAIR yang Dilantik
Menjadi Pj Bupati Pasuruan
```

```
                    Kisah Mahasiswa UNAIR Eksplor Kanada Melalui
IISMA di University of Waterloo



                    Perjalanan Menantang Mahasiswa UNAIR Ikuti IISMA
di University of Szeged Hungaria
```

Penjelasan code :

1. **import bs4**: Mengimpor modul bs4 dari pustaka BeautifulSoup
2. **import requests**: Mengimpor modul requests untuk melakukan pengiriman permintaan HTTP ke situs web dan mengambil kembali responsenya.
3. **from urllib.request import urlopen as uReq**: Mengimpor fungsi urlopen dari modul urllib.request yang diberi nama uReq agar lebih mudah digunakan.
4. **from bs4 import BeautifulSoup**: Mengimpor kelas BeautifulSoup dari modul bs4.
5. **response = requests.get('https://unair.ac.id/news/')**: Mengirimkan permintaan GET ke URL 'https://unair.ac.id/news/' dan menyimpan responsenya di dalam variabel response.
6. **rawhtml = response.text**: Mengambil teks dari respons HTTP lalu menyimpannya ke dalam variabel rawhtml.
7. **soup = BeautifulSoup(rawhtml, 'html.parser')**: Membuat objek BeautifulSoup dari teks HTML yang telah diambil sebelumnya (rawhtml). untuk melakukan analisis dan pencarian pada struktur HTML.
8. **for i in soup.find_all('h2'):**: Menggunakan metode find_all untuk mencari dan mengambil semua elemen HTML yang memiliki tag h2. Selanjutnya akan menghasilkan daftar dari semua elemen dengan tag h2 dalam halaman web.
9. **print(i.get_text())**: Mencetak teks yang terkandung dalam setiap elemen h2. get_text() adalah metode BeautifulSoup yang digunakan untuk mendapatkan teks dari elemen HTML tanpa tag atau atribut HTML.

#NO. 2

```
import requests
from bs4 import BeautifulSoup

for i in range(1, 4):
    url = "https://unair.ac.id/category/featured/page/"+str(i)+"/"
    response = requests.get(url)
    rawhtml = response.text
    soup = BeautifulSoup(rawhtml,'html.parser')
```

```
    for j in soup.find_all('h3'):
        print(j.get_text())
```

```
                        Dr Andriyanto, Alumnus UNAIR yang Dilantik
Menjadi Pj Bupati Pasuruan


                        UNAIR Raih 4,5 Trees Rating pada UI GreenMetric
World University Ranking


                        Pakar Politik UNAIR Sebut Pengusungan Gibran
Jadi Strategi Jangka Panjang


                        Berkomitmen Tingkatkan Transparansi Informasi,
UNAIR Gabung JDIH


                        Komitmen Tingkatkan Kualitas Pendidikan, Rektor
UNAIR Kukuhkan Enam Guru Besar


                        Tambah Lagi, UNAIR Kini Miliki 11 Jurnal Ilmiah
Terindeks Scopus


                        UNAIR Raih Anugerah Jatim Bangkit Awards Berkat
Sukseskan Pemulihan Pandemi


                        Kukuhkan Tujuh Guru Besar, Rektor UNAIR Ajak
Akademisi Tingkatkan Daya Kritis


                        Beri Kuliah Tamu di UNAIR, Mahfud MD Tekankan
Pentingnya Politik Kebangsaan


                        UNAIR Anugerahi Khofifah Gelar Doktor Honoris
Causa
```

UNAIR Luluskan 1.382 Wisudawan, Rektor: Anda Orang-Orang Terpilih

Rektor Beri Pesan Gubes untuk Bumikan Ilmu Pengetahuan

UNAIR Bagikan 1974 Sertifikat Halal Gratis untuk UMKM

Kukuhkan Tujuh Guru Besar, Rektor UNAIR Tekankan Kerja Sama Riset Internasional

UNAIR Tingkatkan Kontribusi dengan Tambah 7 Guru Besar Baru

UNAIR Duduki Peringkat Kedua Nasional Versi THE WUR

Dukung Merdeka Belajar, UNAIR Berikan Ruang Eksplorasi Bagi Mahasiswa

Pengukuhan Gubes Wujudkan UNAIR Jadi SMART University

Jadi Tuan Rumah The 6th ASEAN+3 Rector’s Conference, UNAIR Tekankan Pentingnya Kolaborasi

Intip Kisah Guru Besar UNAIR yang 18 Tahun Mengabdi Jadi Dokter Forensik

Kukuhkan Gubes, Rektor: Semoga Target Segera Tercapai

Singkirkan Puluhan Peserta, Mahasiswa UNAIR Bawa Pulang Medali Emas

Jadi Pionir, FKH UNAIR Adakan Program MBKM di 10 Wilayah Jatim

Resmi! UNAIR Bakal Miliki Plaza Airlangga

Menkes Luncurkan Permenkes Rumah Sakit Kapal, Berobat di RSTKA Kini Bisa Pakai BPJS

Pengukuhan Guru Besar Jadi Tambahan Energi bagi UNAIR

Kukuhkan Empat Guru Besar FK, Rektor UNAIR Tekankan Kecintaan pada Ilmu

Masuk Deretan Top 100 Peneliti Indonesia, Wakil Rektor UNAIR Bagikan Tips bagi Peneliti Pemula

UNAIR Siap Tingkatkan Kontribusi dengan 12 Guru Besar Baru

UNAIR Sambut Kedatangan 390 Peserta Pertukaran Mahasiswa Merdeka

Janis Rosalita, Alumnus UNAIR jadi Atlet Terbaik

```
Putri SIWO 2023


                        Rektor UNAIR Beri Tanggapan Kebijakan Baru
Penghapusan Skripsi


                        IKA UNAIR Wilayah Inggris Gelar Silaturahmi dan
Konsolidasi Alumni


                        Produk Penurun Glukosa dalam Darah Gubes UNAIR
Tembus Pameran Internasional di Pakistan


                        Mahasiswa UNAIR Ciptakan Alat Deteksi Stress
pada E-Sport


                        Tanam Jahe dan Sereh untuk Tingkatkan Kesehatan
Warga Flat Kenari
```

Penjelasan code

1. **import request**s: mengimpor pustaka requests, untuk melakukan permintaan HTTP ke situs web.
2. **from bs4 import BeautifulSoup**: mengimpor kelas BeautifulSoup dari pustaka BeautifulSoup,untuk melakukan analisis dan manipulasi dokumen HTML.
3. **for i in range(1, 4):**: melakukan iterasi tiga kali, dengan nilai i mulai dari 1 hingga 3 (inklusif). Tujuan dari loop ini adalah untuk mengakses tiga halaman berbeda dari situs web.
4. **url = "https://unair.ac.id/category/featured/page/"+str(i)+"/"**: Di setiap iterasi, kode akan membangun URL dengan menggabungkan bagian tetap "https://unair.ac.id/category/featured/page/" dengan nilai i yang saat itu. Misal, untuk i = 1, URL akan menjadi "https://unair.ac.id/category/featured/page/1/".
5. **response = requests.get(url)**: mengirimkan permintaan GET ke URL yang telah dibuat sebelumnya dan menyimpan respons HTTP di dalam variabel response.
6. **rawhtml = response.text**: mengambil teks dari respons HTTP dan menyimpannya dalam variabel rawhtml.

7. **soup = BeautifulSoup(rawhtml,'html.parser')**: Membuat objek BeautifulSoup dari teks HTML yang telah diambil sebelumnya. Ini memungkinkan kita untuk melakukan analisis dan pencarian pada struktur HTML.

8. **for j in soup.find_all('h3'):**: Menggunakan metode find_all untuk mencari dan mengambil semua elemen HTML yang memiliki tag h3. Lalu akan menghasilkan daftar dari semua elemen dengan tag h3 dalam halaman web.

9. **print(j.get_text())**: Memunculkan output yang terkandung dalam setiap elemen h3. get_text() adalah metode BeautifulSoup yang digunakan untuk mendapatkan teks dari elemen HTML tanpa tag atau atribut HTML.

#NO. 3

Berikut adalah kode scraping pada halaman web Playstation store untuk mengambil judul dan harga game dan menyimpannya dalam file csv

```
import scrapy
import pandas as pd

class QuotesSpider(scrapy.Spider):
    name = "quotes"

    start_urls = [
        "https://store.playstation.com/en-id/category/05a2d027-cedc-
4ac0-abeb-8fc26fec7180/"
    ]

    def parse(self, response):
        games = response.css('div.psw-product-tile.psw-interactive-
root')

        data = []

        for game in games:
            judul = game.css('span.psw-t-body.psw-c-t-1.psw-t-
truncate-2::text').extract_first()
            harga = game.css('div.psw-fill-x.psw-price.psw-l-
inline.psw-l-line-left-top > div > span::text').extract_first()

            if judul and harga:
                harga_bersih = harga.replace("Rp", "").replace("\xa0",
"").strip()
                data.append({
                    "Judul": judul,
                    "Harga": "Rp " + harga_bersih
                })

        # Menyimpan data dalam file CSV
```

```
        df = pd.DataFrame(data)
        df.to_csv("output.csv", index=False)

        print(df)
```

Penjelasan Code

1. **import scrapy dan import pandas as pd:**
   - mengimpor pustaka Scrapy untuk mengekstrak data dari situs web.
   - mengimpor pustaka Pandas,untuk manipulasi dan analisis data / menyimpan data dalam bentuk DataFrame dan menulisnya ke file CSV.
2. **class QuotesSpider(scrapy.Spider)**:
   - definisi kelas untuk *spider* adalah entitas utama dalam Scrapy yang melakukan penarikan data dari situs web.
3. **name = "quotes"**:
   - membuat variabel untuk memanggil *spider* saat menjalankan perintah Scrapy.
4. **start_urls = [...]:**
   - membuat variabel URl untuk data dari web yang ingin di scrapy
5. **def parse(self, response):**
   - kelas *spider* yang akan dieksekusi pertama kali ketika *spider* dimulai. response adalah objek yang berisi konten dari halaman web yang diunduh.

6.** games = response.css('div.psw-product-tile.psw-interactive-root'):**

- menggunakan format css untuk memilih elemen HTML yang mengandung informasi tentang game. kemudian akan diiterasi untuk mengekstrak judul dan harga game.

1. **for game in games:**
   - **judul = game.css('span.psw-t-body.psw-c-t-1.psw-t-truncate-2::text').extract_first():** mengekstrak teks dari elemen yang berisi judul game.
   - **harga = game.css('div.psw-fill-x.psw-price.psw-l-inline.psw-l-line-left-top > div > span::text').extract_first():** mengekstrak teks dari elemen yang berisi harga game. -**if judul and harga:** Memastikan bahwa baik judul maupun harga ditemukan sebelum melanjutkan.
   - **harga_bersih = harga.replace("Rp", "").replace("\xa0", "").strip():** Membersihkan teks harga dari karakter tambahan seperti "Rp" dan spasi.
   - **data.append({"Judul": judul, "Harga": "Rp " + harga_bersih}):** Menambahkan judul dan harga yang telah diekstrak ke dalam list `data`.
2. **df = pd.DataFrame(data):** Membuat DataFrame menggunakan Pandas dengan data yang telah diekstrak.
3. **df.to_csv("output.csv", index=False):** Menyimpan DataFrame ke dalam file CSV bernama "output.csv".

4. **print(df):** Mencetak DataFrame ke konsol.