**PRAKTIKUM ALGORITMA STRUKTUR DATA**

**MODUL 7**

***Regular Expressions***

****

**Disusun oleh:**

**Adinda Aulia Hapsari**

**L200220037**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS KOMUNIKASI DAN INFORMATIKA**

**UNIVERSITAS MUHAMMADIYAH SURAKARTA**

**TAHUN 2024**

|  |  |
| --- | --- |
| Setelah kegiatan selesai, lembar kerja ini dicetak (di-print) dan dikumpulkan ke asisten.  NIM : L200220037  Nama : Adinda Aulia Hapsari  Nama Asisten :  Tanggal Praktikum : 24 Mei 2024 | (Diisi oleh Asisten)  Nilai Praktek :  Tanda Tangan : |

**Soal-soal untuk Mahasiswa**

**SUDAH DINILAI DI KELAS**

1. Mengekstrak kata-kata dengan awalan ‘me’. Bukalah halaman web id.wikipedia. org/wiki/Indonesia. Salin seluruh tubuh artikel itu dan tuang ke Notepad3. Simpan sebagai Indonesia.txt. Tugasmu adalah mengekstrak semua kata yang mempunyai awalan ‘me’, lalu menyimpannya dalam sebuah tuple. Perhatikan bahwa ‘me’ bisa muncul diawal, tengah, ataupun akhir kalimat.

print('\n--- Oleh L200220037 ---')

import re

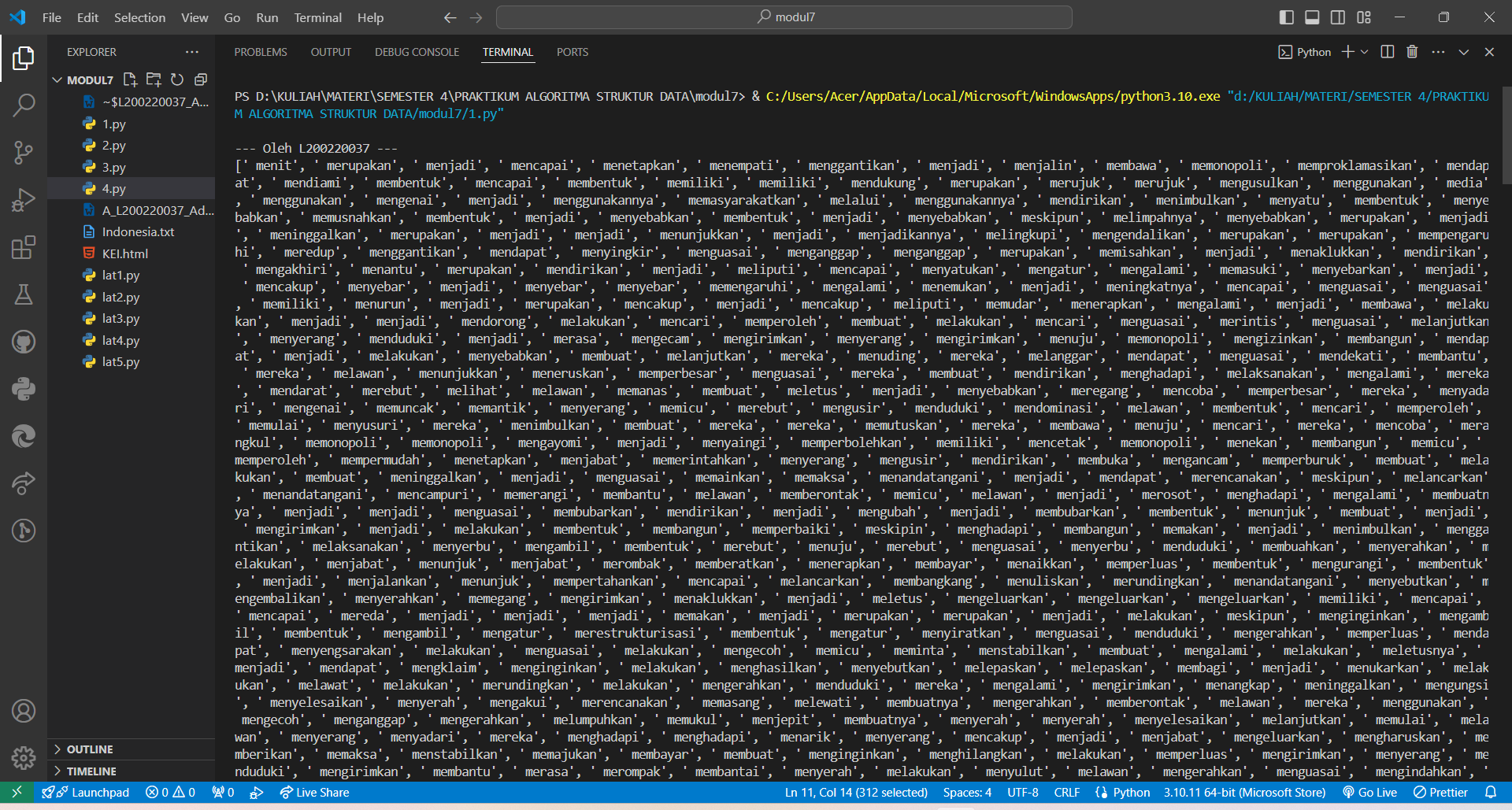
a = open('Indonesia.txt','r')

baca = a.read()

a.close

cocok= re.findall("\sme\w+",baca)

print(cocok)



1. Mengekstrak kata-kata dengan awalan ‘di’. Sama seperti diatas, tapi yang kamu ekstrak adalah kata-kata dengan awalan ‘di’. Perhatikan, yang dimaksud disini adalah awalan ‘di’ (menunjukkan kata kerja pasif), bukan kata depan ‘di’ (menunjukkan kete rangan tempat), yang kita bahas di nomer di bawah.

print('\n--- Oleh L200220037 ---')

import re

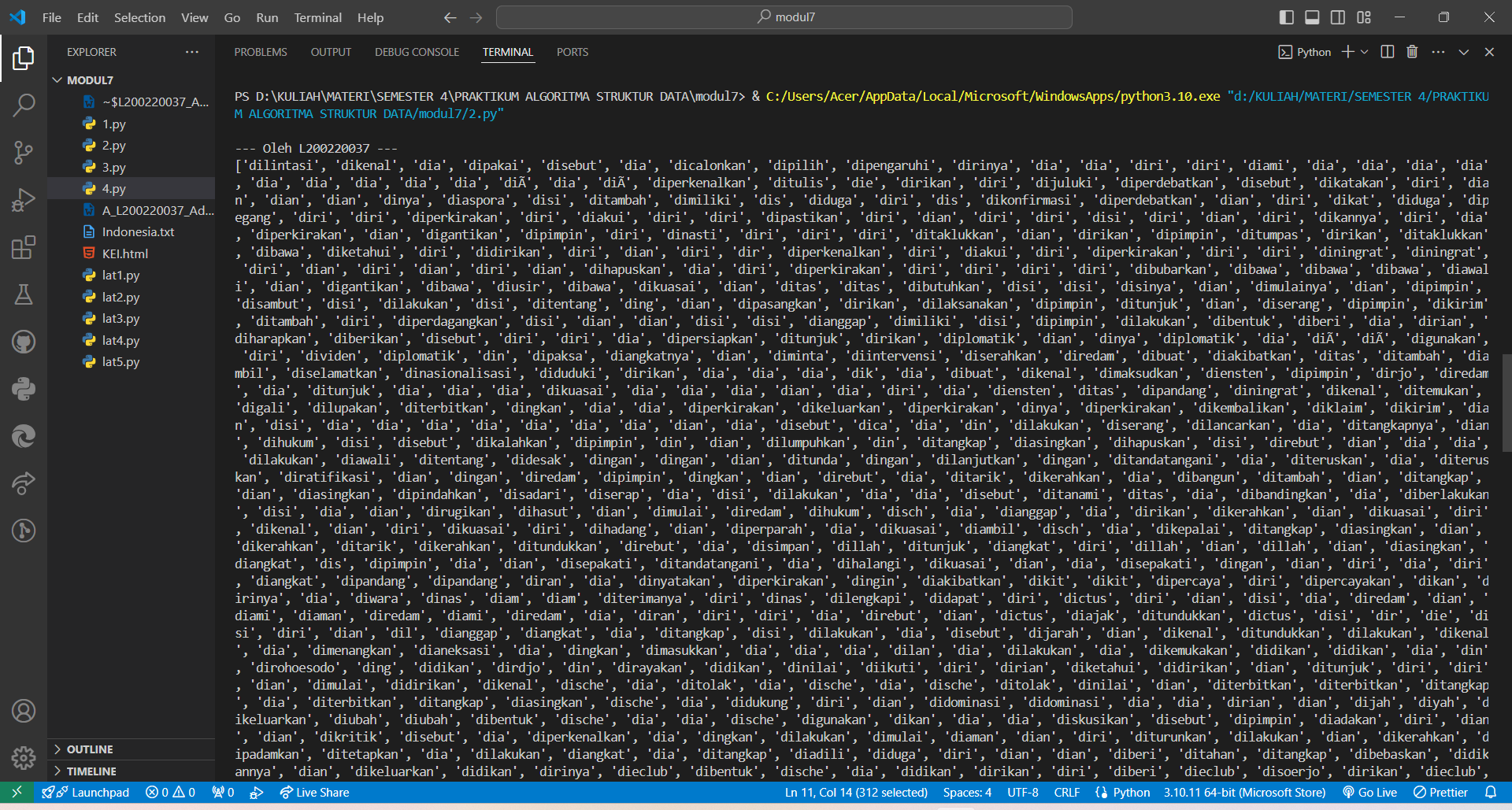
a = open('Indonesia.txt','r')

baca = a.read()

a.close

cocok= re.findall(r"di\w+",baca)

print(cocok)



1. Mengekstrak kata depan ‘di’ dan tempat yang ditunjuknya. Di sini tugasmu adalah mengekstrak, masih dari berkas Indonesia.txt, kata depan ‘di’ beserta satu kata yang mengikutinya. Misal, dari string

’Saya dilahirkan dan dibesarkan di sini, di Indonesia, sebuah negeri indah permai di mana para saudagar di masa lampau berdatangan.’

Akan terekstrak (’di sini’,’di Indonesia’,’di mana’,’di masa’)

print('\n--- Oleh L200220037 ---')

import re

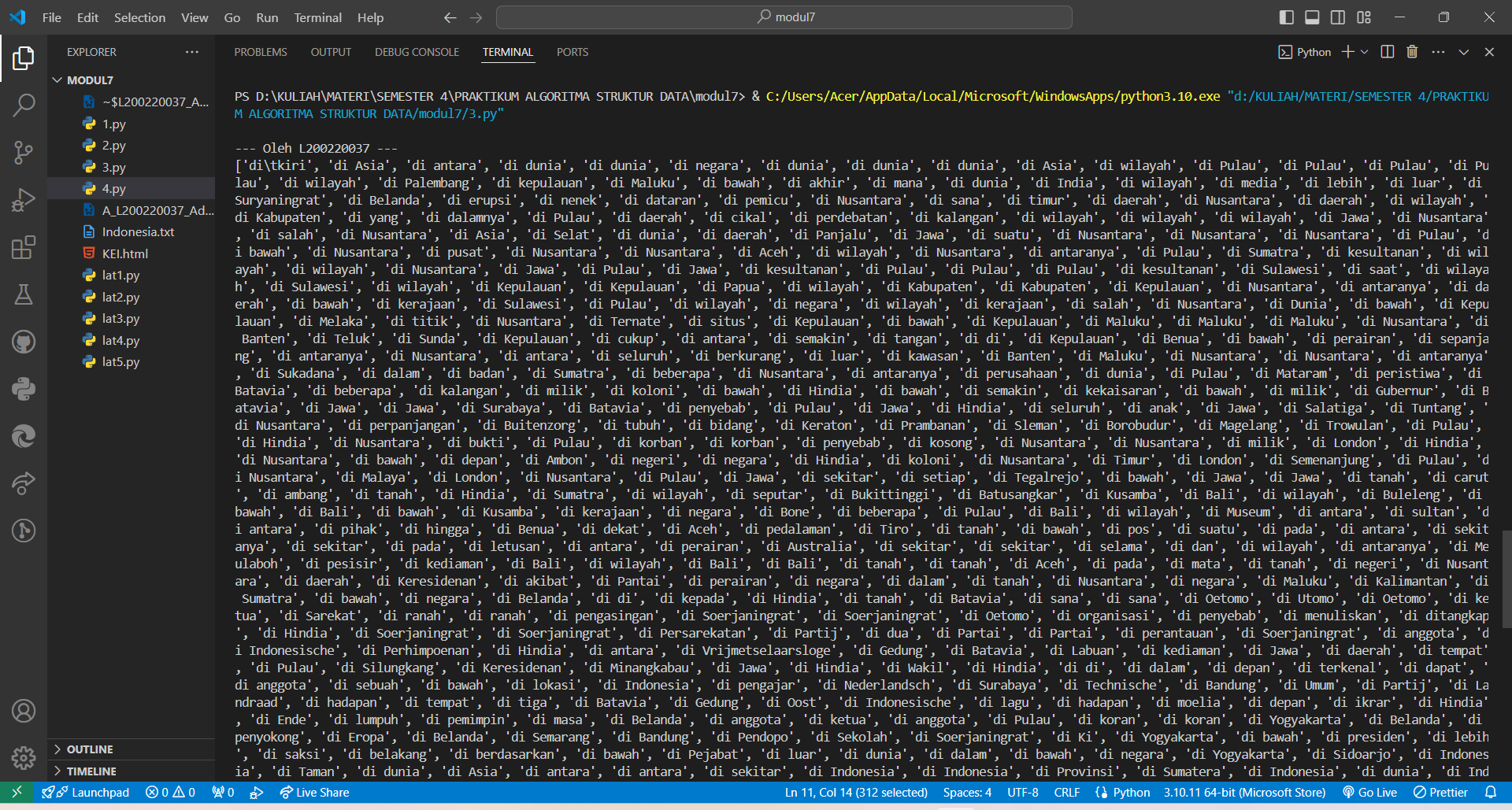
a = open('Indonesia.txt','r')

baca = a.read()

a.close

cocok= re.findall(r"di\s\w+",baca)

print(cocok)



1. Mengekstrak data dari berkas .html. Tugasmu di sini adalah membuat sebuah list of tuples yang memuat nama-nama negara beserta Innovation index-nya. Kunjungilah halaman en.wikipedia.org/wiki/Knowledge\_Economic\_Index. Pastikan halaman itu memuat sebuah table besar hasil penelitian. Kerjakan langkah-langkah berikut.

* Simpanlah halaman itu sebagai berkas html murni, dengan nama KEI.html
* Bukalah berkas KEI.html itu dengan Notepad++ atau Notepad. Perhatikan segmen segmen defile itu yang mempunyai isi seperti dibawah.



* Akses file itu menggunakan f=open (’KEI.html’,’r’,encoding=’latin1’) dan baca file menggunakan teks=f.read(). Lalu tutup aksesnya menggunakan f.close().
* Sebagai awalan, ekstraklah semua nama-nama negara yang ada diberkas itu memakai regex, dengan memperhatikan pola di sekitaran nama-nama negara.
* Sekarang ekstraklah kolom “Innovation” untuk semua negara di tabel itu (untuk contoh Belgium di atas, kolom ini adalah kolom yang angkanya 8.96 ) memakai regex7. Sebagai tambahan, ubahlah string ’8.96’ menjadi float.
* Dengan konsep group (memakai tanda kurung (...) di tempat yang tepat, lihat Section 7.4 dihalaman 70), ekstraklah nama negara beserta Innovation Index-nya, lalu buatlah kode seperlunya untuk memodifikasinya, sehingga akhirnya kamu mempunyai listof tuples seperti berikut.

[(’Belgium’, 8.96), (’Malaysia’, 6.83), (’Indonesia’, 3.32), ...]

print('\n--- Oleh L200220037 ---')

import re

f = open('KEI.html', 'r', encoding='latin1')

teks = f.read()

f.close()

i = r'\s<td>[\d.\w/]+</td>'

p = r'(\w+)</a></td>' + i + i + i + r'\s<td>([\d.\w/]+)</td>'

cocok = re.findall(p, teks)

cocok1 = [(i[0], float(i[1])) for i in cocok]

print(cocok1)

