

PRAKTIKUM ALGORITMA STRUKTUR DATA

MODUL 7

Regular Expressions



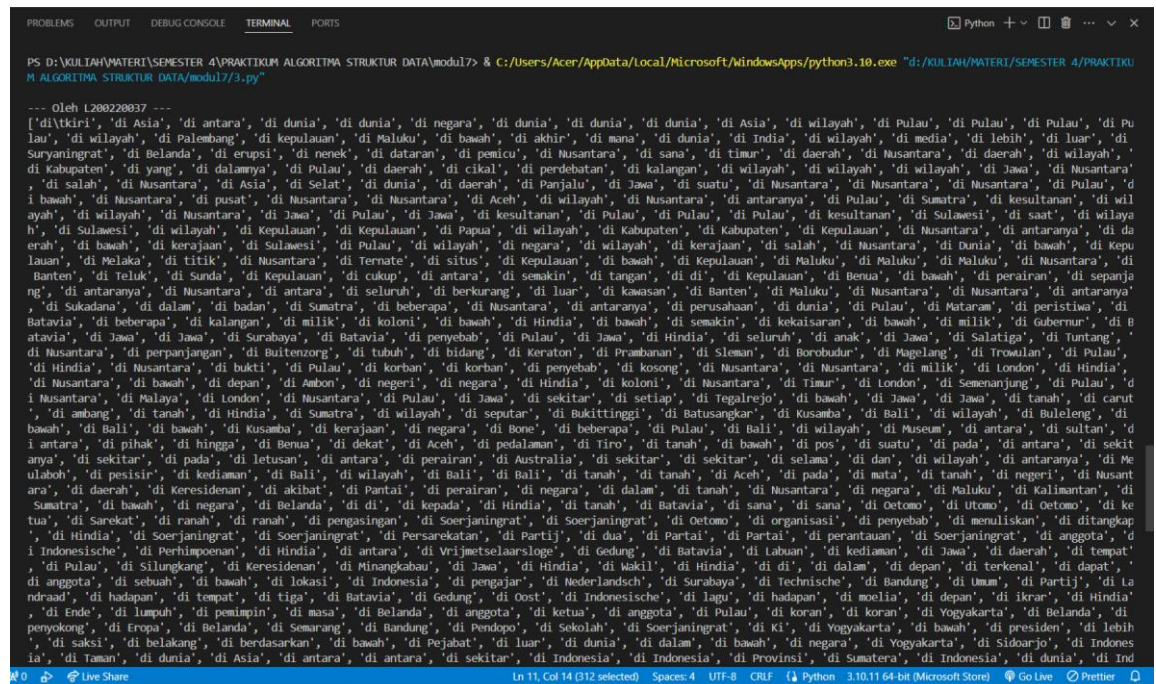
Disusun oleh:

Adinda Aulia Hapsari

L200220037

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS MUHAMMADIYAH SURAKARTA
TAHUN 2024**


```
print(cocok)
```



```
PS D:\KULIAH\WATERI\SEMESTER 4\PRAKTIKUM ALGORITMA STRUKTUR DATA\modul7> & C:/Users/Acer/AppData/Local/Microsoft/WindowsApps/python3.10.exe "d:/KULIAH/WATERI/SEMESTER 4/PRAKTIKUM ALGORITMA STRUKTUR DATA/modul7/3.py"

--- Oleh L200220937 ---

['di\tikiri', 'di Asia', 'di antara', 'di dunia', 'di dunia', 'di negara', 'di dunia', 'di dunia', 'di dunia', 'di Asia', 'di wilayah', 'di Pulau', 'di Pulau', 'di Pulau', 'di Pu
lau', 'di wilayah', 'di Palembang', 'di kepulauan', 'di Maluku', 'di bawah', 'di akhir', 'di mana', 'di dunia', 'di India', 'di wilayah', 'di media', 'di lebih', 'di luar', 'di
Suryaningrat', 'di Belanda', 'di erupsi', 'di nenek', 'di dataran', 'di pemicu', 'di Nusantara', 'di sana', 'di timur', 'di daerah', 'di Nusantara', 'di daerah', 'di wilayah',
di Kabupaten', 'di yang', 'di dalamnya', 'di Pulau', 'di daerah', 'di cikal', 'di perdebatan', 'di kalangan', 'di wilayah', 'di wilayah', 'di wilayah', 'di Jawa', 'di Nusantara
', 'di salah', 'di Nusantara', 'di Asia', 'di Selat', 'di dunia', 'di daerah', 'di Panjalu', 'di Jawa', 'di suatu', 'di Nusantara', 'di Nusantara', 'di Nusantara', 'di Pulau', 'd
i bawah', 'di Nusantara', 'di pusat', 'di Nusantara', 'di Nusantara', 'di Aceh', 'di wilayah', 'di Nusantara', 'di antaranya', 'di Pulau', 'di Sumatra', 'di kesultanan', 'di wil
ayah', 'di wilayah', 'di Nusantara', 'di Jawa', 'di Pulau', 'di Jawa', 'di kesultanan', 'di Pulau', 'di Pulau', 'di Pulau', 'di kesultanan', 'di Sulawesi', 'di saat', 'di wilaya
h', 'di Sulawesi', 'di wilayah', 'di Kepulauan', 'di Kepulauan', 'di Papua', 'di wilayah', 'di Kabupaten', 'di Kabupaten', 'di Kepulauan', 'di Nusantara', 'di antaranya', 'di da
erah', 'di bawah', 'di kerajaan', 'di Sulawesi', 'di Pulau', 'di wilayah', 'di negara', 'di wilayah', 'di kerajaan', 'di salah', 'di Nusantara', 'di dunia', 'di bawah', 'di Kepu
lauan', 'di Melaka', 'di titik', 'di Nusantara', 'di Ternate', 'di situs', 'di Kepulauan', 'di bawah', 'di Kepulauan', 'di Maluku', 'di Maluku', 'di Maluku', 'di Nusantara', 'di
Banten', 'di Teluk', 'di Sunda', 'di Kepulauan', 'di cukup', 'di antara', 'di semakin', 'di tangan', 'di di', 'di Kepulauan', 'di Benua', 'di bawah', 'di perairan', 'di sepanja
ng', 'di antaranya', 'di Nusantara', 'di antara', 'di seluruh', 'di berkurang', 'di luar', 'di kawasan', 'di Banten', 'di Maluku', 'di Nusantara', 'di Nusantara', 'di antaranya
', 'di Sukadana', 'di dalam', 'di badan', 'di Sumatra', 'di beberapa', 'di Nusantara', 'di antaranya', 'di perusahaan', 'di dunia', 'di Pulau', 'di Mataran', 'di peristiwa', 'di
Batavia', 'di beberapa', 'di kalangan', 'di milik', 'di koloni', 'di bawah', 'di Hindia', 'di bawah', 'di semakin', 'di kekaisaran', 'di bawah', 'di milik', 'di gubernur', 'di B
atavia', 'di Jawa', 'di Jawa', 'di Surabaya', 'di Batavia', 'di penyebab', 'di Pulau', 'di Jawa', 'di Hindia', 'di seluruh', 'di anak', 'di Jawa', 'di Salatiga', 'di Tuntang',
di Nusantara', 'di perpanjangan', 'di Buitenzorg', 'di tubuh', 'di bidang', 'di Keraton', 'di Prambanan', 'di Sleman', 'di Borobudur', 'di Magelang', 'di Trowulan', 'di Pulau',
di Hindia', 'di Nusantara', 'di bukti', 'di Pulau', 'di korban', 'di korban', 'di penyebab', 'di kosong', 'di Nusantara', 'di Nusantara', 'di milik', 'di London', 'di Hindia',
di Nusantara', 'di bawah', 'di depan', 'di Ambon', 'di negeri', 'di negara', 'di Hindia', 'di koloni', 'di Nusantara', 'di Timur', 'di London', 'di Semenanjung', 'di Pulau', 'd
i Nusantara', 'di Malaya', 'di London', 'di Nusantara', 'di Pulau', 'di Jawa', 'di sekitar', 'di setiap', 'di Tegalarjo', 'di bawah', 'di Jawa', 'di Jawa', 'di tanah', 'di carut
', 'di ambang', 'di tanah', 'di Hindia', 'di Sumatra', 'di wilayah', 'di seputar', 'di Bukittinggi', 'di Batusangkar', 'di Kusamba', 'di Bali', 'di wilayah', 'di Buleleng', 'di
bawah', 'di Bali', 'di bawah', 'di Kusamba', 'di kerajaan', 'di negara', 'di Bore', 'di beberapa', 'di Pulau', 'di Bali', 'di wilayah', 'di Museum', 'di antara', 'di sultan', 'di
i antara', 'di pihak', 'di hinga', 'di Benua', 'di dekat', 'di Aceh', 'di pedalaman', 'di Tiro', 'di tanah', 'di bawah', 'di pos', 'di pada', 'di antara', 'di sekit
anya', 'di sekitar', 'di pada', 'di letusan', 'di antara', 'di perairan', 'di Australia', 'di sekitar', 'di sekitar', 'di selama', 'di dan', 'di wilayah', 'di antaranya', 'di Me
ulaboh', 'di pesisir', 'di kodiaman', 'di Bali', 'di wilayah', 'di Bali', 'di Bali', 'di tanah', 'di tanah', 'di Aceh', 'di pada', 'di mata', 'di tanah', 'di negeri', 'di Nusan
ara', 'di daerah', 'di keresidenan', 'di akibat', 'di Pantai', 'di perairan', 'di negara', 'di dalam', 'di tanah', 'di Nusantara', 'di negara', 'di Maluku', 'di Kalimantan', 'di
Sumatra', 'di bawah', 'di negara', 'di di', 'di kepada', 'di Hindia', 'di tanah', 'di Batavia', 'di sana', 'di sana', 'di Oetomo', 'di Oetomo', 'di Oetomo', 'di ke
tua', 'di Sarekat', 'di ranah', 'di ranah', 'di pengasingan', 'di Soerjaningrat', 'di Soerjaningrat', 'di Oetomo', 'di organisasi', 'di penyebab', 'di menuliskan', 'di ditangka
n', 'di Hindia', 'di Soerjaningrat', 'di Soerjaningrat', 'di Persarekatan', 'di Partij', 'di dua', 'di Partai', 'di Partai', 'di Partai', 'di Soerjaningrat', 'di anggota', 'di
i Indonesische', 'di Perhimpunan', 'di Hindia', 'di antara', 'di Vrijzetselaarsloge', 'di Gedung', 'di Batavia', 'di Labuan', 'di kediaman', 'di Jawa', 'di daerah', 'di tempat
', 'di Pulau', 'di silungkang', 'di keresidenan', 'di Jawa', 'di Hindia', 'di Wakil', 'di Hindia', 'di di', 'di dalam', 'di depan', 'di terkenal', 'di dapat',
di anggota', 'di sebuah', 'di bawah', 'di lokasi', 'di Indonesia', 'di pengajar', 'di Nederlandsch', 'di Surabaya', 'di Technische', 'di Bandung', 'di Umm', 'di Partij', 'di La
ndraad', 'di hadapan', 'di tempat', 'di tiga', 'di Batavia', 'di Gedung', 'di di', 'di Oost', 'di Indonesische', 'di lagu', 'di hadapan', 'di moelia', 'di depan', 'di ikrar', 'di Hindia
', 'di Ende', 'di lumpuh', 'di pemimpin', 'di masa', 'di Belanda', 'di anggota', 'di ketua', 'di anggota', 'di Pulau', 'di koran', 'di koran', 'di Yogyakarta', 'di Belanda', 'di
penyokong', 'di Eropa', 'di Belanda', 'di Semarang', 'di Bandung', 'di Sekolah', 'di Soerjaningrat', 'di Ki', 'di Yogyakarta', 'di bawah', 'di presiden', 'di lebih
', 'di saksi', 'di belakang', 'di berdasarkan', 'di bawah', 'di Pejabat', 'di luar', 'di dunia', 'di dalam', 'di bawah', 'di negara', 'di Yogyakarta', 'di Sidoarjo', 'di Indone
sia', 'di Taman', 'di dunia', 'di Asia', 'di antara', 'di antara', 'di sekitar', 'di Indonesia', 'di Indonesia', 'di Provinsi', 'di Sumatera', 'di Indonesia', 'di dunia', 'di Ind
```

4. 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.

```
...
<td>8.73</td>
<td>8.70</td>
<td>8.82</td>
<td>8.96</td>
<td>9.14</td>
<td>8.02</td>
<td>16</td>
</tr>
<tr>
...

```

- 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 di halaman 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)
```

```
PS D:\KULIAH\MATERI\SEMESTER 4\PRAKTIKUM ALGORITMA STRUKTUR DATA\modul7> & C:/Users/Acer/AppData/Local/Microsoft/WindowsApps/python3.10.exe "d:/KULIAH/MATERI/SEMESTER 4/PRAKTIKUM ALGORITMA STRUKTUR DATA/modul7/4.py"

--- Oleh L200220037 ---
[('Denmark', 9.57), ('Sweden', 9.79), ('Finland', 9.66), ('Netherlands', 9.48), ('Norway', 9.06), ('Canada', 9.43), ('Switzerland', 9.89), ('Kingdom', 9.18), ('States', 9.45), ('Australia', 8.72), ('Ireland', 9.04), ('Austria', 8.9), ('Iceland', 2.98), ('Germany', 9.0), ('Zealand', 8.65), ('Belgium', 8.96), ('Taiwan', 9.24), ('Luxembourg', 8.91), ('Japan', 9.15), ('France', 8.61), ('Estonia', 7.49), ('Slovenia', 8.31), ('Spain', 8.14), ('Singapore', 9.56), ('Israel', 9.34), ('China', 8.64), ('Italy', 8.04), ('Hungary', 8.14), ('Republic', 7.6), ('Lithuania', 6.59), ('Korea', 8.47), ('Latvia', 6.4), ('Cyprus', 7.65), ('Portugal', 7.43), ('Greece', 7.63), ('Poland', 6.92), ('Slovakia', 6.86), ('Barbados', 7.51), ('Croatia', 7.54), ('Chile', 6.81), ('Bulgaria', 6.43), ('Emirates', 6.74), ('Romania', 5.66), ('Uruguay', 5.26), ('Qatar', 5.77), ('Dominica', 3.76), ('Rica', 6.24), ('Malaysia', 6.83), ('Federation', 6.89), ('Bahrain', 4.2), ('Kuwait', 5.05), ('Ukraine', 5.77), ('Argentina', 6.85), ('Tobago', 6.02), ('Brazil', 6.07), ('Turkey', 5.67), ('Africa', 6.92), ('Jordan', 5.66), ('Armenia', 6.17), ('Mexico', 5.82), ('Thailand', 5.98), ('Oman', 4.95), ('Macedonia', 4.76), ('Mauritius', 3.7), ('Arabia', 4.04), ('Jamaica', 5.36), ('Moldova', 4.39), ('Kazakhstan', 3.77), ('Belarus', 5.54), ('Lebanon', 4.69), ('Tunisia', 4.58), ('Panama', 5.45), ('Georgia', 5.38), ('Peru', 3.88), ('Mongolia', 2.06), ('Colombia', 4.26), ('China', 5.12), ('Guyana', 4.47), ('Philippines', 3.63), ('Venezuela', 5.73), ('Namibia', 3.3), ('Lanka', 4.44), ('Albania', 3.1), ('Egypt', 4.55), ('Botswana', 4.34), ('Republic', 2.91), ('Salvador', 3.19), ('Azerbaijan', 3.05), ('Kyrgyzstan', 2.7), ('Paraguay', 3.47), ('Ecuador', 3.55), ('Morocco', 3.67), ('Bolivia', 3.05), ('Iran', 3.02), ('Uzbekistan', 3.51), ('Algeria', 3.48), ('Verde', 2.25), ('Indonesia', 3.32), ('Honduras', 3.3), ('India', 3.97), ('Guatemala', 2.47), ('Vietnam', 2.83), ('Swaziland', 4.55), ('Republic', 3.44), ('Nicaragua', 1.99), ('Kenya', 3.87), ('Tajikistan', 2.33), ('Senegal', 2.77), ('Zimbabwe', 4.09), ('Ghana', 2.08), ('Uganda', 2.72), ('Madagascar', 2.54), ('Mauritania', 1.75), ('Tanzania', 2.39), ('Pakistan', 2.75), ('Lesotho', 2.7), ('Benin', 2.33), ('Nigeria', 2.72), ('Yemen', 1.68), ('Mali', 1.69), ('Mozambique', 1.86), ('Angola', 2.44), ('Cameroon', 2.49), ('Faso', 2.15), ('Nepal', 2.04), ('Malawi', 2.11), ('Laos', 1.43), ('Bangladesh', 1.71), ('Myanmar', 1.17), ('Rwanda', 1.47), ('Ethiopia', 1.57), ('Djibouti', 1.29), ('Eritrea', 1.56), ('Leone', 1.7), ('Herzegovina', 3.29), ('Montenegro', 4.85), ('Haiti', 1.15), ('Ivoire', 2.52), ('Sudan', 1.97), ('Zambia', 2.37)]
PS D:\KULIAH\MATERI\SEMESTER 4\PRAKTIKUM ALGORITMA STRUKTUR DATA\modul7>
```