

MODUL 1: PENGENALAN INFORMATION RETRIEVAL DENGAN PYTHON

1.1 Deskripsi Singkat

Information retrieval adalah pencarian suatu materi (biasanya dokumen) dalam bentuk tidak terstruktur (biasanya teks) untuk memenuhi kebutuhan informasi dari sekumpulan besar (biasanya tersimpan di komputer). Contoh pengaplikasiannya diantaranya yaitu mesin pencarian pada website.

1.2 Tujuan Praktikum

Setelah praktikum pada modul 1 ini diharapkan mahasiswa mempunyai kompetensi sebagai berikut:

- 1) Dapat memahami konsep dasar information retrieval dan contoh aplikasinya.
- 2) Dapat memanfaatkan Anaconda sebagai tools untuk manajemen *environment* dan *package*.
- 3) Dapat membuat file python sederhana dan menjalankannya.
- 4) Dapat membuat program untuk membaca file di dalam suatu folder dan memfilter file dengan query tertentu.

1.3 Material Praktikum

Pada kegiatan modul 1 diperlukan beberapa material berupa file, yaitu:

- 1) File installer *text editor* Notepad++ atau Sublime
- 2) File installer Anaconda
- 3) Folder "berita" yang berisi beberapa file text

1.4 Kegiatan Praktikum

A. Pengantar Python

Python adalah bahasa pemrograman yang banyak digunakan oleh *data scientist*. Adapun kelebihan dari Python yaitu sebagai berikut.

1. Mudah dipahami
Python memiliki sintaks sederhana yang mirip dengan bahasa Inggris sehingga mudah dipahami.
2. Skalabilitas
Python berjalan di sistem interpreter tanpa proses kompilasi, yang berarti kode program akan dieksekusi langsung setelah ditulis. Oleh karena itu, python dapat mempercepat proses *prototyping*.
3. Pilihan *library*
Python memiliki banyak pilihan *library* yang mempermudah analisis data dan pengembangan model dengan *machine learning*, seperti pandas, numpy, scipy, dan

scikit-learn. Selain itu, python juga banyak memiliki library yang mempermudah pengembangan *deep neural network* seperti tensorflow, keras, dan pytorch.

4. Grafik dan visualisasi

Python menyediakan beragam *library* untuk visualisasi data seperti matplotlib dan seaborn.

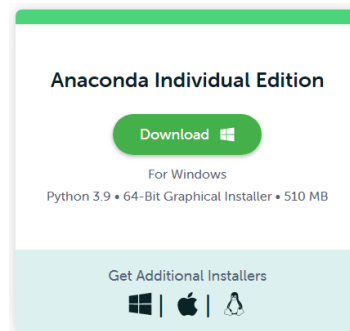
B. Instalasi Python melalui Anaconda

Untuk memudahkan manajemen *package* dan *environment* dalam pemrograman Python, pada praktikum ini digunakan Anaconda dalam proses instalasinya. Anaconda adalah *package distribution* yang dibangun untuk *data science*. Kelebihan utama Anaconda diantaranya dapat membuat environment yang mengisolasi proyek tertentu dengan Python yang berbeda versi ataupun *package* yang berbeda versi. Anaconda mencakup Python dan beragam *scientific package* beserta dependensinya sehingga membutuhkan *disk space* cukup besar untuk diinstal. Bagi yang terkendala *disk space*, dapat menggunakan Miniconda sebagai alternatif.

File instalasi anaconda untuk versi individual (*open source*) dapat ditemukan di halaman resminya menggunakan link berikut.

<https://www.anaconda.com/products/individual>

Setelah itu, klik tombol Download atau pilih file instalasi yang sesuai sistem operasi yang digunakan: Windows, MacOS, atau Linux.



Kemudian lanjutkan dengan melakukan instalasi Anaconda dengan mengklik file instalasi yang telah didownload hingga selesai.

Jika menggunakan Windows, command prompt untuk Anaconda dapat ditemukan dengan mengetikkan “Anaconda Prompt” pada kotak pencarian. Kemudian jalankan Anaconda Prompt sehingga muncul jendela berikut.



Untuk system operasi lain, command untuk Anaconda dapat langsung dijalankan di Terminal. Untuk mengecek versi conda dan python yang terinstal, jalankan command berikut.

```
conda -version
```

```
python -version
```

Agar proyek yang dikerjakan terisolasi dari proyek lainnya, pembuatan environment perlu dilakukan. Pada Anaconda Prompt, ketikkan command untuk membuat environment baru dengan versi Python tertentu, misalnya Python 3.8 dengan nama environment "ir_env".

```
conda create --name ir_env python=3.8
```

Lanjutkan dengan mengetikkan "y" sebagai konfirmasi bahwa akan menginstal Python 3.8 dengan daftar package sesuai dengan yang tertulis.

Untuk mengaktifkan environment tersebut, jalankan command:

```
conda activate ir_env
```

Kata dalam tanda kurung berubah dari "base" yang merujuk versi original environment, menjadi "ir_env" sebagai environment yang saat ini sedang diaktifkan.



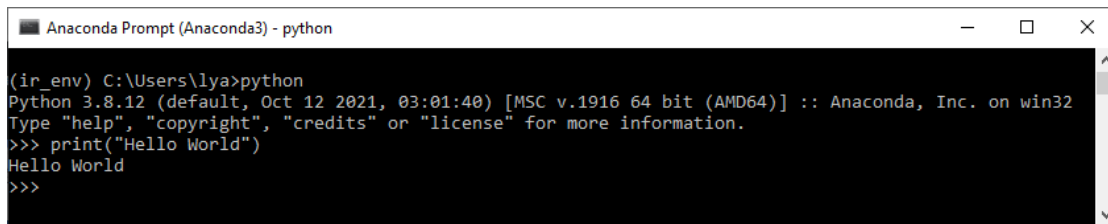
Untuk menonaktifkan environment yang saat ini sedang dijalankan, jalankan command berikut.

```
conda deactivate
```

C. Membuat dan Menjalankan kode Python

Kode python dapat langsung dibuat pada jendela *command prompt* dengan mengetikkan perintah "python". Baris perintah selanjutnya akan dianggap sebagai kode python, ditandai dengan ">>>", dan akan langsung dijalankan ketika menekan tombol Enter. Contohnya

adalah mengetikkan kode untuk mencetak *string* "Hello World" seperti gambar di bawah ini.



```
Anaconda Prompt (Anaconda3) - python
(in_env) C:\Users\lya>python
Python 3.8.12 (default, Oct 12 2021, 03:01:40) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello World")
Hello World
>>>
```

Untuk keluar dari kode python, ketikkan `exit()`.

Selanjutnya, kita akan membuat kode python di file terpisah yang berekstensi .py. Gunakan sublime sebagai *text editor*, atau dapat menggunakan notepad++ sebagai alternatif. Ketikkan kembali baris kode untuk mencetak string di file tersebut.

```
print("Hello World")
```

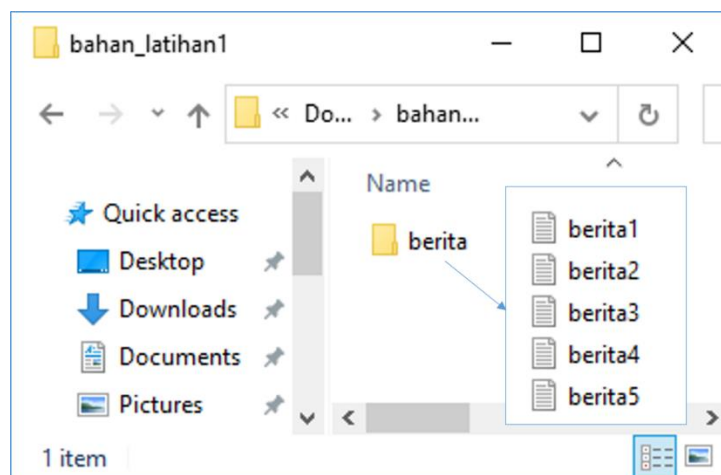
Simpan file dan jalankan file python di atas dengan mengetikkan command berikut di Anaconda Command Prompt.

```
python nama_file.py
```

Jangan lupa untuk memindahkan path ke direktori tempat file python disimpan dengan perintah `cd direktori_tujuan` sebelum menjalankan kode di atas.

D. Kode untuk Membaca File di dalam Folder

Buat file baru, misalnya latihan1_1.py kemudian buat kode python untuk membaca file di dalam folder "berita".



Dalam python, suatu baris dianggap komentar jika diawali dengan #. Jika ingin membuat beberapa baris komentar, awali baris pertama komentar dan akhiri baris terakhirnya dengan `"""`. Berikut adalah contoh dari penulisan komentar.

```
# ini komentar satu baris

"""ini baris komentar 1
baris komentar 2
...
baris terakhir komentar"""
```

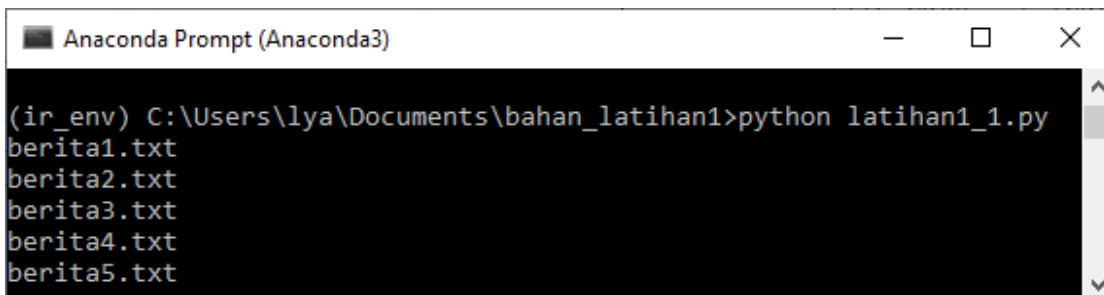
Berikut adalah fungsi untuk mencetak nama file dari suatu folder dengan meng-import terlebih dahulu module `os`.

```
# Import Module
import os

# Folder Path
path = "Enter Folder Path"

# List all files in a directory using os.listdir
for file in os.listdir(path):
    if os.path.isfile(os.path.join(path, file)):
        print(file)
```

Jalankan file python `latihan1_1.py` di atas dengan mengetikkan command berikut di Anaconda Command Prompt sehingga menampilkan output sebagai berikut.



```
Anaconda Prompt (Anaconda3)
(ir_env) C:\Users\lya\Documents\bahan_latihan1>python latihan1_1.py
berita1.txt
berita2.txt
berita3.txt
berita4.txt
berita5.txt
```

Kemudian ubah kode pada `latihan1_1.py`, tambahkan fungsi untuk membaca file `.txt` yang ada di suatu folder. Fungsi dalam python dibuat dengan perintah `def`. Berikut adalah kode fungsinya.

```
def read_text_file(file_path):
    with open(file_path, 'r') as f:
        print(f.read())
```

Panggil fungsi tersebut untuk membaca isi semua file yang ada pada folder "berita".

```
# iterate through all file
for file in os.listdir(path):
    # Check whether file is in text format or not
    if file.endswith(".txt"):
        file_path = f"{path}\\{file}"

        # call read text file function
        read_text_file(file_path)
```

Kemudian jalankan kembali latihan1_1.py pada Anaconda Command Prompt, sehingga mencetak semua isi file pada folder berita.

E. Kode untuk Mencari Informasi dengan Query Tertentu

Secara sederhana, *information retrieval* dapat diterapkan untuk mencari informasi dengan query tertentu. Jalankan kode berikut untuk mengecek apakah suatu query ada dalam string tertentu.

```
text_1 = "Wilayah Kamu Sudah 'Bebas' COVID-19? Cek 34 Kab/Kota  
Zona Hijau Terbaru"  
text_2 = "Vaksin COVID-19 Bakal Rutin Setiap Tahun? Tergantung,  
Ini Penjelasannya"  
text_3 = "RI Mulai Suntikkan Booster di 2022, Masihkah Ampuh  
Lawan Varian Delta Cs?"  
query = "COVID-19"  
docs = [text_1, text_2, text_3]  
for doc in docs:  
    if query in doc:  
        print(doc)
```

Selain menggunakan kode di atas, Regular Expression (regex) dapat juga digunakan untuk melakukan pencarian dengan kode berikut.

```
import re  
for doc in docs:  
    if re.search(query, doc):  
        print(doc)
```

1.5 Penugasan

1. Buat kode python untuk menampilkan daftar nama dokumen pada folder "berita" yang terdapat query "corona".