

Dasar-Dasar Pemrograman 1

Tugas Pemrograman 02

Deadline: 13 Oktober 2022 23:55 Waktu SCeLe



FAKULTAS
ILMU
KOMPUTER

Tujuan

- Memperoleh pengalaman praktis dari topik-topik yang sudah dipelajari terkait operasi string, f-string, baca/tulis file teks, exception, dan konsep fungsi.
- Eksplorasi library python yang tidak disampaikan di kelas namun masih terkait seperti library **os** untuk melakukan **scanning** pada sebuah direktori atau folder (<https://docs.python.org/3/library/os.html>).

Pendahuluan

Sistem operasi Unix/Linux mempunyai perintah **grep** yang bisa mencari sebuah pola string ke sebuah file atau ke seluruh file yang ada di sebuah direktori. Sebagai gambaran, peserta bisa mempelajari terlebih dahulu melalui link berikut:

1. <https://www.geeksforgeeks.org/grep-command-in-unixlinux/>
2. <https://www.gnu.org/software/grep/manual/grep.html>
3. [OS Module in Python with Examples - GeeksforGeeks](#) (`chdir()`, `path.exists()`, `path.isfile()` mungkin akan berguna)

Penjelasan Program

Pada Tugas Pemrograman 02 kali ini, peserta akan membuat simulasi dari perintah **grep** tersebut menggunakan bahasa pemrograman Python dengan sedikit modifikasi atau versi yang lebih sederhana.

! Petunjuk: Download file mydir.zip dan mydir2.zip kemudian ekstrak zip masing-masing untuk dapat digunakan sebagai contoh direktori TP ini. Link download: [Direktori TP2](#)

Perhatikan isi dari folder **mydir** berikut:

```
mydir
|---- file1.txt
|---- olahraga
|        |---- bulutangkis
|        |        |----- file2.txt
|        |        |----- file3.txt
|        |---- sepakbola
|        |        |----- file4.txt
|---- komputasi
|        |---- file5.txt
```

Direktori **mydir** bisa digunakan untuk menguji fungsionalitas program peserta. Program Python yang peserta kembangkan nantinya dapat dieksekusi dengan format berikut:

```
python grep.py [options (-w / -i)] [string pattern yang dicari] [nama file / direktori]
```

Biru : Opsional (pilih salah satu atau tidak ada sama sekali; tidak boleh dua-duanya)

Merah: Wajib ada

Catatan: Sistem berbasis Unix tidak akan menerima karakter **backslash** (\) sebagai argumen ketika mengeksekusi program. Silahkan menggunakan **forward slash** (/) sebagai separator direktori alternatif. ([Sumber](#))

Tanpa argumen options

Program akan melakukan proses pencarian pola string secara **case sensitive** sebagai **substring** ke semua baris pada sebuah file atau pada semua file di sebuah direktori. Contoh:

```
python grep.py "mat pag" mydir\file1.txt
```

```
mydir\file1.txt          line 1  Halo selamat pagi semoga sehat selalu
mydir\file1.txt          line 4  Selamat pagi untuk kita semua
```

```
python grep.py "mat pag" mydir
```

```
mydir\file1.txt          line 1  Halo selamat pagi semoga sehat selalu
mydir\file1.txt          line 4  Selamat pagi untuk kita semua
mydir\olahraga\sepakbola\file4.txt line 3  Ucapan selamat pagi dari pemain sepakbol
```

Untuk setiap baris yang ditemukan, format output yang dicetak di layar adalah sebagai berikut (gunakan **f-string**):

[lokasi file] line **[line number]** **[baris pada file]**

- [lokasi file] dialokasikan sebanyak **40** karakter dan dicetak **left-aligned**,
- [line number] dialokasikan sebanyak **3** karakter dan dicetak **left-aligned**; [line number] dimulai dari satu (one-based)
- [baris pada file] hanya mencetak **40** karakter pertama pada baris dan dicetak **left-aligned**.



Hint : Pola string yang dicari bisa saja terpotong dari 40 karakter pertama pada baris, sehingga tidak muncul di output.

Argumen -w

Program akan melakukan proses pencarian pola string secara **case sensitive** sebagai **sebuah kata penuh (whole word)** ke semua baris pada sebuah file atau pada semua file di sebuah direktori. Definisi **kata penuh** secara umum adalah sebuah *substring* yang berbatasan langsung dengan **whitespace** baik di depan maupun di belakang (termasuk substring paling kiri yang berbatasan dengan *whitespace* di belakang dan substring paling kanan yang berbatasan dengan *whitespace* di depan). Sebagai contoh:

```
python grep.py -w "olahraga" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 1 | Bulutangkis adalah olahraga yang menggun |
| mydir\olahraga\bulutangkis\file2.txt | line 2 | Bulutangkis adalah olahraga rakyat indon |
| mydir\olahraga\bulutangkis\file2.txt | line 3 | Cabang olahraga yang dipertandingkan di |
| mydir\olahraga\bulutangkis\file2.txt | line 5 | Bulutangkis adalah cabang olahraga |
| mydir\olahraga\bulutangkis\file3.txt | line 1 | Indonesia berhasil menyumbang medali ema |
| mydir\olahraga\bulutangkis\file3.txt | line 3 | Bulutangkis merupakan olahraga ketangkas |
| mydir\olahraga\sepakbola\file4.txt | line 2 | Sepakbola adalah olahraga yang dimainkan |

```
python grep.py -w "mat pag" mydir
```

Perhatikan bahwa dengan string pattern **"mat pag"**, program tidak menemukan satu baris pun yang cocok jika menggunakan mode **whole world**. Jika tidak ada satu baris pun yang cocok, program tidak menampilkan apa-apa di layar (hanya baris kosong).

Argumen -i

Program akan melakukan proses pencarian pola string secara **case insensitive** sebagai **substring** ke semua baris pada sebuah file atau pada semua file di sebuah direktori. Sebagai contoh:

```
python grep.py -i "PaGi" mydir\file1.txt
```

| | | |
|-----------------|--------|---------------------------------------|
| mydir\file1.txt | line 1 | Halo selamat pagi semoga sehat selalu |
| mydir\file1.txt | line 4 | Selamat pagi untuk kita semua |

```
python grep.py -i "MaT PaG" mydir\file1.txt
```

| | | |
|-----------------|--------|---------------------------------------|
| mydir\file1.txt | line 1 | Halo selamat pagi semoga sehat selalu |
| mydir\file1.txt | line 4 | Selamat pagi untuk kita semua |

```
python grep.py -i "olahraga" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 1 | Bulutangkis adalah olahraga yang menggun |
| mydir\olahraga\bulutangkis\file2.txt | line 2 | Bulutangkis adalah olahraga rakyat indon |
| mydir\olahraga\bulutangkis\file2.txt | line 3 | Cabang olahraga yang dipertandingkan di |
| mydir\olahraga\bulutangkis\file2.txt | line 4 | Olahraga bulutangkis sangat populer di i |
| mydir\olahraga\bulutangkis\file2.txt | line 5 | Bulutangkis adalah cabang olahraga |
| mydir\olahraga\bulutangkis\file3.txt | line 1 | Indonesia berhasil menyumbang medali ema |
| mydir\olahraga\bulutangkis\file3.txt | line 3 | Bulutangkis merupakan olahraga ketangkas |
| mydir\olahraga\sepakbola\file4.txt | line 1 | Olahraga sepakbola paling favorit |
| mydir\olahraga\sepakbola\file4.txt | line 2 | Sepakbola adalah olahraga yang dimainkan |

Coba Anda perhatikan dan bandingkan pencarian string pattern **"olahraga"** dengan mode **-i** dan mode **-w**.

Pola String

Tanpa wildcard

Pola string berikut merupakan bentuk pattern biasa seperti "mat pag", "olahraga", "halo", dsb.

Dengan wildcard (*)

Pola string juga bisa mengandung **sebuah** karakter wildcard (*). Ingat, **hanya sebuah karakter wildcard**. Tidak boleh ada lebih dari satu karakter (*). Contohnya adalah "**bul*kis**". Jika tidak sesuai dengan ketentuan ini, program akan berhenti dan menampilkan pesan kesalahan "**Argumen program tidak benar.**"

Pola string **X*Y** akan **match** dengan baris yang mengandung **X** kemudian diikuti oleh beberapa karakter (boleh juga string kosong) dan diikuti lagi dengan **Y**. Contoh "**bul*kis**" *match* dengan baris "**taufik hidayat adalah atlit bulutangkis terkenal**"; "**pag*sem**" *match* dengan baris "**Selamat pagi untuk kita semua**" (untuk mode tanpa options atau dengan -i); namun "**pag*sem**" **tidak match** dengan baris "**Selamat pagi untuk kita semua**" jika mode adalah -w.

Perhatikan bahwa **X*** dan ***X** sama dengan **X**.

```
python grep.py -w "Bul*kis" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 1 | Bulutangkis adalah olahraga yang menggun |
| mydir\olahraga\bulutangkis\file2.txt | line 2 | Bulutangkis adalah olahraga rakyat indon |
| mydir\olahraga\bulutangkis\file2.txt | line 5 | Bulutangkis adalah cabang olahraga |
| mydir\olahraga\bulutangkis\file3.txt | line 2 | Bulutangkis menyumbang 1 medali emas di |
| mydir\olahraga\bulutangkis\file3.txt | line 3 | Bulutangkis merupakan olahraga ketangkas |

```
python grep.py -w "bul*kis" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 3 | Cabang olahraga yang dipertandingkan di |
| mydir\olahraga\bulutangkis\file2.txt | line 4 | Olahraga bulutangkis sangat populer di i |
| mydir\olahraga\bulutangkis\file3.txt | line 1 | Indonesia berhasil menyumbang medali ema |

```
python grep.py -w "Bulutangkis*" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 1 | Bulutangkis adalah olahraga yang menggun |
| mydir\olahraga\bulutangkis\file2.txt | line 2 | Bulutangkis adalah olahraga rakyat indon |
| mydir\olahraga\bulutangkis\file2.txt | line 5 | Bulutangkis adalah cabang olahraga |
| mydir\olahraga\bulutangkis\file3.txt | line 2 | Bulutangkis menyumbang 1 medali emas di |
| mydir\olahraga\bulutangkis\file3.txt | line 3 | Bulutangkis merupakan olahraga ketangk |

```
python grep.py "hari*berjalan" mydir
```

| | | |
|-----------------|--------|--|
| mydir\file1.txt | line 2 | Semoga hari ini berjalan dengan menyenangkan |
|-----------------|--------|--|

```
python grep.py -i "Ol*gA" mydir
```

| | | |
|--------------------------------------|--------|--|
| mydir\olahraga\bulutangkis\file2.txt | line 1 | Bulutangkis adalah olahraga yang menggun |
| mydir\olahraga\bulutangkis\file2.txt | line 2 | Bulutangkis adalah olahraga rakyat indon |
| mydir\olahraga\bulutangkis\file2.txt | line 3 | Cabang olahraga yang dipertandingkan di |
| mydir\olahraga\bulutangkis\file2.txt | line 4 | Olahraga bulutangkis sangat populer di i |
| mydir\olahraga\bulutangkis\file2.txt | line 5 | Bulutangkis adalah cabang olahraga |
| mydir\olahraga\bulutangkis\file3.txt | line 1 | Indonesia berhasil menyumbang medali ema |
| mydir\olahraga\bulutangkis\file3.txt | line 3 | Bulutangkis merupakan olahraga ketangk |
| mydir\olahraga\sepakbola\file4.txt | line 1 | Olahraga sepakbola paling favorit |
| mydir\olahraga\sepakbola\file4.txt | line 2 | Sepakbola adalah olahraga yang dimainkan |

Direktori mydir2

Selain direktori **mydir**, dokumen tugas juga disertai dengan **mydir2** yang berisi **1033 dokumen** berbahasa Inggris dari **Medline** yang berasal dari domain kesehatan atau medis (http://ir.dcs.gla.ac.uk/resources/test_collections/medl/). Peserta juga bisa menguji program atau melakukan eksperimen dengan koleksi yang lebih besar ini. Sebagai contoh:

```
python grep.py -w "albuminoid" mydir2
```

| | | |
|------------------|---------|--|
| mydir2\2\165.txt | line 3 | of p-32 into the albuminoid rna fraction |
| mydir2\2\165.txt | line 16 | p-32 and c-14-adenine into albuminoid rn |
| mydir2\2\165.txt | line 19 | the effect of formaldehyde and heating o |
| mydir2\2\165.txt | line 20 | was much less than on the albuminoid rna |
| mydir2\2\183.txt | line 3 | of p-32 into the albuminoid rna fraction |
| mydir2\2\183.txt | line 16 | p-32 and c-14-adenine into albuminoid rn |
| mydir2\2\183.txt | line 19 | the effect of formaldehyde and heating o |
| mydir2\2\183.txt | line 20 | was much less than on the albuminoid rna |
| mydir2\6\511.txt | line 1 | 1747. the problem of albuminoid |
| mydir2\6\511.txt | line 2 | albuminoid is the main constituent of th |
| mydir2\6\511.txt | line 4 | albuminoid from the cortex of ox lens ha |
| mydir2\6\511.txt | line 9 | albuminoid can be regarded as an insolub |
| mydir2\6\511.txt | line 12 | albuminoid is present in the normal lens |
| mydir2\6\511.txt | line 15 | perhaps more likely that the albuminoid |

Format Output Jika Error

Format argumen tidak sesuai

Print "Argumen program tidak benar." dan berhentikan program.

Path tidak ditemukan

Print "Path <nama_file / nama_direktori> tidak ditemukan" dan berhentikan program.

Petunjuk

1. Fungsi `walk()` pada library `os` dapat digunakan untuk mengunjungi semua files yang berada pada sebuah direktori/folder.
2. Argumen program seperti `[-w` atau `-i]`, `[string pattern]`, dan `[nama file / nama folder]` dapat diakses menggunakan `sys.argv` pada library `sys` (<https://www.geeksforgeeks.org/how-to-use-sys-argv-in-python/>).
3. Peserta perlu membangun program secara modular dengan menggunakan **konsep fungsi**. Peserta bisa membuat beberapa fungsi seperti, misal fungsi `print_line(...)` yang bertugas mencetak satu baris yang match dengan pola string di layar; fungsi `scan_file(...)` yang bertugas melakukan pemeriksaan ke setiap baris di suatu file dan memanggil `print_line(...)` jika suatu baris match dengan pola string; dan fungsi-fungsi yang lainnya.

Komponen Penilaian

Sesuai dengan ketentuan yang ada di [Rubrik Penilaian TP DDP1](#).

Plagiarisme tidak akan ditoleransi. Anda diperbolehkan berdiskusi dengan teman terkait ide implementasi tugas ini. Harap menuliskan kolaborator jika berdiskusi dengan mahasiswa lain. Perlu diingat bahwa **implementasi kode dilakukan secara individu**. Tim pengajar akan melakukan *code similarity checking* pada implementasi kode mahasiswa, dan jika terbukti melakukan kecurangan/plagiarisme akan mendapat **sanksi berupa nilai 0** pada TP ini.

Pengumpulan

Berkas yang perlu dikumpulkan:

- [Kelas]_[KodeAsdos]_[NPM]_[NamaLengkap]_TP02.py

Kumpulkan berkas .py yang **telah di-zip** dengan format penamaan seperti berikut.

[Kelas]_[KodeAsdos]_[NPM]_[NamaLengkap]_TP02.zip

Contoh:

A_ABC_2206123456_ThamiEndamora_TP02.zip