

Warning

This page is located in archive. Go to the latest version of this [course pages](#). Go the latest version of [this page](#).

HW 07 - Hledání textu v souborech

Termín odevzdání	04.12.2021 23:59 PST
Povinné zadání	3b
Volitelné zadání	3b
Bonusové zadání	Není
Počet uploadů	10
Podpůrné soubory	b0b36prp-hw07.zip [/b211/_media/courses/b0b36prp/hw/b0b36prp-hw07.zip]

Povinné zadání

Bez použití knihovny `<string.h>` (nebo podobné) napište program, který ze zadaného souboru vypíše všechny řádky obsahující zadaný řetězec. Budete tak implementovat zjednodušenou verzi programu `grep` pro příkazový řádek, který je dostupný ve většině operačních systémů Unixového typu.

Program bude volán ve tvaru:

```
./program PATTERN FILE
```

kde PATTERN je hledaný řetězec, FILE název souboru.

V případě jakékoli chyby vypíše smysluplný důvod chyby na standardní chybový výstup a program bude ukončen s nenulovou návratovou hodnotou.

Program implementujte v souboru `grep.c`

Pro rozpoznání konce řádku je možné používat znak `'\n'`.

V případě prázdného výstupu ukončete program s návratovou hodnotou 1.

Příklad 1

První příklad je motivován praktickou ukázkou, jak v Ubuntu zjistit velikost volné operační paměti z příkazové řádky.

```
grep MemFr /proc/meminfo
```

```
./program MemFr /proc/meminfo
```

Obsah vstupního souboru	Řetězec PATTERN	Očekávaný výstup	Očekávaný chybový výstup	Návratová hodnota
MemTotal: 16430396 kB MemFree: 4848016 kB MemAvailable: 11697864 kB Buffers: 1139932 kB Cached: 5662728 kB ...	MemFr	MemFree: 4848016 kB	žádný	0

Příklad 2

Prázdný výstup

```
grep NeexistujiciText /proc/meminfo
```

Obsah vstupního souboru	Řetězec PATTERN	Očekávaný výstup	Očekávaný chybový výstup	Návratová hodnota
<pre>MemTotal: 16430396 kB MemFree: 4848016 kB MemAvailable: 11697864 kB Buffers: 1139932 kB Cached: 5662728 kB ...</pre>	NeexistujiciText	žádný	žádný	1

Volitelné zadání

Ve volitelné části můžete přidat tři různé funkcionality, za které můžete získat body. Počet argumentů tedy již nebude konstantní a program bude volán ve tvaru:

```
./program [OPTIONS] PATTERN [FILE]
```

kde PATTERN je hledaný řetězec, FILE název souboru a OPTIONS jsou volitelné argumenty pro volitelné zadání.

A - Přesměrování na standardní vstup (0.5b)

Pokud nebude zadán argument volitelný argument FILE, bude jako vstup použit **stdin**. Díky tomu je možné přesměrovat výstup jiného programu a filtrovat tak jeho výstup.

```
dmesg | grep dvd
```

Příklad výstupu:

```
[ 2.653604] sr 5:0:0:0: [sr0] scsi3-mmc drive: 48x/48x writer dvd-ram cd/rw xa/form2 cdda tray
```

B - Regulární výrazy (1b)

Pouze speciální symboly `[*,+,?]` bez závorekání. Speciální symbol se vyskytuje maximálně jednou.

```
./program -E PATTERN [FILE]
```

Regulární výrazy [https://en.wikipedia.org/wiki/Regular_expression]

Příklad 2

Obsah vstupního souboru	Řetězec 'colou?r'	Řetězec 'colou*r'	Řetězec 'colou+r'
<pre>color colour colouur colouuur</pre>	<pre>color colour</pre>	<pre>color colour colouur colouuur</pre>	<pre>colour colouur colouuur</pre>

C - Obarvení výstupu (1.5b)

Obarvení PATTERN ve výstupu červenou barvou, stejně jako originální grep. Text je možné obarvit v terminálu pomocí přidání přesné sekvence znaku (Escape sequence) před a za obarvovaný text, viz. [ANSI escape code](https://en.wikipedia.org/wiki/ANSI_escape_code) [https://en.wikipedia.org/wiki/ANSI_escape_code]. Tato funkcionality se zapíná následujícím přepínačem:

```
./program --color=always PATTERN [FILE]
```

Používejte prosím stejné escape sekvence jako systémový grep, aby bylo možné výstup jednoduše zkontrolovat. ESC reprezentuje speciální znak 0x1b (27 dekadicky).^[1]

	Před obarvený test	Za obarvený text
Textový zápis	ESC[01;31m + ESC[K	ESC[m + ESC[K
Hexadecimální zápis	1b 5b 30 31 3b 33 31 6d + 1b 5b 4b	1b 5b 6d + 1b 5b 4b

Příklad pub05-C

Obsah vstupního souboru	Řetězec PATTERN	Očekávaný výstup	Očekávaný výstup (hexdump -C)
Hello, hello Bye	el	Hello, hello	00000000 48 1b 5b 30 31 3b 33 31 6d 1b 5b 4b 65 6c 1b 5b H. [01;31m. [Ke1. [00000010 6d 1b 5b 4b 6c 6f 2c 20 68 1b 5b 30 31 3b 33 31 m. [K1o, h. [01;31 00000020 6d 1b 5b 4b 65 6c 1b 5b 6d 1b 5b 4b 6c 6f 0a m. [Ke1. [m. [K1o. 0000002f

Varianty

Veřejné příklady + Makefile: [b0b36prp-hw07.zip](#) [/b211/_media/courses/b0b36prp/hw/b0b36prp-hw07.zip]

	Povinné zadání	Volitelné zadání
Název v BRUTE	HW07	
Odevzdávané soubory	grep.c	
Argumenty při spuštění	žádné	- -color=always -E
Kompilace pomocí	clang -pedantic -Wall -Werror -std=c99	
Očekávaná časová složitost ²⁾	$\mathcal{O}(n)$	$\mathcal{O}(n^2)$ ³⁾

¹⁾
Doplněno 2. 12. 2019 na základě studentského dotazu.

²⁾
Kde n je velikost vstupního souboru.

³⁾
Složitost je dána použitím reulárních výrazů, jinak $\mathcal{O}(n)$.