

Примерни задачи за курс “Операционни системи”, СУ, ФМИ

19 февруари 2023 г.

Семинарните упражнения на курса “Операционни системи” разглеждат следните теми, групирани по “Работа в GNU/Linux shell” и “Използване на системни примитиви в програми на C”:

1. Въведение (shell)
2. Файлова система и работа с файлове (shell)
3. Обработка на текст (shell)
4. Процеси (shell)
5. Командни интерпретатори и скриптове (shell)
6. Системни примитиви за вход и изход (C)
7. Системни примитиви за работа с процеси (C)
8. Системни примитиви за работа с pipe-ове (C)

Тук са събрани примерни практически задачи по различните теми, както и някои теоретични въпроси, с надеждата те да бъдат полезни на студентите при тяхната работа в курса. За практическите задачи (shell и C) се очаква студентите да са разгледали някои по-базови задачи по дадена тема преди да преминат към изложените тук.

Съдържание

1 Работа в GNU/Linux shell

Забележка: За всички задачи, освен ако не е указано друго, в имената на файловете и директориите няма специални символи. Във файловата система може да съществуват директории, до които нямате достъп.

1.1 Поточна обработка (теми 1,2,3)

Зад. 1 2016-SE-01 Даден е текстов файл с име `philip-j-fry.txt`. Напишете shell script и/или серия от команди, които извеждат броя редове, съдържащи поне една четна цифра и несъдържащи малка латинска буква от `a` до `w`.

Примерно съдържание на файла:

```
123abv123
123zz123
MMU_2.4
```

Примерен изход:

Броят на търсените редове е 2

Зад. 2 2017-IN-01 Напишете серия от команди, извеждащи на екрана само броя на всички обекти във файловата система, чиито собственик е текущият потребител.

Забележка: Във файловата система със сигурност съществуват директории, до които нямате достъп.

Зад. 3 2017-IN-02 Напишете серия от команди, които изтриват:

- а) всички файлове в текущата директория и нейните поддиректории, които са с нулева дължина.
- б) 5-е най-големи файла в `home` директорията на текущия потребител и нейните поддиректории.

Зад. 4 2017-IN-03 Напишете серия от команди, които от файла `/etc/passwd` да вземат под-низ, състоящ се от втора и трета цифра на факултетния номер на студентите от специалност Информатика, чиито фамилии завършват на "а". Изведете коя комбинация от цифри се среща най-често и коя е тя.

Примерно съдържание на файла:

```
s45194:x:1255:502:Elizabet Mihaylova, Inf, k3, g1:/home/Inf/s45194:/bin/bash
s45139:x:1261:502:Vasilena Peycheva:/home/Inf/s45139:/bin/bash
s81257:x:1079:503:Vasilena Nikolova, KN, 2kurs, 5gr:/home/KN/s81257:/bin/bash
s81374:x:1117:503:Ivan Kamburov, KN, 2kurs, 7gr:/home/KN/s81374:/bin/bash
kiril:x:508:500:Kiril Varadinov:/home/kiril:/bin/bash
s61812:x:1128:504:Vladimir Genchev:/home/SI/s61812:/bin/bash
user:x:1000:99:Inactive user just to start UID from 1000:/home/user:/sbin/nologin
s81254:x:1077:503:Mariela Tihova, KN, 2kurs, 5gr:/home/KN/s81254:/bin/bash
s81386:x:1121:503:Daniela Ruseva, KN, 2kurs, 7gr:/home/KN/s81386:/bin/bash
s45216:x:1235:502:Aleksandar Yavashev, Inf, k3, g3:/home/Inf/s45216:/bin/bash
```

Примерен изход:

2 51

Зад. 5 2017-SE-01 Намерете имената на топ 5 файловете в текущата директория с най-много hardlinks.

Зад. 6 2018-SE-01 Променете правата на всички директории, намиращи се някъде във вашата `home` директория, така че да станат такива, каквито биха се получили, ако ги бяхте създали с маска `0022`.

Зад. 7 2018-SE-02 Напишете серия от команди, извеждащи на екрана *само* inode-а на най-скоро променения (по съдържание) файл, намиращ се в home директорията на потребител pesho (или нейните под-директории), който има повече от едно име.

Зад. 8 2018-SE-03 При поддръждане в нарастващ ред на числовите потребителски идентификатори (UID) на акаунтите, дефинирани в системата, 201-ят акаунт е от групата, запазена за акаунти от специалност СИ.

Изведете списък с имената (име и фамилия) и home директории на всички акаунти от специалност СИ, подреден по факултетен номер.

За справка:

```
s61988:x:1219:504:Stoian Genchev,SI,2,5:/home/SI/s61988:/bin/bash
s81430:x:1234:503:Iordan Petkov, KN, k2, g7:/home/KN/s81430:/bin/bash
s61807:x:1248:504:Elica Venchova:/home/SI/s61807:/bin/bash
s62009:x:1254:504:Denitsa Dobрева, 2, 6:/home/SI/s62009:/bin/bash
s61756:x:1258:504:Katrin Kartuleva, SI, 4, 1:/home/SI/s61756:/bin/bash
s855287:x:1195:504:Vaska Kichukova,SI,2,5:/home/SI/s855287:/bin/bash
```

Примерен изход:

```
Katrin Kartuleva:/home/SI/s61756
Elica Venchova:/home/SI/s61807
Stoian Genchev:/home/SI/s61988
Denitsa Dobрева:/home/SI/s62009
Vaska Kichukova:/home/SI/s855287
```

Зад. 9 2019-SE-01 Даден е текстовият файл `planets.txt`, който съдържа информация за гравитационно закръглените обекти в дадена слънчева система. На всеки ред има информация за точно един обект в следните колони, разделени с ';' :

- име на обекта
- тип на обекта (един знак)
 - T - земен тип
 - G - газов гигант
 - I - леден гигант
- средно разстояние на обекта до локалната звезда
- маса на обекта (относителна величина)
- обем на обекта (относителна величина)
- плътност (g/cm^3)
- средна орбитална скорост (km/s)

Първият ред във файла е header, който описва имената на колоните.

Данните за обектите не са сортирани.

Намерете и изведете разделени с таб името и масата на обекта, който е **едновременно**:

- най-близкият до локалната звезда
- от същия тип като типа на най-далечният до локалната звезда обект

Примерен входен файл:

```
name;type;distance;mass;volume;density;speed
earth;T;1.00000011;1;1;5.52;29.7859
mars;T;1.52366231;0.107;0.151;3.94;24.1309
saturn;G;9.53707032;95;763.62;0.7;9.6724
mercury;T;0.38709893;0.055;0.056;5.43;47.8725
venus;T;0.72333199;0.815;0.857;5.24;35.0214
jupiter;G;5.20336301;318;1321.3;1.33;13.0697
neptune;I;30.06896348;17;57.747;1.76;5.4778
uranus;I;19.19126393;14.5;63.102;1.3;6.8352
```

Зад. 10 2019-SE-02 Вие сте асистент по ОС. На първото упражнение казвате на студентите да си напишат данните на лист, взимате го и им правите акаунти. След упражнението обаче, забравяте да вземете листа със себе си - сещате се половин час по-късно, когато трябва да въведете имената на студентите в таблица, но за зла беда в стаята вече няма ни помен от листа (вероятно иззет от спешния отряд на GDPR-полицията)

Сещате се, че в началото на упражнението UNIX-часовникът е показвал 1551168000, а в края 1551176100.

Напишете команда, която изкарва разделени с таб факултетните номера и имената на потребителите от специалност СИ, чиито home директории са променили статуса си (status change time) в зададения времеви интервал.

Приемете, че всички потребители от СИ имат home директории под /home/SI.

Примерен изход:

```
62198 Ivaylo Georgiev
62126 Victoria Georgieva
62009 Denitsa Dobрева
62208 Trayana Nedelcheva
```

Няколко реда от /etc/passwd за справка:

```
s62136:x:1302:503:Alexander Ignatov, SI, 2, 2:/home/KN/s62136:/bin/bash
s62171:x:1031:504:Deivid Metanov:/home/SI/s62171:/bin/bash
s62126:x:1016:504:Victoria Georgieva:/home/SI/s62126:/bin/bash
s62009:x:1170:504:Denitsa Dobрева,SI,3,3:/home/SI/s62009:/bin/bash
s62196:x:1221:504:Elena Tuparova,SI,2,1:/home/SI/s62196:/bin/bash
```

Зад. 11 2019-SE-03 От всички файлове в home директорията на потребителя velin, изведете дълбочината на файл, който:

- има същия inode като този на най-скоро променения файл сред тях
- има минимална дълбочина

Пояснение Под "дълбочина" да се разбира дълбочина в дървото на файловата система: например файлът /foo/bar/baz има дълбочина 3.

Зад. 12 2020-SE-01 За всички файлове, които (едновременно):

- се намират някъде във вашата home директория;
- имат права, каквито биха имали, ако ги бяхте създали с маска 0022

променете правата им така, че group owner-а на файла да има право да пише в него.

Зад. 13 2020-SE-02 Даден е текстовият файл `spacex.txt`, който съдържа информация за полети на ракети на SpaceX. На всеки ред има информация за един такъв полет в следните колони, разделени с вертикална черта '|':

- дата на полета в UNIX формат
- космодрум
- успешност на полета (две възможни стойности)
 - **Success** - успешен полет
 - **Failure** - неуспешен полет
- полезен товар

Първият ред във файла е header, който описва имената на колоните. Данните във файла не са сортирани.

Намерете и изведете разделени с двоеточие (':') успешността и информацията за полезния товар на най-скорошния полет, който е изстрелян от космодрума с най-много неуспешни полети.

Примерен входен файл:

```
date|launch site|outcome|payload
1291759200|CCAFS|Success|Dragon demo flight and cheese
```

```
1435438800|CCAFS|Failure|SpaceX CRS-7
1275666300|CCAFS|Success|Dragon Spacecraft Qualification Unit
1452981600|VAFB|Success|Jason-3
1498165200|KSC|Success|BulgariaSat-1
1473454800|CCAFS|Failure|Amos-6
1517868000|KSC|Success|Elon Musk's Tesla
1405285200|CCAFS|Success|Orbcomm
```

Зад. 14 2022-CE-01

Напишете серия от команди, които намират всички обикновени файлове, отговарящи едновременно на следните условия:

- намират се в home директорията на текущия потребител, но не и в нейните под-директории;
- собственик им е текущият потребител.

Серията от команди трябва да променя правата на намерените файлове така, че да станат такива, каквито биха се получили, ако при създаването на файловете ефективната маска е била 0002.

Гарантирайте, че серията от команди няма изход.

1.2 Скриптове (теми 1,2,3,4,5)

Зад. 15 2016-SE-01 Напишете shell скрипт, който по подаден един позиционен параметър, ако този параметър е директория, намира всички symlink-ове в нея и под-директориите ѝ с несъществуващ destination.

Зад. 16 2016-SE-02 Напишете shell скрипт, който приема един позиционен параметър - число. Ако скриптът се изпълнява като root, да извежда обобщена информация за общото количество активна памет (*RSS - resident set size, non-swapped physical memory that a task has used*) на процесите на всеки потребител. Ако за някой потребител обобщеното число надвишава подадения параметър, да изпраща подходящи сигнали за прекратяване на процеса с най-много активна памет на потребителя.

Забележка: Приемаме, че изхода в колоната RSS е число в същата мерна единица, като числото, подадено като аргумент. Примерен формат:

USER	PID	%CPU	%MEM	VSZ	RSS	TTY	STAT	START	TIME	COMMAND
root	1	0.0	0.0	15816	1884	?	Ss	May12	0:03	init [2]
root	2	0.0	0.0	0	0	?	S	May12	0:00	[kthreadd]
root	3	0.0	0.0	0	0	?	S	May12	0:02	[ksoftirqd/0]

Забележка: Алтернативно може да ползвате изхода от `ps -e -o uid,pid,rss`

Зад. 17 2016-SE-03 Напишете shell скрипт който, ако се изпълнява от root, проверява кои потребители на системата нямат homedir или не могат да пишат в него.

Примерен формат:

```
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
```

Зад. 18 2016-SE-03 В текущата директория има само обикновени файлове (без директории). Да се напише bash script, който приема 2 позиционни параметъра – числа, които мести файловете от текущата директория към нови директории (a, b и c, които трябва да бъдат създадени), като определен файл се мести към директория 'a', само ако той има по-малко редове от първи позиционен параметър, мести към директория 'b', ако редове са между първи и втори позиционен параметър и в 'c' в останалите случаи.

Зад. 19 2016-SE-04 Файловете във вашата home директория съдържат информация за музикални албуми и имат специфична структура. Началото на всеки ред е годината на издаване на албума, а непосредствено, след началото на всеки ред следва името на изпълителя на песента. Имената на файловете се състоят от една дума, която съвпада с името на изпълнителя.

Примерно съдържание на файл с име "Bonnie":

2005г. Bonnie - "God Was in the Water" (Randall Bramblett, Davis Causey) - 5:17
 2005г. Bonnie - "Love on One Condition" (Jon Cleary) - 3:43
 2005г. Bonnie - "So Close" (Tony Arata, George Marinelli, Pete Wasner) - 3:22
 2005г. Bonnie - "Trinkets" (Emory Joseph) - 5:02
 2005г. Bonnie - "Crooked Crown" (David Batteau, Maia Sharp) - 3:49
 2005г. Bonnie - "Unnecessarily Mercenary" (Jon Cleary) - 3:51
 2005г. Bonnie - "I Will Not Be Broken" - "Deep Water" (John Capek, Marc Jordan) - 3:58

Да се напише shell скрипт приемащ два параметъра, които са имена на файлове от вашата home директория. Скриптът сравнява, кой от двата файла има повече на брой редове, съдържащи неговото име (на файла). За файла победител изпълнете следните действия:

- извлекете съдържанието му, без годината на издаване на албума и без името на изпълнителя
- сортирайте лексикографски извлеченото съдържание и го запишете във файл с име 'изпълнител.songs'

Примерен изходен файл (с име Bonnie.songs):

"Crooked Crown" (David Batteau, Maia Sharp) - 3:49
 "God Was in the Water" (Randall Bramblett, Davis Causey) - 5:17
 "I Will Not Be Broken" - "Deep Water" (John Capek, Marc Jordan) - 3:58
 "Love on One Condition" (Jon Cleary) - 3:43
 "So Close" (Tony Arata, George Marinelli, Pete Wasner) - 3:22
 "Trinkets" (Emory Joseph) - 5:02
 "Unnecessarily Mercenary" (Jon Cleary) - 3:51

Зад. 20 2016-SE-06 Имате текстов файл със следното съдържание (всяка книга е на един ред):

1979 г. - „Синият тайфун“ (сборник съветски научнофантастични разкази за морето)
 1979 г. - „Двойната звезда“ - Любен Дилов
 1979 г. - „Завръщане от звездите“ - Станислав Лем (Превод: Веселин Маринов)
 1979 г. - „Среща с Рама“ - Артър Кларк (Превод: Александър Бояджиев)
 1979 г. - „Алиби“ - Димитър Пеев (криминален роман)
 1979 г. - „Тайнственият триъгълник“ (сборник НФ разкази за морето)
 1979 г. - „Второто нашествие на марсианците“ - Аркадий и Борис Стругацки
 1979 г. - „Гробищен свят“ - Клифърд Саймък (Превод: Михаил Грънчаров)
 1979 г. - „Чоки“ - Джон Уиндъм (Превод: Теодора Давидова)
 1979 г. - „Спускане в Маелстрьом“ - Едгар Алан По (Превод: Александър Бояджиев)
 1980 г. - „Допълнителна примамка“ - Робърт Ф. Ъънг (Превод: Искра Иванова, ...)
 1980 г. - „Кристалното яйце“ - Хърбърт Уелс (Превод: Борис Миндов, ...)
 1980 г. - „Онирофил“ (сборник италиански НФ разкази) (Превод: Никола Иванов, ...)

Напишете shell script (приемащ аргумент име на файл) който извежда:

- всеки ред от файла с добавен пореден номер във формат "1. ", "2. ", ... "11. " ...
- махат данните за годината на издаване
- сортират изхода по заглавие (лексикографски, възходящо)

Примерен изход (показани са само първите 4 реда):

5. „Алиби“ - Димитър Пеев (криминален роман)
 7. „Второто нашествие на марсианците“ - Аркадий и Борис Стругацки
 8. „Гробищен свят“ - Клифърд Саймък (Превод: Михаил Грънчаров)
 2. „Двойната звезда“ - Любен Дилов

Зад. 21 2017-IN-01 Напишете скрипт, който приема три задължителни позиционни аргумента:

- име на файл
- *низ1*
- *низ2*

Файлът е текстови, и съдържа редове във формат:

ключ=стойност

където *стойност* може да бъде:

- празен низ, т.е. редът е *ключ=*
- низ, състоящ се от един или повече термове, разделени с интервали, т.е., редът е *ключ=t₁ t₂ t₃*

Някъде във файла:

- се съдържа един ред с *ключ* първия подаден низ (*низ1*);
- и може да се съдържа един ред с *ключ* втория подаден низ (*низ2*).

Скриптът трябва да променя реда с *ключ* *низ2* така, че обединението на термовете на редовете с ключове *низ1* и *низ2* да включва всеки терм еднократно.

Примерен входен файл:

```
$ cat z1.txt
FOO=73
BAR=42
BAZ=
ENABLED_OPTIONS=a b c d
ENABLED_OPTIONS_EXTRA=c e f
```

Примерно извикване:

```
$ ./a.sh z1.txt ENABLED_OPTIONS ENABLED_OPTIONS_EXTRA
```

Изходен файл:

```
$ cat z1.txt
FOO=73
BAR=42
BAZ=
ENABLED_OPTIONS=a b c d
ENABLED_OPTIONS_EXTRA=e f
```

Зад. 22 2017-IN-02 Напишете скрипт, който приема задължителен позиционен аргумент - име на потребител *FOO*. Ако скриптът се изпълнява от root:

- а) да извежда имената на потребителите, които имат повече на брой процеси от *FOO*, ако има такива;
- б) да извежда средното време (в секунди), за което са работили процесите на всички потребители на системата (*TIME*, във формат *HH:MM:SS*);
- в) ако съществуват процеси на *FOO*, които са работили над два пъти повече от средното време, скриптът да прекратява изпълнението им по подходящ начин.

За справка:

```
$ ps -e -o user,pid,%cpu,%mem,vsz,rss,ttty,stat,time,command | head -5
USER      PID %CPU %MEM    VSZ   RSS TTY      STAT    TIME COMMAND
root         1  0.0  0.0  15820  1920 ?        Ss     00:00:05 init [2]
root         2  0.0  0.0      0     0 ?        S       00:00:00 [kthreadd]
root         3  0.0  0.0      0     0 ?        S       00:00:01 [ksoftirqd/0]
root         5  0.0  0.0      0     0 ?        S<      00:00:00 [kworker/0:0H]
```

Зад. 23 2017-IN-03 Напишете скрипт, който извежда името на потребителския акаунт, в чиято home директория има най-скоро променен обикновен файл и кой е този файл. Напишете скрипта с подходящите проверки, така че да бъде валиден инструмент.

Зад. 24 2017-SE-01 Напишете скрипт, който получава задължителен първи позиционен параметър – директория и незадължителен втори – число. Скриптът трябва да проверява подадената директория и нейните под-директории и да извежда имената на:

- а) при подаден на скрипта втори параметър – всички файлове с брой hardlink-ове поне равен на параметъра;
- б) при липса на втори параметър – всички symlink-ове с несъществуващ destination (счупени symlink-ове).

Забележка: За удобство приемаме, че ако има подаден втори параметър, то той е число.

Зад. 25 2017-SE-02 Напишете скрипт, който приема три задължителни позиционни параметра - директория *SRC*, директория *DST* (която не трябва да съдържа файлове) и низ *ABC*. Ако скриптът се изпълнява от root потребителя, то той трябва да намира всички файлове в директорията *SRC* и нейните под-директории, които имат в името си като под-низ *ABC*, и да ги мести в директорията *DST*, запазвайки директориината структура (но без да запазва мета-данни като собственик и права, т.е. не ни интересуват тези параметри на новите директории, които скриптът би генерирал в *DST*).

Пример:

- в *SRC* (/src) има следните файлове:

```
/src/foof.txt
/src/1/bar.txt
/src/1/foo.txt
/src/2/1/foobar.txt
/src/2/3/barf.txt
```

- *DST* (/dst) е празна директория
- зададения низ е foo

Резултат:

- в *SRC* има следните файлове:

```
/src/1/bar.txt
/src/2/3/barf.txt
```

- в *DST* има следните файлове:

```
/dst/foof.txt
/dst/1/foo.txt
/dst/2/1/foobar.txt
```

Зад. 26 2017-SE-03 Напишете скрипт, който ако се изпълнява от root потребителя:

- а) извежда обобщена информация за броя и общото количество активна памет (*RSS - resident set size, non-swaped physical memory that a task has used*) на текущите процеси на всеки потребител;
- б) ако процесът с най-голяма активна памет на даден потребител използва два пъти повече памет от средното за потребителя, то скриптът да прекратява изпълнението му по подходящ начин.

За справка:

```
$ ps aux | head -5
USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
root         1  0.0  0.0  15820  1052 ?        Ss   Apr21    0:06 init [2]
root         2  0.0  0.0      0      0 ?        S    Apr21    0:00 [kthreadd]
root         3  0.0  0.0      0      0 ?        S    Apr21    0:02 [ksoftirqd/0]
root         5  0.0  0.0      0      0 ?        S<   Apr21    0:00 [kworker/0:0H]
```

Алтернативно, може да ползвате изхода от ps -e -o uid,pid,rss

Зад. 27 2017-SE-04 Напишете shell script, който получава задължителен първи позиционен параметър – директория и незадължителен втори – име на файл. Скриптът трябва да намира в подадената директория и нейните под-директории всички symlink-ове и да извежда (при подаден аргумент файл – добавяйки към файла, а ако не е – на стандартния изход) за тях следната информация:

- ако destination-а съществува – *името на symlink-a -> името на destination-a*;
- броя на symlink-овете, чийто destination не съществува.

Примерен изход:


```
lbaz -> /foo/bar/baz
lqux -> ../../../../qux
lquux -> /foo/quux
Broken symlinks: 34
```

Зад. 28 2017-SE-05 Напишете скрипт, който получава два задължителни позиционни параметъра – директория и низ. Сред файловете в директорията би могло да има такива, чиито имена имат структура `vmlinuz-x.y.z-arch` където:

- `vmlinuz` е константен низ;
- тиретата “-” и точките “.” присъстват задължително;
- *x* е число, version;
- *y* е число, major revision;
- *z* е число, minor revision;
- наредената тройка *x.y.z* формира глобалната версия на ядрото;
- *arch* е низ, архитектура (платформа) за която е съответното ядро.

Скриптът трябва да извежда само името на файла, намиращ се в подадената директория (но не и нейните поддиректории), който:

- спазва гореописаната структура;
- е от съответната архитектура спрямо параметъра-низ, подаден на скрипта;
- има най-голяма глобална версия.

Пример:

- Съдържание на `./kern/`:

```
vmlinuz-3.4.113-amd64
vmlinuz-4.11.12-amd64
vmlinuz-4.12.4-amd64
vmlinuz-4.19.1-i386
```

- Извикване и изход:

```
$ ./task1.sh ./kern/ amd64
vmlinuz-4.12.4-amd64
```

Зад. 29 2017-SE-06 Напишете скрипт, който ако се изпълнява от root потребителя, намира процесите на потребителите, които не са root потребителя и е изпълнено поне едно от следните неща:

- имат зададена несъществуваща home директория;
- не са собственици на home директорията си;
- собственика на директорията не може да пише в нея.

Ако общото количество активна памет (*RSS - resident set size, non-swaped physical memory that a task has used*) на процесите на даден такъв потребител е по-голямо от общото количество активна памет на root потребителя, то скриптът да прекратява изпълнението на всички процеси на потребителя.

За справка:

```
$ ps aux | head -n 5
USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
root         1  0.0  0.0  15820  1052 ?        Ss   Apr21    0:06 init [2]
root         2  0.0  0.0      0     0 ?        S    Apr21    0:00 [kthreadd]
root         3  0.0  0.0      0     0 ?        S    Apr21    0:02 [ksoftirqd/0]
root         5  0.0  0.0      0     0 ?        S<   Apr21    0:00 [kworker/0:0H]
```

Алтернативно, може да ползвате изхода от `ps -e -o uid,pid,rss`

```
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
s61934:x:1177:504:Mariq Cholakova:/home/SI/s61934:/bin/bash
```

Зад. 30 2018-SE-01 Нека съществува програма за моментна комуникация (Instant messaging), която записва логове на разговорите в следния формат:

- има определена директория за логове (*LOGDIR*)
- в нея има директорийна структура от следния вид:
LOGDIR/протокол/акаунт/приятел/
като на всяко ниво може да има няколко екземпляра от съответния вид, т.е. няколко директории *протокол*, във всяка от тях може да има няколко директории *акаунт*, и във всяка от тях – няколко директории *приятел*
- във всяка от директориите *приятел* може да има файлове с имена от вида *yyyy-mm-dd-hh-mm-ss.txt* – година-месец-ден и т.н., спрямо това кога е започнал даден разговор
- всеки такъв файл представлява лог на даден разговор със съответния приятел, като всяка разменена реплика между вас е на отделен ред
- даден идентификатор *приятел* може да се среща няколко пъти в структурата (напр. през различни ваши акаунти сте водили разговори със същия приятел)

Напишете скрипт, който приема задължителен позиционен аргумент - име на лог директория (*LOGDIR*). Скриптът трябва да извежда десетимата приятели, с които имате най-много редове комуникация глобално (без значение протокол и акаунт), и колко реда имате с всеки от тях. Опишете в коментар как работи алгоритъмът ви.

Зад. 31 2018-SE-02 Напишете скрипт, който приема два позиционни аргумента – име на текстови файл и директория. Директорията не трябва да съдържа обекти, а текстовият файл (US-ASCII) е стенограма и всеки ред е в следния формат:

ИМЕ ФАМИЛИЯ (уточнения): Реплика

където:

- ИМЕ ФАМИЛИЯ присъстват задължително;
- ИМЕ и ФАМИЛИЯ се състоят само от малки/главни латински букви и тирета;
- (уточнения) не е задължително да присъстват;
- двоеточието ':' присъства задължително;
- Репликата не съдържа знаци за нов ред;
- в стринга преди двоеточието ':' задължително има поне един интервал между ИМЕ и ФАМИЛИЯ;
- наличието на други интервали където и да е на реда е недефинирано.

Примерен входен файл:

```
John Lennon (The Beatles): Time you enjoy wasting, was not wasted.
Roger Waters: I'm in competition with myself and I'm losing.
John Lennon:Reality leaves a lot to the imagination.
Leonard Cohen:There is a crack in everything, that's how the light gets in.
```

Скриптът трябва да:

- създава текстови файл *dict.txt* в посочената директория, който на всеки ред да съдържа:
ИМЕ ФАМИЛИЯ;НОМЕР
където:
– ИМЕ ФАМИЛИЯ е уникален участник в стенограмата (без да се отчитат уточненията);
– НОМЕР е уникален номер на този участник, избран от вас.
- създава файл *НОМЕР.txt* в посочената директория, който съдържа всички (и само) редовете на дадения участник.

Зад. 32 2018-SE-03 Напишете скрипт, който приема два позиционни аргумента – имена на текстови файлове в CSV формат:

```
8,foo,bar,baz
2,quz,,foo
12,1,3,foo
3,foo,,
5,,bar,
7,,,
4,foo,bar,baz
```

Валидни са следните условия:

- CSV файловете представляват таблица, като всеки ред на таблицата е записан на отделен ред;
- на даден ред всяко поле (колона) е разделено от останалите със запетая;
- броят на полетата на всеки ред е константа;
- в полетата не може да присъства запетая, т.е., запетаята винаги е разделител между полета;
- ако във файла присъстват интервали, то това са данни от дадено поле;
- първото поле на всеки ред е число, което представлява идентификатор на реда (ID).

Примерно извикване: `./foo.sh a.csv b.csv`

Скриптът трябва да чете `a.csv` и на негова база да създава `b.csv` по следния начин:

- някои редове във файла се различават само по колоната ID, и за тях казваме, че формират множество A_i
- за всяко такова множество A_i да се оставя само един ред - този, с най-малка стойност на ID-то;
- редовете, които не са членове в някое множество A_i се записват в изходния файл без промяна.

Зад. 33 2019-SE-01 Напишете два скрипта (по един за всяка подточка), които четат редове от STDIN. Скриптовите трябва да обработват само редовете, които съдържат цели положителни или отрицателни числа; останалите редове се игнорират. Скриптовите трябва да извежда на STDOUT:

- а) всички уникални числа, чиято абсолютна стойност е равна на максималната абсолютна стойност сред всички числа
- б) всички най-малки уникални числа от тези, които имат максимална сума на цифрите си

Примерен вход:

```
We don't
n11d n0
educat10n
12.3
6
33
-42
-42
111
111
-111
```

Примерен изход за а):

```
-111
111
```

Примерен изход за б):

```
-42
```

Зад. 34 2019-SE-02 Напишете шел скрипт, който приема множество параметри. Общ вид на извикване:

`./foo.sh [-n N] FILE1...`

В общия случай параметрите се третираат като имена на (`.log`) файлове, които трябва да бъдат обработени от скрипта, със следното изключение: ако първият параметър е стрингът `-n`, то вторият параметър е число, дефиниращо стойност на променливата `N`, която ще ползваме в скрипта. Въвеждаме понятието *идентификатор на файл* (ИДФ), което се състои от името на даден файл без разширението `.log`. За удобство приемаме, че скриптът:

- ще бъде извикван с аргументи имена на файлове, винаги завършващи на `.log`
- няма да бъде извикван с аргументи имена на файлове с еднакъв ИДФ.

Лог файловете са текстови, като всеки ред има следния формат:

- време: timestamp във формат `YYYY-MM-DD HH:MM:SS`

- интервал
- данни: поредица от символи с произволна дължина

За удобство приемаме, че редовете във всеки файл са сортирани по време възходящо.

Примерно съдържание на даден лог файл:

```
2019-05-05 06:26:54 orthanc rsyslogd: rsyslogd was HUPed
2019-05-06 06:30:32 orthanc rsyslogd: rsyslogd was HUPed
2019-05-06 10:48:29 orthanc kernel: [1725379.728871] Chrome_~dThread[876]: segfault
```

Скриптът трябва да извежда на `STDOUT` последните `N` реда (ако `N` не е дефинирано - 10 реда) от всеки файл, в следния формат:

- timestamp във формат `YYYY-MM-DD HH:MM:SS`
- интервал
- ИДФ
- интервал
- данни

Изходът трябва да бъде глобално сортиран по време възходящо.

Зад. 35 2019-SE-03 За удобство приемаме, че разполагате със системен инструмент `sha256sum`, който приема аргументи имена на файлове като за всеки файл пресмята и извежда уникална хеш стойност, базирана на съдържанието на файла. Изходът от инструмента е текстови, по един ред за всеки подаден като аргумент файл, в следния формат:

- хеш стойност с дължина точно 64 знака
- два интервала
- име на файл

Примерна употреба и изход:

```
$ sha256sum /var/log/syslog /var/log/user.log README.md
b2ff8bd882a501f71a144b7c678e3a6bc6764ac48eb1876fb5d11aac11014b78 /var/log/syslog
e3b0c44298fc1c149afb4c8996fb92427ae41e4649b934ca495991b7852b855 /var/log/user.log
e4702d8044b7020af5129fc69d77115fd4306715bd678ba4bef518b2edf01fb9 README.md
```

Напишете скрипт, който приема задължителен параметър име на директория (ДИР1). Някъде в директорията ДИР1 може да съществуват архивни файлове с имена `NAME_report-TIMESTAMP.tgz`, където:

- `NAME` е низ, който не съдържа символ `'_'`
- `TIMESTAMP` е във формат Unix time (POSIX time/UNIX Epoch time)

На всяко пускане на скрипта се обработват само новосъздадените или модифицираните по съдържание спрямо предното пускане на скрипта архивни файлове от горния тип. За всеки такъв архивен файл се изпълнява следното:

- ако архивният файл съдържа файл с име `meow.txt`, то този текстови файл да бъде записан под името `/extracted/NAME_TIMESTAMP.txt`, където `NAME` и `TIMESTAMP` са съответните стойности от името на архивния файл.

Зад. 36 2020-SE-01 Напишете shell скрипт, който получава два задължителни позиционни параметъра - име на файл (`bar.csv`) и име на директория. Директорията може да съдържа текстови файлове с имена от вида `foobar.log`, всеки от които има съдържание от следния вид:

Пример 1 (loz-gw.log):

Licensed features for this platform:

```
Maximum Physical Interfaces    : 8
VLANs                          : 20
Inside Hosts                   : Unlimited
Failover                       : Active/Standby
VPN-3DES-AES                   : Enabled
```

```
*Total VPN Peers      : 25
VLAN Trunk Ports      : 8
```

This platform has an ASA 5505 Security Plus license.

Serial Number: JMX00000000

Running Activation Key: 0e268e0c

Пример 2 (border-lozenets.log):

Licensed features for this platform:

```
Maximum Physical Interfaces : 4
VLANs                       : 16
Inside Hosts                : Unlimited
Failover                    : Active/Active
VPN-3DES-AES                : Disabled
*Total VPN Peers            : 16
VLAN Trunk Ports            : 4
```

This platform has a PIX 535 license.

Serial Number: PIX5350007

Running Activation Key: 0xd11b3d48

Имената на лог файловете (loz-gw, border-lozenets) определят даден hostname, а съдържанието им дава детайли за определени параметри на съответният хост.

Файлът `bar.csv`, който трябва да се генерира от вашия скрипт, е т.н. CSV (comma separated values) файл, тоест текстови файл - таблица, на който полетата на всеки ред са разделени със запетая. Първият ред се ползва за определяне на имената на колоните.

Скриптът трябва да създава файла `bar.csv` на база на лог файловете в директорията. Генерираният CSV файл от директория, която съдържа само `loz-gw.log` и `border-lozenets.log` би изглеждал така:

```
hostname,phy,vlans,hosts,failover,VPN-3DES-AES,peers,VLAN Trunk Ports,license,SN,key
loz-gw,8,20,Unlimited,Active/Standby,Enabled,25,8,ASA 5505 Security Plus,JMX00000000,0e268e0c
border-lozenets,4,16,Unlimited,Active/Active,Disabled,16,4,PIX 535,PIX5350007,0xd11b3d48
```

Полетата в генерирания от скрипта CSV файл не трябва да съдържат излишни trailing/leading интервали. За улеснение, приетете, че всички whitespace символи във входните файлове са символа "интервал".

Зад. 37 2020-SE-02 Напишете shell скрипт, който приема задължителен параметър - име на файл. Файлът е log файл на HTTP сървър, в който се записват всички получени от сървъра request-и, които клиентите са изпратили. Файлът е текстови, като на всеки ред има информация от следния вид:

```
35.223.122.181 dir.bg - [03/Apr/2020:17:25:06 -0500] GET / HTTP/1.1 302 0 "-" "Zend_Http_Client"
94.228.82.170 del.bg - [03/Apr/2020:17:25:06 -0500] POST /auth HTTP/2.0 400 153 "foo bar" "<UA>"
```

Всеки ред на файла се състои от полета, разделени с интервал. Описание на полетата с пример спрямо първият ред от горните:

- адрес на клиент - 35.223.122.181
- име на виртуален хост (сайт) - dir.bg
- име на потребител - -
- timestamp на заявката - [03/Apr/2020:17:25:06 -0500]

- заявка - GET / HTTP/1.1 - състои се от три компонента, разделени с интервал: метод на заявката (за удобство приемаме, че може да има само GET и POST заявки), ресурсен идентификатор, и протокол (приемаме, че може да има само HTTP/1.0, HTTP/1.1 и HTTP/2.0 протоколи)
- код за статус на заявката - 302
- брой байтове - 0
- referer - "-" - ограден в двойни кавички, подава се от HTTP клиента, произволен низ
- user agent - "Zend_Http_Client" - ограден в двойни кавички, подава се от HTTP клиента, произволен низ

За всеки от top 3 сайта, към които има най-много заявки, скриптът трябва да изведе в долния формат:

- брой на HTTP/2.0 заявките
- брой на не-HTTP/2.0 заявките
- top 5 клиента, направили най-много заявки, завършили с код, по-голям от 302 (и броя на съответните им заявки)

```
dir.bg HTTP/2.0: 0 non-HTTP/2.0: 5
del.bg HTTP/2.0: 5 non-HTTP/2.0: 0
      5 94.228.82.170
      2 34.73.112.204
      1 185.217.0.138
```

Зад. 38 2020-SE-03 Под *пакет* ще разбирате директория, която има следната структура:

```
<name>
|-- version
|-- tree
|...
```

Където *<name>* е името на пакета, *version* е текстов файл, който съдържа низ от вида 1.2.3-4 и нищо друго, а *tree* е директория с произволно съдържание.

За да получим *архив на пакет*, архивираме (*tar*) и компресиране (*xz*) съдържанието на директорията *tree*.

Под *хранилище* ще разбирате директория, която има следната структура:

```
<repo name>
|-- db
|-- packages
|...
```

Където *<repo name>* е името на хранилището, *db* е текстов файл, чиито редове имат вида *<package name>-<package version> <package checksum>* и са сортирани лексикографски. Директорията *packages* съдържа архиви с имена *<package checksum>.tar.xz*, които съответстват на редове в *db*. Под *<package checksum>* имаме предвид sha256 сумата на архива на пакета.

Напишете скрипт *repo_add.sh*, който приема два аргумента - път до хранилище и път до пакет, който добавя пакета в хранилището. Ако същата версия на пакет вече съществува, архивът се заменя с новата версия. В противен случай, новата версия се добавя заедно с другите.

Заб: Първо си проектирайте общия алгоритъм на работа.

Примерно хранилище:

```
myrepo
|-- db
|-- packages
    |-- 6e3549438bc246b86961b2e8c3469321ca22eabd0a6c487d086de7a43a0ef766.tar.xz
    |-- 66b28e48161ba01ae25433b9ac4086a83b14d2ee49a62f2659c96514680ab6e8.tar.xz
    |-- 99c934ad80bd9e49125523c414161e82716b292d4ed2f16bb977d6db7e13d9bc.tar.xz
```

Със съдържание на *db*:

```
glibc-2.31-2 6e3549438bc246b86961b2e8c3469321ca22eabd0a6c487d086de7a43a0ef766
zlib-1.1.15-8 66b28e48161ba01ae25433b9ac4086a83b14d2ee49a62f2659c96514680ab6e8
zlib-1.2.11-4 99c934ad80bd9e49125523c414161e82716b292d4ed2f16bb977d6db7e13d9bc
```

Примерен пакет:

```
zlib
|-- version          # contains '1.2.11-3'
|-- tree
| ...
```

Съдържание на хранилището след изпълнение на `./repo-add.sh myrepo zlib`

```
myrepo
|-- db
|-- packages
| -- 6e3549438bc246b86961b2e8c3469321ca22eabd0a6c487d086de7a43a0ef766.tar.xz
| -- 66b28e48161ba01ae25433b9ac4086a83b14d2ee49a62f2659c96514680ab6e8.tar.xz
| -- b839547ee0aed82c74a37d4129382f1bd6fde85f97c07c5b705eeb6c6d69f162.tar.xz
| -- 99c934ad80bd9e49125523c414161e82716b292d4ed2f16bb977d6db7e13d9bc.tar.xz
```

Със съдържание на db:

```
glibc-2.31-2 6e3549438bc246b86961b2e8c3469321ca22eabd0a6c487d086de7a43a0ef766
zlib-1.1.15-8 66b28e48161ba01ae25433b9ac4086a83b14d2ee49a62f2659c96514680ab6e8
zlib-1.2.11-3 b839547ee0aed82c74a37d4129382f1bd6fde85f97c07c5b705eeb6c6d69f162
zlib-1.2.11-4 99c934ad80bd9e49125523c414161e82716b292d4ed2f16bb977d6db7e13d9bc
```

Зад. 39 2020-SE-04 Напишете скрипт, който приема два аргумента - имена на директории. Първата (*SRC*) съществува, докато втората (*DST*) трябва да бъде създадена от скрипта. Директорията *SRC* и нейните поддиректории може да съдържат файлове, чиито имена завършат на `.jpg`. Имената на файловете може да съдържат интервали, както и поднизове, оградени със скоби, например:

```
A single (very ugly) tree (Outdoor treks) 2.jpg
Falcons.jpg
Gorgonzola (cheese).jpg
Leeches (two different ones) (Outdoor treks).jpg
Pom Pom Pom.jpg
```

За даден низ ще казваме, че е *почистен*, ако от него са премахнати leading и trailing интервалите и всички последователни интервали са сведени до един.

За всеки файл дефинираме следните атрибути:

- *заглавие* - частта от името преди `.jpg`, без елементи оградени в скоби, почистен. Примери:

```
A single tree 2
Falcons
Gorgonzola
Leeches
Pom Pom Pom
```

- *албум* - последният елемент от името, който е бил ограден в скоби, почистен. Ако *албум* е празен стринг, ползваме стойност по подразбиране `misc`. Примери:

```
Outdoor treks
misc
cheese
Outdoor treks
misc
```

- *data* - времето на последна модификация на съдържанието на файла, във формат YYYY-MM-DD
- *хеш* - първите 16 символа от sha256 сумата на файла. *Забележка: приемаме, че в тази идеална вселена първите 16 символа от sha256 сумата са уникални за всеки файл от тези, които ще се наложи да обработваме.*

Скриптът трябва да създава в директория *DST* необходимата структура от под-директории, файлове и symlink-ове, така че да са изпълнени следните условия за всеки файл от *SRC*:

- *DST/images/хеш.jpg* - копие на съответния файл
- следните обекти са относителни symlink-ове към *хеш.jpg*:
 - *DST/by-date/дата/by-album/албум/by-title/заглавие.jpg*
 - *DST/by-date/дата/by-title/заглавие.jpg*
 - *DST/by-album/албум/by-date/дата/by-title/заглавие.jpg*
 - *DST/by-album/албум/by-title/заглавие.jpg*
 - *DST/by-title/заглавие.jpg*

Зад. 40 2020-SE-05 Напишете shell скрипт, който приема 3 позиционни аргумента – две имена на файлове и име на директория. Примерно извикване:

```
$ ./foo.sh foo.pwd config.cfg cfgdir/
```

В директорията *cfgdir/* и нейните под-директории може да има файлове с имена завършващи на *.cfg*. За да са *валидни*, тези файлове трябва да съдържат редове само в следните формати (редовете започващи с # са коментари):

```
# internal laboratory
{ no-production };

{ volatile };

# meow configs
{ run-all; };
```

Във файла *foo.pwd* има описани потребителски имена (username) и MD5 хеш суми на паролите им, с по един запис на ред, в следният формат:

```
username:password_hash
```

Също така, разполагате с команда **pwgen**, която генерира и извежда на **STDOUT** случайни пароли, и знаете, че поддържа следните два аргумента:

```
$ pwgen [ password_length ] [ number_of_passwords ]
```

Вашият скрипт трябва да валидира *cfg* файловете в директорията, и за тези, които не са валидни, да извежда на **STDOUT** името на файла и номерирани редовете, които имат проблем, в следния формат:

```
Error in filename.cfg:
Line 1:XXXX
Line 37:YYYY
```

където *XXXX* и *YYYY* е съдържанието на съответния ред.

За валидните файлове, скриптът трябва да:

- генерира *config.cfg* като обединение на съдържанието им;
- името на файла, без частта *.cfg* дефинира потребителско име. Ако във файла с паролите не съществува запис за този потребител, то такъв да се добави и на стандартния изход да се изведе потребителското име и паролата (поне 16 символа) разделени с един интервал.

Зад. 41 2020-SE-06 Под *конфигурационен файл* ще разбираме файл, в който има редове от вида `key=value`, където `key` и `value` могат да се състоят от букви, цифри и знак "долна черта" ("_"). Освен това, във файла може да има празни редове; може да има произволен whitespace в началото и в края на редовете, както и около символа "=". Също така са допустими и коментари в даден ред: всичко след символ "#" се приема за коментар.

Под `<date>` ще разбираме текущото време, върнато от командата `date` без параметри; под `<user>` ще разбираме името на текущият потребител.

Напишете shell скрипт `set_value.sh`, който приема 3 позиционни аргумента – име на конфигурационен файл, ключ (`foo`) и стойност (`bar`). Ако ключът:

- присъства във файла с друга стойност, скриптът трябва да:
 - да закоментира този ред като сложи # в началото на реда и добави в края на реда `# edited at <date> by <user>`
 - да добави нов ред `foo = bar # added at <date> by <user>` точно след стария ред
- не присъства във файла, скриптът трябва да добави ред от вида `foo = bar # added at <date> by <user>` на края на файла

Примерен foo.conf:

```
# route description

from = Sofia
to   = Varna   # my favourite city!
type = t2_1
```

Примерно извикване:

```
./set_value.sh foo.conf to Plovdiv
```

Съдържание на foo.conf след извикването:

```
# route description

from = Sofia
# to   = Varna   # my favourite city! # edited at Tue Aug 25 15:48:29 EEST 2020 by human
to = Plovdiv # added at Tue Aug 25 15:48:29 EEST 2020 by human
type = t2_1
```

Зад. 42 2021-SE-01

Разполагате с машина, на която е инсталиран специализиран софтуер, който ползва два потребителски акаунта – `oracle` и `grid`. Всеки от потребителите *би трябвало* да има `environment` променлива `ORACLE_HOME`, която указва абсолютен път до директория във формат `/path/to/dir`. В поддиректория `bin` на зададената директория *би трябвало* да има изпълним файл с име `adrci`. Всеки от двата потребителя има собствена директория `diag_dest`, която е във вида `/u01/app/потребител`. Когато някой от потребителите изпълни неговото копие на командата `adrci` с параметър `exes="show homes"` може да получи на `STDOUT` един от следните два изхода:

- вариант 1: (неуспех): `No ADR homes are set`
- вариант 2: (успех):

```
ADR Homes:
diag/rdbms/orclbi/orclbi1
diag/rdbms/orclbi/orclbi2
```

И в двата случая командата приключва с `exit code 0`. Ако командата се изпълни успешно, тя връща списък с един или повече `ADR Homes`, които са релативни имена на директории спрямо `diag_dest` на съответния потребител.

Напишете скрипт, който може да се изпълнява само от някой от тези два акаунта, и извежда на `STDOUT` размера в мегабайти и абсолютният път на всеки `ADR Home`.

Примерен изход:

```
0      /u01/app/oracle/diag/rdbms/orclbi/orclbi1
389    /u01/app/oracle/diag/rdbms/orclbi/orclbi2
```

Зад. 43 2021-SE-02

Един от често използваните DNS сървъри е BIND9, при който описанието на DNS зоните обикновено стои в текстови файлове, наричани *зонални файлове*. За улеснение, в рамките на задачата, ще ползваме опростено описание на зоналните файлове.

Под *whitespace* разбираме произволна комбинация от табове и интервали.

Под *FQDN* разбираме низ, който има допустими символи малки латински букви, цифри и точка; не може да започва с точка, не може да има две или повече съседни точки, задължително завършва с точка.

Зоналните файлове съдържат *ресурсни записи*, по един на ред. Общият вид на даден ресурсен запис е <ключ> <TTL> <клас> <тип> <RDATA>, разделени с whitespace, например:

```
astero.openfmi.net. 3600 IN A 185.117.82.99
```

Където:

- ключ (*astero.openfmi.net.*) – FQDN
- TTL (3600) – цифри; полето може да липсва
- клас (IN) – главни латински букви; класът винаги е IN
- тип (A) – главни латински букви; някои от SOA, NS, A, AAAA
- RDATA (185.117.82.99) – данни на записа; различни за различните типове ресурсни записи; всичко след типа до края на реда.

Знакът точка-и-запетая ; е знак за коментар, и всичко след него до края на реда се игнорира.

Във всеки зонален файл трябва да има точно един SOA запис, и той трябва да е първият запис във файла. Пример за едноредов SOA запис:

```
openfmi.net. 3600 IN SOA nimbus.fccf.net. root.fccf.net. 2021041901 86400 7200 3024000 3600
```

RDATA-та на SOA запис се състои от два FQDN-а и пет числа, разделени с whitespace.

Въпреки, че горното е валиден SOA запис, за прегледност в зоналните файлове често се ползва следният синтаксис (многоредов SOA запис, еквивалентен на горния):

```
openfmi.net. 3600 IN SOA nimbus.fccf.net. root.fccf.net. (
    2021041901 ; serial
    86400      ; refresh
    7200       ; retry
    3024000    ; expire
    3600       ; negative TTL
)
```

т.е., поредицата от числа се разбива на няколко реда, оградени в обикновени скоби, и за всяко число се слага коментар какво означава.

Първото от тези числа (serial) представлява серийният номер на зоната, който трябва да се увеличава всеки път, когато нещо в зоналният файл се промени. Изключително важно е това число само да нараства, и никога да не намалява.

Един от често използваните формати за серийен номер, който показва кога е настъпила последната промяна в зоналния файл представлява число във вида YYYYMMDDTT, т.е., четири цифри за година, две цифри за месец, две цифри за дата и още две цифри за поредна промяна в рамките на описания ден. За последните две цифри (TT) има ограничение да са от 00 до 99 (естествено, така не може да има повече от 100 промени в рамките на един ден).

За удобство приемаме, че конкретен серийен номер (точната поредица цифри) се среща само на едно място в зоналния файл.

Напишете шел скрипт, който по подадени имена на зонални файлове променя серийният номер в SOA записа на всеки файл по следният алгоритъм:

- ако датата в серийният номер е по-стара от днешната, новият сериен номер трябва да е от вида `днешната00`
- ако датата в серийният номер е равна на днешната, серийният номер трябва да се увеличи с единица

Важат следните условия:

- скриптът трябва да може да обработва и едноредови, и многоредови SOA записи
- за всеки зонален файл, който не е успял да обработи, скриптът трябва да вади съобщение за грешка, което включва и името на зоналния файл. Съобщенията трябва да са лесно обработваеми с познатите инструменти за обработка на текст.

Зад. 44 2021-SE-03 Напишете shell скрипт, който приема два позиционни параметъра – имена на файлове. Примерно извикване:

```
$ ./foo.sh input.bin output.h
```

Файлът `input.bin` е двоичен файл с елементи `uint16_t` числа, създаден на *little-endian* машина.

Вашият скрипт трябва да генерира C хедър файл, дефиниращ масив с име `arr`, който:

- съдържа всички елементи от входния файл;
- няма указана големина;
- не позволява промяна на данните.

Генерираният хедър файл трябва да:

- съдържа и `uint32_t` променлива `arrN`, която указва големината на масива;
- бъде валиден и да може да се `#include`-ва без проблеми от C файлове, очакващи да “виждат” `arr` и `arrN`.

За да е валиден един входен файл, той трябва да съдържа не повече от 524288 елемента.

За справка, `dump` на съдържанието на примерен `input.bin`:

```
00000000: 5555 5655 5955 5a55 6555 6655 6955 6a55  UVUUYUZUeUfUiUjU
00000010: 9555 9655 9955 9a55 a555 a655 a955 aa55  .U.U.U.U.U.U.U.U
```

Зад. 45 2021-SE-04

Разполагате с машина, на която е инсталиран специализиран софтуер, който ползва два потребителски акаунта – `oracle` и `grid`. За всеки от двата акаунта съществува директория, която ще наричаме `diag_dest` и тя е от вида `/u01/app/потребител`.

Всеки от потребителите *би трябвало* да има `environment` променлива `ORACLE_HOME`, която указва абсолютен път до директория във формат `/път/до/дир`. В поддиректорията `bin` на зададената директория *би трябвало* да има изпълним файл с име `adrci`.

Всеки от потребителите може да подаде серия от подкоманди, които неговото копие на `adrci` да изпълни, като го извика с параметър `exec="cmd1; cmd2; cmd3"`. Отделните подкоманди се разделят с точка и запетая (;). Командата `adrci` винаги приключва с `exit code 0`.

Нека дефинираме следните подкоманди:

- **SET BASE** – за да се гарантира правилна работа на командата `adrci`, при всяко нейно извикване първата подкоманда трябва да бъде от вида `SET BASE diag_dest`, където `diag_dest` е абсолютният път на съответната директория за дадения потребител
- **SHOW HOMES** – подкоманда `SHOW HOMES` извежда на `STDOUT` един от следните два изхода:
 - вариант 1: (неуспех): `No ADR homes are set`
 - вариант 2: (успех):

```
ADR Homes:
diag/rdbms/orclbi/orclbi1
diag/rdbms/orclbi/orclbi2
```

Ако командата се изпълни успешно, тя връща списък с един или повече ADR Homes, които са релативни имена на директории спрямо `diag_dest` на съответният потребител.

От полученият списък с релативни пътища *интересни* за нас са само тези, които за име на директория на второ ниво имат една от следните директории: `crs`, `tnslsnr`, `kfod`, `asm` или `rdbms`.

- **SET HOMEPATH** – подкоманда **SET HOMEPATH** път задава *активна* работна директория, спрямо която ще се изпълняват последващи подкоманди в рамките на същото изпълнение на **adrci**; *път* е релативен път до директория, спрямо изхода на **SHOW HOMES**
- **PURGE** – подкоманда **PURGE -AGE** минути изтрива определени файлове в текущо активната работна директория, по-стари от дефинираното време в минути. *Забележка: изтриват се само безопасни файлове, т.е. такива, чието изтриване няма да доведе до проблем. Дефиницията на безопасни файлове е извън обхвата на тази задача.*

Напишете shell скрипт, който може да се изпълнява само от някой от указаните два акаунта и приема задължителен първи позиционен аргумент число (в часове, минимална стойност 2 часа). Скриптът трябва да почиства безопасните файлове по-стари от зададените часове в интересните ADR Home-ове на съответния потребител.

Зад. 46 2022-CE-01

Описание на формат на CSV (текстови) файл:

- CSV файлът представлява таблица, като всеки ред на таблицата е записан на отделен ред;
- на даден ред всяко поле (колона) е разделено от останалите със запетая;
- в полетата не може да присъства запетая, т.е. запетаята винаги е разделител между полетата;
- броят на полетата на всеки ред е константа;
- първият ред във файла е header, който описва имената на колоните.

Разполагате с два CSV файла със следното примерно съдържание:

- **base.csv**:

```
unit name,unit symbol,measure
second,s,time
metre,m,length
kilogram,kg,mass
ampere,A,electric current
kelvin,K,thermodynamic temperature
mole,mol,amount of substance
candela,cd,luminous intensity
```
- **prefix.csv**:

```
prefix name,prefix symbol,decimal
tera,T,1000000000000
giga,G,1000000000
mega,M,1000000
mili,m,0.001
nano,n,0.000000001
```

Където смисълът на колоните е:

- за **base.csv**
 - **unit name** – име на мерна единица
 - **unit symbol** – съкратено означение на мерната единица
 - **measure** – величина
- за **prefix.csv**
 - **prefix name** – име на представка
 - **prefix symbol** – означение на представка
 - **decimal** – стойност

Забележка: Във файловете може да има и други редове, освен показаните в примера. Приемаме, че файловете спазват описания по-горе формат, т.е. не е необходимо да проверявате формата.

Напишете shell скрипт, който приема три задължителни параметъра: *число*, *prefix symbol* и *unit symbol*.

Скриптът, ако може, трябва да извежда числото в основната мерна единица без представка, добавяйки в скоби величината и името на мерната единица.

Примерен вход и изход:

```
$ ./foo.sh 2.1 M s
```

2100000.0 s (time, second)

Забележка: За изчисления може да ползвате `bc`.

Зад. 47 2022-CE-02 Съвременните компютри могат да влизат в различни режими за енергоспестяване (suspend) и излизат от този режим (wake-up) при настъпването на определени събития. Linux kernel предоставя специален виртуален файл `/proc/acpi/wakeup`, чрез който може да се проверява или променя настройката за “събуждане” при настъпване на различни събития. Тъй като този файл е интерфейс към ядрото, “четенето” от файла и “писането” във файла изглеждат различно.

За улеснение на задачата ще ползваме опростено описание на този файл.

Под *whitespace* разбираме произволна комбинация от табове и интервали.

При четене от файла изходът представлява четири колони, разделени с *whitespace*, в полетата не може да има *whitespace*; първият ред е header на съответната колона.

Примерно съдържание на файла:

Device	S-state	Status	Sysfs node
GLAN	S4	*enabled	pci:0000:00:1f.6
XHC	S3	*enabled	pci:0000:00:14.0
XDCI	S4	*disabled	
LID	S4	*enabled	platform:PNPOCOD:00
HDAS	S4	*disabled	pci:0000:00:1f.3
RP01	S4	*enabled	pci:0000:00:1c.0

където:

- първата колона дава уникално име на устройство, което може да събуди машината, като името е ограничено до четири знака главни латински букви и цифри;
- третата колона описва дали това устройство може да събуди машината. Възможните стойности са *enabled/disabled*, предхождани от звездичка;
- втората и четвъртата колона ги игнорираме в рамките на задачата.

При записване на име на устройство във файла `/proc/acpi/wakeup`, неговото състояние се променя от *disabled* на *enabled* или обратно.

Например, ако `/proc/acpi/wakeup` изглежда както примера по-горе, при изпълнение на командата `echo XHC > /proc/acpi/wakeup`, третият ред ще се промени на:

XHC	S3	*disabled	pci:0000:00:14.0
-----	----	-----------	------------------

При изпълнение на командата `echo HDAS > /proc/acpi/wakeup`, шестият ред ще се промени на:

HDAS	S4	*enabled	pci:0000:00:1f.3
------	----	----------	------------------

Напишете shell скрипт, който при подаден първи параметър име на устройство (напр. `LID`) настройва съответното устройство да **не може** да събуди машината.

Зад. 48 2022-IN-01

Както знаете, при отваряне на файл с редактора `vi`, той създава в същата директория временен файл с име в следния формат: точка, името на оригиналния файл, точка, `swp`. Например, при редактиране на файл с име `foo.txt` ще се създаде временен файл с име `.foo.txt.swp`.

Напишете shell скрипт, който приема два задължителни позиционни аргумента – имена на директории. Примерно извикване: `./foo.sh ./dir1 /path/to/dir2/`

В `dir1` може да има файлове/директории, директорията `dir2` трябва да е празна.

Скриптът трябва да копира всички обикновени файлове от `dir1` (и нейните под-директории) в `dir2`, запазвайки директорийната структура, но без да копира временните файлове, създадени от редактора `vi` (по горната дефиниция).

Забележка: За удобство приемаме, че не ни вълнува дали метаданните на обектите (owner, group, permissions, etc.) ще се запазят.

Примерни обекти:

```
dir1/  
dir1/a  
dir1/.a.swp  
dir1/b  
dir1/c/d  
dir1/c/.bar.swp
```

Обекти след изпълнението:

```
dir2/  
dir2/a  
dir2/b  
dir2/c/d  
dir2/c/.bar.swp
```

Зад. 49 2022-IN-02

Името на дадена машина можете да вземете с командата `hostname -s`.

Разполагате с машина, на която е инсталиран специализиран софтуер, който ползва два потребителски акаунта – `oracle` и `grid`.

Всеки от потребителите *би трябвало* да има `environment` променлива `ORACLE_BASE`, която указва абсолютен път до директория във формат `/път/до/дир`.

Всеки от потребителите *би трябвало* да има `environment` променлива `ORACLE_HOME`, която указва абсолютен път до директория във формат `/път/до/дир`. В поддиректорията `bin` на зададената директория *би трябвало* да има изпълним файл с име `sqlplus`.

Всеки от потребителите *би трябвало* да има `environment` променлива `ORACLE_SID` с някакъв низ като стойност.

Ако горните три `environment` променливи съществуват, всеки от потребителите може да изпълнява неговото копие на командата `sqlplus` със следните параметри: `sqlplus -SL "/ as роля" @foo.sql` където `роля` трябва да бъде низът `SYSDBA` при изпълнение от `oracle` и `SYSASM` при изпълнение от `grid`. И в двата случая `sqlplus` изпълнява SQL заявките от файла (`foo.sql`, името на файла няма значение) и извежда изхода на `stdout`. Ако съдържанието на `sql` файла е:

```
SELECT value FROM v$parameter WHERE name = 'diagnostic_dest';  
EXIT;
```

изходът ще бъде стойността на търсения параметър `diagnostic_dest` в следния вид:

```
oracle@astero:~$ sqlplus -SL "/ as sysdba" @a.sql
```

```
VALUE
```

```
-----  
/u01/app/oracle
```

```
oracle@astero:~$
```

Параметърът `diagnostic_dest` може да няма стойност, в който случай изведеният низ ще е празен. Изходът винаги е 5 реда, стойността винаги е на 4-и ред. Ако командата `sqlplus` не се изпълни успешно, тя ще върне ненулев `exit code`.

За всеки от двата акаунта съществува директория, която ще наричаме `diag_base`. Конкретната директория е:

- същата като `ORACLE_BASE`, ако `diagnostic_dest` няма стойност
- същата като `diagnostic_dest`, ако `diagnostic_dest` има стойност

За всеки от двата акаунта *би трябвало* да съществува под-директория на `diag_base` с име `diag`, която ще наричаме `diag_dir`.

Съществуват три множества *интересни* за нас файлове:

- множество **crs** – за потребител **grid**, в *diag_dir* има под-директория **crs**, в която има под-директория с името на машината, в която има под-директория **crs**, в която има под-директория **trace**. *Интересни* за нас файлове в тази директория са такива, чието име завършва на подчертавка-число и имат разширение **trc** или **trm**, например **foo_356.trc**, **bar_40001.trm**.
- множество **tnslsnr** – за потребител **grid**, в *diag_dir* има под-директория **tnslsnr**, в която има под-директория с името на машината, в която има *няколко* директории с различни имена. Във всяка от тези директории има под-директории **alert** и **trace**. *Интересни* за нас са файловете в **alert**, чието име завършва на подчертавка-число и имат разширение **xml** (напр. **baz_78.xml**) и файловете в **trace**, чието име завършва на подчертавка-число и имат разширение **log** (напр. **qux_88231.log**).
- множество **rdbms** – за потребител **oracle**, в *diag_dir* има под-директория **rdbms**, в която има *няколко* под-директории, във всяка от които може да има *няколко* под-директории. *Интересни* за нас са *само* файловете в тези директории на второ ниво, чието име завършва на подчертавка-число и имат разширение **trc** или **trm**, например **corge_27.trc**, **grault_1024.trm**.

Напишете скрипт, който може да се изпълнява само от някой от тези два акаунта, и приема задължителен първи позиционен аргумент число (в дни). В зависимост от това кой потребител изпълнява скрипта, той трябва да извежда на stdout за всяко множество на съответния потребител общият размер (в килобайти) на описаните по-горе *интересни* файлове за които времето на последната промяна (по съдържание) на файла е по-голямо от зададеното като параметър на скрипта.

Примерно изпълнение и изход:

```
oracle@astero:~$ ./foo.sh 42
rdbms: 14400
grid@astero:~$ ./foo.sh 73
crs: 28800
tnslsnr: 33600
```

Забележка: Правилното ползване на временни файлове е разрешено.

Зад. 50 2022-IN-03 Вашите колеги от съседната лаборатория работят по проект, в който до сега се е ползвала *нестандартна* имплементация на командата **java**. По време на миграцията с цел използване на *стандартна* имплементация на командата се оказало, че на много места командата се вика по стария, грешен начин, който вече е невалиден. Вашата задача е да напишете shell скрипт **iwojima.sh**, който работи подобно на старата имплементация на **java**, но вътрешно вика новата имплементация по правилния начин.

Файловият формат JAR (Java ARchive) представлява пакет, който обединява няколко Java class файлове, допълнителни мета-данни и различни ресурси. Файлове в този формат обикновено (но не задължително) имат разширение **.jar**. Част от метаданните във файла указват кой е главният клас, чийто метод **main()** се изпълнява по подразбиране. Въпросният **main()** метод ще получи списък с аргументите, указани при извикването на командата **java** (подобно на **main()** функцията в езика C).

Едно от извикванията на *стандартната* имплементация на командата **java** (единственото, което разглеждаме в тази задача) има следният синтаксис: **java [options] -jar filename [args]** където:

- **options** – незадължителна поредица от опции на командата **java**, разделени с интервал; всяка опция задължително започва с тире;
- **filename** – име на JAR файл, който да бъде изпълнен, използва се винаги с опция **-jar**;
- **args** – незадължителни аргументи, с които ще бъде стартиран **main()** методът на класа по подразбиране (аргументи на самата Java програма).

Като пример разполагаме с JAR файл (**app.jar**), с един клас в него, чийто **main()** метод реализира единствено функционалност, еквивалента на следният bash код:

```
if [ -z "${qux}" ]; then
    qux=default
fi
echo "${qux}"
```

Общият вид на изпълнение с командата `java` би бил:

```
$ java -jar ./app.jar
default
```

Една от стандартните опции на командата `java` е `-Dproperty=value`, с което се дефинира променлива с име `property` и със стойност `value`. При нужда от няколко променливи, за всяка се ползва отделна опция `-D`. Пример за правилно ползване на опцията в *стандартната* имплементация:

```
$ java -Dqux=foo -jar /path/to/app.jar
foo
```

За съжаление, *нестандартната* имплементация на `java`, която е ползвана до момента, работи по друг начин. Пример:

```
$ java -jar ./app.jar -Dqux=foo
default
```

```
$ java -Dqux=foo -jar ./app.jar -hoge fuga
default
```

```
$ java -jar -Dqux=foo app.jar
foo
```

Ако трябва да обобщим как работи *нестандартната* имплементация:

- всичко след името на JAR файла са аргументи за `main()` метода (напр. `-hoge` и `fuga`);
- името на JAR файла е *някъде* след опцията `-jar`;
- опции за `java` командата може да има и преди, и след опцията `-jar`;
- опцията `-Dproperty=value` работи само ако е след опцията `-jar`.

Забележки: Вашият скрипт `iwojima.sh` няма нужда да поддържа извиквания, които не подават опция `-jar`. За удобство приемаме, че опциите (вкл. `property` и `value`), имената на JAR файловете и аргументите не могат да съдържат интервали.

Зад. 51 2022-IN-04

Вашите колеги от съседната лаборатория работят със специализирана система `Fuga`, която се базира на два текстови файла – за автентификация (`foo.pwd`) и основен конфигурационен файл (`foo.conf`). Двата файла би трябвало да се съхраняват в главната директория на системата, която ще наричаме `fuga`.

На всеки ред във първия файл има потребителско име (малки латински букви) и паролата за този потребител, разделени с двоеточие, като паролата не е в чист вид, а е *хеширана*.

Форматът на конфигурационния файл е недефиниран. Тъй като конфигурационният файл е твърде голям за удобна работа, вашите колеги са решили да го разделят на части в отделни *малки* конфигурационни файлове, които държат в директория `fuga/cfg` и нейните под-директории. Всеки такъв файл има име във формат `bar.cfg`, където `bar` е име на потребител на системата. Колегите ви са написали скрипт (`fuga/validate.sh`), който приема един задължителен позиционен аргумент – име на конфигурационен файл. Скриптът проверява указания конфигурационен файл за валиден синтаксис и може да приключи с някой от следните `exit code`-ове:

- 0 – файлът е валиден;
- 1 – файлът не е валиден;
- 2 – настъпила е грешка при изпълнение на скрипта.

Ако файлът не е валиден, скриптът извежда на `stdout` редовете, на които има грешка, предхождани от `Line x`: където `x` е номера на реда. В останалите случаи скриптът няма изход.

Разполагате с команда `pwgen`, която генерира и извежда на `stdout` случайни пароли (букви и цифри), и знаете, че поддържа следните два аргумента: `pwgen [password_length] [number_of_passwords]`

Също така разполагате с командата `mkpasswd`, която по подаден аргумент – парола в чист вид извежда нейният *хеш* на `stdout`.

Помогнете на колегите си, като напишете шел скрипт, който приема параметър – име на директория `fuga`. Скриптът трябва да:

- извежда на `stderr` грешните редове от *малките* конфигурационни файлове, във формата изведен от валидиращия скрипт, като всеки ред трябва да започва с името на конфигурационния файл и знак двоеточие `':'`;
- (ре-)генерира `foo.conf` като конкатенация на валидните *малки* конфигурационни файлове;
- провери за всеки валиден *малък* конфигурационен файл дали потребителят съществува в автентификационния файл и ако не – да го добави по подходящия начин, като изведе на `stdout` потребителското име и паролата в чист вид, разделени с двоеточие.

Зад. 52 2022-SE-01 Съвременните компютри могат да влизат в различни режими за енергоспестяване (`suspend`) и излизат от този режим (`wake-up`) при настъпването на определени събития. Linux kernel предоставя специален виртуален файл `/proc/acpi/wakeup`, чрез който може да се проверява или променя настройката за “събуждане” при настъпване на различни събития, в общия случай - при активност на някое устройство. Тъй като този файл е интерфейс към ядрото, “четенето” от файла и “писането” във файла изглеждат различно.

За улеснение на задачата ще ползваме опростено описание на този файл.

Под *whitespace* разбираме произволна комбинация от табове и интервали. При четене от файла изходът представлява четири колони, разделени с `whitespace`, в полетата не може да има `whitespace`; първият ред е `header` на съответната колона. Примерно съдържание на файла:

Device	S-state	Status	Sysfs node
GLAN	S4	*enabled	pci:0000:00:1f.6
XHC	S3	*enabled	pci:0000:00:14.0
XDCI	S4	*disabled	
LID	S4	*enabled	platform:PNPOCOD:00
HDAS	S4	*disabled	pci:0000:00:1f.3
RP01	S4	*enabled	pci:0000:00:1c.0

където:

- първата колона дава уникално име на устройство, което може да събуди машината, като името е ограничено до четири знака главни латински букви и цифри;
- третата колона описва дали това устройство може да събуди машината. Възможните стойности са `enabled/disabled`, предхождани от звездичка;
- втората и четвъртата колона ги игнорираме в рамките на задачата.

При записване на име на устройство във файла `/proc/acpi/wakeup`, неговото състояние се променя от `disabled` на `enabled` или обратно. Например, ако файлът изглежда както примера по-горе, при запис на `XHC` в него, третият ред ще се промени на:

XHC	S3	*disabled	pci:0000:00:14.0
-----	----	-----------	------------------

При запис на `HDAS`, шестият ред ще се промени на:

HDAS	S4	*enabled	pci:0000:00:1f.3
------	----	----------	------------------

Дефиниран е формат на конфигурационен файл, който описва желан комплект от настройки на `wakeup` събития. Примерен файл:

```
# comment bar
GLAN    disabled
LID     enabled    # comment foo
```

където:

- знакът `диез (#)` е знак за коментар до края на реда;
- редовете *би трябвало* да са комбинация от име на устройство и желаното състояние на настройката за събуждане при събития от това устройство, разделени с `whitespace`.

Напишете скрипт, който при подаден първи параметър име на конфигурационен файл в горния формат осигурява исканите настройки за събуждане. Ако във файла има ред за устройство, което не съществува, скриптът да извежда предупреждение. Обърнете внимание на обработката за грешки и съобщенията към потребителя – искаме скриптът да бъде удобен и валиден инструмент.

Зад. 53 2022-SE-02 Дефинирана е система за бекъпване на сървъри, която държи направените архиви в главна директория (която ще наричаме *fubar*), в която има четири под-директории за различни *класове* бекъпи:

- 0 – съдържа годишни бекъпи
- 1 – съдържа месечни бекъпи
- 2 – съдържа седмични бекъпи
- 3 – съдържа дневни бекъпи

Всяка директория съдържа архивни файлове с имена във формат `hostname-area-YYYYMMDD.tar.xz`, където:

- `hostname` е името на някаква машина, която е бекъпвана;
- `area` е типът бекъп за съответната машина;
- `YYYYMMDD` е датата, на която е направен бекъпа;
- никое от горните полета не може да съдържа тире или точка;
- някои от файловете могат да са symlink-ове.

Примерни имена на файлове:

```
astero-etc-20220323.tar.xz   stratios-etc-20220428.tar.xz   nestor-db-20220404.tar.xz
gila-srv-20220405.tar.xz    catalyst-var-20220406.tar.xz   drake-home-20220404.tar.xz
dominix-var-20220404.tar.xz
```

Комбинацията от `hostname` и `area` дефинира уникален *обект* за бекъпване. Всички архивни файлове са *пълноценни* бекъпи на даден *обект* и са достатъчни за неговото възстановяване при нужда (*заб. извън обхвата на задачата*).

Ако даден файл е symlink, то той може да е валиден или счупен. Symlink-овете са създадени за удобство и не ги считаме за *пълноценни* бекъпи.

Политиката ни за бекъп казва, че за да имаме *валиден* бекъп на даден *обект*, за него трябва да имаме минимум 1 годишен, 2 месечни, 3 седмични и 4 дневни *пълноценни* бекъпа.

Важност:

- *обектите* са равни по важност помежду си;
- важността на *класовете* бекъпи е във възходящ ред по горния списък, т.е. при равни други условия дневните бекъпи са по-ценни от седмичните и т.н.;
- в рамките на един *клас* бекъпи по-новите бекъпи са по-важни от по-старите.

Напишете shell скрипт, който приема два два задължителни позиционни аргумента – име на директория и число в интервала [1, 99]. Примерно извикване: `./foo.sh ./bar/ 30` където:

- директорията представлява главна директория (*fubar*) на описаната система за бекъпване;
- числото дефинира колко процента е максималното допустимо използвано място на файловата система, в която се намира подадената директория.

За удобство приемаме, че директорията *fubar* и всички обекти в нея се намират в една и съща файлова система.

Упътване: Командата `df` извикана с аргумент име на файл/директория връща информация за файловата система, в която той се намира. Пример:

```
$ df ./README.md
Filesystem      1K-blocks      Used Available Use% Mounted on
/dev/mapper/o7-hm 100663296 61735732  37288764   63% /home
```

Скриптът трябва да изтрива *минимален* брой *пълноценни* архивни файлове така, че:

- всеки *обект* да има *валиден* бекъп;
- обръща внимание на описаните по-горе важности;

- процентите използвано място върху файловата система да е *не повече* от подаденият параметър (а ако това е невъзможно, скриптът да освободи колкото може повече място, без да нарушава *валидностите* на бекъпите на *обектите*);
- не е допустимо след работата на скрипта да останат счупени symlink-ове.

Зад. 54 2023-IN-01

Вашите колеги от съседната лаборатория имат няколко GNU/Linux машини със специфични RAID контролери и ползват специализирана система за наблюдение Hugin. Помогнете на колегите си, като напишете shell скрипт, който взима температурата на дисковете, закачени към тези контролери, и генерира изход, удобен за ползване от системата за наблюдение.

За достъп до данните на дисковете, колегите ви казват, че root потребителя може да изпълни следната команда, която взима информация от даден контролер:

```
ssaccli ctrl slot=x pd all show detail
```

където *x* е номер на PCI слот, в който е включен контролера. Примерен изход от командата:

```
Smart Array P420i in Slot 0 (Embedded)
```

```
Array A
```

```
physicaldrive 1I:2:1
  Rotational Speed: 15000
  Current Temperature (C): 35
```

```
physicaldrive 1I:2:2
  Interface Type: SAS
  Current Temperature (C): 36
  SMR Support: None
```

```
physicaldrive 1I:2:3
  Status: OK
  Drive Type: Spare Drive
  Current Temperature (C): 34
  Maximum Temperature (C): 42
```

Още един пример:

```
Smart Array P822 in Slot 1
```

```
Array J
```

```
physicaldrive 2E:1:19
  Current Temperature (C): 29
```

```
physicaldrive 2E:1:20
  Current Temperature (C): 28
```

```
Unassigned
```

```
physicaldrive 2E:1:21
  Current Temperature (C): 30
```

```
physicaldrive 2E:1:22
  Current Temperature (C): 29
```

Ако обобщим, изходът изглежда така:

- обща информация за контролера в съответния слот, и частност, *модел на контролера* (P420i/P822)
- една или повече секции на второ ниво, дефиниращи (нула или повече) масиви или (нула или една) секция с неизползвани дискове (Array A/Array J/Unassigned), като буквата (A, J, ...) дефинира *името на масива* в рамките на контролера
- във всяка такава секция има една или повече подсекции за дискове в съответния масив (physicaldrive 1I:2:1, physicaldrive 2E:1:22, и т.н.), където низът след physicaldrive е *име на този диск* в рамките на контролера

- във всяка подсекция за даден диск има *множество* параметри във вида *key name: value* като за колегите ви е интересен само параметърът **Current Temperature (C)**

Нека въведем следните термини:

- *идентификатор на диск* – низ във вида “SSAсmmmappp”, където:
 - SSA – литерал
 - с – номер на слот
 - mmm – модел на контролер
 - a – име на масив или UN при неизползван диск
 - ppp – име на диск, без двуеточия
- *label на диск* – низ във вида “SSAc mmm a qqq”, където
 - qqq – име на диск
 - останалите полета са по горната дефиниция
- *забележка*: кавичките не са част от идентификатора или от label-a на даден диск.

Възможно е да съществува *environment* променлива CTRLSLOTS, чиято стойност да е един или повече номера на PCI слотове, разделени с интервали, и в този случай скриптът трябва да извежда информация от тези контролери. Ако такава променлива не е дефинирана, скриптът по подразбиране да работи с контролера в слот 0.

Скриптът трябва да поддържа следните видове изпълнение:

- с един параметър **autoconf** – скриптът трябва да извежда низа **yes**
- с един параметър **config** – скриптът трябва да извежда следния изход:

```
graph_title SSA drive temperatures
graph_vlabel Celsius
graph_category sensors
graph_info This graph shows SSA drive temp
SSA0P420iA1I21.label SSA0 P420i A 1I:2:1
SSA0P420iA1I21.type GAUGE
SSA0P420iA1I22.label SSA0 P420i A 1I:2:2
SSA0P420iA1I22.type GAUGE
SSA0P420iA1I23.label SSA0 P420i A 1I:2:3
SSA0P420iA1I23.type GAUGE
SSA1P822J2E119.label SSA1 P822 J 2E:1:19
SSA1P822J2E119.type GAUGE
SSA1P822J2E120.label SSA1 P822 J 2E:1:20
SSA1P822J2E120.type GAUGE
SSA1P822UN2E121.label SSA1 P822 UN 2E:1:21
SSA1P822UN2E121.type GAUGE
SSA1P822UN2E122.label SSA1 P822 UN 2E:1:22
SSA1P822UN2E122.type GAUGE
```

- без параметри – скриптът трябва да извежда следния изход:

```
SSA0P420iA1I21.value 35
SSA0P420iA1I22.value 36
SSA0P420iA1I23.value 34
SSA1P822J2E119.value 29
SSA1P822J2E120.value 28
SSA1P822UN2E121.value 30
SSA1P822UN2E122.value 29
```

В горните изходи:

- низовете, започващи с **graph** са литерали
- низовете, в които се среща **.label**, **.type** или **.value** започват с *идентификатор на диск*
- низът след **.label** е *label на диск*
- **.type** винаги е GAUGE
- числото след **.value** е температурата на въпросния диск

Вашите колеги от съседната лаборатория ползват специализиран софтуер за оптометрични изследвания, който записва резултатите от всяко измерване в отделен файл. Файловете имат уникално съдържание, по което се определя за кое измерване се отнася файла. За съжаление, тъй като колегите ви ползват бета версия на софтуера, той *понякога* записва по няколко пъти резултатите от дадено измерване в произволна комбинация от следните варианти:

- нула или повече отделни обикновени файла с еднакво съдържание;
- нула или повече групи от hardlink-ове, като всяка група съдържа две или повече имена на даден файл с измервания.

Помогнете на колегите си, като напишете shell скрипт, който приема параметър – име на директория, съдържаща файлове с измервания. Скриптът трябва да извежда на стандартния изход списък с имена на файлове, кандидати за изтриване, по следните критерии:

- ако измерването е записано само в отделни файлове, трябва да остане един от тях;
- ако измерването е записано само в групи от hardlink-ове, всяка група трябва да се намали с едно име;
- ако измерването е записано и в групи, и като отделни файлове, за групите се ползва горния критерий, а всички отделни файлове се премахват.

2 Използване на системни примитиви в програми на C

Забележки:

- Полезни man страници:

open(2)	close(2)	read(2)	write(2)	lseek(2)
fork(2)	wait(2)	exec(3)		
pipe(2)	dup(2)	mkfifo(3)		
err(3)	exit(3)	printf(3)	setbuf(3)	
malloc(3)	stat(2)	time(2)	sleep(3)	
strlen(3)	strcmp(3)	qsort(3)	strtol(3)	
- Задачите ви трябва да се компилират с осигурения **Makefile**.
- Обърнете внимание на коментарите, именуването на променливи и подреждането на кода.
- За всички задачи, освен ако не е указано друго:
 - за удобство приемаме, че работим само с Little-endian машини;
 - не е гарантирано, че разполагате с достатъчно памет, в която да заредите всички данни.

2.1 Вход и изход (тема 6)

Зад. 56 2016-SE-01 Напишете програма на C, която приема параметър - име на (двоичен) файл с байтове. Програмата трябва да сортира файла.

Зад. 57 2016-SE-02 Двоичните файлове `f1` и `f2` съдържат 32 битови числа без знак (`uint32_t`). Файлът `f1` е съдържа n двойки числа, нека i -тата двойка е $< x_i, y_i >$. Напишете програма на C, която извлича интервалите с начало x_i и дължина y_i от файла `f2` и ги записва залепени в изходен файл `f3`.

Пример:

- `f1` съдържа 4 числа (2 двойки): 30000, 20, 19000, 10
- програмата записва в `f3` две поредици 32-битови числа, взети от `f2` както следва:
- най-напред се записват числата, които са на позиции 30000, 30001, 30002, ... 30019.
- след тях се записват числата от позиции 19000, 19001, ... 19009.

Забележка: C пълен брой точки ще се оценяват решения, които работят със скорост, пропорционална на размера на изходния файл `f3`.

Зад. 58 2016-SE-03 Напишете програма на C приемаща параметър – име на (двоичен) файл с `uint32_t` числа. Програмата трябва да сортира файла. Ограничения:

- Числата биха могли да са максимум 100 000 000 на брой.
- Програмата трябва да работи на машина със същия `endianness`, както машината, която е създала файла.
- Програмата трябва да работи на машина с 256 MB RAM и 8 GB свободно дисково пространство.

Зад. 59 2017-IN-01 Напишете програма на C, която приема четири параметъра – имена на двоични файлове.

Примерно извикване:

```
$ ./main f1.dat f1.idx f2.dat f2.idx
```

Първите два (`f1.dat` и `f1.idx`) и вторите два (`f2.dat` и `f2.idx`) файла са *входен* и *изходен комплект* със следния смисъл:

- DAT-файловете (`f1.dat` и `f2.dat`) представляват двоични файлове, състоящи се от байтове (`uint8_t`);
- IDX-файловете представляват двоични файлове, състоящи се от наредени тройки от следните елементи (и техните типове), които дефинират поредици от байтове (низове) от съответния DAT файл:
 - *отместване* `uint16_t` – показва позицията на първия байт от даден низ спрямо началото на файла;
 - *дължина* `uint8_t` – показва дължината на низа;

– запазен `uint8_t` – не се използва.

Първата двойка файлове (`f1.dat` и `f1.idx`) съществува, а втората трябва да бъде създадена от програмата по следния начин:

- трябва да се копират само низовете (порединици от байтове) от входния комплект, които започват с главна латинска буква (A – 0x41, Z – 0x5A).
- ако файловете са неконсистентни по някакъв начин, програмата да прекратява изпълнението си по подходящ начин.

Забележка: За удобство приемаме, че DAT файлът съдържа текстови данни на латински с ASCII кодова таблица (един байт за буква).

Примерен вход и изход:

```
$ xxd f1.dat
00000000: 4c6f 7265 6d20 6970 7375 6d20 646f 6c6f  Lorem ipsum dolo
00000010: 7220 7369 7420 616d 6574 2c20 636f 6e73  r sit amet, cons
00000020: 6563 7465 7475 7220 6164 6970 6973 6369  ectetur adipisci
00000030: 6e67 2065 6c69 742c 2073 6564 2064 6f20  ng elit, sed do
00000040: 6569 7573 6d6f 6420 7465 6d70 6f72 2069  eiusmod tempor i
00000050: 6e63 6964 6964 756e 7420 7574 206c 6162  ncididunt ut lab
00000060: 6f72 6520 6574 2064 6f6c 6f72 6520 6d61  ore et dolore ma
00000070: 676e 6120 616c 6971 7561 2e20 5574 2065  gna aliqua. Ut e
00000080: 6e69 6d20 6164 206d 696e 696d 2076 656e  nim ad minim ven
00000090: 6961 6d2c 2071 7569 7320 6e6f 7374 7275  iam, quis nostru
000000a0: 6420 6578 6572 6369 7461 7469 6f6e 2075  d exercitation u
000000b0: 6c6c 616d 636f 206c 6162 6f72 6973 206e  llamco laboris n
000000c0: 6973 6920 7574 2061 6c69 7175 6970 2065  isi ut aliquip e
000000d0: 7820 6561 2063 6f6d 6d6f 646f 2063 6f6e  x ea commodo con
000000e0: 7365 7175 6174 2e20 4475 6973 2061 7574  sequat. Duis aut
000000f0: 6520 6972 7572 6520 646f 6c6f 7220 696e  e irure dolor in
00000100: 2072 6570 7265 6865 6e64 6572 6974 2069  reprehenderit i
00000110: 6e20 766f 6c75 7074 6174 6520 7665 6c69  n voluptate veli
00000120: 7420 6573 7365 2063 696c 6c75 6d20 646f  t esse cillum do
00000130: 6c6f 7265 2065 7520 6675 6769 6174 206e  lore eu fugiat n
00000140: 756c 6c61 2070 6172 6961 7475 722e 2045  ulla pariatu. E
00000150: 7863 6570 7465 7572 2073 696e 7420 6f63  xcepteur sint oc
00000160: 6361 6563 6174 2063 7570 6964 6174 6174  caecat cupidatat
00000170: 206e 6f6e 2070 726f 6964 656e 742c 2073  non proident, s
00000180: 756e 7420 696e 2063 756c 7061 2071 7569  unt in culpa qui
00000190: 206f 6666 6963 6961 2064 6573 6572 756e  officia deserun
000001a0: 7420 6d6f 6c6c 6974 2061 6e69 6d20 6964  t mollit anim id
000001b0: 2065 7374 206c 6162 6f72 756d 2e0a      est laborum..
```

```
$ xxd f1.idx
00000000: 0000 0500 4f01 0200 4e01 0300      ....O...N...
```

```
$ xxd f2.dat
00000000: 4c6f 7265 6d45 78      LoremEx
```

```
$ xxd f2.idx
00000000: 0000 0500 0500 0200      .....
```

Зад. 60 2017-SE-01 Напишете програма на C, която приема три параметъра – имена на двоични файлове.

Примерно извикване:

```
$ ./main f1.bin f2.bin patch.bin
```

Файловете `f1.bin` и `f2.bin` се третираат като двоични файлове, състоящи се от байтове (`uint8_t`). Файлът `f1.bin` е “оригиналният” файл, а `f2.bin` е негово копие, което е било модифицирано по

някакъв начин (извън обхвата на тази задача). Файлът `patch.bin` е двоичен файл, състоящ се от наредени тройки от следните елементи (и техните типове):

- *отместване* (`uint16_t`) – спрямо началото на `f1.bin/f2.bin`
- *оригинален байт* (`uint8_t`) – на тази позиция в `f1.bin`
- *нов байт* (`uint8_t`) – на тази позиция в `f2.bin`

Вашата програма да създава файла `patch.bin`, на базата на съществуващите файлове `f1.bin` и `f2.bin`, като описва вътре само разликите между двата файла. Ако дадено отместване съществува само в единия от файловете `f1.bin/f2.bin`, програмата да прекратява изпълнението си по подходящ начин.

Примерен f1.bin:

```
00000000: f5c4 b159 cc80 e2ef c1c7 c99a 2fb0 0d8c ...Y...../...
00000010: 3c83 6fed 6b46 09d2 90df cf1e 9a3c 1f05 <.o.kF.....<..
00000020: 05f9 4c29 fd58 a5f1 cb7b c9d0 b234 2398 ..L).X...{...4#.
00000030: 35af 6be6 5a71 b23a 0e8d 08de def2 214c 5.k.Zq.:.....!L
```

Примерен f2.bin:

```
00000000: f5c4 5959 cc80 e2ef c1c7 c99a 2fb0 0d8c ..YY...../...
00000010: 3c83 6fed 6b46 09d2 90df cf1e 9a3c 1f05 <.o.kF.....<..
00000020: 05f9 4c29 fd58 a5f1 cb7b c9d0 b234 2398 ..L).X...{...4#.
00000030: afaf 6be6 5a71 b23a 0e8d 08de def2 214c ..k.Zq.:.....!L
```

Примерен patch.bin:

```
00000000: 0200 b159 3000 35af ...Y0.5.
```

Зад. 61 2017-SE-02 Напишете програма на C, която да работи подобно на командата `cat`, реализирайки само следната функционалност:

- общ вид на изпълнение: `./main [OPTION] [FILE]...`
- ако е подаден като **първи** параметър `-n`, то той да се третира като опция, което кара програмата ви да номерира (глобално) всеки изходен ред (започвайки от 1).
- програмата извежда на `STDOUT`
- ако няма подадени параметри (имена на файлове), програмата чете от `STDIN`
- ако има подадени параметри – файлове, програмата последователно ги извежда
- ако някой от параметрите е тире (`-`), програмата да го третира като специално име за `STDIN`

Примерно извикване:

```
$ cat a.txt
a1
a2

$ cat b.txt
b1
b2
b3

$ echo -e "s1\ns2" | ./main -n a.txt - b.txt
1 a1
2 a2
3 s1
4 s2
5 b1
6 b2
7 b3
```


Забележка: Погледнете `setbuf(3)` и `strcmp(3)`.

Зад. 62 2017-SE-03 Напишете програма на C, която приема три параметъра, имена на двоични файлове.

Примерно извикване:

```
$ ./main patch.bin f1.bin f2.bin
```

Файловете `f1.bin` и `f2.bin` се третират като двоични файлове, състоящи се от байтове (`uint8_t`). Файлът `patch.bin` е двоичен файл, състоящ се от наредени тройки от следните елементи (и техните типове):

- *отместване* `uint16_t`
- *оригинален байт* `uint8_t`
- *нов байт* `uint8_t`

Програмата да създава файла `f2.bin` като копие на файла `f1.bin`, но с отразени промени на базата на файла `patch.bin`, при следния алгоритъм:

- за всяка наредена тройка от `patch.bin`, ако на съответното *отместване* (в байтове) спрямо началото на файла е записан байта *оригинален байт*, в изходния файл се записва *нов байт*. Ако не е записан такъв *оригинален байт* или такова *отместване* не съществува, програмата да прекратява изпълнението си по подходящ начин;
- всички останали байтове се копират директно.

Забележка: Наредените тройки във файла `patch.bin` да се обработват последователно.

Примерен f1.bin:

```
00000000: f5c4 b159 cc80 e2ef c1c7 c99a 2fb0 0d8c ...Y...../...
00000010: 3c83 6fed 6b46 09d2 90df cf1e 9a3c 1f05 <.o.kF.....<..
00000020: 05f9 4c29 fd58 a5f1 cb7b c9d0 b234 2398 ..L).X...{...4#.
00000030: 35af 6be6 5a71 b23a 0e8d 08de def2 214c 5.k.Zq.:.....!L
```

Примерен patch.bin:

```
00000000: 0200 b159 3000 35af ...Y0.5.
```

Примерен f2.bin:

```
00000000: f5c4 5959 cc80 e2ef c1c7 c99a 2fb0 0d8c ..YY...../...
00000010: 3c83 6fed 6b46 09d2 90df cf1e 9a3c 1f05 <.o.kF.....<..
00000020: 05f9 4c29 fd58 a5f1 cb7b c9d0 b234 2398 ..L).X...{...4#.
00000030: afaf 6be6 5a71 b23a 0e8d 08de def2 214c ..k.Zq.:.....!L
```

Зад. 63 2017-SE-04 Напишете програма на C, която да работи подобно на командата `cat`, реализирайки само следната функционалност:

- програмата извежда на `STDOUT`
- ако няма подадени параметри, програмата чете от `STDIN`
- ако има подадени параметри – файлове, програмата последователно ги извежда
- ако някой от параметрите започва с тире (-), програмата да го третира като специално име за `STDIN`

Примерно извикване:

```
$ ./main f - g
```

– извежда съдържанието на файла `f`, после `STDIN`, след това съдържанието на файла `g`.

Зад. 64 2018-SE-01 Напишете програма на C, която да работи подобно на командата `tr`, реализирайки само следната функционалност:

- програмата чете от стандартния вход и пише на стандартния изход

- общ вид на изпълнение: `./main [OPTION] SET1 [SET2]`
- `OPTION` би могъл да бъде или `-d`, или `-s`, или да липсва
- `SET1` и `SET2` са низове, в които знаците нямат специално значение, т.е. всеки знак “означава” съответния знак. `SET2`, ако е необходим, трябва да е същата дължина като `SET1`
- ако е подаден като *първи* параметър `-d`, програмата трие от входа всяко срещане на знак μ от `SET1`; `SET2` не е необходим
- ако е подаден като *първи* параметър `-s`, програмата заменя от входа всяка поредица от повтोरения знак μ от `SET1` с еднократен знак μ ; `SET2` не е необходим
- в останалите случаи програмата заменя от входа всеки знак μ от `SET1` със съответстващият му позиционно знак ν от `SET2`.

Примерно извикване:

```
$ echo asdf | ./main -d 'd123' | ./main 'sm' 'fa' | ./main -s 'f'
af
```

Зад. 65 2018-SE-02 Напишете програма на C, която приема два параметъра – имена на файлове:

- примерно извикване: `./main input.bin output.bin`
- файловете `input.bin` и `output.bin` се третират като двоични файлове, състоящи се от `uint32_t` числа
- файлът `input.bin` може да съдържа максимум 4194304 числа
- файлът `output.bin` трябва да бъде създаден от програмата и да съдържа числата от `input.bin`, сортирани във възходящ ред
- *endianness*-ът на машината, създава файла `input.bin` е същият, като на текущата машина
- *ограничения на ресурси*: програмата трябва да работи с употреба на максимум 9 MB RAM и 64 MB дисково пространство.

Зад. 66 2018-SE-03 Напишете програма на C, която да работи подобно на командата `cut`, реализирайки само следната функционалност:

- програмата трябва да чете текст от стандартния вход и да извежда избраната част от всеки ред на стандартния изход;
- ако първият параметър на програмата е низът `-c`, тогава вторият параметър е или едноцифрено число (от 1 до 9), или две едноцифрени числа N и M ($N \leq M$), разделени с тире (напр. 3-5). В този случай програмата трябва да изведе избраният/избраните символи от реда: или само символа на указаната позиция, или няколко последователни символи на посочените позиции.
- ако първият параметър на програмата е низът `-d`, тогава вторият параметър е низ, от който е важен само първият символ; той се използва като разделител между полета на реда. Третият параметър трябва да бъде низът `-f`, а четвъртият – или едноцифрено число (от 1 до 9), или две едноцифрени числа N и M ($N \leq M$), разделени с тире (напр. 3-5). В този случай програмата трябва да разглежда реда като съставен от няколко полета (може би празни), разделени от указания символ (първият символ от низа, следващ параметъра `-d`), и да изведе или само указаното поле, или няколко последователни полета на указаните позиции, разделени от същия разделител.
- ако някой ред няма достатъчно символи/полета, за него програмата трябва да изведе каквото (докъдето) е възможно (или дори празен ред).

Зад. 67 2018-SE-04 Напишете програма на C, която приема два параметъра – имена на файлове:

- примерно извикване: `./main input.bin output.bin`
- файловете `input.bin` и `output.bin` се третират като двоични файлове, състоящи се от `uint16_t` числа
- файлът `input.bin` може да съдържа максимум 65535 числа
- файлът `output.bin` трябва да бъде създаден от програмата и да съдържа числата от `input.bin`, сортирани във възходящ ред
- *endianness*-ът на машината, създава файла `input.bin` е същият, като на текущата машина
- *ограничения на ресурси*: програмата трябва да работи с употреба на максимум 256 KB RAM и 2 MB дисково пространство.

Зад. 68 2019-SE-01 В приложната статистика често се използват следните описателни статистики за дадена извадка:

- *средна стойност* $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
- *дисперсия* $D = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$

където $\{x_1, x_2, \dots, x_N\}$ са стойностите на извадката, а N е техният брой.

Напишете програма на C, която приема задължителен параметър – име на двоичен файл. Файлът съдържа информация за потребители, които са влизали в системата, и се състои от наредени петорки от следните елементи и техните типове:

- *уникален потребителски идентификатор (UID)* `uint32_t`
- *запазено1* `uint16_t` – не се използва в задачата
- *запазено2* `uint16_t` – не се използва в задачата
- *време1* `uint32_t` – момент на започване на сесията (Unix time)
- *време2* `uint32_t` – момент на завършване на сесията (Unix time)

За потребители, които са имали сесии, квадратът на продължителността на които е по-голям от дисперсията D на продължителността на всички сесии във файла, програмата трябва да изведе на `STDOUT` потребителският им идентификатор и продължителността на най-дългата им сесия.

Можете да разчитате на това, че във файла ще има не повече от 16384 записа и че в тях ще се срещат не повече от 2048 различни потребителски идентификатора.

Зад. 69 2020-IN-01 Напишете програма на C, която приема три параметъра – имена на двоични файлове.

Примерно извикване:

```
$ ./main patch.bin f1.bin f2.bin
```

Файловете `patch.bin` и `f1.bin` съществуват, и на тяхна база програмата трябва да създаде `f2.bin`.

Файлът `patch.bin` се състои от две секции – 16 байтов хедър и данни. На базата на хедъра програмата трябва да може да интерпретира съдържанието на файла. Структурата на хедъра е:

- `uint32_t, magic` – магическа стойност `0xEFBEADDE`, която дефинира, че файлът следва тази спецификация
- `uint8_t, header version` – версия на хедъра, с единствена допустима стойност за момента `0x01`, която дефинира останалите байтове от хедъра както следва:
 - `uint8_t, data version` – версия (описание) на използваните структури в секцията за данни на файла
 - `uint16_t, count` – брой записи в секцията за данни
 - `uint32_t, reserved 1` – не се използва
 - `uint32_t, reserved 2` – не се използва

Възможни структури в секцията за данни на файла спрямо *data version*:

- при версия `0x00`
 - `uint16_t, offset`
 - `uint8_t, original byte`
 - `uint8_t, new byte`
- при версия `0x01`
 - `uint32_t, offset`
 - `uint16_t, original word`
 - `uint16_t, new word`
- *забележка:* и при двете описани версии *offset* е отместване в брой елементи спрямо началото на файла

Двоичните файлове `f1.bin` и `f2.bin` се третират като състоящи се от елементи спрямо *data version* в `patch.bin`.

Програмата да създава файла `f2.bin` като копие на файла `f1.bin`, но с отразени промени на базата на файла `patch.bin`, при следния алгоритъм:

- за всяка наредена тройка от секцията за данни на `patch.bin`, ако на съответният *offset* в оригиналния файл `f1.bin` е записан елементът *original byte/word*, в изходният файл се записва *new byte/word*. Ако не е записан такъв елемент или той не съществува, програмата да прекратява изпълнението си по подходящ начин;

- всички останали елементи се копират директно.

Наредените тройки в секцията за данни на файла `patch.bin` да се обработват последователно. Обърнете внимание на обработката за грешки и съобщенията към потребителя – искаме програмата да бъде удобен и валиден инструмент.

Зад. 70 2020-SE-01 Напишете програма на C, която приема шест задължителни позиционни параметъра – имена на файлове. Примерно извикване:

```
$ ./main affix postfix prefix infix suffix crucifixus
```

Всички файлове започват с хедър с фиксирана дължина от 16 байта. Пети и шести (спрямо \mathbb{Z}^+) байт от хедъра дефинират `uint16_t` число *count*, което описва броя на елементите във файла. Файловете `affix` и `infix` се състоят от елементи от тип `uint16_t`, файловете `prefix` и `crucifixus` – от елементи от тип `uint8_t`, `postfix` – от `uint32_t`, а `suffix` – от `uint64_t`.

Интервал наричаме наредена двойка числа, която дефинира номер (спрямо \mathbb{Z}) на начален елемент и брой елементи от даден файл. *Комплект* наричаме наредена четворка от интервали, които позиционно се отнасят спрямо файловете `{post,pre,in,suf}fix`.

Елементите на файла `affix` дефинират серия от комплекти, на чиято база програмата трябва да генерира изходния файл `crucifixus`.

Зад. 71 2020-SE-02 Инженерите от съседната лаборатория работят с комплекти SCL/SDL файлове, напр. `input.scl/input.sdl`.

В SCL файла са записани нива на сигнали (ниско 0 или високо 1), т.е., файлът се третира като състоящ се от битове. В SDL файла са записани данни от тип `uint16_t`, като всеки елемент съответства позиционно на даден бит от SCL файла.

Помогнете на колегите си, като напишете програма на C, която да бъде удобен инструмент за изваждане в нов файл само на тези SDL елементи, които са имали високо ниво в SCL файла, запазвайки наредбата им.

Зад. 72 2021-SE-01 Инженерите от съседната лаборатория ползват специализиран хардуер и софтуер за прехвърляне на данни по радио, но за съжаление имат два проблема:

- в радио частта: дълги поредици битове само 0 или само 1 чупят преноса;
- в софтуерната част: софтуерът, който ползват, може да прехвърля само файлове с четен брой байтове дължина.

Помогнете на колегите си, като напишете програма на C, която решава тези проблеми, като подготвя файлове за прехвърляне. Програмата трябва да приема два задължителни позиционни аргумента – имена на файлове. Примерно извикване:

```
$ ./main input.bin output.bin
```

Програмата чете данни от `input.bin` и записва резултат след обработка в `output.bin`. Програмата трябва да работи като `encoder`, който имплементира вариант на Manchester code, т.е.:

- за всеки входен бит 1 извежда битовите 10, и
- за всеки входен бит 0 извежда битовите 01

Например, следните 8 бита вход

```
1011 0110 == 0xB6
```

по описаният алгоритъм дават следните 16 бита изход

```
1001 1010 0110 1001 == 0x9A69
```

Зад. 73 2021-SE-02 Вашите колеги от съседната лаборатория са написали програма на C, която може да обработва подаден входен двоичен файл и на негова база генерира изходен двоичен файл. Програмата работи като `encoder`, който имплементира вариант на Manchester code, т.е.:

- за всеки входен бит 1 извежда битовите 10, и
- за всеки входен бит 0 извежда битовите 01

Например следните 8 бита вход

```
1011 0110          == 0xB6
```

по описания алгоритъм дават следните 16 бита изход

```
1001 1010 0110 1001    == 0x9A69
```

Напишете програма на C, която извършва обратния процес, т.е., декодира файлове, създадени от горната програма.

Примерно извикване:

```
$ ./main input.bin output.bin
```

Зад. 74 2021-SE-03

Напишете програма на C, която приема два позиционни параметъра – имена на файлове. Примерно извикване:

```
$ ./main input.bin output.h
```

Файлът `input.bin` е двоичен файл с елементи `uint16_t` числа, създаден на *little-endian* машина.

Вашата програма трябва да генерира C хедър файл, дефиниращ масив с име `arr`, който:

- съдържа всички елементи от входния файл;
- няма указана големина;
- не позволява промяна на данните.

Генерираният хедър файл трябва да:

- съдържа и `uint32_t` променлива `arrN`, която указва големината на масива;
- бъде валиден и да може да се `#include`-ва без проблеми от C файлове, очакващи да “виждат” `arr` и `arrN`.

За да е валиден един входен файл, той трябва да съдържа не повече от 524288 елемента.

Зад. 75 2022-IN-01

Вашите колеги от съседната лаборатория работят със специализирана система Hoge, която съхранява данните си в `binary` файлове. За съжаление обаче им се налага да работят с две различни версии на системата (стара и нова) и разбира се, те ползват различни файлови формати.

Вашата задача е да напишете програма на C (`./main list.bin data.bin out.bin`), която конвертира файлове с данни от стария (`list.bin + data.bin`) в новия (`out.bin`) формат.

Всички файлове се състоят от две секции – 8 байтов хедър и данни. Структурата на хедъра е:

- `uint16_t, magic` – магическа стойност `0x5A4D`, която дефинира, че файлът е от системата Hoge и следва тази спецификация;
- `uint16_t, filetype` – дефинира какъв тип файл е това, с допустими стойности 1 за `list.bin`, 2 за `data.bin` и 3 за `out.bin`;
- `uint32_t, count` – дефинира броя на елементите в секцията с данни на файла.

Секцията за данни на всички файлове съдържа елементи – цели числа без знак, от съответните типове:

- `list.bin` – `uint16_t`;
- `data.bin` – `uint32_t`;
- `out.bin` – `uint64_t`.

Във файлове `data.bin` и `out.bin` елементи, чиято стойност е 0, са валидни елементи, но се игнорират от системата Hoge (тяхното наличие във файла не пречи на работата на системата).

Всеки елемент в секцията за данни на `list.bin` има две характеристики – *позиция* и *стойност*, като *позиция* е отместването на съответния елемент, докато *стойност* е неговата стойност. Тези две числа семантично дефинират отмествания във файловете с данни:

- *позиция* дефинира отместване в `data.bin`;
- *стойност* дефинира отместване в `out.bin`.

На базата на тези числови характеристики елементите на `list.bin` дефинират правила от вида: “елементът на отместване *позиция* в `data.bin` трябва да отиде на отместване *стойност* в `out.bin`”.

Забележка: Всички отмествания са спрямо N_0 в брой елементи.

Програмата трябва да интерпретира по ред дефинираните правила и на тяхна база да генерира изходния файл.

Обърнете внимание на проверките, обработката за грешки и съобщенията към потребителя – искаме програмата да бъде удобен инструмент.

Забележка: Не е гарантирано, че разполагате с достатъчно памет, в която да заредите всички елементи на които и да от файловете.

Зад. 76 2022-SE-01 Напишете програма на C, която приема два позиционни аргумента – имена на двоични файлове. Примерно извикване: `./main data.bin comparator.bin`

Файлът `data.bin` се състои от две секции – 8 байтов хедър и данни. Структурата на хедъра е:

- `uint32_t, magic` – магическа стойност `0x21796F4A`, която дефинира, че файлът следва тази спецификация;
- `uint32_t, count` – описва броя на елементите в секцията с данни.

Секцията за данни се състои от елементи – `uint64_t` числа.

Файлът `comparator.bin` се състои от две секции – 16 байтов хедър и данни. Структурата на хедъра е:

- `uint32_t, magic1` – магическа стойност `0xAFBC7A37`;
- `uint16_t, magic2` – магическа стойност `0x1C27`;
- комбинацията от горните две `magic` числа дефинира, че файлът следва тази спецификация;
- `uint16_t, reserved` – не се използва;
- `uint64_t, count` – описва броя на елементите в секцията с данни.

Секцията за данни се състои от елементи – наредени 6-торки:

- `uint16_t, type` – възможни стойности: 0 или 1;
- 3 бр. `uint16_t, reserved` – възможни стойности за всяко: 0;
- `uint32_t, offset1`;
- `uint32_t, offset2`.

Двете числа `offset` дефинират отместване (спрямо N_0) в брой елементи за `data.bin`; `type` дефинира операция за сравнение:

- 0: “по-малко”;
- 1: “по-голямо”.

Елементите в `comparator.bin` дефинират правила от вида:

- “елементът на `offset1`” трябва да е “по-малък” от “елементът на `offset2`”;
- “елементът на `offset1`” трябва да е “по-голям” от “елементът на `offset2`”.

Програмата трябва да интерпретира по ред дефинираните правила в `comparator.bin` и ако правилото не е изпълнено, да разменя in-place елементите на съответните отмествания. Елементи, които са равни, няма нужда да се разменят.

Зад. 77 2023-IN-01 Вашите колеги от съседната лаборатория са попаднали на вогонски* крипто-вирус и вашата задача е да напишете програма, която възстановява файлове, криптирани от вируса. Примерно извикване: `./main input.encrypted output.decrypted`

Вирусът не бил особено добре проектиран и след анализ се оказало, че подменя оригиналните файлове с криптирани със следните особености:

- вогоните явно предпочитат да работят с данни от по 128 бита, като всеки такъв елемент ще наричаме *unit* (U). Както входния (оригинален) файл, така и изходния (криптиран) файл се третират като състоящи се от unit-и. Криптираният файл винаги е точен брой unit-и (номерирани спрямо N_0), като при нужда към оригиналния файл преди криптиране се добавят байтове `0x00` (padding), за да стане точен брой unit-и.
- unit-ите в криптирания файл могат да се ползват за съхранение на следните типове данни (и

*Adams, Douglas. The Ultimate Hitchhiker’s Guide to the Galaxy: Five Novels in One Outrageous Volume. Del Rey, 2010.

техните големина): *header* (4U), *section* (2U), *cryptdata* (1U), *unused data* (1U).

- типове данни *section* и *cryptdata* са обработени с *побитово изключващо или* (XOR) със 128 битов ключ (ключът е голям 1U), като за *section* ключът се ползва два пъти – за първият и за вторият *unit* по отделно.
- криптираният файл винаги започва с **header** (4U) и има следната структура:
 - `uint64_t, magic` – магическа стойност `0x0000534f44614c47`, която дефинира, че файлът е криптиран спрямо тази спецификация
 - `uint32_t, cfsb` – размер на криптирания файл в байтове
 - `uint32_t, cfsu` – размер на криптирания файл в брой *unit*-и
 - `uint32_t, ofsb` – размер на оригиналния файл в байтове
 - `uint32_t, ofsu` – размер на оригиналния файл в брой *unit*-и
 - `uint32_t, unused1` – не се използва в тази версия на вируса
 - `uint32_t, cksum` – checksum-a на оригиналния файл, получена по следния алгоритъм:
 - * оригиналният файл се третира като състоящ се от `uint32_t` елементи
 - * при нужда от padding на последния елемент се ползват байтове `0x00`
 - * всички елементи се XOR-ват един с друг и полученият резултат е `cksum`
 - 128 бита[†], `sectionkey` – ключ, с който са кодирани *section*-ите
 - четири “слота”, описващи начална позиция в U на *section*
 - * `uint32_t, s0`
 - * `uint32_t, s1`
 - * `uint32_t, s2`
 - * `uint32_t, s3`
 - * тъй като *unit*-и с номера в интервала $[0,3]$ се ползват за самия *header*, в горните слотове 0 означава, че този слот не се използва (няма такъв *section*)
 - * *unit*-ите в изходния (декриптиран) файл са по реда на секциите
- *section* (2U) е кодиран със `sectionkey` от *header*-а и след декодиране има следната структура:
 - `int64_t, relative_offset` – дефинира от кое U започват *unit*-ите с данни за тази секция. Числото е относително спрямо първият *unit* на *section*, например, ако *section* е записан в U 4/5 и стойността на `offset` е 2, то първите данни на секцията ще са в U 6.
 - `uint64_t, len` – дефинира брой *unit*-и в този *section*
 - 128 бита, `datakey` – ключ, с който са кодирани *unit*-ите с данни в този *section*
 - *unit*-ите в изходния (декриптиран) файл са по реда на *unit*-ите в секцията
- *cryptdata* (1U) – *unit* в който има записани данни, кодирани с `datakey` на секцията, към която е този *unit*
- *unused data* (1U) – *unit*, който не се ползва. Колегите ви предполагат, че това е един от начините чрез които вогоните са вкарвали “шум” в криптирания файл, но без да правят декриптирането невъзможно.

2.2 Процеси, pipe-ове, вход и изход (теми 6,7,8)

Зад. 78 2016-SE-01 Напишете програма на C, която по подадено име на (текстови) файл като параметър, извежда съдържанието на файла сортирано, чрез употреба на външните програми `cat` и `sort` през `pipe()`.

Зад. 79 2016-SE-02 Напишете програма на C, която реализира `simple command prompt`. Тя изпълнява в цикъл следната поредица действия:

1. Извежда промпт на стандартния изход.
2. Прочита име на команда.
3. Изпълнява без параметри прочетената команда.

Командите се търсят в директорията `/bin`. За край на програмата се смята въвеждането на `exit`.

Зад. 80 2017-IN-01 Напишете програма на C, която използвайки външни shell команди през `pipe()` да извежда статистика за броя на използване на различните shell-ове от потребителите, дефинирани в системата. Изходът да бъде сортиран във възходящ ред според брой използвания на shell-овете.

Примерно извикване и изход:

```
$ ./main
```

[†] Ако искате да работите с 128 бита наведнъж удобно по познатия начин, ползвайте `#define uint128_t __uint128_t`

```
1 /bin/sync
3 /bin/bash
7 /bin/false
17 /usr/sbin/nologin
```

Зад. 81 2017-IN-02 Напишете програма на C, която приема незадължителен параметър – име на команда. Ако не е зададена команда като параметър, да се ползва командата `echo`. Максималната допустима дължина на командата е 4 знака.

Програмата чете низове (с максимална дължина 4 знака) от стандартния си вход, разделени с интервали (0x20) или знак за нов ред (0x0A). Ако някой низ е с дължина по-голяма от 4 знака, то програмата да терминира със съобщение за грешка.

Подадените на стандартния вход низове програмата трябва да третира като множество от параметри за дефинираната команда. Програмата ви трябва да изпълни командата колкото пъти е необходимо с *максимум два* низа като параметри, като изчаква изпълнението да приключи, преди да започне ново изпълнение.

Примерни вход, извиквания и изходи:

```
$ cat f1
a1
$ cat f2
a2
$ cat f3
a3
$ echo -e "f1\nf2 f3" | ./main cat
a1
a2
a3
$ echo -e "f1\nf2 f3" | ./main
f1 f2
f3
```

Зад. 82 2018-SE-01 Напишете програма на C, която приема параметър – име на директория. Програмата трябва да извежда името на най-скоро променения (по съдържание) файл в тази директория и нейните под-директории, чрез употреба на външни шел команди през `pipe()`.

Зад. 83 2019-SE-01 Напишете програма-наблюдател *P*, която изпълнява друга програма *Q* и я рестартира, когато *Q* завърши изпълнението си. На командния ред на *P* се подават следните параметри:

- *праг за продължителност* в секунди – едноцифрено число от 1 до 9
- *Q*
- *незадължителни параметри на Q*

P работи по следния алгоритъм:

- стартира *Q* с подадените параметри
- изчаква *Q* да завърши изпълнението си
- записва в текстов файл `run.log` един ред с три полета - цели числа (разделени с интервал):
 - момент на стартиране на *Q* (Unix time)
 - момент на завършване на *Q* (Unix time)
 - код за грешка, с който *Q* е завършила (exit code)
- проверява дали е изпълнено *условието за спиране* и ако не е, преминава отново към стартирането на *Q*

Условие за спиране: Ако наблюдателят *P* установи, че при две последователни изпълнения на *Q* са били изпълнени и двете условия:

1. кодът за грешка на *Q* е бил различен от 0;
2. разликата между момента на завършване и момента на стартиране на *Q* е била по-малка от подадения като първи параметър на *P* праг;

то *P* спира цикъла от изпълняване на *Q* и сам завършва изпълнението си.

Текущото време във формат Unix time (секунди от 1 януари 1970 г.) можете да вземете с извикване

на системната функция `time()` с параметър `NULL`; функцията е дефинирана в `time.h`. Ако изпълнената програма е била прекъсната от подаден сигнал, това се приема за завършване с код за грешка 129.

Зад. 84 2020-SE-01 Напишете две програми на C (`foo` и `bar`), които си комуникират през *наименована тръба*. Програмата `foo` приема параметър - име на файл, програмата `bar` приема параметър - команда като абсолютен път до изпълним файл.

Примерни извиквания и ред на изпълнение (в отделни терминали):

```
./foo a.txt
./bar /usr/bin/sort
```

Програмата `foo` трябва да изпълнява външна команда `cat` с аргумент името на подадения файл, така че съдържанието му да се прехвърли през тръбата към програмата `bar`, която от своя страна трябва да изпълни подадената и като аргумент команда (без параметри; `/usr/bin/sort` в примера), която да обработи получените през тръбата данни, четейки от стандартен вход. Еквивалент на горния пример би било следното изпълнение:

```
cat a.txt | /usr/bin/sort
```

Зад. 85 2020-SE-02 При изграждане на система за пренасяне на сериен асинхронен сигнал върху радиопреносна мрежа се оказало, че големи поредици от битове само нули или само единици смущават сигнала, поради нестабилно ниво. Инженерите решили проблема, като:

- в моментите, в които няма сигнал от серийният порт, вкарвали изкуствено байт `0x55` в потока;
- реалните байтове `0x00`, `0xFF`, `0x55` и `0x7D` се кодирали посредством XOR-ване (побитова обработка с *изключващо-или*) с `0x20`, като полученият байт се изпращал през потока, предхождан от `0x7D`, който играе ролята на *escape character*.

Разполагате със запис от такъв поток. Напишете програма на C, която приема два параметъра - имена на файлове. Примерно извикване:

```
$ ./main input.lfld output.bin
```

Програмата трябва да обработва записа и да генерира `output.bin`, който да съдържа оригиналните данни. Четенето на входните данни трябва да става посредством изпълнение на външна shell команда.

Зад. 86 2020-SE-03 Напишете програма на C, която приема един задължителен позиционен параметър - име на файл. Файлът се състои от не повече от 8 наредени тройки елементи:

- *име на файл* – точно 8 байта, последният от които задължително е `0x00`. Ако името е по-късо от 7 знака, излишните байтове са `0x00`;
- *offset* – `uint32_t`, който дава пореден номер на елемент (спрямо N_0) във файла;
- *length* – `uint32_t`, който дава брой елементи.

За всяка наредена тройка програмата трябва да пусне child процес, който да XOR-не (обработи с *изключващо-или*) елементите (`uint16_t`) от съответния файл един със друг, и да върне резултата на parent процеса, който от своя страна трябва да XOR-не всички получените резултати и да изведе полученото число в следния формат:

```
result: 573B
```

Забележка: C пълен брой точки се оценяват решения, в които child процесите работят паралелно.

Зад. 87 2021-SE-01 Ваши колеги - асистенти по ОС се оплакват, че студентите упорито ползват командата `sudo`, въпреки че им е обяснено да не я ползват. Вашата задача е да подготвите фалшиво `sudo`, което:

- добавя ред в текстови файл `foo.log` с командата, която потребителят се е опитал да изпълни във формат **ДАТА ПОТРЕБИТЕЛ ДАННИ**, където:
 - ДАТА е във формат `YYYY-MM-DD HH:MM:SS.UUUUUU` (`UUUUUU` са микросекунди)
 - ПОТРЕБИТЕЛ е login name (username) на потребителя
 - ДАННИ е всичко, което потребителят е написал, например `./main ls -l /root`
 - горните полета са разделени с интервал

– пример: 2021-05-10 14:38:17.584344 s99999 ./main ls -l /root

- заключва акаунта, който е изпълнил програмата
- приключва изпълнението на всички процеси, стартирани от този потребител

За да може да изпълни тези неща, фалшивото `sudo` ще има нужда от `root` права, а в същото време трябва да знае кой акаунт изпълнява командата. Традиционно това може да се реши със `SUID`, което обаче не важи за `shell` скриптове. За да заобиколим този проблем, напишете програма на `C` (`main`), която за заключването на акаунта и терминирането на процесите извиква външни `shell` команди през `pipe()`.

Полезни `man` страници:

- `gettimeofday(2)`
- `localtime(3)`
- `strftime(3)`
- `getuid(2)`
- `getpwuid(3)`
- `passwd(1)`

Опишете в `doc.txt` как бихте настроили компилираната вече програма така, че студентите да ползват нея, вместо оригиналното `sudo`.

Естествено, по време на разработка на програмата можете да я тествате само с вашият акаунт и нямате `root` права. За да може все пак да тествате, подменете в кода финалните команди, които програмата изпълнява (тези, които наистина изискват `root` права), като им добавите едно `echo`. Например, ако трябва да изпълните командата

```
foo -m pesho
```

вместо това в кода ви трябва да е дефинирано изпълнение на

```
echo foo -m pesho
```

Зад. 88 2022-IN-01 Ваши колеги - асистенти по ОС имат нужда от демонстрационна програма на `C`, която да служи като пример за конкурентност и синхронизация на процеси. Напишете такава програма, приемаща два задължителни позиционни параметъра – едноцифрени числа. Примерно извикване: `./main N D`

Общ алгоритъм на програмата:

- началният (родителски) процес създава процес-наследник
- N на брой пъти се изпълнява:
 - родителският процес извежда на `stdout` низа “DING ”
 - процесът-наследник извежда на `stdout` низа “DONG”
 - родителският процес изчаква D секунди

Гарантирайте, че:

- процесът-наследник винаги извежда “DONG” след като родителския процес е извел “DING ”
- родителският процес винаги започва изчакването след като процеса-наследник е извел “DONG”

Забележка: За изчакването погледнете `sleep(3)`.

Зад. 89 2023-IN-01

Ваши колеги – асистенти по ОС – имат нужда от демонстрационна програма на `C`, която да служи като пример за конкурентност и синхронизация на процеси.

Дефиниран е *нареден* списък с три думи $L = ("tic", "tac", "toe \backslash n")$, като всяка дума е с дължина четири знака. Напишете програма, приемаща две числа като аргументи (`./main NC WC`), като NC е ограничено в интервала $[1, 7]$, а WC – в интервала $[1, 35]$. Програмата трябва *задължително* да работи по следния общ алгоритъм:

- началният (родителски) процес създава NC на брой процеси-наследници;
- групата процеси извеждат на `stdout` общо WC на брой думи от горния списък при следните правила:
 - ако броят на думите за извеждане WC е по-голям от общия брой думи в L , след изчерпване на думите в L програмата започва списъка отначало колкото пъти е нужно;

- всеки процес извежда само една дума;
- първата дума се извежда от родителския процес;
- ако броят на думите за извеждане WC е по-голям от *общия* брой процеси, след изчерпване на процесите програмата започва да ги ползва от начало. Например при родителски процес P и два процеса-наследници C_1 и C_2 редът на използване на процесите ще бъде $P, C_1, C_2, P, C_1, C_2, P, C_1, \dots$
- изведените думи *гарантирано* трябва да са по реда, дефиниран в L ;
- всеки процес *гарантирано* започва да извежда на `stdout` поредната дума, след като предходния процес е приключил с извеждането на предишната.

Забележка: За конвертиране на аргументите погледнете `strtol(3)`.

3 Теоретични задачи

Упътвания за семафорите:

- Семафорът е обект за синхронизация, локалните му данни са брояч cnt и списък на приспаните процеси L . Конструкторът му $init(n)$ присвоява начална стойност на брояча ($cnt = n$), списъкът се инициализира празен. Семафорът има два метода – $wait()$ и $signal()$. Методът $wait()$ намаля с единица брояча cnt и ако стойността на брояча стане отрицателна, добавя в списъка L информация за текущия процес и го спира временно (процесът бива приспан, блокиран). Методът $signal()$ увеличава cnt и ако стойността на брояча преди увеличението е отрицателна, изважда процес от списъка L и го събужда. Ако от L се вади най-рано приспания процес, наричаме семафора силен. Всяка друга стратегия на събуждане реализира слаб семафор.
- Приемете, че инициализацията на семафорите се прави от процес, който поражда процесите, обсъждани в условието на задачата, преди тяхното стартиране.

3.1 Единични процеси

Зад. 90 2016-SE-01 Всеки от процесите P и Q изпълнява поредица от три инструкции:

process P	process Q
p_1	q_1
p_2	q_2
p_3	q_3

Осигурете чрез семафори синхронизация на P и Q така, че инструкцията p_1 да се изпълни преди q_2, а q_2 да се изпълни преди p_3.

Зад. 91 2017-SE-01 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че инструкцията p_1 да се изпълни преди q_2 и r_2.

Забележка: Решения на задачата с повече от един семафор не носят пълен брой точки.

Зад. 92 2017-SE-04 Всеки от процесите P и Q изпълнява поредица от две инструкции:

process P	process Q
p_1	q_1
p_2	q_2

Осигурете чрез семафори синхронизация на P и Q така, че инструкцията p_1 да се изпълни преди q_2, а q_1 да се изпълни преди p_2.

Зад. 93 2017-SE-05 Всеки от процесите P и Q изпълнява поредица от три инструкции:

process P	process Q
p_1	q_1
p_2	q_2
p_3	q_3

Осигурете чрез два семафора синхронизация на P и Q така, че отделните инструкции да се изпълнят в следния времеви ред: p_1, q_1, p_2, q_2, p_3, q_3.

Зад. 94 2018-CS-01 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- инструкция p_1 да се изпълни преди q_2 и r_2
- инструкция r_2 да се изпълни преди p_3

Забележка: Решения на задачата с повече от два семафора не носят пълен брой точки.

Зад. 95 2018-CS-02 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- Инструкция p_1 да се изпълни преди q_2.
- Инструкция q_1 да се изпълни преди r_2.
- Инструкция r_1 да се изпълни преди p_2.
- Инструкция r_3 да се изпълни след p_2 и q_2.

Зад. 96 2018-SE-01 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- някоя от инструкциите p_2 и q_2 да се изпълни преди r_2;
- ако инструкция p_2 се изпълни преди r_2, то q_2 да се изпълни след r_2;
- ако инструкция q_2 се изпълни преди r_2, то p_2 да се изпълни след r_2;

Забележка: Решения на задачата с повече от два семафора не носят пълен брой точки.

Забележка: Решения на задачата при които p_2 или q_2 винаги изчакват r_2 носят 0 точки.

Зад. 97 2019-CS-01 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- инструкция p_1 да се изпълни преди q_2 и r_2
- инструкция p_3 да се изпълни след q_2 и r_2

Зад. 98 2019-CS-02 Всеки от процесите P, Q и R изпълнява поредица от две инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2

Осигурете чрез три семафора синхронизация на P, Q и R така, че отделните инструкции да се изпълнят в следния времеви ред: p_1, q_1, r_1, p_2, q_2, r_2.

Зад. 99 2019-SE-01 Всеки от процесите P и Q изпълнява поредица от три инструкции:

process P	process Q
p_1	q_1
p_2	q_2
p_3	q_3

Осигурете чрез два семафора синхронизация на P и Q така, че да са изпълнени едновременно следните времеви зависимости:

1. инструкция p_1 да се изпълни преди q_2
2. инструкция q_2 да се изпълни преди p_3
3. инструкция q_1 да се изпълни преди p_2
4. инструкция p_2 да се изпълни преди q_3

Забележка: Решения на задачата с повече от два семафора не носят пълен брой точки.

Зад. 100 2020-CS-01 Процесите P, Q и R изпълняват поредица от две инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- инструкцията p_1 да се изпълни преди q_2 и r_2
- инструкцията q_1 да се изпълни преди p_2 и r_2
- инструкцията r_1 да се изпълни преди p_2 и q_2

Зад. 101 2020-SE-03 Всеки от процесите P, Q и R изпълнява поредица от три инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2
p_3	q_3	r_3

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- инструкция p_1 да се изпълни преди q_2 и r_2;
- ако q_2 се изпълни преди r_2, то и q_3 да се изпълни преди r_2;
- ако r_2 се изпълни преди q_2, то и r_3 да се изпълни преди q_2.

Обосновете отговора си.

3.2 Процеси с много копия

Зад. 102 2017-SE-02 Преди стартиране на процеси P и Q са инициализирани два семафора и брояч:

```
semaphore e, m
e.init(1); m.init(1)
int cnt = 0
```

Паралелно работещи няколко копия на всеки от процесите P и Q изпълняват поредица от инструкции:

process P	process Q
m.wait()	e.wait()

```

        cnt=cnt+1                q_section
        if cnt=1 e.wait()        e.signal()
m.signal()

p_section

m.wait()
cnt=cnt-1
if cnt=0 e.signal()
m.signal()

```

Дайте обоснован отговор на следните въпроси:

- Могат ли едновременно да се изпълняват инструкциите `p_section` и `q_section`?
- Могат ли едновременно да се изпълняват няколко инструкции `p_section`?
- Могат ли едновременно да се изпълняват няколко инструкции `q_section`?
- Има ли условия за deadlock или starvation за някой от процесите?

Упътване:

- Ще казваме, че `P` е в критична секция, когато изпълнява инструкцията си `p_section`. Същото за `Q`, когато изпълнява `q_section`.
- Изяснете смисъла на брояча `cnt` и какви процеси могат да бъдат приспани в опашките на двата семафора.
- Покажете, че в опашката на семафора `e` има най-много едно копие на `P` и произволен брой копия на `Q`.
- Покажете, че в момента на изпълнение на `e.signal()` в кой да е от процесите, никой процес не е в критичната си секция.

Зад. 103 2017-SE-03

- Няколко копия на процеса `P` изпълняват поредица от три инструкции:

```

process P
  p_1
  p_2
  p_3

```

Осигурете чрез семафор синхронизация на копията така, че най-много един процес да изпълнява инструкция `p_2` във всеки един момент.

- Опишете разликата при реализация на слаб и силен семафор.
- Възможно ли е в зависимост от начина на реализация на семафора в подусловие а) да настъпят условия за deadlock или starvation? Ако да, опишете сценарий за поява на неприятната ситуация.

Зад. 104 2017-SE-08

Множество паралелно работещи копия на всеки от процесите `P` и `Q` изпълняват поредица от две инструкции:

```

process P                process Q
  p_1                    q_1
  p_2                    q_2

```

Осигурете чрез семафори синхронизация на работещите копия така, че да са изпълнени едновременно следните условия:

- в произволен момент от времето да работи най-много едно от копията;
- работещите копия да се редуват във времето – след изпълнение на копие на `P` да следва изпълнение на копие на `Q` и обратно;
- първоначално е разрешено да се изпълни копие на `P`.

Зад. 105 2018-SE-03

Множество паралелно работещи копия на всеки от процесите `P` и `Q` изпълняват поредица от три инструкции:

```

process P                process Q

```

p_1	q_1
p_2	q_2
p_3	q_3

Осигурете чрез семафори синхронизация на работещите копия, така че три инструкции – p_1, q_2 и p_3 се редуват циклично:

- първа се изпълнява инструкция p_1 на някое от работещите копия на процес P;
- след завършването ѝ се изпълнява инструкция q_2 на някое копие на Q;
- след нея – p_3 на някое копие на P;
- с това едно минаване през цикъла завършва и отново може да се изпълни инструкция p_1 на някое от работещите копия на процес P.

Зад. 106 2020-CS-01 Множество паралелно работещи копия на всеки от процесите P, Q и R изпълняват поредица от две инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- в произволен момент от времето да работи най-много едно от копията.
- работещите копия да се редуват във времето – след изпълнение на копие на P да се изпълни копие на Q, после копие на R. Следва ново изпълнение на P и т.н.
- първоначално е разрешено да се изпълни копие на P.

Зад. 107 2020-SE-02 Множество паралелно работещи копия на всеки от процесите P, Q, R и W изпълняват поредица от две инструкции:

process P	process Q	process R	process W
p_1	q_1	r_1	w_1
p_2	q_2	r_2	w_2

Осигурете чрез семафори синхронизация на работещите копия така, че да са изпълнени едновременно следните условия:

- В произволен момент от времето да работи най-много едно от копията (един-единствен процес глобално).
- Работещите копия да се редуват във времето – първо се изпълнява копие на P или Q. След това трябва да се изпълни копие на R или W. Следва ново изпълнение на копие на P или Q и т.н.

Обосновете отговора си.

Зад. 108 2021-SE-01 Всеки от процесите P, Q и R изпълнява поредица от инструкции:

process P	process Q	process R
p_1	q_1	r_1
p_2	q_2	r_2

Процесите P и Q са единични, процесът R се изпълнява в много копия.

Осигурете чрез семафори синхронизация на P, Q и R така, че да се изпълнят едновременно следните изисквания:

- всички инструкции на P и Q да се изпълнят преди инструкция r_1 на всяко копие на R;
- процесите P и Q да се изпълнят ефикасно, т.е. да е възможно паралелното им изпълнение, без да се изчакват.

Обосновете отговора си.

3.3 Ползване на брояч

Зад. 109 2019-SE-01 Множество паралелно работещи копия на процеса P изпълняват поредица от две инструкции:

```
process P
  p_1
  p_2
```

Осигурете чрез семафори синхронизация на работещите копия така, че да се изпълнят едновременно следните изисквания:

- инструкцията p_2 на всяко от работещите копия да се изпълни след като инструкция p_1 е завършила изпълнението си в поне 3 работещи копия.

Упътване: Освен семафори, ползвайте и брояч.

Зад. 110 2020-SE-01 Множество паралелно работещи копия на всеки от процесите P и Q изпълняват поредица от три инструкции:

```
process P                process Q
  p_1                    q_1
  p_2                    q_2
  p_3                    q_3
```

Осигурете чрез семафори синхронизация на P и Q така, че поне една инструкция p_1 да се изпълни преди всички q_2, и поне една инструкция q_1 да се изпълни преди всички p_2.

Обосновете отговора си.

Зад. 111 2022-CS-01 Паралелно работещи копия на всеки от процесите P и Q изпълняват поредица от две инструкции:

```
process P                process Q
  p_1                    q_1
  p_2                    q_2
```

Осигурете чрез семафори синхронизация на работещите копия така, че да се изпълнят едновременно следните изисквания:

- в произволен момент от времето да работи най-много едно от копията;
- работещите копия да се редуват във времето – след изпълнение на копие на P, два пъти се изпълнява копие на Q, и цикълът се повтаря – веднъж P, два пъти Q и т.н.;
- първоначално е разрешено да се изпълни копие на P.

Упътване: Освен семафори, ползвайте и брояч.

Зад. 112 2022-IN-01

Ваксинационен център работи на следния принцип:

- множество доктори поставят ваксини;
- има маса, на която може да има максимум M на брой ампули с ваксина;
- всеки доктор взима една ампула от масата (ако на масата има ампули) и поставя ваксината на пациент;
- ако на масата няма ампули, докторът казва на медицинската сестра да донесе нова партида ампули и тя “презарежда” масата с нови M на брой ампули;
- работният ден започва с “пълна” маса с M на брой ампули.

Множество паралелно работещи копия на процеса D (лекари) изпълняват следния код:

```
while true:
  getVaccineFromTable()
  administerVaccineToPatient()
```

Единичен процес N (мед. сестра) изпълнява следния код:

```
while true:
    reloadTableWithVaccines(M)
```

Ограничения:

- процес D не може да изпълни `getVaccineFromTable()`, ако на масата няма ампули с ваксина;
- процес N може да изпълни `reloadTableWithVaccines(M)` само ако на масата няма ампули с ваксина.

Осигурете чрез семафори синхронизация на процесите така, че да са изпълнени синхронизационните ограничения. Обосновете решението си.

Упътване: Освен семафори, ползвайте и брояч.

Зад. 113 2022-IN-02 Паралелно работещи копия на всеки от процесите P и Q изпълняват поредица от две инструкции:

process P	process Q
p_1	q_1
p_2	q_2

Осигурете чрез семафори синхронизация на работещите копия така, че да се изпълнят едновременно следните изисквания:

- в произволен момент от времето да работи най-много едно от копията;
- работещите копия да се редуват във времето – след изпълнение на две копия на P, три пъти се изпълнява копие на Q, и цикълът се повтаря – два пъти P, три пъти Q и т.н.;
- първоначално е разрешено да се изпълни копие на P.

Упътване: Освен семафори, ползвайте и брояч(и).

Зад. 114 2023-IN-01 Множество от N паралелно работещи копия на процеса P изпълняват следният цикъл с две инструкции:

```
WHILE true
    p_1
    p_2
ENDWHILE
```

Осигурете чрез семафори и броячи синхронизация на работещите копия така, че на всяка итерация на цикъла всички процеси се изчакват взаимно, а именно: инструкцията `p_2` се изпълнява, след като всички копия са изпълнили `p_1` (на същата итерация).

3.4 Анализ на race condition

Зад. 115 2019-SE-03 Процесите P и Q се изпълняват паралелно. Споделената променлива A има начална стойност 4. Променливата R е локална за двата процеса.

process P	process Q
R=A	R=A
R=R+3	R=R+2
A=R	A=R

Каква е стойността на A след изпълнението на процесите? Дайте обоснован отговор.

3.5 Програми на C

Зад. 116 2018-CS-01 Дадена е програма за ОС Linux, написана на езика C:

```
#include <unistd.h>
#include <string.h>
int main(void)
{
    char* buff = "Hello world!\n";
    int p;
    if (fork()==0) write(1,buff,strlen(buff));
    p=fork();
    write(1,buff,strlen(buff));
}
```

- Колко пъти ще се отпечата текста **Hello world!** при изпълнението на програмата? Обосновете отговора си.
- Нарисувайте кореновото дърво с върхове процесите, които ще се стартират в резултат от изпълнението на програмата и ребра двойките родител-наследник.

Зад. 117 2018-SE-02 Дадена е програма за ОС Linux, написана на езика C:

```
#include <unistd.h>
#include <stdio.h>
int main(void)
{
    int p1, p2;
    p1=fork();
    p2=fork();
    printf("Hello world!\n");
}
```

- а) Колко пъти ще се отпечата текста **"Hello world!"** при изпълнението на програмата? Обосновете отговора си.
- б) Как работи системното извикване **fork()**?
- в) Нарисувайте кореновото дърво с върхове процесите, които ще се стартират в резултат от изпълнението на програмата и ребра двойките родител-наследник.

Зад. 118 2019-CS-02 Дадена е програма за ОС Linux, написана на езика C:

```
#include <unistd.h>
#include <stdio.h>
int main(void)
{
    int p1, p2, p3;
    p1=fork();
    if (p1==0) {
        p2=fork();
        if (p2>0) p3=fork();
    }
    printf("Hello world!\n");
}
```

- а) Колко пъти ще се отпечата текста **Hello world!** при изпълнението на програмата? Обосновете отговора си.
- б) Как работи системното извикване **fork()**?
- в) Нарисувайте кореновото дърво с върхове процесите, които ще се стартират в резултат от изпълнението на програмата и ребра двойките родител-наследник.

Зад. 119 2022-IN-01

Дадена е следната програма на C за GNU/Linux операционна система:

```
1 #include <sys/types.h>
2 #include <unistd.h>
```

```

3 #include <stdlib.h>
4 int
5 main(void)
6 {
7     pid_t p1,p2,p3;
8     write(1,"A",1);
9     p1 = fork();
10    if (p1 == 0)
11    {
12        write(2,"BB",2);
13        p2 = fork();
14        if (p2 > 0)
15        {
16            write(1,"C",1);
17            p3 = fork();
18            write(2,"DD",2);
19            if (p3 > 0) exit(0);
20        }
21    }
22    write(1,"A",1);
23    exit(0);
24 }

```

Забележка: Номерата на редовете са за илюстративни цели и не са част от кода на програмата.

При компилиране и стартиране на програмата се създава процес с PID 42. Изберете примерни стойности за PID на всеки създаден процес-наследник.

- а) Опишете обосновано изпълнението на програмата, като включите следната информация:
 - процес <PID>, ред <n>, създава процес <PID>
 - процес <PID>, ред <n>, извежда...
 - процес <PID>, ред <n>, терминира (прекратява изпълнението си)
- б) Нарисувайте диаграма, която визуализира взаимоотношенията между процесите. Например, може да нарисувате кореново дърво с върхове процесите, които ще се стартират в резултат от изпълнението на програмата и ребра двойките родител-наследник. Маркирайте върховете със съответния PID.
- в) Какъв ще бъде изхода на `stdout` след приключване на процесите? Обосновете отговора си.

Упътване: Не е гарантирано, че системното извикване `fork()` приключва успешно.

3.6 “Разказвателни” задачи

Зад. 120 2016-SE-02 Опишете накратко основните процедури и структури данни, необходими за реализация на семафор.

Каква е разликата между слаб и силен семафор?

Опишете максимално несправедлива ситуация, която може да се получи в избирателна секция, ако на входа на секцията пазач – член на изборната комисия пуска гласоподавателите вътре така:

- във всеки момент в секцията може да има най-много двама гласоподаватели.
- пазачът работи като слаб семафор.

Зад. 121 2017-SE-06 Да приемем, че в съвременната операционна система процесът има 4 състояния:

- *R* – работещ (*running*, използва CPU)
- *A* – активен (*ready*, очаква CPU)
- *S* – блокиран (*sleeping*, очаква вход/изход)
- *T* – изчакващ време (*sleeping*, очаква времеви момент)

Нарисувайте диаграма на състоянията и преходите между тях. Диаграмата е ориентиран граф с върхове отделните състояния и ребра – възможните преходи.

Опишете накратко събитията, предизвикващи преход по всяко ребро на графа.

Зад. 122 2017-SE-07 Процесът P създава тръба (pipe) с извикване на функцията `pipe(int pipefd[2])` в ОС GNU/Linux.

- а) Кои процеси не могат да ползват тръбата?
- б) Опишете друг метод за изграждане на комуникационен канал, който дава възможност на произволни процеси да изградят и ползват канала. Допълнително искаме новоизградения канал да е достъпен само за процесите, които са го създали.

Упътване: Прочетете ман-страницата за функцията `pipe()`.

Зад. 123 2018-CS-03 При споделено ползване на памет от няколко процеса е възможно да настъпи надпревара за ресурси (race condition).

- а) (2k точки) Дефинирайте понятието race condition.
- б) (k точки) Възможно ли е да настъпи race condition в еднопроцесорна система? Ако да, при какви условия.
- в) (3k точки) Какви инструменти ползваме, за да избегнем race condition?

Зад. 124 2018-CS-04 Една от класическите задачи за синхронизация се нарича *Задача за читателите и писателите (Readers-writers problem)*.

- а) (k точки) Опишете условието на задачата.
- б) (2k точки) Опишете решение, използващо семафори.

Зад. 125 2018-CS-05 Какви са възможните състояния на процес.

Нарисувайте диаграма на състоянията и преходите между тях.

Опишете накратко ситуациите, предизвикващи преходи между състояния.

Зад. 126 2018-SE-04 Опишете накратко кои системни извиквания изграждат стандартните комуникационни канали в UNIX – неименувана тръба (pipe), връзка процес-файл, двустранна връзка процес-процес (connection).

Зад. 127 2018-SE-05 Опишете какви изисквания удовлетворява съвременна файлова система, реализирана върху блоково устройство (*block device*). Опишете накратко реализацията и целта на следните инструменти:

- а) отлагане на записа, алгоритъм на асансьора;
- б) поддържане на журнал на файловата система.

Зад. 128 2019-CS-01 При реализация на файлова система върху твърд диск файловете и директориите се записват върху сектори от диска. Времето за достъп до секторите зависи от текущото положение на механичните компоненти на диска – над коя пътечка е главата за четене/запис и каква е позицията ѝ над пътечката.

Защо се прави разместване във времето на операциите по четене и запис върху диска?

Опишете накратко реализацията и целта на алгоритъма на асансьора.

Зад. 129 2019-CS-03 Опишете как се изгражда комуникационен канал (connection) между процес-сървер и процес-клиент със следните системни извиквания в стандарта POSIX:

`socket()`, `bind()`, `connect()`, `listen()`, `accept()`

Зад. 130 2019-CS-04 Опишете накратко основните комуникационни канали в ОС Linux. Кои канали използват пространството на имената и кои не го правят?

Зад. 131 2019-SE-04 Опишете разликата между синхронни и асинхронни входно-изходни операции. Дайте примери за програми, при които се налага използването на асинхронен вход-изход.

Зад. 132 2019-SE-06 Опишете реализацията на комуникационна тръба (pipe) чрез семафори. Предполагаме, че тръбата може да съхранява до n байта, подредени в обикновена опашка. Тръбата се ползва от няколко паралелно работещи изпращачи/получатели на байтове. Процесите изпращачи слагат байтове в края на опашката, получателите четат байтове от началото на опашката.

Упътване: В теорията на конкурентното програмиране задачата е известна като “producer-consumer problem”.

Зад. 133 2020-CS-01 Определете понятията *взаимно блокиране* (deadlock) и *гладуване* (livelock, resource starvation). Дайте примери.

Зад. 134 2020-IN-01 Кои системни извиквания работят с файлов дескриптор, създаден с `open(2)` на

обикновен файл, но винаги ще завършат с грешка за дескриптор, създаден с `pipe(2)`?

Опишете накратко функционалността ѝм.

Зад. 135 2021-CS-01 Каква задача решава инструментът `spinlock` (активно изчакване)?

Опишете хардуерните инструменти, необходими за реализацията на `spinlock`.

В кои ситуации не бива да се ползва `spinlock`?