

1. Úloha (2016)

Popis úlohy

Cílem úlohy je vytvořit skript, který prozkoumá adresář a vytvoří report o jeho obsahu. Předmětem rekurzivního zkoumání adresáře je počet souborů a adresářů, hloubka adresářové struktury, typy souborů (podle přípony) a velikosti souborů. Výstupem skriptu je textový report. Skript je konfigurovatelný pomocí příkazové řádky.

Specifikace chování skriptu

JMÉNO

- `dirstat` - statistika obsahu adresáře

POUŽITÍ

```
dirstat [-i FILE_ERE] [DIR]
```

POPIS

1. Pokud byl skriptu zadán adresář (DIR), bude předmětem zkoumání. Bez zadaného adresáře se prozkoumává aktuální adresář.
2. Přepínač `-i` způsobuje ignorování souborů a adresářů, jejichž název odpovídá rozšířenému regulárnímu výrazu `FILE_ERE`. `FILE_ERE` nesmí pokrývat název kořenového adresáře.

Součástí reportu je:

1. informace o adresářích:
 - a. počet všech adresářů,
 - b. největší hloubka adresářové struktury,
 - c. průměrný počet souborů v adresáři,
2. informace o všech souborech a o jednotlivých typech souborů (podle přípony souborů):
 - a. počet všech souborů,
 - b. velikost největšího souboru,
 - c. průměrná velikost souboru,
 - d. medián velikostí souborů.

FORMÁT REPORTU

```
Root directory: DIR
Directories: ND
Max depth: DD
Average no. of files: AF
All files: NF
  Largest file: LF
  Average file size: AS
  File size median: MS
File extensions: EL
Files .EXT: NEXT
  Largest file .EXT: LEXT
  Average file size .EXT: AEXT
  File size median .EXT: MEXT
```

1. Formát reportu musí přesně (vč. mezer) odpovídat předcházejícímu textu s odpovídajícími čísly.
2. Formát reportu nezahrnuje statistiky ignorovaných souborů či adresářů.
3. DIR je zkoumaný adresář předaný na příkazové řádce při spuštění skriptu. V případě, že skriptu nebyl zadán adresář, je DIR aktuální pracovní adresář (cwd).
4. ND (>0) je počet adresářů v adresářovém stromu.
5. DD (>0) je největší adresářová hloubka (prázdný kořenový adresář má hloubku 1).

6. AF (celé číslo) je aritmetický průměr počtů obyčejných souborů v každém adresáři.
7. NF (≥ 0) je počet všech obyčejných souborů.
8. LF (> 0) je velikost největšího souboru. Pokud neexistuje žádný soubor s nenulovou velikostí, místo čísla LF a cesty se vypíše 'N/A'.
9. AS (celé číslo) je průměrná velikost souborů. Pokud je počet souborů NF 0, vypíše se místo čísla AS řetězec N/A.
10. MS (celé číslo) je medián velikostí souborů. Pokud je počet souborů NF 0, vypíše se místo čísla MS řetězec N/A.
11. EL je čárkou oddělený, abecedně seřazený seznam přípon všech obyčejných souborů. Necht' přípona souboru je suffix názvu souboru za znakem '.' (tečka) obsahující pouze alfanumerické znaky.
12. Pro každou příponu EXT v seznamu EL (podle pořadí v seznamu) bude v reportu vypsána samostatná sekce, ve které:
 - a. EXT je přípona souborů.
 - b. NEXT (> 0) je počet souborů s danou příponou.
 - c. LEXT (≥ 0) je velikost největšího souboru s danou příponou.
 - d. AEXT (celé číslo) je průměrná velikost souborů s danou příponou.
 - e. MEXT (celé číslo) je medián velikostí souborů s danou příponou.
 - f. Pokud ve zkoumaném adresáři neexistuje žádný soubor s příponou, report žádnou sekci obsahovat nebude.

NÁVRATOVÁ HODNOTA

1. Skript vrací úspěch v případě úspěšného zjištění všech informací. Vrací chybu v případě chyby při zjišťování informací o souborech či adresářích. V takovém případě skript skončí také s chybovým hlášením.

Implementační detaily

- Skript by měl běžet na všech běžných shellech (dash, ksh, bash). Ve školním prostředí můžete použít základní (POSIX) /bin/sh.
- Referenční stroj neexistuje. Skript musí běžet na běžně dostupných OS GNU/Linux a *BSD. Ve školním prostředí máte k dispozici počítače v laboratořích (CentOS), stroj merlin (CentOS) a eva (FreeBSD). Pozor, na stroji merlin je shell /bin/ksh symbolický odkaz na bash (tj. nechová se jako Korn shell jako na obvyklých strojích).
- Skript nesmí používat dočasné soubory.

Příklady použití

Příklady předpokládají skript `dirstat` v cestě spustitelných programů (PATH).

1. Statistiky jednoho adresáře:

```
$ dirstat path/to/izp/proj3
Root directory: path/to/izp/proj3
Directories: 1
Max depth: 1
Average no. of files: 3
All files: 3
  Largest file: 14288
  Average file size: 11412
  File size median: 10352
File extensions: c,o
Files .c: 1
  Largest file .c: 9595
  Average file size .c: 9595
  File size median .c: 9595
Files .o: 1
  Largest file .o: 10352
  Average file size .o: 10352
  File size median .o: 10352
```

2. Statistiky ve větším adresáři:

```
$ dirstat /etc
Root directory: /etc
Directories: 405
Max depth: 4
Average no. of files: 5
```

All files: 2258
Largest file: 6961203
Average file size: 12743
File size median: 634
File extensions: cnf,conf,desktop,repo,sh
Files .cnf: 3
Largest file .cnf: 10923
Average file size .cnf: 3833
File size median .cnf: 345
Files .conf: 400
Largest file .conf: 82209
Average file size .conf: 1888
File size median .conf: 619
Files .desktop: 26
Largest file .desktop: 11420
Average file size .desktop: 3683
File size median .desktop: 4137
Files .repo: 14
Largest file .repo: 1328
Average file size .repo: 1037
File size median .repo: 1217
Files .sh: 34
Largest file .sh: 2703
Average file size .sh: 642
File size median .sh: 388