

Отчет по курсу “Операционные системы”

Задание 1.10

Владимир Руцкий, 4057/2

21 декабря 2009 г.

Постановка задачи

“Напишите скрипт, отыскивающий дублирующиеся файлы в домашнем каталоге пользователя или в каталогах, указанных в качестве параметров командной строки. Дублирование файлов можно, в частности, проверить по совпадению имени и контрольной суммы MD5 (используйте команду `md5sum`).”

Выбранный метод решения

1. Разбор аргументов командной строки будет произведён с помощью конструкции **while-if**. Каждая введённая в параметрах директория будет обработана отдельно.
2. Для каждой обрабатываемой директории будет построен список файлов с помощью команды **find**.
3. Для каждого файла будет вычислена его MD5-сумма, базовое имя и полное имя. Эти данные будут сохранены именно в таком порядке в одной строке, разделённые нулевым-символом.
4. Строки описания для всех файлов из обрабатываемой директории будут отсортированы в алфавитном порядке, в результате чего строки для файлов с одинаковыми MD5-суммами и именами станут находиться рядом в рабочем потоке.
5. Из рабочего потока будут построчно считаны все описания файлов, и в случае нахождения рядом описаний, с одинаковыми MD5-суммами и именами файлов, имена соответствующих файлов будут выведены как результат.

Исходный код

Исходный код 1: task_1_10.sh

```
1 #!/bin/bash
2 # task_1_10.sh
3 # Task 1.10 on Unix course.
4 # Vladimir Rutsky <altsysrq@gmail.com>
5 # 20.12.2009
6
7 #HOME_DIR=~
8 HOME_DIR=~ /public /semester07
9
10 function print_usage
11 {
12     echo "Usage:"
13     echo "  _$_$_h, _$_help _$_print_help_and_exit "
14     echo "  _$_$_directory1 _$_[ directory1 _$_... ] "
15 }
16
17 function process_dir
18 {
19     if [ ! -d "$1" ]; then
20         echo "Directory _'$1'_ doesn't exists."
21         exit 1
22     fi
23
24     PREV_HASH=
25     PREV_BASENAME=
26     PREV_FILE=
27     find "$1" -type f -exec /bin/sh -c "md5sum _'{ }'_ | _$_awk _'{ print _$_$1 }'_ | _$_tr _$_
28         -d _$_'\n' " \; -printf '\0%f\0%P\n' | sort | tr '\0' '\n' | \
29     while true; do
30         if [ "$PREV_HASH" != "" ] && [ "$HASH" = "$PREV_HASH" ] && [ "
31             $BASENAME" = "$PREV_BASENAME" ]; then
32             echo "Similar_files:"
33             echo $PREV_FILE
34             while [ "$HASH" = "$PREV_HASH" ] && [ "$BASENAME" = "$PREV_BASENAME
35                 " ]; do
36                 echo $FILE
37                 read HASH && read BASENAME && read FILE || break 2
38             done
39             fi
40
41             PREV_HASH=$HASH
42             PREV_BASENAME=$BASENAME
43             PREV_FILE=$FILE
44             read HASH && read BASENAME && read FILE || break
45         done
46     }
47
48     if [ $# == 0 ]; then
49         process_dir "$HOME_DIR"
50     elif [ "$1" == "-h" ] || [ "$1" == "--help" ]; then
51         print_usage
52     else
53         exit 0
54     fi
55 }
```

```
50 else
51     while [ "$1" != "" ]; do
52         process_dir "$1"
53         shift
54     done
55 fi
```