Lab #2: IO Lab

Prof. Jae W. Lee (jaewlee@snu.ac.kr)
Department of Computer Science and Engineering
Seoul National University

TA (snu-arc-sysprog-ta@googlegroups.com)

Contents

Important Dates

- 19 Mar. - Lab Hand-out Session (Today!)
- 26 Mar. - Live Q&A Session (1)
- 2 Apr. - Live Q&A Session (2)
- 3 Apr. 23:59 - Submission Deadline

- Questions about the lab will be conducted through github issue. Feel free to post questions (except your code!)
- Live Q&A Session is totally optional

Before the Presentation::

- **All content on this slide is sourced from README file.**

- **For detailed information, please refer to REAME.**
  - https://github.com/SNU-ARC/2024_spring_sysprog_Lab2/blob/main/README.md

Goal of this Lab(1/2)

- **We implement a tool that lists all files in a directory and all its subdirectories.**

```
$ dirtree -v -s demo
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo
  subdir1                                               sysprog:sysprog       4096 rwxrwxr-x  d
    sparsefile                                          sysprog:sysprog       8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic...  sysprog:sysprog       1000 rw-rw-r--
  subdir2                                               sysprog:sysprog       4096 rwxrwxr-x  d
    brokenlink                                          sysprog:sysprog          8 rwxrwxrwx  l
    symboliclink                                        sysprog:sysprog          6 rwxrwxrwx  l
  subdir3                                               sysprog:sysprog       4096 rwxrwxr-x  d
    pipe                                                sysprog:sysprog          0 rw-rw-r--  f
    socket                                              sysprog:sysprog          0 rwxrwxr-x  s
  one                                                   sysprog:sysprog          1 rw-rw-r--
  two                                                   sysprog:sysprog          2 rw-rw-r--
-------------------------------------------------------------------------------------------------
4 files, 3 directories, 2 links, 1 pipe, and 1 socket                         21497
```

Goal of this Lab(2/2)

- **You will learn**
  - how to iterate through all files in a directory
  - how to retrieve the metadata of a file
  - how to print nicely formatted output
  - that error handling requires a significant effort
  - that string handling is not one of C's strengths
  - and a bunch of other useful programming tricks and C library functions

Environment setup(1/3)

- **You can get skeleton code and test bench from git repo**
  - `git clone https://github.com/SNU-ARC/2024_spring_sysprog_Lab2.git`

Environment setup(2/3)<Optional>

- **If you want to keep your own repository, you should keep the lab's visibility to private. Otherwise, others would see your work.**

- **Changing visibility**
  - After cloning the repository, you should change the push remote URL to your own repository.

1. **Create an empty repository that you're going to manage (again, keep it private)**
2. **Copy the url of that repository**
3. **On your terminal in the cloned directory, type**
   `git remote set-url --push origin <repo_url>`
4. **Check with git remote -v if the push URL has changed to yours while the fetch URL remains the same (this repo)**

Environment setup(3/3)

- **The handout contains the following files and directories.**

File/Directory	Description
README.md	this file
Makefile	Makefile driver program
src/dirtree.c	Skeleton for dirtree.c. Implement your solution by editing this file.
reference/	Reference implementation
tools/	Tools to generate directory trees for testing

→

File/Directory	Description
gentree.sh	Driver script to generate a test directory tree.
mksock	Helper script to generate a Unix socket.
*.tree	Script files describing the directory tree layout.

Overview

- Our tool is called *dirtree.*
- Dirtree recursively traverses a directory tree and prints out a sorted list of all files.

```
$ dirtree demo
demo
  subdir1
    sparsefile
    thisisanextremelylongfilenameforsuchasimplisticfile
  subdir2
    brokenlink
    symboliclink
  subdir3
    pipe
    socket
  one
  two
```

Overview

- ***Dirtree can also show details..***

```
$ dirtree -v demo
demo
  subdir1                                          sysprog:sysprog        4096 rwxrwxr-x  d
    sparsefile                                     sysprog:sysprog        8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic... sysprog:sysprog    1000 rw-rw-r--
  subdir2                                          sysprog:sysprog        4096 rwxrwxr-x  d
    brokenlink                                     sysprog:sysprog           8 rwxrwxrwx  l
    symboliclink                                   sysprog:sysprog           6 rwxrwxrwx  l
  subdir3                                          sysprog:sysprog        4096 rwxrwxr-x  d
    pipe                                           sysprog:sysprog           0 rw-rw-r--  f
    socket                                         sysprog:sysprog           0 rwxrwxr-x  s
  one                                              sysprog:sysprog           1 rw-rw-r--
  two                                              sysprog:sysprog           2 rw-rw-r--
```

Overview

- ***Dirtree can also show details or a summary of a directory***

```
$ dirtree -v -s demo
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo
  subdir1                                               sysprog:sysprog       4096 rwxrwxr-x  d
    sparsefile                                          sysprog:sysprog       8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic...  sysprog:sysprog       1000 rw-rw-r--
  subdir2                                               sysprog:sysprog       4096 rwxrwxr-x  d
    brokenlink                                          sysprog:sysprog          8 rwxrwxrwx  l
    symboliclink                                        sysprog:sysprog          6 rwxrwxrwx  l
  subdir3                                               sysprog:sysprog       4096 rwxrwxr-x  d
    pipe                                                sysprog:sysprog          0 rw-rw-r--  f
    socket                                              sysprog:sysprog          0 rwxrwxr-x  s
  one                                                   sysprog:sysprog          1 rw-rw-r--
  two                                                   sysprog:sysprog          2 rw-rw-r--
-------------------------------------------------------------------------------------------------
4 files, 3 directories, 2 links, 1 pipe, and 1 socket                         21497
```

Overview

- ***Dirtree can also show only directories***

```
$ dirtree -d -v -s demo
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo
  subdir1                                           sysprog:sysprog         4096 rwxrwxr-x   d
  subdir2                                           sysprog:sysprog         4096 rwxrwxr-x   d
  subdir3                                           sysprog:sysprog         4096 rwxrwxr-x   d
-------------------------------------------------------------------------------------------------
3 directories
```

Overview

- ***Dirtree can generate aggregate totals over several directories***

```
$ dirtree -v -s demo/subdir1 demo/subdir2
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo/subdir1                                                                                    
  sparsefile                                            sysprog:sysprog     8192 rw-rw-r--      
  thisisanextremelylongfilenameforsuchasimplisticfile   sysprog:sysprog     1000 rw-rw-r--      
-------------------------------------------------------------------------------------------------
2 files, 0 directories, 0 links, 0 pipes, and 0 sockets                     9192                

Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo/subdir2                                                                                    
  brokenlink                                            sysprog:sysprog        8 rwxrwxrwx  l    
  symboliclink                                          sysprog:sysprog        6 rwxrwxrwx  l    
-------------------------------------------------------------------------------------------------
0 files, 0 directories, 2 links, 0 pipes, and 0 sockets                       14                

Analyzed 2 directories:
  total # of files:                 2
  total # of directories:           0
  total # of links:                 2
  total # of pipes:                 0
  total # of sockets:               0
  total file size:               9206
```

demo/subdir1

demo/subdir2

Dirtree Specification(1/13)

- **Command line arguments**
  - *Dirtree* accepts the following command line arguments

```
dirtree [Options] [Directories]
```

> Options

Option	Description
-h	Help screen
-d	Turn on directory only mode
-v	Turn on detailed mode
-s	Turn on summary mode

> Directories

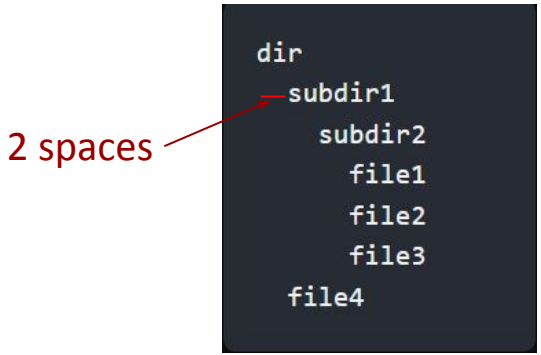
- A list of directories that are to be traversed.
- *Dirtree* accepts up to 64 directories.
- If no directory is given, then the current directory is traversed.

Dirtree Specification(2/13)

- **Operation**
  - ***Dirtree*** traverses each directory in the list *[Directories]* **recursively**.
  - In each directory, it enumerates all directory entries and prints them **in alphabetical order**.
    - Directories are listed before files.
    - The special entries '.' and '..' are ignored.
  - A summary is printed after each directory.
    - If several directories are traversed, **an aggregate total** is printed at the end.

Dirtree Specification(3/13)

- **Output**
  - As *dirtree* traverses the directory tree, it prints the names of the sorted entities in a directory.
  - The names **<u>are indented</u>** according to the level of the subdirectory.
  - For each additional level, the names are printed after **<u>two spaces</u>** to allow for easy visual identification of the directory structure.



Dirtree Specification(4/13)

- **Detailed mode**
  - In detailed mode, dirtree prints out the following additional details for each entry:
    - **User and group**
      - Each file in Unix belongs to a user and a group.
      - Detailed mode prints the names of the user and the group separated by a colon (:).
    - **Size**
      - The size of the file in bytes.
    - **Permissions**
      - The read/write/execute permissions for user owner, for group owner, and for others.
    - **File type**
      - Indicates the type of file by a single character

```
$ dirtree -v demo
demo
  subdir1                                          sysprog:sysprog        4096 rwxrwxr-x  d
    sparsefile                                     sysprog:sysprog        8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic... sysprog:sysprog    1000 rw-rw-r--
  subdir2                                          sysprog:sysprog        4096 rwxrwxr-x  d
    brokenlink                                     sysprog:sysprog           8 rwxrwxrwx  l
    symboliclink                                   sysprog:sysprog           6 rwxrwxrwx  l
  subdir3                                          sysprog:sysprog        4096 rwxrwxr-x  d
    pipe                                           sysprog:sysprog           0 rw-rw-r--  f
    socket                                         sysprog:sysprog           0 rwxrwxr-x  s
  one                                              sysprog:sysprog           1 rw-rw-r--
  two                                              sysprog:sysprog           2 rw-rw-r--
```

Type	Character
File	*(empty)*
Directory	d
Link	l
Character device	c
Block device	b
Fifo	f
Socket	s

Dirtree Specification(5/13)

- **Directory mode**
  - In summary mode, directory typed entries are printed only.

```
$ dirtree -d test1
test1
  a
    b
      c
        d
          e
  dir1
  dir2
  dir3
```

Dirtree Specification(6/13)

- **Summary mode**
  - In summary mode, dirtree prints **a header and footer** around each directory and a one-liner containing **statistics** about the directory.
    - The number of files, directories, links, pipes and socket.
    - Total file size
  - If there are more than one directories provided on the command line, **an aggregate total** of all listed directories is shown.

```
$ dirtree -v -s demo
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo
  subdir1                                               sysprog:sysprog       4096 rwxrwxr-x  d
    sparsefile                                          sysprog:sysprog       8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic...  sysprog:sysprog       1000 rw-rw-r--
  subdir2                                               sysprog:sysprog       4096 rwxrwxr-x  d
    brokenlink                                          sysprog:sysprog          8 rwxrwxrwx  l
    symboliclink                                        sysprog:sysprog          6 rwxrwxrwx  l
  subdir3                                               sysprog:sysprog       4096 rwxrwxr-x  d
    pipe                                                sysprog:sysprog          0 rw-rw-r--  f
    socket                                              sysprog:sysprog          0 rwxrwxr-x  s
  one                                                   sysprog:sysprog          1 rw-rw-r--
  two                                                   sysprog:sysprog          2 rw-rw-r--
-------------------------------------------------------------------------------------------------
4 files, 3 directories, 2 links, 1 pipe, and 1 socket                         21497
```

```
$ dirtree -v -s demo/subdir1 demo/subdir2
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo/subdir1
  sparsefile                                            sysprog:sysprog       8192 rw-rw-r--
  thisisanextremelylongfilenameforsuchasimplisticfile   sysprog:sysprog       1000 rw-rw-r--
-------------------------------------------------------------------------------------------------
2 files, 0 directories, 0 links, 0 pipes, and 0 sockets                       9192

Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo/subdir2
  brokenlink                                            sysprog:sysprog          8 rwxrwxrwx  l
  symboliclink                                          sysprog:sysprog          6 rwxrwxrwx  l
-------------------------------------------------------------------------------------------------
0 files, 0 directories, 2 links, 0 pipes, and 0 sockets                         14

Analyzed 2 directories:
  total # of files:                2
  total # of directories:          0
  total # of links:                2
  total # of pipes:                0
  total # of sockets:              0
  total file size:              9206
```

Dirtree Specification(7/13)

- **Summary mode**
  - If directory mode is also enabled, dirtree **only counts the number of directories.** Any other statistics including size will not be printed in the summary line.

```
$ dirtree -d -v -s test1
Name                                                User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------
test1
  a                                                 sysprog:sysprog     4096 rwxrwxr-x  d
    b                                               sysprog:sysprog     4096 rwxrwxr-x  d
      c                                             sysprog:sysprog     4096 rwxrwxr-x  d
        d                                           sysprog:sysprog     4096 rwxrwxr-x  d
          e                                         sysprog:sysprog     4096 rwxrwxr-x  d
  dir1                                              sysprog:sysprog     4096 rwxrwxr-x  d
  dir2                                              sysprog:sysprog     4096 rwxrwxr-x  d
  dir3                                              sysprog:sysprog     4096 rwxrwxr-x  d
-------------------------------------------------------------------------------------------
8 directories
```

Dirtree Specification(8/13)

- **Output formatting**
  - The output prints all elements with **the correct indentation.**

```
        1         2         3         4         5         6         7         8         9         10
1........0.........0.........0.........0.........0.........0.........0.........0.........0.........0

Name                                                User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
<path and name                                   >  <  user>:<group >  <    size> <  perms>  t
<path and name                                   >  <  user>:<group >  <    size> <  perms>  t
...
-------------------------------------------------------------------------------------------------
<summary                                                          >   <  total size>
```

Output element	Width	Alignment	Action on overflow
Path and name	54	left	cut and end with three dots
User name	8	right	ignore
Group name	8	left	ignore
File size	10	right	ignore
Permission	9	right	ignore
Type	1		
Summary line	68	left	limit to 68 characters
Total size	14	right	ignore

Dirtree Specification(9/13)

- **Output formatting**
  - The output in simple mode prints all elements with **the correct indentation.**
  - In detailed mode, the output is nicely formatted and filenames that are too long are **cut** and end with **three dots (...)**.
  - Unless explicitly specified, you can decide for yourself whether and how you are formatting exceptional cases (error messages, etc.)

```
$ dirtree -v -s demo2
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo2
  subdir1                                               sysprog:sysprog       4096 rwxrwxr-x  d
    subdir2                                             sysprog:sysprog       4096 rwxrwxr-x  d
      fifo                                              sysprog:sysprog          0 rw-rw-r--  f
      link                                              sysprog:sysprog         11 rwxrwxrwx  l
      socket                                            sysprog:sysprog          0 rwxrwxr-x  s
      unreasonablyextremelylongfilenamethatdoesntfi...  sysprog:sysprog          0 rw-rw-r--
  file1                                                 sysprog:sysprog          0 rw-rw-r--
-------------------------------------------------------------------------------------------------
2 files, 2 directories, 1 link, 1 pipe, and 1 socket                          8203
```

Dirtree Specification(10/13)

- **Output formatting**
  - The output in simple mode prints all elements with **the correct indentation.**
  - In detailed mode, the output is nicely formatted and filenames that are too long are **cut** and end with **three dots (...).**
  - Unless explicitly specified, you can decide for yourself whether and how you are formatting exceptional cases (error messages, etc.)
  - *Dirtree* takes great care to output **grammatically correct** English.
    - Zero or >=2 elements are output in plural form, while for exactly one element the singular form is used. Compare the two summary lines:

```
0 files, 2 directories, 1 link, 1 pipe, and 1 socket

1 file, 1 directory, 2 links, 0 pipes, and 5 sockets
```

Dirtree Specification(11/13)

- **Error handling**
  - Errors that occur when processing a directory are reported in place of the entries of that directory:

```
$ dirtree -v /etc/cups
/etc/cups
  ...
  interfaces                                  root:lp          4096 rwxr-xr-
  ppd                                         root:lp          4096 rwxr-xr-
    .keep_net-print_cups-0                    root:root           0 rw-r--r-
  ssl                                         root:lp          4096 rwx-----
    ERROR: Permission denied
  client.conf                                 root:root          31 rw-r--r-
  ...
```

Dirtree Specification(12/13)

- **Error handling**
  - Errors that occur when processing a directory are reported in place of the entries of that directory
  - If an error occurs when retrieving the metadata of a file, the error message is printed **in place of the file's meta data**:

```
$ dirtree -v -s demo3
Name                                                User:Group          Size     Perm
-------------------------------------------------------------------------------------
demo3
  dir1                                             sysprog:sysprog      4096 --x-----
    ERROR: Permission denied
  dir2                                             sysprog:sysprog      4096 r-------
    file1                                          Permission denied
    file2                                          Permission denied
    file3                                          Permission denied
  dir3                                             sysprog:sysprog      4096 rwx-----
    file4                                          sysprog:sysprog         0 -----x--
    file5                                          sysprog:sysprog         0 ----w---
    file6                                          sysprog:sysprog         0 ---r----
-------------------------------------------------------------------------------------
6 files, 3 directories, 0 links, 0 pipes, and 0 sockets                    12288
```

Dirtree Specification(13/13)

- **Error handling**
  - Errors that occur when processing a directory (permission errors) are reported in place of the entries of that directory
  - If an error occurs when retrieving the metadata of a file, the error message is printed **in place of the file's meta data**
  - For any other errors, **you can choose what to do.** The reference implementation aborts on most errors

```
$ dirtree -s -v demo
Name                                                User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------
demo
  subdir1                                           arcuser:users       4096 rwxrwxr-x  d
    sparsefile                                      arcuser:users       8192 rw-rw-r--
Out of memory.
```

Code & Test(1/4)

- **The skeleton provides data structures**
  - to manage the statistics of a directory,
  - a function to read the next entry from a directory while ignoring the '.' and '..' entries,
  - a comparator function to sort the entries of a directory using quicksort,
  - and full argument parsing and syntax helpers.

- **You have to implement the following two parts:**
  - in `main`()
    - Iterate through the list of directories stored in directories.
    - For each directory, call `processDir`() with the appropriate parameters.
  - in `processDir`()
    - Open, enumerate, sort, and close the directory. Print elements one by one. Update statistics.
    - If the element is a directory, call `processDir`() recursively.

Code & Test(2/4)

- **C library calls**
  - To help you get started, we provide a list of C library calls / system calls grouped by topic that you may find helpful to solve this lab.

Topic	C library call	Description
String operations	`strcmp()`	compare two strings
	`strncpy()`	copy up to n characters of one string into another
	`strdup()`	create a copy of a string. Use `free()` to free it after use
	`asprintf()`	asprintf() is extremely helpful to print into a string and allocate memory for it at the same time. We will show some examples during the lab session.
Directory management	`opendir()`	open a directory to enumerate its entries
	`closedir()`	close an open directory
	`readdir()`	read next entry from directory
File meta data	`stat()`	retrieve meta data of a file, follow links
	`lstat()`	retrieve meta data of a file, do not follow links
User/group information	`getpwuid()`	retrieve user information (including their name) for a given user ID
	`getgrgid()`	retrieve group information (including its name) for a given group ID
Sorting	`qsort()`	quick-sort an array
Error Handling	`strerror()`	Get string pointer for given error code. Refer `errno` man page for detailed error code description.

Code & Test(3/4)

- **The tools directory contains tools to generate test directory trees to test your solution. Invoke *gentree.sh* with a script file to generate one of the provided test directory trees.**

File/Directory	Description
gentree.sh	Driver script to generate a test directory tree.
mksock	Helper script to generate a Unix socket.
*.tree	Script files describing the directory tree layout.

> Assuming you are located in the root directory of your I/O lab repository, use the following command to generate the demo directory tree

```
$ ls
dirtree.c  Makefile  README.md  reference  tools
$ tools/gentree.sh tools/demo.tree
Generating tree from 'tools/demo.tree'...
Done. Generated 4 files, 2 links, 1 fifos, and 1 sockets. 0 errors reported.
```

Code & Test(4/4)

- **The tools directory contains tools to generate test directory trees to test your solution.**

```
$ reference/dirtree -v -s demo/
Name                                                    User:Group          Size     Perms Type
-------------------------------------------------------------------------------------------------
demo/
  subdir1                                               sysprog:sysprog       4096 rwxrwxr-x  d
    sparsefile                                          sysprog:sysprog       8192 rw-rw-r--
    thisisanextremelylongfilenameforsuchasimplistic...  sysprog:sysprog       1000 rw-rw-r--
  subdir2                                               sysprog:sysprog       4096 rwxrwxr-x  d
    brokenlink                                          sysprog:sysprog          8 rwxrwxrwx  l
    symboliclink                                        sysprog:sysprog          6 rwxrwxrwx  l
  subdir3                                               sysprog:sysprog       4096 rwxrwxr-x  d
    pipe                                                sysprog:sysprog          0 rw-rw-r--  f
    socket                                              sysprog:sysprog          0 rwxrwxr-x  s
  one                                                   sysprog:sysprog          1 rw-rw-r--
  two                                                   sysprog:sysprog          2 rw-rw-r--
-------------------------------------------------------------------------------------------------
4 files, 3 directories, 2 links, 1 pipe, and 1 socket                         21497
```

Grading Policy

- **Test bench : 80 %**
  - There will be hidden test cases than given example trees
- **Report** : **20 %**
  - Explain C library call that you used in your code
  - Briefly explain your code including following information
    - How to iterate through all files in a directory
    - how to retrieve the metadata of a file
    - how to print nicely formatted output
- **For late submission:**
  - A deduction of 20% p per 24 hours

Submission(via eTL)

- **Write-up**
  - Briefly describe your implementation.
  - Filename: [student_id].pdf (example: 2024-12345.pdf)
  - **Please** submit it in **pdf** format. Other formats are not accepted.
- **Compress your source code and write-up into a single file**
  - Compress **dirtree.c** and your report
  - Filename should be [student_id].tar (example: 2024-12345.tar).
  - **Please** submit it in **tar** format. Other formats are not accepted.
  - Refer README.md for submission instructions.
- **Submission deadline: by 23:59 on April 3, 2024**

Questions?