Lecture 07
File Input & Output

CS211 – Fundamentals of Computer Programming II

Branden Ghena – Winter 2023

Slides adapted from:
Jesse Tov

Administrivia

- Homework 2 due tonight
  - Remember that slip days exist

  - Beware: office hours are overloaded
    - Prepare for long delays until you can get help, and only high-level help
    - Feel free to ask questions on Piazza too
      - I'll be checking it frequently

- No more exercises for two weeks!
  - Get started on Homework 3 early instead

- Homework 3 has two parts
  - Part 1 due next week
  - Part 2 due in two weeks

Homework 2 hint: comparing strings

```
char* a = “abc”
char b[4] = {‘a’, ‘b’, ‘c’, ‘\0’}
if (a == b) {
  print(“They match!\n”);
} else {
  print(“They do not match\n”);
}
```

This code prints: “They do not match\n”. Why?

What does `a == b` compare?

Two pointers!

Strings must be compared with strcmp()

- https://www.cplusplus.com/reference/cstring/strcmp/

- int strcmp(const char* str1, const char* str2)
  - Compares two strings character-by-character until reaching a '\0'
  - Returns an integer value of the following:
    - <0 str1 comes before str2 alphabetically
    - 0 str1 is equal to str2
    - >0 str1 comes after str2 alphabetically

SEGV is a null pointer dereference

```
Check failed (test/test_vc.c:36):
  assertion: cp
test/test_vc.c:37:9: runtime error: load of null pointer of type 'size_t'
AddressSanitizer:DEADLYSIGNAL
=================================================================
==1167490==ERROR: AddressSanitizer: SEGV on unknown address 0x000000000000 (pc 0x000000400cbe bp 0x000000000000 sp 0x7ff
e84837c00 T0)
==1167490==The signal is caused by a READ memory access.
==1167490==Hint: address points to the zero page.
SCARINESS: 10 (null-deref)
    #0 0x400cbd in test_2_candidates test/test_vc.c:37
    #1 0x400cbd in main test/test_vc.c:66
    #2 0x7f544b789492 in __libc_start_main (/lib64/libc.so.6+0x23492)
    #3 0x400a8d in _start (/home/slc8828/cs211/hw03/.bin/test_vc-16+0x400a8d)

AddressSanitizer can not provide additional info.
SUMMARY: AddressSanitizer: SEGV test/test_vc.c:37 in test_2_candidates
==1167490==ABORTING
```

- This AddressSanitizer error is due to dereferencing a NULL pointer
  - Often in Homework 3, it's because you tried to read a NULL candidate name
  - Possibly with `strcmp()`

Today’s Goals

- Practice dynamic memory allocation with arrays
  - How do we make an array the dynamically changes size?
- Introduce and explore concept of linked lists
  - What are they and what are their advantages?
  - How do we write code that uses them?
- Discuss concept of pointers to pointers

Getting the code for today

`cd ~/cs211/lec/` (or wherever you put stuff)

```
tar -xkvf ~cs211/lec/07_fileio.tgz
cd 07_fileio/
```

Outline

- **Ownership Review**
- File Input & Output (I/O)
- Standard I/O
- Dynamic Arrays

Review: ownership idea

- `malloc()` creates memory objects (chunks of heap memory)
  - MUST later be freed

- The “owner” of a memory object is responsible for it
  - Must either `free()` it

  - Or transfer ownership to something else
    - Pass into another function
    - Store it in some data structure for later

The full ownership protocol

- The owner of a heap-allocated object is responsible for deallocating it
  - No one else may do so
- Borrowers of an object may access or modify it
  - But they may not hold on to a reference to it or deallocate it
- Passing or returning a pointer *may or may not* transfer ownership
  - Transfer: caller must have owned it previously and now give up ownership
  - No transfer: caller could also be borrowing. New function is borrowing

Borrowing example

Function A

- **borrows** memory object

Can use and modify memory
Cannot free() memory
Cannot store memory for later

Calls Function B and passes the memory

Function B

- MUST also be borrowing

Can use and modify memory
Cannot free() or store memory

After return
Can still use and modify memory

Ownership example, return transferring ownership

Function A

- has no memory object

Calls Function B

After return
Can use and modify memory
Must free(), store, or transfer memory

Function B

- **creates** memory object (malloc)

Can use and modify memory
Must free(), store, or transfer memory

Returns object, transferring ownership

Ownership example, calling a borrowing function

Function A
- **owns** memory object

Can use and modify memory
Must free(), store, or transfer memory

Calls Function B and passes the memory

Function B
- **borrows** memory object

Can use and modify memory
Cannot free() or store memory

After return
Can still use and modify memory
Must free(), store, or transfer memory

Ownership example, transferring ownership

Function A

- **owns** memory object

Can use and modify memory
Must free(), store, or transfer memory

Calls Function B and passes the memory

Function B

- **takes ownership** of memory object

Can use and modify memory
Must free(), store, or transfer memory

After return
Cannot access memory

Break + Practice

- Example from Homework 3
  - `void ballot_insert(ballot_t ballot, char* name)`
    - Borrows `ballot` transiently
    - Takes ownership of `name`

- What is `ballot_insert()` allowed to do to `ballot`?

- What is `ballot_insert()` allowed to do to `name`?

Break + Practice

- Example from Homework 3
  - `void ballot_insert(ballot_t ballot, char* name)`
    - Borrows `ballot` transiently
    - Takes ownership of `name`

- What is `ballot_insert()` allowed to do to `ballot`?
  - Can modify and use. Cannot `free()` or store.

- What is `ballot_insert()` allowed to do to `name`?
  - Can modify and use. MUST `free()` or store.

Outline

- Ownership Review
- **File Input & Output (I/O)**
- Standard I/O
- Dynamic Arrays

Files

- Collections of data
  - Usually in permanent storage on your computer

- Types of files
  - Regular files
    - Arbitrary data
    - Think of each file as a big array of bytes (just like memory)

  - Directories
    - Collections of regular files

  - Special files
    - Links, pipes, devices (see CS343)

How do we interact with files?

- Analogy: think of a file as a book
  - Big array of characters (bytes)

1. Open the book, starting at the first page
2. Read from the book
3. Write to the book
4. Change pages (without reading everything in between)
5. Close the book when finished

System calls for interacting with files

1. Open the book, starting at the first page
   - fopen()
2. Read from the book
   - fread()
3. Write to the book
   - fwrite()
4. Change pages (without reading everything in between)
   - fseek()
5. Close the book when finished
   - fclose()

References

- https://www.cplusplus.com/reference/cstdio/
  - Explanation of and links for everything in <stdio.h>

Opening files

```
FILE* fopen(const char* filename, const char* mode);
```

- `filename` is the string path for the file
  - “/home/branden/cs211/s23/hw/hw1/src/tr.c”
  - “./arguments.c”
  - “arguments.c”

- `mode` specifies what you intend to do with the file
  - “r” - read only (must exist)
  - “w” - write (overwrites if exists)
  - “a” - append (starts writing at end of file if exists)

Open returns a FILE object

```
FILE* fopen(const char* filename, const char* mode);
```

- Pointer type for an object used to interact with the file
  - A “handle” to the file
- Other file interaction functions will take in a `FILE*` as an argument
  - Don’t need to remember the file path and look it up every time
- `NULL` instead specifies an error attempting to open the file

Reading files

```
size_t fread(void* ptr, size_t size, size_t count, FILE* stream);
```

- `ptr` is a pointer to an array to read into
  - At least `size×count` bytes in length
- `size` is the number of bytes for each element in the array
- `count` is the number of elements to read
- `stream` is the file pointer returned from a previous call to `fopen()`

- Note: nowhere do we specify where to *start* reading
  - Library keeps track of a file offset with the file
  - Updated on each read
    - First read of 100 bytes starts at zero, next starts 100 bytes in

How do we know when we finished the file?

```
size_t fread(void* ptr, size_t size, size_t count, FILE* stream);
```

- Return from read is the count of elements *actually* read
  - Less than `count` means there was either an error or end-of-file was reached

- `feof()` lets you check if end-of-file was reached
- `ferror()` lets you check for particular errors

Writing files looks a lot like reading

```
size_t fwrite(const void* ptr, size_t size, size_t count,
                FILE* stream);
```

- Array to write from, size of elements in the array, number of elements to write, and a file pointer

- Returns number of elements *actually* written

- Write occurs at the current file offset

Moving the file offset

```
int fseek(FILE* stream, long int offset, int origin);
```

- Moves to offset for this file descriptor based on origin:
  - SEEK_SET – set to offset (essentially start of file plus offset)
  - SEEK_CUR – current location plus the offset
  - SEEK_END – end of file plus the offset (which should be negative)
- Returns zero if successful
  - Anything else means an error occurred
- `ftell()` gets the current location in a file
  - So you can seek back there later

Closing a file

```
int fclose(FILE* stream);
```

- Closes the file
- Returns zero on success
- It is an error to keep using the file descriptor after it is closed
    - Just like with dynamic memory management

Buffered I/O

- C standard library buffers your interactions to make them more efficient
  - One big write to a file is MUCH faster than many small writes

- Sometimes you want to write to output *right now*
  - `fflush()` guarantees that the buffer is written *now*
  - Otherwise no write is guaranteed until `fclose()` is called

- Example: `printf()` buffers until a newline is reached
  - So a print right before a fault might not appear unless it includes a '\n' 😱

Example: kitten tool

- Command line tool: `cat` – prints out the contents of files
  - Does so very efficiently

- Our program: `kitten` – prints out the contents of one file
  - No efficiency promises

- Implementing `kitten` only requires file I/O calls we've discussed!

Live coding: implement kitten

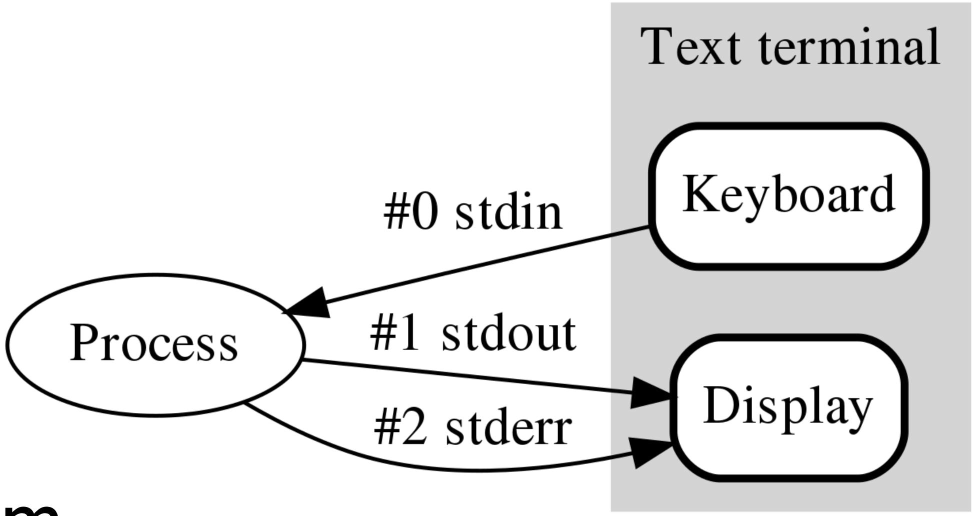
kitten-starter.c
kitten-complete.c

- Requirements
  - Parse argv[] to find file to open
  - Open the file
  - Read in lines from the file repeatedly
    - If end-of-file is reached, break (`feof()`)
    - Print contents of file
  - Handle errors

Outline

- Ownership Review
- File Input & Output (I/O)
- **Standard I/O**
- Dynamic Arrays

How do programs talk to users?



- We glossed over this before
  - `printf()`
  - `scanf()`
- Work through the same file mechanism
  - Three special files created for each program
  - stdin – standard input
  - stdout – standard output
  - stderr – standard error
- `printf()` -> `fprintf(stdout)` -> handle arguments & `fwrite(stdout)`

Standard I/O is a process thing, not a C thing

- You can access them in Python, for instance
  - https://docs.python.org/3/library/sys.html#sys.stdin

`sys.stdin`
`sys.stdout`
`sys.stderr`

File objects used by the interpreter for standard input, output and errors:

- `stdin` is used for all interactive input (including calls to `input()`);
- `stdout` is used for the output of `print()` and expression statements and for the prompts of `input()`;
- The interpreter's own prompts and its error messages go to `stderr`.

These streams are regular text files like those returned by the `open()` function. Their parameters are chosen as follows:

Standard I/O is configured by the shell

- When you run a program in command line, the shell attaches a standard input, standard output, and standard error to it

- Defaults
  - stdin - read from terminal
  - stdout - write to terminal
  - stderr - write to terminal

Live coding: kitten upgrades

kitten-starter.c
kitten-complete.c

- Errors should be written to stderr
- Output can be written to stdout directly using `fwrite()`
  - Instead of using `printf()` in a loop to do it for us

Redirecting standard I/O

- Shells by default setup standard I/O to connect to the keyboard and the screen
  - But any file will also work

- Shell I/O redirection commands
  - COMMAND < filename
    - Connect standard input to filename
  - COMMAND > filename
    - Connect standard output to filename (overwrite)
  - COMMAND >> filename
    - Connect standard output to filename (append)

Piping commands

- A command shell desire is to run multiple commands where the output of the first feeds into the second

- COMMAND1 | COMMAND2
  - Connects stdout of COMMAND1 to stdin of COMMAND2

- Example: print out files and sort by size
  - ls –lah | sort –h

Sidebar: super useful command for testing

- **tee** [*OPTION*]... [*FILE*]...
  - Reads from stdin and write to **both** stdout and file

- Example: prints out a list of files and saves results
  - ls –lah | tee results.txt

- I run this with various programs I'm testing, so I can record the results, but also seem them in real-time.

Example: redirection with kitten

- Standard I/O redirection is handled when the process is created
  - So it does not need to be aware of it at all
- Our kitten tool works with redirection automatically!
  - ./kitten arguments.c > OUTPUT_FILE

Break + Thinking Excercise

- Take a look at the `cat` command to see the other flags it supports

```
-A, --show-all
       equivalent to -vET

-b, --number-nonblank
       number nonempty output lines, overrides -n

-e     equivalent to -vE

-E, --show-ends
       display $ at end of each line

-n, --number
       number all output lines

-s, --squeeze-blank
       suppress repeated empty output lines

-t     equivalent to -vT

-T, --show-tabs
       display TAB characters as ^I

-u     (ignored)

-v, --show-nonprinting
       use ^ and M- notation, except for LFD and TAB
```

How hard would these be to implement in `kitten`?

Outline

- Ownership Review
- File Input & Output (I/O)
- Standard I/O
- **Dynamic Arrays**

Dealing with dynamic input

- What if you want to read in data, but you don't know how much data there might be?

- Arrays in C are a fixed size
- But you can `malloc()` as many times as needed
  - Request some memory
  - Use until you run out
  - Request more memory and copy existing values over

  - `realloc()` makes this simple

Example of dynamic memory: read_line()

```
char* read_line(void)
```

- Reads an entire line at a time from stdin
  - Can't know in advance how many bytes there will be to read

  - Keeps reading in bytes until '\n' character or end-of-file
  - Needs to request more memory until it holds the entire line

- Note: part of the 211 library, not standard C

Live coding: implement read_line()

readline-starter.c
readline-complete.c

```
char* read_line(void)
```

- Requirements
  - Read from stdin until '\n' or end-of-file (EOF)
  - Allocate an array to hold the read characters
    - Make sure to end it with a '\0'
  - Returns
    - NULL pointer if EOF was reached immediately
    - Pointer to string otherwise (not including the newline character)

Realloc versus malloc

- We could just `malloc()` and copy ourselves, what does `realloc()` add?

- `realloc()` can be far more efficient
  - Doesn't have to copy data at all if there is room in the heap to expand

- Also simpler for programmers
  - Can't forget to free the old memory if `realloc()` does it for you

Default string size will change efficiency

- Memory efficiency
  - Pointer returned could have way more memory than characters
  - User might hold on to memory for a while before freeing
  - The less wasted memory, the less memory the program needs

- Runtime speed
  - `malloc()` and `realloc()` are slow
  - The fewer times we call them, the faster the program will run

- Need to pick a sweet spot to balance the two of these
  - Real program: starts at 80 characters, doubles size when reallocating

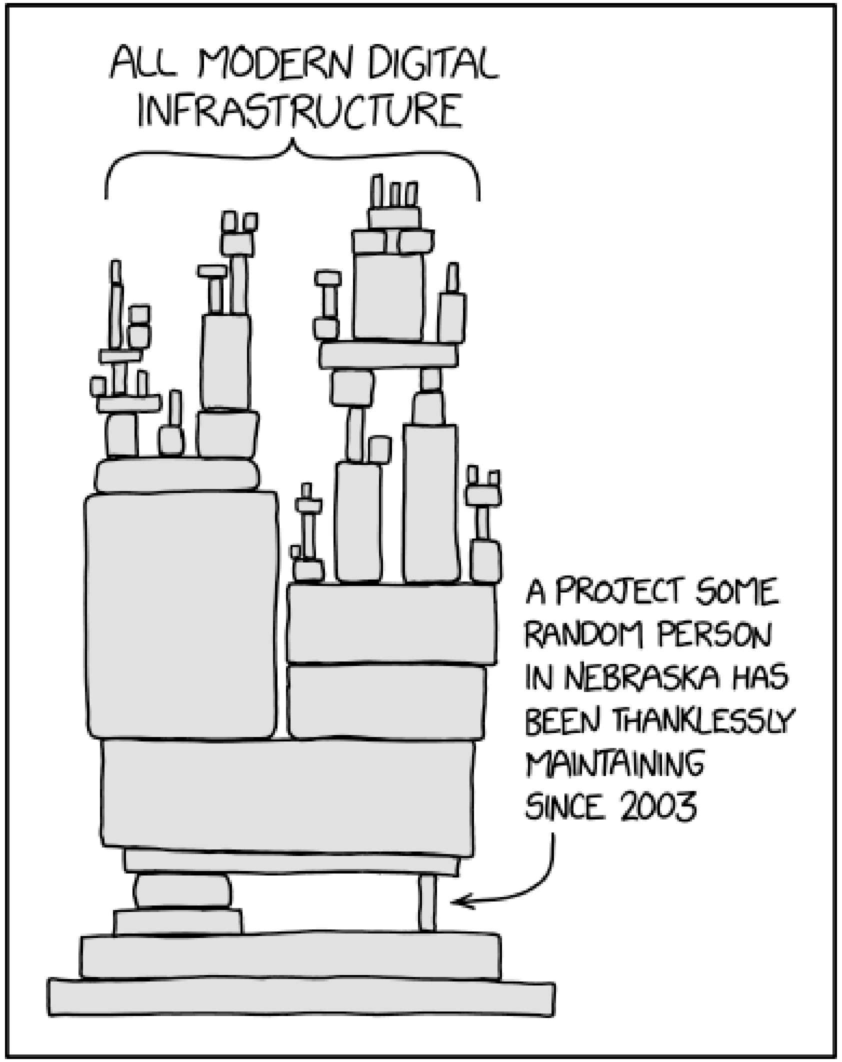
Does efficiency really matter though?

- If you’re writing a CS211 homework: **no**

- If you’re writing a Javascript interpreter for Firefox,
  - Which has millions of users
  - times hundreds of websites per day for each user
  - times hundreds of lines of code per website
  - and each line of code is read with `read_line()`

  - **YES**

Break + relevant xkcd



https://xkcd.com/2347/

Outline

- Ownership Review
- File Input & Output (I/O)
- Standard I/O
- Dynamic Arrays