AI INSTITUTE #AIISC
UNIVERSITY OF SOUTH CAROLINA

UNIVERSITY OF SOUTH CAROLINA

CSCE 240: Advanced Programming Techniques
Lecture 3: Input and Output

PROF. BIPLAV SRIVASTAVA, AI INSTITUTE

16TH JANUARY 2024

Carolinian Creed: "I will practice personal and academic integrity."

Credits: Some material reused with permission of Dr. Jeremy Lewis. Others used as cited with thanks.

Organization of Lecture 3

- Introduction Section
  - Recap of Lecture 2
  - Logistics: Piazza, Github, Spreadsheet
- Main Section
  - Review additional tasks: Sorting of numbers
  - Concept: Data types
  - Concept: Strings
  - Concept: Streams, Input and Output
  - Concept: Error Handling
  - Numeric processing: Calculator
  - Team activity: code review, testing
- Concluding Section
  - About next lecture – Lecture 4
  - Ask me anything

Introduction Section

Recap of Lecture 2

- We discussed
  - Computer architecture
  - Concepts of pointers and iteration
- Looked at sorting of numbers in different languages

Logistics: Slack, Github, Spreadsheet

- Piazza – please sign-in - https://piazza.com/sc/spring2024/csce240h
- Github –
  - Create and share your private repo on “Students and Code Links” spreadsheet
  - Name the repo: CSCE240H-Spring2024-<your-first-name>. Example: CSCE240H-Spring2023-Bob
  - Additionally
    - Put a readme file (Readme.md) and give you full name
    - Create a sub-folder called “home-works”. Future home works will be here
    - Create a sub-folder called “project”. Your project will reside here.
- Spreadsheet for tracking student information (‘StudentEditable - Students and Code Links’)

Main Section

Review: Sorting of Numbers

- C++, Java, Python
- Using peer-reviewing

Concept: Data Types

Common C++ types

Type	Typical Bit Width	Typical Range
char	1byte	-127 to 127 or 0 to 255
unsigned char	1byte	0 to 255
signed char	1byte	-127 to 127
int	4bytes	-2147483648 to 2147483647
unsigned int	4bytes	0 to 4294967295
signed int	4bytes	-2147483648 to 2147483647
short int	2bytes	-32768 to 32767
unsigned short int	2bytes	0 to 65,535
signed short int	2bytes	-32768 to 32767
long int	8bytes	-2,147,483,648 to 2,147,483,647
signed long int	8bytes	same as long int
unsigned long int	8bytes	0 to 4,294,967,295
long long int	8bytes	-(2^63) to (2^63)-1
unsigned long long int	8bytes	0 to 18,446,744,073,709,551,615
float	4bytes	
double	8bytes	
long double	12bytes	
wchar_t	2 or 4 bytes	1 wide character

Credit and Reference: https://www.tutorialspoint.com/cplusplus/cpp_data_types.htm

Concept: Strings in Languages

- C: unsigned array of characters
  - Methods to copy, find length,…
  - Array operators
- C++: standard class
  - Methods to find length, compare
- Java: built-in data type
- Python: built-in data type; sequence of Unicode

Reference:

- Python: https://docs.python.org/3/library/stdtypes.html#textseq, https://docs.python.org/3/library/string.html

Concept: Streams, Input, Output

- **Streams**: an abstraction of end-point (file, terminal, string, web resource - source, destination) with characters in sequential order of any length.
- Input
  - cin: command line for input
  - ifstream: file for input
- Output
  - cout: command line for output
  - ofstream: file for output
- Both
  - fstream, sstream

- Pattern for using streams
  1. Open a stream
  2. Do operation (read, write or both)
  3. Close the stream

Reference:

- C++ IO classes - https://www.cplusplus.com/reference/iolibrary/, https://www.cprogramming.com/tutorial/c++-iostreams.html

Concept: Streams, Input, Output

- Pattern for using streams
  1. Open a stream
  2. Do operation (read, write or both)
  3. Close the stream

open (filename, mode);

`ios::in`	Open for input operations.
`ios::out`	Open for output operations.
`ios::binary`	Open in binary mode.
`ios::ate`	Set the initial position at the end of the file. If this flag is not set, the initial position is the beginning of the file.
`ios::app`	All output operations are performed at the end of the file, appending the content to the current content of the file.
`ios::trunc`	If the file is opened for output operations and it already existed, its previous content is deleted and replaced by the new one.

Reference:

- C++ IO classes - https://www.cplusplus.com/reference/iolibrary/, https://www.cprogramming.com/tutorial/c++-iostreams.html
- Table courtesy - https://www.cplusplus.com/doc/tutorial/files/

Concept: Error Handling

- Error is unavoidable. Think ahead.
- Good programs have stable behavior; they handle known and unknown situations
- Error can be tested on streams anytime.
  - bad(): Returns true if a reading or writing operation fails.
  - fail(): Returns true in the same cases as bad(), but also in the cases of a format.
  - eof(): Returns true if a file open for reading has reached the end.
  - good(): It is the most generic state flag returning false if any of the previous functions would return true.

Reference:

- Table courtesy - https://www.cplusplus.com/doc/tutorial/files/

Illustrate Concepts

- C++
  - Class3and4_C++_IO.cpp
  - Notice: code and data are in separate folders

Illustration: Java

- Notice the three libraries for input and output, respectively
- Notice try/ catch to handle errors

Illustration: Python

- File: L3_DemoReadWrite.py
- Main code is 3 lines: open, read/ write, close
- No data typing

Review C++ With Peers – In Class

- Code walk through
- Unit testing

Practice Programming (10 mins)

- Implement sorting of numbers in C++
- Implement sorting in another language (Java or Python)
- Add code on personal GitHub
- Update Instructor when done

Peer Review: Homework Assignment #1

1. Go to spread sheet and on ”Homework Assignments - Peer Review” tab
2. Go to the row with your name
3. Peer review (10 mins)
    1. Enter roll number of person on your **LEFT** under “ID of code reviewer”
    2. Share code for the reviewer to see
    3. Reviewer: enter review (1-5)
    4. **Note**: negotiate – review code of neighbor or get own’s code reviewed
4. Peer test (10 mins)
    1. Enter roll number of person on your **RIGHT** under “ID of code tester”
    2. Share command line for the tester to see
    3. Tester: enter review (1-5)
    4. **Note**: negotiate – test code of neighbor or get own’s code tested

Peer Reviewing Guideline (10 mins)

- Look out for
  - Can you understand what the code is doing ?
  - Can you explain the code to someone else (non-coder) ?
  - Can you spot possible issues without running it?
    - Are the variables initialized ?
    - Are files closed?
    - Is their unnecessary code bloat ?
- What not to judge
  - Usage of language features, unless they are inappropriate
- Assign rating
  - 1: code not available
  - 2: code with major issues
  - 3: code with minor issues
  - 4:
  - 5: no issues

Peer Testing Guideline (10 mins)

- Look out for
  - Does the program run as the coder wanted it to be (specification) ?
  - Does the program run as the instructor wanted it to be (requirement - customer) ?
  - Does the program terminate abruptly ?
  - Any special feature?
- What not to judge
  - Person writing the code
- Assign rating
  - 1: code not available
  - 2: code runs with major issues (abnormal termination, incomplete features)
  - 3: code runs with minor issues
  - 4:
  - 5: No issues

Programming Home Work (#1) – C++

- Write a program called FileBasedCalculator.
  - It reads three lines from an input file (called input.txt): the operation to be done (add, subtract, multiply or divide), and two integer numbers.
  - It writes two lines to an output file (called output.txt). The first line says – ”The result of <operation> on <num1> and <num2> is below”. The second line has the result.
- Modify the program so that it can handle missing input file name.

Discussion: Course Project

2024 Project Choices

- Choice 1: (Chatbot to) Recommend courses to students in a department for a semester
  - User: students
  - Data: course catalog, course description
  - Student choice: department and its courses
  - AI: recommendation, chatbot
- Choice 2: (Chatbot to) answer questions about airline compensation and flight delays
  - User: students
  - Data: airlines' policies, government regulations
  - Student choice: airline, countries of travel
  - AI: data cleaning and organization, chatbot
- Choice 3: …

2024 Project Choice

- Choice 3: (Chatbot to) know about companies from their government filings
  - User: anyone
  - Data: 10-k filings
  - Student choice: companies
  - AI: data cleaning and organization, chatbot
- Choice 4: …

Course Project – Companies Data

- Libraries of Congress: https://guides.loc.gov/company-research/financials
- EDGAR: https://www.sec.gov/edgar/searchedgar/companysearch
  - Apple: https://www.sec.gov/edgar/browse/?CIK=320193&owner=exclude
    - 10-k: https://www.sec.gov/ix?doc=/Archives/edgar/data/320193/000032019323000106/aapl-20230930.htm#i1cb1ba018cb1455aa66bd3f9ab0c5b1a_175
  - Pfizer info: https://www.sec.gov/edgar/browse/?CIK=78003&owner=exclude
    - 10-k: https://www.sec.gov/ix?doc=/Archives/edgar/data/78003/000007800323000024/pfe-20221231.htm#i8050b09ca8a0411dbcb0b6576ce1fc7a_298

Comparing 10-Ks

- Table of Contents for Pfizer and Apple
- Note 4-parts of the report

PFIZER	Apple	
PART I	Part I	
ITEM 1. BUSINESS	Item 1.	Business
About Pfizer	Item 1A.	Risk Factors
Commercial Operations	Item 1B.	Unresolved Staff Comments
Research and Development	Item 1C.	Cybersecurity
Collaboration and Co-Promotion Agreements	Item 2.	Properties
International Operations	Item 3.	Legal Proceedings
Sales and Marketing	Item 4.	Mine Safety Disclosures
Patents and Other Intellectual Property Rights		
Competition	Part II	
		Market for Registrant's Common Equity, Related Stockholder Matters and
Pricing Pressures and Managed Care Organizations	Item 5.	Issuer Purchases of Equity Securities
Raw Materials	Item 6.	[Reserved]
		Management's Discussion and Analysis of Financial Condition and Results of
Government Regulation and Price Constraints	Item 7.	Operations
Environmental Matters	Item 7A.	Quantitative and Qualitative Disclosures About Market Risk
Human Capital	Item 8.	Financial Statements and Supplementary Data
		Changes in and Disagreements with Accountants on Accounting and Financial
ITEM 1A. RISK FACTORS	Item 9.	Disclosure
ITEM 1B. UNRESOLVED STAFF COMMENTS	Item 9A.	Controls and Procedures
ITEM 2. PROPERTIES	Item 9B.	Other Information
ITEM 3. LEGAL PROCEEDINGS	Item 9C.	Disclosure Regarding Foreign Jurisdictions that Prevent Inspections
ITEM 4. MINE SAFETY DISCLOSURES		
INFORMATION ABOUT OUR EXECUTIVE OFFICERS	Part III	
	Item 10.	Directors, Executive Officers and Corporate Governance
PART II	Item 11.	Executive Compensation
ITEM 5. MARKET FOR THE COMPANY'S COMMON EQUITY, RELATED		Security Ownership of Certain Beneficial Owners and Management and
STOCKHOLDER MATTERS AND ISSUER PURCHASES OF EQUITY SECURITIES	Item 12.	Related Stockholder Matters
ITEM 6. [RESERVED]	Item 13.	Certain Relationships and Related Transactions, and Director Independence
ITEM 7. MANAGEMENT'S DISCUSSION AND ANALYSIS OF FINANCIAL		
CONDITION AND RESULTS OF OPERATIONS	Item 14.	Principal Accountant Fees and Services
ITEM 7A. QUANTITATIVE AND QUALITATIVE DISCLOSURES ABOUT MARKET		
RISK		
ITEM 8. FINANCIAL STATEMENTS AND SUPPLEMENTARY DATA	Part IV	
ITEM 9. CHANGES IN AND DISAGREEMENTS WITH ACCOUNTANTS ON		
ACCOUNTING AND FINANCIAL DISCLOSURE	Item 15.	Exhibit and Financial Statement Schedules
ITEM 9A. CONTROLS AND PROCEDURES	Item 16.	Form 10-K Summary
ITEM 9B. OTHER INFORMATION		
ITEM 9C. DISCLOSURE REGARDING FOREIGN JURISDICTIONS THAT PREVENT INSPECTIONS		
PART III		
ITEM 10. DIRECTORS, EXECUTIVE OFFICERS AND CORPORATE GOVERNANCE		
ITEM 11. EXECUTIVE COMPENSATION		
ITEM 12. SECURITY OWNERSHIP OF CERTAIN BENEFICIAL OWNERS AND		
MANAGEMENT AND RELATED STOCKHOLDER MATTERS		
ITEM 13. CERTAIN RELATIONSHIPS AND RELATED TRANSACTIONS, AND		
DIRECTOR INDEPENDENCE		
ITEM 14. PRINCIPAL ACCOUNTING FEES AND SERVICES		
PART IV		
ITEM 15. EXHIBITS, FINANCIAL STATEMENT SCHEDULES		
15(a)(1) Financial Statements		
15(a)(2) Financial Statement Schedules		
15(a)(3) Exhibits		
ITEM 16. FORM 10-K SUMMARY		
SIGNATURES		
N/A = Not Applicable		

Course Project – Knowing About Companies

- **Project**: Develop collaborative assistants (chatbots) that offer useful information about companies
- Specifically, use the EDGAR dataset on companies at: https://www.sec.gov/edgar/searchedgar/companysearch.
  - For Apple, it is: https://www.sec.gov/edgar/browse/?CIK=320193&owner=exclude
  - Each student will choose two companies (from thousand available).
- Programming assignment programs will: (1) extract data about two companies from 10-k, (2) process it, (3) make content available in a command-line interface, (4) handle any user query and (5) report on interaction statistics.

Discussion: Nature and Simplifications

- Once you select a company, the content is also fixed.
  - Enter selection in column F of spreadsheet
- Some simplifications
  - **Download local copy** v/s web query
  - **Read static content first**
  - **Handle a subset of content**
  - **Have default handling for questions** the chatbot does not understand
- Do project in a language you are most comfortable with
- Use all advanced programming concepts to simplify coding

Discussion: Chatbot Loop

- Input: from user (called utterance)
  - Problem specific query (i.e., about company risk factors)
  - Chitchat
  - Unrelated
- Output: from system (response)
  - Handle unrelated
  - Handle chitchat
  - Answer to query

- **Do it until user say over!**

Concluding Section

Lecture 3: Concluding Comments

- We discussed the concepts of data types, strings
- We discussed the concepts of streams and error handling
- We looked at programs in C++, Java and python on file handling
- Homework assignment - FileBasedCalculator
- Discussed projects

About Next Lecture – Lecture 4

Lecture 4: I/O

- Will review Home Assignment-1 (FileBasedCalculator) in class
- Handling mixed data types
- Error handling
- Printing values with formatting