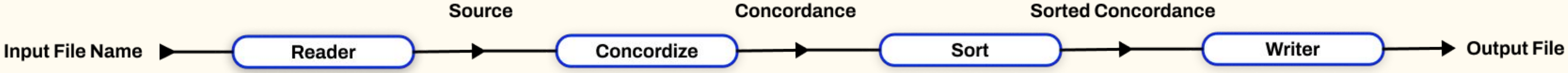
KWIC

TIC3001 Group 13

Roshan A0094621H
Noel A0211270Y

11 April 2022

Assignment 1 - KWIC



Input File Name → Reader → Source → Concordize → Concordance → Sort → Sorted Concordance → Writer → Output File

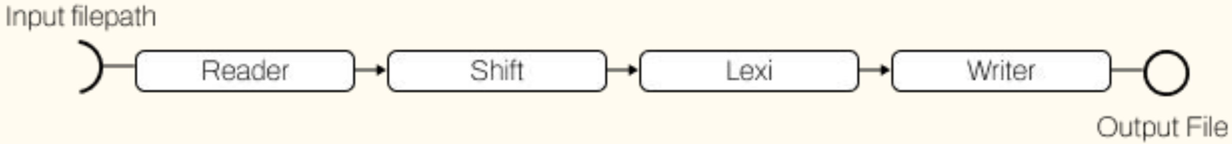
1) Reader accepts a filename and reads the file content as list of lines.
2) Concordance is populated from the source (Circular shifting).
3) Concordance is sorted.
4) Writer buffers concordance into an output file.

Architectural Designs

1. Pipe & Filter

2. Abstract Data Types and Objects

KWIC - Pipe & Filter



Filter Components

Reader Takes in input file data and process the source data as an array of lines.
A line is an array of words.

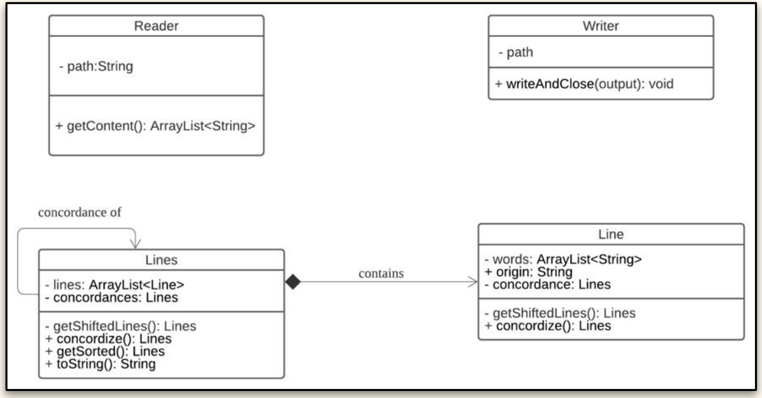
Shift Produce concordance from source.

Lexi Outputs concordance in lexicographical order.

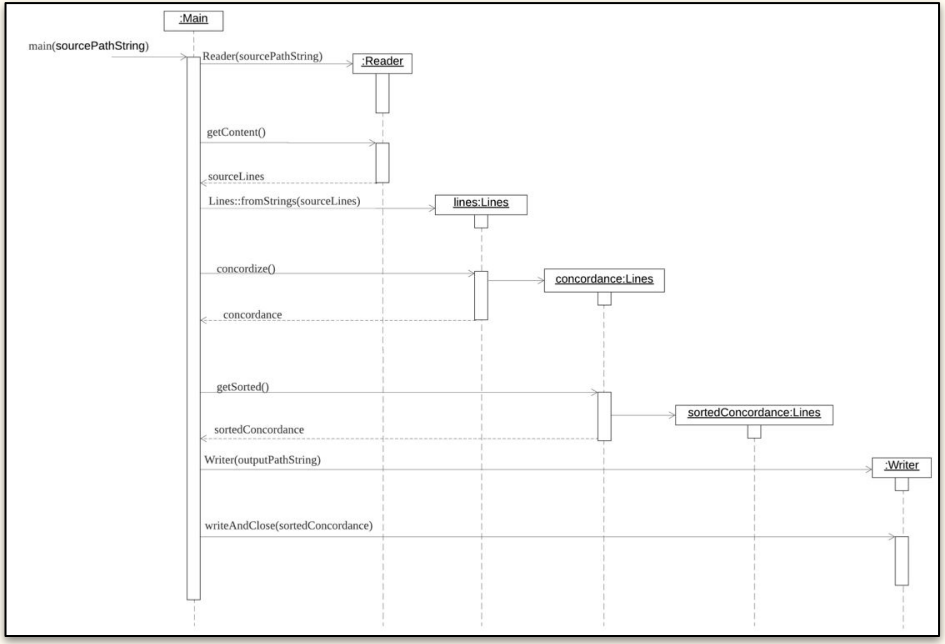
Write Write concordance to file.

KWIC - Abstract Data Types and Objects

Class Diagram



Sequence Diagram



Reader/Writer

- SRP

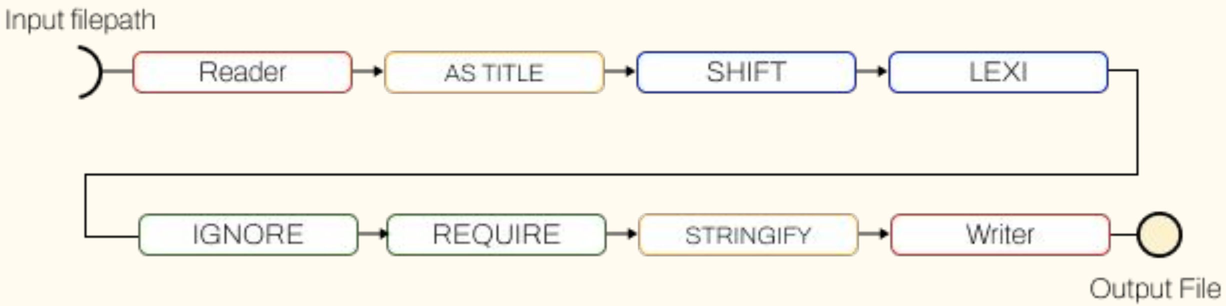
Lines

- Reusability (origin, concordance, sortedConcordance…)
- Data Cohesion

Assignment 2 - KWIC Extension

1) Read data from file.
2) Parse data as an array of lines, with each line an array of words.
3) Concordance is populated from the source (Circular shifting).
4) Sort lines by alphabetical order.
5) **Filter by omitting lines starting with ignore keywords.**
6) **Filter by selecting lines starting with require keywords.**
7) Combine lines as text.
8) Buffer test into output file.

KWIC Extension - Pipe & Filter 🚰



Motivation

- Main execution is a sequence of data transformations.
- Each filter is a unit of responsibility.
- Defined input and output interface (value & type) of filter components.

KWIC Extension - Pipe & Filter

1) Read…

2) Parse…

3) … Circular shifting

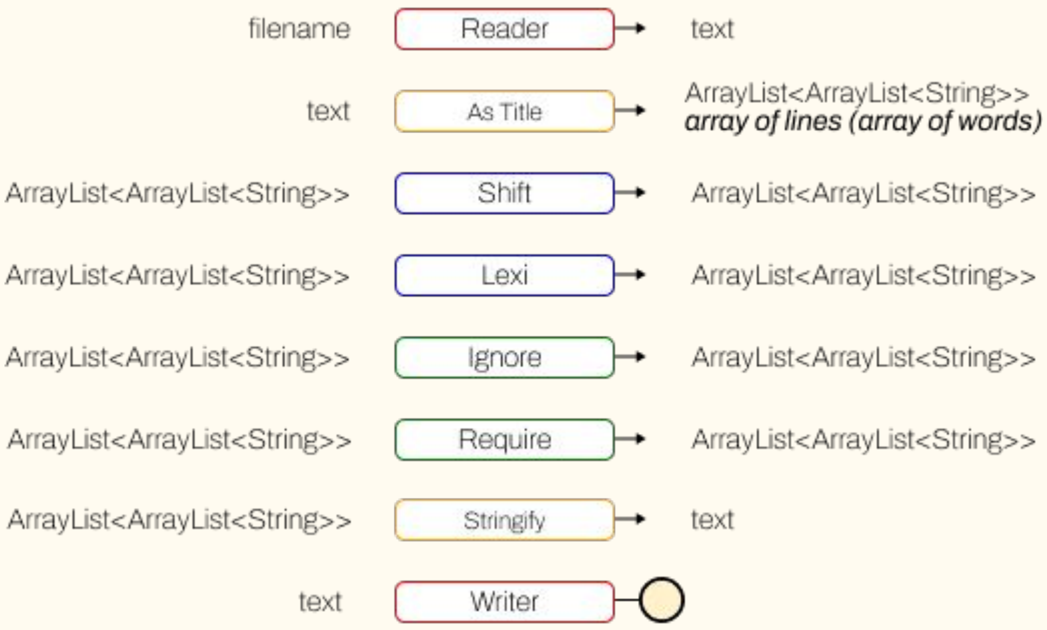
4) Sort …

5) … omitting … ignore keywords.

6) … selecting … require keywords.

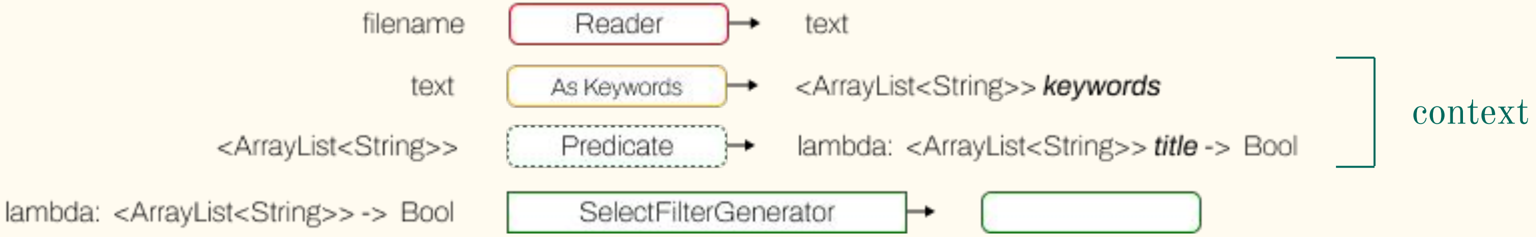
7) … as text.

8) Buffer…



filename	Reader	text
text	As Title	ArrayList<ArrayList<String>> *array of lines (array of words)*
ArrayList<ArrayList<String>>	Shift	ArrayList<ArrayList<String>>
ArrayList<ArrayList<String>>	Lexi	ArrayList<ArrayList<String>>
ArrayList<ArrayList<String>>	Ignore	ArrayList<ArrayList<String>>
ArrayList<ArrayList<String>>	Require	ArrayList<ArrayList<String>>
ArrayList<ArrayList<String>>	Stringify	text
text	Writer	

KWIC Extension - Filter Generators



filename → Reader → text

text → As Keywords → <ArrayList<String>> ***keywords***

<ArrayList<String>> → Predicate → lambda: <ArrayList<String>> ***title*** -> Bool

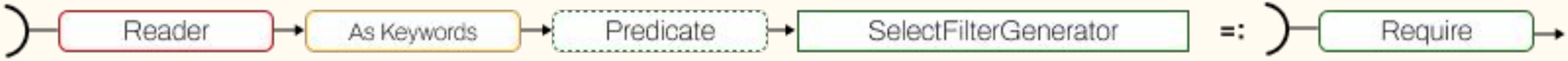
context

lambda: <ArrayList<String>> -> Bool → SelectFilterGenerator →

Pipeline Example

requireFilter

Titles selection from a predicate in the context of query keyword



Reader → As Keywords → Predicate → SelectFilterGenerator =: Require →

Assignment 3 - KWIC Search

1. **Data Loading**

a) Read filenames from ListOfFiles.in file.
b) Parse filenames as a list of Manual filenames.
c) Read titles from each Manual file.
d) Parse titles as an array of lines, with each line an array of words.
e) Concordance is populated from titles for each Manual (Circular shifting).
f) Store the concordance of the Manuals.

2. **Search Feature**

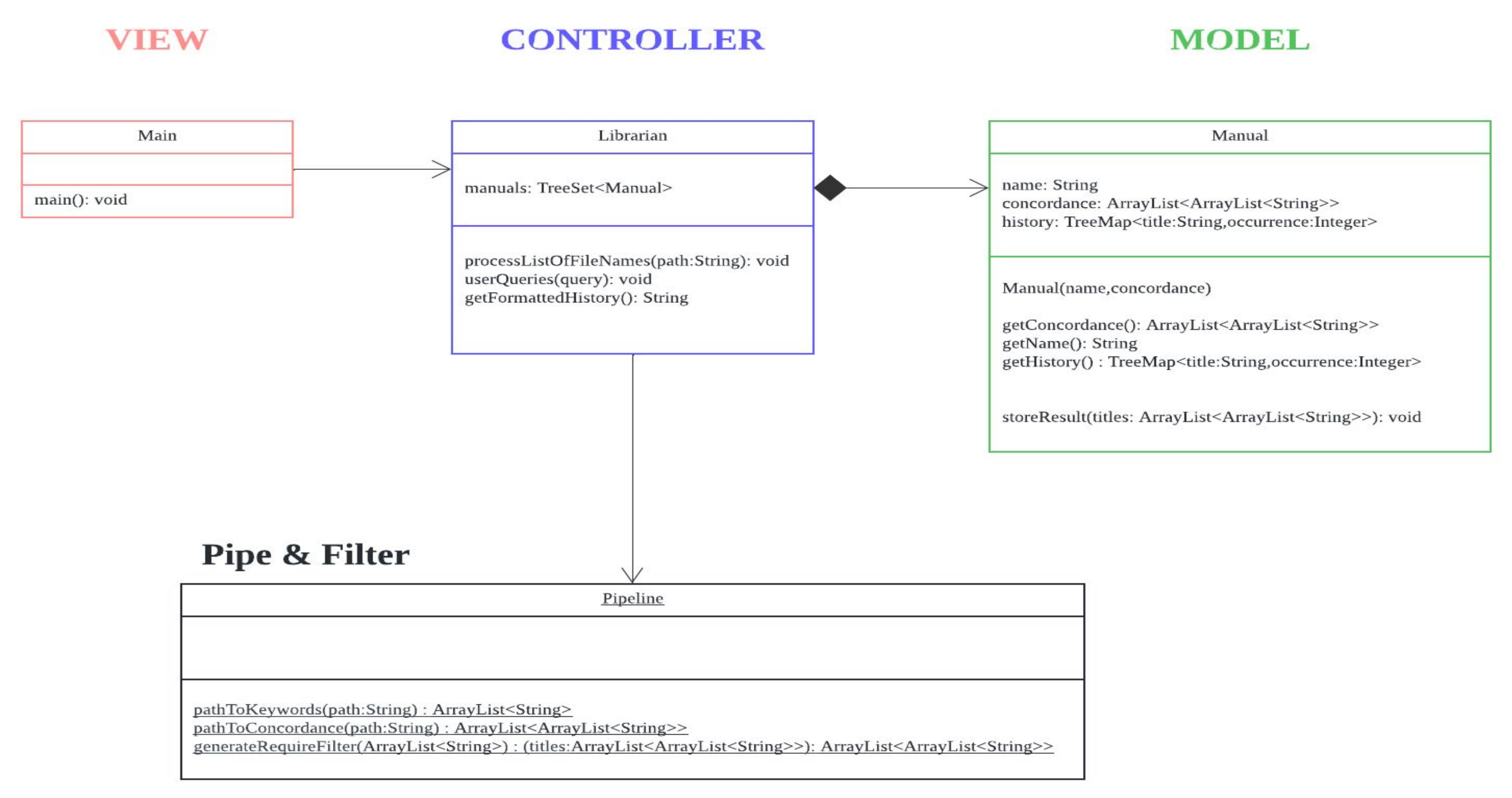
a) Accepts user query.
b) Execute query and store result.
c) Retrieve and format history as output.
d) Display output.

* Program loops until user inputs exit keyword.

KWIC Search - Architectural Design

MVC + Pipeline & Filter

From Design to Class



VIEW

CONTROLLER

MODEL

Main

main(): void

Librarian

manuals: TreeSet<Manual>

processListOfFileNames(path:String): void
userQueries(query): void
getFormattedHistory(): String

Manual

name: String
concordance: ArrayList<ArrayList<String>>
history: TreeMap<title:String,occurrence:Integer>

Manual(name,concordance)

getConcordance(): ArrayList<ArrayList<String>>
getName(): String
getHistory() : TreeMap<title:String,occurrence:Integer>

storeResult(titles: ArrayList<ArrayList<String>>): void

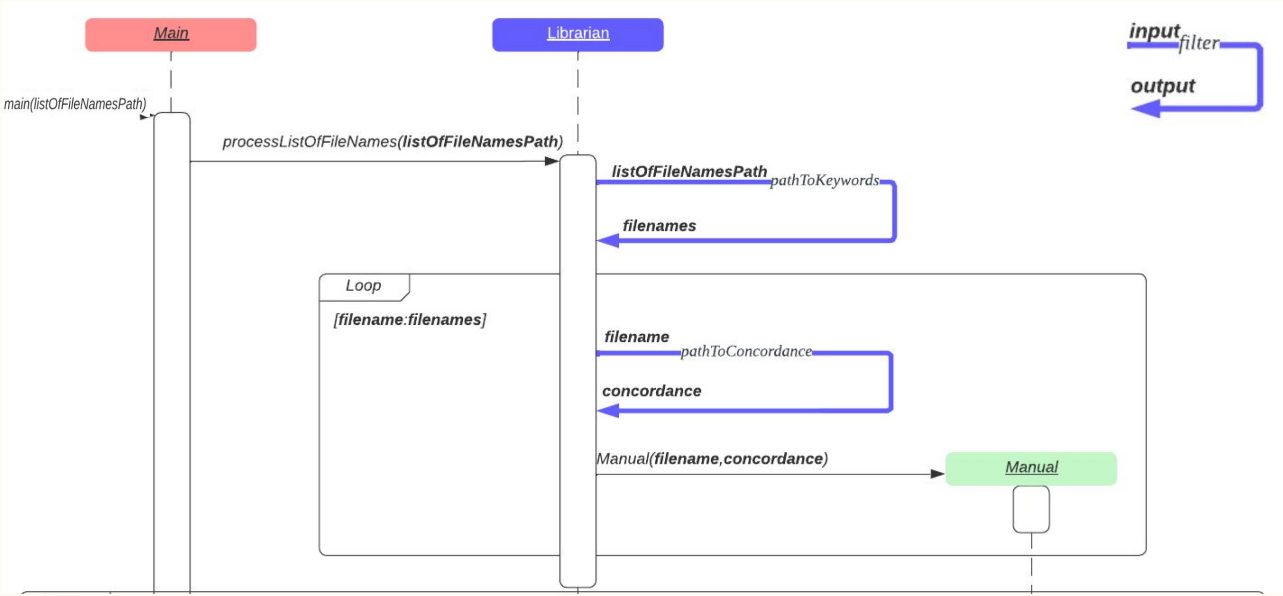
Pipe & Filter

Pipeline

pathToKeywords(path:String) : ArrayList<String>
pathToConcordance(path:String) : ArrayList<ArrayList<String>>
generateRequireFilter(ArrayList<String>) : (titles:ArrayList<ArrayList<String>>): ArrayList<ArrayList<String>>

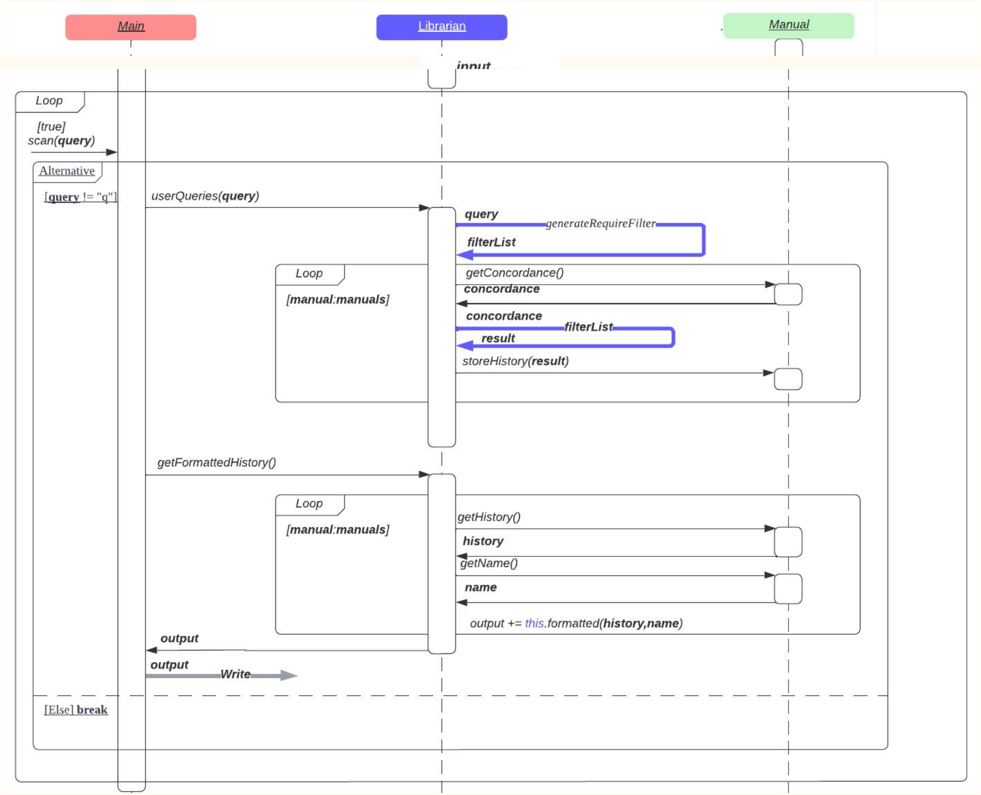
KWIC Search - Data Loading

Sequence Diagram

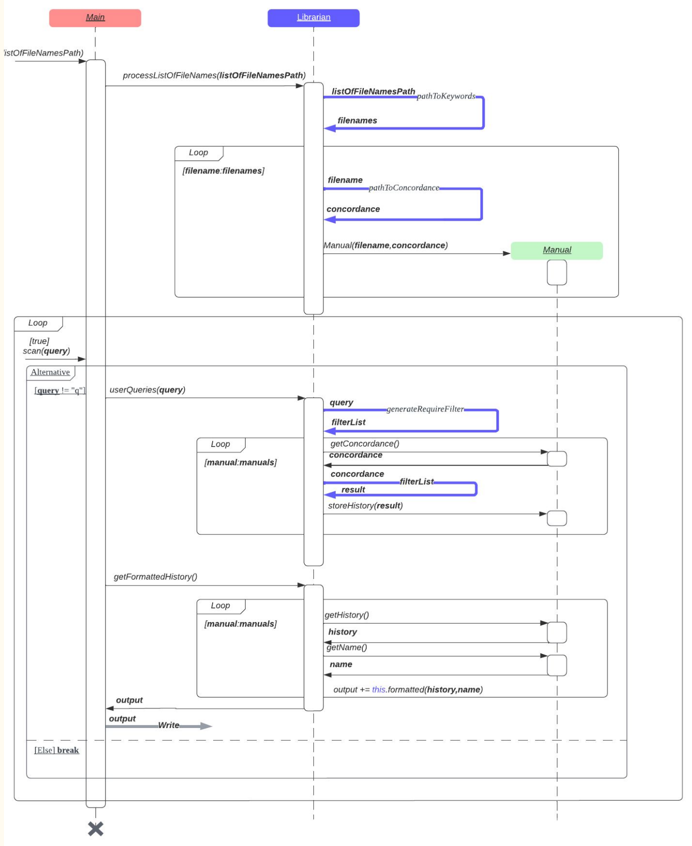


KWIC Search - Search Feature

Sequence Diagram



Conclusion - Functionality



Main (View) - Accepts input and display output

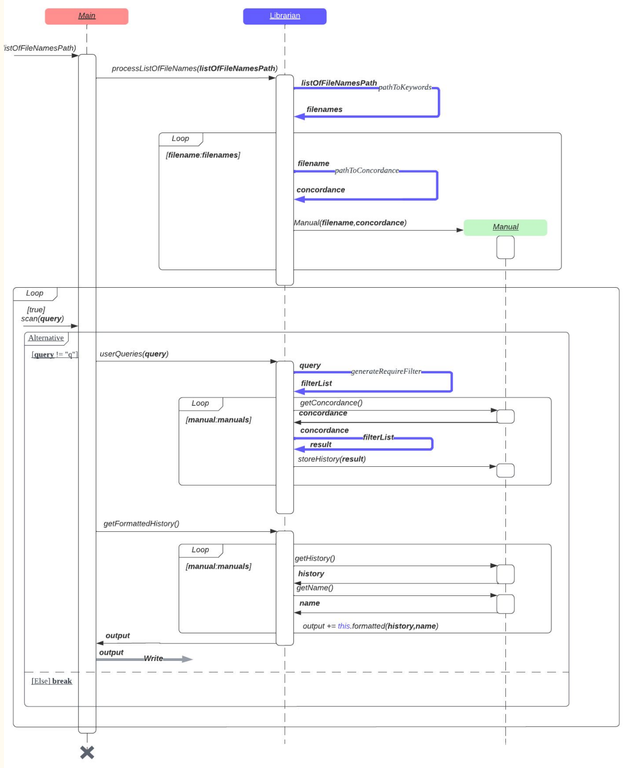
Librarian (Controller) - Data interaction and data (string) formatting.

Pipeline (Pipe & Filter) - Data transformation

Manual (Model) - Data storage

Conclusion - Principles

SOC + SRP



Main (View)

- Abstraction (from View and Controller)

Librarian (Controller)

- Cohesion of logical operations.

Pipeline

- Separation of Concern
- Reusability (Composition of filter components)
- Open-Closed Principle

Manual (Model)

- Information Hiding

Thank you