Project Overview

Project Title   : Identrix

Project Type   : GUI

Problem Domain   : Bioinformatics sequence similarity programs (e.g., BLAST, HMMER) compare biological sequences to identify matches, predict functions, and study evolution. They are used in research, medicine (diagnostics, drug discovery), agriculture (genetic improvement), and forensics for tasks like gene annotation, pathogen detection, and DNA profiling.

Applications   : Academic Research
- Evolutionary Biology
- Structural Biology
- Microbiology

Biotech & Pharmaceutical Companies
- Drug Discovery
- Vaccine Development
- Personalized Medicine

Clinical & Medical Diagnostics
- Genetic Testing
- Pathogen Detection

Agricultural Science
- Crop Improvement
- Livestock Genomics

Forensic Science
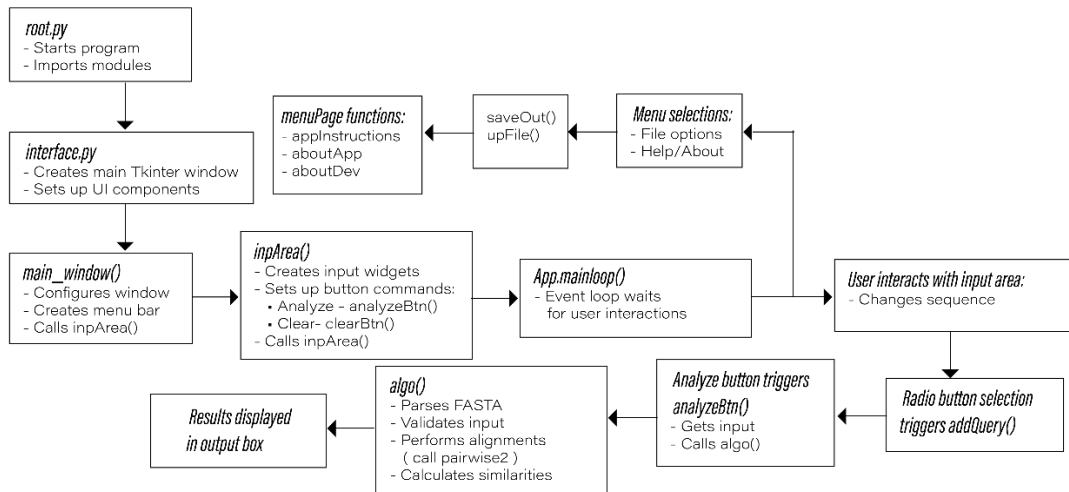- DNA Profiling

Technical Architecture

Programming Language   :   Python 3.13.3
Libraries/   Packages   :
- PIL (For image handling)
- tkinter (For the UI/ UX design)
- customTkinter (For better UI/ UX)

- Biopython (For sequence alignment)

Design Pattern   :   Structured Programming
                                     Modular Programming

Code Structure/
Program Flow   :

root.py
- Starts program
- Imports modules

interface.py
- Creates main Tkinter window
- Sets up UI components

main_window()
- Configures window
- Creates menu bar
- Calls inpArea()

inpArea()
- Creates input widgets
- Sets up button commands:
  • Analyze – analyzeBtn()
  • Clear– clearBtn()
- Calls inpArea()

menuPage functions:
- appInstructions
- aboutApp
- aboutDev

saveOut()
upFile()

Menu selections:
- File options
- Help/About

App.mainloop()
- Event loop waits
  for user interactions

User interacts with input area:
- Changes sequence

algo()
- Parses FASTA
- Validates input
- Performs alignments
  (call pairwise2)
- Calculates similarities

Analyze button triggers analyzeBtn()
- Gets input
- Calls algo()

Radio button selection triggers addQuery()

Results displayed in output box

I/O Specifications

User Input   : FASTA File / String
                                      (and optionally similar amino acids for protein sequences)

Output Format   : Sequence length (after alignment)
                                      Similar residues (in number & percentage)
                                      Identical residues (in number & percentage)
                                      Aligned sequences

Interaction Flow   : User Input sequences as file or string
                                      (and optionally similar amino acids for protein sequences)
➔ Inputted sequences aligned by pairwise2 in biopython module (Needleman-Wunsch algorithm)
➔ Aligned sequences compared in algo function in button_func.py
➔ If any error is encountered, error message shown
➔ Otherwise, the result is shown in a text-area