P.D.E.A's
Prof. Ramkrishna More Arts, Commerce & Science College
Akurdi, Pune-44
"Accredited with A+ grade (CGPA 3.46) by NAAC."

Synopsis

Group Details:

Roll No.	Name
71229	Saikh Sohail Chand
71226	Shinde Pratik Chandrakant
71228	Rasal Dipak Barku

1. Project Title:

Duplicate File Remover - An Efficient Storage Management Solution

2. Abstract

The **Duplicate File Remover** is a Java-based desktop application designed to help users **identify and remove duplicate files** from their system efficiently. It scans selected directories, compares file content using **MD5 checksum algorithms**, and provides an intuitive **GUI for preview and confirmation** before deletion. To ensure safety, it maintains a **history log in JSON format** for tracking deleted files. With **multi-threading support**, **system tray integration**, and **real-time progress visualization**, the tool offers an optimized user experience for effective storage management.

Introduction

Motivation

With the increasing accumulation of digital files, users often store multiple copies unknowingly, leading to wasted disk space. Studies show that up to **30% of storage can be occupied by duplicates**. Manually identifying and removing them is tedious and error-prone, making an **automated, user-friendly solution essential**.

Problem Statement

Finding and deleting duplicate files manually is inefficient, especially in large directories. Many existing tools rely only on filenames, which can be misleading. A robust solution is needed to accurately detect duplicates based on file content, provide a review mechanism before deletion, and maintain a history log to prevent accidental data loss.

Objectives & Goals

Goal

To develop a **fast, reliable, and user-friendly** tool that accurately detects and removes duplicate files while ensuring data integrity.

Objectives

- Design an **intuitive GUI** for easy file selection and operation.
- Implement **MD5 checksum-based comparison** for precise duplicate detection.
- Provide a **preview & confirmation system** to prevent accidental deletions.
- Maintain a **JSON-based history log** for tracking deleted files.
- Enable **system tray integration** for seamless desktop experience.
- Use **multi-threading** for fast and responsive performance.

Project Scope and Limitations

Scope

- Scans selected directories and their subdirectories.
- Identifies duplicate files based on **MD5 hash comparison**.
- Allows users to review and confirm before deletion.
- Stores deletion records in JSON format.
- Provides system tray integration for quick access.
- Displays **real-time progress indicators** during scanning and deletion.

Limitations

- Performance may slow down with **extremely large directories**.
- Not optimized for **files larger than 2GB**.
- Does not support **file recovery after deletion**.
- Requires **Java Runtime Environment (JRE)** to run.
- Scanning network drives may **affect performance**.

Functionality Modules

1. **File Scanning Module** – Traverses directories and calculates checksums.
2. **Duplicate Detection Module** – Compares file content using MD5 hashes.
3. **User Interface Module** – Provides an interactive GUI for easy usage.
4. **Confirmation System** – Allows users to review duplicates before deletion.
5. **File Deletion Module** – Removes duplicates safely upon confirmation.
6. **History Tracking Module** – Maintains a JSON log of deleted files.
7. **System Tray Integration** – Runs in the background for quick access.
8. **Progress Visualization** – Displays real-time scanning and deletion status.

Technology Stack

Frontend

- **Java Swing & AWT** – For GUI and system tray integration.
- **Custom UI Components** – For an enhanced user experience.
- **Event-Driven Programming** – For responsive interactions.
- **Multi-threading (SwingWorker)** – Ensures UI responsiveness.

Backend

- **Java I/O & NIO** – For efficient file system operations.
- **Java Security (MD5)** – For accurate file content comparison.
- **Java Collections Framework** – For optimized data handling.
- **JSON Processing** – For structured deletion records.
- **Exception Handling** – For robust error prevention.

Conclusion

The **Duplicate File Remover** is an efficient, user-friendly solution for managing storage clutter. By utilizing **checksum-based duplicate detection, real-time preview, and history tracking**, it ensures safe and effective file management. With **multi-threading and system tray integration**, it provides a **seamless and optimized experience**, making it an essential tool for **anyone looking to free up disk space effortlessly**.