भारतीय प्रौद्योगिकी संस्थान जम्मू

INDIAN INSTITUTE OF TECHNOLOGY JAMMU

विद्याधनं सर्वधन प्रधानम्

Student Research Conclave

- Problem Statement Title - : Redaction Tool “Empowering Privacy with Smart Redaction Solutions”
- Team Leader - Bhagwan Ji Jha
- College - GOVT. Engineering College Ajmer

Problem statement

“RE-DACT" is a simple and secure tool that helps **remove or hide sensitive information** from documents like text, images, PDFs, and videos. Users can choose how much information to hide, from just names to completely anonymized data. The tool **ensures data privacy by not storing or sharing user data**. Over time, it will also generate realistic fake data for safe sharing and learning. It’s easy to use, **works both online and offline, and keeps data secure.**

Proposed Solution

Develop a scalable and secure redaction tool to anonymize sensitive data across various formats. The tool will support **Hindi and English Language**, offer **automated and manual redaction options**, handle **multi-page PDFs with filter**, and allow **customizable redaction levels**. It ensures data security through HTTPS and input validations, making it ideal for privacy compliance and secure data sharing.

The inspiration behind RE-DACT

- **Affordable Tools:** Many advanced redaction tools are expensive, so we aim to make redaction easy and affordable for everyone.
- **Hindi Support:** Currently, no tools automatically redact Hindi text, so we want to fill that gap.
- **Simplified Manual Redaction:** We make redaction faster and easier by allowing users to choose how much redaction they need.
- **Data Security:** Unlike other tools, RE-DACT ensures your data is never saved and stays in your control.
- **Multiple Formats:** Many tools only handle basic formats, but RE-DACT works with text, images, PDFs, and videos.
- **Synthetic Data:** Unlike other tools, RE-DACT can generate realistic anonymized data that keeps sensitive information safe.

Process Flow

Step 1: Upload & Validate

- Upload a PDF, image, or video file.
- The system checks the file.
  - Invalid files are rejected.
  - Valid files move to the next step.

Step 2: Choose Redaction Settings

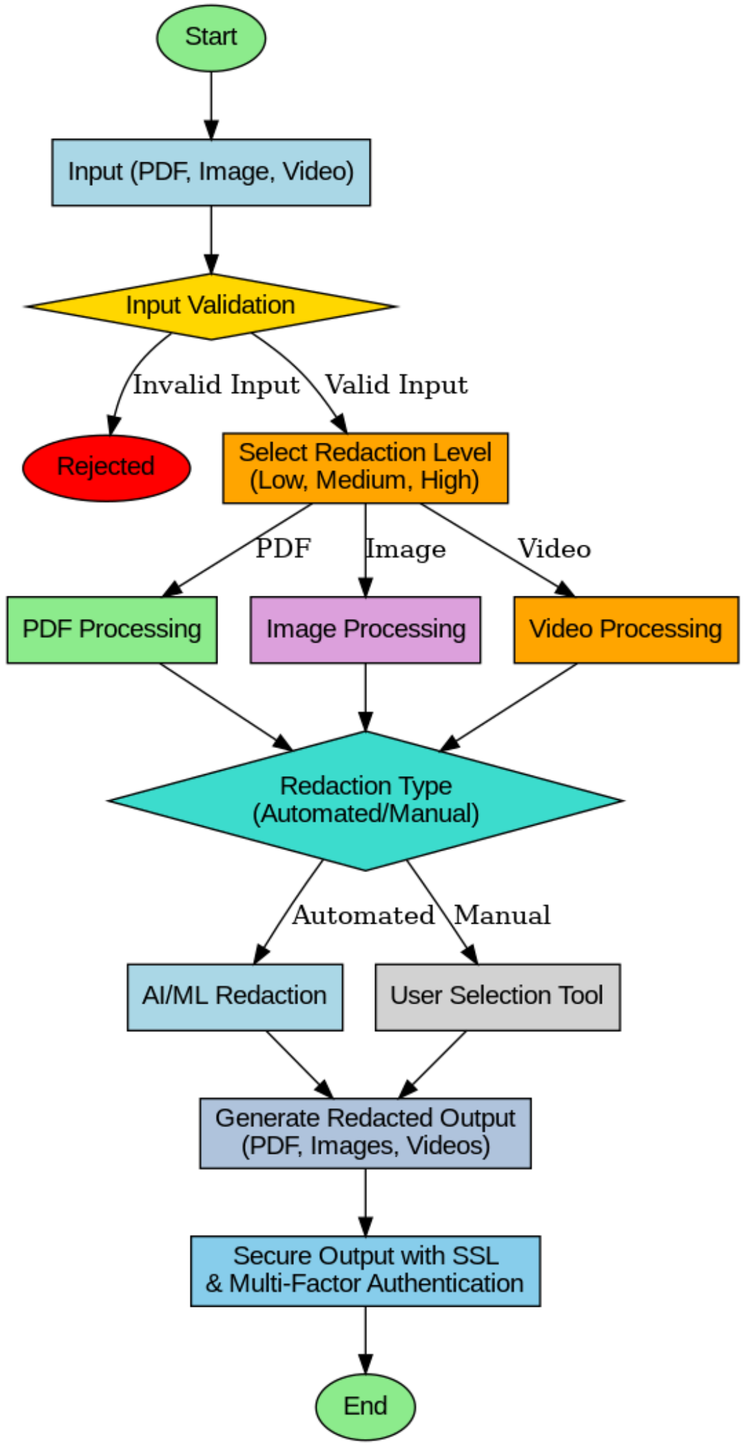
- Set a password for extra security before redaction.
- Choose a redaction level (Low, Medium, High).
- The system processes the file based on type (PDF, Image, or Video).

Step 3: Redact Data

- Select Automated Redaction (AI/ML) or Manual Redaction (User Selection).
- The chosen method removes sensitive information.

Step 4: Secure & Download

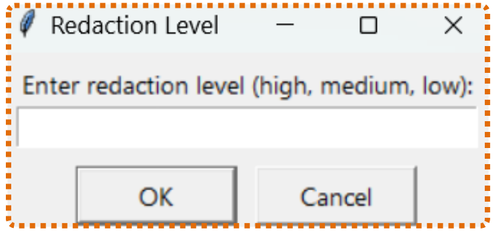
- The system generates the redacted file.
- The file is secured with SSL, Multi-Factor Authentication, and the user-set password.
- The user can download the protected redacted file safely.



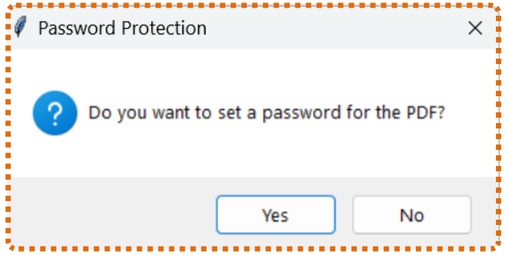
Technical Approach

PDF Redaction Tool

User Redaction Level



Password Protection



UI Development

- React.js: For building a user-friendly interface that supports multilingual input (Hindi and English).

Core Frameworks

- **TensorFlow/PyTorch:** For implementing AI/ML models for redaction and OCR.
- **Tesseract OCR:** For extracting text from images and PDFs.
- **PyPDF2 :** For PDF handling and processing multiple pages.
- **HTTPS and SSL Libraries:** HTTPS(for encrypting data transmitted between the client and server) to keep data transfer secure .
- **Django:** a more robust, all-in-one framework with built-in features like user authentication.
- **Natural Language Processing:** Spacy

Databases

- Non-Relational Database(MONGODB)/Relational Database(MYSQL)

Development and Deployment

- GIT (for version control & collaboration) , AWS/Microsoft Azure(Cloud services for hosting, scaling, and managing the application)

Potential Challenges and Risks

Strategic Solutions

Missing or Over-Redaction

The tool might fail to redact some sensitive data or remove too much.

Train the tool with diverse data and improve it over time based on user feedback.

Data Privacy

Users may worry about their sensitive data being stored or shared.

Process all data locally, don't save user data, and use encryption for added security.

Supporting Languages and Formats

Handling multiple languages like Hindi and supporting files like PDFs and videos can be challenging.

Start with popular formats and languages, then gradually add more based on demand.

Slow Processing for Large Files

Large files might make the tool slow or crash.

Use efficient algorithms and cloud services to handle bigger files smoothly.

Feasibility and Viability

TECHNICAL FEASIBILITY

- **Django** handles file uploads, backend logic, and integration smoothly.
- Python Libraries like **Tesseract-OCR, PyPDF2, and OpenCV** support text detection, file processing, and redaction.
- Multilingual Support is achievable using **Tesseract and APIs like Google Translate.**
- Deployment can be done easily on cloud platforms like **AWS or Heroku.**
- Performance Optimization tools in Django ensure smooth handling of large files

ECONOMIC FEASIBILITY

- **Development Costs:** Minimal as Django and Python libraries are open-source, reducing software expenses.
- **Infrastructure:** Affordable cloud hosting options like AWS, Heroku, or local servers can be used.
- **Team Expertise:** Python and Django are beginner-friendly, lowering training or hiring costs.
- **Revenue Potential:** Strong demand in legal, healthcare, and government sectors for data privacy tools ensures monetization through subscriptions or pay-per-use models.
- **Maintenance:** Regular updates and optimizations are cost-effective with an experienced team.

Operational Feasibility

- It uses free, open-source tools like **Django and Python libraries**.
- Hosting can be done on cost-effective platforms like AWS or Heroku.
- Python and Django are easy to learn, so no big spending on training or hiring.
- There's strong demand for privacy tools, making it easy to earn through subscriptions or pay-per-use plans.
- Maintaining and updating the tool will not be too expensive

REFRENCE AND RESEARCH

Redaction Tools for Data Privacy and Protection

- https://www.redactable.com/
- https://www.pdffiller.com/
- http://surl.li/tsidxs

Adobe Acrobat Premium

- https://www.youtube.com/watch?v=lSB5iXpEyPY

भारतीय प्रौद्योगिकी संस्थान जम्मू

INDIAN INSTITUTE OF TECHNOLOGY JAMMU

विद्याधनं सर्वधन प्रधानम

THANK YOU

Ask any Question(s) ?

PRESENTATION SRC IIT JAMMU - 2K25