Conceptual Framework

The conceptual framework of this study focuses on the assimilation of Optical Character Recognition (OCR), machine learning, and template matching algorithms. OCR is the mainly responsible for the text extraction from submitted documents. The Machine learning algorithms are responsible for recognizing patterns, classifying document elements, and improving the system’s accuracy over time. Template matching algorithms which involve comparing a document’s structure and content to predefined guidelines, is absolutely necessary to assess the compliance.

The way these technologies work together provides a structured approach to document compliance analysis. This framework is essential for automating the compliance assessment process and generating usable recommendations.

System components

1. Input Layer

Purpose: Accept reports in various formats for processing.

Components:

Document Upload Interface: Allows students to upload reports in formats like PDF, DOCX, or images.

File Preprocessor: Converts all input files into a standardized format (e.g., image or text) for further processing.

2. Template Repository

Purpose: Store and manage predefined report templates.

Components:

Template Storage: Stores various templates based on requirements (e.g., thesis report, research proposal, dissertation).

Template Specifications: Metadata for each template, including layout, field positions, text formatting, and section organization.

3. OCR and Content Extraction Module

Purpose: Extract text and layout data from uploaded reports.

Components:

* Optical Character Recognition (OCR): Converts scanned images or PDFs to editable text and layout information.
* Text Segmentation: Breaks the document into logical sections (title, abstract, chapters, references, etc.).
* Layout Detection: Identifies formatting elements such as margins, font sizes, headers, and footers.

4. Matching and Validation Layer

Purpose: Compare extracted data with the stored templates to identify discrepancies.

Components:

Template Matching Engine:

* Compares document content and layout with the predefined template.
* Checks for presence, position, and accuracy of key fields like student name, title, abstract, and bibliography.

Content Validation:

* Verifies adherence to content-specific guidelines (e.g., citation style, page limits, chapter order).
* Ensures required sections are included (e.g., "Acknowledgements," "Literature Review").

Compliance Checks:

* Validates formatting (font size, line spacing, margins).
* Ensures academic integrity by integrating with plagiarism detection tools.

5. Feedback and Output Module

Purpose: Provide actionable feedback to users.

Components:

* Error Reports: Highlight deviations from the template, such as missing sections or incorrect formatting.
* Suggestions for Correction: Provide specific instructions for making the report compliant.
* Validation Certificate: Generate a certificate of compliance if the report meets the template requirements.

6. User Management and Monitoring

Purpose: Manage access and track usage.

Components:

* Student Profiles: Store past submissions and compliance history.
* Administrator Dashboard: Allow faculty or administrators to monitor compliance trends and manage templates.