



**Progressive Education Society's  
Modern College of Engineering, Shivajinagar,  
Pune – 411 005**

**DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS**

**Course Name : Elective III (C) - Information Retrieval**

**Course Code : MCA10601C**

**Project Name**

Efficient Search in Large eBook Libraries

---

**Name : Sandesh Chaitram Bagde**

**Roll No. : 52104**

**Div : B**

---

**Name : Pavan Lahu Barde**

**Roll No. : 52105**

**Div : B**

## **Abstract :**

The **eBook Library** is a full-featured, web-based digital library management system developed using Python and the Flask framework. It enables users to upload, store, search, view, download, and delete PDF eBooks in a secure and organized manner. The system integrates **Whoosh**, a powerful full-text search engine, to index the content of uploaded PDFs, allowing users to perform keyword-based searches across the entire text of all documents. Key features include content-based search with highlighted snippets, file categorization, metadata extraction (page count, upload date), pagination, and responsive user interface powered by Bootstrap. The application ensures robust error handling, logging, and automatic re-indexing to maintain data integrity. Designed for local deployment, it runs on port 5001 and supports cross-platform usage, making it an ideal solution for personal or small-scale digital libraries.

---

## **Project Flow – Text Description (Sequential Steps)**

### **1. Application Startup**

- Run python app.py
- Validates templates/, static/style.css, books/, and indexdir/
- Creates or opens Whoosh index
- Indexes all PDFs in books/ if index is missing or outdated
- Starts Flask server on http://127.0.0.1:5001

### **2. Home Page**

- Lists all PDFs with title, upload date, and category
- Supports pagination (10 books/page)

### **3. Upload Workflow**

- User selects PDF and optional category
- File saved to books/

- Text extracted using PyPDF2
- If text found → indexed in Whoosh
- If no text (image-based PDF) → file deleted, warning shown

#### 4. Search Workflow

- User enters keyword
- Whoosh searches indexed content
- Returns results with highlighted snippets
- Supports sorting: Relevance, Title, Date
- Pagination: 5 results/page
- Auto re-indexes if no results found

#### 5. View Details

- Shows page count and excerpt from first page

#### 6. Download / Delete

- Download: Sends file as attachment
- Delete: Removes file + updates index

#### 7. Re-indexing

- Manual trigger via /index\_books
- Ensures search accuracy after bulk changes

---

## Key Features in Flow

- **Real-time Indexing:** Every upload updates the search index instantly
- **Fault Tolerance:** Handles missing files, permissions, and corrupted PDFs
- **User Feedback:** Flash messages for success, warning, and errors
- **Logging:** All actions logged to app.log for debugging