

PDF-Textausleser mit Filterfunktion in Java

Entwickeln Sie eine Java-Anwendung, die folgende Funktionalitäten bietet:

1. Textauslesen aus PDF-Dateien:

- Die Anwendung sollte in der Lage sein, Texte aus PDF-Dokumenten zu extrahieren.
- Verwenden Sie eine geeignete Bibliothek oder API, um den Text aus den PDF-Dateien zu lesen. Hier sind einige Optionen:
 - **iText**: Eine Open-Source-Bibliothek zum Erstellen, Bearbeiten und Lesen von PDF-Dateien in Java. Sie ermöglicht das Extrahieren von Text aus PDFs.
 - **GroupDocs.Viewer**: Diese Bibliothek kann ebenfalls Text aus PDF-Dateien extrahieren.
 - **PDFBox**: Eine weitere Open-Source-Bibliothek von Apache, die PDF-Dateien verarbeiten kann.

2. Filterfunktion für bestimmte Informationen:

- Implementieren Sie einen Filtermechanismus, um spezifische Informationen aus dem extrahierten Text zu filtern.
- Beispielsweise könnten Sie nach Schlüsselwörtern suchen oder reguläre Ausdrücke verwenden, um relevante Daten zu identifizieren.
- Definieren Sie, welche Informationen gefiltert werden sollen (z. B. Datum, Namen, Preise usw.).

3. Benutzerfreundliche Oberfläche:

- Erstellen Sie eine einfache Benutzeroberfläche (CLI oder GUI), um PDF-Dateien auszuwählen und die Filterung durchzuführen.
- Geben Sie die gefilterten Informationen übersichtlich aus.

Hinweise

- Bei Fragen oder Unsicherheiten wenden Sie sich an Ihren Betreuer oder Ihre Betreuerin.