

ZIELE

Was wir bereits umgesetzt haben:

1. PDF-Verarbeitung:
 - Wir haben eine Möglichkeit entwickelt, PDF-Dokumente zu lesen und den Text daraus zu extrahieren.
 - Der extrahierte Text wird bereinigt, indem wir Seitenzahlen und Sonderzeichen entfernen, damit der Text für die Analyse besser geeignet ist.
2. Textanalyse:
 - Wir extrahieren wichtige Begriffe (wie z.B. „Prüfungsordnung“ oder „Hausordnung“) aus dem Text und kennzeichnen diese als „Entitäten“.
 - Der Text wird in einzelne Sätze unterteilt, um diese später gezielt verwenden zu können.
3. Verarbeitung mehrerer Dokumente:
 - Wir haben bereits eine Funktion, die alle PDFs in einem Ordner durchsucht, den Text extrahiert und analysiert.

Unsere nächsten Schritte:

1. Verbesserung der Textverarbeitung

- Seitenzahlen speichern: Wir sollten die Seitenzahlen beim Extrahieren des Textes speichern, damit wir später genau angeben können, auf welcher Seite wir die Information gefunden haben.
- Kategorisierung des Textes: Wir könnten den Text noch besser strukturieren, indem wir ihn in verschiedene Themenbereiche unterteilen, z.B. „Prüfungsordnung“, „Hausordnung“ etc. Das hilft uns, schneller die relevanten Passagen zu finden.

2. Bessere Suche nach passenden Antworten

- Wenn jemand eine Frage stellt, müssen wir in der Lage sein, die besten Antworten aus dem Text zu finden. Wir könnten eine Technik einbauen, die prüft, welche Sätze aus dem Dokument am meisten mit der Frage übereinstimmen, etwa durch den Vergleich der Frage mit den Sätzen im Text.

3. Aufbau eines Frage-Antwort-Systems

- Wir sollten eine Funktion entwickeln, mit der Nutzer Fragen stellen können, und der Bot ihnen eine passende Antwort aus den Dokumenten gibt.
- Wichtig ist, dass wir nicht nur die richtige Antwort liefern, sondern auch die Quelle nennen – also welche Seite oder welcher Abschnitt des Dokuments die Information enthält.

4. Benutzeroberfläche erstellen

- *VII, eine einfache Oberfläche entwickeln, auf der Nutzer PDFs hochladen und Fragen stellen kann?*

5. Tests und Optimierung

- Wir sollten das System mit verschiedenen Dokumenten testen und überprüfen, ob der Bot die richtigen Informationen findet. Falls er Fehler macht, müssen wir den Prozess weiter verbessern und optimieren.