GEN AI MINI PROJECT

Neema Shrivastava

PES2UG23CS377

6F

Problem Statement:

Study Buddy (PDF Quizzer)

Goal: Paste a textbook chapter (Context) and ask "What is the definition of X?".

Tech: Extractive QA pipeline.

Abstract:

This project implements a PDF-based question-answering system designed to help students study textbook chapters more effectively. The system uses DistilBERT, a state-of-the-art natural language processing model, to perform extractive question answering on PDF documents. Users can upload any textbook chapter in PDF format and ask factual questions to receive instant answers extracted directly from the text. The implementation leverages the Hugging Face Transformers library with the distilbert-base-cased-distilled-squad model. To handle long documents that exceed the model's 512-token context window, the system employs a text chunking strategy that splits PDFs into manageable segments with overlapping chunks, then searches through all chunks to return the answer with the highest confidence score.

Documentation

I understood that the project required building an extractive question-answering system where answers are pulled directly from the source text rather than generated, using the distilbert-base-cased-distilled-squad model which is optimized for speed and efficiency. The main challenge was handling textbook chapters that exceed DistilBERT's token limit, so I implemented a chunking approach that splits documents into 2000-character segments with 200-character overlap to preserve context. I built a complete Jupyter notebook with an interactive interface featuring PDF file upload, dynamic question input via a text area, and automatic answer extraction. The system processes each uploaded PDF by extracting text using PyPDF2, splitting it into chunks, running the QA pipeline on each chunk, and returning the best answer.

Study Buddy - PDF Quizzer

Upload a PDF textbook chapter and ask questions to get instant answers.

Step 1: Upload Your PDF

Upload PDF (1)

Step 2: Ask Your Question

Question: Type your question here (e.g., "What is the definition of photosynthesis?")

Get Answer

PDF Loaded Successfully

File: example.pdf

Characters extracted: 764

Text chunks created: 1

Ready for questions. Type your question below and click "Get Answer".

Study Buddy - PDF Quizzer

Upload a PDF textbook chapter and ask questions to get instant answers.

Step 1: Upload Your PDF

Upload PDF (1)

Step 2: Ask Your Question

Question: what is entropy?

Get Answer

Answer Found

Question: what is entropy?

Answer: degree of disorder or randomness in a system

File C:/Users/shriv/Downloads/example.pdf

1 of 1

Entropy is a fundamental concept in thermodynamics that describes the degree of disorder or randomness in a system. In a physical system, entropy quantifies the number of microscopic configurations that correspond to a macroscopic state.

According to the second law of thermodynamics, the total entropy of an isolated system always increases over time. This means that natural processes tend to move toward a state of greater disorder. Entropy provides a direction for spontaneous processes and explains why certain reactions occur naturally while others do not.

In information theory, entropy is used as a measure of uncertainty or information content. A system with higher entropy contains more uncertainty, while a system with lower entropy is more predictable.