

# GenAi assignment 3

---



Samyan Ahmed 22i-2639

Hassan Ilyas 22i-2414

## Table of Contents

---

Chunking & Ingestion.....	0
Retrieval & Answering.....	1
Models & Tools.....	1
Screenshots.....	2
GenAi assignment 3.....	0

## FAST-NUCES FYP Handbook Assistant – Implementation Note

## Chunking & Ingestion

- Source: 3. FYP-Handbook-2023.pdf (FAST-NUCES BS Final Year Project Handbook 2023).
- Parser: PyMuPDF via PyMuPDFLoader, which preserves page numbers in metadata.
- Chunking: RecursiveCharacterTextSplitter with chunk\_size = 1000 characters ( $\approx$ 300 words) and chunk\_overlap = 300. Each chunk stores:
- page: zero-based loader page
- section\_hint: first heading/few meaningful words from the chunk
- chunk\_id: sequential label (chunk\_0, chunk\_1, ...)
- Indexing: Embeddings from sentence-transformers/all-MiniLM-L6-v2 (Hugging Face)  $\rightarrow$  FAISS local index (faiss\_index). Ingestion script: ingest.py.

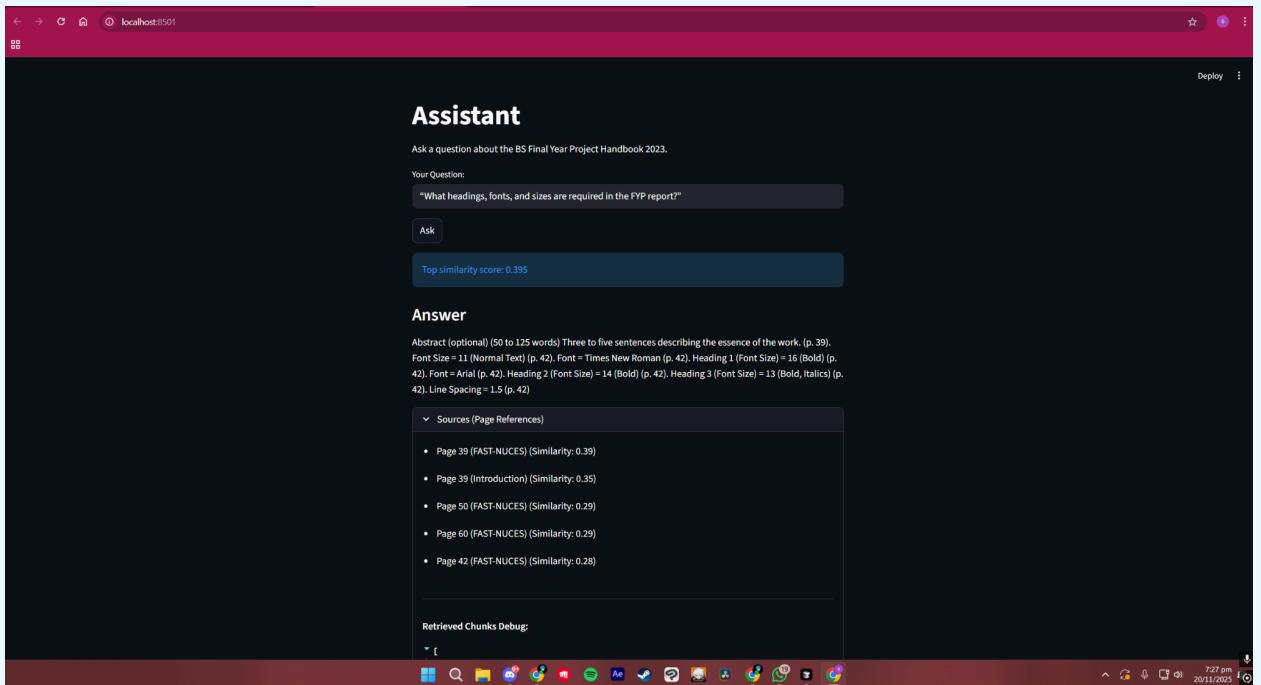
## Retrieval & Answering

- Query handling: User input is passed through expand\_query() to append domain-specific synonyms (chapters, margins, spacing, endnotes, etc.).
- Retrieval settings:
- k = 5 (top-5 cosine similarity hits from FAISS)
- Similarity threshold = 0.22. If no chunk exceeds the threshold, the app replies “I don’t have that information in the handbook.”
- Formatting-focused questions (margins/spacing/fonts) drop the threshold guard but display a warning when similarity is low.
- Grounding: The UI assembles the retrieved chunks (with page metadata) into an answer context, then a deterministic extraction layer builds the final response while citing (p. X) references.
- Precautions:
- Out-of-scope questions inherit the “no information” guard.
- Every answer includes page references and optional quotes pulled directly from retrieved chunks.
- Users can inspect all retrieved chunks in the “Sources (Page References)” expander (includes page numbers, section hints, similarity scores, and chunk previews).

## Models & Tools

- Embeddings: sentence-transformers/all-MiniLM-L6-v2
- Vector store: FAISS (local)
- UI: Streamlit (app.py) with a single textbox + “Ask” button + answer panel + collapsible sources list
- Prompt template embedded directly in app.py (no external LLM calls; responses come from local extraction logic)
- Prompt log: prompt\_log.txt

# Screenshots



**Assistant**

Ask a question about the BS Final Year Project Handbook 2023.

Your Question:

"What headings, fonts, and sizes are required in the FYP report?"

Ask

Top similarity score: 0.395

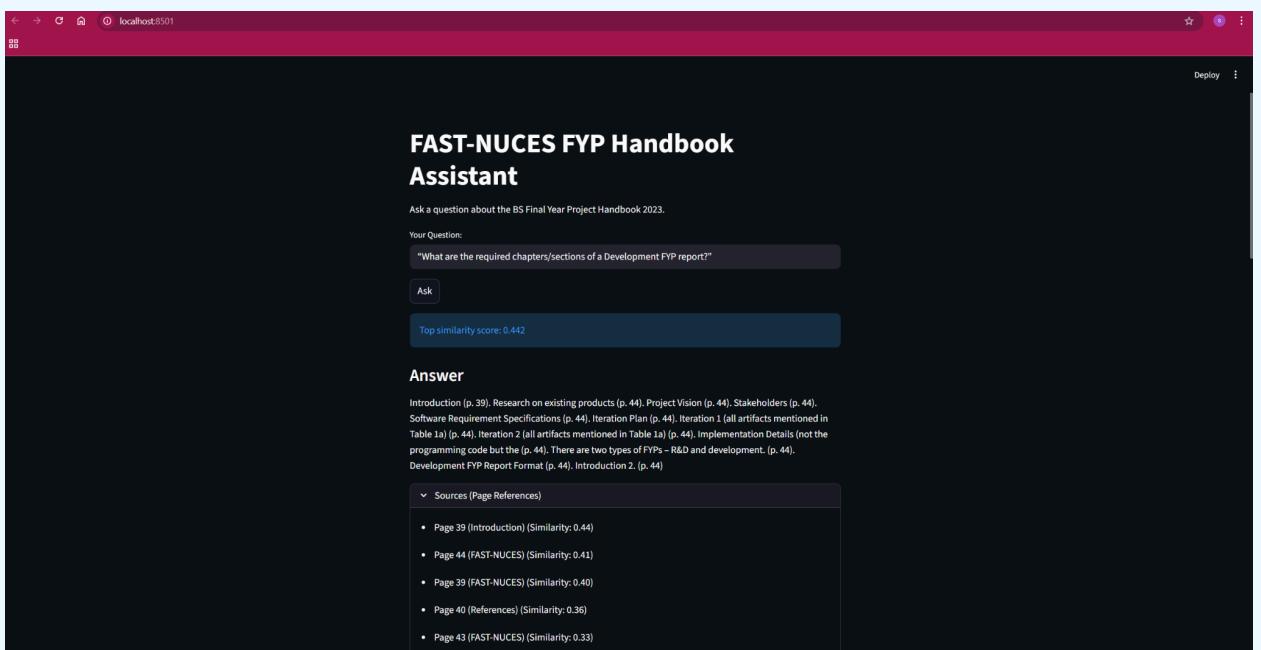
**Answer**

Abstract (optional) (50 to 125 words) Three to five sentences describing the essence of the work. (p. 39).  
 Font Size = 11 (Normal Text) (p. 42), Font = Times New Roman (p. 42), Heading 1 (Font Size) = 16 (Bold) (p. 42), Font = Arial (p. 42), Heading 2 (Font Size) = 14 (Bold) (p. 42), Heading 3 (Font Size) = 13 (Bold, Italic) (p. 42), Line Spacing = 1.5 (p. 42)

▼ Sources (Page References)

- Page 39 (FAST-NUCES) (Similarity: 0.39)
- Page 39 (Introduction) (Similarity: 0.35)
- Page 50 (FAST-NUCES) (Similarity: 0.29)
- Page 60 (FAST-NUCES) (Similarity: 0.29)
- Page 42 (FAST-NUCES) (Similarity: 0.28)

Retrieved Chunks Debug:



**FAST-NUCES FYP Handbook Assistant**

Ask a question about the BS Final Year Project Handbook 2023.

Your Question:

"What are the required chapters/sections of a Development FYP report?"

Ask

Top similarity score: 0.442

**Answer**

Introduction (p. 39), Research on existing products (p. 44), Project Vision (p. 44), Stakeholders (p. 44).  
 Software Requirement Specifications (p. 44), Iteration Plan (p. 44), Iteration 1 (all artifacts mentioned in Table 1a) (p. 44), Iteration 2 (all artifacts mentioned in Table 1a) (p. 44), Implementation Details (not the programming code but the (p. 44), There are two types of FYPs – R&D and development. (p. 44), Development FYP Report Format (p. 44), Introduction 2. (p. 44)

▼ Sources (Page References)

- Page 39 (Introduction) (Similarity: 0.44)
- Page 44 (FAST-NUCES) (Similarity: 0.41)
- Page 39 (FAST-NUCES) (Similarity: 0.40)
- Page 40 (References) (Similarity: 0.36)
- Page 43 (FAST-NUCES) (Similarity: 0.33)