Maurizio Scibilia
AI Engineer | Retrieval, RAG Evaluation & Evidence-first AI

ParancU Demo Script and Test Queries

Evidence-first review flow for version 0.1.0

1. Recommended setup

- Enter a user-supplied OpenAI API key before preparing a new corpus.
- Use 06_question_answering_text.txt for the recommended first test.
- Confirm that corpus preparation completes before opening Ask.
- Evaluate evidence containment and retrieval focus, not writing style.

2. Suggested demo flow

1. Prepare the sample text. During preparation, OpenAI is used only to generate a short summary and one guiding question per source chunk.
2. Ask a broad question and inspect the result.
3. Ask a specific question and compare focus.
4. Ask a query with a rare term or exact name.
5. Try a vague query and observe the confidence and result quality.
6. Optionally repeat the workflow using Scan/OCR.

3. General QA queries

- What is question answering?
- What is the difference between open-book and closed-book question answering?
- Why is retrieval useful for question answering?
- What were BASEBALL and LUNAR?
- How effective was LUNAR in its domain?

4. Evidence-containment queries

- Which passage explains how question-answering systems use unstructured documents?
- What evidence describes the domain limitations of early QA systems?
- Which passage explains why restricted-domain systems were effective?

5. Lexical/anchor sensitivity

- What does the document say about SHRDLU?
- Where are the Apollo Moon missions mentioned?
- What percentage of domain questions could LUNAR answer?
- Which passage mentions ELIZA and DOCTOR?

6. Vague-query checks

- Tell me about QA.
- What is this about?
- Explain the system.

Vague queries may produce broader evidence. They are useful for observing ambiguity, but should not be the main quality measure.

7. What to inspect

- Retrieved source passage.
- Guiding question and summary associated with the passage.
- Semantic, lexical/anchor, and combined retrieval scores.
- Whether the passage actually contains evidence that could support an answer.

8. Review notes template

- Query tested
- Retrieved evidence useful: Yes / Partial / No
- Evidence contained: Yes / Partial / No
- OCR/input issue observed: Yes / No
- Short note on what worked or failed

9. Local-first boundary

After preparation, corpus persistence, multilingual E5-small query embedding, semantic/lexical scoring, ranking, and evidence inspection run on-device. The OpenAI step is confined to summary and guiding-question generation during corpus preparation.