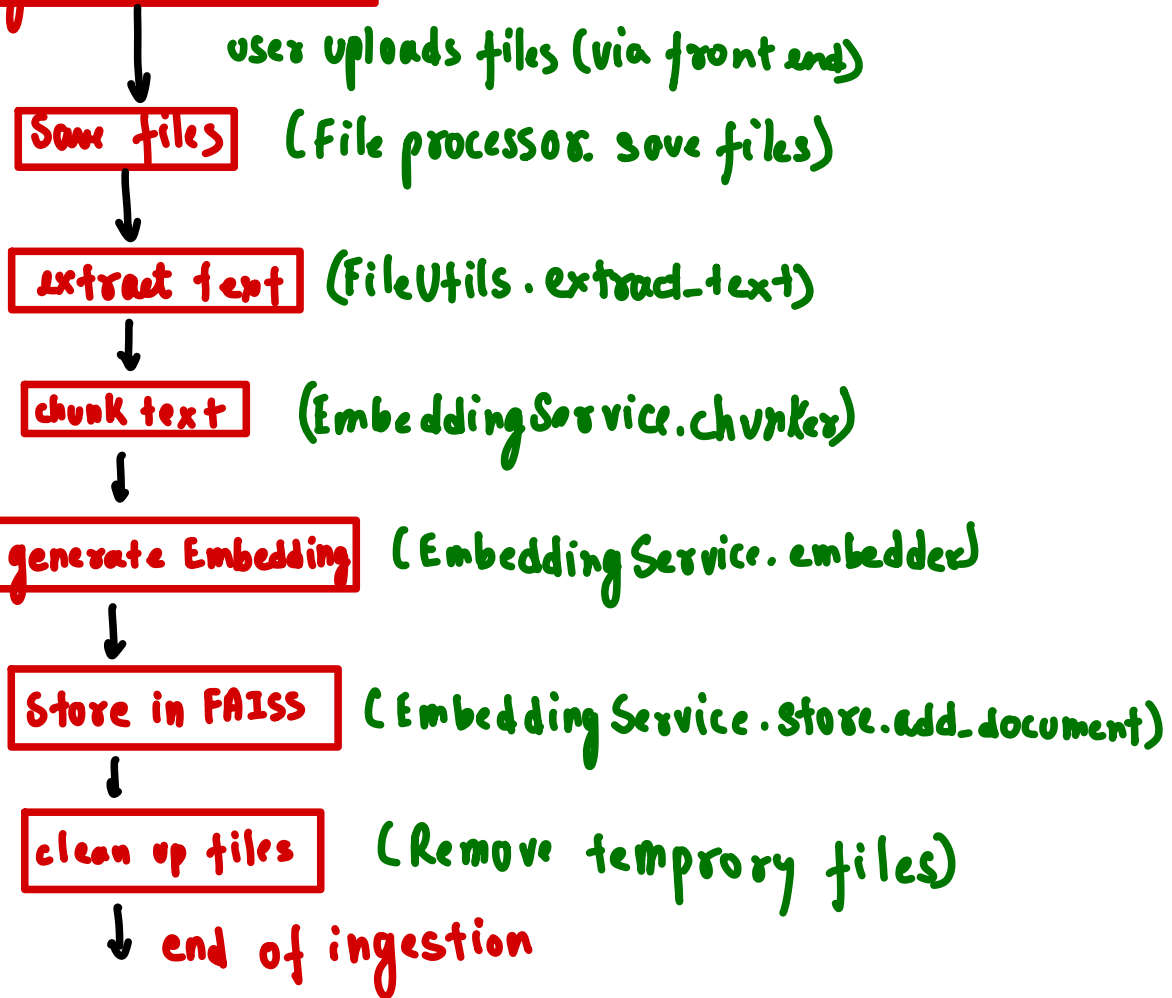
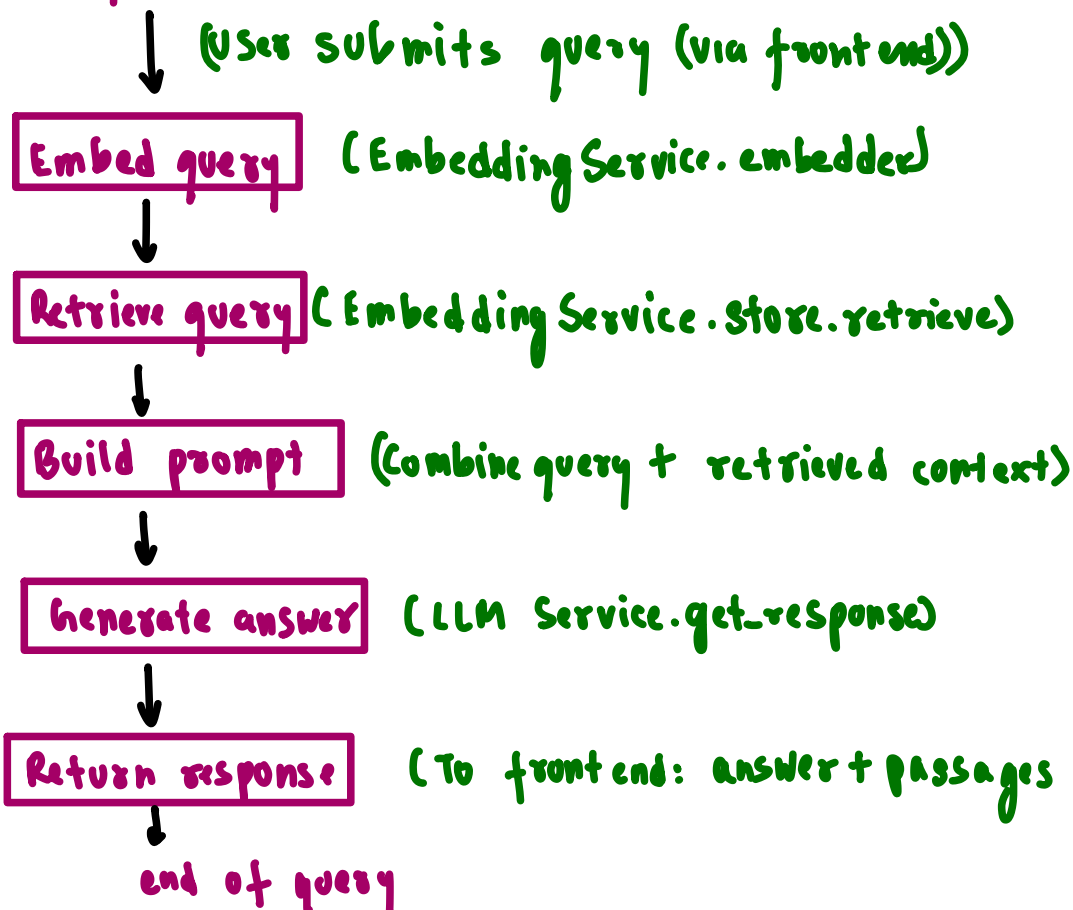


WORK FLOW

Ingestion workflow



Query workflow



Other important algorithm

Sliding window technique is implemented in `get_default_chunker()` in `embedding-service`

Say chunk size = 500
overlap is 100

chunk 1 → words 0 → 500

chunk 2 → words 400 → 900

⋮

Its importance is, if a sentence or idea falls b/w 2 chunks, it could get cut off.

The sliding overlap ensures the context flow smoothly, so the AI doesn't lose meaning when embedding and later retrieving related text.