Historia konwersacji i kolejne pytania

Rozwiązanie odniesienia i nowe wyszukiwanie dokumentów

Czas	Pracujesz z	Rezultat
około 45 min	Ollama lub checkpoint	conversation_tests.jsonl

Prześledzisz cztery tury rozmowy: pytanie o konfigurację, dwa pytania zależne i prośbę o źródła. Sprawdzisz także wyczyszczenie rozmowy oraz niejednoznaczne odniesienie.

Cel i rezultat rozdziału

Przypiszesz historię, pytanie samodzielne i źródła do właściwej tury, nie zastępując dokumentów historią.

Po zakończeniu masz

- cztery tury oraz dwa przypadki kontrolne;
- zapis pytań oryginalnych i samodzielnych;
- dowód zachowania źródeł oraz resetu rozmowy.

Zanim zaczniesz

Rozpakuj cały pakiet Student. Zachowaj razem notebooki, dane, kod, checkpointy i wyniki. Ścieżki w karcie są liczone od Student. Przygotowanie stanowiska opisuje instrukcja pracy z modułem 5.

Otwórz notebook

`notebooki/M12_M5_R3_historia_konwersacji_start.ipynb`

Materiał wejściowy	Do czego służy
dane/scenariusze_rozmowy.json	Wiadomości demonstracji DocAssist.
kod/m5_tools.py	Resolver, history_chat_step() i stan rozmowy.

Ollama lub checkpoint. Czas karty dotyczy ćwiczenia, bez instalacji i oglądania prezentacji. Materiał: M12-M5-v5, slajdy 87–138.

Praca krok po kroku

Krok 1 Przygotuj rozmowę — Wybierz tryb, uruchom Przygotowanie i odczytaj dokumentację resolvera.

- Wariant podstawowy rozpoznaje ograniczone wzorce DocAssist; nie jest ogólnym rozumieniem języka.
- Kod pakietu używa next_turn_id od 1; migrate_r3_state() mapuje wcześniejsze turn_id + 1.

Krok 2 Prześledź zapis historii — Otwórz get_recent_history() oraz finalize_turn().

- Historia przechowuje maksymalnie trzy ostatnie pary wiadomości.
- last_result oznacza ostatni RAG, a turn_log obejmuje wszystkie zakończone tury.

Krok 3 Wykonaj dialog — Uruchom Próby i odczytaj tury od 1 do 4.

- Dla pytań zależnych porównaj original_question, standalone_question i history_used.
- Każde pytanie merytoryczne ponownie uruchamia RAG; odczyt źródeł go nie uruchamia.

Krok 4 Sprawdź źródła i reset — Porównaj źródła tury 3 z source_origin tury 4 oraz wynik po /clear.

- Po wyczyszczeniu pojawia się nowe conversation_id, a numeracja zaczyna się od 1.
- Przypadek niejednoznaczny ma prosić o doprecyzowanie bez wywołania RAG.

Krok 5 Uzupełnij ocenę — Dla każdego result_id zapisz obserwację, dowód i wniosek.

- W dowodzie podaj konkretne pola stanu lub identyfikatory fragmentów.
- Opisz różnicę między użyciem historii a odczytem zapisanych źródeł.

Krok 6 Zachowaj wariant rozmowy — Zapisz ocenę, notebook i katalog próby.

- Opcjonalna rozmowa swobodna używa run_assistant(use_selector=False) i trybu Ollama.
- Jeżeli resolver nie obsługuje własnego sformułowania, opisz granicę zamiast zgadywać obiekt.

Ważne: ponowne Przygotowanie tworzy nowy katalog. Nowe wykonanie oceń osobno. Po zmianie komentarzy uruchom wyłącznie ocenę i zapis; zachowaj surowe wyniki.

Punkty kontrolne pracy

Po etapie	Masz sprawdzić	Gdy wynik nie pasuje
Po turze 2	Pytanie samodzielne wskazuje usługę.	Sprawdź ostatnie wiadomości i ograniczenie resolvera.
Po turze 4	Źródła pochodzą z właściwej wcześniejszej tury.	Porównaj source_origin i chunk_id.
Po /clear	Nowa rozmowa nie używa starego odniesienia.	Sprawdź conversation_id i resolution_status.

Nazewnictwo i zapis wyników

Wyniki zapisują się w wyniki/R3/<data_id>/. Odczytaj rzeczywisty katalog z komunikatu Wyniki. Główny rezultat: conversation_tests.jsonl.

Zapisz notebook przez Ctrl+S. Oceny zapisują się jako ocena_<id>.json; szablon ocena_do_uzupelnienia.json pozostaje bez zmian. Sposób oddania określa prowadzący.

Lista kontroli przed zakończeniem

- □ Rozróżniam resolution_status i rag.status.
- □ Wskazałem nowe wywołania RAG w dialogu.
- □ Odczyt źródeł nie zmienia last_result.
- □ Znam ograniczone wzorce resolvera.
- □ Sprawdziłem niejednoznaczność i /clear.
- □ Oceny i wyniki dotyczą tej samej próby.

Gdy pojawi się problem

Objaw	Najpierw sprawdź
Resolver prosi o doprecyzowanie	Sprawdź, czy obiekt jest jednoznaczny w trzech ostatnich turach.
Swobodne pytanie nie działa offline	Checkpoint obejmuje zapisane żądania; własny dialog wymaga Ollama.
Zmienione źródła po nowym pytaniu	To oczekiwane: każde pytanie wykonuje nowe wyszukiwanie.
Brak źródeł po resetowaniu	/clear usuwa stan poprzedniej rozmowy, lecz nie pliki wyników.

Zasada końcowa: zachowaj oryginalny wynik i własne uzasadnienie. Błędna odpowiedź lub decyzja modelu jest materiałem do analizy, nie powodem do poprawiania surowego zapisu.