Integracja lokalnego asystenta z RAG

Jeden przepływ od wiadomości do odpowiedzi i śladu

Czas	Pracujesz z	Rezultat
około 60 min	Ollama lub checkpoint	functional_tests.jsonl i assistant_trace.jsonl

Połączysz interfejs, historię, selektor, kontrolowane narzędzia i RAG w jednym prototypie. Pięcioturowa demonstracja oraz przypadki graniczne pokażą zachowanie stanu i rozdzielenie statusów.

Cel i rezultat rozdziału

Uruchomisz assistant_step(), prześledzisz wszystkie zakończenia i przygotujesz wyniki do późniejszej oceny w M6.

Po zakończeniu masz

- jedenaście wyników scenariusza funkcjonalnego;
- powiązane ślady tur i prób narzędzi;
- własny raport integracji i ocenę wyników.

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_R6_docassist_start.ipynb
```

Materiał wejściowy	Do czego służy
dane/scenariusze_rozmowy.json	Pięć tur: procedura, stan, awaria, źródła, metadane.
kod/m5_tools.py	assistant_step(), begin_turn(), finalize_turn() i komendy.

Ollama lub checkpoint. Czas karty dotyczy ćwiczenia, bez instalacji i oglądania prezentacji. Materiał: M12-M5-v5, slajdy 242–330.

Praca krok po kroku

Krok 1 Przygotuj integrację — Wybierz tryb, uruchom Przygotowanie i przeczytaj kontrakt z R1.

- Zestaw wymaga funkcji M3/M4 dołączonych w kod; nie przenoś samych notebooków.
- Pełne zapisy diagnostyczne dotyczą wyłącznie fikcyjnych danych ćwiczenia.

Krok 2 Prześledź koordynator — Otwórz assistant_step() i wskaż etapy walidacji, resolvera, selektora i wykonawcy.

- Znajdź normalize_action_proposal() między surową propozycją a decyzją.
- Wspólna finalize_turn() zapisuje także odmowę, błąd i doprecyzowanie.

Krok 3 Uruchom demonstrację — Wykonaj Próby i odczytaj pierwszych pięć tur.

- Każda tura ma conversation_id, turn_id i unikatowe result_id.
- Porównaj źródła trzeciej i czwartej tury; metadane nie nadpisują last_result.

Krok 4 Sprawdź granice — Odczytaj oba TLS, żądanie restartu, puste wejście i dwa przypadki po /clear.

- Dla TLS sprawdź rzeczywisty status i reason; nie wymuszaj oceny na podstawie słowa TLS.
- Po /clear wcześniejsze odniesienie i źródła są niedostępne, ale pliki wyników pozostają.

Krok 5 Oceń i opisz integrację — Uzupełnij oceny wyników i trzy pola raportu w notebooku.

- W polach granice, dowody i przekazanie_do_M6 odwołaj się do konkretnych result_id.
- Oddziel poprawność techniczną, decyzję selektora i poparcie odpowiedzi dokumentami.

Krok 6 Sprawdź pakiet do przekazania — Zapisz raport, oceny i notebook, a następnie otwórz notebook 07.

- Wskaż katalog R6 w zmiennej PROBA; kontroler sprawdzi powiązania i kompletność.
- Test automatyczny nie zastępuje oceny treści. Zachowaj wyniki do audytu w M6.

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

Punkty kontrolne pracy

Po etapie	Masz sprawdzić	Gdy wynik nie pasuje
Po dialogu	Źródła i metadane mają właściwe pochodzenie.	Sprawdź rag, source_origin i last_result.
Po granicach	Każde zakończenie ma spójny wynik i ślad.	Połącz pliki po result_id oraz conversation_id.
Po kontroli	Raport i oceny wskazują tę samą próbę.	Uzupełnij znaczniki i uruchom tylko komórki zapisu.

Nazewnictwo i zapis wyników

Wyniki zapisują się w wyniki/R6/<data_id>/. Odczytaj rzeczywisty katalog z komunikatu Wyniki. Główny rezultat: functional_tests.jsonl i assistant_trace.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. Zachowaj także raport_integracji_<id>.md i surowe_wyniki.jsonl.

Lista kontroli przed zakończeniem

- □ Prześledziłem pięć tur oraz przypadki graniczne.
- □ Zachowałem statusy decyzji, narzędzia, RAG i tury.
- □ Po /clear powstała nowa rozmowa.
- □ Pełne wyniki są oddzielone od skróconego śladu.
- □ Raport wskazuje dowody i ograniczenia prototypu.
- □ Kontrola notebooka 07 potwierdza kompletność zapisu.

Gdy pojawi się problem

Objaw	Najpierw sprawdź
Błąd selektora przed narzędziem	Sprawdź surową propozycję oraz parse_error; narzędzie nie powinno zostać wywołane.
Niezgodne źródła po nowym RAG	Sprawdź aktualizację last_sources także dla pustej listy.
Komenda zwiększa numer tury	Komendy obsłuż przed assistant_step(); naturalna prośba jest osobną turą.
Kontroler nie znajduje oceny	Uzupełnij wszystkie pola, zapisz ocenę i wskaż dokładny katalog PROBA.

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