MODUŁ 6 • ROZDZIAŁ 3

Ocena wyszukiwania

Karta pracy uczestnika

Czas pracy	Materiały	Rezultat
Około 60 min	Notebook R3 i karta	Pliki opisane w karcie

Najpierw sprawdź, czy system znalazł właściwe fragmenty i przekazał je do kontekstu. W tym ćwiczeniu nie generujesz odpowiedzi modelu językowego.

Cel i rezultat rozdziału

Zapiszesz pełny ranking, wynik filtrowania i kontekst. Oddzielisz obecność dowolnej grupy źródeł od pokrycia wszystkich wymaganych grup.

Po zakończeniu masz

- Surowe wyniki wyszukiwania powiązane z testem i konfiguracją.
- Ręczne oceny kontekstu z uzasadnieniami.
- Wyniki połączone z recenzjami przez result_id.

Zanim zaczniesz

Zakończ R2. W trybie checkpoint użyj wyłącznie listy checkpoint_subset.json. Własne pytania i pełny zestaw wymagają nowych embeddingów w trybie Ollama.

Otwórz notebook

```
notebooki/M12_M6_R3_ocena_retrieval_start.ipynb
```

Plik lub materiał	Do czego służy
evaluation/test_cases.jsonl	Jawne retrieval_question i oczekiwane źródła.
dane/przypadki/checkpoint_subset.json	Sprawdzony podzbiór wyszukiwania offline.

Tryb: checkpoint lub Ollama. Slajdy M12-M6-v6: 148–213. Czas orientacyjny: 60 min.

Praca krok po kroku

1. Wybierz zakres

- Uruchom podzbiór checkpointów lub pełny zestaw pytań wymagających wyszukiwania.
- Zapisz wybrane test_id. Testów narzędzi bez RAG nie oceniaj jako nietrafnego wyszukiwania.

2. Wykonaj wyszukiwanie

- prepare_search_question pobiera jawne pytanie, a evaluate_retrieval_turn uruchamia wyszukiwanie.
- Zapis jest natychmiastowy: nowy run_id i result_id dla każdego wyniku.

3. Sprawdź ranking

- Porównaj pełny ranking top_k z wymaganymi grupami źródeł.
- Trafienie jednej grupy nie oznacza pokrycia całej odpowiedzi wielоźródłowej.

4. Sprawdź kontekst

- Porównaj ranking, próg i fragmenty faktycznie przekazane do kontekstu.
- W T106 potwierdź warunek progu. Brak oczekiwanych źródeł oznacza metryki N/D, nie zero.

5. Oceń kontekst

- Wypełnij sześć kryteriów 0–2 lub N/D zgodnie z kontraktem oraz uzasadnij każdą ocenę.
- Oceń osobno błąd krytyczny. Zapisuj tylko ukończone recenzje.

6. Połącz i sprawdź

- Połącz wyniki z recenzjami przez result_id i zgodny kontrakt.
- Zachowaj pełny ranking. Sam wykaz źródeł odpowiedzi nie pozwala go odtworzyć.

R3 mierzy izolowane wyszukiwanie. Nie jest dowodem rankingu użytego w późniejszej generacji odpowiedzi.

Punkty kontrolne pracy

Kontrola	Pytanie	Dowód
Ranking	Czy zapisano wszystkie top_k?	chunk_id, score, tekst i metadane.
Kontekst	Co usunął próg i filtr?	Osobny zapis ranking oraz context.
Recenzja	Czy ocena dotyczy tego wyniku?	Zgodne result_id i contract_hash.

Nazewnictwo i zapis wyników

evaluation/retrieval_runs.jsonl; retrieval_reviews.jsonl; retrieval_results.jsonl (w evaluation)

W R4 ranking tego samego wywołania będzie połączony z odpowiedzią przez retrieval_result_id.

Lista kontroli przed zakończeniem

□ Nie uruchamiałem generacji w R3.
□ Zachowałem pełny ranking.
□ Rozróżniam any_group i all_groups.
□ Uzasadniłem oceny i N/D.
□ Nie zamieniłem błędu wykonania na zero.
□ Połączyłem recenzje z właściwymi wynikami.

Gdy pojawi się problem

Problem	Co zrobić
Brak checkpointu pytania	Użyj sprawdzonej listy lub trybu Ollama.
Oczekiwany fragment poza top_k	Zapisz błąd retrieval; zmianę zbadasz w R5.
Pusty kontekst	Sprawdź próg i powód odrzucenia fragmentów.
Nie da się zapisać recenzji	Uzupełnij oceny, uzasadnienia i autora.

Zachowaj wyniki i uzasadnienia. Oznacz pochodzenie danych oraz próby niewykonane.