Od listy embeddingów do lokalnego indeksu

Spójny zapis, walidacja i odczyt po restarcie

Czas orientacyjny	Pracujesz z	Rezultat
45 min	NumPy, metadanymi i trzema plikami indeksu	folder semantic_index gotowy do przekazania do M4

Zapiszesz wektory oddzielnie od treści, zachowując ich wzajemne powiązanie. Potwierdzisz, że wyszukiwanie działa po odczycie indeksu w świeżej sesji.

Cel i rezultat rozdziału

Zbudujesz i zwalidujesz lokalny indeks z jednoznaczną relacją między wierszem macierzy a rekordem fragmentu.

Po zakończeniu masz

- embeddings.npy, chunks.jsonl i index_metadata.json;
- jawnie wywołane validate_manifest() i validate_index();
- porównanie wyników przed zapisem i po restarcie.

Zanim zaczniesz

Najpierw wykonaj notebook 00. Otwórz wskazany starter, wybierz MODE i uruchom Przygotowanie. Pola UZUPEŁNIJ wypełnij samodzielnie. SOURCE wybiera własne R2 albo checkpoint; nie ma automatycznej zamiany wejścia.

```
notebooki/M12_M3_R5_zapis_indeksu_start.ipynb
```

Materiał	Do czego służy
chunks_with_embeddings.jsonl z R2	Wektory, tekst i pełne metadane.
embedding_manifest.json z R2	Pochodzenie i konfiguracja modelu.
Pytanie kontrolne z R4	Porównanie wyszukiwania przed i po zapisie.

Gdzie wrócić w prezentacji

W R5 odszukaj „Ćwiczenie krok po kroku: zbuduj i zapisz lokalny indeks” oraz „Ponowne wczytanie indeksu według manifestu”. Format wyników sprawdź w R4 na slajdzie „Budowanie listy wyników”.

Praca krok po kroku

Krok 1 Zbuduj macierz — Uruchom build_index() i sprawdź macierz float32 o kształcie N × 1024.

- Odrzuć wektory zerowe i wartości nieskończone lub NaN.
- Przeczytaj normalize_vector(); sprawdź normy wierszy z tolerancją numeryczną.

Krok 2 Zbuduj rekordy treści — Utwórz chunk_records bez pola embedding, w tej samej kolejności co macierz.

- Zachowaj chunk_id, text oraz całe metadata.
- Dodaj vector_index od 0 do N−1. Nie zastępuj nim trwałego chunk_id.

Krok 3 Zapisz komplet — W jednym folderze semantic_index zapisz macierz, rekordy i index_metadata.json.

- Użyj nazw embeddings.npy, chunks.jsonl i index_metadata.json.
- Manifest indeksu ma inny cel niż embedding_manifest.json z R2; nie zamieniaj plików.

Krok 4 Wywołaj walidację — Uruchom validate_manifest() i validate_index() na zapisanym komplecie.

- Sprawdź wymagane pola, model, wymiar, liczność, typ float32, skończoność i normy.
- Sprawdź unikatowe chunk_id, ciągłość vector_index oraz wymagane teksty i metadane.

Krok 5 Sprawdź świeżą sesję — Uruchom w notebooku kontrolę indeksu w niezależnym procesie Python.

- Proces czyta pliki z dysku oraz zapisany wektor pytania; nie korzysta ze zmiennych notebooka.
- Dodatkowo otwórz notebook 07 w nowym kernelu i wskaż tę samą próbę R5.

Krok 6 Porównaj wyniki — Porównaj identyfikatory i score przed zapisem oraz po odczycie.

- Wynik {score, chunk} przekształć przez normalize_search_results() do formatu R4.
- Drobne różnice float32 porównuj z tolerancją. Opisz remisy i niezgodności.

Zapis próby: komórka Przygotowanie automatycznie tworzy nowy folder wyniki/R5/<proba_id>/. Zachowaj wyświetloną ścieżkę.

Punkty kontrolne pracy

Kiedy	Co sprawdzasz	Co zrobić przy problemie
Mapowanie	Wiersz i vector_index odnoszą się do tego samego chunk_id.	Porównaj rekordy z pierwotną kolejnością R2.
Walidacja	Funkcje zostały wywołane i sprawdzają dane z dysku.	Sama definicja walidatora nie jest kontrolą.
Restart	Wyniki odtwarzają się w świeżej sesji.	Odtwórz brakujące definicje; sprawdź zgodność modelu.

Nazewnictwo i zapis wyników

Notebook zapisuje semantic_index/ i kontrola_indeksu.md w wyniki/R5/<proba_id>/. Do M4 przekazujesz cały folder semantic_index, a nie sam embeddings.npy.

Identyfikator <proba_id> nadaje notebook. Pliki są chronione przed nadpisaniem; nowy pomiar rozpocznij od Przygotowania. Zapisz notebook przez Ctrl+S i wskaż próbę do oceny.

Lista kontroli przed zakończeniem

- □ Zapisałem wszystkie trzy pliki w jednym folderze.
- □ Macierz ma typ float32 i jednostkowe normy.
- □ Zachowałem chunk_id oraz pełne metadane.
- □ Jawnie wywołałem oba walidatory.
- □ Sprawdziłem mapowanie z wynikiem R2.
- □ Sprawdziłem odczyt w niezależnej sesji.

Gdy pojawi się problem

Objaw	Co sprawdzić
NameError po restarcie	Odtwórz importy, stałe, funkcje modelu, normalizacji, walidacji i wyszukiwania.
Liczności zgodne, teksty błędne	Sprawdź kolejność względem R2; zgodne wymiary nie dowodzą poprawnego powiązania.
Niezgodny model lub wymiar	Wczytaj zgodny komplet. Nie łącz plików z różnych prób.

Walidacja struktury nie dowodzi zgodności znaczenia tekstu z wektorem. Potrzebne są także kontrola mapowania i porównanie wyników. Zapisz także uzupełniony notebook.