KARTA PRACY UCZESTNIKA MODUŁ 3 ROZDZIAŁ 1

Od tekstu do reprezentacji wektorowej

Ocena znaczenia fragmentów przed uruchomieniem modelu

Czas orientacyjny	Pracujesz z	Rezultat
20 min	pytaniem, czterema fragmentami i własnym uzasadnieniem	hipotezy_podobienstwa.json

Porównasz dopasowanie słów z podobieństwem znaczenia. Najpierw sformułujesz własną ocenę, aby później móc odróżnić oczekiwania od wyników modelu.

Cel i rezultat rozdziału

Wskażesz fragment odpowiadający na pytanie oraz odróżnisz parafrazę od tekstu tylko podobnego tematycznie.

Po zakończeniu masz

- ranking czterech fragmentów z własnym uzasadnieniem;
- przykłady wspólnych słów i parafraz;
- zapis hipotezy oraz opis ograniczeń embeddingów.

Zanim zaczniesz

Otwórz wskazany notebook i wykonuj komórki kolejno. R1 nie wymaga modelu. Pola UZUPEŁNIJ wymagają własnych ocen przed zapisem.

```
notebooki/M12_M3_R1_intuicja_embeddingow_start.ipynb
```

Materiał	Do czego służy
dane/R1_fragmenty.md	Pytanie i fragmenty A–D bez klucza odpowiedzi.
Prezentacja R1	Pojęcia embeddingu, modelu i wyszukiwania.

Gdzie wrócić w prezentacji

W rozdziale 1 odszukaj „Porównanie dwóch sposobów wyszukiwania” i „Ćwiczenie — rozpoznaj podobieństwo znaczeniowe”.

Praca krok po kroku

Krok 1 Przeczytaj materiał — Otwórz dane/R1_fragmenty.md i przeczytaj pytanie oraz wszystkie cztery fragmenty.

- Pracuj bez podpowiedzi modelu i bez klucza prowadzącego.
- Ustal, jakiej czynności i jakiego warunku dotyczy pytanie.

Krok 2 Rozpoznaj wspólne słowa — W każdym fragmencie zaznacz słowa lub ich odmiany występujące w pytaniu.

- Nie traktuj liczby wspólnych wyrazów jako dowodu trafności.
- Zapisz przykład, w którym wspólne słowo nie wystarcza do odpowiedzi.

Krok 3 Porównaj znaczenie — Wskaż wyrażenia opisujące podobną czynność innymi słowami.

- Zwróć uwagę na obiekt, warunek wykonania i oczekiwany rezultat.
- Odróżnij informację przydatną od jedynie związanego z nią tematu.

Krok 4 Ułóż ranking — Uporządkuj fragmenty od najbardziej do najmniej przydatnego.

- Dla każdego zapisz ocenę przydatności i jednozdaniowe uzasadnienie.
- Dopuszczalny remis wyjaśnij; nie wymyślaj wartości podobieństwa liczbowego.

Krok 5 Zapisz hipotezę — Utwórz hipotezy_podobienstwa.json z pytaniem i listą ocen.

- Dla każdego fragmentu zapisz: fragment_id, pozycja, uzasadnienie i zawiera_odpowiedz.
- Wynik oznacz jako analiza_reczna; własny wniosek zapisz w komórce wniosek.

Krok 6 Sprawdź rozumienie — Wyjaśnij własnymi słowami, czym jest embedding i czego nie gwarantuje.

- Rozróżnij model embeddingowy od modelu generującego odpowiedź.
- Zachowaj hipotezę bez późniejszego dopasowywania jej do wyników modelu.

Zapis próby: komórka Przygotowanie automatycznie tworzy nowy folder wyniki/R1/<proba_id>/. Zachowaj wyświetloną ścieżkę.

Punkty kontrolne pracy

Kiedy	Co sprawdzasz	Co zrobić przy problemie
Po lekturze	Ocena uwzględnia czynność i warunek z pytania.	Wróć do całego pytania, nie tylko słów kluczowych.
Po rankingu	Każdy fragment ma własne uzasadnienie.	Uzupełnij brakujące oceny lub uzasadnij remis.
Po zapisie	JSON otwiera się i zawiera cztery oceny.	Sprawdź cudzysłowy, przecinki i wartości true/false.

Nazewnictwo i zapis wyników

Notebook zapisuje hipotezy_podobienstwa.json i wnioski.md w wyniki/R1/<proba_id>/. Zachowaj także własny notebook.

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

- □ Oceniłem wszystkie cztery fragmenty.
- □ Wskazałem wspólne słowa i parafrazy.
- □ Uzasadniłem ranking na podstawie treści.
- □ Nie przedstawiam hipotez jako pomiaru modelu.
- □ Rozróżniam embedding i generowanie odpowiedzi.
- □ Zapisałem hipotezy_podobienstwa.json.

Gdy pojawi się problem

Objaw	Co sprawdzić
Kilka fragmentów wydaje się trafnych	Sprawdź, czy każdy zawiera informację potrzebną do odpowiedzi.
Trudno ustalić kolejność	Zapisz remis i wskaż, czego brakuje do rozstrzygnięcia.
Błąd przy odczycie JSON	W JSON użyj podwójnych cudzysłowów oraz true/false, nie True/False.

Zachowaj ocenę sprzed obliczeń. Rozbieżność z przyszłym rankingiem jest materiałem do analizy. Zapisz także uzupełniony notebook.