

Zaawansowane bazy danych i hurtownie danych

Praca domowa #5: MongoDB

Rozwiąż poniższe zadania implementując je w dowolnym języku programowania. Do sprawozdania dołącz również kod programu, który odpowiedzialny jest za połączenie z bazą danych MongoDB. Sprawozdanie (plik PDF) powinno zawierać numer i treść zadania, rozwiązanie zadania oraz zrzut ekranu prezentujący wywoływane polecenie oraz wynik. Jeśli wynikiem zadania będzie duża liczba dokumentów, zrzut ekranu powinien prezentować fragment tego wyniku i informację o liczbie zwróconych dokumentów. Zadania powinny być rozwiązane w oparciu o jedno polecenie.

Korzystając z `mongoimport` załaduj dane o filmach:

- (a) z pliku `title.basics.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Title`
- (b) z pliku `title.ratings.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Rating`
- (c) z pliku `name.basics.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Name`

Zadanie 1

Sprawdź liczbę dokumentów w kolekcjach `Title`/`Rating`/`Name`.

Zadanie 2

Wybierz 4 pierwsze dokumenty z kolekcji `Title`, które były wyprodukowane w roku 2020, są z kategorii filmów `Romance`, ich czas trwania jest większy niż 90 minut, ale nie przekracza 120 minut. Zwracane dokumenty powinny zawierać tytuł, rok produkcji, kategorię oraz czas trwania. Dane uporządkuj rosnąco wg tytułu filmu. Sprawdź również, ile dokumentów zwróciłoby zapytanie po wyłączeniu ograniczenia w postaci 4 pierwszych dokumentów. Wyszukując łańcuchy skorzystaj z operatora `$regex`.

Zadanie 3

Sprawdź ile filmów różnego typu (pole `titleType`) było wyprodukowanych w roku 2000. Wynik zapytania powinien zwracać nazwę typu oraz liczbę filmów.

Zadanie 4

W oparciu o kolekcje `Title` oraz `Rating` sprawdź średnią ocenę filmów dokumentalnych wyprodukowanych w latach 2010-2012. Wyświetl tytuł filmu, rok produkcji oraz jego średnią ocenę. Dane uporządkuj malejąco wg średniej oceny.

- sprawdź, ile takich dokumentów zwróci zapytanie
- wyświetl tylko 5 pierwszych dokumentów spełniających powyższe warunki

Zadanie 5

Utwórz indeks tekstowy dla pola `primaryName` w kolekcji `Name`. Następnie używając tego indeksu znajdź dokumenty opisujące osoby o nazwisku `Fonda` oraz `Coppola`. Przy wyszukiwaniu włącz opcję, która będzie uwzględniać wielkie/małe litery.

a) Ile dokumentów zwraca zapytanie?

b) Wyświetl 5 pierwszych dokumentów (dokument powinien zwracać dwa pola: `primaryName`, `primaryProfession`).

Zadanie 6

Utwórz indeks z porządkiem malejącym dla pola `birthYear` (kolekcja `Name`). Następnie wyświetl listę indeksów w kolekcji `Name`. Ile indeksów posiada ta kolekcja?

Zadanie 7

Dla każdego filmu (kolekcja `Title`), który ma najwyższą średnią ocenę (10.0), dodaj pole `max` z wartością równą 1. W poleceniu skorzystaj z kolekcji `Rating`, która zawiera informacje o średniej ocenie filmu.

Zadanie 8

W oparciu o kolekcje `Title` oraz `Rating` sprawdź średnią ocenę dowolnego filmu o określonym tytule oraz roku produkcji. Zapytanie powinno zwrócić nazwę filmu, rok produkcji oraz średnią ocenę.

Zadanie 9

Do filmu `Blade Runner` z roku 1982 dodaj pole `rating`, które będzie tablicą dokumentów z polami: `averageRating` oraz `numVotes`. Wartości pól w zagnieżdżonym dokumencie powinny posiadać odpowiednie wartości pobrane z kolekcji `Rating`.

Zadanie 10

Zmodyfikuj pole `rating` w dokumencie z Zadania 9, dodając jeszcze jeden dokument z polami `averageRating` oraz `numVotes` oraz z wartościami: 10 oraz 55555.

Zadanie 11

Usuń pole `rating` dodane do filmu `Blade Runner` w Zadaniu 10.

Zadanie 12

Do filmu `Pan Tadeusz` z 1999 roku dodaj pole `avgRating` z wartością 9.1. Jeśli nie ma takiego filmu, polecenie powinno zadziałać jak `upsert`.

Zadanie 13

Z kolekcji `Title` usuń dokumenty, które reprezentują filmy wyprodukowane przed 1989 rokiem. Ile takich dokumentów zostało usuniętych?