

Zaawansowane bazy danych i hurtownie danych

Zadanie#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 wynik. Jeśli wynikiem zadania będzie duża liczba dokumentów, zrzut ekranu powinien prezentować fragment tego wyniku.

Zadanie 1

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.principals.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Cast`
- (c) z pliku `title.crew.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Crew`
- (d) z pliku `title.ratings.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Rating`
- (e) z pliku `name.basics.tsv.gz` (<https://datasets.imdbws.com/>) do bazy IMDB oraz kolekcji `Name`

Zadanie 2

Sprawdź liczbę dokumentów w kolekcjach `Title/Cast/Crew/Rating/Name`. Wyświetl pierwszy dokument każdej kolekcji.

Zadanie 3

Wybierz 5 pierwszych dokumentów z kolekcji `Title`, które były wyprodukowane w roku 2005, są z kategorii filmów `Romance`, ich czas trwania jest większy niż 100 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 5 pierwszych dokumentów.

Zadanie 4

Wybierz filmy z kolekcji `Title` z roku 1930, które są z kategorii filmów `Comedy`. Zwracane dokumenty powinny zawierać oryginalny tytuł filmu, czas trwania filmu oraz kategorię. Dane uporządkuj malejąco wg czasu trwania filmu.

Zadanie 5

Korzystając z kolekcji `Cast` (lub `Crew`), `Title` oraz `Name`, sprawdź kto był reżyserem filmu `Casablanca` z 1942 roku. Podaj imię i nazwisko reżysera oraz jego datę urodzenia. Jeśli to możliwe, skorzystaj tylko z jednego polecenia.

Zadanie 6

Sprawdź ile filmów różnego typu (pole `titleType`) było wyprodukowanych w latach 2007-2009. Wynik zapytania powinien zwracać nazwę typu oraz liczbę filmów.

Zadanie 7

W oparciu o kolekcje `Title` oraz `Rating` sprawdź średnią ocenę filmów dokumentalnych wyprodukowanych w latach 1994-1996. Wyświetl tytuł filmu, rok produkcji oraz jego średnią ocenę. Dane uporządkuj malejąco wg średniej oceny.

- a) sprawdź, ile takich dokumentów zwróci zapytanie
- b) wyświetl tylko 10 pierwszych dokumentów spełniających powyższe warunki

Zadanie 8

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 9

Stwórz indeks tekstowy dla pola `primaryProfession` w kolekcji `Name`. Następnie używając tego indeksu znajdź 5 pierwszych osób urodzonych w latach 1950-1980, które są aktorami lub reżyserami. Sprawdź również liczbę zwracanych dokumentów po wyłączeniu ograniczenia w postaci 10 pierwszych dokumentów. Zapytanie powinno zwrócić imię i nazwisko osoby, datę urodzenia oraz profesję.

Zadanie 10

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`).
- c) Sprawdź, ile indeksów posiada kolekcja `Name`?