

Zadanie «Nagroda Nobla» (nob)

Wprowadzenie

Co roku, w październiku, świat nauki z wielkim napięciem i zainteresowaniem kieruje swoją uwagę na prace Komitetu Noblowskiego, który decyduje o uhonorowaniu wybitnych jednostek, zespołów lub organizacji za ich znaczący wkład w rozwój wiedzy należącej do jednej z pięciu kategorii. Osobną, szóstą, kategorię stanowi Nagroda Pokojowa, której laureaci są nagradzani za działalność na rzecz pokoju.

Poza społecznością naukową, również i opinia publiczna na bieżąco śledzi statystyki i ciekawostki związane z laureatami, takie jak np. liczba osób nagrodzonych wielokrotnie, najczęściej nagradzane kraje i instytucje, czy np. lista najmłodszych/najstarszych wśród osób nagrodzonych w poszczególnych kategoriach.

Źródłem wiedzy o laureatach Nagrody Nobla może być krajowy portal danych Szwecji, w którego publicznych zasobach cyfrowych są utrzymywane stosowne pliki zapisane w różnych formatach danych.

Zadanie

W folderze tego zadania jest plik `nob.csv`, zawierający szczegółowe informacje o laureatach Nagrody Nobla.

Do rozwiązywania zadań potrzebne Ci będą takie dane zapisane w tym pliku:

- imię i nazwisko laureata lub jego nazwa (pole `name`),
- nazwisko laureata (pole `known_name`),
- płeć laureata (pole `gender`),
- data i kraj urodzenia laureata (pola `birth_date` i `birth_country`),
- data śmierci laureata (pole `death_date`),
- afiliacja laureata (pole `affiliation_1`),
- kategoria, w której laureat otrzymał nagrodę (pole `category`),
- tytuł osiągnięcia (pole `motivation`),
- rok otrzymania wyróżnienia (pole `awardYear`),
- data wyřęczenia wyróżnienia (pole `dateAwarded`),

- wysokość nagrody pieniężnej (pole `prizeAmount`),
- informacja o tym, czy dana nagroda była indywidualna, czy też została nadana dla laureata będącego organizacją (pole `ind_or_org`),
- informacja o tym, jaki był udział/wkład danej osoby w osiągnięciu (pole `portion`), jeżeli nagroda była wieloosobowa.

Część 1 Zaprojektuj relacyjną bazę danych, w której możliwe będzie zapisanie wymienionych wyżej danych (interesują Cię **tylko** podane wyżej dane) i zapewnij jej spójność poprzez definicję odpowiednich więzów integralności. Analizując zapisane dane, samodzielnie zdecyduj, jakie więzy integralności powinny zostać nałożone. Rozpatrz zasadność nałożenia takich więzów, jak:

- klucz główny,
- klucz obcy,
- nie dopuszczanie wystąpienia wartości `NULL`,
- ograniczenie `CHECK`,
- więzy klucza jednoznacznego `UNIQUE`.

Część 2 Napisz program w dowolnym języku wysokiego poziomu, który zaimportuje wskazane dane z pliku CSV do utworzonej w poprzednim kroku bazy danych.

Część 3 Wykorzystując zaprojektowaną i wypełnioną danymi bazę danych, wykonaj następujące zapytania SQL.

1. Wypisz imiona i nazwiska kobiet – laureatek, które na przestrzeni lat 1940-2019 otrzymały nagrody w więcej niż jednej kategorii.
2. Dla każdej kategorii nagrody, wypisz nazwy krajów, w których urodziła się największa liczba laureatów.
3. Napisz zapytanie sprawdzające, czy istnieje zależność pomiędzy kategorią nagrody a miejscem urodzenia laureatów. W tym celu:
 - (a) W pierwszym kroku stwórz perspektywę (widok), który agreguje liczbę laureatów w poszczególnych kategoriach z podziałem na kraje urodzenia.

- (b) W drugim kroku, korzystając z napisanego przez siebie w punkcie 3a widoku, wykonaj zapytanie zwracające listę nazw krajów, które dominują w poszczególnych kategoriach nagród.

Uwagi dodatkowe

Projektując bazę danych wnikliwie przeanalizuj plik CSV i zwróć m. in. uwagę na to, że:

1. Niektóre dane powtarzają się – np. informacja o imieniu i nazwisku laureata, jeżeli otrzymał on Nagrodę Nobla kilka razy. Tymczasem w bazie nie powinno być redundancji danych – czyli dane nie powinny w niej być powielane.

Przykładowo, informacja o imieniu i nazwisku laureata nie powinna się powtarzać nawet wtedy, gdy na przestrzeni lat otrzymał on więcej niż jedną Nagrodę Nobla.

2. Niektóre pola mogą mieć pustą zawartość.

Przykładowo, jeżeli nagrodę otrzymuje jakaś organizacja, wtedy pole `family_name` jest niewypełnione.

3. Niektóre pola są od siebie logicznie zależne.

Przykładowo, data urodzenia laureata nie jest późniejsza niż rok uzyskania nagrody.

Ustalenia techniczne

1. Rozwiązaniem zadania są trzy pliki (z rozszerzeniami odpowiednimi do zawartości):

- plik tekstowy o nazwie `IKU-nob-1` z zestawem poleceń DDL tworzących stosowne struktury na potrzeby części 1,
- plik o nazwie `IKU-nob-2` z kodem źródłowym programu z części 2, który importuje wskazane dane z pliku CSV do zaprojektowanej bazy danych,
- plik tekstowy o nazwie `IKU-nob-3` z trzema zapytaniami SQL, wyświetlającymi stosowne informacje podane w części 3.

Wszystkie powyższe pliki należy spakować do pliku `IKU-nob.zip`, gdzie `IKU` jest *indywidualnym kodem uczestnika*.

2. Plik `IKU-nob.zip` należy przesłać przez stronę konkursu «Złoty Indeks» Platformy Zdalnej Edukacji korzystając z łącza do przesyłania rozwiązań zadania «nob».
3. Zadanie jest oceniane w skali 0-15 punktów.