

Wyjściowy schemat danych

Dana jest baza danych o zasobach archiwalnych:

Archiwa (Archives) - informacje o archiwach, w których znajdują się dokumenty.

```
CREATE TABLE Archives (  
    id SERIAL PRIMARY KEY,  
    name VARCHAR(255) NOT NULL,  
    location TEXT,  
    description TEXT);
```

Autorzy (Authors) - Informacje o autorach dokumentów archiwalnych.

```
CREATE TABLE Authors (  
    id SERIAL PRIMARY KEY,  
    first_name VARCHAR(100) NOT NULL,  
    last_name VARCHAR(100) NOT NULL,  
    birth_date DATE,  
    death_date DATE,  
    nationality VARCHAR(100));
```

Dokumenty (Documents) - informacje o dokumentach archiwalnych.

```
CREATE TABLE Documents (  
    id SERIAL PRIMARY KEY,  
    title VARCHAR(255) NOT NULL,  
    creation_date DATE,  
    archive_id INT REFERENCES Archives(id) ON DELETE SET NULL,  
    description TEXT,  
    type VARCHAR(100),  
    language VARCHAR(50),  
    digital_copy_url TEXT);
```

Znaczniki (Tags) - słowa kluczowe opisujące dokumenty archiwalne.

```
CREATE TABLE Tags (  
    id SERIAL PRIMARY KEY,  
    name VARCHAR(100) UNIQUE NOT NULL  
);
```

```
CREATE TABLE DocumentAuthors (  
    document_id INT REFERENCES Documents(id) ON DELETE CASCADE,  
    author_id INT REFERENCES Authors(id) ON DELETE CASCADE,  
    PRIMARY KEY (document_id, author_id));
```

```
CREATE TABLE DocumentTags (  
    document_id INT REFERENCES Documents(id) ON DELETE CASCADE,  
    tag_id INT REFERENCES Tags(id) ON DELETE CASCADE,  
    PRIMARY KEY (document_id, tag_id));
```

Zadanie 1

Przeprowadzić proces inżynierii wstecz (ang. *reverse engineering*) tej bazy danych i przedstawić jej model relacji w postaci diagramu.

Zaproponować rozszerzenie tego schematu o informacje o osobach, które są wymienione w dokumentach, np. list króla Jana II Kazimierza Wazy do elektora saskiego Jana Jerzego i Wettyna. Osoby powinny być łatwe do wyszukiwania bez konieczności analizy tekstu dokumentu. Musi być też możliwość szybkiego znalezienia wszystkich lub wybranych dokumentów ze wzmiankami o danej osobie, np. "wszystkie dokumenty adresowane do Jana Kazimierza", "wszystkie dokumenty wysłane przez Jana Kazimierza", "wszystkie dokumenty, w których wspomniano o Janie Kazimierz po jego śmierci w 1672 roku" itp.

Rozwiązanie ma mieć postać diagramu relacji i skryptu SQL tworzącego nowe tabele i/lub modyfikującego tabele istniejące.

Zadanie 2

Napisać następujące zapytania SQL do wyjściowego schematu danych ze strony 1:

- Dla każdego autora podać znacznik (znaczniki jeśli jest ich więcej niż 1) najczęściej występujące w dokumentach jego autorstwa.
- Dla autora Krzysztof Stencel podać archiwa, w których nie znajdują się dokumenty jego autorstwa.
- Zakładając, że lista dopuszczalnych wartości w kolumnach `Authors.nationality` i `Documents.language` jest taka sama, ustalić listę autorów, którzy kiedykolwiek stworzyli dokument w nie swoim języku.

Zadanie 3

Archiwum prowadzi dziennik zmian w archiwalnej bazy danych w postaci tabeli relacyjnej o następujących kolumnach:

- ~~login, first_name, last_name, email, password, role,~~
~~change_id, document_id, document_title, archive_id,~~
~~archive_name, archive_location, change_timestamp, changes,~~
~~approving_officer_login, approving_officer_first_name,~~
~~approving_officer_last_name~~

Przeanalizuj tę tabelę, ustal zależności funkcyjne i wskaż, w której jest postaci normalnej. Następnie przeprowadź normalizację do 3NF lub BCNF i narysuj diagram ERD powstałego zestawu tabel.

Jeśli trzeba przyjmij rozsądne założenia odnośnie do tej tabeli. Wypisz je.