

# TASS Projekt 2

Radosław Kostrzewski, Bartosz Latosek, Patryk Będkowski

## 1. Temat

*Dla wskazanego profilu wykonawcy, wyszukaj w Spotify lub Discogs innych wykonawców, którzy z nim współpracowali i jednocześnie spełniają dodatkowe kryterium dostępne na ich stronie w Wikipedii lub innym serwisie.*

Zdecydowaliśmy się na zrealizowanie aplikacji interaktywnej, za pomocą której użytkownicy przy pomocy specyficznych filtrów mogą wyszukiwać informacje o interesujących ich artystach wraz z muzykami, z którymi współpracowali.

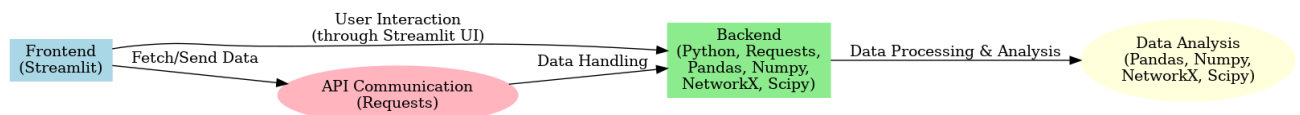
## 2. Źródła danych

1. **Spotify Web API:** Spotify oferuje dość intuicyjne i proste w obsłudze API pozwalające na wyszukiwanie artystów, albumów i utworów, co na pewno będzie przydatne w identyfikacji współpracy między wykonawcami.
2. **MusicBrainz:** Szczegółowe informacje o artystach, z których można pobierać interesujące nas informacje zadane przez użytkowników aplikacji. Strona udostępnia API, ale dodatkowo przez ułożoną strukturę strony będzie dobrym kandydatem do *webscraping'u*.

## 3. Technologie wykorzystane w aplikacji

1. **Frontend:** Do obsługi aplikacji przez użytkownika wykorzystamy narzędzie `streamlit` (Biblioteka języka Python udostępniana przez *Snowflake*), które umożliwia szybkie prototypowanie aplikacji.
2. **Pobieranie i przetwarzanie danych:** Do komunikacji z API Spotify planujemy wykorzystać bibliotekę `requests` w Pythonie. W przypadku pobierania danych z serwisu MusicBrainz zastosujemy biblioteki `Selenium` która służy do scrapowania danych z serwisów.
3. **Analiza danych:** Jako główne narzędzie analizy będziemy używać języka Python wraz z modułami takimi jak `pandas`, `numpy`, `networkx`, `scipy`. Do wizualizacji grafów posłużymy się narzędziem `pyviz`.

## 4. Architektura aplikacji



## 5. Metodyka analizy danych i uzyskania zadowalającego złączenia

**Klucz łączący:** Podstawowym kluczem łączącym dane z różnych źródeł będzie unikalna nazwa artysty, aby poprawnie zidentyfikować i połączyć odpowiednie rekordy.

**Analiza współpracy:** Na podstawie danych z Spotify zidentyfikujemy artystów, którzy współpracowali z wybranym wykonawcą.

**Weryfikacja dodatkowych kryteriów:** Korzystając z informacji z MusicBrainz, sprawdzimy, czy artyści spełniają określone dodatkowe kryteria, takie jak gatunek muzyczny, kraj pochodzenia, data urodzenia czy okres aktywności.

## 6. Scenariusz użycia aplikacji

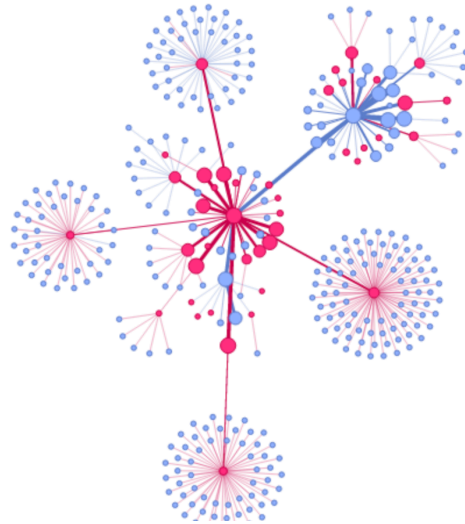
Opracujemy aplikację która po wpisaniu nazwy artysty zaprezentuje użytkownikowi listę współtwórców spełniających podane przez niego kryteria.

Użytkownik wpisuje nazwę artysty, którego działalność chce przeanalizować. Dodatkowo może zaznaczyć filtry, aby przeanalizować tylko wybrane współpracy z artystami, takie jak:

- Gatunek muzyczny (np. jazz, rock, pop).
- Kraj pochodzenia (np. USA, Polska).
- Okres aktywności (np. lata 80., obecnie aktywny).
- Data urodzenia artystów (np. po 1990 roku).

Aplikacja wysyła zapytanie do Spotify Web API, aby znaleźć informacje o współpracach wskazanego artysty. Pobiera dane o artystach współpracujących z MusicBrainz przy pomocy webscrapingu. Filtruje wyniki według kryteriów określonych przez użytkownika. Łączy dane z obu źródeł za pomocą unikalnej nazwy artysty jako klucza.

Na tej podstawie tworzony jest graf połączeń gdzie węzłem centralnym jest wyszukiwany pierwotnie twórca, z połączeniami do innych artystów, z którymi współpracował. Krawędzie między artystami posiadają wagę oznaczającą ilość współprac. Filtrowanych jest Top 10 artystów względem współprac. Otrzymujemy graf podobny do poniższego:



Źródło: Pyvis: A Python Library for Neo4j Graph Visualization. Medium.com

Stworzony graf będzie zaprezentowany użytkownikowi w aplikacji i będzie on pozwalał na interaktywne opisywanie każdego z węzłów.

#### Metryki do analizy współpracy

1. **Liczba unikalnych współprac**
  - Liczba unikalnych artystów współpracujących z pierwotnie wyszukiwanym twórcą.
2. **Top najczęstszych współpracowników:**
  - 10 najczęściej pojawiających się węzłów w grafie
  - Najczęstsze gatunki muzyczne wśród współpracowników.
3. **Najbardziej wpływowa współpraca:**
  - Węzeł z najwyższą wagą (najwięcej wspólnych projektów).
4. **Dominujące kraje pochodzenia współpracowników**
  - Procentowy udział artystów z różnych krajów pochodzenia wśród współpracowników.
5. **Proporcja współprac według okresu aktywności**
  - Rozkład współprac na różne lata działalności.
6. **Centralność w grafie (Betweenness Centrality)**
  - Metryka pokazująca, który artysta pełni kluczową rolę jako „pośrednik” w grafie współprac.
7. **Gęstość grafu współpracy**
  - Miara określająca, jak wiele połączeń istnieje w stosunku do liczby możliwych połączeń w grafie.
8. **Najczęściej współpracująca grupa gatunkowa**
  - Grupowanie współpracowników według gatunków i identyfikacja, który gatunek dominuje w współpracach.