



POLITECHNIKA
POZNAŃSKA

Zastosowania Technologii Informatycznych

Projekt:

**“Gdzie mogę znaleźć
odpowiedź na moje pytanie?”**

Spis treści:

1. Założenia projektu	3
a. Założenie 1	3
b. Założenie 2	3
c. Założenie 3	3
2. Wykorzystane technologie.....	3
3. Źródła danych	4
4. Kryteria wyszukiwania	5
5. Problemy	6
a. Opis problemu	6
b. Statystyki	7

1. Założenia projektu:

Użytkownik będzie mógł wyszukać odpowiedzi na zdefiniowane już pytania: “Gdzie mogę znaleźć moich przodków?”, “Gdzie mogę znaleźć utwór muzyczny/wykonawcę” oraz przeszukiwać endpointy strony wikidata. Do uzyskania odpowiedzi wykorzystamy zapytania SPARQL na dwóch bazach danych: dbpedia.org oraz europeana.eu.

Założenie 1: Tworzona przez nas aplikacja ma za zadanie połączyć się z bazami: dbpedia, europeana.eu oraz zdefiniowanymi przez użytkownika endpointami i pobrać z nich znajdujące się tam informacje.

Założenie 2: Użytkownik ma możliwość zdefiniowania poszczególnych danych dotyczących szukanych przez niego osób (np. miejsce urodzenia/śmierci). W ten sposób użytkownicy mogą szukać swoich potencjalnych przodków - porównując miejsce i czas życia osób, które aplikacja zwróci im jako pasujący do ich kryteriów. Podobnie może filtrować i wyszukiwać utwory/wykonawców.

Założenie 3: Wyniki zapytania mają zostać wyświetlone na ekranie użytkownika jako rezultat przeszukania bazy z możliwością ich przeglądania.

Założenie 4: Użytkownik może także zdefiniować własne zapytania SPARQL oraz endpointy do których je wysyła. W odpowiedzi powinien otrzymać listę stosowną do swojego zapytania.

2. Wykorzystane technologie:

Do realizacji naszego projektu wykorzystaliśmy SPARQL.NET - API napisane w języku C#. Jako dodatkowe zadanie postanowiliśmy dodać owo API do silnika Unity 3D, aby uzyskać szeroki wachlarz możliwości wzbogacenia wizualnego wyników otrzymywanych do naszych zapytań. Unity bardzo dobrze współpracuje z językiem C# oraz oferuje ogromne możliwości graficzne, dlatego właśnie wszystkie te elementy tak bardzo do siebie pasowały i tworzyły bardzo ciekawy do realizacji projekt.

3. Źródła danych:

Źródłem danych do naszego projektu jest baza wiedzy dbpedia.org - specjalny rodzaj bazy danych zwany bazą wiedzy (ang. knowledge graph), który zapewnia możliwość zbierania, organizowania, dzielenia oraz przeszukiwania informacji. Dbpedia jest własnością firmy Google, która używa podobnego podejścia przy wyszukiwaniu danych w swojej wyszukiwarce.

Dodatkowo skorzystamy też z bazy europeana.eu, która zawiera informacje o utworach muzycznych oraz ich twórcach.

4. Kryteria wyszukiwania:

1. Wyszukiwanie przodków:

Wyszukiwanie w bazie umożliwia filtrowanie 4 podstawowych informacji o osobach tam się znajdujących:

- miejsca urodzenia
- miejsca śmierci
- rok przed którym osoba się urodziła
- rok przed którym osoba zmarła

Dzięki takiemu podziałowi możemy:

- sprecyzować region w którym poszukujemy naszych przodków - zawężając wyszukiwanie do miasta z którego pochodzą nasi przodkowie, lub w jakim umarli
- wyszukiwać przodków tylko w okresie do roku z którego posiadamy dane - w większości przypadków użytkownik zna dwa pokolenia wstecz, dlatego aplikacja oferuje możliwość pominięcia wyszukiwania w tym okresie

2. Wyszukiwanie utworów muzycznych:

Wyszukiwanie w bazie umożliwia filtrowanie:

- nazwy utworów muzycznych
- imię/nazwisko wykonawcy

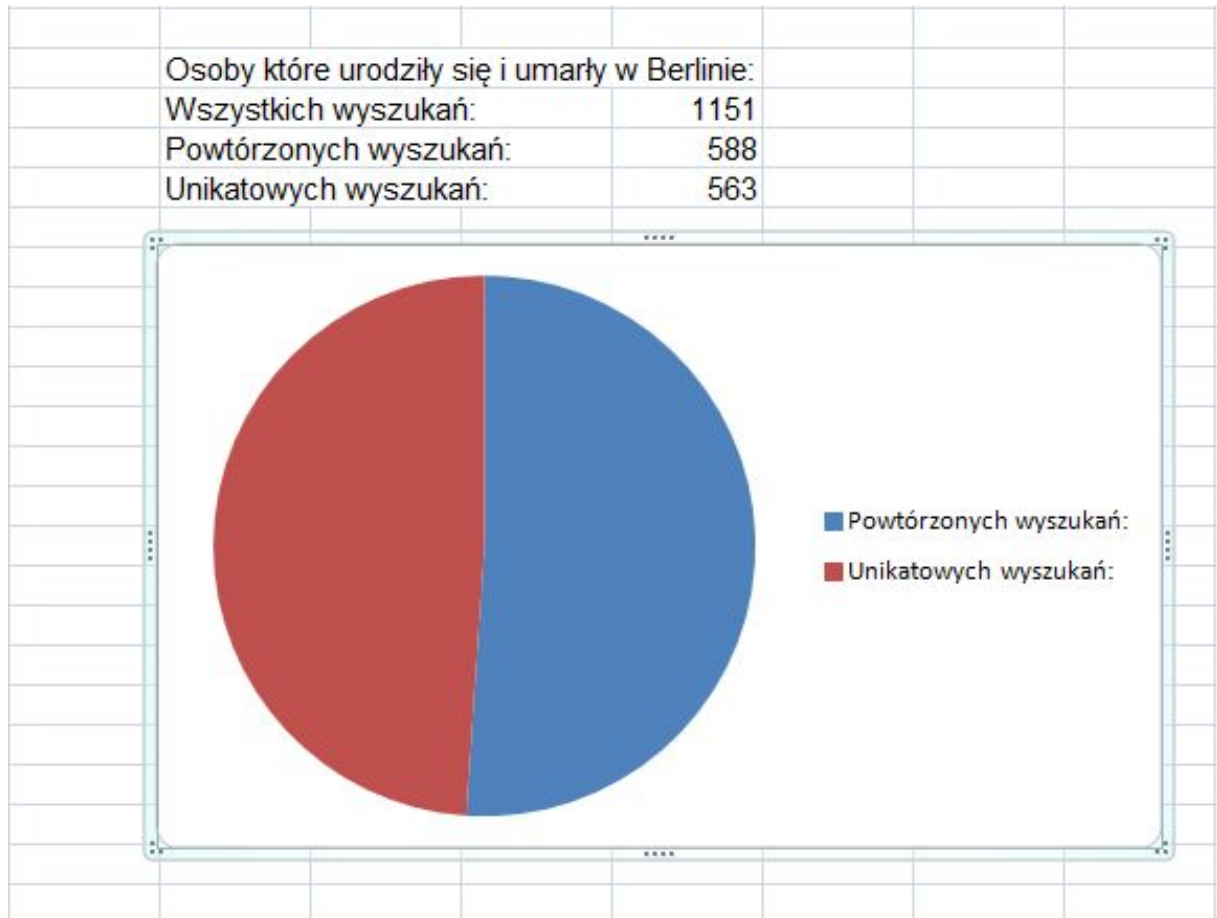
3. Wyszukiwanie endpointów:

Wyszukiwanie w bazie umożliwia filtrowanie:

- części pytania

5. Problemy:

Podstawowym problemem jaki napotkaliśmy to wielorazowe zwracanie tej samej osoby. Aby rozwiązać ten problem wprowadziliśmy kontrolę powtarzających się imion oraz odnośników do miejsc w bazie, ale nawet takie sprawdzanie otrzymywanych rekordów nie okazało się w 100% skuteczne. Poniżej przykłady powtarzających się wyników dla przykładowych zapytań:

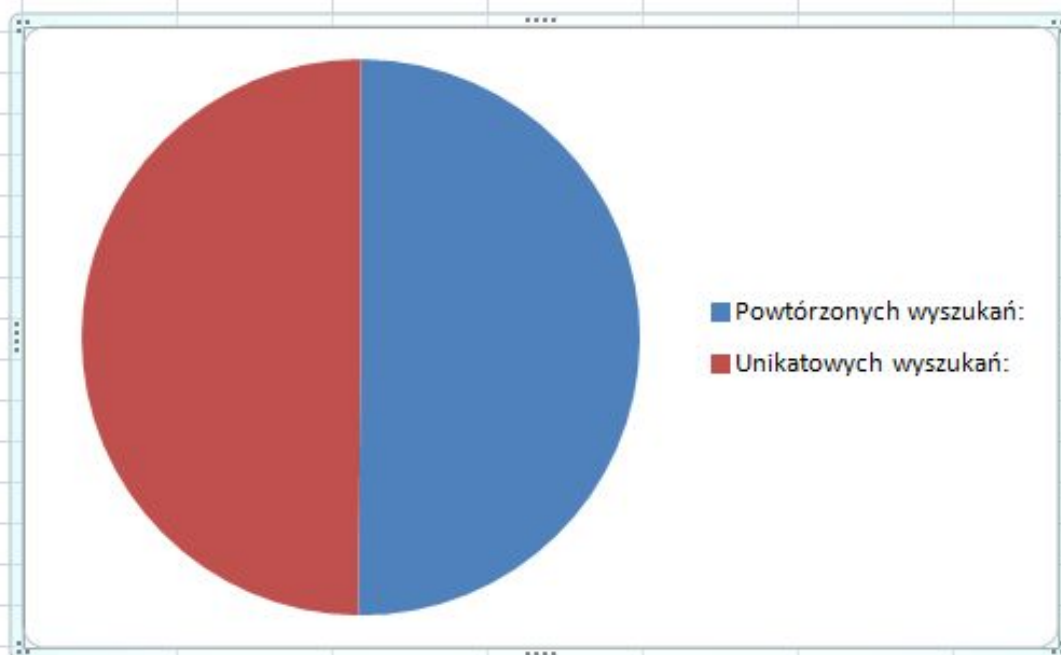


Osoby które urodziły się w Warszawie przed 1990 rokiem:

Wszystkich wyszukiń: 937

Powtórzonych wyszukiń: 470

Unikatowych wyszukiń: 467

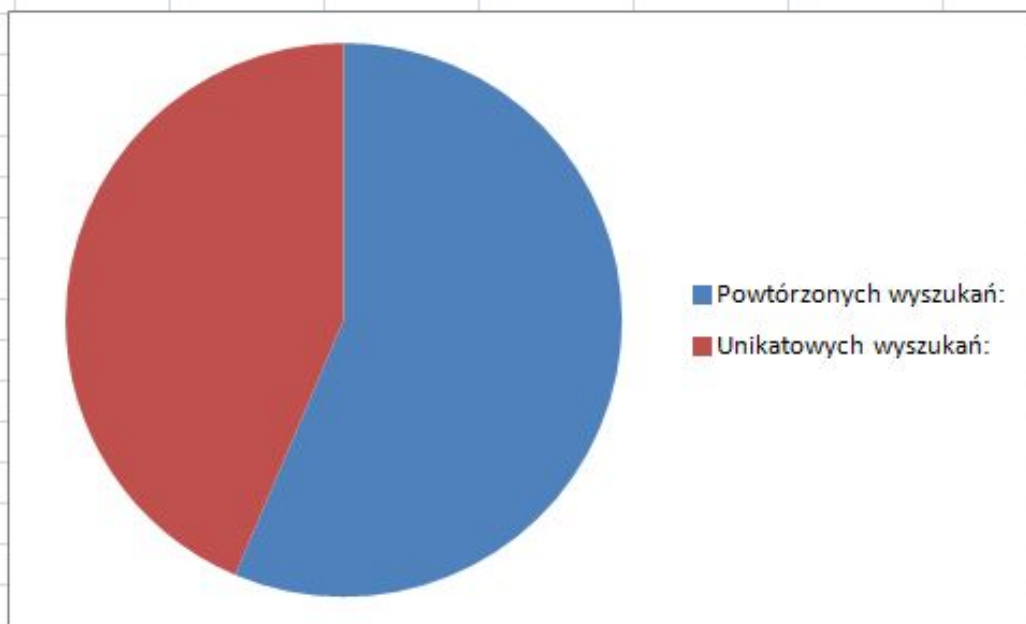


Osoby urodzone w Londynie przed 1900 rokiem:

Wszystkich wyszukiń: 1810

Powtórzonych wyszukiń: 1021

Unikatowych wyszukiń: 789



Jak widać w zależności od zapytania prawie połowa wyników zostaje zidentyfikowana jako duplikaty i nie jest wyświetlana jako ostateczna odpowiedź.