



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	4
5. Problemy	4
a. Opis problemu	4
b. Statystyki	5

1. Założenia projektu:

Pytanie na które będziemy szukać odpowiedzi to “Gdzie mogę znaleźć moich przodków?”, odpowiedzi zaś szukać będziemy za pomocą zapytań SPARQL w bazie danych dbpedia.org.

Założenie 1: Tworzona przez nas aplikacja ma za zadanie połączyć się z bazą dbpedia i pobrać z niej informację o znajdujących się tam osobach.

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.

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

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.

4. Kryteria wyszukiwania:

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:

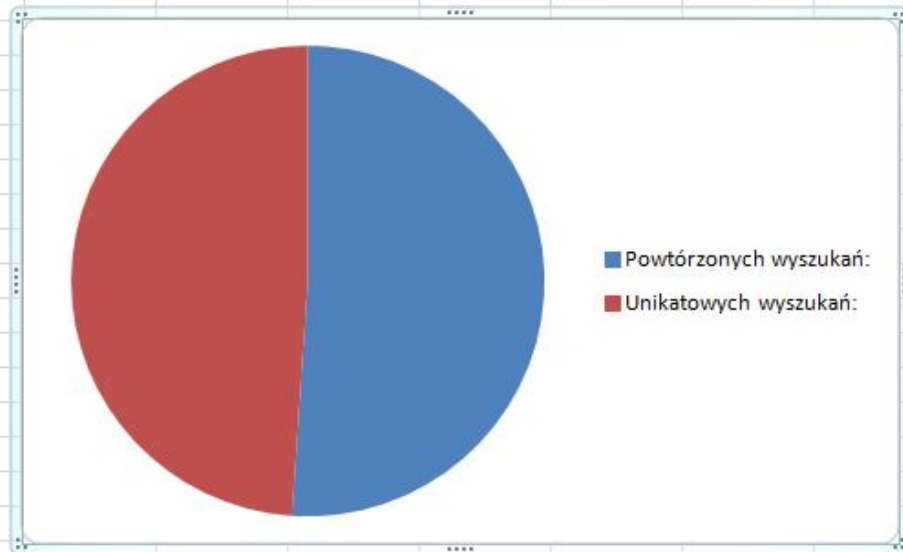
- sprecyzować region w którym poszukujemy naszych przodków - zawężając wyszukiwanie do miasta z którego pochodzą nasi przodkowie, lub w jakim umarli
- wyszukiwać przodków tylko w okresie do roku z którego posiadamy dane - w większości przypadków użytkownik zna dwa pokolenia wstecz, dlatego aplikacja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

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:

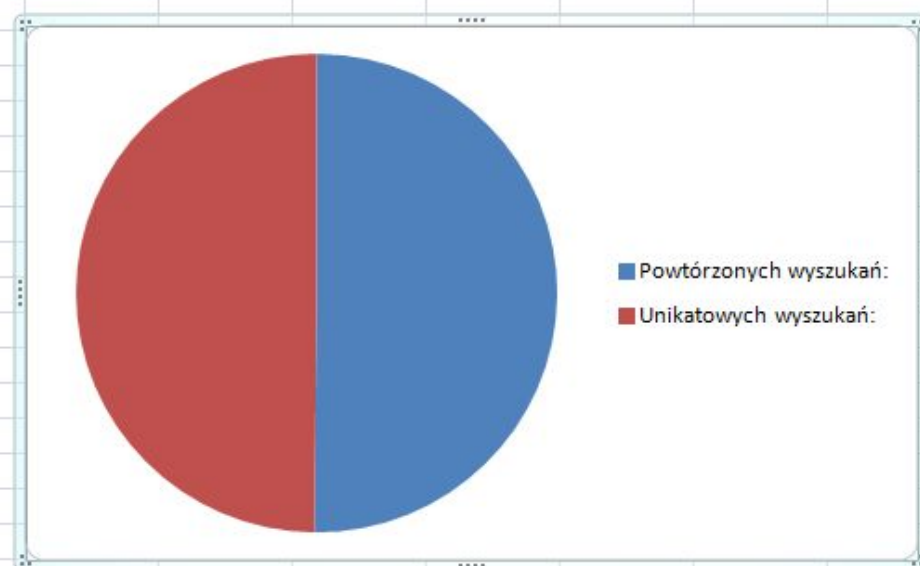
Osoby które urodziły się i umarły w Berlinie:

Wszystkich wyszukiń:	1151
Powtórzonych wyszukiń:	588
Unikatowych wyszukiń:	563

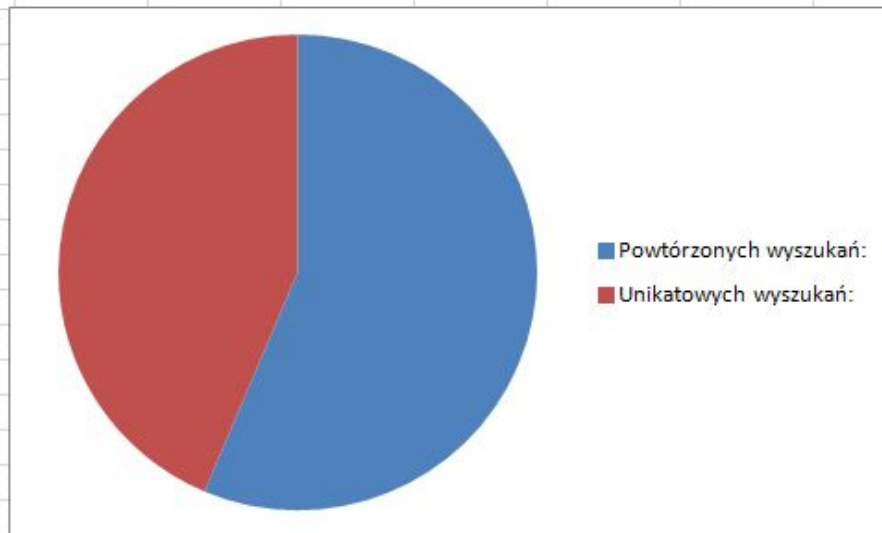


Osoby które urodziły się w Warszawie przed 1990 rokiem:

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.