

Generator sztucznych danych osobowych

Ewa Pasterz

Grudzień 2022

Spis treści

1	Opis projektu	2
2	Wymagania funkcjonalne	2
2.1	Historie użytkownika	2
2.2	Opis kluczowych funkcjonalności	3
3	Wymagania niefunkcjonalne	3
3.1	Użyteczność	3
3.2	Niezawodność	3
3.3	Wydajność	4
3.4	Wsparcie	4
4	Technologie	4
5	Źródła	4

1 Opis projektu

Celem projektu jest stworzenie aplikacji generującej sztuczne dane osobowe. Generator będzie dostępny jako aplikacja okienkowa napisana w języku Java w środowisku IntelliJ IDEA Community.

Komisja Europejska definiuje dane osobowe jako „wszelkie informacje dotyczące zidentyfikowanej lub możliwej do zidentyfikowania żyjącej osoby fizycznej. Poszczególne informacje, które w połączeniu ze sobą mogą prowadzić do zidentyfikowania tożsamości danej osoby, także stanowią dane osobowe.” Przykładowymi wymienianymi danymi osobowymi są imię i nazwisko, adres zamieszkania, adres e-mail taki jak imię.nazwisko@mail.com (ale już nie typu info@firma.com), dane o lokalizacji czy adres IP. Na podstawie definicji Komisji Europejskiej zdecydowałam, że dane generowane w moim projekcie będą uwzględniały:

- imię,
- nazwisko,
- płeć,
- kraj pochodzenia,
- adres zamieszkania (kraj, miasto, ulica, numer domu, kod pocztowy),
- numer telefonu,
- adres e-mail generowany na podstawie imienia i nazwiska.

Część danych będzie generowana na podstawie innych, już określonych, danych, np. generowane imię będzie pasowało do płci i kraju pochodzenia. Numer telefonu będzie zgodne z zasadami tworzenia tego numeru w wygenerowanym kraju zamieszkania.

Wygenerowane dane będzie można indywidualnie ponownie generować, bez zmieniania pozostałych informacji. Dane będzie można zapisać na dysku w postaci pliku JSON.

2 Wymagania funkcjonalne

2.1 Historie użytkownika

1. Jako użytkownik chcę móc wygenerować sztuczne dane osobowe, aby mieć przykładowe dane osobowe.
2. Jako użytkownik chcę móc wygenerować spójne i logiczne dane, aby przypominały one jak najbardziej prawdziwe dane osobowe.
3. Jako użytkownik chcę móc zmienić poszczególne dane, aby mieć większy wpływ na wygenerowany wynik.
4. Jako użytkownik chcę móc zapisać wygenerowane dane, aby móc je wykorzystać w przyszłości.

2.2 Opis kluczowych funkcjonalności

Na podstawie historii użytkownika wyodrębniłam następujące funkcjonalności:

- Aplikacja będzie generowała sztuczne dane osobowe zgodne z realnym światem, tzn. generowane kraje, adresy i imiona będą prawdziwe, a numer telefonu będzie generowany zgodnie z przepisami konkretnych państw.
- Użytkownik będzie uruchamiał generowanie danych poprzez naciśnięcie konkretnego przycisku.
- Naciśnięcie przycisku do generowania, gdy są już wygenerowane dane, spowoduje wygenerowanie nowych danych i nadpisanie poprzednich.
- Użytkownik będzie miał możliwość wygenerowania ponownie poszczególnych danych.
- Użytkownik będzie proszony o potwierdzenie swojej decyzji dotyczącej ponownego wygenerowania danych, zarówno gdy ponownie będzie generował wszystkie, jak i pojedyncze dane.
- Użytkownik będzie mógł zapisać wygenerowane dane na dysku.
- Wygenerowane imię i nazwisko będą pasować do wygenerowanego kraju pochodzenia i płci.
- Wygenerowany numer telefonu będzie pasować do kraju zamieszkania.
- Ponowne generowanie pojedynczych danych nie będzie wpływało na pozostałe dane, czyli np. ponownie wygenerowana płeć nie zmieni imienia.

3 Wymagania niefunkcjonalne

Wymagania zostały podzielone zgodnie z metodą FURPS. Wymagania funkcjonalne zostały opisane w poprzedniej sekcji, a pozostałe poniżej.

3.1 Użyteczność

- Interfejs aplikacji będzie czytelny i przejrzysty.
- Generowane dane będą ułożone w dwie kolumny.
- Aplikacja będzie utrzymana w stonowanej kolorystyce.
- Aplikacja będzie w języku angielskim.

3.2 Niezawodność

- Aplikacja nie będzie się zawieszać.
- Aplikacja będzie zależna od dostępu do internetu.

3.3 Wydajność

- Dane będą generowane natychmiastowo.

3.4 Wsparcie

- Użytkownik będzie proszony o potwierdzenie ponownego generowania danych, aby zapobiec niechcianej utracie wyników.
- W przypadku ponownego generowanie pojedynczych danych użytkownik będzie ostrzegany, że potencjalnie nowy wynik może być niespójny z pozostałymi wynikami.

4 Technologie

Aplikacja będzie napisana w języku Java w środowisku IntelliJ IDEA Community 2022.2.2. Interfejs graficzny aplikacji będzie zaimplementowany przy użyciu Java Swing.

Adresy będą pobierane przy użyciu Bing Map API lub pobrane ze zbioru OpenAdresses. Imiona i nazwiska będą pobierane przy użyciu UiNames API. Numer telefonu będzie generowany przy użyciu Randommer API. Jeśli wystąpią problemy przy używaniu któregoś z wymienionych API, zostanie ono zastąpione przez inne, jeśli takowe istnieje, lub możliwe do wygenerowania dane będą zapisane i pobierane w programie.

5 Źródła

1. Reforma unijnych przepisów o ochronie danych - Czym są dane osobowe?