Gabriela Lewandowska 01.04.2019

Paweł Prusałowicz

Prowadzący: Łukasz Neumann

**Dokumentacja wstępna**

Heurystyczne naprawianie atrybutów tekstowych

Zaawansowane Programowanie w C++

**Zadanie projektowe**

Zadanie projektowe polega na stworzeniu biblioteki z interfejsem w języku Python, która pozwoli na czyszczenie danych rzeczywistych na podstawie słownika.

**Opis funkcjonalności**

Biblioteka dostarcza użytkownikowi narzędzi niezbędnych do zastosowania we własnym programie narzędzi do czyszczenia danych rzeczywistych, które mogą potem zostać poddane dalszej analizie.

Podstawowymi funkcjami biblioteki są:

* możliwość załadowania własnego słownika, na podstawie którego może zostać naprawiona dowolna ilość zestawów danych
* możliwość załadowania własnego zestawu danych w postaci jednowymiarowego numpy.array, pandas.Series lub Pythonowej listy, gdzie typem elementu jest string
* wykonanie czyszczenia danych

**Technologie**

Biblioteka zaimplementowana zostanie w języku Python, jednak z uwagi na relatywnie dużą ilość danych, które mają zostać