

Opis wykonania projektu zaliczeniowego

Projekt końcowy realizowany w ramach problematyki projektowania nowoczesnych systemów informatycznych, wspierających analizę danych i procesy decyzyjne oparte na algorytmach data science, obejmuje zaprojektowanie, implementację i dokumentację systemu eksplorującego dane tekstowe z wykorzystaniem poznanych na zajęciach metod analizy danych oraz podejścia Reproducible Research.

Cele projektu i zakres

- stworzenie **działającego skryptu w języku R**, który analizuje dane tekstowe (z jednego lub wielu plików .txt, .csv) przy pomocy algorytmów używanych podczas zajęć
 - zidentyfikowanie i zbadanie **trendów w danych tekstowych** (np. eksploracja częstości słów, analiza sentymentu, klastrowanie, modelowanie tematów itp.)
 - opracowanie kompletnej **dokumentacji systemu informatycznego** zgodnej z podejściem inżynierii oprogramowania
 - spełnienie standardów **Reproducible Research**: kod oraz dane dostępne w formacie możliwym do powtórzenia i odtworzenia
-

Kryteria zaliczenia (0 - 50 punktów)

Część I. Reproducible Research (25 pkt)

- Plik z **kodelem źródłowym w R** oraz zestawem danych (txt) – 20 pkt
- **Raport HTML** wygenerowany ze skryptu R – 5 pkt

Część II. Dokumentacja Specyfikacji Wymagań (25 pkt)

Zawartość dokumentacji projektowej w formie Dokumentacji **SRS**:

- **Wprowadzenie** – 2,5 pkt
- **Cele systemu** – 2,5 pkt
- **Wymagania funkcjonalne** – 5 pkt
- **Wymagania niefunkcjonalne** – 5 pkt
- **Interfejsy użytkownika i wymagania dotyczące danych** – 2,5 pkt
- **Słownictwo dokumentacji** – 2,5 pkt
- **Przypadki użycia (use cases)** – 2,5 pkt
- **Scenariusze użytkownika (user stories)** – 2,5 pkt

Wymagania techniczne projektu

- Analiza danych tekstowych powinna wykorzystywać przynajmniej **jedną z omawianych technik**: np. analiza częstości słów, analiza sentymentu, klastrowanie, itp.
- Użytkownik powinien mieć możliwość **uruchomienia analizy na dostarczonych danych** (plikach).
- W projekcie powinny znaleźć się elementy **wizualizacji wyników** (np. wykresy ggplot2, chmury słów).
- Kod musi być **czytelnie udokumentowany** (komentarze, podział na sekcje).

Zalecenia organizacyjne

- Cały projekt powinien zostać umieszczony w repozytorium **GitHub** (link do repozytorium wkleimy w udostępniony google sheet)
- Projekt może być realizowany w zespole maksymalnie 3-osobowym (z zaznaczeniem zakresu pracy każdej osoby).
- Termin oddania projektu: **ostatnie zajęcia w semestrze.**