

Dokumentacja Specyfikacji Wymagań (SRS)

Projekt: Analiza Sentymentu w Czasie

Wersja dokumentu: 1.0

Data: 18.03.2025

Autor: [Janina Kowalska, Jan Kowalski]

1. Wprowadzenie:

Niniejszy dokument opisuje specyfikację wymagań dla skryptu R, który przeprowadza analizę sentymentu danych tekstowych w czasie. Skrypt wykorzystuje bibliotekę SentimentAnalysis oraz różne słowniki sentymentów (GI, HE, LM, QDAP) do analizy i wizualizacji ewolucji sentymentu w tekście.

2. Cele systemu:

- Przeprowadzenie analizy sentymentu tekstu z wykorzystaniem różnych słowników.
- Wizualizacja skumulowanego sentymentu dla każdego słownika.
- Porównanie sentymentu na podstawie różnych słowników.
- Analiza zmian sentymentu w czasie.

3. Wymagania funkcjonalne:

- **Wczytywanie danych:**
 - Skrypt powinien umożliwiać wczytanie danych tekstowych z lokalnego pliku .txt.
- **Analiza sentymentu:**
 - Skrypt powinien przeprowadzać analizę sentymentu tekstu z wykorzystaniem biblioteki SentimentAnalysis.
 - Skrypt powinien wykorzystywać słowniki sentymentów: GI, HE, LM, QDAP.
 - Skrypt powinien konwertować ciągłe wartości sentymentu na wartości kierunkowe.
- **Wizualizacja danych:**
 - Skrypt powinien generować wykresy skumulowanego sentymentu dla każdego słownika (wykresy domyślne i wykresy ggplot2).
 - Skrypt powinien generować wykres porównujący sentyment na podstawie różnych słowników (ggplot2).
 - Skrypt powinien generować wykresy przedstawiające ewolucję sentymentu w czasie (wykresy liniowe i wygładzone).
- **Agregacja danych:**
 - Skrypt powinien agregować sentyment z różnych słowników w jednej ramce danych.
 - Skrypt powinien usuwać brakujące wartości (NA).

4. Wymagania niefunkcjonalne:

- **Wydajność:**
 - Analiza pliku o długości 1000 zdań powinna trwać nie dłużej niż 10 sekund.
- **Bezpieczeństwo:**
 - System powinien zapewnić poprawność danych wyjściowych.
- **Niezawodność:**
 - Skrypt powinien poprawnie obsługiwać różne formaty danych tekstowych.
 - Skrypt powinien poprawnie obsługiwać brakujące wartości.
- **Użyteczność:**
 - Wykresy powinny być czytelne i zawierać odpowiednie etykiety.
- **Kompatybilność:**
 - Skrypt powinien być kompatybilny z R w wersji 4.0 lub nowszej.
 - Skrypt powinien korzystać z bibliotek SentimentAnalysis, ggplot2, ggthemes i tidyverse.

5. Interfejsy użytkownika:

- **Wejście:**
 - Plik tekstowy .txt.
- **Wyjście:**
 - Wykresy skumulowanego sentymentu (domyślne i ggplot2).
 - Wykres porównujący sentyment (ggplot2).
 - Wykresy ewolucji sentymentu w czasie (ggplot2).

6. Wymagania dotyczące danych:

- Skrypt zakłada, że dane tekstowe są w języku angielskim.
- Skrypt wykorzystuje słowniki sentymentów dostępne w pakiecie SentimentAnalysis.
- Skrypt nie obsługuje analizy sentymentu dla innych języków.
- Skrypt nie obsługuje analizy sentymentu dla danych tekstowych z innych źródeł niż pliki .txt.
- Skrypt nie obsługuje plików o rozmiarze powyżej 100 MB.

Słownictwo dokumentacji:

- **Sentyment:** emocjonalne nastawienie w tekście.
- **Słownik sentymentów:** lista słów i ich ocen wg sentymentu.
- **Skumulowany sentyment:** suma ocen sentymentu dla całego tekstu.
- **Wartości kierunkowe:** konwersja ciągłych wartości sentymentu na kategorie (np. pozytywny, negatywny, neutralny).
- **Ewolucja sentymentu:** zmiana sentymentu w czasie (wzdłuż czasu narracyjnego).

Przypadki użycia (use cases)

- Użytkownik:
 - wczytuje plik .txt.
 - uruchamia analizę
 - wyświetla wyniki
- Skrypt/system:
 - przetwarza tekst
 - analizuje sentyment tekstu
 - generuje wykresy skumulowanego sentymentu
 - generuje wykres porównujący sentyment
 - generuje wykresy ewolucji sentymentu

Testowe przypadki użycia:

- Test z plikiem .txt zawierającym tekst o pozytywnym sentymencie.
- Test z plikiem .txt zawierającym tekst o negatywnym sentymencie.
- Test z plikiem .txt zawierającym tekst o neutralnym sentymencie.
- Test z plikiem .txt zawierającym tekst o mieszanym sentymencie.
- Test z plikiem .txt zawierającym brakujące wartości.

Scenariusze użytkownika (user stories)

Scenariusz 1: Analiza opinii klientów o produkcie

- **Jako:** Analityk marketingowy
- **Chcę:** Przeanalizować opinie klientów o nowym produkcie z pliku tekstowego
- **Aby:** Zrozumieć ogólny sentyment klientów i zidentyfikować obszary, które wymagają poprawy.

Kryteria akceptacji:

- Użytkownik może wczytać plik tekstowy z opiniami klientów.
- Skrypt przeprowadza analizę sentymentu za pomocą różnych słowników.
- Skrypt generuje wykresy skumulowanego sentymentu i porównuje wyniki z różnych słowników.
- Skrypt generuje wykresy ewolucji sentymentu w czasie.
- Użytkownik może zidentyfikować ogólny sentyment klientów i obszary, które wymagają poprawy.

Scenariusz 2: Monitorowanie sentymentu w mediach społecznościowych

- **Jako:** Specjalista ds. mediów społecznościowych
- **Chcę:** Monitorować sentyment w mediach społecznościowych
- **Aby:** Reagować na negatywne opinie i wzmacniać pozytywny wizerunek marki.

Kryteria akceptacji:

- Użytkownik może wczytać dane tekstowe z mediów społecznościowych (np. z Twittera) do pliku tekstowego.
- Skrypt przeprowadza analizę sentymentu i generuje wykresy ewolucji sentymentu w czasie.
- Użytkownik może monitorować zmiany sentymentu.
- Użytkownik może identyfikować nagłe zmiany sentymentu i reagować na nie.
- Użytkownik może generować raporty z analizy sentymentu.