

Dokumentacja Specyfikacji Wymagań

Projekt: Zależność między zabarwieniem emocjonalnym tekstów, a ich popularnością

Wersja dokumentu: 1.0

Data: 08.06.2025

Autor: Anastasiia Mashchenko, Oliwia Turkowska

1. Wprowadzenie

Niniejszy dokument opisuje specyfikację dla skryptu R, który przeprowadza analizę między zabarwieniem emocjonalnym tekstów, a ich popularnością. Miarą popularności jest liczba like'ów oraz komentarzy. Skrypt wykorzystuje bibliotekę SentimentAnalysis, w tym słownik QDAP do analizy i wizualizacji sentymentu w tekście oraz korelację do przedstawienia zależności między zabarwieniem emocjonalnym tekstów, a liczbą polubień i komentarzy.

2. Cele systemu

- Przeprowadzenie analizy sentymentu z wykorzystaniem słownika QDAP.
- Wizualizacja sentymentu dla słownika.
- Wyliczenie korelacji.
- Wizualizacja korelacji między sentymentem danego tekstu, a liczbą polubień i komentarzy.
- Wyliczenie uśrednionych danych.

3. Wymagania funkcjonalne

- **Wczytywanie danych**
 - Skrypt powinien umożliwiać wczytanie danych z lokalnego pliku .csv zawierających treści tekstów oraz dane - liczba polubień oraz komentarzy.
- **Analiza sentymentu**
 - Skrypt powinien przeprowadzać analizę sentymentu tekstu z wykorzystaniem biblioteki SentimentAnalysis.
 - Skrypt powinien wykorzystywać słownik QDAP.
- **Wizualizacja danych**
 - Skrypt powinien generować wykresy sentymentu dla słownika.
 - Skrypt powinien generować wykresy przedstawiające korelację między sentymentem, a liczbą polubień oraz komentarzy (osobno).

4. Wymagania niefunkcjonalne

- **Wydajność**
 - Analiza pliku o długości 1000 zdań powinna trwać nie dłużej niż 10 sekund.
- **Bezpieczeństwo**
 - System powinien zapewniać poprawność danych wyjściowych.

- **Niezawodność**
 - Skrypt powinien poprawnie obsługiwać brakujące wartości (np. brak popularności lub tekstu)
- **Użyteczność**
 - Wszystkie wykresy powinny być czytelne oraz zawierać odpowiednie etykiety.
- **Kompatybilność**
 - Skrypt powinien być kompatybilny z R w wersji 4.0 lub nowszej.

5. Interfejs użytkownika

- **Wejście**
 - Plik .csv
- **Wyjście**
 - Wykres sentymentu (ggplot2)
 - Wykres korelacji sentymentu z liczbą polubień (ggplot2)
 - Wykres korelacji sentymentu z liczbą komentarzy (ggplot2)
 - Dane dotyczące uśrednionego sentymentu oraz korelacji z polubieniami i komentarzami.

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 .csv.
- Skrypt nie obsługuje plików o rozmiarze powyżej 100 MB.

7. Słownictwo dokumentacji

- **Sentyment:** emocjonalne nastawienie w tekście.
- **Słownik sentymentów:** lista słów i ich ocen wg sentymentu.
- **Korelacja:** statystyczna relacji między dwiema zmiennymi.
- **Popularność:** liczba like'ów oraz komentarzy.

8. Przypadki użycia (use cases)

- **Użytkownik:**
 - wczytuje plik .csv.
 - uruchamia analizę
 - wyświetla wyniki
- **Skrypt/system:**
 - przetwarza tekst
 - analizuje sentyment tekstu
 - generuje wykresy sentymentu
 - generuje wykres korelacji

9. Scenariusze użytkownika (user stories)

Scenariusz 1: Śledzenie popularności w mediach społecznościowych

Jako: Specjalista ds. rozwoju w mediach społecznościowych

Chcę: Sprawdzić jakie treści bardziej wpływają na użytkowników oraz na ich aktywność pod określonymi postami.

Aby: Rozwijać popularność w mediach społecznościowych.

Kryteria akceptacji:

- Użytkownik może wczytać wpisy z mediów społecznościowych (np. z Twittera) oraz liczby like'ów i komentarzy do pliku .csv.
- Skrypt przeprowadza analizę sentymentu i korelacji oraz generuje wykresy.
- Użytkownik może sprawdzać jakie wpisy powodują zwiększoną aktywność obserwujących.
- Użytkownik może dostosować nastawienie emocjonalne wpisów na mediach społecznościowych, aby były chętniej wyświetlane przez obserwujących.

Scenariusz 2: Analiza recenzji filmów.

Jako: Specjalista ds. zwiększenia sprzedaży biletów

Chcę: Sprawdzić jakie recenzje przyciągają więcej uwagi.

Aby: Określić jakie treści zwiększają popularność danego filmu.

Kryteria akceptacji:

- Użytkownik może wczytać recenzje filmów oraz liczbę like'ów i komentarzy pod nimi.
- Skrypt przeprowadza analizę sentymentu i korelacji oraz generuje wykresy.
- Użytkownik może sprawdzić zależność między pozytywnym lub negatywnym sentymentem, a popularnością danego filmu.
- Użytkownik może dostosować nastawienie emocjonalne recenzji, aby były częściej czytane przez klientów.