Multimedialna baza danych tekstów z identyfikacja plagiatów.

Algorytm identyfikacji plagiatów składać się będzie z X sprawdzeń pomiędzy tekstami. Sprawdzenia te będą kolejnymi algorytmami wykorzystującymi różne kryteria oceny. Ogólny algorytm sterujący będzie korzystał z algorytmów sprawdzających. Następnie z cząstkowych ocen, wygenerowanych przez algorytmy wyliczał będzie wynik końcowy. Każde ze sprawdzeń będzie miało przypisaną odpowiednią wagę.

Wynik pełnego algorytmu:

1. Algorytm: Sprawdzenie powtarzania się pełnych zdań.

Kroki:

* Podział tekstów na zdania
* Konwersja liter na małe
* Usuniecie białych znaków na początku i końcu zdania
* Porównanie każdego zdanie z tekstu z każdym zdaniem z tekstu 2.
* Zanotowanie każdego powtarzającego się zdania bez duplikatów.

Waga do wyliczania wyniku końcowego: 3

Sposób wyliczania wyniku:

1. Algorytm: Sprawdzenie powtarzania się sekwencji X słów w jednym zdaniu

Kroki:

* Podział tekstów na zdania.
* Odfiltrowanie zdań które mają mniej niż X słów.
* Konwersja liter na małe
* Podział każdego zdania na słowa z zachowaniem kolejności
* Usuniecie białych znaków na początku i końcu słowa
* Porównanie słów każdego zdania z tekstu 1 ze słowami z każdego zdania z tekstu 2 z zachowaniem kolejności.
* Zanotowanie każdej powtarzającego się X słów w zdaniach z tekstu 1 i tekstu 2 bez duplikatów.

Sposób wyliczania wyniku:

Waga do wyliczania wyniku końcowego:

1. Algorytm: Sprawdzenie powtarzania się X słów w jednym zdaniu bez względu na

Kroki:

* Podział tekstów na zdania.
* Odfiltrowanie zdań które mają mniej niż X słów.
* Konwersja liter na małe
* Podział każdego zdania na słowa
* Usuniecie białych znaków na początku i końcu słowa
* Porównanie słów każdego zdania z tekstu 1 ze słowami z każdego zdania z tekstu 2
* Zanotowanie każdej powtarzającego się X słów w zdaniach z tekstu 1 i tekstu 2 bez duplikatów.

Sposób wyliczania wyniku:

Waga do wyliczania wyniku końcowego:

1. Ilość liter w tekstach.
2. Ilość zdań w tekstach.
3. Ilość słów w tekście.