

## PROJEKT

### „System antyplagiatowy.”

Cel projektu :

Stworzenie oprogramowania wspomagającego wykrywanie plagiatów w dokumentach tekstowych (.doc) i kodach źródłowych aplikacji.

Wymagania funkcjonalne i niefunkcjonalne :

1. System działa dla przynajmniej 40 prac w bazie.
2. Złożoność obliczeń nie jest traktowana priorytetowo.
3. System działa dla plików .doc i plików zawierających kod źródłowy.
4. System generuje raport z wynikami algorytmu , wskazując miejsca w których istnieje podejrzenie plagiatu.
5. Interfejs powinien być jak najbardziej intuicyjny/prosty.

Diagram sekwencji :

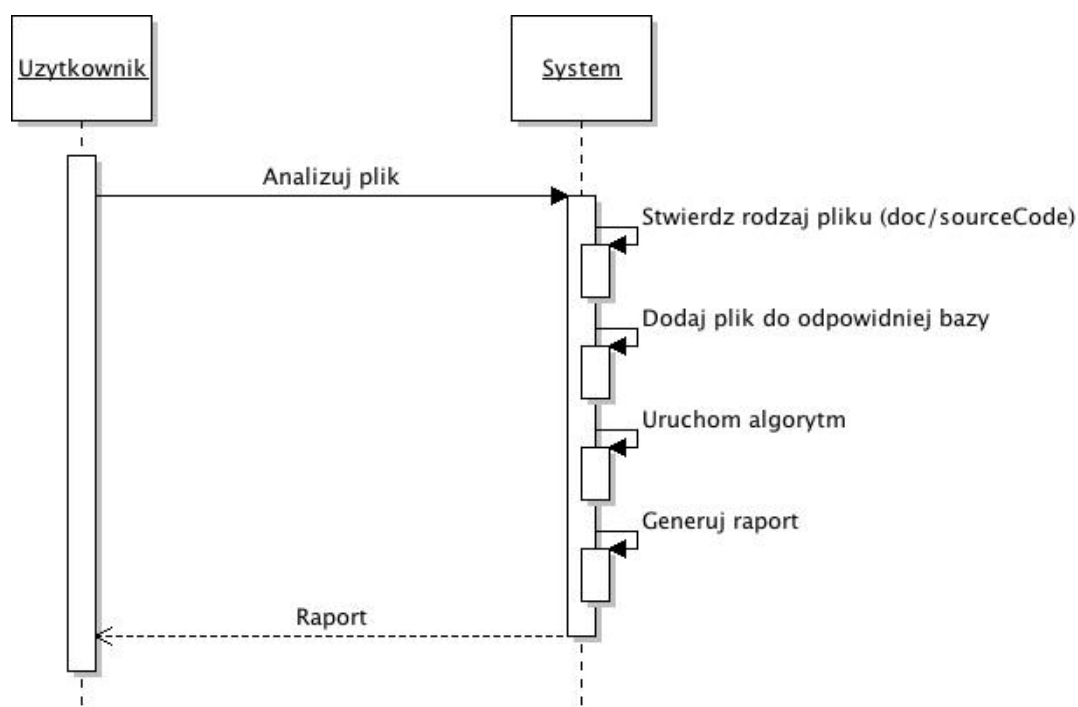


Diagram przypadków użycia :



Algorytm dla dokumentacji :

1. Podział tekstu na paragrafy.
2. Usunięcie cytatów , znaków interpunkcyjnych i przyimków. Konwersja tekstu do małych liter. Zapisanie przetworzonego dokumentu w bazie.
3. Obliczenie częstotliwości występowania każdego ze słów w paragrafie.
4. Wybranie ~20% najczęściej powtarzających się słów oraz ~20% najrzadziej – tworząc słownik.
5. Szukamy zestawu słów (punkt powyżej) we wszystkich słownikach prac w bazie.

Jeżeli powyższy algorytm da wynik pozytywny (podejrzenie plagiatu) uruchamiamy algorytm Karpa-Rabina dla losowych wzorców (po 4-5 słów w sekwencji).

Algorytm dla kodu:

1. Konwersja białych znaków na pojedyncze spacje.
2. Usunięcie komentarze.
3. Tworzymy słownik zawierający ilość wystąpień każdej zmiennej , funkcji + lista załączonych bibliotek.
4. Sprawdzamy podobieństwo użytych bibliotek +/- 10%.
5. Sprawdzamy czy któryś słownik w bazie ma podobną ilość słów ze słownika stworzonego w punkcie 3 , +/- 5%.