

***Ewa Nowakowska***

***106565***

**Informatyka, studia stacjonarne I stopnia, semestr 5**

Projekt z przedmiotu Wybrane zagadnienia współczesnej informatyki pt.:

***Aplikacja webowa do akwizycji i analizy aktualności ze strony Kolegium Nauk Społecznych Uniwersytetu Rzeszowskiego.***

**Rzeszów, 2021 r.**

*Link do repozytorium:* [*https://github.com/EwkaN/WZWI\_Projekt*](https://github.com/EwkaN/WZWI_Projekt)

**1. Zakres projektu.**

Zakresem projektu jest aplikacja webowa z zaawansowanym GUI (możliwość ustawiania parametrów) zrealizowanym z wykorzystaniem pakietu Shiny, możliwością pobierania wiadomości i magazynowania ich w Apache Solr, analizy częstości występowania słów i wizualizacji chmury słów w wiadomościach zmagazynowanych w Apache Solr.

**2. Charakterystyka analizowanych danych.**

Dane użyte w projekcie to treść aktualności ze strony <https://www.ur.edu.pl/kolegia/kolegium-nauk-spolecznych/>. Dane pobierane są za pomocą funkcji read\_html, oraz przesyłane są do SOLRa, a następnie pobierane. Usuwane są znaki interpunkcyjne oraz tzw. “stop words”. Następnie dane poddawane są lematyzacji.

**3. Wykorzystane narzędzia.**

Projekt został stworzony w RStudio z wykorzystaniem języka R i pakietu Shiny, wiadomości w projekcie magazynowane są za pomocą Apache Solr.

Następnie generowana jest chmura słów, do tej czyności posłużył pakiet ‘wordcloud2’.

**4. Architektura i implementacja aplikacji.**

Projekt pobiera dane w formacie html za pomocą funkcji read\_html i wyodrębnia div z klasą news-content z aktualności, a następnie zapisuje ją do odpowiednie z pobranych z SOLR danych usuwane są znaki interpunkcyjne oraz stop words. Po przygotowany w opisany sposób danych są one lematyzowane odpowiednią funkcją z wykorzystaniem <http://ws.clarin-pl.eu/nlprest2/base/process>. Za pomocą funkcji TermDocumentMatrix() tworzona jest macierz ze słowami, z których generowana będzie chmura.

**5. Interfejs użytkownika.**

Użytkownik może ustawiać takie wartości na jakie tylko ma ochotę, poprzez kształt, rozmiar chmury oraz minimalną częstotliwość występowania słowa.



**6. Podsumowanie.**

RStudio umożliwia wykonanie aplikacji webowej Shiny generującej chmurę słów za pomocą pakietu ‘wordcloud2’. Aplikacja umożliwia użytkownikowi zobaczenie w ciekawy sposób jak pobierane są dane i układane w konkretne kształty.