# Filmy - faza 2

Justyna Jankowiak, Emilia Momotko, Adrianna Sudoł

### Podstawowe informacje:

Liczba zebranych filmów:

## [1] 124706

Wielkość zebranych danych: 1172 MB

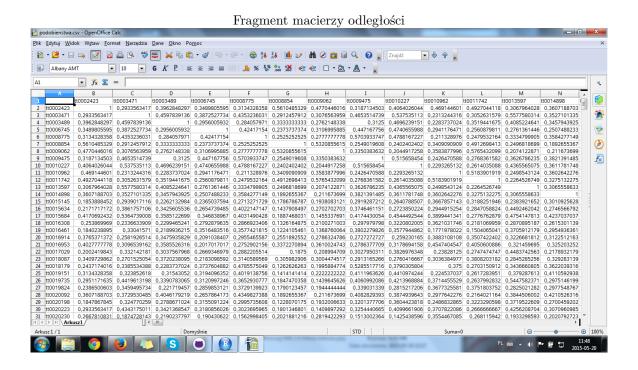
Szczegółowe opisy funkcji można znaleźć w pakiecie FilmyJMS dostępnym na https://github.com/tyskaj/Filmy/tree/master/filmy pakiet

#### Wskaźniki:

- frakcja wskaźnik podobieństwa między dwoma pojedynczymi wektorami tekstowymi w którym słowa są oddzielone separatorem '@'.
- frakcja\_powtarzalnych\_slow wskaźnik podobieństwa między dwoma tekstami
- czas funkcja sprawdzająca czy dwa czasy filmów zawierają się w tych samych przedziałach.
- rok funkcja sprawdzająca czy lata wydania dwóch filmów zawierają się w tych samych przedziałach.
- porownaj\_mediany sprawdzenie czy dwie wartości są do siebie odpowiednio bliskie (na podstawie mediany).
- porownaj\_rozrzut funkcja wyznacza wskaźnik podobieństwa między dwiema wartościami rozrzutu.
- czy zawiera wskaźnik sprawdzający czy jeden element zawiera się w drugim.
- roznica funkcja służąca do oceny czy dany film ma podobną ocenę, liczbę użytkowników oceniających oraz liczbę recenzji. W zależności od tego do czego ją stosujemy należy zmienić wartość param, która określa dopuszczalną różnicę między dwoma wartościami. Dla oceny jest to 0.3, liczby użytkowników: 50000, liczby recenzji: 20.
- czy to samo wskaźnik określający czy porównywane filmy otrzymały Oscary bądź inne nagrody.

## Macierz odległości

Macierz została zbudowana za pomocą funkcji tworz\_macierz oraz odleglosci z pakietu FilmyJMS w której użyto w różnych kombinacjach wyżej wymienione wskaźniki. W sumie badamy podobieństwa na podstawie 21 różnych cech filmu, które można odpowiednio ważyć pod względem ich istotności, w zależności od preferencji użytkownika.



### Mapa filmów

