Wizualizacja treści zamieszczanych na profilu facebookowym iitv.info

Marcin Kosiński kosinskim@student.mini.pw.edu.pl



Wizualizacji danych dokonano dzięki narzędziu IBM MANY EYES. Obróbki danych dokonano przy pomocy pakietu statystycznego \mathcal{R} .

Dane przedstawione na wizualizacjach zostały wczytane z konkursowego zbioru danych wypowiedzi-FB, z którego wykorzystano jedynie kolumnę text. Zatem przedstawiona wizualizacja dotyczy jedynie treści publikowanych przez internautów i moderatorów na profilu iitv.info założonym na popularnym portalu społecznościowym facebook.

Aby uniknąć ewentualnego problemu ze SPAMem, z komentarzy wybrano tylko unikatowe komentarze, co oznacza, że jeżeli ta sama treść pojawiła się pod wieloma komentarzami lub wpisami, to została uwzględniona do wizualizacji tylko raz.

Spowodowało to, że baza danych została pomniejszona z około 46 tys wierszy do 11 582 wierszy. Tyle dokładnie było unikatowych komentarzy i wpisów na tym profilu.



Powyższa chmura tagów, jak także ta z tytułowej strony, ukazują najczęściej pojawiające się słowa w komentarzach i wpisach publikowanych na profilu iitv.info. Widać, że najczęściej wpisy dotyczą słów takich jak: **odcinek, napisy, sezon** oraz **serial**, a także ich typowych odmian.

Zbiór danych po zmniejszeniu do 11 582 wierszy składał się z **178 357** słów i znaków przestankowych. Z tej liczby wybrano **150** najczęściej pojawiających się oraz przedstawiona na wspomnianych dwóch chmurach tagów. Im słowo częściej pojawiało się na profilu, tym większe jest w stosunku do innych na chmurze tagów.

Wizualizacja skupiająca uwagę na **50** najważniejszych, czyli najczęściej występujących, słowach znajduje się poniżej.



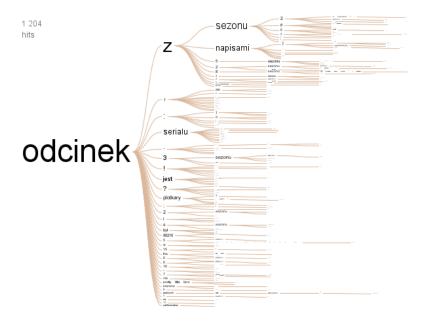
Z kolei 20 najczęściej wspominanych słów prezentuje się tak.



Wiedząc jakie słowa są najczęściej używane przez internautów na tym profilu, można przebadać kontekst, w jakim były one używane. Można sprawdzić jakie słowa je poprzedzały i jakie słowa po nich następowały.

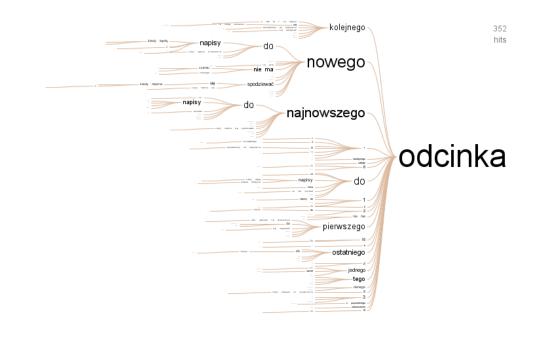


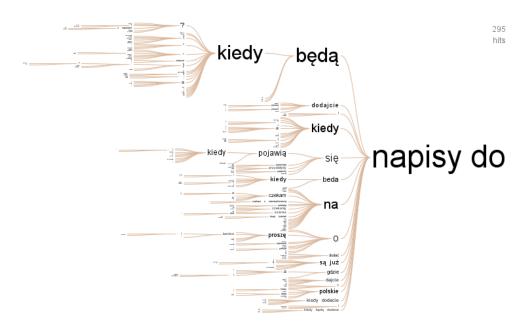
Poniżej prezentacja fraz następujących po słowach, które pojawiały się najczęściej - napisy do, odcinek.





Poniżej prezentacja fraz poprzedzających słowach, które pojawiały się najczęściej - napisy do, odcinek.





Tak stworzone drzewa słów można podziwiać i interaktywnie modyfikować wedle uznania na moim profilu na stornie IBM.