

Raport - wybory 2015

Adrianna Sudol i Marcin Rdzanowski

14 maja 2015

Dane:

Ilość zebranych różnych tekstów o kandydatach biorących udział w wyborach prezydenckich 2015:

```
## [1] 109294
```

Dla poszczególnych portali liczba zebranych danych jest następująca:

##	portal	liczba_artykulow
## 1	onet.pl	344
## 2	onet.pl/wybory	209
## 3	wp.pl	265
## 4	newsweek.pl	1235
## 5	gazeta.pl	262
## 6	tvn24.pl	1278
## 7	tvn24.pl/wybory	3480
## 8	tvpinfo.pl	400
## 9	twitter-kandydat	42412
## 10	twitter-nasluch	59409

Objętość zebranych danych spakowanych do pliku .zip wynosi:

```
## 156.99 MB
```

Dane były zbierane w dniach: od 16.03.2015 do 11.05.2015.

Po uporządkowaniu i wyczyszczeniu zebranych danych wszystkie zostały zapisane w ramce danych o nazwie *wybory2015*, którą można znaleźć w pakiecie **wsk** (<https://github.com/sudola>)

Opis ramki danych:

- *tekst* - treść artykułu lub tweeta,
- *portal* - nazwa portalu z którego są pobierane dane,
- *data* - data dnia w którym pojawił się artykuł lub tweet,
- *waga* - zmienna 0-1 czy dany artykuł był głównym tematem czy nie (tweety wszystkie mają automatycznie wagę 1)
- *l_retweet* - liczba kolejnych 'podaj dalej' tweetów (dla artykułów NA)
- *l_polubien* - liczba polubień tweeta (dla artykułów NA)
- *komorowski* - kolumna 0-1 czy dany tweet lub artykuł jest o Bronisławie Komorowskim (1) czy nie (0),
- *duda* - kolumna 0-1 czy dany tweet lub artykuł jest o Andrzeju Dudzie (1) czy nie (0),

- *jarubas* - kolumna 0-1 czy dany tweet lub artykuł jest o Adamie Jarubasie (1) czy nie (0),
- *ogorek* - kolumna 0-1 czy dany tweet lub artykuł jest o Magdalenie Ogórek (1) czy nie (0),
- *palikot* - kolumna 0-1 czy dany tweet lub artykuł jest o Januszu Palikocie (1) czy nie (0),
- *korwin* - kolumna 0-1 czy dany tweet lub artykuł jest o Januszu Korwin - Mikke (1) czy nie (0),
- *kukiz* - kolumna 0-1 czy dany tweet lub artykuł jest o Pawle Kukizie (1) czy nie (0),
- *sentyment_1* - sentyment wyliczony dla danego tekstu na podstawie słownika 1
- *sentyment_2* - sentyment wyliczony dla danego tekstu na podstawie słownika 2

Wskaźniki:

Dostępne wskaźniki są następujące:

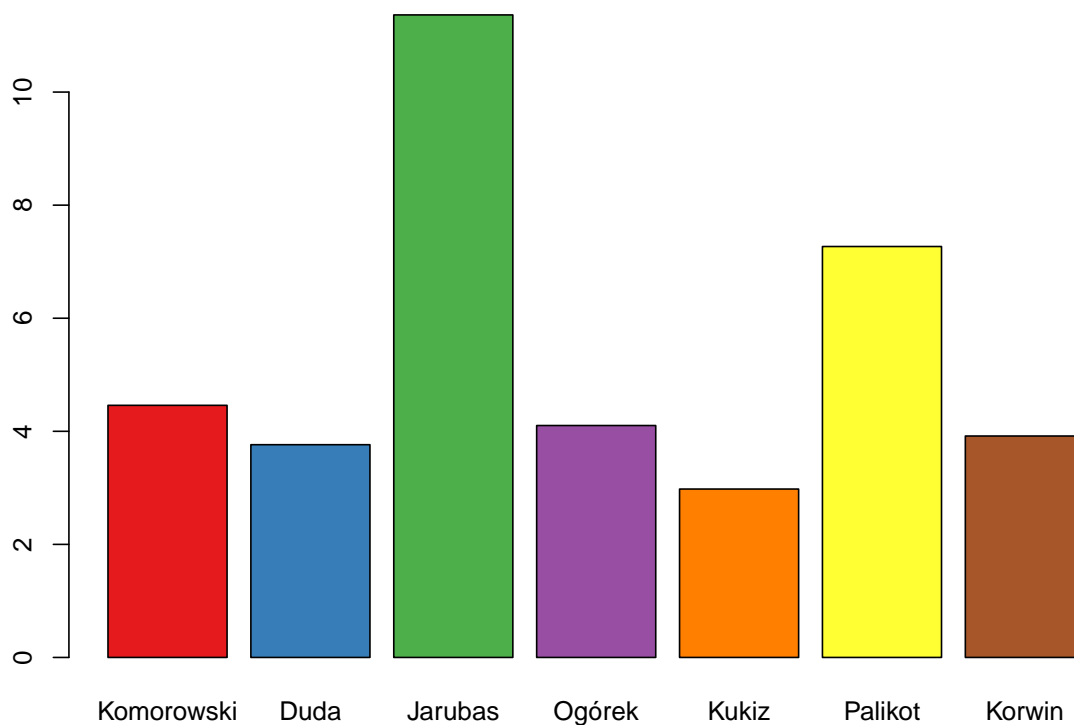
- *poz_do_neg* - stosunek artykułów o wydźwięku pozytywnym do tych o wydźwięku negatywnym,
- *pstwo_neg* - prawdopodobieństwo wystąpienia negatywnego tekstu o kandydacie w mediach,
- *pstwo_neu* - prawdopodobieństwo wystąpienia neutralnego tekstu o kandydacie w mediach,
- *pstwo_poz* - prawdopodobieństwo wystąpienia pozytywnego tekstu o kandydacie w mediach,
- *pstwo_widoczności* - prawdopodobieństwo widoczności kandydata w mediach,
- *szansa_neg* - szansa wystąpienia tekstu o wydźwięku negatywnym,
- *szansa_poz* - szansa wystąpienia tekstu o wydźwięku pozytywnym,
- *szansa_neu* - szansa wystąpienia tekstu o wydźwięku neutralnym,
- *twitt_poz_neg* - różnica średniej liczby ‘retweetów’ (‘polubień’) pozytywnych i negatywnych.

Ponadto każdy ze wskaźników można zastosować osobno dla każdego z kandydatów, osobno dla każdego portalu oraz dla osobnej daty oraz ich kombinacje.

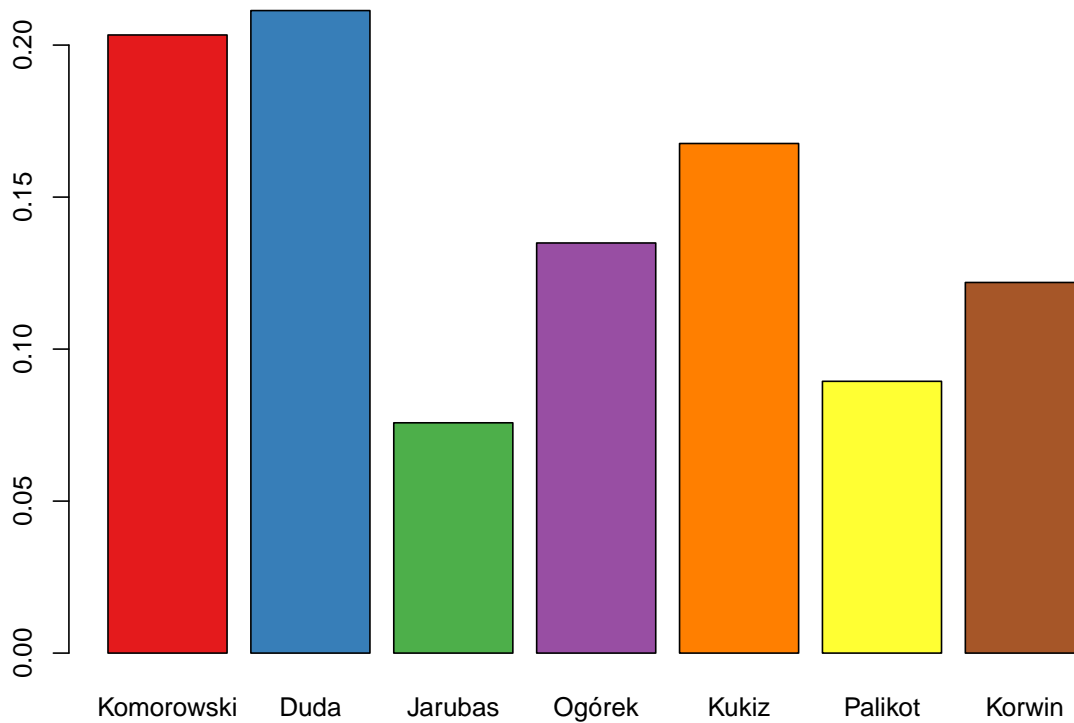
Wizualizacja danych:

Analiza sentymentu oparta jest o dwa różne słowniki: *słownik1*, *słownik2*, których opis można znaleźć w pakiecie **wsk**. Wykresy dla zależności stosunku pozytywnych do negatywnych tekstów na temat danego kandydata:

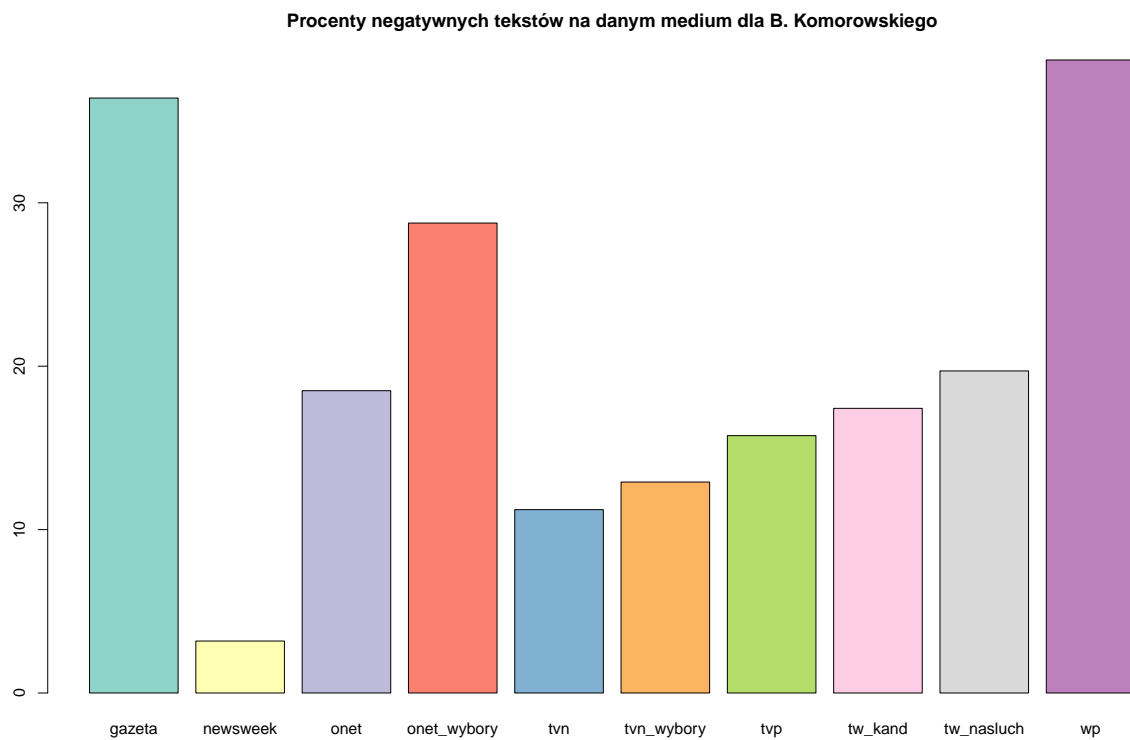
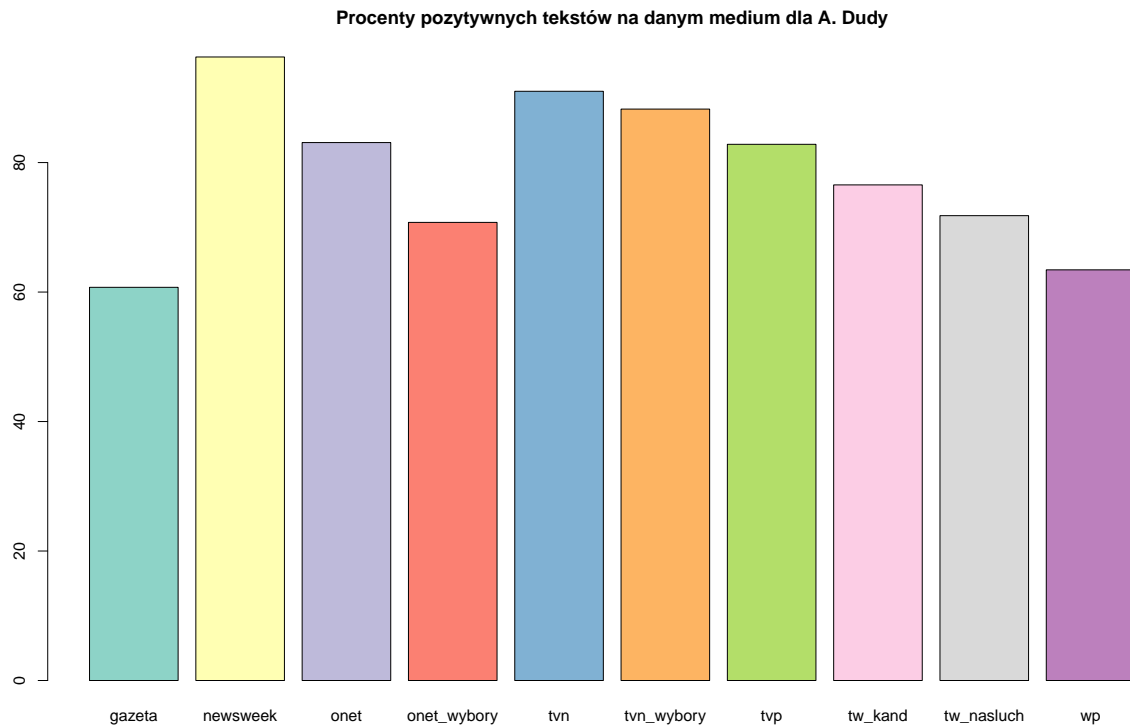
Stosunek tekstów pozytywnych do negatywnych dla sentymentu 1



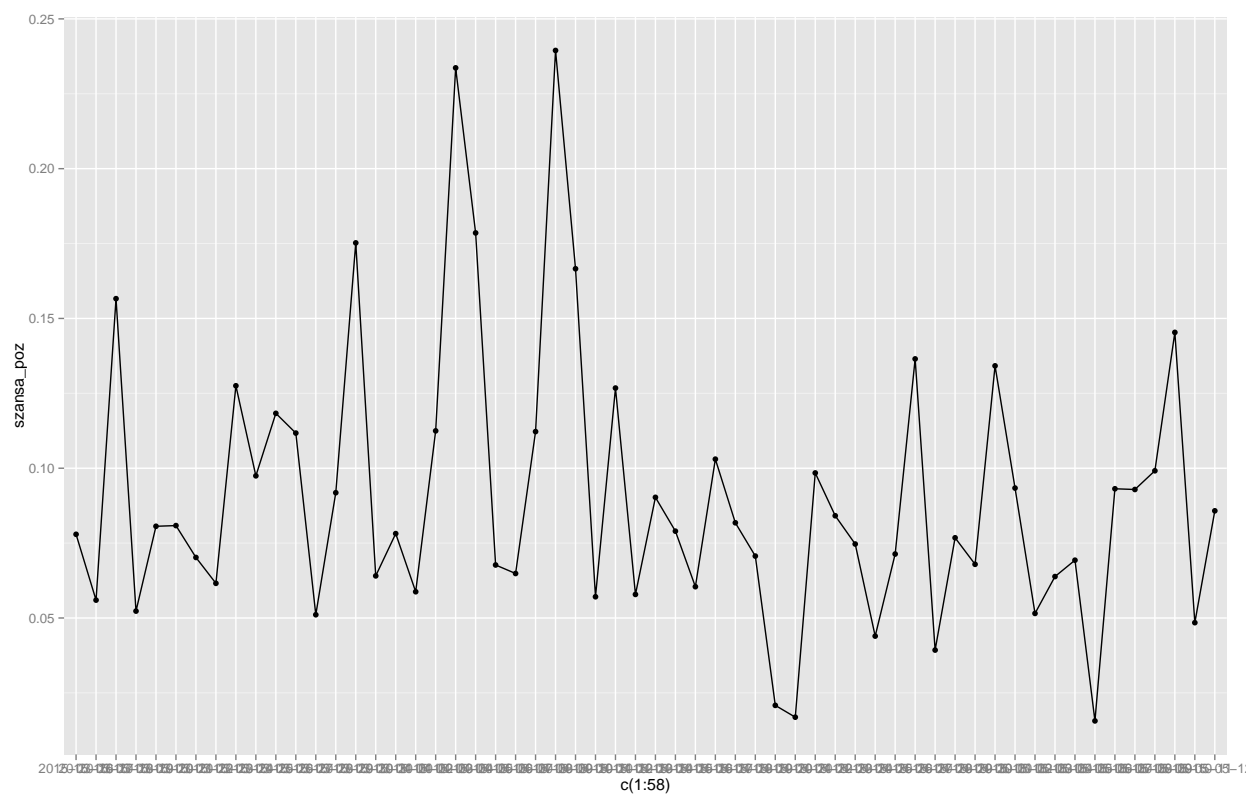
Stosunek tekstów pozytywnych do negatywnych dla sentymentu 2



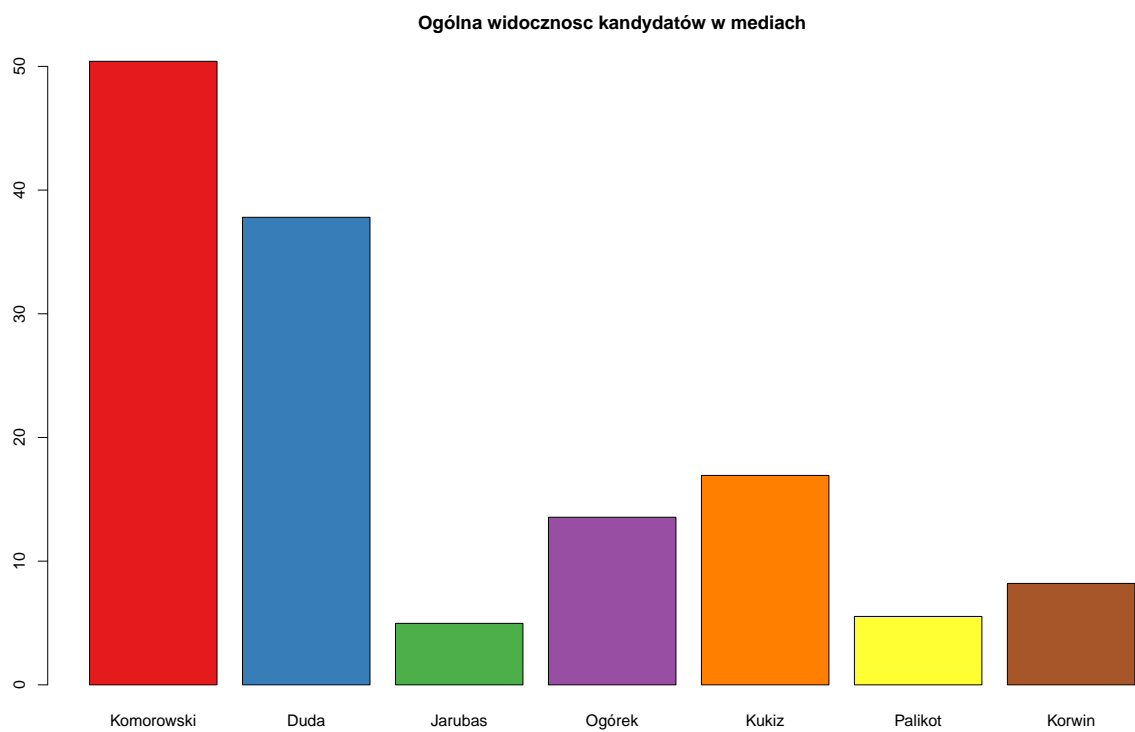
Wykresy przedstawiające procent pozytywnych, negatywnych i neutralnych tekstów:



Szansa pozytywnych artykułów dla Magdaleny Ogórek na przełomie całego okresu zbierania danych:



Ogólna widoczność kandydatów:



Wykres zależności obecności Pawła Kukiza w mediach na przełomie pełnego okresu zbierania danych:

