

## Porównanie wyników metod BOW.

Dla przygotowanych danych po zastosowaniu samego stemmingu, samej lematyzacji, kolejność: lematyzacja, stemming, kolejność: stemming, lematyzacja. Najlepsze wyniki były dla zastosowania lematyzacji i stemmingu w kolejności lematyzacja, stemming.

Stworzyliśmy dwa sposoby liczenia metody BOW (liczyliśmy dla zbioru treningowego i testowego oddzielnie):

1.  $(\sum \frac{a}{b})/c$ ;

- a - ilość słów, które powtarzają się w przewidzianym opisie, z tymi które należały do worka stworzonego dla zdań treningowych całego obrazka;
- b - ilość słów zdania przewidzianego;
- c - ilość obrazów zbioru treningowego/testowego

2. Policzyliśmy ilość powtórzeń w zdaniu przewidywanym do ilości powtórzeń z worka treningowego danego obrazka i dla wszystkich innych worków zbioru. Sprawdzaliśmy najlepszy wynik, a potem porównywaliśmy go z wynikiem dla zbioru treningowego danego obrazka. Liczyliśmy ile wyników zgadza się.

zdanie 1, worek treningowy 1, worek treningowy 2, ..., worek treningowy n

$$\frac{\text{zdanie 1}}{\text{worek treningowy 1}}, \dots, \frac{\text{zdanie 1}}{\text{worek treningowy 1}} \rightarrow \max$$

sprawdzamy, czy  $\max == \frac{\text{zdanie 1}}{\text{worek treningowy 1}}$  (liczymy ile jest takich sytuacji w zbiorze)

ilość tych sytuacji dzielimy przez ilość obrazów zbioru treningowego/testowego

b. jeśli  $\frac{\text{zdanie 1}}{\text{worek treningowy 1}}$  jest 0, 1, 2 lub 3 na uszeregowanej liście najlepszych wyników również dodajemy go do poprawnej sytuacji

c. jeśli  $\frac{\text{zdanie 1}}{\text{worek treningowy 1}}$  jest 0, 1 lub 2 na uszeregowanej liście najlepszych wyników również dodajemy go do poprawnej sytuacji

d. jeśli  $\frac{\text{zdanie 1}}{\text{worek treningowy 1}}$  jest w 40% najlepszych wyników i jest większy= niż 0.5 na uszeregowanej liście najlepszych wyników również dodajemy go do poprawnej sytuacji

e. jeśli  $\frac{\text{zdanie 1}}{\text{worek treningowy 1}}$  jest w 30% najlepszych wyników i jest większy= niż 0.5 na uszeregowanej liście najlepszych wyników również dodajemy go do poprawnej sytuacji

	Stemming	lematyzacja	ls	sl
BOW 2 train	0.431	0.4211666666666667	0.4321666666666666	0.432
BOW 2b train	0.4411(6)	0.4303333333333333	0.4428333333333333	0.4426666666666666
BOW 2c train	0.4391(6)	0.429	0.4408333333333333	0.4406666666666666
BOW 2d train			0.5556666666666666	
BOW 2e train			0.5006666666666666	
BOW 1 train	0.4594	0.44279497955747954	0.46224704184704185	0.4622887085137085
BOW 2 test	0.262	0.3	0.3	0.264
BOW 2b train	0.293		0.397	0.295
BOW 2c test				
BOW 2d test			0.335	
BOW 2e test			0.306	
BOW 1 test	0.3974624098124098	0.40107590187590186	0.40121875901875903	0.40107590187590186

