

Pytania otwarte.....pkt,
Pytania zamknięte.....pkt
Razem..... pkt

Imię i nazwisko.....
Numer indeksu.....

K O Ł O K W I U M

Eksploracja Tekstu i Wyszukiwanie Informacji w Mediach Społecznościowych
3 grudnia 2018 r.

Pytania otwarte (5 pkt. każde):

1. Opisz znane prawa lingwistyczne ze szczególnym uwzględnieniem prawa Zipfa.
2. Omów wybrany algorytm analizy sentymentu bez nadzoru.
3. Omów model przestrzeni wektorowej.

Pytania zamknięte (0.5 pkt. każde, jedna odpowiedź prawidłowa)

Zad 1. Jaki proces służy do modelowania dystrybucji rozkładu klastrów emocjonalnych w np. danych BBC Forum:

- (a) losowy i.i.d.,
- (b) związany z potęgowym wzrostem klastrów,
- (c) Markowa z jednokrokową pamięcią.

Zad 2. Pierwszy proces zamienił formy „washed” i „washing” na „wash”, a „baby” na „babi”, natomiast drugi formy „was”, „are” na „to be”.

- (a) pierwszy proces to lematyzacja,
- (b) drugi proces do lematyzacja,
- (c) oba procesy to stemowanie.

Zad 3. Rozpatrujemy **jeden** dokument. Czy IDF dla dowolnego wyrazu:

- (a) może przyjmować dowolną wartość,
- (b) wynosi 1,
- (c) wynosi 0.

Zad 4. Jaką w ogólności metodę wykorzystuje model tematyczny LSA do rozdziału tekstu na tematy i przypisaniu słów do poszczególnych tematów

- (a) PCA – analizę składowych głównych ,
- (b) SVM – maszynę wektorów nośnych,
- (c) SVD – rozkładu macierzy na wartości osobliwe.

Zad 5. Ile 3-gramów uzyskamy ze zdania „Piszę właśnie proste kolokwium”?

- (a) 2 ,
- (b) 3,
- (c) 1.