

MAY

2020

WEEK 20

13

WEDNESDAY

DAY 134-232

Q4 TF-IDF score

Document 1: study passing algorithm article  
NLP blogDocument 2: study pos tagging article NLP  
blog

| words     | TF (doc1) | TF (doc2) | IDF             | TF-IDF                  | TF-IDF |
|-----------|-----------|-----------|-----------------|-------------------------|--------|
| study     | 1/6       | 1/6       | $\log(2) = 0.3$ | $0.3 \times 1/6 = 0.05$ | 0      |
| passing   | 1/6       | 0         | $\log(2) = 0.3$ | $0.3 \times 1/6 = 0.05$ | 0      |
| algorithm | 1/6       | 0         | $\log(2) = 0.3$ | $0.3 \times 1/6 = 0.05$ | 0      |
| article   | 1/6       | 1/6       | $\log(2) = 0.3$ | 0                       | 0      |
| NLP       | 1/6       | 1/6       | $\log(4) = 2$   | 0                       | 0      |
| blog      | 1/6       | 1/6       | $\log(4) = 2$   | 0                       | 0      |
| pos       | 0         | 1/6       | $\log(2) = 0.3$ | 0                       | 0.05   |
| tagging   | 0         | 1/6       | $\log(2) = 0.3$ | 0                       | 0.05   |

$$TF-IDF (doc1) = [0 \ 0.05 \ 0.05 \ 0 \ 0 \ 0 \ 0 \ 0]$$

$$TF-IDF (doc2) = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0.05 \ 0.05]$$

Cosine similarity b/w them

$$= 0$$

M T W T F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S

• • • 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31