Jongho Jung
Nandita Jha

1	2	3	Σ

Assignment 10

(Due: 30.01.2023)

Task 1 Word Cloud

(a) Impl

(b) Impl

(c) The smaller the slope and step size, the tighter the word cloud is built, but takes longer. To mitigate this, we can use a spatial index like KD-TREE or remember the last collision and recheck it after moving.

Task 2 Bag-of-words Model

(a)
- (240 a and at based calendar cluster data diagnosis dynamics gait in laboratory of patients series tests the time visualization)

1. (1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 2, 1, 1, 1, 1, 1, 0)
2. (0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1)
3. (0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 2, 1, 1, 0, 0, 1, 1)

(b) 1.-2. 0.3442651863295481

2.-3. 0.502518907629606

3.-1. 0.6227991553292184

The more same words a sentence has, the greater the similarity.

(c)
- (240 based calendar cluster data diagnosis dynamics gait laboratory patients series tests time visualization)

1. (1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0)
2. (0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1)
3. (0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1)

1.-2. 0.2857142857142857

2.-3. 0.4629100498862757

3.-1. 0.4629100498862757

By removing all stopwords, the similarity is low. Eliminating stopwords in the bag-of-words model, can more accurately represent the similarity of semantic elements in sentences.

Task 3 TF-IDF

(a) The words with the highest term frequency are stopwords such as a or is.

(b)
- a 0.98
- congress 11.51
- elections 10.10
- freedom 11.51

- government 17.32
- is 6.48
- machu 3.91
- people 17.61
- resignation 8.44
- ruins 3.91
- saturday 13.76
- the 0

The highest score is for a word that is used significantly and frequently. The lowest score is for simple frequently used words such as stopwords.

```
const data = {
    a: 49,
    congress: 5,
    elections: 4,
    freedom: 5,
    government: 25,
    is: 43,
    machu: 1,
    people: 24,
    resignation: 3,
    ruins: 1,
    saturday: 7,
    the: 50,
};
const values = Object.values(data);
const N = Math.max.apply(Math, values);
for (const [key, value] of Object.entries(data)) {
    let tf = value;
    let idf = Math.log(N / value);
    let tfidf = tf * idf;
    console.log(key, tfidf);
}
```

Listing 1: Javascript Code