Edwin Alejandro Cobos Fonseca - 986553

Q1. Write an in-mapper combiner algorithm modifying Co-Ocurrence Matrix (Pair Approach) algorithm.

```
Class Mapper
    method Initialized()
        H = new AssociationArray

    method Map(docId a, doc r)
        for all u in r do
            for all w in window(u) do
                if (H{u,w} is null)
                    H{u,w} = 1
                else
                    H{u,w} = H{u,w} + 1

    method Close()
        for all (u,w) in H do
            Emit((u,w), H{u,w})
```

Edwin Alejandro Cobos Fonseca - 986553

Q2. Write an in-mapper combiner algorithm modifying Co-Occurrence Matrix (Stripe approach) algorithm.

```
Class Mapper
    method Initialize( )
        H = new AssociativeArray

    method Map(docId a, doc r)
        for all u in d do
            h = new AssociativeArray
            for all w in Window(u) do
                h{w} = h{w} + 1

            if (H{u} == null)
                H{u} = h
            else
                H{u} = H{u} + h

    method Close( )
        for all u in H do
            Emit(u, H{u})
```

Q3. Edwin Alejandro Cobos Fonseca - 986553



a. Illustrate Pair approach:

- Input-Split 1:
  {Cat, mat, rat, cat}
  {cat, bat, cat, pat}
  {cat, bat, rat, bat}

- Input-Split 2:
  {Cat, rat, bat, rat}
  {bat, map, pat, bat}
  {pat, cat, bat, mat}

- Mapper Output 1: (Input-Split 1)
  - Record 1
    ((cat, mat), 1)
    ((cat, rat), 1)
    ((mat, rat), 1)
    ((mat, cat), 1)
    ((rat, cat), 1)
  - Record 2
    ((cat, bat), 1)
    ((bat, cat), 1)
    ((bat, pat), 1)
    ((cat, pat), 1)
  - Record 3
    ((cat, bat), 1)
    ((cat, rat), 1)
    ((cat, bat), 1)
    ((rat, bat), 1)
    ((bat, rat), 1)

- Mapper Output 2: (Input-Split 2)
  - Record 1
    ((cat, rat), 1)
    ((cat, bat), 1)
    ((cat, rat), 1)
    ((rat, bat), 1)
    ((bat, rat), 1)
  - Record 2
    ((bat, map), 1)
    ((bat, pat), 1)
    ((mat, pat), 1)
    ((mat, bat), 1)
    ((pat, bat), 1)
  - Record 3
    ((pat, cat), 1)
    ((pat, bat), 1)
    ((pat, mat), 1)
    ((cat, bat), 1)
    ((cat, mat), 1)
    ((bat, mat), 1)

5-5

- Reducer 1 (Input) → $x < K$

((bat, cat), [1])
((bat, mat), [1, 1])
((bat, pat), [1, 1])
((bat, rat), [1, 1])
((cat, bat), [1, 1, 1, 1, 1])
((cat, mat), [1, 1])
((cat, pat), [1])
((cat, rat), [1, 1, 1, 1])

- Reducer 1 (Output)

((bat, cat), 1)
((bat, mat), 2)
((bat, pat), 2)
((bat, rat), 2)
((cat, bat), 5)
((cat, mat), 2)
((cat, pat), 1)
((cat, rat), 4)

- Reducer 2 (Input) → $x > K$

((mat, bat), [1])
((mat, cat), [1])
((mat, pat), [1])
((mat, rat), [1])
((pat, bat), [1, 1])
((pat, cat), [1])
((pat, mat), [1])
((rat, bat), [1, 1])
((rat, cat), [1])

- Reducer 2 (Output)

((mat, bat), 1)
((mat, cat), 1)
((mat, pat), 1)
((mat, rat), 1)
((pat, bat), 2)
((pat, cat), 1)
((pat, mat), 1)
((rat, bat), 2)
((rat, cat), 1)

Edwin Alejandro Cobos Fonseca - 986553

b. Illustrate In-Mapper Combining Version of the 'Pair approach'.

- Mapper Output 1: (Input-Split 1)

((cat, map), 1)
((cat, rat), 2)
((mat, rat), 1)
((mat, cat), 1)
((rat, cat), 1)
⋮

⋮
((cat, bat), 3)
((bat, cat), 1)
((bat, pat), 1)
((cat, pat), 1)
⋮

⋮
((bat, rat), 1)
((rat, bat), 1)

- Mapper Output2(Input-Split 2)

((cat, rat), 2)
((cat, bat), 2)
((rat, bat), 1)
((bat, rat), 1)
⋮

⋮
((bat, mat), 2)
((bat, pat), 1)
((mat, pat), 1)
((mat, bat), 1)
((pat, bat), 2)
⋮

⋮
((pat, cat), 1)
((pat, mat), 1)
((cat, mat), 1)

[5-5]

- Reducer 1 (Input) → $x < K$

((bat, cat), [1])
((bat, mat), [2])
((bat, pat), [1, 1])
((bat, rat), [1, 1])
((cat, mat), [1, 1])
((cat, bat), [3, 2])
((cat, pat), [1])
((cat, rat), [2, 2])

- Reducer 2 (Input) → $x > K$

((mat, bat), [1])
((mat, cat), [1])
((mat, pat), [1])
((mat, rat), [1])
((pat, bat), [2])
((pat, cat), [1])
((pat, mat), [1])
((rat, bat), [1, 1])
((rat, cat), [1])

- Reducer 1 (Output)

((bat, cat), 1)
((bat, mat), 2)
((bat, pat), 2)
((bat, rat), 2)
((cat, bat), 5)
((cat, mat), 2)
((cat, pat), 1)
((cat, rat), 4)

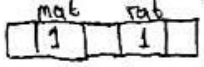
- Reducer 2 (Output)

((mat, bat), 1)
((mat, cat), 1)
((mat, pat), 1)
((mat, rat), 1)
((pat, bat), 2)
((pat, cat), 1)
((pat, mat), 1)
((rat, bat), 2)
((rat, cat), 1)

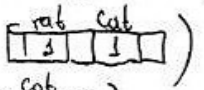
Edwin Alejandro Cobos Fonseca - 986553

c. Illustrate Stripe Approach:

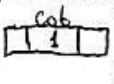
- Mapper Output 1

(cat, 

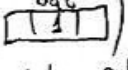
)

(mat, 

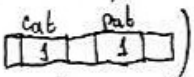
)

(rat, 

)

(cat, 

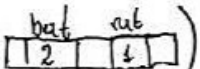
)

(bat, 

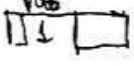
)

(cat, 

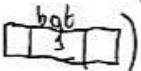
)

(cat, 

)

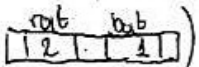
(bat, 

)

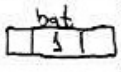
(rat, 

)

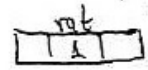
- Mapper Output 2

((cat, 

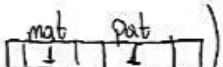
)

(rat, 

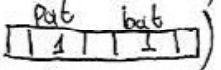
)

(bat, 

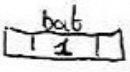
)

(bat, 

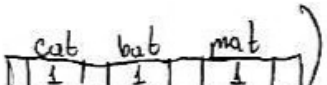
)

(mat, 

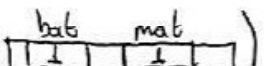
)

(pat, 

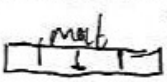
)

(pat, 

)

(cat, 

)

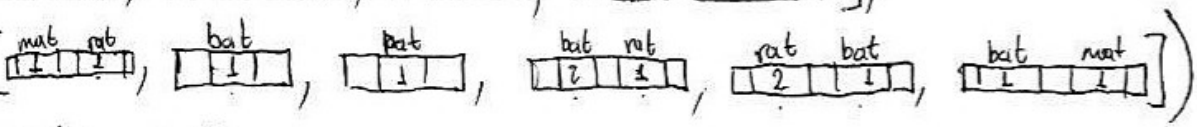
(bat, 

)

- Reducer 1 (Input 1)

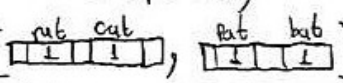
(bat, [ 

])

(cat, [ 

])

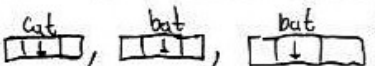
- Reducer 2 (Input 2)

(mat, [ 

])

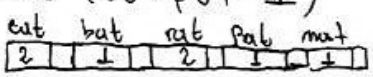
(pat, [ 

])

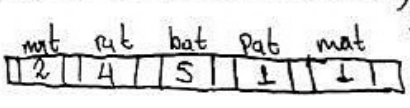
(rat, [ 

])

- Reducer (Output 1)

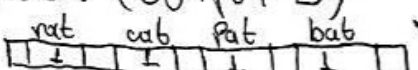
(bat, 

)

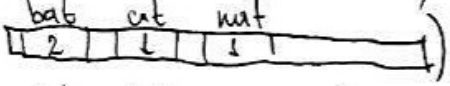
(cat, 

)

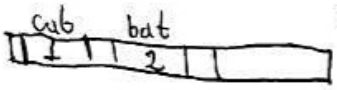
- Reducer (Output 2)

(mat, 

)

(Pat, 

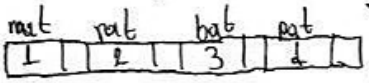
)

(rat, 

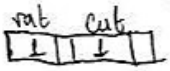
)

d. Illustrate In-Mapper Combining Version of the Stripe Approach

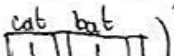
-Mapper Output 1

(Cat, 

)

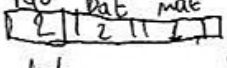
(mat, 

)

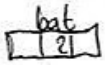
(rat, 

)

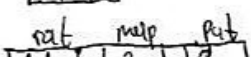
-Mapper Output 2

(Cat, 

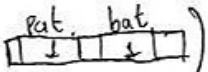
)

(rat, 

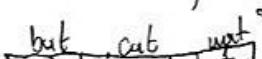
)

(bat, 

)

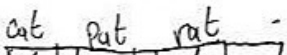
(mat, 

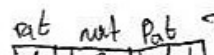
)

(pat, 

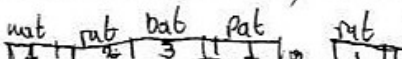
)

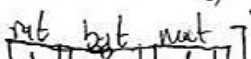
-Reducer Input 1

(bat, [ 

, 

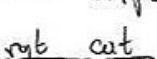
])

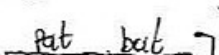
(cat, [ 

, 

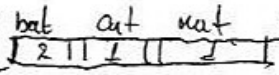
])

-Reducer Input 2

(mat, [ 

, 

])

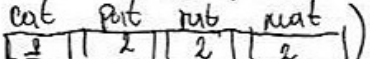
(pat, [ 

])

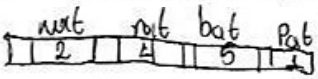
(rat, [ 

])

-Reducer Output 1

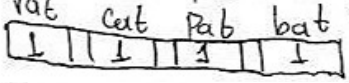
(bat, 

)

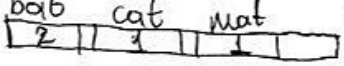
(cat, 

)

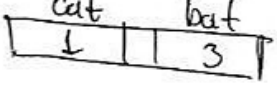
-Reducer Output 2

(mat, 

)

(pat, 

)

(rat, 

)