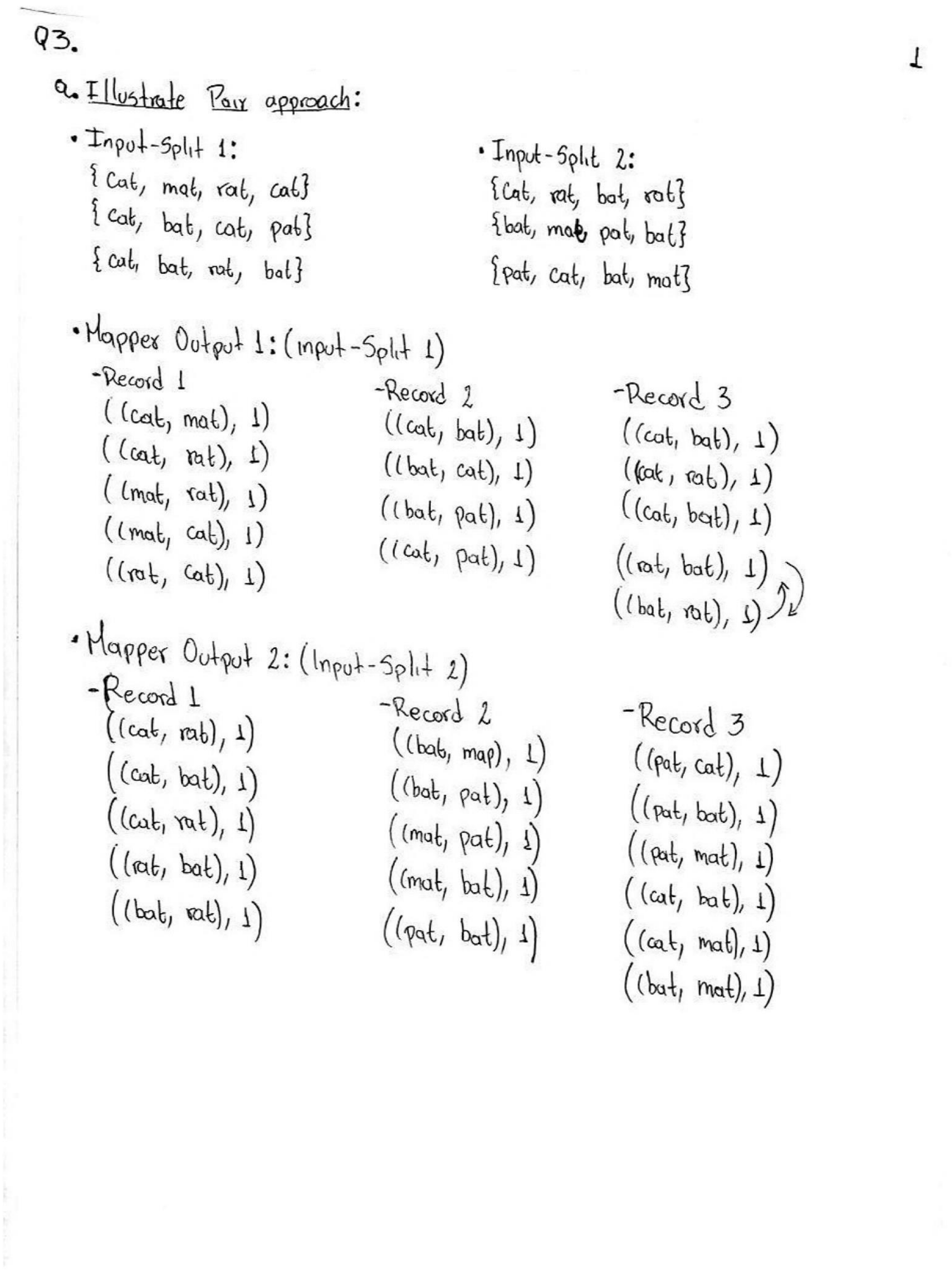
Q1. Write an in-mapper combiner **algorithm** modifying Co-occurrence Matrix (pairs approach) algorithm.

```
Class Mapper
    method initialized()
        H = new AssociationArray
    Method map(docId a, doc r)
        for all v in r do
            for all w in window(v) do
                if(H{v,w} is null)
                    H{v,w} = 1
                else
                    H{v,w} = H{v,w} + 1
    Method close()
        for all (v,w) in H do
            emit((v,w),H{v,w})
```

Q2. Write an in-mapper combiner **algorithm** modifying Co-occurrence Matrix (stripe approach) algorithm.

```
Class Mapper
    method initialized()
        H = new AssociationArray
    Method map(docId a, doc r)
        for all v in r do
            h = new AssociationArray
            for all w in window(v) do
                h{w} = h{w} + 1
            if(H{v} == null)
                H{v} = 1
            else
                H{v} = H{v} + 1
    Method close()
        for all v in H do
            emit(v,H{v})
```

Q3. Assume that there are two input spits and two reducers. Note that Mapper 1 and Reducer 1 run on the same machine. Mapper 2 and Reducer 2 run on the same machine.



Q3.

1

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, mat, 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)

-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)

$\boxed{5-5}$ -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)

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)

-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])

$\boxed{5-5}$

-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)

C. Illustrate <u>Stripe</u> <u>Approach:</u>

- Mapper Output 1

(Cat, $\overset{\text{mat}}{1}\ \overset{\text{rat}}{1}$)

(mat, $\overset{\text{rat}}{1}\ \overset{\text{cat}}{1}$)

(rat, $\overset{\text{cat}}{1}$)

(Cat, $\overset{\text{bat}}{1}$)

(bat, $\overset{\text{cat}}{1}\ \overset{\text{pat}}{1}$)

(Cat, $\overset{\text{pat}}{1}$)

(Cat, $\overset{\text{bat}}{2}\ \overset{\text{rat}}{1}$)

(bat, $\overset{\text{rat}}{1}$)

(rat, $\overset{\text{bat}}{1}$)

- Mapper Output 2

(cat, $\overset{\text{rat}}{2}\ \overset{\text{bat}}{1}$)

(rat, $\overset{\text{bat}}{1}$)

(bat, $\overset{\text{rat}}{1}$)

(bat, $\overset{\text{mat}}{1}\ \overset{\text{pat}}{1}$)

(mat, $\overset{\text{pat}}{1}\ \overset{\text{bat}}{1}$)

(pat, $\overset{\text{bat}}{1}$)

(Pat, $\overset{\text{cat}}{1}\ \overset{\text{bat}}{1}\ \overset{\text{mat}}{1}$)

(Cat, $\overset{\text{bat}}{1}\ \overset{\text{mat}}{1}$)

(bat, $\overset{\text{mat}}{1}$)

- Reducer 1 (Input 1)

(bat, [$\overset{\text{rat}}{1}$, $\overset{\text{cat}}{1}\ \overset{\text{pat}}{1}$, $\overset{\text{rat}}{1}$, $\overset{\text{cat}}{1}\ \overset{\text{bat}}{1}\ \overset{\text{mat}}{1}$])

(Cat, [$\overset{\text{mat}}{1}\ \overset{\text{rat}}{1}$, $\overset{\text{bat}}{1}$, $\overset{\text{pat}}{1}$, $\overset{\text{bat}}{2}\ \overset{\text{rat}}{1}$, $\overset{\text{rat}}{2}\ \overset{\text{bat}}{1}$, $\overset{\text{bat}}{1}\ \overset{\text{mat}}{1}$])

- Reducer 2 (Input 2)

(mat, [$\overset{\text{rat}}{1}\ \overset{\text{cat}}{1}$, $\overset{\text{pat}}{1}\ \overset{\text{bat}}{1}$])

(pat, [$\overset{\text{bat}}{1}$, $\overset{\text{cat}}{1}\ \overset{\text{bat}}{1}\ \overset{\text{mat}}{1}$])

(rat, [$\overset{\text{cat}}{1}$, $\overset{\text{bat}}{1}$, $\overset{\text{bat}}{1}$])

- Reducer (Output 1)

(bat, $\overset{\text{cat}}{2}\ \overset{\text{bat}}{1}\ \overset{\text{rat}}{2}\ \overset{\text{pat}}{1}\ \overset{\text{mat}}{1}$)

(Cat, $\overset{\text{mat}}{2}\ \overset{\text{rat}}{4}\ \overset{\text{bat}}{5}\ \overset{\text{pat}}{1}\ \overset{\text{mat}}{1}$)

- Reducer (Output 2)

(mat, $\overset{\text{rat}}{1}\ \overset{\text{cat}}{1}\ \overset{\text{pat}}{1}\ \overset{\text{bat}}{1}$)

(Pat, $\overset{\text{bat}}{2}\ \overset{\text{cat}}{1}\ \overset{\text{mat}}{1}$)

(rat, $\overset{\text{cat}}{1}\ \overset{\text{bat}}{2}$)

d. Illustrate In-Mapper Combining Version of the <u>Stripe</u> <u>Approach</u>

-Mapper Output 1

(Cat, $\boxed{\overset{\text{mat}}{1} \mid \overset{\text{rat}}{2} \mid \overset{\text{bat}}{3} \mid \overset{\text{pat}}{1}}$)

(mat, $\boxed{\overset{\text{rat}}{1} \mid \overset{\text{cat}}{1}}$)

(rat, $\boxed{\overset{\text{cat}}{1} \mid \overset{\text{bat}}{1}}$)

- Mapper Output 2

(Cat, $\boxed{\overset{\text{rat}}{2} \mid \overset{\text{bat}}{2} \mid \overset{\text{mat}}{1}}$)

(rat, $\boxed{\overset{\text{bat}}{2}}$

(bat, $\boxed{\overset{\text{rat}}{1} \mid \overset{\text{map}}{2} \mid \overset{\text{pat}}{1}}$)

(mat, $\boxed{\overset{\text{pat}}{1} \mid \overset{\text{bat}}{1}}$)

(pat, $\boxed{\overset{\text{bat}}{2} \mid \overset{\text{cat}}{1} \mid \overset{\text{mat}}{1}}$)

-Reducer Input 1

(bat, [$\boxed{\overset{\text{cat}}{1} \mid \overset{\text{pat}}{1} \mid \overset{\text{rat}}{1}}$, $\boxed{\overset{\text{rat}}{1} \mid \overset{\text{mat}}{2} \mid \overset{\text{Pat}}{1}}$])

(cat, [$\boxed{\overset{\text{mat}}{1} \mid \overset{\text{rat}}{2} \mid \overset{\text{bat}}{3} \mid \overset{\text{Pat}}{1}}$, $\boxed{\overset{\text{rat}}{1} \mid \overset{\text{bat}}{2} \mid \overset{\text{mat}}{1}}$])

-Reducer Input 2

(mat, [$\boxed{\overset{\text{rat}}{1} \mid \overset{\text{cat}}{1}}$, $\boxed{\overset{\text{pat}}{1} \mid \overset{\text{bat}}{1}}$])

(Pat, [$\boxed{\overset{\text{bat}}{2} \mid \overset{\text{cat}}{1} \mid \overset{\text{mat}}{1}}$])

(rat, [$\boxed{\overset{\text{cat}}{1} \mid \overset{\text{bat}}{1}}$, $\boxed{\overset{\text{bat}}{2}}$])

-Reducer Output 1

(bat, $\boxed{\overset{\text{cat}}{1} \mid \overset{\text{pat}}{2} \mid \overset{\text{rat}}{2} \mid \overset{\text{mat}}{2}}$)

(cat, $\boxed{\overset{\text{mat}}{2} \mid \overset{\text{rat}}{4} \mid \overset{\text{bat}}{5} \mid \overset{\text{Pat}}{1}}$)

-Reducer Output 2

(mat, $\boxed{\overset{\text{rat}}{1} \mid \overset{\text{cat}}{1} \mid \overset{\text{Pab}}{1} \mid \overset{\text{bat}}{1}}$)

(pat, $\boxed{\overset{\text{bab}}{2} \mid \overset{\text{cat}}{1} \mid \overset{\text{mat}}{1}}$)

(rat, $\boxed{\overset{\text{cat}}{1} \mid \overset{\text{bat}}{3}}$)