

Edwin Alejandro Cobos Fonseca - 986553

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

Class Mapper

method Initialized()

$H = \text{new AssociationArray}$

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 (stnpe approach) algorithm.

Class Mapper

method Initialize()

$H = \text{new AssociativeArray}$

method Map(doc d, doc r)

for all  $u$  in  $d$  do

$h = \text{new AssociativeArray}$

for all  $w$  in  $\text{Window}(u)$  do

$h\{w\} = h\{w\} + 1$

if ( $H\{u\} := \text{null}$ )

$H\{u\} = h$

else

$H\{u\} = H\{u\} + h$

method Close()

for all  $u$  in  $H$  do

Emit( $u$ ,  $H\{u\}$ )

Q3. Edwin Alejandro Cobos Fonseca - 986553

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)

((rat, 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, mat), 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)  $\rightarrow x < K$ 

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

- Reducer 1 (Output)

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

5-5 - Reducer 2 (Input)  $\rightarrow x > K$ 

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

- 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)	((cat, bat), 3)	((bat, rat), 1)
((mat, rat), 1)	((bat, cat), 1)	((rat, bat), 1)
((mat, cat), 1)	((bat, pat), 1)	
((rat, cat), 1)	((cat, pat), 1)	
⋮	⋮	

- Mapper Output 2 (Input-Split 2)

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

-Reducer 1 (Input)  $\rightarrow x < K$ S-S-Reducer 2 (Input)  $\rightarrow x > K$ 

((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])

((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)

## Illustrate Stripe Approach:

### - Mapper Output 1

(cat, 

mat	rat
1	1

)

(mat, 

rat	cat
1	1

)

(rat, 

cat
1

)

(cat, 

bat
1

)

(bat, 

cat	pat
1	1

)

(cat, 

pat
1

)

(cat, 

bat	rat
1	1

)

(bat, 

pat
1

)

(rat, 

bat
1

)

### - Mapper Output 2

(cat, 

rat	bat
1	1

)

(rat, 

bat
1

)

(bat, 

rat
1

)

(bat, 

mat	pat
1	1

)

(mat, 

pat	bat
1	1

)

(pat, 

bat
1

)

(pat, 

cat	bat	mat
1	1	1

)

(cat, 

bat	mat
1	1

)

(bat, 

mat
1

)

### - Reducer 1 (Input 1)

(bat, [

rat
1

, 

cat	pat
1	1

, 

rat
1

, 

cat	bat	mat
1	1	1

])

(cat, [

mat	rat
1	1

, 

bat
1

, 

pat
1

, 

bat	rat
1	1

, 

rat	bat
1	1

, 

bat	mat
1	1

])

### - Reducer 2 (Input 2)

(mat, [

rat	cat
1	1

, 

pat	bat
1	1

])

(pat, [

bat
1

, 

cat	bat	mat
1	1	1

])

(rat, [

cat
1

, 

bat
1

, 

bat
1

])

### - Reducer (Output 1)

(bat, 

cat	bat	rat	pat	mat
2	1	2	1	1

)

(cat, 

mat	rat	bat	pat	mat
2	1	5	1	1

)

### - Reducer (Output 2)

(mat, 

rat	cat	pat	bat
1	1	1	1

)

(pat, 

bat	cat	mat
2	1	1

)

(rat, 

cat	bat
1	2

)

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

-Mapper Output 1

(cat, 

1	2	3	4
---	---	---	---

)  
 (mat, 

1	2
---	---

)  
 (rat, 

1	2
---	---

)

-Mapper Output 2

(cat, 

2	2	1	1
---	---	---	---

)  
 (rat, 

2
---

)  
 (bat, 

1	2	2
---	---	---

)  
 (mat, 

1	2
---	---

)  
 (pat, 

2	1	1
---	---	---

)

-Reducer Input 1

(bat, [

cat	pat	rat
1	1	1

, 

rat	mat	pat
1	2	1

])  
 (cat, [

mat	rat	bat	pat
1	2	3	1

, 

rat	bat	mat
1	2	1

])

-Reducer Input 2

(mat, [

rat	cat
1	1

, 

pat	bat
1	1

])  
 (pat, [

bat	cat	mat
2	1	1

])  
 (rat, [

cat	bat
1	1

, 

bat
2

])

-Reducer Output 1

(bat, 

cat	pat	rat	mat
1	2	2	2

)  
 (cat, 

mat	rat	bat	pat
2	4	5	2

)

-Reducer Output 2

(mat, 

rat	cat	pat	bat
1	1	1	1

)  
 (pat, 

bat	cat	mat
2	1	1

)  
 (rat, 

cat	bat
1	3

)