

Name: EROMONSEI SAMSON
COMP 5317

1.) Retrieval system (S_1)

Predicted

TP		\bar{F}_N
Actual	4	16
	6	
		T_N

$$\text{Precision} = \frac{TP}{TP + \bar{F}P} = \frac{4}{10} = 0.4$$

$$\text{Recall} = \frac{TP}{TP + \bar{F}N} = \frac{4}{20} = 0.2$$

(S_2) Retrieval system 2

Predicted

TP		\bar{F}_N
Actual	5	15
	5	
		T_N

$$\text{Precision} = \frac{5}{10} = 0.5$$

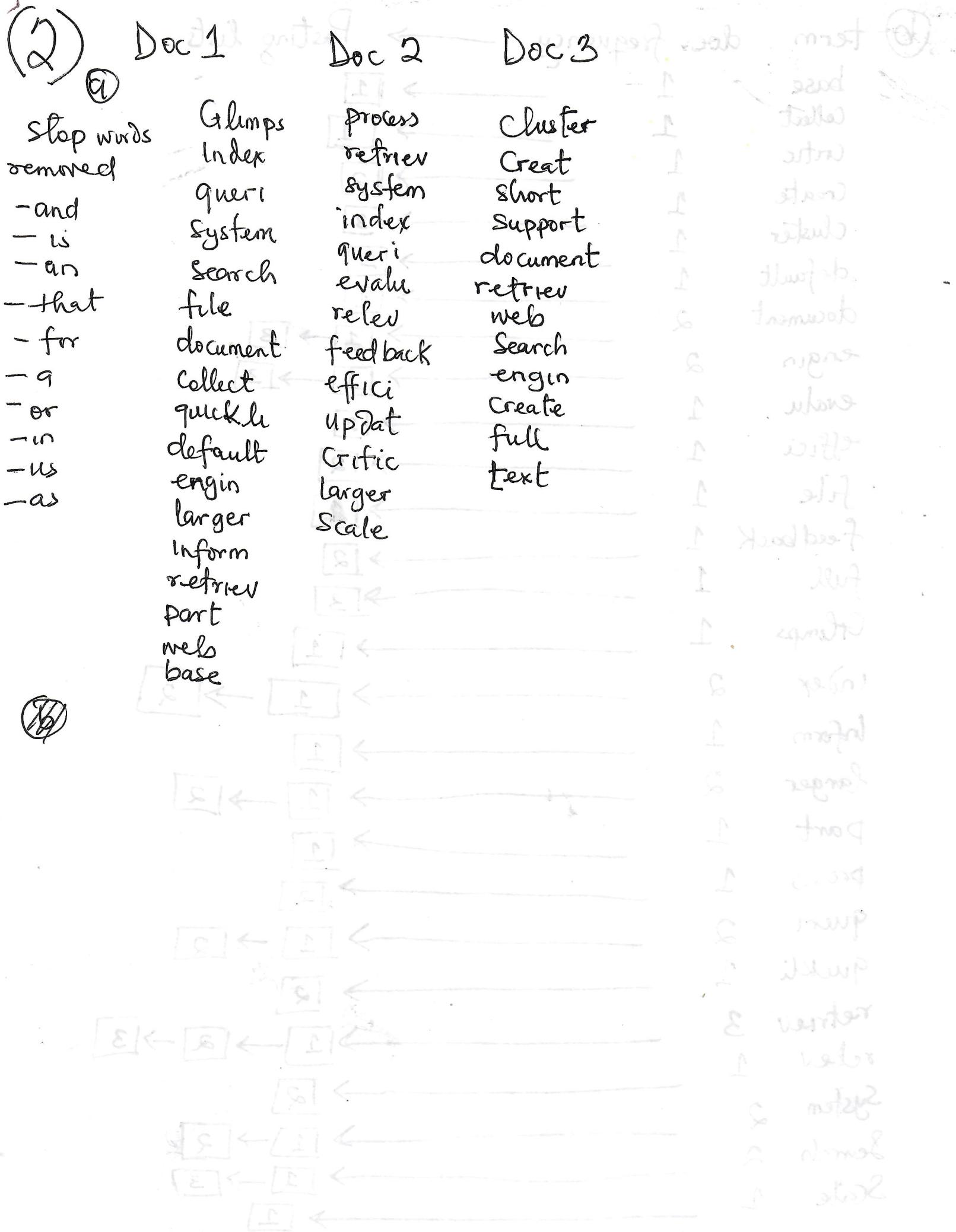
$$\text{Recall} = \frac{5}{20} = 0.25$$

1.b.) Compute the F-Score (S_1) where $\beta=1$

$$\frac{(1+\beta^2) \times P \times R}{P + R} = \frac{(1+1^2) \times 0.4 \times 0.2}{0.4 + 0.2} = 0.27$$

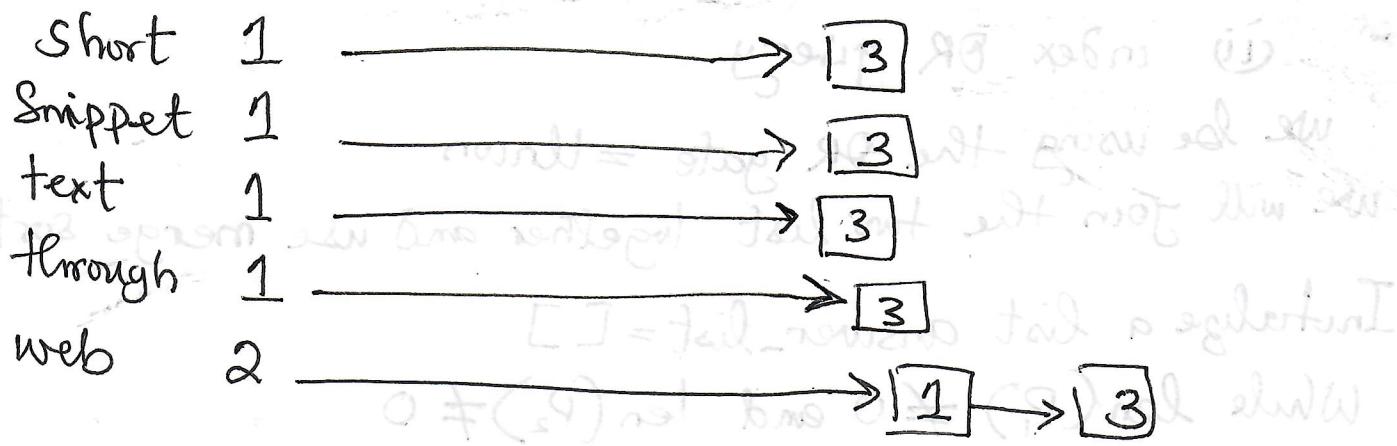
F-Score (S_2)

$$\frac{(1+1^2) \times 0.5 \times 0.25}{0.5 + 0.25} = 0.33$$



(b) term doc. frequency → Posting lists

base	1	→ [1]
Collect	1	→ [1]
Critic	1	→ [2]
Create	1	→ [3]
cluster	1	→ [3]
default	1	→ [3]
document	2	→ [1] → [3]
engin	2	→ [1] → [3]
evalu	1	→ [2]
effici	1	→ [2]
file	1	→ [1]
feedback	1	→ [2]
full	1	→ [3]
Glimps	1	→ [1]
Index	2	→ [1] → [2]
Inform	1	→ [1]
larger	2	→ [1] → [2]
Part	1	→ [1]
process	1	→ [2]
Query	2	→ [1] → [2]
Quickli	1	→ [2]
retriev	3	→ [1] → [2] → [3]
reliev	1	→ [1] → [2] → [3]
System	2	→ [2]
Search	2	→ [1] → [2]
Scale	1	→ [1]



2(c) index AND Query

(i) P_1 index → [1] → [2]
 P_2 query → [2]

Using the intersection (P_1, P_2) with the condition the lists
 answer_list = empty List [] one sorted
 while $P_1 \neq \text{NIL}$ and $P_2 \neq \text{NIL}$
 if $\text{docID}(P_1) == \text{docID}(P_2)$
 then add $(\text{answer_List} \geq \text{docID}(P_1))$
 set $P_1 \leftarrow \text{next}(P_1)$
 set $P_2 \leftarrow \text{next}(P_2)$
 else if $\text{docID}(P_1) < \text{docID}(P_2)$:
 then $P_1 \leftarrow \text{next}(P_1)$
 else:
 $P_2 \leftarrow \text{next}(P_2)$

return answer_list

P_1 index → [1] → [2] ← 3
 P_2 query → [1] ← 2
 Index AND Query → [1] ← 1

ii) index OR query

We will be using the OR gate = Union

We will join the two list together and use merge sort algorithm

Initialize a list answer_list = []

While $\text{len}(P_1) \neq 0$ and $\text{len}(P_2) \neq 0$

If $\text{DocID}(P_1) = \text{DocID}(P_2)$

then Add (answer_List DocID(P_1))

$P_1 \leftarrow \text{next}(P_1)$

$P_2 \leftarrow \text{next}(P_2)$

Else If $\text{DocID}(P_1) < \text{DocID}(P_2)$

then Add (answer_List DocID(P_1))

$P_1 \leftarrow \text{next}(P_1)$

Else

Add (answer P_2)

$P_2 \leftarrow \text{next}(P_2)$

$\text{len}(P_1)-1$ and $\text{len}(P_2)-1$

If $\text{len}(P_1)=0$

ADD (answer + list of $\text{len}(P_1)-1$)

Else

ADD (answer + list of $\text{len}(P_2)-1$)

P_1 index \rightarrow 

P_2 index \rightarrow 

$P_1 \cup P_2 \rightarrow$ 

(iii) index AND (NOT query)

In this scenario we are trying to generate a list that satisfy index and exclude those that satisfy query

Index $\rightarrow [1] \rightarrow [2]$

Query $\rightarrow [2] \sqcap [1] \sqcap [3] \sqcap [4] \sqcap [5] \sqcap [6] \sqcap [7]$

[1] satisfy only index but [2] satisfy index and query
and since we need to exclude 2

Index AND (NOT query) $\rightarrow [1]$

[iv] (Search AND Query) OR (Search AND retrieve)

(Search \sqcap Query) \sqcup (Search \sqcap retrieve)

Search P₁ $\rightarrow [1] \rightarrow [3]$

Query P₂ $\rightarrow [1]$

retrieve P₃ $\rightarrow [1]$

Using previous algorithm in 2(c)

Search \sqcap Query \sqcup Search \sqcap retrieve

\downarrow
[1]

\sqcup

\downarrow
[1]

Final list $\rightarrow [1]$

(v) (Index OR cluster) AND (Web OR System)

Index \rightarrow [1] \rightarrow [2]

Web \rightarrow [1] \rightarrow [3]

Cluster \rightarrow [3]

System \rightarrow [1] \rightarrow [2]



[1] \rightarrow [2] \rightarrow [3]

∩

[1] \rightarrow [2] \rightarrow [3]

[2] \leftarrow [1] \leftarrow [3]

[1] \rightarrow [2] \rightarrow [3]

||

2d) (i) Small impact on the non positional posting size reduction, is because only term posting one consider with the position. The ~~case folding~~ stemming will reduce duplicate differences. the case folding convert all characters into the same and same words are not repeated in the posting.

example. in Doc 3, Clusters and clusters.

(ii) No impact on the positional posting size, because

Clusters \rightarrow doc3 \rightarrow position 1

Clusters \rightarrow doc3 \rightarrow position 9

the memory allocation will not change and not be removed.