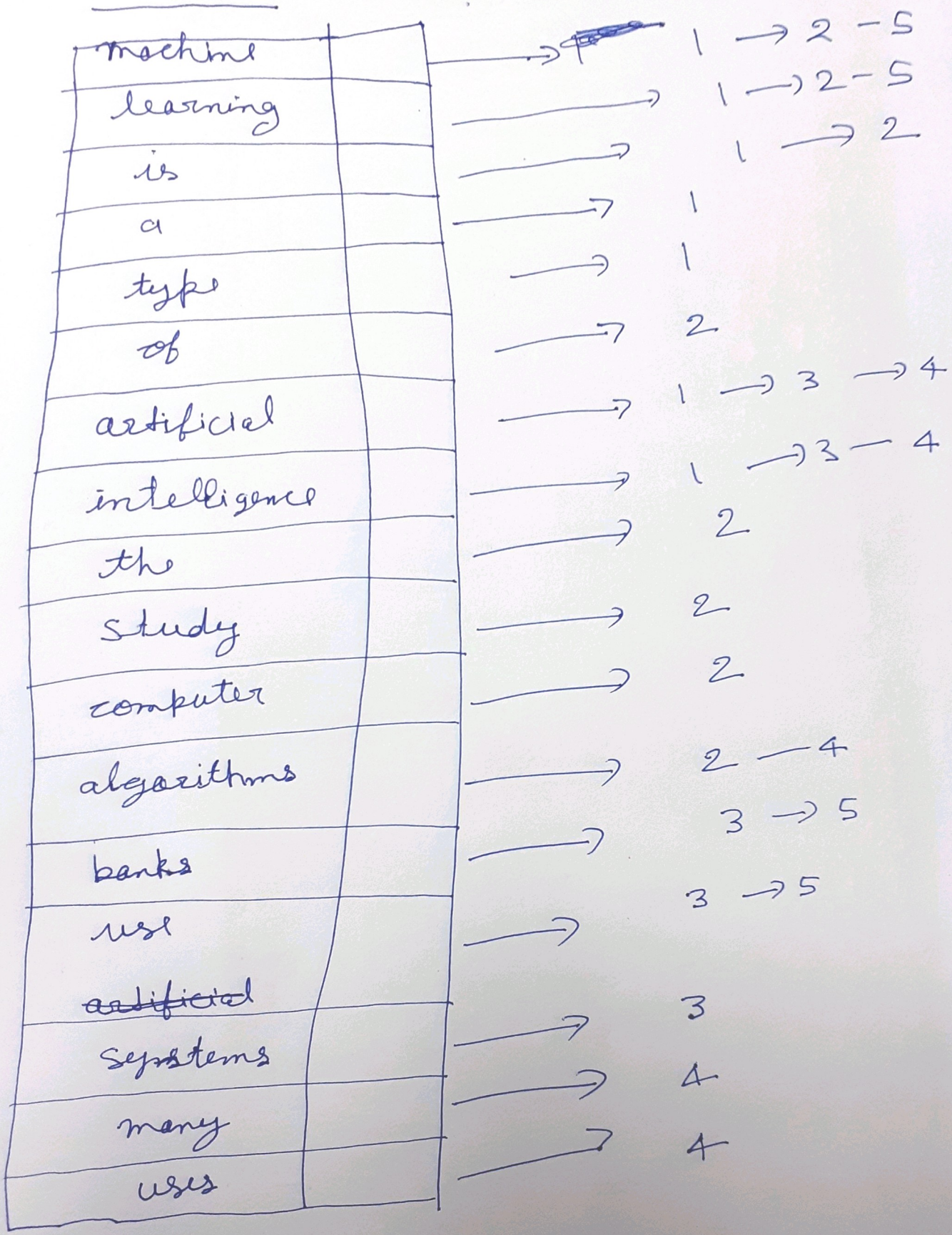
Ans-1 Dictionary



(2)

(information OR learning) AND
(machine OR intelligence) AND
(artificial OR Retrieval) ;

Term	Posting size
artificial	10,77,00
information	15,58,00
intelligence	9,93,00
learning	14,56,00
machine	13,27,00
Retrieval	38,60,00

information OR learning = O(15,5800 + 14,56,00)
= O(3,01,400)

machine OR intelligence = O(13,2700 + 9,93,00)
= O(2,32,000)

artificial OR Retrieval = O(10,77,00 + 38,60,00)
= O(4,93,700)

So order is depends on AND operation, which gives recommendation as :—

(machine OR intelligence) AND (information OR learning) AND (artificial OR Retrieval)

Ans 6 Good Attributes for inverted index.

→ Time complexity / Space complexity / maintainence.

On average cost above attributes are good for Inverted Index.

Good Attribute for boolean model

→ For worst case searching in inverted index is takes more time than boolean model. For example, if query terms are present in all the documents.

Ans-9 Factoid questions are like Answer of these questions are expected to be a span of context. generally in few words like 1, 2, or 3.

or

Its answers are facts.

Ans 7 (a) Term_2 and term_3 and (7 term_4)

Term_2 and Term_3 ⇒

t_2 (1 0 1 0 1 0 1)

t_3 (1 1 1 1 1 0 1)

(and) (1 0 1 0 1 0 1)

7(T_4) (1 1 1 1 0 1 0)

(1 0 1 0 0 0 0)

So Doc_1 and Doc_3

(b) Term_2 or Term_3 but not Term_4

T_2 1 0 1 0 1 0 1

T_3 1 1 1 1 1 0 1

or → 1 1 1 1 1 0 1

7 t_4 1 1 1 1 0 1 0

and → 1 1 1 1 0 0 0

Doc 1, 2, 3, 4

Ans 8 -

term 2 → 0 1 0 0 0 0 0 0 0 0

sentence 1 → 1 1 1 0 0 0 0 0 0 0

sentence 2 → 0 1 1 0 0 1 0 0 0 0

$$\text{cosine} = \frac{S1 \cdot S2}{|S1|\,|S2|} = \frac{2}{3}$$