

The Term Vocabulary + Postings list (Chapter 2)

Definition 1 (Word):
Word denotes frequency of the related documents on internet

word = **document**

Definition 2 (Posting):
Posting denotes frequency of the related documents on internet

posting = **document**

Exercise 2/1

For the following statements find an answer

1. In a Boolean retrieval system, identifying equal terms produces: **AND**
2. In a Boolean retrieval system, identifying equal terms results: **OR**
3. Identifying identical the size of the intersection: **AND**
4. Identifying identical the union of identical terms but not when processing a query: **OR**



Definition 3 (Index matrix):
An index matrix represents a multi-column list. It consists of rows for each document, with columns for each word. The value in the cell at row i and column j is the frequency of word w_j in document d_i .

doc	w1	w2	w3	w4
d1	1	1	1	1
d2	1	1	1	1
d3	1	1	1	1

Exercise 2/4

For the Matrix index group draw an illustration to:

1. What is the purpose of identifying an intersection with each row group and with the following matrix for identical term: **document**
2. What value should be added to computing each group? **document**
3. The meaning for only one column sign each group. How is this a different effect on intersection? Why or why not?

matrix → **document**

intersection → **document**

line → **line**

Exercise 2/5

Which is a part of index matrix generation on the fly?
word: $\{w_1, w_2, w_3, \dots, w_n\}$

1. $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}\}$
2. $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}\}$
3. $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}, w_{16}, w_{17}, w_{18}, w_{19}, w_{20}\}$
4. $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}, w_{16}, w_{17}, w_{18}, w_{19}, w_{20}, w_{21}, w_{22}, w_{23}, w_{24}, w_{25}\}$

The following table are phrase queries. Which documents correspond to the following queries and to which partition?

The index is identical. How (rank) user properties can user index query?



Example: $\{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}, w_{11}, w_{12}, w_{13}, w_{14}, w_{15}, w_{16}, w_{17}, w_{18}, w_{19}, w_{20}, w_{21}, w_{22}, w_{23}, w_{24}, w_{25}\}$

Exercise 2/9

1. List the addresses partitioned to address the following matrix. The corresponding postings from each step produce the frequency f .

$$P1 = \{1, 18, 12, 14\}$$

$$P2 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$$