

MCQ:

1. Information Retrieval is a software system that finds material of a data that satisfies information needs from large collections.
 - a. Structured.
 - b. Unstructured.**
 - c. Both.
 - d. None of them.
2. Which of the following is an example of IR system?
 - a. Web search.
 - b. E-mail search.
 - c. Searching on laptop.
 - d. All of them**
3. Web search is an example of large IR systemscale that stored in more than one computer.
 - a. Yes**
 - b. No
4. E-mail search& laptop search are examples of personal IR system scalethat stored in one computer.
 - a. Yes.**
 - b. No.
5. Legal information retrieval & corporate knowledge bases are examples of enterprise IR system scale.
 - a. Yes.**
 - b. No.
6. A is a set of document.
 - a. Collection.**
 - b. Goal.
 - c. Task.
- ➔ 7. The goal of the IR systems is to retrieve documents with information that is to user's information need & help user to complete task.
 - a. Irrelevant.
 - b. Relevant.**
8. The information need defined as query to the search engine.
 - a. True.**
 - b. False.
9. There are different types of queries, which of the following is/are used in IR systems?
 - a. Boolean query. ✓
 - b. Phrase query. ✓
 - c. Proximity query. ✓
 - d. Free text query. ✓
 - e. Parametric query
 - f. All of them.**
10. Ad hoc retrievalis finding documents in a collection of documents(corpus), relevant to certain user need.
 - a. True.**

- b. False.
11. model is a model that queries posed on Boolean expressions.
- a. **Boolean retrieval model.**
- b. Ranked retrieval model.
- c. Bag of words retrieval model.
- d. Vector space model.
12. AND, OR, NOT are examples of Boolean retrieval model.
- a. **True.**
- b. False.
13. Brute Force Approach & Organize beforehand are from Boolean retrieval model approaches.
- a. **True.**
- b. False.

14. Brute Force Approach use liner search.

- a. **True.**
- b. False.

15. Which of the following is/are from Brute Force Approach disadvantages.

- a. Slow for large corpus.
- b. NOT is non-trivial.
- c. Other operations not feasible (ex. NEAR)
- d. **All of them.**

16. In term document incidence matrices, represented in row, while represented in columns.

- a. **Terms, documents.**
- b. Documents, terms.

17. In term document incidence matrices, value is if document didn't contain term, If document contains it.

- a. **0, 1.**
- b. 1, 0.
- c. 1, -1.
- d. 1, 1.

18. Using following term document incidence matrix, what is result of query "**Brutus**AND

CaesarNOT**Calpurnia**"?

- a. 10011010.
- b. 01110011.
- c. 010000.
- d. **100100.**

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
mercy	1	0	1	1	1	1
worser	1	0	1	1	1	0

19. Term document incidence matrix are very sparse.

- a. **True.**
- b. False.

20. In for each term 't', we must store a list of all documents that contain term 't'.

- a. Term document incidence.
- b. **Inverted index.**

21. In inverted index, stored in memory, while stored on hard disk.

- a. **Dictionary, posting lists.**
- b. Posting lists, dictionary.
- c. Terms, lists.
- d. Pointers, terms.

22. use linked lists or variable length array.

- a. **memory.**
- b. disk.
- c. Posting list.

23. is cutting character sequence into word tokens.

- a. Stemming.
- b. **Tokenization.**
- c. Normalization.

24.map text and query term into same form.

- a. Stemming.
- b. Tokenization.
- c. **Normalization.**

25. match different forms of root.

- a. **Stemming.**
- b. Tokenization.
- c. Normalization.

26. Stop words are very common words like(the, a, to, of)

- a. **True.**
- b. False.

27. Which of the following is from inverted index steps?

- a. Tokenization.
- b. Sorting.
- c. Dictionary & Posting.
- d. **All of them.**

28. Token sequence is a sequence of pairs.

- a. **Term, Document ID.**
- b. Document Frequency, Document ID.
- c. Term, Document Frequency.
- d. Document ID, Document Frequency.

#Note: Modified token = Term.#

29. In inverted index, for sort step, we sort by

- a. Term then frequency.
- b. **Term then document id.**
- c. Document id then term.
- d. Frequency then term.

△, >

Friendship
↓
Friend → Modified
Token

30. In step we add document frequency.

- a. Sort.
- b. Tokenization.
- c. **Dictionary & Posting.**

31. "LIMIT! /3 STATUTE", /3 means.

- a. Search for words limit & statute with no words between them.
- b. Search for words limit & statute with 3 words exactly between them.
- c. **Search for words limit & statute with 3 words at most between them.**

32. For following query, which two terms should we process first?

(tangerine OR trees) AND (marmalade OR skies) AND (kaleidoscope OR eyes).

- a. tangerine OR trees. -> 363465
- b. marmalade OR skies. -> 379571
- c. **kaleidoscope OR eyes.** -> **300321**

Term	Freq
eyes	213312
kaleidoscope	87009
marmalade	107913
skies	271658
tangerine	46653
trees	316812