

Information Retrieval (IS322) / Data Storage and Retrieval (IS313)

midterm exam – April 2023

Student Name: _____

Student ID: _____

Use the answer sheet give the (best) choice — 30 question 2/3 each – submit both sheets

1- The process that involves retrieval of data from various sources in order to process it further is called:

- a- Data Mining b- Data Analysis **c- Data Extraction** d- Information retrieval e- Web Mining

X 2- The automated retrieval of specific information related to a selected topic from bodies of text is called

- a- Crawling b- Data Extraction c- Data Mining d- Information Extraction e- Data Analysis

3- The Goldberg machine is a ----- Machine that searched for a pattern of dots or letters across catalog entries stored on a roll of microfilm.

- a- Mechanical** b- Electronic c- Laser d- Digital e- Magnetic Tape

4- ----- is the topic about which the user desires to know more

- a- A query **b- An information need** c- A user task d- A misconception e- A misformulation

5- ----- is what the user conveys to the computer in an attempt to communicate the information need.

- a- A query** b- An information need c- A user task d- A misconception e- A misformulation

6- if the result is called ----- that means the user perceives as containing information of value with respect to his information need.

- a. valid b. complete c. reasonable **d. relevant** e. incomplete

7- The fraction of the relevant documents in the collection were returned by the IR system is called ----

- a. recall** b. precision c. f-measure d. relevance e. soundness

8- The fraction of the returned results are relevant to the information need is called ----

- a. recall **b. precision** c. f-measure d. relevance e. soundness

9- Consider Grepping: It is NOT true that:

- a. It is a very effective process b. grep is a UNIX command c. Impractical for near queries
d. good for ranked retrieval e. allows useful possibilities for wildcard pattern matching

10- The Boolean Retrieval model is a -----

- a. model for information retrieval** b. model that views a document as a set of sentences
c. data model d. good model for ranked retrieval e. a model for ranked retrieval

Given the following Term-Document Incidence Matrix for questions (11-15)

	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6
Egypt	1	1	1	0	1	1
Syria	0	0	1	1	0	1
Russia	1	0	1	0	1	0
France	0	1	1	0	0	0
Iraq	1	1	0	1	1	1

11- The query Russia and Egypt and France will result to
a. 110110 b. 111011 c. 100010 d. 001001 e. 001000

12- The query Russia and Egypt not France will result to
a. 110110 b. 111011 c. 100010 d. 001001 e. 001000

13- Which document has Syria and Iraq but not Egypt
a. 1 b. 2 c. 3 d. 4 e. 5

14- The posting list 1,3,5 is for
a. Doc 1 b. Doc 2 c. Egypt d. Syria e. Russia

15- The given matrix is not typical because it
a. has a big collection b. has too many terms c. is sparse d. is not sparse e. has a lot of zero's

16- If the collection is 1,000,000 and the number of terms is 100,000 and the number of terms in a query is 5, what is the maximum size of any posting list
a. 500,000 b. 5,000,000 c. 1,000,000 d. 100,000 e. 20,000

17- If the collection is 1,000,000 and the number of terms is 100,000 and the number of terms in a query is 5, what is the maximum number of posting lists. (assume no phrases)
a. 500,000 b. 5,000,000 c. 1,000,000 d. 100,000 e. 20,000

18- If the Term-Document Incidence Matrix is sparse then the equivalent inverted index
a. contains fewer terms b. contains more terms
c. contains shorter posting lists d. contains longer posting lists e. use more memory

19- In a Boolean retrieval system, stemming -----
a. increase the size of the vocabulary b. never lowers precision. c. can increase the retrieved set
d. increase the number of relevant documents e. should not be invoked at indexing

20- In a Boolean retrieval system, stemming never lowers recall because stemming-----
a. will decrease the retrieved set b. can increase the retrieved set c. increase the size of the vocabulary
d. decrease the size of the vocabulary e. increase the number of relevant documents

21- In the initial stages of text processing Tokenization is the process of:

- a. cut character sequence into words b. mapping text and query terms to the same form
c. omitting very common words d. matching different forms of a root e. authorization

22- In the initial stages of text processing Stemming is the process of:

- a. cut character sequence into words b. mapping text and query terms to the same form
c. omitting very common words d. matching different forms of a root e. authorization

23- The goal of the Extended Boolean model is to overcome the drawbacks of the Boolean model that has been used in information retrieval which mainly was -----

- a. always too few results b. always too much results c. always wrong results
d. bad ranking of the result set e. the result set is often too small or too big

24- WestLaw is NOT -----

- a. an example of Extended Retrieval Model b. a type of a Boolean model c. legal search service
d. a model that require special query language e. for western Diplomacy

Given the following portion of a positional index (FOR 25,26)

1 angels: 2: (36,174,252,651); 4: (12,22,102,432); 7: (17);
fools: 2: (1,17,74,222); 4: (2,18,78,108,458); 7: (3,13,23,193);
2 fear: 2: (87,704,722,901); 4: (13,43,113,433); 7: (18,328,528);
in: 2: (3,37,76,444,851); 4: (3,10,20,110,470); 6: (5,15,25,195);
3 rush: 2: (2,66,194,321,702); 4: (6,9,19,69,114,429,569); 7: (4,14,404);

25- Which document(s) if any meet the positional query "fools rush in"

- a. 2, 4, 7 b. 2, 4, 6 c. 4, 7 d. 2, 4 e. none of them

26- Which document(s) if any meet the positional query "angels fear rush"

- a. 2, 4, 7 b. 2, 4, 6 c. 4, 7 d. 2, 4 e. none of them

X 27- which of the following westlaw queries will find the following sentence
happiness is an emotional state characterized by feelings of joy

- a. happ! /s emot! /p joy satis! b. happ! /s emot! /2 joy satis! c. happ! /p emot! /2 joy satis!
d. happy! /s emot! /p joy satis! e. happ! /s emot! /p satis!

28- Not Knowing what to search for in order to get your information need is called

- a. False information b. miscommunication c. misformulation d. Misconception e. fake information

29- The main issues for biword indexes

- a. slower than positional indexes b. famous names such as "Mohamed Ali"
c. complicated inverted index d. False positives e. stop words

30- Not Knowing how to write suitable query for your information need is called

- a. False information b. miscommunication c. misformulation d. Misconception e. fake information