Information Retrieval

1. What is information retrieval?
   - ○ **A.** The creation and populating of a database
   - ○ **B.** The restoration of data that is accidentally lost or corrupted
   - ○ **C.** The design and implementation of search algorithms
   - ○ **D.** The collection of useful information from datasets

2. What does precision refer to in information retrieval?
   - ○ **A.** The ability of an algorithm to avoid retrieving non-relevant information
   - ○ **B.** The ability of an algorithm to separate useful from useless information
   - ○ **C.** The ability of an algorithm to strike it's intended target
   - ○ **D.** The ability of an algorithm to retrieve all relevant documents

3. Which one of the following is not a boolean search operator?
   - ○ **A.** AND
   - ○ **B.** IF
   - ○ **C.** NOT
   - ○ **D.** OR

4. Given two search terms separated by a boolean operator (e.g. New AND Mexico) which boolean will return the most results?
   - ○ **A.** AND
   - ○ **B.** OR
   - ○ **C.** NOT
   - ○ **D.** IF

5. The C-reactive protein in human is associated with the increased risk of thrombosis and is also associated with fibrosis. What will be the keywords to search for this protein that will give more accurate results?

a. C-reactive protein, thrombosis, fibrosis
b. C-reactive protein, human, increased risk of thrombosis, fibrosis
c. C-reactive protein, human
d. protein, human

○ **A.** a

○ **B.** b

○ **C.** a and b

○ **D.** all of the above

6. The purpose of alignment is:

○ **A.** To search for homologues

○ **B.** To identify functional domains

○ **C.** To compare cross species

○ **D.** All of the above

7. Which of the following is NOT an alignment of ATTACG and TTAG

○ **A.**

```
A T - T A C G
| | | | | | |
T T A G - - -
```

○ **B.**

```
- - A T T A C G - -
| | | | | | | | | |
T T - - - - - - A G
```

○ **C.**

```
A T A C G
| | | | |
T T A G -
```

○ **D.**

```
A T T A C G
| | | | | |
- T T A - G
```

8. To align the entire domain into the sequence use:

○ **A.** Global alignment

○ **B.** Local alignment

○**C.** Fit alignment

○**D.** Optimal alignment

9. Assume that match score is +1, mismatch is -2, and indel is -1. What is the score for the following alignment?

```
A T T A C
| | | | |
G T - A C
```

○**A.** -2

○**B.** -1

○**C.** 0

○**D.** +1

10. Which of the following describe types of scoring models?

a. Optimization based on minimizing the distance between sequences
b. Optimization based on maximizing the similarity between sequences
c. Comparision of sequences that show the regions of greatest statistical similarity

○**A.** c only

○**B.** a and b only

○**C.** a and c only

○**D.** a, b, and c

11. Which of the following is a local alignment of AATACG and TTTACT?

○**A.**

```
-TAC-
 |||
-TAC-
```

○**B.**

```
AAT A C G
   | | |
- - T A C -
```

○**C.**

```
AATACG -
|||||||
TT TAC - T
```

○**D.**

```
 - T A C -
| | |  | |
T T A C T
```

12. Consider the given un-gapped alignment of two DNA sequences, with a score of +1 for a match and -1 for a mismatch.

DNA 1	A	G	T	A	C	T	T
DNA 2	A	T	G	C	C	A	T
SCORE	1	0	-1	-2	-1	-2	-1

Determine the ladder points, and excursions from the above table. Which of the following lists the maximum heights achieved by the excursions?

○ **A.** 1, 0, 1

○ **B.** 2, 1, 2

○ **C.** 1, 2, 1, 2

○ **D.** 2, 4, 0, 1

13. To blast "NSLNKVIPSPPTHSLG" we use:

○ **A.** Blastp

○ **B.** Blastn

○ **C.** tBlast

○ **D.** A or B

14. Bit score is:

○ **A.** The measure of similarity between the query and hit

○ **B.** The measure of probability that the similarity is by chance

○ **C.** The number of identities

○ **D.** All of the above

15. Each of the following is an example of a Scoring Matrix used for protein alignments, EXCEPT

○ **A.** PAM 100

○ **B.** BLOSUM 50

○ **C.** PAM 250

○ **D.** PFAM

16. Assume we have a scoring matrix for proteins. For the row corresponding to amino acid, "R", we have corresponding values in the columns for "A" and "W" as -10 and -1 respectively. What inference may be drawn from this observation.

○ **A.** Over time R is more likely to mutate to "A" compared to "W".

○ **B.** Over time R is less likely to change to "A" compared to "W".

○ **C.** R is a functionally more similar to "A" compared to its similarity to "W".

○ **D.** None of the above.

17. Consider the scoring matrix below, with a gap "indel"cost of -4.

	A	R	N	D	C	Q	...
A	2	-2	0	0	-2	0	...
R	-2	6	0	-1	-4	1	...
N	0	0	2	2	-4	1	...
D	0	-1	2	4	-5	2	...
C	-2	-4	-4	-5	12	-5	...
Q	0	1	1	2	-5	4	...
...							

What is the score of the following alignment:

```
ADAD-RQQ
|||:  :|
ADACN-DQ
```

○ **A.** -2

○ **B.** -1

○ **C.** 0

○ **D.** +1