

## Indian Institute of Information Technology Allahabad

Review Test - First Component (C1) (September 2022)  
B.Tech (IT, ECE), M.Tech

Course Name	Course Code	Date of Exam	MM	Time
Information Retrieval	INR	September 15th, 2022	15	1 Hr

*Important Instructions: All questions are compulsory.*

1. Draw the inverted index for the following collection of documents . (2 marks)
- Doc 1: machine learning is a type of artificial intelligence  
 Doc 2: machine learning is the study of computer algorithms  
 Doc 3: banks use artificial intelligence systems  
 Doc 4: artificial intelligence uses many algorithms  
 Doc 5: banks use machine learning
2. Recommend a query processing order for (information OR learning) AND (machine OR intelligence) AND (artificial OR Retrieval ) ; and below table gives the information about size of posting for a particular term: (2 marks)

Term	Postings 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

3. Write down differences between wikidata and wikipedia. (1 marks)
4. What are properties and items in wikidata. Write down one example of both. (1 marks)
5. Describe "Wikidata Query Service". (1 marks)
6. Compare Boolean model and inverted index model with the help attributes for which one is better than other? (2 mark)
7. Write down Boolean query for searching relevant documents for given two requirements. Use given term-document matrix and find out relevant documents for both the generated queries by Boolean operations. (3 marks)
- (a) documents which contain Term\_2 and Term\_3 but not term\_4  
 (b) documents which contain Term\_2 or Term\_3 but not term\_4

Term/Document	Docum <sub>1</sub>	Docum <sub>2</sub>	Docum <sub>3</sub>	Docum <sub>4</sub>	Docum <sub>5</sub>	Docum <sub>6</sub>	Docum <sub>7</sub>
term <sub>1</sub>	0	1	0	1	0	1	0
term <sub>2</sub>	1	0	1	0	1	0	1
term <sub>3</sub>	1	1	1	1	1	0	1
term <sub>4</sub>	0	0	0	0	1	0	1
term <sub>5</sub>	1	1	0	0	1	1	0
term <sub>6</sub>	0	0	1	1	0	0	1

8. Suppose a vocabulary has total 10 terms and terms are stored in vocabulary in ordered manner (term\_1, term\_2, term\_3 ... term\_10). Write down one-hot vector of term\_2. Suppose a sentence sentence\_1 = "term\_1 term\_2 term\_3" and sentence\_2 = "term\_2 term\_3 term\_6". Write down one-hot representation of sentence\_1 and sentence\_2 and calculate cosine similarity between them. (2 marks)

9. Generate a factoid question which has answer in given paragraph. (1 marks)

Information retrieval (IR) in computing and information science is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science[1] of searching for information in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds.