University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement an inverted index construction algorithm for the following two documents document1 = "The computer science students are appearing for practical examination." document2 = "computer science practical examination will start tomorrow." Build a simple document retrieval system using the constructed index to find the documents containing terms "computer science". [20] B. Build a question-answering system using techniques such as information extraction. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement an inverted index construction algorithm for the following two documents: [20]
	 document1 = "The quick brown fox jumped over the lazy dog"
	 document2 = "The lazy dog slept in the sun"
	 Build a simple document retrieval system using the constructed index to find the documents containing terms "lazy sun"
	
	 B. Write a program to calculate precision, recall and F-measure where true positive is 60, false positive is 30 and false negative is 20. [20]
	     Recall= TP /TP+FN
	     Precision = TP / TP+FP
	     F-score = 2 * (Precision * Recall) / (Precision + Recall)
		40

2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Develop a spelling correction module using edit distance algorithms and find the edit distance between strings "nature" and "creature". [20] B. Implement the Boolean retrieval model for the following corpus. [20] Document 1:The cat chased the dog around the garden. Document2: She was sitting in the garden last night. Document 3: I read the book the night before. Process the query "garden or night".	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Develop a web crawler to fetch and index web pages. [20] B. Handle challenges such as robots.txt, dynamic content, and crawling delays. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider a simplified web graph with the following link structure: • Page A has links to pages B, C, and D. • Page B has links to pages C and E. • Page C has links to pages A and D. Apply the PageRank algorithm and analyze the results. [20] B. Implement a text summarization algorithm (e.g., extractive or abstractive). [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement cosine similarity to find similarity between following query and document. [20] query="gold silver truck" document="shipment of gold damaged in a gold fire" B. Build a question-answering system using techniques such as information extraction. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Develop a web crawler to fetch and index web pages. [20] B. Handle challenges such as robots.txt, dynamic content, and crawling delays. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement a text summarization algorithm (e.g., extractive or abstractive). [20] B. Implement the Boolean retrieval model for the following corpus [20] Document 1:BSc lectures start at 7. Document 2:My lectures are over. Document 3: Today is a holiday. Process the query "not lectures"	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement the vector space model with TF-IDF weighting for the following corpus: [20] Document 1: "Document about python programming language and data analysis.", Document 2: "Document discussing machine learning algorithms and programming techniques.", Document3: "Overview of natural language processing and its applications." query = "python programming" B. Calculate the cosine similarity between the query and each document from the above problem. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Write a program to calculate precision, recall and F-measure where true positive is 60, false positive is 30 and false negative is 20. [20] Recall = Recall= TP /TP+FN Precision =Precision = TP / TP+FP F-score = 2 * (Precision * Recall) / (Precision + Recall) B. Use an evaluation toolkit to measure average precision and other evaluation metrics. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Given a dataset of 20_newsgroups, build a classification model using Naïve Bayes classifier that can accurately classify each review as either positive or negative. [20] B. Given a dataset of 20_newsgroups, build a classification model using SVM classifier that can accurately classify documents into categories like "alt.atheism," "soc.religion.christian," "comp.graphics," or "sci.med," [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider the following corpus:
	  "India has the second-largest population in the world.",
	 " It is surrounded by oceans from three sides which are Bay Of Bengal in the east, the Arabian Sea in the west and Indian oceans in the south.",
	 "Tiger is the national animal of India.",
	 "Peacock is the national bird of India.",
	 "Mango is the national fruit of India."
	  Build a question-answering system and query for "Which is the national bird of India?" [20]
	
	B. Write a program to calculate precision, recall and F-measure where true positive is 60, false positive is 30 and false negative is 20. [20]
	Recall = Recall= TP /TP+FN
	Precision =Precision = TP / TP+FP
	 F-score = 2 * (Precision * Recall) / (Precision + Recall)
		40

2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider a simplified web graph with the following link structure: • Page A has links to pages B, C, and D. • Page B has links to pages C and E. • Page C has links to pages A and D. Apply the PageRank algorithm and analyze the results. [20] B. Write a program to calculate precision, recall and F-measure where true positive is 20, false positive is 10 and false negative is 30. [20] Recall = Recall= TP /TP+FN Precision =Precision = TP / TP+FP F-score = 2 * (Precision * Recall) / (Precision + Recall)	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider a web graph with the following link structure: [20] • Page A has links to pages B and C. • Page B has links to pages C and D. • Page C has links to pages A and D. • Page D has a link to page B Apply the PageRank algorithm and analyze the results. B. Implement a text summarization algorithm [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement the vector space model with TF-IDF weighting for the following corpus: [20] Document 1: "Document about python programming language and data analysis.", Document 2: "Document discussing machine learning algorithms and programming techniques.", Document3: "Overview of natural language processing and its applications." query = "python programming" B. Implement cosine similarity to find similarity between following query and document. [20] query="gold silver truck" document="shipment of gold damaged in a gold fire"	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Apply the clustering algorithm(e.g., K-means or hierarchical clustering) for a given set of documents. [20] "Machine learning is the study of computer algorithms that improve through experience.", "Deep learning is a subset of machine learning.", "Natural language processing is a field of artificial intelligence.", "Computer vision is a field of study that enables computers to interpret the visual world.", "Reinforcement learning is a machine learning algorithm.", "Information retrieval is the process of obtaining information from a collection.", "Text mining is the process of deriving high-quality information from text.", "Data clustering is the task of dividing a set of objects into groups.", "Hierarchical clustering builds a tree of clusters.", "K-means clustering is a method of vector quantization." B. Given a set of predicted binary labels (y_pred) and their corresponding true binary labels (y_true), evaluate the performance by calculating precision, recall, f1-score and average precision of a binary classification model using standard metrics from scikit-learn module [20] y_true = [0, 1, 1, 0, 1] y_scores = [0.1, 0.8, 0.6, 0.3, 0.9]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement an inverted index construction algorithm for the following two documents: [20] document1 = "our class meeting starts soon" document2 = "my class starts at 6." B. Build a simple document retrieval system using the constructed index to find the documents containing terms "class meeting". [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Implement the vector space model with TF-IDF weighting for the following corpus: [20] Document 1: "The sun is the star at the center of the solar system.", Document2: "She wore a beautiful dress to the party last night." Document 3: "The book on the table caught my attention immediately." query = "solar system" B. Implement a clustering algorithm (e.g., K-means or hierarchical clustering). Also, apply the clustering algorithm to a set of documents and evaluate the clustering results. [20]	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider a simplified web graph with the following link structure: • Page A has links to pages B, C, and D. • Page B has links to pages C and E. • Page C has links to pages A and D. Apply the PageRank algorithm and analyze the results. [20] B. Implement the Boolean retrieval model for the following corpus [20] Document 1:The university exam is scheduled next week. Document2: The university of mumbai has declared the result. Process the query "university and mumbai"	40
2.	Viva	05
3.	Journal	05

University Of Mumbai

T.Y. B.Sc. Sem VI (Computer Science) Rev 21 Practical Examination – April 2025

Duration: 2 hrs   **Marks:** 50

Date: 15 April 2025   **Time:** 9:00 - 11:00

Course: USCSP6032 Information Retrieval – Practical

Candidate's University Seat Number: _____

1.	A. Consider the following corpus: "India has the second-largest population in the world.", " It is surrounded by oceans from three sides which are Bay Of Bengal in the east, the Arabian Sea in the west and Indian oceans in the south.", "Tiger is the national animal of India.", "Peacock is the national bird of India.", "Mango is the national fruit of India." Build a question-answering system and query for "Which is the national bird of India?" [20] B. Create a system that can summarize text data. The given text data is as follows: [20] Natural language processing (NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction. Many challenges in NLP involve natural language understanding, natural language generation, and machine learning. Text summarization is the process of distilling the most important information from a source (text) to produce an abridged version for a particular user or task. Automatic text summarization methods are greatly needed to address the ever-growing amount of text data available online to both better help discover relevant information and to consume the vast amount of text data available more efficiently.	40
2.	Viva	05
3.	Journal	05