

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Баламирзоев Назим Лиодинович
Должность: Ректор
Дата подписания: 25.03.2026 16:00:30
Уникальный программный ключ:
5cf0d6f89e80f49a334f6a4ba58e91f3326b9926



МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
Институт кибербезопасности и цифровых технологий
Региональный партнёр
ФГБОУ ВО
«Дагестанский государственный технический университет»



1. .08

Направление подготовки 09.03.01 «Информатика и вычислительная техника»

Направленность (профиль подготовки): «Прикладной искусственный интеллект»

Квалификация выпускника бакалавр

Форма обучения очная

Махачкала 2023

ПАСПОРТ фонда оценочных средств

по дисциплине Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ

1. Результаты обучения по дисциплине:

Код	Наименование компетенции	Индикатор достижения компетенции	В результате освоения дисциплины обучающийся должен:	Другая дисциплина (дисциплины) / практика, участвующая в формировании компетенции
ПК-6	Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК 6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях	<p>Знать:</p> <ul style="list-style-type: none"> - концептуальные положения современной науки о данных (Data Science); - модели представления больших данных и способы организации структур их хранения; - области применения больших данных и их преимущества; - основные источники поиска и добычи больших данных; - основы логической и физической организации распределенных файловых систем для хранения больших данных; - основы обработки и анализа потоковых данных; - методы машинного обучения и интеллектуального анализа больших данных. - этапы обработки и аналитики больших данных; - принципы формирования поисковых запросов для открытых источников, специализированных библиотек и репозиториев. <p>Уметь:</p> <ul style="list-style-type: none"> - применять методы обработки и интерпретации больших данных; - применять методы 	Основы обучаемых алгоритмов, Практики: Производственная (технологическая) практика

			<p>обработки потоковых данных;</p> <ul style="list-style-type: none">- использовать распределенную инфраструктуру для поиска в открытых источниках, специализированных библиотеках и репозиториях;- решать задачи анализа ссылок и потоков данных;- решать задачи частых предметных наборов;- использовать основные алгоритмы кластеризации и поиска похожих объектов в структурах больших данных;- применять языки программирования, скриптовые языки и специализированное ПО для обработки больших данных. <p>Владеть:</p> <ul style="list-style-type: none">- основными современными технологиями поиска данных в открытых источниках, специализированных библиотеках и репозиториях;- языком формирования поисковых запросов;- навыками представления и обработки больших социальных графов;- технологией настройки поисковых критериев и применения автоматических поисковых систем;- современными технологиями хранения и обработки больших данных.	
--	--	--	--	--

2. Программа оценивания контролируемой компетенции:

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции / индикатора	Наименование оценочного средства
1	Раздел 1. Введение в проблематику сбора и обработки больших данных (Big Data).	ПК-6	Собеседование при защите лабораторных работ 1, проверка заданий для самостоятельной работы.
2	Раздел 2. Программный стек для обработки больших данных.	ПК-6	Собеседование при защите лабораторных работ 2, проверка заданий для самостоятельной работы.
3	Раздел 3. Задачи обработки больших данных	ПК-6	Собеседование при защите лабораторных работ 3-8, проверка заданий для самостоятельной работы.
Форма промежуточной аттестации в 6 семестре – экзамен			

Вопросы для экзамена по дисциплине **Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

Вопросы:

1. Наука о данных (Data Science) – история возникновения, этапы развития, цели, методы и области применения.
2. Понятие больших данных (Big Data). Актуальность и сферы применения технологии больших данных.
3. Подходы к структурированию и типы данных в Data Science.
4. Структурированные и неструктурированные данные. Примеры из различных предметных областей.
5. Виды больших данных.
6. Классификация источников данных. Проблемы поиска и представления больших данных.
7. Задача предварительной оценки и подготовки данных.
8. Витрины данных (data marts), склады данных (data warehouses), озера данных (data lakes).
9. Поставщики открытых данных. Проверка качества данных.
10. Основы статистического моделирования.
11. Добыча данных как алгоритмическая задача.
12. Проблема обобщения данных и выделения признаков.
13. Статистические пределы сбора больших данных. Принцип Бонферрони. Мера важности данных.
14. Хэш-функции и их применение в задачах индексации больших данных.
15. Особенности работы с внешней памятью при обработке больших данных.

16. Логическая и физическая организация вычислительных узлов для обработки больших данных.
17. Обзор существующих распределенных файловых систем. Принципы организации больших файловых систем.
18. Общая характеристика технологии MapReduce. Схема вычислений по технологии MapReduce.
19. Задачи-распределители Map. Группировка по ключу.
20. Задачи-редукторы Reduce. Комбинаторы функций Reduce.
21. Схема выполнения MapReduce-программы. Обработка отказов узлов.
22. Основные алгоритмы и операции с использованием MapReduce.
23. Системы потоков работ. Понятие потокового графа и функции потока.
24. Системы Clusters и Hadoop.
25. Рекурсивные обобщения MapReduce.
26. Реализация транзитивного замыкания с помощью набора рекурсивных задач.
27. Задачи соединения и задачи устранения дубликатов.
28. Система Pregel и графовое представление данных.
29. Супершаговые вычисления Pregel.
30. Модели измерения качества алгоритмов, работающих в вычислительном кластере.
31. Понятие коммуникационной стоимости задачи. Многопутевые соединения.
32. Основы теории сложности MapReduce.
33. Постановка задачи о сходстве объектов. Понятие меры расстояния (метрики).
34. Сходства множеств по Жаккару. Сходство текстовых документов.
35. Понятие коллаборативной фильтрации.
36. Задача разбиения документов на шинглы.
37. Хэширование документов с учетом близости. Теория функций, учитывающих близость.
38. LSH-семейство хэш-функций.
39. Поточковая модель данных. Системы управления потоками данных.
40. Проблема выборки данных из потока. Фильтрация потоков.
41. Задача подсчета элементов в потоке.
42. Задача вычисления распределения частоты элементов в потоке (задача оценивания моментов).
43. Задача подсчет единиц в окне и затухающие окна.
44. Постановка задачи анализа ссылок в поисковых системах. Функция PageRank и особенности ее алгоритмической реализации.
45. Графовая структура веб-пространства. Проблема тупиков и паучьих ловушек.
46. Варианты использования PageRank в поисковой системе. Эффективное вычисление PageRank.
47. Тематический PageRank и алгоритм TrustRank.
48. Алгоритм HITS.
49. Понятие частого предметного набора. Модель корзины покупок.
50. Алгоритм Apriori.
51. Использование оперативной памяти для подсчета предметных наборов. Обработка больших наборов данных в оперативной памяти.
52. Алгоритм Парка-Чена-Ю (PCY).
53. Алгоритм поиска частых предметных наборов с ограниченным числом проходов.
54. Задача подсчета частых предметных наборов в потоке.
55. Обзор методов и стратегий кластеризации. Классификация алгоритмов кластеризации.
56. Иерархическая кластеризация в евклидовом пространстве и оценка ее эффективности.
57. Иерархическая кластеризация в неевклидовых пространствах.
58. Кластеризация в неевклидовых пространствах (общий обзор методов).
59. Применение алгоритма k-средних для кластеризации больших данных.

60. Алгоритм кластеризации больших наборов данных CURE (Clustering Using REpresentatives).
61. Графовые модели социальных сетей. Метрики для графов социальных сетей.
62. Задача кластеризации графа социальной сети.
63. Алгоритм Гирвана-Ньюмана.
64. Промежуточность и ее использование для нахождения сообществ. Прямое нахождение сообществ.
65. Задача разрезания графа социальной сети.
66. Нахождение пересекающихся сообществ.
67. Метод анализа графов социальных сетей Simrank.
68. Задача подсчета треугольников и других простых подграфов.
69. Задачи о путях и окрестностях в больших графах. Понятие диаметра графа.

Описание показателей и критериев оценивания с указанием шкалы оценивания для очной и других форм обучения (с применением балльно-рейтинговой системы и/или без ее использования):

Зачет по дисциплине складывается из текущего рейтинга и зачетного рейтинга.

Минимальный балл текущего рейтинга в соответствии с положением о рейтинге равен 36, максимальный – 60.

Текущий рейтинг складывается по результатам сдачи лабораторных работ (максимум – 60 баллов)

Экзаменационный рейтинг определяется ответами на вопросы – до 40 баллов.

36-40 баллов выставляется, если студент демонстрирует полное понимание вопросов, правильность ответов, полное и логически последовательное изложение материала.

30-35 баллов выставляется, если студент демонстрирует: значительное понимание вопросов, правильность, но недостаточную полноту ответов на заданные теоретические вопросы; допущение неточности ответа;

15-29 баллов выставляется, если студент демонстрирует: понимание вопросов, по существу излагает материал, но не усвоил его деталей, есть погрешности в ответах; допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении материала;

Менее 15 баллов выставляется, если студент демонстрирует: непонимание вопросов; студент не знает значительной части материала, не ответил на дополнительные вопросы или отказался от ответов на вопросы и задания.

В итоге по курсу, суммируя итоги текущего рейтинга и рейтинга, полученного за ответ на экзамене, студент может набрать от 60 баллов до 100 баллов

- оценка «отлично» выставляется обучающемуся, если он набрал от 86 до 100 баллов;
- оценка «хорошо» выставляется обучающемуся, если он набрал от 75 до 85 баллов;
- оценка «удовлетворительно» выставляется обучающемуся, если он набрал от 60 до 74 баллов;
- оценка «неудовлетворительно» выставляется обучающемуся, если он набрал менее 61 балла.

ПРИМЕРНЫЙ ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ

№ п/ п	Наименование оценочного средств	Краткая характеристика оценочного средства	Представлен ие оценочного средства
1	2	3	4
1	Собеседование	Средство контроля, организованное как специальная беседа преподавателя с обучающимся в ходе сдачи лабораторных работ на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний, умений и навыков обучающегося по определенному разделу, теме, проблеме и т.п.	Вопросы для собеседования при сдаче лабораторных работ

Вопросы для собеседований при защите лабораторных работ по дисциплине **Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

Вопросы для собеседования при защите лабораторной работы №1

1. Приведите примеры использования больших данных в различных предметных областях.
2. Сформулируйте набор типовых требований, предъявляемых для аппаратного обеспечения комплекса для работы с большими данными.
3. Сформулируйте набор типовых требований, предъявляемых для программного обеспечения комплекса для работы с большими данными.
4. Сформулируйте понятие и назовите основные характеристики больших данных (Big Data). Назовите основные проблемы и трудности при работе с большими данными.
5. Какие работы обычно выполняются на этапе предварительной обработки данных?
6. Назовите основные подходы к структурированию данных.
7. В чем принципиальное отличие структурированных и неструктурированных данных?
8. В чем принципиальное отличие машинных данных и данных на естественном языке?
9. Какие работы обычно выполняются на этапе разведывательного анализа данных?
10. Перечислите основные типы мультимедиа данных.
11. Что понимается под термином «поточковые данные»? Приведите примеры.
12. Назовите основные положительные и отрицательные свойства языка Python.
13. Какие факторы, на Ваш взгляд, способствуют удобству и популярности языка Python в сфере анализа данных?
14. Какие основные типы данных используются в Python? Какие виды числовых данных Вы знаете?
15. Перечислите наиболее популярные математические библиотеки Python и их назначение.
16. Опишите основные синтаксические особенности работы с переменными в языке Python.
17. Что такое список? Какими способами можно добавлять и удалять элементы списка?
18. Могут ли содержаться в списке элементы разных типов?
19. Что понимается под срезом списка?
20. Есть ли в Python возможность индексации с обратного конца списка? Если да, то каким образом?
21. Назовите простейший способ объединения двух списков.
22. Назовите основное отличие между определенными и неопределенными циклами.
23. Как реализуются определенные циклы в Python?
24. Как реализуются неопределенные циклы в Python?
25. Назовите основной способ реализации вложенных конструкций в Python.
26. Что понимается под словарем (диктом)? Как осуществляется доступ к элементам словаря?
27. Назовите известные Вам способы итерирования словарей.
28. Что такое пара ключ-значение и какова ее роль в ассоциативных типах данных? Приведите пример.
29. Назовите простейший способ добавление пары ключ-значение в словарь.
30. Какие типы данных могут использоваться для ключей словаря? Можно ли использовать в качестве ключа другой словарь?
31. Какие типы данных могут использоваться для значений словаря? Можно ли использовать в качестве значения другой словарь?
32. В чем основное назначение библиотеки Pandas?
33. Что такое «серия» в Pandas? Какая связь между сериями Pandas и словарями Python?

34. Что понимается под термином «датафрейм»?
35. Опишите известные Вам способы объединения двух датафреймов.
36. Какие режимы объединения датафреймов Вам известны и в чем их отличие?
37. В чем отличие между методами и атрибутами датафрейма? Приведите примеры известных Вам методов и атрибутов.
38. Возникла задача переименования некоторых строк и столбцов датафрейма. Методы или атрибуты Вы будете использовать для ее решения? Какие?
39. Необходимо определить тип каждого столбца датафрейма. Методы или атрибуты Вы будете использовать для решения этой задачи? Какие?
40. Что понимается под векторизацией в Pandas?
41. Перечислите возможные ошибки данных, подлежащие очистке.
42. Приведите примеры ошибок данных при вводе.
43. Приведите примеры физически невозможных данных.
44. Приведите примеры отсутствующих значений в массивах данных.
45. Выбросы, пробелы и опечатки – примеры и способы очистки.
46. Отклонения от свода правил - примеры и способы очистки.
47. Агрегирование данных – суть, примеры и способы реализации.
48. Экстраполяция – суть, примеры и способы реализации.
49. Производные метрики – суть, примеры и способы реализации.
50. Опишите процедуру создания вспомогательных переменных.
51. В чем необходимость процесса сокращения количества переменных?

Вопросы для собеседования при защите лабораторной работы №2

1. Назовите известные вам способы логической организации вычислительных узлов. Дайте их характеристики (преимущества/недостатки).
2. Назовите известные вам способы физической организации вычислительных узлов. Дайте их характеристики (преимущества/недостатки).
3. Есть ли зависимость физической организации вычислительного узла от выбранной логической схемы?
1. Назовите основные критерии принадлежности данных.
2. Охарактеризуйте основные свойства внутренних и внешних данных. В чем их особенности хранения и использования?
3. Что такое «витрина данных»? Приведите примеры использования витрин данных.
4. Что такое «склад данных»? Приведите примеры использования складов данных.
5. Что такое «озеро данных»? Приведите примеры использования озер данных.
6. В чем суть инвентаризации информационных ресурсов?
7. Что такое «открытые данные»? Приведите примеры.
8. Особенности работы с поставщиками открытых данных.
9. Назовите основные качественные критерии при проверке открытых данных.
10. Платформа Hadoop – назначение и характеристики.
11. Требования к аппаратному обеспечению для развертывания платформы Hadoop.
12. Основные этапы загрузки в платформу больших данных Hadoop.
13. Какие вам известны технологии для реализации параллелизма в Hadoop?
14. Охарактеризуйте суть технологии MapReduce.
15. Что понимается под задачей-распределителем Map? В чем суть фазы отображения (Map) данных?
16. Что такое группировка по ключу?
17. Что понимается под задачей-редуктором Reduce? В чем суть фазы свертки (Reduce) данных?
18. Что понимается под комбинатором функций Reduce?
19. Перечислите основные алгоритмы и операции с использованием MapReduce.

20. Как обеспечивается хранение данных в Apache Hadoop?
21. С какими типами СУБД возможна работа в Apache Hadoop?
22. Каким образом в кластерах осуществляется параллельная обработка информации?
Расскажите о структуре узлов в кластере.
23. Опишите стандартный вариант схемы выполнения MapReduce-программы.
24. Каким образом осуществляется обработка отказов узлов MapReduce-процесса?
25. Каковы недостатки подхода MapReduce? Какие пути их устранения существуют на данный момент?
26. Как в Hadoop обеспечивается управление ресурсами?
27. Каким образом в Hadoop обеспечивается мониторинг состояний компонентов системы и используемых ими ресурсов?
28. Дайте определение потокового графа и функции потока.
29. Какие вам известны основные рекурсивные обобщения технологии MapReduce.
30. Раскройте понятие коммуникационной стоимости задачи.

Вопросы для собеседования при защите лабораторной работы №3

1. По каким критериям можно оценить сходство двух множеств?
2. Объясните суть метода оценки сходства множеств «по Жаккару».
3. В чем математический смысл коэффициента Жаккара?
4. Можно ли оценить сходство двух документов, используя метод Жаккара?
5. Можно ли решить задачу поиска фактов плагиата, используя метод Жаккара?
6. Назовите основные практические задачи, которые можно решить с помощью метода Жаккара.
7. Объясните связь проблемы коллаборативной фильтрации и задачи о сходстве множеств.
8. Назовите основные практические задачи, которые можно решить с помощью метода коллаборативной фильтрации.
9. Вычислить коэффициент Жаккара для каждой пары следующих трех множеств: $\{1, 2, 3, 4\}$, $\{2, 3, 5, 7\}$, $\{2, 4, 6\}$.
10. Вычислить коэффициент Жаккара для каждой пары следующих трех мультимножеств: $\{1, 1, 1, 2\}$, $\{1, 1, 2, 2, 3\}$, $\{1, 2, 3, 4\}$.
11. Пусть имеется универсальное множество U с n элементами, и мы случайным образом выбираем из него два подмножества S и T по m элементов в каждом. Каково математическое ожидание коэффициента Жаккара S и T ?
12. В чем заключается подход разбиения на шинглы при решении задачи идентификации лексически похожих документов?
13. Что такое k -шингл?
14. Каким образом выбирается размер шингла?
15. Выпишите первые десять 3-шинглов из первого предложения «The most effective way to represent documents as sets, for the purpose of identifying lexically similar documents is to construct from the document the set of short strings that appear within it».
16. Каково максимальное возможное количество k -шинглов в документе, содержащем n байтов? Можете предполагать, что размер алфавита достаточно велик, так что число возможных строк длины k не меньше n .

Вопросы для собеседования при защите лабораторной работы №4

1. В чем различие между потоками и базами данных?
2. Опишите структуру типовой системы управления потоками данных.

3. Приведите примеры источников потоков данных.
4. Назовите два основных способа предъявить запрос к потоку данных.
5. Назовите основные проблемы обработки потоков данных.
6. Перечислите основные известные методы получения надежных выборок из потока данных.
7. Сформулируйте общую постановку задачи о выборке.
8. Что понимается под динамическим изменением размера выборки?
9. Что понимается под фильтрацией потоков данных?
10. В чем назначение фильтра Блума?
11. В чем суть проблемы Count-Distinct?
12. Перечислите основные шаги алгоритма Флажолле-Мартена.
13. Вычислите меру неожиданности (второй момент) для потока 3, 1, 4, 1, 3, 4, 2, 1, 2. Чему равен третий момент этого потока?
14. Пусть в потоке n элементов и m из них различны. Выразите минимальное и максимальное значение меры неожиданности в виде функции от m и n .
15. Существует несколько способов разбить битовый поток 1001011011101 на интервалы. Найдите все.

Вопросы для собеседования при защите лабораторной работы №5

1. Что такое PageRank?
2. В чем основные причины эффективности PageRank?
3. Что такое спам термов?
4. Какая web-страница называется тупиковой?
5. Назовите основные подходы к решению проблемы web-тупиков?
6. Что понимается под термином «паучья ловушка»?
7. Случайно или намеренно образуются структуры типа «паучья ловушка»?
8. Что понимается под термином «телепортация»?
9. Назовите известные вам усовершенствования PageRank. В чем их отличительные особенности?
10. Что такое тематический PageRank? Для чего он нужен?
11. Какие шаги необходимо предпринять для того, чтобы включить тематический PageRank в поисковую систему?
12. Что понимается под названием «ссылочный спам»?
13. Что такое спам-ферма?
14. Как выглядит web-пространство с точки зрения ссылочного спамера?
15. Какие вам известны методы борьбы со ссылочным спамом?
16. Что такое TrustRank?
17. В чем заключается идея «спамной массы»?
18. Существует ли выгода от объединения спам-ферм?

Вопросы для собеседования при защите лабораторной работы №6

1. В чем практическая ценность задачи обнаружения частых предметных наборов? Что понимается под частым предметным набором?
2. Для чего применяется модель корзины покупок? В чем ее суть?
3. Можно ли представить частый предметный набор с помощью ассоциативных правил вида «если-то»?
4. В чем суть свойства монотонности предметных наборов?
5. Как устроен алгоритм Apriori? В чем его назначение?
6. Как применить алгоритм Apriori для поиска частых предметных наборов?

7. Пусть количество предметов n равно 20. Если воспользоваться треугольной матрицей для подсчета пар, то счетчик какой пары будет храниться в элементе $a[100]$?
8. Как устроен алгоритм РСУ? В чем его назначение?
9. Как можно улучшить алгоритм РСУ?
10. Назовите известные вам двухпроходные алгоритмы поиска всех или большинства частых предметных наборов.
11. Назовите основные пути адаптации алгоритма SON к параллельным вычислениям?
12. Пусть имеется 8 предметов A, B, \dots, H и следующие максимальные частые предметные наборы: $\{A, B\}$, $\{B, C\}$, $\{A, C\}$, $\{A, D\}$, $\{E\}$, $\{F\}$. Найдите негативную кайму.
13. Пусть предмет i встречается ровно s раз в файле, содержащем n корзин, где s – пороговая поддержка. Если взять выборку из $n/100$ корзин и уменьшить порог для выборки до $s/100$, то какова вероятность, что i окажется частым? Можете предполагать, что s и n делятся на 100.
14. В чем основное различие между потоками и файлами с точки зрения частых предметных наборов?
15. Предположим, что подсчитываются частые предметные наборы в затухающем окне с коэффициентом затухания c . Предположим также что с вероятностью p данный элемент потока (корзина) содержит оба предмета i и j . Кроме того, с вероятностью p корзина содержит i , но не содержит j , и с такой же вероятностью p содержит j , но не содержит i . Выразите в виде функции от c и p долю времени, в течение которого у пары $\{i, j\}$ будет существовать оценка.

Вопросы для собеседования при защите лабораторной работы №7

1. Какие наборы данных подходят для кластеризации?
2. Приведите примеры метрик в пространствах, пригодных для кластеризации?
3. В чем отличия иерархических алгоритмов кластеризации от алгоритмов на основе отнесения точек?
4. По каким признакам можно классифицировать алгоритмы кластеризации?
5. Что понимается под термином «проклятие размерности» в многомерных евклидовых пространствах?
6. Докажите, что математическое ожидание расстояния между двумя случайными независимыми и равномерно распределенными точками отрезка прямой длиной 1 равно $1/3$?
7. Каково математическое ожидание евклидова расстояния между двумя случайными независимыми и равномерно распределенными точками единичного квадрата?
8. Каким образом работают алгоритмы иерархической кластеризации?
9. Какие подходы существуют для принятия решения об остановке процесса кластеризации?
10. Как повысит эффективность базового алгоритма иерархической кластеризации?
11. Возможна ли иерархическая кластеризация в неевклидовых пространствах?
12. Что такое кластроид?
13. Какими способами выбирается кластроид?
14. Выполните иерархическую кластеризацию одномерного множества точек 1, 4, 9, 16, 25, 36, 49, 64, 81 в предположении, что кластер представлен своим центроидом (средней точкой) и на каждом шаге объединяются кластеры с наименьшим расстоянием между центроидами?
15. Опишите и прокомментируйте структуру алгоритма k-средних?
16. Каким образом происходит инициализация кластеров в алгоритме k-средних?
17. Как правильно выбрать значение k в алгоритме k-средних?
18. Объясните суть алгоритма CURE?

19. Перечислите основные этапы инициализации в алгоритма CURE.
20. Как решается задача кластеризации потока? Какие вам известны алгоритмы решения этой задачи?
21. Как решается задача кластеризации в параллельной среде?

Вопросы для собеседования при защите лабораторной работы №8

1. Что означает свойство локальности социальной сети?
2. Перечислите основные отличительные особенности социальной сети.
3. Что такое социальный граф?
4. Назовите основные разновидности социальных сетей.
5. Перечислите основные типы вершин социального графа.
6. Для каких практических целей решается задача кластеризации социальных графов?
7. Что такое метрика графа социальной сети? Приведите примеры таких метрик.
8. Дают ли приемлемое решение задачи кластеризации социальных графов стандартные методы кластеризации? Обоснуйте ответ.
9. Что понимается под термином промежуточности ребер социального графа?
10. Опишите назначение и основные шаги алгоритма Гирвана-Ньюмана (GN)?
11. Как использовать свойство промежуточности ребер для нахождения сообществ? Есть ли ограничения у этого метода?
12. Опишите прямой способ нахождения сообществ посредством поиска подмножеств вершин?
13. Как можно использовать задачу разрезания графа к организации графов социальных сетей?
14. Что понимается под нормализованным разрезом графа?
15. Сформулируйте постановку задачи нахождения пересекающихся сообществ?
16. Опишите модель графа принадлежности (affiliation-graph model)?
17. Что такое Simrank?
18. Что понимается под случайным блужданием в социальном графе?
19. Что такое случайное блуждание с перезапуском?
20. В чем заключается операция подсчета треугольников, применительно к графам социальных сетей? Можно ли решать эту задачу с помощью MapReduce?
21. Назовите известные вам алгоритмы решения задач о путях и окрестностях в больших графах?
22. Что понимается под диаметром орграфа?
23. Что понимается под транзитивным замыканием графа? Как с этим связано понятие достижимости вершин?
24. Опишите процесс вычисления транзитивного замыкания методом рекурсивного удвоения.
25. В чем основные отличия интеллектуального транзитивного замыкания в большом графе?

Описание показателей и критериев оценивания с указанием шкалы оценивания для очной и других форм обучения (с применением балльно-рейтинговой системы и /или без ее использования).

По результатам собеседования при сдаче каждой из лабораторных работ студенту начисляется от 0 до 10 баллов:

Оцениваются следующие показатели:

– 8-10 баллов выставляется, если студент выполнил работу в установленный срок, правильно и полно отвечает на вопросы, объясняет их на примерах, связывает с программной реализацией.

– 5-7 баллов выставляется, если студент отвечает на вопросы, объясняет их на примерах, связывает с программной реализацией, но недостаточно полно и с некоторыми неточностями.

– 2-4 балла выставляется, если студент отвечает на большинство из заданных вопросов, и может объяснить ход их решения на примере

– менее 2 баллов выставляется, если студент не отвечает на большую часть заданных вопросов, не может объяснить их на примере.

Сведения о дополнениях и изменениях, внесенных в ФОС дисциплины

Учебный год	Решение кафедры (№ протокола, дата)	Внесенные в ФОС дополнения и изменения	Подпись заведующего кафедрой