

Демонстрационный вариант и методические рекомендации
по направлению «Бизнес-информатика»

Профиль:

«Системы больших данных / Big Data Systems»

ДЕМОНСТРАЦИОННЫЙ ВАРИАНТ

Время выполнения задания – 180 мин.

Ответы могут быть даны на русском или
английском языке

**Методика проверки и предварительные критерии оценивания работ участников
олимпиадных состязаний**

Результаты олимпиадного задания оцениваются по 100-балльной системе. Проверяется только чистовик. Максимальные баллы по пунктам задания:

1. Дайте развернутые ответы

1.1 – 15 баллов

1.2 – 15 баллов

2. Решите задачи

2.1. – 15 баллов

2.2. – 10 баллов

2.3. – 15 баллов

2.4. – 15 баллов

2.5. – 15 баллов

3. Дифференцирование оценок. Дифференцированные оценки возможны только в заданиях типа 1 (развернутые ответы). Для заданий типа 2 (решение задач) дифференцированные оценки в основном не предусмотрены: выставляется либо максимальная оценка при правильном решении и ответе, либо ноль баллов. Дифференцирование возможно только в задачах, состоящих из нескольких частей, при которых каждая часть может быть оценена независимо, либо в заданиях, предполагающих наличие вариантов правильных решений.

4. Ответы на вопросы могут быть даны на русском или английском языке. При оценке ответа выбор языка не учитывается.

1. Дайте развернутые ответы

1.1. Advantages and disadvantages of using models of cloud computing in the enterprise

1.2. Виды информации и формы ее представления, используемые при принятии управленческих решений на предприятии

2. Решите задачи.

2.1. В реляционной базе данных предприятия хранятся данные о сотрудниках и выполненных ими проектах. В базе хранится следующая информация о сотрудниках: ФИО; дата рождения; пол; адрес проживания; телефон; наименование подразделения, в котором работает сотрудник; должность. Каждый сотрудник может работать только в одном подразделении, однако, может участвовать в одном или нескольких проектах. Информация о проектах включает: наименование проекта, дату начала, дату окончания, руководителя проекта, исполнителей проекта. Требуется:

Олимпиада для студентов и выпускников вузов – 2015 г.

1. используя любую общепринятую нотацию, нарисовать схему базы данных, удовлетворяющую третьей нормальной форме, с выделением первичных и внешних ключей, типа и направления связи;
2. сделать подробное описание таблиц с расшифровкой имен полей, типов и свойств данных;
3. используя операторы языка SQL, написать один запрос для получения следующей информации:
наименование подразделения, список сотрудников подразделения, с указанием выполняемых ими проектов и роли в проекте (руководитель/исполнитель) за определенный период времени

2.2. Компания использует четыре источника поставки деталей одного вида для производства устройств N и M – фирмы A, B, C и D. Фирма A поставляет 40% деталей, B – 30%; C – 20%; D – 10%. Статистика использования поставок показывает, что среди деталей компании A – 3% бракованных деталей; у компании B – 5%; у компании C – 7%; у компании D – 8%. В каждом устройстве N используется 5 деталей. В каждом устройстве M – 3 детали. При использовании хотя бы одной бракованной детали устройства оказываются неработоспособными. В месяц на предприятии изготавливается 100 устройств N и 200 устройств M.

Какова вероятность того, что изготовленное устройство окажется неработоспособным?
Каков процент выпуска годных устройств на предприятии?

2.3. В 1963 году Б. Мандельброт, исследуя цены на хлопок, предположил, что распределение доходности согласуется с устойчивым распределением Леви. Оказалось, что усечённые блуждания Леви очевидным образом наблюдаются в экономике для флуктуаций различных величин на фондовом рынке: ценных бумаг, биржевых фондовых индексов, которые очень хорошо интерпретируются как случайные процессы с независимыми приращениями. Наиболее простой вид плотности двустороннего распределения Леви:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \cos(tx) \exp(-|t|^\alpha) dt, \text{ где } 0 < \alpha \leq 2.$$

При каком целочисленном значении α данное распределение соответствует распределению Гаусса?

2.4. Обозначим: d – некий символ, например, цифра или буква латинского алфавита; d в степени 1 (обозначим d^1) – это другое обозначение символа d .

Если n – натуральное число и n больше 1, то d в степени n (обозначим d^n) – это последовательность, состоящая из n вхождений символа d .

Таким образом, $b^3 g^2 c^4$ – это выражение (или, другими словами, цепочка) $bbbgccccc$.

Выражения вида $x \rightarrow y$, где x – это символ, y – символ или цепочка символов, будем понимать как подстановки, когда левая часть x заменяется на правую часть y .

Пусть задана следующая система подстановок:

$S \rightarrow A C$, (1)

$A \rightarrow p B$, (2)

$C \rightarrow w D$, (3)

$B \rightarrow p A$, (4)

$E \rightarrow w$, (5)

$$D \rightarrow w E, \quad (6)$$

$$B \rightarrow p, \quad (7)$$

$$E \rightarrow w C \quad (8).$$

Требуется, используя введенные выше обозначения и теоретико-множественные обозначения, описать множество L , состоящее из ВСЕХ цепочек в алфавите $\{p, w\}$, которые можно получить подстановками из указанной системы, если всегда первой применяется подстановка (1), т.е. подстановка с левой частью S .

2.5. В 1885 году немецкий ученый Герман Эббингауз опубликовал результаты экспериментов, которые легли в основу теории обучения. Испытуемым предъявлялось на некоторое время множество бессмысленных триграмм (трехбуквенных сочетаний). Спустя определенное количество времени проверялось, насколько точно запоминалось это множество. Оказалось, что относительное количество сохраненной в памяти информации (I) зависело от времени (t) как

$$I \approx \exp(-t/s), \text{ где } s - \text{сила памяти.}$$

Получите зависимость вероятности ошибки запоминания от числа предъявлений множества триграмм.

Программа и методические материалы для подготовки к олимпиаде (магистерская программа «Системы больших данных / Big Data Systems»)

Основы информационных технологий и систем

Понятие информации, ее основные свойства и особенности. Понятие сообщения и его формы, знаки, алфавиты, понятие формального языка. Информация и данные. Конечный вероятностный источник сообщений. Энтропия источника.

Представление информации. Обработка информации. Понятие алгоритма и его свойства. Способы формальной записи алгоритмов. Распределенная обработка информации и проблемы взаимодействия параллельно выполняемых процессов обработки.

Современные методы анализа структурированной и неструктурированной информации. Обработка текстов на естественном языке.

Современные методы хранения и обработки информации.

Основные классы информационных систем. Архитектура информационной системы. Основные элементы информационных систем. Структура и принципы функционирования информационных систем.

Основная литература

1. Новиков Ф.А. Дискретная математика для программистов – СПб.: Питер 2006.
2. Джордж Риз. Облачные вычисления. Изд-во: БХВ-Петербург, 2011.
3. Брандт З. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров. М.: Мир, 2003.
4. В.Грекул, Г. Денищенко, Н.Коровкина. Проектирование информационных систем М.: Изд-во: Бинوم. Лаборатория знаний. Сер.: Основы информационных технологий. 2008.

Дополнительная литература

1. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989.
2. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование. М.: Финансы и статистика, 2010.
3. Е. Горяинова, А.Панков, Е.Платонов. Прикладные методы анализа статистических данных. М.: Изд-во: Высшая Школа Экономики, 2012.
4. В.А.Фомичев. Математические основы представления содержания посланий компьютерных интеллектуальных агентов. М.: Изд-во: Высшая Школа Экономики, 2007.
5. А. Е. Пентус, М. Р. Пентус. Математическая теория формальных языков. Изд-во: Бинوم. Лаборатория знаний, Интернет-университет информационных технологий. Серия: Основы информатики и математики. 2006

Моделирование и оптимизация бизнес-процессов

Методологии описания деятельности. Понятие о моделировании

деятельности. Моделирование деятельности и моделирование процессов. Предметные области в деятельности организации. Уровни описания. Общие принципы моделирования деятельности. Эволюция развития методологий описания. Моделирование организации, модели организационной структуры, моделирование и анализ бизнес-процессов, процессные модели проекта. Методологии и стандарты SADT,

Инструментальные системы для моделирования бизнеса.

Основная литература

1. Хаммер М., Чампи Дж. Реинжиниринг корпорации: Манифест революции в бизнесе. – М.: Изд-во «Манн, Иванов и Фербер», 2011.
2. Шеер А.-Б. Бизнес-процессы. Основные понятия. Теория. Методы. – М.: Весть-МетаТехнология, 1999.

Дополнительная литература

1. Марка Д.А., Мак Гоуэн К. Методология структурного системного анализа и проектирования SADT. – М.: МетаТехнология, 1993.
2. Андерсен Б. Бизнес-процессы. Инструменты совершенствования. – М.: РИА Стандарты и качество, 2007.
3. Каменнова М.С., Громов А.И., Ферапонтов М.М., Шматалюк А.Е. Моделирование бизнеса. Методология ARIS. – М.: Весть-МетаТехнология, 2001.

Базы данных

Виды моделей данных. Структурированные модели данных: сетевые, иерархические, реляционные. Особенности, преимущества и недостатки. Влияние типа модели на особенности проектирования. Понятие предметной области. Способы описания предметной области. ER-модели. Объекты и классы объектов. Атрибуты объектов. Типы объектов. Виды связей. Классы членства.

Проектирование баз данных на основе использование ER-моделей. Алгоритм проектирования (алгоритм перехода от базовой ER-модели). Проектирование с использованием CASE-систем. Целостность баз данных. Классификация ограничений целостности. Причины, вызывающие нарушение ограничений целостности. Процедурный и декларативный способы задания ограничений целостности.

Основная литература

1. Диго С.М. Базы данных. Проектирование и использование. Учебник - М.: «Финансы и статистика», 2005.- 592 с.
2. Мишенин А.И. Теория экономических информационных систем. - М.: «Финансы и статистика», 2008.

Дополнительная литература

1. Диго С. М. Access: учебно-практич пособие. - М.: ТК Велби, изд-во Проспект, 2006
2. Когаловский М. Р. Энциклопедия технологий баз данных. – М.:

«Финансы и статистика», 2008.

3. Харрис Энди. PHP/MySQL для начинающих. Пер. с англ., Кудиц-образ, 2005.
4. Грабер М. SQL. – М.: ЛОРИ, 2001.
5. Дейт К.Дж. Введение в системы баз данных, 6-е изд.: Пер. с англ. – СПб.: Издательский дом «Вильямс», 2000.
6. Виктор Гольцман. MySQL 5.0. Изд-во: Питер, 2009

Теория вероятностей и математическая статистика

Основные понятия. Независимость. Функции распределения. Числовые характеристики случайных величин. Условные вероятности. Биномиальное и нормальное распределения вероятностей. Распределение Пуассона и другие распространенные распределения вероятностей. Устойчивые распределения. Применение вероятностного аппарата к статистической информации. Исследование выборками. Статистические оценки. Их свойства и методы построения

Доверительные интервалы. Проверка гипотез. Дисперсионный анализ. Примеры многомерных распределений, многомерное нормальное распределение. Моменты, ковариационная матрица. Информация Фишера. Свойства оценок максимального правдоподобия. Порядковые статистики.

Основная литература

1. Шведов А.С. Теория вероятностей и математическая статистика. 2-е изд. М.: Изд- во ГУ-ВШЭ, 2005.
2. Шведов А.С. Теория вероятностей и математическая статистика – 2 (пром
3. Мантенья Р., Стенли Х. Введение в эконофизику. Корреляции и сложность в финансах. Пер. с англ., М.: УРСС, 2007.
4. Пугачев В.С. Теория вероятностей и математическая статистика. М.: Физматлит, 2002.
5. Borak S., Hardle W., Weron R. Statistical Tools for Finance and Insurance. Stable Distributions. Springer, 2005.

Дополнительная литература

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и матем
2. Романовский М.Ю., Романовский Ю.М. Введение в эконофизику. Статистические и динамические модели. М.: 2007.