

Документ подписан простой электронной подписью
Информация о владельце:
ФИС: Комарова Светлана Юриевна
Должность: Проректор по образовательной деятельности
Дата подписания: 05.09.2024 12:35:12
Уникальный программный ключ:
43ba42f5deae4116bbfcb9ac98e39108031227e81add207cbee4149f2098d7a~

**Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Омский государственный аграрный университет имени П.А. Столыпина»
Экономический факультет**

ОПОП по направлению подготовки
09.04.02 Информационные системы и технологии

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине**

Б1.В.02 Аналитическая обработка информации

Направленность (профиль) «Информационные системы и технологии»

Обеспечивающая преподавание дисциплины кафедра	Кафедра экономики, бухгалтерского учета и финансового контроля
Разработчик, канд. экон. наук	О.П. Зайцева

ВВЕДЕНИЕ

1. Фонд оценочных средств по дисциплине является обязательным обособленным приложением к Рабочей программе дисциплины.

2. Фонд оценочных средств является составной частью нормативно-методического обеспечения системы оценки качества освоения обучающимися указанной дисциплины.

3. При помощи ФОС осуществляется контроль и управление процессом формирования обучающимися компетенций, из числа предусмотренных ФГОС ВО в качестве результатов освоения дисциплины.

4. Фонд оценочных средств по дисциплине включает в себя: оценочные средства, применяемые для входного контроля; оценочные средства, применяемые в рамках индивидуализации выполнения, контроля фиксированных видов ВАРС; оценочные средства, применяемые для текущего контроля и оценочные средства, применяемые при промежуточной аттестации по итогам изучения дисциплины.

5. Разработчиками фонда оценочных средств по дисциплине являются преподаватели кафедры экономики, бухгалтерского учета и финансового контроля, обеспечивающей изучение обучающимися дисциплины в университете. Содержательной основой для разработки ФОС послужила Рабочая программа дисциплины.

1. ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ ИЗУЧЕНИЯ
учебной дисциплины, персональный уровень достижения которых проверяется
с использованием представленных в п. 3 оценочных средств

Компетенции, в формировании которых задействована дисциплина		Код и наименование индикатора достижений компетенции	Компоненты компетенций, формируемые в рамках данной дисциплины (как ожидаемый результат ее освоения)		
код	наименование		знать и понимать	уметь делать (действовать)	владеть навыками (иметь навыки)
1			2	3	4
Профессиональные компетенции					
ПК-1	Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации	ИД-1 _{ПК-1} Знать: принципы, методы и средства анализа и структурирования профессиональной информации; методы системного и критического анализа; методики разработки стратегии действий для выявления и решения проблемной ситуации; методы исследований; принципы построения математических моделей процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений	основные принципы, методы и средства системного и критического анализа профессионально й информации; методики выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	выбирать принципы, методы и средства системного и критического анализа профессиональной информации; методики выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	применения основных принципов, методов и средств системного и критического анализа профессиональной информации; методики выявления и решения проблемной ситуации; принципов построения математических моделей процессов и объектов принятия решений
		ИД-2 _{ПК-1} Уметь: анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров; применять на практике новые научные принципы и методы исследований; разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений	аспекты анализа профессионально й информации; новые научные принципы и методы исследований; математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем	анализировать профессиональную информацию; применять на практике новые научные принципы и методы исследований; разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем	всестороннего анализа профессиональной информации; применения на практике новых научных принципов и методов исследований; разработки и применения математических моделей процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений
		ИД-3 _{ПК-1} Владеть: навыками подготовки научных докладов, публикаций и аналитических	этапы подготовки научных докладов, публикаций и аналитических обзоров с обоснованными	осуществлять подготовку научных докладов, публикаций и аналитических обзоров с	разработки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями;

		обзоров обоснованными выводами и рекомендациями; методологией системного и критического анализа проблемных ситуаций; методиками постановки цели, определения способов ее достижения, разработки стратегий действий; применения новых научных принципов и методов исследования для решения профессиональных задач.	выводами и рекомендациями; методологию системного и критического анализа проблемных ситуаций; методики разработки стратегий действий для решения профессиональных задач.	обоснованными выводами и рекомендациями; проводить системный и критический анализ проблемных ситуаций; осуществлять разработку стратегий действий для решения профессиональных задач.	методологию системного и критического анализа проблемных ситуаций; методики разработки стратегий действий для решения профессиональных задач.
ПК-4	Способен составлять структуру программного средства, определять необходимые информационные потоки и исследовать варианты структур	ИД-1 _{ПК-4} Знать: методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	разбираться в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	применения методов планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований
		ИД-2 _{ПК-4} Уметь: выбирать методики и шаблоны; планировать проектные работы; описывать бизнес-процессы	разнообразие методик и шаблонов; аспекты планирования проектных работ; бизнес-процессы	выбирать методики и шаблоны; планировать проектные работы; описывать бизнес-процессы	применения выбираемых методик и шаблонов; планирования проектных работ; описания бизнес-процессов
		ИД-3 _{ПК-4} Владеть: навыками разработки методик выполнения аналитических работ; разработки техникокоммерческого предложения и участие в его защите; планирования, организации и контроля аналитических работ в ИТ-проекте	этапы разработки методик выполнения аналитических работ; техникокоммерческого предложения; планирования, организации и контроля аналитических работ в ИТ-проекте	участвовать в разработке методики выполнения аналитических работ; техникокоммерческого предложения и его защите; планирования, организации и контроля аналитических работ в ИТ-проекте	навыками разработки методик выполнения аналитических работ; техникокоммерческого предложения и его защите; планирования, организации и контроля аналитических работ в ИТ-проекте

ЧАСТЬ 2. ОБЩАЯ СХЕМА ОЦЕНИВАНИЯ ХОДА И РЕЗУЛЬТАТОВ ИЗУЧЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Общие критерии оценки и реестр применяемых оценочных средств

2.1 Обзорная ведомость-матрица оценивания хода и результатов изучения учебной дисциплины в рамках педагогического контроля

Категория контроля и оценки		Режим контрольно-оценочных мероприятий				
		само- оценка	взаимо- оценка	Оценка со стороны		Комис- сионная оценка
				препода- вателя	предста- вителя произво- дства	
		1	2	3	4	5
Входной контроль	1			Входное тестирование		
Индивидуализация выполнения*, контроль фиксированных видов ВАРС:	2					
- презентация	2.1			Проверка отчетных материалов		
Текущий контроль:	3					
- Самостоятельное изучение тем		вопросы		опрос		
- в рамках практических (лабораторных) занятий и подготовки к ним	3.1	задания		Проверка отчетных материалов Собеседование		
- в рамках обще- университетской системы контроля успеваемости	3.2			Контроль текущей успеваемости по контрольным неделям, установленным в университете		
Промежуточная аттестация* обучающихся по итогам изучения дисциплины	4			Дифференцированный зачет		
* данным знаком помечены индивидуализируемые виды учебной работы						

2.2 Общие критерии оценки хода и результатов изучения учебной дисциплины

1. Формальный критерий получения обучающимися положительной оценки по итогам изучения дисциплины:	
1.1 Предусмотренная программа изучения дисциплины обучающимся выполнена полностью до начала процесса промежуточной аттестации	1.2 По каждой из предусмотренных программой видов работ по дисциплине обучающийся успешно отчитался перед преподавателем, демонстрируя при этом должный (не ниже минимально приемлемого) уровень сформированности элементов компетенций
2. Группы неформальных критериев качественной оценки работы обучающегося в рамках изучения дисциплины:	
2.1 Критерии оценки качества хода процесса изучения обучающимся программы дисциплины (текущей успеваемости)	2.2. Критерии оценки качества выполнения конкретных видов ВАРС
2.3 Критерии оценки качественного уровня итоговых результатов изучения дисциплины	2.4. Критерии аттестационной оценки качественного уровня результатов изучения дисциплины

2.3 РЕЕСТР
элементов фонда оценочных средств по учебной дисциплине

Группа оценочных средств	Оценочное средство или его элемент
	Наименование
1. Средства для входного контроля	Тестовые вопросы для проведения входного контроля
	Критерии оценки ответов на тестовые вопросы входного контроля
2. Средства для индивидуализации выполнения, контроля фиксированных видов ВАРС	Перечень тем для написания презентации.
	Процедура выбора темы обучающимся
	Критерии оценки презентации
3. Средства для текущего контроля	Вопросы для самостоятельного изучения темы
	Общий алгоритм самостоятельного изучения темы
	Критерии оценки самостоятельного изучения темы
	Вопросы и задания для самоподготовки по темам лабораторных занятий
	Критерии оценки самоподготовки по темам лабораторных занятий
4. Средства для промежуточной аттестации по итогам изучения дисциплины	Программа для сдачи дифференцированного зачета по учебной дисциплине
	Плановая процедура проведения дифференцированного зачета
	Тестовые вопросы для проведения итогового контроля
	Критерии оценки ответов на тестовые вопросы итогового контроля
	Вопросы и практические задания для проведения промежуточной аттестации по дисциплине и критерии оценки

2.4 Описание показателей, критериев и шкал оценивания и этапов формирования компетенций в рамках дисциплины

Индекс и название компетенции	Код индикатора достижений компетенции	Индикаторы компетенции	Показатель оценивания – знания, умения, навыки (владения)	Уровни сформированности компетенций				Формы и средства контроля формирования компетенций
				компетенция не сформирована	минимальный	средний	высокий	
				Оценки сформированности компетенций				
				2	3	4	5	
				Оценка «неудовлетворительно»	Оценка «удовлетворительно»	Оценка «хорошо»	Оценка «отлично»	
				Характеристика сформированности компетенции				
				Компетенция в полной мере не сформирована. Имеющихся знаний, умений и навыков недостаточно для решения практических (профессиональных) задач	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач	Сформированность компетенции в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных практических (профессиональных) задач	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных практических (профессиональных) задач	
Критерии оценивания								
ПК-1	ИД-1пк1	Полнота знаний	Основных принципов, методов и средств системного и критического анализа профессиональной информации; методики выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	не знает основные принципы, методы и средства системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	есть представление о принципах, методах и средствах системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципах построения математических моделей процессов и объектов	ориентируется в принципах, методах и средствах системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципах построения математических моделей процессов и объектов	хорошо разбирается в принципах, методах и средствах системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципах построения математических моделей процессов и объектов	Опрос Проверка выполненных лабораторных работ Презентация Тест Проверочная работа
		Наличие умений	Выбора принципов, методов и средств системного и критического анализа профессиональной информации; методики выявления и решения проблемной ситуации; принципов построения математических моделей процессов и объектов	не умеет выбирать принципы, методы и средства системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	слабо ориентируется в выборе принципов, методов и средств системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципов построения математических моделей процессов и объектов	может выбрать принципы, методы и средства системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципы построения математических моделей процессов и объектов	отлично ориентируется в выборе принципов, методов и средств системного и критического анализа профессиональной информации; методике выявления и решения проблемной ситуации; принципов построения математических моделей процессов и объектов	
		Наличие навыков	Применения	не может применить	с трудом может применить	может применить принципы.	может грамотно применить	

[illegible]

[illegible]

			выводами и рекомендациями; методологию системного и критического анализа проблемных ситуаций; методики разработки стратегий действий для решения профессиональных задач.	критического анализа проблемных ситуаций; методику разработки стратегий действий для решения профессиональных задач.	методологию системного и критического анализа проблемных ситуаций; методику разработки стратегий действий для решения профессиональных задач.	методологию системного и критического анализа проблемных ситуаций; методику разработки стратегий действий для решения профессиональных задач.	методологию системного и критического анализа проблемных ситуаций; методику разработки стратегий действий для решения профессиональных задач.	
ПК-4	ИД-1пк4	Полнота знаний	Методов планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	не знает методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	поверхностно знаком с методами планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	знает методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	знает различные методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	Опрос Проверка выполненных лабораторных работ Презентация Тест Проверочная работа
		Наличие умений	Разбираться в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	не разбирается в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	слабо разбираться в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	разбирается в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	отлично разбирается в методах планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	
		Наличие навыков (владение опытом)	Применения методов планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	не умеет применять методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	может ограниченно применять методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	может применять методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	может грамотно применять методы планирования проектных работ, методы управления, процесс аналитической обработки информации; знает возможности систем поддержки разработки и сопровождения требований	
	ИД-2пк4	Полнота знаний	Разнообразие методик и	не знает разнообразия методик и шаблонов; аспектов	поверхностно знаком с разнообразием методик и	знает разнообразия методик и шаблонов; аспектов	Отлично ориентируется в разнообразии методик и	

ЧАСТЬ 3 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков, характеризующих этапы формирования компетенций

Часть 3.1. Типовые контрольные задания, необходимые для оценки знаний, умений, навыков

3.1.1 . Средства для индивидуализации выполнения, контроля фиксированных видов ВАРС

Место электронной презентации в структуре учебной дисциплины

Разделы дисциплины, освоение которых обучающимися сопровождается или завершается выполнением электронной презентации		Компетенции, формирование/развитие которых обеспечивается в ходе выполнения электронной презентации
№	Наименование	
3	Основные этапы и методы информационно-аналитической работы	ПК-1 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации
4	Работа с источниками информации	ПК-1 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации ПК-4 Способен составлять структуру программного средства, определять необходимые информационные потоки и исследовать варианты структур
5	Способы оценки информации	ПК-1 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации
6	Обеспечение безопасности и защиты информации	ПК-1 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации ПК-4 Способен составлять структуру программного средства, определять необходимые информационные потоки и исследовать варианты структур
7	Прикладные программы для обработки эмпирической социологической информации	ПК-1 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики, определять качество проводимых исследований, составлять отчеты о проделанной работе, обзоры, готовить публикации ПК-4 Способен составлять структуру программного средства, определять необходимые информационные потоки и исследовать варианты структур

Перечень примерных тем электронной презентации

Тема презентации выбирается студентами самостоятельно по согласованию с преподавателем. Материал подготавливается студентами на основе индивидуальной проработки рекомендованной преподавателем и самостоятельно подобранной основной и дополнительной учебной литературы по теме, ее анализа.

Примерные темы:

- Принципы работы алгоритма Кнута, Морриса и Пратта.
- Принципы работы алгоритма Бойера и Мура.
- Принципы работы алгоритма Стемминга.
- Аналогии Стеммер Портера.
- Латентно-семантический анализ
- Алгоритмы кластеризации
- Факторное пространство
- Принципы работы алгоритма инвертированного индекса.
- Особенности булевого и вероятностного поиска.
- Хеширование
- Принципы работы алгоритма шинглов.
- Принципы работы алгоритма CRC.
- CRC-арифметика
- Канонизация текстов

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ

– оценка «отлично» по презентации присваивается за глубокое раскрытие темы, качественное оформление работы, содержательность презентации, за понимание студентом отражённого в презентации материала;

– оценка «хорошо» по презентации присваивается при соответствии выше перечисленным критериям, но при наличии в содержании работы и ее оформлении небольших недочетов или недостатков;

– оценка «удовлетворительно» по презентации присваивается за неполное раскрытие темы, выводов и предложений, носящих общий характер, отсутствие наглядного представления работы;

– оценка «неудовлетворительно» по презентации присваивается за слабое и неполное раскрытие темы, несамостоятельность изложения материала, выводы и предложения, носящие общий характер, отсутствие наглядного представления работы.

3.1.2. ВОПРОСЫ для проведения входного контроля

Тест для входного контроля

1. Как исполнитель обработки информации, человек действует:

- 1) всегда формально и однозначно
- 2) не всегда формально и однозначно+
- 3) всегда творчески
- 4) формально и творчески

2. Что представляет собой перевод текста с немецкого языка на русский язык?

- 1) поиск информации
- 2) структурирование данных
- 3) изменение формы представления информации+
- 4) получение новых сведений

3. Что происходит по команде $n \vee m$?

- 1) запись метки в текущую пустую клетку
- 2) запись метки в произвольную клетку
- 3) запись метки в текущую пустую клетку и удаление метки из соседней
- 4) запись метки в текущую пустую клетку и переход к выполнению команды m +

4. Что такое Машина Тьюринга?

- 1) универсальное устройство, использующее языки программирования высокого уровня
- 2) универсальный исполнитель обработки любых символьных последовательностей в любом алфавите+
- 3) работает с двоичным алфавитом
- 4) является частным случаем машины Поста

5. Что представляет собой Система команд исполнителя алгоритмов (СКИ)?

- 1) совокупность некоторых команд языка исполнителя
- 2) совокупность команд, которые придумывает каждый человек, работающий с исполнителем
- 3) совокупность самых главных команд исполнителя+
- 4) совокупность всех команд языка исполнителя

6. Информация, которая представляется в виде исходных данных:

- 1) должна быть получена
- 2) сохраняется
- 3) подвергается обработке+
- 4) передаётся

7. Когда возникла теория алгоритмов?

- 1) в 20-х годах XX века
- 2) в 30-х годах XX века+
- 3) в 40-х годах XX века
- 4) в 50-х годах XX века

8. Что такое Машина Поста?

- 1) универсальное устройство, использующее языки программирования высокого уровня
- 2) универсальный исполнитель обработки любых символьных последовательностей в любом алфавите
- 3) работает с двоичным алфавитом+
- 4) машина Тьюринга является частным случаем машины Поста

9. Свойство присущее процессу выполнения каждого шага алгоритма отдельно от других называется

- 1) дискретность+
- 2) понятность
- 3) точность
- 4) конечность

10. Определение понятия «шаг алгоритма»:

- 1) перемещение исполнителя на одну позицию вправо или влево
- 2) отдельная инструкция в описании алгоритма

- 3) отдельное действие, которое исполнитель выполняет по команде+
- 4) одна математическая операция

11. Что включает в себя схема обработки информации?

- 1) исходные данные, правила обработки, исполнитель, результаты+
- 2) исходные данные и правила их обработки
- 3) исходные данные и результаты
- 4) исходные данные, исполнитель, правила обработки

12. Что ассоциируется с понятием алгоритма в математике?

- 1) способ вычисления корней квадратного уравнения
- 2) способ вычисления НОД+
- 3) способ деления дробей
- 4) способ умножения дробей

13. Что представляет собой составление картотеки учебников для 10 класса?

- 1) поиск информации
- 2) получение новой информации
- 3) изменение формы представления информации
- 4) систематизация данных+

14. Что означает действие $2 \rightarrow 3$?

- 1) сдвиг вправо на один шаг+
- 2) сдвиг вниз на один шаг
- 3) сдвиг влево на один шаг
- 4) запись метки в клетку №3

15. Исходные данные – это

- 1) результат работы алгоритма
- 2) информация, которая подвергается обработке+
- 3) информация, которая получается после обработки
- 4) информация, которая хранится на внешнем носителе

16. Для чего предназначена машина Поста?

- 1) производить прием информации
- 2) производить хранение информации
- 3) производить преобразование информации на внешнем носителе
- 4) производить преобразования на информационной ленте+

17. На какие числа распространяются правила выполнения вычислений, описанные Мухаммедом аль-Хорезми?

- 1) многозначные десятичные числа+
- 2) интегралы
- 3) производные
- 4) только натуральные числа

18. Каретка – это:

- 1) оперативное запоминающее устройство машины Поста
- 2) процессор и считывающее устройство машины Поста+
- 3) процессор машины Поста
- 4) считывающее устройство машины Поста

19. Что совершает исполнитель?

- 1) создает информацию
- 2) хранит информацию
- 3) обрабатывает информацию+
- 4) изобретает информацию

20. Что такое Алгоритм Евклида?

- 1) способ вычисления наименьшего общего кратного (НОК) двух натуральных чисел
- 2) способ вычисления наибольшего общего делителя (НОД) двух натуральных чисел+
- 3) способ нахождения общего знаменателя двух обыкновенных дробей
- 4) способ нахождения частного от деления двух чисел

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ ответов на тестовые вопросы входного контроля

- оценка «зачтено» выставляется обучающемуся, если получено более 61% правильных ответов.
- оценка «не зачтено» выставляется обучающемуся, если количество правильных ответов на вопросы 60% и менее.

3.1.3 Средства для текущего контроля

ВОПРОСЫ для самостоятельного изучения темы «Содержание информационной и аналитической работы»

1. Понятие «информация»?

2. Дайте определение понятию «информационная работа»?
3. Дайте определение понятию «аналитическая работа»?
4. Что означает «качество информации»?
5. Что понимается «под обработкой информации»?

ВОПРОСЫ
для самостоятельного изучения темы
«Методика информационно-аналитической работы»

1. Назовите два характерных уровня научного исследования?
2. С чем связан эмпирический уровень исследования?
3. С чем связан теоретический уровень исследования?
4. Понятие «факты действительности»?
5. Понятие «научные факты»?
6. Какова главная задача исследователя?
7. Что понимается под исследовательским процессом?
8. Понятие «структура»?
9. Понятие «аргументирование»?
10. Дайте определение термину «тавтология»?

Общий алгоритм самостоятельного изучения темы

1) Ознакомиться с рекомендованной учебной литературой и электронными ресурсами по теме (ориентируясь на вопросы для самоконтроля).
2) На этой основе составить развёрнутый план изложения темы
3) Провести самоконтроль освоения темы по вопросам, выданным преподавателем
4) Подготовиться к предусмотренному контрольно-оценочному мероприятию по результатам самостоятельного изучения темы (опрос)
5) Принять участие в указанном мероприятии в установленное для внеаудиторной работы время

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ

Оценку «отлично» выставляют студенту, глубоко и прочно освоившему теоретический материал дисциплины. Ответ должен быть логичным, грамотным. Студенту необходимо показать знание не только основного, но и дополнительного материала.

Оценку «хорошо» заслуживает студент, твердо знающий программный материал дисциплины, грамотно и по существу излагающий его.

Оценку «удовлетворительно» получает студент, который имеет знания только основного материала, но не усвоил его детали. В ответах на поставленные вопросы студентом допущены неточности, даны недостаточно правильные формулировки, нарушена последовательность в изложении программного материала.

Оценка «неудовлетворительно» говорит о том, что студент не знает значительной части материала по дисциплине, допускает существенные ошибки в ответах.

ВОПРОСЫ и ЗАДАНИЯ
для самоподготовки к лабораторным занятиям

В процессе подготовки к лабораторному занятию обучающийся изучает представленные ниже вопросы по темам. На занятии обучающийся демонстрирует свои знания по изученным вопросам, выполняя практическое задание. Представляет отчет. Для усвоения материала по теме занятия обучающийся решает задачи.

Вопросы

1. Содержание информационной и аналитической работы
2. Методика информационно-аналитической работы
3. Основные этапы и методы информационно-аналитической работы
4. Работа с источниками информации
5. Способы оценки информации
6. Обеспечение безопасности и защиты информации
7. Прикладные программы для обработки информации

Задания

Лабораторная работа 1

Краткие теоретические сведения

Прямой поиск

Основная идея алгоритма прямым поиском заключается в посимвольном сравнении строки с подстрокой. В начальный момент происходит сравнение первого символа строки с первым символом подстроки, второго символа строки со вторым символом подстроки и т. д. Если произошло совпадение всех символов, то фиксируется факт нахождения подстроки. В противном случае производится сдвиг подстроки на одну позицию вправо и повторяется посимвольное сравнение (табл. 1). Символы, которые сравниваются, на рисунке выделены жирным. Рассматриваемые сдвиги подстроки повторяются до тех пор, пока конец подстроки не достиг конца строки или не произошло полное совпадение символов подстроки со строкой, то есть найдется подстрока.

Таблица.1 Демонстрация алгоритма прямого поиска

Строка	A	B	C	A	B	C	A	A	B	C	A	B	D
Подстрока	A	B A	C B A	A C B A	B A C B A	D B A C B A	D B A C B A	D B A C B A	D B A C B	D B A C	D B A	D B	D

Данный алгоритм является малозатратным и не нуждается в предварительной обработке и в дополнительном пространстве. Большинство сравнений алгоритма прямого поиска являются лишними.

Поэтому в худшем случае алгоритм будет малоэффективен, так как его сложность будет пропорциональна $O((n-m+1)*m)$, где n и m – длины строки и подстроки соответственно.

Алгоритм Кнута, Морриса и Пратта

Основным отличием алгоритма Кнута, Морриса и Пратта от алгоритма прямого поиска заключается в том, что сдвиг подстроки выполняется не на один символ на каждом шаге алгоритма, а на некоторое переменное количество символов. Следовательно, перед тем как осуществлять очередной сдвиг, необходимо определить величину сдвига. Для повышения эффективности алгоритма необходимо, чтобы сдвиг на каждом шаге был бы как можно большим (табл. 2). На рисунке символы, подвергшиеся сравнению, выделены жирным шрифтом.

Если для произвольной подстроки определить все ее начала, одновременно являющиеся ее концами, и выбрать из них самую длинную (не считая, конечно, саму строку), то такую процедуру принято называть префикс-функцией. В реализации алгоритма Кнута, Морриса и Пратта используется предобработка искомой подстроки, которая заключается в создании префикс-функции на ее основе. При этом используется следующая идея: если префикс (он же суффикс) строки длиной i длиннее одного символа, то он одновременно и префикс подстроки длиной $i-1$. Таким образом, проверяем префикс предыдущей подстроки, если же тот не подходит, то префикс ее префикса, и т.д. Действуя так, находим наибольший искомый префикс.

Таблица 2 Демонстрация алгоритма Кнута, Морриса и Пратта

Строка	A	B	C	A	B	C	A	A	B	C	A	B	D
Подстрока	A	B	C	A A	B B	D C	A A	B B A	D C B	A C	B A	D B	D

Точный анализ рассматриваемого алгоритма весьма сложен. Д. Кнут, Д. Моррис и В. Пратт доказывают, что для данного алгоритма требуется порядка $O(m+n)$ сравнений символов (где n и m – длины строки и подстроки соответственно), что значительно лучше, чем при прямом поиске.

Алгоритм Бойера и Мура

Существует множество вариаций алгоритма Бойера и Мура, рассмотрим простейший из них, который состоит из следующих шагов.

Первоначально строится таблица смещений для искомой подстроки. Далее идет совмещение начала строки и подстроки и начинается проверка с последнего символа подстроки. Если последний символ подстроки и соответствующий ему при наложении символ строки не совпадают, подстрока сдвигается относительно строки на величину, полученную из таблицы смещений, и снова проводится сравнение, начиная с последнего символа подстроки. Если же символы совпадают, производится сравнение предпоследнего символа подстроки и т.д. Если все символы подстроки совпали с наложенными символами строки, значит, найдена подстрока и поиск окончен. Если же какой-то (не последний) символ подстроки не совпадает с соответствующим символом строки, далее производим сдвиг подстроки на один символ вправо и снова начинаем проверку с последнего символа. Весь алгоритм выполняется до тех пор, пока либо не будет найдено вхождение искомой подстроки, либо не будет достигнут конец строки (табл. 3). На рисунке символы, подвергшиеся сравнению, выделены жирным шрифтом.

Величина сдвига в случае несовпадения последнего символа вычисляется, исходя из следующего: сдвиг подстроки должен быть минимальным, таким, чтобы не пропустить вхождение подстроки в строке. Если данный символ строки встречается в подстроке, то смещаем подстроку таким образом, чтобы символ строки совпал с самым правым вхождением этого символа в подстроку. Если же подстрока вообще не содержит этого символа, то сдвигаем подстроку на величину, равную ее длине, так что первый символ подстроки накладывается на следующий за проверявшимся символом строки.

Величина смещения для каждого символа подстроки зависит только от порядка символов в подстроке, поэтому смещения удобно вычислить заранее и хранить в виде одномерного массива, где каждому символу алфавита соответствует смещение относительно последнего символа подстроки.

Таблица 3 Демонстрация алгоритма Бойера и Мура

Строка	A	B	C	A	B	C	A	A	B	C	A	B	D
Подстрока	A	B	C	A	B	D	B	C	A	B	D	B	D

Таким образом, данный алгоритм является наиболее эффективным в обычных ситуациях, а его быстродействие повышается при увеличении подстроки или алфавита. В наихудшем случае трудоемкость рассматриваемого алгоритма $O(m+n)$.

Задание к лабораторной работе

Написать программу на выбранном языке программирования, реализующую описанные выше алгоритмы для поиска подстроки в строке. Программа должна запрашивать имя входного файла. Оценить трудоемкость рассматриваемых алгоритмов.

Лабораторная работа 2

Краткие теоретические сведения

Стеммер Портера – алгоритм стемминга (нахождения основы слова для заданного исходного слова), опубликованный Мартином Портером. Алгоритм не использует баз основ слов, а работает, последовательно применяя ряд правил отсечения окончаний и суффиксов.

Вводим ряд определений:

Гласные буквы – а, е, и, о, у, ы, э, ю, я (буква ё считается равнозначной букве е).

RV – область слова после первой гласной. Она может быть пустой, если гласные в слове отсутствуют.

R1 – область слова после первого сочетания "гласная- согласная".

R2 – область R1 после первого сочетания "гласная- согласная".

Пример В слове

противоестественном: RV = тивоеестественном.

R1 =

ивоеестественном. R2 = остественном.

Теперь определим несколько классов окончаний слова (табл. 4).

Таблица 4 Классы окончаний слова

Название класса	Окончания слов
PERFECTIVE	Группа 1*: в, вши, вшись.
GERUND	Группа 2: ив, ивши, ившись, ыв, ывши, ывшись.
ADJECTIVE	ее, ие, ые, ое, ими, ыми, ей, ий, ый, ой, ем, им, ым, ом, его, ого, ему, ому, их, ых, ую, юю, ая, яя, ою, ею.
PARTICIPLE	Группа 1*: ем, нн, вш, ющ, щ. Группа 2: ивш, ывш, ующ.
REFLEXIVE	ся, сь.
VERB	Группа 1*: ла, на, ете, йте, ли, й, л, ем, н, ло, но, ет, ют, ны, ть, ешь, нно. Группа 2: ила, ыла, ена, ейте, уйте, ите, или, ыли, ей, уй, ил, ыл, им, ым, ен, ило, ыло, ено, ят, ует, уют, ит, ыт, ены, ить, ыть, ишь, ую, ю.
NOUN	а, ев, ов, ие, ье, е, иями, ями, ами, еи, ии, и, ией, ей, ой, ий, й, иям, ям, ием, ем, ам, ом, о, у, ах, иях, ях, ы, ь, ию, ью, ю, ия, ья, я.
SUPERLATIVE	ейш, ейше.
DERIVATIONAL	ост, ость.
ADJECTIVAL	ADJECTIVAL определяется как ADJECTIVE или PARTICIPLE + ADJECTIVE. Например: бегавшая = бега + вш + ая.

* Окончаниям из группы 1 должна предшествовать буква а или я.

Правила:

1. При поиске окончания из всех возможных выбирается наиболее длинное. Например, в слове величие выбираем окончание ие, а не е.
2. Все проверки производятся над областью RV. Так, при проверке на PERFECTIVE GERUND предшествующие буквы а и я также должны быть внутри RV. Буквы перед RV не участвуют в проверках вообще.

Шаг 1

Найти окончание PERFECTIVE GERUND. Если оно существует – удалить его и завершить этот шаг.

Иначе, удаляем окончание REFLEXIVE (если оно существует). Затем в следующем порядке пробуем удалить окончания: ADJECTIVAL, VERB, NOUN. Как только одно из них найдено – шаг завершается.

Шаг 2

Если слово оканчивается на и – удаляем и.

Шаг 3

Если в R2 найдется окончание DERIVATIONAL – удаляем его.

Шаг 4

Возможен один из трех вариантов:

1. Если слово оканчивается на нн – удаляем последнюю букву.
2. Если слово оканчивается на SUPERLATIVE – удаляем его и снова удаляем последнюю букву, если слово оканчивается на нн.
3. Если слово оканчивается на ь – удаляем его.

Задание к лабораторной работе

Написать программу на выбранном языке программирования, реализующую описанный выше алгоритм для поиска по тексту однокоренных слов. Программа должна запрашивать имя входного файла и слово для поиска. Результатом работы программы должен быть файл, содержащий список однокоренных слов.

Лабораторная работа 3

Краткие теоретические сведения

Латентно-семантический анализ

Латентно-семантический анализ отображает документы и отдельные слова в так называемое «семантическое пространство», в котором и производятся все дальнейшие сравнения. При этом

делаются следующие предположения:

- 1) Документы это просто набор слов. Порядок слов в документах игнорируется. Важно только то, сколько раз то или иное слово встречается в документе.
- 2) Семантическое значение документа определяется набором слов, которые как правило идут вместе. Например, в биржевых сводках, часто встречаются слова: «фонд», «акция», «доллар»
- 3) Каждое слово имеет единственное значение. Это, безусловно, сильное упрощение, но именно оно делает проблему разрешимой.

Пример

Для примера возьмем несколько заголовков из различных новостей. На первом шаге из этих заголовков исключаем, так называемые, стоп-символы. Это слова которые встречаются в каждом тексте и не несут в себе смысловой нагрузки, это, прежде всего, все союзы, частицы, предлоги и множество других слов. Полный список использованных стоп-символов можно посмотреть в Приложении 1.

Далее необходимо выполнить операцию стемминга. Она не является обязательной, если набор текстов достаточно большой, или если тексты на английском языке, в силу того, что количество вариаций той или иной словоформы в английском языке существенно меньше чем в русском. В нашем же случае, пропускать этот шаг не стоит т.к. это приведет к существенной деградации результатов.

Дальше исключаем слова встречающиеся в единственном экземпляре. Это тоже необязательный шаг, он не влияет на конечный результат, но сильно упрощает математические вычисления. В итоге у нас остались, так называемые, индексируемые слова, они выделены жирным шрифтом:

1. Британская **полиция** знает о местонахождении **основателя WikiLeaks**
2. В **суде США** начинается процесс **против** россиянина, рассылавшего спам
3. **Церемонию вручения Нобелевской премии** мира бойкотируют 19 **стран**
4. В **Великобритании арестован основатель** сайта **Wikileaks** Джулиан Ассандж
5. Украина игнорирует **церемонию вручения Нобелевской премии**
6. Шведский **суд** отказался рассматривать апелляцию **основателя Wikileaks**
7. НАТО и **США** разработали планы обороны **стран** Балтии **против** России
8. **Полиция Великобритании** нашла **основателя WikiLeaks**, но, не **арестовала**
9. В Стокгольме и Осло сегодня состоится **вручение Нобелевских премий Латентно**

семантический анализ

На первом шаге требуется составить частотную матрицу индексируемых слов. В этой матрице строки соответствуют индексируемым словам, а столбцы — документам. В каждой ячейке матрицы указано какое количество раз слово встречается в соответствующем документе.

Таблица 5 Частотная матрица индексируемых слов

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Wikileaks	1	0	0	1	0	1	0	1	0
Арестова	0	0	0	1	0	0	0	1	0
Великобритан	0	0	0	1	0	0	0	1	0
Вручен	0	0	1	0	1	0	1	0	0
Нобелевск	0	0	1	0	1	0	0	0	1
Основател	1	0	0	1	0	1	0	1	0
Полиц	1	0	0	0	0	0	0	1	0
Прем	0	0	1	0	1	0	0	0	1
Прот	0	1	0	0	0	0	1	0	0
Стран	0	0	1	0	0	0	1	0	0
суд	0	1	0	0	0	1	0	0	0
Сша	0	1	0	0	0	0	1	0	0
Церемон	0	0	1	0	1	0	0	0	0

Следующим шагом необходимо провести сингулярное разложение полученной матрицы. Т.е. исходную матрицу M мы представляем в виде:

$$M = U * W * V^t,$$

где U и V^t — ортогональные матрицы, а W — диагональная матрица. Причем диагональные элементы матрицы W упорядочены в порядке убывания. Диагональные элементы матрицы W называются сингулярными числами.

Таблица 6 Сингулярное разложение частотной матрицы индексируемых слов (матрица

U)

Wikileaks	0.57	-0.01	0.01	-0.2	0.13	0.16	-0.16	-0.25	-0.64
Арестова	0.34	0	0.07	0.41	-0.42	-0.02	0.1	0.17	0.01
Великобритан	0.34	0	0.07	0.41	-0.42	-0.02	0.1	0.17	0.01
Вручен	0	0.52	0.07	-0.06	-0.08	-0.15	-0.17	0.2	-0.07
Нобелевск	0	0.52	0.07	-0.06	-0.08	-0.15	-0.17	0.2	0.32
Основател	0.57	-0.01	0.01	-0.2	0.13	0.16	-0.16	-0.25	0.64
Полиц	0.31	0	0.05	0.07	0.57	-0.6	0.29	0.37	0
Прем	0	0.52	0.07	-0.06	-0.08	-0.15	-0.17	0.02	-0.25
Прот	0.02	0.03	-0.06	0.13	-0.05	-0.22	0	-0.25	0
Стран	0.01	0.22	-0.31	0.39	0.41	0.56	-0.22	0.4	0
суд	0.12	0.01	-0.38	-0.62	-0.3	0.12	0.21	0.55	0
Сша	0.02	0.03	-0.61	0.13	-0.05	-0.22	0	-0.25	0
Церемон	0	0.38	0.03	0.02	0.08	0.31	0.82	-0.29	0

Таблица 7 Сингулярное разложение частотной матрицы индексируемых слов (матрица

W)

3.41	0	0	0	0	0	0	0	0
0	3.3	0	0	0	0	0	0	0
0	0	2.27	0	0	0	0	0	0
0	0	0	1.49	0	0	0	0	0
0	0	0	0	1.19	0	0	0	0
0	0	0	0	0	0.98	0	0	0
0	0	0	0	0	0	0.71	0	0
0	0	0	0	0	0	0	0.43	0
0	0	0	0	0	0	0	0	0

Таблица 8 Сингулярное разложение частотной матрицы индексируемых слов (матрица

V^t)

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.43	0.05	0.01	0.54	0	0.37	0.01	0.63	0
0	0.02	0.65	-0.01	0.59	0	0.09	-0.01	0.47
0.03	-0.7	-0.04	0.06	0.1	-0.16	-0.67	0.09	0.09
-0.22	-0.24	0.15	0.28	-0.11	-0.68	0.44	0.33	-0.13
0.69	-0.32	0.22	-0.49	-0.12	-0.03	0.27	-0.02	-0.19
-0.27	-0.34	0.44	0.29	-0.13	0.45	0.12	-0.31	-0.45
-0.03	0.3	0.14	-0.17	0.44	-0.15	-0.3	0.24	-0.71
-0.3	0.12	0.4	-0.39	-0.53	0.12	-0.23	0.46	0.13
0.35	0.35	0.35	0.35	-0.35	-0.35	-0.35	-0.35	0

Согласно простым правилам произведения матриц, видно, что столбцы и строки, соответствующие меньшим сингулярным значениям, дают наименьший вклад в итоговое произведение. Например, мы можем отбросить последние столбцы матрицы U и последние строки матрицы V^t, оставив только первые 2. Важно, что при этом гарантируется, оптимальность полученного произведения. Разложение такого вида называют двумерным сингулярным разложением:

Таблица 9 Двумерное сингулярное разложение (матрица U)

Wikileaks	0.57	-0.01
Арестова	0.34	0
Великобритан	0.34	0
Вручен	0	0.52
Нобелевск	0	0.52
Основател	0.57	-0.01
Полиц	0.31	0
Прем	0	0.52
Прот	0.02	0.03
Стран	0.01	0.22

суд	0.12	0.01
Сша	0.02	0.03
Церемон	0	0.38

Таблица 10 Двумерное сингулярное разложение (матрица W)

3.41	0
0	3.3

Таблица 11 Двумерное сингулярное разложение (матрица Vt)

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.43	0.05	0.01	0.54	0	0.37	0.01	0.63	0
0	0.02	0.65	-0.01	0.59	0	0.09	-0.01	0.47

На практике, конечно, количество групп будет намного больше, пространство будет не двумерным, а многомерным, но сама идея остается той же. Можно определять местоположения слов и статей в нашем пространстве и использовать эту информацию для, например, определения тематики статьи.

В полученном факторном пространстве документы и термины концентрируются областями, имеющими общий семантический и латентный смысл.

Применительно к кластеризации, получаемые области и есть кластеры. С помощью математических преобразований можно определить центры кластеров.

Задание к лабораторной работе

Написать программу на выбранном языке программирования, реализующую описанный выше алгоритм для кластеризации заголовков по темам. Количество заголовков задать не более 10 (оптимально от 7 до 10), количество кластеров не более 3. Программа должна запрашивать заголовки (они могут храниться в файлах). После получения двумерного сингулярного разложения осуществить прямую кластеризацию самостоятельно выбранным и изученным методом (иерархические алгоритмы, поиск k-средних и т.д.). Результатом работы программы должен быть файл, содержащий заголовки, разбитые по кластерам.

Лабораторная работа 4

Краткие теоретические сведения

Инвертированный индекс

Инвертированный индекс — структура данных, в которой для каждого слова коллекции документов в соответствующем списке перечислены все места в коллекции, в которых оно встретилось. Инвертированный индекс используется для поиска по текстам.

Булев поиск

Булев поиск опирается на использование инвертированного индекса ключевых слов, т. е. таблицы, в которой для каждого ключевого слова перечисляются все документы, где оно встречается. Главным достоинством этого алгоритма является связывание слов запроса логическими операциями. К недостаткам этого алгоритма следует отнести невозможность определения релевантности запросу полученной выборки документов и, как следствие, невозможность ее сортировки.

Опишем как решается задача нахождения документов в которых встречаются все слова из поискового запроса. При обработке однословного поискового запроса, ответ уже есть в инвертированном индексе — достаточно взять список соответствующий слову из запроса. При обработке многословного запроса берутся списки, соответствующие каждому из слов запроса и пересекаются.

Пример.

Пусть у нас есть корпус из трех текстов $T_0 = \text{"it is what it is"}$, $T_1 = \text{"what is it"}$ и $T_2 = \text{"it is a banana"}$, тогда инвертированный индекс будет выглядеть следующим образом:

"a": {2}
 "banana": {2}
 "is": {0, 1, 2}
 "it": {0, 1, 2}
 "what": {0, 1}

Здесь цифры обозначают номера текстов, в которых встретилось соответствующее слово. Тогда обработка поискового "what is it" запроса даст следующий результат * + * + * + * +.

Таблицы индекса

Для эффективной организации поиска документов необходимо задействовать несколько таблиц в базе данных. В самом простом случае используются следующие три.

Таблица документов Documents. В ней хранится информация обо всех документах, проиндексированных системой, а именно название документа, его авторы, тип файла, путь к файлу/URL и т. д. При этом каждому документу необходимо присвоить уникальный идентификатор Doc_id.

Таблица ключевых слов/словарь Words. Здесь хранятся все ключевые слова системы и соответствующие им номера Word_id.

Инвертированный индекс Inverse, используемый для поиска. В этой таблице хранится идентификатор слова Word_id и соответствующий ему список документов, содержащих это слово.

Эффективная организация словаря

Одна из самых важных и трудных проблем индексации текстов связана с созданием и пополнением словаря ключевых слов. Главная сложность ее заключается в том, что для эффективной работы системы необходимо рассматривать только базовые словоформы ключевых слов..

Еще одна проблема индексирования связана с выявлением и удалением из текста так называемых стоп-слов. Они не несут смысловой нагрузки в текущей предметной области, и для эффективной работы системы их следует удалять при индексировании. Как правило, стоп-словами являются предлоги, союзы, артикли, вводные слова и т. п. Они очень часто встречаются в документах, но малоинформативны. Для их удаления можно либо использовать отдельный словарь стоп-слов, либо считать все слова с высокими частотами встречаемости в базе данных текстов стоп-словами и удалять их при индексировании.

Ранжирование

Обычно в поисковых системах после построения с помощью инвертированного индекса списка документов, содержащих слова из запроса, идет ранжирование документов из списка.

Для каждого запроса необходимо вычислить значение Score документа – показатель релевантности документа запросу, на основании которого и производится ранжирование.

Для расчета Score предлагается использовать аддитивную модель. В качестве слагаемых в данной модели предлагаются следующие:

встречаемость слов из запроса в документе (W_{single}), встречаемость пар слов из запроса в документе (W_{pair}) и встречаемость текста запроса целиком (W_{phrase}). Помимо этого есть два слагаемых, дающих преимущество за наличие всех слов запроса в документе ($W_{AllWords}$). Итоговая формула выглядит следующим образом:

$$Score = W_{single} + W_{pair} + k_1 * W_{AllWords} + k_2 * W_{phrase}$$

где $k_1 = 1$, $k_2 = 1/350$

Задание к лабораторной работе

Написать программу на выбранном языке программирования, реализующую индексацию документов (не менее 5 документов) описанным выше алгоритмом и осуществить поиск документа, удовлетворяющего заданный запрос. Осуществить ранжирование документов по их релевантности запросу. Программа должна запрашивать имена входных файлов и выводить заголовки и выдержки из найденных по запросу документов.

Лабораторная работа 5

Краткие теоретические сведения

Проблема обнаружения нечетких дубликатов является одной из наиболее важных и трудных задач анализа данных и поиска информации. Актуальность этой проблемы определяется разнообразием приложений, в которых необходимо учитывать «похожесть», например, текстовых документов — это и улучшение качества индекса и архивов поисковых систем за счет удаления избыточной информации, и объединение новостных сообщений в сюжеты на основе сходства этих сообщений по содержанию, и фильтрация спама (как почтового, так и поискового), и установление нарушений авторских прав при незаконном копировании информации (проблема плагиата или копирайта), и ряд других.

Алгоритм шинглов

Алгоритм шинглов — алгоритм, разработанный для поиска копий и дубликатов рассматриваемого текста в веб-документе, мощный инструмент, призванный бороться с проявлениями плагиата в интернете.

Реализация алгоритма подразумевает несколько этапов:

- канонизация текстов;
- разбиение текста на шинглы;
- нахождение контрольных сумм;
- поиск одинаковых подпоследовательностей.

Канонизация текста

Канонизация текста - приведение оригинального текста к единой нормальной форме через очищение его от всех вспомогательных единиц текста (предлогов, союзов, знаков препинания, тегов и прочее), которые не должны участвовать в сравнении. Часто предполагается также удаление имен прилагательных, поскольку они, как правильно, несут эмоциональную, а не смысловую нагрузку.

Канонизация текста также требует приведения имен существительных в именительный падеж, единственное число, а иногда – оставление только их корневых значений.

После проведения всех указанных операций получается «чистый» текст, пригодный для сравнения.

Контрольная сумма

В самом общем своем виде контрольная сумма представляет собой некоторое значение, построенное по определенной схеме на основе кодируемого сообщения.

Алгоритм CRC базируется на свойствах деления с остатком двоичных многочленов, то есть многочленов над конечным полем $GF(2^N)$. Значение CRC является по сути остатком от деления многочлена, соответствующего входным данным, на некий фиксированный порождающий многочлен.

Каждой конечной последовательности битов $a_0, a_1, \dots, a_{N-1}$ взаимно однозначно сопоставляется двоичный полином, последовательность коэффициентов которого представляет собой исходную последовательность. Например, последовательность битов 1011010 соответствует многочлену:

$$P(x) = 1 \cdot x^6 + 0 \cdot x^5 + 1 \cdot x^4 + 1 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x^1 = x^6 + x^4 + x^3 + x^1$$

Значение контрольной суммы в алгоритме с порождающим многочленом $G(x)$ степени N определяется как битовая последовательность длины N , представляющая многочлен $R(x)$, получившийся в остатке при делении многочлена $P(x)$, представляющего входной поток бит, на многочлен $G(x)$:

$$R(x) = P(x) \cdot x^N \bmod G(x),$$

где $R(x)$ — многочлен, представляющий значение CRC, $P(x)$ — многочлен, коэффициенты которого представляют входные данные, $G(x)$ — порождающий многочлен, N — степень порождающего многочлена.

Рассмотрим общую схему алгоритма расчета CRC:

1. Выбрать полином P , в результате автоматически становится известна его степень N .
2. Добавить к исходной двоичной последовательности сообщения N нулевых битов. Это добавление делается для гарантированной обработки всех битов исходной последовательности.
3. Выполнив деление дополненной N нулями исходной строки S на полином P по правилам CRC арифметики. Запомнить остаток от деления, который и будет являться CRC.

Задание к лабораторной работе

Написать программу на выбранном языке программирования, реализующую поиск нечетких дубликатов заданных текстов описанным выше алгоритмом. Программа должна запрашивать имена входных файлов и выводить схожие документы и степень их схожести (в процентах).

Лабораторная работа 6

Краткие теоретические сведения

Цифровой сигнал — сигнал данных, у которого каждый из представляющих параметров описывается функцией дискретного времени и конечным множеством возможных значений.

Сигналы представляют собой дискретные электрические или световые импульсы. При таком способе вся емкость коммуникационного канала используется для передачи одного сигнала. Цифровой сигнал использует всю полосу пропускания кабеля. Полоса пропускания — это разница между максимальной и минимальной частотой, которая может быть передана по кабелю. Каждое устройство в таких сетях посылает данные в обоих направлениях, а некоторые могут одновременно принимать и передавать. Узкополосные системы (baseband) передают данные в виде цифрового сигнала одной частоты.

Дискретный цифровой сигнал сложнее передавать на большие расстояния, чем аналоговый сигнал, поэтому его предварительно модулируют на стороне передатчика, и демодулируют на стороне приёмника информации. Использование в цифровых системах алгоритмов проверки и

восстановления цифровой информации позволяет существенно увеличить надёжность передачи информации.

Замечание. Следует иметь в виду, что реальный цифровой сигнал по своей физической природе является аналоговым. Из-за шумов и изменения параметров линий передачи он имеет флуктуации по амплитуде, фазе/частоте (джиттер), поляризации. Но этот аналоговый сигнал (импульсный и дискретный) наделяется свойствами числа. В результате для его обработки становится возможным использование численных методов (компьютерная обработка).

Для того, чтобы представить аналоговый сигнал последовательностью чисел конечной разрядности, его следует сначала превратить в дискретный сигнал, а затем подвергнуть квантованию. Квантование является частным случаем дискретизации, когда дискретизация происходит по одинаковой величине называемой квантом. В результате сигнал будет представлен таким образом, что на каждом заданном промежутке времени известно приближённое (квантованное) значение сигнала, которое можно записать целым числом.

Если записать эти целые числа в двоичной системе, получится последовательность нулей и единиц, которая и будет являться цифровым сигналом (рисунок).

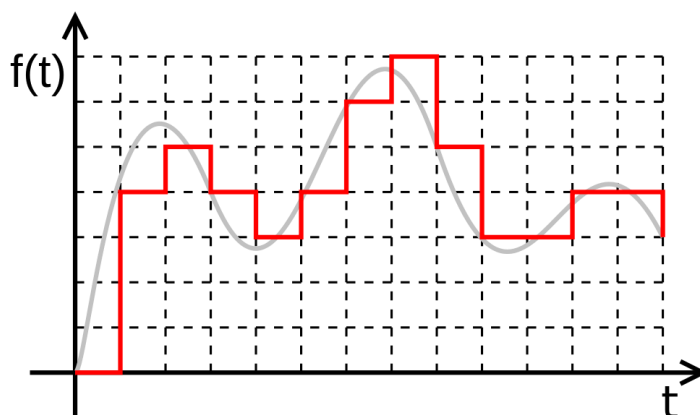


Рис. Преобразование аналогового сигнала в цифровой сигнал

Задание к лабораторной работе

На выбранном языке программирования реализовать алгоритм преобразования аналогового сигнала $F(x)$ в цифровой сигнал, а также отобразить аналоговый и цифровой сигнал графически на промежутке от 0 до 10 сек. Параметры задания аналогового сигнала и частоту дискретизации взять из табл.

Таблица 12 Параметры задания аналогового сигнала и частоту дискретизации

Последняя цифра номера студенческого билета	0	1	2	3	4	5	6	7	8	9
$F(x)$	sin	cos	sin	cos	sin	cos	sin	cos	sin	cos
Период $F(x)$, сек	3	4	5	3	4	5	3	4	5	3
Частота дискретизации, сек	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25	0,25

Лабораторная работа 7

Краткие теоретические сведения

Внедрение компьютеров буквально во все сферы человеческой деятельности является на сегодняшний день, наверное, самым наглядным результатом научного прогресса. Компьютер сейчас используется для выполнения такой работы, которая раньше считалась не только самой скучной и утомительной, но и рутинной – самой объемной во всех статистических исследованиях: учет и организация исходных данных, вычисление различных показателей и многие другие статистически-математические операции, сложные для понимания и вычисления. С появлением многочисленных

статистических и вычислительных программ исследователи получили долгожданную возможность быстро и легко справляться с различной

вычислительной работой и больше времени уделять анализу и интерпретации полученных данных.

Для статистического анализа необходима обработка очень большого объема данных, и часто сводные таблицы данных могут достигать практически необозримых размеров. Многофункциональные пакеты для обработки и анализа данных, такие как Excel, Access, STADIA, STATA, JMR (StatView), SAS (компания SAS Institute), STATGRAPICS, STATISTICA, SPSS и бесчисленное множество других нашли широкое распространение в самых разных областях исследований.

Statistika- это современный пакет статистического анализа, в котором реализованы все новейшие компьютерные и математические методы анализа данных.

Система *STATISTICA* состоит из набора модулей, в каждом из которых собрана тематически связанная группа процедур. При переключении модулей можно либо оставлять открытым только одно окно приложения *STATISTICA*, либо все вызванные ранее модули, поскольку каждый из них может выполняться в отдельном окне.

Архитектура пакета *STATISTICA* и реализованный в нем аппарат многозадачности позволяет не только без каких-либо усилий сравнивать результаты анализа, полученные разными процедурами, но и одновременно выполнять их в разных окнах приложений.

Одним из самых незаменимых помощников для исследователя, работающего в области статистического анализа данных, является программа SPSS - очень мощный и широко распространенный пакет профессионального компьютерного анализа данных SPSS для Windows (аббревиатура SPSS расшифровывается как StatisticalPackageforSocialScience – Статистический пакет для социальных наук) - это модульный, полностью интегрированный, обладающий всеми необходимыми возможностями программный комплекс, охватывающий все этапы аналитического процесса: планирование, сбор данных, доступ к данным и управление данными, анализ, создание отчетов и распространение результатов. Ближайшим аналогом SPSS является программа Statistika.

Основные функциональные возможности программы по обработке статистических данных и презентации результатов: кодирование статистических данных; основные определения и понятия из области статистики: шкалы статистических измерений, выборки и распределения; первичная обработка и подготовка данных для применения основных статистических методов, вычислительные возможности SPSS; построение таблиц сопряженности и расчет статистических характеристик для них; интерпретация таблиц сопряженности и ее практическое значение; методы анализа множественных вариантов ответов и условия их применения; множественные варианты ответов и частотный анализ; множественные варианты ответов и таблицы сопряженности; и многое другое.

RapidMiner (ранее известный как YALE) - наверное мощнейшая на сегодняшний день комплексная система для DataMining (Интеллектуальный Анализ Данных, ИАД) и статистического анализа. Обладает очень серьезным набором алгоритмов для обработки и анализа, включая обработку больших массивов данных. Пакет обладает достаточно оригинальной концепцией. Работа с любым набором данных представляет собой процесс древовидного типа, в который можно, как в конструкторе, добавлять различные операторы ввода/вывода, обработки, визуализации, анализа и т.п.

Weka – это Java-приложение, предназначенное для интеллектуального анализа данных.

Алгоритмы могут быть применены непосредственно к набору данных или вызваны из вашего собственного кода Java. Особенности программы «Weka»:

1. Программа написана на Java
2. Инструменты для предварительной обработки данных
3. Классификация данных
4. Кластеризация данных
5. Задание ассоциативных правил
6. Weka хорошо подходит для разработки новых схем машинного обучения.

Задание к лабораторной работе

1. Познакомьтесь на сайте <http://www.spss.ru> с деятельностью российского офиса SPSS - официального представителя SPSS в России. Более подробно изучите направления использования SPSS на практике.

2. Соберите сведения о предприятиях, расположенных на территории вашего города (или сведения о членах вашей студенческой группы и т. д.). Используйте для их характеристики признаки, измеренные по разным шкалам. Внесите исходные данные в программу IBM SPSS STATISTICS 22.0. Созданный файл данных используйте для изучения различных возможностей SPSS.

3. Представьте количественный признак в вашем учебном файле данных в виде пяти равных интервалов, новому признаку присвойте метку и создайте метки для всех его

значений. Используйте для этой цели программу SPSS. Все Ваши действия с помощью меню SPSS фиксируйте в редакторе Синтаксиса с помощью кнопки Вставка.

4. Ниже приведены значения средней начисленной заработной платы за несколько периодов по субъектам РФ. Дополните имеющиеся сведения актуальными данными, воспользовавшись сайтом Госкомстата. Рассчитайте с помощью SPSS разницу в уровне заработной платы по субъектам РФ и проведите сравнительный анализ. Транспонируйте с помощью SPSS матрицу данных и рассчитайте с помощью SPSS темпы прироста заработной платы по годам. Сделайте содержательные выводы. Какими командами SPSS Вам пришлось воспользоваться для выполнения данного задания?

Таблица 13 Значения средней начисленной заработной платы за несколько периодов по субъектам РФ

	Кемеровская область	Томская область	Новосибирская область	Омская область	Алтайский край	СФО
1990	0,351	0,358	0,316	0,290	0,299	0,327
1995	671,3	546,0	391,4	396,1	332,2	553,3
2000	2425,4	2544,9	1819,1	1466,1	1364,8	2269,6
2001	3312,7	4055,1	2657,8	2294,8	1870,4	3190,8
2002	4248,3	5235,2	3708,8	3452,9	2567,6	4309,5
2003	5263,6	6685,4	4623,5	4426,9	3147,5	5325,3
2004	6706,6	7972,2	5833,4	5482,5	3955,3	6507,8
2005	8653,6	9609,9	7264,3	7124,3	4913,8	8109,7
2006	10407,7	11317,2	9165,6	8866,6	6147,0	9877,5
2007	12554,9	14429,0	12017,0	11003,6	7804,7	12344,8
2008	15410,0	17675,3	15713,6	13524,8	9731,5	15381,4
2009	15995,0	19340,0	16798,5	14780,5	10871,6	16606,4
2010	18027,8	21450,	18229,7	16708,2	12050,7	18657,6

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ

Оценку «отлично» выставляют студенту, глубоко и прочно освоившему теоретический и практический материал дисциплины. Ответ должен быть логичным, грамотным. Студенту необходимо показать знание не только основного, но и дополнительного материала. Студент должен свободно справляться с поставленными задачами, правильно обосновывать принятые решения.

Оценку «хорошо» заслуживает студент, твердо знающий программный материал дисциплины, грамотно и по существу излагающий его. Необходимо правильно применять теоретические положения при решении практических задач, владеть определенными навыками и приемами их выполнения.

Оценку «удовлетворительно» получает студент, который имеет знания только основного материала, но не усвоил его детали, испытывает затруднения при решении практических задач. В ответах на поставленные вопросы студентом допущены неточности, даны недостаточно правильные формулировки, нарушена последовательность в изложении программного материала.

Оценка «неудовлетворительно» говорит о том, что студент не знает значительной части материала по дисциплине, допускает существенные ошибки в ответах, не может решить практические задачи или решает их с затруднениями.

3.1.4. Средства для промежуточной аттестации по итогам изучения дисциплины

Нормативная база проведения промежуточной аттестации обучающихся по результатам изучения дисциплины:	
1) действующее «Положение о текущем контроле успеваемости, промежуточной аттестации обучающихся по программам высшего образования (бакалавриат, специалитет, магистратура) и среднего профессионального образования в ФГБОУ ВО Омский ГАУ»	
Основные характеристики промежуточной аттестации обучающихся по итогам изучения дисциплины	
Цель промежуточной аттестации -	установление уровня достижения каждым обучающимся целей и задач обучения по данной дисциплине, изложенным в п.2.2 настоящей программы
Форма промежуточной аттестации -	дифференцированный зачет
Место процедуры получения зачёта в графике учебного процесса	1) участие обучающегося в процедуре получения зачёта осуществляется за счёт учебного времени (трудоемкости), отведённого на изучение дисциплины
	2) процедура проводится в рамках ВАРО, на последней неделе семестра
Основные условия получения обучающимся зачёта:	1) обучающийся выполнил все виды учебной работы (включая самостоятельную) и отчитался об их выполнении в сроки, установленные графиком учебного процесса по дисциплине; 2) прошёл заключительное тестирование и выполнил проверочную работу

Программа для сдачи дифференцированного зачета по учебной дисциплине

Для получения допуска к зачету студент должен прослушать лекции, регулярно посещать лабораторные занятия и выполнять на них задания.

Студенты, пропустившие занятия должны отработать в порядке, согласованном с преподавателем. По всем видам текущего контроля (отчеты по лабораторным заданиям) должны быть получены положительные оценки. Кроме этого, должны быть сдано задание ВАРС (презентация).

По итогам изучения дисциплины, обучающиеся проходят заключительное тестирование и выполняет проверочную работу.

Тестирование является формой контроля, направленной на проверку владения терминологическим аппаратом, современными информационными технологиями и конкретными знаниями в области фундаментальных и прикладных дисциплин.

Тестирование осуществляется по всем темам и разделам дисциплины, включая темы, выносимые на самостоятельное изучение.

Плановая процедура проведения дифференцированного зачета

Процедура тестирования ограничена во времени и предполагает максимальное сосредоточение обучающегося на выполнении теста, содержащего несколько тестовых заданий.

Тестирование проводится в письменной форме (на бумажном носителе). Тест включает в себя 30 вопросов. Время, отводимое на выполнение теста - 30 минут. В каждый вариант теста включаются вопросы в следующем соотношении: закрытые (одиночный выбор) – 25-30%, закрытые (множественный выбор) – 25-30%, открытые – 25-30%, на упорядочение и соответствие – 5-10%

На тестирование выносятся по 10 вопросов из каждого раздела дисциплины.

Бланк теста

Образец

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Омский государственный аграрный университет имени П.А. Столыпина»

Тестирование по итогам освоения дисциплины Б1.В.02 Аналитическая обработка информации в составе ОПОП 09.04.02 Информационные системы и технологии

ФИО _____ группа _____

Дата _____

Уважаемые обучающиеся!

Прежде чем приступить к выполнению заданий внимательно ознакомьтесь с инструкцией:

1. Отвечая на вопрос с выбором правильного ответа, правильный, на ваш взгляд, ответ (ответы) обведите в кружок.

2. В заданиях открытой формы впишите ответ в пропуск.

3. В заданиях на соответствие заполните таблицу.

4. В заданиях на правильную последовательность впишите порядковый номер в квадрат.

4. Время на выполнение теста – 30 минут

5. За каждый верный ответ Вы получаете 1 балл, за неверный – 0 баллов.

Максимальное количество полученных баллов 30.

Желаем удачи!

Вопросы теста

1. Аналитик это ...

- а) специалист в области анализа и моделирование
- б) специалист в предметной области;
- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

2 Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

3 Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

4 Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

5 Задача кластеризации заключается в ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

6 Целью поиска ассоциативных правил является ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

7 До предполагаемых моделей относятся такие модели данных:

- а) модели классификации и последовательностей;
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

8 В описательных моделях относятся следующие модели данных:

- а) модели классификации и последовательностей;
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

9 Модели классификации описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

10 Модели последовательностей описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

11 Регрессивные модели описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

12. Виды лингвистической неопределенности:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами;
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);
- г) неоднозначность смысла фраз (Синтаксическая и семантическая).

13. Модели исключений описывают ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

14 Итоговые модели обнаружат ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

15 Модели ассоциации проявляют ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

16 Виды физической неопределенности данных:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью);
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);
- г) неоднозначность смысла фраз (Синтаксическая и семантическая).

17 Очистка данных — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитической задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их

информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

18 Обогащение — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

19 Консолидация — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

20 Транзакция — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

21 Метаданные — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

22 Классификация — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

23 Регрессия — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) выявление закономерностей между связанными событиями
- г) это установление зависимости дискретной выходной переменной от входных переменных

24 Кластеризация — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) выявление закономерностей между связанными событиями
- г) это установление зависимости дискретной выходной переменной от входных переменных.

25 Ассоциация — ...

- а) это установление зависимости непрерывной выходной переменной от входных переменных

- б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов
- в) выявление закономерностей между связанными событиями
- г) это установление зависимости дискретной выходной переменной от входных переменных

26 Машинное обучение — ...

- а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

27 Аналитическая платформа — ...

- а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

28 Обучающая выборка — ...

- а) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- б) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

29 Ошибка обучения — ...

- а) это ошибка, допущенная моделью на учебной множества.
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества
- в) имена, типы, метки и назначения полей исходной выборки данных
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

30 Ошибка обобщения — ...

- а) это ошибка, допущенная моделью на учебной множества.
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества
- в) имена, типы, метки и назначения полей исходной выборки данных
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

ШКАЛА И КРИТЕРИИ ОЦЕНКИ

ответов на тестовые вопросы тестирования по итогам освоения дисциплины

- оценка «отлично» выставляется обучающемуся, если получено более 81% правильных ответов.
- оценка «хорошо» - получено от 71 до 80% правильных ответов.
- оценка «удовлетворительно» - получено от 61 до 70% правильных ответов.
- оценка «неудовлетворительно» - получено менее 61% правильных ответов.

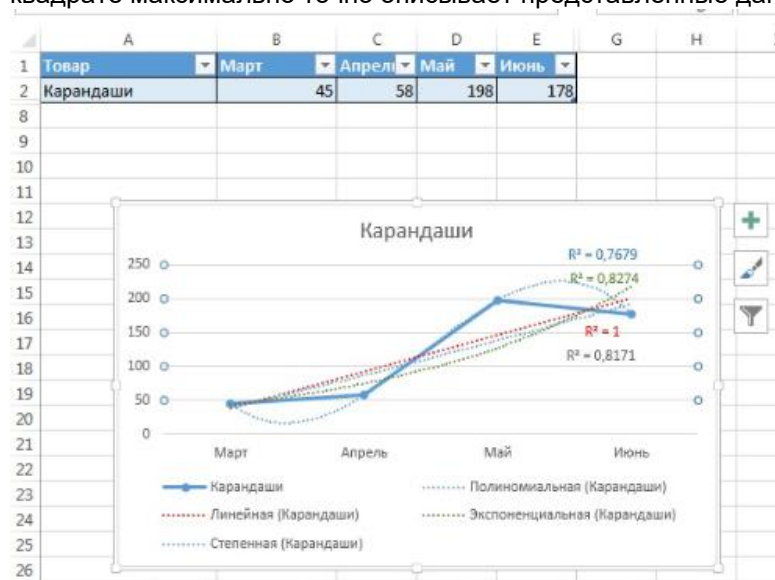
**Вопросы и практические задания
для проведения промежуточной аттестации по дисциплине «Аналитическая обработка
информации»
(Дифзачет)**

Вопросы

1. Основные задачи, решаемые в системах поддержки принятия решений при аналитической обработке профессиональной информации
2. Подсистемы, входящие в обобщенную архитектуру системы поддержки принятия решений
3. Подходы к анализу данных с помощью информационных систем
4. Базовые методы анализа
5. Задачи Data Mining
6. Способы визуализации данных
7. Назначение OLAP-систем
8. Компоненты OLAP-системы
9. Двенадцать правил Кодда, определяющих OLAP
10. Назначение программы Deductor Studio
11. Основные модули по анализу данных в Deductor Studio
12. Основные классы информационно-аналитических систем
13. Понятие и свойства хранилища данных
14. Источники данных, используемые для создания хранилища данных в Deductor Studio
15. Назначение многомерных хранилищ данных
16. Базовые понятия многомерной модели данных
17. Ошибки данных в источниках, которые могут быть обнаружены средствами аналитической платформы Deductor Studio
18. Методы восстановления пропущенных значений, которые реализованы в аналитической платформе Deductor Studio
19. Основные цели и назначение факторного и корреляционного анализа в предварительном анализе данных.
20. Целью проведения анализа «что – если»: что он позволяет исследовать?

Практические задания

1. Укажите этапы подведения промежуточных итогов профессиональной информации в MS Excel
2. Разработка стратегии действий для решения проблемной ситуации: линия тренда с каким R в квадрате максимально точно описывает представленные данные?



3. Разработка стратегии действий для решения проблемной ситуации: какое значение вернет функция СЧЁТ в процессе анализа данных, представленных на рисунке?

В1

А B C D E F G H I J K L

1	Названия месяцев	Коэффициент										
2	Январь	10										
3	Февраль	20										
4	Март	30										
5	Апрель	Отсутствует										
6	Май	50										
7	Июнь	60										
8	Июль	Отсутствует										
9	Август	80										
10	Сентябрь	90										
11	Октябрь	100										
12	Ноябрь	110										
13	Декабрь	Отсутствует										

Аргументы функции

СЧЁТ

Значение1: B1:B13 = {"Коэффициент";10;20;30;"Отсутств..."

Значение2: = число

Подсчитывает количество ячеек в диапазоне, который содержит числа.

Значение: значение1;значение2,... от 1 до 255 аргументов, которые могут содержать или ссылаться на данные различных типов, но учитываются только числовые значения.

Справка по этой функции

OK Отмена

4. При анализе данных сводной таблицы, в какую ее область добавляется вычисляемое поле?

ФАЙЛ ГЛАВНАЯ ВСТАВКА РАЗМЕТКА СТРАНИЦЫ ФОРМУЛЫ ДАННЫЕ РЕЦЕНЗИРОВАНИЕ ВИД

D14

104542

А B C D E F G H I

3	Сумма по полю Продажи	Названия столбцов						
4	Названия строк	Восточный	Западный	Южный	Общий итог			
5	январь	259416	99833	171897	531146			
6	февраль	255487	100333	135497	491317			
7	март	296958	107884	147425	552267			
8	апрель	248956	110628	131401	490985			
9	май	293192	144889	132165	570246			
10	июнь	281641	133153	122156	536950			
11	июль	263899	147484	110844	522227			
12	август	283917	176325	107935	568177			
13	сентябрь	252049	181518	101233	534800			
14	октябрь	273592	212932	104542	591066			
15	ноябрь	292585	232032	98041	622658			
16	декабрь	288378	239514	95986	623878			
17	Общий итог	3290070	1886525	1459122	6635717			

Поля сводной табл...

Выберите поля для добавления в отчет:

☒ Дата

☒ Регион

☒ Продажи

☐ Кварталы

ДРУГИЕ ТАБЛИЦЫ...

Перетащите поля в нужную область:

ФИЛЬТРЫ КОЛОННЫ

Регион

СТРОКИ

Дата

Σ ЗНАЧЕНИЯ

Сумма по полю...

5. В процессе аналитической обработки профессиональной информации в Excel, какую информацию выведет на экран инструмент «Промежуточные итоги» при заданных значениях на рисунке?:

A1

Штат

А B C D E F G H I

1	Штат	Регион	Месяц	Квартал	Продажи			
2	Калифорния	Запад	Январь	Qtr-1	1 118			
3	Калифорния	Запад	Февраль	Qtr-1	1 960			
4	Калифорния	Запад	Март	Qtr-1				
5	Калифорния	Запад	Апрель	Qtr-2				
6	Калифорния	Запад	Май	Qtr-2				
7	Калифорния	Запад	Июнь	Qtr-2				
8	Вашингтон	Запад	Январь	Qtr-1				
9	Вашингтон	Запад	Февраль	Qtr-1				
10	Вашингтон	Запад	Март	Qtr-1				
11	Вашингтон	Запад	Апрель	Qtr-2				
12	Вашингтон	Запад	Май	Qtr-2				
13	Вашингтон	Запад	Июнь	Qtr-2				
14	Орегон	Запад	Январь	Qtr-1				
15	Орегон	Запад	Февраль	Qtr-1				
16	Орегон	Запад	Март	Qtr-1				
17	Орегон	Запад	Апрель	Qtr-2				
18	Орегон	Запад	Май	Qtr-2				
19	Орегон	Запад	Июнь	Qtr-2				
20	Аризона	Запад	Январь	Qtr-1				
21	Аризона	Запад	Февраль	Qtr-1	1 375			

Промежуточные итоги

При каждом изменении в:

Штат

Операция:

Количество

Добавить итоги по:

☒ Штат

☐ Регион

☐ Месяц

☐ Квартал

☐ Продажи

☒ Заменить текущие итоги

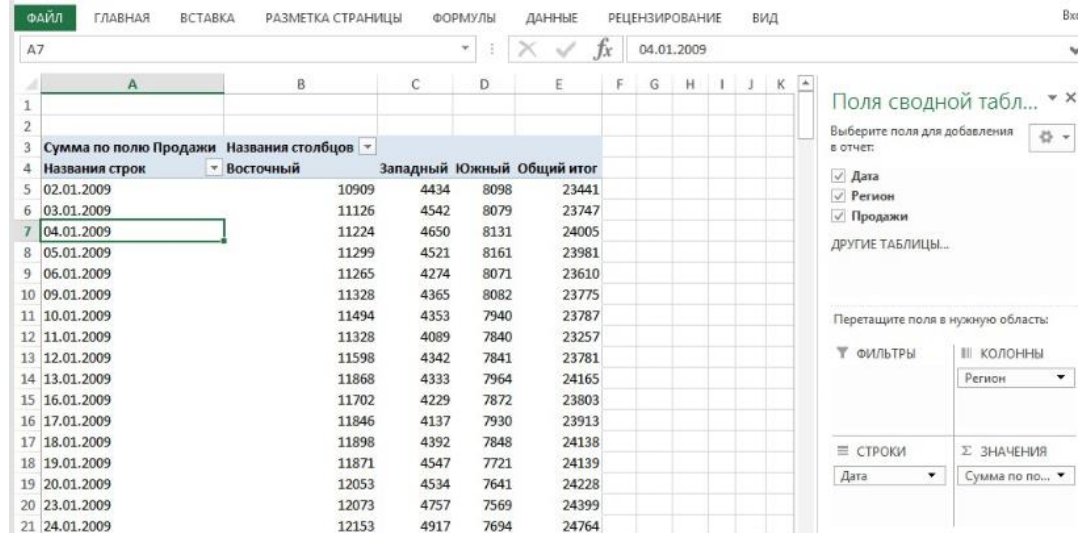
☐ Конец страницы между группами

☒ Итоги под данными

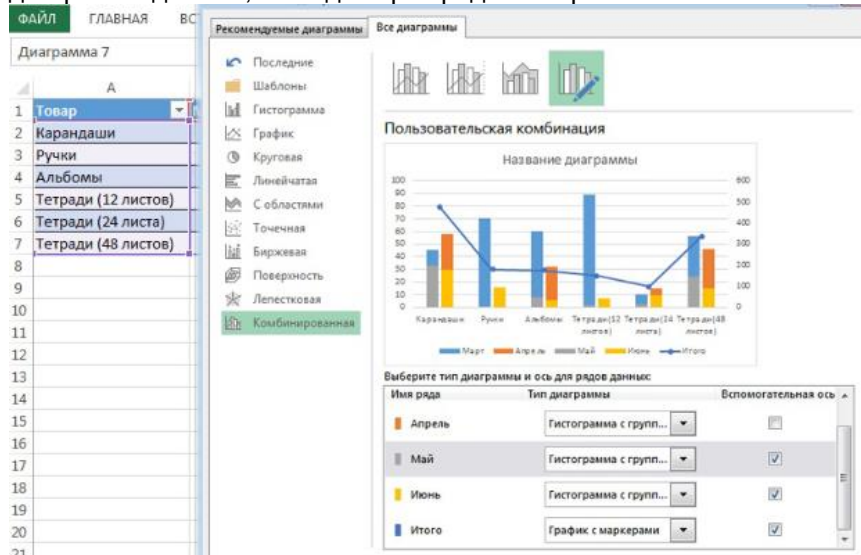
Убрать все

OK Отмена

6. Для формирования аналитического отчета, каким образом можно выполнить группировку данных по месяцам в сводной таблице?



7. Разработка стратегии действий для решения проблемной ситуации: что будет отображать диаграмма данных, если для трех рядов выбрать показ вспомогательной оси?



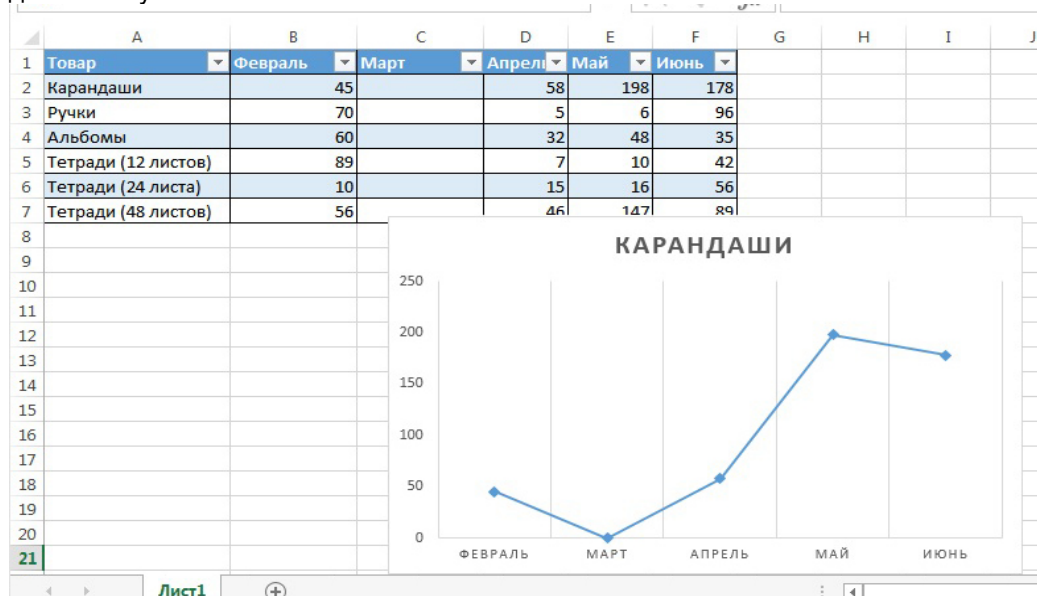
8. В этом подходе при анализе профессиональной информации предполагается, что данные порождены в соответствии с некоторой вероятностной моделью и являются средством ее оценки. Модель может быть наиболее точной формой знания о данном явлении. Однако, модель можно предположить и при слабом уровне знаний, а потом проводить эксперименты, чтобы ее подтвердить. Как называется такая модель поиска решений?

9. Разработка стратегии действий для решения проблемной ситуации: какие товары будут на экране при применении расширенного фильтра с данным условием?

Товар	Цена	Количество	Заказ
Карандаши	10,00р.	45	450
Ручки	15,00р.	70	1050
Альбомы	150,00р.	60	9000
Тетради (12 листов)	25,00р.	89	2225
Тетради (24 листа)	45,00р.	10	450
Тетради (48 листов)	80,00р.	56	4480

Товар	Цена	Количество	Формула	Цена
	<50			>85

10. Какой способ аналитической обработки информации при пропущенных значениях применен в данном случае?



ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ

- «зачтено» выставляется, если обучающийся смог всесторонне раскрыть теоретическое содержание вопроса, способен применить полученные знания при решении практических задач;

- «не зачтено» выставляется, если обучающийся не смог всесторонне раскрыть теоретическое содержание вопроса, не способен применить полученные знания при решении практических задач.

ЛИСТ РАССМОТРЕНИЙ И ОДОБРЕНИЙ

**Фонд оценочных средств учебной дисциплины Б1.В.02 Аналитическая обработка информации
в составе ОПОП 09.04.02 Информационные системы и технологии**

1. Рассмотрен и одобрен в качестве базового варианта:	
а) На заседании обеспечивающей кафедры экономики, бухгалтерского учета и финансового контроля протокол № <u>11</u> от <u>19.05</u> .2022.	
Зав. кафедрой, канд. экон. наук, доцент	 О.А. Блинов
б) На заседании методической комиссии по направлению 09.04.02 Информационные системы и технологии протокол № <u>9</u> от <u>24.05</u> .2022.	
Председатель МКН – 09.04.02, канд. экон. наук	 С.А. Нардина
2. Рассмотрен и одобрен внешним экспертом	
Заместитель генерального директора ООО ФТО «Центр разработки»   Д.В. Малыгин	

ИЗМЕНЕНИЯ И ДОПОЛНЕНИЯ
к фонду оценочных средств учебной дисциплины
Б1.В.02 Аналитическая обработка информации
в составе ОПОП 09.04.02 Информационные системы и технологии

Ведомость изменений

Срок, с которого вводится изменение	Номер и основное содержание изменения и/или дополнения	Отметка об утверждении/ согласовании изменений	
		инициатор изменения	руководитель ОПОП или председатель МКН