

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

ИНСТИТУТ ЭМИТ
Кафедра Системного анализа и информатики

УТВЕРЖДЕНА
на заседании кафедры
Системного анализа и информатики
Протокол от «15» июня 2021 г. № 7

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Индекс Б1.В.06 «Интеллектуальный информационный поиск»

по направлению подготовки 38.04.01 «Экономика»

направленность «Системы больших данных в экономике»

квалификация магистр

очная форма обучения

Год набора - 2021

Москва, 2021 г.

Автор(ы)—составитель(и):

к.т.н., доцент кафедры Системного анализа и информатики Стефановский Д.В.

Заведующий кафедрой

Системного анализа и информатики, к.т.н., доцент, Маруев С.А.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.....	4
2. Объем и место дисциплины в структуре образовательной программы.....	5
3. Содержание и структура дисциплины.....	5
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине.....	7
5. Методические указания для обучающихся по освоению дисциплины	12
6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине	15
6.1. Основная литература	15
6.2. Дополнительная литература	15
6.3. Учебно-методическое обеспечение самостоятельной работы	15
6.4. Нормативные правовые документы	15
6.5. Интернет-ресурсы	15
6.6. Иные источники	15
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы	15

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Дисциплина Б1.В.05 «Интеллектуальный информационный поиск» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-1	Способен подготовить данные для проведения аналитических работ по исследованию больших данных	ПКс-1.1	способен к выбору методов и инструментальных средств для проведения аналитических работ

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

Профессиональные действия	Код этапа освоения компетенции	Результаты обучения
ведение аналитической деятельности в области экономики	ПКс-1.1	на уровне знаний: - Технологии анализа данных: статистический анализ, семантический анализ, методы сравнения средних, частотный анализ, кластерный анализ, факторный анализ, деревья классификации, многомерное шкалирование - Математическое моделирование - Статистические методы: параметрические, непараметрические, управляемые, неуправляемые, полууправляемые - Семантический анализ: обработка естественного языка, sentimentный анализ, анализ текста
		на уровне умений: - Проводить сравнительный анализ методов и инструментальных средств анализа данных - Разрабатывать и оценивать модели - Проводить аналитические работы

2. Объем и место дисциплины в структуре ОП ВО

Общий объем дисциплины 5 ЗЕ (180 ак.ч./135 астр.ч.). В целях планирования, при расчете структуры дисциплины применены академические часы (ак.ч.).

Количество академических часов, выделенных на занятия лекционного типа – 16 ак.ч., занятия семинарского типа – 48 ак.ч.; на самостоятельную работу обучающихся – 78 ак.ч. и промежуточную аттестацию – 36 ак.ч.

Место дисциплины в структуре образовательной программы:

- дисциплина Б1.В.05 «Интеллектуальный информационный поиск» изучается в 1 семестре на 1 курсе;

- дисциплина реализуется после изучения дисциплин:
микроэкономика (в объеме бакалавриата),
математический анализ (в объеме бакалавриата),
линейная алгебра (в объеме бакалавриата),
экономическая информатика (в объеме бакалавриата),
дискретная математика (в объеме бакалавриата),
теория вероятности и математическая статистика (в объеме бакалавриата);
- дисциплина может реализоваться частично или полностью с применением ЭО и/или ДОТ. Учебные материалы дисциплины размещаются по адресу lms.ranepa.ru
- форма промежуточной аттестации – экзамен.

3. Содержание и структура дисциплины

Таблица 2.

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак.час./час						Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Очная форма обучения								
Тема 1	Системы поддержки принятия решений	18	2		6		10	Опрос
Тема 2	Источники данных	24	2		8		14	Опрос
Тема 3	Интеллектуальный информационный поиск	30	4		10		16	Опрос
Тема 4	Визуальный анализ структурированных данных — Visual Mining	18	2		6		10	Опрос
Тема 5	Анализ текстовой информации — Text Mining	18	2		6		10	Опрос
Тема 6	Стандарты Data Mining	18	2		6		10	Опрос
Тема 7	Стандартные диаграммы для признакового описания объектов. Кросс- таблицы. Транзакционная и иерархическая модели данных.	18	2		6		10	Опрос, ДЗ (1-8), КР (1-8)
		2						Консультация
Промежуточная аттестация		36				-	-	Экзамен
Всего:		180\135	16/12			48/36	78/58,5	

Примечание: * формы текущего контроля успеваемости: опрос (О), контрольная работа (КР)

Примечание **: в рамках указанной контактной работы с обучающимися учебные занятия могут проводиться с использованием ДОТ и/или ЭО

Содержание дисциплины

Тема 1. Системы поддержки принятия решений.

Задачи систем поддержки принятия решений. Базы данных — основа СППР. Неэффективность использования OLTP-систем для анализа данных.

Тема 2. Источники данных.

Концепция хранилища данных. Организация ХД. Очистка данных. Концепция хранилища данных и анализ.

Тема 3. Интеллектуальный информационный поиск

Задачи Data Mining: классификация задач; задача классификации и регрессии; задача поиска ассоциативных правил; задача кластеризации. Практическое применение Data Mining: интернет-технологии, торговля, телекоммуникации, промышленное производство, медицина, банковское дело, страховой бизнес и другие области применения. Модели Data Mining: предсказательные модели, описательные модели. Методы Data Mining: базовые методы; нечеткая логика; генетические алгоритмы; нейронные сети. Процесс обнаружения знаний: основные этапы анализа, подготовка исходных данных. Управление знаниями (Knowledge Management). Средства Data Mining.

Тема 4. Визуальный анализ структурированных данных — Visual Mining

Выполнение визуального анализа данных. Характеристики средств визуализации данных. Методы визуализации: методы геометрических преобразований; отображение иконок; методы, ориентированные на пиксели; иерархические образы.

Тема 5. Анализ текстовой информации — Text Mining

Задача анализа текстов: этапы анализа текстов; предварительная обработка текста; задачи Text Mining. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста; стадия локального анализа; стадия интеграции и вывода понятий. Классификация текстовых документов: описание задачи классификации текстов; методы классификации текстовых документов. Методы кластеризации текстовых документов: представление текстовых документов; иерархические методы кластеризации текстов; бинарные методы кластеризации текстов. Задача аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации. средства Orange 3.0 или Knime или Полианалист.

Тема 6. Стандарты Data Mining

Кратко о стандартах. Стандарт CWM: назначение, структура и состав CWM, пакет Data Mining. Стандарт CRISP: появление, структура, фазы и задачи. Стандарт PMML. Другие стандарты Data Mining: стандарт Microsoft Data Mining eXtensions (DMX); стандарт Java Data Mining.

Тема 7. Стандартные диаграммы для признакового описания объектов. Кросс-таблицы. Транзакционная и иерархическая модели данных.

Серии диаграмм. Визуализация наборов точек из многомерного пространства. Иконографический подход.

Стандартные диаграммы для многомерной модели данных. Понятие о семантическом слое. Системы отчётности.

Формальные контексты. Частично-упорядоченные множества. Диаграммы Хассе, TreeMap, диаграммы Эйлера-Венна. Карты, различные виды картограмм.

Понятие о содержательных и эстетических требованиях в визуализации. Первые этапы разработки нового метода визуализации.

Разработка математических процедур для нового метода визуализации. Разбор полного цикла создания нового метода визуализации на примере карты науки.

Распределение фундаментальных задач ИАД и основных инструментов статистики по моделям данных: в разрезе исходных данных, в разрезе результатов. Модели источника (source models)

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости.

4.1.1. В ходе реализации дисциплины Б1.В.05 «Интеллектуальный информационный поиск» используются следующие методы текущего контроля успеваемости обучающихся: домашние задания и контрольные работы.

Тема	Методы текущего контроля успеваемости
Тема 1	Опрос
Тема 2	Опрос
Тема 3	Опрос
Тема 4	Опрос
Тема 5	Опрос
Тема 6	Опрос
Тема 7	Опрос
Тема 8	Опрос, домашнее задание по темам 1-8, контрольная работа по темам 1-8

4.1.2. Экзамен проводится с применением следующих методов (средств): в письменной форме в виде контрольной работы.

4.2. Материалы текущего контроля успеваемости обучающихся

Типовые оценочные материалы по теме 1

Опрос:

- Перечислите основные задачи систем поддержки принятия решений. Поясните на примере.
- Опишите применение баз данных в СППР. Поясните на примере.
- Обоснуйте неэффективность использования OLTP-систем для анализа данных. Поясните на примере.

Типовые оценочные материалы по теме 2

Опрос:

- Опишите концепцию хранилищ данных. Приведите общую схему организации ХД.
- Опишите процедуру очистки данных. Поясните ее на примере.
- Поясните связь концепции хранилищ данных и анализа данных.

Типовые оценочные материалы по теме 3

Опрос:

- Опишите следующие задачи Data Mining: классификация задач; задача классификации и регрессии; задача поиска ассоциативных правил; задача кластеризации. Поясните на примерах назначение и особенности использования.
- Опишите основные особенности практического применения Data Mining в интернет-технологии. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в торговле. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в телекоммуникации. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в промышленное производство. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в банковском деле. Поясните на примере.

- Опишите следующие модели Data Mining: предсказательные модели, описательные модели. Поясните на примерах.
- Опишите следующие методы Data Mining: базовые методы; нечеткая логика; генетические алгоритмы; нейронные сети. Поясните на примерах.
- Опишите основные особенности процесса обнаружения знаний. Поясните на примерах.

Типовые оценочные материалы по теме 4

Опрос:

- Опишите основные этапы и особенности выполнения визуального анализа данных. Поясните на примерах.
- Перечислите основные характеристики средств визуализации данных. Поясните на примере пути их улучшения.
- Перечислите следующие методы визуализации: методы геометрических преобразований; отображение иконок; методы, ориентированные на пиксели; иерархические образы. Приведите примеры.

Типовые оценочные материалы по теме 5

Опрос:

- Опишите следующие задачи анализа текстов: этапы анализа текстов; предварительная обработка текста; задачи Text Mining. Поясните на примерах.
- Опишите основные принципы алгоритмов извлечения ключевых понятий из текста. Поясните на примере.
- Опишите основные этапы и задачи классификация текстовых документов. Поясните на примере.
- Перечислите основные задачи классификации текстов. Поясните на примерах.
- Перечислите основные методы классификации текстовых документов. Поясните на примерах.
- Перечислите основные особенности кластеризации текстовых документов. Поясните на примерах.
- Опишите суть задач аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации. Поясните на примерах.
- Перечислите существующие средства анализа текстовой информации. Поясните на примерах их достоинства и недостатки

Типовые оценочные материалы по теме 6

Опрос:

- Опишите особенности стандарта CWM. Поясните на примере.
- Стандарт CRISP: появление, структура, фазы и задачи. Поясните на примере задачи и этапы.
- Стандарт PMML. Поясните суть и особенности на примерах.

Типовые оценочные материалы по теме 7

Опрос:

- Опишите особенности и суть визуализация наборов точек из многомерного пространства. Поясните на примерах
- Опишите особенности иконографического подхода. Поясните на примере. Дайте понятие о семантическом слое. Поясните на примере использование семантических слоев.
- Дайте понятие о формальных контекстах. Поясните на примерах.

Домашнее задание.

1. Дайте сравнительное описание процесса извлечения ключевых понятий из текста.
2. Предложите алгоритм тематического моделирования корпуса текстов, каждому студенту выдается уникальный набор новостных текстов. Алгоритм должен быть реализован на одном из языков программирования и в результате определять сколько и какие темы присутствуют в корпусе.

Контрольная работа:

Вопрос. Расскажите о стандарт CRISP: появление, структура, фазы и задачи?

Задача. Дайте укрупненное описание решения задач анализа данных согласно стандарту CRISP.

4.3. Оценочные средства для промежуточной аттестации.

4.3.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования.

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-1	Способен подготовить данные для проведения аналитических работ по исследованию больших данных	ПКс-1.1	способен к выбору методов и инструментальных средств для проведения аналитических работ

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ПКс-1.1	способен к выбору методов и инструментальных средств для проведения аналитических работ, разработке, проверке, оценке используемых работ	понимает и может обосновать целесообразность выбранных методов и инструментальных средств для проведения аналитических работ

4.3.2. Типовые оценочные средства.

Вопросы к экзамену:

- Перечислите основные задачи систем поддержки принятия решений. Поясните на примере.
- Опишите применение баз данных в СППР. Поясните на примере.
- Обоснуйте неэффективность использования OLTP-систем для анализа данных. Поясните на примере.
- Опишите концепцию хранилищ данных. Приведите общую схему организация ХД.
- Опишите процедуру очистки данных. Поясните ее на примере.
- Поясните связь концепции хранилищ данных и анализа данных.
- Опишите многомерную модель данных и дайте определение OLAP-систем. Поясните на примерах.
- Поясните суть концептуальное многомерное представление. Поясните на примере.

- Опишите следующие задачи Data Mining: классификация задач; задача классификации и регрессии; задача поиска ассоциативных правил; задача кластеризации. Поясните на примерах назначение и особенности использования.
- Опишите основные особенности практического применения Data Mining в интернет-технологии. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в торговле. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в телекоммуникации. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в промышленном производстве. Поясните на примере.
- Опишите основные особенности практического применения Data Mining в банковском деле. Поясните на примере.
- Опишите следующие модели Data Mining: предсказательные модели, описательные модели. Поясните на примерах.
- Опишите следующие методы Data Mining: базовые методы; нечеткая логика; генетические алгоритмы; нейронные сети. Поясните на примерах.
- Опишите основные особенности процесса обнаружения знаний. Поясните на примерах.
- Опишите основные этапы и особенности выполнения визуального анализа данных. Поясните на примерах.
- Перечислите основные характеристики средств визуализации данных. Поясните на примере пути их улучшения.
- Перечислите следующие методы визуализации: методы геометрических преобразований; отображение иконок; методы, ориентированные на пиксели; иерархические образы. Приведите примеры.
- Опишите следующие задачи анализа текстов: этапы анализа текстов; предварительная обработка текста; задачи Text Mining. Поясните на примерах.
- Опишите основные принципы алгоритмов извлечения ключевых понятий из текста. Поясните на примере.
- Опишите основные этапы и задачи классификации текстовых документов. Поясните на примере.
- Перечислите основные задачи классификации текстов. Поясните на примерах.
- Перечислите основные методы классификации текстовых документов. Поясните на примерах.
- Перечислите основные особенности кластеризации текстовых документов. Поясните на примерах.
- Опишите суть задач аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации. Поясните на примерах.
- Перечислите существующие средства анализа текстовой информации. Поясните на примерах их достоинства и недостатки.
- Опишите особенности стандарта CWM. Поясните на примере.
- Стандарт CRISP: появление, структура, фазы и задачи. Поясните на примере задачи и этапы.
- Стандарт PMML. Поясните суть и особенности на примерах.
- Опишите особенности и суть визуализации наборов точек из многомерного пространства. Поясните на примерах.

- Опишите особенности иконографического подхода. Поясните на примере. Дайте понятие о семантическом слое. Поясните на примере использование семантических слоев.
- Дайте понятие о формальных контекстах. Поясните на примерах.
- Дайте сравнительное описание процесса извлечения ключевых понятий из текста.
- Предложите алгоритм тематического моделирования корпуса текстов, каждому студенту выдается уникальный набор новостных текстов. Алгоритм должен быть реализован на одном из языков программирования и в результате определять сколько и какие темы присутствуют в корпусе.
- Расскажите о стандарт CRISP: появление, структура, фазы и задачи?
- Дайте укрупненное описание решения задач анализа данных согласно стандарту CRISP.

Шкала оценивания.

Оценка определяется по формуле: $\frac{1}{4}$ опрос + $\frac{1}{4}$ д.з. + $\frac{1}{4}$ к.р. + $\frac{1}{4}$ сдача зачета.

10-бальная шкала	Традиционная шкала	«Зачтено»/«Не зачтено»	Определение
10	Отлично	Зачтено	Полные, глубокие и систематические знания, полный и правильный ответ на теоретический вопрос, полное и правильное решение задачи.
9	Отлично	Зачтено	Глубокие и систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
8	Отлично	Зачтено	Систематические знания, правильный ответ на теоретический вопрос, правильное решение задачи.
7	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи.
6	Хорошо	Зачтено	Систематические знания, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи с незначительными неточностями.
5	Удовлетворительно	Зачтено	Ответ на теоретический вопрос неполный, правильное решение задачи с незначительными неточностями.
4	Удовлетворительно	Зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит арифметические ошибки, не влияющие на правильность хода решения задачи.
3	Неудовлетворительно	Не зачтено	Ответ на теоретический вопрос неполный, решение задачи содержит идеологические ошибки.
2	Неудовлетворительно	Не зачтено	Ответ на теоретический вопрос неверный и/или решение задачи содержит идеологические ошибки.
1	Неудовлетворительно	Не зачтено	Ответ на теоретический вопрос неверный и решение задачи отсутствует.
0	Неудовлетворительно	Не зачтено	Ответ на теоретический вопрос отсутствует и решение задачи отсутствует.

4.4. Методические материалы по проведению промежуточной аттестации

Экзамен проводится в аудитории. Отсчет времени, отведенного на письменную работу, идет по завершении процедуры размещения студентов и раздачи заданий.

Студент обязан являться на письменный контроль в указанное в расписании время. В случае опоздания время, отведенное на письменный контроль знаний, не продлевается.

При себе студенты могут иметь только письменные принадлежности. Необходимую для выполнения работы бумагу выдает преподаватель.

Преподаватель раздает варианты работы, содержащий 2 вопроса. Листы с заданиями должны быть повернуты текстом вниз, чтобы студенты до окончания процедуры раздачи не могли начать выполнение работы. По окончании раздачи вариантов студентам разрешается перевернуть текст задания и одновременно приступить к выполнению работы. По окончании отведенного времени студенты одновременно заканчивают выполнение работы. Если работа завершена существенно раньше срока, то по разрешению преподавателя студент может покинуть аудиторию досрочно.

Мобильные телефоны должны быть выключены и убраны со столов, допускается использование калькуляторов, выполняющих только простые арифметические вычисления.

Во время проведения письменного контроля знаний студентам не разрешается пользоваться учебными программами, справочниками и прочими источниками информации.

Использование материалов, а также попытка общения с другими студентами или иными лицами, в том числе с применением электронных средств связи, несанкционированные перемещения и т.п. являются основанием для удаления студента из аудитории и последующего проставления в ведомость оценки «неудовлетворительно».

Во время проведения письменного контроля знаний студентам разрешается покинуть аудиторию только при условии сдачи работы в объеме, выполненном к моменту выхода из аудитории. Дальнейшее продолжение работы запрещается.

Ответы в работе без объяснений не засчитываются. Рисунки должны быть четкими, все линии графиков, используемых при ответах на вопросы задач, должны быть подписаны.

Продолжительность экзаменационной письменной работы 120 минут.

5. Методические указания для обучающихся по освоению дисциплины

Любой вид занятий, создающий условия для зарождения самостоятельной мысли, познавательной и творческой активности студента связан с самостоятельной работой. В широком смысле под самостоятельной работой понимают совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие. Самостоятельная работа может реализовываться: непосредственно в процессе аудиторных занятий – на лекциях, практических и семинарских занятиях, при выполнении контрольных и лабораторных работ и др.; в контакте с преподавателем вне рамок аудиторных занятий – на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре и других местах при выполнении студентом учебных и творческих заданий.

Лекции

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности

и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции. Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось присить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Семинар и проведение опроса

Каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений студенту необходимо самостоятельно выбрать тему доклада по проблеме семинара и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Работа с литературными источниками.

В процессе подготовки к семинарским занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также

научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Методические указания по выполнению самостоятельной работы:

Тема 1. Системы поддержки принятия решений.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Задачи систем поддержки принятия решений. Базы данных — основа СППР. Неэффективность использования OLTP-систем для анализа данных.

Тема 2. Источники данных.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Концепция хранилища данных. Организация ХД. Очистка данных. Концепция хранилища данных и анализ.

Тема 3. Интеллектуальный информационный поиск.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Задачи Data Mining: классификация задач; задача классификации и регрессии; задача поиска ассоциативных правил; задача кластеризации. Практическое применение Data Mining: интернет-технологии, торговля, телекоммуникации, промышленное производство, медицина, банковское дело, страховой бизнес и другие области применения. Модели Data Mining: предсказательные модели, описательные модели. Методы Data Mining: базовые методы; нечеткая логика; генетические алгоритмы; нейронные сети. Процесс обнаружения знаний: основные этапы анализа, подготовка исходных данных. Управление знаниями (Knowledge Management). Средства Data Mining.

Тема 4. Визуальный анализ структурированных данных – Visual Mining.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Выполнение визуального анализа данных. Характеристики средств визуализации данных. Методы визуализации: методы геометрических преобразований; отображение иконок; методы, ориентированные на пиксели; иерархические образы.

Тема 5. Анализ текстовой информации — Text Mining

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Задача анализа текстов: этапы анализа текстов; предварительная обработка текста; задачи Text Mining. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста; стадия локального анализа; стадия интеграции и вывода понятий. Классификация текстовых документов: описание задачи классификации текстов; методы классификации текстовых документов. Методы кластеризации текстовых документов: представление текстовых документов; иерархические методы кластеризации текстов; бинарные методы кластеризации текстов. Задача аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации. Средства анализа текстовой информации: средства Oracle — Oracle Text; средства от IBM

— Intelligent Miner for Text; средства SAS Institute — Text Miner; средства Мегапьютер Интеллидженс — TextAnalyst.

Тема 6. Стандарты Data Mining

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Кратко о стандартах. Стандарт CWM: назначение, структура и состав CWM, пакет Data Mining. Стандарт CRISP: появление, структура, фазы и задачи. Стандарт PMML. Другие стандарты Data Mining: стандарт SQL/MM; стандарт Microsoft Data Mining eXtensions (DMX); стандарт Java Data Mining.

Тема 7. Стандартные диаграммы для признакового описания объектов. Кросс-таблицы. Транзакционная и иерархическая модели данных.

Студент изучает предложенные материалы темы, решает задачи по теме и готовится к опросу по следующей тематике:

Серии диаграмм. Визуализация наборов точек из многомерного пространства. Иконографический подход.

Стандартные диаграммы для многомерной модели данных. Понятие о семантическом слое. Системы отчётности.

Формальные контексты. Частично-упорядоченные множества. Диаграммы Хассе, TreeMap, диаграммы Эйлера-Венна. Карты, различные виды картограмм.

Понятие о содержательных и эстетических требованиях в визуализации. Первые этапы разработки нового метода визуализации.

Разработка математических процедур для нового метода визуализации. Разбор полного цикла создания нового метода визуализации на примере карты науки.

Распределение фундаментальных задач ИАД и основных инструментов статистики по моделям данных: в разрезе исходных данных, в разрезе результатов. Модели источника (source models).

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Основная литература.

1. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с.: ил. + CD-ROM — (Учебная литература для вузов) ISBN 978-5-9775-0368-6.
2. Эдвард Тафти. Визуальное представление больших объемов информации.

6.2. Дополнительная литература.

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. — М.: МИЭМ, 2011. — 272 с.
2. Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. Анализ больших наборов данных

6.3. Учебно-методическое обеспечение самостоятельной работы.

Отдельное обеспечение не предусмотрено.

6.4. Нормативные правовые документы.

Не предусмотрены.

6.5. Интернет-ресурсы.

1. <http://citeseer.ist.psu.edu/> — основной источник знаний по Computer Science, по многим статьям есть полные тексты
2. <http://citeseer.cs.msu.su/> — русскоязычная электронная библиотека научных статей
3. <http://arxiv.org/> — библиотека электронных публикаций, в основном по физике, но доля «Computer Science» в последнее время стремительно увеличивается
4. <http://rexa.info/> — библиографическая поисковая система по статьям, авторам и грантам
5. <http://elibrary.ru/> — российская научная электронная библиотека
6. <http://iinwww.ira.uka.de/bibliography/index.html>
7. библиографическая база данных для работы с BibTeX
8. <http://www.gotai.net/> -- русскоязычный сайт об искусственном интеллекте
9. Math-Net.ru -- общероссийский математический портал

6.6. Иные источники.

Не предусмотрены.

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Для лекций:

1. Персональный компьютер
2. Мультимедийный проектор
3. Доска, мел или маркеры

Для лабораторных занятий:

1. Компьютерный класс,
2. Виртуальная машина Ubuntu 15.04 b выше с установленным Postgresql и MongoDB
3. Мультимедийный проектор
4. Доска, маркеры
5. Компилятор R-2.15.1 – GNU - <http://www.r-project.org/> либо интегрированная среда разработки RStudio – GNU AGP - <http://www.rstudio.com/ide/>.
6. Jupyter Notebook - бесплатная интерактивная оболочка для языка программирования Python, позволяющая объединить код, текст и диаграммы.
7. Компилятор Scala – <http://www.scala-lang.org/>
8. Программный комплекс анализа новостного сайта - "Crawler-Persona"
9. База данных "Централизация государственных закупок в 2014 г".
10. База данных учебно-методических материалов по дисциплине "Макроэкономика".
11. База данных Бюджетная и социально-экономическая статистика субъектов Российской Федерации.