

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ
И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

УТВЕРЖДАЮ
Зав. кафедрой АОИ, профессор
_____ Ю.П. Ехлаков

“ ___ ” _____ 2016 г.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

по выполнению самостоятельной работы по дисциплине:

**МЕТОДЫ И АЛГОРИТМЫ РАСПОЗНАВАНИЯ И ОБРАБОТКИ
ДАННЫХ**

(Дисциплина вариативной части)

Уровень основной образовательной программы: **магистратура**

Направление подготовки магистра: **09.04.04 - Программная инженерия**

Магистерская программа: **"Методы и технологии индустриального
проектирования программного обеспечения"**

Проф. Н.В. Замятин

Содержание

1. Введение	3
2. Рекомендации по выполнению самостоятельной работы.....	4
3. Вопросы по темам для самостоятельной подготовки.....	7
4. Учебно-методические материалы	8

Введение

Дисциплина «**Методы и алгоритмы распознавания и обработки данных**» представляет систематическое изложение материала по методам распознавания и обработки данных, и дает базовые знания, необходимые специалисту по программной инженерии независимо от его специализации. Наряду с изучением способов обработки данных и их распознавания, как единого целого рассматриваются основные понятия и наиболее важные характеристики программных и аппаратных компонентов, используемых для обработки данных, вопросы сжатия данных и распознавания текстов.

Большой объем знаний по данной теме не позволяет изучать весь материала в рамках аудиторных занятий. Поэтому часть материала выносится для самостоятельной работы. С одной стороны такое обучение позволяет студенту осваивать материал самостоятельно и углубленно, а с другой стороны и готовить себя к постоянному обучению в течении всей жизни после университета.

Данное методическое пособие должно помочь студенту правильно выбрать тему, выделить проблемные места, сформулировать вопросы, по которым студент может оценить степень усвоения материала, а также указать необходимую литературу для самостоятельного изучения разделов данной дисциплины.

Распределение самостоятельной работы (всего – 90 часов)

Виды самостоятельной работы	Контроль выполнения работы				
	Всего по виду работы				Виды контроля
	1	2	3		
1. Проработка лекционного материала	6	6	6	18	Тестовый опрос,

					проверка конспекта, мини-лекция
2. Подготовка к лабораторным работам	9	9	9	27	Отчет по ЛР
3. Подготовка к контрольным работам	1	2	2	5	Контрольная работа
4. Задачи распознавания и обработки данных:	10	–	–	10	
5. Статистические методы обработки данных.		10	-	10	
6. Основные методы многомерной обработки данных	–	–	10	10	
7. Подготовка мини-доклада			10	10	
Итого по 1 семестру (разделы 1–3)	26	27	37	90	
Итого по разделу дисциплины					
Подготовка и сдача экзамена				36	Экзамен

2. Рекомендации по выполнению самостоятельной работы

Самостоятельная работа студентов выражается в освоении необходимого объема учебной программы по дисциплине, выработке навыков профессиональной деятельности при изучении вынесенных на самостоятельную работу вопросов. Это выражается в подготовке к лекционным и лабораторным занятиям, подготовке рефератов, выполнении соответствующих заданий. В связи с большим объемом материала по методам обработки данных и их распознаванию того минимального времени, отведенного для их изучения учебным планом, явно недостаточно. Поэтому студентам предлагается провести самостоятельное углубленное изучение ряда тем, а результаты его выразить в письменных ответах на вопросы заданий в виде рефератов или обзорных лекций.

Прежде чем приступить к изложению ответов на вопросы задания, студент получает необходимые знания об основных понятиях, терминах, общих вопросах обработки и распознавания данных из лекций. Эти знания также можно получить в процессе самоподготовки по предлагаемым вопросам к теме по дисциплине.

Самостоятельная подготовка состоит в подборке и изучении предлагаемой в настоящем указании учебно-методической литературы, а также использовании

дополнительной литературы. В связи с быстрым развитием технологий обработки данных и знаний, литература, на которую делаются ссылки на момент изучения данной дисциплины может устареть. Поэтому при выполнении работы целесообразно использовать Интернет.

Темы рефератов, контрольные вопросы, письменные ответы на которые являются основой для оценки результатов самостоятельной работы, определяются текущей темой лекции. Приветствуется инициативное предложение варианта темы реферата студентом, в рамках перспективных разработок программного обеспечения и информационных сетей, как объектов продаж рынка ИКТ.

При непосредственном написании рефератов недопустимо дословное копирование материала из интернета, а также с использованием сканеров фрагментов текста из учебников, учебных пособий и брошюр, за исключением отдельных научных и практически положений. В этом случае необходимо сделать ссылку на первоисточник.

При подготовке реферата должны соблюдаться такие требования, как логическая последовательность изложения ответа по каждому вопросу, убедительность аргументации, если она присутствует, краткость, конкретность и точность формулировок, исключающих возможность неоднозначного толкования.

Объем реферата 10–12 страниц 14 шрифтом (Times New Roman).

В тексте необходимо применять научные и технические термины, обозначения и определения, установленные соответствующими стандартами, а при их отсутствии – общепринятые в научной и технической литературе. Не допускается применение для одного и того же понятия без соответствующих пояснений различных научных терминов, близких по смыслу (синонимов) и иностранных слов, сокращений слов, обозначений, кроме установленных правилами русской орфографии или соответствующими государственными стандартами.

Ответы на вопросы реферата по текущей теме должны быть даны в электронном и письменном виде. Электронный вариант высылается преподавателю по электронной почте в течение недели после получения текущей темы. Письменный вариант приносится на лекцию или контрольную работу. Подсчет страниц (листов) начинается с титульного листа, но номер на нем не ставится. Номер страницы проставляется в правом верхнем углу. Нумерация страниц текста работы и приложений должна быть сквозной.

Реферат должен включать следующие обязательные структурные части: титульный лист, оглавление (содержание), основная часть ответов по каждому вопросу, список использованной литературы, приложения. Возможна краткая оценка современного состояния исследуемой проблемы (введение) по всем или в отдельности по каждому вопросу задания. Введение располагается после оглавления.

При наличии затруднений в подборе учебной и нормативной литературы для выполнения задания, в изучении отдельных вопросов по дисциплине **«Методы и алгоритмы распознавания и обработки данных»** студенты могут получить необходимую консультацию у преподавателя (кабинет 429 корпуса ФЭТ) в специально определенные для этого дни.

Письменная работа студента, являющаяся основой для оценки результатов его самостоятельной работы, состоит из ответа на контрольные вопросы и решения теста. Она выполняется студентом самостоятельно по каждой теме, определенной рабочей программой, отдельно и лично им предоставляется преподавателю в указанный им срок, но не позднее 15 дней до экзамена.

Если реферат не зачтен, то с учетом замечаний преподавателя подлежит доработке и повторной сдаче.

Студенты, не предоставившие письменной работы и не доработавшие ее после замечаний преподавателя, к экзамену или зачету по дисциплине не допускаются.

Письменная работа подписывается студентом ее выполнившим, с указанием фамилии, инициалов, группы, даты и сдается для проверки и рецензирования преподавателю.

3. Вопросы по темам для самостоятельной подготовки

1. Понятие анализа данных, его основные задачи. Роль анализа данных в современном бизнесе.
2. Специфика работы с информацией. Свойства информации. Информация и данные. Виды и методы анализа данных.
3. Понятие OLAP-технологии. Многомерная модель данных. Схемы реализации OLAP-куба.
4. Архитектура OLAP-систем. Виды OLAP- систем: преимущества и недостатки.
5. Сферы применения OLAP-систем. Преимущества и недостатки OLAP-систем.
6. Понятие интеллектуального анализа данных (Data Mining). Задачи и стадии DM.
7. Модели Data Mining. Сферы применения технологии Data Mining.
8. Методики обнаружения нового знания в хранилищах данных (KDD). Этапы KDD.
9. Алгоритмы и задачи таксономии: динамическая и иерархическая таксономии, таксономия с суперцелью. Алгоритмы таксономии класса FOREL.
10. Методы Data Mining: классификация. Основные алгоритмы классификационного анализа.
11. Методы Data Mining: регрессия. Основные алгоритмы регрессионного анализа.
12. Методы Data Mining: поиск ассоциативных правил. Основные алгоритмы ассоциативного анализа.
13. Методы Data Mining: кластеризация. Основные алгоритмы кластерного анализа.
14. Визуальный анализ данных (Visual Mining): этапы, преимущества и недостатки.
15. Характеристики средств визуализации: типы данных, инструменты визуализации.
16. Методы визуализации, их преимущества и недостатки.
17. Анализ текстовой информации (Text Mining), его задачи.
18. Этапы анализа текстов.
19. Извлечение ключевых понятий: стадия локального анализа.
20. Извлечение ключевых понятий: стадия интеграции и вывода понятий.
21. Аннотирование текстов. Современные инструменты для анализа текстовой информации.

22. Системы мобильных агентов. Использование мобильных агентов для анализа данных.
23. Система анализа распределенных данных.

4. Учебно-методические материалы

Рекомендуемая литература

Основная

1. Воскобойников, Ю.Е. Регрессионный анализ данных в пакете MATHCAD + CD [Электронный ресурс] : учебное пособие. — Электрон. дан. — СПб. : Лань, 2011. — 224 с. — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id
2. Крутиков, В.И. Анализ данных [Электронный ресурс] : учебное пособие / В.И. Крутиков, В.В. Мешечкин. — Электрон. дан. — Кемерово : Издательство КемГУ (Кемеровский государственный университет), 2014. — 138 с. — Режим доступа: http://e.lanbook.com/books/element.php?pl1_id

Дополнительная

1. Красильников Н.Н. Цифровая обработка 2D- и 3D-изображений. Учеб. пособие. СПб.: БХВ-Петербург, 2011. - 608 с.
2. Миркин, Б.Г. (2014). Введение в анализ данных: учебник и практикум для бакалавриата и магистратуры. Издательство Юрайт.
3. <http://habrahabr.ru/post/172043/> – корреляционный анализ (доступно, русском);
4. <http://habrahabr.ru/post/167341/> – бинарные тесты в R (на русском);
5. <http://habrahabr.ru/post/168877/> – статистические тесты в R для категориальных переменных (на русском);
6. <http://habrahabr.ru/post/176795/> – сравнение количественных переменных в R (на русском).

Для обеспечения дисциплины используются следующие УМП:

1. Замятин Н.В. Методические указания к выполнению лабораторных работ по дисциплине «Методы и алгоритмы распознавания и обработки данных». - Томск: ТУСУР, каф. АОИ, 2016 - 100 с. [Электронный ресурс]: сайт кафедры АОИ. – URL: http://aoi.tusur.ru/upload/methodical_materials/MU_BI_SRV_lab_ubg_8_file__82_4967.pdf
2. Замятин Н.В. Методические указания по организации самостоятельной работы студента по дисциплине «Методы и алгоритмы распознавания и обработки данных». - Томск : ТУСУР, каф. АОИ, 2012 - 10 с. [Электронный ресурс]: сайт каф. АОИ. – URL: http://aoi.tusur.ru/upload/methodical_materials/MU_BI_SRV_sam_ubg_5_file__79_4913.pdf.