

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Рector

Дата подписания: 28.06.2023 14:24:59

Уникальный программный ключ:

c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Директор Института магистратуры

Иванова Е.А.

«01» июня 2023 г.

Рабочая программа дисциплины

Информационный поиск и обработка естественного языка

Направление 09.04.03 Прикладная информатика

магистерская программа

09.04.03.03 Машинное обучение и технологии больших данных

Для набора 2023 года

Квалификация

магистр

Кафедра Информационных систем и прикладной информатики

Составители рабочей программы:

доцент Хаймин Евгений Сергеевич

СОДЕРЖАНИЕ

| | |
|---|----|
| I. Цели и задачи освоения дисциплины | 4 |
| II. Место дисциплины в структуре образовательной программы | 4 |
| III. Требования к результатам освоения дисциплины | 6 |
| IV. Содержание и структура дисциплины | 7 |
| 4.1. Содержание дисциплины, структурированное по темам | 7 |
| 4.2. План внеаудиторной самостоятельной работы | 8 |
| 4.3. Содержание учебного материала | 10 |
| V. Образовательные технологии | 12 |
| VI. Учебно-методическое обеспечение дисциплины | 12 |
| 6.1. Основная литература | 12 |
| 6.2. Дополнительная литература | 12 |
| 6.3. Периодические издания | 13 |
| 6.4. Перечень ресурсов сети Интернет | 13 |
| VII. Материально-техническое обеспечение дисциплины | 14 |
| VIII. Методические указания для обучающихся по освоению дисциплины | 14 |
| IX. Учебная карта дисциплины | 15 |
| X. Фонд оценочных средств | 16 |
| 10.1. Паспорт фонда оценочных средств | 16 |
| 10.2. Практические работы №№ 1–6 (выполнение, подготовка отчёта, защита отчёта) | 16 |
| 10.3. Задания для контрольной работы | 16 |
| 10.4. Индивидуальное проектное задание | 20 |

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Информационный поиск и обработка естественного языка» (ИПиОЯ) является ознакомление обучающихся с методами обработки текста на естественном языке, а также методами обработки слабоструктурированных данных и извлечения информации. Предполагается знакомство с методами извлечения отношений, анализа тональности, аннотирования и кластеризации текстов, а также с существующими программными реализациями этих методов.

Основными задачами при этом являются:

- получение теоретических знаний и практических навыков информационного поиска и обработки естественно-языковых текстов;
- получение представления о сложностях, связанных с применением существующих методов информационного поиска и обработки естественно-языковых текстов;
- развитие умений использовать полученные знания по разработке, адаптации и использованию новейших средств информатики для информационного поиска и обработки текстов на естественных языках.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к модулю профессиональных дисциплин, формируемому участниками образовательных отношений, части образовательной программы, формируемой участниками образовательных отношений.

В соответствии с рабочим учебным планом данная дисциплина изучается в третьем семестре, на освоение дисциплины отводится 54 часа аудиторной работы (18 часов лекционных и 36 часов практических занятий), 162 часа самостоятельной работы студента.

Для изучения данной учебной дисциплины необходимы следующие знания и умения, формируемые предшествующими элементами образовательной программы:

| Наименование дисциплины (модуля), практики | Требуемые знания, умения, навыки |
|--|---|
| Методы машинного обучения | <p><i>Знания:</i></p> <ul style="list-style-type: none">– Знает классы методов и алгоритмов машинного обучения.– Знает возможности современных инструментальных средств и систем программирования для решения задач машинного обучения.– Знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения. <p><i>Умения:</i></p> <ul style="list-style-type: none">– Умеет ставить задачи и адаптировать методы и алгоритмы машинного обучения.– Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения.– Умеет применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения. |

| Наименование дисциплины (модуля), практики | Требуемые знания, умения, навыки |
|--|--|
| Современные проблемы и методы прикладной информатики | <p><i>Знания:</i></p> <ul style="list-style-type: none"> – Современного состояния и тенденций научно-технического развития информационного общества. – Направлений современных исследований и аналитические инструменты в прикладной информатике. – Основных тенденции и прогноза научно-технического развития в области информационно-коммуникационных технологий. – Современных методов и средств информатики для решения прикладных задач. – Научные фронтиры в области компьютерных наук: последние достижения, современные вызовы и открытые вопросы. – Знает содержание, объекты и субъекты информационного общества и цифровой экономики, критерии эффективности функционирования информационного общества, теоретические проблемы информатики, искусственного интеллекта, современные методы, средства, стандарты информатики для решения прикладных задач различных классов; правовые, экономические, социальные и психологические аспекты информатизации деятельности организационно-экономических систем. – Знает состав современных методов и средств информатики, передовые методы искусственного интеллекта для решения задач профессиональной деятельности. <p><i>Умения:</i></p> <ul style="list-style-type: none"> – Анализировать особенности и состояние современного информационного общества и пути его развития. – Анализировать возможности и выбирать современные методы и средства информатики для решения прикладных задач. – Умеет применять при решении задач профессиональной деятельности критерии эффективности функционирования информационного общества и цифровой экономики; структуру интеллектуального капитала, методы оценки эффективности. – Умеет проводить анализ современных методов и средств информатики и искусственного интеллекта для решения прикладных задач различных классов. <p><i>Навыки:</i></p> <ul style="list-style-type: none"> – Применения аналитических инструментов и методов для исследования современного состояния и тенденций научно-технического развития информационного общества. – Обобщения результатов проведенного анализа и исследования. |

Знания, умения и навыки, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы:

- производственная практика, преддипломная практика;
- выполнение и защита выпускной квалификационной работы.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины направлено на формирование следующих компетенций в соответствии с образовательной программой:

Перечень планируемых результатов обучения по дисциплине, соотнесённых с индикаторами достижения компетенций

| Компетенция | Индикаторы достижения компетенции | Результаты обучения |
|--|--|---|
| ПК-1. Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях | ПК-1.1. Ставит задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области | <i>Знания:</i> <ul style="list-style-type: none">– Знает классы методов и алгоритмов машинного обучения.– Знает процедуры критического анализа, методики решения комплекса задач предметной области. <i>Умения:</i> <ul style="list-style-type: none">– Умеет ставить задачи и адаптировать методы и алгоритмы машинного обучения. <i>Владеть:</i> <ul style="list-style-type: none">– Владеет навыками постановки задач по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области.– Владеет навыками использования методов и алгоритмов машинного обучения. |

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоёмкость дисциплины составляет 6 зачётных единиц, 216 часов,
 Форма промежуточной аттестации: дифференцированный зачет

4.1. Содержание дисциплины, структурированное по темам

| № п/п | Темы дисциплины | Семестр | Виды учебной работы и их трудоёмкость, часы (в том числе с использованием онлайн-курсов) | | | | Наименования оценочных средств |
|--|---|---------|---|-------------------------|-------------------------|--------------------------------|--|
| | | | Контактная работа | | | Самостоя- тельная работа | |
| | | | Лекции | Практические занятия | Лабораторные занятия | | |
| Модуль 1. Информационный поиск | | | | | | | |
| 1 | Введение в информационный поиск | 3 | 2 | 4 | - | 18 | Практическая работа № 1 (собеседование по результатам выполнения практических работ) |
| 2 | Словари и нечёткий поиск | 3 | 2 | 4 | - | 18 | Практическая работа № 2 (собеседование по результатам выполнения практических работ) |
| 3 | Ранжирование, взвешивание терминов и модель векторного пространства | 3 | 2 | 4 | | 18 | Практическая работа № 3 (собеседование по результатам выполнения практических работ) |
| 4 | Оценка информационного поиска | 3 | 2 | 4 | | 18 | Индивидуальное проектное задание. Контрольная работа №1 |
| Модуль 2. Обработка естественного языка | | | | | | | |
| 5 | Введение в обработку естественно- языковых текстов | 3 | 2 | 4 | - | 18 | Практическая работа № 4 (собеседование по результатам выполнения практических работ) |
| 6 | Методы обработки естественных языков | 3 | 2 | 4 | - | 18 | Практическая работа № 5 (собеседование по результатам выполнения практических работ) |
| 7 | Вопросно-ответные системы | 3 | 2 | 8 | | 18 | Практическая работа № 6 (собеседование по результатам выполнения практических работ) |

| № п/п | Темы дисциплины | Семестр | Виды учебной работы и их трудоёмкость, часы (в том числе с использованием онлайн-курсов) | | | | Наименования оценочных средств |
|--|--|---------|---|-------------------------|-------------------------|--------------------------------|--|
| | | | Контактная работа | | | Самостоя- тельная работа | |
| | | | Лекции | Практические занятия | Лабораторные занятия | | |
| 8 | Программирование и проектирование систем обработки естественных языков | 3 | 4 | 4 | | 36 | Индивидуальное проектное задание. Контрольная работа №2 |
| Промежуточная аттестация (для дисциплин с экзаменом) | | 3 | – | – | – | – | – |
| Итого часов | | | 18 | 36 | – | 162 | – |

4.2. План внеаудиторной самостоятельной работы

| № п/п | Темы дисциплины | Семестр | Вид самостоятельной работы | Сроки выполнения (нед.) | Затраты времени (часы) | Учебно-методическое обеспечение |
|--|---|---------|---|-------------------------------|------------------------------|--|
| Модуль 1. Информационный поиск | | | | | | |
| 1 | Введение в информационный поиск | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 1-2 неделя | 18 | Основная [1-4] и дополнительная [5, 8] литература |
| 2 | Словари и нечёткий поиск | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 3-4 неделя | 18 | Основная [1-4] и дополнительная [5, 6, 9] литература |
| 3 | Ранжирование, взвешивание терминов и модель векторного пространства | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 5-6 неделя | 18 | Основная [1-4] и дополнительная [5, 6, 9] литература |
| 4 | Оценка информационного поиска | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 7-8 неделя | 18 | Основная [1-4] и дополнительная [5, 6, 9] литература |
| Модуль 2. Обработка естественного языка | | | | | | |
| 5 | Введение в обработку естественно-языковых текстов | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 9-10 неделя | 18 | Основная [1-4] и дополнительная [7] литература |

| № п/п | Темы дисциплины | Семестр | Вид самостоятельной работы | Сроки выполнения (нед.) | Затраты времени (часы) | Учебно-методическое обеспечение |
|--|--|---------|---|-------------------------|------------------------|--|
| 6 | Методы обработки естественных языков | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 11-12 неделя | 18 | Основная [1-4] и дополнительная [5-6] литература |
| 7 | Вопросно-ответные системы | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 13-14 неделя | 18 | Основная [1-4] и дополнительная [5, 6, 9] литература |
| 8 | Программирование и проектирование систем обработки естественных языков | 3 | Проработка конспектов лекций, работа с учебной литературой и подготовка к практическим занятиям | 15-18 неделя | 36 | Основная [1-4] и дополнительная [5-10] литература |
| Подготовка к экзамену (для дисциплин с экзаменом) | | | | | - | |
| Общая трудоёмкость самостоятельной работы по дисциплине | | | | | 162 | – |

4.3. Содержание учебного материала

Модуль 1. Информационный поиск

Тема 1. Введение в информационный поиск

Введение в информационный поиск: основные понятия, практическая значимость, задачи. Булев поиск: пример информационного поиска, первая попытка создать инвертированный индекс, обработка булевых запросов. Лексикон и список словопозиций: схематизация документа и декодирование последовательности символов, определение лексикона.

Тема 2. Словари и нечёткий поиск

Словари и нечёткий поиск: поисковые структуры для словарей, запросы с джокером, исправление опечаток, фонетические исправления. Построение индекса: блочное индексирование, основанное на сортировке, однопроходное индексирование в оперативной памяти. Сжатие индекса: статистические характеристики терминов в информационном поиске, сжатие словаря, сжатие инвертированного индекса.

Тема 3. Ранжирование, взвешивание терминов и модель векторного пространства

Ранжирование, взвешивание терминов и модель векторного пространства: параметрические и зонные индексы, частота термина и взвешивание. Ранжирование в полнофункциональной поисковой системе: эффективное ранжирование, компоненты информационно-поисковой системы.

Тема 4. Оценка информационного поиска

Оценка информационного поиска: оценка информационно-поисковой системы, стандартные текстовые коллекции, оценка неранжированных результатов поиска, оценка ранжированных результатов поиска, оценка релевантности, качество системы и её полезность для пользователя, снипеты. Обратная связь по релевантности и расширение запроса: обратная связь по релевантности и псевдорелевантность, глобальные методы для переформулирования запроса.

Модуль 2. Обработка естественного языка

Тема 5. Введение в обработку естественно-языковых текстов

Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика. Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» как фундамент для построения систем автоматической обработки текста.

Тема 6. Методы обработки естественных языков

Анализ и синтез текста. Морфологический и синтаксический анализ. Парсинг. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Интерактивное разрешение лексической и синтаксической неоднозначности. Правильные и статистические подходы к автоматической обработке текста. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы. Синтаксическая структура предложения.

Тема 7. Вопросно-ответные системы

Вопросно-ответные системы: основы вопросно-ответной системы, архитектуры вопросно-ответной системы, установление смысла вопроса и порождение ответов. Распознавание имён людей, географических названий и других сущностей, различные подходы к распознаванию именованных сущностей

Тема 8. Программирование и проектирование систем обработки естественных языков

Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация. Понятия лексемы, словоформы, леммы, морфемы, псевдо-основы и псевдо-окончания. Грамматические категории. Словоизменительная парадигма. Морфотактика. Структура данных морфологического словаря, лексикона. Грамматические модели русского языка в контексте автоматической обработки. Минимальное расстояние редактирования. Алгоритм подсчета расстояния Левенштейна. Практика по подсчету минимального расстояния Левенштейна. Понятие статистической языковой модели. Области применения. N-граммы.

Перечень тем практических занятий

| № п/п | Тема практического занятия | Количество часов |
|--------------------|--|---------------------|
| Модуль 1. | | |
| 1 | Морфологический анализ | 6 |
| 2 | Автоматическое построение рефератов текстовых документов | 6 |
| 3 | Анализ тональности | 6 |
| Модуль 2. | | |
| 5 | Извлечение информации | 6 |
| 6 | Машинный перевод | 6 |
| 7 | Программирование и проектирование систем обработки естественных языков | 6 |
| Всего часов | | 36 |

У. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

По дисциплине предусмотрены следующие методы обучения и интерактивные формы проведения занятий:

- визуализации учебного материала (презентации лекционного материала доступны в системе электронного обучения);
- дискуссионные (обсуждение новых информационных технологий);
- групповой работы (работа в малых группах на практических занятиях при проведении поиска информационных источников и выявлении научных трендов);

Наряду с традиционными образовательными технологиями, для реализации дисциплины могут использоваться технологии электронного обучения и дистанционные образовательные технологии в электронной информационно-образовательной среде университета. Лекционные занятия и другие формы контактной работы обучающихся с преподавателем могут проводиться с использованием платформ Microsoft Teams, Cisco, Moodle (BigBlueButton) и др., что позволяет обеспечить онлайн и офлайн взаимодействие преподавателя с обучающимися в рамках дисциплины.

Основными методами текущего контроля являются электронный учёт и контроль учебных достижений студентов (использование средств сервиса балльно-рейтинговой системы; ведение электронного журнала успеваемости, проведение электронного тестирования и применение других средств контроля с использованием системы электронного обучения).

У. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Основная литература

1. Винокурский Д. Л. Инструментальные средства информационных систем: курс лекций: учебное пособие / Д.Л. Винокурский, Б.В. Крахоткина - Ставрополь: СКФУ, 2018. - 165 с. <http://biblioclub.ru/index.php?page=book&id=562702>

2. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python: учебное пособие / В.М. Волкова, М.А. Семенова, Е.С. Четвертакова, С.С. Вожов; Новосибирский государственный технический университет - Новосибирск: Новосибирский государственный технический университет, 2017. - 74 с. <http://biblioclub.ru/index.php?page=book&id=576496>

3. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>

4. Кокорина, И.В. Основы математической обработки информации в филологии: комбинаторика, теория вероятностей и математическая статистика : учебно-методическое пособие / И.В. Кокорина ; Министерство образования и науки Российской Федерации, Федеральное государственное автономное образовательное учреждение высшего профессионального образования Северный (Арктический) федеральный университет им. М.В. Ломоносова. - Архангельск : ИД САФУ, 2014. - 115 с. : ил. - Библиогр. в кн. - ISBN 978-5-261-00928-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=312317>

6.2. Дополнительная литература

5. Ромм Я. Е. Детерминированный информационный поиск на основе сортировки с распараллеливанием базовых операций / Я.Е. Ромм; С.С. Белоконова - М.: Издательство Научный мир, 2014. - 197 с. <http://biblioclub.ru/index.php?page=book&id=468725>

6. Миркин Б. Г. Введение в анализ данных. Учебник и практикум - М.: Издательство Юрайт, 2020. - 174 с. <https://urait.ru/book/vvedenie-v-analiz-dannyh-450262>

7. Информационный поиск и обработка естественного языка: методические указания по выполнению курсового проекта для студентов всех форм обучения направления 09.04.03 Прикладная информатика / Сост.: Е.А. Симоненко; Кубан. гос. технол. ун-т; кафедра информационных систем и программирования. – Краснодар: КубГТУ, 2019. – 13 с. Алексеев, Г.В. Численное экономико-математическое моделирование и оптимизация [Электронный ресурс]: учебное пособие / Г.В. Алексеев, И.И. Холявин. - Саратов: Вузовское образование, 2019. - 195 с. - ЭБС «IPRbooks» - Режим доступа: <http://www.iprbookshop.ru/79692.html>

8. Костюкова, Н.И. Комбинаторные алгоритмы для программистов / Н.И. Костюкова. - 2-е изд./, исправ./ - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 217 с. : ил. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=429067>

9. Гусякова, А.В. Информационные технологии и лингвистика XXI века : учебное пособие / А.В. Гусякова ; Министерство образования и науки Российской Федерации. - Москва : МПГУ, 2016. - 96 с. : ил. - Библиогр. в кн. - ISBN 978-5-4263-0398-0 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=469675>

6.3. Периодические издания

- [IEEE Spectrum](https://spectrum.ieee.org/) <https://spectrum.ieee.org/>
- Intelligent Enterprise/RE (журнал «Корпоративные системы») <https://www.iemag.ru/about/>
- BYTE Россия <https://www.bytemag.ru/about/>

6.4. Перечень ресурсов сети Интернет

- Образовательная платформа Юрайт <https://urait.ru/>
- ЭБС «Университетская библиотека онлайн» <http://biblioclub.ru>.
- IBM Academic Initiative http://ictis.sfedu.ru/ibm_academic_initiative/ (учебные материалы)
- НОУ «Интуит» <https://www.intuit.ru/>
- Образовательная платформа Юрайт <https://urait.ru/>
- ЭБС «Университетская библиотека онлайн» <http://biblioclub.ru>.
- Znanium.com. Базовая коллекция: электронно-библиотечная система: сайт / ООО "Научно-издательский центр Инфра-М". – Москва. – URL: <http://znanium.com/catalog>. – Режим доступа: для зарегистрир. пользователей. – Текст: электронный.
- IPRBooks. Базовая коллекция: электронно-библиотечная система: сайт / Общество с ограниченной ответственностью Компания "Ай Пи Ар Медиа". – Саратов. – URL: <http://www.iprbookshop.ru/586.html> – Режим доступа: для зарегистрир. пользователей. – Текст электронный.
- Национальная электронная библиотека (НЭБ): федеральная государственная информационная система: сайт / Министерство культуры Российской Федерации, Российская государственная библиотека. – Москва. – URL: <https://нэб.рф/>. – Режим доступа: для зарегистрир. пользователей. – Текст: электронный.
- Электронная библиотека: библиотека диссертаций: сайт / Российская государственная библиотека. – Москва: РГБ. – URL: <http://diss.rsl.ru/?lang=ru>. – Режим доступа: для зарегистрир. пользователей. – Текст: электронный.
- eLIBRARY.RU: научная электронная библиотека: сайт. – Москва. – URL: <https://elibrary.ru/defaultx.asp>. – Режим доступа: для зарегистрир. пользователей. – Текст: электронный.
- CYBERLENINKA: научная электронная библиотека: сайт. – Москва. – URL: <https://cyberleninka.ru/> – Режим доступа: для зарегистрир. пользователей. – Текст: электронный.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

При реализации дисциплины используются следующие помещения, оборудование и программное обеспечение:

Лаборатория машинного обучения и технологий больших данных

Персональные компьютеры (8 шт.), проектор, экран. Windows 10, Microsoft Office 365, Adobe Acrobat Reader (Бесплатное проприетарное ПО, <https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader/volume-distribution.html>), Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Foxit (Бесплатное проприетарное ПО, <https://www.foxitsoftware.com/ru/>), i2 Analyst's Notebook (Бесплатная лицензия для образовательных целей), <https://developer.ibm.com/academic/>, Notepad++, Бесплатное ПО (GNU GPL 2), <https://notepad-plus-plus.org/>, Total Commander 7.x, WinRAR, XAMPP, Бесплатное ПО (GNU GPL), <http://www.apachefriends.org/en/xampp.html>, Team Foundation Server 2015, Visual Studio 2015, Android Studio, Операционная система на базе Linux; Офисный пакет Open Office, актуальные версии браузеров Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Edge, Safari с поддержкой протокола WebRTC, PyCharm 2017.1.2 <https://www.jetbrains.com/pycharm/> Свободное ПО, <https://www.python.org/>, Evolus Pencil, Свободное ПО (GNU GPL 2), <https://pencil.evolus.vn/>

VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Курс ИПиОЯ направлен на развитие профессиональных компетенций, служащих основой готовности магистрантов к профессиональной деятельности.

Самостоятельная работа магистрантов предполагает изучение теоретического материала (проработку конспектов лекций, работу с учебной литературой и информационно-образовательными ресурсами), подготовку к практическим занятиям, выполнение практических заданий, написание отчета по выполненному проекту.

Для успешного освоения дисциплины необходимы:

- открытость познавательной позиции как тип познавательного отношения к миру;
- особое отношение к парадоксам и противоречиям, вариативность субъективных способов восприятия и осмысления событий;
- направленность на переоценку собственного опыта, обнаружение противоречий между усвоенными знаниями и поступающей информацией, гибкости в построении собственной познавательной деятельности и др.

Теоретический материал следует изучать последовательно, в соответствии с приведенным содержанием курса и содержанием основной литературы. Приступать к выполнению заданий рекомендуется после того, как усвоены основные понятия и базовые идеи соответствующего раздела. Для своевременной помощи в выполнении заданий преподаватель проводит еженедельные консультации.

Если учебные занятия проводятся с использованием ЭО и ДОТ, то при их организации и проведении необходимо руководствоваться соответствующими методическими рекомендациями и инструкциями по работе в ЭИОС университета.

IX. УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ

Курс 2, семестр 3, очная форма обучения

| № п/п | Виды контрольных мероприятий (наименования оценочных средств) | Количество баллов | |
|---|--|---|-------------------|
| | | Текущий контроль | Рубежный контроль |
| Модуль 1. | | | |
| 1 | Практические работы №№ 1–3 (собеседование по результатам выполнения практических работ) | 30 (3 работы × 10 баллов) | – |
| 2 | Контрольная работа | – | 10 |
| Модуль 2. | | | |
| 3 | Практические работы №№ 4–6 (собеседование по результатам выполнения практических работ) | 30 (3 работы × 10 баллов) | – |
| 4 | Индивидуальное проектное задание | | 20 |
| 5 | Контрольная работа | – | 10 |
| Всего | | 60 | 40 |
| Бонусные баллы | | Не предусмотрены | |
| Промежуточная аттестация в форме диф. зачета | | Оценка по дисциплине выставляется по сумме баллов за текущий контроль, рубежный контроль: <ul style="list-style-type: none"> – 85–100 баллов – оценка «отлично»; – 71–84 балла – оценка «хорошо»; – 60–70 баллов – оценка «удовлетворительно»; – менее 60 баллов – оценка «неудовлетворительно» | |

Х. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

10.1. Паспорт фонда оценочных средств

| № п/п | Индикатор достижения компетенции | Наименование оценочного средства |
|-------|--|---|
| 1 | ПК-1.1. Ставит задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области | – практические работы № 1-6 (собеседование по результатам выполнения практических работ); – индивидуальное проектное задание; – контрольные работы. |

10.2. Практические работы №№ 1–6 (выполнение, подготовка отчёта, защита отчёта)

Темы практических занятий:

1. Морфологический анализ
2. Автоматическое построение рефератов текстовых документов
3. Анализ тональности
4. Извлечение информации
5. Машинный перевод
6. Программирование и проектирование систем обработки естественных языков

Методические рекомендации по выполнению практических (семинарских) занятий

Практические работы выполняются после освоения соответствующего теоретического материала. Работы выполняются индивидуально как на учебном занятии, так и во время самостоятельной работы. После выполнения работы, полученные результаты оформляются в виде отчета. Каждый отчет должен включать титульный лист с наименованием работы, ФИО и группой студента, краткую теоретическую справку о выполняемом в работе задании, скриншоты среды моделирования с процессом и результатом работы.

Критерии оценки:

Всего за практические занятия студент может набрать 60 баллов ($6 \times 10 = 60$)

- 9-10 баллов выставляется студенту, если все требования, предъявляемые к заданию, выполнены, работа подготовлена и представлена в срок, студент продемонстрировал в процессе защиты работы и участия в обсуждении других работ требуемые качества;
- 7-8 балла выставляется студенту, если все требования, предъявляемые к заданию, выполнены, но есть существенные замечания по ряду характеристик выполнения и/или защиты работы;
- 5-6 балла выставляется студенту, если большинство требований, предъявляемых к заданию, выполнены, но студент не защитил работу в срок или не продемонстрировал в процессе защиты работы и участия в обсуждении других работ большинства требуемых качеств;
- работа не зачтена (0 баллов), если разработанное задание репродуктивного уровня, студент демонстрирует недостаточные знания по теоретическим аспектам работы, требования к работе выполнены частично. Небрежно оформленные иллюстрации, грамматические ошибки в отчете.

10.3. Задания для контрольной работы

Модуль 1. Информационный поиск. Контрольная работа №1

1. Основные понятия информационного поиска
2. Булев поиск

3. Лексикон и список словопозиций
4. Словари и нечёткий поиск: поисковые структуры для словарей, запросы с джокером, исправление опечаток, фонетические исправления.
5. Построение индекса: блочное индексирование, основанное на сортировке, однопроходное индексирование в оперативной памяти.
6. Сжатие индекса: статистические характеристики терминов в информационном поиске, сжатие словаря, сжатие инвертированного индекса.
7. Ранжирование, взвешивание терминов и модель векторного пространства: параметрические и зонные индексы, частота термина и взвешивание.
8. Ранжирование в полнофункциональной поисковой системе: эффективное ранжирование, компоненты информационно-поисковой системы.
9. Оценка информационного поиска: оценка информационно-поисковой системы, стандартные текстовые коллекции, оценка неранжированных результатов поиска, оценка ранжированных результатов поиска, оценка релевантности, качество системы и её полезность для пользователя, снипеты.
10. Обратная связь по релевантности и расширение запроса: обратная связь по релевантности и псевдорелевантность, глобальные методы для переформулирования запроса.

Модуль 2. Обработка естественного языка. Контрольная работа №2

11. Лингвистика как наука о языке.
12. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика.
13. Лингвистика и прагматика.
14. Лингвистическое моделирование.
15. Действующие модели языка.
16. Теория «Смысл – Текст» как фундамент для построения систем автоматической обработки текста.
17. Анализ и синтез текста.
18. Морфологический и синтаксический анализ.
19. Парсинг.
20. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх».
21. Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста.
22. Интерактивное разрешение лексической и синтаксической неоднозначности.
23. Правильные и статистические подходы к автоматической обработке текста.
24. Алгоритм синтаксического анализа.
25. Синтаксические отношения.
26. Синтагмы.
27. Синтаксическая структура предложения.
28. Вопросно-ответные системы: основы вопросно-ответной системы, архитектуры вопросно-ответной системы, установление смысла вопроса и порождение ответов.
29. Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация.
30. Понятия лексемы, словоформы, леммы, морфемы, псевдо-основы и псевдо-окончания.
31. Грамматические категории.
32. Словоизменительная парадигма.
33. Морфотактика.
34. Структура данных морфологического словаря, лексикона.
35. Грамматические модели русского языка в контексте автоматической обработки.
Минимальное расстояние редактирования.
36. Алгоритм подсчета расстояния Левенштейна.
37. Практика по подсчету минимального расстояния Левенштейна.

38. Понятие статистической языковой модели.
39. Области применения. N-граммы.
40. Основные особенности использования методов теории игр.

Требования к контрольной работе

Контрольная работа представляет собой один из видов самостоятельной работы обучающихся. По сути – это изложение ответов на определенные теоретические вопросы по учебной дисциплине, а также решение практических задач. Контрольные проводятся для того, чтобы развить у обучающихся способности к анализу научной и учебной литературы, умение обобщать, систематизировать и оценивать практический и научный материал, укреплять навыки овладения понятиями определенной науки и т. д.

При оценке контрольной преподаватель руководствуется следующими критериями:

- работа была выполнена автором самостоятельно;
- обучающийся подобрал достаточный список литературы, который необходим для осмысления темы контрольной;
- автор сумел составить логически обоснованный план, который соответствует поставленным задачам и сформулированной цели;
- обучающийся проанализировал материал;
- контрольная работа отвечает всем требованиям четкости изложения и аргументированности, объективности и логичности, грамотности и корректности;
- обучающийся сумел обосновать свою точку зрения;
- контрольная работа оформлена в соответствии с требованиями;
- автор защитил контрольную и успешно ответил на все вопросы преподавателя.

Контрольная работа, выполненная небрежно, не по своему варианту, без соблюдения правил, предъявляемых к ее оформлению, возвращается без проверки с указанием причин, которые доводятся до обучающегося. В этом случае контрольная работа выполняется повторно.

Вариант контрольной работы выдается в соответствии с порядковым номером в списке магистрантов.

Критерии оценки знаний при написании контрольной работы

Отметка «отлично» выставляется обучающемуся, показавшему всесторонние, систематизированные, глубокие знания вопросов контрольной работы и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Отметка «хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности, которые может устранить с помощью дополнительных вопросов преподавателя.

Отметка «удовлетворительно» выставляется обучающемуся, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными понятиями выносимых на контрольную работу тем, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

Отметка «неудовлетворительно» выставляется обучающемуся, который не знает большей части основного содержания выносимых на контрольную работу вопросов тем дисциплины, допускает грубые ошибки в формулировках основных понятий и не умеет использовать полученные знания.

Требования к выполнению тестового задания

Тестирование является одним из основных средств формального контроля качества обучения. Это метод, основанный на стандартизированных заданиях, которые позволяют измерить психофизиологические и личностные характеристики, а также знания, умения и навыки испытуемого.

Основные принципы тестирования, следующие:

- связь с целями обучения - цели тестирования должны отвечать критериям социальной полезности и значимости, научной корректности и общественной поддержки;
- объективность - использование в педагогических измерениях этого принципа призвано не допустить субъективизма и предвзятости в процессе этих измерений;
- справедливость и гласность - одинаково доброжелательное отношение ко всем обучающимся, открытость всех этапов процесса измерений, своевременность ознакомления обучающихся с результатами измерений;

– систематичность – систематичность тестирований и самопроверок каждого учебного модуля, раздела и каждой темы; важным аспектом данного принципа является требование репрезентативного представления содержания учебного курса в содержании теста; гуманность и этичность - тестовые задания и процедура тестирования должны исключать нанесение какого-либо вреда обучающимся, не допускать ущемления их по национальному, этническому, материальному, расовому, территориальному, культурному и другим признакам;

Важнейшим является принцип, в соответствии с которым тесты должны быть построены по методике, обеспечивающей выполнение требований соответствующего федерального государственного образовательного стандарта.

В тестовых заданиях используются четыре типа вопросов:

- закрытая форма - является наиболее распространенной и предлагает несколько альтернативных ответов на поставленный вопрос. Например, обучающемуся задается вопрос, требующий альтернативного ответа «да» или «нет», «является» или «не является»,

«относится» или «не относится» и т.п. Тестовое задание, содержащее вопрос в закрытой форме, включает в себя один или несколько правильных ответов и иногда называется выборочным заданием. Закрытая форма вопросов используется также в тестах-задачах с выборочными ответами. В тестовом задании в этом случае сформулированы условие задачи и все необходимые исходные данные, а в ответах представлены несколько вариантов результата решения в числовом или буквенном виде. Обучающийся должен решить задачу и показать, какой из представленных ответов он получил.

- открытая форма - вопрос в открытой форме представляет собой утверждение, которое необходимо дополнить. Данная форма может быть представлена в тестовом задании, например, в виде словесного текста, формулы (уравнения), графика, в которых пропущены существенные составляющие - части слова или буквы, условные обозначения, линии или изображения элементов схемы и графика. Обучающийся должен по памяти вставить соответствующие элементы в указанные места («пропуски»).

- установление соответствия - в данном случае обучающемуся предлагают два списка, между элементами которых следует установить соответствие;

- установление последовательности - предполагает необходимость установить правильную последовательность предлагаемого списка слов или фраз.

Критерии оценки знаний при проведении тестирования

34-40 баллов - выставляется студенту при условии правильного ответа не менее чем 85% тестовых заданий;

28-33 балла - выставляется студенту при условии правильного ответа не менее чем 70% тестовых заданий;

22-32 балла - выставляется студенту при условии правильного ответа не менее 50 %;

менее 22 балла - выставляется студенту при условии правильного ответа менее чем на 50 % тестовых заданий.

Результаты текущего контроля используются при проведении промежуточной аттестации.

Примеры тестовых заданий

1. На технологическом уровне обработка естественного языка может быть реализована с помощью существующих библиотек и пакетов программ с открытым исходным кодом. Назовите верные примеры

- Apertium
- GATE (General Architecture for Text Engineering)

- Learning Based Java
- LinguaStream
- LingPipe
- Mallet
- MontyLingua
- Apache OpenNLP
- UIMA
- Stanford Parser
- Caffee/Caffee2
- CNTK
- TensorFlow
- Theano
- Torch/Pytorch

2. В методах NLP задача классификация текста имеет следующие цели:

- обнаружение спама
- установление авторства
- анализ эмоциональной окраски текста
- определение возраста и пола автора
- определение главной темы документа
- идентификация языка

3. Как называется класс методов контент-анализа в компьютерной лингвистике, предназначенный для автоматизированного выявления в текстах эмоционально окрашенной лексики и эмоциональной оценки авторов (мнений) по отношению к объектам, речь о которых идёт в тексте? Укажите все правильные варианты, включая синонимы термина

- Анализ тональности текста
- Определение мнения/отношения
- Анализ субъективности текста
- Контент-анализ
- Машинное обучение с учителем
- Анализ опорных векторов

10.4. Индивидуальное проектное задание

Для выполнения задания студент выбирает одну из предложенных в рабочей программе тем, либо согласует с преподавателем свою.

Темы для выполнения индивидуального проектного задания:

- Проектирование информационной системы с вопросно-ответной подсистемой.
- Проектирование информационной системы с рекомендательной подсистемой.
- Разработка информационно-поисковой системы на языке Haskell.
- Разработка информационно-поисковой системы на языке Clojure.
- Разработка информационно-поисковой системы на языке R.
- Разработка графовой базы данных с системой нечёткого поиска.
- Разработка приложения для определения эмоциональной окраски диалога в социальной сети.
- Разработка приложения для определения темы диалога в социальной сети.

- Разработка приложения для определения тематики сообщества в социальной сети.
- Разработка приложения для определения тематики интересов пользователя по его участию в сообществах социальной сети.

В работе должны быть продемонстрированы:

- владение терминологией дисциплины;
- соблюдение рекомендованной структуры разделов и содержания плана
- умение структурировать информацию.

Баллы за задание снижаются на 1-3 балла за каждую ошибку в зависимости от критичности для корректности и реализуемости плана тестирования.

Критерии итоговой оценки проектного задания:

- 19-20 баллов выставляется студенту, если все требования, предъявляемые к заданию, выполнены, работа подготовлена и представлена в срок, студент продемонстрировал в процессе защиты и участия в обсуждении других работ требуемые качества;
- 14-18 баллов выставляется студенту, если все требования, предъявляемые к заданию, выполнены, но есть существенные замечания по ряду характеристик выполнения и/или защиты;
- 12-13 баллов выставляется студенту, если большинство требований, предъявляемых к заданию, выполнены, но студент не защитил работу в срок или не продемонстрировал в процессе защиты и участия в обсуждении других работ большинства требуемых качеств;
- 0 баллов выставляется студенту, если разработанное задание репродуктивного уровня, студент демонстрирует поверхностные представления об основных терминах дисциплины, задание выполнено менее чем на 50 %.