

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ В.Ф. УТКИНА»**

Кафедра «Вычислительная и прикладная математика»

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ
«Машинная обработка текстов на ЕЯ»**

Направление подготовки
09.03.04 «Программная инженерия»

Направленность (профиль) подготовки
«Программное обеспечение систем искусственного интеллекта»

Уровень подготовки – бакалавриат

Квалификация выпускника – бакалавр

Форма обучения – очная

Срок обучения – 4 года

Рязань

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов и процедур, предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретенных компетенций, обучающихся целям и требованиям основной образовательной программы в ходе проведения текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций и индикаторов их достижения, приобретаемых обучающимся в соответствии с этими требованиями.

Контроль знаний обучающихся проводится в форме текущего контроля и промежуточной аттестации.

Текущий контроль успеваемости и промежуточная аттестация проводятся с целью определения степени усвоения учебного материала, своевременного выявления и устранения недостатков в подготовке обучающихся, организации работы обучающихся в ходе учебных занятий и оказания им индивидуальной помощи.

К контролю текущей успеваемости относятся проверка знаний, умений и навыков обучающихся на практических занятиях по результатам выполнения и защиты обучающимися индивидуальных заданий, по результатам выполнения контрольных работ и тестов, по результатам проверки качества конспектов лекций и иных материалов.

В качестве оценочных средств на протяжении семестра используется устные и письменные ответы студентов на индивидуальные вопросы, письменное тестирование по теоретическим разделам курса, реферат. Дополнительным средством оценки знаний и умений студентов является отчет о выполнении практических заданий и его защита.

По итогам курса обучающиеся сдают экзамен. Форма проведения – устный ответ с письменным подкреплением по утвержденным билетам, сформулированным с учетом содержания дисциплины. В билет для экзамена включается два теоретических вопроса и задача. В процессе подготовки к устному ответу студент должен составить в письменном виде план ответа.

1. Перечень компетенций с указанием этапов их формирования

При освоении дисциплины формируются следующие компетенции: ПК-8 (индикаторы ПК-8.1, ПК-8.2), ПК-10 (индикаторы ПК-10.2, ПК-10.4).

Указанные компетенции формируются в соответствии со следующими этапами:

- формирование и развитие теоретических знаний, предусмотренных указанными компетенциями (лекционные занятия, самостоятельная работа студентов);
- приобретение и развитие практических умений, предусмотренных компетенциями (практические занятия, самостоятельная работа студентов);
- закрепление теоретических знаний, умений и практических навыков, предусмотренных компетенциями, в ходе решения конкретных задач на занятиях, выполнения индивидуальных заданий на практических занятиях и их защиты, а также в процессе сдачи экзамена.

2 Показатели и критерии оценивания компетенций (результатов) на различных этапах их формирования, описание шкал оценивания

Сформированность каждой компетенции в рамках освоения данной дисциплины оценивается по трехуровневой шкале:

- пороговый уровень является обязательным для всех обучающихся по завершении освоения дисциплины;

- продвинутый уровень характеризуется превышением минимальных характеристик сформированности компетенций по завершении освоения дисциплины;
- эталонный уровень характеризуется максимально возможной выраженностью компетенций и является важным качественным ориентиром для самосовершенствования.

При достаточном качестве освоения более 80% приведенных знаний, умений и навыков преподаватель оценивает освоение данной компетенции в рамках настоящей дисциплины на эталонном уровне, при освоении более 60% приведенных знаний, умений и навыков – на продвинутом, при освоении более 40% приведенных знаний, умений и навыков – на пороговом уровне. При освоении менее 40% приведенных знаний, умений и навыков компетенция в рамках настоящей дисциплины считается неосвоенной.

Уровень сформированности каждой компетенции на различных этапах ее формирования в процессе освоения данной дисциплины оценивается в ходе текущего контроля успеваемости и представлено различными видами оценочных средств.

Оценке сформированности в рамках данной дисциплины подлежат компетенции/индикаторы:

Показатели достижения планируемых результатов обучения и критерии их оценивания на разных уровнях формирования компетенций приведены в таблице 1.

Таблица 1. Показатели достижения индикаторов компетенции

1	2	3	4
Компетенция: код по ФГОС 3++, формулировка	Индикаторы	Этап	Наименование оценочного средства
ПК-8 (09.03.04/02 Программное обеспечение систем искусственного интеллекта) Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК-8.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях <u>Знать</u>: методы поиска данных. <u>Уметь</u>: отделять достоверные источники данных от сомнительных, осуществлять критический отбор данных, проверять их на целостность и непротиворечивость. <u>Владеть</u>: методологией поиска данных из разных источников. ПК-8.2 Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения <u>Знать</u>: методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных; методы планирования вычислительного экспе-	1	Экзамен

1	2	3	4
	римента, формирования обучающей и контрольной выборок. <u>Уметь</u>: выявлять и исключать из массива данных ошибочные данные и выбросы; выделять входные и выходные переменные с целью использования предиктивных моделей. <u>Владеть</u>: основами подготовки структурированных и неструктурированных данных для машинного обучения.		
ПК-10 (09.03.04/02 Программное обеспечение систем искусственного интеллекта) Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий ИИ	ПК-10.2 Участвует в реализации проектов в области сквозной цифровой субтехнологии "Обработка естественного языка" <u>Знать</u>: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка». <u>Уметь</u>: решать задачи по выполнению коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка». <u>Владеть</u>: методологией сквозной цифровой субтехнологии «Обработка естественного языка».	1	Экзамен
	ПК-10.4 Участвует в реализации проектов в области сквозной цифровой субтехнологии «Распознавание и синтез речи» <u>Знать</u>: принципы построения систем распознавания и синтеза речи, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии	1	Экзамен

1	2	3	4
	«Распознавание и синтез речи»». <u>Уметь</u>: решать задачи по выполнению коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Распознавание и синтез речи»». <u>Владеть</u>: методологией сквозной цифровой субтехнологии «Распознавание и синтез речи».		

Преподавателем оценивается содержательная сторона и качество материалов, приведенных в отчетах студента по практическим занятиям. Кроме того, преподавателем учитываются ответы студента на вопросы по соответствующим видам занятий при текущем контроле:

- контрольные опросы;
- задания для практических занятий.

Принимается во внимание **знания** обучающимися:

- подходов в обработке естественного языка;
- графематический, морфологический, синтаксический, семантический этапов анализа текстов;
- особенностей текстов, таких как омонимия и сочетаемость слов;
- грамматики Хомского в синтаксическом этапе обработки текста.

наличие **умений**:

- обрабатывать текст на ЕЯ;
- выполнять синтаксический анализ текста с использованием лингвистического взгляда и грамматик Хомского;
- анализировать текст методом "мешка слов" и использованием N-грамм.

обладание навыками:

- обработки и анализа текстов на ЕЯ;
- работы с различными этапами обработки текста;
- работы с методами TF-IDF, "мешок слов".

Критерии оценивания уровня сформированности компетенции в процессе выполнения практических работ:

41%-60% правильных ответов соответствует пороговому уровню сформированности компетенции на данном этапе ее формирования;

61%-80% правильных ответов соответствует продвинутому уровню сформированности компетенции на данном этапе ее формирования;

81%-100% правильных ответов соответствует эталонному уровню сформированности компетенции на данном этапе ее формирования.

Сформированность уровня компетенций не ниже порогового является основанием для допуска обучающегося к промежуточной аттестации по данной дисциплине.

Формой промежуточной аттестации по данной дисциплине является экзамен, оцениваемый по принятой в ФГБОУ ВО «РГРТУ» четырехбалльной системе: «неудовлетворительно», «удовлетворительно», «хорошо» и «отлично».

Критерии оценивания промежуточной аттестации представлены в таблице.

Шкала оценивания	Критерии оценивания
«отлично»	студент должен: продемонстрировать глубокое и прочное усвоение знаний материала; исчерпывающе, последовательно, грамотно и логически стройно изложить теоретический материал; правильно формулировать определения; уметь сделать выводы по излагаемому материалу; безупречно ответить не только на вопросы билета, но и на дополнительные вопросы в рамках рабочей программы дисциплины; выполнить все практические задания, предусмотренные программой
«хорошо»	студент должен: продемонстрировать достаточно полное знание материала; продемонстрировать знание основных теоретических понятий; достаточно последовательно, грамотно и логически стройно излагать материал; уметь сделать достаточно обоснованные выводы по излагаемому материалу; ответить на все вопросы билета; выполнить все практические задания, предусмотренные программой.
«удовлетворительно»	студент должен: продемонстрировать общее знание изучаемого материала; знать основную рекомендуемую программой дисциплины учебную литературу; уметь строить ответ в соответствии со структурой излагаемого вопроса; показать общее владение понятийным аппаратом дисциплины; уметь устранить допущенные погрешности в ответе на теоретические вопросы; выполнить все практические задания, предусмотренные программой.
«неудовлетворительно»	ставится в случае: невыполнения практических занятий; незнания значительной части пройденного материала; не владения понятийным аппаратом дисциплины; существенных ошибок при изложении учебного материала; неумения строить ответ в соответствии со структурой излагаемого вопроса; неумения делать выводы по излагаемому материалу. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение по образовательной программе без дополнительных занятий по соответствующей дисциплине (формирования и развития компетенций, закрепленных за данной дисциплиной). Оценка «неудовлетворительно» выставляется также, если студент после начала экзамена отказался его сдавать или нарушил правила сдачи экзамена (списывал, подсказывал, обманом пытался получить более высокую оценку и т.д.).

3. Типовые контрольные задания или иные материалы

ФОС по дисциплине содержит следующие оценочные средства, позволяющие оценить знания, умения и уровень приобретенных компетенций при текущем контроле и промежуточной аттестации, разбитые по модулям дисциплины:

- перечни экзаменационных вопросов;
- макеты билетов к экзамену.

Средства для оценки различных уровней формирования компетенций по категориям «знать», «уметь», «владеть» обеспечивают реализацию основных принципов контроля, таких, как объективность и независимость, практико-ориентированность, междисциплинарность.

С учетом этого, контрольные вопросы (задания, задачи,) входящие в ФОС, для различных категорий и уровней освоения компетенций имеют следующий вид:

Уровень ЗНАТЬ

Дескрипторы	Пример задания из оценочного средства
виды представления данных, методы поиска и парсинга данных	1. Опишите этапы обработки текстов на ЕЯ, особенности каждого. Приведите примеры используемых инструментов.
уровни представления данных (ODS, DDL, семантический слой, модель данных)	1. Опишите этапы обработки текстов на ЕЯ, особенности каждого. Приведите примеры используемых инструментов.
основные инструменты, библиотеки и технологии Data Science	1. Опишите этапы обработки текстов на ЕЯ, особенности каждого. Приведите примеры используемых инструментов.
методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и не-структурированных данных	1. Вектор документа. Применение. Основные подходы к формированию. Приведите пример.
методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок	1. Вектор документа. Применение. Основные подходы к формированию. Приведите пример. 2. Корпуса текстов, назначение. Назовите особенности создания, приведя примеры использования.
принципы построения систем обработки ЕЯ, методы и технологии ИИ для анализа ЕЯ, методы и подходы к планированию и реализации проектов по созданию СИИ на основе сквозной субтехнологии "Обработка естественного языка"	1. Опишите этапы обработки текстов на ЕЯ, особенности каждого. Приведите примеры используемых инструментов.

Уровень УМЕТЬ

Дескрипторы	Пример задания из оценочного средства
отделять достоверные источники данных от сомнительных, осуществлять критических отбор данных, проверять их на целостность и непротиворечивость	1. Лингвистические онтологии и тезаурусы. Что изучает онтология? Что такое тезаурус? Применение. 2. Грамматики Хомского, особенности.
использовать инструменты и библиотеки для DataScience для поиска данных в открытых источниках, специализированных библиотеках и ре-	1. Назовите способы выделения терминов из текстов, приведя алгоритмы. Опишите особенности каждого.

позиториях	
выявлять и исключать из массива данных ошибочные данные и выбросы	1. Лингвистические онтологии и тезаурусы. Что изучает онтология? Что такое тезаурус? Применение.
выявлять входные и выходные переменные с целью использования предиктивных моделей	1. Опишите существующие архитектуры для нейросетевых моделей ЕЯ. Приведите характерные особенности для каждой из них.
осуществлять разметку структурированных и неструктурированных данных	1. Отличительные особенности текстов на русском языке. Приведите пример.
использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения	1. Опишите алгоритмы получения краткого содержания текстов на ЕЯ. 2. Алгоритмы генерирования текстов. Опишите существующие модели и приведите примеры возможного их использования, учитывая особенности.
использовать методы и технологии массово параллельной обработки и анализа данных	1. Опишите существующие архитектуры для нейросетевых моделей ЕЯ. Приведите характерные особенности для каждой из них.
применять методы и подходы к планированию и реализации проектов по созданию систем ИИ на основе сквозной цифровой субтехнологии "Обработка ЕЯ"	1. Чат-боты, назначение, примеры использования. Опишите алгоритм взаимодействия составляющих элементов, этапы обработки.

Перечень вопросов к экзамену и макет экзаменационного билета

1. Этапы обработки текстов на ЕЯ и их особенности.
2. Основные подходы к формированию вектора документа.
3. Способы выделения терминов из текстов
4. Корпуса текстов, примеры, назначение, особенности создания
5. Лингвистические онтологии и тезаурусы
6. Грамматики Хомского
7. Достоинства и недостатки аналитического и нейросетевого подходов к анализу текстов на ЕЯ
8. Существующие архитектуры для нейросетевых моделей ЕЯ
9. Тест Тьюринга сильные и слабые стороны
10. Чат-боты, назначение, примеры использования
11. Алгоритмы получения краткого содержания текстов на ЕЯ
12. Отличительные особенности текстов на русском языке
13. Алгоритмы генерирования текстов
14. Задача Коши для ОДУ. Метод Рунге - Кутты 2-го порядка точности. Оценка точности.

Макет экзаменационного билета

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования «Рязанский государственный радиотехнический университет»
(РГРТУ)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

по дисциплине «**Машинная обработка текстов на ЕЯ**»

1. Этапы обработки текстов на ЕЯ и их особенности.
2. Достоинства и недостатки аналитического и нейросетевого подходов к анализу текстов на ЕЯ.
3. Алгоритмы генерирования текстов

Билет рассмотрен и утвержден на заседании кафедры «__» _____20 г.

Перечень лабораторных работ

Лабораторная работа 1.1. Графематический и морфологический этапы обработки текста на ЕЯ.

Цель работы. Знакомство с этапами и принципами обработки текста на ЕЯ, а также получение практических навыков использования инструментов. Знакомство с графематическим и морфологическим этапами обработки текстов, стеммингом, лемматизацией, омонимией.

Задание:

Необходимо настроить рабочее окружение. В лабораторной работе потребуется определить морфологические особенности слов, их склонение, а также описание правил сочетания в роде, числе и падеже, используя изученные инструменты (pymorphy2, msystem и др.)

Источник данных задается индивидуально по варианту.

Лабораторная работа 1.2. Синтаксический этап обработки текста. Лингвистический взгляд. Грамматики Хомского.

Цель работы. Знакомство с этапами и принципами обработки текста на ЕЯ, а также получение практических навыков использования инструментов. Знакомство с синтаксическим этапом обработки текста, лингвистической постановкой задачи.

Задание:

Необходимо настроить рабочее окружение. В лабораторной работе потребуется определять связи слов в предложениях, используя изученные инструменты (nltk, pymorphy2, msystem и др.) Используя GC-grammar, определить части предложения в тек-

сте.

Источник данных задается индивидуально по варианту.

Лабораторная работа 1.3. Семантический этап обработки текста. Онтологии и тезаурусы.

Цель работы. Знакомство с этапами и принципами обработки текста на ЕЯ, а также получение практических навыков использования инструментов. Знакомство с семантическим этапом обработки текста, онтологией. Изучить применение тезаурусов.

Задание:

Необходимо настроить рабочее окружение. В лабораторной работе потребуется создавать онтологии, используя изученные инструменты. Используя семантические падежи, разметить части текста.

Источник данных задается индивидуально по варианту.

Лабораторная работа 1.4 Корпуса текстов. TF-IDF. Поиск терминов.

Цель работы. Изучить применение корпусов текстов, TF-IDF, а также поиск терминов на примере работы со статистическими инструментами.

Задание:

Необходимо настроить рабочее окружение. В лабораторной работе потребуется статистически определять термины в наборе документов, используя изученные инструменты (RusCorpora и др.). Использовать аналитическое выделение терминов в наборе документов.

Лабораторная работа 1.5. Анализ текстов методом «Мешок слов». N-граммы.

Цель работы. Знакомство с методом анализа текстов «Мешок слов». Изучить применение N-грамм.

Задание:

Необходимо настроить рабочее окружение. В лабораторной работе потребуется составить «Мешок слов», определить тематически сходные документы, используя изученные инструменты.

Источник данных задается индивидуально по варианту.

Лабораторная работа 2.1 Статистический анализ слов. Word2Vec для хранения контекста.

Цель работы. Практическое использование Word2Vec для хранения контекста при статистическом анализе слов в тексте. Изучить особенности выходных данных.

Задание:

Необходимо настроить рабочее окружение. Подготовить текстовые данные для дальнейшей обработки методом контекстного представления в модели Word2Vec. Полученную модель протестировать на корректность и адекватность.

Источник данных задается индивидуально по варианту.

Лабораторная работа 2.2 Автоматическое выделение ключевых элементов текста.

Цель работы. Применение алгоритмов выделения ключевых элементов к векторам документов. Изучить особенности выходных данных.

Задание:

Необходимо настроить рабочее окружение. Подготовить текстовые данные для дальнейшей обработки методом выделения ключевых элементов. Полученные выходные данные объяснить.

Источник данных задается индивидуально по варианту.

Лабораторная работа 2.3 Обработка разговорного ЕЯ.

Цель работы. Практическое использование инструментов для задач поиска шаблонов высказываний.

Задание: Необходимо настроить рабочее окружение. Подготовить текстовые данные для дальнейшего поиска шаблонов, используя изученные инструменты (SpaCy, OpenNLTK, StanfordNLP и др.).

Источник данных задается индивидуально по варианту.

Лабораторная работа 2.4 Чат-боты. Тест Тьюринга и «китайская комната».

Цель работы. Изучение облачных технологий Dialogflow. Реализация чат-бота для изучения экспериментов Тьюринга и «китайской комнаты»

Задание:

Разработать чат-бота Dialogflow. Изучить особенности реализации. Индивидуальное задание.

Лабораторная работа 2.5 Краткое содержание текстов.

Цель работы. Применение алгоритмов суммаризации текста.

Задание:

Необходимо настроить рабочее окружение. Подготовить текстовые данные для дальнейшей обработки методом суммаризации. В лабораторной работе потребуется реализовать алгоритмы суммаризации и получить содержание текста.

Источник данных задается индивидуально по варианту.