Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 12.10.2023 17:28:14
Уникальный программный ключ:
8db180d1a3f02ac9e60521a672742f7f5e10b1b6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение высшего образования

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ
Директор
Полиграфического института
_____ /И.В. Нагорнова/
«_30_» ___июня___ 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Теоретические основы переработки текстовой информации в автоматизированных системах»

Направление подготовки
22.03.01 «Материаловедение и технологии материалов»

Профиль
«Материаловедение и цифровые технологии»

Квалификация (степень) выпускника
Бакалавр

Форма обучения
очная

Москва 2022

1. Цели освоения дисциплины

Для обучающихся по направлению 22.03.01 «Материаловедение и технологии материалов» курс «Теоретические основы переработки текстовой информации в автоматизированных системах» является обязательной дисциплиной вариативной части, устанавливаемой высшим учебным заведением.

Основными целями освоения дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» являются получение представления о преобразовании текстовой информации на этапах технологического процесса производства печатной продукции, о применяемых программных и технических средствах, качественных и количественных характеристиках оценки обработки текста в автоматизированных системах принтмедиаиндустрии.

Задачами изучения дисциплины является изучение теоретических основ преобразования информации в процессе производства печатной продукции, основных методов редакционно-издательской работы с текстом, получение представления об уровне автоматизации процессов принтмедиаиндустрии, изучение информационных основ переработки текста, характеристик и информационных свойств текста, знакомство с программными средствами ввода, обработки текста в принтмедиаиндустрии, с методами и количественными характеристиками оценки качества печатной продукции.

2. Место дисциплины в структуре ООП бакалавриата

Дисциплина «Теоретические основы переработки текстовой информации в автоматизированных системах» относится к блоку учебного плана, части, формируемой участниками образовательных отношений.

Изучение данной дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» базируется на

- знаниях по математике и информатике, полученных обучающимися в школьном курсе,
- дисциплинах базовой части учебного плана 22.03.01 профиля «Материаловедение и цифровые технологии» подготовки бакалавров, параллельно изучаемых обучающимися в первом семестре:
  - — Высшая математика,
  - — Цифровая грамотность.

Для освоения учебной дисциплины, обучающиеся:

- иметь навыки выполнения математических расчетов, действий с матрицами, построения графиков, логарифмирования, решения систем уравнений, дифференцирования;

- иметь общее представление об основных положениях теории информации;
- знать методы представления информации в ЭВМ;
- знать автоматизированные программные средства выполнения математических расчетов.

Основные положения дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» используются в дальнейшем при изучении следующих дисциплин учебного плана подготовки бакалавров направления 22.03.01 профиля «Материаловедение и цифровые технологии»:

- Современные программные средства моделирования процессов и объектов,
- Основы деловой коммуникации,
- Методы анализа и контроля показателей качества среды в принтмедиаиндустрии,
- Полиграфические технологии в производстве печатной продукции.

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

В результате освоения дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» у обучающихся формируются следующие компетенции и должны быть достигнуты следующие результаты обучения как этап формирования соответствующей компетенции:

Коды компе-тенции	Результаты освоения ООП Содержание компетенций*	Перечень планируемых результатов обучения по дисциплине**
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИУК-1.1. Анализирует задачу, выделяя ее базовые составляющие ИУК-1.2. Осуществляет поиск, критически оценивает, обобщает, систематизирует и ранжирует информацию, требуемую для решения поставленной задачи ИУК-1.3. Рассматривает и предлагает рациональные варианты решения поставленной задачи, используя системный подход, критически оценивает их достоинства и недостатки.
УК- 4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке	ИУК-4.1. Учитывает особенности деловой коммуникации на государственном и иностранном языках в зависимости от особенностей вербальных и невербальных средств общения ИУК-4.2. Умеет вести обмен деловой

Коды компе-тенции	Результаты освоения ООП Содержание компетенций*	Перечень планируемых результатов обучения по дисциплине**
	Российской Федерации и иностранном(ых) языке(ах)	информацией в устной и письменной формах на государственном и иностранном языках с учетом своеобразия стилистики официальных и неофициальных писем, а также социокультурных различий в формате корреспонденции

4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 часа), в том числе самостоятельная работа обучающегося в объёме 54 часов (36 из которых – подготовка к экзамену) для очной формы обучения. Лекционные занятия планируются в объеме 18 часов, лабораторные занятия - в объеме 36 часов.

Изучение дисциплины происходит в течение одного (первого) семестра.

Форма обучения	**курс**	**семестр**	**Трудоемкость дисциплины в часах**							**Форма итогового контроля**
			Всего час./ зач. ед	**Аудиторных часов (контактная работа)**	**Лекции**	**Семинарские (практические) занятия**	**Лабораторные работы**	**Самостоятельная работа**	**Контроль (промежуточная аттестация)**	
Очная	1	1	108/3	54	18	–	36	18	36	**экзамен**
Очно-заочная	–	–	–	–	–	–	–	–	–	–
Заочная	–	–	–	–	–	–	–	–	–	–

Структура и содержание дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» по срокам и видам работы отражены в Приложении 1.

Содержание разделов дисциплины

Тема 1. Введение. Общие сведения о принтмедиаиндустрии и месте полиграфии в культурной жизни общества.

Тема 2. Общие сведения об информации.

Общие сведения об информации и информатике. Сведения об информации и ее видах. Содержание информатики как научной дисциплины. Перспективы развития средств информации. Автоматизация как основное направление развития. Информационные основы переработки текста и изображений средств переработки информации.

Тема 3. Количественные характеристики информации.

Количественные характеристики информации. Энтропия. Энтропия взаимосвязанных событий. Количество информации в непрерывных сигналах и дискретных сообщениях.

Тема 4. Информационные свойства текстов.

Общая характеристика текстов. Шрифты, предъявляемые к ним требования. Роль статистики языка при обработке текстов. Формы информации, содержащейся в текстах. Символьная, синтаксическая и семантическая информация. Прогнозирование появления букв и слов в текстах. Избыточность текста.

Тема 5. Математические основы технологии верстки изданий.

Математические основы технологии верстки изданий. Общие сведения о технологии набора и верстки изданий. Программное обеспечение. Организационно-техническое обеспечение.

Тема 6. Качество переработки текста.

Требования к качеству полиграфической продукции. Некоторые требования технических правил набора и верстки. Методы редакционно-издательской работы с текстом. Источники и виды ошибок в текстах. Количественные характеристики ошибок. Методы оценки количества ошибок в тексте.

Тема 7. Методы автоматического чтения и распознавания текста.

Место читающих автоматов в технологической цепи переработки текста. Распознание текстов с помощью стандартных программных средств. Распознавание печатных текстов. Особенности распознавания рукописных текстов. Особенности распознавания голосовых сообщений.

Тема 8. Преобразование количества ошибок в тексте при автоматизированном процессе переработки текстовой информации: компьютерном наборе и распознавании.

Моделирование процесса автоматизированного набора. Общие сведения о случайных марковских процессах. Векторы состояния текстов и их преобразование. Переходные и установившиеся режимы марковских процессов. Преобразование текста при автоматизированном вводе. Преобразование текста при наборе.

Тема 9. Кодирование и корректура текстов как процесс управления потоками ошибок.

Потоки событий и их свойства. Процесс накопления ошибок в текстах. Динамика идеальной корректуры. Динамика корректуры с пропуском ошибок. Моделирование процессов последовательных корректур. Оценка качества редакционно-издательского процесса.

5. Образовательные технологии

Методика преподавания дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» и реализация компетентностного подхода в изложении и восприятии

материала предусматривает использование следующих форм проведения групповых, индивидуальных, контактных (аудиторных) занятий в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся:

- проведение занятий лекционного типа;
- подготовка к выполнению лабораторно-практических работ в лабораториях и компьютерных классах вуза;
- оформление и защита лабораторных работ;
- организация и проведение текущего контроля знаний обучающихся в форме опроса или бланочного тестирования;
- контрольная работа.

При проведении лекционных и лабораторных занятий, промежуточной и семестровой аттестации по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах» целесообразно использовать следующие образовательные технологии:

1. На лабораторных занятиях использовать современное программное обеспечение, применяемое в принтмедиаиндустрии для обработки текста и иллюстраций, что позволяет формировать практические навыки.
2. Процедуры промежуточного контроля по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах» допускается проводить в форме бланочного или компьютерного тестирования.
3. В течение семестра в рамках самостоятельной работы обучающиеся выполняют индивидуальные задания, состоящее из теоретической (доклад) и практической частей (индивидуальный вариант контрольного задания в лабораторной работе).
4. Проведение лекционных занятий, содержащих таблицы и рисунки в качестве иллюстраций, целесообразно осуществлять с использованием слайдов, подготовленных в программе Microsoft Power Point. Лекционная часть проводится в форме онлайн конференций в системе Webinar.ru по ссылке, указанной в расписании учебных занятий.
5. Самостоятельная проработка дополнительного материала на площадке дистанционного образования Московского Политеха https://online.mospolytech.ru

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы обучающихся

В процессе обучения используются следующие оценочные формы самостоятельной работы обучающихся: оценочные средства текущего контроля успеваемости и промежуточных аттестаций, подготовка к лабораторным занятиям и их выполнение.

Оценочные средства текущего контроля успеваемости включают контрольные вопросы и задания в форме тестирования, для контроля освоения обучающимися разделов дисциплины, защиты лабораторных работ, решение контрольных работ.

Образцы тестовых заданий, контрольных вопросов и заданий для проведения текущего контроля, вопросов и заданий для экзамена, приведены в приложении 2.

Формы текущего контроля успеваемости по разделам дисциплины приведены в содержании разделов (см. п. 6 настоящей рабочей программы).

6.1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

6.1.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

В результате освоения дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» формируются следующие универсальные компетенции:

Код компетенции	В результате освоения образовательной программы обучающийся должен обладать
УК-1	способностью осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач
УК- 4	способностью осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)

В процессе освоения образовательной программы данные компетенции, в том числе их отдельные компоненты, формируются в ходе освоения обучающимися дисциплин (модулей), практик в соответствии с учебным планом и календарным графиком учебного процесса. Дисциплина «Теоретические основы переработки текстовой информации в автоматизированных системах» участвует в формировании перечисленных компетенций.

6.1.2. Описание показателей и критериев оценивания компетенций, формируемых по итогам освоения дисциплины, описание шкал оценивания

Показателем оценивания компетенций на различных этапах их формирования является достижение обучающимися планируемых результатов обучения по дисциплине.

УК-1 – Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач				
Код и индикаторы достижения компетенции	Критерии оценивания			
	2	3	4	5
ИУК-1.1. Анализирует задачу, выделяя ее базовые составляющие	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие ИУК -1.1.	Обучающийся демонстрирует неполное соответствие ИУК -1.1. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие ИУК -1.1. Допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие ИУК -1.1. Свободно оперирует приобретенными знаниями.
ИУК-1.2. Осуществляет поиск, критически оценивает, обобщает, систематизирует и ранжирует информацию, требуемую для решения поставленной задачи	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие ИУК-1.2.	Обучающийся демонстрирует неполное соответствие ИУК-1.2. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие ИУК-1.2. Допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие ИУК-1.2. Свободно оперирует приобретенными знаниями.
ИУК-1.3. Рассматривает и предла-	Обучающийся демонстрирует полное отсут-	Обучающийся демонстрирует неполное соответ-	Обучающийся демонстрирует частичное соот-	Обучающийся демонстрирует полное соот-

гает рациональные варианты решения поставленной задачи, используя системный подход, критически оценивает их достоинства и недостатки.	ствие или недостаточное соответствие ИУК-1.3.	ствие ИУК-1.3. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	ветствие ИУК-1.3. Допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	ветствие ИУК-1.3. Свободно оперирует приобретенными знаниями.
УК-4 – Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)				
ИУК-4.1. Учитывает особенности деловой коммуникации на государственном и иностранном языках в зависимости от особенностей вербальных и невербальных средств общения	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие ИУК-4.1.	Обучающийся демонстрирует неполное соответствие ИУК-4.1. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.	Обучающийся демонстрирует частичное соответствие ИУК-4.1. Допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие ИУК-4.1. Свободно оперирует приобретенными знаниями.
ИУК-4.2. Умеет вести обмен деловой информацией в устной и письменной формах на государственном и иностранном языках с учетом	Обучающийся демонстрирует полное отсутствие или недостаточное соответствие ИУК-4.2.	Обучающийся демонстрирует неполное соответствие ИУК-4.2. Допускаются значительные ошибки, проявляется недостаточность знаний, по ряду показателей, обучающийся	Обучающийся демонстрирует частичное соответствие ИУК-4.2. Допускаются незначительные ошибки, неточности, затруднения при аналитических операциях.	Обучающийся демонстрирует полное соответствие ИУК-4.2. Свободно оперирует приобретенными знаниями.

своеобразия стилистики официальных и неофициальных писем, а также социокультурных различий в формате корреспонденции		испытывает значительные затруднения при оперировании знаниями при их переносе на новые ситуации.		

Шкалы оценивания результатов промежуточной аттестации и их описание
Форма промежуточной аттестации: экзамен.

Промежуточная аттестация обучающихся в форме экзамена проводится по результатам выполнения всех видов учебной работы, предусмотренных учебным планом по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах», при этом учитываются результаты текущего контроля успеваемости в течение семестра. По итогам промежуточной аттестации выставляется оценка «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно».

Примерный алгоритм оценки результатов ответа обучающегося на экзамене выглядит следующим образом:

Ответ на каждый вопрос экзаменационного билета оценивается по шкале: «отлично», «хорошо», «удовлетворительно» или «неудовлетворительно». Оценка ответу обучающегося на вопрос билета присваивается следующим образом:

Качество ответа студента	оценка
Отсутствует ответ на вопрос / дан полностью неверный ответ / ответ не по теме вопроса	неудовлетворительно
Дан краткий ответ с существенными (большим количеством) ошибками / неточностями	
Дан полный ответ, содержащий ошибки / неточности. На наводящие вопросы даны неверные (неполные) ответы	удовлетворительно
Дан развёрнутый ответ, содержащий ошибки / неточности. На наводящие вопросы даны неполные ответы	хорошо
Дан развернутый ответ, содержащий ошибки / неточности. На наводящие вопросы даны верные, развёрнутые ответы	отлично
Дан правильный развернутый ответ на вопрос билета	

Балл ответа на экзаменационный билет рассчитывается как среднее из общего количества вопросов и задач в экзаменационном билете.

Оценка работы обучающегося в семестре осуществляется с учётом прохождения точек промежуточного контроля знаний.

Контрольная точка 1 - Количественные характеристики информационного сообщения, как случайного процесса. Энтропия случайных дискретных сообщений.

Контрольная точка 2 - Количественные характеристики информационного сообщения, как случайного процесса. Энтропия взаимосвязанных сообщений.

Контрольная точка 3 - Программные средства чтения и распознавания текста. Сравнительная характеристика. Программные средства набора и верстки. (Доклад, сообщение)

Контрольная точка 4 - Анализ качества переработки текста в процессе операции набора и верстки текста. Моделирование технологической операции методом дискретных марковских процессов. Создание математической вероятностной (матричной) модели процесса набора. Анализ качества переработки текста в процессе операции корректуры текста. Создание математической вероятностной (матричной) модели процесса корректуры.

Во время лабораторно-практических занятий преподаватель оценивает активность студента, учитывая работу у доски, ответы на поставленные вопросы, ответы во время опроса, подготовку отчетов по лабораторным работам и сдачу указанные сроки, по шкале:

«Неудовлетворительно/Удовлетворительно/Хорошо/Отлично».

Оценка «Неудовлетворительно» соответствует 0 баллам (как и отсутствие студента на занятиях).

Шкала оценки работы студента на лабораторном занятии следующая:

- Неудовлетворительно - обучающийся не работал в течение занятия, или отсутствовал (менее 65 баллов),
- Удовлетворительно - обучающийся не смог правильно объяснить выполнение задания, выполнил не все запланированные задания или не в полном объеме (65-70 баллов),
- Хорошо - обучающийся, работая активно, выполнил не все запланированные задания, не предоставил обоснования (71-84 балла),
- Отлично - обучающийся выполнил все задания и правильно отвечал на поставленные по заданиям вопросы (85-100 баллов).

Фонд и образцы оценочных средств представлены в приложении 2 к рабочей программе дисциплины.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Винокурова О.А. Теоретические основы переработки информации в полиграфии. Задания для практических занятий и для самостоятельной работы студентов – М.: МГУП, 2010; 105 с.
2. Самарин, Ю.Н. Технологические процессы автоматизированных производств (Полиграфическое производство): учебник / Ю.Н. Самарин; Моск. гос. ун-т печати имени Ивана Федорова. - М. : МГУП, 2015. - 556 с.

7.2. Дополнительная литература:

1. Мильчин, А.Э. Справочник издателя и автора, редакционно-издательское оформление издания; М., Изд-во Студии Артемия Лебедева, 2014;1010 с.
2. Самарин, Ю.Н. История автоматизации допечатных процессов, учебное пособие для студентов, обучающихся по спец.: 220201.65 «Управление и информатика в технических системах»; 220301.65 «Автоматизация технологических процессов и производств (полиграфия)»; 220200.62 «Автоматизация и управление»; М.,МГУП, 2008; 240 с.

3. MATHCAD в обучении информатике и математике, учеб. пособие для студентов высш. учеб. заведений, обучающихся по направлению подготовки ВПО 010300 - Математика. Компьютерные науки; Тула, Изд-во ТГПУ им. Л.Н. Толстого, 2009; 363 с.;

7.3. Программное обеспечение и Интернет-ресурсы:

1. Банк тестовых заданий в системе тестирования LMS Moodle по курсу дисциплины или бланочное тестирование.
2. Mathcad-14. Пакет компьютерной математики. Договор № 24/08 от 19.05.2008 г.; LibreOffice 5.0 Бесплатная версия Adobe Acrobat Reader. Бесплатная версия; Microsoft Office Стандартный 2007 (Word, Excel, PowerPoint) Договор № 1981-M87 от 03.02.2014 г;
   Adobe Acrobat Reader. Бесплатная версия;
   FireFox. Интернет-браузер. Бесплатная версия.
3. Полезные учебно-методические и информационные материалы представлены на сайтах:

http://freeanalogs.ru
http://newocr.com

8. Материально-техническое обеспечение дисциплины

1. Аудитории общего фонда для лекционных, лабораторных занятий г. Москва, ул. Прянишникова, д. 2а.
2. Персональные компьютеры, мониторы, столы, интерактивная доска или проектор. Рабочее место преподавателя: персональный компьютер, стол, стул. Комплекс технических средств, позволяющих проецировать изображение из программных средств подготовки презентаций (экран, проектор, ноутбук или компьютер с подключенным оборудованием).
3. Возможность доступа в интернет.

9. Методические рекомендации для самостоятельной работы обучающихся

Рабочим учебным планом предусмотрено изучение дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» в 1 семестре при очной форме обучения (1-й год обучения). По дисциплине проводятся лекционные, лабораторные и практические занятия.

Лекционные занятия проводятся в соответствии с содержанием настоящей рабочей программы и представляют собой изложение основ теории информации, теоретических основ переработки текстовой и изобразительной информации в принтмедиасистемах и комплексах, основных положений преобразования информации в процессе производства и анализ качества выполнения технологических операций, сравнительная характеристика применяемых программных средств, изложение перспектив дальнейшего развития.

Посещение лекционных занятий является обязательным. Допускается конспектирование лекционного материала как письменным, так и компьютерным способом.

Регулярная проработка материала конспектов лекций по каждому разделу в рамках подготовки к промежуточным и итоговым формам аттестации по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах» является одним из важнейших видов самостоятельной работы обучающегося в течение семестра, необходимой для качественной подготовки к промежуточной и семестровой аттестации по дисциплине. В рамках подготовки к лабораторным занятиям рекомендуется повторение теоретического материала по соответствующей теме, изучение литературы.

Промежуточная аттестация по дисциплине проходит в форме экзамена. Примерный перечень вопросов к экзамену по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах» приведен в приложении 2 к настоящей рабочей программе.

10. Методические рекомендации для преподавателя

Дисциплина «Теоретические основы переработки текстовой информации в автоматизированных системах» является дисциплиной к вариативной части учебного плана, устанавливаемой высшим учебным заведением, обеспечивает формирование представлений о преобразовании текстовой информации на этапах технологического процесса производства печатной продукции, о применяемых программных и технических средствах, качественных и количественных характеристиках оценки обработки текста в автоматизированных системах принтмедиаиндустрии.

В условиях конструирования образовательных систем на принципах компетентностного подхода произошло концептуальное изменение роли преподавателя, который наряду с традиционной ролью носителя знания выполняет функцию организатора научно-поисковой работы студента, консультанта в процедурах выбора, обработки и интерпретации информации, необходимой для практического действия и дальнейшего развития. Это должно обязательно учитываться при проведении лекционных, лабораторных и практических занятий по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах».

Преподавание теоретического (лекционного) материала по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах» осуществляется по последовательной схеме на основе учебного плана по направлению 22.03.01 «Материаловедение и технологии материалов».

Подробное содержание отдельных разделов дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» рассматривается в разделе 4 рабочей программы.

Структура и последовательность проведения лабораторных и практических занятий по дисциплине представлена в приложении 1 к настоящей рабочей программе.

Целесообразные к применению в рамках дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах» образовательные технологии изложены в п. 5 настоящей рабочей программы.

Перечень основной и дополнительной литературы и нормативных документов, необходимых в ходе преподавания дисциплины «Теоретические основы переработки текстовой информации в автоматизированных системах», приведен в п. 7 настоящей рабочей программы. Преподавателю следует ориентировать студентов на использование при подготовке к промежуточной аттестации по дисциплине, материалов лекций. Предпочтение работы с лекциями чтению учебной литературы формирует у обучающегося навыки самостоятельной работы.

При проведении занятий рекомендуется использование активных и интерактивных форм занятий (устных опросов, докладов, сообщений) в сочетании с внеаудиторной работой.

Программа составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки бакалавров 22.03.01 Материаловедение и технологии материалов, утвержденным приказом МНиВО РФ от 2 июня 2020 г. № 701.

Программу составила:
доцент, к.т.н., Винокурова О.А. /_____/.

Программа утверждена на заседании кафедры «Полиграфические системы» «23» июня 2022 г., протокол № 11.

Заведующий кафедрой
доцент, к.т.н. /Суслов М.В./

Согласовано:
Директор ПИ /Нагорнова И.В./

Приложение 1.

Структура и содержание дисциплины
«Теоретические основы переработки текстовой информации в автоматизированных системах»
по направлению подготовки
22.03.01 – «Материаловедение и технологии материалов»

П1.1. Тематический план дисциплины (для очной формы обучения)

№ п/п	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу обучающихся, и трудоемкость в часах				Виды самостоятельной работы обучающихся					Формы аттестации	
				Л	П/С	Лаб	СРС	КР	КП	РГР	Реферат	К/Р	Э	З
1.	*Введение.* Общие сведения о принтмедиаиндустрии и месте полиграфии в культурной жизни общества.	1	1	2			2						2	
2.	*Тема 1.* Общие сведения об информации.	1	2	2			2						2	
3.	*Лабораторная работа* 1 Энтропия как мера неопределенности результатов наблюдений. Энтропия дискретных сообщений.	1	3			4							4	
4.	*Тема 2.* Количественные характеристики информации.	1	4	2			2						2	
5.	*Лабораторная работа* 2 Оценка количественных показателей взаимозависимых информационных сообщений. Энтропия дискретных взаимосвязанных сообщений.	1	5			4							2	

№ п/п	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу обучающихся, и трудоемкость в часах				Виды самостоятельной работы обучающихся					Формы аттестации	
				Л	П/С	Лаб	СРС	КР	КП	РГР	Реферат	К/Р	Э	З
	Условная энтропия. Оценка взаимозависимости событий Применение программных средств для выполнения индивидуального задания.													
6.	*Тема 3.* Информационные свойства текстов.	1	6	2			2						2	
7.	*Лабораторная работа 3* Частотный анализ фрагмента адаптированного текста русского языка.	1	7			4								
8.	*Лабораторная работа* 4 Частотный анализ фрагмента литературного текста русского языка	1	8			4							1	
9.	*Лабораторная работа* 5 Частотный анализ фрагмента литературного англоязычного текста	1	9			4							1	
10.	*Тема 4.* Общие сведения о технологии набора и верстки изданий. Математические основы технологии верстки изданий	1	10	2			2				2		2	
11.	*Лабораторная работа 6.* Форматирование фрагмента литературного текста по заданным параметрам в Microsoft Word		11			4							2	

№ п/п	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу обучающихся, и трудоемкость в часах				Виды самостоятельной работы обучающихся					Формы аттестации	
				Л	П/С	Лаб	СРС	КР	КП	РГР	Реферат	К/Р	Э	З
12.	*Тема 5.* Качество переработки текста. Требования к качеству полиграфической продукции	1	12	2			2						2	
13.	*Тема 6.* Методы и программное обеспечение автоматического чтения и распознавания текста.	1	13	2			2						2	
14.	*Лабораторная работа* 7 Программные средства чтения и распознавания текста. Сравнительная характеристика.	1	14			4							2	
15.	*Тема 7.* Преобразование количества ошибок в тексте в автоматизированном процессе.	1	15	2			2						3	
16.	*Лабораторная работа* 8 Моделирование технологической операции набора текста методом дискретных марковских процессов. Создание математической (матричной) модели процесса набора.	1	16			4							2	
17.	*Тема 8.* Кодирование и корректура текстов как процесс управления потоками ошибок.	1	17	2			2					2	3	

№ п/п	Раздел	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу обучающихся, и трудоемкость в часах				Виды самостоятельной работы обучающихся					Формы аттестации	
				Л	П/С	Лаб	СРС	КР	КП	РГР	Реферат	К/Р	Э	З
18.	*Лабораторная работа* 9 Моделирование технологической операции последовательных корректур текста методом дискретных марковских процессов. Создание математической (матричной) модели процесса корректуры.	1	18			4							2	
19.	***Форма промежуточной аттестации***	1	18										**Э**	
20.	Всего часов по дисциплине	**108**		**18**		**36**	**18**						**36**	

Приложение 2.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

(МОСКОВСКИЙ ПОЛИТЕХ)

Направление подготовки: 22.03.01 МАТЕРИАЛОВЕДЕНИЕ И ТЕХНОЛОГИИ МАТЕРИАЛОВ

ООП (профиль): «Материаловедение и цифровые технологии»

Форма обучения: очная

Вид профессиональной деятельности: научно-исследовательская

Кафедра «Полиграфические системы»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

Теоретические основы переработки текстовой информации в автоматизированных системах

Состав:

1. Паспорт фонда оценочных средств
2. Показатель сформированности компетенций
3. Примерный перечень оценочных средств
4. Описание оценочных средств (образцы тестовых заданий, контрольных и экзаменационных вопросов по курсу «Теоретические основы переработки текстовой информации в автоматизированных системах»)

Составитель:

доц., к.т.н. Винокурова О.А.

Москва 2022 г.

П2.1 Паспорт фонда оценочных средств по дисциплине

«Теоретические основы переработки текстовой информации в автоматизированных системах»

№ п/п	Контролируемые разделы дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1.	*Тема 1. Введение.* Общие сведения о принтмедиа-индустрии и месте полиграфии в культурной жизни общества.	УК-1	УО Т Э
2.	*Тема 2.* Общие сведения об информации.	УК-1	УО Т К/Р Э
3.	*Тема 3.* Количественные характеристики информации.	УК-1	УО Т ОЛР К/Р Э
4.	*Тема 4.* Информационные свойства текстов.	УК-1	УО Т ОЛР Э
5.	*Тема 5* Общие сведения о технологии набора и верстки изданий. Математические основы технологии верстки изданий.	УК-1 УК-4	УО ОЛР Т Э
6.	*Тема 6.* Качество переработки текста. Требования к качеству полиграфической продукции	УК-1 УК-4	УО Э
7.	*Тема 7.* Методы и программное обеспечение автоматического чтения и распознавания текста.	УК-1 УК-4	УО ОЛР ДС Э
8.	*Тема 8.* Преобразование количества ошибок в тексте в автоматизированном процессе.	УК-1 УК-4	УО ОЛР Т Э
9.	*Тема 9.* Кодирование и корректура текстов как процесс управления потоками ошибок.	УК-1 УК-4	УО ОЛР Т Э

П2.2. Показатель уровня сформированности компетенций

Теоретические основы переработки текстовой информации в автоматизированных системах
ФГОС ВО 22.03.01 – «Материаловедение и технологии материалов»

В процессе освоения данной дисциплины студент формирует и демонстрирует следующие компетенции

Компетенции		Перечень индикаторов достижения компетенции	Технология формирования компетенций	Форма оценочного средства**	Степени уровней освоения компетенций
Индекс	Формулировка				
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИУК-1.1. Анализирует задачу, выделяя ее базовые составляющие ИУК-1.2. Осуществляет поиск, критически оценивает, обобщает, систематизирует и ранжирует информацию, требуемую для решения поставленной задачи ИУК-1.3. Рассматривает и предлагает рациональные варианты решения поставленной задачи, используя системный подход, критически оценивает их достоинства и недостатки.	Лекция Лабораторная работа Самостоятельная работа	УО Т ОЛР К/Р ДС Э	**Базовый уровень:** Анализирует задачу, выделяя ее базовые составляющие Осуществляет поиск, критический анализ и синтез информации. Рассматривает варианты решения поставленной задачи, оценивает их достоинства и недостатки. **Повышенный уровень** Самостоятельно применяет системный подход для решения поставленных задач. Критически оценивает, обобщает, систематизирует и ранжирует информацию, требуемую для решения поставленной задачи Рассматривает и предлагает рациональные варианты решения поставленной задачи, используя системный подход, критически оценивает их достоинства и недостатки.

УК-4	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)	ИУК-4.1. Учитывает особенности деловой коммуникации на государственном и иностранном языках в зависимости от особенностей вербальных и невербальных средств общения ИУК-4.2. Умеет вести обмен деловой информацией в устной и письменной формах на государственном и иностранном языках с учетом своеобразия стилистики официальных и неофициальных писем, а также социокультурных различий в формате корреспонденции	Лекция Лабораторная работа Самостоятельная работа	УО ОЛР ДС Э	**Базовый уровень:** Использует информационно-коммуникационные технологии при поиске необходимой информации в процессе решения стандартных задач. Учитывает особенности деловой коммуникации при решении профессиональных задач. Знает наиболее используемые программные средства подготовки оригинала в процессе производства; Владеет навыками использования программных средств обработки и преобразования текстовой информации. **Повышенный уровень** Умеет вести обмен деловой информацией в устной и письменной формах. Использует информационно-коммуникационные технологии для самостоятельного решения профессиональных задач обработки текстовой информации.

**- Сокращения форм оценочных средств см. в приложении 2.3 к РП.

П2.3 Примерный перечень оценочных средств (ОС) по дисциплине

«Теоретические основы переработки текстовой информации в автоматизированных системах»

№ ОС	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1.	Отчет по лабораторной работе (ОЛР)	Продукт самостоятельной работы обучающегося, представляющий собой средство проверки умений применять полученные знания для решения поставленной задачи по заранее определенной методике и краткое изложение в письменном виде полученных результатов экспериментального и теоретического анализа определенной учебно- исследовательской темы.	Перечень и темы лабораторных работ
2.	Тест (Т)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.	Фонд тестовых заданий
3.	Устный опрос собеседование, (УО)	Средство контроля, организованное как специальная беседа педагогического работника с обучающимся на темы, связанные с изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п.	Вопросы по темам/разделам дисциплины
4.	Контрольная работа (К/Р)	Средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу	Комплект контрольных заданий по вариантам
5.	Доклад, сообщение (ДС)	Продукт самостоятельной работы студента, представляющий собой публичное выступление по представлению полученных результатов решения определенной учебно-практической, учебно-исследовательской или научной темы	Темы докладов, сообщений
6.	Экзамен (Э)	Форма промежуточной аттестации студента, определяемые учебным планом подготовки по направлению	Комплект экзаменационных билетов

П2.4. Описание оценочных средств по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

П2.4.1 Критерии оценки работы обучающегося на лабораторных работах (ЛР)

«5» (отлично): оформлен отчет (ОЛР) и выполнена лабораторная работа, предусмотренная планом; проведены все необходимые расчеты, корректно написан вывод.

Обучающийся:

на высоком уровне владеет знанием теоретических основ преобразования

информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«4» (хорошо): оформлен отчет (ОЛР) и выполнена лабораторная работа с корректирующими замечаниями преподавателя; проведены все необходимые расчеты, корректно написан вывод.

Обучающийся:

хорошо владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«3» (удовлетворительно): оформлен отчет (ОЛР) и выполнена лабораторная работа, проведены все необходимые расчеты, написан вывод с корректирующими замечаниями преподавателя.

Обучающийся:

на удовлетворительном уровне владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«2» (неудовлетворительно): обучающийся не оформил отчет (ОЛР), не выполнил или выполнил неправильно лабораторные работы, предусмотренные планом.

Обучающийся:

не владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных

программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

П2.4.2. Критерии оценки компьютерного тестирования (Т) обучающегося

Компьютерное тестирование проводится для текущего контроля знаний студентов, оценивается в соответствии с процентом правильных ответов, данных студентом на вопросы теста.

Стандартная шкала соответствия результатов компьютерного тестирования выставляемой балльной оценке:

– «отлично» - свыше 85% правильных ответов;
– «хорошо» - от 70,1% до 85% правильных ответов;
– «удовлетворительно» - от 55,1% до 70% правильных ответов;
– от 0 до 55% правильных ответов – «неудовлетворительно»

Стандартный регламент тестирования включает:

– количество вопросов – 30;
– продолжительность тестирования – 30 минут;
– генерация теста из БТЗ – методом случайной выборки;
– режим контроля – жесткий (отсутствие возможности тестируемым увидеть результат ответа на вопрос теста в процессе тестирования).

«отлично»: тестируемый демонстрирует системные теоретические знания, владеет терминами и обладает способностью быстро реагировать на вопросы теста.

Обучающийся:

на высоком уровне владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета.

«хорошо»: тестируемый в целом демонстрирует системные теоретические знания, владеет большинством терминов и обладает способностью быстро реагировать на вопросы теста.

Обучающийся:

хорошо владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета.

«удовлетворительно»: системные теоретические знания у тестируемого отсутствуют, он владеет некоторыми терминами, на вопросы теста реагирует достаточно медленно.

Обучающийся:

на удовлетворительном уровне владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета.

«неудовлетворительно»: системные теоретические знания у тестируемого отсутствуют, терминологией не владеет, на вопросы теста реагирует медленно.

Обучающийся:

не владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета.

П2.4.3. Критерии оценки устного опроса обучающегося (УО)

Устный опрос (контрольные точки) по текущей теме лабораторных работ проводится во время лабораторных занятий в виде собеседования по теме и результатам выполнения лабораторной работы.

Оценивается:

«максимум» - 3 балла, «минимум» - 2 балла, «неудовлетворительно» - менее 2 баллов.

«максимум»: обучающийся четко и без ошибок или с корректирующими замечаниями преподавателя ответил на все контрольные вопросы по теме лабораторной работы.

Обучающийся:

на высоком уровне или хорошо владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«минимум»: обучающийся ответил на все контрольные вопросы по данной теме лабораторной работы с замечаниями.

на удовлетворительном уровне владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«неудовлетворительно»: обучающийся ответил на контрольные вопросы по теме лабораторной работы с ошибками или не ответил на контрольные вопросы.

Обучающийся:

не владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов

оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

П2.4.4. Критерии оценки обучающегося на контрольной работе (К/Р)

Контрольные работы проводятся в виде вариантов заданий по темам дисциплины.
Контрольная работа №1: минимум («удовлетворительно») – 16 баллов, максимум («отлично») – 31 балл;
Контрольная работа №2: минимум – 16 баллов, максимум –31 балл;
Контрольная работа №3: минимум – 16 баллов, максимум –31 балл;
Контрольная работа №4: минимум – 16 баллов, максимум –31 балл;

«отлично»: обучающийся выполнил все задания, предусмотренные контрольной работой, без ошибок выполнены все расчеты и построения, сделаны корректные выводы на основании расчетов.

Обучающийся:

на высоком уровне или хорошо владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«хорошо»: обучающийся выполнил все задания, предусмотренные контрольной работой, в расчетах допущены ошибки, при указании на них, самостоятельно исправляются обучающимся, сделаны корректные выводы на основании расчетов.

Обучающийся:

хорошо владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«удовлетворительно»: обучающийся выполнил не все задания, предусмотренные контрольной работой, сделаны корректные выводы на основании расчетов.

Обучающийся:

на удовлетворительном уровне владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции,

информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

«неудовлетворительно»: обучающийся выполнил одно из заданий контрольной работы, или выполнил неверно, ответил на контрольные вопросы по темам с ошибками или не ответил на контрольные вопросы.

Обучающийся:

не владеет знанием теоретических основ преобразования информации в процессе производства печатной продукции, информационных свойств текста, его количественных характеристик, специальной терминологией, математических методов оценки качества набора и корректуры текста; навыками оценки информационной емкости текста с использованием математических методов и программных средств расчета, использования статистических и вероятностных методов моделирования технологических операций набора и корректуры текста, использования современных программных средств редакционно-издательской работы с текстовыми оригиналами, современных программных средств чтения и распознавания текста.

П2.5. Примерный перечень элементов ФОС для проверки уровня сформированности компетенций по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

Для проверки уровня сформированности компетенций УК-1, УК-4 согласно установленным показателям (см. приложение П2.2) используются следующие формы оценочного средства: устный опрос, контрольная работа, тематика докладов, тестовые задания.

П2.5.1 Контрольные вопросы по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

Приведённый ниже перечень контрольных вопросов используется в качестве вопросов при устном опросе обучающихся.

1. Что такое «информация»?
2. Что такое «бит»?
3. Что такое «байт» или чему он равен?
4. Что называется сигналом?
5. Дать классификацию сигнала по физической природе.
6. Перечислите фазы преобразования получаемой информации.
7. Изобретения, ознаменовавшие важнейшие этапы совершенствования и создания средств приема, передачи и хранения информации.
8. Перечислите массовые средства письменного и иного хранения информации.
9. Дайте определение энтропии сообщения.
10. Перечислите свойства энтропии.

11. Запишите формулу для определения энтропии источника дискретных сообщений.
12. Как связана энтропия дискретного сообщения и его дисперсия.
13. Запишите формулу для определения энтропии непрерывного сигнала.
14. Запишите формулу для определения количества информации в дискретном сообщении.
15. Как определить количество информации во взаимосвязанных сообщениях.
16. Что такое «текст»?
17. Перечислите основные характеристики шрифта.
18. Что такое «кегль шрифта»? Перечислите кегли, имеющие названия.
19. Каковы единицы измерения типографских шрифтов?
20. Что такое «гарнитура шрифта»? Перечислите несколько.
21. Что такое «технологичность шрифта»?
22. Что такое «технологичность шрифта»?
23. Что такое «экономичность шрифта»?
24. Что понимается под удобочитаемостью шрифта»?
25. Дайте характеристику компьютерным шрифтам.
26. Перечислите виды компьютерных шрифтов.
27. Каков алгоритм оценки меры неопределенности текстовых сообщений (текста)?
28. Что такое «избыточность информации»?
29. Перечислите виды информации, содержащейся в текстах.
30. Охарактеризуйте символьную информацию текстового сообщения.
31. Охарактеризуйте синтаксическую информацию текстового сообщения.
32. Охарактеризуйте семантическую информацию текстового сообщения.
33. Какой вид информации называется Шенноновской?
34. Что такое «иллюстрационная информация»?
35. Что называют растрированием изображения?
36. Что такое «линиатура растра»?
37. Дайте определение коду.
38. Что называется кодированием?
39. Что называется декодированием?
40. Что такое «эффективное кодирование»?
41. Назовите пути борьбы с помехами при кодировании.
42. Перечислите основные этапы технологического процесса полиграфического производства.
43. Что называется редактированием?
44. Что называется корректурой текста?
45. На какие группы делится набор текста по степени сложности?
46. Какие требования предъявляются к полиграфическому воспроизведению текста.
47. Перечислите единицы объема
48. Перечислите основные этапы редакционно-издательской работы с изданием.
49. Поясните алгоритмы формирования строки и переноса слов в процессе автоматизированной верстки текста.
50. Какое программное обеспечение используется для автоматического чтения и распознавания текста?
51. Какое программное обеспечение используется для верстки издания?
52. Дайте характеристику ошибок, возникающих в процессе работы с текстом.

53. Дайте характеристику дискретным марковским процессам.
54. Что такое «вектор состояния текста» с точки зрения модели?
55. Что такое «матрица перехода текста» с точки зрения модели этапа технологического процесса?
56. Что такое «стохастическая матрица»?
57. Какое состояние называется поглощающим?
58. Запишите типовую идеальную матрицу кодирования текста (набора, чтения).
59. Запишите типовую идеальную матрицу корректуры текста.
60. Какая матрица называется треугольной? Моделью какого этапа технологического процесса она может являться?
61. Какая матрица называется единичной? Моделью какого этапа технологического процесса она может являться?

П2.5.2. Примерные варианты задания для контрольных работ по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

1. Найти основные характеристики случайной величины *X*, которая принимает различные целочисленные значения и характеризуется исходами x_i, появляющимися с относительной частотой (вероятностью) P_i:

Среднее значение (математическое ожидание),

дисперсию,

среднеквадратическое отклонение.

x_i	-1	0	1	2
P_i	0,1	0,7	0,1	0,1

2. Даны варианты относительных частот наступления пяти событий. Определить событие с максимальной энтропией.

Первое событие - $P_0=0;\ P_1=0{,}1;\ P_2=0;\ P_3=0{,}9$

Второе событие - $P_0=0;\ P_1=0{,}2;\ P_2=0;\ P_3=0{,}8$

Третье событие - $P_0=0;\ P_1=0{,}4;\ P_2=0;\ P_3=0{,}6$

Четвертое событие - $P_0=0{,}3;\ P_1=0{,}7;\ P_2=0;\ P_3=0$

Пятое событие - $P_0=0{,}5;\ P_1=0;\ P_2=0{,}5;\ P_3=0$

3. Найти совместную и условные энтропии случайных сигналов X и Y некоторого источника, совместные относительные частоты которых приведены в таблице. Оценить взаимозависимость этих сигналов.

вероятности	*x1*	*x2*	*x3*	*x4*
y1	0,24	0,16	0,06	0,05
y2	0,03	0,11	0,04	0,01
y3	0,05	0,05	0,1	0,1

4. Создать модель процесса переработки текстовой информации на этапе ввода информации (набор, сканирование, чтение и распознавание). Найти элементы матрицы (математической модели) процесса преобразования количества ошибок оператором

компьютерного набора при допущении о пуассоновском распределении процесса накопления ошибок. Среднее значение числа ошибок равно mx = 3.

5. Создать модель процесса переработки текстовой информации на этапе корректуры ошибок при условии, что оператор является абсолютно внимательным и следует отметкам корректора. Вероятность обнаружения одной единственной ошибки корректором составляет 0,9.

6. Произвести анализ качества переработки текстовой информации в соответствии с технологической последовательностью операций:

Авторский оригинал, имеющий следующие вектор наличия ошибок в тексте авторского оригинала Равт=$\|0.05, 0.5, 0.45\|$.

Набор текста оператором компьютерного набора со средним количеством ошибок 2 на 2,5 тыс. знаков.

Распечатка на лазерном принтере, при условии отсутствия технических сбоев и сбоев программного обеспечения.

Корректура выполняется при условии, что оператор является абсолютно внимательным и следует отметкам корректора. Вероятность обнаружения одной единственной ошибки корректором составляет 0,95.

Рассчитать необходимое количество последовательных корректур, при условии, что допустимое количество ошибок – не более 2-х на 5 тыс. знаков.

П2.5.3 Образцы тестовых заданий

1. Дополните

Мера неопределенности результатов наблюдения называется ...

Правильные варианты ответа: энтропия; энтропией;

2. Дополните

Единица измерения энтропии называется ...

Правильные варианты ответа: бит; битом;

3. Наука о связи, переработке информации и управлении в технических системах, живой природе и обществе называется ...

Правильные варианты ответа: кибернетика; кибернетикой;

4. Наука о связи, переработке информации и управлении в технических системах называется ...

Правильные варианты ответа: техническая кибернетика; технической кибернетикой;

5. Расположите события в порядке убывания энтропии

1 $P_i = \frac{1}{8}$

2 $P_i = \frac{1}{2}$

3 $P_i = \frac{1}{6}$

4 $P_i = \frac{1}{7}$

5 $P_i = \frac{1}{4}$

*Правильные варианты ответа:*1; 4; 3;5;2

6. Дополните

Зависимость $H(x,y) = H(x) + H(y)$ справедлива для ... событий,

где $H(x,y)$ – совместная энтропия,

$H(x)$, $H(y)$ – энтропия наступления каждого из событий.

Правильные варианты ответа: независим#$#;

7. Дополните

Зависимость $H(x)>H_y(x)$, $H(y)>H_x(y)$ справедлива для … событий

Правильные варианты ответа: зависим#$#;

8. Отметьте правильный(е) ответ(ы)

Совместная энтропия $H(x, y)$ это ...

1. $H(y)-H_y(x)$
2. $H(x)-H_x(y)$
3. $H(x)+H(y)$
4. $H(x)+H_x(y)$
5. $H(y)+H_y(x)$

Правильные варианты ответа: 4,5

9. Если вероятности событий равны

$P_0 = 0$, $P_1 = 1$, $P_2 = 0$, $P_3 = 0$
Энтропия равна….
Правильные варианты ответа: нулю; 0; нолю; ноль; нуль;

10.Вероятности событий равны

$P_0 = 0.5$, $P_1 = 0.5$
Энтропия равна…
Правильные варианты ответа: единице; 1; одному;

11.Вероятности событий равны

$P_0=0.25$, $P_1=0.25$, $P_2=0.25$, $P_3=0.25$

Энтропия равна…

Правильные варианты ответа: двум; 2; два;

12.Вероятности событий равны

$P_0=0.125$, $P_1=0.125$, $P_2=0.125$, $P_3=0.125$,

$P_4=0.125$, $P_5=0.125$, $P_6=0.125$, $P_7=0.125$

Энтропия равна…

Правильные варианты ответа: 3; трем;

13. Вероятности событий равны

$P_i = \frac{1}{8}$

Энтропия равна…

Правильные варианты ответа: 3; трем; 3.0; 3,0

14. Вероятности событий равны $Pi = \frac{1}{16}$

Энтропия равна…

Правильные варианты ответа: четырем; четыре; 4;

15. Вероятности событий равны $Pi = \frac{1}{32}$

Энтропия равна…

Правильные варианты ответа: пяти; 5; пять;

16. Вероятности событий равны $Pi = \frac{1}{64}$

Энтропия равна…

Правильные варианты ответа: 6; шести; шесть;

17. Вероятности событий равны $Pi = \frac{1}{128}$

Энтропия равна…

Правильные варианты ответа: 7; семи; семь;

Дополните

18. Вероятности событий равны $Pi = \frac{1}{256}$

Энтропия равна…

Правильные варианты ответа: 8; восемь; восьми;

19. Заданы равновероятные события. Расположите номера событий в порядке возрастания энтропии

1. $P_i = \frac{1}{8}$

2. $P_i = \frac{1}{2}$

3. $P_i = \frac{1}{6}$

4. $P_i = \frac{1}{7}$

5. $P_i = \frac{1}{4}$

Правильные варианты ответа:2; 5; 3;4;1

20.Отметьте событие с максимальной энтропией, если заданы равновероятные события

1. $P_i = \frac{1}{2}$

2. $P_i = \frac{1}{4}$

3. $P_i = \frac{1}{8}$

4. $P_i = \frac{1}{7}$

5. $P_i = \frac{1}{6}$

Правильные варианты ответа:3

21.Отметьте событие с минимальной энтропией, если заданы равновероятные события

1. $P_i = \frac{1}{7}$

2. $P_i = \frac{1}{6}$

3. $P_i = \frac{1}{8}$

4. $P_i = \frac{1}{2}$

5. $P_i = \frac{1}{4}$

Правильные варианты ответа: 4

22. Энтропия события с вероятностью $P_j = \frac{1}{256}$ больше энтропии события с вероятностью $P_i = \frac{1}{4}$ в … раза.

Правильные варианты ответа: четыре; 4; 4.0; 4,0

23. Энтропия события с вероятностью $P_j = \frac{1}{64}$ больше энтропии события с вероятностью $P_i = \frac{1}{2}$ в … раза.

Правильные варианты ответа: шесть; 6; 6.0; 6,0

24. Энтропия события с вероятностью $P_j = \frac{1}{64}$ больше энтропии события с вероятностью $P_i = \frac{1}{4}$ в … раза.

Правильные варианты ответа: три; 3;

25. Отметьте сигнал с максимальной энтропией, если задано пять случайных сигналов с дисперсиями

1. $D_x = 0.5$
2. $D_x = 0.4$
3. $D_x = 0.6$
4. $D_x = 0.1$
5. $D_x = 0.2$

Правильные варианты ответа: 3

26. Отметьте сигнал с минимальной энтропией, если задано пять случайных сигналов с дисперсиями

1. $D_x = 0.5$
2. $D_x = 0.4$
3. $D_x = 0.6$
4. $D_x = 0.1$
5. $D_x = 0.2$

Правильные варианты ответа: 4

27. Дополните

$H(x) = \log_2 2 = \ldots$ бит.

Правильные варианты ответа: 1; 1.0; 1,0

28. Дополните

$H(x) = \log_2 4 = \ldots$ бит(а).

Правильные варианты ответа: 2; 2.0; 2,0

29. Дополните

$H(x) = \log_2 8 = \ldots$ бит(а).

Правильные варианты ответа: 3; 3.0; 3,0

30.Дополните

$H(x)=\log_2 16=...$бит(а).

Правильные варианты ответа: 4;

31.Дополните

$H(x)=\log_2 32=...$бит(а).

Правильные варианты ответа: 5;

32.Дополните

$H(x)=\log_2 64=...$бит(а).

Правильные варианты ответа: 6;

33.Дополните

$H(x)=\log_2 128=...$бит(а).

Правильные варианты ответа: 7;

34. Графическая форма знаков алфавитной системы письма называется:

Текстом

Шрифтом

Кеглем

Интерлиньяжем

Гарнитурой

Правильные варианты ответа: шрифт, шрифтом.

35. Установите соответствие между названием кегля шрифта и его размером:

1.	Цицеро	A.	6 пунктов
2.	Текст	B.	8 пунктов
3.	Корпус	C.	10 пунктов
4.	Петит	D.	12 пунктов
5.	Нонпарель	E.	20 пунктов

Правильные варианты ответа: 1 - D

2 - E

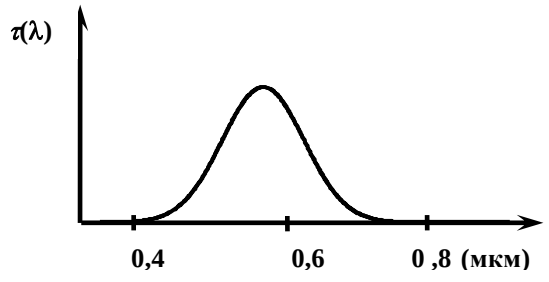
3 - C

4 - B

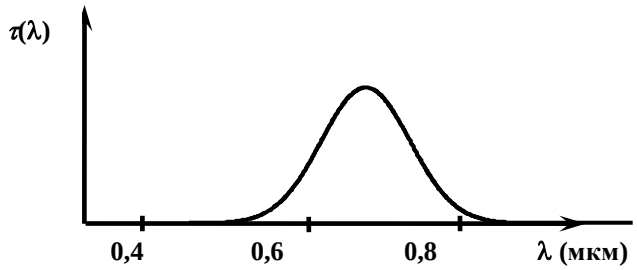
5 - A

36.Спектру излучения, изображенному на рисунке соответствует ... цвет

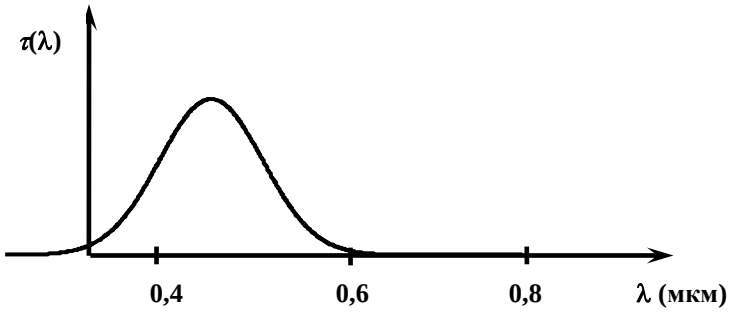
Правильные варианты ответа: красный;



37.Спектру излучения, изображенному на рисунке соответствует ... цвет



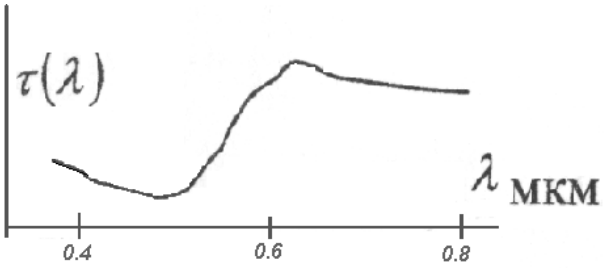
Правильные варианты ответа: зеленый;



38.Спектру излучения, изображенному на рисунке соответствует ... цвет

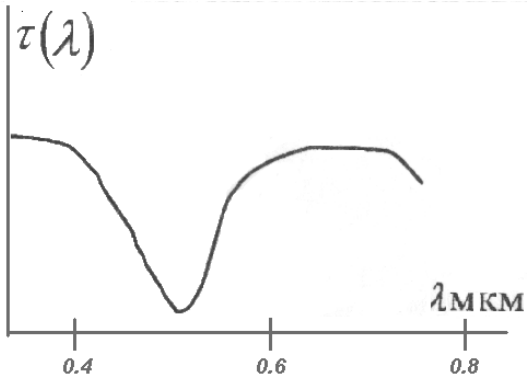
Правильные варианты ответа: синий;

39.На графике изображен спектр излучения ... краски



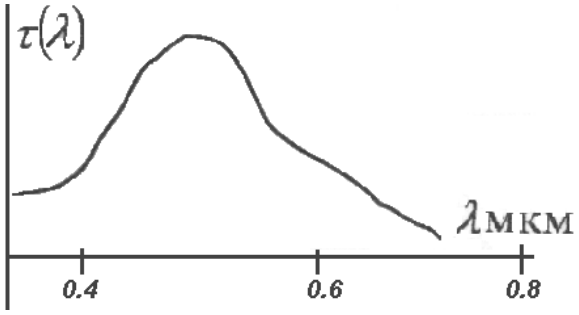
Правильные варианты ответа: желтый;

40.На графике изображен спектр излучения ... краски



Правильные варианты ответа: пурпурной; пурпурная;

41. На графике изображен спектр излучения ... краски



Правильные варианты ответа: голубой; голубая;

П2.5.4. Примерная тематика докладов и сообщений по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

Сравнительная характеристика программных средств чтения и распознавания текстовой информации, использующих технологию OCR. Изучение программного пакета ABBYY Finereader.

Сравнительная характеристика программных средств чтения и распознавания текстовой информации, использующих технологию OCR. Изучение программного пакета OmniPage.

Сравнительная характеристика программных средств чтения и распознавания текстовой информации, использующих технологию OCR. Изучение программного пакета Readiris Pro.

Сравнительная характеристика программных средств чтения и распознавания текстовой информации, использующих технологию OCR. Изучение программного пакета OCR Cuneiform.

Сравнительная характеристика программных средств набора и верстки текста. Особенности программного пакета Page Maker.

Сравнительная характеристика программных средств набора и верстки текста. Особенности программного пакета Adobe InDesign.

Сравнительная характеристика программных средств набора и верстки текста. Особенности программного пакета QuarkXPress.

Сравнительная характеристика программных средств набора и верстки текста. Особенности программного пакета CorelDRAW Graphics.

П2.5.5 Пример задания по лабораторной работе по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

Частотный анализ фрагмента литературного текста

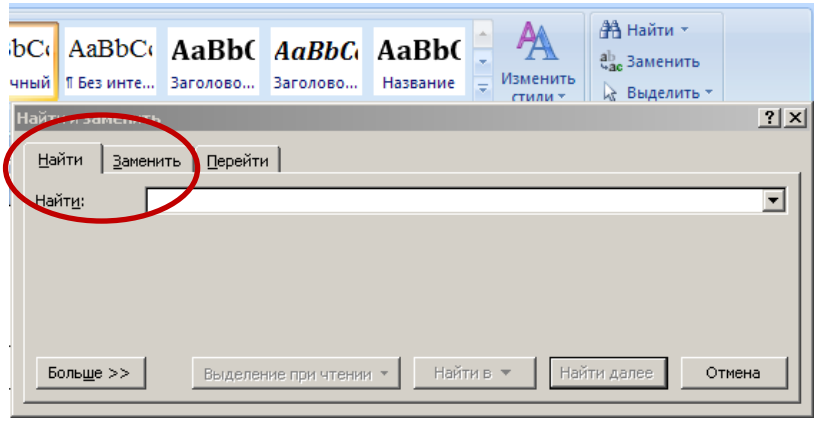
2. Распечатать исходный текст по индивидуальному варианту (из соответствующего файла ТОПТИАС__35).
3. Предложенный текст адаптировать для частотного анализа:
   - удалить колонтитул,
   - удалить знаки препинания,
   - удалить цифры,

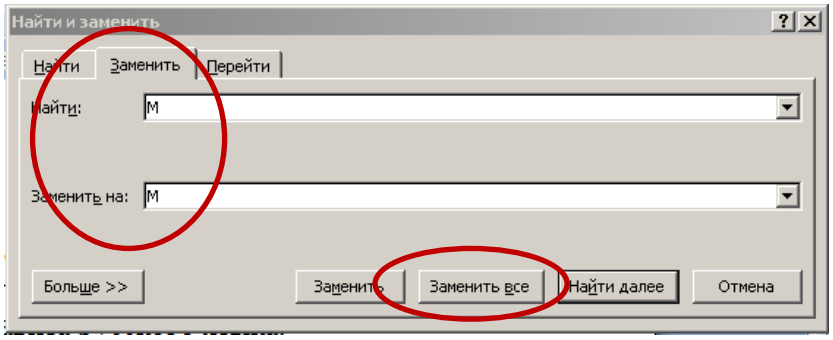
- оставить между словом и частицей вместо дефиса один пробел,
- удалить абзацные отступы,
- удалить двойные пробелы.

4. Удаление знаков производится при помощи поисковой системы Microsoft Word в закладке меню «ГЛАВНАЯ» «Найти – Заменить»:
   «Найти **НАБРАТЬ ЗНАК** – Заменить на **ПУСТОЕ МЕСТО**»

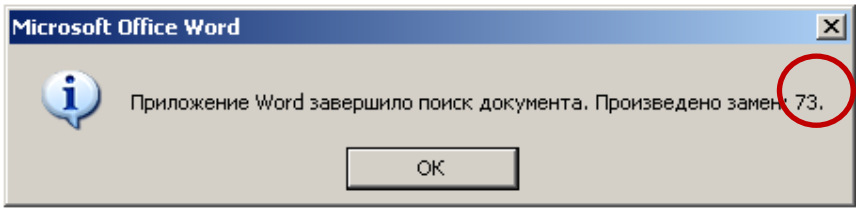
5. Распечатать адаптированный текст и сохранить текст в новый файл, например,
   ТОПТИАС_35-1. Эта операция выполняется при помощи закладки в строке меню «ФАЙЛ»→ «Сохранить как» → Указать из раскрывающегося списка папку или место, куда выполняется сохранение.

6. Выполнить частотный анализ адаптированного текстового фрагмента литературного произведения при помощи поисковой системы Microsoft Word в закладке меню «ГЛАВНАЯ» «Найти – Заменить», например
   «Найти **М** – Заменить на **М**»





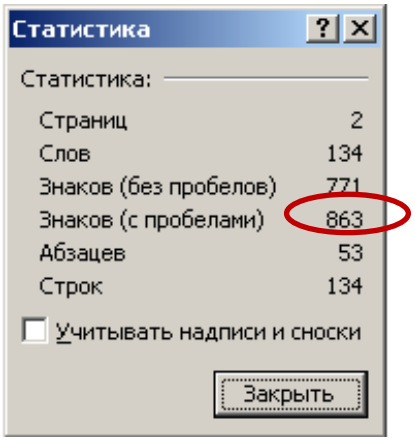
Нажать на кнопку «заменить все»:



В некоторых версиях Microsoft Word сразу при поиске выдается количество найденных символов.

7. Полученное значение записать в ячейку с буквой «Абсолютная частота ni». Выполнить такую операцию для каждой буквы и пробела.

8. Проверить общее количество знаков можно при помощи закладки «Статистика», оно должно совпасть с ячейкой СУММА Σni= (**) в столбце «Абсолютная частота» таблицы.

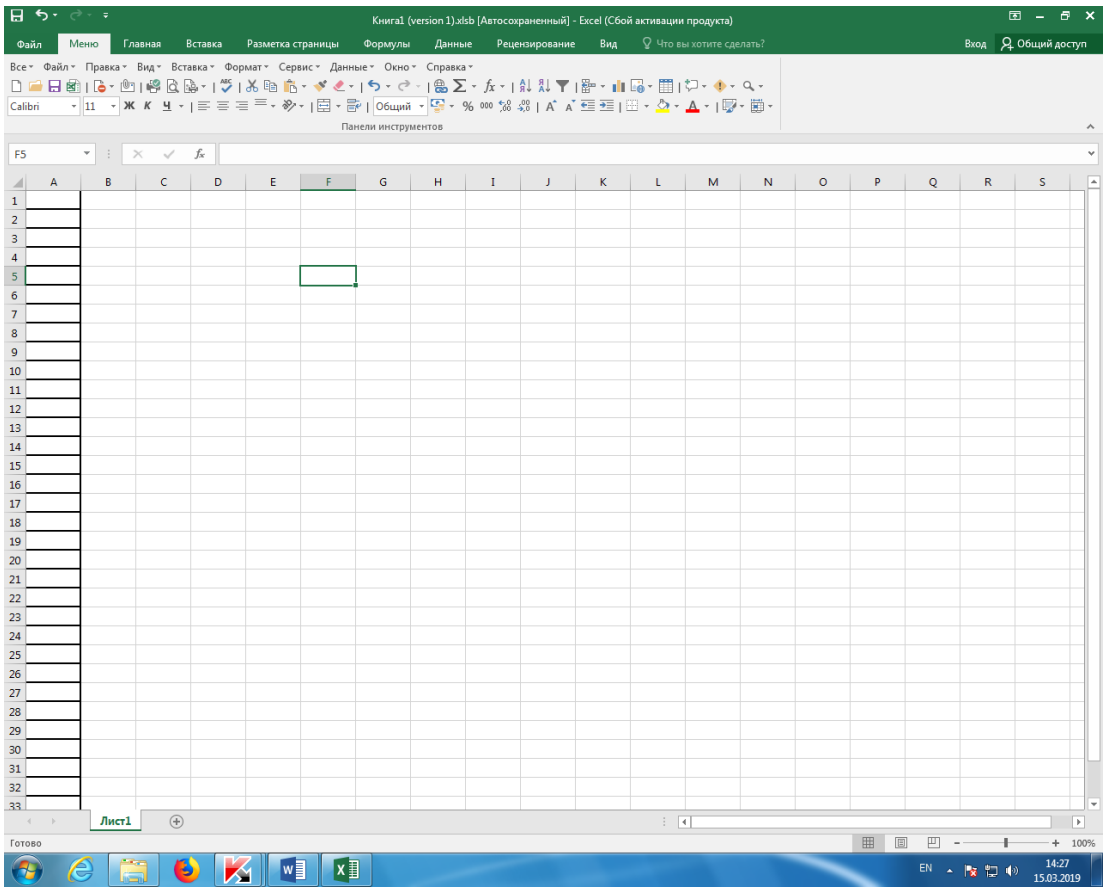


9. Выполнить расчет относительной частоты употребления каждого знака в тексте wi=ni/ Σni . Оценить частоту ni, относительную частоту wi употребления знаков алфавита (округлить по математическим правилам до тысячных долей 0,000), набранного буквами кириллического алфавита русского языка
10. Полученное значение записать в ячейку с буквой в столбце «Относительная частота wi».
11. Проверить корректность выполненных расчетов – близость суммарных значений относительной частоты Σwi к 1.
12. Рассчитать экспериментальное значение энтропии hi= - wi·log 2 wi для каждого знака*.
13. Рассчитать суммарное значение энтропии H= Σhi для варианта текстового фрагмента.
14. Результаты анализа внести в таблицу и письменно СДЕЛАТЬ ВЫВОДЫ о ЧАСТОТЕ употребления знаков. Сравнить с частотой употребления знаков (букв) в первом контрольном задании (Контрольное задание_1_ООТАСП_).

Знак (буква)	Абсолютная частота *n*i	Относительная частота *w*i=*n*i/ Σ*n*i	Экспериментальное значение энтропии *h*i= - *w*i·log $_2$ *w*i
Аа			
Бб			
Вв			
Гг			
Дд			
Ее, Ёё			
Жж			
Зз			
Ии			
Йй			
Кк			
Лл			
Мм			
Нн			

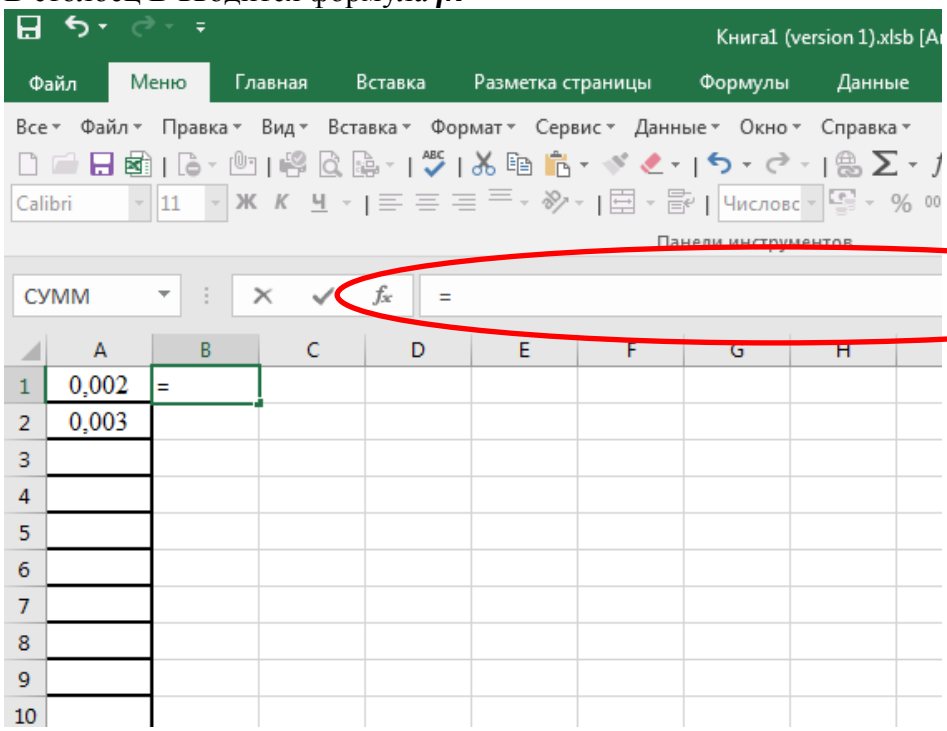
Знак (буква)	Абсолютная частота ni	Относительная частота wi=ni/ Σni	Экспериментальное значение энтропии hi= - wi·log $_2$ wi
Оо			
Пп			
Рр			
Сс			
Тт			
Уу			
Фф			
Хх			
Цц			
Чч			
Шш			
Щщ			
Ьь			
Ыы			
Ъъ			
Ээ			
Юю			
Яя			
пробел			
СУММА	Σni= **(**)**	Σwi=	H= Σhi=

15. Оформить отчет на листах формата А4 с титульным листом, предоставить к защите и оцениванию в распечатанном виде.
16. Расчет энтропии целесообразно выполнить при помощи табличного редактора Microsoft Excel. Создать файл для расчета можно через меню «ПУСК»



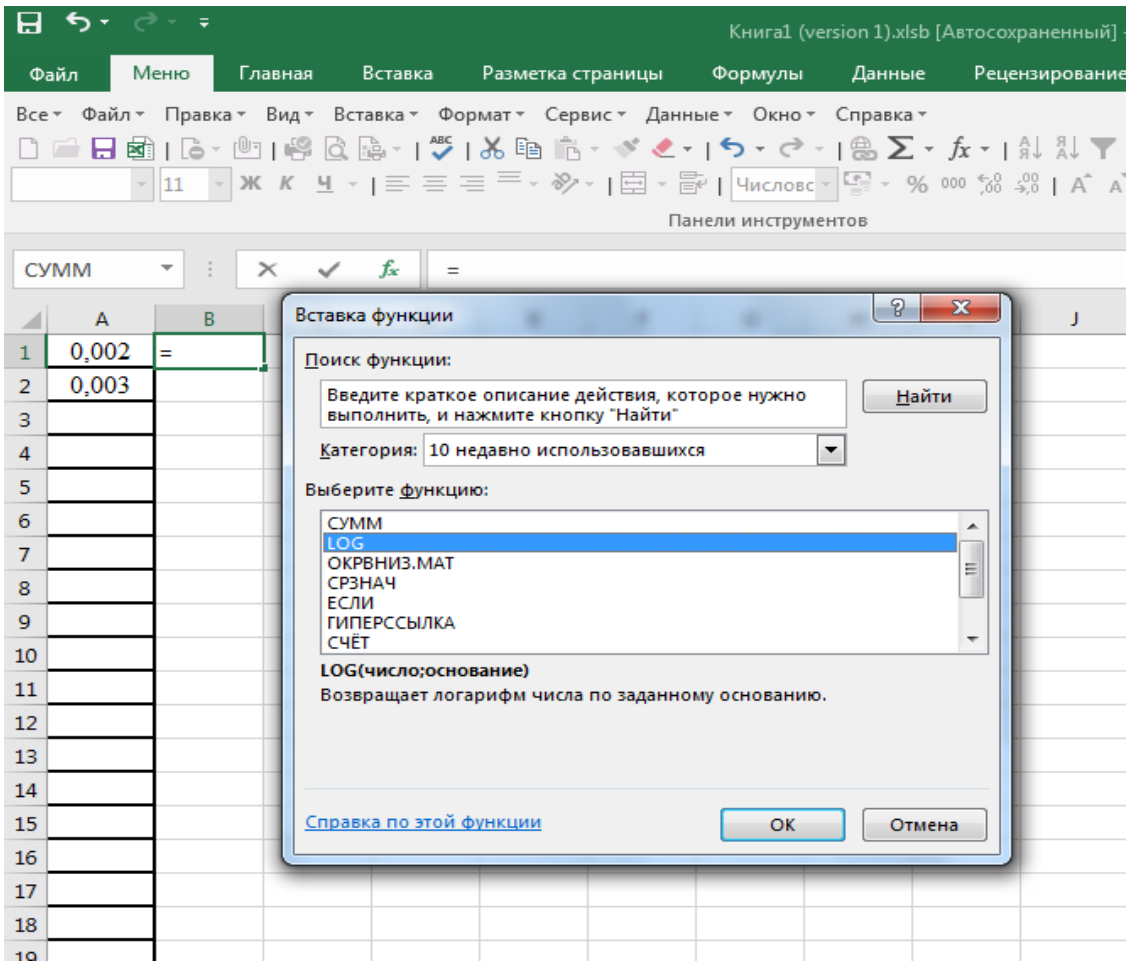
В столбец А вносятся данные из таблицы «Относительная частота» из файла Microsoft Word.

В столбец В вводится формула ***fx***

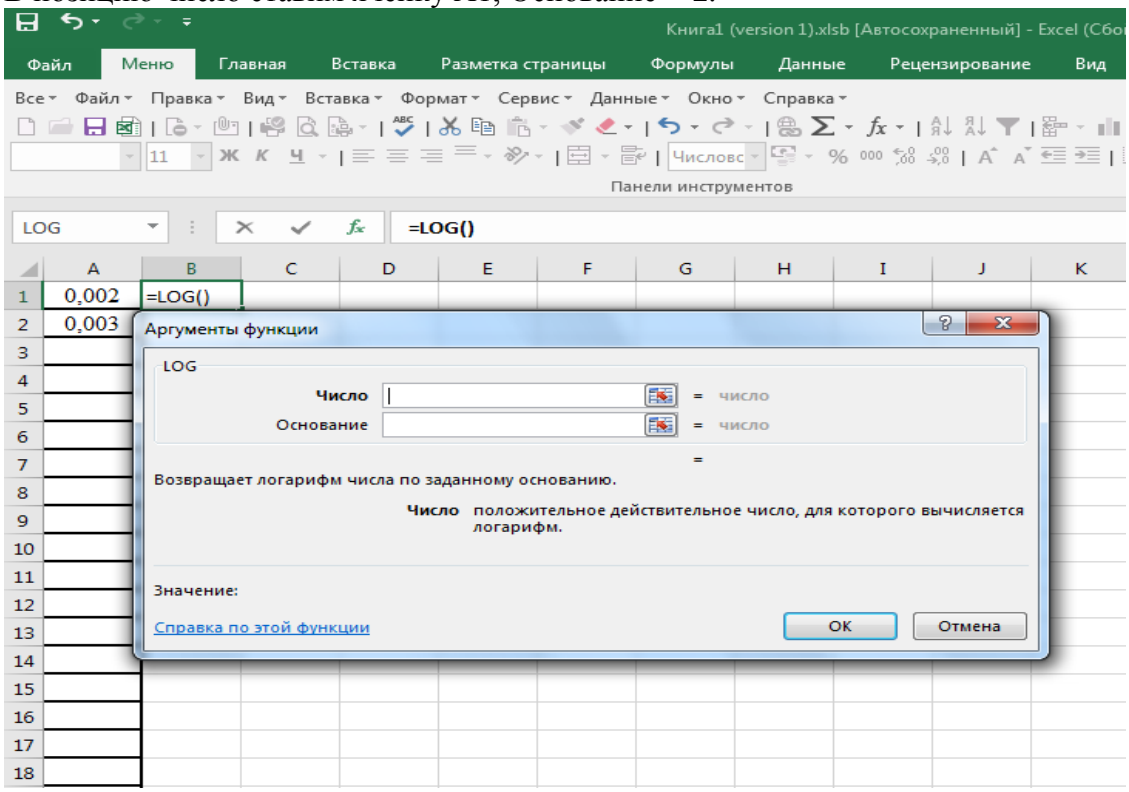


Курсор ставится в позицию В1, в строке набора текста или формул набирается знак равенства (признак дальнейшего ввода формулы);

Выбрать из списка формул *fx* в позиции «математические» LOG();



В позицию число ставим ячейку A1, Основание - 2.



Далее копируется формула из ячейки B1 в оставшиеся ячейки столбца B.

Полученные расчетные данные переносятся копированием в таблицу файла Microsoft Word столбец «Экспериментальное значение энтропии».

В нижней ячейке столбца рассчитывается суммарная энтропия H по фрагменту текста.

П2.5.6 Перечень экзаменационных вопросов по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах»

1. Общие сведения об информации и информатике.
2. Сведения об информации и ее видах.
3. Информационные основы переработки текста. Понятие текста.
4. Информационные основы переработки изображений. Свойства изобразительной информации.
5. Понятие принтмедиаиндустрии. Изобретения, ознаменовавшие важнейшие этапы совершенствования и создания средств приема, передачи и хранения информации. Массовые средства письменного и иного хранения информации.
6. Формы информации, содержащейся в текстах. Прогнозирование появления букв и слов в текстах. Избыточность текста.
7. Формы информации, содержащейся в текстах. Символьная, синтаксическая и семантическая информация.
8. Понятие энтропии. Энтропия текстовых сообщений
9. Роль статистики языка при обработке текстов. Количество символьной информации в текстах.
10. Свойства шрифтов. Общая классификация шрифтов.
11. Свойства шрифтов. Понятие кегля, гарнитуры, начертания.
12. Единицы типографской системы мер, применяемые для обозначения кегля.
13. Классификация типографских шрифтов.
14. Шрифты, основные требования, предъявляемые к ним.
15. Шрифты, понятие трекинга и кернинга при наборе.
16. Шрифты, понятие удобочитаемости, технологичности, экономичности.
17. Единицы учета объема текстов. Понятие формата издания.
18. Понятие группы сложности набора текста.
19. Схемы движения заказа на допечатной стадии технологического процесса подготовки издания.
20. Этапы движения заказа на допечатной стадии технологического процесса подготовки издания. Понятие редактирования.
21. Этапы движения заказа на допечатной стадии технологического процесса подготовки издания. Корректура.
22. Общая характеристика программных средств набора и верстки.
23. Общая характеристика программных средств чтения и распознавания текстового материала.
24. Показатели качества печатной продукции.
25. Ошибки как показатель качества печатной продукции. Источники и виды ошибок в тексте.
26. Виды ошибок в тексте. Приблизительная оценка ошибок по грамматическим категориям.
27. Переработка текстовой информации с точки зрения дискретного марковского процесса преобразования количества ошибок в тексте. Дискретная матричная модель.

28. Переработка текстовой информации с точки зрения дискретного марковского процесса преобразования количества ошибок в тексте. Общие сведения о дискретных марковских процессах.
29. Переработка текстовой информации с точки зрения дискретного марковского процесса преобразования количества ошибок в тексте. Преобразование текста при его вводе в компьютер или наборе.
30. Переработка текстовой информации с точки зрения дискретного марковского процесса преобразования количества ошибок в тексте. Преобразование текста при корректуре.

П2.5.7 Образец экзаменационного билета по дисциплине «Теоретические основы переработки текстовой информации в автоматизированных системах».

министерство науки и высшего образования российской федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(МОСКОВСКИЙ ПОЛИТЕХ)

Институт *полиграфический* Кафедра *ПС*
Дисциплина *Теоретические основы переработки текста в автоматизированных системах*
Направление подготовки *22.03.01 «Материаловедение и технологии материалов»*
курс *1* группа *221-741* Форма обучения *очная*

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № ____

1. Формы информации, содержащейся в текстах. Символьная, синтаксическая и семантическая информация

2. Общая характеристика программных средств чтения и распознавания текстового материала.

3. Найти совместную, условную энтропии случайных сигналов X и Y, совместные вероятности которых приведены в таблице, сделать вывод о взаимозависимости сигналов X и Y .

вероятности	*x1*	*x2*	*x3*	*x4*
y1	0,24	0,16	0,06	0,05
y2	0,03	0,11	0,04	0,01
y3	0,05	0,05	0,1	0,1

Утверждено на заседании кафедры ПС

«_ __ _» _____20___ г., протокол № _____

Зав. кафедрой _____ /Суслов М.В./