Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гриб Владислав Валерьевич
Должность: Ректор
Дата подписания: 19.11.2026 07:23:39
Уникальный программный ключ:
637517d24e103c3db032acf37e839d98ec1c5bb2f5eb89c29abfcd7f43985447

Образовательное частное учреждение высшего образования
«МОСКОВСКИЙ УНИВЕРСИТЕТ ИМЕНИ А.С. ГРИБОЕДОВА»
(ИМПЭ им. А.С. Грибоедова)

МЕЖДУНАРОДНЫЙ ИНСТИТУТ ИНОСТРАННЫХ ЯЗЫКОВ
И МЕЖКУЛЬТУРНЫХ КОММУНИКАЦИЙ

УТВЕРЖДЕНО:
Директор международного института иностранных языков и межкультурных коммуникаций
_____ /Л.В. Парамонова/
«11» сентября 2026 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

АВТОМАТИЗИРОВАННАЯ ОБРАБОТКА ТЕКСТОВЫХ МАССИВОВ

по направлению подготовки:
45.03.03 Фундаментальная и прикладная лингвистика
(уровень бакалавриата)

профилю:
«Лингвистика в современной информационно-коммуникационной цифровой среде»

форма обучения: **очная**

Москва

Фонд оценочных средств для дисциплины «Автоматизированная обработка текстовых массивов». Направление подготовки 45.03.03 Фундаментальная и прикладная лингвистика, профиль: «Лингвистика в современной информационно-коммуникационной цифровой среде» / Н.В. Груздева – М.: ИМПЭ им. А.С. Грибоедова.

Фонд оценочных средств является неотъемлемой частью рабочей программы дисциплины.

Разработчик: Н.В. Груздева

Заведующий кафедрой: _____/Д.В. Парамонова

1. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Настоящий Фонд оценочных средств (ФОС) по дисциплине «Автоматизированная обработка текстовых массивов» является неотъемлемым приложением к рабочей программе дисциплины (РПД) «Автоматизированная обработка текстовых массивов». На данный ФОС распространяются все реквизиты утверждения, представленные в РПД по данной дисциплине.

2. ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ

Для определения качества освоения обучающимися учебного материала по дисциплине используются следующие оценочные средства:

№ п/п	Оценочное средство	Краткая характеристика оценочного средства	Представление оценочного средства в ФОС
1	Тестирование	Вид контроля, позволяющий оценить изученный теоретический материал.	Вопросы для проведения тестирования
2	Практические задания	Вид контроля, позволяющий оценить умение обучающегося применять осваиваемую компетенцию в практических ситуациях и при решении производственных задач	Задания к практическому занятию
3	Контрольная работа	Вид контроля, позволяющий определить результат освоения компетенций по дисциплине в рамках рассматриваемой темы, оцениваемый с помощью соответствующих индикаторов достижения компетенций	Задания контрольной работы
4	Самостоятельная работа	Вид контроля, позволяющий оценить проработку теоретического материала, изучение рекомендуемой литературы, выполнение практико-ориентированных заданий (заполнение таблиц, проведение сравнительного анализа, составление схем и др.), решение практических задач, создание презентаций, написание рефератов, подборку нормативного и иного материала и выполнение других заданий	Задания самостоятельной работы
5	Курсовая работа	Вид контроля, позволяющий выявить степень владения базовыми знаниями, умениями и навыками, необходимыми для обучения, и определить уровень владения новым материалом	Индивидуальные задания (темы) для курсовой работы
6	Зачет/Зачет с оценкой/Экзамен	Вид контроля, позволяющий выявить степень овладения знаниями, умениями и навыками,	Вопросы для подготовки к

		необходимыми для дальнейшего освоения образовательной программы подготовки	зачету/зачету с оценкой/экзамену

3. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ДИСЦИПЛИНЕ

3.1. Сопроводительная информация.

Разработчик	Н.В. Груздева
Кафедра	Кафедра европейских языков и межкультурной коммуникации
Наименование дисциплины	Автоматизированная обработка текстовых массивов
Факультет / институт	Международного института иностранных языков и межкультурных коммуникаций
Направление подготовки / специальность	45.03.03 Фундаментальная и прикладная лингвистика
Количество вопросов в оценочных заданиях (диапазон)	45-90
Общее время тестирования (мин)	90
Общее количество вопросов/заданий в ФОС	90
Размещенность на сайте Университета примерного перечня вопросов, заданий ФОС – для подготовки обучающихся к прохождению оценки (да / нет)	да

3.2. Характеристика оцениваемых компетенций.

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
ОПК-5	Способен создавать и редактировать тексты профессионального назначения	**ИОПК-5.1. Знает** специфику создания и редактирования профессиональных текстов на иностранном языке, нормы межкультурной коммуникации и требования к переводным и оригинальным иноязычным материалам. **ИОПК-5.2. Умеет** создавать, переводить и редактировать профессиональные тексты на иностранном языке, обеспечивая их эквивалентность, аутентичность и соответствие целевой аудитории. **ИОПК-5.3. Владеет** технологиями профессионального иноязычного письма, перевода и редактирования, включая работу со специализированными словарями и электронными базами данных.

6. МАТЕРИАЛЫ ДЛЯ ДИАГНОСТИЧЕСКОЙ РАБОТЫ

Задания для диагностической работы должны обеспечивать оценку полностью или частично сформированных компетенций. Каждое задание должно быть привязано к тому или иному индикатору сформированности компетенций.

При формировании заданий для диагностической работы необходимо использовать тестовые задания следующих типов:

Тип задания 1. Задания закрытого типа на установление соответствия.

Тип задания 2. Задания закрытого типа на установление последовательности.

Тип задания 3. Задания комбинированного типа, предполагающие выбор одного правильного Ответа из предложенных с последующим объяснением своего выбора.

Тип задания 4. Задания комбинированного типа, предполагающие выбор нескольких Ответов из предложенных с последующим объяснением своего выбора.

Тип задания 5. Расчетные задачи.

Все типы заданий должны быть представлены не менее одного раза.

<table>
<tr><th rowspan="2">№ п/п</th><th rowspan="2">Тема занятия</th><th rowspan="2">Код компетенции</th><th rowspan="2">Индикатор</th><th rowspan="2">Тип задания</th><th colspan="2">Задание</th></tr>
<tr><th>Вариант 1</th><th>Вариант 2</th></tr>
<tr><td>1.</td><td>Тема 1. Текстовый массив как объект автоматизированной обработки</td><td>ОПК-5 Способен создавать и редактировать тексты профессионального назначения</td><td>ИОПК-5.1. Знает специфику создания и редактирования профессиональных текстов на иностранном языке, нормы межкультурной коммуникации и требования к переводным и оригинальным иноязычным материалам.</td><td>1</td><td>Задание 1. Установите соответствие между форматом файла и его характеристикой
<table><tr><td>А) txt</td><td>1. Формат с жесткой фиксацией верстки, сложный для извлечения чистого текста.</td></tr><tr><td>Б) xml</td><td>2. Простой текстовый файл без форматирования, идеальный для первичной обработки</td></tr><tr><td>В) pdf</td><td>3. Формат на основе языка разметки, хранящий и данные, и их структуру</td></tr><tr><td>Г) docx</td><td>4. Современный формат текстового процессора, представляющий собой архив XML-файлов</td></tr></table>Ответ: А-2, Б-3, В-1, Г-4.</td><td>Задание 1. Установите соответствие между форматом файла и его характеристикой
<table><tr><td>А) csv</td><td>1. Формат для обмена табличными данными, где значения разделены запятыми</td></tr><tr><td>Б) html</td><td>2. Язык разметки для создания веб-страниц, требующий очистки от тегов</td></tr><tr><td>В) json</td><td>3. Легковесный формат обмена данными, основанный на парах «ключ-значение</td></tr><tr><td>Г) rtf</td><td>4. Кроссплатформенный формат, сохраняющий базовое форматирование текста</td></tr></table>Ответ: А-1, Б-2, В-3, Г-4</td></tr>
</table>

				3	*Задание 2. Что такое «токенизация» в контексте обработки текстовых массивов?* Ответ: Процесс разбиения сплошного текста на минимальные значимые единицы (токены), обычно слова и знаки препинания. Объяснение: это базовый шаг, без которого компьютер не может анализировать структуру текста.	*Задание 2. Что такое «лемматизация»?* Ответ: Приведение словоформы к ее начальной словарной форме (лемме) с учетом контекста и части речи. Объяснение: в отличие от простого отсечения окончаний, лемматизация дает корректное словарное слово (например, «бежал» → «бежать»).
2.	Тема 2. Терминология и лексикографические базы данных	ОПК-5 Способен создавать и редактировать тексты профессионального назначения	ИОПК-5.3. Владеет технологиями профессионального иноязычного письма, перевода и редактирования, включая работу со специализированными словарями и электронными базами данных.	3	*Задание 3. Почему для обмена терминологией между разными системами предпочтительнее использовать формат TBX, а не Excel?* Ответ: TBX является строгим международным стандартом, который сохраняет иерархию, метаданные и связи между терминами, теряющиеся в плоской таблице Excel. Объяснение: это гарантирует корректный импорт без потери структуры данных.	*Задание 3. Что такое «контекст» в карточке термина и зачем он нужен переводчику?* Ответ: Пример предложения из реального текста, где используется термин. Объяснение: Контекст помогает понять точное значение термина, его грамматическое окружение и избежать ошибок многозначности.
				5	*Задание 4. В таблице глоссария 4 столбца (Термин, Перевод, Определение, Контекст). Если в глоссарии 150 терминов, сколько всего ячеек с данными нужно заполнить (не считая заголовков)?* Ответ B1: 600 (4 × 150).	*Задание 4. При автоматическом извлечении терминов программа выдала 200 кандидатов. Лингвист отбраковал 35% из них как нерелевантные. Сколько терминов осталось в финальном списке?* Ответ B2: 130 (200 - 70 = 130).

<table>
<tr>
<td rowspan="2">3.</td>
<td rowspan="2">Тема 3. Корпусные ресурсы в создании и редактировании иноязычных текстов</td>
<td rowspan="2">ОПК-5 Способен создавать и редактировать тексты профессионального назначения</td>
<td rowspan="2">ИОПК-5.2. Умеет создавать, переводить и редактировать профессиональные тексты на иностранном языке, обеспечивая их эквивалентность, аутентичность и соответствие целевой аудитории.</td>
<td>4</td>
<td>Задание 5. Выберите два типа метаданных, которые полезны для фильтрации запросов в корпусе.
А) Цвет шрифта в исходном документе,
Б) Год публикации текста,
В) Жанр или тематика текста,
Г) Имя пользователя, создавшего файл.
Ответ: Б, В. Объяснение: Год и жанр являются лингвистически значимыми параметрами, влияющими на лексику и стиль.</td>
<td>Задание 5. Выберите два преимущества использования специализированного корпуса перед общим.
А) Меньший размер файла,
Б) Высокая релевантность найденных примеров для узкой тематики,
В) Наличие только художественной литературы,
Г) Отсутствие общеупотребительной лексики, засоряющей поиск.
Ответ: Б, Г. Объяснение: Специализированный корпус отсекает шум и дает точные отраслевые соответствия.</td>
</tr>
<tr>
<td>1</td>
<td>Задание 6. Установите соответствие между типом корпуса и его описанием
<table>
<tr><td>А) Параллельный</td><td>1. Тексты, созданные изучающими иностранный язык, для анализа типичных ошибок.</td></tr>
<tr><td>Б) Монолингвальный</td><td>2. Тексты одной узкой предметной сферы (например, только медицинские или юридические)</td></tr>
<tr><td>В) Учебный (Learner)</td><td>3. Оригинальные тексты и их профессиональные переводы, выровненные по предложениям</td></tr>
<tr><td>Г) Специализированный</td><td>4. Большой массив текстов на одном языке для изучения</td></tr>
</table>
</td>
<td>Задание 6. Установите соответствие между типом корпуса и его описанием
<table>
<tr><td>А) Поиск по маске (wildcard)</td><td>1. Отображение искомого слова в центре строки с левым и правым контекстом</td></tr>
<tr><td>Б) KWIC (Key Word In Context)</td><td>2. Поиск слов с вариативным написанием (например, "электр*")</td></tr>
<tr><td>В) Частотный список</td><td>3. Ограничение выборки текстов только определенным годом или жанром</td></tr>
<tr><td>Г) Фильтрация по метаданным</td><td>4. Рейтинг слов в порядке убывания количества их</td></tr>
</table>
</td>
</tr>
</table>

					частотности и коллокаций Ответ: А-3, Б-4, В-1, Г-2	употреблений в корпусе Ответ: А-2, Б-1, В-4, Г-3
4.	Тема 4. Системы автоматизированного перевода (CAT-системы)	ОПК-5 Способен создавать и редактировать тексты профессионального назначения	ИОПК-5.2. Умеет создавать, переводить и редактировать профессиональные тексты на иностранном языке, обеспечивая их эквивалентность, аутентичность и соответствие целевой аудитории.	2	*Задание 7. Установление последовательности* Типичный рабочий процесс (workflow) в CAT-системе: 1. Контроль качества (QA), 2. Анализ файла и pre-translation, 3. Экспорт переведенного файла, 4. Собственно перевод и редактирование сегментов. Ответ: 2, 4, 1, 3.	*Задание 7. Установление последовательности* Этапы работы с памятью переводов (TM): 1. Обновление TM новыми утвержденными сегментами, 2. Поиск совпадений (match) для нового сегмента, 3. Вставка перевода из TM или ручной перевод, 4. Сохранение сегмента в локальную или облачную TM. Ответ: 2, 3, 4, 1
				3	*Задание 8. Что означает «100% match» (полное совпадение) в памяти переводов?* Ответ: Новый сегмент в точности (посимвольно) совпадает с сегментом, уже хранящимся в памяти переводов. Объяснение: это позволяет перевести сегмент мгновенно, гарантируя терминологическую и стилистическую консистентность.	*Задание 8. В чем главная цель этапа MTPE (Post-Editing)?* Ответ: Доведение машинного перевода до качества, приемлемого для конечного пользователя, за меньшее время, чем полный перевод с нуля. Объяснение: Переводчик не пишет текст, а исправляет ошибки машины, выступая в роли редактора.
5.	Тема 5. Машинный перевод и его место в профессиональном процессе	ОПК-5 Способен создавать и редактировать тексты профессионального назначения	ИОПК-5.1. Знает специфику создания и редактирования профессиональных текстов на иностранном языке, нормы межкультурной	5	*Задание 9. Переводчик делает легкое постредактирование со скоростью 1000 слов в час. Сколько часов ему потребуется, чтобы отредактировать текст объемом 4500 слов?* Ответ: 4.5 (4500 / 1000).	*Задание 9. Из 2000 слов, переведенных нейросетью, переводчик при проверке обнаружил 30 критических смысловых ошибок. Каков процент критических ошибок в тексте?* Ответ: 1.5% (30 / 2000 × 100).
				2	*Задание 10. Установление последовательности*	*Задание 10. Установление последовательности*

			коммуникации и требования к переводным и оригинальным иноязычным материалам.		Этапы процесса MTPE (Machine Translation Post-Editing): 1. Финальная проверка качества (QA), 2. Получение сырого машинного перевода, 3. Полное или легкое постредактирование текста человеком, 4. Оценка пригодности текста для МП (MT suitability assessment). Ответ: 4, 2, 3, 1.	Эволюция качества машинного перевода для сложного текста: 1. Полностью неприемлемый результат, требующий полного переписывания, 2. Результат, требующий полного пост-редактирования (Full PE), 3. Результат, требующий только легкого пост-редактирования (Light PE), 4. Результат, пригодный для публикации без правок (Raw MT). Ответ: 1, 2, 3, 4.
6.	Тема 6. Автоматизированное редактирование и контроль качества текста	ОПК-5 Способен создавать и редактировать тексты профессионального назначения	ИОПК-5.3. Владеет технологиями профессионального иноязычного письма, перевода и редактирования, включая работу со специализированными словарями и электронными базами данных.	3	*Задание 11. Что такое «несоответствие тегов» (tag mismatch) в CAT-системе и почему это критическая ошибка?* Ответ: это отсутствие, дублирование или изменение порядка служебных кодов форматирования. Объяснение: это критическая ошибка, так как она может привести к «падению» верстки, невозможности скомпилировать файл или невидимости текста в конечном продукте.	*Задание 11. В чем ограничение автоматических средств проверки правописания (spell checkers)?* Ответ: они выявляют только слова, отсутствующие в словаре, но не могут обнаружить контекстуальные ошибки (например, «компания» вместо «кампания»). Объяснение: для выявления таких ошибок нужны более сложные контекстные анализаторы или человек.
				4	*Задание 12. Выберите два преимущества использования автоматизированных QA-инструментов перед ручной проверкой.* А) Способность оценивать художественную ценность текста, Б) 100% гарантия нахождения всех смысловых ошибок, В) Высокая скорость проверки больших объемов текста,	*Задание 12. Выберите два типа ошибок, которые типичный QA-инструмент в CAT-системе может найти автоматически.* А) Неправильный выбор стилистического регистра, Б) Пропуск перевода целого сегмента (untranslated segment), В) Несовпадение чисел в источнике и переводе, Г) Неудачная игра слов.

					Г) Исключение человеческого фактора (усталости, невнимательности) при поиске формальных несовпадений. Ответ: В, Г. Объяснение: Машины идеально подходят для рутинной проверки формальных параметров (числа, теги, глоссарий), где человек часто ошибается из-за усталости.	Ответ: Б, В. Объяснение: это формальные, алгоритмизируемые проверки, которые являются основой QA-инструментов.

Критерии оценивания тестового задания:

Оценка	Критерии оценивания
отлично	от 90 до 100 % правильно выполненных заданий
хорошо	от 70 до 89 % правильно выполненных заданий
удовлетворительно	от 50 до 69 % правильно выполненных заданий
неудовлетворительно	менее 50 % правильно выполненных заданий