

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)**

ИНСТИТУТ ЛИНГВИСТИКИ
УНЦ компьютерной лингвистики

АВТОМАТИЧЕСКИЙ ПЕРЕВОД

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.03.03 Фундаментальная и прикладная лингвистика

Код и наименование направления подготовки/специальности

Фундаментальная и прикладная лингвистика

Наименование направленности (профиля)/ специализации

Уровень высшего образования: *бакалавриат*

Форма обучения: *Очная*

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2022

Автоматический перевод

Рабочая программа дисциплины

Составитель:

к. филол. н., профессор УНЦ компьютерной лингвистики Л.Л. Иомдин

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики

№_5__ от_____31.03.2022_____

ОГЛАВЛЕНИЕ

1. Пояснительная записка	4
1.1. Цель и задачи дисциплины	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	5
1.3. Место дисциплины в структуре образовательной программы	7
2. Структура дисциплины.....	7
3. Содержание дисциплины.....	7
4. Образовательные технологии	9
5. Оценка планируемых результатов обучения.....	12
5.1 Система оценивания	12
5.2 Критерии выставления оценки по дисциплине	13
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	14
6. Учебно-методическое и информационное обеспечение дисциплины.....	17
6.1 Список источников и литературы	17
7. Материально-техническое обеспечение дисциплины	17
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов.....	17
9. Методические материалы.....	18
9.1 Планы семинарских/ практических занятий	18
9.2 Иные материалы	20
Приложение 1. Аннотация рабочей программы дисциплины	22

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель дисциплины - освоение студентами базовых понятий и методов автоматического перевода, в первую очередь лингвистических понятий и методов, знакомство с историей автоматического перевода, принципов многоуровневого моделирования естественного языка, основных подходов к морфологическому, синтаксическому и семантическому анализу текстов, к соотношению грамматики и словаря.

Задачи:

- ознакомить студентов с общей проблематикой и принципиальными подходами к реализации автоматического перевода между естественными языками;
- указать студентам на место автоматического перевода в ряду задач автоматической обработки текста;
- обучить студентов осуществлять самостоятельный анализ морфологических и синтаксических характеристик текста, релевантных для выполнения автоматического перевода, а также оценивать результаты автоматической разметки;
- обучить студентов самостоятельно формулировать правила обработки естественного текста для нужд автоматического анализа и синтеза текста.

В результате освоения дисциплины обучающийся должен:

знать:

- основные понятия современной лингвистики, лежащей в основе автоматического анализа и синтеза текстов;
- базовые принципы и методы автоматической обработки естественного текста для целей автоматического перевода и смежных задач (таких, как извлечение знаний из текстов и глубоко семантического анализа);
- основные принципы создания словаря для автоматического перевода;

уметь:

- определять морфологическую структуру предложения, определять основные элементы синтаксической структуры предложения в соответствии с задачей автоматического перевода;
- осуществлять самостоятельный анализ предложения в заданной лингвистической парадигме;
- обнаруживать в реальном предложении ситуации разного типа неоднозначностей, существенных для анализа текста и автоматического перевода;
- строить базовые правила автоматической обработки текста в заданной лингвистической парадигме);
- составлять простые статьи словаря для автоматического перевода (на примере конкретных естественных языков, в первую очередь, русского и английского);

владеть:

- современной терминологией автоматической обработки текста;

- навыками анализа простых предложений для задачи автоматического анализа и синтеза текстов;
- навыками создания простых правил автоматического перевода в рамках выбранной лингвистической парадигмы.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ПК-10. Способен пользоваться лингвистически ориентированными программными продуктами	ПК-10.1	Знает: основные типы систем, использующих модули лингвистического анализа; основные принципы и методы компьютерного моделирования лингвистических задач.
	ПК-10.2	Умеет: анализировать работу различных систем обработки текста и звучащей речи для выявления основных лингвистических компонентов и основных типов обработки текста, используемых в данных системах; подбирать необходимые лингвистические ресурсы для различных задач лингвистического обеспечения систем (например, лексикографических, задач морфологического анализа и т.п.).
	ПК-10.3	Имеет практический опыт работы с различными системами автоматической и экспертной обработки текста и звучащей речи.
ПК-11. Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умеет пользоваться такими ресурсами	ПК-11.1	Знает: основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.
	ПК-11.2	Умеет: пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными

		программными продуктами.
	ПК-11.3	Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации.
ПК-12. Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем	ПК-12.1	Знает: основные системы автоматической обработки звучащей речи и текстов на естественном языке; базовые принципы автоматической обработки языковых данных; основные интеллектуальные и информационные электронные системы и принципы работы с ними.
	ПК-12.2	Умеет: пользоваться существующими системами автоматической обработки текста и звучащей речи, интеллектуальными и информационными электронными системами; проводить их сравнительный анализ; проектировать модули данных систем, составлять технические задания.
	ПК-12.3	Имеет практический опыт работы с системами автоматической обработки текста и звучащей речи; проектирования модулей таких систем.
ПК-13. Способен проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем	ПК-13.1	Знает: типы, характеристики и особенности основных доступных в Интернете лингвистических ресурсов.
	ПК-13.2	Умеет: сравнивать данные, полученные с использованием различных электронных лингвистических ресурсов и систем; применять методы математического анализа и моделирования в профессиональной деятельности.
	ПК-13.3	Имеет практический опыт тестирования электронных лингвистических ресурсов, систем и компонентов.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Автоматический перевод» относится к части, формируемой участниками образовательных отношений, блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Общая морфология», «Общий синтаксис», «Общая семантика», «Общая и компьютерная лексикография», «Программирование в лингвистике», «Компьютерная и корпусная лингвистика», «Грамматика конструкций и микросинтаксис».

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 2 з.е., 72 академических часа.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
	Лекции	8
	Семинары	20
	Всего:	28

Объем дисциплины в форме самостоятельной работы обучающихся составляет 44 академических часа.

3. Содержание дисциплины

Раздел I. История автоматического перевода

Джорджтаунский эксперимент 1954 г. Распространение результатов в мировую практику. Автоматический перевод в мире и в СССР. Взлеты и падения. Ограниченность возможностей языкового описания. Кристаллизация подходов к правилочному машинному переводу.

Раздел II. Лингвистическое моделирование

Действующие модели языка.

Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.

Многоуровневый подход к анализу и синтезу. Поверхностный и глубинный уровни представления языка и текста.

Раздел III. Грамматика и словарь естественного языка

Представление об интегральном описании языка. Гармонизация словаря и грамматики.

Синтаксис составляющих и синтаксис зависимостей. Гибридные синтаксические представления.

Теоретическая лингвистика и задача автоматической обработки текстов.

Раздел IV. Автоматический анализ и синтез текста

Морфологический анализ текста. Морфологические категории и грамемы. Алгоритмы морфологического анализа.

Синтаксический анализ текста. Понятие синтагмы. Алгоритмы синтаксического анализа. Правилые и статистические подходы к автоматической обработке текста. Ограниченная статистическая поддержка правилых систем.

Трансфер. Типология правил перевода. Общие и трафаретные правила.

Раздел V. Языковая неоднозначность и методы ее разрешения при автоматической обработке текста

Языковая неоднозначность как неотъемлемое свойство языка и способы ее разрешения. Типы неоднозначности. Морфологическая, лексическая и синтаксическая неоднозначность.

Методы разрешения неоднозначности при автоматической обработке текста.

Применение онтологии и методов машинного обучения и интерактивных методов для разрешения неоднозначности.

Раздел VI. Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке

Система автоматического перевода как механизм обратной связи и источник новых лингвистических знаний.

Уроки конкретных систем машинного перевода для теоретической лингвистики и лексикографии.

Раздел VII. Типология систем автоматического перевода

Типы систем машинного перевода.

Правилый и статистический подходы к машинному переводу.

Поколения машинного перевода.

Интерлингва. Виды интерлингвы (естественные и искусственные семантические языки). UNL как разновидность интерлингвы.

Раздел VIII. Словарь систем автоматического перевода

Словари системы автоматического перевода. Морфологические и комбинаторные словари.

Структура словарной статьи. Одноязычные и переводные зоны.

Синтаксические признаки.

Семантические признаки, дескрипторы и онтологические концепты.

Теория валентностей. Семантические и синтаксические валентности.

Модель управления.

Раздел IX. Лексические функции в машинном переводе

Типология лексических функций. Коллокаты и субституты. Составные лексические функции. Суперпозиция лексических функций.

Применение лексических функций для идиоматического перевода

Лексические функции и перифразирование.

Использование лексических функций для снятия лексической и синтаксической неоднозначности текста.

Раздел X. Смежные задачи прикладной лингвистики

Методы общие для автоматического перевода и других задач прикладной лингвистики.

Информационный поиск.

Интеллектуальный анализ данных.

Вопросно-ответные системы.

Лингвистические онтологии. Семантический анализ с учетом внеязыковой действительности.

Раздел XI. Цифровые лингвистические ресурсы и их роль в автоматической обработке текстов

Аннотированные корпуса текстов.

Word Net. Frame Net.

Датасеты.

Векторная семантика.

4. Образовательные технологии

Наименование раздела	Виды учебных занятий	Образовательные технологии
Раздел I. Введение. История автоматического перевода. Уроки Джорджтаунского эксперимента	Лекция 1	
Раздел II. Лингвистическое моделирование.	Лекция 2. Постановка проблемы	

	Семинар 1. Действующие модели языка: общий подход	Коллективный анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел III. Грамматика и словарь естественного языка.	Лекция 3, часть 1.	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 3, часть 1. Введение в интегральное описание языка	
Раздел IV. Анализ и синтез текста	Лекция 3, часть 2.	
	Семинар 3, часть 2. Морфологический анализ	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 4. Синтаксический анализ. Статистические подходы	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел V. Языковая неоднозначность	Семинар 5 Эмпирические Методы разрешения неоднозначности при АОТ	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 6. Онтологические, машинные и интерактивные методы	
Раздел VI. Автоматический перевод как источник новых лингвистических знаний.	Семинар 7. Уроки машинного перевода	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел VII. Типология систем машинного перевода	Семинар 8. Интерлингва и ее разновидности, включая UNL	Анализ языковых явлений; самостоятельное создание элементов UNL
Раздел VIII. Словарь системы машинного перевода	Лекция 4, часть 1. Введение в компьютерную лексикографию.	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора

	Семинар 9. Структура словарной статьи.	
	Семинар 10. Синтаксические и семантические признаки.	
	Семинар 11. Теория валентностей.	
	Семинар 12, часть 1. Модель управления	
Раздел IX. Лексические функции в машинном переводе	Лекция 4, часть 2.	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 12, часть 2. Типология лексических функций	
	Семинар 13. Применение лексических функций для задач машинного перевода и АОР	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел X. Смежные задачи прикладной лингвистики.	Семинар 14. Информационный поиск. Интеллектуальный анализ данных. Вопросно-ответные системы. Лингвистические онтологии.	Демонстрация и обсуждение излагаемых задач
Раздел XI. Цифровые лингвистические ресурсы	Лекция 5. Общее введение в цифровые ресурсы	
	Семинар 15. Аннотированные корпуса текстов. Word Net, Frame Net	Демонстрация и обсуждение ресурсов
Обсуждение индивидуальных решений задач	Семинар 16. Представление и решение индивидуальных задач	Представление и решение индивидуальных задач. Консультации по курсу
Аттестация	Зачёт	Ответы на контрольные вопросы по курсу. Решение практических заданий

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Распределение баллов по видам учебной деятельности таково:

- 1) уровень активности студента на семинарах (выполнение домашних заданий и участие в их обсуждении; выступления по прочитанной литературе; участие в обсуждении и выполнении коллективных заданий; представление промежуточных результатов индивидуальных исследований и участие в их обсуждении) — всего до 20 баллов;
- 2) написание мини-тестов по темам предшествующих занятий — до 15 баллов;
- 3) написание промежуточной письменной контрольной — до 25 баллов;
- 4) защита индивидуального исследовательского проекта — до 30 баллов;
- 5) ответы на контрольные вопросы по курсу — до 10 баллов

Баллы за участие в семинарах. Преподавание дисциплины строится на параллельном обсуждении теоретических вопросов и выполнении индивидуальных и коллективных практических заданий. Все эти задания в обязательном порядке обсуждаются на семинарах. Активное и разумное участие в обсуждении этих заданий, а также доклады о прочитанной литературе и обсуждение индивидуальных исследовательских проектов может принести студенту до 20 баллов.

Баллы за выполнение текущих контрольных работ. В начале большинства семинаров проводятся краткие тесты, включающие в себя базовые вопросы по темам предшествующих занятий. Суммарно за семестр выполнение этих тестов может принести студенту до 15 баллов.

Баллы за промежуточную аттестацию. Зачёт состоит из двух частей. В первой части студенты представляют ответы на практические задания. Оценивается степень овладения тематикой, корректность выбранного метода исследования, наличие содержательных результатов и умение их представить аудитории (в сумме — до 30 баллов). Во второй части студенты устно отвечают на контрольные вопросы по курсу (до 10 баллов).

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала	Шкала ECTS
95 – 100	отлично	A

83 – 94		зачтено	B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	хорошо/ зачтено	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	удовлетво- рительно/ зачтено	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	неудовлет- ворительно/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Контрольные вопросы по теоретическому материалу

Контрольные вопросы преимущественно затрагивают теоретический материал; проверка умений и навыков осуществляется в первую очередь с помощью практических заданий. Контрольные вопросы также используются на зачете.

Алгоритм автоматического перевода

1. Уровни представления текста. Поверхностные и глубинные уровни.
2. Каковы основные этапы автоматического перевода?
3. Этап трансфера в автоматическом переводе.
4. Что такое интерлингва?
5. Основные принципы статистического машинного перевода.

Морфология

6. Что такое морфологический анализ?
7. Что такое морфологический синтез?
8. Морфологические позиции слова и их роль в морфологическом анализе.
9. Основные средства и этапы морфологического анализа текста.
10. Словоизменительные грамматические категории
11. Классифицирующие грамматические категории: для чего они нужны?
12. Каковы основные грамматические категории английских глаголов.
13. Каковы основные категории русских прилагательных и их значения?

Синтаксис

14. Дерево составляющих и дерево зависимостей – сходства и различия.
15. Что такое синтаксическое отношение?
16. Повторимые и неповторимые синтаксические отношения
17. Что такое синтаксический анализ?
18. Что такое нормализованная синтаксическая структура?
19. Что такое расширенная синтаксическая структура?
20. Что такое проективность синтаксической структуры?
21. Согласование в узком и широком смысле

Словарь

22. Для чего служат синтаксические признаки слов?
23. Семантические признаки слов и концепты,
24. Что такое модель управления слова?
25. Актанты и сирконстанты предикатных слов.
26. Типичная модель управления русского глагола.
27. Типичная модель управления русского предлога.
28. Что такое сильноуправляемые предлоги?
29. Что такое сильноуправляемые союзы?
30. Каковы различия между синтаксическими и семантическими валентностями слова?
31. Активные, пассивные и дистантные валентности слов.

Ресурсы

32. Что такое лексическая функция?
33. Как могут использоваться в автоматическом переводе лексические функции?
34. Что такое размеченный корпус текстов?
35. Как устроен WordNet и для чего он может использоваться?

Образцы вопросов для кратких тестов

- 1) Построить МорфС предложения *Он не спал и трех часов вчера.*
Построить СинтС предложения *В Питере снегу напало, все деревья в снегу.*
- 2) Построить МорфС предложения *Он взял за правило просто говорить о сложных вещах.*
Построить СинтС этого предложения.

Образцы домашних заданий

- 1) Домашнее задание № 3 по автоматическому переводу. Валентная структура слова.
Вариант 1.

Ответьте на следующие вопросы:

- а) Сколько валентностей глагола *сменять* насыщено в предложении
Одна заря сменить другую спешит, дав ночи полчаса? (А.С.Пушкин):

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

- б) Сколько валентностей глагола *сменять* насыщено в предложении

Предложения сменить кириллицу на латиницу раздавались и в нашей стране.

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

- в). Сколько валентностей глагола *сменять* насыщено в предложении

Рауль Кастро, сменивший на посту президента Кубы брата Фиделя, разрешил простым кубинцам пользоваться мобильными телефонами?

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

г). Сколько валентностей глагола *сменять* насыщено в предложении

А уж совсем потом — то есть в третьем классе — нам сменили школьную форму (Д. Драгунский)?

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

д) Сколько валентностей глагола *сменять* насыщено в предложении

Мы мечтали выиграть звание чемпиона еще в 2001-м, когда высшая лига проходила в Томске (даже сменили название с «Сибири» на «Томск-400») (НКРЯ)

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

Все ли валентности глагола *сменять* насыщены в предложениях 1-5?

Все ответы прокомментируйте!

2) Домашнее задание № 4 по автоматическому переводу. Актантные конструкции.
Вариант 2.

Найдите актантные связи (т.е. связи, оформляемые предикативным, агентивным, квазиагентивным, 1-м (2-м, 3-м...) комплетивным и предложным или дательным-субъектным СинТО) в следующем тексте (<https://www.newsru.com/cinema/14oct2020/embrace.html>):

Российский фотограф победил в престижном мировом конкурсе со снимком тигриных "Объятий", на создание которого ушел почти год. «Свет, цвета и текстура здесь как на картине, написанной маслом. Тигрица как бы превратилась в часть леса. Ее хвост сливается с корнями дерева. Они стали одним целым», - описала фото председатель жюри конкурса Wildlife Photographer of the Year Роз Кидманд-Кокс в разговоре с ВВС. По словам исполнительного директора Музея естествознания по научной работе Тома Литлвуда, кадр, на котором представить исчезающего вида переживает подобный момент спокойствия, "вселяет надежду".

3) Домашнее задание № 8 по автоматическому переводу. Состав компьютерного словаря.
Вариант 3.

Постарайтесь определить, какие словарные статьи следовало бы ввести в компьютерный словарь русского языка для многозначных слов, леммы которых таковы:

а) *просыпать*

б) *просыпаться*

в) *единица*

г) *дежурный*

Ответы прокомментируйте!

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Основная:

Бутусова, А. С. Машинный и автоматизированный перевод : учебное пособие / А. С. Бутусова, Ю. В. Бец ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. - 106 с. - ISBN 978-5-9275-3982-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2057597>

Хорошилов, А. А. Фразеологический машинный перевод текстов. Теоретические основы и технологические решения : монография / Ал-др А. Хорошилов, А. В. Кан, Ал-ей А. Хорошилов ; под науч. ред. К. К. Колина. - Москва : Берлин : Директ-Медиа, 2019. - 466 с. - ISBN 978-5-4499-0089-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1908602>

Дополнительная:

Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537001>

Шунейко, А. А. Корпусная лингвистика : учебник для вузов / А. А. Шунейко. — Москва : Издательство Юрайт, 2024. — 222 с. — (Высшее образование). — ISBN 978-5-534-13603-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/543746>

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером и проектором для демонстрации учебных материалов.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные

задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.
- для глухих и слабослышащих: в печатной форме, в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1 Планы семинарских/ практических занятий

Семинар 1. Действующие модели языка: общий подход	Коллективный анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
---	---

Семинар 2. Действующие модели языка: общий подход	Коллективный анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 3, часть 1. Морфологический анализ	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 4. Синтаксический анализ. Статистические подходы	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 5 Эмпирические Методы разрешения неоднозначности при АОТ	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 6. Онтологические, машинные и интерактивные методы	
Семинар 7. Уроки машинного перевода	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 8. Интерлингва и ее разновидности, включая UNL	Анализ языковых явлений; самостоятельное создание элементов UNL
Семинар 9. Структура словарной статьи.	
Семинар 10. Синтаксические и семантические признаки.	
Семинар 11. Теория валентностей.	
Семинар 12. Модель управления и типология лексических функций	
Семинар 13. Применение лексических функций для задач машинного перевода и АОТ	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Семинар 14. Информационный поиск. Интеллектуальный анализ данных. Вопросно-ответные системы. Лингвистические онтологии.	Демонстрация и обсуждение излагаемых задач
Семинар 15. Аннотированные корпуса текстов. Word Net, Frame Net	Демонстрация и обсуждение ресурсов
Семинар 16. Представление и решение индивидуальных задач	Представление и решение индивидуальных задач. Консультации по курсу

9.2 Иные материалы

Рекомендуемая литература для более глубокого освоения программы.

Основная литература

Apresjan Ju, Boguslavsky I., Iomdin L et al. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning – Text Theory (June 16-18 2003). Paris: École Normale Supérieure, 2003. P. 279-288.

Апресян Ю.Д. Избранные труды. Том 2: Интегральное описание языка и системная лексикография. М.: Школа « Языки русской культуры», 1995.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистический процессор для сложных информационных систем. М.: Наука, 1992.

Мельчук И.А. Опыт теории лингвистических моделей класса «Смысл – Текст». – М.: Наука, 1974/М.: Школа « Языки русской культуры», 1999.

Дополнительная литература

Баранов А.Н. Введение в прикладную лингвистику. Серия "Новый лингвистический учебник". М.: Эдиториал УРПС. 2001. Глава 2, раздел 1.3.1. Моделирование общения (с. 20-31); Глава 4, разделы 1.3.1. – 1.3.4. «Естественный» перевод: лингвистические проблемы (с. 143-163); 1.4. Машинный перевод (с. 168-178).

Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие /. — М.: МИЭМ, 2011. — 272 с.

Кулагина О. С. Исследования по машинному переводу. – М., 1979.

W. John Hutchins and Harold L. Somers. An introduction to machine translation. London: Academic Press, 1992.

Carl Pollard, Ivan A. Sag: Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press. 1994.

Boguslavsky I, Iomdin L. Nivre J. Parsing the Russian Dependency Treebank. Proceedings of COLING-2008. Manchester, 2008.

Philipp Koehn. Statistical Machine Translation. Cambridge University Press. 2009.

Jury D. Apresjan, Igor M. Boguslavsky, Leonid L. Iomdin, Leonid L. Tsinman. Lexical Functions in Actual NLP-Applications // Selected Lexical and Grammatical Issues in the Meaning–Text Theory. In honour of Igor Mel'čuk. (Ed. by Leo Wanner). John Benjamins, Studies in Language Companion. Series 84. ISBN 978 90 272 3094 2. 2007. P. 199-230.

New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems. Eds: Alessandro Oltramari, Piek Vossen, Lu Qin, Eduard Hovy. Springer, Theory and Applications in Natural Language Processing Series. 2013.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина реализуется в УНЦ компьютерной лингвистики Института лингвистики.

Цель дисциплины - освоение студентами базовых понятий и методов автоматического перевода, в первую очередь лингвистических понятий и методов, знакомство с историей автоматического перевода, принципов многоуровневого моделирования естественного языка, основных подходов к морфологическому, синтаксическому и семантическому анализу текстов, к соотношению грамматики и словаря.

Задачи:

- ознакомить студентов с общей проблематикой и принципиальными подходами к реализации автоматического перевода между естественными языками;
- указать студентам на место автоматического перевода в ряду задач автоматической обработки текста;
- обучить студентов осуществлять самостоятельный анализ морфологических и синтаксических характеристик текста, релевантных для выполнения автоматического перевода, а также оценивать результаты автоматической разметки;
- обучить студентов самостоятельно формулировать правила обработки естественного текста для нужд автоматического анализа и синтеза текста.

Дисциплина направлена на формирование следующих компетенций:

- ПК-10. Способен пользоваться лингвистически ориентированными программными продуктами
- ПК-11. Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умеет пользоваться такими ресурсами
- ПК-12. Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем
- ПК-13. Способен проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем

В результате освоения дисциплины обучающийся должен:

знать:

- основные понятия современной лингвистики, лежащей в основе автоматического анализа и синтеза текстов;

- базовые принципы и методы автоматической обработки естественного текста для целей автоматического перевода и смежных задач (таких, как извлечение знаний из текстов и глубоко семантического анализа;
- основные принципы создания словаря для автоматического перевода;

уметь:

- определять морфологическую структуру предложения, определять основные элементы синтаксической структуры предложения в соответствии с задачей автоматического перевода;
- осуществлять самостоятельный анализ предложения в заданной лингвистической парадигме;
- обнаруживать в реальном предложении ситуации разного типа неоднозначностей, существенных для анализа текста и автоматического перевода;
- строить базовые правила автоматической обработки текста в заданной лингвистической парадигме);
- составлять простые статьи словаря для автоматического перевода (на примере конкретных естественных языков, в первую очередь, русского и английского);

владеть:

- современной терминологией автоматической обработки текста;
- навыками анализа простых предложений для задачи автоматического анализа и синтеза текстов;
- навыками создания простых правил автоматического перевода в рамках выбранной лингвистической парадигмы.

По дисциплине предусмотрена промежуточная аттестация в форме *зачета*.

Общая трудоемкость освоения дисциплины составляет 2 зачетные единицы.