

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ



Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «РГГУ»)

Институт лингвистики

УНЦ компьютерной лингвистики

Рабочая программа дисциплины

«Введение в компьютерную лингвистику»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Магистерская программа: Фундаментальная и компьютерная лингвистика

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

Введение в компьютерную лингвистику
Рабочая программа дисциплины

Составитель:

к.ф.н., доцент А.Ч.Пиперски

Ответственный редактор:

д. филол. н., профессор В.И.Подлеская

УТВЕРЖДЕНО

**Протокол заседания УНЦ компьютерной
лингвистики**

№ 1 от «28» августа 2019г.

Оглавление

1. Пояснительная записка

- 1.1. Предмет
- 1.2. Цель и задачи дисциплины
- 1.3. Формируемые компетенции и результаты освоения дисциплины
- 1.4. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

- 5.1. Система оценивания
- 5.2. Критерии выставления оценок
- 5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

6. Учебно-методическое и информационное обеспечение дисциплины

- 6.1. Список литературы

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

9. Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

1. Пояснительная записка

1.1 Предмет

Предметом дисциплины (модуля) является ознакомление с методами и лингвистическими технологиями, применяемыми при создании компьютерных систем обработки текстов в научно-практической области исследований «компьютерная лингвистика», и рассматриваемых в сопоставлении с лингвистическими и культурно-лингвистическими свойствами языковых произведений – предложений и текстов, а также в связи с задачами обработки текстов как социальными запросами общества. Подача материала частично увязана с историей компьютерной лингвистики, что позволяет лучше представить качественную составляющую процессов, моделируемых современными лингвистическими технологиями, изучаемыми в других курсах по профилю «Компьютерная лингвистика», основанных преимущественно на эмпирических, в частности, статистических методах.

1.2 Цель и задачи курса

Курс направлен на решение следующих задач:

- создать представление о компьютерной лингвистике как новейшей научно-практической области исследований, ее возникновении в контексте смежных наук и ее современной организации;
- познакомить магистрантов с основными лингвистическими технологиями, реализующими анализ предложения (текста) по уровням лингвистической разметки и основными приемами автоматической генерации текстов;
- познакомить магистрантов с основными типами ресурсов, создающимися и использующимися компьютерными программами для решения конкретных задач в исследовательских целях, при разработке лингвистических технологий и в приложениях;
- соединить интуитивные и традиционные представления о свойствах естественно-языковых текстов со способами их формализации и моделирования в работах по компьютерной лингвистике;
- выработать у магистрантов элементарные практические навыки по применению компьютерно-лингвистических методов к языковому материалу и использованию лингвистических технологий.

1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к абстрактному мышлению, анализу, синтезу (ОК-1);

готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

готовностью к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);
способностью анализировать, сопоставлять и критически оценивать различные лингвистические направления, теории и гипотезы (ОПК-5);
способностью адаптироваться к новым теориям и результатам мировой науки и расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук (ОПК-6);
способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);
способностью проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики (ПК-1);
способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);
способностью разрабатывать лингвистические компоненты электронных языковых ресурсов (лингвистические корпуса, словари, онтологии, базы данных) (ПК-7).

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать:

- структуру научно-практической области исследований «компьютерная лингвистика» и ее место в контексте смежных наук, цели этой области и условия ее появления и развития;
- основные методы компьютерной лингвистики и роль эмпирических методов на современном этапе;
- состав уровней технологий компьютерной лингвистики в соответствии со свойствами и организацией лингвистических объектов;
- основные методы и технологии анализа и генерации текстов;
- основные типы лингвистических ресурсов, используемых лингвистическими технологиями;
- существенные для передачи информации свойства текстов и их моделирование в компьютерной лингвистике.

Уметь:

- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- применять полученные знания в области моделирования конкретных процессов анализа и генерации текстов в научно-исследовательской и других видах практической деятельности;
- осуществлять тестирование и оценку основных ресурсов и лингвистических технологий анализа, и генерации текстов.

Владеть:

- самыми общими методами и ресурсами обработки текстов при решении задач компьютерной лингвистики в исследовательской и практической работе.

1.4 Место дисциплины в структуре образовательной программы

Дисциплина (модуль) «Введение в компьютерную лингвистику» является базовой частью цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная

лингвистика», и адресована студентам 1 курса (1 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение домашних заданий; тестовое задание; защита исследовательского проекта; промежуточная аттестация в форме: экзамен.

Общая трудоемкость освоения дисциплины (модуля) составляет 3 зачетные единицы, 108 часов.

Программой дисциплины (модуля) предусмотрены: лекционные занятия – 8 часов; практические занятия – 22 часа; самостоятельная работа студента – 60 часов, контроль – 18 часов

2. Структура дисциплины

№ п/ п	Раздел Дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				лекц ии	семи- нары	самос тояте льная работ а	ко нтр оль	
1.	Введение. Краткая история и предмет компьютерной лингвистики (КЛ). Общая структура и основные задачи КЛ. Теоретическая и прикладная КЛ.	1	1	2		4		ДЗ1. Знакомство с организацией научно-практической области КЛ. Описание термина (по тезаурусу)
2.	Уровни языка в традиционной лингвистике и общие уровневые модели естественного языка (ЕЯ) в КЛ. Задачи анализа и синтеза предложения. Проблема неоднозначности.	1	2	2	2	4		ДЗ2. Синтаксическая разметка фрагмента текста.
3.	«Понимание» текстов в узкой предметной области.	1	3		2	4		ДЗ3. Тестирование программы Элиза. Выделение видов

	Метод шаблонов. Семантически ориентированный метод анализа.							шаблонов.
4.	Машинный перевод (МП). Схемы МП. Синтаксический анализ (отечественная традиция). Фильтровый метод.	1	4		2	4		ДЗ4. Тестирование систем МП, организованных по разным схемам. Сравнение результатов.
5.	Формализмы синтаксического анализа в англоязычной традиции.	1	5		2	4	4	(Проверка синтаксической разметки)
6.	Корпуса текстов. Распространение эмпирических методов решения задач в КЛ. Метод n-gram – понятие статистической модели языка. Современное состояние МП.	3	6	2		6		ДЗ5. Исследовательский проект (часть 1): корпусное исследование семантики русского глагола.
7.	Задачи семантики и типы семантических моделей в КЛ, связь с лексикографией. Лексико-семантические базы.	1	7		4	6		ДЗ6. Исследовательский проект (часть 2): Описание значений русского глагола в виде фреймов.
8.	Автоматическая генерация текстов на ЕЯ. Теория риторических структур.	1	8		4	4		ДЗ7. Упражнение на построение модели дискурса конкретного текста в терминах теории риторических отношений.
9.		1	9				2	(Проверка структуры текста. Обсуждение русских фреймов)
10.	Общие модели дискурса. Моделирование стиля и жанра текстов в ресурсах и системах КЛ.	1	10	2		6		ДЗ8. Исследовательский проект (часть 3): Сопоставление русских фреймов с фреймами базы FrameNet.
11.	Тематический аспект текста и поиск информации. Диалоговые и		11		4	4		ДЗ9. Тестирование современных диалоговых систем.

	интерактивные системы.							
12.	Автоматическая обработка устной речи и ее приложения.	1	12		2	4		ДЗ10. Оформление отчета по исследовательскому проекту.
13.	Обсуждение курсовой работы	1	13			4		
14.	Экзамен	1	14			4		Контрольные вопросы
				8	22	60	18	

3. Содержание дисциплины

Содержание программы учебной дисциплины:

Часть 1. Задачи и методы обработки языковых произведений – анализ предложения.

1.1. Компьютерная лингвистика (КЛ) в системе смежных наук. Краткая история и предмет КЛ.

Возникновение компьютерной лингвистики как новейшего научно-практического направления исследований, ее объект и предмет. КЛ и формы передачи информации. КЛ в системе смежных наук, общая структура и основные разделы КЛ. Теоретическая и прикладная КЛ.

1.2. Уровни языка в традиционной лингвистике и уровневые модели естественного языка (ЕЯ) в КЛ. Задачи анализа предложения. Проблема неоднозначности.

Уровни языка в традиционной лингвистике и уровневые модели естественного языка в КЛ: уровневая модель «Смысл-Текст» И.А. Мельчука как структурная модель ЕЯ. Понятие лингвистического представления. «Системно-функциональная грамматика» М.А.К. Хэллiday как функциональная модель ЕЯ. Виды лингвистической информации, выражаемой в предложении: морфологическая, лексическая, синтаксическая, дискурсная, прагматическая и ее выявление методами КЛ. Сегментация текста и морфологический анализ. Конечные автоматы и регулярные выражения. Задачи синтаксического компонента и синтаксические представления. Задачи дискурсивного анализа. Мнозначность элементов ЕЯ и проблема неоднозначности в КЛ, виды неоднозначности.

1.3. «Понимание» текстов в узкой предметной области и неоднозначность. Метод шаблонов. Семантически ориентированный метод анализа.

Моделирование «понимания» в узкой предметной области как способ избежать решения проблемы неоднозначности. Составляющие «понимания» текста компьютером на примере ранних систем искусственного интеллекта (70ые годы 20 в): Student, Sir, Eliza. SHDRU - Робот Винóграда – первая система КЛ с развитым лингвистическим компонентом. Метод шаблонов и продукционные правила. Понятие лингвистического процессора. Метод семантически ориентированного анализа А.С. Нариньяни как средство решения проблемы доступа на ЕЯ к структурированным источникам информации. Распознавание именованных сущностей.

1.4. Машинный перевод (МП). Схемы МП. Анализ и синтез предложения. Понятие языка-посредника. Синтаксический анализ (отечественная традиция). Фильтровый метод.

Человеческий и машинный перевод (МП). Проблема несоответствия между языками. Типологические и контрастивные виды несоответствий между языками. Понятие языка-

посредника. Прямая схема МП. Схема МП «интерлингва» и виды языков-посредников, понятия анализа и синтеза предложения. Схема МП «трансфер». Преимущества схемы «трансфер» перед схемой «интерлингва». Полнота синтаксического анализа и его применение в системах обработки текстов. Грамматика зависимостей как способ представления синтаксической структуры предложения. Фильтровые технологии.

1.5 Формализмы синтаксического анализа в англоязычной традиции.

Метод шаблонов (Pattern matching), его применение в системах КЛ. Грамматика составляющих - Phrase Structure Grammar (PSG) - как способ представления синтаксической структуры предложения. Стратегии синтаксического анализа. Синтаксические формализмы: трансформационная грамматика Н. Хомского, Расширенные сети переходов (Augmented Transition Networks (ATN)); расширение и обобщение грамматики PSG: APSG и GPSG. Грамматики свойств и операция унификации. HPSG – Вершинная грамматика непосредственных составляющих.

Часть 2. Лингвистические ресурсы. Задачи и методы обработки языковых произведений – анализ и генерация текстов. Прикладные направления.

2.6. Корпуса текстов. Распространение эмпирических методов решения задач в КЛ. Метод n-gram – понятие статистической модели языка. Современное состояние МП.

Системы МП, основанные на примерах - example-based MT (EBMT), и системы типа «Переводческая память» (translation memory). Стадии обработки текста в системе МП, основанной на примерах (EBMT). Статистические системы МП и понятие статистической модели языка. Развитие и современное состояние прямых систем МП на примере системы ПРОМПТ. Развитие систем типа «трансфер» в сторону гибридных систем. Развитие систем, основанных на языке-посреднике, в сторону систем, основанных на знаниях (KBMT).

2.7. Задачи семантики и типы семантических моделей в КЛ, связь с лексикографией. Лексико-семантические базы как технологии описания лексики.

Семантические типы и семантические примитивы. Падежные грамматики. Концептуализации в теории Р. Шенка, и семантическая классификация процессов. Классификация процессов по М. Хэллiday. Связь семантических типов процессов с семантическими ролями. Виды онтологий по особенностям описываемых понятий: философская, когнитивная, лексикографическая, лексическая, информационно-поисковые тезаурусы. Лексико-семантические отношения как средство описания парадигматического аспекта лексических систем: проект WordNet; принципы выделения отношений. Лексико-семантические базы на основе WordNet. Лексико-семантические базы как средство описания синтагматического аспекта лексических систем. Отечественные семантические словари. Проект FrameNet и иерархические отношения между фреймами.

2.8. Автоматическая генерация текстов на ЕЯ. Теория риторических структур.

Виды входных данных для систем автоматической генерации текстов. Шаблонные и лингвистически мотивированные подходы. Схема идеализированной системы генерации текстов. Соответствие между типами входных данных и типами генерируемых текстов. Процессы автоматической генерации текстов: макро и микропланирование, лексикализация и грамматикализация, вставка ссылочных конструкций, агрегация, языковое оформление. Понятие дискурса и риторической структуры текста. Теория риторических структур (TPC) как модель структуры дискурса текста: основные положения, определение риторического отношения, типы риторических отношений.

Универсальный «план-оператор» для планирования текстов на основе определения риторического отношения.

2.10. Общие модели структуры дискурса. Моделирование жанра и стиля текстов в ресурсах и системах КЛ.

Сравнение ТРС с другими моделями структуры дискурса: Теория связности дискурса Хоббса, модель структуры дискурса английского языка на основе системно-функциональной грамматики, моделирование методами формальной семантики – Discourse Representation Theory. Сравнение общих моделей дискурса. Понятие жанра. Классификация жанров в европейских корпусах текстов. Метаинформация о жанре текста в Национальном корпусе русского языка. Моделирование сюжета для текстов разных жанров: структура волшебных сказок (В.Я. Пропп); структура новеллы (В. Ленерт); структура специального текста (инструкции в системе генерации текстов AGILE). Понятия стиля в традиционной лингвистике. Попытки формализации понятия стиля в КЛ. Моделирование функциональных стилей в виде «конструкций» на основе статистической обработки текстов.

2.11. Тематический аспект текста и поиск информации. Диалоговые и интерактивные системы.

Проявление тематического аспекта в текстах. Задачи информационного поиска: информационный поиск (ИП), фактографический поиск, поиск новой информации, распознавание именованных сущностей, отношений и тональности. Параметры полноты и точности ИП. Метод индексирования, основанный на частоте встречаемости слов как основа ИП. Метод кластеризации как способ классификации документов и терминов. Методы обработки неоднословных терминов. Использование информационно-поисковых тезаурусов. Лингвистические способы улучшения результатов статистических методов ИП.

2.12. Автоматическая обработка устной речи и ее приложения.

Распознавание речи или преобразование «звук-буква»: преобразование речи в сеть измерений, преодоление вариаций, моделирование звуков речи, применение языковых ограничений, интеграция речевых и текстовых технологий. Синтез речи или преобразование «буква-звук». Статистические модели синтеза речи, основанные на звуковых корпусах. Применение речевых технологий.

4. Образовательные технологии

В соответствии с требованиями ФГОС по направлению 45.04.03 «Фундаментальная и прикладная лингвистика» и с учетом специфики магистерской программы «Фундаментальная и компьютерная лингвистика». Занятия лекционного типа составляют не более 20% аудиторных занятий, а удельный вес занятий, проводимых в интерактивных формах, составляют не менее 40% аудиторных занятий. Интерактивные формы обучения в данном курсе предполагают:

- систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с отчетом по результатам исследования);
- он-лайн демонстрации работы с лингвистическими интернет-источниками (и др.);
- он-лайн использование лингвистических ресурсов (Национальный корпус русского языка, Лексико-семантические базы и др.);

- обсуждения курсовых исследовательских проектов;
- работа в группах по выполнению домашних практических заданий.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой 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	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,
- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдаёт зачёт по всему курсу и предъявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

5.2. Критерии выставления оценок

При выставлении оценки преподаватель ориентируется на следующие содержательные критерии.

Количество баллов	Критерии оценки
95–100 (A)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество их выполнения оценено числом баллов, близким к</p>

Количество баллов	Критерии оценки
	<p>максимальному.</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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
83–94 (B)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, почти все задания, предусмотренные рабочей программой дисциплины, выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.</p> <p>Обучающийся адекватно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Достаточно свободно ориентируется в учебной и профессиональной литературе.</p> <p>Почти все компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
68–82 (C)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.</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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «хороший».</p>
56–67 (D)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.</p> <p>Обучающийся испытывает определённые затруднения в</p>

Количество баллов	Критерии оценки
	<p>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>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
50–55 (E)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные рабочей программой дисциплины учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.</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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».</p>
21–49 (FX)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>
0–20 (F)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины не освоено. Необходимые практические навыки работы не сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены с грубыми ошибками. Дополнительная самостоятельная работа над материалом дисциплины не приведет к какому-либо значимому повышению качества выполнения</p>

Количество баллов	Критерии оценки
	<p>учебных заданий.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

В качестве домашних заданий предлагаются задачи следующих типов

- Д31. Знакомство с организацией научно-практической области КЛ. Описание термина (по тезаурусу)
- Д32. Синтаксическая разметка фрагмента текста.
- Д33. Тестирование программы Элиза. Выделение видов шаблонов.
- Д34. Тестирование систем МП, организованных по разным схемам. Сравнение результатов.
(Проверка синтаксической разметки)
- Д35. Исследовательский проект (часть 1): корпусное исследование семантики русского глагола.
- Д36. Исследовательский проект (часть 2): Описание значений русского глагола в виде фреймов.
- Д37. Упражнение на построение модели дискурса конкретного текста в терминах теории риторических отношений.
(Проверка структуры текста. Обсуждение русских фреймов)
- Д38. Исследовательский проект (часть 3): Сопоставление русских фреймов с фреймами базы FrameNet.
- Д39. Тестирование современных диалоговых систем.
- Д310. Оформление отчета по исследовательскому проекту.

Экзамен ориентирован на следующие контрольные вопросы

КЛ как новая научно-практическая область исследований в контексте смежных наук.
Задачи теоретической КЛ и приложений.
Уровни языка и уровневые общие модели языка в КЛ.
Методы анализа текста.
Составляющие понятия «понимание». Понятие лингвистического процессора.
Семантически ориентированный метод анализа текстов.
Машинный перевод: схемы МП, понятие языка-посредника.
Синтаксический анализ предложения и фильтровый метод. Различия отечественной и англоязычной традиций в области синтаксического анализа.
Эволюция формализмов синтаксического анализа в англоязычной традиции.
Корпусы текстов и распространение эмпирических методов.
Типы систем МП, основанных на эмпирических методах. Понятие статистической «модели языка».

Падежные грамматики и семантические примитивы в КЛ. Связь синтаксиса и лексикологии.

Лексико-семантические базы как лексические ресурсы для систем обработки текстов.

Моделирование дискурса в ресурсах и системах КЛ.

Методы автоматической генерации текстов.

Моделирование параметров текстов (жанр, стиль) в ресурсах и системах КЛ.

Задачи поиска информации и мера терминологичности слова в тексте.

Виды поиска информации в КЛ.

Направления и приложения моделирования устной речи.

Планы семинарских занятий

Занятие 1(1)

Уровни языка в традиционной лингвистике и общие уровневые модели естественного языка (ЕЯ) в КЛ. Понятие лингвистического представления.

Цель занятия: На конкретных примерах показать различие принципов уровневого деления в традиционной лингвистике и общих моделях ЕЯ в КЛ. Ввести понятие лингвистического представления языкового объекта.

Основная литература:

Солнцев В. М. Уровни языка // Лингвистический энциклопедический словарь

Статья доступна в Интернет: <http://tapemark.narod.ru/les/539a.html>

Мельчук И. А. Об одной лингвистической модели типа «Смысл-Текст». Уровни представления языковых высказываний // Изв. АН СССР, Серия литературы и языка, 1974. Т. 33, №5-6. Статья доступна в Интернет: <http://feb-web.ru/feb/izvest/1974/05/745-436.htm>

Halliday, M.A.K. *Introduction to Functional Grammar*, London: Edward Arnold. 1985.

Скачивается по адресу:

<http://gigabitwarez.com/download/An%20introduction%20to%20functional%20grammar%20halliday.html>

Соколова Е.Г. Синтаксическая разметка в терминах грамматики зависимостей и синтаксических функций [Электронный ресурс]: метод. пособие. М.: РГГУ, 2011. -33 с. – Библиогр.: с. 33(5 назв.). – Режим доступа: <http://elibrary.ru/elib/000003603.pdf>

Основные ресурсы:

Dialing : <http://aot.ru/> (морфологическое представление)

Основной корпус и синтаксический подкорпус НКРЯ: <http://www.ruscorpora.ru/search-syntax.html>

Занятие 2(1)

Задачи анализа предложения. Проблема неоднозначности.

Цель занятия: Познакомить студентов с системами КЛ, реализующими процессы «уровневого анализа»: сегментация, морфологический и синтаксический анализ, установление анафорических отношений. Сформировать у студентов представление о неоднозначности как специфической проблеме КЛ и познакомить с типами неоднозначности, возникающих при анализе на разных уровнях языка.

Основная литература:

Анна А. Зализняк. Многозначность и смежные понятия. // Анна А. Зализняк. Многозначность в языке и способы ее представления. М: Языки славянских культур, 2006. Глава 1, 1.1. с. 20-34.

Буторов В.Д. Моделирование синтаксиса естественного языка // Прикладное языкознание. Учебник. – СПб., 1996. с. 142-161.
Виноград Т. Система, понимающая естественный язык. М: Мир, 1976. Раздел: 1.2. Образец диалога, с. 21-32.

Занятие 3(2)

«Понимание» текстов в узкой предметной области. Метод шаблонов. Понятие лингвистического процессора.

Цель занятия: Познакомить студентов с моделированием понимания в узких предметных областях, выявляющим «информационные составляющие» процесса понимания. Продемонстрировать применение метода шаблонов. Дать понятие «лингвистического процессора».

Основная литература:

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. Лингвистический процессор для сложных информационных систем/ М.: Наука, 1992.
Баранов А.Н. Введение в прикладную лингвистику. М., 2003. Раздел: Моделирование общения. С. 20–25.
Виноград Т. Система, понимающая естественный язык. М: Мир, 1976.

Дополнительно – презентация по ранним системам понимания текстов.

Интернет источники:

ELIZA: <http://www.manifestation.com/neurotoys/eliza.php3> - первая программа-собеседник Дж. Вейценбаума («компьютерный психотерапевт»)

Занятие 4(2)

Семантически ориентированный метод анализа. Распознавание именованных сущностей.

Цель занятия: Познакомить студентов с моделированием понимания в узких предметных областях и семантически-ориентированным методом А.С. Нариньяни как способами «обойти» проблему неоднозначности. Распознавание именованных сущностей как элемент семантически-ориентированного анализа текстов.

Основная литература:

Жигалов В.А., Соколова Е.Г. InBASE: технология построения ЕЯ интерфейсов к базам данных // Труды Международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям Том 2, Аксаково, Июнь 2001 с. 123-135. Доступна с сайта: <http://dialog-21.ru/digest/archive/2001/?year=2001&vol=22725&id=6900>
Нариньяни А.С. Лингвистические процессоры ЗАПСИБ. Препринт N 199. Вычислительный центр СО АН СССР. Новосибирск. 1979.
Цибульский В.В., Ежов А.С., Поляков Г.А., Феклистов В.В. Анализ и классификация времени и сроков в российских нормативных и правовых актах. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 20012» (в печати).
Jurafsky, Daniel, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice-Hall. 2009. P. 761-767.

Занятие 5(3)

Машинный перевод (МП). Схемы МП. Анализ и синтез предложения. Понятие языка-посредника.

Цель занятия: Обсудить основы машинного перевода по сравнению с человеческим. Рассмотреть МП, основанный на правилах, как процесс преодоления несоответствий между языками и разрешения неоднозначности. Рассмотреть схему МП «интерлингва» и виды языков-посредников, UNL как амбициозный международный проект интерлингвы, понятия анализа и синтеза предложения. Преимущества схемы «трансфер» перед схемой «интерлингва».

Основная литература:

- Баранов А.Н. Введение в прикладную лингвистику. М., 2003. Раздел: «Естественный» перевод: лингвистические проблемы. С. 143–167.
- Hutchins J. Machine translation general overview // *The Oxford handbook of computational linguistics* (R. Mitkov ed.) N.Y.: Oxford university press, 2003. P. 501-511.
- Jurafsky, Daniel, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice-Hall. 2009. P. 895-909.

Интернет источники:

UNL (Universal Networking Language): <http://www.undl.org/>.

Занятие 6(3)

Синтаксический анализ (отечественная традиция). Фильтровый метод.

Цель занятия: На примере синтаксического анализа познакомить студентов с фильтровым методом. Связать с проблемой неоднозначности в КЛ. Обсудить различие в эволюции синтаксического анализа в отечественных работах и за рубежом.

Основная литература:

- Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. Лингвистический процессор для сложных информационных систем/ М.: Наука, 1992. 256 с.
- Шаров С.А. Средства компьютерного представления лингвистической информации //1996. <http://www.ksu.ru/eng/science/itc/vol000/002/>

Занятие 7(4)

Формализмы синтаксического анализа в англоязычной традиции.

Цель занятия: Познакомить студентов с особенностями англоязычной традиции синтаксического анализа и кратко показать историю развития синтаксических формализмов от PSG до грамматик унификаций. Формализм как технология.

Основная литература:

- Salton, Gerard. *Automatic text processing. The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Publishing Company, Inc., 1989.
- Jurafsky, Daniel, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice-Hall. 2009. “Feature structures”, “Unification of feature structures” pp. 526-531.
- Шаров С.А. Средства компьютерного представления лингвистической информации //1996. <http://www.ksu.ru/eng/science/itc/vol000/002/>

Интернет-источники для аудиторной работы:

HPSG (Вершинная грамматика непосредственных составляющих):

<http://www2.lingsoft.fi/cgi-bin/engcg?snt=Baby%2C+I+love+you.&h=on>.

Он-лайн демо: <http://erg.emmtee.net>.

LFG (Лексическая Функциональная Грамматика): <http://decentius.aksis.uib.no/logon/xle.xml>

Занятие 8(5)

Корпуса текстов. Распространение эмпирических методов решения задач в КЛ. Метод n-gram – понятие статистической модели языка. Современное состояние МП.

Цель занятия: Системы МП, основанные на примерах - example-based MT (ЕВМТ), и системы типа «Переводческая память» (translation memory). Стадии обработки текста в системе МП, основанной на примерах (ЕВМТ). Статистические системы МП и понятие статистической модели языка. Развитие и современное состояние прямых систем МП на примере системы ПРОМПТ. Развитие систем типа «трансфер» в сторону гибридных систем. Развитие систем, основанных на языке-посреднике, в сторону систем, основанных на знаниях (КВМТ).

Основная литература:

Somers H. Machine translation: latest development // The Oxford handbook of computational linguistics (R. Mitkov ed.) N.Y.: Oxford university press, 2003. P. 512-528.

Интернет-источники для аудиторной работы:

ПРОМПТ: <http://translate.ru>

ЭТАПЗ : <http://proling.iitp.ru/>

GOOGLE. Translate : http://translate.google.com/translate_t#

Занятие 9(6)

Задачи семантики и типы семантических моделей в КЛ, связь с лексикографией.

Цель занятия: Семантические типы и семантические примитивы. Падежные грамматики. Концептуализации в теории Р. Шенка, и семантическая классификация процессов. Классификация процессов по М. Хэллiday. Связь семантических типов процессов с семантическими ролями.

Основная литература:

Кобозева И.М. Лингвистическая семантика. Эдиториал УРСС. Москва. 2000. Раздел: Семантические валентности лексемы как семантические отношения, обусловленные ее лексическим значением. С. 140-146.

Salton, Gerard. Automatic text processing. The transformation, analysis, and retrieval of information by computer. Addison-Wesley Publishing Company, Inc., 1989.

Занятие 10(6)

Лексико-семантические базы как технологии описания лексики.

Цель занятия: Представить онтологии как средство организации понятий в КЛ и рассмотреть основные традиции описания понятий: философская, когнитивная, лексикографическая, лексическая, информационно-поисковые тезаурусы. Познакомить с

основными технологиями описания парадигматических и синтагматических лингвистических отношений в проектах WordNet и FrameNet.

Основная литература:

Ляшевская О.Н., Кузнецова Ю.Л. Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312.

Интернет-источники для аудиторной работы:

и источники для аудиторной работы:

WordNet (онтология значений полнзначных слов английского языка и лексико-семантические парадигматические отношения : <http://wordnet.princeton.edu/> .

Corelex: (типы регулярной полисемии английских существительных) : <http://www.cs.brandeis.edu/~paulb/CoreLex/overview.html>.

FrameNet (значения и лексико-семантические синтагматические отношения английских предикатных слов в виде фреймов. Онтологические отношения фреймов) : <http://framenet.icsi.berkeley.edu/>.

Занятие 11(7)

Автоматическая генерация текстов на ЕЯ.

Цель занятия: Обсудить виды входных данных для систем автоматической генерации текстов и подходы к моделированию выходных текстов, схему идеализированной системы генерации текстов, соответствие между типами входных данных и типами генерируемых текстов. Познакомить с процессами автоматической генерации текстов: макро и микропланирование, лексикализация и грамматикализация, вставка ссылочных конструкций, агрегация, языковое оформление.

Основная литература:

Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке - состояние вопроса и прикладные системы // НТИ, Серия 2, №10, 2005, с.12-22.

Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке – теории, методы, технологии // НТИ, Серия 2, №7, 2006, с.1-15.

Занятие 12(7)

Теория риторических структур.

Цель занятия: Понятие дискурса и риторической структуры текста. Теория риторических структур (ТРС) как модель структуры дискурса текста: основные положения, определение риторического отношения, типы риторических отношений.

Основная литература:

Кибрик, А. А. Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. - 2009. - N 2. - С. 2-21.

Степанов Ю.С. Стил. // Лингвистический энциклопедический словарь. М., 1990. С.494-495.

Интернет-источники для аудиторной работы:

RST – (Rhetorical Structure Theory - Теория Риторических Структур): <http://www.sfu.ca/rst/>

Занятие 13(8)

Общие модели дискурса в КЛ.

Цель занятия: Обсудить общие модели структуры дискурса. Сравнение ТРС с другими моделями структуры дискурса: Теория связности дискурса Хоббса, модель структуры дискурса английского языка на основе системно-функциональной грамматики, моделирование методами формальной семантики – Discourse Representation Theory. Сравнение общих моделей дискурса.

Интернет-источники для аудиторной работы:

RST – (Rhetorical Structure Theory - Теория Риторических Структур): <http://www.sfu.ca/rst/>

Занятие 14(8)

Моделирование стиля и жанра текстов в ресурсах и системах КЛ.

Цель занятия: Обсудить моделирование понятий жанра для классификации текстов в европейских корпусах текстов и в НКРЯ, моделирование сюжета текстов разных жанров: волшебных сказок (В.Я. Пропп), новеллы (В. Ленерт), структуры специального текста - инструкции в системе генерации текстов AGILE. Представить лингвистическое понятие стиля и попытки формализации понятия стиля в КЛ, а также в системах генерации текстов. Попытки моделирования функциональных стилей в виде «конструкций» на основе статистической обработки текстов.

Основная литература:

Баранов А.Н. Введение в прикладную лингвистику. М., 2003. Раздел: «Моделирование структуры сюжета». С. 25–30.

Ю.С. Степанов. Стилъ // Лингвистический энциклопедический словарь. М., 1990. с. 494-495.

Володин А.П. и Храковский В.С. Семантика и типология императива. Русский императив. Л.: Наука, 1986.

Занятие 15(9)

Тематический аспект текста и поиск информации. Диалоговые и интерактивные системы.

Цель занятия: Выделить тематический аспект в текстах. Информационный поиск документов на основе метайнформации и автоматическая тематическая атрибуция текстов. Ввести понятие меры терминологичности $tf*idf$. Задачи информационного поиска: информационный поиск (ИП), фактографический поиск, поиск новой информации, распознавание именованных сущностей, отношений и тональности. Параметры полноты и точности ИП. Метод кластеризации как способ классификации документов и терминов. Лингвистические способы улучшения результатов статистических методов ИП. Применение диалоговых и интерактивных систем.

Основная литература:

Salton, Gerard. *Automatic text processing. The transformation, analysis, and retrieval of information by computer.* Addison-Wesley Publishing Company, Inc., 1989.

Занятие 16(10)

Автоматическая обработка устной речи.

Цель занятия: Сообщить основные принципы автоматической обработки устной речи, показать важность этого направления КЛ в современном обществе на примере приложений в быту, криминалистике, коммерции, производстве и т.д.

Основная литература:

Cole R. (ed.) Survey of the state of the Art in Human Language Technology. 1996. (<http://cslu.cse.ogi.edu/HLTsurvey/>). Chapter 1. Spoken language input.

“Речевые технологии” научно-практический журнал (2008-...) <http://speechtechnology.ru/general>

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список литературы

Основная литература

1. Боярский К.К. Введение в компьютерную лингвистику. Учебное пособие. – СПб ИТМО: 2013 – 73 стр. <http://elib.ict.nsc.ru/jspui/bitstream/ICT/1452/1/1470.pdf>
2. Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «ДИАЛОГ» - Москва, 2018 http://www.dialog-21.ru/media/4529/_dialog2018scopus.pdf
3. Соколова Е.Г. Синтаксическая разметка в терминах грамматики зависимостей и синтаксических функций [Электронный ресурс]: метод. пособие. М.: РГГУ, 2011. - 33 с. – Библиогр.: с. 33(5 назв.). – Режим доступа: <http://elib.lib.rsuh.ru/elib/000003603.pdf>
4. Chacon Thiago Costa. Improved **computational** models of sound change shed light on the history of the Tukanoan languages = Вопросы языкового родства / Thiago Costa Chacon, Johann-Mattis List // Вестник РГГУ. Серия "Филология. Вопросы языкового родства". - 2015. - № 3 (13). - С. 177-203. - Библиогр.: с. 200-203. - ил.
5. Church K.W., Mercer R.L. Introduction to the Special Issue on Computational Linguistics Using Large Corpora. // CL, vol. 19, 1993 – 24 p. <https://aclanthology.info/pdf/J/J93/J93-1001.pdf>
6. Stephen Hansen Michael McMahon Andrea Prat Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach *The Quarterly Journal of Economics*, Volume 133, Issue 2, 1 May 2018, Pages 801–870, <https://doi.org/10.1093/qje/qjx045>

Рекомендованная литература

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. Пособие /Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. -272 с.
2. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. Лингвистический процессор для сложных информационных систем/ М.: Наука, 1992. 256 с.
3. Баранов А.Н. Введение в прикладную лингвистику. М., 2003. (Разделы: Моделирование общения.С. 20-25, и Моделирование структуры сюжета. С. 25-30.)
4. Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке – теории, методы, технологии // НТИ, Серия 2, №7, 2006, с.1-15.
5. Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке - состояние вопроса и прикладные системы // НТИ, Серия 2, №10, 2005, с.12-22.
6. Буторов В.Д. Моделирование синтаксиса естественного языка // Прикладное языкознание. Учебник. (ред. А.С.Гердт). СПб., 1996. С. 142-160.

7. Виноград Т. Система, понимающая естественный язык. М: Мир, 1976. (Разделы: Образец диалога. С. 21-32; Синтаксис и значение. С. 32-41; Базовый подход к представлению значений. С. 44-47.)
8. Дерюгина О. Программы-собеседники. НТИ серия 1, N 6, стр. 31-35.
9. Жигалов В.А., Соколова Е.Г. InBASE: технология построения ЕЯ интерфейсов к базам данных // Труды Международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям Том 2, Аксаково, Июнь 2001 с. 123-135. Доступна с сайта: <http://dialog-21.ru/digest/archive/2001/?year=2001&vol=22725&id=6900>
10. Зализняк А.А. Русский грамматический словарь, Изд. 2-е. М.: Рус. словарь, 2003.
11. Зализняк Анна А. Многозначность и смежные понятия. //Анна А. Зализняк. Многозначность в языке и способы ее представления. М: Языки славянских культур, 2006. Глава 1, 1.1. с. 20-34.
12. Искусственный интеллект. Справочник: В 3 кн. Кн. 1-2. М.: Радио и связь, 1990.
13. Кибрик А.А. Анализ дискурса в когнитивной перспективе. [Электронный ресурс] 2003 // http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/DA_cognitive_perspective@Diss_2003.pdf
14. Кибрик, А. А. Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. - 2009. - N 2. - С. 2-21.
15. Кобозева И.М. Лингвистическая семантика. Эдиториал УРСС. Москва. 2000. Раздел: Семантические валентности лексемы как семантические отношения, обусловленные ее лексическим значением. С. 140-146.
16. Леонтьева Н.Н. Автоматическое понимание текстов. Системы. Модели. Ресурсы. М.: Academia, 2006.
17. Лингвистический энциклопедический словарь. М., 1990.
18. Ляшевская О.Н., Кузнецова Ю.Л. Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312.
19. Мельчук И.А. Об одной лингвистической модели типа «Смысл-Текст». Уровни представления языковых высказываний // Изв. АН СССР, Серия литературы и языка, 1974. Т. 33, №5-6. <http://feb-web.ru/feb/izvest/1974/05/745-436.htm>
20. Непейвода Н.Н. Квазиискусственный язык // Диалог'2002, Т 1, Москва: Наука. 2002. С.314-318. <http://www.dialog-21.ru/materials/archive.asp?id=7361&y=2002&vol=6077>.
21. Ножов И. Морфогическая и синтаксическая обработка текста (модели и программы)", 2003 год (диссертация). Глава 2. Доступна с сайта <http://aot.ru/technology.html>.
22. Соколова Е.Г. Синтаксическая разметка в терминах грамматики зависимостей и синтаксических функций [Электронный ресурс]: метод. пособие. М.: РГГУ, 2011. - 33 с. – Библиогр.: с. 33(5 назв.). – Режим доступа: <http://elib.lib.rsuh.ru/elib/000003603.pdf>
23. Тестелец Я.Г. Введение в общий синтаксис. М.: Изд. центр РГГУ, 2001. С. 213-215, 722-747.
24. Цибульский В.В., Ежов А.С., Поляков Г.А., Феклистов В.В. Анализ и классификация времени и сроков в российских нормативных и правовых актах. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 20012» (в печати).
25. Шаров С.А. Средства компьютерного представления лингвистической информации //1996. <http://www.ksu.ru/eng/science/ittc/vol000/002/>

26. Шенк Р. Обработка концептуальной информации. М.: Энергия, 1980.
27. Chris Manning and Hinrich Schütze, Chapter 7. Word sense disambiguation. Chapter 8. Lexical acquisition. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing
28. Hutchins J. Machine translation: general overview // *The Oxford handbook of computational linguistics* (R. Mitkov ed.) N.Y.: Oxford university press, 2003. P. 501-511.
29. Jurafsky, Daniel, and James H. Martin. 2009. Chapter 10-12 in: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice-Hall. Доступна с сайта <http://lib.mexmat.ru/books/10138>
30. Jurafsky, Daniel, and James H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd edition. Prentice-Hall. 2009.
31. Salton, Gerard. *Automatic text processing. The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Publishing Company, Inc., 1989.
32. *Survey of the state of the Art in Human Language Technology* (Ronald Cole, editor in chief) Cambridge University Press. 1997. См. также: (<http://cslu.cse.ogi.edu/HLTSurvey/>).
33. *The Oxford handbook of computational linguistics* (R. Mitkov ed.) N.Y.: Oxford university press, 2003.
34. Uszkoreit H. What is computational linguistics? http://www.coli.uni-saarland.de/~hansu/what_is_cl.html

Программное обеспечение и Интернет-ресурсы

Учебная дисциплина должна быть обеспечена учебно-методической документацией и материалами. Обязательная литература должна быть представлена в библиотеке ВУЗа, сети Интернет или локальной сети ВУЗа (факультета). Для обучающихся должен быть обеспечен онлайн доступ к Интернет источникам и системам. В частности, в процессе обучения используются следующие Интернет-ресурсы:

Интернет-источники:

А. Ресурсы:

- Русско-английский тезаурус по компьютерной лингвистике. Подобласть знаний: <http://uniserv.iis.nsk.su/thes/index.php?ent=74>.
- Национальный корпус русского языка (НКРЯ): <http://www.ruscorpora.ru>
Синтаксический подкорпус НКРЯ: <http://www.ruscorpora.ru/search-syntax.html>
Параллельный подкорпус НКРЯ:
- Малый Академический Словарь (МАС): <http://feb-web.ru/feb/mas/mas-abc/default.asp>.
- Лексико-семантические базы:
 - WordNet** (онтология значений полнзначных слов английского языка и лексико-семантические парадигматические отношения : <http://wordnet.princeton.edu/> .
 - Corelex**: (типы регулярной полисемии английских существительных) : <http://www.cs.brandeis.edu/~paulb/CoreLex/overview.html>.
 - FrameNet** (значения и лексико-семантические синтагматические отношения английских предикатных слов в виде фреймов. Онтологические отношения фреймов) : <http://framenet.icsi.berkeley.edu/>.

Б. Системы анализа предложения по уровням:

- морфологический, синтаксический, поверхностный семантический:
Dialing : <http://aot.ru/>
- семантико-синтаксический:
HPSG (Вершинная грамматика непосредственных составляющих):
<http://www2.lingsoft.fi/cgi-bin/engcg?snt=Baby%2C+I+love+you.&h=on>.
Он-лайн демо: <http://erg.emmtee.net>.
LFG (Лексическая Функциональная Грамматика):
<http://decentius.aksis.uib.no/logon/xle.xml>

В. Прикладные системы он-лайн:

- Программы-собеседники:
Eliza («компьютерный психотерапевт») Дж. Вейценбаума:
<http://www.manifestation.com/neurotoys/eliza.php3>.
- Вопросно-ответные системы:
START (в MIT) : <http://start.csail.mit.edu/>
- Машинный перевод:
ПРОМПТ (коммерческая, прямой): <http://translate.ru>
ЭТАПЗ (экспериментальная, трансфер): <http://proling.iitp.ru/>
Dialing Translator (экспериментальная) : <http://aot.ru/cgi-bin/translate.cgi>
GOOGLE. Translate (коммерческая, статистический):
http://translate.google.com/translate_t#
- Анализ и поиск текстовой информации): <http://demo.rco.ru/>
RCO : <http://demo.rco.ru/>.

Г. Модели и теории (на Интернет сайтах):

- **RST** – (Rhetorical Structure Theory - Теория Риторических Структур):
<http://www.sfu.ca/rst/>.
- **UNL** (Universal Networking Language): <http://www.undl.org/>.

7. Материально-техническое обеспечение дисциплины

Занятия по курсу «Введение в компьютерную лингвистику» можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимо также наличие доски, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

При необходимости рабочая программа дисциплины может быть адаптирована для обеспечения образовательного процесса лицам с ограниченными возможностями здоровья, в том числе для дистанционного обучения. Для этого от студента требуется представить заключение психолого-медико-педагогической комиссии (ПМПК) и личное заявление (заявление законного представителя).

В заключении ПМПК должно быть прописано:

- рекомендуемая учебная нагрузка на обучающегося (количество дней в неделю, часов в день);
- оборудование технических условий (при необходимости);
- сопровождение и (или) присутствие родителей (законных представителей) во время учебного процесса (при необходимости);
- организация психолого-педагогического сопровождение обучающегося с указанием специалистов и допустимой нагрузки (количества часов в неделю).

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся, при необходимости могут быть созданы фонды оценочных средств, адаптированные для лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе.

Форма проведения текущей и итоговой аттестации для лиц с ограниченными возможностями здоровья устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно (на бумаге, на компьютере), в форме тестирования и т.п.). При необходимости студенту предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

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. Приложения

Приложение 1. Аннотация дисциплины

Предметом дисциплины (модуля) является ознакомление с методами и лингвистическими технологиями, применяемыми при создании компьютерных систем обработки текстов в научно-практической области исследований «компьютерная лингвистика», и рассматриваемых в сопоставлении с лингвистическими и культурно-

лингвистическими свойствами языковых произведений – предложений и текстов, а также в связи с задачами обработки текстов как социальными запросами общества. Подача материала частично увязана с историей компьютерной лингвистики, что позволяет лучше представить качественную составляющую процессов, моделируемых современными лингвистическими технологиями, изучаемыми в других курсах по профилю «Компьютерная лингвистика», основанных преимущественно на эмпирических, в частности, статистических методах.

Курс направлен на решение следующих задач:

- создать представление о компьютерной лингвистике как новейшей научно-практической области исследований, ее возникновении в контексте смежных наук и ее современной организации;
- познакомить магистрантов с основными лингвистическими технологиями, реализующими анализ предложения (текста) по уровням лингвистической разметки и основными приемами автоматической генерации текстов;
- познакомить магистрантов с основными типами ресурсов, создающимися и используемыми компьютерными программами для решения конкретных задач в исследовательских целях, при разработке лингвистических технологий и в приложениях;
- соединить интуитивные и традиционные представления о свойствах естественно-языковых текстов со способами их формализации и моделирования в работах по компьютерной лингвистике;
- выработать у магистрантов элементарные практические навыки по применению компьютерно-лингвистических методов к языковому материалу и использованию лингвистических технологий.

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к абстрактному мышлению, анализу, синтезу (ОК-1);

готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

готовностью к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью анализировать, сопоставлять и критически оценивать различные лингвистические направления, теории и гипотезы (ОПК-5);

способностью адаптироваться к новым теориям и результатам мировой науки и расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук (ОПК-6);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и

компьютерной лингвистики (ПК-1);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью разрабатывать лингвистические компоненты электронных языковых ресурсов (лингвистические корпуса, словари, онтологии, базы данных) (ПК-7)

и соотношенных с ними результатов освоения дисциплины (модуля):

Знать:

- структуру научно-практической области исследований «компьютерная лингвистика» и ее место в контексте смежных наук, цели этой области и условия ее появления и развития;
- основные методы компьютерной лингвистики и роль эмпирических методов на современном этапе;
- состав уровней технологий компьютерной лингвистики в соответствии со свойствами и организацией лингвистических объектов;
- основные методы и технологии анализа и генерации текстов;
- основные типы лингвистических ресурсов, используемых лингвистическими технологиями;
- существенные для передачи информации свойства текстов и их моделирование в компьютерной лингвистике.

Уметь:

- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- применять полученные знания в области моделирования конкретных процессов анализа и генерации текстов в научно-исследовательской и других видах практической деятельности;
- осуществлять тестирование и оценку основных ресурсов и лингвистических технологий анализа, и генерации текстов.

Владеть:

- самыми общими методами и ресурсами обработки текстов при решении задач компьютерной лингвистики в исследовательской и практической работе.

Дисциплина (модуль) «Введение в компьютерную лингвистику» является базовой частью цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика: компьютерная лингвистика», «Фундаментальная и прикладная лингвистика: теория языка» и адресована студентам 1 курса (1 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение домашних заданий; тестовое задание; защита исследовательского проекта; промежуточная аттестация в форме: экзамен.

Общая трудоемкость освоения дисциплины (модуля) составляет 3 зачетные единицы,

108 часов.

Программой дисциплины (модуля) предусмотрены: лекционные занятия – 8 часов; практические занятия – 22 часа; самостоятельная работа студента – 60 часов, контроль – 18 часов.

Приложение 2. Лист изменений

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение к листу изменений №1	31.08.2020г	1

Приложение к листу изменений №1

1. Структура дисциплины (к п. 2 РПД на 2020)

Общая трудоёмкость дисциплины составляет 3 з.е., 114 ч., в том числе контактная работа обучающихся с преподавателем 30 ч., самостоятельная работа обучающихся 66 ч.

2. Образовательные технологии (к п.4 на 2020 г.)

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.

3. Перечень БД и ИСС (к п. 6 на 2020 г.)

№п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru

4. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)

1	Microsoft Share Point 2010	Microsoft	лицензионное
2	Windows 10 Pro	Microsoft	лицензионное
3	Kaspersky Endpoint Security	Kaspersky	лицензионное
4	Microsoft Office 2016	Microsoft	лицензионное
5	Zoom	Zoom	лицензионное