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

«Программирование лингвистических задач. Основные алгоритмы лингвистического анализа»

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

Программирование лингвистических задач.

Основные алгоритмы лингвистического анализа

Рабочая программа дисциплины

Составитель:

В.П.Селегей

Ответственный редактор:

д. филол. н., профессор В.И.Подлесская

УТВЕРЖДЕНО
Протокол заседания УНЦ компьютерной лингвистики

№ 1 от «28» августа 2019г.

Оглавление

1. Пояснительная записка

1.1 Предмет

Предметом дисциплины (модуля) является изучение основных алгоритмов лингвистического анализа, предназначенных для компьютерной обработки лингвистических данных, а также формальных математических моделей, лежащих в основе данных методов. Курс частично увязан с курсом «Машинное обучение», прикладные программы, реализующие методы и принципы, изучаемые в настоящей дисциплине, рассматриваются в курсе «Прикладные пакеты для лингвистических исследований». В курсе подробно разбирается то, как соотносятся лингвистические и технические соображения при решении конкретных прикладных задач, какие математические методы лучше всего подходят для той или иной проблемы, каким образом реальный языковой материал определяет выбор метода и его последующую реализацию, изучаются как подходы, основанные на лингвистически мотивированных правилах, так и статистические методы, привлекающие лингвистику лишь в качестве дополнительного инструментов.

1.2 Цель и задачи курса

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными математическими методами, применяемыми для решения лингвистических задач, а также с программными продуктами, реализующими данные методы;
- познакомить магистрантов с основными подходами к решению задач прикладной лингвистики (правиловым и статистическим), а также изучить соотношение данных подходов для конкретных проблем;
- познакомить магистрантов с математическими методами, лежащими в основе алгоритмов лингвистического анализа и влиянием лингвистического материала на выбор метода, а также влиянием выбранного метода на полученные результаты;
- научить магистрантов как предварительно выбирать алгоритм решения для прикладных лингвистических задач, так и дорабатывать выбранный алгоритм в зависимости от специфики задачи;
- выработать у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя как научные статьи, так и более специализированные технические материалы.

1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);
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 изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);
способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8)

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 алгоритмы, используемые для решения стандартных задач компьютерной лингвистики, таких как автоматический морфологический и синтаксический анализ, анализ тональности, исправление опечаток и т. д., а также структуру данных, используемых в данных алгоритмов;
- математические модели, лежащие в основе основных алгоритмов анализа лингвистических данных, а также применимость данных алгоритмов на материале конкретных задач для разных языков;
- существенные с вычислительной точки зрения лингвистические свойства текстов и другого языкового материала;
- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии применяются в алгоритмах лингвистического анализа;

Уметь:
- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм для решения той или иной задачи прикладной лингвистики, а также обосновывать его выбор;
- анализировать результаты применения компьютерных алгоритмов к лингвистическим данным;
- модифицировать выбранный алгоритм в зависимости от результатов его работы
- подбирать данные для обучения выбранного алгоритма в случае, если он основан на статистических методах

Владеть:
- основными методами обработки лингвистических данных в зависимости от предметной области.

1.4 Место дисциплины в структуре образовательной программы

Дисциплина (модуль) *«Программирование лингвистических задач. Основные алгоритмы лингвистического анализа»* является *вариативной* частью *профессионального* цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам *1 курса (2 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 *2* зачетные единицы, *72* часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – *20 часов*; самостоятельная работа студента – *34 часа*, 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)*
				лекции	семинары	самостоятельная работа	контроль	
1.	Теория формальных языков. Регулярные выражения.	1	1		2	4		ДЗ1. Описание языковых явлений с помощью регулярных выражений
2.	Конечные автоматы. Конечные преобразователи	1	2	2	1	2		ДЗ2. Описание языковых явлений с помощью конечных автоматов и преобразователей
3.	Контекстно-свободные грамматики. Алгоритмы проверки выводимости.	1	3		2	4		.
4.	Грамматики зависимостей. Применение формальных грамматик для описания синтаксических явлений.	1	4		2	2		ДЗ3. Проверка выводимости в грамматике зависимостей.
5.	Энграммные модели. Методы сглаживания.	1	5		1	2		ДЗ4. Моделирование в программе FOMA системы глагольного словоизменения.
6.	Скрытые марковские модели. Основные проблемы для	1	6	2	2	4		ДЗ5. Нахождение оптимальных параметров скрытой мар-

	скрытых марковских моделей.							ковской модели.
7.	Применение скрытых марковских моделей для автоматического морфологического анализа	1	7	2	2	2		
8.	Расстояние Левенштейна и автоматическое исправление опечаток	1	8		2	4		ДЗ6. Вычисление расстояния Левенштейна
9.	Применения марковских моделей для решения различных задач компьютерной лингвистики. Условные случайные поля.	1	9	2	2	4		
10.	Алгоритмы для автоматического извлечения парадигм и автоматического деления на морфемы.	1	10		2	4		
11.	Экзамен	1	11		2	4		Контрольные вопросы
	Итого:				20	34	18	

3. Содержание дисциплины

1. Теория формальных языков. Регулярные выражения.

Формальные языки, основные операции над ними. Понятие регулярного выражения. Задание формальных языков регулярными выражениями.

2. Конечные автоматы. Конечные преобразователи

Конечные автоматы. Детерминированные и недетерминированные конечные автоматы. Алгоритм детерминизации. Эквивалентность регулярных выражений и конечных автоматов. Языки, не задаваемые конечными автоматами.

3. Контекстно-свободные грамматики. Алгоритмы проверки выводимости.

Контекстно-свободные грамматики. Вывод в контекстно-свободной грамматике. Нормальная форма Хомского. Алгоритм Кока-Янгера-Касами проверки выводимости. Алгоритм «перенос-свёртка».

4. Грамматики зависимостей. Применение формальных грамматик для описания синтаксических явлений.

Грамматики зависимостей. Проективные и непроективные зависимости. Связь грамматик зависимостей с контекстно-свободными грамматиками. Алгоритм «перенос-свёртка» для грамматик зависимостей.

5. Энграммные модели. Методы сглаживания.

Определение вероятности текста, энграммные модели. Определение контекстных вероятностей, алгоритмы сглаживания (аддитивное, Уиттена-Белла, Кнезера-Нея). Применение энграммных моделей. Автоматическая генерация текста.

6. Скрытые марковские модели. Основные проблемы для скрытых марковских моделей.

Марковские цепи, определение вероятности последовательности в марковской цепи и наиболее вероятного состояния. Скрытые марковские модели. Связь энграммыных моделей и скрытых марковских моделей. Нахождение вероятности последовательности, алгоритм Витерби. Алгоритм максимизации ожидания для определения параметров модели.

7. Применение скрытых марковских моделей для автоматического морфологического анализа. Интерпретация параметров марковской модели в терминах морфологического анализа. Достоинства и недостатки марковской модели.

8. Расстояние Левенштейна и автоматическое исправление опечаток.

Расстояние Левенштейна. Динамический алгоритм для его вычисления. Варианты расстояния Левенштейна. Применение расстояния Левенштейна для исправления опечаток: порождение вариантов. Выбор наиболее вероятного исправления, контекстные вероятности и другие подходы.

9. Применения марковских моделей для решения различных задач компьютерной лингвистики. Условные случайные поля.

Модель канала связи, её связь с марковскими моделями. Интерпретация алгоритмов исправления опечаток и машинного перевода в терминах марковских моделей. Интеграция признаков в марковские модели: условные случайные поля. Алгоритм Витерби для условных случайных полей.

10. Алгоритмы для автоматического извлечения парадигм и автоматического деления на морфемы.

Применение условных случайных полей для автоматического деления на морфемы. Автоматическое определение словоизменительной парадигмы, метод наибольшей общей подпоследовательности. Символьные энграммы и их применение для автоматической лемматизации.

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:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с отчетом по результатам исследования);

2. он-лайн демонстрации работы с лингвистическими интернет-источниками (и др.);
3. он-лайн использование лингвистических ресурсов (Национальный корпус русского языка, Лексико-семантические базы и др.);
4. обсуждения курсовых исследовательских проектов;
5. 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:

<table>
<tr><th>100-балльная шкала</th><th colspan="2">Традиционная шкала</th><th>Шкала ECTS</th></tr>
<tr><td>95 – 100</td><td rowspan="2">отлично</td><td rowspan="5">зачтено</td><td>A</td></tr>
<tr><td>83 – 94</td><td>B</td></tr>
<tr><td>68 – 82</td><td>хорошо</td><td>C</td></tr>
<tr><td>56 – 67</td><td rowspan="2">удовлетворительно</td><td>D</td></tr>
<tr><td>50 – 55</td><td>E</td></tr>
<tr><td>20 – 49</td><td rowspan="2">неудовлетворительно</td><td rowspan="2">не зачтено</td><td>FX</td></tr>
<tr><td>0 – 19</td><td>F</td></tr>
</table>

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)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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 максимальному. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой,

Количество баллов	Критерии оценки
	справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Свободно ориентируется в учебной и профессиональной литературе. Компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».
83−94 (B)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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. Обучающийся адекватно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения. Достаточно свободно ориентируется в учебной и профессиональной литературе. Почти все компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».
68–82 (C)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Компетенции, закреплённые за дисциплиной, сформированы на уровне «хороший».
56–67 (D)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками

Количество баллов	Критерии оценки
	и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
50–55 (E)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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. 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. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».
21–49 (FX)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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. 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. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.
0–20 (F)	Оценка выставляется с учетом текущей и промежуточной аттестации. 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 учебных заданий. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических

Количество баллов	Критерии оценки
	задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Текущий контроль успеваемости студентов проводится в следующих формах: выполнение домашних заданий (6 заданий – 60 баллов максимум); теоретический зачёт (40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов.

В качестве домашних заданий предлагаются задания следующих типов

ДЗ1. Описание языковых явлений с помощью регулярных выражений.
ДЗ2. Описание языковых явлений с помощью конечных автоматов и преобразователей.
ДЗ3. Проверка выводимости в грамматике зависимостей.
ДЗ4. Моделирование в программе FOMA системы глагольного словоизменения.
ДЗ5. Нахождение оптимальных параметров скрытой марковской модели.
ДЗ6. Вычисление расстояния Левен-штейна.

Экзамен ориентирован на следующие контрольные вопросы

Регулярные выражения.
Конечные автоматы, связь с регулярными выражениями.
Примеры языков, не задаваемых конечными автоматами.
Контекстно-свободные грамматики.
Алгоритм проверки выводимости в контекстно-свободной грамматике.
Грамматики зависимостей.
Алгоритм проверки выводимости в грамматике зависимостей.
Энграммные модели, методы сглаживания вероятностей.
Скрытые марковские модели, определения, примеры.
Нахождение вероятности выходной последовательности в марковской модели.
Нахождение наиболее вероятной скрытой последовательности в марковской модели.
Расстояние Левенштейна.
Поиск близких слов в словаре с помощью расстояния Левенштейна.
Контекстное исправление опечаток.
Условные случайные поля и алгоритм Витерби для них.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список литературы

Основная литература

1. Филенко К. В Лингвистический анализ названий кинопродукции III рейха в контексте политического дискурса[Текст] = Linguistic analysis of the Third Reich movie titles in the context of political discourse / К. В. Филенко // Политическая лингвистика. - 2015. - № 1 (51). - С. 231-237. - Библиогр.: с. 236 (17 назв.).
2. Быкова В. В. Алгоритмы концептуального моделирования и классификации текстов в корпусе тувинского языка[Текст] = Algorithms of conceptual modeling and text classification in the tuvan language corpus / В. В. Быкова, Ч. М. Монгуш // Программные продукты и системы. - 2017. - Т. 30, № 3. - С. 487-495. - Библиогр.: с. 495 (14 назв.). - ил.: 2 рис., 4 табл.
3. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge.
4. Jeffrey Friedl, Mastering Regular Expressions, published by O'Reilly
5. Ilya Segalovich. "A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine"
6. Helmut Schmid and Florian Laws. "Estimation of Conditional Probabilities With Decision Trees and an Application to Fine-Grained POS Tagging"
7. Thorsten Brants. "TnT -- A Statistical Part-of-Speech Tagger"
8. Tomita Parser, руководство по использованию. https://tech.yandex.ru/tomita/
9. MaltParser - a data-driven dependency parser. http://www.maltparser.org/
10. MULTEXT-East Morphosyntactic Specifications, Version 4. http://nl.ijs.si/ME/V4/msd/html/msd.html
11. О. Н. Ляшевская, С. А. Шаров. Новый частотный словарь русской лексики. http://dict.ruslang.ru/freq.pdf
12. Numpy and Scipy Documentation: https://docs.scipy.org/doc/
13. Manning C. D. et al. Foundations of statistical natural language processing. – Cambridge: MIT press, 1999

Рекомендованная литература

1. Steven Bird, Ewan Klein, and Edward Loper. "Natural Language Processing with Python", O'Reilly, 2nd Edition
2. Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge.
3. Jeffrey Friedl, Mastering Regular Expressions, published by O'Reilly
4. Ilya Segalovich. "A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine"
5. Helmut Schmid and Florian Laws. "Estimation of Conditional Probabilities With Decision Trees and an Application to Fine-Grained POS Tagging"
6. Thorsten Brants. "TnT -- A Statistical Part-of-Speech Tagger"
7. Tomita Parser, руководство по использованию. https://tech.yandex.ru/tomita/
8. MaltParser - a data-driven dependency parser. http://www.maltparser.org/
9. MULTEXT-East Morphosyntactic Specifications, Version 4. http://nl.ijs.si/ME/V4/msd/html/msd.html
10. О. Н. Ляшевская, С. А. Шаров. Новый частотный словарь русской лексики. http://dict.ruslang.ru/freq.pdf
11. Numpy and Scipy Documentation: https://docs.scipy.org/doc/
12. М. Р. Пентус, А. Е. Пентус. Математическая теория формальных языков.М., Интуит, 2006.

13. Прикладная и компьютерная лингвистика, Под. ред. А. В. Митрениной. М., УРСС, 2016
14. Manning C. D. et al. Foundations of statistical natural language processing. – Cambridge: MIT press, 1999
15. Jurafsky D., Martin, J. H. Speech & language processing. – Pearson Education India, 2000.

7. Материально-техническое обеспечение дисциплины

Занятия по курсу «Программирование лингвистических задач. Основные алгоритмы лингвистического анализа» 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, регулируемыми эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.

9. Приложения

Приложение 1. Аннотация дисциплины

Предметом дисциплины (модуля) является изучение основных алгоритмов лингвистического анализа, предназначенных для компьютерной обработки лингвистических данных, а также формальных математических моделей, лежащих в основе данных методов. Курс частично увязан с курсом «Машинное обучение», прикладные программы, реализующие методы и принципы, изучаемые в настоящей дисциплине, рассматриваются в курсе «Прикладные пакеты для лингвистических исследований». В курсе подробно разбирается то, как соотносятся лингвистические и технические соображения при решении конкретных прикладных задач, какие математические методы лучше всего подходят для той или иной проблемы, каким образом реальный языковой материал определяет выбор метода и его последующую реализацию, изучаются как подходы, основанные на лингвистически мотивированных правилах, так и статистические методы, привлекающие лингвистику лишь в качестве дополнительного инструментов.

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными математическими методами, применяемыми для решения лингвистических задач, а также с программными продуктами, реализующими данные методы;
- познакомить магистрантов с основными подходами к решению задач прикладной лингвистики (правиловым и статистическим), а также изучить соотношение данных подходов для конкретных проблем;
- познакомить магистрантов с математическими методами, лежащими в основе алгоритмов лингвистического анализа и влиянием лингвистического материала на выбор метода, а также влиянием выбранного метода на полученные результаты;
- научить магистрантов как предварительно выбирать алгоритм решения для прикладных лингвистических задач, так и дорабатывать выбранный алгоритм в зависимости от специфики задачи;
- выработать у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя как научные статьи, так и более специализированные технические материалы.

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);
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 изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);
способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты

интеллектуальных и информационных электронных систем (ПК-8)

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 алгоритмы, используемые для решения стандартных задач компьютерной лингвистики, таких как автоматический морфологический и синтаксический анализ, анализ тональности, исправление опечаток и т. д., а также структуру данных, используемых в данных алгоритмов;
- математические модели, лежащие в основе основных алгоритмов анализа лингвистических данных, а также применимость данных алгоритмов на материале конкретных задач для разных языков;
- существенные с вычислительной точки зрения лингвистические свойства текстов и другого языкового материала;
- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии применяются в алгоритмах лингвистического анализа;

Уметь:

- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм для решения той или иной задачи прикладной лингвистики, а также обосновывать его выбор;
- анализировать результаты применения компьютерных алгоритмов к лингвистическим данным;
- модифицировать выбранный алгоритм в зависимости от результатов его работы
- подбирать данные для обучения выбранного алгоритма в случае, если он основан на статистических методах

Владеть:

- основными методами обработки лингвистических данных в зависимости от предметной области.

Дисциплина (модуль) *«Программирование лингвистических задач. Основные алгоритмы лингвистического анализа»* является *вариативной* частью *профессионального* цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам *1 курса (2 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 *2* зачетные единицы, *72* часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – *20 часов*; самостоятельная работа студента – *34 часа*, 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 2 з.е., 76 ч., в том числе контактная работа обучающихся с преподавателем 20 ч., самостоятельная работа обучающихся 38 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