



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Дальневосточный федеральный университет»
(ДФУ)
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

«СОГЛАСОВАНО»

Руководитель ОП

Подпись

Дремлюга Р.И.

УТВЕРЖДАЮ

И.о директора Академии цифровой трансформации

Еременко А.С.

«26» января 2022 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Компьютерная лингвистика и обработка естественных языков

Направление подготовки - 09.04.01 Информатика и вычислительная техника
(Искусственный интеллект и большие данные (совместно с ПАО Сбербанк))

Форма подготовки очная

курс 2 семестр 3
лекции 18 час.
практические занятия 36 час.
лабораторные работы 00 час.
в том числе с использованием МАО практические занятия 00 час
всего часов аудиторной нагрузки 54 час.
самостоятельная работа 90 час.
в том числе на подготовку к экзамену 36 час.
контрольные работы (количество) не предусмотрены
курсовая работа / курсовой проект не предусмотрены
зачет с оценкой не предусмотрен
экзамен 3 семестр

Рабочая программа дисциплины разработана в соответствии с федеральным государственным образовательным стандартом (ФГОС) высшего образования (ВО) – магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника, утвержденным приказом Министерства образования и науки Российской Федерации от 19.09.2017 г. № 918 (с изменениями и дополнениями).

Рабочая программа обсуждена на заседании Академии цифровой трансформации протокол № 1 от 25 января 2022 г.

И.о директора Академии цифровой трансформации

Еременко А.С.

Составители: к.т.н. Еременко А.С., Загумённов А.А.

Владивосток
2022

Оборотная сторона титульного листа РПУД

I. Рабочая программа пересмотрена на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

II. Рабочая программа пересмотрена на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

III. Рабочая программа пересмотрена на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

IV. Рабочая программа пересмотрена на на заседании департамента:

Протокол от « ____ » _____ 20__ г. № _____

Директор департамента _____
(подпись) (И.О. Фамилия)

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель курса - изучение различных моделей автоматической обработки текста на естественном языке, применяемых в современных информационных системах и затрагивающих несколько языковых уровней обрабатываемого текста, включая уровни морфологии, синтаксиса, дискурса и семантики.

Задачи курса:

• знать основные уровни анализа и синтеза текста на естественном языке, существующие модели статистического, морфологического и синтаксического анализа текстов и их применение в типичных прикладных программных системах обработки текстов;

• понимать существенные отличия естественных языков от искусственных и особенности компьютерных моделей естественного языка;

• познакомить с принципами построения различных лингвистических ресурсов, включая корпуса текстов, терминологические словари, тезаурусы, онтологии;

• изучить виды лингвистических ресурсов, используемых при обработке текстов, и методы их создания;

• научиться решать прикладные задачи, требующие многоуровневого анализа и синтеза текста (такие как машинный перевод, генерация текста, извлечение информации и знаний из текста).

В результате данной дисциплины у обучающихся формируются следующие универсальные и профессиональные компетенции:

Наименование категории (группы) компетенций	Код и наименование компетенции выпускника	Код и наименование индикатора достижения компетенции
Коммуникация	УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1 Создает различные типы письменных и устных текстов на русском и иностранном языке для академического и профессионального взаимодействия

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
УК-4.1 Создает различные типы письменных и устных текстов на русском и иностранном языке для академического и профессионального взаимодействия	Знает основные специальные термины и грамматические конструкции для работы с оригинальными текстами академического и профессионального характера
	Умеет строить лексически правильно, грамотно, логично и последовательно устные и письменные высказывания в ситуациях академического и профессионального взаимодействия
	Владеет навыками построения лексически правильного, грамотного, логичного и последовательного устного и письменного высказывания в ситуациях академического и профессионального взаимодействия на английском языке

Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
ПК-1 Способен проектировать и разрабатывать системные и прикладные решения по анализу больших данных	ПК-1.1 Владеет инструментарием получения, хранения, передачи и обработки больших данных
	ПК-1.2 Формулирует и решает системные и прикладные задачи анализа больших данных для конкретных предметных областей
	ПК-1.3 Способен управлять разработкой продуктов, услуг и решений на основе больших данных
ПК-2 Способен разрабатывать методики выполнения аналитических работ	ПК-2.1 Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации; описывать методики выполнения аналитических работ
	ПК-2.2 Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
ПК-1.1 Владеет инструментарием получения, хранения, передачи и обработки больших данных	Знает технологии, методы и инструментальные средства обработки больших данных
	Умеет использовать архитектуры и модели баз и хранилищ данных, адаптированные к технологиям больших данных
	Владеет навыками разработки предложений по развитию и совершенствованию системы получения, хранения, передачи, обработки больших данных
ПК-1.2 Формулирует и решает системные и прикладные задачи	Знает существующий опыт разработки и использования продуктов и услуг на основе технологий больших данных

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
анализа больших данных для конкретных предметных областей	Умеет разрабатывать проектную и рабочую документацию на разработку аналитических услуг на основе технологий больших данных
	Владеет навыками решения прикладных задач анализа больших данных для конкретных предметных областей
ПК-1.3 Способен управлять разработкой продуктов, услуг и решений на основе больших данных	Знает существующий опыт разработки и использования продуктов и услуг на основе технологий больших данных
	Умеет управлять исполнением проектных работ в области больших данных
	Владеет навыками создания прототипа сервиса на основе аналитики больших данных
ПК-2.1 Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации; описывать методики выполнения аналитических работ	Знает основные методики и практики выполнения аналитических работ
	Умеет выявлять проблемы и сложности в существующих практиках выполнения аналитических работ в организации
	Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах
ПК-2.2 Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах	Знает методы, применяемые для функционального и оперативного управления предприятиями
	Планировать проектные работы
	Владеет навыками выполнения аналитических работ, их апробации и доработки на выбранных проектах

II. ТРУДОЁМКОСТЬ ДИСЦИПЛИНЫ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ ПО ДИСЦИПЛИНЕ

Общая трудоемкость дисциплины составляет 4 зачётных единицы (144 академических часа).

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Лек	Лекции
Пр	Практические занятия
СР	Самостоятельная работа обучающегося в период теоретического обучения

Структура дисциплины:

Форма обучения – очная.

№	Наименование раздела дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося					Формы текущего контроля успеваемости и промежуточной аттестации	
			Лек	Лаб	Пр	ОК	СР		Контроль
1	Лингвистическое моделирование	3	3		6		54	36	УО-1, ПР-7;
2	Основные задачи и проблемы анализа естественно-языковых текстов	3	6		12				
3	Машинный перевод и другие прикладные задачи компьютерной лингвистики	3	6		12				
4	Современные методы и средства глубокого семантического анализа текста	3	3		6				УО-1, ПР-9;
Итого:			18		36		54	36	Экзамен

III. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА (18 часов)

Тема 1. Лингвистическое моделирование

Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика.

Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» как фундамент для построения систем автоматической обработки текста.

Тема 2. Основные задачи и проблемы анализа естественно-языковых текстов

Грамматика и словарь естественного языка. Представление об интегральном описании языка. Представление о лексических функциях.

Краткий обзор формальных грамматик. Порождающие грамматики. Грамматики составляющих и грамматики зависимостей. Гибридные грамматики.

Анализ и синтез текста. Морфологический и синтаксический анализ. Парсинг. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх».

Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Интерактивное разрешение лексической и синтаксической неоднозначности.

Правилковые и статистические подходы к автоматической обработке текста.

Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы. Синтаксическая структура предложения.

Тема 3. Машинный перевод и другие прикладные задачи компьютерной лингвистики

Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке. Система машинного перевода как механизм обратной связи и источник новых лингвистических знаний.

Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва (на примере UNL-универсального сетевого языка). Правилковый, статистический и гибридный перевод.

Морфологический компонент системы автоматической обработки текстов. Морфологическая структура слова и предложения.

Словарь системы автоматической обработки текстов. Словарь системы машинного перевода.

Структура словарной статьи. Синтаксические признаки. Семантические признаки (дескрипторы). Теория валентностей. Модель управления.

Аннотированные корпуса текстов и их роль в задачах автоматической обработки текстов.

Синонимическое перифразирование высказываний и его прикладное значение.

Обзор задач прикладной лингвистики.

Современные цифровые лингвистические ресурсы (Word Net, Frame Net, Treebanks).

Тема 4. Современные методы и средства глубокого семантического анализа текста

Современные методы глубокого семантического анализа текста с участием лингвистических онтологий. Умозаключения на основе здравого смысла (common sense reasoning).

IV. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Практическая часть курса включает в себя практические работы (36 часов), в процессе которых решаются конкретные прикладные задачи анализа и обработки естественного языка.

№	Тема (раздел) дисциплины	Виды учебных занятий, включая самостоятельную работу				
		Лекции	Практич. (семинар.) занятия	Лаборат. работы	Задания, курсовые работы	Самост. работа
1	Лингвистическое моделирование	3	6			13
2	Основные задачи и проблемы анализа естественно-языковых текстов	6	12			14
3	Машинный перевод и другие прикладные задачи компьютерной лингвистики	6	12			14
4	Современные методы и средства глубокого семантического анализа текста	3	6			13
Итого часов		18	36			54

Практическая работа № 1. Разработка и тестирование морфологического анализатора.

Практическая работа № 2. Разработка и тестирование метода автоматического реферирования.

Практическая работа № 3. Разработка и тестирование определения тональности текста.

Практическая работа № 4. Разработка и тестирование методов выделения упоминаний персон и организаций в тексте.

Практическая работа № 5. Обучение модели перевода на основе параллельного корпуса и ее оценка.

V. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

План-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию:

Примерная дата проведения	Наименование контрольного мероприятия	Форма контроля	Нормы времени на выполнение
1-6 недели	Работа с конспектом	Просмотр и проверка выполнения самостоятельной работы преподавателем, обсуждение результатов выполненной работы на занятии	20 час.
7-12 недели	Подготовка практических работ	Отчет о практической работе	20 час.
13-18 недели	Подготовка к зачету/экзамену	Зачет с оценкой	14 час.
Итого			54 часа

Учебно-методическое обеспечение самостоятельной работы студента по дисциплине «Компьютерная лингвистика и обработка естественных языков» предусматривает:

- поиск дополнительной литературы, к которой студенты могут прибегать при возникновении особой заинтересованности в конкретной теме;
- определение перечня контрольных вопросов, позволяющих студентам самостоятельно проверить качество полученных знаний;
- организацию консультаций преподавателя со студентами для разъяснения вопросов, вызывающих у студентов затруднения при самостоятельном освоении учебного материала.

Дополнительными формами самостоятельной работы являются групповые и индивидуальные задания, выступающие продолжением аудиторных занятий и направленные на овладение практическими навыками по основным разделам дисциплины.

Методические рекомендации к работе с литературными источниками

В процессе подготовки к практическим занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной)

литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Критерии оценки выполнения самостоятельной работы

Контроль самостоятельной работы студентов предусматривает:

- соотнесение содержания контроля с целями обучения;
- объективность контроля;
- валидность контроля (соответствие предъявляемых заданий тому, что предполагается проверить);
- дифференциацию контрольно-измерительных материалов.

Формы контроля самостоятельной работы

1. Просмотр и проверка выполнения самостоятельной работы преподавателем.
2. Самопроверка, взаимопроверка выполненного задания в группе.
3. Обсуждение результатов выполненной работы на занятии.
4. Текущее тестирование.

Критерии оценки результатов самостоятельной работы

Критериями оценок результатов внеаудиторной самостоятельной работы студента являются:

- уровень освоения студентами учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- умения студента активно использовать электронные образовательные ресурсы, находить требующуюся информацию, изучать ее и применять на практике;
- обоснованность и четкость изложения ответа;
- оформление материала в соответствии с требованиями;
- умение ориентироваться в потоке информации, выделять главное;
- умение четко сформулировать проблему, предложив ее решение, критически оценить решение и его последствия;

- умение показать, проанализировать альтернативные возможности, варианты действий;
- умение сформировать свою позицию, оценку и аргументировать ее

Критерии оценки выполнения контрольных заданий для самостоятельной работы

Процент правильных ответов	Оценка
От 95% до 100%	отлично
От 76% до 95%	хорошо
От 61% до 75%	удовлетворительно
Менее 61 %	неудовлетворительно

Самостоятельная работа при подготовке к экзамену включает изучение теоретического материала с использованием лекционных материалов, рекомендуемых источников, материалов по практическим занятиям и лабораторным работам.

Перечень примерных вопросов и заданий для самостоятельной работы

1. Перечислить направления компьютерной лингвистики.
2. Сформулировать общие принципы построения автоматизированных систем обработки текстов.
3. Разъяснить принципы работы графематического и морфологического анализаторов.
4. Перечислить методы задания синтаксической структуры предложений.
5. Разъяснить принципы работы фрагментационного и синтаксического анализаторов. Описать принцип их взаимодействия.
6. Изложить основные идеи подхода И. Мельчука к семантическому анализу.
7. Привести примеры мер семантической близости.
8. Дать определения отмеченных последовательностей, контекста, дистрибутивных классов.
9. Дать формальные определения частей речи, грамматического рода и категории падежа в терминах модели языка, предложенной С. Маркусом.
10. Изложить основные идеи теории речевых действий.
11. Привести классификацию речевых действий.
12. Суть теоретико-модельного подхода к исследованию семантики текстов.
13. Сформулировать принципы построения тезаурусов и онтологий. Сходства и отличия.
14. Дать определения семантических сетей, фреймов.
15. Неточные рассуждения. Что такое логика Заде?
16. Привести примеры искусственных языков и нотаций.
17. Что такое корпусная лингвистика?

18. Применение частотных методов в компьютерной лингвистике. Перечислить, описать, привести примеры.
19. В чем отличие между классификацией и кластеризацией текстов?
20. Перечислить методы классификации и кластеризации текстовой информации. Сформулировать основные принципы.
21. Разъяснить принципы работы автоматических систем извлечения информации.
22. Сформулировать принципы обработки неструктурированной и плохо структурированной информации. Индексация текстов.
23. Перечислить формальные методы атрибуции текстов.
24. Дать определения лингвостатистических параметров, авторского инварианта и лингвистических спектров.
25. Привести примеры использования методов кластеризации и классификации для определения авторства текстов.
26. Перечислить основные направления исследований социальных сетей.
27. Дать определения центральностей разного типа.
28. Описать методы анализа социальных сетей.
29. Перечислить основные методы обнаружения спам-сообщений. Привести примеры.
30. Пояснить принцип работы байесовского классификатора.

Перечень примерных тем рефератов и докладов

1. Проблемы автоматизации синтаксического анализа предложений.
2. Проблемы обнаружения кореференций и анафор в текстах на ЕЯ.
3. Применение алгоритмов и методов обработки текстовой информации в технике.
4. Применение алгоритмов и методов обработки текстовой информации в медицине.
5. Применение алгоритмов и методов обработки текстовой информации в системах безопасности.
6. Возможности программных приложений для анализа социальных сетей.
7. Проблемы автоматической идентификации авторов текстов.
8. Методы автоматического построения онтологий.
9. Сравнение алгоритмов обнаружения и исправления ошибок и опечаток.
10. Сравнение алгоритмов морфологического анализа.

VI. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

Для текущей аттестации при изучении дисциплины «Компьютерная лингвистика и обработка естественных языков» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР):

Практическое задание (ПР-1)

Тест (ПР-2)

Эссе (ПР-3)

Реферат (ПР-4)

Конспект (ПР-7)

Проект (ПР-9)

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Модуль I	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование ПР-1 Практическое задание	Вопросы к экзамену
			умеет		
			владеет		
2	Модуль II	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-1 Практическое задание ПР-3 Эссе	
3	Модуль III	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-4 Реферат ПР-2 Тест ПР-1 Практическое задание	
4	Модуль IV	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-1 Практическое задание	

VII. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(печатные и электронные издания)

- 1. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с.
- 2. Языкознание: От Аристотеля до компьютерной лингвистики: Научно-популярное / Алпатов В. - М.:Альпина нон-фикшн, 2018. - 253 с.: 60x90 1/16. - (Научно-популярная литература) (Переплёт) ISBN 978-5-91671-804-1 - Режим доступа: <http://znanium.com/catalog/product/1003471>
- 3. Просто о больших данных : пер. с англ. / Джудит Гурвиц, Алан Ньюджент, Ферн Халпер [и др.]. - Москва : Сбербанк, : [Эксмо], 2015. - 395 с. - <http://lib.dvfu.ru:8080/lib/item?id=chamo:826169&theme=FEFU>

Дополнительная литература

(печатные и электронные издания)

1. Daniel Jurafsky and James H. Martin. Speech and Language Processing. Pearson Prentice Hall, 2009. Доступ: <https://nlp.stanford.edu/~manning/xyzzzy/JurafskyMartinEd2book.pdf>.
2. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2009. Доступ: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
3. Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце. Введение в информационный поиск. Вильямс, 2011. ISBN 978-5-8459-1623-5, 5 экз.
4. Рябушкина, С. В. Морфология современного русского языка. Флинта, 2009 ISBN 978-5- 9765-0771-5 .— . 3. Копотев М. Введение в корпусную лингвистику. Animedia Company, 2014. ISBN 978- 80-7499-067-0
5. Николаев И.С., Митрина О.В., Ландо Т.М. (Ред.) Прикладная и компьютерная лингвистика, URSS, 2016. ISBN 978-5-9710-3472-8
6. Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс] : учебное пособие / С.В. Пальмов. – Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. – 127 с. – 2227-8397. – Режим доступа: <http://www.iprbookshop.ru/75376.html>

7.Петрунин, Ю. Ю. Информационные технологии анализа данных. Data Analysis : учебное пособие для вузов по управленческим и экономическим специальностям и направлениям / Ю. Ю. Петрунин ; Московский государственный университет, Факультет государственного управления. – 3-е изд. – М. : Университет, 2014 – 291 с. – Каталог НБ ДВФУ: <http://lib.dvfu.ru:8080/lib/item?id=chamo:734307&theme=FEFU>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. «ИТ-образование в Рунете». Образовательные ресурсы Рунета: <http://ifets.ieee.org/russian/depository/resource.htm>
2. «Российский общеобразовательный портал»: <http://www.school.edu.ru/>
3. «Издание литературы в электронном виде»: <http://www.magister.msk.ru/library/library.htm>
4. Annual Review: <http://www.annualreviews.org/ebvc>
5. Scopus - мультидисциплинарная реферативная база данных: <http://www.scopus.com/>
6. Единая коллекция образовательных ресурсов: <http://school-collection.edu.ru/catalog/>
7. Информационные ресурсы Российской Библиотечной Ассоциации (РБА): <http://www.rba.ru/>
8. Каталог электронных ресурсов научной библиотеки ДВФУ: <http://www.dvfu.ru/web/library/elib>
9. Коллекция журналов издательства Elsevier на портале ScienceDirect: <http://www.sciencedirect.com/>.
10. Научная электронная библиотека (НЭБ): <http://www.elibrary.ru/>
11. Портал «Гуманитарное образование» <http://www.humanities.edu.ru/index.html>
12. Российская государственная библиотека (электронный каталог): <http://www.rsl.ru/>
13. Университетская информационная система Россия (УИС Россия): <http://uisrussia.msu.ru>
14. Электронная библиотечная система «Айбукс»: <http://ibooks.ru/>
15. Электронная библиотечная система «Университетская библиотека»: www.biblioclub.ru.
16. Электронная библиотечная система издательства «Лань»: <http://e.lanbook.com/>

Перечень информационных технологий и программного обеспечения

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используется следующее программное обеспечение: Microsoft Office (Access, Excel, PowerPoint, Word и т. д), Open Office, Skype, программное обеспечение электронного ресурса сайта ДВФУ, включая ЭБС ДВФУ.

VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Работа с теоретическими материалами. Изучение дисциплины следует начинать с проработки тематического плана лекций, уделяя особое внимание структуре и содержанию темы и основных понятий. Изучение «сложных» тем следует начинать с составления логической схемы основных понятий, категорий, связей между ними. Целесообразно прибегнуть к классификации материала, в частности при изучении тем, в которых присутствует большое количество незнакомых понятий, категорий, теорий, концепций, либо насыщенных информацией типологического характера. Студенты должны составлять конспекты лекций, систематически готовиться к практическим занятиям, вести глоссарий и быть готовы ответить на контрольные вопросы в ходе лекций и аудиторных занятий. Успешное освоение программы курса предполагает прочтение ряда оригинальных работ и выполнение практических заданий.

Подготовка и выполнение практических заданий. По каждой теме дисциплины предлагаются вопросы и практические задания. Перед выполнением заданий изучите теорию вопроса, предполагаемого к исследованию. Самостоятельная работа студентов заключается:

- в подготовке к практическим занятиям в форме консультаций и дискуссий;
- в выполнении индивидуальных и групповых заданий,
- в подготовке к защите курсовой работы,
- в подготовке к итоговому собеседованию.

Цель практических (семинарских) занятий – научить студентов самостоятельно анализировать учебную и научную литературу и вырабатывать у них опыт самостоятельного мышления по проблемам курса, а также выработать навыки практического применения теоретических знаний. Как правило, семинары проводятся в виде практик-консультаций с элементами дискуссии. При этой форме работы отдельным студентам могут поручаться сообщения по тому или иному вопросу, а также ставя

дополнительные вопросы, как всей аудитории, так и определенным участникам обсуждения.

Самостоятельная работа студентов, предусмотренная учебным планом, соответствует более глубокому усвоению изучаемого курса, формирует навыки исследовательской работы и ориентирует на умение применять теоретические знания на практике.

Материалом для подготовки могут стать конспекты лекций, профессиональная литература, учебно-методическое обеспечение дисциплины.

Методические рекомендации для написания конспектов

Конспекты, написанные от руки, предоставляются преподавателю для оценки (зачёт/незачёт). Учитывая, что в большинстве случаев тексты первоисточников весьма объёмные, для конспектирования можно выбрать только страницы, разделы или главы (30-50 стр. печатного текста). Объём законспектированного текста в тетради определяется самим студентом.

IX. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные занятия по дисциплине проводятся в помещениях, оснащенных соответствующим оборудованием и программным обеспечением.

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

Материально-техническое и программное обеспечение дисциплины

Наименование специальных* помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы
Учебная аудитория для занятий семинарского типа 690922, Приморский край, г. Владивосток, остров Русский, полуостров Саперный, поселок Аякс, 10 Здание ФЭК корпус А, лит О, ауд. 468	Комплект специализированной мебели: доска аудиторная – 1 шт.; парты – 30 шт.; стул -30 шт.; Проектор DLP, 3000 ANSI Lm, WXGA 1280x800, 2000:1 EW330U Mitsubishi,; Системный блок с монитором. Процессор: Intel I5-8600k 3.6Ghz, оперативная память: 32gb, жесткий диск: 1ТБ, графический ускоритель: Nvidia GTX 1080 Беспроводные ЛВС для обучающихся обеспечены системой на базе точек доступа 802.11a/b/g/n 2x2 MIMO(2SS).

Рабочие места для людей с ограниченными возможностями здоровья оснащены дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими и читающими машинами, видеоувеличителем с возможностью регуляции

цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками.

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенные туалетными комнатами, табличками информационно-навигационной поддержки.

X. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценочных средств

Для текущей аттестации при изучении дисциплины «Компьютерная лингвистика и обработка естественных языков» используются следующие оценочные средства:

1) Устный опрос (УО):

Собеседование (консультация с преподавателем) (УО-1)

2) Письменные работы (ПР): практическое задание (ПР-1); тест (ПР-2); эссе (ПР-3); реферат (ПР-4); конспект (ПР-7)

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Модуль I	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование ПР-1 Практическое задание	Вопросы к экзамену
			умеет		
			владеет		
2	Модуль II	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-1 Практическое задание ПР-3 Эссе	
3	Модуль III	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-4 Реферат ПР-2 Тест ПР-1 Практическое задание	
4	Модуль IV	УК-4.1, ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2	знает	УО-1 Собеседование	Вопросы к экзамену
			умеет	УО-1 Собеседование	
			владеет	ПР-1 Практическое задание	

	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде
УО-1	Устный опрос	Средство контроля усвоения учебного материала темы, раздела или разделов дисциплины, организованное как учебное занятие в виде собеседования преподавателя с обучающимися.	Вопросы по темам/разделам дисциплины
ПР-1	Задача (практическое задание)	Средство оценки умения применять полученные теоретические знания в практической ситуации. Задача (задание) должна быть направлена на оценивание тех компетенций, которые подлежат освоению в данной дисциплине, должна содержать четкую инструкцию по выполнению или алгоритм действий.	Комплект задач и заданий
ПР-2	Тест	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.	Фонд тестовых заданий
ПР-3	Эссе	Средство, позволяющее оценить умение обучающегося письменно излагать суть поставленной проблемы, самостоятельно проводить анализ этой проблемы с использованием концепций и аналитического инструментария соответствующей дисциплины, делать выводы, обобщающие авторскую позицию по поставленной проблеме.	Тематика эссе
ПР-4	Реферат	Продукт самостоятельной работы обучающихся, представляющий собой краткое изложение в письменном виде полученных результатов теоретического анализа определенной научной (учебно-исследовательской) темы, где автор раскрывает суть исследуемой проблемы, приводит различные точки зрения, а также собственные взгляды на нее.	Темы рефератов
ПР-7	Конспект	Продукт самостоятельной работы обучающегося, отражающий основные идеи заслушанной лекции, сообщения	Литература для конспектирования

Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины

Текущая аттестация студентов. Текущая аттестация студентов по дисциплине проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация по дисциплине проводится в форме контрольных мероприятий (защита эссе, защита лабораторных работ, тестирование) по

оцениванию фактических результатов обучения студентов осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- учебная дисциплина (активность на занятиях, своевременность выполнения различных видов заданий, посещаемость всех видов занятий по аттестуемой дисциплине);

- степень усвоения теоретических знаний;

- уровень овладения практическими умениями и навыками по всем видам учебной работы;

- результаты самостоятельной работы.

Оценивание результатов освоения дисциплины на этапе текущей аттестации проводится в соответствии с используемыми оценочными средствами и критериями.

Процедура и критерии оценивания эссе/реферата

Оценивание защиты эссе проводится при представлении эссе в электронном виде, по двухбалльной шкале: «зачтено», «незачтено».

Оценка «зачтено» выставляется студенту, если он представляет к защите эссе, удовлетворяющее поставленным к эссе требованиям (использование данных отечественной и зарубежной литературы, источников Интернет, информации нормативно-правового характера и передовой практики, представление краткого терминологического словаря по теме), по оформлению, если студент демонстрирует владение методами и приемами теоретических аспектов работы, не допускает фактических ошибок, связанных с пониманием проблемы.

Оценка «незачтено» выставляется студенту, если он не владеет методами и приемами теоретических аспектов работы, допускает существенные ошибки в работе, связанные с пониманием проблемы, представляет эссе с существенными отклонениями от правил оформления письменных работ.

Процедура и критерии оценивания отчетов по лабораторным работам

Оценивание защиты лабораторной работы проводится при представлении отчета в электронном виде, по двухбалльной шкале: «зачтено», «незачтено».

Оценка «зачтено» выставляется студенту, если он представляет к защите отчет по лабораторной работе, удовлетворяющий требованиям по поставленным заданиям, по оформлению, демонстрирует владение методами и приемами теоретических и/или практических аспектов работы.

Оценка «незачтено» выставляется студенту, если он не владеет методами и приемами теоретических и/или практических аспектов работы, допускает существенные ошибки в работе, представляет отчет с существенными отклонениями от правил оформления письменных работ.

Промежуточная аттестация студентов. Промежуточная аттестация студентов по дисциплине «Компьютерная лингвистика и обработка естественных языков» проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Промежуточная аттестация по дисциплине проводится в виде экзамена, форма экзамена - «устный опрос в форме ответов на вопросы», «практические задания по типам».

Порядок проведения экзамена, форма экзаменационного билета определены локальным нормативным актом ДВФУ «Положение о текущем контроле успеваемости, текущей и промежуточной аттестации студентов, обучающихся по программам высшего образования (бакалавриата, специалитета и магистратуры) в ДВФУ».

В экзаменационный билет входят два вопроса и одно практическое задание.

Критерии выставления оценки студенту на экзамене по дисциплине

Баллы (рейтинговой оценкой)	Оценка экзамена (стандартная)	Требования к сформированным компетенциям
86 -100	«отлично»	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал монографической литературы, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76 - 85	«хорошо»	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.

61 -75	«удовлетворительно»	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.
0 -60	«неудовлетворительно»	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного «не материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

Оценочные средства для промежуточной аттестации

Список вопросов к экзамену

1. Что такое уровни представления языковых выражений? Какие бывают уровни?
2. Морфологический анализ и синтез текстов. Поверхностная и глубинная морфология. Анализ композитов.
3. Основные типы представления синтаксической структуры предложения. Зависимости и составляющие. Дерево зависимостей.
4. Понятие синтаксического правила (синтагмы).
5. Грамматика и словарь.
6. Синтаксические признаки слова.
7. Валентностная структура предиката. Синтаксические и семантические валентности. Модель управления слова.
8. Основные типы компьютерных синтаксических ресурсов. Словари и корпусы текстов.
9. Глубокий анализ лингвистических данных: постановка задачи, основные методы и подходы.
10. Снятие неоднозначности морфологического разбора на основе скрытых марковских моделей.
11. Методы обработки неизвестных слов в морфологическом анализе.

12. Использование сглаживания Гуда-Тьюринга в языковом моделировании.
13. Основные метрики оценки качества информационного поиска.
14. Источники данных для задачи автоматического реферирования.
15. Методы повышения производительности формирования веб-снипкетов.
16. Подходы к повышению качества статистических парсеров на основе составляющих.
17. Оценки вычислительной сложности вероятностного синтаксического анализа на основе составляющих и деревьев зависимостей.
18. Формирование словарей тонально окрашенных слов и фраз с помощью запросов к машинам поиска интернета.
19. Группы признаков для выделения именованных сущностей из текста.
20. Подходы к извлечению отношений между именованными сущностями из больших корпусов текстов.
21. Методы оценки вопросно-ответного поиска.
22. Сингулярное разложение матриц в задаче векторного представления семантики слов.
23. Недостатки метрики BLEU.