




МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

«СОГЛАСОВАНО»


Руководитель ОП



(подпись) Спицына Н.А.
(Ф.И.О. рук.ОП)
«11» июля 2019г.

«УТВЕРЖДАЮ»

Заведующая кафедрой прикладной математики, механики,
управления и программного обеспечения



(подпись) Артемьева И.Л.
(ФИО зав. кафедрой)
«11» июля 2019г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Автоматическая обработка естественного языка

Направление подготовки 45.03.03 «Фундаментальная и прикладная лингвистика»

Форма подготовки очная

курс 4 семестр 7,8
лекции 30час.
практические занятия ___ час.
лабораторные работы 30 час.
в том числе с использованием МАО лек ___ - / пр. ___ / лаб. 18__ час
всего часов аудиторной нагрузки – 60 час.
в том числе с использованием МАО _18_-час
самостоятельная работа 84 час.
контрольные работы (количество)
курсовая работа / курсовой проект нет
зачет 7,8 семестр
экзамен _нет__ семестр

Рабочая программа составлена в соответствии с требованиями образовательного стандарта, самостоятельно установленного ДВФУ, утвержденного приказом ректора ДВФУ №12-13-235 от 18.02.2016 г.

Рабочая программа обсуждена на заседании кафедры прикладной математики, механики, управления и программного обеспечения, протокол № 11 от «9» июля 2019 г.

Заведующая кафедрой прикладной математики, механики, управления и программного обеспечения Артемьева И.Л., д.т.н., профессор

Составитель: доцент кафедры прикладной математики, механики, управления и программного обеспечения Прудникова Л.И., к.ф.-м.н., доцент

АННОТАЦИЯ

Рабочая программа дисциплины «Автоматическая обработка естественного языка» разработана для студентов 4 курса, обучающихся по направлению 45.03.03 «Фундаментальная и прикладная лингвистика», все профили.

Трудоемкость дисциплины 4 зачетных единиц (144 часов). Дисциплина реализуется в 7 и 8 семестрах. В 7 семестре дисциплина содержит 18 часов лекций, 0 часов практических занятий, 18 часов лабораторных работ (из них 18 часов лабораторных работ в интерактивной форме), 36 часов самостоятельной работы. Во 8 семестре дисциплина содержит 12 часов лекций, 0 часов практических занятий, 12 часов лабораторных работ, 48 часов самостоятельной работы.

Дисциплина «Автоматическая обработка естественного языка» базируется на дисциплинах «Математические основы лингвистики», «Основы технологии программирования», «Современные информационные технологии», «Формальные модели в лингвистике». Знания, полученные при ее изучении, будут использованы при подготовке выпускной квалификационной работы.

Цель дисциплины – сформировать у бакалавров системное представление о методах автоматической обработки естественно-языкового текста на всех уровнях лингвистического анализа; ознакомление студентов с практическими методами автоматической обработки естественно-языкового текстов.

Задачи дисциплины:

1. Освоить общие принципы построения систем обработки речи (устной и письменной).
2. Дать представление о возможностях автоматического порождения речи.
3. Ознакомить с лингвистическим обеспечением различных прикладных систем (автоматического индексирования, реферирования, и т.д.).
4. Раскрыть структуру лингвистических процессоров.

Для успешного изучения дисциплины «Автоматическая обработка естественного языка» у обучающихся должны быть сформированы следующие предварительные компетенции: способность осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных; владение основными концепциями, принципами, теориями и фактами, связанными с информатикой; готовность к использованию методов разработки лингвистических систем.

В результате изучения данной дисциплины у обучающихся формируются следующие общепрофессиональные и профессиональные компетенции (элементы компетенций). ОК-5 способностью использовать современные методы и технологии (в том числе информационные) в профессиональной деятельности ОК-14 способностью к самоорганизации и самообразованию ОПК-2 владением основами математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур ОПК-7 способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-1 владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	Знает	как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;
	Умеет	проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет	методами поиска информации по работе с естественным языком;
ПК-10 способность пользоваться лингвистически ориентированными программными продуктами	Знает	особенности реализации естественно-языковых систем, назначение лингвистического процессора;
	Умеет	эксплуатировать и сопровождать информационные системы и сервисы
	Владеет	методикой использования компьютерных технологий в предметной области;
ПК-12 способностью использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке,	Знает	лингвистические технологии в сфере АОТ; виды программной продукции АОТ;
	Умеет	осуществлять и обосновывать выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;

лингвистических компонентов интеллектуальных и информационных электронных систем	Владеет	навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);
ПК-13 способность проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем	Знает	ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;
	Умеет	применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;
	Владеет	навыками работы со специальными программными средствами АОТ;

Для формирования вышеуказанных компетенций в рамках дисциплины «Автоматическая обработка естественного языка» применяются следующие методы активного/ интерактивного обучения: метод новых вариантов.

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Раздел.1.Язык как объект лингвистики(18ч.).

Тема 1. Введение.(2ч.)

Язык как знаковая система. Язык, речь, речевая деятельность. Компьютерная лингвистика.

Тема 2.Корпусная лингвистика (4 ч.)

Корпусные методы и применение корпусов в лингвистических исследованиях. Национальный корпус русского языка. Корпус и Web: сходства и различия. Корпус как репрезентативная выборка текстов. Webкак несбалансированный набор текстов. Возможности Web'а в лингвистических исследованиях.

Тема 3. Автоматическая обработка письменного текста.(6 ч.)

Автоматическое индексирование. Создание словников, конкордансов, частотных словарей. Законы Зипфа. Автоматическое реферирование. Автоматическое порождение текста. Цепи Маркова SIMP-таблицы. Метод фокуса внимания. Метод с использованием словарей. Программы автоматического порождения текстов.

Тема 4. Автоматическая обработка звучащей речи.(6 ч.)

Речевые технологии. Системы распознавания речи. Акустическая модель. Лингвистическая модель. Механизмы распознавания речи. Голосовое управление. Автоматический синтез речи. Основные направления и методы современных разработок (экспертный и вероятностный подходы).

Раздел.2. Компьютерная лингвистика(12 часов)

Тема 1. Язык и компьютер.(2ч.)

Задачи компьютерной лингвистики. Особенности системы ЕЯ: уровни и связи. Моделирование в компьютерной лингвистике. Лингвистические ресурсы. Приложения компьютерной лингвистики.

Тема 2. Инструментальные системы разработки приложений по автоматической обработке текстов на естественном языке.(4ч.)

Программные средства лингвистической обработки. Представление лингвистических данных. Подходы к представлению данных. Лингвистическая разметка. Лингвистические аннотации. Представления, основанные на абстракции. Архитектура инструментальных ЕЯ-систем. Процессы обработки текста.

Тема 3. Системы обработки ЕЯ-текстов.(3ч.)

Системы на базе разметки. Системы на базе аннотаций. Системы интеграции поверхностной и глубокой обработки. Системы, развивающие отдельные аспекты обработки текста. Прочие системы.

Тема 4.Тезаурус.(3ч.)

Структура информационно-поискового тезауруса. Построение информационно-поисковых тезаурусов. Математические модели в поиске. Основные аспекты и понятия поиска информации. Сфера применения и перспективы развития информационно-поисковых тезаурусов. Применение тезаурусного метода поиска в Интернет.

II. СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Лабораторные работы (30 часов)

Лабораторная работа №1.(4час.).

Системы обработки естественного языка (NL-Processing).(3ч.)

Лабораторная работа №2. (3час.).

Информационно-поисковые языки (на ограниченном естественном языке).

Лабораторная работа №3. (3 час.).

Поисковый образ запроса и поисковое предписание.

Лабораторная работа №4. (4 час.).

Универсальные среды для систем автоматической обработки русского текста

Лабораторная работа №5.(3 час.).

Смысловый и нечеткий поиск документов на русском и/или английском языках.

Лабораторная работа №6.(4 час.).

Анализ текстов для построения систем автоматического перевода с русского на английский язык (и наоборот).

Лабораторная работа №7. (3 час.).

Программы для синтаксического и морфологического анализа русскоязычных текстов.

Лабораторная работа №8. (4 час.).

Словари и тезаурусы.

Лабораторная работа №9. (2 час.).

Психолингвистические программы.

III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Автоматическая обработка естественного языка» представлено в Приложении 1 и включает в себя:

- план-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию;
- характеристика заданий для самостоятельной работы обучающихся и методические рекомендации по их выполнению;
- требования к представлению и оформлению результатов самостоятельной работы;
- критерии оценки выполнения самостоятельной работы.

IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел.1.Язык как объект лингвистики	ПК-10; ПК-11; ПК-12; ПК-13	Знает	Конспект(ПР-7)	Вопросы к экзамену 1-7
2.	Раздел.2. Компьютерная лингвистика	ПК-10; ПК-11; ПК-12; ПК-13	Знает	Конспект(ПР-7)	Вопросы к экзамену 8-14
3.	Лабораторные работы №1-9	ПК-10; ПК-11; ПК-12; ПК-13	умеет, владеет	Лабораторная работа (ПР-6)	Отчет

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта

деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 2.

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

1. Введение в прикладную лингвистику: учебное пособие / А. Н. Баранов; Московский государственный университет, Филологический факультет. А. Н. Баранов. - М.: URSS [Либроком], 2013. - 367 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:779256&theme=FEFU>
2. Корпусная лингвистика / Е.В.Грудева. —М.: ФЛИНТА, 2012. - 165 с.
3. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие / Н.Н. Леонтьева. – М.: Академия, 2006. - 304с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:235710&theme=FEFU>
4. Тезаурусы в задачах информационного поиска. / Н.В. Лукашевич. - М.: Изд-во Московского университета, 2011. - 512с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:403174&theme=FEFU>
5. Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс]: учеб.пособие / Л. Ю. Щипицина. — М.: ФЛИНТА, 2013. — 128 с. - ISBN 978-5-9765-1431-7
<http://znanium.com/catalog.php?bookinfo=462989>
6. Информатика и лингвистика: Учебное пособие/ВолосатоваТ.М., ЧичваринН.В. - М.: НИЦ ИНФРА-М, 2016. - 196 с.: 60x90 1/16. - (Высшее образование:Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-
<http://znanium.com/catalog.php?bookinfo=508097>
7. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов [Электронный ресурс] / Рубашкин В.Ш. - М.: ФИЗМАТЛИТ, 2012. -
<http://www.studentlibrary.ru/book/ISBN9785922114394.html>
8. Компьютерная лингвистика. Учебное пособие / Ю.Н. Марчук.- М.:Изд.АСТ, 2007. - 319 с.

Дополнительная литература

1. Интернет-коммуникация как новая речевая формация / [С. И. Агаюлова, Е. Н. Галичкина, Е. И. Горошко и др.; науч. ред. Т.Н.Колокольцева, О.В. Луговина] – М.: Флинта: Наука, 2012.- 323 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:675726&theme=FEFU>
2. Информационные технологии в лингвистике / А.В. Зубов, И.И. Зубова. – М.: Академия, 2004. - 207с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:245160&theme=FEFU>
3. Современные информационные технологии для гуманитария: практическое руководство / А.Т. Хроленко, А.В. Денисов. - М.: Флинта: Наука, 2010. - 128 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:308500&theme=FEFU>
4. Интернет-коммуникация как новая речевая формация / [Агаюлова С. И., Галичкина Е. Н., Горошко Е. И. и др.; научн. ред.: Т.Н.Колокольцева, О.В. Луговина] М.: Флинта Наука, 2012. - 323 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:675726&theme=FEFU>
5. Компьютерная лингвистика. Учебное пособие / Марчук Ю.Н. - М.:Изд.АСТ, 2007. - 317 с.
6. Корпусная лингвистика / Е.В. Грудева. -М.:ФЛИНТА, 2012. - 165 с.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://window.edu.ru/resource/387/80387> Введение в компьютерную лингвистику. Учебное пособие. / К.К Боярский. - СПб: НИУ ИТМО, 2013. - 72 с.
2. <http://window.edu.ru/library/pdf2txt/465/78465/59324> Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е.И. Большакова, Э.С.Клышинский, Д.В. Ландэ, А.А.Носков, О.В. Пескова, Е.В.- Ягунова М.: МИЭМ, 2011. -272с.
3. <http://prog-school.ru/> Школа программирования
4. <http://www.hardforum.ru/f22/> Форум программистов
5. <http://bib.com.ua/> ВІВлиотекапрограммиста
6. Научный форум dxdy>> Моделирование русского языка
<http://dxdy.ru/topic4210-60.html>
7. Лингвофорум>>Компьютерная лингвистика
<http://lingvoforum.net/index.php?topic=29965.0>
8. Форум Диалог <http://forum.dialog-21.ru/actualthread.aspx?tid=363>

9. Научно-образовательный портал "Лингвистика в России: ресурсы для исследователей"http://uisrussia.msu.ru/linguist/B_comput_ling.jsp
10. Форум myTTS <http://mytts.forum2x2.ru/f45-forum>
11. Книги по лингвистике <http://www.antic-r.ru/bib15.htm>

Перечень информационных технологий и программного обеспечения

1. Операционная система Windows.
2. MicrosoftOffice.
3. Интернет.
4. Лингвистические программы: ClearTextReader'sEdition, Словогрыз, xReplacer; RussianMorphologicalDictionary, Рабочее Место Лингвиста, Galaktika-ZOOM; SakramentText-to-SpeechEnginev2.0, Аки Росс, Письмовник; NeurOKSemanticSuite, ARMEngine 4.0, Следопыт 3.0; Гиперсловарь Ариадна, Словарь сокращений русского языка, Обратные и частотные словари русского языка; Гиперсловарь Ариадна, Словарь сокращений русского языка, Обратные и частотные словари русского языка.

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используются следующие информационно-справочные системы:

1. Научная электронная библиотека eLIBRARY.
2. Электронно-библиотечная система издательства «Лань».
3. Электронная библиотека "Консультант студента".
4. Электронно-библиотечная система IPRbooks.
5. Информационная система "ЕДИНОЕ ОКНО доступа к образовательным ресурсам".
6. Доступ к электронному заказу книг в библиотеке ДВФУ, доступ к нормативным документам ДВФУ, расписанию, рассылке писем.

Лекции проводятся с использованием проектора и мультимедийного комплекса для проведения лекций внутренней системы портала ДВФУ. Лабораторные занятия проводятся в специализированном компьютерном классе. Для составления документации используется текстовый процессор (LibreOffice или MicrosoftWord).

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Дисциплина изучается в следующих организационных формах: лекции, лабораторное занятие; самостоятельное изучение теоретического материала; самостоятельное выполнение индивидуального задания; индивидуальные консультации.

Работа на лекции

Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое «конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось просить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Вопросы для самоконтроля по разделу 1

1. Автоматическая обработка звучащей речи.
2. Системы распознавания речи.
3. Механизмы распознавания речи.

4. Автоматический синтез речи.
5. Автоматическое индексирование.
6. Создание словарей, конкордансов, частотных словарей.
7. Законы Зипфа.

Вопросы для самоконтроля по разделу 2

8. Автоматическое реферирование.
9. Автоматическое порождение текста.
10. Структура информационно-поискового тезауруса.
11. Построение информационно-поисковых тезаурусов.
12. «Мягкое» понимание текста. Семантической структуры.
13. Получение компрессированного текста.
14. Основные характеристики текста.

Лабораторные работы

При выполнении лабораторной работы необходимо следовать методическим рекомендациям по ее выполнению. Результатом лабораторной работы является созданный документ, который демонстрируется преподавателю в конце работы. Студент должен уметь отвечать на вопросы преподавателя, поясняя процесс создания документа и выполнения работы.

Самостоятельная работа студента

Основной формой самостоятельной работы студента является изучение теоретического материала, его дополнение рекомендованной литературой, выполнение индивидуальных заданий, а также активная работа на практических и лабораторных занятиях. Целью выполнения индивидуальных заданий является закрепление практических навыков, полученных в процессе выполнения лабораторной работы.

Контроль за выполнением самостоятельной работы студента производится в виде контроля каждого этапа работы (см. приложение 1). Студент должен планировать график самостоятельной работы по дисциплине и придерживаться его.

VIII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Лекции проводятся с использованием проектора и внутренней системы портала ДВФУ. Лабораторные занятия проходят в аудиториях, оборудованных компьютерами типа Lenovo C360G-i34164G500UDK с

лицензионными программами MicrosoftOffice 2013 и аудио-визуальными средствами проектор Panasonic DLPProjectorPT-D2110XE, плазма LG FLATRON M4716ССВАМ4716СJ. Для выполнения самостоятельной работы студенты в жилых корпусах ДВФУ обеспечены Wi-Fi.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

**УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ
ОБУЧАЮЩИХСЯ
по дисциплине «Автоматическая обработка естественного
языка»**

**Направление подготовки 45.03.03 «Фундаментальная и прикладная лингвистика»
Форма подготовки очная**

**Владивосток
2019**

План-график выполнения самостоятельной работы по дисциплине

№ п/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
1.	1-я неделя	Самоорганизация подгрупп, выбор индивидуальных тем	2ч.	Утверждение в ходе практического занятия
2.	2-8 недели	Работа с первоисточниками	10ч.	Опрос (анализ и обсуждение проработанных первоисточников в ходе практического занятия)
3.	9-16 недели	Работы с лингвистическими программами по ИТ	16ч.	Опрос в ходе практического занятия
4.	17-18 недели	Подготовка презентации и демонстрации работы с программами	8ч.	Демонстрация в ходе практического занятия
	19-26 недели	Работы с лингвистическими программами по ИТ	40ч.	Опрос в ходе практического занятия
	27-30 недели	Подготовка презентации и демонстрации работы с программами	8ч.	Демонстрация в ходе практического занятия

Рекомендации по самостоятельной работе студентов

Самостоятельная работа студентов состоит в работе с литературой, подготовке к лабораторным работам и выполнении индивидуальных заданий по темам.

Работа с литературой

В процессе подготовки к лабораторным работам студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами Интернета 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.

к лабораторным работам

Подготовку к каждой лабораторной работе каждый студент должен начать с изучения теоретического материала и ознакомления с имеющимися реализациями.

Индивидуальные задания

По теме лабораторной работы студентам выдаются индивидуальные задания, предназначенные для закрепления умений и навыков, полученных при выполнении лабораторной работы. Задания выполняются самостоятельно. По каждому заданию предполагается подготовка индивидуального документа (проекта). Выполненный проект демонстрируется преподавателю в начале следующей лабораторной работы. Структура индивидуального задания определяется темой лабораторной работы.

Критерии оценки индивидуальных заданий (проектов)

– 100-86 баллов выставляется, если содержание и составляющие части соответствуют выданному заданию. Продемонстрировано владение навыками подготовки документа по теме. Фактических ошибок, связанных с пониманием структуры и содержания задания нет.

– 85-76 - баллов выставляется, если при выполнении задания допущено не более одной ошибки. Продемонстрировано владение навыками подготовки документа по теме. Фактических ошибок, связанных с пониманием структуры и содержания задания нет.

– 75-61 балл выставляется, если при выполнении задания допущено не более двух ошибок. Продемонстрировано знание и владение навыками подготовки документа по теме. Допущено не более 2 ошибок, связанных с пониманием структуры и содержания задания.

– 60-50 баллов - если структура и содержание задания не соответствуют требуемым.

Шкала оценивания

Менее 60 баллов	незачтено	неудовлетворительно
От 61 до 75 баллов	зачтено	удовлетворительно
От 76 до 85 баллов	зачтено	хорошо
От 86 до 100 баллов	зачтено	отлично



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

**по дисциплине «Автоматическая обработка естественного языка»
Направление подготовки 45.03.03 «Фундаментальная и прикладная лингвистика»**

Форма подготовки очная

**Владивосток
2019**

Паспорт ФОС

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-11 владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	Знает	как устроена современная электронная языковая среда, основные уровни обработки текста наЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;
	Умеет	проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет	методами поиска информации по работе с естественным языком;
ПК-10 способность пользоваться лингвистически ориентированными программными продуктами	Знает	особенности реализации естественно-языковых систем, назначение лингвистического процессора;
	Умеет	эксплуатировать и сопровождать информационные системы и сервисы
	Владеет	методикой использования компьютерных технологий в предметной области;
ПК-12 способностью использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем	Знает	лингвистические технологии в сфере АОТ; виды программной продукции АОТ;
	Умеет	осуществлять и обосновывать выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет	навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);
ПК-13 способность проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических	Знает	ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;
	Умеет	применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;

компонентов интеллектуальных и информационных электронных систем	Владеет	навыками работы со специальными программными средствами АОТ;
--	---------	--

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел.1.Язык как объект лингвистики	ПК-10 ПК-11 ПК-12 ПК-13	Знает	Конспект(ПР-7)	Вопросы к экзамену 1-7
2.	Раздел.2. Компьютерная лингвистика	ПК-10 ПК-11 ПК-12 ПК-13	Знает	Конспект(ПР-7)	Вопросы к экзамену 8-14
3.	Лабораторные работы №1-9	ПК-10 ПК-11 ПК-12 ПК-13	умеет, владеет	Лабораторная работа (ПР-6)	Отчет

Шкала оценивания уровня сформированности компетенций

Код и формулировка компетенции	Этапы формирования компетенции		критерии	показатели
ПК-13 способность проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем	знает (пороговый уровень)	ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;	Знание основных законов развития компьютерной лингвистики и программных технологий	Способность дать ответы на вопросы
	умеет (продвинутый)	применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;	Умение выбрать и применить программное обеспечение, соответствующее требованиям задачи	Умение продемонстрировать способность использования полученных знаний при выполнении практических заданий

	владеет (высокий)	навыками работы со специальными программными средствами АОТ;	Владение методами использования средств при выполнении практических работ	Наличие выполненных заданий
ПК-11 владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимедальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	знает (пороговый уровень)	как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;	Знание устройства электронной языковой среды и задач, решаемых при обработке текстов на ЕЯ	Способность дать ответы на вопросы
	умеет (продвинутый)	проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	Умение использовать программные продукты АОТ в своей профессиональной деятельности;	Умение продемонстрировать способность использования полученных знаний при выполнении практических заданий
	владеет (высокий)	методами поиска информации по работе с естественным языком;	Владение набором операций, предоставляемых программными продуктами	Наличие выполненных заданий

ПК-10 способность пользоваться лингвистиче ски ориентирова нными программны ми продуктами	знает (порогов ый уровень)	особенности реализации естественно- языковых систем, назначение лингвистическ ого процессора;	Знание лингвистическ их компонентов интеллектуальн ых и информационн ых электронных систем	Способность дать ответы на вопросы
	умеет (продвин утый)	эксплуатирова ть и сопровождать информационн ые системы и сервисы	Умение осуществлять и обосновывать выбор проектных решений по видам обеспечения информационн ых систем	Умение продемонстрировать способность использования полученных знаний при выполнении практических заданий
	владеет (высокий)	методикой использования компьютерных технологий в предметной области;	Владение методикой работы в сфере АОТ	Наличие выполненных заданий
ПК-12 способность использоват ь лингвистиче ские технологии для проектирова ния систем автоматичес кой обработки звучащей речи и письменного текста на естественно м языке, лингвистиче ских компонентов интеллектуа льных и информацио	знает (порогов ый уровень)	лингвистическ ие технологии в сфере АОТ; виды программной продукции АОТ;	Знание основных принципов использования лингвистическ их технологий	Способность дать ответы на вопросы
	умеет (продвин утый)	осуществлять и обосновывать выбор проектных решений по видам обеспечения информационн ых систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействи ют лингвистическ	Умение исследовать современные системы;	Умение продемонстрировать способность использования полученных знаний при выполнении практических заданий

нных электронных систем		ие и иные данные, откуда их можно получить;		
	владеет (высокий)	навыками вести исследовательс кую лингвистическ ую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);	Владение навыками освоения современных технических средств и информационн ых технологий, служащие для обеспечения лингвистическ ой деятельности	Наличие выполненных заданий

Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины

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

Текущая аттестация проводится в форме вопросов к экзамену для проверки теоретических знаний, а также в форме защиты проекта, выполняемого в рамках самостоятельной работы параллельно с лабораторными работами и осуществляется ведущим преподавателем.

Объектами оценивания выступают:

- степень усвоения теоретических знаний - оценивается в форме вопросов к зачету;
- уровень овладения практическими умениями и навыками – оценивается в форме защиты индивидуального задания (проекта).

Промежуточная аттестация студентов. Промежуточная аттестация студентов проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

По дисциплине предусмотрен экзамен, который проводится в устной форме.

Критерии выставления оценки студенту на экзамене

Баллы (рейтинговой оценки)	Оценка зачета (стандартная)	Требования к сформированным компетенциям
86-100	«зачтено»/ «отлично»	Оценка «зачтено» / «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, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76-85	«зачтено»/ «хорошо»	Оценка «зачтено» / «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	«зачтено»/ «удовлетворительно»	Оценка «зачтено» / «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	«не зачтено»/ «неудовлетворительно»	Оценка «не зачтено» / «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.
------	--	--

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ВОПРОСЫ К ЗАЧЕТУ

Раздел.1.Язык как объект лингвистики.

1. Автоматическая обработка звучащей речи.
2. Системы распознавания речи.
3. Механизмы распознавания речи.
4. Автоматический синтез речи.
5. Автоматическое индексирование .
6. Создание словарей, конкордансов, частотных словарей.
7. Законы Зипфа.

Раздел.2.Компьютерная лингвистика

- 8.Автоматическое реферирование.
9. Автоматическое порождение текста.
10. Структура информационно-поискового тезауруса.
11. Построение информационно-поисковых тезаурусов.
12. «Мягкое» понимание текста. Семантической структуры.
13. Получение компрессированного текста.
14. Основные характеристики текста.

ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ АТТЕСТАЦИИ

Критерии оценки:

86-100 баллов - выставляется студенту, если вопрос полностью раскрыт без ошибок и даны ответы на все дополнительные вопросы

76-85баллов - выставляется студенту, если вопрос полностью раскрыт без ошибок и даны ответы не на все дополнительные вопросы

61-75 баллов - выставляется студенту, если вопрос не полностью раскрыт без ошибок и даны ответы на все дополнительные вопросы

50-60 баллов - выставляется студенту, если вопрос не полностью раскрыт без ошибок и даны ответы не на все дополнительные вопросы