



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

**ШКОЛА ЕСТЕСТВЕННЫХ НАУК**

«СОГЛАСОВАНО»

Руководитель образовательной программы

  
28.08

И.Л. Артемьева

2015 г.



**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ (РПУД)  
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА**

**Направление подготовки – 02.03.03 «Математическое обеспечение и  
администрирование информационных систем»**

**профиль «Технология программирования»**

**Форма подготовки (очная)**

курс 4 семестр 7, 8  
лекции 30 (час.)  
практические занятия 0 час.  
лабораторные работы 30 час.  
в том числе с использованием МАО лек. 0 /пр. 0 /лаб. 30 час.  
в том числе в электронной форме лек.      /пр.      /лаб. 18 час.  
всего часов аудиторной нагрузки 60 (час.)  
в том числе с использованием МАО 30 час.  
самостоятельная работа 84 (час.)  
в том числе на подготовку к экзамену      час.  
курсовая работа / курсовой проект      семестр  
зачет 7,8 семестр  
экзамен нет

Рабочая программа составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего образования, утвержденного приказом Министерства образования и науки РФ от 15 марта 2015 г. № 222

Рабочая программа обсуждена на заседании кафедры прикладной математики, механики, управления и программного обеспечения, протокол № 7 от «4» июля 2015 г.

Зав. кафедрой, д.т.н., профессор И.Л.Артемьева

Составитель: к.ф.-м.н., доцент Л.И. Прудникова

**Оборотная сторона титульного листа РПУД**

**I. Рабочая программа пересмотрена на заседании кафедры:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**II. Рабочая программа пересмотрена на заседании кафедры:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_  
(подпись) (И.О. Фамилия)

## ABSTRACT

**Bachelor's/Specialist's/Master's degree in 02.03.03 – Software and Administration of Information Systems**

**Study profile/ Specialization/ Master's Program “Title” Programming technology**

**Course title:** Computational linguistics

**Variable part of Block 1, 4 credits**

**Instructor:** Prudnikova L.

**At the beginning of the course a student should be able to:** apply main methods and tools of software development; speak about content, main stages and development trends of programming, mathematical software and information technologies; use main models of information technologies and the methods of its application to solve problems in subject spheres.

**Learning outcomes:** readiness to analyse the problems and the development trends of programming technologies; ability to determine the problems and the development trends of the software market; readiness to use main models of information technologies and the methods of its application to solve problems in subject spheres; readiness to use modern system program means: operating systems, operating and network environments, service programs.

**Course description:** solving of the problems of automatic processing of data presented on natural language; application of language computer model in linguistics and related disciplines.

**Main course literature:**

1. Baranov A.N. Vvedenie v prikladnuyu lingvistiku [Introduction to applied linguistics]. Moscow, URSS [Librokom], 2013. 367 p.

<https://lib.dvfu.ru:8443/lib/item?id=chamo:779256&theme=FEFU>

2. Boyarskiy K.K. Vvedenie v prikladnuyu lingvistiku [Introduction to applied linguistics]. Saint Petersburg, ITMO University, 2013. 72 p.

<http://window.edu.ru/resource/387/80387>

3. Lukashevich N.V. Tezaurusy v zadachakh informatsionnogo poiska [Thesauruses in information search problems]. Moscow, MSU Publ., 2011. 512 p.

<https://lib.dvfu.ru:8443/lib/item?id=chamo:403174&theme=FEFU>

**Form of final control:** pass-fail exam.

## АННОТАЦИЯ

Рабочая программа дисциплины «Компьютерная лингвистика» разработана для студентов 4 курса бакалавриата, обучающихся по направлению 02.03.03 «Математическое обеспечение и администрирование информационных систем», профиль «Технология программирования». Дисциплина является обязательной дисциплиной вариативной части учебного плана: Б1.В.ОД.6.

Трудоемкость дисциплины 4 зачетных единиц (144 часа). 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. На самостоятельную работу студентов отводится 72 часа. В 8 семестре дисциплина содержит 12 часов лекций, 0 часов практических занятий, 12 часов лабораторных работ, из них 12 часов лабораторных работ в интерактивной форме. На самостоятельную работу студентов отводится 12 часов.

Дисциплина «Компьютерная лингвистика» базируется на дисциплинах, в которых изучается теория формальных языков, русский язык и культура речи, методы проектирования и разработки компьютерных программ. Знания, полученные при ее изучении, будут использованы в дисциплинах, связанных с созданием приложений различных типов, в которых требуется организация работы с текстами или фрагментами речи на естественном языке.

**Цель** курса «Компьютерная лингвистика» - сформировать у бакалавров систему знаний, связанных с решением задач автоматической обработки информации, представленной на естественном языке, а также со всей сферой применения компьютерных моделей языка в лингвистике и смежных дисциплинах.

### **Задачи дисциплины:**

1. Обучение студентов методам формального представления естественных языков.
2. Освоение современных теорий построения систем, поддерживающих естественно-языковые интерфейсы.
3. Обучение студентов алгоритмам и методам, применяемых в естественно-языковых системах.

Для успешного изучения дисциплины «Компьютерная лингвистика» у обучающихся должны быть сформированы следующие предварительные компетенции: готовность применять основные методы и инструменты разработки программного обеспечения; владение знаниями о содержании, основных этапах и тенденциях развития программирования, математического

обеспечения и информационных технологий. Иметь навыки использования основных моделей информационных технологий и способов их применения для решения задач в предметных областях.

В результате изучения данной дисциплины у обучающихся формируются следующие общекультурные/ общепрофессиональные/ профессиональные компетенции (элементы компетенций).

Код и формулировка компетенции	Этапы формирования компетенции	
ОПК-3 готовность анализировать проблемы и направления развития технологий программирования	Знает	технологии автоматизированной обработки текстовой информации, основные принципы представления знаний о предметной области в виде рубрикаторов, тезаурусов, онтологий особенности реализации естественно-языковых систем, назначение лингвистического процессора
	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; определять основные классы ЕЯ-систем;
	Владеет	широким диапазоном различных информационно-коммуникационных технологий; методами поиска информации по работе с естественным языком
ОПК-6 способность определять проблемы и тенденции развития рынка программного обеспечения	Знает	сферу применения и перспективы развития информационно-поисковых тезаурусов, применение тезаурусного метода поиска в Интернет
	Умеет	интерпретировать результаты автоматической обработки лингвистических данных
	Владеет	методикой использования компьютерных технологий в предметной области
ПК-2 готовность к использованию основных моделей информационных технологий и способов их применения для решения задач в предметных областях	Знает	существенные отличия естественных языков от искусственных и особенности компьютерных моделей естественного языка, какие лингвистические технологии следует применять в зависимости от условий задачи, понимать, на какие лингвистические данные и ресурсы может опираться технология
	Умеет	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте
	Владеет	приемами работы с прикладным программным обеспечением
ПК-5 готовность к использованию современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ	Знает	типичные программные системы (приложения) в области АОТ и их архитектурные особенности
	Умеет	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте

	Владеет	навыками работы со специальными программными средствами автоматизированной обработки текстов
--	---------	--

Для формирования вышеуказанных компетенций в рамках дисциплины «Компьютерная лингвистика» применяются следующие методы активного/интерактивного обучения: метод новых вариантов, метод круглого стола.

## **I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА**

### **Раздел.1 Компьютерная лингвистика. (8ч.)**

#### **Тема 1. Моделирование в компьютерной лингвистике. (4ч.)**

Задачи компьютерной лингвистики. Особенности системы ЕЯ: уровни и связи. Моделирование в компьютерной лингвистике. Автоматическая обработка звучащей речи. Автоматическая обработка письменного текста. Представление о модели "смысл  $\Leftrightarrow$  текст".

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

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

### **Раздел.2 Модель лингвистической онтологии для автоматической обработки текстов (10ч.)**

#### **Тема 1. Типы онтологий: верхнего уровня, предметных областей, прикладные онтологии. Лексические онтологии. (4ч.)**

Определение понятий: онтология, концепт, отношение, аксиомы. Типы онтологий: верхнего уровня, предметных областей, прикладных онтологий. Лексические онтологии. Примеры онтологий.

#### **Тема 2. Задачи, решаемые с помощью онтологий и тезаурусов (информационный поиск, интеграция гетерогенных источников данных, SemanticWeb) (6ч.)**

Назначение онтологий. Информационный поиск. Интеграция разнородных источников данных. SemanticWeb. Онтологии предметных областей и прикладные онтологии: назначение, отличительные черты, решаемые задачи (примеры проектов). Языки описания онтологий. Основные

синтаксические структуры: классы, отношения, аксиомы. Тезаурусы. Основные принципы разработки, создания и использования традиционных информационно-поисковых тезаурусов. Примеры тезаурусов. Информационно-поисковые тезаурусы в условиях сверхбольших электронных коллекций и автоматической обработки текстов. Тезаурус для автоматического концептуального индексирования как особый вид тезауруса.

### **Раздел.3 Примеры систем (12ч.)**

#### **Тема 1. Программы анализа и лингвистической обработки текстов.(2ч.)**

Программа для синтаксического и морфологического анализа русскоязычных текстов. Автоматизированная система поиска и аналитической обработки информации.

#### **Тема 2. Программы преобразования текстов.( 2ч.)**

Программы переформатирования текстовых или html-файлов (Windows). Программы поиска и замены текста по заданным маскам. Массовые операции поиска и замены во множестве текстовых файлов.

#### **Тема 3. Генераторы текстов и "говорящие" программы.(2ч.)**

Система общения на русском языке. Автоматическое составление и перевод деловых писем и документов.

#### **Тема 4. Системы обработки естественного языка (NL-Processing).(2ч.)**

Программы поиска и анализ в текстовых массивах информации, реализованные с использованием технологии нейронных сетей. Встраивание в разрабатываемые приложения функции полнотекстового поиска и морфологического анализа текстов.

#### **Тема 5. Словари и тезаурусы.(2ч.)**

Универсальные среды для систем автоматической обработки русского текста. Словари исходных словоформ русского языка в виде zip-архивов.

#### **Тема 6. Психолингвистические программы.(2ч.)**

Программы психолингвистического анализа текстов. Лексический и контент-анализ текстов.

## **II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА**

### **Практические занятия (0час.)**

Не предусмотрены

### **Лабораторные работы (30ч.)**

**Лабораторная работа №1. (2 час.).**

Информационно-поисковые системы (ИПС).

**Лабораторная работа №2.** (4 час.).

Информационно-поисковые языки (на ограниченном естественном языке).

**Лабораторная работа №3.** (8 час.).

Индексирование. Поисковый образ документа.

**Лабораторная работа №4.** (4 час.).

Поисковый образ запроса и поисковое предписание.

**Лабораторная работа №5.** (6 час.).

Алгоритмы информационного поиска.

**Лабораторная работа №6.** (6 час.).

Программное обеспечение для информационно-поисковых систем.

### **III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ**

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Компьютерная лингвистика» 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 Компьютерная лингвистика.	ОПК-3, ПК-2	знает	Конспект (ПР-7)	Зачет, Вопросы № 1-11

	Раздел.2 Модель лингвистической онтологии для автоматической обработки текстов	ОПК-3, ПК-2	знает	Конспект (ПР-7)	Зачет, Вопросы № 12-20
2	Раздел 3. Программы анализа и лингвистической обработки текстов, Программы преобразования текстов, Генераторы текстов и "говорящие" программы.	ОПК-6, ПК-5	знает	Собеседование (УО-1),	Зачет, Вопросы № 21-24
		ПК-2, ПК-5	умеет	ПР-6 лабораторная работа 1-4	
	владеет				
3	Раздел 3. Системы обработки естественного языка (NL-Processing), Словари и тезаурусы, Психолингвистические программы.	ОПК-6, ПК-5	знает	Собеседование (УО-1),	Зачет, Вопросы № 25-27
		ПК-2, ПК-5	умеет	ПР-6 лабораторная работа 5, 6	
	владеет				

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**

*(электронные и печатные издания)*

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. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие / Н.Н. Леонтьева. – М.: Академия, 2006г.-304с.  
<https://lib.dvfu.ru:8443/lib/item?id=chamo:235710&theme=FEFU>

3. 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>
4. Онтологии и тезаурусы: модели, инструменты, приложения: Учебное пособие. / Б.В. Добров, В.В. Иванов, Н.В. Лукашевич, В.Д. Соловьев. - М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2009. - 173 с.: ил. - (Серия "Основы информационных технологий"). <http://www.iprbookshop.ru/22417>
5. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов [Электронный ресурс] / Рубашкин В.Ш. - М. : ФИЗМАТЛИТ, 2012. – 348 с. <http://www.studentlibrary.ru/book/ISBN9785922114394.html>
6. Гаврилова, Т. А. Интеллектуальные технологии в менеджменте: инструменты и системы [Электронный ресурс] : Учеб. пособие. 2-е изд. / Т. А. Гаврилова, Д. И. Муромцев; Высшая школа менеджмента СПбГУ : СПб.: Изд-во «Высшая школа менеджмента»; Издат. дом С.-Петерб. гос. ун-та, 2008. . 488 с. - ISBN 978-5-9924-0017-5 <http://znanium.com/bookread2.php?book=504514>

#### Дополнительная литература

1. Интернет-коммуникация как новая речевая формация / [Агагюлова 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>
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. Компьютерная семантика русского языка / В. А. Тузов - СПб.: Изд-во СПбГУ, 2003. – 400 с.
4. Компьютерная лингвистика. Учебное пособие / Марчук Ю.Н. - М.:Изд.АСТ, 2007. - 317 с.
5. Корпусная лингвистика / Е.В. Грудева. - М.: ФЛИНТА, 2012. - 165 с.
6. Начала компьютерной лингвистики: Учеб. пособие. / Ю.И. Шемякин. - М.: Изд-во МГОУ, А/О "Росвузнаука", 1992. - 113с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:37257&theme=FEFU>
7. 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. <http://window.edu.ru/library/pdf2txt/465/78465/59324>

8. Введение в компьютерную лингвистику. Учебное пособие. / К.К, Боярский. - СПб: НИУ ИТМО, 2013. - 72 с.  
<http://window.edu.ru/resource/387/80387>

#### **Текущие публикации в следующих изданиях:**

1. Журнал «Вопросы языкознания».
2. Журнал «Известия Академии Наук, сер. Литературы и языка.»
3. Журнал «Computational Linguistics (CL)».
4. Журнал «Computational Linguistics in the Netherlands (CLIN)».
5. Журнал «Linguistic Issues in Language Technology (LiLT)».
6. Журнал « Web journal of Formal, Computational & Cognitive Linguistics» [электронный ресурс] .-Режим доступа:<http://fccl.ksu.ru>

#### **Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:**

1. <http://dxdy.ru/topic4210-60.html> Научный форум dxdy. Моделирование русского языка
2. <http://lingvoforum.net/index.php?topic=29965.0>  
Лингвофорум>>Компьютерная лингвистика
3. <http://forum.dialog-21.ru/actualthread.aspx?tid=363> Форум Диалог
4. [http://uisrussia.msu.ru/linguist/\\_B\\_comput\\_ling.jsp](http://uisrussia.msu.ru/linguist/_B_comput_ling.jsp) Научно-образовательный портал "Лингвистика в России: ресурсы для исследователей"
5. <http://mytts.forum2x2.ru/f45-forum> Форум myTTS
6. <http://www.antic-r.ru/bibl5.htm> Книги по лингвистике
7. <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, 2011г. -272 с.
8. <http://window.edu.ru/resource/387/80387> Введение в компьютерную лингвистику. Учебное пособие. / К.К, Боярский. - СПб: НИУ ИТМО, 2013. - 72 с.

#### **Перечень информационных технологий и программного обеспечения**

1. Операционная система Windows.
2. Microsoft Office.
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.

## **VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Дисциплина изучается в следующих организационных формах: лекции, практические занятия, 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. Лингвистическая разметка.
8. Автоматическое реферирование.

### **Вопросы для самоконтроля по разделу 2**

1. Структура информационно-поискового тезауруса.
2. Построение информационно-поисковых тезаурусов.
3. Определение понятий: онтология, концепт, отношение, аксиомы.
4. Лексические онтологии.
5. Языки описания онтологий
6. Основные синтаксические структуры: классы, отношения, аксиомы.

### **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 M4716CCBAM4716CJ. 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 "КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА"**

**Направление подготовки – 02.03.03 «Математическое обеспечение и  
администрирование информационных систем»  
профиль «Технология программирования»**

**Форма подготовки очная**

**Владивосток  
2015**

### 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	Выбор темы ИПС	2ч.	Утверждение в ходе лабораторной работы
2.	2-8 недели	Поисковый образ документа. Информационно-поисковые языки Поисковый образ запроса и поисковое предписание.	10ч.	Опрос (анализ и обсуждение проработанных первоисточников в ходе лабораторных работ)
3.	9-18 недели	Алгоритмы информационного поиска.	10ч.	Опрос (анализ и обсуждение проработанных первоисточников в ходе лабораторных работ)
4.	19-28 недели	ПО для ИПС	20ч.	Демонстрация в ходе лабораторной работы
5.	1-я неделя	Самоорганизация подгрупп, выбор темы	2ч.	Утверждение в ходе практического занятия
6.	2-7 недели	Работа с первоисточниками	15ч.	Опрос (анализ и обсуждение проработанных первоисточников в ходе практического занятия)
7.	8-10 недели	Работы с лингвистическими программами	15ч.	Опрос в ходе практического занятия
8.	11-12 недели	Подготовка презентации и демонстрации работы с программами	10ч.	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.

### **Подготовка 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 "КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА"

**Направление подготовки – 02.03.03 «Математическое обеспечение и  
администрирование информационных систем»  
Профиль «Технология программирования»**

**Форма подготовки очная**

**Владивосток  
2015**

## Паспорт ФОС

Код и формулировка компетенции	Этапы формирования компетенции	
ОПК-3 готовность анализировать проблемы и направления развития технологий программирования	Знает	технологии автоматизированной обработки текстовой информации, основные принципы представления знаний о предметной области в виде рубрикаторов, тезаурусов, онтологий; особенности реализации естественно-языковых систем, назначение лингвистического процессора;
	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; определять основные классы ЕЯ-систем;
	Владеет	широким диапазоном различных информационно-коммуникационных технологий; методами поиска информации по работе с естественным языком
ОПК-6 способность определять проблемы и тенденции развития рынка программного обеспечения	Знает	сферу применения и перспективы развития информационно-поисковых тезаурусов, применение тезаурусного метода поиска в Интернет.
	Умеет	интерпретировать результаты автоматической обработки лингвистических данных;
	Владеет	методикой использования компьютерных технологий в предметной области;
ПК-2 готовность к использованию основных моделей информационных технологий и способов их применения для решения задач в предметных областях	Знает	существенные отличия естественных языков от искусственных и особенности компьютерных моделей естественного языка, какие лингвистические технологии следует применять в зависимости от условий задачи, понимать, на какие лингвистические данные и ресурсы может опираться технология ;
	Умеет	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте;
	Владеет	приемами работы с прикладным программным обеспечением;
ПК-5 готовность к использованию современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ	Знает	типичные программные системы (приложения) в области АОТ и их архитектурные особенности;
	Умеет	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте;

	Владеет	навыками работы со специальными программными средствами автоматизированной обработки текстов
--	---------	--

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел.1 Компьютерная лингвистика.	ОПК-3, ПК-2	знает	Конспект (ПР-7)	Зачет, Вопросы № 1-11
	Раздел.2 Модель лингвистической онтологии для автоматической обработки текстов	ОПК-3, ПК-2	знает	Конспект (ПР-7)	Зачет, Вопросы № 12-20
2	Раздел 3. Программы анализа и лингвистической обработки текстов, Программы преобразования текстов, Генераторы текстов и "говорящие" программы.	ОПК-6, ПК-5	знает	Собеседование (УО-1),	Зачет, Вопросы № 21-24
		ПК-2, ПК-5	умеет	ПР-6 лабораторная работа 1-4	
	владеет				
3	Раздел 3. Системы обработки естественного языка (NL-Processing), Словари и тезаурусы, Психолингвистические программы.	ОПК-6, ПК-5	знает	Собеседование (УО-1),	Зачет, Вопросы № 25-27
			умеет		
		владеет	ПР-6 лабораторная работа 5, 6		

### Шкала оценивания уровня сформированности компетенций

Код и формулировка компетенции	Этапы формирования компетенции		критерии	показатели
ОПК-3 готовность	знает (пороговый	технологию автоматизир	Знание представлений	Способность перечислить и раскрыть суть технологии

анализировать проблемы и направления развития технологий программирования	уровень)	ованной обработки текстовой информации, основные принципы представления знаний о предметной области в виде рубрикаторов, тезаурусов, онтологий	знаний о предметной области. Знание о тезаурусах, онтологиях.	автоматизированной обработки текстовой информации. Способность определить подходящие онтологии.
	умеет (продвинутой)	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 работать с лингвистическими данными	Способность получать лингвистические данные
	владеет (высокий)	широким диапазоном различных информационно-коммуникационных технологий;	Владение различными информационно-коммуникационными технологиями	Способность работать с различными информационно-коммуникационными технологиями
ОПК-3 готовность анализировать проблемы и	знает (пороговый уровень)	особенности реализации естественно-языковых систем,	Знание реализации лингвистических процессоров	Способность реализовать лингвистический процессор

направления развития технологий программирования		назначение лингвистического процессора;		
	умеет (продвинутой)	определять основные классы ЕЯ-систем;	Умение исследовать современные ЕЯ-системы	Способность работать с современными ЕЯ-системами
	владеет (высокий)	методами поиска информации по работе с естественным языком	Владение набором операций, предоставляемых современными программными средствами создания и редактирования документов	Способность выполнить практические задания по поиску информации по работе с естественным языком
ОПК-6 способность определять проблемы и тенденции развития рынка программного обеспечения	знает (пороговый уровень)	сферу применения и перспективы развития информационно-поисковых тезаурусов, применение тезаурусного метода поиска в Интернет.	Знание основные принципы разработки, создания и использования традиционных информационно-поисковых тезаурусов	Способность перечислить технические и программные средства, предназначенные для работы с тезаурусами
	умеет (продвинутой)	интерпретировать результаты автоматической обработки лингвистических данных;	Умение работать с корпусом как репрезентативная выборкой текста	Способность выполнить разметку корпуса
	владеет (высокий)	методикой использования компьютерных технологий в	Владение методикой работы с технологиями обработки ЕЯ	Способность выполнить практические задания по работе с технологиями обработки ЕЯ

		предметной области;		
ПК-2 готовность к использованию основных моделей информационных технологий и способов их применения для решения задач в предметных областях	знает (пороговый уровень)	существенные отличия естественных языков от искусственных и особенности компьютерных моделей естественного языка, какие лингвистические технологии следует применять в зависимости от условий задачи, понимать, на какие лингвистические данные и ресурсы может опираться технология ;	Знание моделей ЕЯ	Способность перечислить лингвистические технологии, дать понятие отличие ЕЯ от искусственных
	умеет (продвинутой)	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте;	Умение разрабатывать системы взаимодействия с ЭВМ на ЕЯ	Способность выбрать подходящие программные средства для выполнения разработок
	владеет (высокий)	приемами работы с прикладными программным обеспечением	Владение набором операций, предоставляемых ППО	Способность выполнить практические задания в ППО

		М;		
ПК-5 готовность к использованию современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ	знает (пороговый уровень)	типичные программные системы (приложения) в области АОТ и их архитектурные особенности	Знание архитектуры систем АОТ	Способность перечислить особенности архитектуры систем АОТ
	умеет (продвинутой)	разрабатывать системы, которые позволяли бы взаимодействовать с ЭВМ в конкретной проблемной области на естественном языке или каком-то его ограниченном варианте	Умение разрабатывать системы работающие на ЕЯ	Способность выбрать подходящие программные средства для выполнения разработок
	владеет (высокий)	навыками работы со специальными программными средствами автоматизированной обработки текстов	Владение набором операций, предоставляемых современными автоматизированной обработки текстов	Способность выполнить практические задания по работе с специальными программными средствами автоматизированной обработки текстов , продемонстрировать результаты и пояснить процесс выполнения задания

### **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 зачету 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 зачет, который проводится в устной форме.

### Критерии выставления оценки студенту на зачете

Баллы (рейтингов ой 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**

**по дисциплине «Компьютерная лингвистика»**

#### **Раздел.1 Компьютерная лингвистика.**

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. Автоматическое реферирование.
8. Автоматическое порождение текста.
9. Получение компрессированного текста.
10. Основные характеристики текста.
11. Модель "смысл  $\Leftrightarrow$  текст".

#### **Раздел.2 Модель лингвистической онтологии для автоматической обработки текстов.**

12. Назначение онтологий.
13. Онтологии предметных областей и прикладные онтологии.
14. Лексические онтологии.
15. Языки описания онтологий.
16. Основные синтаксические структуры: классы, отношения, аксиомы.
17. SemanticWeb.
18. Структура информационно-поискового тезауруса.
19. Построение информационно-поисковых тезаурусов.
20. Тезаурус для автоматического концептуального индексирования.

#### **Раздел.3 Примеры систем**

21. Программа для синтаксического и морфологического анализа русскоязычных текстов.
22. Автоматизированная система поиска и аналитической обработки информации.
23. Программы переформатирования текстовых или html-файлов
24. Автоматическое составление и перевод деловых писем и документов.
25. Программы поиска и анализ в текстовых массивах информации, реализованные с использованием технологии нейтронных сетей.
26. Универсальные среды для систем автоматической обработки русского текста.
27. Программы психолингвистического анализа текстов.

**Критерии оценки:**

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

**76-85 баллов** - 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