

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное учреждение  
высшего профессионального образования  
"Казанский (Приволжский) федеральный университет"  
Институт вычислительной математики и информационных технологий



подписано электронно-цифровой подписью

**Программа дисциплины**  
**Компьютерная лингвистика БЗ.В.1**

Направление подготовки: 230400.62 - Информационные системы и технологии

Профиль подготовки: Информационные системы в образовании

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

**Автор(ы):**

Гильмуллин Р.А. , Хайруллина Л.Э.

**Рецензент(ы):**

Сулейманов Д.Ш.

**СОГЛАСОВАНО:**

Заведующий(ая) кафедрой: Галимянов А. Ф.

Протокол заседания кафедры No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Учебно-методическая комиссия Института вычислительной математики и информационных технологий:

Протокол заседания УМК No \_\_\_\_ от " \_\_\_\_ " \_\_\_\_\_ 201\_\_ г

Регистрационный No 9118314

Казань  
2014

## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре основной образовательной программы
3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля
4. Структура и содержание дисциплины/ модуля
5. Образовательные технологии, включая интерактивные формы обучения
6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов
7. Литература
8. Интернет-ресурсы
9. Материально-техническое обеспечение дисциплины/модуля согласно утвержденному учебному плану

Программу дисциплины разработал(а)(и) Гильмуллин Р.А. , Rinat.Gilmullin@kpfu.ru ; доцент, к.н. Хайруллина Л.Э. Кафедра информационных систем отделение фундаментальной информатики и информационных технологий , Liliya.Hajrullina@kpfu.ru

### 1. Цели освоения дисциплины

Освоение методов и алгоритмов компьютерной лингвистики применительно к задачам обработки текстовой информации в информационных системах.

### 2. Место дисциплины в структуре основной образовательной программы высшего профессионального образования

Данная учебная дисциплина включена в раздел " Б3.В.1 Профессиональный" основной образовательной программы 230400.62 Информационные системы и технологии и относится к вариативной части. Осваивается на 2 курсе, 4 семестр.

Дисциплина относится к вариативной части профессионального цикла.

### 3. Компетенции обучающегося, формируемые в результате освоения дисциплины /модуля

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-5 (общекультурные компетенции)	умение применять методы и средства познания, обучения и самоконтроля для интеллектуального развития, повышения культурного уровня, профессиональной компетенции, сохранения своего здоровья, нравственного и физического самосовершенствования
ОК-7 (общекультурные компетенции)	умение критически оценивать свои достоинства и недостатки, наметить пути и выбрать средства развития достоинств и устранения недостатков
ПК-10 (профессиональные компетенции)	готовность разрабатывать, согласовывать и выпускать все виды проектной документации
ПК-17 (профессиональные компетенции)	готовность проводить подготовку документации по менеджменту качества информационных технологий
ПК-2 (профессиональные компетенции)	способность проводить техническое проектирование

В результате освоения дисциплины студент:

1. должен знать:

Методы компьютерной лингвистики применительно к обработке текстов на естественном языке.

2. должен уметь:

Определять различные вероятностные характеристики текстов на естественном языке. Строить различного типа словари и методы их использования.

3. должен владеть:

Методами анализа текстов на естественном языке.

4. должен демонстрировать способность и готовность:

применять полученные знания в профессиональной деятельности

#### 4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) 180 часа(ов).

Форма промежуточного контроля дисциплины экзамен в 4 семестре.

Суммарно по дисциплине можно получить 100 баллов, из них текущая работа оценивается в 50 баллов, итоговая форма контроля - в 50 баллов. Минимальное количество для допуска к зачету 28 баллов.

86 баллов и более - "отлично" (отл.);

71-85 баллов - "хорошо" (хор.);

55-70 баллов - "удовлетворительно" (удов.);

54 балла и менее - "неудовлетворительно" (неуд.).

#### 4.1 Структура и содержание аудиторной работы по дисциплине/ модулю

##### Тематический план дисциплины/модуля

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Введение	4	1-2	4	0	4	домашнее задание
2.	Тема 2. Информационные измерения в текстах на естественном языке	4	3-4	4	0	4	домашнее задание
3.	Тема 3. Статистические законы для лингвистических величин.	4	5-6	4	0	4	домашнее задание
4.	Тема 4. Корпусная лингвистика	4	7-8	4	0	4	контрольная точка
5.	Тема 5. Морфологические модели	4	9-10	4	0	4	домашнее задание
6.	Тема 6. Электронные словари	4	11-12	4	0	4	домашнее задание
7.	Тема 7. Синтаксические модели	4	13-14	4	0	4	домашнее задание
8.	Тема 8. Семантический анализ текстов	4	15-16	4	0	4	домашнее задание
9.	Тема 9. Задача автоматического перевода текстов на естественных языках	4	17-18	4	0	4	контрольная работа

N	Раздел Дисциплины/ Модуля	Семестр	Неделя семестра	Виды и часы аудиторной работы, их трудоемкость (в часах)			Текущие формы контроля
				Лекции	Практические занятия	Лабораторные работы	
Итого				36	0	36	

## 4.2 Содержание дисциплины

### Тема 1. Введение

#### **лекционное занятие (4 часа(ов)):**

Цели и проблемы компьютерной лингвистики. Основные задачи и методологии их решения. История возникновения и развития компьютерной лингвистики.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 1. Работа с переводчиками. Часть 1.

### Тема 2. Информационные измерения в текстах на естественном языке

#### **лекционное занятие (4 часа(ов)):**

Энтропия лингвистического эксперимента. Определение количества информации в текстах на естественном языке. Контекстная обусловленность и избыточность текста. Измерение смысловой информации в тексте.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 2. Работа с переводчиками. Часть 2.

### Тема 3. Статистические законы для лингвистических величин.

#### **лекционное занятие (4 часа(ов)):**

Частотные характеристики словника. Смысловые интерпретации законов, описывающих частотные характеристики словарей и их универсальность в задачах информатики.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 3. Работа с синтаксическими анализаторами. Часть 1.

### Тема 4. Корпусная лингвистика

#### **лекционное занятие (4 часа(ов)):**

Возникновение и развитие корпусной лингвистики. Разметка корпусов текстов. Теоретические вопросы корпусной лингвистики. Противопоставление корпусной и когнитивной лингвистик. Применение в задачах автоматической обработки текстов.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 4. Работа с синтаксическими анализаторами. Часть 2.

### Тема 5. Морфологические модели

#### **лекционное занятие (4 часа(ов)):**

Морфологический разбор слов естественного языка. Модели построения компьютерных морфологий. Словарные морфологии, модели и проблемы их использования. Аналитические методы морфологического разбора обработки текстов. Теория ?Смысл-текст?. Лексические функции.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 5. Работа с морфологическими анализаторами. Часть 1.

### Тема 6. Электронные словари

#### **лекционное занятие (4 часа(ов)):**

Организация электронных словарей. Эффективные алгоритмы и форматы хранения словарей. Пословные переводчики. Псевдоморфология в задаче поиска по словарю.

#### **лабораторная работа (4 часа(ов)):**

Лабораторная работа 6. Работа с морфологическими анализаторами. Часть 2.

### Тема 7. Синтаксические модели

#### **лекционное занятие (4 часа(ов)):**

Задача синтаксического разбора. Фрагментация. Синтаксический граф предложения на естественном языке и проблема множественности.

**лабораторная работа (4 часа(ов)):**

Лабораторная работа 7. Работа с семантическими анализаторами. Часть 1.

**Тема 8. Семантический анализ текстов**

**лекционное занятие (4 часа(ов)):**

Семантический и синтаксический граф предложения на естественном языке и проблема множественности. Задача выделения смысла текста. Функциональные модели описания естественного языка.

**лабораторная работа (4 часа(ов)):**

Лабораторная работа 8. Работа с семантическими анализаторами. Часть 1.

**Тема 9. Задача автоматического перевода текстов на естественных языках**

**лекционное занятие (4 часа(ов)):**

Автоматический перевод текстов с одного естественного языка на другой как центральная проблема компьютерной лингвистики. Математические аспекты. Различные типы систем машинного перевода. Перспективы развития систем искусственного интеллекта для анализа текстов на естественном языке.

**лабораторная работа (4 часа(ов)):**

Лабораторная работа 9. Построение лингвопрограмм.

**4.3 Структура и содержание самостоятельной работы дисциплины (модуля)**

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Введение	4	1-2	подготовка домашнего задания	6	домашнее задание
2.	Тема 2. Информационные измерения в текстах на естественном языке	4	3-4	подготовка домашнего задания	6	домашнее задание
3.	Тема 3. Статистические законы для лингвистических величин.	4	5-6	подготовка домашнего задания	6	домашнее задание
4.	Тема 4. Корпусная лингвистика	4	7-8	подготовка к контрольной точке	6	контрольная точка
5.	Тема 5. Морфологические модели	4	9-10	подготовка домашнего задания	6	домашнее задание
6.	Тема 6. Электронные словари	4	11-12	подготовка домашнего задания	6	домашнее задание
7.	Тема 7. Синтаксические модели	4	13-14	подготовка домашнего задания	6	домашнее задание

N	Раздел Дисциплины	Семестр	Неделя семестра	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
8.	Тема 8. Семантический анализ текстов	4	15-16	подготовка домашнего задания	6	домашнее задание
9.	Тема 9. Задача автоматического перевода текстов на естественных языках	4	17-18	подготовка к контрольной работе	6	контрольная работа
	Итого				54	

## 5. Образовательные технологии, включая интерактивные формы обучения

Самостоятельно студенты готовятся к лабораторным работам. Используются разборы конкретных ситуаций.

## 6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

### Тема 1. Введение

домашнее задание , примерные вопросы:

Повторение материала: Цели и проблемы компьютерной лингвистики. Основные задачи и методологии их решения. История возникновения и развития компьютерной лингвистики. подготовка к лабораторной работе

### Тема 2. Информационные измерения в текстах на естественном языке

домашнее задание , примерные вопросы:

Повторение материала: Энтропия лингвистического эксперимента. Определение количества информации в текстах на естественном языке. Контекстная обусловленность и избыточность текста. Измерение смысловой информации в тексте, подготовка к лабораторной работе

### Тема 3. Статистические законы для лингвистических величин.

домашнее задание , примерные вопросы:

Повторение материала: Частотные характеристики словника. Смысловые интерпретации законов, описывающих частотные характеристики словарей и их универсальность в задачах информатики , подготовка к лабораторной работе

### Тема 4. Корпусная лингвистика

контрольная точка , примерные вопросы:

Контрольный опрос по темам 1-4. Повторение темы Возникновение и развитие корпусной лингвистики. Разметка корпусов текстов. Теоретические вопросы корпусной лингвистики. Противопоставление корпусной и когнитивной лингвистик. Применение в задачах автоматической обработки текстов.

### Тема 5. Морфологические модели

домашнее задание , примерные вопросы:

Повторение материала: Морфологический разбор слов естественного языка. Модели построения компьютерных морфологий. Словарные морфологии, модели и проблемы их использования. Аналитические методы морфологического разбора обработки текстов. Теория ?Смысл-текст?. Лексические функции. лабораторной работе

### Тема 6. Электронные словари

домашнее задание , примерные вопросы:

Повторение материала: Организация электронных словарей. Эффективные алгоритмы и форматы хранения словарей. Пословные переводчики. Псевдоморфология в задаче поиска по словарю. , подготовка к лабораторной работе

### **Тема 7. Синтаксические модели**

домашнее задание , примерные вопросы:

Повторение материала: Задача синтаксического разбора. Фрагментация. Синтаксический граф предложения на естественном языке и проблема множественности. , подготовка к лабораторной работе

### **Тема 8. Семантический анализ текстов**

домашнее задание , примерные вопросы:

Повторение материала: Семантический и синтаксический граф предложения на естественном языке и проблема множественности. Задача выделения смысла текста. Функциональные модели описания естественного языка. , подготовка к лабораторной работе

### **Тема 9. Задача автоматического перевода текстов на естественных языках**

контрольная работа , примерные вопросы:

Контрольная работа по темам 5-9

Примерные вопросы к экзамену:

Примерные вопросы:

1. Цели и проблемы компьютерной лингвистики.
2. История возникновения и развития компьютерной лингвистики.
3. Определение количества информации в текстах на естественном языке. Контекстная обусловленность и избыточность текста. Измерение смысловой информации в тексте.
4. Частотные характеристики словника.
5. Смысловые интерпретации законов, описывающих частотные характеристики словарей и их универсальность в задачах информатики.
6. Возникновение и развитие корпусной лингвистики.
7. Разметка корпусов текстов. Теоретические вопросы корпусной лингвистики. Вероятностные характеристики словоупотреблений в текстах на естественных языках.
8. Лингвистическая типология в компьютерной лингвистике.
9. Морфологический разбор слов естественного языка.
10. Модели построения компьютерных морфологий.
11. Словарные морфологии, модели и проблемы их использования.
12. Организация электронных словарей.
13. Эффективные алгоритмы и форматы хранения словарей. Пословные переводчики.
14. Задача синтаксического разбора. Фрагментация.
15. Синтаксический граф предложения на естественном языке и проблема множественности.
16. Семантический и синтаксический граф предложения на естественном языке и проблема множественности.
17. Задача выделения смысла текста.
18. Функциональные модели описания естественного языка.
19. Автоматический перевод текстов с одного естественного языка на другой.
20. Различные типы систем машинного перевода.
21. Перспективы развития систем искусственного интеллекта для анализа текстов на естественном языке.

#### **7.1. Основная литература:**

Информатика. Базовый курс, Симонович, С. В., 2008г.

Новые информационные технологии и лингвистика, Потапова, Родмонга Кондратьевна, 2012г.



1. Грудева Е. В. Корпусная лингвистика [Электронный ресурс] : учеб. пособие / Е. В. Грудева. - 2-е изд., стер. - М. : ФЛИНТА, 2012. - 165 с. <http://znanium.com/bookread.php?book=455049>
2. Червяков Н.И., Евдокимов А.А., Галушкин А.И. Применение искусственных нейронных сетей и системы остаточных классов в криптографии. - - М.: Физматлит, 2012. - 280 с. URL: [http://e.lanbook.com/books/element.php?pl1\\_id=5300](http://e.lanbook.com/books/element.php?pl1_id=5300)
3. Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс] : учеб. пособие / Л. Ю. Щипицина. ? М. : ФЛИНТА, 2013. ? 128 с. <http://znanium.com/bookread.php?book=462989>

## 7.2. Дополнительная литература:

1. Курош А.Г. Курс высшей алгебры/Лань,2013-332с. <http://e.lanbook.com/view/book/30198/>
2. Осипов, Д.Л.. Базы данных и Delphi: теория и практика: [+ пробные версии ПО] / Дмитрий Осипов.?Санкт-Петербург: БХВ-Петербург, 2011.?733 с.- Режим доступа: <http://znanium.com/bookread.php?book=355202>

## 7.3. Интернет-ресурсы:

- Автоматическая обработка текста - <http://aot.ru/>
- Материалы кафедры компьютерной лингвистики - <http://il.rsuh.ru/section.html?id=9661>
- Методы лингвистической семантики для повышения точности и полноты поиска - <http://www.intuit.ru/studies/courses/579/435/info>
- Модели поведения, восприятия и мышления - <http://www.intuit.ru/studies/courses/2191/423/info>
- Проектирование систем искусственного интеллекта - <http://www.intuit.ru/studies/courses/1122/167/info>

## 8. Материально-техническое обеспечение дисциплины(модуля)

Освоение дисциплины "Компьютерная лингвистика" предполагает использование следующего материально-технического обеспечения:

Мультимедийная аудитория, вместимостью более 60 человек. Мультимедийная аудитория состоит из интегрированных инженерных систем с единой системой управления, оснащенная современными средствами воспроизведения и визуализации любой видео и аудио информации, получения и передачи электронных документов. Типовая комплектация мультимедийной аудитории состоит из: мультимедийного проектора, автоматизированного проекционного экрана, акустической системы, а также интерактивной трибуны преподавателя, включающей тач-скрин монитор с диагональю не менее 22 дюймов, персональный компьютер (с техническими характеристиками не ниже Intel Core i3-2100, DDR3 4096Mb, 500Gb), конференц-микрофон, беспроводной микрофон, блок управления оборудованием, интерфейсы подключения: USB, audio, HDMI. Интерактивная трибуна преподавателя является ключевым элементом управления, объединяющим все устройства в единую систему, и служит полноценным рабочим местом преподавателя. Преподаватель имеет возможность легко управлять всей системой, не отходя от трибуны, что позволяет проводить лекции, практические занятия, презентации, вебинары, конференции и другие виды аудиторной нагрузки обучающихся в удобной и доступной для них форме с применением современных интерактивных средств обучения, в том числе с использованием в процессе обучения всех корпоративных ресурсов. Мультимедийная аудитория также оснащена широкополосным доступом в сеть интернет. Компьютерное оборудование имеет соответствующее лицензионное программное обеспечение.

Компьютерный класс, представляющий собой рабочее место преподавателя и не менее 15 рабочих мест студентов, включающих компьютерный стол, стул, персональный компьютер, лицензионное программное обеспечение. Каждый компьютер имеет широкополосный доступ в сеть Интернет. Все компьютеры подключены к корпоративной компьютерной сети КФУ и находятся в едином домене.

Используется компьютерный класс

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 230400.62 "Информационные системы и технологии" и профилю подготовки Информационные системы в образовании .

Автор(ы):

Гильмуллин Р.А. \_\_\_\_\_

Хайруллина Л.Э. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.

Рецензент(ы):

Сулейманов Д.Ш. \_\_\_\_\_

"\_\_" \_\_\_\_\_ 201\_\_ г.