

**ГОСУДАРСТВЕННОЕ НАУЧНОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
АКАДЕМИЯ НАУК РЕСПУБЛИКИ ТАТАРСТАН**

СОГЛАСОВАНО

Вице-президент

В.В. Хоменко

« 12 » _____ 2017 г.



УТВЕРЖДЕНО

Директор института

Д.Ш. Сулейманов

« 12 » _____ 2017 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА»**

Уровень: подготовка научно-педагогических кадров (аспирантура)

Направление подготовки кадров высшей квалификации:

02.06.01 Компьютерные и информационные науки

05.13.17 - Теоретические основы информатики

Квалификация выпускника: Исследователь. Преподаватель-исследователь

Нормативный срок освоения программы: 3 года.

Форма обучения: очная

Казань


Рабочая программа составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (уровень подготовки кадров высшей квалификации) по направлению подготовки 02.06.01 Компьютерные и информационные науки, утвержденным приказом Минобрнауки РФ от 30.07.2014 г. № 864; паспортом специальности научных работников 05.13.17 – Теоретическая информатика; учебным планом Института «Прикладная семиотика» Академии наук Республики Татарстан.

Составитель рабочей программы:

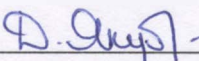
В.Н.С., К.Ф.-М.Н.  А.Ф. Галимянов
(должность, ученая степень) (подпись) (Ф.И.О.)

СОГЛАСОВАНО:

Заместитель директора по научной работе

« 12 » сентября 2017 г.  О.А. Невзорова
(подпись) (Ф.И.О.)

Ученый секретарь

« 12 » сентября 2017 г.  Д.Д. Якубова
(подпись) (Ф.И.О.)

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 " Б1.В.ОД.2 Профессиональный" основной образовательной программы 05.13.17 «Теоретические основы информатики» и относится к базовой (общепрофессиональной) части. Осваивается на 1 и 2 курсах.

Данная дисциплина является базовой для цикла общепрофессиональной подготовки в области использования информационных технологий, и является основой для построения всех остальных смежных дисциплин данной области.

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

В результате освоения дисциплины формируются следующие компетенции:

Шифр компетенции	Расшифровка приобретаемой компетенции	Большое влияние в приобретении	Среднее влияние в приобретении	Малое влияние в приобретении
УК-1	способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях	x		
УК-2	способность проектировать и осуществлять комплексные исследования, в том числе междисциплинарные, на основе целостного системного научного мировоззрения с использованием знаний в области истории и философии науки		x	
УК-3	готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и образовательных задач	x		
УК-4	готовность использовать современные методы и технологии научной коммуникации на государственном и иностранных языках		x	
УК-5	способность планировать и решать задачи собственного профессионального и личностного развития			x
ОПК-1	способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий		x	
ОПК-2	готовность к преподавательской деятельности по основным образовательным программам высшего образования	x		
Профессиональные компетенции				
ПК-1	Готовность использовать для решения конкретных исследовательских задач методы современных информационных технологий	x		
ПК-2	Готовность к педагогической деятельности в поликультурной среде с использованием		x	

Шифр компетенции	Расшифровка приобретаемой компетенции	Большое влияние в приобретении	Среднее влияние в приобретении	Малое влияние в приобретении
	современных информационных технологий			

В результате освоения дисциплины аспирант:

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.

применять полученные знания в профессиональной деятельности

4. Структура и содержание дисциплины/ модуля

Общая трудоемкость дисциплины составляет 6 зачетных(ые) единиц(ы) 216 часа(ов).

Форма промежуточного контроля дисциплины зачет во 2-м курсе.

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

№	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	1	1-2	4	4		домашнее задание
2.	Тема 2. Информационные измерения в текстах на естественном языке	1	3-4	4	4		домашнее задание
3.	Тема 3. Статистические законы для лингвистических величин.	1	5-6	8	4		домашнее задание
4.	Тема 4. Корпусная лингвистика	1	7-8	8	4		контрольная точка
5.	Тема 5. Морфологические модели	1	9-10	8	8		домашнее задание
6.	Тема 6. Электронные словари	2	11-12	8	8		домашнее задание
7.	Тема 7. Синтаксические модели	2	13-14	8	8		домашнее задание
8.	Тема 8. Семантический анализ текстов	2	15-16	8	8		домашнее задание
9.	Тема 9. Задача автоматического перевода текстов на естественных языках	2	17-18	8	8		контрольная работа
	Тема . Итоговая форма контроля	2		0	8		экзамен
	Итого			64	64		

4.2 Содержание дисциплины

Тема 1. Введение

лекционное занятие

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

Практическая работа 1

Тема 2. Информационные измерения в текстах на естественном языке

лекционное занятие

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

Практическая работа 2

Тема 3. Статистические законы для лингвистических величин.

лекционное занятие

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

Практическая работа 3

Тема 4. Корпусная лингвистика

лекционное занятие

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

Практическая работа 4

Тема 5. Морфологические модели

лекционное занятие

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.

Практическая работа 5

Тема 6. Электронные словари

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

Практическая работа 6

Тема 7. Синтаксические модели

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

Практическая работа 7

Тема 8. Семантический анализ текстов

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

Практическая работа 8

Тема 9. Задача автоматического перевода текстов на естественных языках

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)):

Практическая работа 9

4.3 Структура и содержание самостоятельной работы дисциплины (модуля)

N	Раздел Дисциплины	Курс	Виды самостоятельной работы студентов	Трудоемкость (в часах)	Формы контроля самостоятельной работы
1.	Тема 1. Введение	1	подготовка домашнего задания	8	домашнее задание
2.	Тема 2. Информационные измерения в текстах на естественном языке	1	подготовка домашнего задания	10	домашнее задание
3.	Тема 3. Статистические законы для лингвистических величин.	1	подготовка домашнего задания	10	домашнее задание
4.	Тема 4. Корпусная лингвистика	1	подготовка к контрольной точке	10	контрольная точка
5.	Тема 5. Морфологические модели	1	подготовка домашнего задания	10	домашнее задание
6.	Тема 6. Электронные словари	2	подготовка домашнего задания	10	домашнее задание
7.	Тема 7. Синтаксические модели	2	подготовка домашнего задания	10	домашнее задание
8.	Тема 8. Семантический анализ текстов	2	подготовка домашнего задания	10	домашнее задание
9.	Тема 9. Задача автоматического перевода текстов на естественных языках	2	подготовка к контрольной работе	10	контрольная работа
	Итого			88	

5. Образовательные технологии, включая интерактивные формы обучения

Самостоятельно аспиранты готовятся к практическим работам. Используются разборы конкретных ситуаций.

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

самостоятельной работы аспирантов

Тема 1. Введение

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 2. Информационные измерения в текстах на естественном языке

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 3. Статистические законы для лингвистических величин.

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 4. Корпусная лингвистика

контрольная точка , примерные вопросы:

Темы 1-4

Тема 5. Морфологические модели

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 6. Электронные словари

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 7. Синтаксические модели

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 8. Семантический анализ текстов

домашнее задание , примерные вопросы:

Повторение материала лекции, подготовка к лабораторной работе

Тема 9. Задача автоматического перевода текстов на естественных языках

контрольная работа , примерные вопросы:

Темы 5-9

Фонд оценочных средств по предмету с приобретаемыми компетенциями

№	Вопросы/контрольные	УК-1	УК-2	УК-3	УК-4	УК-5	ОПК-1	ОПК-2	ПК-1	ПК-2
1.	Цели и проблемы компьютерной лингвистики.		х	х		х	х		х	
2.	История возникновения и развития компьютерной лингвистики.	х		х	х		х	х	х	
3.	Определение количества информации в текстах на естественном языке. Контекстная обусловленность и		х	х		х		х		

	избыточность текста. Измерение смысловой информации в тексте.								
4.	Частотные характеристики словаря.	х		х	х		х	х	х
5.	Смысловые интерпретации законов, описывающих частотные характеристики словарей и их универсальность в задачах информатики.		х		х	х		х	
6.	Возникновение и развитие корпусной лингвистики.	х		х		х	х		х
7.	Разметка корпусов текстов. Теоретические вопросы корпусной лингвистики. Вероятностные характеристики словоупотреблений в текстах на естественных языках.	х	х		х		х	х	х
8.	Лингвистическая типология в компьютерной лингвистике.		х	х		х	х		х
9.	Морфологический разбор слов естественного языка.	х		х	х		х	х	х
10.	Модели построения компьютерных морфологий.	х	х			х	х		х
11.	Словарные морфологии, модели и проблемы их использования.	х		х	х		х		
12.	Организация электронных словарей.		х	х		х			х
13.	Эффективные алгоритмы и форматы хранения словарей. Пословные переводчики.	х		х		х		х	х
14.	Задача синтаксического разбора. Фрагментация.	х	х		х		х		
15.	Синтаксический граф предложения на естественном языке и проблема множественности.			х		х		х	х
16.	Семантический и синтаксический граф предложения на естественном языке и проблема множественности.		х		х		х	х	х
17.	Задача выделения смысла		х		х	х		х	х

	текста.								
18.	Функциональные модели описания естественного языка.	х		х	х		х		х
19.	Автоматический перевод текстов с одного естественного языка на другой.		х		х	х		х	х
20.	Различные типы систем машинного перевода.	х		х		х	х		х
21.	Перспективы развития систем искусственного интеллекта для анализа текстов на естественном языке.		х		х	х		х	

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

7.2. Дополнительная литература:

1. Современные информационные технологии для гуманитария: Практическое руководство / А.Т. Хроленко, А.В. Денисов. - М.: Флинта: Наука, 2007. - 128 с. <http://znanium.com/catalog.php?bookinfo=320764>

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 и находятся в едином домене.

Используется компьютерный класс

Программа составлена в соответствии с требованиями ФГОС ВПО и учебным планом по направлению 02.06.01 Компьютерные и информационные науки и профилю 05.13.17 «Теоретические основы информатики»..

Автор(ы):

Галимянов А.Ф. 

"12" сентября 2017 г.

Рецензент(ы):

Невзорова О.А. 

"12" сентября 2017 г.