

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
"Казанский (Приволжский) федеральный университет"  
Институт искусственного интеллекта, робототехники и системной инженерии



**УТВЕРЖДАЮ**

Проректор по образовательной деятельности КФУ

\_\_\_\_\_ Турилова Е.А.

"\_\_" \_\_\_\_\_ 20\_\_ г.

## **Программа дисциплины**

Компьютерная лингвистика и семиотика

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

## Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
  - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
  - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): заместитель директора центра Кокунин П.А. (Научно-исследовательский центр Центр превосходства Специальная робототехника и искусственный интеллект, Институт вычислительной математики и информационных технологий), PAKokunin@kpfu.ru

### **1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО**

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- основные понятия компьютерной лингвистики и семиотики;
- алгоритмы и методы анализа текста, включая синтаксический, семантический и прагматический анализ;
- основы корпусного анализа, включая сбор, разметку и анализ больших объемов текстовых данных;
- методы машинного обучения и искусственного интеллекта;

Должен уметь:

- работать с программными инструментами для обработки естественного языка;
- анализировать и обрабатывать текстовые данные с использованием методов компьютерной лингвистики;
- использовать алгоритмы машинного обучения для решения конкретных задач.

Должен владеть:

- навыками понимания и интерпретации результатов анализа текста, полученных с помощью компьютерных программ;
- навыками практического использования языков программирования и различных библиотек;
- навыками разработки и создания программ для обработки языковых данных;
- навыками работы с алгоритмами машинного обучения.

Должен демонстрировать способность и готовность:

- применять полученные знания и навыки в практической деятельности;
- совершенствовать свои знания и навыки в области компьютерной лингвистики и семиотики;
- адаптироваться к новым технологиям и методам работы.

### **2. Место дисциплины (модуля) в структуре ОПОП ВО**

Данная дисциплина (модуль) включена в раздел "Б1.В.14 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.04 "Программная инженерия (Разработчик искусственного интеллекта и когнитивных систем)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 семестре.

### **3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся**

Общая трудоемкость дисциплины составляет 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 36 часа(ов), в том числе лекции - 12 часа(ов), практические занятия - 24 часа(ов), лабораторные работы - 0 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 99 часа(ов).

Контроль (зачёт / экзамен) - 9 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 3 семестре.

### **4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий**

#### **4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)**

N	Разделы дисциплины / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная ра- бота
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
N	Разделы дисциплины / модуля	Се- местр	Виды и часы контактной работы, их трудоемкость (в часах)						Само- стоя- тель- ная ра- бота
			Лекции, всего	Лекции в эл. форме	Практи- ческие занятия, всего	Практи- ческие в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
1.	Тема 1. Обработка естественного языка	3	1	0	1	0	0	0	8
2.	Тема 2. Корпусная лингвистика	3	1	0	2	0	0	0	6
3.	Тема 3. Анализ настроений	3	1	0	2	0	0	0	6
4.	Тема 4. Машинный перевод	3	1	0	2	0	0	0	10
5.	Тема 5. Поиск информации	3	1	0	2	0	0	0	8
6.	Тема 6. Распознавание именованных объектов	3	1	0	2	0	0	0	9
7.	Тема 7. Вычислительная семантика	3	1	0	2	0	0	0	10
8.	Тема 8. Анализ дискурса	3	1	0	2	0	0	0	8
9.	Тема 9. Генерация языка	3	1	0	1	0	0	0	8
10.	Тема 10. Лингвистические ресурсы и лексическое приобретение	3	1	0	3	0	0	0	6
11.	Тема 11. Прагматика и моделирование диалога	3	1	0	3	0	0	0	10
<b>4.2 Содержание дисциплины (модуля)</b>									
	Тема 12. Вычислительная фонетика и фонология	3	1	0	2	0	0	0	10
Семиотические основы обработки естественного языка. Разработка алгоритмов для автоматической обработки текстов и определения их тематики. Исследование синтаксиса и грамматики языков для улучшения качества обработки компьютерами. Анализ символических систем и их использование в компьютерной лингвистике и семиотике.									99

**Тема 2. Корпусная лингвистика**

Корпусная лингвистика и ее роль в компьютерной лингвистике и семиотике. Создание и использование корпусов для анализа естественного языка и обучения машинного перевода. Инструменты для работы с текстовыми корпусами. Корпусы и анализ дискурса. Создание и использование корпусов для изучения диалектов и различий между языками.

**Тема 3. Анализ настроений**

Анализ эмоциональной окраски текстов в компьютерной лингвистике и семиотике. Использование корпусов для определения тональности и эмоциональной окраски текстов. Анализ языковых выражений эмоций и их использование в машинном обучении. Классификация настроений и эмоций с использованием машинного обучения. Анализ настроений и искусственный интеллект: перспективы развития.

**Тема 4. Машинный перевод**

Анализ методов машинного перевода и их применение в современных системах. Использование статистических и нейронных методов для улучшения качества машинного перевода. Проблемы машинного перевода, связанные с многоязычностью и культурными различиями. Оценка качества машинного перевода с помощью метрик и экспертных оценок.

**Тема 5. Поиск информации**

Источники и методы поиска информации в компьютерной лингвистике и семиотике. Анализ и классификация текстов для эффективного поиска информации. Индексно-поисковые системы и их применение в компьютерной лингвистике. Семантический поиск и анализ текста. Извлечение ключевых слов и фраз для поиска информации. Проблемы поиска информации в многоязычном контексте и методы их решения.

**Тема 6. Распознавание именованных объектов**

Методы выделения именованных объектов из текста. Обработка именованных объектов с использованием машинного обучения и искусственного интеллекта. Применение методов распознавания именованных объектов для автоматического перевода. Распознавание людей, организаций и мест в тексте. Интеграция систем распознавания именованных объектов с вопросно-ответными системами.

**Тема 7. Вычислительная семантика**

Семантические представления языковых единиц в компьютерной лингвистике. Методы вычисления семантической близости слов и предложений. Применение семантического анализа в вопросно-ответных системах и автоматическом переводе. Семантический поиск информации и извлечение знаний. Корпусная семантика и анализ больших объемов текстовых данных. Применение вычислительной семантики в обучающих системах и анализе тональности текста.

#### **Тема 8. Анализ дискурса**

Дискурс-анализ и компьютерная лингвистика. Семиотический подход к анализу дискурса. Грамматика и синтаксис дискурса в компьютерной лингвистике. Методы анализа дискурса и их применение в обработке естественного языка. Анализ тональности дискурса и извлечение субъективных мнений.

#### **Тема 9. Генерация языка**

Семиотические основы генерации связного текста: символы, знаки и коды. Генерация текста с использованием нейросетевых моделей: возможности и ограничения. Применение алгоритмов машинного обучения для создания связных и контекстуальных текстов. Адекватность и связность текста: подходы к измерению и улучшению.

#### **Тема 10. Лингвистические ресурсы и лексическое приобретение**

Разработка и использование лингвистических баз данных. Интеграция лингвистических ресурсов в системы автоматического перевода и генерации текстов. Проблемы оцифровки и доступности лингвистических ресурсов. Создание и использование онтологий языка для расширения лингвистических ресурсов.

#### **Тема 11. Прагматика и моделирование диалога**

Семиотика как основа моделирования диалога в компьютерной лингвистике. Методы и инструменты анализа диалога в компьютерной лингвистике. Диалог как система знаков и его моделирование в компьютерной лингвистике. Моделирование диалога с учетом его прагматики.

#### **Тема 12. Вычислительная фонетика и фонология**

Математическое моделирование звуковых сигналов в вычислительной фонетике. Методы автоматической сегментации и классификации звуков в вычислительном фонетическом анализе. Применение вычислительной фонетики в области распознавания речи. Фонотактика в вычислительных фонологических моделях. Вычислительная фонология и автоматический анализ текста. Применение вычислительных фонетических и фонологических методов в обработке естественного языка.

### **5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)**

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

Журнал - [http://www.creativeconomy.ru/mag\\_gp/](http://www.creativeconomy.ru/mag_gp/)

Журнал - [http://www.basw-ngo.by/page.php?issue\\_id=2855](http://www.basw-ngo.by/page.php?issue_id=2855)

Правительство РФ - <http://government.ru/>

### **6. Фонд оценочных средств по дисциплине (модулю)**

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемому результату обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

### 7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

### 8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Журнал - [http://www.creativeconomy.ru/mag\\_rp/](http://www.creativeconomy.ru/mag_rp/)

Журнал - [http://www.basw-ngo.by/page.php?issue\\_id=2855](http://www.basw-ngo.by/page.php?issue_id=2855)

Правительство РФ - <http://government.ru/>

### 9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	Лекционные занятия проходят в интерактивной форме, предполагающей вовлечение обучающихся в обсуждение всех предложенных тем. Применяются такие формы лекционных занятий как лекция-презентация, лекция-дискуссия, проблемная лекция, видео-лекция. Студенты активно участвуют в конструировании знаний во время круглых столов, дискуссионных площадок.
практические занятия	Практические занятия, семинары являются одной из основных форм образовательного процесса, ориентированной на усвоение студентами теоретического материала и выработку практических компетенций. Основной целью практических занятий является комплексный контроль усвоения пройденного материала, хода выполнения студентами самостоятельной работы и рассмотрение наиболее сложных и спорных вопросов в рамках темы занятия. Подготовка к семинарам предполагает самостоятельную работу студентов по изучению материала по конкретной теме.
самостоятельная работа	Самостоятельная работа преследует цель закрепить, углубить и расширить знания, полученные студентами в ходе аудиторных занятий, а также сформировать навыки работы с научной, учебной и учебно-методической литературой, развивать творческое, продуктивное мышление обучаемых, их креативные качества, формирование общекультурных и профессиональных компетенций.

Вид работ	Методические рекомендации
экзамен	Экзамен проводится в письменной форме. В билет включаются теоретические вопросы и задачи из перечня вопросов для подготовки к зачету. Студенту дается 90 минут для выполнения своего варианта задания. По завершению основной части экзамена обучающийся может добрать необходимые баллы в ходе устного опроса студента преподавателем.

**10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)**

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

**11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)**

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

**12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья**

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи;
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.03.04 "Программная инженерия" и профилю подготовки "Разработчик искусственного интеллекта и когнитивных систем".

*Приложение 2*  
*к рабочей программе дисциплины (модуля)*  
*Б1.В.14 Компьютерная лингвистика и семиотика*

**Перечень литературы, необходимой для освоения дисциплины (модуля)**

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

**Основная литература:**

1. Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. - 3-е изд., перераб. - Санкт-Петербург : СПбГУ, 2020. - 234 с. - ISBN 978-5-288-05997-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1244746> (дата обращения: 29.01.2024). - Режим доступа: по подписке.
2. Шишкин, А. Г. Методы цифровой обработки и распознавания речи : монография / А.Г. Шишкин. - Москва : ИНФРА-М, 2023. - 347 с. - (Научная мысль). - DOI 10.12737/1904325. - ISBN 978-5-16-018017-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1904325> (дата обращения: 29.01.2024). - Режим доступа: по подписке.
3. Плас, Дж. В. Python для сложных задач: наука о данных и машинное обучение : практическое руководство / Дж. В. Плас. - Санкт-Петербург : Питер, 2021. - 576 с. - (Серия 'Бестселлеры O'Reilly'). - ISBN 978-5-4461-0914-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1739601> (дата обращения: 29.01.2024). - Режим доступа: по подписке.

**Дополнительная литература:**

1. Лимановская, О. В. Основы машинного обучения : учебное пособие / О. В. Лимановская, Т. И. Алферьева. - 2-е изд., стер. - Москва : ФЛИНТА : Изд-во Урал. ун-та, 2022. - 88 с. - ISBN 978-5-9765-5006-3 (ФЛИНТА) ; ISBN 978-5-7996-3015-7 (Изд-во Урал. ун-та). - Текст : электронный. - URL: <https://znanium.com/catalog/product/1891377> (дата обращения: 29.01.2024). - Режим доступа: по подписке.
2. Груздев, А. В. Предварительная подготовка данных в Python: Том 1. Инструменты и валидация : практическое руководство / А. В. Груздев. - Москва : ДМК Пресс, 2023. - 816 с. - ISBN 978-5-93700-156-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2109509> (дата обращения: 29.01.2024). - Режим доступа: по подписке.
3. Филиппов, К. А. Лингвистика текста и современный анализ устной речи : учебное пособие / К. А. Филиппов. - Санкт-Петербург : СПбГУ, 2016. - 228 с. - ISBN 978-5-288-05654-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1244754> (дата обращения: 29.01.2024). - Режим доступа: по подписке.



*Приложение 3*  
*к рабочей программе дисциплины (модуля)*  
*Б1.В.14 Компьютерная лингвистика и семиотика*

**Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем**

Направление подготовки: 09.03.04 - Программная инженерия

Профиль подготовки: Разработчик искусственного интеллекта и когнитивных систем

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2024

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows