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

УТВЕРЖДАЮ
Проректор по образовательной деятельности КФУ
Проф. Д.А. Таюрский
_____ «01» июня 2021 г.

подписано электронно-цифровой подписью

Программа дисциплины

Компьютерная лингвистика

Направление подготовки: 09.03.02 - Информационные системы и технологии
Профиль подготовки: Информационные системы и технологии
Квалификация выпускника: бакалавр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2022

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): доцент, к.н. Гатиатуллин А.Р. (Кафедра информационных систем, отделение фундаментальной информатики и информационных технологий), 1agatiat@int.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:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-1	Проверка работоспособности и рефракторинг кода программного обеспечения, интеграция программных модулей и компонент и верификация выпусков программного обеспечения
ПК-3	Проверка и отладка программного кода, тестирование информационных ресурсов с точки зрения логической целостности (корректность ссылок, работа элементов форм

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

Методы компьютерной лингвистики применительно к обработке текстов на естественном языке.

Должен уметь:

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

Должен владеть:

Методами анализа текстов на естественном языке.

Должен демонстрировать способность и готовность:

Применять полученные знания в профессиональной деятельности.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.02 Дисциплины (модули)" основной профессиональной образовательной программы 09.03.02 "Информационные системы и технологии (Информационные системы и технологии)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 3 курсе в 5 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 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 68 часа(ов), в том числе лекции - 34 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 34 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 58 часа(ов).

Контроль (зачёт / экзамен) - 54 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 5 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	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)						Самостоятельная работа
			Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
1	Тема 1. Введение. Предмет и задачи. История								

становления компьютерной лингвистики.

ЭЛЕКТРОННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА КФУ

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)						Самостоятельная работа
			Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
2.	Тема 2. Основные направления и предмет компьютерной лингвистики. Когнитивный инструментарий компьютерной лингвистики.	5	6	0	0	0	6	0	9
3.	Тема 3. Естественные и искусственные языки.	5	6	0	0	0	6	0	10
4.	Тема 4. Современный машинный перевод.	5	6	0	0	0	6	0	10
5.	Тема 5. Компьютерная лексикография и корпусная лингвистика.	5	6	0	0	0	6	0	10
6.	Тема 6. Информационные технологии в научной деятельности (этапы конструирования логики научного исследования). Компьютерная лингводидактика.	5	6	0	0	0	6	0	10
	Итого		34	0	0	0	34	0	58

4.2 Содержание дисциплины (модуля)

Тема 1. Введение. Предмет и задачи. История становления компьютерной лингвистики.

Цели и проблемы компьютерной лингвистики. Основные задачи и методологии их решения. Компьютерная лингвистика как междисциплинарное научное направление. Предмет компьютерной лингвистики. История становления и основные тенденции развития. Джоржтаунский эксперимент как начало создания компьютерной лингвистики.

Тема 2. Основные направления и предмет компьютерной лингвистики. Когнитивный инструментарий компьютерной лингвистики.

Когнитивный инструментарий компьютерной лингвистики. Направления КЛ: автоматизированный анализ, распознание и автоматическая обработка текста; искусственный интеллект и машинный перевод; компьютерная лексикография (электронные словари); корпусная лингвистика и создание электронной базы данных; компьютерная лингводидактика.

Тема 3. Естественные и искусственные языки.

Компьютерное обеспечение представления знаний. Естественные и искусственные языки. Автоматизированный анализ: распознавание и синтез устной и письменной речи. Проблемы и перспективы, морфологический анализ, проблемы семантического анализа, синтаксический анализ, реферирование и аннотирование текста. Лингвистические базы данных: модели и типы данных. Создания общих искусственных языков для представления информации; усовершенствования языков доступа к информации.

Тема 4. Современный машинный перевод.

Предпосылки возникновения и развитие машинного перевода (МП). Последовательность формальных операций, обеспечивающих анализ и синтез в системе машинного перевода.

Стратегии машинного перевода. Автоматизированные и полуавтоматизированные системы перевода. Области использования машинного перевода. Важнейшие системы машинного перевода.

Тема 5. Компьютерная лексикография и корпусная лингвистика.

Типология традиционных и электронных словарей. Словарные процессоры. Типология словарей. Тезаурусы и терминологические словари. Компьютерные технологии составления и эксплуатации словарей. Специальные программы - базы данных, компьютерные картотеки, программы обработки текста. Корпусная лингвистика, ориентированная на создание национальных корпусов естественных языков. Типы лингвистической разметки.

Тема 6. Информационные технологии в научной деятельности (этапы конструирования логики научного исследования). Компьютерная лингводидактика.

Применение компьютера при частотных статистических анализах текстов на материалах разных языков. Накопление знаний и фактов, поиск литературы, опытно-экспериментальная работа при фиксации лингвистических данных. Дидактическая концепция обучения на основе компьютерных технологий. Принципы создания электронных учебных средств. Классификация электронных средств учебного назначения. Электронные учебники, словари, тренажеры, практикумы, учебно-методические материалы. Мультимедиа в помощь филологу.

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 №14-55-996ин/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

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 планируемым результатам 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)

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

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

Вид работ	Методические рекомендации
лекции	Лекция является основной формой обучения в высшем учебном заведении. В ходе лекционного курса дается целостное представление о курсе. Записи лекций в конспектах должны быть избирательными, полностью следует записывать только определения. В конспекте рекомендуется применять сокращение слов, что ускоряет запись. Вопросы, возникающие в ходе лекции, рекомендуется записывать на полях и после окончания лекции обратиться за разъяснением к преподавателю. Необходимо активно работать с конспектом лекции: после окончания лекции рекомендуется перечитать свои записи, внести поправки и дополнения на полях. Конспекты лекций следует использовать при подготовке к семинарам, при подготовке к экзамену, при выполнении самостоятельных заданий и домашних работ.
лабораторные работы	Лабораторные работы составлены в соответствии с программой дисциплины и предназначены для закрепления теоретического материала, полученного на лекциях и практических занятиях, и приобретения студентами способности самостоятельно решать стандартные задачи профессиональной деятельности с применением информационно-коммуникационных технологий. При подготовке к лабораторным работам студент должен самостоятельно повторить теоретический материал.
самостоя-тельная работа	Самостоятельная работа проводится с целью углубления знаний по дисциплине и предусматривает: чтение студентами рекомендованной литературы и усвоение теоретического материала дисциплины; подготовку к практическим занятиям, работу с Интернет-источниками; подготовку к сдаче экзамена. Планирование времени на самостоятельную работу, необходимого на изучение настоящей дисциплины, студентам лучше всего осуществлять на весь семестр, предусматривая при этом регулярное повторение пройденного материала. Материал, законспектированный на лекциях, необходимо регулярно дополнять сведениями из литературных источников, представленных в рабочей программе дисциплины. По каждой из тем для самостоятельного изучения, приведенных в рабочей программе дисциплины следует сначала прочитать рекомендованную литературу и при необходимости составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в этой теме и для освоения последующих разделов курса. Для расширения знаний по дисциплине рекомендуется использовать Интернет-ресурсы: проводить поиск в различных системах, сайтов и обучающих программ, рекомендованных преподавателем на лекционных занятиях. Для лучшего усвоения учебного материала и подготовки к семинарским занятиям предполагается активная внеаудиторная самостоятельная работа студентов с учебной литературой, с нормативными, методическими и справочными материалами.
экзамен	В период подготовки к экзамену обучающиеся вновь обращаются к пройденному учебному материалу. При этом они не только закрепляют полученные знания, но и получают новые. Подготовка обучающегося к экзамену включает в себя три этапа: самостоятельная работа в течение процесса обучения; непосредственная подготовка в дни, предшествующие экзамену по темам курса; подготовка к ответу на вопросы, содержащиеся в билетах. Литература для подготовки к экзамену рекомендуется преподавателем. Экзамен проводится по билетам, охватывающим весь пройденный по данной теме материал. По окончании ответа преподаватель может задать обучающемуся дополнительные и уточняющие вопросы.

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.

Мультимедийная аудитория.

Компьютерный класс.

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 09.03.02 "Информационные системы и технологии" и профилю подготовки "Информационные системы и технологии".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.02 Компьютерная лингвистика

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.03.02 - Информационные системы и технологии
Профиль подготовки: Информационные системы и технологии
Квалификация выпускника: бакалавр
Форма обучения: очное
Язык обучения: русский
Год начала обучения по образовательной программе: 2022

Основная литература:

1. Информатика и лингвистика: Учебное пособие / ВолосатоваТ.М., ЧичваринН.В. - М.: НИЦ ИНФРА-М, 2016. - 196 с. - (Высшее образование: Бакалавриат) ISBN 978-5-16-010977-0 - Режим доступа: http://znanium.com/catalog/product/508097. - Дата обращения: 15.06.2022.

2. Компьютерный практикум по курсу 'Информатика' : учеб. пособие / В.Т. Безручко. - 3-е изд., перераб. и доп. - М. : ИД 'ФОРУМ' : ИНФРА-М, 2017. - 368 с. (Высшее образование). - Режим доступа: http://znanium.com/catalog/product/756204. - Дата обращения: 15.06.2022.

Дополнительная литература:

1. Компьютерная математика: Учебное пособие/К.В.Титов - М.: ИЦ РИОР, НИЦ ИНФРА-М, 2016. - 261 с. - (Высшее образование) ISBN 978-5-369-01470-7 - Режим доступа: http://znanium.com/catalog/product/523231. - Дата обращения: 15.06.2022.

2. Языкознание: От Аристотеля до компьютерной лингвистики: Научно-популярное / Алпатов В. - М.:Альпина нон-фикшн, 2018. - 253 с. - (Научно-популярная литература) ISBN 978-5-91671-804-1 - Режим доступа: http://znanium.com/catalog/product/1003471. - Дата обращения: 15.06.2022.

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.02 Компьютерная лингвистика

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

Направление подготовки: 09.03.02 - Информационные системы и технологии

Профиль подготовки: Информационные системы и технологии

Квалификация выпускника: бакалавр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2022

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

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань" , доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "Консультант студента", доступ к которой предоставлен обучающимся. Многопрофильный образовательный ресурс "Консультант студента" является электронной библиотечной системой (ЭБС), предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Полностью соответствует требованиям федеральных государственных образовательных стандартов высшего образования к комплектованию библиотек, в том числе электронных, в части формирования фондов основной и дополнительной литературы.